PREDICTION OF ICEBERG-SEABED INTERACTION USING MACHINE LEARNING ALGORITHMS

by

© Hamed Azimisiahchaghaei

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Civil Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

October 2023

St. John's Newfoundland Canada

ABSTRACT

Every year thousands of icebergs are born out of glaciers in the Arctic zone and carried away by the currents and winds into the North Atlantic. These icebergs may touch the sea bottom in shallow waters and scratch the seabed, an incident called “ice-gouging”. Ice-gouging may endanger the integrity of the buried subsea pipelines and power cables because of subgouge soil displacement. In other words, the shear resistance of the soil causes the subgouge soil displacement to extend much deeper than the ice keel tip. This, in turn, may cause the displacement of the pipelines and cables buried deeper than the most possible gouge depth. Determining the best burial depth of the pipeline is a key design aspect and needs advanced continuum numerical modeling and costly centrifuge tests. Empirical equations suggested by design codes may be also used but they usually result in an over-conservative design.

Iceberg management, i.e., iceberg towing and re-routing, is currently the most reliable approach to protect the subsea and offshore structures, where the approaching icebergs are hooked and towed in a safe direction. Iceberg management is costly and involves a range of marine fleets and advanced subsea survey tools to determine the iceberg draft, etc.

The industry is constantly looking for cost-effective and quick alternatives to predict the iceberg draft and subgouge soil displacements. In this study, powerful machine learning (ML) algorithms were used as an alternative cost-effective approach to first screen the threatening icebergs by determining their drafts and then to predict the subgouge soil displacement to be fed into the structural integrity analysis.

Developing a reliable solution to predict the iceberg draft and subgouge soil displacement requires a profound understanding of the problem's dominant parameters. Therefore, the present study started with dimensional analyses to identify the dimensionless groups of key parameters

governing the physics of the problem. Two comprehensive datasets were constructed using the monitored characteristics of the real icebergs for draft prediction and experimental studies for the subgouge soil displacements reported in the literature. Using the constructed database, 14 ML algorithms ranging from neural network-based (NN-based) to three-based methods were sequentially used to predict the iceberg draft and the subgouge soil displacement. The studies were conducted both in clay and sand seabed. By different combinations of the input parameters, several ML models were developed and assessed by performing sensitivity analysis, error analysis, discrepancy analysis, uncertainty analysis, and partial derivative sensitivity analysis to identify the superior ML models along with the most influential input parameters. The best ML model was able to predict the iceberg drafts alongside the subgouge soil features with the highest level of precision, correlation, and lowest degree of complexity. A set of ML-based explicit equations were also derived from the wide range of field and experimental measurements for the estimation of iceberg drafts, subgouge soil deformations, and ice keel reaction forces, which outperformed the existing empirical equations. The study resulted in developing a set of tools that can be used for both a cost-effective screening of the threatening icebergs and the prediction of the corresponding subgouge soil displacements. The outcome of the study can effectively contribute to a significant reduction of iceberg management costs and greenhouse gas (GHG) emissions through the mitigation of the marine spread operation.

Keywords: Iceberg draft, Iceberg-seabed interaction process, Dimensionless analysis, Machine learning, Simulation, Subsea assets

STATEMENT OF AUTHORSHIP

I have been the first author and the main contributor to the conducted work, including mathematical works, modeling, data analysis, post-processing, and preparation of the paper draft. Dr. Hodjat Shiri was my supervisor providing the research topic, overall approach, checking the results and providing advice, reviewing and editing the paper drafts, etc. Dr. Masoud Mahdianpari from C-CORE was my co-supervisor, who joined the project midway, and provided advice on the selection and implementation of different ML algorithms, reviewing and editing the paper drafts, etc. Also, I had a chance to collaborate on a few papers with Dr. Sohrab Zendehboudi, a member of my supervisory committee, and Dr. Eduardo Malta, a post-doctoral fellow in Dr. Shiri's lab, who provided advice on the conducted research. The outcome of this research includes 20 journal and conference papers as follows:

No.	Title	Type of paper	Status
P1	Azimi, H. and Shiri, H., 2020. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98, 102120.	Journal	Published
P2	Azimi, H. and Shiri, H., 2021. Modeling subgouge sand deformations by using multi-layer perceptron neural network. In The 31st International Ocean and Polar Engineering Conference. Rhodes, Greece, ISOPE-I-21-2150.	Conference	Published
P3	Azimi, H. and Shiri, H., 2020. Dimensionless groups of parameters governing the ice-seabed interaction process.	Journal	Published

	Journal of Offshore Mechanics and Arctic Engineering, 142(5), 051601.		
P4	Azimi, H., Shiri, H. and Zendehboudi, S., 2023. Prediction of Ice-Induced Subgouge Soil Deformation in Sand Using Group Method of Data Handling–Based Neural Network. Journal of Cold Regions Engineering (ASCE), 37(2), 04023006.	Journal	Published
P5	Azimi, H., Shiri, H. and Zendehboudi, S., 2022. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology, 193, 103426.	Journal	Published
P6	Azimi, H. and Shiri, H., 2021. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3), 2307-2335.	Journal	Published
P7	Azimi, H. and Shiri, H., 2021. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04), 411-420.	Journal	Published
P8	Azimi, H. and Shiri, H., 2021. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239, 109795.	Journal	Published

P9	Azimi, H., Shiri, H. and Malta, E.R., 2021. A non-tuned machine learning method to simulate ice-seabed interaction process in clay. Journal of Pipeline Science and Engineering, 1(4), 379-394.	Journal	Published
P10	Azimi, H., Shiri, H. and Mahdianpari, M., 2022, April. Simulation of Subgouge Sand Deformations Using Robust Machine Learning Algorithms. Offshore Technology Conference (OTC), Houston, Texas, USA, OTC-31937-MS.	Conference	Published
P11	Azimi, H., Shiri, H. and Mahdianpari, M., 2022. Iceberg-seabed interaction analysis in sand by a random forest algorithm. Polar Science, 34, 100902.	Journal	Published
P12	Azimi, H., Shiri, H. and Mahdianpari, M., 2022. Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. Journal of Pipeline Science and Engineering, 2(4), 100075.	Journal	Published
P13	Azimi, H., Mahdianpari, M. and Shiri, H., 2023. Determination of Parameters Affecting the Estimation of Iceberg Draft. China Ocean Engineering, 37(1), 62-72.	Journal	Published
P14	Azimi, H., Shiri, H. and Mahdianpari, M., 2023. Sensitivity Analysis of Parameters Governing the Iceberg Draft through Neural Network-Based Models. Journal of Ocean Engineering and Marine Energy, 9.	Journal	Published

P15	Azimi, H., Shiri, H. and Mahdianpari, M., 2023. Generalized Structure of the Group Method of Data Handling for Modeling Iceberg Drafts. Submitted as journal manuscript.	Journal	Under review
P16	Azimi, H., Shiri, H. and Mahdianpari, M., 2023. Iceberg Draft Prediction using Gradient Boosting Regression Algorithm. Submitted as journal manuscript.	Journal	Under review
P17	Azimi, H., Shiri, H. and Mahdianpari, M., 2023. Iceberg Draft Evaluation through Robust Tree-Based Machine Learning Algorithms. Submitted as journal manuscript.	Journal	Revise requested
P18	Azimi, H., Shiri, H. and Mahdianpari, M., 2023. Iceberg Drafts Assessment using Decision Tree Regression (DTR), Artificial Neural Network (ANN), and Support Vector Regression (SVR) algorithms. In 27th International Conference on Port and Ocean Engineering under Arctic Conditions (POAC), Glasgow, UK.	Conference	Published
P19	Azimi, H., Shiri, H. and Mahdianpari, M., 2023. Evaluation of Iceberg Draft and Iceberg-Seabed Interaction using Random Forest Regression Algorithm. In 33rd International Ocean and Polar Engineering Conference (ISOPE), 2023, Ottawa, Canada.	Conference	Published
P20	Azimi, H., Shiri, H. and Mahdianpari, M., 2023. Simulation of the Iceberg draft and subgouge soil characteristics using Extra Tree Regression Algorithm. Submitted as journal manuscript.	Journal	Under review

ACKNOWLEDGMENT

I would love to extensively appreciate the endless support, meticulous supervision, infinite patience, and invaluable advice of my supervisor, Dr. Hodjat Shiri during the completion of the current research. Special gratitude is extended to Dr. Masoud Mahdianpari, my co-supervisor, who always guided me to boost my research objectives toward a new set of challenges. I would like also to sincerely acknowledge the time and guidance provided by Dr. Sohrab Zendehboudi as my supervisory committee.

Moreover, I gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)," and the "Newfoundland Research and Development Corporation (RDC) (now DIET)" through "Collaborative Research and Developments Grants (CRD)." Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

Last but not least, I genuinely appreciate my mother and father who raised me with love and helped me in all my pursuits. I am grateful to my beloved, Dr. Mina Sanahmadi for her loving support.

List of Figures

List of Tables

List of Abbreviations and Symbols

Abbreviations

AF	Activation function
AI	Artificial intelligence
AIC	Akaike information criteria
ALE	Arbitrary Lagrangian Eulerian
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
AUV	Autonomous underwater vehicle
BP	Backpropagation
CART	Classification and Regression Trees
CEL	Coupled Eulerian-Lagrangian
CI	Confidence interval
CRM	Coefficient of residual mass
DE	Differential evolution
DT	Decision tree
DTR	Decision tree regression
DR	Discrepancy ratio
ELM	Extreme learning machine
ETR	Extra tree regression
ETs	Expression trees
FE	Finite element

FFNN	Feedforward neural network
GBR	Gradient boosting regression
GEP	Gene expression programming
GMDH	Group method of data handling
GP	Genetic programming
GS-GMDH	Generalized structure group method of data handling
Hardlim	Hard limit activation function
Hypertan	Hyperbolic tangent
KNR	K-nearest neighbors regression
LM	Levenberg-Marquardt
LR	Linear regression
LS	Least square
MAE	Mean absolute error
Mean	Mean calculated errors
ML	Machine learning
MLP	Multilayer perceptron
MLPNN	Multi-layer perceptron neural network
MSE	Mean square error
NHN	Number of hidden neurons
NN	Neural network
NSC	Nash-Sutcliff efficiency coefficient
PDSA	Partial derivative sensitivity analysis
PEI	Prediction error interval

R	Correlation coefficient
Radbas	Radial basis activation function
RBF	Radial basis functions
RD	Relative density
RFR	Random forest regression
RMSE	Root mean squared error
ROV	Remote operating vehicle
SA	Sensitivity analysis
SAELM	Self-adaptive extreme learning machine
SaE-ELM	Self-adaptive evolutionary extreme learning machine
SC	Soft computing
Sig	Sigmoidal activation function
SLFFNN	Single-layer feed-forward neural network
SI	Scatter index
Sin	Sinusoidal activation function
StDev	Standard deviation of errors
SVR	Support vector regression
Tribas	Triangular basis activation function
UA	Uncertainty analysis
VAF	Variance accounted for
VC	Vapnik-Chervonenkis
WI	Willmott index
WUB	Width of uncertainty bound

English Symbols

a and b	Learning parameters
BHI	Matrix of bias of hidden layer
C	Penalty Parameter
CC	Coefficient of crossover
C (i, j)	The ith chromosome value
c	Cohesion of clay
D	Iceberg draft
Ds	Gouge depth
d_h	Horizontal subgouge soil deformation
d_v	Vertical subgouge soil deformation
dh0	maximum horizontal subgouge soil deformation
dhy	Horizontal subgouge deformation
e_j	Estimated error
Fa	Friction force on the bottom of the ridge
Fb	Buoyancy force
Fc	Coulomb's passive friction force
Fda	Drag force from wind
Fdw	Drag force from the current
Fi	Driving force from the surrounding floe
Fw	Ridge weight
f	Latent function

fi	Functional symbol
$G(\,)$	Activation function
g	Gravitational acceleration
H	Iceberg height
H(k,ϖ)	Extreme learning machine feature mapping matrix
H^{+}	Moore-Penrose generalized inverse matrix
h’	Height of the berm
I	Individuals of the DE algorithm
InW	Matrix of input weight
InV	Matrix of input parameters
I_r	Dilation index
ith	Neurons in the input layer
jth	Neurons in the hidden layer or Fitness case
j_{rand}	Parameter between 1 and D
k	Number of coefficients
L	Iceberg length
l	Size of the hidden layer
Lh	Resultant of horizontal forces
Lv	Resultant of vertical forces
M	Iceberg mass
m	Number of GMDH algorithm’s input
m'	Number of the principal quantities
N	Seabed reaction

n	Number of observed values
$nf_{l,\varpi}$ and $n\varrho_{l,\varpi}$	Trial vectors
$\bar{O}$	Average of observed values
O_i	Observed value or Output of machine learning algorithm
$OutW$	Matrix of output weights
Pa	Atmospheric pressure
P_i	Predicted value
$\bar{P}$	Average predicted values
$P_{l,\varpi}$	Probability procedure
P1	Number of generations
P2	Number of chromosomes
P3	Number of genes
P4	Head size
P5	Linking function
P6	Mutation rate
P7	Inversion rate
P8	One-point recombination rate
P9	Two-point recombination rate
P10	Gene recombination rate
P11	Gene transportation rate
P15	Function set
P1	Number of generations
P2	Number of chromosomes

P3	Number of genes
P4	Head size
P5	Linking function
P6	Mutation rate
P7	Inversion rate
P8	One-point recombination rate
rand_j	Random variable
S_e	Standard deviation of the estimated error
Sf	Iceberg shape factor
T	Transpose the matrix
Tj	Observed value for the jth fitness case
V	Velocity of the plow or velocity of iceberg keel
V_e	Maximum vertical extent of subgouge deformation
W	Gouge width
w	Iceberg width
$w_{i,j}$	Input weight
x	Position of iceberg along the scour axis
Y	Output of the ELM model
y	Soil depth

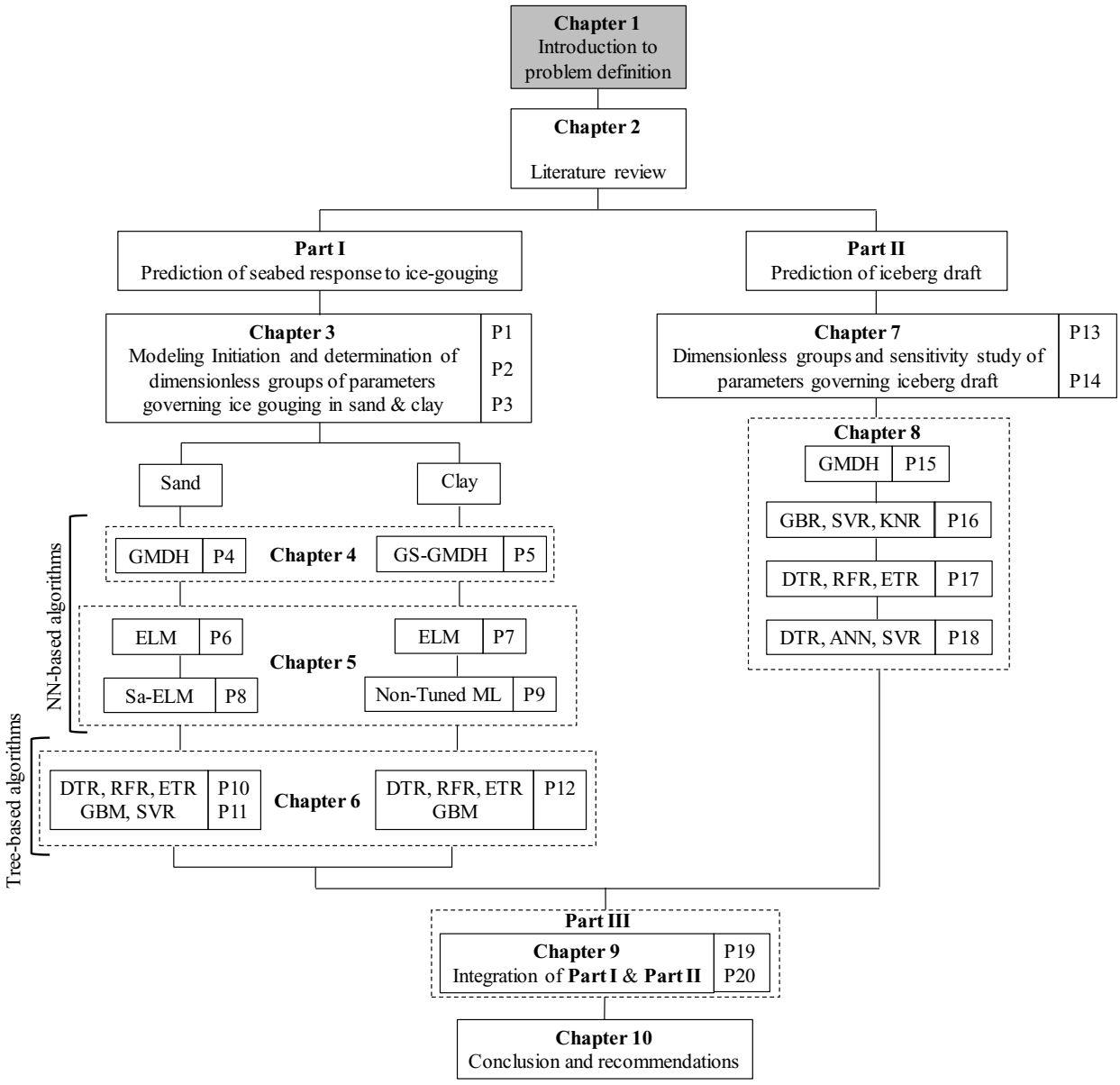
Greek Symbols

α	Keel attack angle
β_j	Coefficient to link the jth neuron in the hidden layer

Ɗ	Dimensions of problem
Γ_i	Unknown coefficients stemming from regression approaches
γ_s	Soil specific weight
Δ	Matrix of unknown coefficients
$\Delta_{i,\varpi}$	Trial vector
δ1, δ2, δ3, …, δn	Primarily parameters in the dimensionless analysis
ζ	Number of optimized parameters during the training procedure
η	Ice-seabed interaction parameters
ϑ	Computational vector of DE algorithm
ɭ	Learning rate
κ	Parameter to lessen the overfitting in the ETR algorithm
κ_j	Biases of the hidden layer neurons
Ɱ	Selection range
Ŋ	A set of a_j, b_j
ɲ	Parameter to lessen the overfitting in the ETR algorithm
Λ	Scaling factor
μsw	Seawater viscosity
Π1, Π2,…, Πn	Dimensionless groups
π	Positive constant value to hinder the zero improvement rate
π1, π2, π3, …, πn-m	Dimensional groups regarding primary parameters
ρi	Density of an iceberg
ρ_s	Seabed Soil density
ρsw	Density of seawater

σ_b	bearing pressure
σ_k^i	Randomly calculated from an individual
φ	Internal friction angle of sand
$\chi_1, \chi_2, \chi_3, \dots, \chi_k$	Input parameters of GMDH algorithm
χ_i	Inputs of the ELM model
$\Psi_{i,\varpi}$	mutant vectors
ω	Angle of the surcharged soil slope
ϖ	Generation number of DE algorithm
Ϙ	Output of the problem
Ϙ	Transformation Function

1. Chapter 1: Introduction



1.1. Background

Climate change is causing the polar ice caps and glaciers in the Arctic and subarctic waters to melt faster, leading to an increase in the number of icebergs that break off each year. These traveling icebergs are a potential threat and one of the governing design factors of subsea pipelines. If the iceberg draft is greater than the ocean's depth, the iceberg tip may scour the seabed which is called “Ice-gouing”. The ice-gouging may impair the serviceability and even the structural integrity of the buried pipelines by creating large subgouge soil displacements and consequently large pipeline deformations. Figure 1-1 schematically shows an iceberg scouring the seabed.

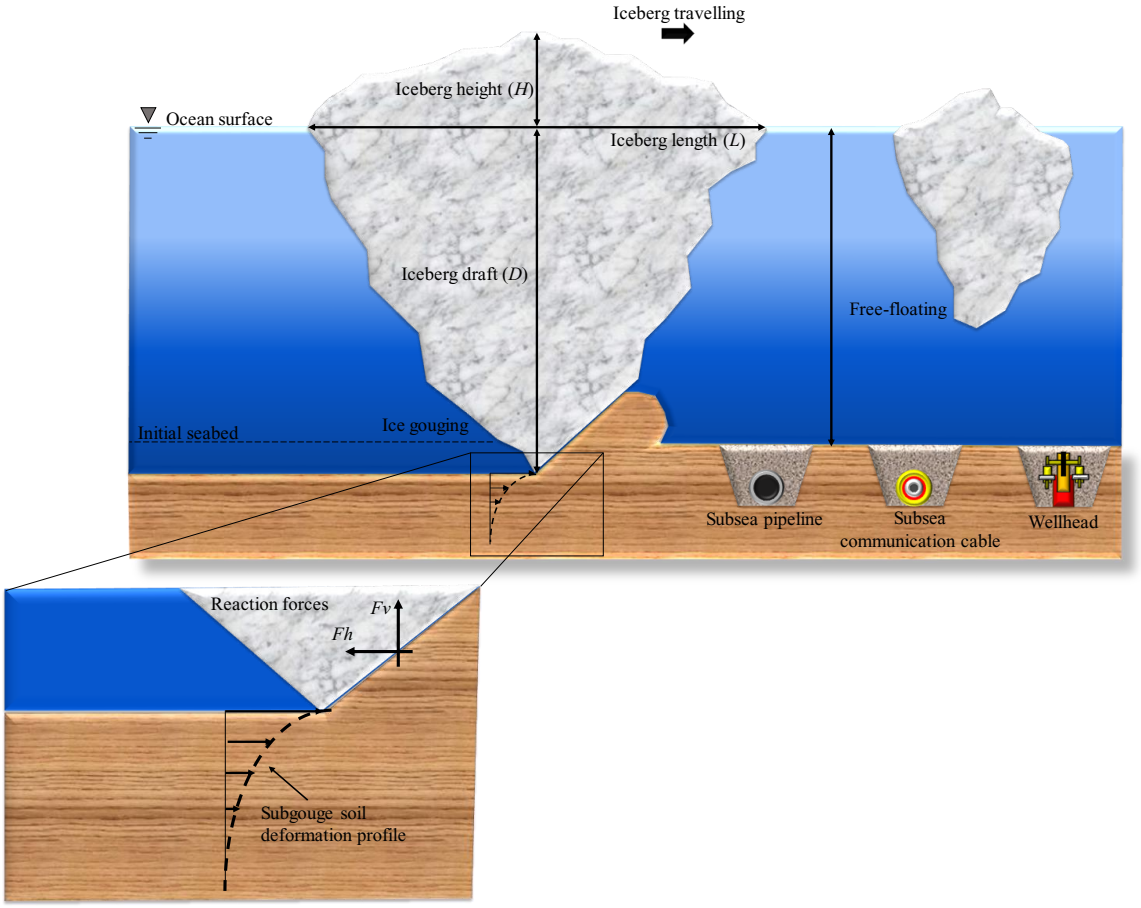


Figure 1-1. Iceberg in the scouring and free-floating conditions

1.2. Motivation

Iceberg management operations with significant downtime and expenses, such as iceberg towing and re-routing, are currently performed to protect the subsea and offshore infrastructures. The side scan sonar mounted on autonomous underwater vehicles (AUVs), and remote operating vehicles (ROVs) are presently used to support the iceberg management operation by measuring the iceberg's geometrical and physical properties. However, these operations are costly and time-consuming because of needing vessels, experienced crew, expensive logistics, and advanced pieces of equipment.

Despite the offshore structures such as ships, wind turbines, and floating or fixed platforms that can be directly attacked by icebergs, the threat that goes to the buried pipelines and power cables is much more complicated involving the ice-soil-structure interactions. During the ice-gouging, the subgouge soil displacement is largely extended down the seabed much deeper than the gouge depth. This may cause large deformations in the pipelines buried even below the gouge depth. Investigation of the subgouge soil deformation and the best trench depth to bury the pipelines and power cables have been a hot topic in the literature over the past few decades.

On the other hand, long-running numerical finite element (FE) simulations and costly centrifuge testing programs are conducted to explore the subgouge soil displacement that, in turn, governs the structural response of the buried subsea pipelines and power cables.

The industry is still demanding more cost-effective, reliable, and environmental-friendly solutions that can assess the iceberg draft and evaluate the iceberg-seabed interaction characteristics using the easily measurable exposed parameters, particularly in the initial steps of iceberg management projects and subsea structure designs.

The current study has tried to respond to this demand by using powerful Machin Learning (ML) algorithms both for the iceberg screening through the prediction of the iceberg drafts using the above-water visible parameters and also the prediction of the resultant subgouge soil displacements. In computer science, machine learning (ML) is a subfield of artificial intelligence (AI) that is largely described as the ability of a machine to imitate intelligent human manners. ML algorithms are applied to accomplish complicated tasks in a way that is identical to how humans solve problems. ML is utilized in different domains such as engineering, education, medicine, etc. Computer algorithms perform the entire data processing, interpretation, and analysis; hence, no human intervention exists for the modeling and simulation of data. An ML algorithm is able to learn from data and identify patterns and trends in problems. ML technology is a precise, reliable, cost-effective tool for solving various linear and nonlinear issues.

The key objectives of the current research were as below which were addressed throughout the forthcoming chapters:

- Determination of the parameters governing the iceberg draft and iceberg-seabed interaction process both in sand and clay.
- Simulation of the iceberg draft and iceberg-seabed interaction outcome using ML algorithms as a cost-effective, quick, and precise alternative.
- Simulation of maximum horizontal subgouge deformations using ML algorithms as a design factor of submarine structures against ice-gouging attacks.
- Identifying the superior ML algorithms in the iceberg draft estimation and prediction of the subgouge soil displacements.
- Combining the developed ML-based models to screen the threatening icebergs and predict the resultant subgouge soil displacements by those causing ice.

The stepwise flowchart of the objectives is depicted in Figure 1-2.

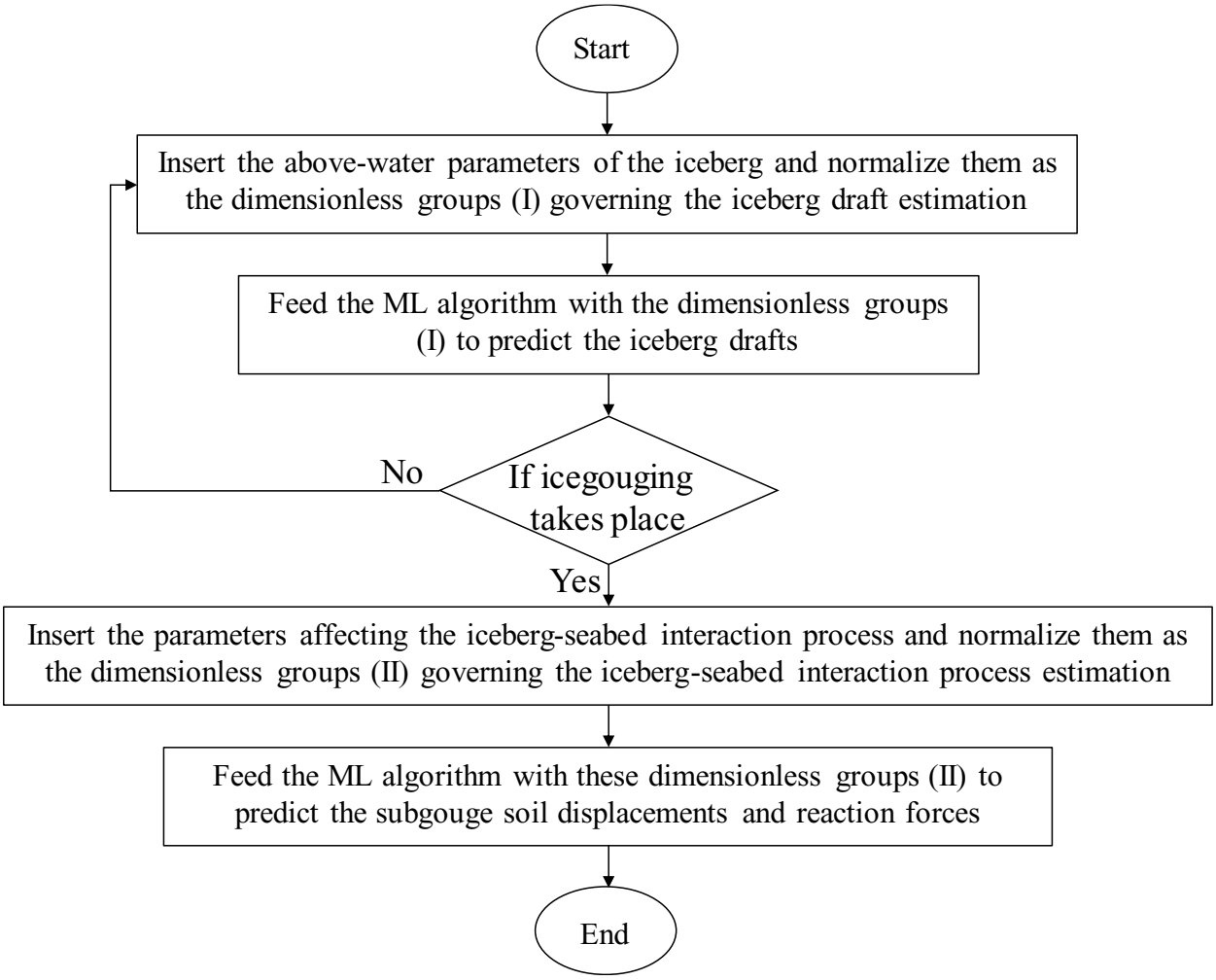


Figure 1-2. Flowchart of the current methodology

First, the above-water parameters of the potentially threatening icebergs are normalized as the dimensionless groups governing the iceberg draft prediction using the π-Buckingham theorem. These parameters may include iceberg mass, length, width, height, and shape factors. Subsequently, the ML algorithm is fed with these dimensionless groups of parameters to predict the iceberg draft. The predicted draft is compared with the water depth to determine whether an

ice-gouging happens. For the icebergs which scour the seabed, the parameters affecting the iceberg-seabed interaction process are normalized as the dimensionless groups of parameters governing the iceberg-seabed interaction process. Lastly, the horizontal and vertical subgouge soil displacement (d_h & d_v) along with the horizontal and vertical ice keel reaction forces (F_h & F_v) are predicted using the ML algorithms. This information can be used by design engineers to feed into the decoupled pipeline beam-spring models that are used in industry to obtain the structural pipe response to ice-induced soil displacements.

1.3. Organization of the Thesis

This is a paper-based dissertation that comprises three main parts (i.e., Part I, Part II, and Part III) accommodating 10 chapters that present 16 journal and 4 conference papers. Part I which accommodates Chapters 3 to 6 represents several ML algorithms and associated studies to predict the seabed response to the ice-gouging in sand and clay. Part II comprises Chapters 7 and 8 which are allocated to investigations performed on ML-based prediction of the iceberg drafts. Ultimately, the outcome of Part I and Part II are integrated into Part III as Chapter 9 to make a complete loop of screening the icebergs, identifying the threatening ice-gouges, and predicting the seabed response to it. Figure 1-3 depicts the sequence of study and their logical connection with chapters as well as the objectives. This flowchart with highlighted boxes has been used on the first page of each chapter to facilitate tracing the flow of the content topics.

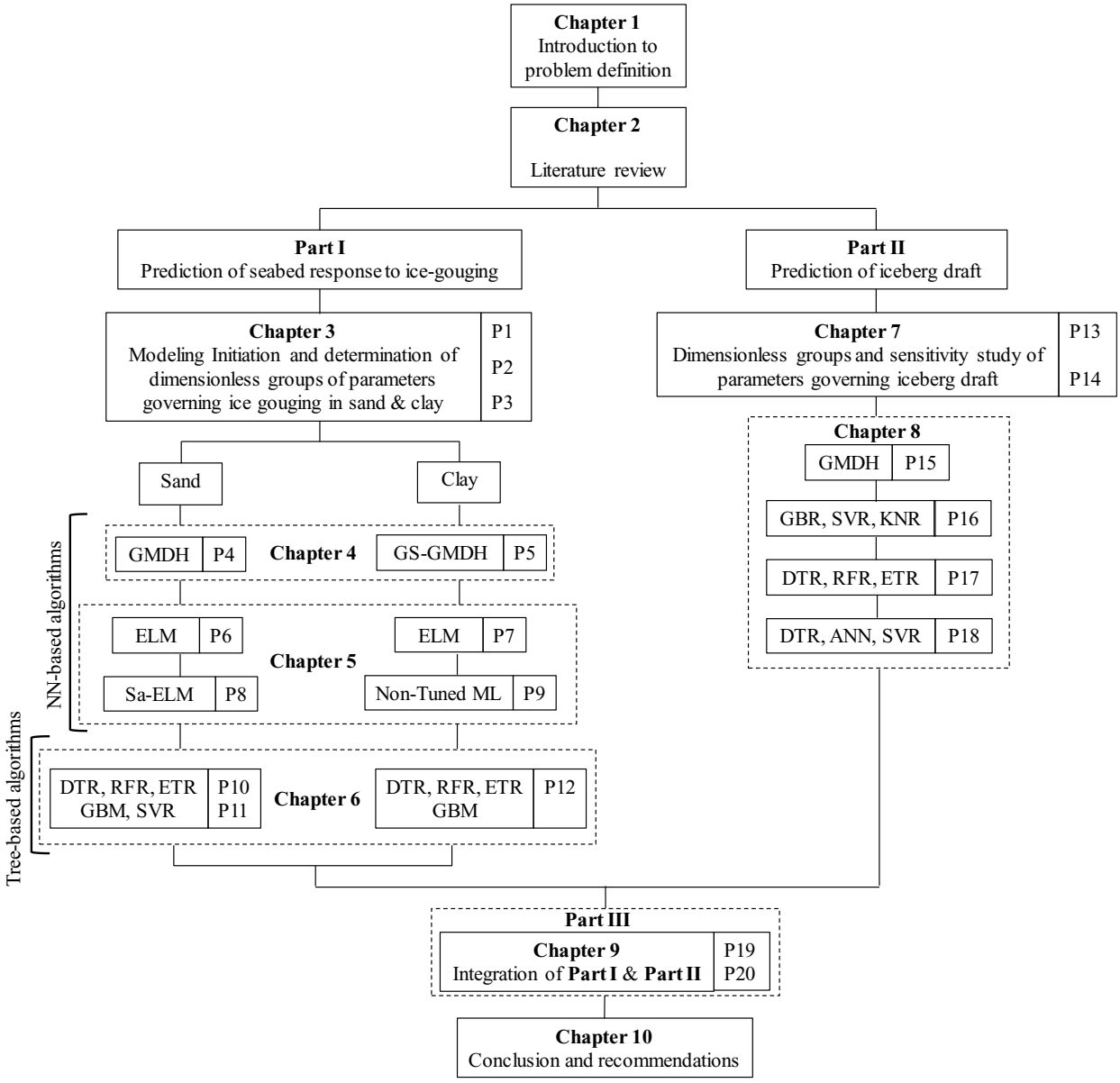


Figure 1-3. The sequence of study and their logical connection with chapters as well as the objectives of chapters

Chapter 1 is an introduction to the problem that covers the background and motivation of the study along with the key objectives and outcomes. All of the produced papers have their independent literature review. However, a short review of the literature published in these realms has been provided in Chapter 2 to facilitate reading the background research. Chapter 3 includes three

published papers, which is an initiation of the modeling subgouge soil displacement in this research. The maximum horizontal subgouge soil deformations in the sand seabed were modeled using gene expression programming (GEP) and artificial neural network (ANN). Even though the performance of GEP was promising, solely the effect of gouge depth, the maximum vertical extent of subgouge deformation, the attack angle, the dilation index, and soil depth were used (Azimi and Shiri 2020a). Subsequently, the maximum horizontal subgouge soil displacements in the sandy seabed were also estimated by using the simplest ML algorithm, e.g., artificial neural network (ANN). The ANN algorithm could not give an explicit model and similar to the prior work, the effect of all governing parameters is not taken into account to compute the objective function (Azimi and Shiri 2021a). To overcome these limitations, the dimensionless groups of parameters governing the subgouge soil displacement in the sandy and clay seabed were identified in section 3 of Chapter 3. The linear regression (LR), as the basic regression analysis, was applied to approximate the maximum subgouge soil displacements in both sandy and clay seabed (Azimi and Shiri 2020b).

Chapter 4 includes two published journal articles in which, using the dimensionless groups of parameters governing the subgouge soil displacement introduced in the prior step, the group method of data handling (GMDH) was employed to estimate the horizontal and vertical subgouge soil deformations in the sand. This algorithm provided a set of explicit equations which outperforms the empirical and artificial neural network (ANN) models (Azimi et al. 2023a). The GMDH algorithm showed some limitations such as using a quadratic polynomial, limiting the inputs of each neuron to two, and limiting the inputs of each neuron to the use of the adjacent layer neurons. To overcome these challenges, the generalized structure group method of data handling (GS-GMDH) was applied in the prediction of horizontal and vertical subgouge soil displacements

in the clay seabed. The GS-GMDH-based explicit equation demonstrated an acceptable accuracy and a better performance in comparison with the previous empirical models (Azimi et al. 2022a). However, the need for a robust ML algorithm with lower computation time and the ability in providing a computational matrix to calculate the objective function encouraged the use of the extreme learning machine (ELM) algorithm in Chapter 5. Chapter 5 comprises four published journal articles. It is worth mentioning that a computational matrix is a useful tool since it is simple to apply and can be implemented by all programming languages. Thus, the ice keel reaction forces and subgouge soil displacement in the sandy and clay seabed were modeled through the ELM algorithms in this chapter (Azimi and Shiri 2021b; 2021c).

Though ELM showed an acceptable outcome, the manual selection of optimized hidden layer neurons plays a crucial role in the simulation procedure. To overcome this drawback, the self-adaptive evolutionary extreme learning machine (SaE-ELM) as a hybrid model of the ELM network and differential evolution (DE) algorithm was employed to model the subgouge soil characteristics in the sandy and clay seabed in this chapter. In the SaE-ELM algorithm, the number of neurons in the hidden layer is optimized by using the DE algorithm. During the training procedure, the control coefficients in the DE algorithm and the strategies for trial vector generation are determined through a self-adaptive mechanism. In other words, during the simulation process, the output weights are estimated utilizing the ELM network, whilst the input weights and biases of the hidden layer neurons are determined through the DE algorithm. The hybrid ML algorithm showed an acceptable ability to model the objective functions (Azimi and Shiri 2021d and Azimi et al. 2021).

Until this stage of the research, only neural network-based (NN-based) algorithms had been used to simulate the subgouge soil parameters. Drawbacks include overfitting issues and higher

computation time that led to the use of other ML algorithms. Hence, it was decided to apply the tree-based ML algorithms for the prediction of the subgouge soil displacement in Chapter 6 which is presented in two published papers. In this chapter, the capability of several robust tree-based algorithms such as decision tree regression (DTR), random forest regression (RFR), gradient boosting regression (GBR), and extra tree regression (ETR) was examined. The obtained results revealed that the ETR model was the premium ML algorithm in terms of precision, correlation, discrepancy, complexity, and computation time to simulate the subgouge soil parameters in both clay and sandy seabed (Azimi et al. 2022b; 2022c; 2022d).

After achieving all of the objectives in Part I and using the lessons learned, Part II was started by determining the dimensionless groups of parameters governing the iceberg draft prediction. These governing parameters and their sensitivity analysis are presented in Chapter 7. The basic regression-based model, e.g., LR, was then employed to derive a set of LR-based equations to surmise the iceberg drafts. Additionally, the best combination of dimensionless groups alongside the most influential inputs in the simulation of the iceberg draft was recognized. In other words, a comprehensive sensitivity analysis was conducted by three NN-based ML algorithms comprising ANN, ELM, and Sa-ELM. A Sa-ELM-based matrix was obtained to calculate the iceberg drafts in Chapter 7 (Azimi et al. 2023c).

Chapter 8 contains four papers, in which the iceberg drafts were predicted utilizing GMDH and GS-GMDH algorithms. An explicit equation for the estimation of iceberg draft in practical applications was driven (Azimi et al, 2023d). Furthermore, to overcome the limitations of the applied NN-based ML algorithms and enhance the simulation performance, a set of non-NN-based algorithms including Support Vector Regression (SVR), K-Nearest Neighbors Regression (KNR), DTR, RFR, GBR, and ETR were utilized for the simulation of the iceberg drafts. A comparison

between ANN, DTR, and SVR models showed that the SVR algorithm was more robust to estimate the iceberg drafts in comparison with the ANN and DTR algorithms. However, the ETR algorithm showed excellent efficiency in dealing with the iceberg draft modeling, with the highest degree of accuracy, correlation, and simplicity.

Ultimately, the outcomes of Part I and II were integrated into Part III by modeling the iceberg drafts and the resultant subgouge soil displacements using superior ML algorithms, e.g., RFR and ETR models. The results of this part were discussed in Chapter 9, which comprises two papers. In this chapter, the iceberg drafts and iceberg-seabed interaction parameters were predicted through the RFR and ETR algorithms. Chapter 10 concludes the key findings of the current research and provides several recommendations for future studies.

References

Azimi, H., Shiri, H., 2020a. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Modeling subgouge sand deformations by using multi-layer perceptron neural network. 31st International Ocean and Polar Engineering Conference. Rhodes, Greece. ISOPE-I-21-2150.

Azimi, H., Shiri, H., 2020b. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., Zendehboudi, S., 2023a. Prediction of Ice-Induced Subgouge Soil Deformation in Sand Using Group Method of Data Handling–Based Neural Network. Journal of Cold Regions Engineering, 37(2): 04023006.

Azimi, H., Shiri, H., Zendehboudi, S., 2022a. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology, 193: 103426.

Azimi, H., Shiri, H., 2021b. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021c. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04): 411-420.

Azimi, H., Shiri, H., 2021d. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., Malta, E.R., 2021. A non-tuned machine learning method to simulate ice-seabed interaction process in clay. Journal of Pipeline Science and Engineering, 1(4): 379-394.

Azimi, H., Shiri, H., Mahdianpari, M., 2022b. Simulation of Subgouge Sand Deformations Using Robust Machine Learning Algorithms. Offshore Technology Conference, Houston, Texas, USA, OTC-31937-MS.

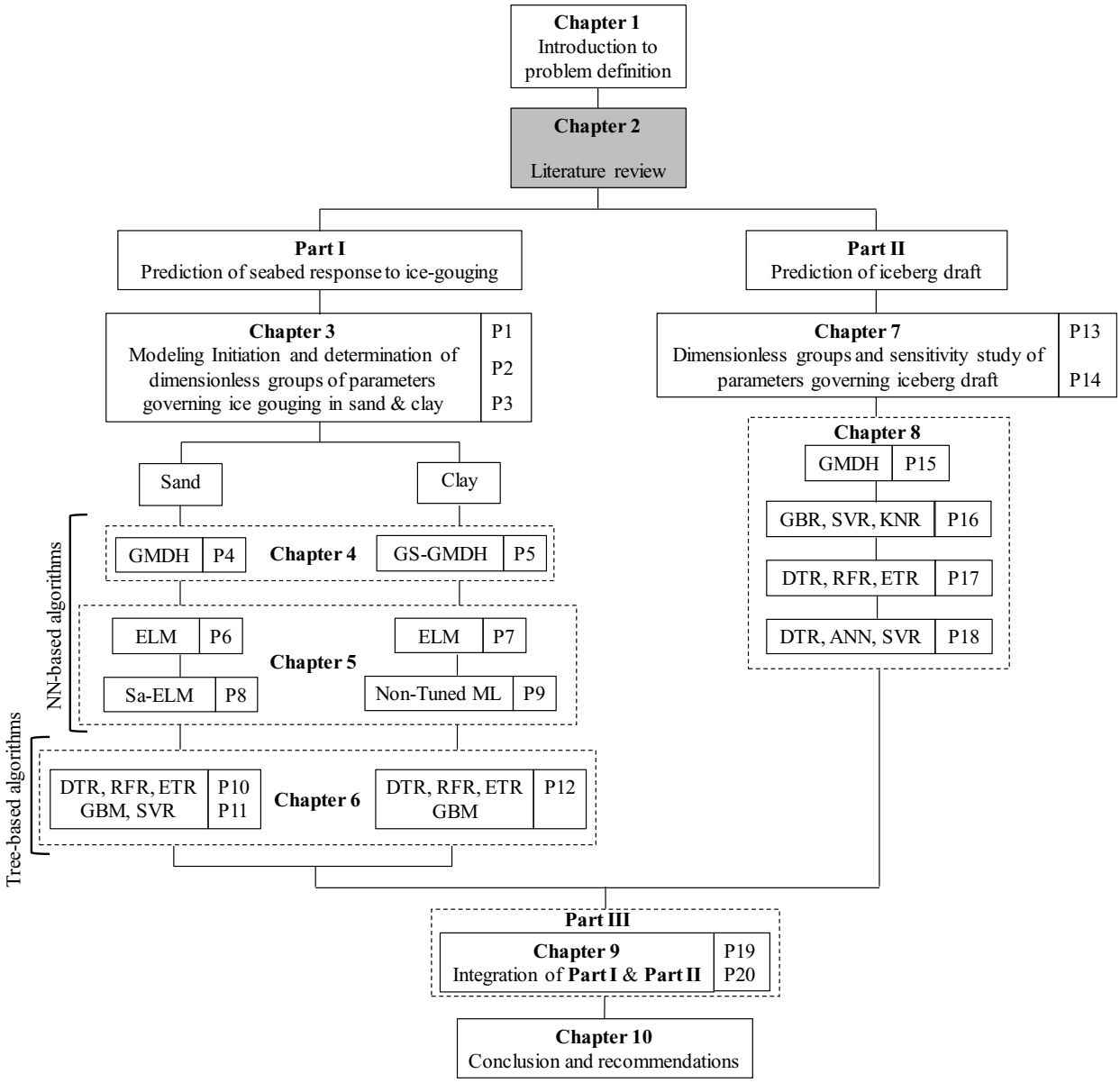
Azimi, H., Shiri, H., Mahdianpari, M., 2022c. Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. Journal of Pipeline Science and Engineering, 2(4): 100075.

Azimi, H., Shiri, H., Mahdianpari, M., 2022d. Iceberg-seabed interaction analysis in sand by a random forest algorithm. Polar Science, 34: 100902.

Azimi, H., Mahdianpari, M., Shiri, H., 2023b. Determination of Parameters Affecting the Estimation of Iceberg Draft. China Ocean Engineering, 37(1): 62-72.

Azimi, H., Shiri, H., Mahdianpari, M., 2023c. Sensitivity analysis of parameters governing the iceberg draft through neural network-based models. Journal of Ocean Engineering and Marine Energy, 1-16.

2. Chapter 2: Literature review



2.1. Overview

This thesis is paper-based research and each chapter has its literature review; hence, a succinct literature review from chapters 3 to 9 was summarized in Chapter 2.

2.1.1. Literature review

2.1.1.1. Iceberg draft estimation

The efficient iceberg management designs and the guaranteed operational integrity of the sea bottom-funded infrastructure against iceberg attacks in ice-prone areas demand the appropriate iceberg draft appraisal, which may lead to a potential decrease in operating expenses and downtime. Earlier investigations have tended to focus on modeling the iceberg draft by using the iceberg length or iceberg mass. For instance, Allaire (1972) analyzed the stability status of various icebergs regarding the sail features. They demonstrated that the minimum stable ratio of iceberg width to sail height for tabular, dry dock, and dome bergs were 6:1, 4:1, and 1.8:1, respectively. Similarly, Robe and Farmer (1976) measured the drafts of tabular, broken tabular, pinnacle, dry dock, and domed by the sonar technology. They correlated the iceberg draft with the iceberg height and proposed a regression model for the estimation of the iceberg draft.

Bass (1980) analytically evaluated the stability of icebergs in different configurations. The study highlighted that the volume of the underwater section was almost seven times greater than the volume of the above-water section. In addition, Brooks (1980) analyzed the stability of theoretical icebergs and provided a hypothesis to estimate the iceberg draft. The investigation demonstrated that the draft of icebergs was smaller than the length of icebergs. In another study, Hotzel and Miller (1983) assessed the iceberg dimensions and supposed that the planimetric section of icebergs was circular, while the length of a berg was twice the circle radius. The iceberg draft was

estimated in terms of iceberg length by using the power-law method. Furthermore, the International Ice Patrol (IIP) has been annually monitoring the characteristics of traveling icebergs in the Atlantic and Arctic Oceans. In 1987, the IIP reported that the iceberg draft was surmised as 3.95 times the height of a sail.

Liang (2001) presented a model to determine the relationship between the dynamics and stability of icebergs using the size and shape of bergs. The author asserted that the draft and hydrostatic force distribution affected the iceberg's stability. Several investigations were performed by the Centre for Cold Ocean Resources Engineering (C-CORE) to study the shape, behavior, and dynamics of icebergs over the years. The above-water shape of bergs comprising the blocky, domed, dry dock, tabular, pinnacle, and wedge icebergs were reported regarding the standard categories. The iceberg dimensions were recorded using the scan sonar technology and marine sextant. The study showed that the iceberg draft was correlated with iceberg length (C-CORE, 2001).

In another study, Barker et al. (2004) evaluated the geometry of iceberg sails and keels. They estimated the cross-sectional areas of the berg at different water depth intervals from a particular waterline length. To approximate the iceberg draft, a set of models in terms of the iceberg length were suggested using the power curve and regression analysis. McKenna (2004) studied the spatial correlation between the above-water and below-water sections of the iceberg. The relationships between the length, width, height, and draft of icebergs were determined, and the three-dimensional shape of icebergs was simulated. Dowdeswell and Bamber (2007) examined the keel depths of traveling icebergs in Antarctic waters. The authors estimated the depth of the keel through the ice thickness and surface elevation. The research concluded that a tiny minority of icebergs in the Antarctica and Greenland waters had drafts of greater than 650 m.

Moreover, Stuckey (2008) modeled the iceberg drift speed using the probabilistic approach. The author stated that the above-water dimension, below-water dimension, and shape factor of the icebergs impacted the environmental driving forces. McKenna and King (2009) simulated the deterioration mechanisms of various icebergs by considering the incremental changes in draft, mass, and shape of bergs. The study highlighted that the draft and length of icebergs were reduced by decreasing the iceberg mass. King (2012) modeled the iceberg's characteristics including the draft, length, and mass through the Monte Carlo simulation. The author outlined that the survey of iceberg drafts greater than 150 m was quite limited. In addition, Sacchetti et al. (2012) analyzed the features of different icebergs, such as wedged, domed, tabular, and pinnacle and ice scouring in the Northeast Atlantic Ocean. The investigation reported that the bimodal distribution of the scour depth was observed across the study area.

Turnbull et al. (2015) forecasted the drift trajectory of the traveling icebergs in Northwest Greenland using the hindcast simulation. It was revealed that the trajectory of icebergs was sensitized to the iceberg drafts. King et al. (2016) performed a field investigation to calculate the iceberg rolling rate. The iceberg drafts were estimated utilizing a calving analysis, with a calculated standard deviation of draft variations from 19% to 34%. The iceberg drafts corresponded with the mass of the icebergs. In another investigation, Talimi et al. (2016) simulated the iceberg-structure interaction using computational fluid dynamics (CFD). The modeling results demonstrated that the iceberg was shifted upward by almost 10% of the iceberg draft. Turnbull et al. (2018) proposed a model for the drift estimation of moving icebergs on the Grand Banks of Newfoundland. This model approximated the draft of icebergs roughly 1.3 times more than the real values.

McKenna et al. (2019) have recently simulated ice scouring on the Grand Banks of Canada using the Monte Carlo method. The iceberg draft alterations were utilized to lessen the size of draft

variations in this modeling. Most recently, Stuckey et al. (2021) simulated the three-dimensional iceberg shapes by adopting field measurements. The investigation demonstrated that the draft and mass of the iceberg were estimated in terms of the iceberg length by utilizing the power curve. Azimi et al. (2023a) determined the dimensionless groups of the parameters affecting the iceberg draft estimation for the first time. The Authors developed a set of linear regression (LR) models using the identified dimensionless groups and recognized the most influencing parameters affecting along with the premium LR models. They presented several LR-based models to approximate the iceberg drafts in daily engineering practices. The best LR model outperformed the previous empirical models. Next, Azimi et al. (2022b) simulated the iceberg drafts through the Generalized Structure of Group Method of Data Handling (GS-GMDH) algorithms. The comparison between the GS-GMDH model with the GMDH and artificial neural network (ANN) models demonstrated the superiority of the GS-GMDH model. An explicit GS-GMDH-based model was then presented to estimate the iceberg draft in real situations. Azimi et al. (2022c) performed a sensitivity analysis of parameters governing the iceberg draft through neural network-based models. They applied three neural network-based machine learning algorithms: self-adaptive extreme learning machine (Sa-ELM), extreme learning machine (ELM), and ANN. The study showed that the Sa-ELM algorithm had a reasonable performance in terms of accuracy, correlation, and complexity in estimating the iceberg drafts. The authors presented a Sa-ELM-based matrix to calculate the iceberg drafts in practical applications. Azimi et al. (2022d) utilized the gradient boosting regression (GBR) algorithm to simulate the iceberg drafts. The premium GBR model and the most significant input parameters were identified by performing a sensitivity analysis. The comparison of the best GBR model's performance with Support Vector Regression (SVR) and K-Nearest Neighbors Regression (KNR) algorithms demonstrated that the GBR

algorithm had the highest level of precision and correlation along with the lowest degree of complexity in predicting the iceberg drafts.

2.1.1.2. Iceberg seabed interaction process

Despite the offshore structures such as ships, wind turbines, and floating or fixed platforms that can be directly attacked by icebergs, the threat that goes to the buried pipelines and power cables is much more complicated involving the ice-soil-structure interactions. During the ice-gouging, the subgouge soil displacement is vastly extended down the seabed much deeper than the gouge depth. This may cause large deformations in the pipelines buried even below the gouge depth. Investigation of the subgouge soil deformation and the best trench depth to bury the pipelines and power cables have been a hot topic in the literature over the past few decades.

Hefty field and experimental investigations, as well as time-consuming numerical simulations, are currently utilized to evaluate the ice-seabed interaction process (Comfort and Graham 1986; Machemehl and Jo 1989; Kioka et al. 2000); however, the industry is constantly looking for alternative approaches to predict the behavior of the subsea pipelines to lessen the collision risk of icebergs with the subsea assets. This means that the subgouge soil deformations and reaction forces are key design factors that should be less than the allowable values suggested by the current codes of practice.

In the past three decades, many researchers have sought to study the parameters affecting the ice-seabed interaction process. For instance, Paulin (1991) performed several experimental ice-gouging tests in sand and clay seabed. The author showed that the vertical reaction forces were more influential for experimental models with a smaller angle of attack, and the horizontal reaction force was the dominant factor for a greater attack angle. Paulin (1992) also implemented some

laboratory tests for submerged and dry sand to quantify the reaction forces and deformations during the ice-scouring event. The author concluded that the ice scouring mechanism in submerged and dry circumstances was almost identical; however, the loads and reaction forces for the submerged sand were smaller than in the dry seabed condition.

Lach (1996) investigated the ice scouring problem experimentally and numerically. The author performed centrifuge tests to determine the ice-clay interaction parameters. The study also showed that the scouring force was a function of undrained shear strength, keel geometry, and subgouge deformation. A series of the Pressure Ridge Ice Scour Experiment (PRISE) joint-industry tests were conducted by Woodworth-Lynes et al. (1996). The author demonstrated that the subgouge deformations could be estimated in terms of the soil depth, gouge depth, and gouge width. At the Center for Cold Ocean Resources Engineering (C-CORE), several centrifuge tests were conducted to measure the ice-induced deformations and reaction forces in both sand and clay seabed (C-CORE 1995; C-CORE 1996). The experimental results showed that the magnitude of scour force increased with increasing the soil shear strength, scour width, and scour depth, while this parameter reduced when the keel attack angle increased. The investigations proved that the horizontal component of the reaction forces grew by increasing the width of the scouring, whereas the horizontal reaction forces reduced with the steeper attack angle. Hynes (1996) investigated the ice keel scour problem for a sandy seabed in a centrifuge study. The author reported the values of scour-induced reaction force, displacement, and pore pressure. They concluded that the relation between the scour loads and scour depths could be described linearly.

Schoonbeek et al. (2006) measured the subgouge clay deformations through centrifuge tests. The authors presented an empirical equation in terms of gouge geometry and soil parameters to estimate the subgouge depth. Been et al. (2008) studied the mechanism of clay failure in ice

scouring problems. The authors concluded that greater deformations occurred in a clay mass with higher undrained shear strength, meaning that the scour depth was a function of the soil properties. Several Pipeline Ice Risk Assessment and Mitigation (PIRAM) tests in the sand condition were conducted by Yang (2009). The effects of gouge depth, frontal berm height, and gouge rate were assessed in this investigation. The author asserted that the gouge geometry and attack angle had a remarkable effect on the reaction forces.

Liu et al. (2015) simulated the ice keel-seabed interaction employing a discrete element method (DEM), and the cohesive frictional model (CFM) was applied to model the freeze bonds between blocks of ice. The numerical results were validated by PIRAM tests, and the authors asserted that the numerical results improved by considering the water drag. Arnau Almirall (2017) conducted several 1g laboratory tests to perform the subgouge sand features in saturated and dry circumstances. The author studied the effect of velocity change, scour geometry, and soil conditions on the subgouge parameters. The study demonstrated that the ice-induced sand deformations in the 1g test are less than those resulting from the centrifuge tests. Shin et al. (2019) developed a three-dimensional model to simulate the ice-gouging phenomenon. The authors showed that the numerical model had a better performance by considering the geostatic stress and contact condition. Nematzadeh and Shiri (2020) modeled, using the modified Mohr-Coulomb (MMC) scheme, the ice-gouging problem in a sandy seabed. The authors concluded that the magnitude of subgouge sand displacements and reaction forces enhanced with the growing value of the unit weight and relative density. Hashemi and Shiri (2022) simulated the iceberg-seabed interaction in clay by incorporating the strain rate and strain-softening effects. The authors conducted a parametric study to assess the impact of different model parameters on the seabed response to iceberg intrusion. This study concluded that strain softening had a significant influence

on a larger amount of plastic shear strain below and in front of the iceberg tip. Hashemi et al. (2022) modeled the impact of the layered seabed, e.g., soft over stiff clay, in the ice-gouging problem. The keel reaction force was reduced by growing the attack angle in this study.

Artificial intelligence (AI) and machine learning (ML) models have been limitedly used for the analysis of iceberg-seabed interaction. Kioka et al. (2003, 2004) used a Neural Network (NN) model for the simulation of the ice-gouging problem in the sand. The NN model was validated using a mechanical approach, and the author concluded that the NN results had a strong correlation with this mechanical approach. Azimi and Shiri (2020a) introduced the dimensionless parameters affecting the iceberg-seabed interaction in the sand by using Buckingham's theory for the first time. The authors proposed some linear regression (LR) models to estimate the maximum subgouge soil displacements. Azimi and Shiri (2020b) simulated the horizontal subgouge soil displacements in the sand through Gene Expression Programming (GEP). The gouge depth and the dilation angle were found as the most significant input variables to predict the objective function. Azimi and Shiri (2021a) utilized the extreme learning machine (ELM) for modeling the iceberg-scoured parameters. The authors suggested a set of ELM-based formulas for estimating the subgouge deformations and reaction forces. The shear strength of the seabed soil and the gouge depth ratio was the most influential input to predict the reaction forces and the subgouge deformations, respectively. Azimi and Shiri (2021b) applied a multi-layer perceptron neural network (MLPNN) to simulate horizontal ice-intrusion displacements. The authors concluded that the soil depth and the bearing pressure had a remarkable impact on the estimation of the target parameter.

Moreover, Azimi and Shiri (2021c) optimized the ELM model using the differential evolution (DE) algorithm to simulate the iceberg-seabed interaction mechanism in the sand. The study

highlighted that the berm height ratio, gouge depth ratio, and vertical component of the load had a significant effect on the simulation of the subgouge soil parameters. Azimi et al. (2021) applied a non-tuned machine learning algorithm to model the sub-gauge soil displacements in clay seabed along with the reaction forces. The authors concluded that the horizontal load factor and the gouge depth ratio were recognized as the most influential input parameters, and a set of equations was provided to estimate the sub-gauge soil features. Azimi and Shiri (2021d) assessed the iceberg-seabed interaction process in clay mass through the ELM model. The soil depth was the most significant input factor governing the subgouge soil deformations. Azimi et al. (2022a) utilized an evolutionary design of the generalized group method of data handling (GS-GMDH) to model the iceberg-seabed interaction mechanism. The comparison of the best GS-GMDH model with the artificial neural network (ANN) and the GMDH algorithm showed the better performance of the GS-GMDH model. Azimi et al. (2022b) simulated the horizontal and vertical subgouge soil deformations in clay by using the Decision Tree Regression (DTR), Random Forest Regression (RFR), and Extra Tree Regression (ETR) models. The simulation results demonstrated that the ETR model possessed the highest degree of precision and correlation with the experimental values. Azimi et al. (2022c) estimated the iceberg-seabed interaction characteristics in clay seabed through decision tree regression (DTR), random forest regression (RFR), and gradient boosting regression (GBR) algorithms. The authors stated that the ETR algorithm was able to predict the target parameters with a better performance.

References

Allaire, P.E., 1972. Stability of simply shaped icebergs. Journal of Canadian Petroleum Technology, 11(04).

Arnau Almirall, S., 2017. Ice-gouging in sand and the associated rate effects. Doctoral dissertation, University of Aberdeen, UK.

Azimi, H., Mahdianpari, M., Shiri, H., 2023a. Determination of Parameters Affecting the Estimation of Iceberg Draft. China Ocean Engineering, 37(1): 62-72.

Azimi, H., Shiri, H., Mahdianpari, M., 2022b. Generalized Structure of Group Method of Data Handling For Modeling Iceberg Drafts. Under review in Ocean Modelling.

Azimi, H., Shiri, H., Mahdianpari, M., 2022c. Sensitivity Analysis of Parameters Governing the Iceberg Draft through Neural Network-Based Models. Journal of Ocean Engineering and Marine Energy, 1-16.

Azimi, H., Shiri, H., Mahdianpari, M., 2022d. Iceberg Draft Prediction using Gradient Boosting Regression Algorithm. Marine Systems & Ocean Technology.

Azimi, H., Shiri, H., 2020a. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021b. Modeling subgouge sand deformations by using multi-layer perceptron neural network. 31st International Ocean and Polar Engineering Conference. Rhodes, Greece. ISOPE-I-21-2150.

Azimi, H., Shiri, H., 2021c. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., Malta, E.R., 2021. A non-tuned machine learning method to simulate ice-seabed interaction process in clay. Journal of Pipeline Science and Engineering, 1(4): 379-394.

Azimi, H., Shiri, H., 2021d. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04): 411-420.

Azimi, H., Shiri, H., Zendehboudi, S., 2022a. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology, 193: 103426.

Azimi, H., Shiri, H., Mahdianpari, M. 2022b. Simulation of Subgouge Sand Deformations Using Robust Machine Learning Algorithms. Offshore Technol. Conference. OTC-31937-MS.

Azimi, H., Shiri, H., Mahdianpari, M., 2022c. Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. Journal of Pipeline Science and Engineering, 2(4): 100075.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In The fourteenth international offshore and polar engineering conference. ISOPE-I-04-116.

Bass, D.W., 1980. Stability of icebergs. Annals of Glaciology, 1: 43-47.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour displacement in clays from physical model tests. In International Pipeline Conference, 48609: 239-245.

Brooks, L.D., 1980. Another hypothesis about iceberg draft. In POAC 79, Proceedings of the 5th International Conference on Port and Ocean Engineering under Arctic Conditions, Vol. 1, pp. 241-252.

Campbell, P.J., 2014. Grand Banks Scour Catalogue (GBSC) GeoDatabase. Geological Survey of Canada, Atlantic.

C-CORE, (2001. Documentation of iceberg grounding events from the 2000 season. C-CORE publication 01-C10 (Revision 0). Report submitted to the Geological Survey of Canada, Atlantic.

C-CORE, 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2.

C-CORE, 1996. PRISE Phase 3c: Extreme LEE Gouge Event-Modeling and Interpretation, C-CORE Report 96-C32.

Comfort, G., Graham, B., Dunwoody, B., 1986. Evaluation of Sea Bottom Ice Scour Models. Environmental Studies Revolving Funds, Vol. 37.

Dowdeswell, J.A., Bamber, J.L., 2007. Keel depths of modern Antarctic icebergs and implications for sea-floor scouring in the geological record. Marine Geology, 243(1-4): 120-131.

Hashemi, S., Shiri, H., 2022. Numerical Modeling of Ice–Seabed Interaction in Clay by Incorporation of the Strain Rate and Strain-Softening Effects. Journal of Offshore Mechanics and Arctic Engineering, 144(4): 042101.

Hashemi, S., Shiri, H., Dong, X., 2022. The influence of layered soil on ice-seabed interaction: Soft over stiff clay. Applied Ocean Research, 120: 103033.

Hotzel, I.S., Miller, J.D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand, Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

King, T., 2012. Iceberg Interaction Frequency Model for Subsea Structures. Arctic Technology Conference. OTC-23787-MS.

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Arctic Technology Conference. OTC-27358-MS.

Kioka, S., Yasunaga, Y., Watanabe, Y., Saeki, H., 2000. Evaluation of ice forces acting on sea bed due to ice scouring. 10th International Offshore and Polar Engineering Conference, ISOPE-I-00-111.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. 13th International Offshore and Polar Engineering Conference. ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. 14th International Offshore and Polar Engineering Conference. ISOPE-I-04-107.

Konuk, I., Gracie, R., 2004. A 3-Dimensional Eulerian Finite Element Model for Ice Scour. Proceedings of the Fifth International Pipeline Conference, pp. 1911–1918.

Lach, P.R., 1998. Centrifuge Modelling of Large Soil Deformation due to Ice Scour. Doctoral dissertation, Memorial University of Newfoundland. St. John's. NL. Canada.

Liang, B., 2001. Iceberg stability and deterioration. Doctoral dissertation, Memorial University of Newfoundland. St. John's. NL. Canada.

Liu, L., Bailey, E., Sarracino, R., Taylor, R., Power, C., Stanbridge, C., 2015. Numerical simulation of ice ridge gouging. 34th International Conference on Ocean, Offshore and Arctic Engineering, 56567: V008T07A022.

Machemehl, J.L., Jo, C.H., 1989. Note on nearshore ice gouge depths in Alaskan Beaufort Sea. Journal of cold regions engineering, 3(3): 150-153.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report: 20473.

McKenna, R., King, T., 2009. Modelling Iceberg Shape, Mass and Draft Changes. In Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions, No. POAC09-107.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. In Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, pp. 9-19.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Nobahar, A., Kenny, S., King, T., McKenna, R., Phillips, R. (2007). Analysis and design of buried pipelines for ice-gouging hazard: A probabilistic approach. Journal of Offshore Mechanics and Arctic Engineering, 129(3): 219-228.

Minerals Management Service., 2002. Evaluation of Sub–Sea Physical Environmental Data for the Beaufort Sea OCS and Incorporation into a Geographic Information System (GIS) Database. OCS Study MMS 2002–017.

Papusha, A., Kazunin, D., Gontarev, D., Vasilevich, V., Kyalbieva, S., 2013. Stability and strength of the subsea pipeline under iceberg load in Arctic. SPE Arctic and Extreme Environments Technical Conference and Exhibition, Moscow, Russia. SPE-166941-MS.

Paulin, M.J., 1991. Preliminary Results of Physical Model Tests of ice Scour, Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering, St. John's, NL, Canada.

Paulin, M.J., 1992. Physical model analysis of iceberg scour in dry and submerged sand. Master thesis, Memorial University of Newfoundland, St. John's, NL, Canada.

Phillips, R., Nobahar, A., Zhou, J., 2004. Trench Effects on Pipeline/Soil Interaction. ASME International Pipeline Conference, IPC04-0141.

Phillips, R., Barrett, J. A., Al-Showaiter, A., 2010. Ice Keel–Seabed Interaction: Numerical Modelling Validation. Offshore Technology Conference, Vol. 20696: pp. 13.

Robe, R.Q., Farmer, L.D., 1976. Physical Properties of Icebergs. Part I. Height to Draft Ratios of Icebergs. Part II. Mass Estimation of Arctic Icebergs. Coast Guard Research and Development Center Groton Conn.

Sacchetti, F., Benetti, S., Cofaigh, C.Ó., Georgiopoulou, A., 2012. Geophysical evidence of deep-keeled icebergs on the Rockall Bank, Northeast Atlantic Ocean. Geomorphology, 159: 63-72.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip Line Field Solutions as an Approach to Understand Ice Subgouge Deformation Patterns. 16th International Offshore and Polar Engineering Conference. ISOPE-I-06-289.

Shin, M.B., Park, D.S., Seo, Y.K., 2019. Comparative study of ice gouge simulation considering ice keel-seabed interactions. Journal of Ocean Engineering and Technology, 33(6): 556-563.

Stuckey, P.D., 2008. Drift speed distribution of icebergs on the grand banks and influence on design loads. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., King, T., 2021. Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. In International Conference on Offshore Mechanics and Arctic Engineering, 85178: V007T07A018.

Talimi, V., Ni, S., Qiu, W., Fuglem, M., MacNeill, A., Younan, A., 2016. Investigation of Iceberg Hydrodynamics. Arctic Technology Conference. OTC-27493-MS.

Turnbull, I.D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Regions Science and Technology, 110: 1-18.

Turnbull, I. D., King, T., Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference. OTC-29109-MS.

Winkler, E., 1867. Die Leher von der Elastizitat und Festigkeit, Dominicus, Prague Dominicus, Prague.

Woodworth-Lynes, C., Nixon, D., Phillips, R., Palmer, A., 1996. Subgouge deformations and the security of Arctic marine pipelines. Offshore Technology Conference. OTC-8222-MS.

Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John’s, NL, Canada

Part I

Prediction of Seabed Response to Ice-Gouging

3. Chapter 3

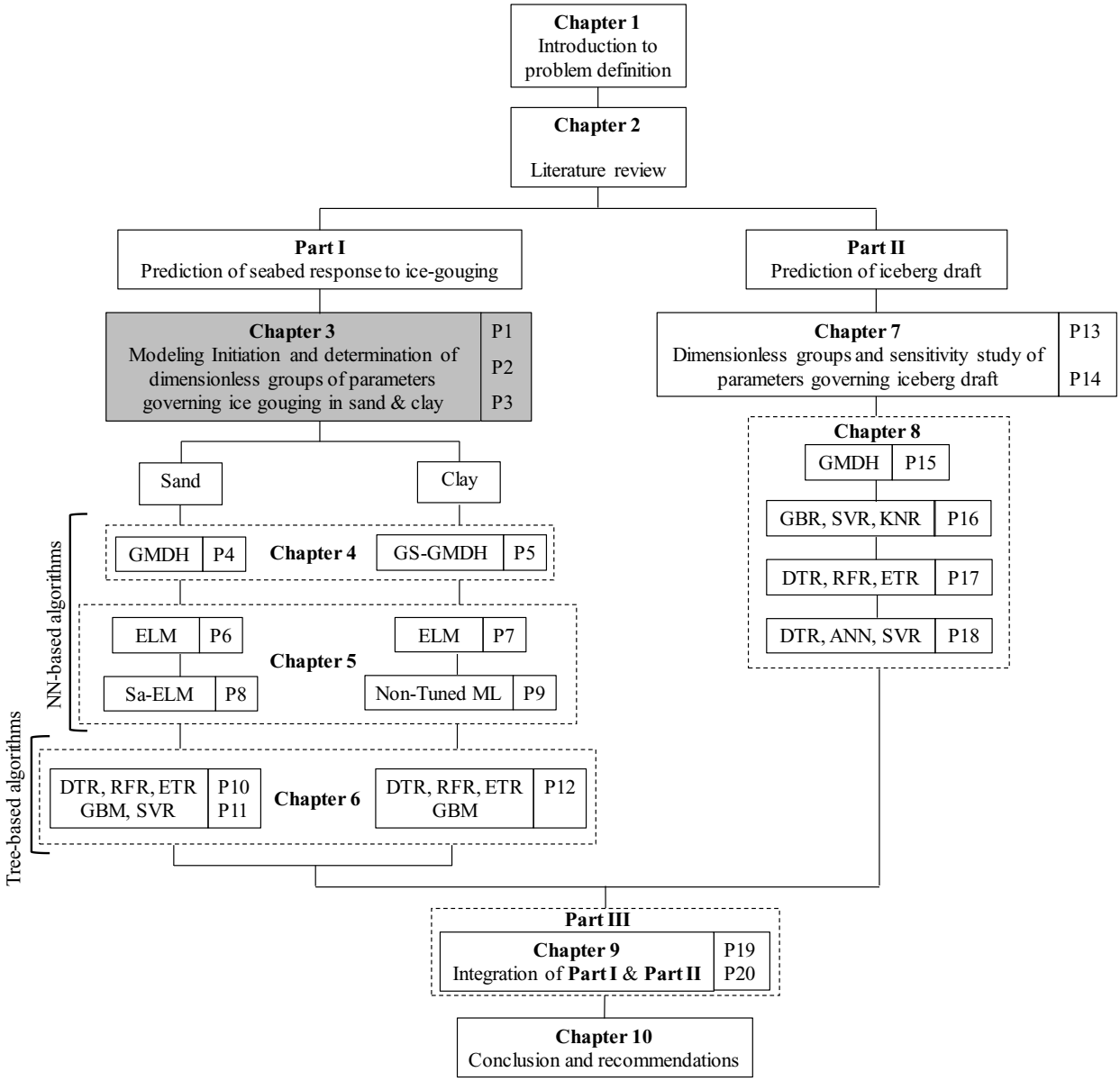
Modeling Initiation and determination of dimensionless groups of parameters governing ice-gouging in sand & clay

This chapter includes three sections as follows:

Section 1: Ice-Seabed Interaction Analysis in Sand Using a Gene Expression Programming-Based Approach

Section 2: Modeling Subgouge Sand Deformations by Using Multi-Layer Perceptron Neural Network

Section 3: Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process


Chapter 1
Introduction to problem definition
Chapter 2
Literature review
Part I
Prediction of seabed response to ice-gouging
Part II
Prediction of iceberg draft
Chapter 3
Modeling Initiation and determination of dimensionless groups of parameters governing ice gouging in sand & clay
P1
P2
P3
Chapter 7
Dimensionless groups and sensitivity study of parameters governing iceberg draft
P13
P14
Chapter 8
GMDH P15
GBR, SVR, KNR P16
DTR, RFR, ETR P17
DTR, ANN, SVR P18
Sand
Clay
NN-based algorithms
GMDH P4
Chapter 4
GS-GMDH P5
ELM P6
ELM P7
Chapter 5
Sa-ELM P8
Non-Tuned ML P9
Tree-based algorithms
DTR, RFR, ETR GBM, SVR P10 P11
Chapter 6
DTR, RFR, ETR GBM P12
Part III
Chapter 9
Integration of Part I & Part II
P19
P20
Chapter 10
Conclusion and recommendations

Section 1

Ice-Seabed Interaction Analysis in Sand Using a Gene Expression Programming-Based Approach

This section is a journal paper which has been published in Applied Ocean Research (2020), Vol. 98, pp. 102120, ISSN: 0141-1187

Abstract

Arctic subsea pipelines are usually buried for physical protection against ice-induced scours. Determination of the maximum horizontal deformations for guaranteed operational integrity and cost-effective design is probably the most challenging aspect of subsea pipelines in ice-prone areas. The large uncertainties associated with the design of an ice-protected subsea pipeline using the existing empirical equations, advanced experimental, and sophisticated numerical studies with significant time and cost impacts are usually preferred to ensure the sufficiency and cost-effectiveness of the pipeline design against the ice attack. This has caused the industry to keep looking for more effective and less-costly solutions for modeling the ice impact on buried pipelines. In this study, a Gene Expression Programming (GEP) model representing the Artificial Intelligence (AI) approaches was used for the first time to simulate the subgouge soil deformation in the sand. A database was constructed using some of the published experimental studies identifying the key input parameters including soil depth, bearing pressure, the maximum vertical extent of subgouge deformation, attack angle, and dilation index. Subsequently, six GEP models were developed and validated by using a K-fold cross-validation method. The performance of the GEP method was compared with an Artificial Neural Network (ANN) model, and uncertainty analysis (UA) along with a partial derivative sensitivity analysis (PDSA) was conducted to assess the influence domain of the key parameters. The study showed that the evolutionary numerical methods could be used as an accurate and cost-effective alternative for modeling ice-induced subgouge deformations.

Keywords: Ice-gouging; Subgouge sand deformation; Gene expression programming (GEP); Artificial neural network (ANN); K-fold cross-validation; Partial derivative sensitivity analysis (PDSA)

3.1.1. Introduction

Arctic territories contain a huge amount of hydrocarbon deposits such as crude oil and natural gases that increased energy demand is the cause of exploration in the Arctic regions. Subsea pipelines are widely used to transfer hydrocarbon and other exploration and production-related contents between the onshore and offshore facilities (Emmerson and Lahn 2012; Alba 2015). Marine pipelines are threatened by the ice-gouging, and pressure ridges attack crossing the pipeline route in shallow waters. Subsea trenching and backfilling are commonly used to bury the pipeline for physical protection against the ice scour (see Figure 3-1).

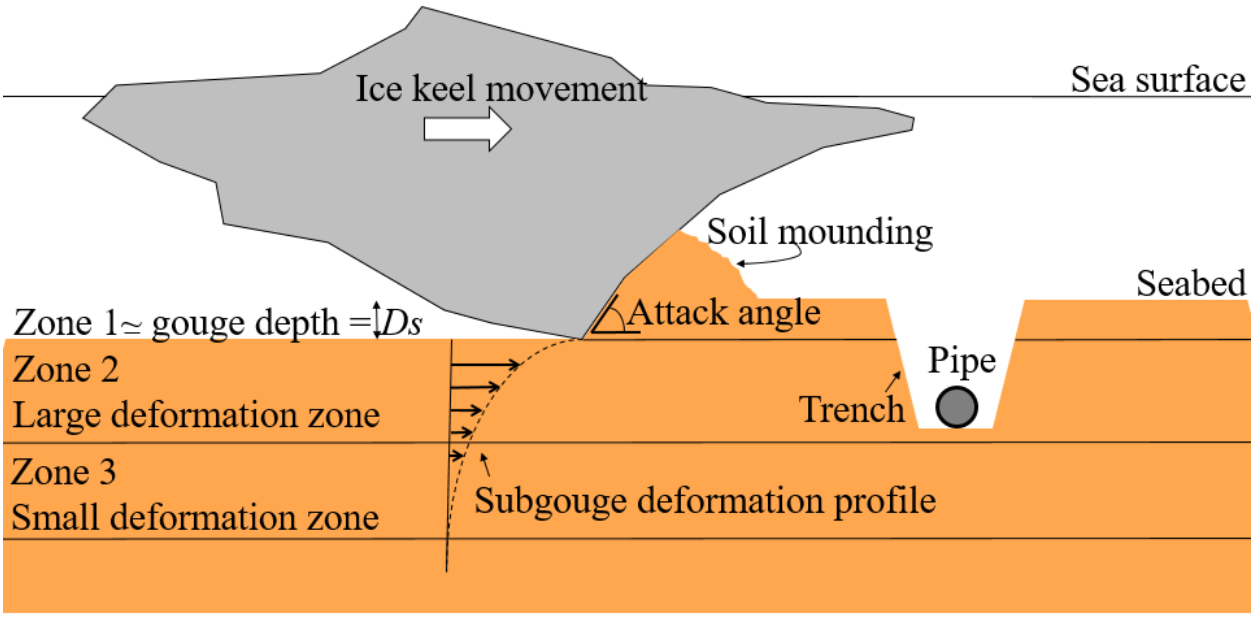


Figure 3-1. Schematic layout of the ice-scoured seabed

Identifying the maximum deformations for safe and cost-effective protection of the pipeline is a challenging problem. Costly experimental and advanced numerical simulation is mandatory for accurate modeling of the subgouge soil deformation and consequently, the pipe response. Semi-

empirical equations are often used in industry for a faster and less expensive assessment through a decoupled analysis by scarifying some level of accuracy. However, this approach also needs a costly free-field ice-gouging analysis and still imposes oversimplification expenses on the practice. The industry is still looking for alternative approaches to mitigate the computational and experimental efforts and improve the accuracy of simulation at the same time.

In this study, the application of an evolutionary method, i.e., Gene Expression Programming (GEP), was investigated as an alternative approach for the assessment of the subgouge soil deformation in the sand. In contrast, artificial intelligence (AI) approaches and soft computing (SC) techniques such as Artificial Neural Networks (ANNs), Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Machine (SVM), Group Method of Data Handling (GMDH) and Gene Expression Programming (GEP) are extensively applied to model complex and nonlinear problems with a large number of influential parameters (Rajaee 2011; Guven, and Kisi 2013; Kaydani et al. 2014; Haghiabi 2016; Azimi et al. 2017; Ebtehaj et al. 2018; Najafzadeh et al. 2018a; Najafzadeh et al. 2018b; Shaghaghi et al. 2018; Yavari et al. 2018; Moradi et al. 2019). For instance, Kioka and Kubouchi (2003) estimated the characteristics of the ice scour problem by adopting the neural network (NN). The attack angle, the ice velocity, the mean grain size of sediment, and the sea-bottom slope were assumed as the input parameters, whereas parameters of scour curve were considered as the output of the model. Moreover, Kioka et al. (2004) combined a neural network (NN) approach with a mechanical model to estimate the ice-induced scour depth. A sensitivity analysis was performed and then the bottom shape of the ridge and the ice condition surrounding the ridge were introduced as the most important factors affecting during an ice scouring event. Therefore, these cost-effective and accurate tools have many privileges, such as high modeling accuracy and enhanced simulation speed.

In this study, the GEP method was adopted by using the data collected from the existing experimental studies. The GEP method has been successfully applied in the literature to model the suspended sediment and sediment scour (Azamathulla et al. 2013; Shabanlou et al. 2018). However, this eminent technique has never been applied for simulation of the subgouge soil deformation, and the current study is considered the first time application of its kind. The GEP is quite flexible and versatile compared to other evolutionary methods like a genetic algorithm (GA), and Genetic Programming (GP). Furthermore, chromosomes can be genetically manipulated easily since they are pretty small, compact, and linear (Ferreira 2001). Therefore, this facilitated the useful adoption of this methodology to assess the ice-gouging process.

The commonly known key factors affecting the subgouge deformations in the sand were considered as the input parameters to define six GEP models. Additionally, the K-fold cross-validation method was applied to train and test the GEP models. The GEP models were analyzed and the best model with the corresponding most influential input parameters was identified. Simultaneously, the horizontal subgouge sand deformation was modeled using the ANN, and the results were compared with the GEP. Ultimately, an uncertainty analysis (UA) and a Partial Derivative Sensitivity Analysis (PDSA) were conducted to scrutinize the model performance.

The study resulted in highly promising outcomes; for instance, the GEP method had a reasonable accuracy in modeling ice-induced seabed soil deformations. In addition, the approach significantly reduced the analysis cost and study conducting time. Moreover, the PDSA analysis revealed invaluable insight into the effect of various parameters affecting the seabed response to the ice-gouging phenomenon. Finally, compared to an empirical model, the evolutionary method was found to be an appropriate alternative for the assessment of ice-induced scour in the sandy seabed.

3.1.2. Published experimental studies

The current GEP model was fed by the results of selected experimental studies published in the literature. Some of these experimental studies are shortly reviewed in this section to facilitate presenting the results obtained from the current investigation. For instance, Harrison (1973), Chari (1975), Green and Chari (1981), Prasad (1985), and Longva and Bakkejord (1990) investigated different aspects of the ice-gouging problem.

Subsequently, a joint industry and government-sponsored research program entitled the Pressure Ridge Ice Scour Experiment (PRISE) was conducted at the Centre for Cold Ocean Resources Engineering (C-CORE 1995) to understand the requirement for the safe and cost-effective design of subsea pipelines against the ice scour. The study showed the importance of the "dead wedge" underneath the ice keel with respect to variation in scour loads and subscour soil displacements. Analysis of the results served to support Been's (1190) shear-dragging hypothesis. Hynes (1996) carried out a centrifuge ice-gouging study in the sand and observed a linear relationship between the scour loads and depth. The author showed that sand deformation affected simple direct shear due to the stress-strain behavior of soil. Eventually, it was suggested numerical studies should be performed to simulate the centrifuge ice-gouging modeling. Woodworth-Lynas et al. (1998) investigated the soil deformation caused by an ice keel scour event in a centrifuge study. The authors showed that the changes in horizontal load, vertical load, and pore pressure could be easily measured in the experimental model, which was arduous to be observed in the field study. Barker and Timco (2003) extended their experimental study on the ice scour processes to assess the seabed response and loads for a buoyant ice model at the Canadian Hydraulics Centre. The authors concluded that the size and shape of the ice block could influence the results of the study. Phillips et al. (2005) reviewed the pressure ridge ice scour experiment (PRISE) studies and the subgouge

deformations. It was concluded that the shear along the dead wedge under the ice keel was identified as the most significant factor of soil resistance. Yang (2009) conducted a centrifuge study to measure deformations of ice-scoured sand. It was shown that the maximum gouge force was a function of the gauge geometry and keel attack angle and value of the frontal berm height. Furthermore, Bruce et al. (2012) carried out a C-CORE-led Joint Industry Program (JIP) to study the failure mechanism, the confining pressure, and the strength limits of a seabed gouging ice ridge keel. The authors indicated that the keel strength was enhanced by increasing the confining pressure. Phillips and Barrett (2012) reviewed the Pipeline Ice Risk Assessment & Mitigation (PIRAM) JIP studies to assess the 3D mechanical behavior of pipelines in response to ice-gouging events. The authors applied a fully coupled pipeline, seabed, and ice interaction model for the calibration of a pipeline design method. Barrette and Sudom (2014) compiled the data and tests related to ice-soil-pipeline interaction studies. The authors suggested that the database should be analyzed and the parameters affecting the ice-gouging events could be investigated to determine some relationships between the parameters. Arnau and Ivanović (2019) carried out 1g floor tests on cohesion-less seabed scour due to the ice attack. Ultimately, drifting velocity on the scouring loads was identified as an important parameter.

The studies reviewed in the literature attempt to provide a better understanding of parameters affecting the ice-soil-pipeline interaction in an ice-gouging event. Although some parameters, including deformations and loads, may be measured in the experimental investigations, all non-linear parameters governing a design method cannot be observed. In addition, despite a large amount of published experimental studies resulting in some empirical design equations, many companies still prefer to go through costly experimental and numerical studies for specific design cases. The reason behind this is the significant amount of uncertainties associated with several

non-linear parameters governing the design outcome. In this study, the capability of an evolutionary artificial intelligence (AI) approach for an accurate and cost-effective assessment of the seabed response to ice scour was investigated as an alternative approach. Further details are provided in the coming sections.

3.1.3. Horizontal subgouge soil deformation in sand

Figure 3-2 shows samples of numerical subgouge sand deformations, the corresponding keel reaction forces, and their comparison with experimental studies for a one-meter gouge depth (Nematzadeh and Shiri 2019).

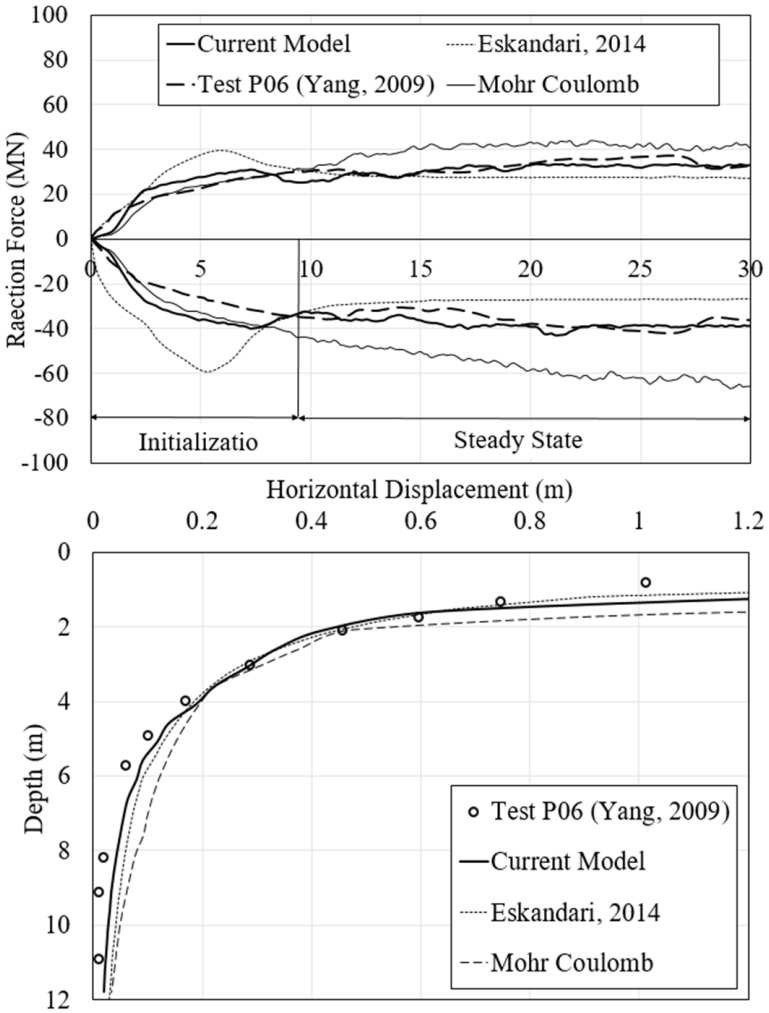


Figure 3-2. Samples of numerical and experimental subgouge sand deformations (Nematzadeh and Shiri 2019)

The maximum horizontal deformation occurs right underneath the keel base. Moving down the seabed, the magnitude of horizontal soil deformation is reduced by a hyperbolic curve. The maximum horizontal subgouge soil deformation (d_{h0}), was obtained using the empirical equations proposed by C-CORE (Yang 2009; C-CORE 2008; C-CORE 2009a; C-CORE 2009b):

$$d_{h0}\sqrt{(\sigma_b/P_a)}/V_e = (0.54/\tan\alpha)\big(1-0.2(I_r+1)\big) \tag{3-1}$$

where, σ_b is bearing pressure (MPa), P_a is atmospheric pressure (0.1 MPa), α is the attack angle (degree), V_e is the maximum vertical extent of subgouge deformation (m), and I_r is dilation index given as:

$$I_r = RD \times (10 - \ln(\sigma_b \times 1000)) - 1 \tag{3-2}$$

where RD is the relative density of sand. The bearing pressure was defined as:

$$\sigma_b(MPa) = 0.09D_s^{1.5} \tag{3-3}$$

where D_s is the gouge depth (m). The maximum vertical extent of subgouge deformation (V_e) was determined based on the range of the dilation index as follows:

$$V_e = \begin{cases} \min[(1+0.5I_r)(D_s+W/5), 1.3(D_s+W/5{,}5D_s)] & if \quad I_r < 2 \\ \min\big[\big(2.6+0.65(I_r-4)\big)(D_s+W/5), 5D_s\big] & if \quad 2 < I_r < 4 \\ 0 & if \quad I_r > 4 \end{cases} \tag{3-4}$$

where W is the gouge width (m). Equations (3-1) to (3-4) denote the horizontal subgouge deformation (d_{h0}) is a function of $\sigma_b, P_a, V_e, \alpha$ and I_r:

$$d_{h0} = f(\sigma_b, P_a, V_e, \alpha, I_r) \tag{3-5}$$

Assuming a constant atmospheric pressure during the tests, so Eq. (3-5) can be written as follows:

$$d_{h0} = f(\sigma_b, V_e, \alpha, I_r) \quad (3\text{-}6)$$

Therefore, incorporating the soil depth (y) effect, the target function of horizontal subgouge deformation (d_{hy}) for the artificial intelligence approaches was defined as follows:

$$d_{hy} = f(\sigma_b, V_e, \alpha, I_r, y) \quad (3\text{-}7)$$

The input parameters given in Eq. (3-7) were combined to construct six different GEP models (see Figure 3-3) with a minimum number of four key parameters.

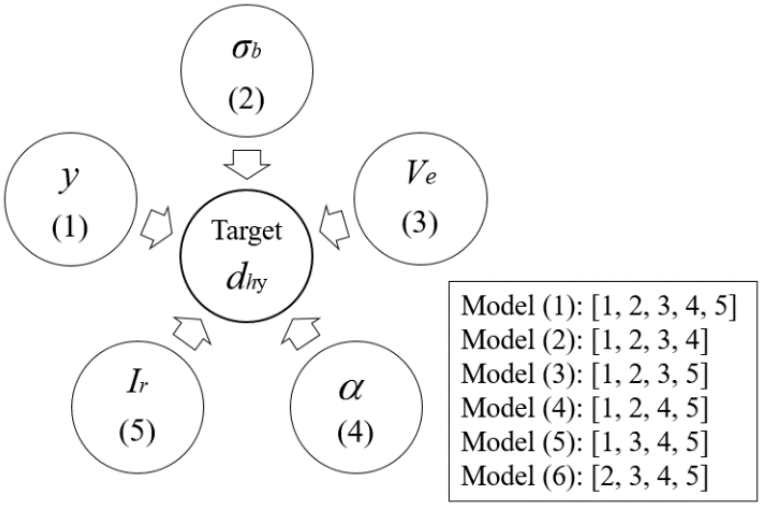


Figure 3-3. Combination of the input parameters for different models

Gaining a good insight into the best input combination (the superior GEP model) and the most important input parameters in an ice-induced scouring problem was one of the key objectives of the current study. To do this, GEP 1 was defined by using all input parameters ($\sigma_b, V_e, \alpha, I_r, y$) and then five GEP models were produced adopting a combination of the other four input parameters. For instance, the σ_b, V_e, α, y parameters for GEP 2, the $\sigma_b, V_e, \alpha, I_r$ parameters for GEP 3, the σ_b, α, I_r, y parameters for GEP 4, the V_e, α, I_r, y parameters for GEP 5, and the $\sigma_b, V_e, \alpha, I_r$ parameters for GEP 6 were utilized. Next, a sensitivity analysis (SA) was performed by comparison of the results from the defined GEP models. An alternative combination of input

parameters could be also defined by using one, two, three, four (GEP 2 to GEP 6), and five input parameters (GEP 1). However, the sensitivity analysis (SA) for models consisting of four and five input parameters identified the best GEP model and the most influencing input parameters without needing other combinations of parameters. As a result, the model GEP 1 simulated the target function by using all of the input parameters and eliminating the effects of other parameters for GEP 2 to GEP 6 including *σb, Ve, α, Ir*, and *y* (Najafzadeh et al. 2018; Shaghaghi et al. 2019). The best GEP model owns the lowest error and the highest correlation. In contrast, the accuracy of the GEP model decreased remarkably by removing the most effective input parameter since the level of effectiveness for the eliminated parameter was significant. Therefore, the applied configuration of input parameters was able to properly perform the sensitivity analysis.

3.1.4. Construction of database

The studies published by C-CORE (1995), Hynes (1996), and Yang (2009) are amongst the most important experimental studies related to ice-induced sand scouring problems in which input parameters $(\sigma_b, V_e, \alpha, I_r, y)$ and the target parameter (d_{hy}) can be expressly adopted to feed the developed GEP models. The key results of the experimental studies conducted by C-CORE (1995) (eight test results in the sand with different relative densities), Hynes (1996) (seven centrifuge tests in clean dry silica sand), and Yang (2009) (seven centrifuge ice-gouging test in AlWhite Silica sand) were used to construct a database for validation of the artificial intelligence numerical models. Using these experimental studies, the attack angle (α), the scour depth (D_s), the scour width (W), the g-level, the relative density (*RD*), and the horizontal subgouge deformation (d_{hy}) at different depths (y) were extracted and summarized in Table 3-1 to Table 3-3

Table 3-1. Characteristics of applied experimental parameters (1995)

Test ID	dhy (m)	y (m)	α (degree)	Ds	W (m)	g-level	RD %
	2.4	0	15	2	15	150	0.374
	1.65	0.75	15	2	15	150	0.374
	1.05	1.5	15	2	15	150	0.374
	0.68	2.25	15	2	15	150	0.374
	0.38	3	15	2	15	150	0.374
	0.08	3.75	15	2	15	150	0.374
PRISE01B	0	4.5	15	2	15	150	0.374
	0.75	0.000	30	1	15	75	0.370
	0.58	0.375	30	1	15	75	0.370
	0.45	0.750	30	1	15	75	0.370
	0.35	1.125	30	1	15	75	0.370
	0.25	1.500	30	1	15	75	0.370
	0.2	1.875	30	1	15	75	0.370
	0.16	2.250	30	1	15	75	0.370
	0.11	2.625	30	1	15	75	0.370
PRISE01C	0.05	3.000	30	1	15	75	0.370
	0.04	3.375	30	1	15	75	0.370
-Drive 1	0.01	3.750	30	1	15	75	0.370
	0	4.125	30	1	15	75	0.370
	0.71	0.000	30	1	15	75	0.370
	0.59	0.375	30	1	15	75	0.370
	0.54	0.750	30	1	15	75	0.370
	0.49	1.125	30	1	15	75	0.370
	0.41	1.500	30	1	15	75	0.370
	0.32	1.875	30	1	15	75	0.370
	0.27	2.250	30	1	15	75	0.370
	0.24	2.625	30	1	15	75	0.370
	0.19	3.000	30	1	15	75	0.370
	0.15	3.375	30	1	15	75	0.370
	0.13	3.750	30	1	15	75	0.370
	0.09	4.125	30	1	15	75	0.370
	0.05	4.500	30	1	15	75	0.370
	0.02	4.875	30	1	15	75	0.370
PRISE01C	0	5.250	30	1	15	75	0.370
	3.45	0.000	15	1	15	150	0.574
	2.4	0.750	15	1	15	150	0.574
	1.65	1.500	15	1	15	150	0.574
	1.2	2.250	15	1	15	150	0.574
PRISE09	0.9	3.000	15	1	15	150	0.574
	0.45	3.750	15	1	15	150	0.574
Drive 1	0.15	4.500	15	1	15	150	0.574
	3.53	0.000	15	2	15	150	0.574

	2.04	0.750	15	2	15	150	0.574
	1.32	1.500	15	2	15	150	0.574
	0.66	2.250	15	2	15	150	0.574
	0.36	3.000	15	2	15	150	0.574

Table 3-2. Characteristics of applied experimental parameters Hynes (1996)

Test ID	dhy (m)	y (m)	α (degree)	Ds (m)	W (m)	g-level	RD %
	0.42	0.750	15	1.70	15	150	0.3
	0.35	1.500	15	1.70	15	150	0.3
	0.18	3.000	15	1.70	15	150	0.3
	0.00	4.500	15	1.70	15	150	0.3
	0.00	6.000	15	1.70	15	150	0.3
	0.00	7.500	15	1.70	15	150	0.3
	0.00	9.000	15	1.70	15	150	0.3
	0.00	10.50	15	1.70	15	150	0.3
Scour B	0.00	12.00	15	1.70	15	150	0.3
	0.00	13.50	15	1.70	15	150	0.3
	0.50	0.150	30	0.98	15	75	0.3
	0.38	1.130	30	0.98	15	75	0.3
	0.30	1.500	30	0.98	15	75	0.3
	0.17	2.250	30	0.98	15	75	0.3
	0.06	3.000	30	0.98	15	75	0.3
	0.00	3.750	30	0.98	15	75	0.3
	0.00	4.500	30	0.98	15	75	0.3
	0.00	5.250	30	0.98	15	75	0.3
	0.00	6.000	30	0.98	15	75	0.3
	0.00	6.150	30	0.98	15	75	0.3
	0.49	0.750	15	1.10	15	75	0.3
	0.44	1.130	15	1.10	15	75	0.3
	0.31	1.500	15	1.10	15	75	0.3
	0.21	2.250	15	1.10	15	75	0.3
	0.14	3.000	15	1.10	15	75	0.3
	0.09	3.750	15	1.10	15	75	0.3
	0.04	4.500	15	1.10	15	75	0.3
	0.00	5.250	15	1.10	15	75	0.3
Scour D	0.00	6.000	15	1.10	15	75	0.3
	0.00	6.750	15	1.10	15	75	0.3
	0.62	0.750	15	1.20	15	150	0.7
	0.35	1.500	15	1.20	15	150	0.7
	0.12	3.000	15	1.20	15	150	0.7
	0.12	4.500	15	1.20	15	150	0.7
	0.00	6.000	15	1.20	15	150	0.7
	0.00	7.500	15	1.20	15	150	0.7

Scour E	0.00	9.000	15	1.20	15	150	0.7
	0.00	10.50	15	1.20	15	150	0.7
	0.00	12.00	15	1.20	15	150	0.7
	0.00	13.50	15	1.20	15	150	0.7
Scour F	0.95	0.750	15	2.14	15	150	0.7
	0.60	1.500	15	2.14	15	150	0.7
	0.20	3.000	15	2.14	15	150	0.7
	0.00	4.500	15	2.14	15	150	0.7
	0.00	6.000	15	2.14	15	150	0.7
	0.00	7.500	15	2.14	15	150	0.7
	0.00	9.000	15	2.14	15	150	0.7
	0.00	10.50	15	2.14	15	150	0.7
	0.00	12.00	15	2.14	15	150	0.7
	0.00	13.50	15	2.14	15	150	0.7

Table 3-3. Characteristics of applied experimental parameters Yang (2009)

Test ID	dhy (m)	y (m)	α ◦	Ds	W (m)	g-level	RD %
P02	0.262	0.0001	30	1.30	10.0	55.6	0.936
	0.140	0.0061	30	1.30	10.0	55.6	0.936
	0.094	0.017	30	1.30	10.0	55.6	0.936
	0.074	0.022	30	1.30	10.0	55.6	0.936
	0.048	0.033	30	1.30	10.0	55.6	0.936
	0.027	0.044	30	1.30	10.0	55.6	0.936
	0.006	0.054	30	1.30	10.0	55.6	0.936
P05	0.065	0	30	1.40	10.0	55.6	0.515
	0.046	0.011	30	1.40	10.0	55.6	0.515
	0.017	0.022	30	1.40	10.0	55.6	0.515
	0.006	0.028	30	1.40	10.0	55.6	0.515
P06	0.048	0	30	0.18	2.20	12.2	0.58
	0.010	0.009	30	0.18	2.20	12.2	0.58
	0.008	0.019	30	0.18	2.20	12.2	0.58
	0.005	0.029	30	0.18	2.20	12.2	0.58
	0.003	0.039	30	0.18	2.20	12.2	0.58
	0.026	0	30	2.30	14.4	80	0.508
	0.014	0.006	30	2.30	14.4	80	0.508
	0.008	0.017	30	2.30	14.4	80	0.508
	0.007	0.022	30	2.30	14.4	80	0.508
	0.004	0.033	30	2.30	14.4	80	0.508
	0.003	0.044	30	2.30	14.4	80	0.508
	0	0.054	30	2.30	14.4	80	0.508
	0.024	0	30	2.40	14.4	80	0.390
	0.018	0.009	30	2.40	14.4	80	0.390
	0.018	0.017	30	2.40	14.4	80	0.390

	0.015	0.025	30	2.40	14.4	80	0.390
	0.009	0.033	30	2.40	14.4	80	0.390
	0.007	0.041	30	2.40	14.4	80	0.390
P08	0.026	0	30	0.19	2.20	12.2	0.681
	0.013	0.008	30	0.19	2.20	12.2	0.681
	0.007	0.015	30	0.19	2.20	12.2	0.681
	0.006	0.023	30	0.19	2.20	12.2	0.681
	0.003	0.031	30	0.19	2.20	12.2	0.681
	0.002	0.038	30	0.19	2.20	12.2	0.681
	0.001	0.046	30	0.19	2.20	12.2	0.681
P09	0.024	0	15	1.20	16.0	80	0.386
	0.018	0.008	15	1.20	16.0	80	0.386
	0.010	0.015	15	1.20	16.0	80	0.386
	0.005	0.022	15	1.20	16.0	80	0.386
	0.003	0.030	15	1.20	16.0	80	0.386
	0.002	0.037	15	1.20	16.0	80	0.386
	0.001	0.045	15	1.20	16.0	80	0.386
	0.001	0.053	15	1.20	16.0	80	0.386
	0.001	0.062	15	1.20	16.0	80	0.386
	6.5E-05	0.069	15	1.20	16.0	80	0.386
	1.9E-05	0.077	15	1.20	16.0	80	0.386

Hynes (1996) used loose and dense sands but did not mention the values of *RD*. Using an existing classification study (e.g., Mitchell and Soga 2005), relative densities of 0.3 and 0.7 were adopted for loose and dense sands of Hynes (1996).

Care should be taken in using the laboratory test data for testing the AI models. The idealized test conditions may affect the reliability of model prediction to some extent. To overcome this limitation, the majority of the data in the current study were selected from centrifuge studies which is one of the most reliable technics for model testing in geotechnical engineering. The centrifugal acceleration simulates gravity and allows for correspondence of stress fields between the model and full-scale resulting in an accurate model prediction within a bandwidth of less than 1.0 times the normalized loads (C-CORE 1995; C-CORE 2009b). This provided a relaxation to build up the current database and feed the artificial intelligence models in the current study. Also, some of the key parameters such as gouge depth are usually determined by field-specific statistical and

probabilistic analysis prior to any ice-gouging analysis, which is considered as a sufficiently reliable method in industrial practice.

3.1.5. Goodness of fit

In order to evaluate the GEP results, correlation coefficient (*R*), Root Mean Squared Error (*RMSE*), Mean Absolute Error (*MAE*), Scatter Index (*SI*), and *BIAS* index were used as key statistical indices given by following a set of equations:

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{3-8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{3-9}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{3-10}$$

$$SI = \frac{RMSE}{\bar{O}} \tag{3-11}$$

$$BIAS = \frac{1}{n}\sum_{i=1}^{n}(P_i - O_i) \tag{3-12}$$

where O_i is the observed horizontal deformation, P_i is the simulated horizontal deformation, $\bar{O}$ is the average of observed horizontal deformations, and *n* is the number of observed horizontal deformations. The closeness of the *RMSE*, *MAE*, and *SI* indices to zero signifies the high accuracy of the numerical model. Moreover, the closeness of the correlation coefficient (*R*) to one means a high correlation of the numerical model. The best model has the lowest *BIAS* index than other models.

3.1.6. Developing Artificial Intelligence numerical models

In this section, the adopted artificial intelligence approaches, including Gene Expression Programming (GEP) and Artificial Neural Networks (ANN), are discussed. The GEP is considered an evolutionary model, whereas the ANN is classified as a classical learning algorithm. In the current study, a based-GEP model was adopted by using JEdit open-source software (JEdit. 2011).

3.1.6.1. Gene Expression Programming (GEP)

The Genetic Algorithm (GA), introduced by Holland in the 1960s, is based on the biological evolution principles to find a solution for computer systems. In GA, the solution is encoded in binary strings between 0 and 1 (Holland 1992). Moreover, the genetic programming (GP) method was presented by Cramer (Cramer 1985) and developed by Koza (1992). The GP produces computer programs in order to solve a problem by adopting the theory of natural selection developed by Darwinians (Gandomi et al. 2013). In the GP solutions, the computer programs are expressed as tree structures and then represented in a functional programming language. Thus, the solution procedure in the GA and GP approaches is the most important difference between these evolutionary methods. In contrast, Gene Expression Programming (GEP) is a strong evolutionary artificial intelligence approach, as an extension to the conventional Genetic Programming that was introduced by (Ferreira 2001). Similar to the GA and GP, the GEP owns populations and individuals that utilize individuals in a specific population and then chooses the best individuals regarding their fitness of them. The genomes in the GEP are coded as chromosomes with a constant length and represent how the chromosomes function, whilst the genes named phenotypes are presented in the forms of expression trees (ETs) (Ferreira 2001). The GEP model has its own function set, terminal set, fitness function, control parameters, and termination condition and the GEP does not borrow from other algorithms. Indeed, the main difference between the GA, GP,

and GEP models is their individuals. The GA individuals are linear strings with a fixed length, whilst the GP individuals are nonlinear entities with different shapes and sizes. The GEP individuals are encoded as linear strings with a fixed length that are expressed as nonlinear entities of various shapes and sizes (Ferreira 2001). In other words, as an evolutionary model, the GEP owns all the advantages of the GA and GP methods and overcomes the drawbacks of these algorithms. The GEP is a "phenotype/genotype" evolved system in which both phenotype and genotype are separately performed, while the GP is just a replicator (Ferreira 2001). Furthermore, simple genetic diversity is a key benefit of the GEP model, meaning that the model uses different genetic operators at various chromosome levels. The GEP have generally five important components including (I) function set (II) terminal set (III) fitness function (IV) control parameters (V) termination condition (Ferreira 2001). Furthermore, the GEP model is able to simulate complex problems by using small population sizes. An example GEP solution is illustrated in Figure 3-4.

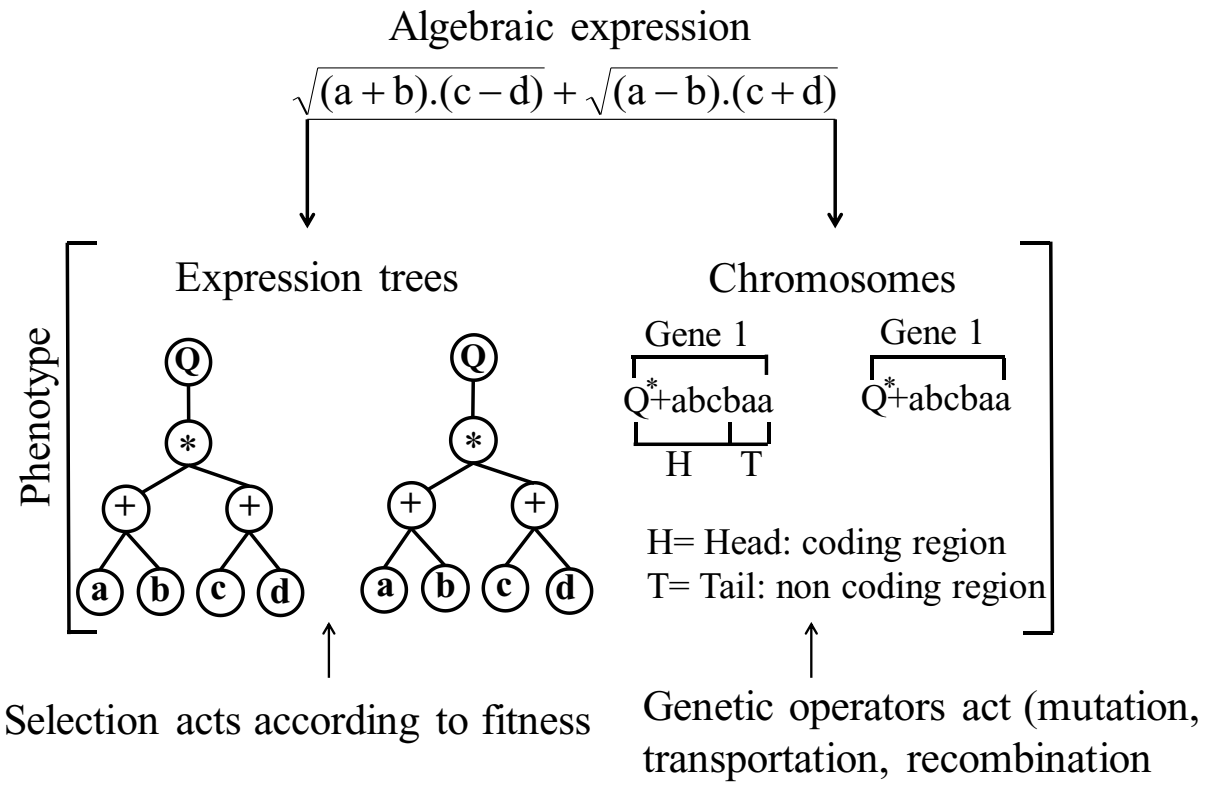


Figure 3-4. Example GEP solution

Therefore, the GEP model was adopted for modeling the subgouge soil deformations as a robust evolutionary system. In the current study, the GEP simulation was commenced with a random initial population in the model. The population included some chromosomes with constant length, where each chromosome may own some genes that have a coding region as the head and a non-coding region as the tail (Ferreira 2001). The organization presented by the genes produced valid programs. Each chromosome in the initial population was evaluated by means of a fitness function that was chosen considering the fitness value by employing the "roulette wheel selection" method (Ferreira 2001). The roulette wheel selection is considered a genetic operator in evolutionary algorithms for choosing potentially best solutions (Blickle and Thiele 1996).

This resulted in the fitter chromosomes having a high chance of producing the next generation. After the selection phase, the chromosomes were modified and reproduced using genetic operators, including mutation, transportation, inversion, and recombination. The mutation is the most important genetic operator that can occur anywhere in a specific chromosome and significantly reshape the expression trees (ETs) (Ferreira 2001). Other genetic operators do not have this capability. A trial and error procedure was employed to suggest the optimized mutation rate value. An initial value of 0.01 was considered and then increased to 0.02. Finally, the mutation rate was set to 0.014 and resulted in the best performance of the GEP. Regarding fitness by using the roulette wheel selection, individuals are chosen and then copied into a new generation. Figure 3-5 illustrates the GEP modeling flowchart, where five main steps were taken to model the horizontal subgouge deformations (d_{hy}).

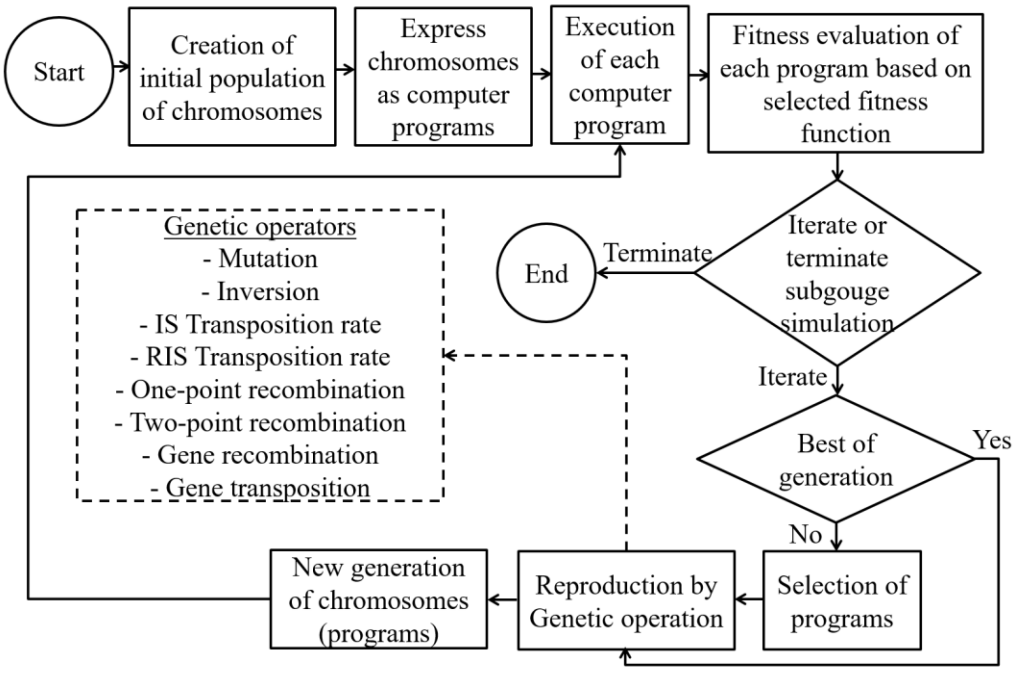


Figure 3-5. GEP flowchart to determine the subgouge soil deformation

Throughout the four main steps, first, a fitness function was determined to modify and reproduce the population. The fitness f_i related to the i^{th} program was considered as follows:

$$f_i = \sum_{j=1}^{c_t} \left(\mathrm{M} - \left| C_{(i,j)} - T_j \right| \right) \tag{3-13}$$

where, M is the selection range, $C\,(i, j)$ is the i^{th} chromosome value for the j^{th} fitness case, and T_j is the observed value for the j_{th} fitness case (Ferreira 2001). The precision becomes zero; when the accuracy $\left(\left| C_{(i,j)} - T_j \right| \right)$ is 0.01. As a result, the *fi* was obtained equal to the minimum value of itself ($fi=f_{min}=(Ct.$ M). Therefore, finding the optimum result was one of the benefits of applying the fitness function (Ferreira 2001).

Second, the function set and terminal set was selected to produce the chromosomes. The terminal set is an independent variable, which was a function of all input parameters. Regarding Eq. (3-7), the "terminal set" consisted of the $\sigma_b, V_e, \alpha, I_r, y$ as input parameters. It is worth mentioning that there is no specific method to determine the function set. Therefore, the function sets such as "+, -, ×, /, √, Ln, Exp" were employed using a trial and error approach. In other words, the accuracy of various function sets was evaluated and then the optimal one was selected. For instance, the values of correlation coefficient (*R*) and scatter index (*SI*) for the "+, -, ×, /, √, Ln, Exp" function sets were respectively obtained to be 0.907 and 0.746. The results from the calculated statistical indices for the applied function sets are arranged in Table 3-4.

Table 3-4. Results from the calculated statistical indices for the applied function sets

Function sets	*R*	*RMSE*	*MAE*	*SI*	*BIAS*
+, -,* , /	0.853	0.358	0.257	0.940	0.003
+, -,* , / , √	0.847	0.382	0.271	0.966	0.002

+, -,* , / , √, Ln	0.893	0.297	0.212	0.825	-0.074
+, -, ×, /, √, Exp	0.907	0.271	0.186	0.746	0.067
+, -, ×, /, √, Ln, Exp	0.929	0.253	0.147	0.673	-0.002
+, -, ×, /, √, Ln, Exp, Sin	0.911	0.261	0.170	0.741	0.003

Third, the chromosome architecture, including head length and the number of genes was determined by assuming an initial value of one and two for the genes and head, respectively. Next, the number of genes and the head length were respectively increased to 8 and 10 and finally, the best results were obtained when the number of genes and head length were equal to 5 and 7, and increasing the number of these parameters was found to have no considerable effect on the modeling results. Thus, the number of optimal genes and the head length were considered equal to 5 and 7 for the numerical model. Fourth, the GEP model owns four linking functions of "Division", "Subtraction", "Multiplication", and "Addition". Regarding the results of the linking functions, the Addition was identified as the superior linking function. In summary, the modeling was conducted by setting the genetic operators given in Table 3-5.

Table 3-5. Genetic operators applied in the study

	Parameters	Setting
P1	Number of generations	200000
P2	Number of chromosomes	80
P3	Number of genes	5
P4	Head size	7
P5	Linking function	Addition

P6	Mutation rate	0.014
P7	Inversion rate	0.1
P8	One-point recombination rate	0.3
P9	Two-point recombination rate	0.3
P10	Gene recombination rate	0.1
P11	Gene transportation rate	0.1
P15	Function set	+, -, ×, /, √, Ln, Exp

In this table, the number of generations, the number of chromosomes, the number of genes, the head size, the linking function, the mutation rate, and the function set were determined by trial and error, whilst the other GEP model parameters including the inversion rate, the one point recombination rate, the two-point recombination rate, gene recombination rate, and gene transportation rate were adjusted as the proposed initial values of GEP tutorial (Azimi et al. 2017; Shabanlou et al. 2018; Ferreira 2001; Karbasi and Azamathulla 2016).

3.1.6.2. Artificial Neural Network (ANN)

Owing to the flexibility of multilayer perceptron (MLP) models to simulate complex problems (Haykin 1994), Artificial Neural Network (ANN) is considered one of the most common neural network (NN) techniques. Each MLP model has some layers, including an input layer, at least a hidden layer, and an output layer. Neurons are in the perceptron layers with a number equal to the input and output variables of the problem, meaning that the number of neurons in the input layer equals five input parameters including $\sigma_b, V_e, \alpha, I_r, y$, whereas one neuron in the output layer is considered the target parameter (d_{hy}). There are no specific rules and regulations to set the number

of neurons in the hidden layer. Therefore, a trial and error method is applied to select the number of neurons in the hidden layer (Bilhan et al. 2010). In the current study, the number of neurons in the hidden layer is initially assumed equal to one and the precision of the ANN model is evaluated by increasing the number to 10. The most optimal number of the hidden layer neurons is selected equal to 5 since the accuracy of the ANN model is insignificantly changed. The number of neurons in the antecedent layer is collected using summed weighted neurons in hidden and output layers and transferred to the next layer by employing an activation function. To choose the optimized activation function, three functions including "linear", "sigmoid", and "hyperbolic tangent" were used to simulate the subgouge deformations. Subsequently, the performance of these activation functions in terms of accuracy was compared together. Lastly, the sigmoid was selected to estimate the deformations. Additionally, the sigmoid activation function has been used in different fields because of its acceptable performance (Smith 1993; Basterretxea et al. 2002; Valipour et al. 2013; Makarynskyy 2004). The applied activation functions were set as follows:

$$linear(x) = x \tag{3-14}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3-15}$$

$$\tan(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{3-16}$$

The values of weighted outputs were summed by means of the MLP throughout a procedure called "model training". The Levenberg-Marquardt (LM) algorithm was applied to train the MLP neural network, where the biases and weights were calculated using the back-propagation algorithm.

3.1.6.3. K-fold Cross-Validation

In this study, the K-fold cross-validation approach was applied to train and test the GEP models (see Figure 3.6). All observations were spilt into five sub-samples (K=5) with an equal size of 22. This resulted in a total number of 110 observations (5 × 22 = 110). First, one sub-sample was randomly selected to test the GEP models and then four remaining sub-samples were utilized to train the GEP models.

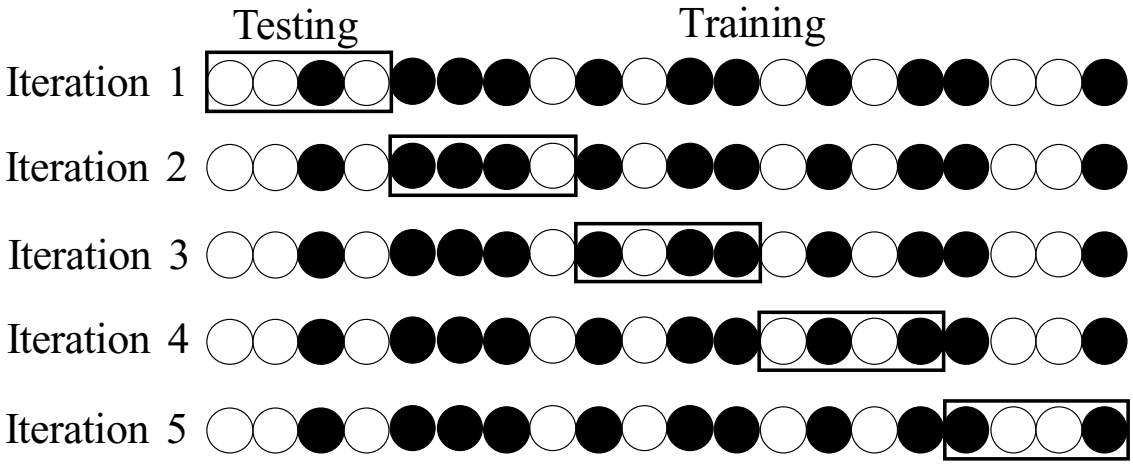


Figure 3-6. Schematic layout of the adopted 5-fold cross-validation method

Subsequently, the procedure was iterated five times for other sub-samples to ensure each sub-sample has been used at least once for training and testing of the GEP models. Eventually, the results of the five cross-validations were averaged as a single estimation for the GEP models (James et al. 2013; Azimi et al. 2021).

3.1.7. Results and discussion

3.1.7.1. Evaluation of GEP models

The computed statistical indices of GEP 1 to GEP 6 models were compared and illustrated in Figure 3-7. GEP 1 resulted in the highest accuracy (the lowest error) with a *BIAS* value of -0.049.

GEP 1 simulated the horizontal deformations of sand using all input parameters $(\sigma_b, V_e, \alpha, I_r, y)$ (see Figure 3-3). The magnitude of calculated *RMSE*, *MAE*, and *SI* indices were 0.232, 0.127, and 0.638, respectively.

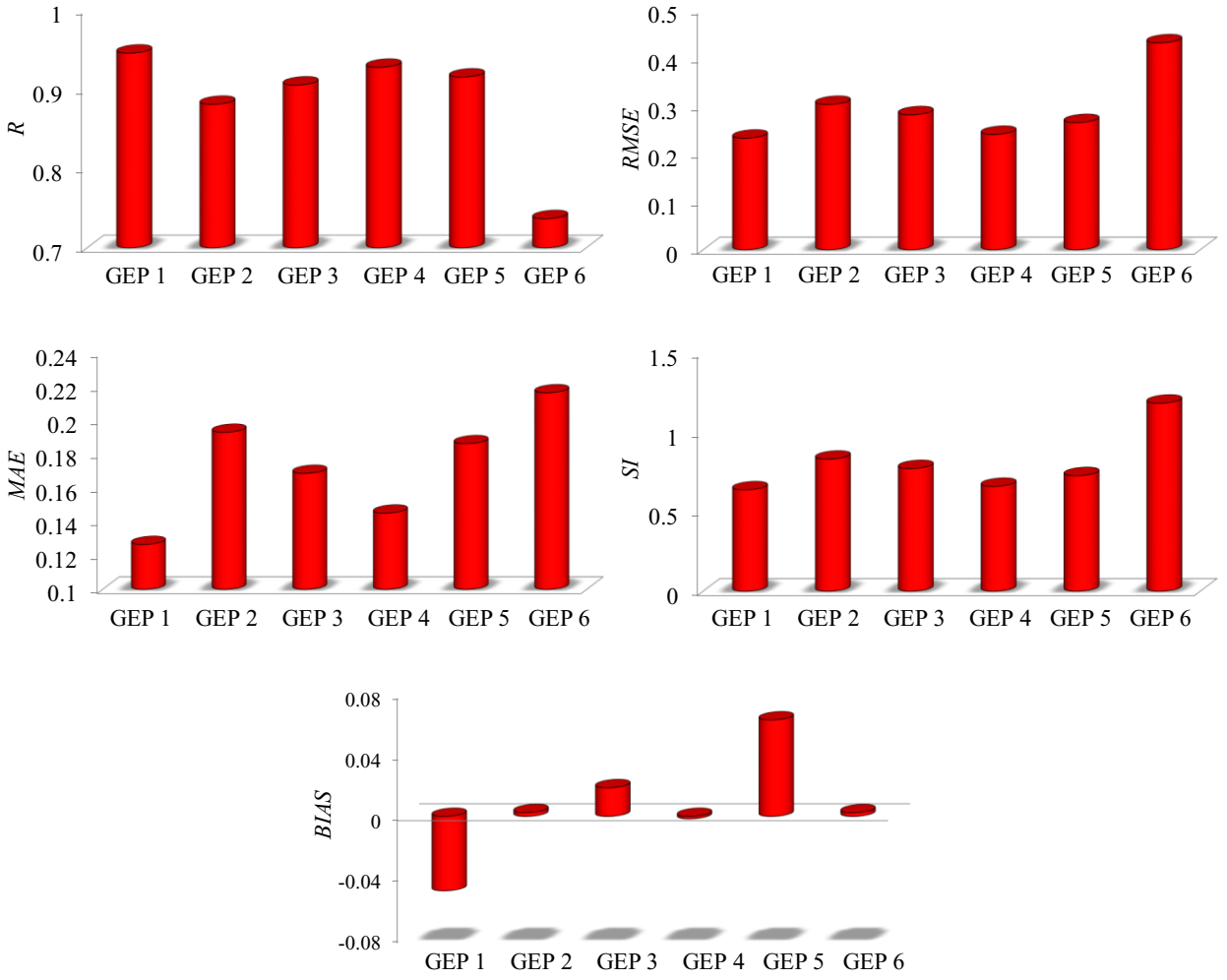


Figure 3-7.. The key statistical indices for GEP 1 to GEP 6

GEP 2 was a function of the σ_b, V_e, α, y parameters, in which the effect of the dilation index (*Ir*) was eliminated in the simulation of the target function (d_{hy}). The scatter index (*SI*) and *RMSE* values for GEP 2 were estimated at 0.831 and 0.302, respectively. For the artificial intelligence

model, the *MAE* and *BIAS* indices were respectively estimated to be 0.193 and 0.002. The attack angle (α) was the eliminated parameter in GEP 3, where the model comprised the input parameters σ_b, V_e, I_r, y. The *MAE*, *RMSE*, and *SI* values for GEP 3 were reckoned to be 0.168, 0.281, and 0.772, respectively, with a *BIAS* index of 0.019. In GEP 4, the key input parameters included σ_b, α, I_r, y, meaning that the maximum vertical extent of subgouge deformation (V_e) was ignored in the estimation of the subgouge deformations. Regarding the results obtained from GEP 4, some regular fluctuations were observed by eliminating the V_e parameter during the simulation of small subgouge deformations. In this case, the scatter index, and *RMSE* values were 0.660 and 0.240, respectively, with a computed *MAE* and *BIAS* indices of 0.145 and -0.002.

The GEP 5 was a function of the V_e, α, I_r, y input parameters and the bearing pressure (σ_b) was the eliminated parameter. The *BIAS*, *RMSE*, and *MAE* values for GEP 5 were respectively equal to 0.063, 0.265, and 0.186, with a scatter index of 0.727. Regarding GEP 3 and GEP 5, there were some similar fluctuations in the simulated subgouge deformations. Furthermore, the calculated statistical indices were approximately the same for GEP 3 and GEP 5 models. The GEP 6 had the highest error and lowest accuracy, where the obtained *SI*, *MAE*, and *BIAS* values were 1.183, 0.216, and 0.002, respectively, with the *RMSE* index of 0.431. The *R* index for GEP 2, GEP 3, and GEP 4 was estimated to be 0.881, 0.905, and 0.927, respectively. Moreover, the correlation coefficient for GEP 5 and GEP 6 was about 0.915 and 0.737, respectively

The subgouge deformations measured by C-CORE (1995) varied in a range of 3.53m (upper limit of the dataset) to 0.0 (lower limit of the dataset), whereas these values for Hynes (1996) and Yang (2009) datasets ranged from 0.95m to zero and 0.262m to zero, respectively. This means that the upper limit of data reported by C-CORE (1995) has shown a meaningful difference with Hynes's (1996) and Yang's (2009) values, signifying that these three datasets are pretty heterogeneous.

Even though the numerical model attempted to estimate the target function with its highest performance in a wide range of experimental values, some discrepancies occurred at the peak points. In other words, the GEP models underestimated these peak points.

For GEP 6, the effect of soil depth (y) was eliminated, and only the parameters $\sigma_b, V_e, \alpha, I_r$ were kept to model the target function, denoting that GEP 6 has four input parameters to estimate the horizontal ice-induced deformations. Thus, the soil depth was a significant parameter in simulating the horizontal subgouge sand deformations. The results of observed and simulated target function distribution in GEP 1 to GEP 6 versus the experiment numbers are shown in Figure 3-8.

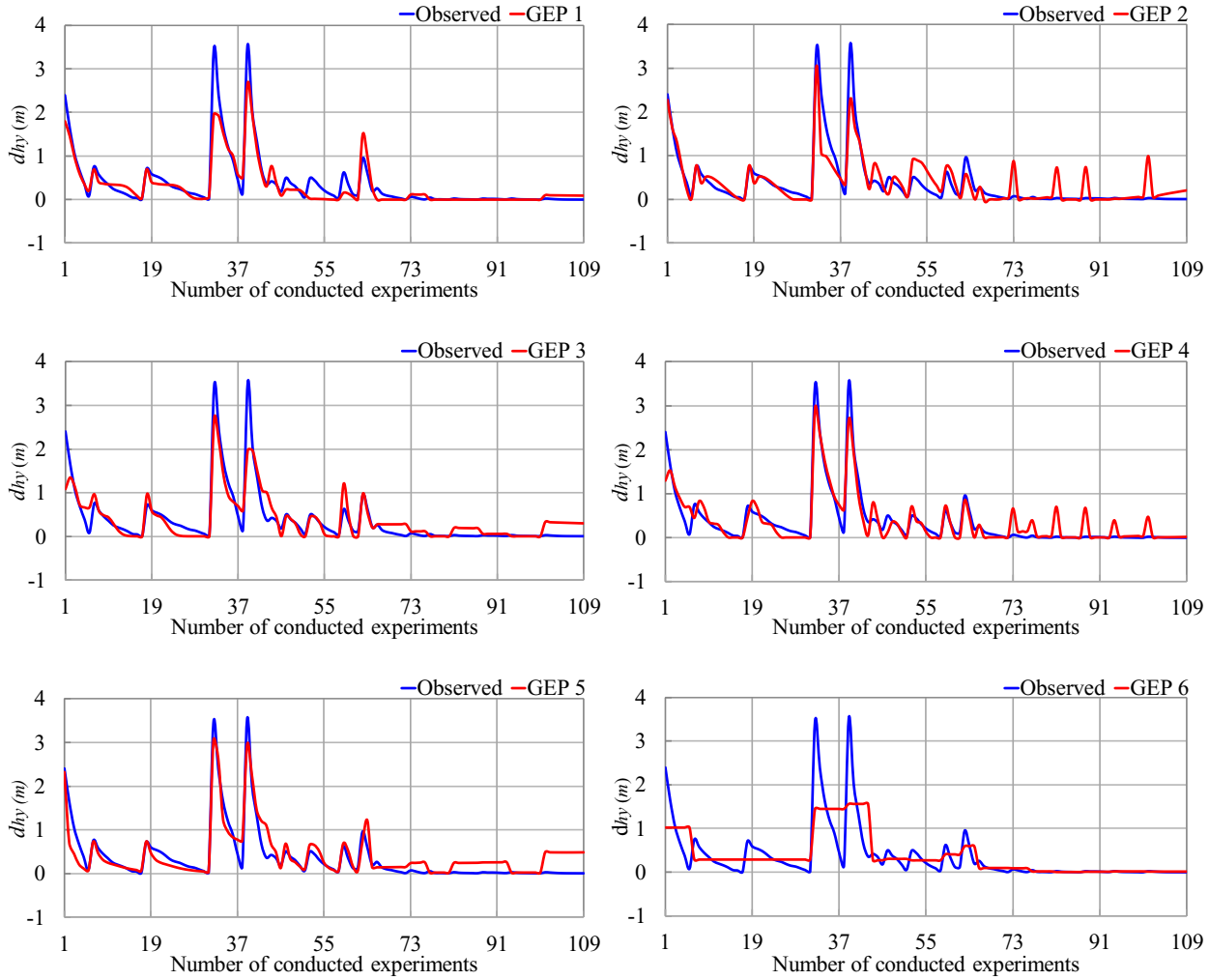
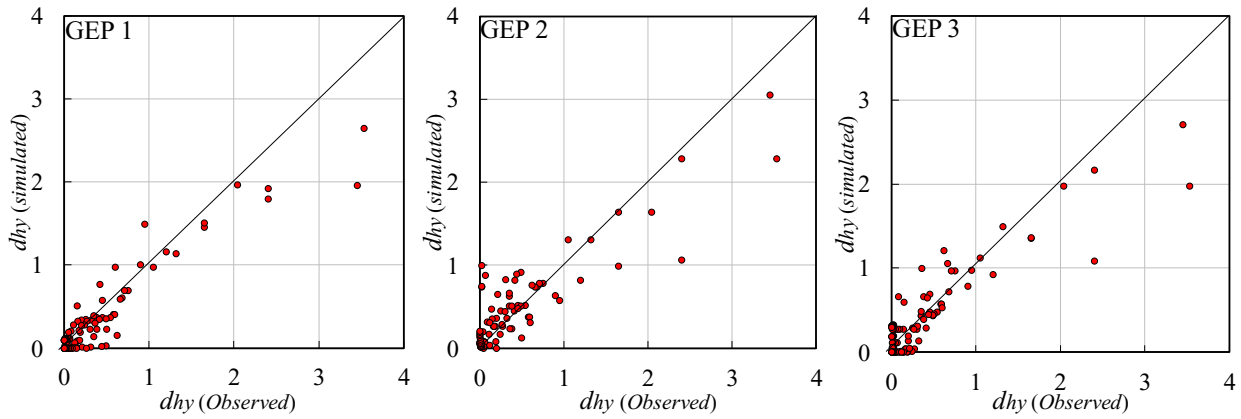


Figure 3-8. Comparison of the observed dhy with the simulated results of GEP 1 to GEP 6

Some oscillatory patterns were observed in GEP 3, GEP 5, and GEP 6. The sensitivity analysis showed that the accuracy of simulations was decreased by eliminating the soil depth (y), as the most significant input parameter. This resulted in a more severe fluctuation in the results of the GEP 6. In GEP 3 and GEP 5, these fluctuations were observed only in small subgouge deformations. Also, having a similar simulation condition in all GEP models, the level of effectiveness for each input parameter affected the results obtained from the GEP models. For instance, the significance of the eliminated inputs (α and σ_b) for GEP 3 and GEP 5 was not as much as the importance of the soil depth (y) removed from GEP 6. This means that some models had better performance (GEP 1 and GEP 4 models), while others had some fluctuations during the simulation procedure (GEP 3, GEP 5, and GEP 6). Furthermore, the scatter plots for GEP 1 to GEP 6 were depicted in Figure 3-9. Based on the numerical modeling, GEP 1 has the highest rate of correlation with a correlation coefficient (R) of 0.945.



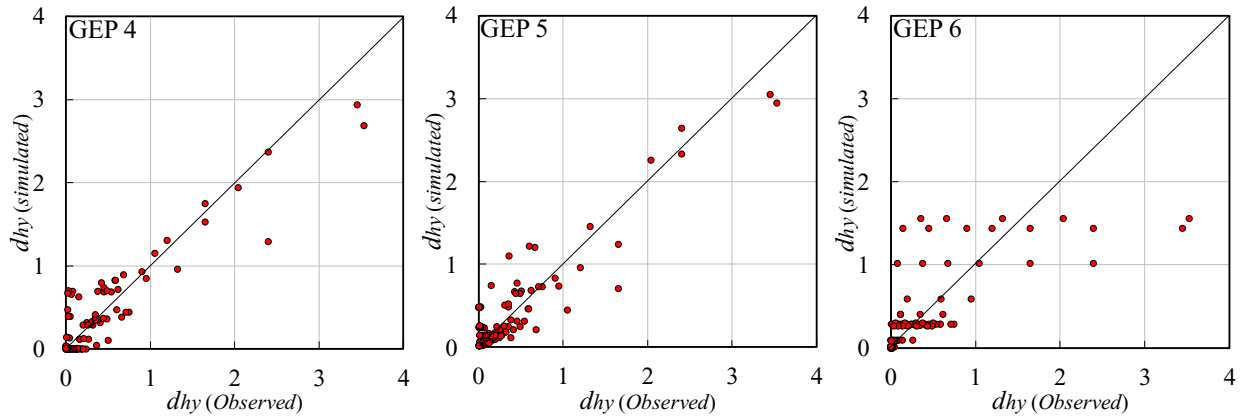


Figure 3-9. Scatter plots for GEP 1 to GEP 6 models

The error distribution of GEP 1 to GEP 6 models is shown in Figure 3-10 in which the effects of removing each input parameter on the error distribution of the GEP models were examined. For instance, about 5% of the results from GEP 1 had an error of less than 4%, while 11% of the estimated horizontal subgouge deformations by the model had an error between 10% and 20%. In contrast, the error distribution for GEP 2 showed that roughly 10% of the modeled deformations had an error of less than 8% though nearly 18% of results from the model had an error of less than 16%. Besides, virtually 18% of results from GEP 2 had an error of less than 16%, whereas this value for GEP 1 was about 15%. Furthermore, approximately 17% of results from GEP 3 produced an error of less than 12%, indicating that the elimination of the attack angle (α) resulted in decreasing accuracy of the model. For GEP 4 in which the effect of V_e was removed, the error distribution indicates that about 22% of results from having an error of less than 24%, whilst this value for GEP 1 was nearly 20%. Likewise, almost 26% of subgouge deformations estimated by GEP 5 had an error of less than 24% with a corresponding value of about 20% for GEP 1. In addition, roughly 19% of estimated sand deformations using GEP 6 had an error of less than 24%, and this value for GEP 1 was about 20%.

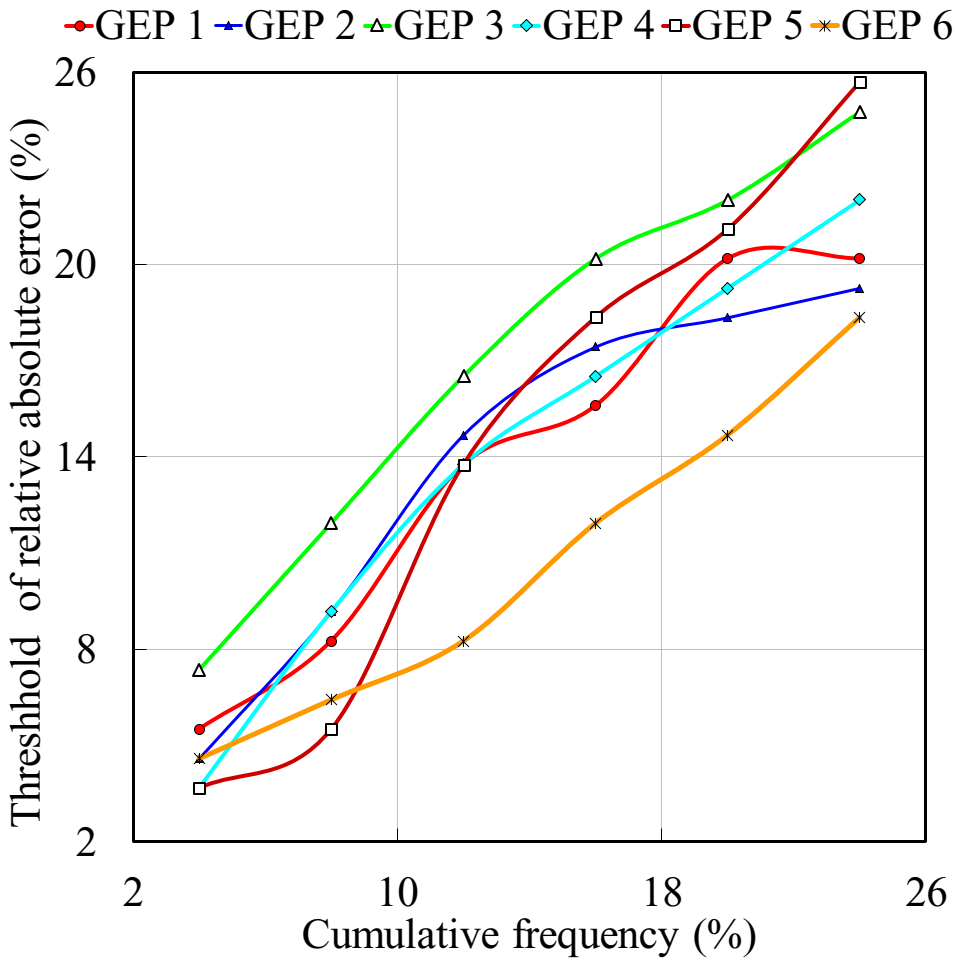


Figure 3-10. Error distribution diagram for GEP models

To further evaluate the GEP models, the discrepancy ratio (*DR*) of the models was defined as follows and presented in Figure 3-11:

$$DR = \frac{d_{hy(simulated)}}{d_{hy(observed)}} \tag{3-17}$$

In general, the magnitude of *DR* approaching the unity denoted the higher performance of the model. Although the maximum and minimum discrepancy ratios are not shown in Figure 3-11, almost all calculated *DR* near the unity can be seen from the figure otherwise, the figure is not able to illustrate the behavior of the *DR* around the unity. The average discrepancy ratio (DR_{ave}) for

GEP 1 was estimated at 9.094, while the value for GEP 2 and GEP 3 was approximately 21.186 and 30.865, respectively. Moreover, the computed value of DR_{ave} for GEP 4 is 4.714 with the corresponding value of 47.763 for the GEP 5 model. Besides, the average discrepancy ratio for GEP 6 was 3.019.

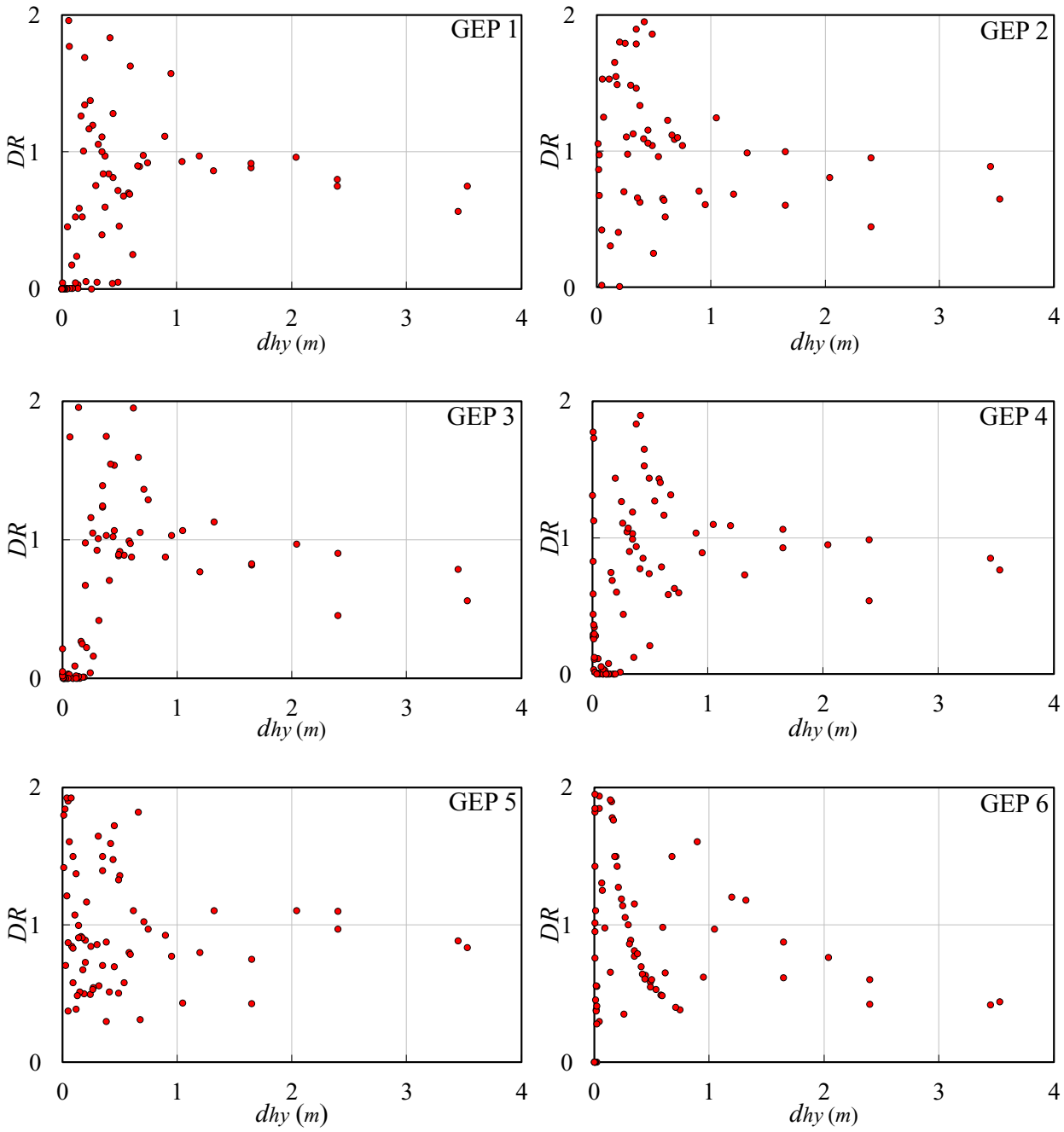


Figure 3-11. Discrepancy ratio (*DR*) for GEP models

Regarding the conducted sensitivity analysis, GEP 1 was identified as the most accurate GEP model to surmise the horizontal deformations of sand. The model estimated the deformations using all input parameters including the bearing pressure (σ_b), the maximum vertical extent of subgouge deformation (V_e), the attack angle (α), the dilation index (I_r), and soil depth (y). Additionally, GEP 4 as a function of σ_b, α, I_r, y was the second accurate model, and GEP 5 was detected as the third accurate GEP model. Among all GEP models, GEP 3 had fourth place in terms of accuracy. Moreover, GEP 2 was considered as fifth most accurate GEP model which estimates the target function using V_e, α, I_r, y input parameters. Ultimately, GEP 6 had the lowest correlation with experimental values. The sensitivity analysis indicated that the soil depth (y) and the dilation index (I_r) are the most significant input parameters, meaning that the elimination of the y and I_r resulted in declining the performance of the GEP model. In addition, the attack angle (α) was identified as the third important input parameter even though the bearing pressure (σ_b) had the fourth place in terms of its impact on the modeling of deformations. Finally, the maximum vertical extent of subgouge deformation (V_e) was detected as an input parameter with the lowest influence.

3.1.7.2. Uncertainty Analysis (UA)

An Uncertainty Analysis (UA) was conducted to evaluate the GEP model's performance by quantification of the uncertainties in the input variables. This methodology has been successfully applied in the literature to describe the error estimated by the GEP models, for example (Karbasi and Azamathulla et al. 2016; Azimi et al. 2021) performed UA for their GEP models. In the current study, the estimated error by the GEP models (e_j) was computed as the difference between the

estimated (P_j) and observed (O_j) subgouge soil deformations $(e_j = P_j - O_j)$. The mean value of the estimated error $(\bar{e})$ and the standard deviation (S_e) were calculated as follows:

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{3-18}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{3-19}$$

The negative sign of $\bar{e}$ signifies the underestimated performance of the GEP model, whereas the positive sign indicates the overestimated performance of the numerical model. Using the $\bar{e}$ and S_e parameters, the confidence bound was defined around an estimated error by adopting the "Wilson score approach" without the continuity correction. The Wilson score interval, developed by Wilson (1927), is an improvement over the normal distribution interval in which an asymmetric normal distribution is used to improve the confidence interval bound (Wallis 2013). This method is applied in the uncertainty analysis of the developed GEP models. A ±1.96Se resulted in a 95% confidence bound, which is represented by a "95% prediction error interval" (95%PEI). Table 3-6 shows the parameters of the UA of GEP models, where the "width of uncertainty bound" is shown by WUB.

Table 3-6. Uncertainty analysis parameters for GEP models

Model	$\bar{e}$	S_e	WUB	95% PEI
GEP 1	0.049	0.228	±0.044	0.005 to 0.092
GEP 2	-0.003	0.304	±0.058	-0.060 to 0.055
GEP 3	-0.019	0.282	±0.054	-0.072 to 0.035
GEP 4	0.002	0.241	±0.046	-0.044 to 0.047
GEP 5	-0.063	0.258	±0.049	-0.112 to -0.014

GEP 6	-0.002	0.433	±0.083	-0.085 to 0.080

Regarding the UA, the 95%PEI for GEP 1 was between 0.005 and 0.092. Also, the width of the uncertainty bound (WUB) for GEP 1 was equal to -0.044. Among all GEP models, GEP 1 had the lowest standard deviation of the estimated error with a corresponding value of 0.228. Moreover, WUB for GEP 2, GEP 3, and GEP 4 was calculated as -0.058, -0.054, and -0.046. For GEP 5, the S_e and WUB values were estimated at 0.258 and -0.049, respectively. Furthermore, the 95%PEI for GEP 6 model was between -0.085 and 0.080. Moreover, the mean value of the estimated error ($\bar{e}$) for GEP 1 and GEP 4 had a positive sign, whereas the parameter for GEP 2, GEP 3, GEP 5, and GEP 6 had a negative sign. Although GEP 4 model estimated the target function (d_{hy}) with an overestimated performance, others including GEP 2, GEP 3, GEP 5, and GEP 6 had an underestimated performance. Finally, the superior model (GEP 1) showed an overestimated performance with reasonable accuracy in order to simulate the ice-induced subgouge deformations.

3.1.7.3. Comparison of the superior model (GEP 1) with ANN

The results from the superior model (GEP 1) were compared with a classical artificial intelligence (AI) approach so-called Artificial Neural Network (ANN). The ANN can model nonlinear and complex problems like subgouge deformation issues with acceptable accuracy and computational time. Besides, ANN does not need knowledge of input and output parameters. Therefore, to assess the performance of the evolutionary system (GEP) and a traditional AI approach, the results from GEP 1 and ANN models were compared. Table 3-7 shows the calculated statistical indices for GEP 1 (the superior model) against the ANN model.

Table 3-7. Statistical indices for GEP 1 and ANN models

Model	*R*	*RMSE*	*SI*	*BIAS*	*MAE*
GEP 1	0.945	0.232	0.638	-0.049	0.127
ANN	0.899	0.280	0.769	0.014	0.220

The estimated correlation coefficient (*R*), *RMSE*, and scatter index (*SI*) for the ANN model were respectively estimated as 0.945, 0.232, and 0.638. The *MAE* and *BIAS* indices for ANN were approximated to be 0.127 and -0.049. The results showed that GEP 1 had a better performance in modeling the subgouge sand deformation compared to ANN. The distribution of observed and simulated target function by GEP 1 and ANN versus the number of experiments is depicted in Figure 3-12.

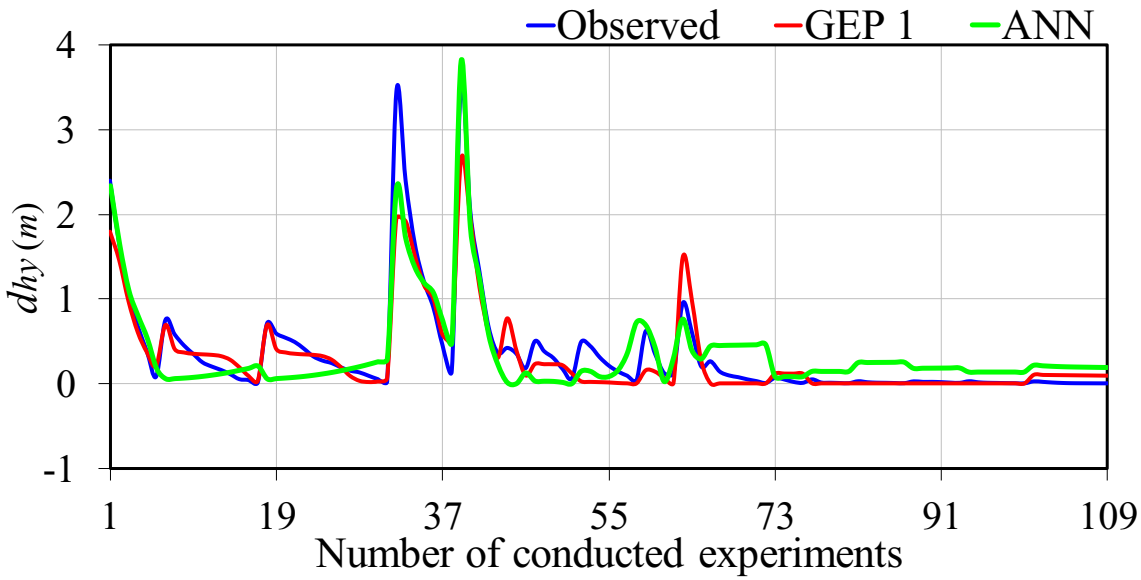


Figure 3-12. Comparison GEP 1 with ANN

The results showed that the GEP model had a better performance for most ice-induced subgouge deformations, whereas both artificial intelligence approaches (GEP and ANN) for large subgouge deformation values had a similar pattern. Although the GEP models underestimated the subgouge deformations at some peak points, the ANN model had an overestimated performance at such values. There were some peak points at which the values of simulated deformations had discrepancies with observed values because of the remarkable difference between the upper limit of adopted datasets.

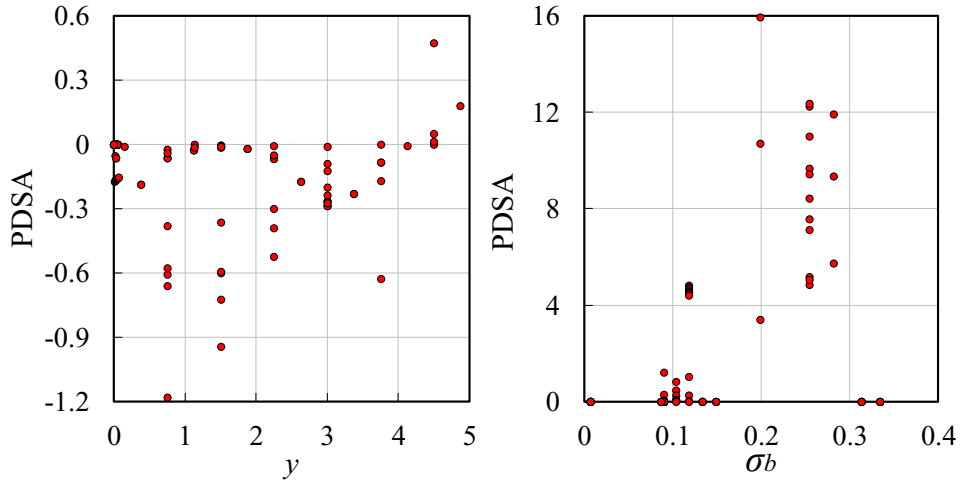
3.1.7.4. Superior GEP model

Regarding the conducted sensitivity analysis and uncertainty analysis, GEP 1 in terms of accuracy and correlation was detected as the superior GEP model. This model simulated the horizontal subgouge sand deformations $(\boldsymbol{d_{hy}})$ by means of all input parameters including $\sigma_b, V_e, \alpha, I_r, y$. Therefore, to estimate the $\boldsymbol{d_{hy}}$ using GEP 1 as the best model, a GEP-based produced equation is presented as follows:

$$d_{hy} = 1/(1 + exp(-((((\sigma_b - y) - y) + 9.339) + ((\alpha - 9.448) \times Ln(\sigma_b))))) + 1/(1 + exp(-((7.008 + ((7.008 - y) - Ir)) + ((\alpha - 6.984) \times Ln(\sigma_b))))) + sqrt(exp(-((Ve + (y - exp(((exp(-((Ir - 2.659)^2))) \times Ir))))^2))) + exp(-((sqrt(y) + (Ve - exp((exp(-((-0.476 + 1.501^2)) + Ir))))^2))exp(-(((Ir - ((exp(-((y - 4.186)^2)) + Ve) + Ve)) + 7.923)^2)) \quad (3\text{-}20)$$

3.1.7.5. Partial Derivate Sensitivity Analysis (PDSA)

A Partial Derivate Sensitivity Analysis (PDSA) was carried out to further assess the effects of input parameters $\sigma_b, V_e, \alpha, I_r$ and y as dependent parameters on the target function (d_{hy} as an independent parameter). The PDSA enabled identifying how the dependent parameters affect the independent ones. The partial derivative of the independent parameter relative to the input parameters $\partial(d_{hy})/(\partial x_i)$ was employed to evaluate the sensitivity of GEP 1 to $x_i(\sigma_b, V_e, \alpha, I_r \text{ and } y)$. A positive PDSA signifies an increasing effect on the d_{hy}, while a negative PDSA indicates a decreasing influence on the subgouge sand deformations. Figure 3-13 shows the PDSA results of the GEP 1 model, where most of the PDSA results for the soil depth parameter (y) are negative, meaning that the subgouge deformation decrements with increasing the soil depth (y).



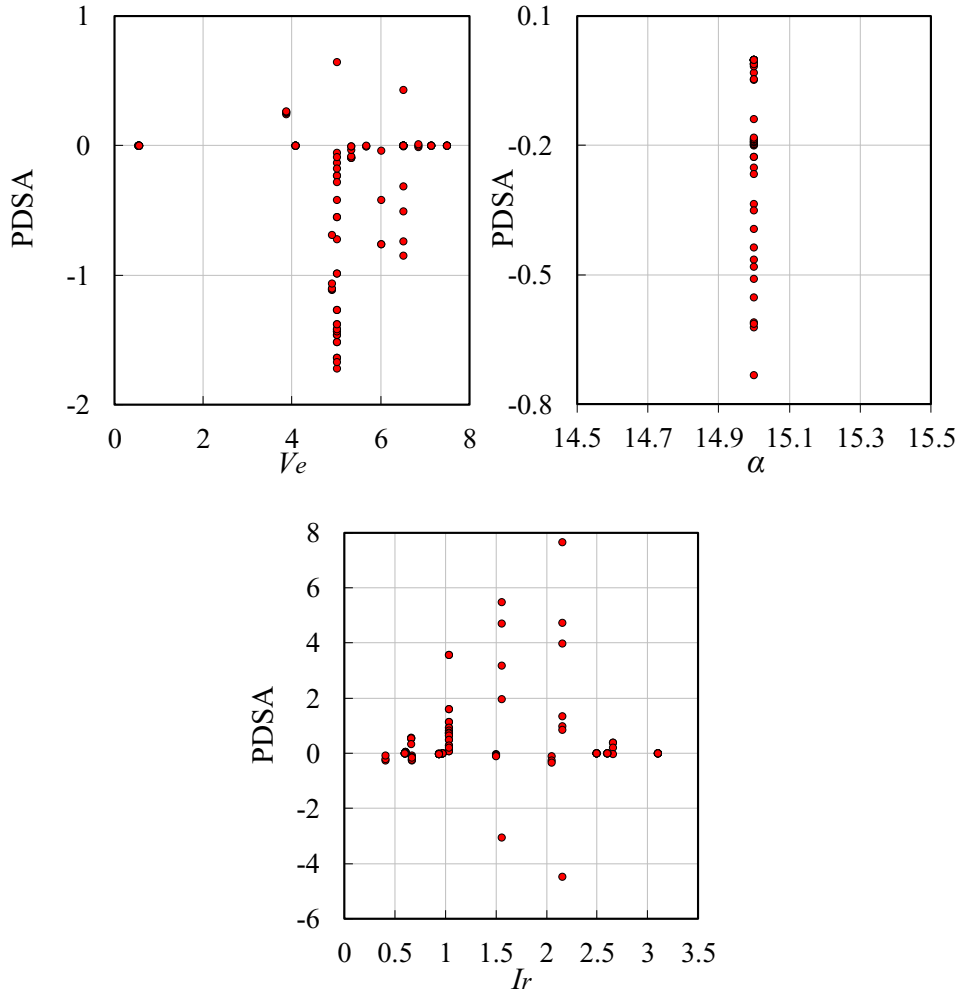


Figure 3-13. PDSA results for GEP 1

Moreover, almost all PDSA results for bearing pressure parameter (σ_b) were positive, denoting that the d_{hy} grows when the σ_b increases. However, the overwhelming majority of the results from the PDSA for V_e and α parameters had a negative sign. For the *Ir* input parameter, all PDSA results were positive. In other words, the subgouge soil deformations were increased with increasing the dilation index.

3.1.7.6. Comparison with C-CORE (2009a)

In this section, the results of GEP 1 are compared with the equation proposed by C-CORE (2009a) (Eq. 3-1), the comparison is arranged in Table 1-8. Regarding the table, the values of the scatter index and correlation coefficient for the C-CORE's (2009a) model were respectively estimated to be 10.581 and 0.638. Moreover, the *MAE*, *BIAS*, and *RMSE* statistical indices for Eq. (3-1) were computed to be 3.121, 3.121, and 3.781, respectively. Thus, GEP 1 model simulated the horizontal ice-induced sand deformations with higher accuracy and correlation compared to the proposed model by C-CORE (2009a).

Table 3-8. Comparison between results of GEP 1 with the C-CORE (2009a) model

Model	*R*	*RMSE*	*SI*	*BIAS*	*MAE*
GEP 1	0.945	0.232	0.638	-0.049	0.127
C-CORE (2009a)	0.146	3.781	10.581	3.121	3.121

A comprehensive numerical study based on artificial intelligence (AI) modeling was carried out for the first time so as to model ice-induced subgouge deformations. To end this, two AI approaches including an evolutionary method (GEP), and a classical learning algorithm (ANN) applied to introduce the best combination of the input parameters as the superior AI model. Furthermore, the most influencing parameters on subgouge deformations were detected by sensitizing the input parameters. Ultimately, an Uncertainty Analysis (UA) was done in order to identify the performance of the best model, and as well as a Partial Derivative Sensitivity Analysis (PDSA) was conducted to assess the impact of the input parameters.

3.1.8. Conclusion

The ice-induced subgouge sand deformation was estimated using two artificial intelligence approaches, including Gene Expression Programming (GEP) as an evolutionary approach and Artificial Neural Network (ANN) as a classical learning machine. Firstly, the main governing parameters were identified and then six GEP models were developed. In addition, the K-fold cross-validation (K=5) method was applied for the training and testing of the numerical models. Subsequently, the GEP models were analyzed, and then the best GEP model with the most effective input parameters was introduced. Moreover, an Uncertainty Analysis (UA) was conducted to evaluate the performance of the GEP models. Then, the results of the best GEP model were compared with the ANN technique. Finally, a Partial Derivative Sensitivity Analysis (PDSA) was carried out. The key results of the study are as follows:

- The GEP could simulate the subgouge soil deformations with a good level of accuracy. For instance, the computed values of *R*, *RMSE*, and *BIAS* for the best GEP model were 0.945, 0.232, and -0.049, respectively. Moreover, about 15% of the results from the best GEP model had an error of less than 16% and an average discrepancy (DR_{ave}) ratio of 9.094.
- The superior model (GEP 1) estimated the subgouge sand deformations using a combination of all input parameters including the bearing pressure (σ_b), the maximum vertical extent of subgouge deformation (V_e), the attack angle (α), the dilation index (I_r), and soil depth (y).
- Regarding the performed sensitivity analysis, the soil depth (y), the dilation index (I_r), the attack angle (α), the bearing pressure (σ_b), and the maximum vertical extent of subgouge deformation (V_e) were respectively prioritized as the most significant input

parameters to estimate the horizontal subgouge sand deformations by using the GEP model.

- The GEP models showed a better performance in comparison with the conventional ANN approach. The results obtained from the GEP were remarkably precise when compared with experimental test measurements.
- Uncertainty analysis (UA) indicated that GEP 1 had an overestimated performance in order to simulate the subgouge sand deformations.
- According to the PDSA, the target function (d_{hy}) decreased with increasing the soil depth (y), whilst the subgouge sand deformations grew with increasing the dilation index (I_r).

The results obtained in this study can not only facilitate proposing new analytical or empirical solutions for the prediction of the subgouge soil deformation but also be beneficial to the planning of costly experimental studies and numerical simulations and mitigate the expenses of future investigations. In the future, the study can be further extended to the clay seabed along with the presence of a buried pipeline and examine other AI techniques such as optimization algorithms and machine learning.

Acknowledgments

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Alba, J.L., 2015. Ice Scour and Gouging Effects with Respect to Pipeline and Wellhead Placement and Design. Bureau of Safety and Environmental Enforcement (BSEE).

Arnau, S., Ivanović, A., 2019. Rate effects during ice scour in sand. Cold Regions Science and Technology, 158: 182-194.

Azamathulla, H.M., Cuan, Y.C., Ghani, A.A., Chang, C.K., 2013. Suspended sediment load prediction of river systems: GEP approach. Arabian Journal of Geosciences, 6: 3469-3480.

Azimi, H., Bonakdari, H., Ebtehaj, I., 2017. A highly efficient gene expression programming model for predicting the discharge coefficient in a side weir along a trapezoidal canal. Irrigation and drainage, 66: 655-666.

Azimi, H., Bonakdari, H. and Ebtehaj, I., 2021. Gene expression programming-based approach for predicting the roller length of a hydraulic jump on a rough bed. ISH Journal of Hydraulic Engineering, 27(sup1): 77-87.

Barker, A., Timco, G., 2003. Laboratory experiments of ice scour processes: buoyant ice model. Cold regions science and technology, 36: 103-114.

Barrette, P.D., Sudom, D., 2014. Ice-soil-pipeline interaction during seabed gouging in physical tests: Database analysis and outstanding issues. Offshore Technology Conference. OTC-24604-MS.

Basterretxea, K., Tarela, J. M., Del Campo, I., 2002. Digital design of sigmoid approximator for artificial neural networks. Electronics Letters, 38(1): 35-37.

Been, K., 1990. Mechanisms of failure and soil deformation during scouring. In Proceedings of the workshop on ice scouring and the design of offshore pipelines. St John's, NL, Canada: C-CORE, pp. 179-191.

Blickle, T., Thiele, L., 1996. A comparison of selection schemes used in evolutionary algorithms. Evolutionary Computation, 4(4): 361-394.

Bilhan, O., Emiroglu, M.E., Kisi, O., 2010. Application of two different neural network techniques to lateral outflow over rectangular side weirs located on a straight channel. Advanced Engineering Software, 41: 831-837.

Bruce, J., Piercey, G., Macneill, A., Phillips, R., Derradji, A., 2012. Physical testing method to study ice keel strength limits during a gouging process. Arctic Technology Conference, OTC-23792-MS.

C-CORE., 1995 Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2.

C-CORE., 2008. PIRAM - Review of Subgouge Defornation. C-CORE rep011 R-07-045-491 v2 December 2008.

C-CORE., 2009a. PIRAM - Gouge Morphology Study, C-CORE Report R -09-013-490 v1, June 2009.

C-CORE., 2009b. PIRAM - Draft Final Report, C-CORE Report R -09-019-490 v1, June 2009.

Chari, T.R., 1975. Some geotechnical aspects of iceberg grounding, Doctoral dissertation, Memorial University of Newfoundland, St. John’s, Canada.

Cramer, N.L., 1985. A representation for the adaptive generation of simple sequential programs. In Proceedings of the first international conference on genetic algorithms, pp. 183-187.

Ebtehaj, I., Bonakdari, H., Zaji, A.H., 2018. A new hybrid decision tree method based on two artificial neural networks for predicting sediment transport in clean pipes. Alexandria engineering journal, 57: 1783-1795.

Emmerson, C., Lahn, G., 2012.. Arctic opening: Opportunity and risk in the high north.

Ferreira, C., 2001. Gene Expression Programming: A New Adaptive Algorithm for Solving Problems. Complex. Syst. 13(2), 87-129.

Gandomi, A.H., Yun, G.J., Alavi, A.H., 2013. An evolutionary approach for modeling of shear strength of RC deep beams. Materials and Structures, 46(12): 2109-2119.

Green, H., Chari, T., 1981. Sediment compaction below iceberg furrows. Oceans, pp. 223-227.

Guven, A., Kisi, O., 2013. Monthly pan evaporation modeling using linear genetic programming. Journal of Hydrology, 503: 178-185.

Haghiabi, A.H., 2016. Prediction of river pipeline scour depth using multivariate adaptive regression splines. Journal of Pipeline Systems Engineering and Practice, 8: 04016015.

Harrison, W.L., 1973. Soil failure under inclined loads-II. Journal of Terramechanics, 10: 11-50.

Haykin, S., 1994. Neural networks: a comprehensive foundation. Prentice Hall PTR.

Holland, J. H., 1992. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press.

Hynes, F., 1996. Centrifuge modelling of ice scour in sand. Doctoral dissertation, Memorial University of Newfoundland, Canada.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An introduction to statistical learning, New York: springer, Vol: 112, pp. 18.

JEdit. 2011. http://sourceforge.net/projects/jedit/.

Karbasi, M., Azamathulla, H.M., 2016. GEP to predict characteristics of a hydraulic jump over a rough bed. KSCE Journal of Civil Engineering, 20: 3006-3011.

Kaydani, H., Najafzadeh, M., Mohebbi, A., 2014. Wellhead choke performance in oil well pipeline systems based on genetic programming. Journal of Pipeline Systems Engineering and Practice, 5: 06014001.

Kioka, S. D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event By A Neural-Network. In The Thirteenth International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. In The Fourteenth International Offshore and Polar Engineering Conference, ISOPE-I-04-107.

Koza, J.R., 1992. Genetic programming: on the programming of computers by means of natural selection. MIT press, Vol. 1.

Longva, O., Bakkejord, K.J., 1990. Iceberg deformation and erosion in soft sediments, southeast Norway. Marine Geology, 92: 87-104.

Makarynskyy, O., 2004. Improving wave predictions with artificial neural networks. Ocean Engineering, 31(5-6): 709-724.

Mitchell, J.K., Soga, K., 2005. Fundamentals of soil behavior. Hoboken, NJ: John Wiley & Sons, Vol. 3.

Moradi, F., Bonakdari, H., Kisi, O., Ebtehaj, I., Shiri, J., Gharabaghi, B., 2019. Abutment scour depth modeling using neuro-fuzzy-embedded techniques. Marine Georesources & Geotechnology, 37(2): 190-200.

Najafzadeh, M., Shiri, J., Rezaie-Balf, M., 2018. New expression-based models to estimate scour depth at clear water conditions in rectangular channels. Marine Georesources & Geotechnology, 36(2): 227-235.

Najafzadeh, M., Shiri, J., Sadeghi, G., Ghaemi, A., 2018. Prediction of the friction factor in pipes using model tree. ISH Journal of Hydraulic Engineering, 24(1): 9-15.

Nematzadeh, A., Shiri, H., 2019. The Influence of Non-linear Stress-Strain Behavior of Dense Sand on Seabed Response to Ice-gouging. Cold Regions Science and Technology. 170: 102929.

Phillips, R., Clark, J.I., Kenny, S., 2005. PRISE studies on gouge forces and subgouge deformations. Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions, pp. 75-84.

Phillips, R., Barrett, J., 2012. PIRAM: Pipeline response to ice-gouging. Arctic Technology Conference, OTC-23790-MS.

Prasad, K.S.R., 1985. Analytical and experimental modellng of iceberg scours. Doctoral dissertation, Memorial University of Newfoundland, Canada.

Rajaee, T., 2011. Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. Science of The Total Environment, 409: 2917-2928.

Shabanlou, S., Azimi, H., Ebtehaj, I., Bonakdari, H., 2018. Determining the scour dimensions around submerged vanes in a 180 bend with the gene expression programming technique. Journal of Marine Science and Application, 17: 233-240.

Shaghaghi, S., Bonakdari, H., Gholami, A., Kisi, O., Shiri, J., Binns, A. D., Gharabaghi, B., 2018. Stable alluvial channel design using evolutionary neural networks. Journal of hydrology, 566: 770-782.

Smith, M., 1993. Neural networks for statistical modeling. Thomson Learning.

Valipour, M., Banihabib, M.E., & Behbahani, S.M.R., 2013. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. Journal of hydrology, 476: 433-441.

Wallis, S., 2013. Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. Journal of Quantitative Linguistics, 20(3): 178-208.

Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association, 22(158): 209-212.

Woodworth-Lynas, C.M.T., Phillips, R., Clark, J.I., Hynes, F., Xiao, X., 1998. Verification of centrifuge model results against field data: results from the Pressure Ridge Ice Scour Experiment (PRISE). 13th International Symposium on Okhotsk Sea and Sea Ice, pp. 123-138.

Yang, W., 2009. Physical modeling of subgouge deformations in sand. Doctoral dissertation, Memorial University of Newfoundland, Canada.

Yavari, S., Maroufpoor, S., Shiri, J., 2018. Modeling soil erosion by data-driven methods using limited input variables. Hydrology Research, 49(5): 1349-1362.

Section 2

Modeling Subgouge Sand Deformations by Using Multi-Layer Perceptron Neural Network

This section was published at the 31st International Ocean and Polar Engineering Conference (ISOPE), 2021, Rhodes, Greece

Abstract

In the Arctic shallow waters, marine pipelines are threatened by traveling icebergs where the seabed may be gouged by these moving masses during warmer months. Estimation of the subgouge soil response is considered a serious design factor for the subsea infrastructures since minimizing the required burial depth for physical protection is quite crucial for the project budget. In this paper, the capability of the Multi-Layer Perceptron Neural Network (MLPNN) is utilized to simulate ice-induced sand deformations. By conducting a sensitivity analysis, the best MLPNN models and the most significant input parameters are identified.

Keywords: Ice-seabed interaction; sandy seabed; Multi-Layer Perceptron Neural Network (MLPNN); Simulation; Sensitivity analysis.

3.1.9. Introduction

The Arctic area contains a huge amount of hydrocarbon deposits such as crude oil and natural gases that increased energy demand is the cause of exploration in the Arctic regions. Subsea pipelines are widely used to transfer hydrocarbon and other exploration and production-related contents between the onshore and offshore facilities (Alba 2015). Marine pipelines are threatened by the ice-gouging, and pressure ridges attack crossing the pipeline route in the Arctic shallow waters. Subsea trenching and backfilling are commonly used to bury the pipeline for physical protection against the ice scour. The schematic layout of an ice-seabed interaction process is illustrated in Figure 3-14. As shown, a hyperbolic curve is produced just beneath the ice keel bottom where the maximum soil displacement occurs at the soil surface.

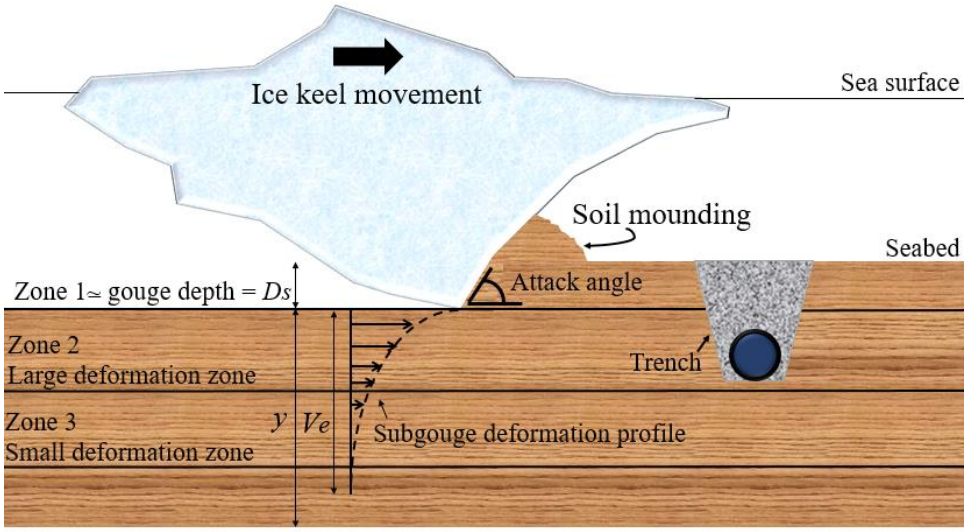


Figure 3-14. Schematic layout of an ice-seabed interaction process

Identifying the maximum deformations for safe and cost-effective protection of the pipeline is a challenging problem. Costly experimental and long-running numerical simulation is mandatory for accurate modeling of the subgouge soil deformation and consequently, the pipe response. For instance, a joint industry and government-sponsored research program entitled the Pressure Ridge

Ice Scour Experiment (PRISE) was conducted at the Centre for Cold Ocean Resources Engineering (C-CORE) to understand the requirement for the safe and cost-effective design of subsea pipelines against the ice scour. The study showed the importance of the "dead wedge" underneath the ice keel with respect to variation in scour loads and subscour soil displacements (C-CORE 1995). Hynes (1996) carried out a centrifuge ice-gouging study in the sand and observed a linear relationship between the scour loads and depth. The author showed that sand deformation affected simple direct shear due to the stress-strain behavior of soil. Eventually, it was suggested that numerical studies should be performed to simulate the centrifuge ice-gouging modeling. Yang (2009) conducted a centrifuge study to measure deformations of ice-scoured sand. It was shown that the maximum gouge force was a function of the gauge geometry and keel attack angle and value of the frontal berm height. Arnau and Ivanović (2019) carried out 1g floor tests on cohesion-less seabed scour due to the ice attack. Ultimately, drifting velocity on the scouring loads was identified as an important parameter. Nematzadeh and Shiri (2019a) developed a CEL model for free-field ice-gouging analysis in sand using ABAQUS/Explicit. The authors incorporated the non-linear strain rate and softening effects through a user-defined subroutine. The study resulted in an improved prediction of the subgouge soil deformation and the keel reaction forces obtained from published experimental studies. Additionally, Nematzadeh and Shiri (2019b) simulated the ice-seabed interaction process using a self-correcting soil model in order to update the shear strength parameters during the pre-peak hardening and the post-peak softening of the sand. The authors showed that the subgouge soil deformation might be overestimated by the conventional decoupled approaches. Additionally, Nematzadeh and Shiri (2020) modeled the effect of the non-linear stress-strain behavior of dense sand in an ice-gouging problem by using a modified Mohr-Coulomb (MMC) model. The authors concluded that the size of side berms and the frontal mound were

affected by the magnitude of the attack angle, where greater subgouge deformations and reaction forces were observed for models with shallower attack angles.

The industry is still looking for alternative approaches to mitigate the computational and experimental efforts and improve the accuracy of simulation at the same time. In recent years, artificial intelligence (AI) approaches have been successfully utilized to model the ice-seabed interaction process since these methods are accurate, quick, and inexpensive. For example, Kioka et al. (2003) and Kioka et al. (2004) modeled an ice-gouging problem using the Neural-Network (NN) approach. The NN showed a high level of accuracy and the authors asserted that this approach could be replaced with the nonlinear multiple-regression methods. Azimi and Shiri (2020a) introduced the parameters governing the ice-seabed interaction process in both clay and sandy conditions. They proposed a set of linear regression (LR) models to estimate the maximum subgouge soil displacements. The authors also concluded that the shear strength parameters of soil and the ratio of gouge depth to gouge width were the most influencing variables. Azimi and Shiri (2020b) simulated the horizontal subgouge deformations in the sand by the gene expression programming (GEP) model. The authors defined six GEP models and then the best GEP model was introduced by conducting a sensitivity analysis. The study showed that the soil depth was considered as the most significant input parameter. Azimi and Shiri (2021) simulated the subgouge sand parameters through robust machine learning (ML). The authors presented a set of matrices in order to estimate the ice-induced sand characteristics.

Although the literature shows that a few research has attempted to model the ice-seabed interaction mechanism by means of AI advancement, there is no study to simulate this problem using the Multi-Layer Perceptron Neural Network (MLPNN). To fill this knowledge gap, the subgouge sand

deformations are simulated by using the MLPNN algorithm in the current study for the first time. More details will be presented in the upcoming sections.

3.1.10. Methodology

3.1.10.1. Horizontal subgouge sand deformation

The maximum horizontal deformation occurs right underneath the keel base. Moving down the seabed, the magnitude of horizontal soil deformation is reduced by a hyperbolic curve. The maximum horizontal subgouge soil deformation (d_{h0}), was obtained using the empirical equations proposed by C-CORE (C-CORE 2009):

$$d_{h0}\sqrt{(\sigma_b/P_a)}/V_e = (0.54/\tan\alpha)\big(1-0.2(I_r+1)\big) \tag{3-21}$$

where, σ_b is bearing pressure (MPa), P_a is atmospheric pressure (0.1 MPa), α is the attack angle (degree), V_e is the maximum vertical extent of subgouge deformation (m), and I_r is dilation index given as:

$$I_r = RD \times (10 - \ln(\sigma_b \times 1000)) - 1 \tag{3-22}$$

where RD is the relative density of sand. The bearing pressure was defined as:

$$\sigma_b(MPa) = 0.09D_s^{1.5} \tag{3-23}$$

where D_s is the gouge depth (m). The maximum vertical extent of subgouge deformation (V_e) was determined based on the range of the dilation index as follows:

$$V_e = \begin{cases} \min[(1+0.5I_r)(D_s+W/5), 1.3(D_s+W/5{,}5D_s)] & if\ \ I_r < 2 \\ \min\big[\big(2.6+0.65(I_r-4)\big)(D_s+W/5), 5D_s\big] & if\ 2 < I_r < 4 \\ 0 & if\ \ I_r > 4 \end{cases} \tag{3-24}$$

where W is the gouge width (m). Equations (3-1) to (3-4) denote the horizontal subgouge deformation (d_{h0}) is a function of $\sigma_b, P_a, V_e, \alpha$ and I_r:

$$d_{h0} = f(\sigma_b, P_a, V_e, \alpha, I_r) \quad (3\text{-}25)$$

Assuming a constant atmospheric pressure during the tests, so Eq. (4-5) can be written as follows:

$$d_{h0} = f(\sigma_b, V_e, \alpha, I_r) \quad (3\text{-}26)$$

Therefore, incorporating the soil depth (y) effect, the target function of horizontal subgouge deformation (d_{hy}) for the artificial intelligence approaches was defined as follows:

$$d_{hy} = f(\sigma_b, V_e, \alpha, I_r, y) \quad (3\text{-}27)$$

The input parameters given in Eq. (3-27) were combined to construct six different MLPNN models with a minimum number of four key parameters. Figure 3-15 portrays the input combination to develop the MLPNN models.

	y	σ_b	V_e	α	I_r
MLPNN 1	■	■	■	■	■
MLPNN 2	■	■	■	■	
MLPNN 3	■	■	■		■
MLPNN 4	■	■		■	■
MLPNN 5	■		■	■	■
MLPNN 6		■	■	■	■

Figure 3-15. Input combination to develop the MLPNN models

3.1.10.2. Construction of database

The studies published by C-CORE (1995), Hynes (1996), and Yang (2009) are amongst the most important experimental studies related to ice-induced sand scouring problems in which input

parameters ($\sigma_b, V_e, \alpha, I_r, y$) and the target parameter (d_{hy}) can be expressly adopted to feed the developed MLPNN models. The key results of the experimental studies conducted by C-CORE (1995) (eight test results in the sand with different relative densities), Hynes (1996) (seven centrifuge tests in clean dry silica sand), and Yang (2009) (seven centrifuge ice-gouging test in AlWhite Silica sand) were used to construct a database for validation of the artificial intelligence numerical models. Using these experimental studies, the attack angle (α), the scour depth (D_s), the scour width (W), the g-level, the relative density (*RD*), and the horizontal subgouge deformation (d_{hy}) at different depths (y) were extracted. Hynes (1996) used loose and dense sands but did not mention the values of *RD*. Using an existing classification study (e.g., Mitchell and Soga 2005), relative densities of 0.3 and 0.7 were adopted for loose and dense sands of Hynes (1996). Care should be taken in using the laboratory test data for testing the AI models. The idealized test conditions may affect the reliability of model prediction to some extent. To overcome this limitation, the majority of the data in the current study were selected from centrifuge studies which is one of the most reliable technics for model testing in geotechnical engineering. The centrifugal acceleration simulates gravity and allows for correspondence of stress fields between the model and full-scale resulting in an accurate model prediction within a bandwidth of less than 1.0 times the normalized loads (C-CORE 1995; C-CORE 2009). This provided a relaxation to build up the current database and feed the artificial intelligence models in the current study. Furthermore, some of the key parameters such as gouge depth are usually determined by field-specific statistical and probabilistic analysis prior to any ice-gouging analysis, which is considered as a sufficiently reliable method in industrial practice. It is worth noting that 70% of the established dataset was applied to training the MLPNN models, whereas 30% of the rest was utilized to test these models.

3.1.10.3. Goodness of fit

In order to evaluate the MLPNN results, correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute error (*MAE*), Nash-Sutcliff efficiency coefficient (*NSC*), and Akaike Information Criteria (*AIC*) were applied as below:

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{3-28}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{3-29}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{3-30}$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{3-31}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{3-32}$$

here, O_i, P_i, $\bar{O}$, $\bar{P}$, and n are the observed ice-induced seabed parameters, the simulated values, the average observed values, the average simulated values, and the number of experimental measurements, respectively. Moreover, the k is the number of hidden layer neurons used in the MLPNN network. The closeness of the correlation coefficient (*R*) and the Nash-Sutcliff efficiency coefficient (*NSC*) to one signified that the MLPNN model owned the highest correlation with the observed values, whilst the closeness of the *RMSE* and *MAE* criteria to zero meant that the particular model had the lowest level of error. To overcome this drawback, the Akaike Information Criteria (*AIC*) was employed. The less complex MLPNN model had the lowest value of the *AIC* index, hence, the superior MLPNN model owned the lowest value of the *AIC* index and error

(*RMSE* and *MAE*) and the highest level of correlation (*R* and *NSC*) with the experimental measurements.

3.1.10.4. Multi-layer perceptron neural network (MLPNN)

Owing to the flexibility of Multi-layer perceptron neural network (MLPNN) to simulate complex problems (Haykin 1994), the MLPNN is considered one of the most common feedforward neural networks (FFNN). Each MLPNN model has a set of layers comprising an input layer, at least a hidden layer, and an output layer (see Figure 3-16).

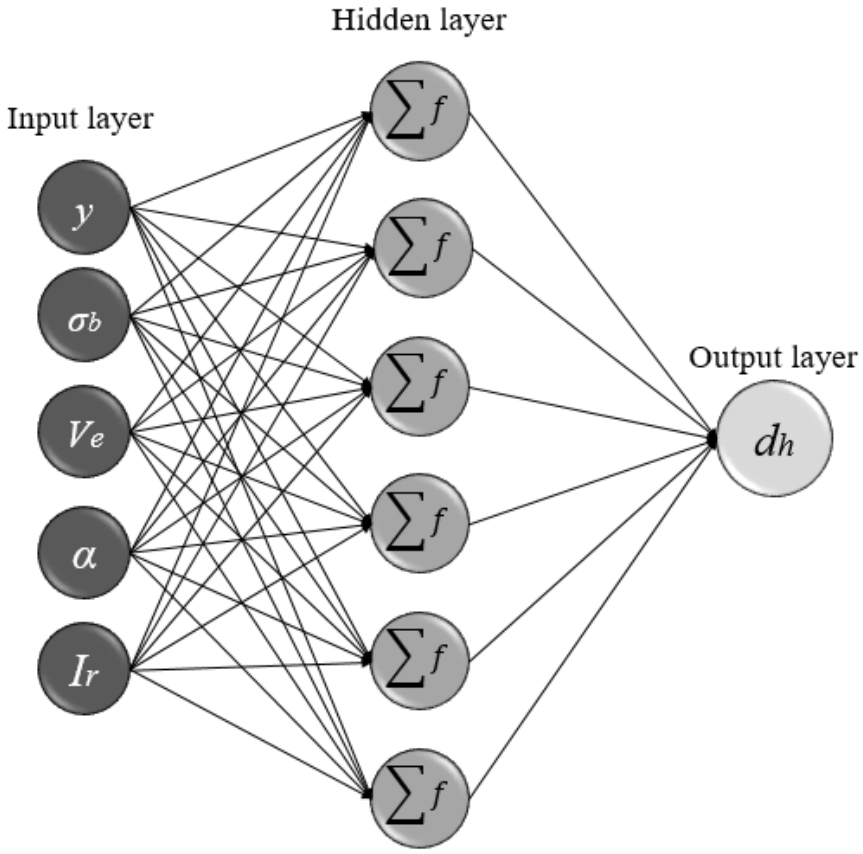


Figure 3-16. Multi-layer perceptron neural network (MLPNN)

Neurons are in the perceptron layers with a number equal to the input and output variables of the problem, meaning that the number of neurons in the input layer equals five input parameters including $\sigma_b, V_e, \alpha, I_r, y$, whereas one neuron in the output layer is considered the target parameter (d_{hy}). There are no specific rules and regulations to set the number of neurons in the hidden layer. Therefore, a trial and error method is applied to select the number of neurons in the hidden layer (Bilhan et al. 2010). In the current study, the number of neurons in the hidden layer is initially assumed equal to one and the precision of the MLPNN model is evaluated by increasing the number to 10. The most optimal number of the hidden layer neurons is selected equal to 6 since the accuracy of the MLPNN model is insignificantly changed. In Figure 3-17, a variation of the number of neurons in the hidden layer against the mean absolute error (*MAE*) index is exhibited. The number of neurons in the antecedent layer is collected using summed weighted neurons in hidden and output layers and transferred to the next layer by employing an activation function. To choose the optimized activation function, three functions including "linear", "sigmoid", and "hyperbolic tangent" were used to simulate the subgouge deformations. Subsequently, the performance of these activation functions in terms of accuracy was compared together. Lastly, the sigmoid activation function was selected to estimate the deformations. Additionally, the sigmoid activation function has been used in different fields because of its acceptable performance (Smith 1993). The values of weighted outputs were summed by means of the MLP throughout a procedure called "model training". The Levenberg-Marquardt (LM) algorithm was applied to train the MLP neural network, where the biases and weights were calculated using the back-propagation algorithm.

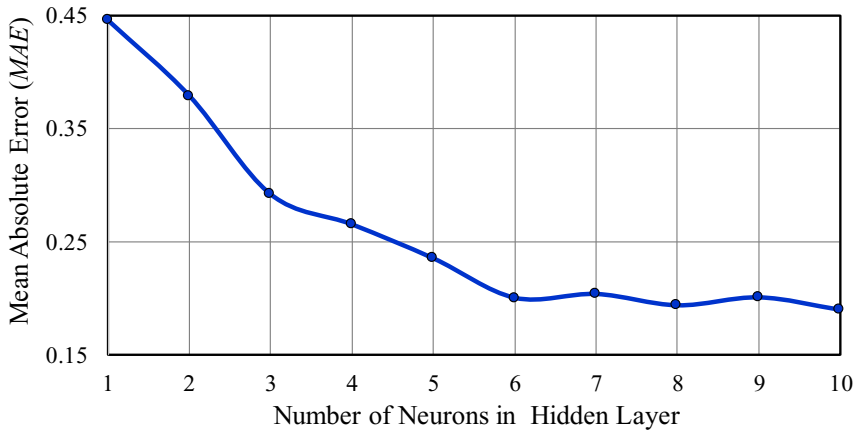


Figure 3-17. Variation of the number of neurons in the hidden layer against mean absolute error (*MAE*) index

3.1.11. Results and discussion

Gaining a good insight into the best input combination (the superior MLPNN model) and the most important input parameters in an ice-induced scouring problem was one of the key objectives of the current study. It is worth noting that the results of MLPNN models in the test mode are evaluated in the current study. Figure 3-18 showes the key statistical indices for MLPNN 1 to MLPNN 6 models. To do this, MLPNN 1 was defined by using all input parameters comprising $\sigma_b, V_e, \alpha, I_r$ and y where the *AIC*, *RMSE*, and *NSC* values were respectively computed to be 4.727, 0.344, and 0.785. Subsequently, five MLPNN models including MLPNN 2 to MLPNN 6 were produced adopting a combination of the other four inputs. For instance, the σ_b, V_e, α, y parameters were utilized to developed MLPNN 2, meaning that the influence of the dilation index parameter (I_r) was ignored. The values of root mean square error and Akaike information criteria for the MLPNN 2 model were equal to 0.879 and 18.150, respectively. The ice keel attack angle (α) was an eliminated factor to simulate the target parameter by means of the MLPNN 3 model, with a Nash-Sutcliff efficiency coefficient (*NSC*) of 0.387. Using the σ_b, α, I_r, y input parameters, the

subgouge sand deformations were predicted by the MLPNN 4 model. It means that the maximum vertical extent of subgouge deformation (V_e) was removed to model the ice-induced through MLPNN 4. The *AIC*, *NSC*, and *RMSE* indices for the MLPNN 4 model were respectively surmised at 9.753, 0.566, and 0.489. Though the V_e, α, I_r, y inputs were utilized to prognosticate the subgouge sand displacements using MLPNN 5, the bearing pressure (σ_b) was a removed parameter. For the MLPNN 5 model, the *RMSE* and *AIC* criteria were approximated as 0.622 and 13.200. To simulate the subgouge sand displacements by MLPNN 6, the $\sigma_b, V_e, \alpha, I_r$ input factors were applied. In other words, the soil depth parameter (y) was a disregarded input to model the target values where the *AIC*, *RMSE*, and *NSC* statistical indices were 15.088, 0.710, and 0.086.

The results of error analysis (*MAE*) for the MLPNN models are depicted in Figure 3-19. Regarding the error analysis, the overwhelming majority of the deformations simulated by MLPNN 1 (nearly 82%) had an error of less than 5%, rather a tiny minority of the predicted displacements (roughly 3%) by this MLPNN model possessed an error of between 5% and 10%. The conducted error analysis showed that almost 18% of the results obtained from MLPNN 2 had an error of greater than 10%. Even though approximately 18% of the subgouge sand deformations modeled by MLPNN 3 possessed an error of greater than 10%, this value for MLPNN 5 was surmised as 12%. The result of error analysis for MLPNN 5 and MLPNN 6 was similar, signifying that about 61% of the ice-induced displacements simulated by these MLPNN models showed an error of less than 5%.

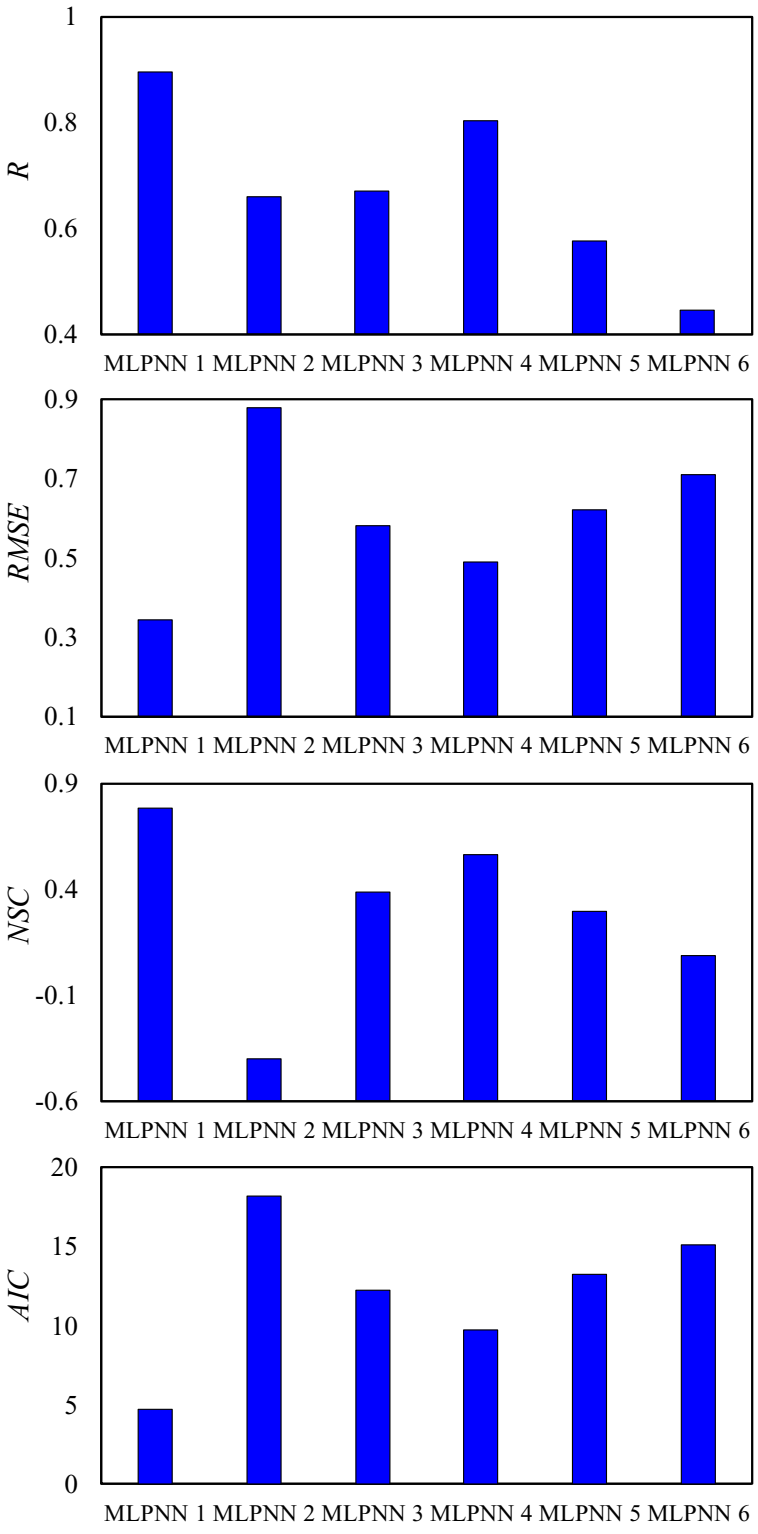


Figure 3-18. The key statistical indices for MLPNN 1 to MLPNN 6

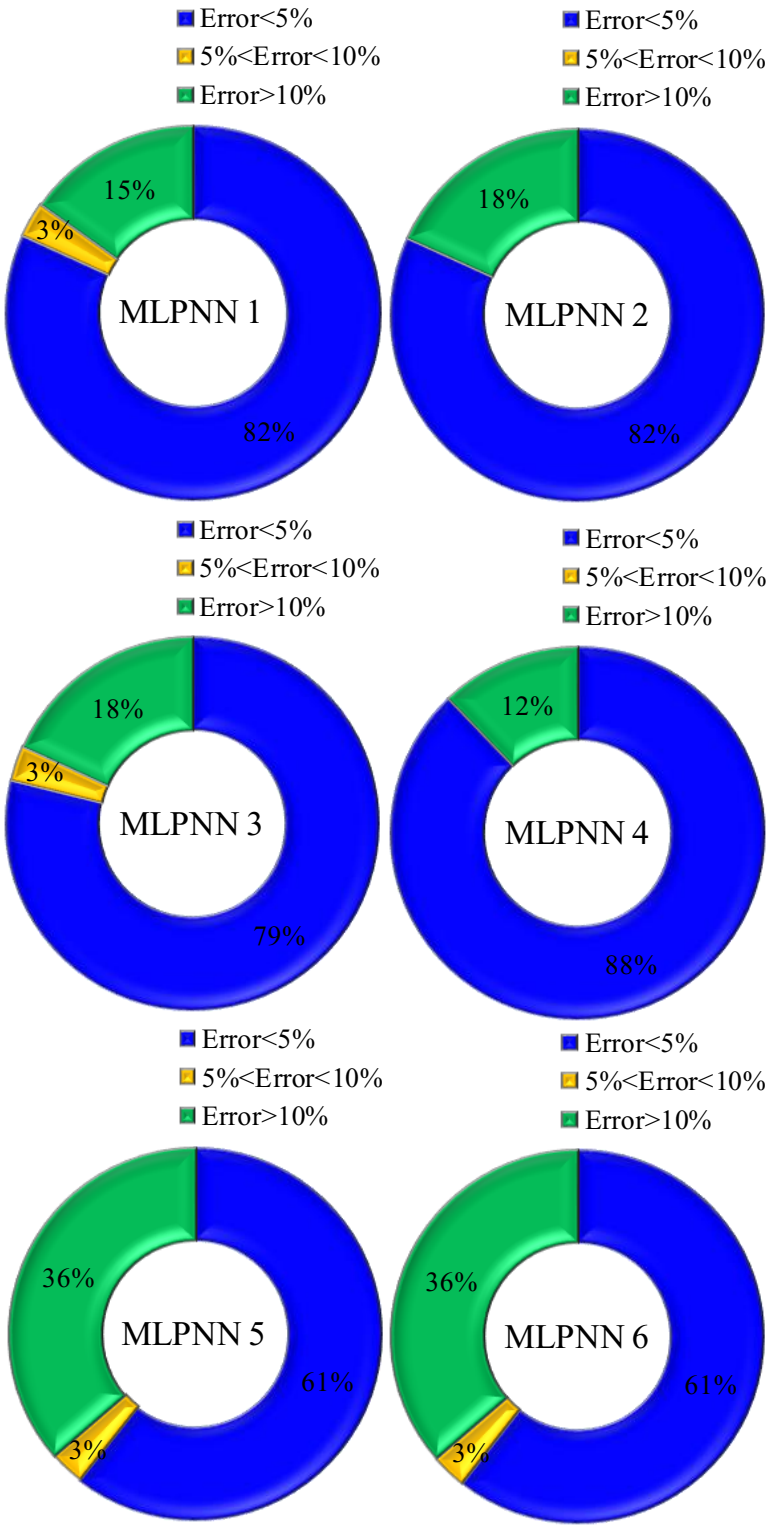


Figure 3-19. Results of error analysis (MAE) for MLPNN models

Figure 3-20 illustrates the scatter plots for the MLPNN models. According to the simulation results, the highest level of correlation was obtained for MLPNN 1, with a correlation coefficient of 0.879, whilst MLPNN 6 demonstrated the lowest level of correlation among the MLPNN model (R=0.447). Moreover, the correlation coefficient value for MLPNN 2, MLPNN 3, MLPNN 4, and MLPNN 5 were respectively computed to be 0.661, 0.671, and 0.805, 0.577.

A sensitivity analysis (SA) was conducted by comparison of the results from the developed MLPNN models. As a result, the MLPNN 1 model simulated the target function by using all of the input parameters and eliminating the effects of other parameters for MLPNN 2 to MLPNN 6 including σ_b, Ve, α, Ir, and y, respectively. The superior MLPNN model (MLPNN 1) had the lowest error and complexity along with the highest correlation. By contrast, the accuracy of the MLPNN model decreased remarkably by removing the input parameters since the level of effectiveness for the eliminated parameter was significant. After the MLPNN 1 model, MLPNN 4 was recognized as the second-best model so as to predict the subgouge sand deformations. Furthermore, MLPNN 3, MLPNN 2, and MLPNN 5 were respectively ranked as the third-best, fourth-best, and fifth-best MLPNN models in estimating the target values. It is worth mentioning that MLPNN 6 had the worst performance among the MLPNN models.

On the other hand, the performed sensitivity analysis indicated that the soil depth parameter (y) was detected as the most influencing input factor for the simulation of the subgouge sand displacements by the MLPNN algorithm as the performance of the MLPNN model was reduced considerably by removing this input parameter. The bearing pressure (σ_b), dilation index (I_r), and keel attack angle (α) were respectively identified as the second-important, third-important, and fourth-important input variables. Based on the simulation results, the maximum vertical extent of

subgouge deformation (V_e) had the lowest level of effectiveness to model the subgouge displacements in the sand.

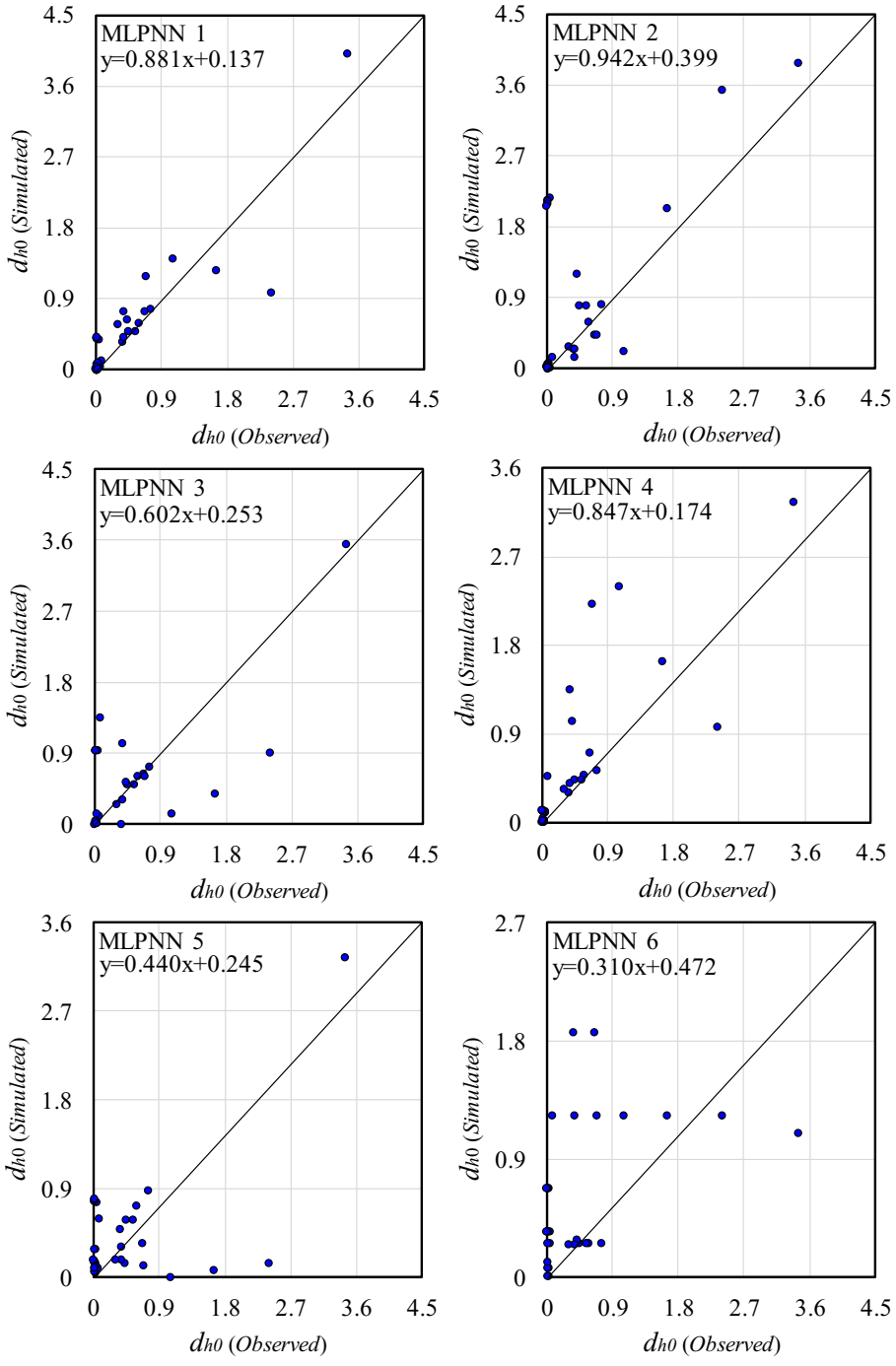


Figure 3-20. Scatter plots for MLPNN 1 to MLPNN 6

As shown, MLPNN 1 was introduced as the superior MLPNN model to simulate the subgouge sand deformations. The profiles of horizontal subgouge deformation simulated by MLPNN 1 (best model) are presented in Figure 3-21. Regarding the simulation results, MLPNN 1 demonstrated the best performance to predict the ice-induced sand displacements, signifying that MLPNN 1 possessed the highest level of precision and correlation along with the lowest level of complexity. As shown, the maximum value of deformations simulated just beneath the iceberg keel and the magnitude of this parameter decreased in greater depth by a hyperbolic curve. Even though the applied datasets (C-CORE 1995, Hynes 1996, and Yang 2009) were quite heterogeneous, the MLPNN 1 managed to simulate the subgouge sand deformations with its highest performance. In other words, MLPNN 1 at dealing with modeling the small deformations had some discrepancies (Yang's model), whereas this model demonstrated a substantial correlation in large displacements (C-CORE and Hynes's models).

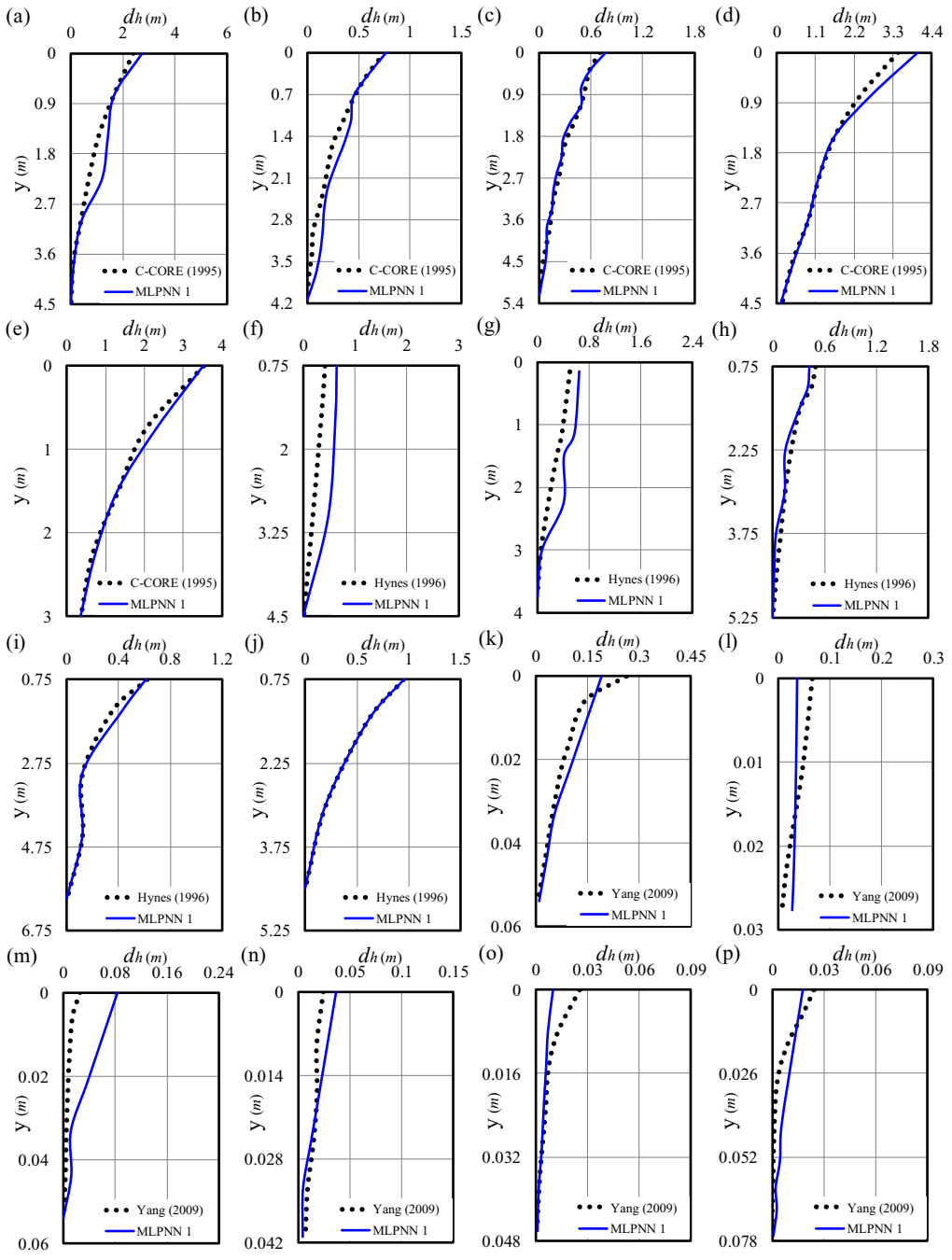


Figure 3-21. Horizontal subgouge deformation profiles simulated by MLPNN 1

MLPNN 1 in terms of all input factors comprising $\sigma_b, V_e, \alpha, I_r$ and y was considered as the premium MLPNN model in order to simulate the subgouge sand displacements. This model demonstrated a good performance at dealing with modeling the target function. The sensitivity

analysis proved that the soil depth (y) and the bearing pressure (σ_b) were regarded as the most effective input factors.

3.1.12. Conclusion

In this study, the subgouge sand deformations were simulated by using the multilayer perceptron neural network (MLPNN) for the first. Initially, the parameters affecting the ice-induced sand displacements were identified, and then by using these factors governing, six MLPNN models ranging from MLPNN 1 to MLPNN 6 were defined. Subsequently, using the published literature, a comprehensive dataset was constructed. It is worth mentioning that 70% of the data were applied to train the MLPNN models and 30% of the remaining was used to test these AI models. After that, the optimized number of neurons in the hidden layer was opted as six, while the sigmoid was selected as the best activation function for the MLPNN algorithm. By performing a set of analyses, the most important results are summarized as follows:

- The MLPNN 1 model a function of all input parameters managed to simulate the subgouge sand deformations with the highest level of accuracy and correlation along with the lowest level of complexity.
- The *R*, *RMSE*, and *AIC* statistical indices for MLPNN 1 were respectively calculated to be 0.897, 0.334, and 4.727.
- Error analysis demonstrated that almost 82% of the ice-induced predicted by MLPNN 1 deformations had an error of less than 5%.
- The performed sensitivity analysis indicated that the soil depth (y), bearing pressure (σ_b), and dilation index (I_r) were detected as the most influencing input factors, rather the

effect of the maximum vertical extent of subgouge deformation (V_e) was quite insignificant to model the subgouge sand displacements.

- Regarding the simulation results, the MLPNN model was able to simulate the large deformations with better performance, whereas some discrepancies were observed in dealing with the small subgouge deformations.

The presented results in the current investigation can facilitate proposing quick and cost-effective solutions for the simulation of the subgouge sand displacements. Furthermore, these outcomes may be beneficial to the planning of costly experimental studies and time-consuming numerical simulations and mitigate the expenditures of future works.

3.1.13. Acknowledgments

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

3.1.14. References

Alba, J.L., 2015. Ice Scour and Gouging Effects With Respect to Pipeline and Wellhead Placement and Design. Bureau of Safety and Environmental Enforcement (BSEE), Wood Group Kenny, Report No.100100.01. PL. REP.004, Houston, TX.

Arnau, S, Ivanović, A., 2019. Rate effects during ice scour in sand. Cold Regions Science and Technology, 158: 182-194.

Azimi, H., Shiri, H., 2020a. Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 1-13.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 105(3): 1-29.

Bilhan, O, Emiroglu, ME, and Kisi, O., 2010. Application of two different neural network techniques to lateral outflow over rectangular side weirs located on a straight channel. Advances in Engineering Software, 41: 831-837.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour. C-CORE Report 95-Cl2, St. John's, NL, Canada.

C-CORE., 2009. PIRAM-Gouge Morphology Study. C-CORE Report R-09-013-490 v1, St. John's, NL, Canada.

Haykin, S., 1994. Neural networks: a comprehensive foundation. Prentice Hall PTR.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand, Doctoral dissertation. Memorial University of Newfoundland, St. John's, NL, Canada.

Kioka, S.D, Kubouchi, A, Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event By A Neural-Network. Proc 13th Int Offshore and Polar Eng Conf, Hawaii, ISOPE-I-03-081.

Kioka, S, Kubouchi, A, Ishikawa, R, Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. Proc 14th Int Offshore and Polar Eng Conf, Toulon, ISOPE-I-04-107.

Nematzadeh, A, Shiri, H., 2019a. Large deformation analysis of ice scour process in dense sand. 10th Congress on Numerical Methods in Engineering, Guimaraes, Portugal.

Nematzadeh, A, Shiri, H., 2019b. Self-correcting soil models for numerical simulation of strain rate dependent ice scour in sand. 9th ECCOMAS Thematic Conference on Smart Structures and Materials, Paris, France.

Nematzadeh, A, Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Smith, M., 1993. Neural networks for statistical modeling, Thomson Learning.

Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Section 3

Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process

This section is a journal paper which has been published in Journal of Offshore Mechanics and Arctic Engineering (2020), Vol. 142, pp. 051601, ISSN: 0892-7219

Abstract

Prediction of subgouge soil deformation during an ice-gouging event is a challenging design factor in Arctic subsea pipelines. An accurate assessment of ice keel-seabed interaction requires expensive model testing and large deformation finite element analysis. Proposing reliable analytical/empirical solutions needs a deep understanding of the key parameters governing the problem. In this study, dimensional analysis of subgouge soil deformations was conducted and eight dimensionless groups of parameters were identified to facilitate proposing potential new solutions. A comprehensive dataset was established for horizontal and vertical subgouge deformations in both sand and clay seabed. Using the identified dimensionless groups, linear regression (LR) models were developed to estimate the horizontal and vertical deformation. Moreover, a sensitivity analysis (SA), as well as an uncertainty analysis (UA), were carried out to identify the superior LR models and the most influential parameter group. A high range of correlation coefficient (R), Nash-Sutcliff efficiency coefficient (NSC), and variance accounted for (VAF) along with a low range of errors was achieved for the best LR model. The results of the superior LR models were also compared with the existing empirical equations. The study showed that the shear strength parameters of the seabed soil and the ratio of gouge depth to gouge width are the governing dimensionless parameters to model the horizontal and vertical subgouge soil deformations.

Keywords: Dimensional analysis, Linear regression (LR), Pipeline, Sensitivity analysis (SA), Subgouge soil deformations, Uncertainty analysis (UA)

3.1.15. Introduction

Over the last decades, oil and gas development projects have been extended toward the Arctic offshore territories. The overwhelming majority of the extracted hydrocarbon deposits are transferred through subsea pipelines from offshore to onshore (Emmerson and Lahn 2012; Alba 2015). In shallow waters, marine pipelines and other facilities like communication cables and wellheads are threatened by traveling icebergs that may gouge the seabed and attack the subsea pipelines and structures (Lanan et al. 2011; Kenny and Jukes 2017). The ice-gouging process causes the subgouge soil to deform and displace the buried pipeline (see Figure 3-22). The maximum horizontal deformation occurs right underneath the keel base (Lach 1996; Nematzadeh and Shiri 2020).

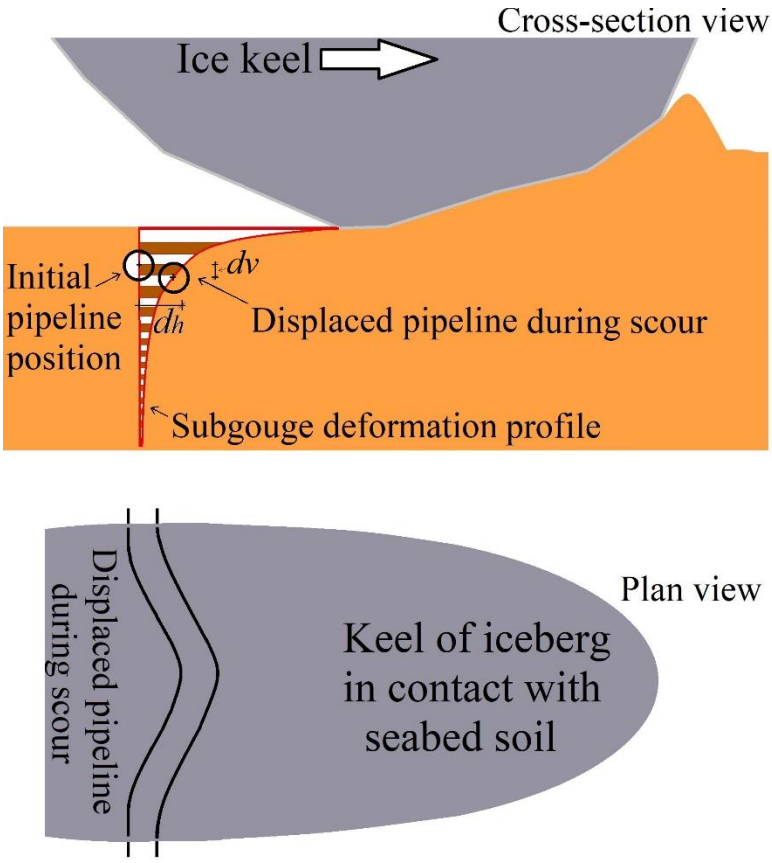


Figure 3-22. A schematic layout of deformations of subgouge soil and buried subsea pipeline during an ice-gouging event

In contrast, the determination of the maximum displacement of pipelines and the resultant safe burial depth is a challenging issue that can simply play with the project budget redlines. The current state of practice for the assessment of the ice's impact on buried pipelines combines the simplicity of the beam-spring approach for modeling the pipeline with the accuracy of the continuum approach for modeling the free field ice-gouging process. Therefore, the free field ice-gouging analysis is the key part of the practical pipeline design against ice impact. By contrast, the identification of the parameters affecting maximum ice-induced soil deformations has crucial importance in ice keel-seabed interaction (Nobahar et al. 2007a; Nobahar et al. 2007b; Kenny et al. 2004).

There are numerous experimental (Paulin 1991; Paulin 1992; C-CORE 1995; Hynes 1996; C-CORE 1996; Yang 2009; Schoonbeek et al. 2006; Been et al. 2008), numerical (Nematzadeh and Shiri 2020; Kenny et al. 2004; Nobahar 2003; Konuk 2009; Moore et al. 2011), and analytical (Chari and Allen 1974; Chari 1975; Chari 1980; Nessim and Jordaan 1985; Comfort and Graham 1986; Prasad and Chari 1986) studies in the literature that have investigated various aspects of the ice-gouging problem. However, there is still no published work (to the authors' knowledge) about the dimensionless group of parameters governing the mechanics of the problem, which is a key knowledge gap that tried to be filled through the objectives of the current study.

Creating a database for the analysis of parameters is quite challenging due to the different approaches and setups in published studies and having limited access to the data. A screening study was first conducted to select the appropriate studies giving access to a sufficient amount of data.

The experimental studies published by Paulin (1991), Paulin (1992), C-CORE (1995), Hynes (1996), C-CORE (1996), and Yang (2009) were used for analyzing the sandy seabed. For clay seabed, the test results published by C-CORE (1995), Lach (1996), C-CORE (1996), Schoonbeek (2006), and Been (2008) were used. Paulin (1991 and 1992) conducted experimental investigations to measure ice-scouring characteristics including surface deformations, subscour displacements, pressure, and reaction forces. The authors used both sand and clay seabed in different conditions of saturation and submergence and concluded that soil deformation is increased by decreasing the relative density of soils. Also, the studies showed that measured loads and forces by the submerged models were less than the dry tests, whereas other values were similar in both the dry and submerged circumstances. Hynes (1996) carried out a series of pressure ridge ice scour experiments (PRISE) to measure horizontal and vertical loads, gouge-induced pore pressures and stresses, and keel-soil interface pressures for a sandy seabed. The results of the study proved the existence of the shear dragging theory, as well as a linear relationship between scour loads and the, scour depth. Yang (2009) investigated subscour deformations on a sandy seabed through a series of testing programs called pipeline-ice risk assessment and mitigation (PIRAM). Besides assessing the pressure and reaction forces, the authors adopted particle image velocimetry (PIV) to measure the subgouge deformations and frontal berms. The study indicated that the maximum horizontal subscour deformations were formed just beneath the ice keel and reduced with the soil depth. Lach (1996) and Schoonbeek (2006) conducted several ice-gouging centrifuge tests in clay to assess the subgouge soil deformations and reaction forces. The deformations were visualized by means of the image processing technique, whilst vertical and horizontal loads were measured using load cells. The authors provided an empirical relationship in which the gouge-induced horizontal deformation was a function of soil properties and the keel geometry. Been et al. (2008) carried out

an experimental study to measure the subgouge forces and deformations in clay. The presence of a pipe, stratified sediments, backfill features, and geometry of the trench was ignored by the authors to provide an idealized subscour model. The authors suggested further studies for an in-depth assessment of the parameters affecting subgouge deformations. The results published within the aforementioned studies were combined for the creation of a comprehensive dataset and to perform the dimensional analysis. The study facilitated proposing potential new solutions for modeling the ice-gouging process.

3.1.16. Methodology

The Buckingham-π theorem was used for dimensional analysis of the subgouge soil deformation problem for the first time in the literature. Figure 3-23 shows the flowchart of the dimensional analysis and determining the dimensionless group of parameters in the current study.

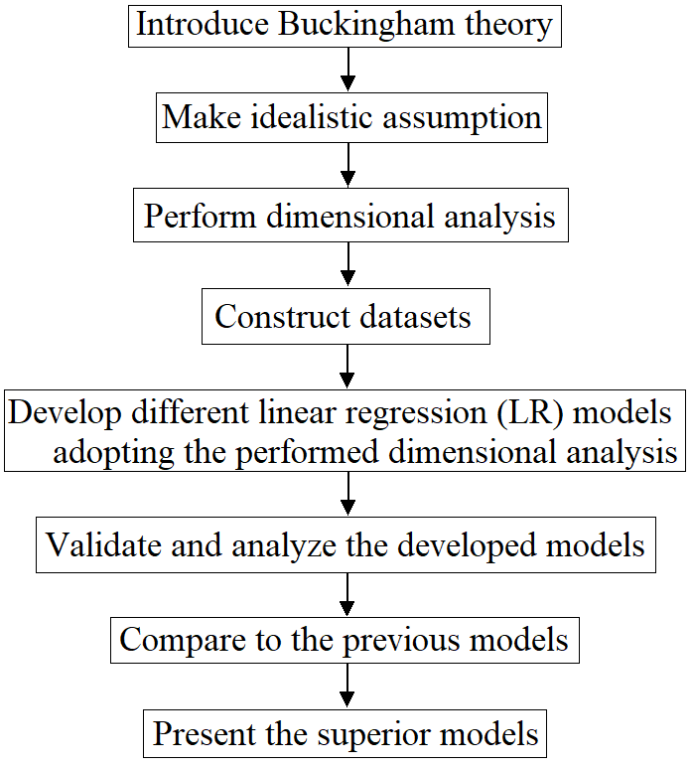


Figure 3-23. Flowchart of the dimensional analysis

Linear regression (LR) models were developed for both sand (nine LR models) and clay (eight LR models) seabed. Through a sensitivity analysis (SA) and an uncertainty analysis (UA), the premium LR models for estimating the maximum horizontal and vertical deformations were introduced and the most important dimensionless parameter groups were also ascertained.

3.1.17. Dimensional analysis

To perform a robust and comprehensive dimensional analysis of subgouge soil deformations, some simplifying suppositions were made as follows to facilitate the implementation of the model in practice:

- Subgouge soil deformations are large (several pipe diameters) so Coulomb's law is able to express the function between shear stress and normal stress. Therefore, the parameters c (cohesion of clay) and φ (internal friction angle of sand) can describe the soil behavior (Schuring and Emori 1965; Bekker 1956, Janosi 1961) during an ice-gouging event,
- The maximum subgouge deformation is formed just under the ice keel in the gouge (C-CORE 1995; Bekker 1956),
- Cohesionless soil (sand or gravel) is assumed in drained conditions owing to quick water drainage during an ice-gouging event (Alba 2015), Cohesive soil (clay and silt) is assumed in undrained conditions owing to gradual water drainage during an ice-gouging event (Alba 2015),
- The contact surface between the ice and the seabed is rough enough since a layer of the soil has adhered to the keel (Schuring and Emori 1965),
- Temperature is constant during an ice scouring problem.

Regarding the C-CORE (2009a; 2009b), subgouge sand deformation ($d_{(sand)}$) is as a function of the gouge depth (D_s), atmospheric pressure (P_a), the relative density of sand (DR), the gouge width (W), and attack angle (α) as follows:

$$d_{(sand)} = f_1(D_s, P_a, DR, W, \alpha) \tag{3-33}$$

However, it should be noted that Eq. (3-33) can be written for any type of soil including sand and clay as follows:

$$d_{(soil)} = f_2(D_s, P_a, c, \varphi, W, \alpha) \tag{3-34}$$

where, c is the cohesion of clay and φ is the internal friction angle of sand. Figure 3-24 schematically shows the force systems in an ice-gouging event (Duplenskiy 2012; Duplenskiy and Gudmestad 2013).

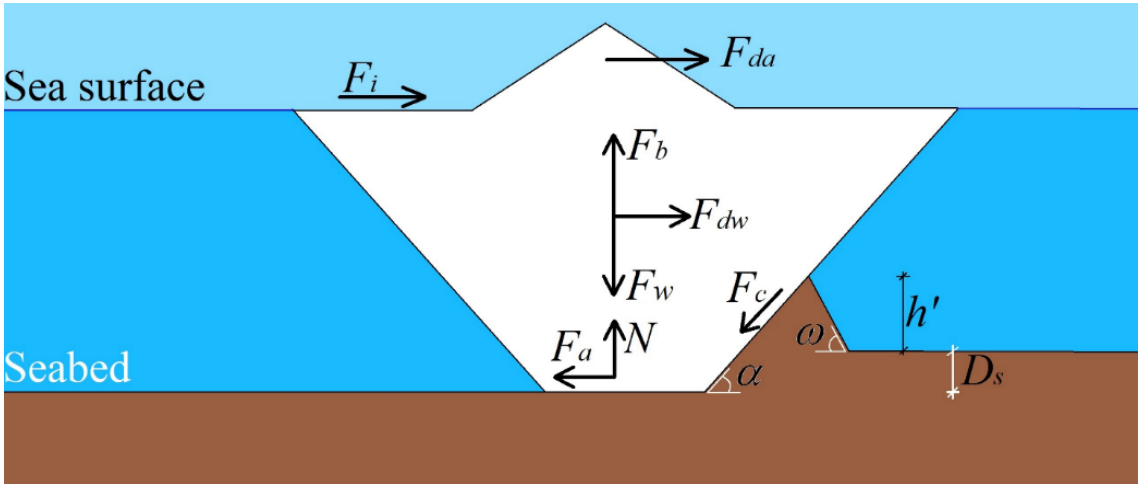


Figure 3-24. Schematic layout of ice-gouging analysis (Duplenskiy 2012)

The parameter F_{da} is the drag force from the wind; F_{dw} is the drag force from the current; F_b is buoyancy force; F_w is ridge weight; N seabed reaction; F_a is friction force on the bottom of the ridge; F_c is Coulomb's passive friction force acting in front and on both sides of the ridge; F_i is driving force from the surrounding floe; ω is the angle of the surcharged soil slope; α is the attack

angle; *h'* is the height of the berm, and D_s is the scour depth. Assuming the force system in Figure 3-24, the resultant horizontal (L_h) and vertical forces (L_v) can be written as follows:

$$L_h = \sum_{i=1}^{n} F_{h_i} = F_i + F_{da} + F_{dw} - F_a - F_c \cos\omega \tag{3-35}$$

$$L_v = \sum_{i=1}^{n} F_{v_i} = F_b + N - W - F_c \sin\omega \tag{3-36}$$

Therefore, equation (3-34) can be rewritten as follows:

$$d_{(soil)} = f_2(D_s, P_a, c, \varphi, W, \alpha, \omega, h', L_h, L_v) \tag{3-37}$$

Furthermore, Schuring and Emori (1965) showed that soil deformations caused by bulldozing, torquing, and penetration are a function of *V* or velocity of the plow or the tractor, *W* or the width of the plow, the plate or the tractor, L_i or driving force, ρ_s or the soil density, *g* or gravitational acceleration, φ and *c* as follows:

$$d_{(soil)} = f_3(V, W, L_i, \rho_s, g, c, \varphi) \tag{3-38}$$

Thus, equation (3-37) can be written as follows:

$$d_{(soil)} = f_2(D_s, P_a, c, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \rho_s, g) \tag{3-39}$$

here, γ_s is the specific weight of soil which equals to $\rho_s \times g$, while the atmospheric pressure (P_a) is constant and can be ignored. Therefore, equation (3-37) is rewritten as follows:

$$d_{(soil)} = f_2(D_s, c, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \tag{3-40}$$

In other words, the subgouge soil deformation can be expressed by using parameters in equation (3-40) and the equation is written in a dimensional form as follows:

$$d_{(soil)} = f_4(\Pi_1, \Pi_2, \Pi_3, \Pi_4, \Pi_5, \Pi_6, \Pi_7, \Pi_8) \tag{3-41}$$

where, Π_1, Π_2,…, Π_8 are dimensionless groups and f_4 is a functional symbol regarding Buckingham's π theory. Therefore, the following dimensionless groups can be defined:

$$\Pi_1 = \frac{D_s}{W} \tag{3-42}$$

$$\Pi_2 = \frac{c}{\gamma_s W} \tag{3-43}$$

$$\Pi_3 = \varphi \tag{3-44}$$

$$\Pi_4 = \alpha \tag{3-45}$$

$$\Pi_5 = \omega \tag{3-46}$$

$$\Pi_6 = \frac{h'}{W} \tag{3-47}$$

$$\Pi_7 = \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3} \tag{3-48}$$

$$\Pi_8 = \frac{V^2}{gW} \tag{3-49}$$

Thus, equation (3-40) is expressed as a function of eight dimensionless groups in the following form:

$$\frac{d_{(soil)}}{W} = f_5\left(\frac{D_s}{W}, \frac{c}{\gamma_s W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{3-50}$$

The $d_{(soil)}$ is considered the maximum subgouge soil deformation which is formed just under the moving ice keel in the gouge centerline. It is noteworthy that the gouge depth and attack angle are functions of some other parameters such as seabed configuration, ice features, and environmental loads. However, these two parameters are usually determined by field-specific statistical and probabilistic analysis before any ice-gouging analysis. Therefore, these two parameters are commonly used as input parameters in experimental and numerical analyses.

3.1.18. Construction of database

Various experimental studies were utilized to analyze the horizontal $\left(d_{h_{(soil)}}\right)$ and vertical $\left(d_{v_{(soil)}}\right)$ subgouge soil deformations for both sand and clay seabed. For the sand, the key values of six experimental studies conducted by Paulin (1991) (one test), Paulin (1992) (four tests), C-CORE (1995) (five tests), Hynes (1996) (five tests), C-CORE (1996) (five tests), and Yang (2009) (seven tests) are employed. For clay, five laboratory investigations performed by C-CORE (1995) (eight tests), Lach (1996) (eight tests), C-CORE (1996) (four tests), Schoonbeek (2006) (one test), and Been (2008) (one test) were used to evaluate the subgouge soil deformations.

The test condition in the iceberg-seabed interaction centrifuge laboratory test depends on the thickness of the soil layer and the size of the model iceberg. According to the web search results, plane strain conditions give the lowest fracture toughness value which is a material property. Plane strain conditions generally prevail when the specimen thickness is greater than or equal to a certain value, which depends on the material and the crack length. Plane stress conditions occur when the specimen thickness is much smaller than the crack length.

For example, in Lach's (1996) and Yang's (2009) centrifuge tests, the plane strain conditions were applied.

Therefore, if the soil layer is thick enough and the model iceberg is small enough, the test condition can be approximated as plane strain. However, if the soil layer is thin or the model iceberg is large, the test condition may deviate from plane strain and have some 3D effects or boundary effects. In that case, the test condition may be closer to real strain, which is the full 3D formulation of strain.

In terms of the boundary conditions in the experimental works, generally, the boundary conditions in iceberg-seabed interaction centrifuge laboratory tests depend on the specific setup and design of the experiments. For example, the top boundary of the soil layer is subjected to a cyclic loading that simulates the wave-induced pressure on the seabed. The bottom boundary of the soil layer is

fixed and impermeable. The lateral boundaries of the soil layer are either free or restrained by a rigid wall. The model iceberg is either fixed or movable and can have different shapes, sizes, velocities, and attack angles.

Table 3-9 shows the summary of the employed dimensionless parameters including maximum, minimum, average, variance, and standard deviation of the experimental measurements.

The statistical significance of correlations was also checked by conducting a t-test and calculation of the P-value, assuming that a P-value of 0.05 or less is statistically significant (Bland and Peacock 2002; Box 1987). This, indeed, was to assess the probability of the relationship between the parameters being affected by an alternative hypothesis (Myers et al. 2010). The calculated P-values for the sand and clay were 0.039 and 0.024, showing that the correlations are statistically significant.

Table 3-9. Range of employed parameters in the current study

Soil	Values	dh/W	dv/W	Ds/W	φ	α	ω	h'/W	$Lh/\gamma s.W^3$	$Lv/\gamma s.W^3$	$V^2/g.W$
	Max	1.356	0.068	0.333	44	30	39.1	1.163	2.235	2.859	0.015
	Min	0.002	0.004	0.065	18	15	-	-	-	-	-
Sand	Ave	0.300	0.019	0.125	34.745	20.556	4.633	0.168	0.586	0.756	0.004
	Var	0.186	0.0002	0.004	28.875	52.469	126.806	0.053	0.378	0.651	3.6E-05
	Std	0.431	0.015	0.064	5.374	7.244	11.261	0.230	0.615	0.807	0.006
	Values	dh/W	dv/W	Ds/W	$c/\gamma s.W$	α	h'/W		$Fh/\gamma s.W^3$	$Fv/\gamma s.W^3$	$V^2/g.W$
	Max	0.396	0.189	0.333	24.059	45	0.077		118.731	243.712	0.031
	Min	0.039	0.013	0.033	0.027	15	-		-	-	-
Clay	Ave	0.166	0.071	0.125	4.791	19.6	0.015		9.907	25.578	0.008

Var	0.008	0.003	0.007	45.999	63.84	0.001	564.852	2720.803	8.01E-05
Std	0.089	0.052	0.084	6.782	7.990	0.025	23.767	52.161	0.009

For sand, eight dimensionless parameters including D_s/W, φ, α, ω, h'/W, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$ were calculated. However, for clay, seven dimensionless parameters were taken into account for the estimation of the subgouge soil deformation including D_s/W, $c/\gamma_s.W$, α, h'/W, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$. The angle of surcharged soil slope (ω) has not been reported in the literature for clay.

It is worth mentioning that all of the data in the current study comes from centrifuge model testing. The centrifugal acceleration simulates the gravity and allows for correspondence of stress fields between the model and full-scale permitting accurate geotechnical modeling. However, to verify the scale effects, modeling of model tests at different "g" levels is usually conducted to ensure the applicability and accuracy of centrifuge modeling. This has been done in many of the studies referred to in the current paper and resulted in an acceptable accuracy with most of the interaction curves within a bandwidth of fewer than 1.0 times the normalized loads. These results provided sufficient relaxation to construct the current database using the published centrifuge test results.

3.1.19. Goodness of fit

To assess the performance and precision of the developed models, some statistical indices including the correlation coefficient (*R*), the variance accounted for (*VAF*), root mean square error (*RMSE*), mean absolute error (*MAE*), scatter index (*SI*), and Nash-Sutcliff efficiency coefficient (*NSC*) were used by the following equations:

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \quad (3\text{-}50)$$

$$VAF = \left(1 - \frac{var(P_i - O_i)}{var(P_i)}\right) \times 100 \quad (3\text{-}51)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (3\text{-}52)$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \quad (3\text{-}53)$$

$$SI = \frac{RMSE}{\bar{O}} \quad (3\text{-}54)$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (3\text{-}55)$$

Here, *Oi*, *Pi*, $\bar{O}$, $\bar{P}$, and *n* are observed values, predicted values, the mean of observed values, the mean of predicted values, and the number of experimental values, respectively. The closeness of the correlation coefficient (*R*) and the Nash-Sutcliff efficiency coefficient (*NSC*) to one indicates that the model has the highest performance, whilst the closeness of the *RMSE*, *MAE*, and *SI* to zero proves that the particular model has the lowest error. Moreover, the best model has the highest value of the *VAF* index.

3.1.20. Results and Discussion

3.1.20.1. Developed models

A schematic presentation of the applied input combinations for sand and clay seabed is illustrated in Figure 3-25. Model 1 estimated the target parameter (soil deformation) through all of the dimensionless input parameters. However, to identify the most effective input parameters governing the subgouge soil deformation, each dimensionless input parameter was eliminated and

then other models were defined including Model 2 to Model 9 for sand and Model 2 to Model 8 for clay.

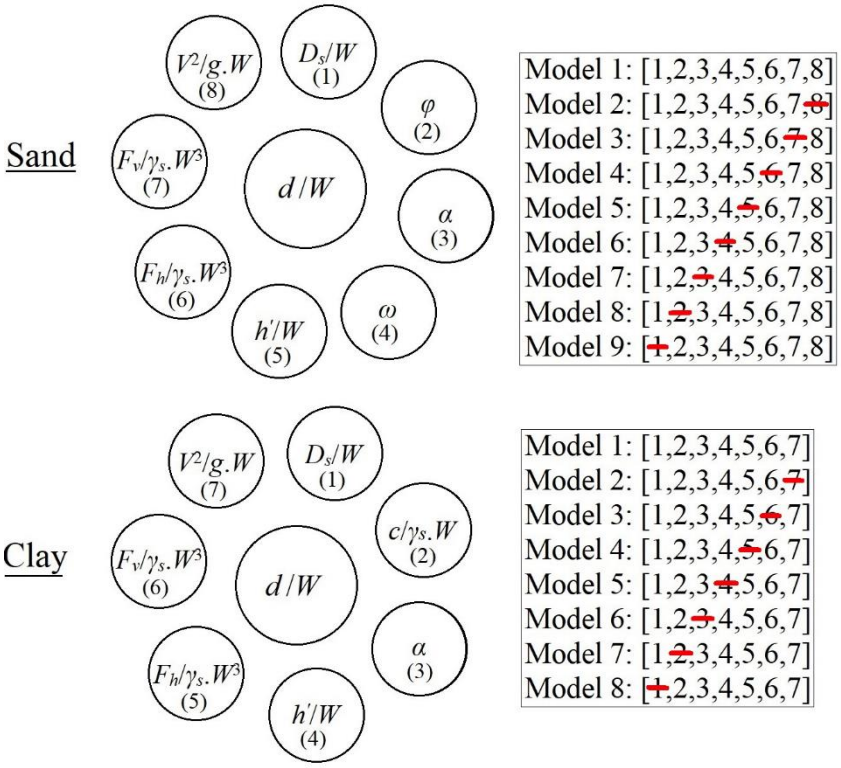


Figure 3-25. A schematic layout of applied developed models for sand and clay seabed

Using the Minitab software, linear regression (LR) was conducted to derive the required equations and introduce the combinations of input parameters. Nine equations for sand and eight equations for clay were obtained to estimate the horizontal (d_h/W) and vertical (d_v/W) subgouge soil deformations. The developed LR models (equations (5-29) to (5-62)) are provided in Appendix I. It is worth mentioning that the LR, which is a linear technique to approximate the relationship between a dependent parameter as a target function and one or more independent parameters as input variables, has been applied in different fields of study in the past (Freedman 2009; Yan and Su 2009; Koç and Barkana 2014; Forbrich et al. 2010).

The simplicity, cost-effectiveness, and low disturbance during the simulation process, are the key advantages of LR technics that have made it a practical method in a range of applications (Bao et al. 2010).

3.1.20.2. Sensitivity analysis in sand

The performance of the developed models for sand seabed was evaluated, meaning that the target parameters (horizontal and vertical deformations) were sensitized to each dimensionless parameter, and the most important parameter and the best models for estimation of the subgouge deformations were introduced. Figure 3-26 compares the calculated statistical indices for the estimated subgouge sand deformations using Model 1 to Model 9.

Model 1 computed the horizontal and vertical deformations through all dimensionless input parameters (D_s/W, φ, α, ω, h'/W, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, $V^2/g.W$). Model 1 with the variance accounted for (*VAF*), the correlation coefficient (*R*), and the scatter index (*SI*), root mean square error (*RMSE*), mean absolute error (*MAE*), and Nash-Sutcliff efficiency coefficient (*NSC*) of 82.974, 0.936, 0.525, 0.012, 0.382, and 0.010, resulting in the highest precision among all of the defined models for sand. The parameters that were eliminated one at a time in Model 2, 3, 4, 5, 6, 7, 8, and 9 were $V^2/g.W$, $L_v/\gamma_s.W^3$, $L_h/\gamma_s.W^3$, h'/W, ω, α, φ, D_s/W, respectively.

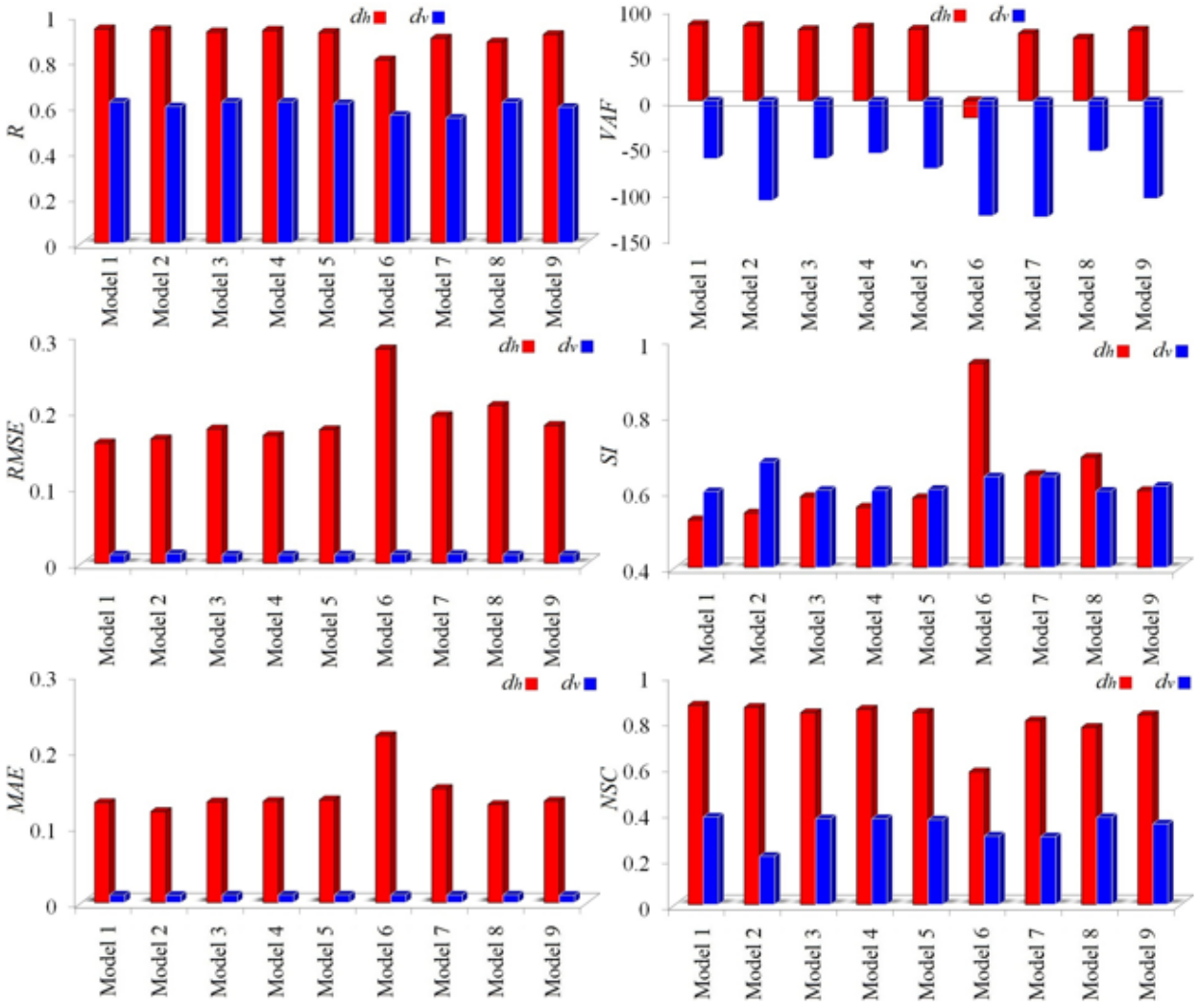


Figure 3-26. Comparison of calculated statistical indices for estimated subgouge sand deformations using Model 1 to Model 9

The error distribution of Model 1 to Model 9 in estimating the subgouge sand deformations is illustrated in 3-27. Based on the presented charts, about 37% of the horizontal deformations estimated by using Model 1 have an error of less than 10%, though 64% of the modeled vertical deformations using Model 1 own an error of less than one percent. Even though 22% of the horizontal deformations simulated adopting Model 1 have an error of more than 20%, almost 4% of the vertical deformations estimated using this model have an error of more than 2%.

Figure 3-27. Error distribution of Model 1 to Model 9 for estimating subgouge sand deformations

The depicted scatter plots for estimated horizontal and vertical sand deformations obtained by Model 1 to Model 9 are shown in Figure 3-27.

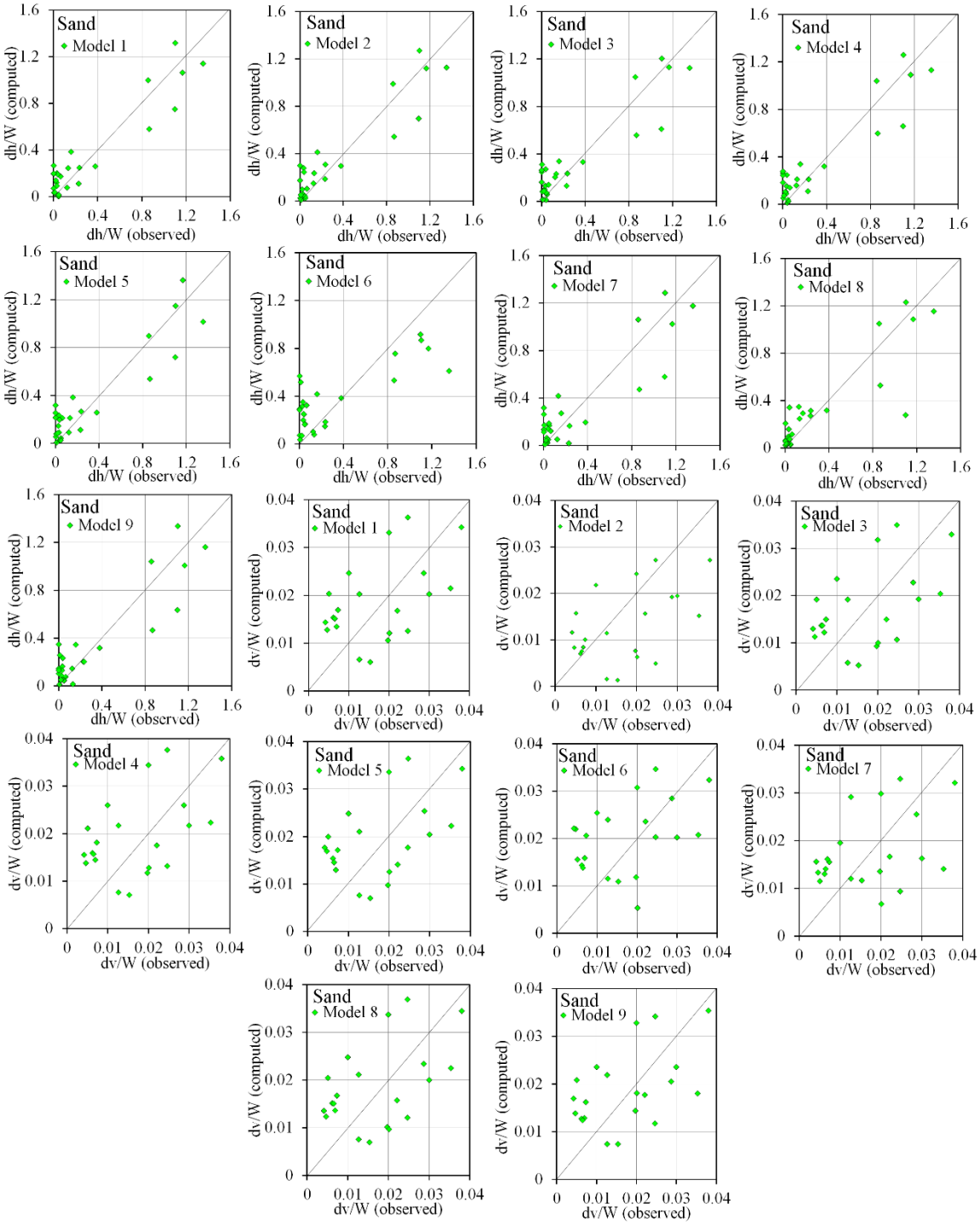


Figure 3-28. Scatter estimated horizontal and vertical subgouge sand deformations by Model 1 to Model 9

Regarding the sensitivity analysis, Model 1 has the highest correlation with the experimental values. Among all developed LR models, Model 1 estimates the horizontal and vertical deformations with the lowest error. Therefore, Model 1 is introduced as the superior model in estimating the subgouge sand deformations. Additionally, Model 2 is detected as the second-best LR model. After Model 1 and Model 2, Model 4 has the highest level of correlation with the experimental measurements. Also, Model 5 and Model 9 in terms of precision and correlation are identified as the fifth and sixth models, respectively. Model 7 and Model 8 are detected as the seventh and eighth-best models among the LR models. Model 6 estimates the subgouge deformations with the lowest level of accuracy and this model has the weakest correlation with the experimental values compared to other LR models.

The internal friction angle of sand (φ) was identified as the most influencing input parameter for estimating the horizontal subgouge sand deformations. Subsequently, the attack angle (α), the ratio of the gouge depth to the gouge width (D_s/W), dimensionless vertical force ($L_v/\gamma_s.W^3$), and dimensionless horizontal force ($L_h/\gamma_s.W^3$) are ascertained as the most effective input parameter in estimating the horizontal subgouge sand deformations, respectively. Furthermore, the attack angle (α) is detected as the most significant input parameter for estimating the vertical subgouge sand deformations. The ratio of the gouge depth to the gouge width (D_s/W), the internal friction angle of sand (φ), dimensionless vertical force ($L_v/\gamma_s.W^3$), and dimensionless horizontal force ($L_h/\gamma_s.W^3$) have the highest impact on the target parameter, respectively.

Evaluation of the dimensionless input parameters reveals that the number of experimental measurements for the surcharged soil slope (ω), the ratio of berm height to the gouge depth (h'/W), and the dimensionless velocity parameter ($V^2/g.W$) are very few; as a result, the implemented sensitivity analysis is not able to identify the level of importance of the ω, h'/W, and $V^2/g.W$ dimensionless groups.

3.1.20.3. Sensitivity analysis in clay

In this section, the subgouge clay deformations estimated by Model 1 to Model 8 are evaluated. The results of the computed statistical indices for the models are depicted in Figure 3-29.

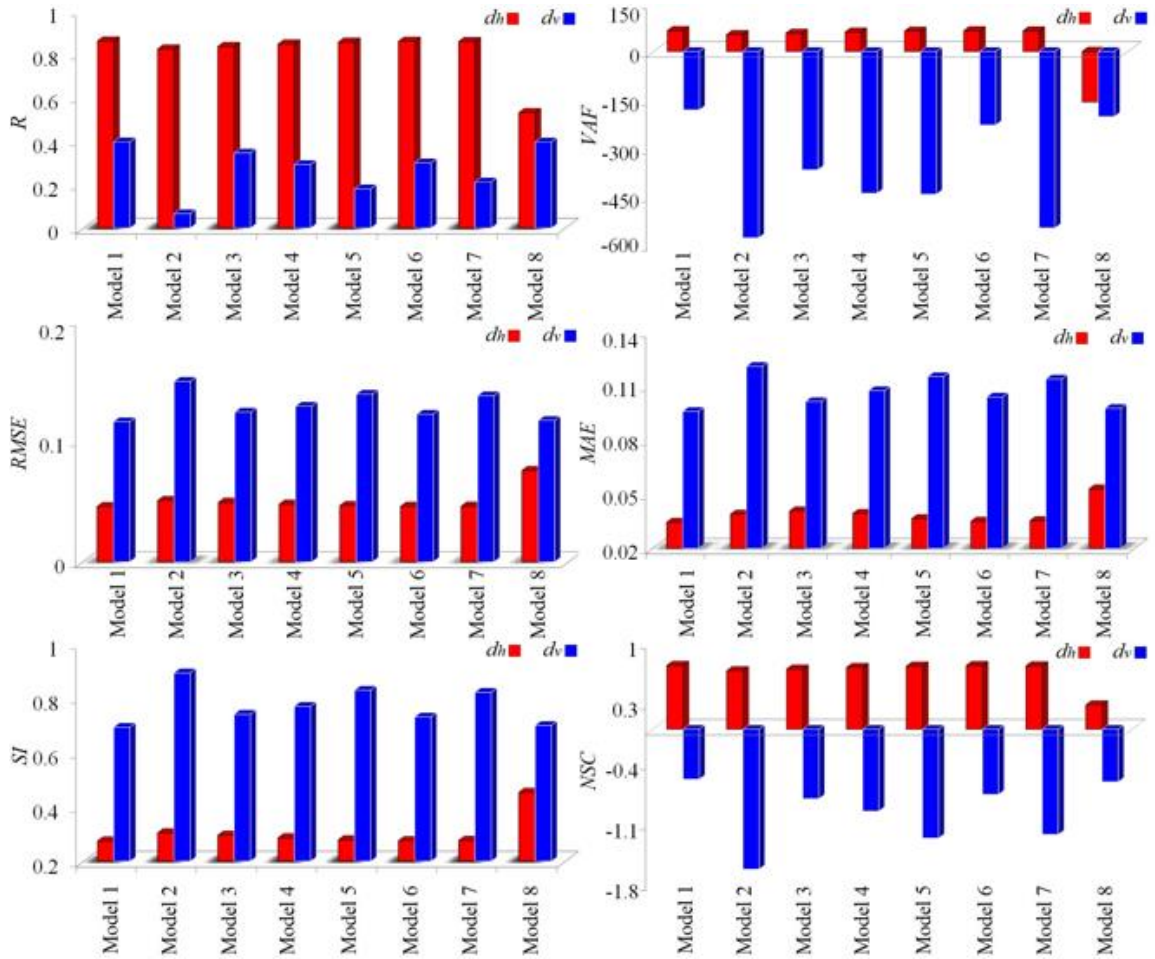


Figure 3-29. Comparison of calculated statistical indices for estimated subgouge clay deformations by Model 1 to Model 8

To estimate the horizontal and vertical subgouge clay deformations, eight distinct models (Model 1 to Model 8) were defined using the dimensionless parameters including D_s/W, $c/\gamma_s.W$,α, h'/W, $L_h/\gamma_s.W^3$, $Lv/\gamma s.W^3$, $V^2/g.W$. Model 1 simulated the target parameter through all dimensionless input parameters with a high level of accuracy in comparison with other LR models. The *NSE*, *VAF*, and *SI* values in Model 1 for estimation of the horizontal subgouge deformations in clay were 0.737, 64.250, and 0.274, respectively. The parameters that were eliminated one at a time in Model 2, 3, 4, 5, 6, 7, and 8 were $V^2/g.W$, $L_v/\gamma_s.W^3$, $L_h/\gamma_s.W^3$, h'/W, α, $c/\gamma_s.W$, D_s/W, respectively. Figure 3-30 shows the error distribution charts of the horizontal and vertical subgouge deformations in clay that were obtained by Model 1 to Model 8.

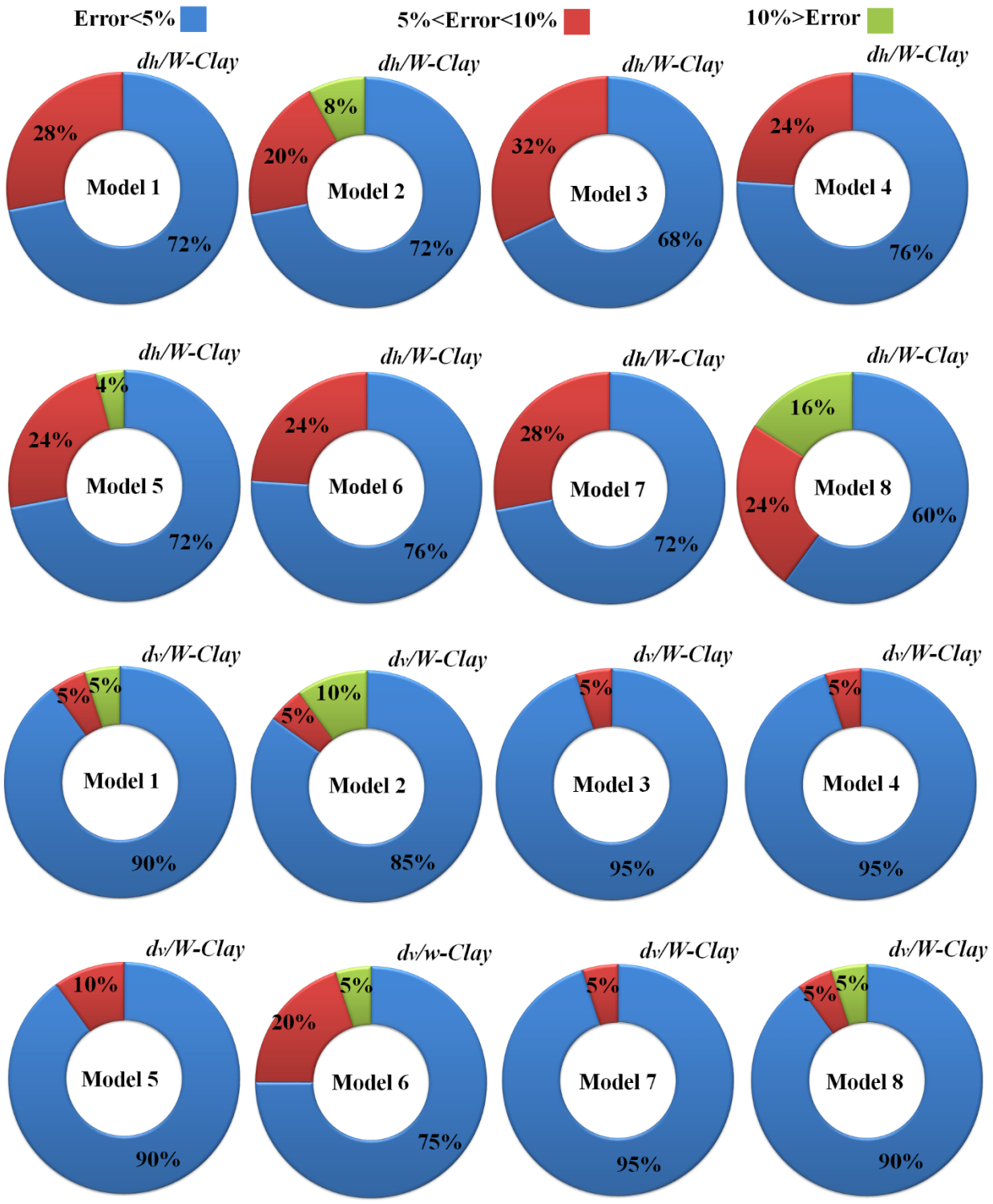


Figure 3-30. Error distribution of Model 1 to Model 8 for estimating the horizontal and vertical subgouge clay deformations

The overwhelming majority of the simulated horizontal deformations obtained by Model 1 have an error of less than 5% and almost 5% of the estimated vertical deformations own an error of more than 10%. In Figure 3-31, the scatter plots for horizontal and vertical ice-induced clay deformations obtained by Model 1 to Model 8 are illustrated.

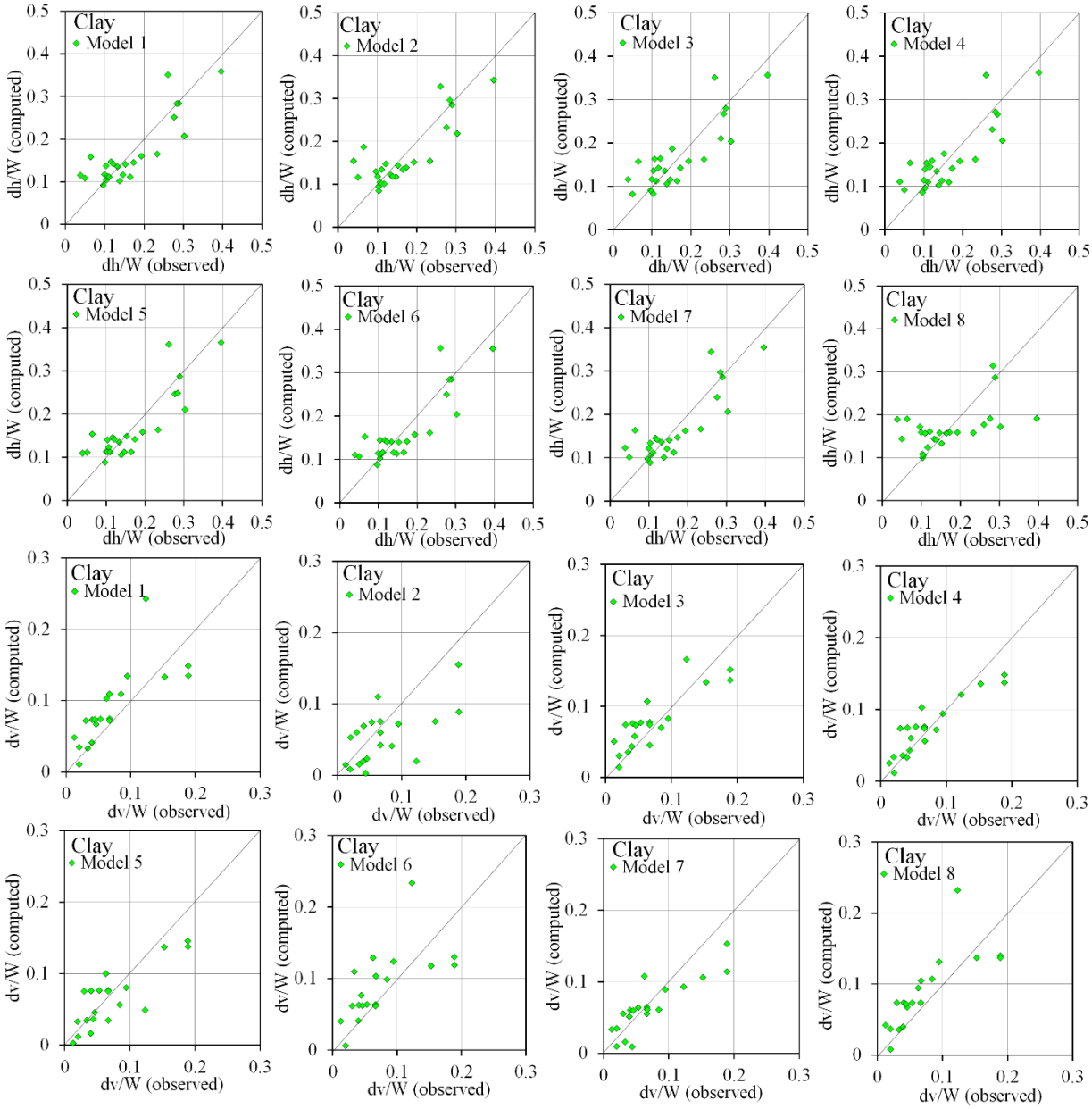


Figure 3-31. Scatter plots of estimated horizontal and vertical subgouge clay deformations by Model 1 to Model 8

Model 1 was identified as the superior model in estimating both the horizontal and vertical subgouge clay deformations. The model was a function of all input parameters having the lowest error and the highest correlation with the experimental values. Furthermore, Model 6, Model 7, and Model 5 were respectively recognized as the second, third, and fourth-best LR models. Moreover, Model 4, Model 3, and Model 2 were identified as the fifth, sixth, and seventh LR models in terms of accuracy and performance. Model 8 was the worst LR model in predicting clay deformations.

The sensitivity analysis indicated that the ratio of the gouge depth to the gouge width (D_s/W) is the most effective input parameter in simulating the horizontal subgouge clay deformations (d_h/W). The dimensionless velocity parameter ($V^2/g.W$), dimensionless vertical force parameter ($L_v/\gamma_s.W^3$), and dimensionless horizontal force parameter ($L_h/\gamma_s.W^3$) were respectively detected as the second, third, and fourth significant input parameters in modeling the horizontal deformations. In contrast, the dimensionless cohesion parameter ($c/\gamma_s.W$) and the attack angle (α) showed a negligible influence on the estimation of the horizontal deformations. The parameter $V^2/g.W$ was detected as the most influencing input parameter in modeling the subgouge vertical deformations in clay. The parameters c/$\gamma_s.W$, $L_h/\gamma_s.W^3$, α, $L_v/\gamma_s.W^3$, and D_s/W were identified as the most important input parameters in the simulation of the horizontal subgouge deformations in clay.

3.1.20.4. Uncertainty analysis (UA)

Uncertainty analysis (UA) was conducted to evaluate the performance of LR models and describe the errors. The errors calculated by LR models (e_j) were computed as the difference between the estimated (P_j) and observed (O_j) subgouge soil deformations $(e_j = P_j - O_j)$. Then, the mean value of the estimated error $(\bar{e})$ and the standard deviation (S_e) were calculated as follows (Crosetto et al. 2000; Azimi et al. 2018; Azimi et al. 2019):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{3-56}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{3-57}$$

The negative sign of $\bar{e}$ means the underestimated performance of the LR model, whereas the positive sign shows the overestimated performance of the model. Using $\bar{e}$ and S_e, confidence bound was formed around an estimated error through the "Wilson score approach" (Wilson 1927; Wallis 2013) without the continuity correction. The Wilson score interval is an improvement over the normal distribution interval in which an asymmetric normal distribution is used to improve the confidence interval bound. Next, ±1.96Se results were expressed in a 95% confidence bound which is a "95% prediction error interval or 95%PEI". Then "width of uncertainty bound or WUB" was calculated. Figure 3-32 shows the results yielded by the UA of the LR models for horizontal and vertical soil deformations. All of the LR models (Model 1 to Model 9) have resulted in an underestimated performance in the estimation of the horizontal subgouge deformations. For predicting the vertical sand deformations, Model 1, Model 4, Model 5, Model 6, Model 8, and Model 9 simulated the target parameter with an underestimated performance, whereas Model 2, Model 3, and Model 7 have an overestimated performance.

To model the horizontal sand deformations, the computed mean value of the standard deviation for Model 1, Model 2, and Model 3 are 0.157, 0.163, and 0.174, respectively. For prognosticating

the horizontal deformations by Model 4, Model 5, and Model 6, the value of S_e were equal to 0.166, 0.175, and 0.280. The mean value of the standard deviation for Model 7, Model 8, and Model 9 were found as 0.196, 0.210, and 0.182, respectively.

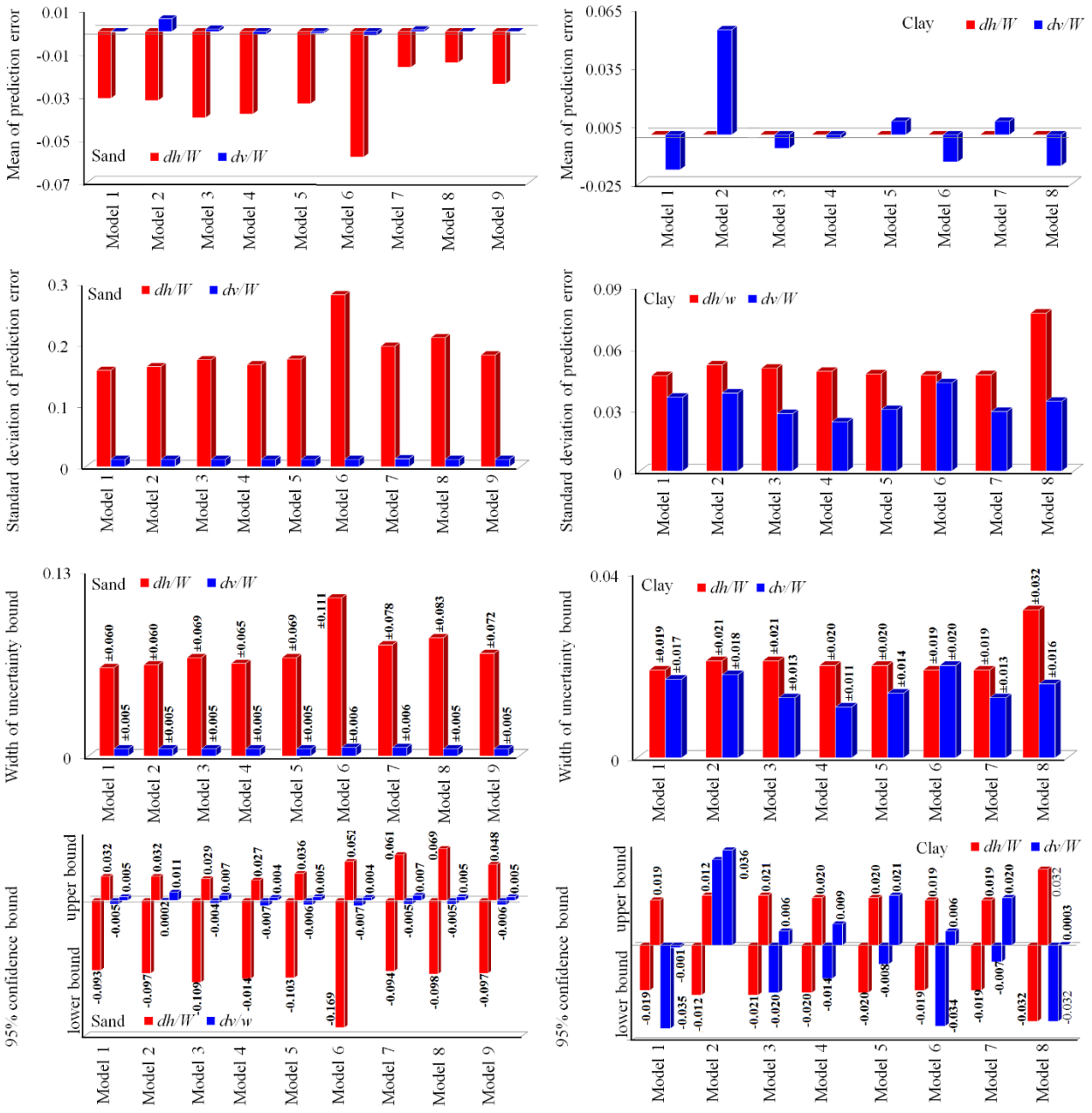


Figure 3-32. Results of uncertainty analysis for LR models

The 95%PEI & WUB for simulating the horizontal sand deformations using Model 1 were (-0.093 to 0.032) & ±0.062. The calculated 95%PEI for Model 2, Model 3, and Model 4 were (-0.097 to 0.032), (-0.109 to 0.029), and (-0.104 to 0.027), whilst the computed WUB values for these LR models are ±0.064, ±0.069, and ±0.065, respectively. The results show that Model 1 owns the narrowest width of the uncertainty bound among these LR models. Also, Figure 3-32 shows that for vertical sand deformations, the narrowest bound belongs to Model 4 and Model 6.

To simulate the horizontal clay deformations, Model 1, Model 2, and Model 7 forecasted the target parameter with an overestimated performance, while Model 3, Model 4, Model 5, Model 6, and Model 8 underestimated the horizontal clay deformations. The vertical clay deformations were underestimated by Model 1, Model 3, Model 4, Model 6, and Model 8, while the other models including Model 2 and Model 7 overestimated the vertical. Model 1 resulted in the narrowest width of uncertainty bound in order to model horizontal clay deformations, while Model 4 had the narrowest width of the uncertainty bound in the estimation of the vertical subgouge deformations among all LR models.

3.1.20.5. Comparison with previous studies

The empirical equations proposed by Woodworth-Lynes et al. (1996) were used to estimate the maximum horizontal (d_h) and vertical (d_v) subgouge clay deformations as follows:

$$d_h = 0.6\sqrt{W.D_s} \tag{3-58}$$

$$d_v = D_s.exp\left(-\frac{1}{3}.\frac{y}{D_s}\right) \tag{3-59}$$

where, d_h, W, D_s, d_v, y are the maximum horizontal subgouge deformation, the width of the gouge, the gouge depth, the maximum vertical subgouge deformation, and the soil depth, respectively. After the normalization of the computed horizontal and vertical deformations by

equations (5-27) and (5-28), a comparison was made between the models proposed by Woodworth-Lynes et al. (1996) and the best LR models (equations (5-29) to (5-32)). The results of the comparison were tabulated in Table 3-10.

Table 3-10. Comparison between empirical relationships and best LR models

Soil	Deformation	Equation	*R*	*VAF*	*RMSE*	*SI*	*MAE*	*NSC*
Sand	Horizontal	(3-58)	0.335	-68.042	2.056	1.677	1.720	-0.147
		(3-60)	0.936	82.974	0.157	0.525	0.131	0.866
	Vertical	(3-59)	-0.222	-2.382	6.388	28.405	3.778	-615.613
		(3-61)	0.618	-63.079	0.012	0.600	0.010	0.382
Clay	Horizontal	(3-58)	0.895	79.647	0.771	0.569	0.411	0.838
		(3-62)	0.858	64.250	0.046	0.274	0.034	0.737
	Vertical	(3-59)	0.718	50.583	1.090	1.550	0.547	-0.218
		(3-63)	0.396	-181.052	0.116	0.693	0.096	-0.578

Table 3-10 shows that Model 1 has resulted in the highest correlation and lowest error in the estimation of both the horizontal and vertical subgouge soil deformations. The values of *R*, *NSC*, and *SI* for Eq. (5-30) were equal to 0.335, -0.147, and 1.677, respectively. Also, for the empirical formula, the *RMSE*, *MAE*, and *VAF* statistical indices were estimated to be 2.056, 1.720, and -65.042. Thus, for predicting the maximum horizontal ice-induced sand deformation, the proposed LR model (Eq. 3-60) showed a better performance than Eq. (3-58). Also, Table 3-10 shows that the performance of the developed LR model for the estimation of the vertical deformations in the sand was better than the empirical formulae. By contrast, the empirical equations better estimated

the horizontal and vertical deformations in clay than the proposed LR models. For instance, the *MAE*, *RMSE*, and *R* indices for Eq. (3-58) were calculated to be 0.411, 0.771, and 0.895. Overall, Model 1, as a function of all dimensionless groups, was introduced as the superior LR model for the simulation of the horizontal and vertical sand deformations. The model forecasted the target parameter accurately with an underestimated performance. The internal friction angle of sand (φ) and the keel attack angle (α) were the most influential dimensionless groups in prognosticating the horizontal and vertical deformations in the sand. For estimation of the subgouge deformation in clay, Model 1 was the premium LR model overestimating the horizontal deformations. However, the LR model showed an underestimated performance in simulating the vertical displacements. Furthermore, the ratio of the gouge depth to the gouge width (Ds/W) had a significant impact on the modeling of the horizontal clay deformations, whilst the dimensionless velocity parameter ($V^2/g.W$) was identified as the most important dimensionless group in simulating the vertical deformations in clay.

3.1.21. Conclusion

A dimensional analysis of the subgouge soil deformation due to an ice-gouging event was performed by using the Buckingham-π theory for the first time. Eight dimensionless groups of key parameters were identified affecting the maximum subgouge soil deformations. Subsequently, using the identified dimensionless groups and developing a comprehensive dataset, nine linear regression (LR) models for the sandy seabed and eight models for the clay seabed were developed. The most significant results of the study are summarized as follows:

- The defined LR models were validated with experimental measurements, signifying that the models estimated the maximum horizontal and vertical subgouge deformations with good precision.
- According to the sensitivity analysis, Model 1 was the best LR model predicting the horizontal and vertical deformations by using all dimensionless input groups including D_s/W, φ,α, ω, h'/W, $F_h/\gamma_s.W^3$, $F_v/\gamma_s.W^3$, $V^2/g.W$. This model had an acceptable level of accuracy, meaning that approximately 32% of the simulated vertical sand deformations showed an error of less than 2%.
- The implemented sensitivity analysis identified the internal friction angle of sand (φ) and the attack angle (α) as the most effective dimensionless groups to estimate the horizontal and vertical subgouge deformations.
- The best LR model (Model 1) simulating the clay deformations was a function of the D_s/W, $c/\gamma_s.W$, α, h'/W, $F_h/\gamma_s.W^3$, $F_v/\gamma_s.W^3$, $V^2/g.W$.
- The ratio of the gouge depth to the gouge width (D_s/W) and the dimensionless velocity parameter ($V^2/g.W$) was respectively recognized as the most influencing dimensionless groups for modeling the horizontal and vertical subgouge clay deformations.
- The uncertainty analysis showed that Model 1 had an underestimated performance in estimating the sand deformations.
- The comparison of the results of the best LR models with the empirical models showed that the LR models predicted the sand deformations more accurately, whilst the empirical models estimated the clay deformations better than the LR models.

The presented dimensional analysis provided a good insight into the key parameter governing ice-induced soil deformations. The results obtained in this study can not only facilitate proposing new

analytical or empirical solutions for the prediction of the subgouge soil deformation but also be beneficial to the planning of costly experimental studies and numerical simulations and mitigate the expenses of future studies.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now InnovateNL)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

Appendix I

The developed linear regression (LR) models (equations (5-29) to (5-62)) are tabulated in Table 3-11.

Table 3-11. Derived equations based on linear regression (LR) using different input combinations

Model	Soil	Relationship	Eq. #
1	Sand	$\|dh/W\|=1.490+2.121(Ds/W)-0.036(\varphi)-0.019(\alpha)+0.030(\omega)-0.466(h'/W)-0.388(Fh/\gamma s.W3)+0.432(Fv/\gamma s.W3)+11.597(V2/g.W)$	(3-60)
		$\|dv/W\|=0.020+0.069(Ds/W)+ 0.0002(\varphi)- 0.00081(\alpha)- 0.0005(\omega)+ 0.008(h'/W)-0.001(Fh/\gamma s.W3)+ 0.0008(Fv/\gamma s.W3)+ 0.727(V2/g.W)$	(3-61)

	Clay	$\|dh/W\|=0.018+1.053(Ds/W)+0.004(c/\gamma s.W)-0.001(\alpha)+1.173(h'/W)+0.004(Fh/\gamma s.W3)-0.003(Fv/\gamma s.W3)+3.333(V2/g.W)$	(3-62)
		$\|dv/W\|=0.174+0.056(Ds/W)-0.008(c/\gamma s.W)-0.003(\alpha)-0.557(h'/W)-0.0009(Fh/\gamma s.W3)+0.002(Fv/\gamma s.W3)-3.932(V2/g.W)$	(3-63)
2	Sand	$\|dh/W\|=1.349+1.868(Ds/W)-0.029(\varphi)-0.020(\alpha)+0.031(\omega)-0.375(h'/W)-0.351(Fh/\gamma s.W3)+0.357(Fv/\gamma s.W3)$	(3-64)
		$\|dv/W\|=-0.006+0.037(Ds/W)+0.001(\varphi)-0.0008(\alpha)- 0.0004(\omega)+0.012(h'/W)+ 0.013(Fh/\gamma s.W3)- 0.012(Fv/\gamma s.W3)$	(3-65)
	Clay	$\|dh/W\|=0.090+0.814(Ds/W)-0.003(c/\gamma s.W)-0.001(\alpha)+2.033(h'/W)+0.004(Fh/\gamma s.W3)-0.002(Fv/\gamma s.W3)$	(3-66)
		$\|dv/W\|=0.087+0.339(Ds/W)-0.001(c/\gamma s.W)-0.003(\alpha)-1.316(h'/W)-0.002(Fh/\gamma s.W3)+0.001(Fv/\gamma s.W3)$	(3-67)
3	Sand	$\|dh/W\|=1.150+1.645(Ds/W)-0.027(\varphi)-0.014(\alpha)+0.031(\omega)-0.354(h'/W)+0.115(Fh/\gamma s.W3)+6.297(V2/g.W)$	(3-68)
		$\|dv/W\|=0.019+0.067(Ds/W)+0.0002(\varphi)-0.0008(\alpha)-0.0005(\omega)+0.008(h'/W)-0.0005(Fh/\gamma s.W3)+0.707(V2/g.W)$	(3-69)
	Clay	$\|dh/W\|=0.019+1.032(Ds/W)+0.001(c/\gamma s.W)-0.0004(\alpha)+0.266(h'/W)-0.001(Fh/\gamma s.W3)+3.104(V2/g.W)$	(3-70)
		$\|dv/W\|=0.174+0.069(Ds/W)-0.007(c/\gamma s.W)-0.003(\alpha)-0.044(h'/W)+0.002(Fh/\gamma s.W3)-3.810(V2/g.W)$	(3-71)
	Sand	$\|dh/W\|=1.318+1.565(Ds/W)-0.032(\varphi)-0.016(\alpha)+0.029(\omega)-0.426(h'/W)+0.144(Fv/\gamma s.W3)+10.092(V2/g.W)$	(3-72)

4		$\|dv/W\|=0.018+ 0.067(Ds/W)+ 0.0003(\varphi)- 0.0008(\alpha)- 0.0005(\omega)+0.008(h'/W)-0.0003(Fv/\gamma s.W3)+0.702$ (V2/g.W)	(3-73)
	Clay	$\|dh/W\|=0.007+1.085(Ds/W)+0.004(c/\gamma s.W)-0.0004(\alpha)+0.360(h'/W)- 0.001(Fh/\gamma s.W3)+3.547$(V2/g.W)	(3-74)
		$\|dv/W\|=0.178+ 0.0458(Ds/W)-0.008(c/\gamma s.W)-0.003(\alpha)- 0.356(h'/W)+0.001(Fv/\gamma s.W3)-4.020$(V2/g.W)	(3-75)
5	Sand	$\|dh/W\|=1.430+2.426(Ds/W)-0.034(\varphi)-0.021(\alpha)+0.027(\omega)- 0.316(Fh/\gamma s.W3)+0.317(Fv/\gamma s.W3)+5.022$(V2/g.W)	(3-76)
		$\|dv/W\|=0.024+0.068(Ds/W)+8.816E-005(\varphi)-0.0008(\alpha)-0.0004(\omega)- 0.004(Fh/\gamma s.W3)+0.004(Fv/\gamma s.W3)+ 0.874$(V2/g.W)	(3-77)
	Clay	$\|dh/W\|=0.003+1.111(Ds/W)+0.007(c/\gamma s.W)- 0.0004(\alpha)+0.003(Fh/\gamma s.W3)-0.002(Fv/\gamma s.W3)+3.703$(V2/g.W)	(3-78)
		$\|dv/W\|=0.182+0.027(Ds/W)-0.010(c/\gamma s.W)-0.003(\alpha)- 0.0004(Fh/\gamma s.W3)+ 0.001(Fv/\gamma s.W3)-4.112$(V2/g.W)	(3-79)
6	Sand	$\|dh/W\|=2.228+0.589(Ds/W)-0.045(\varphi)-0.038(\alpha)-0.078(h'/W)- 0.256(Fh/\gamma s.W3)+0.554(Fv/\gamma s.W3)+21.884$(V2/g.W)	(3-80)
		$\|dv/W\|=0.006+0.081(Ds/W)+0.0004(\varphi)-0.0004(\alpha)+ 0.004(h'/W)- 0.006(Fh/\gamma s.W3)+0.0001(Fv/\gamma s.W3)+ 0.603$(V2/g.W)	(3-81)
	Clay	$\|dh/W\|=0.004+1.057(Ds/W)+0.004(c/\gamma s.W)+1.069(h'/W)+0.003(Fh/\gamma s.W3)-0.003(Fv/\gamma s.W3)+3.444$(V2/g.W)	(3-82)
		$\|dv/W\|=0.114+0.044(Ds/W)- 0.006(c/\gamma s.W)-0.982(h'/W)- 0.001(Fh/\gamma s.W3)+0.002(Fv/\gamma s.W3)-3.571$(V2/g.W)	(3-83)

7	Sand	$\|dh/W\|=0.930+2.259(Ds/W)-0.032(\varphi)+0.036(\omega)-0.547(h'/W)-0.126(Fh/\gamma s.W3)+0.172(Fv/\gamma s.W3)+13.808(V2/g.W)$	(3-84)
		$\|dv/W\|=-0.012+0.051(Ds/W)+0.0007(\varphi)-0.0002(\omega)+0.008(h'/W)+0.007(Fh/\gamma s.W3)-0.010(Fv/\gamma s.W3)+0.692(V2/g.W)$	(3-85)
	Clay	$\|dh/W\|=0.035+0.998(Ds/W)-0.001(\alpha)+1.759(h'/W)+0.004(Fh/\gamma s.W3)-0.002(Fv/\gamma s.W3)+2.887(V2/g.W)$	(3-86)
		$\|dv/W\|=0.131+0.199(Ds/W)-0.003(\alpha)-1.882(h'/W)-0.001(Fh/\gamma s.W3)+0.001(Fv/\gamma s.W3)-2.702(V2/g.W)$	(3-87)
8	Sand	$\|dh/W\|=0.450+0.691(Ds/W)-0.0156(\alpha)+0.032(\omega)-0.400(h'/W)-0.086(Fh/\gamma s.W3)+0.106(Fv/\gamma s.W3)-6.795(V2/g.W)$	(3-88)
		$\|dv/W\|=0.026+0.075(Ds/W)-0.0008(\alpha)-0.0005(\omega)+0.008(h'/W)-0.005(Fh/\gamma s.W3)+0.003(Fv/\gamma s.W3)+0.820(V2/g.W)$	(3-89)
	Clay	$\|dh/W\|=0.209-0.015(c/\gamma s.W)-0.001(\alpha)+4.319(h'/W)+0.006(Fh/\gamma s.W3)-0.002\ (Fv/\gamma s.W3)-2.283(V2/g.W)$	(3-90)
		$\|dv/W\|=0.184-0.009(c/\gamma s.W)-0.003(\alpha)-0.411(h'/W)-0.0008(Fh/\gamma s.W3)+0.002(Fv/\gamma s.W3)-4.274(V2/g.W)$	(3-91)
9	Sand	$\|dh/W\|=1.403-0.026(\varphi)-0.020(\alpha)+0.027(\omega)-0.545(h'/W)-0.134(Fh/\gamma s.W3)+0.307(Fv/\gamma s.W3)+6.914(V2/g.W)$	(3-92)
		$\|dv/W\|=-0.0005+0.001(\varphi)-0.0007(\alpha)-0.0005(\omega)+0.008(h'/W)+0.014(Fh/\gamma s.W3)-0.010(Fv/\gamma s.W3)+0.335(V2/g.W)$	(3-93)

References

Alba, J.L., 2015. Ice Scour and Gouging Effects with Respect to Pipeline and Wellhead Placement and Design. Bureau of Safety and Environmental Enforcement (BSEE).

Azimi, H., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., Khoshbin, F., 2018. Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length. Acta Mechanica, 229(3): 1197-1214.

Azimi, H., Bonakdari, H., Ebtehaj, I., 2021. Gene expression programming-based approach for predicting the roller length of a hydraulic jump on a rough bed. ISH Journal of Hydraulic Engineering, 27(sup1): 77-87.

Bao, F., Zhou, G., Wang, F., Sui, X., 2010. Partitioning soil respiration in a temperate desert steppe in Inner Mongolia using exponential regression method. Soil Biology and Biochemistry, 42(12): 2339-2341.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour displacement in clays from physical model tests. In 2008 7th International Pipeline Conference, pp. 239-245.

Bekker, M.G., 1956. Theory of land locomotion.

Bland, J.M., Peacock, J.L., 2002. Interpreting statistics with confidence. The Obstetrician & Gynaecologist, 4(3): 176-180.

Box, J.F., 1987. Guinness, gosset, fisher, and small samples. Statistical science, 45-52.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2.

C-CORE., 1996. PRISE Phase 3c: Extreme lee Gouge Event - Modeling and Interpretation, C-CORE Report 96-C32.

C-CORE., 2009. PIRAM - Gouge Morphology Study, C-CORE Report R -09-013-490 v1.

C-CORE., 2009. PIRAM - Draft Final Report, C-CORE Report R -09-019-490 v1.

Chari, T., Allen, J., 1974. An analytical model and laboratory tests on iceberg sediment interaction. In Ocean'74-IEEE International Conference on Engineering in the Ocean Environment, pp. 133-136.

Chari, T.R., 1975. Some geotechnical aspects of iceberg grounding. Doctoral dissertation, Memorial University of Newfoundland.

Chari, T.R.A., 1980. Model study of iceberg scouring in the North Atlantic. Journal of Petroleum Technology, 32(12): 2-247.

Comfort, G., Graham, B., 1986. Evaluation of Sea Bottom Ice Scour Models. Environmental Studies Revolving Funds, No. 37.

Crosetto, M., Tarantola, S., Saltelli, A., 2000. Sensitivity and uncertainty analysis in spatial modelling based on GIS. Agriculture, ecosystems & environment, 81(1): 71-79.

Duplenskiy, S., 2012. Protection of subsea pipelines against ice ridge gouging in conditions of substantial surface ice, Master's thesis, University of Stavanger, Norway.

Duplenskiy, S., Gudmestad, O.T., 2013. Draft: Protection of Subsea Pipelines against Ice Ridge Gouging on Conditions of Substantial Surface Ice. Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering.

Emmerson, C., Lahn, G., 2012. Arctic opening: Opportunity and risk in the high north.

Forbrich, I., Kutzbach, L., Hormann, A., Wilmking, M., 2010. A comparison of linear and exponential regression for estimating diffusive CH4 fluxes by closed-chambers in peatlands. Soil Biology and Biochemistry, 42(3): 507-515.

Freedman, D.A., 2009. Statistical models: theory and practice. cambridge university press.

Hynes, F., 1996. Centrifuge modelling of ice scour in sand. Doctoral dissertation, Memorial University of Newfoundland.

Janosi, Z., 1961. The analytical determination of drawbar pull as a function of slip for tracked vehicles in defarmable soils. Proc. of the 1st Int. Conf. of Terrain-Vehicle Systems.

Kenny, S., Bruce, J., King, T., McKenna, R., Nobahar, A., Phillips, R., 2004. Probabilistic design methodology to mitigate ice gouge hazards for offshore pipelines. IPC04-0527. In Proceedings of International Pipeline Conference, 4-8.

Kenny, S., Jukes, P., 2017. Resource Development in Arctic Regions. Encyclopedia of Maritime and Offshore Engineering, 1-21.

Koç, M., Barkana, A. Application of Linear Regression Classification to low-dimensional datasets. Neurocomputing, 131 (2014):331-335.

Konuk, I. Arctic Pipeline Design Challenges and Current Practices: Ice Scour. In ASME 2009 28th International Conference on Ocean, Offshore and Arctic Engineering (pp. 155-162). American Society of Mechanical Engineers. (2009).

Lach, P.R., 1996. Centrifuge modelling of large soil deformation due to ice scour. Doctoral dissertation, Memorial University of Newfoundland.

Lanan, G.A., Cowin, T. G., Johnston, D. K., 2011. Alaskan Beaufort Sea pipeline design, installation and operation. Offshore Technology Conference.

Moore, P., King, T., Sonnichsen, G., 2011. Iceberg Scout Rise-Up Modeling. In Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Condition, No. POAC11-099.

Myers, J.L., Well, A. D., Lorch, J.R., 2010. Developing the fundamentals of hypothesis testing using the binomial distribution. Research Design and Statistical Analysis, 65-90.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Nessim, M.A., Jordaan, I.J., 1985. Protection of arctic submarine pipelines against ice scour.

Nobahar, A., 2003. Effects of soil spatial variability on soil-structure interaction. Doctoral dissertation, Memorial University of Newfoundland.

Nobahar, A., Kenny, S., Phillips, R., 2007a. Buried pipelines subject to subgouge deformations. International Journal of Geomechanics, 7(3): 206-216.

Nobahar, A., Kenny, S., King, T., McKenna, R., Phillips, R., 2007b. Analysis and design of buried pipelines for ice-gouging hazard: a probabilistic approach. Journal of Offshore Mechanics and Arctic Engineering, 129(3): 219-228.

Paulin, M.J., 1991. Preliminary results of physical model tests of ice scour. Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering.

Paulin, M.J., 1992. Physical model analysis of iceberg scour in dry and submerged sand. Doctoral dissertation, Memorial University of Newfoundland.

Prasad, K.S.R., Chari, T.R., 1986. Some factors influencing iceberg scour estimates. Journal of Energy Resources Technology, 108(3): 234-239.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip line field solutions as an approach to understand ice subgouge deformation patterns. In The Sixteenth International Offshore and Polar Engineering Conference. ISOPE-I-06-289.

Schuring, D.J., Emori, R.I., 1965. Soil deforming processes and dimensional analysis. SAE Transactions, 485-494.

Wallis, S., 2013. Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. Journal of Quantitative Linguistics 20(3): 178-208.

Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association, 22(158): 209-212.

Woodworth-Lynes, C., Nixon, D., Phillips, R., Palmer, A. 1996. Subgouge deformations and the security of Arctic marine pipelines. Offshore Technology Conference, OTC-8222-MS.

Yan, X., Su, X., 2009. Linear regression analysis: theory and computing. World Scientific.

Yang, W., 2009. Physical modeling of subgouge deformations in sand. Doctoral dissertation, Memorial University of Newfoundland.

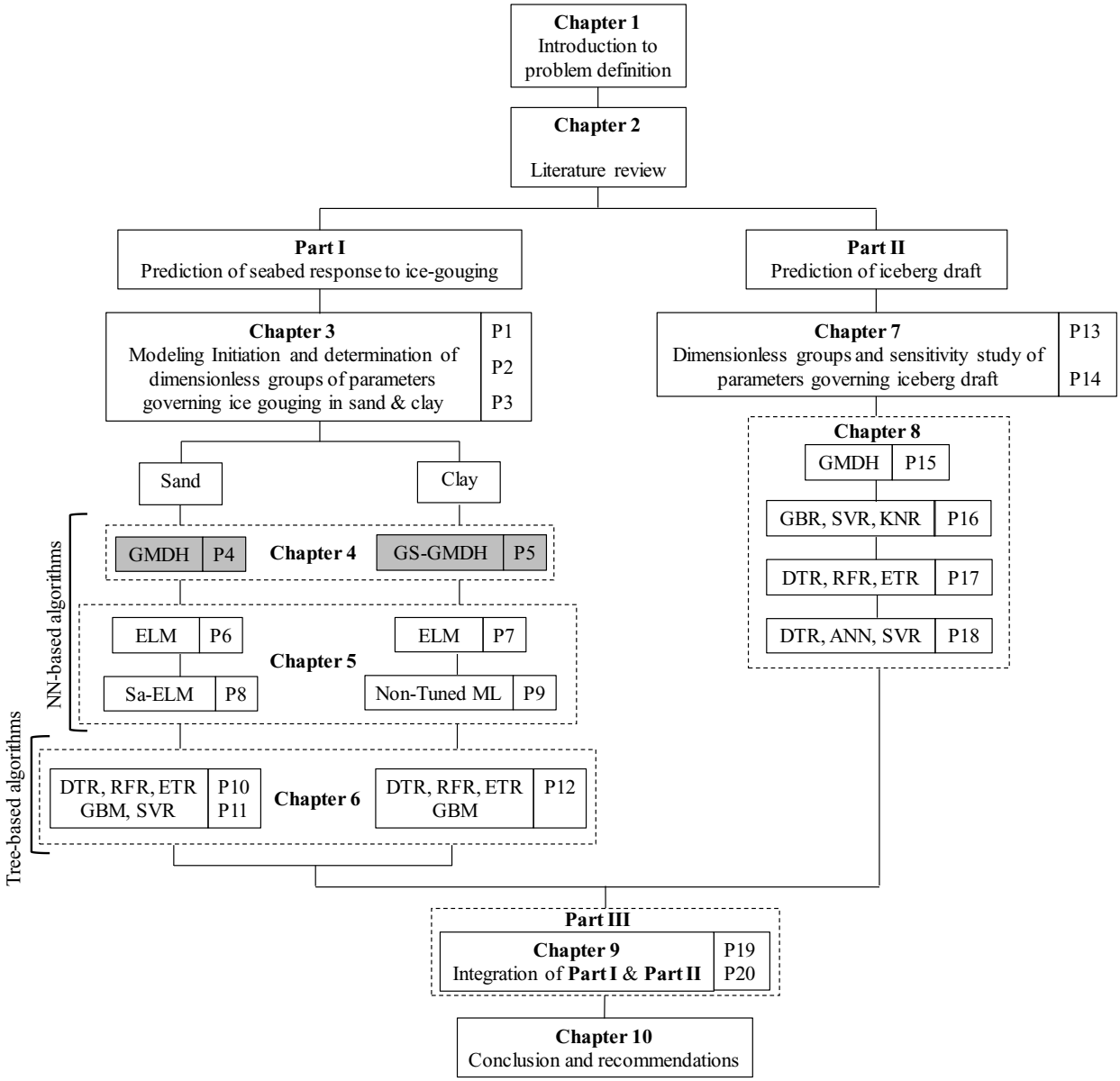
4. Chapter 4

Prediction of Iceberg-Seabed Interaction Process Using Group Method of Data Handling-Based Algorithms

This chapter includes two sections as follows:

Section 1: Prediction of Ice-induced Subgouge Soil Deformation in Sand using Group Method of Data Handling-Based Neural Network

Section 2: Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling


Chapter 1
Introduction to problem definition
Chapter 2
Literature review
Part I
Prediction of seabed response to ice-gouging
Part II
Prediction of iceberg draft
Chapter 3
Modeling Initiation and determination of dimensionless groups of parameters governing ice gouging in sand & clay
P1
P2
P3
Chapter 7
Dimensionless groups and sensitivity study of parameters governing iceberg draft
P13
P14
Chapter 8
GMDH P15
GBR, SVR, KNR P16
DTR, RFR, ETR P17
DTR, ANN, SVR P18
Sand
Clay
NN-based algorithms
GMDH P4
Chapter 4
GS-GMDH P5
ELM P6
ELM P7
Chapter 5
Sa-ELM P8
Non-Tuned ML P9
Tree-based algorithms
DTR, RFR, ETR GBM, SVR P10 P11
Chapter 6
DTR, RFR, ETR GBM P12
Part III
Chapter 9
Integration of Part I & Part II
P19
P20
Chapter 10
Conclusion and recommendations

Section 1

Prediction of Ice-induced Subgouge Soil Deformation in Sand using Group Method of Data Handling-Based Neural Network

This section is a journal paper which has been published in the Journal of Cold Regions Engineering (ASCE) (2023), Vol. 37, pp. 04023006, ISSN: 0887-381X

Abstract

Ice-gouging is one of the critical threats to the subsea pipelines crossing the Arctic and neighboring shallow waters. The burial of subsea pipelines is considered a viable solution to protect them against ice-gouging attacks. The pipeline is usually buried below the deepest recorded ice gouges in that specific geographical location but is still threatened by subgouge soil deformation that is extended down the ice tip due to the shear resistance of the seabed soil. Determination of the subgouge soil deformations is a challenging design aspect that usually requires costly experimental and numerical studies to ensure the structural integrity of the buried pipeline against ice-gouging. In this paper, an alternative and cost-effective methodology has been proposed using the Group Method of Data Handling (GMDH) to simulate the horizontal and vertical subgouge soil deformation profiles in the sandy seabed. Ten GMDH models (GMDH 1 to GMDH 10) were defined by using the governing input parameters comprising the soil depth ratio, the gouge depth ratio, the shear strength of seabed soil, the attack angle, the frontal berm height ratio, the horizontal and vertical loads, and the ice dynamics. The results from the best GMDH models were compared to the artificial neural network (ANN) and empirical approaches showing a robust performance.

Keywords: Ice-gouging analysis; Subgouge soil deformation; Group method of data handling (GMDH); Artificial Neural network (ANN); Sensitivity analysis (SA); Uncertainty analysis (UA)

4.1.1. Introduction

The Arctic area is considered one of the rich regions in terms of energy resources that hold a massive amount of hydrocarbon deposits. Regarding the U.S. Geological Survey (USGS), approximately 13% of the world's undiscovered crude oil (90 billion barrels) and 30% of the world's undiscovered natural gases (47.3 trillion cubic meters) are situated in this cold region (Mouawad 2008). In contrast, the vast majority of the extracted natural resources are transferred using subsea pipelines from offshore to onshore (Alba 2015). During warmer seasons, the seabed may be gouged by traveling icebergs in shallow waters. The soil displacement by ice scour is not limited to the soil body that is in direct contact with the ice keel or front of it. The shear resistance in the soil causes the soil displacement to be extended much deeper than the ice tip which is called subgouge soil deformation (see Figure 4-1).

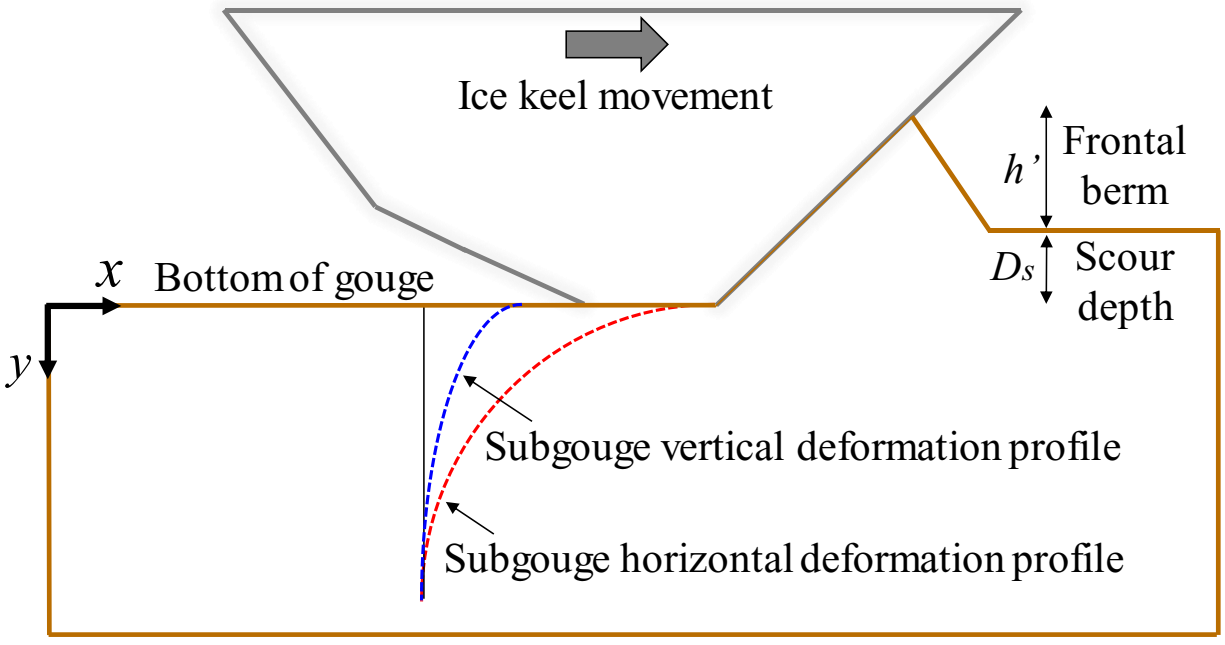


Figure 4-1. A schematic layout of deformations of the subgouge seabed and buried subsea pipeline during an ice-gouging event

Therefore, the subgouge soil deformation may still threaten the pipes that are located deeper than the ice gouge depth or below the ice tip. The subgouge soil deformation profile has a hyperbolic shape. The maximum horizontal deformation occurs directly beneath the bottom of the ice keel and the soil depth for horizontal displacement is measured from the bottom of the ice keel. The subgouge soil displacement may endanger the pipeline integrity in terms of the serviceability limit state (e.g., ovalization) or ultimate limit state (e.g., local buckling or tensile fracture) (Lach 1996, Nematzadeh and Shiri 2020). Figure 4-1 portrays the schematic layout of an ice-gouging event and the resultant subgouge soil deformation.

Commonly, the estimation of the minimum burial depth for pipeline protection against ice-gouging is a challenging design task that can have a significant time and cost impact on the project. In practice, for the evaluation of the ice gouge impact on buried pipelines, a decoupled approach is followed by combining the accuracy of the continuum approach for finite element analysis of the free field ice-gouging event with the simplicity of the beam-spring method for modeling the pipeline. The decoupled approach starts with a free-field ice-gouging analysis without the presence of a pipeline. Then the observed subgouge soil deformations are manually transferred to the springs representing the soil in a beam-spring model to obtain the pipe response to soil movement. The decoupled approach assumes that the soil displacement is completely transferred to the pipe. Therefore, the decoupled approach does not model the slip between the pipe and soil and also the soil's local compression in contact with the pipe. The beam-spring model examines the soil displacements in various depths that have been obtained from a single free-field finite element analysis to come up with the best burial depth. Therefore, free-field ice-gouging is a significant part of the decoupled approach in practical pipeline design against an ice attack that is usually supported by costly computational and experimental studies. The current study introduces a neural

network-based (NN-based) method for fast estimation of the subgouge soil deformations that need to be fed into the beam-spring model. This can be a cost-effective alternative for continuum finite element models that are usually used in the free-field ice-gouging analysis. The proper data is a key object in the success of NN-based analysis. There are a large number of experimental, e.g., (Paulin 1992; Schoonbeek et al. 2008; Been 2008), analytical, e.g., (Nessim and Jordaan 1985, Comfort and Graham 1986), and numerical, e.g., (Nematzadeh and Shiri 2019a, Nematzadeh and Shiri 2019b, Nematzadeh and Shiri 2020) studies in the literature. However, in this paper, the database for training and testing of the NN-based models was extracted from a set of crucial experimental studies published in the literature including Paulin (1992), C-CORE (1995), Hynes (1996), C-CORE (1996), and Yang (2009).

Paulin (1992) performed four centrifuge tests to study the subgouge deformation in the sand under submerged and dry conditions. The authors investigated the surface deformations, sub-scour displacements, pressure, and forces and concluded that the measured loads and forces for the submerged models were less than the dry tests.

Later in 1995-1996, several free field centrifuge laboratory tests of the keel scour problems in various seabed conditions such as loose sand, dense sand, soft clay, and medium clay were implemented in the Centre for Cold Ocean Resources Engineering (C-CORE 1995 and C-CORE 1996). The subgouge soil deformation profiles were measured and also a set of empirical formulas were derived from the experimental measurements to approximate the subgouge soil deformations.

Hynes (1996) carried out eight pressure ridge ice scour experiment (PRISE) tests to measure the horizontal and vertical loads, the gouge-induced pore pressures and stresses, and the keel-soil interface pressures for the sandy seabed. The author proposed a linear relationship between the scour loads and the scour depth based on the shear-dragging theory. Yang (2009) studied the ice-

induced deformations on a sandy seabed by performing seven tests through the pipeline ice risk assessment and mitigation program (PIRAM). The study resulted in proposing a set of equations for the relationship between the horizontal and vertical subgouge deformations, the bearing pressure, the gouge reaction force, the frontal berm height (the height of soil heave in front of the ice keel), and the gouge depth.

NN-based models have been widely applied to simulate different non-linear and complex problems since they are reliable, quick, inexpensive, and accurate (Kaydani et al. 2014, Nazari 2015). There are only a few neural network-based studies that have simulated the ice-gouging event. Kioka et al. (2003) and Kioka et al. (2004) modeled characteristics of the ice scour problem by the neural network (NN) approach. The authors have not studied all parameters affecting a real ice-gouging problem, and the keel failure was examined using a simple model. Additionally, this NN model was just validated by the analytical results.

Azimi and Shiri (2020a) performed a dimensional analysis using Buckingham's theory to identify the parameters affecting the ice-seabed interaction problems for the first time. The authors suggested several linear models (LMs) estimating the maximum ice-induced soil deformations. Moreover, a comprehensive dataset comprising the experimental measurements in sand and clay seabed was constructed, and the LMs were verified by them.

Azimi and Shiri (2020b) simulated the horizontal subgouge sand deformations using gene expression programming (GEP). The authors compared the GEP with a multilayer perceptron (MLP) model, with better performance of the GEP than MLP. The study concluded that the soil depth and the dilation index had a significant impact on subgouge sand deformations. Azimi and Shiri (2021a) simulated the ice-seabed interaction parameters utilizing the extreme learning machine (ELM) model. The gouge depth ratio and the shear strength of the seabed soil were

recognized as the most important input parameters. The multi-layer perceptron neural network was applied to estimate the horizontal subgouge sand deformations by Azimi and Shiri (2021b). The author concluded that the soil depth, bearing pressure, and dilation index were the most significant variables to predict the subgouge sand characteristics.

Among the NN-based models, the Group Method of Data Handling (GMDH) is a robust, fast, and reliable method that has been widely applied in various fields (Ebtehaj et al. 2015, Azimi et al. 2018, Walton et al. 2019). Compared with the classical NN models, the GMDH has several advantages as follows:

- The GMDH is capable of choosing the most influential input parameters and provision of an explicit relationship for the iceberg-seabed interaction problem,
- A hidden neuron in the GMDH structure could simultaneously input from the adjacent and non-adjacent layers resulting in a better performance,
- The GMDH model has a simple organization, where only a few parameters were set to simulate the ice-scoured deformation.

Besides the advantages of the GMDH model in the estimation of the subgouge sand deformations, with a high level of correlation, precision, and simplicity, the model slightly lacks a sufficiently flexible architecture, where the number of layers and neurons in the hidden layers were determined during the simulation instead of predetermination. In terms of pipeline response to ice-gouging, the horizontal and vertical subgouge soil deformations are the ultimate design factors that directly affect the pipeline response. The subgouge soil deformations are resultant of several key input parameters including the attack angle, gouge depth, seabed soil properties, ice dynamics, etc. Similar to almost all of the earlier studies (e.g., Paulin 1991 & 1992; Lach 1996; C-CORE 1995; C-CORE 1996; Hynes 1996; Liferov 2005; Nobahar et al. 2007; Abdalla et al. 2009; Pike et al.

2016; Nematzadeh and Shiri 2019a), the subgouge soil deformation was considered as the key outputs of the ice-gouging analysis in the current study.

In the current study, using the parameters governing the ice-induced sand deformations, ten GMDH models were developed, and then the best models, along with the most influencing parameters, were introduced using sensitivity and uncertainty analysis. The best GMDH models were compared to the artificial neural network (ANN) and empirical approaches. Finally, some GMDH-based formulas were suggested for the first time as a fast and cost-effective alternative approach for calculating the subgouge sand deformations profiles for practical applications. Further details related to the applied methodology and simulation results are provided in the coming sections.

4.1.2. Methodology

First, the development of the models by using the group method of data handling (GMDH) and its comparison with the artificial neural network (ANN) are discussed. Subsequently, the identified parameters governing the subgouge soil deformations are presented by performing dimensional analysis. Eventually, the applied datasets for validation of the NN-based models are introduced.

4.1.2.1. Group method of data handling (GMDH)

The group method of data handling (GMDH) was introduced by Ivakhnenko (1976). The technique has been broadly applied in different fields to approximate linear and non-linear problems, without having sufficient information about the system (Ivakhnenko and Ivakhnenko 1995, Ebtehaj et al. 2015, Azimi et al. 2018, Walton et al. 2019). The GMDH is considered a self-organized technique in which a set of neurons are applied to estimate the subgouge sand deformation like f_i as the output

(subgouge sand deformation) O_i by using the ice-gouging input parameters $\chi = (\chi_1, \chi_2, \chi_3, \dots, \chi_k)$.

In artificial neural network terminology, a neuron defines a mathematical function that simulates the functionality of the biological neuron. The neuron typically calculates the summation of the weighted average of the input parameters and then this sum is multiplied by a nonlinear function entitled the activation function.

This means that a binary connection between different neurons adopting quadratic polynomials results in the development of the GMDH neural network. Therefore, the O_i is expressed for n observed values in each dataset comprising m ice-gouging input parameters and the subgouge sand deformations as output as follows:

$$O_i = f(\chi_{i1,}\chi_{i2,}\chi_{i3,} \dots, \chi_{in,}) \qquad i = (1,2,3, \dots, m) \tag{4-1}$$

During the simulation process, the GMDH algorithm measures the mean square error (*MSE*) criterion as the average squared difference between the simulated subgouge sand deformations and the observed subgouge sand deformations in every single iteration. It is worth noting that the subgouge soil deformation was dimensionless and the *MSE* criterion was dimensionless as well.

The target function (difference between the observed subgouge deformations and the simulated ones) is minimized by calculating the *MSE* between the experimental value of the horizontal or vertical subgouge deformations (o_i) and the estimated values are as follows:

$$MSE = \sum_{i=1}^{m}\left(O_i\left(\chi_{i1,}\chi_{i2,}\chi_{i3,} \dots, \chi_{in,}\right) - o_i\right)^2 \rightarrow min \qquad i = (1,2,3, \dots, m) \tag{4-2}$$

Furthermore, the Corrected Akaike Information Criterion (AIC_c) (Akaike 1973) is employed for the correction of the target function to express the simplest model:

$$AIC_c = n \times log\left[\sum_{i=1}^{m}\left(O_i\left(\chi_{i1,}\chi_{i2,}\chi_{i3,} \dots, \chi_{in,}\right) - o_i\right)\right]^2 + 2\zeta + \frac{2\zeta(2\zeta+1)}{n-\zeta-1} \tag{4-3}$$

where the ζ is the number of optimized parameters during the training procedure. The Akaike information criterion (*AIC*) is an estimator of prediction error and thereby relative quality of statistical models for a given set of data that can estimate the quality of each model, relative to each of the other models. Additionally, the Volterra functional series (Volterra 1887) as a polynomial function is utilized to link between the input ice-gouging parameters and the subgouge sand deformations (output) parameter as follows:

$$O = \Gamma_0 + \sum_{i=1}^{n} \Gamma_i \chi_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \Gamma_{ij} \chi_i \chi_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \Gamma_{ijk} \chi_i \chi_j \chi_k + \cdots \quad (4\text{-}4)$$

The Kolmogorov-Gabor polynomial as a quadratic and bivariate form of the Volterra functional series is applied in the below equation (Ivakhnenko 1971, Ivakhnenko 1976; Ivakhnenko and Ivakhnenko 1995):

$$O = G(\chi_i, \chi_j) = \Gamma_0 + \Gamma_1 \chi_i + \Gamma_2 \chi_j + \Gamma_3 \chi_i^2 + \Gamma_4 \chi_j^2 + \Gamma_5 \chi_i \chi_j \quad (4\text{-}5)$$

where, Γ_i are unknown coefficients stemming from regression approaches to minimize the difference between the observed subgouge deformations (o_i) and the simulated values for each input parameters pair χ_i and χ_j (Farlow 1984). All the unknown coefficients are calculated through the least square (LS) method by solving the following matrix:

$$\Delta\Gamma = O \quad (4\text{-}6)$$

where, the Γ is the unknown coefficient vector of the Kolmogorov-Gabor polynomial, and the *o* is considered as the vector of observed values, so the Δ matrix is expressed in the form of Eq. (4-9):

$$\Gamma = (\Gamma_0, \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5) \quad (4\text{-}7)$$

$$o = \{o_0, o_1, \dots, o_m\}^T \quad (4\text{-}8)$$

$$\Delta = \begin{bmatrix} 1 & \chi_{1p} & \chi_{1q} & \chi_{1p}^2 & \chi_{1q}^2 & \chi_{1p}\chi_{1q} \\ 1 & \chi_{2p} & \chi_{2q} & \chi_{2p}^2 & \chi_{2q}^2 & \chi_{2p}\chi_{2q} \\ m & m & m & m & m & m \\ 1 & \chi_{mp} & \chi_{mq} & \chi_{mp}^2 & \chi_{mq}^2 & \chi_{mp}\chi_{mq} \end{bmatrix} \quad (4\text{-}9)$$

where the superscript *T* donates the transpose of the matrix. The vector of unknown coefficients is estimated through the least-square (LS) technique deriving from the multiple regression analysis and can be expressed as follows (Najafzadeh and Saberi-Movahed 2019):

$$\Gamma = (\Delta^T \Delta)^{-1} \Delta^T o \quad (4\text{-}10)$$

The described procedure is iterated for each specific neuron of the next hidden layer regarding the linkage topology of the GMDH network (Najafzadeh and Saberi-Movahed 2019). Figure 4-2 shows the flowchart of the current study.

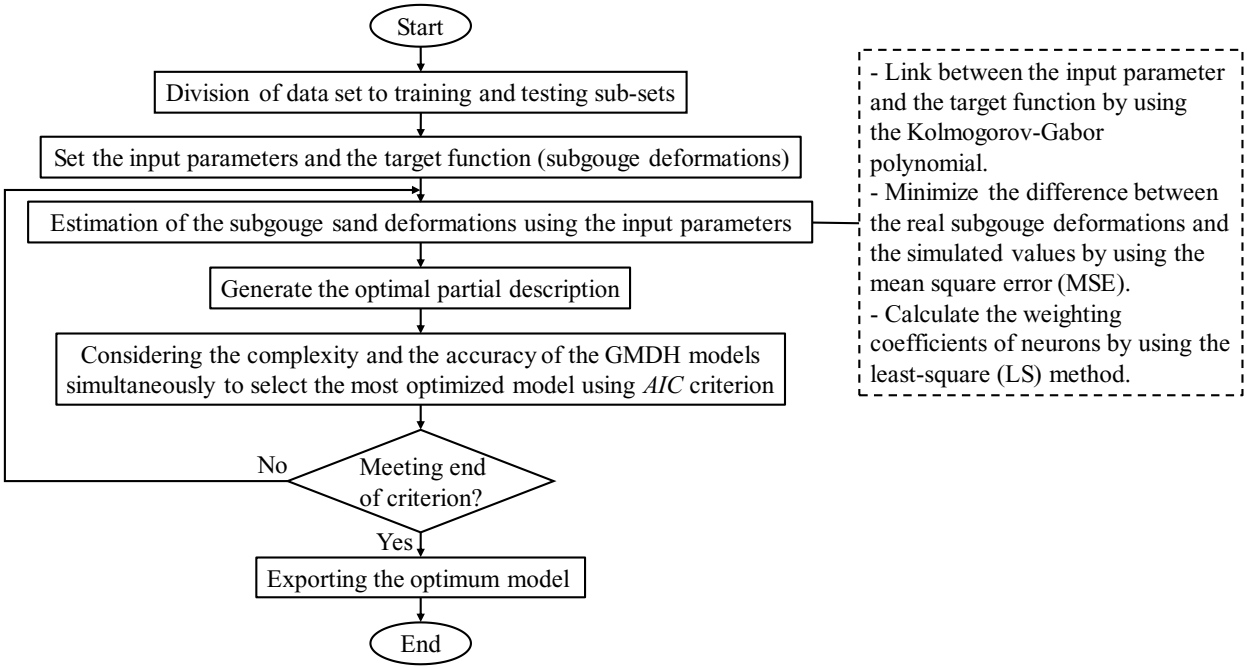


Figure 4-2. Flowchart of this study

Firstly, the constructed dataset was divided into two sub-samples for training and testing the GMDH models. It is noteworthy that four ratios including 50% of the data for training against 50%

of the data for testing, 60% for training against 40% for testing, 70% for training against 30% for testing, and 80% for training against 20% for testing the GMDH models were used in the current study. However, the ratio consists of 60% to train the models, and 40% to test them was utilized since the GMDH models had a better performance during the simulation process. Subsequently, the input and output parameters were respectively determined. After that, the GMDH model optimized the coefficients during the training procedure, and then the AIC_c was controlled and yielded the best outcome in the end.

4.1.2.2. Artificial neural networks (ANN)

Artificial neural networks (ANNs) are inspired by biological neuron networks in which a problem is solved similar to the way the human brain works. Due to the flexibility of the ANN models to predict complex problems, ANNs have been extensively applied to predict different high-dimensional and complicated parameters (Shahin et al. 2001). It is noteworthy that a detailed description of ANNs is beyond the scope of the current study. Generally, each NN model has some layers, including an input layer, at least a hidden layer, and an output layer. Input parameters are connected to the network through the input layer. Thus, the number of neurons in the input layer is equal to the number of input variables of the problem. Besides, the number of neurons in the output layer equals the number of output parameters (subgouge sand deformation parameter). Although there is no regulation to determine the number of neurons in the hidden layer, the number of hidden layer neurons is a function of the problem's complexity (Ebtehaj et al. 2016). In the current study, the number of neurons in the input layer equals the ice-gouging input parameters, while one neuron in the output layer was considered as the target parameter (subgouge sand deformations). Additionally, a trial and error procedure was used to select the number of neurons

in the hidden layers (Ebtehaj et al. 2016). This means that the number of neurons in the hidden layer was initially assumed equal to one, and the precision of the ANN model was evaluated by increasing this value to 10. The most optimal number of hidden layer neurons was selected to be five because the accuracy of the ANN model was insignificantly changed. In the NN-based models, the number of neurons in the antecedent layer is collected using the summation of weighted neurons in the hidden layer and then transferred to the next layer through a function entitled the "activation function" (Ebtehaj et al. 2016). In the current study, three activation functions including the "linear," "sigmoid," and "hyperbolic tangent" functions were used in the ANN model as follows:

$$linear(x) = x \tag{4-11}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4-12}$$

$$\tan(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{4-13}$$

The sigmoid function was selected in the ANN algorithm since this activation function was well simulated the subgouge sand deformations compared with its counterparts. Additionally, the values of weighted outputs were summed through the multilayer perceptron throughout a procedure entitled "model training". This means that the Levenberg-Marquardt (LM) algorithm was applied for training the multilayer perceptron neural network, where the biases and weights were adjusted using the back-propagation algorithm. The number of epochs was set to 2000 when the ANN model converges eventually, where the convergence was reached at epoch number 830.

4.1.2.3. Subgouge deformations

The maximum subgouge soil deformation in the sand is a function of the scour depth (D_s), the internal friction angle of sand (φ), the width of gouge (W), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the frontal berm (h') (the frontal berm is the passive soil heave formed in front of the moving ice keel and above the sea ground, see Figure 4-3), the resultant horizontal load (L_h) and resultant vertical load (L_v), the velocity of ice keel (V), and the specific weight of sand (γ_s) as follows (Azimi and Shiri 2020a):

$$d_{max} = f_1(D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \tag{4-14}$$

where, the L_h and L_v represent the resultant horizontal and vertical loads comprising drag force from wind, drag force from current, buoyancy force, keel weight, friction force on the bottom of the keel, Coulomb's passive friction force acting in front and on both sides of the keel, and driving force from the surrounding floe that is all amongst the input parameters to the problem (Azimi and Shiri 2020a). The d is considered the maximum subgouge deformation in the sand that is formed right under the moving ice keel in the gouge centerline. However, at greater depth on the subgouge centerline, by incorporating the soil depth (y/W), the subgouge sand deformation (d/W) is written as follows (Azimi and Shiri 2020a):

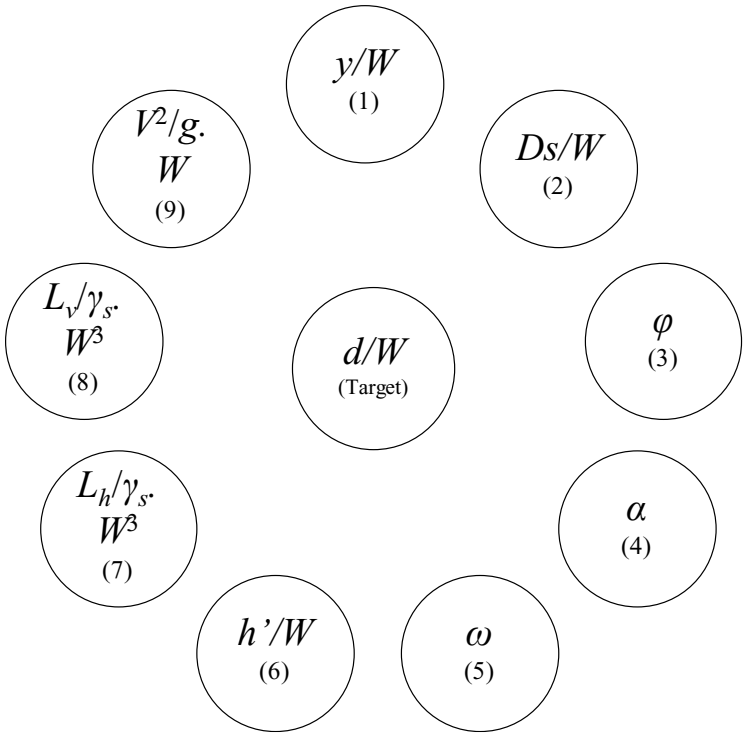
$$\frac{d}{W} = f_2\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{4-15}$$

Similarly, the horizontal (d_h/W) and vertical (d_v/W) components of the subgouge soil deformations can be written as below:

$$\frac{d_h}{W}, \frac{d_v}{W} = f_2\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{4-16}$$

Therefore, the introduced dimensionless groups in equation (4-16) were utilized as the input parameters for the GMDH model. Nondimensionalization determines systematically the characteristic units of a system to use, without relying heavily on prior knowledge of the system's

intrinsic properties. In other words, nondimensionalization can suggest the parameters, which should be used for analyzing a system. Therefore, using nondimensionalized parameters in this study benefitted broadening the applicability of the obtained results independent of physical equations governing the ice-seabed interaction. Figure 4-3 exhibits the applied input combinations in this study.



GMDH 1: [1,2,3,4,5,6,7,8,9]
GMDH 2: [1,2,3,4,5,6,7,8,~~9~~]
GMDH 3: [1,2,3,4,5,6,7,~~8~~,9]
GMDH 4: [1,2,3,4,5,6,~~7~~,8,9]
GMDH 5: [1,2,3,4,5,~~6~~,7,8,9]
GMDH 6: [1,2,3,4,~~5~~,6,7,8,9]
GMDH 7: [1,2,3,~~4~~,5,6,7,8,9]
GMDH 8: [1,2,~~3~~,4,5,6,7,8,9]
GMDH 9: [1,~~2~~,3,4,5,6,7,8,9]
GMDH 10: [~~1~~,2,3,4,5,6,7,8,9]

Figure 4-3. Combinations of input parameters

In other words, ten GMDH models were defined to predict the subgouge sand deformations both horizontally and vertically. GMDH 1, as the yardstick model, included all input variables, and to perform the sensitivity analysis of these variables, the inputs were disregarded one at a time in GMDH 2 to GMDH 10. These GMDH models show different performances in a wide range of precision, correlation, and complexity that makes challenges against using the fundamentals of soil mechanics to decide on the parameter that needs to be dropped. The elimination of the most

influential input parameters in the GMDH architecture resulted in poor performance of the simulation outcome. It is worth noting that implementation of such sensitivity analysis of input parameters in the field, experimental, and finite element studies is quite laborious, costly, and time-consuming.

4.1.2.4. Construction of the database

In this study, different experimental datasets were utilized to simulate the horizontal (d_h) and vertical (d_v) subgouge sand deformations. This means that the key values of five experimental studies reported by Paulin (1992), C-CORE (1995), Hynes (1996), C-CORE (1996), and Yang (2009) were employed for training and testing the GMDH models. These testing programs were carried out by modeling the progressive seabed gouging by the ice keel in centrifuge facilities located at C-CORE. The experimental program consisted of towing a model ice keel across a model testbed at a set gouge depth under various centrifuge accelerations. The test setup consisted of a solid half-width ice keel model mounted on a gantry situated on top of the strong box with an observation window mounted on the side for visualization. Particle image velocimetry (PIV) was used to track the evolution of subgouge deformation in all tests. A large amount of data was acquired from these experimental programs covering most of the key parameters governing the ice-seabed interaction. The quality and sufficiency of these datasets and the statistical significance of the resultant correlations for developing the neural network-based models were checked by performing the t-test and calculation of the P-value. Table 4-1 exhibits the range of the employed dimensionless parameters, including the maximum, minimum, average, variance, and standard deviation of the experimental measurements.

Table 4-1. Range of applied experimental values

Values	y/W	d_h/W	d_v/W	D_s/W	φ	α	ω	h'/W	$L_h/\gamma_s.W^3$	$L_v/\gamma_s.W^3$	$V^2/g.W$
Max	0.838	1.356	0.068	0.333	44	30	39.1	1.163	2.235	2.859	0.015
Min	0	0.002	0.004	0.065	18	15	-	-	-	-	-
Ave	0.220	0.300	0.019	0.125	34.745	20.556	4.633	0.168	0.586	0.756	0.004
Var.	0.071	0.186	0.0002	0.004	28.875	52.469	126.806	0.053	0.378	0.651	3.6E-05
Std.	0.267	0.431	0.015	0.064	5.374	7.244	11.261	0.230	0.615	0.807	0.006

The range of dimensionless horizontal subgouge deformations (d_h/w) for the Paulin's (1992) dataset comprising P-1 to P-4 tests was from 0.002 to 1.356 and these deformations were measured at the dimensionless soil depth up to 0.837. The keel attack angle and the internal friction angle of sand in the P-1 to P-4 tests were reported as 15° and 35°, respectively. The value of the surcharged soil slope angle (ω) in the P-1 to P-4 tests was 39.1°, 30.5°, 29.2°, and 26.3°, respectively. The gouge depth ratio (D_s/W) in P-1 to P-4 were respectively obtained at 0.059, 0.091, 0.088, and 0.093.

The C-1 to C-9 tests were reported in C-CORE's (1996) and C-CORE's (1995) datasets. The value of attack angle in C-1, C-4, C-7, C-8, and C-9 was 15° but this parameter in the C-2, C-5, and C-6 tests was measured as 30°. In the C-1 and C-4 tests, the value of the internal friction angle of sand was 29.022° and 34.542°, whilst this parameter for C-2 and C-4 was similar, with the φ of 28.912°. The internal friction angle of sand seabed in C-5 to C-9 were respectively reported at 41.056°, 33.880°, 41.332°, and 37.882°. The surcharged soil slope angle was not reported in the C-1 to C-9 tests.

Moreover, the H-1 to H-4 tests have measured by Hynes (1996) in which the value of internal friction angle was respectively measured at 38.2°, 36.6°, 37°, and 39.3°. The attack angle in the H-2 test was 30°, whereas this parameter in H-1, H-3, and H-4 was observed to be 15°. The angle of a surcharged soil slope in the H-1 to H-4 tests was not reported and the frontal berm ratio (h'/W) in these tests were at 0.233, 0.207, 0.207, and 0.240, respectively. The value of the ice velocity ratio ($V^2/g.W$) in the H-1 to H-4 tests was constant at 0.015.

The value of dimensionless horizontal subgouge deformations in Yang's model (Y-1 to Y-2 tests) varied from zero to 0.026 and the two models with a keel attack angle of 15° and 30° were applied. Yang (2009) utilized a sandy seabed with the internal friction angles of sand, ranging from 29° to 44°, respectively.

The applied data for training and testing the GMDH model were randomly selected, meaning that some parts of every single experimental dataset were utilized to test the GMDH model, whereas the remaining were employed as testing datasets. Moreover, the statistical significance of correlations was checked by conducting a t-test and calculation of the P-value, assuming that a P-value of 0.05 or less is statistically significant. This, indeed, was to assess the probability of the relationship between the parameters being affected by an alternative hypothesis (Azimi and Shiri 2020a). The calculated P-value for the applied experimental studies was 0.039, showing that the correlation was statistically significant, so the performed t-test showed that the quality of these datasets was acceptable to develop the NN-based models.

Thus, nine dimensionless parameters including y/W, D_s/W, φ, α, ω, h'/W, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$ were calculated to feed the NN-based models. It is noteworthy that 60% of the observed values were applied for training the GMDH models, whereas 40% was for testing these models.

4.1.2.5. Goodness of fit

To evaluate the correlation of the GMDH models, the correlation coefficient (*R*), the variance accounted for (*VAF*), and the Nash-Sutcliff efficiency coefficient (*NSC*) were utilized, whereas the root mean square error (*RMSE*), the mean absolute error (*MAE*), and the scatter index (*SI*) were employed to assess the precision of the NN-based models. In addition, the complexity of these models was examined using the Akaike Information Criterion (*AIC*). The AIC index is a tool for model selection. The best GMDH model should have the lowest complexity and the lowest value of the AIC criterion (Stoica and Selen 2004, Ebtehaj et al. 2014). Presenting an NN-based model with the highest degree of correlation and accuracy along with the lowest level of complexity is the aim of the current study.

The closeness of the *R* and the *NSC* to one indicated that the NN-based model had the highest correlation, whereas the closeness of the *RMSE*, *MAE*, and *SI* to zero signified that the particular model had the lowest error. It should be stated that the premium GMDH model had the highest value of the *VAF* index. Regarding the introduced indices, the premium GMDH model had the lowest AIC index and error (*RMSE*, *SI*, *MAE*), while this model had the highest level of correlation with the observed values (*R* and *NSC*).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{4-17}$$

$$VAF = \left(1 - \frac{var(P_i - O_i)}{var(P_i)}\right) \times 100 \tag{4-18}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{4-19}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{4-20}$$

$$SI = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}}{\bar{O}} \quad (4\text{-}21)$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (4\text{-}22)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (4\text{-}23)$$

where, *Oi, Pi,* $\bar{O}$*,* $\bar{P}$*, n* and *k* are the observed values, the simulated values, the average observed values, the average simulated values, the number of experimental values, and the number of estimated factors included in the numerical model respectively.

4.1.3. Results of sensitivity analysis (SA)

Sensitivity analysis helps to determine how the uncertainty in the output of a mathematical model or system (numerical or otherwise) can be divided and allocated to different sources of uncertainty in its inputs. A related practice is uncertainty analysis, which has a greater emphasis on the quantification and propagation of uncertainty. The uncertainty and sensitivity analysis should be run in tandem. Here, first, a sensitivity analysis (SA) and an uncertainty analysis (UA) for the GMDH models were conducted to simulate the horizontal and vertical deformations. Then, results from the GMDH models were compared to the ANN and the empirical approaches. Lastly, the premium GMDH models and the most influential input parameters were introduced.

To perform the sensitivity analysis, equations (4-17) to (4-23) were applied, whereas the uncertainty analysis was implemented using equations (4-24) to (4-26). In other words, the precision (*RMSE*, *MAE*, and *SI* indices), correlation (*R*, *VAF*, and *NSC* indices), and complexity

(*AIC*) of the NN-based models were evaluated in the sensitivity analysis section, whilst the performance and uncertainty of these models were examined in the uncertainty analysis section.

4.1.3.1. Horizontal soil deformations

For the horizontal subgouge sand deformations, a sensitivity analysis for GMDH 1 to GMDH 10 models was implemented. Regarding the aforementioned sections, ten GMDH models were developed adopting nine dimensionless input parameters including y/W, D_s/W, φ, α, ω, h'/W, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$. To assess the influence of these input parameters, the effect of each dimensionless factor was eliminated in each step, and then GMDH 2 to GMDH 10 were respectively defined. Table 4-2 shows the calculated statistical indices for the simulated horizontal deformations through all GMDH models in training, testing, and combined modes based on their ratios, i.e., 60% to 40% respectively.

Table 4-2. Results of calculated statistical indices for modeling horizontal deformations by GMDH 1 to GMDH 10 models

Indices		Correlation			Precision			Complexity
Mode	Model	*R*	*VAF*	*NSC*	*RMSE*	*SI*	*MAE*	*AIC*
	GMDH 1	0.988	97.541	0.976	0.035	0.400	0.025	-120.769
	GMDH 2	0.900	97.958	0.980	0.032	0.365	0.021	-124.359
	GMDH 3	0.986	97.179	0.973	0.037	0.428	0.025	-118.125
	GMDH 4	0.985	96.873	0.970	0.039	0.449	0.025	-116.189
	GMDH 5	0.988	97.621	0.977	0.034	0.393	0.024	-121.451
	GMDH 6	0.986	97.047	0.971	0.038	0.437	0.024	-117.287

Training	GMDH 7	0.989	97.741	0.978	0.033	0.383	0.023	-122.483
	GMDH 8	0.988	97.612	0.976	0.034	0.394	0.022	-121.329
	GMDH 9	0.983	96.545	0.967	0.041	0.471	0.027	-114.296
	GMDH 10	0.523	-208.403	0.270	0.192	2.205	0.092	-53.297
	GMDH 1	0.955	89.353	0.832	0.065	1.085	0.038	-58.032
	GMDH 2	0.933	44.538	0.778	0.074	1.252	0.034	-54.630
	GMDH 3	0.841	27.883	0.678	0.089	1.507	0.041	-49.877
	GMDH 4	0.821	67.164	0.566	0.104	1.750	0.042	-46.044
	GMDH 5	0.954	89.280	0.910	0.047	0.799	0.026	-66.152
Testing	GMDH 6	0.854	70.802	0.406	0.121	2.047	0.047	-42.030
	GMDH 7	0.951	88.638	0.904	0.049	0.823	0.026	-65.378
	GMDH 8	0.927	68.572	0.826	0.066	1.109	0.032	-57.732
	GMDH 9	0.943	87.494	0.888	0.053	0.888	0.027	-63.435
	GMDH 10	0.461	-79.626	0.125	0.147	2.485	0.078	-37.070
	GMDH 1	0.975	94.266	0.918	0.047	0.674	0.030	-95.674
	GMDH 2	0.913	76.590	0.899	0.049	0.720	0.026	-96.467
	GMDH 3	0.928	69.461	0.855	0.058	0.860	0.031	-90.826
	GMDH 4	0.919	84.989	0.808	0.065	0.969	0.032	-88.131
Combined	GMDH 5	0.974	94.285	0.950	0.039	0.555	0.025	-99.331
training	GMDH 6	0.933	86.549	0.745	0.071	1.081	0.033	-87.184
and	GMDH 7	0.974	94.100	0.948	0.039	0.559	0.024	-99.641

testing	GMDH 8	0.964	85.996	0.916	0.047	0.680	0.026	-95.890
ratios	GMDH 9	0.967	92.925	0.935	0.046	0.638	0.027	-93.952
	GMDH 10	0.498	- 156.892	0.212	0.174	2.317	0.086	-46.806

Regarding the simulation results, the *NSC* for GMDH 1 in the training mode was obtained to be 0.976. GMDH 1 predicts the horizontal subgouge deformations (d_h/W) in terms of all input variables. In GMDH 2, the impact of the dimensionless velocity ratio (ice dynamics parameter or $V^2/g.W$) was omitted for this model. In the testing mode of GMDH 2, the *R* was equal to 0.933.

For GMDH 3, the influence of dimensionless vertical load ($L_v/\gamma_s.W^3$) was removed and the *MAE* index for the testing stage of this model was at 0.0407. The *RMSE* criterion in the testing mode of GMDH 4 equaled 0.104. For GMDH 4, the effect of the dimensionless horizontal load ($L_h/\gamma_s.W^3$) was disregarded. The influence of the h'/W parameter for GMDH 5 was ignored during the simulation process and the value of the *MAE* index in the testing stage of this model was equal to 0.026. During the testing procedure of GMDH 6, the *AIC* index was obtained as -42.030. To simulate the subgouge sand deformations using the GMDH 6 model, the effect of the ω parameter was removed.

For GMDH 7, the effect of the ice keel attack angle (α) was dropped to model the horizontal deformations. In the testing stage of GMDH 7, the value *NSC* was computed as 0.904.

The value of the *AIC* criterion equaled -54.732 during the testing of the GMDH 8 model. To model the horizontal subgouge deformations using GMDH 8, the influence of the internal friction angle of sand (φ) as the shear strength parameter of the seabed soil was removed. In the GMDH 9 model,

the impact of the gouge depth ratio (*Ds/W*) was ignored, with the *MAE* of 0.027 in the testing mode.

Additionally, the *AIC* in the training stage of GMDH 10 was estimated at -53.297. The influence of the dimensionless soil depth parameter (*y/W*), to simulate the horizontal deformations (d_h/W) adopting GMDH 10, was eliminated.

Regarding the simulation results of combined modes, overall, GMDH 1 had the lowest level of error and the highest level of correlation with experimental measurements. Thus, GMDH 1 was identified as the best model in terms of accuracy (*RMSE*, *SI*, and *MAE*), correlation (*R*, *VAF*, and *NSC*), and complexity (*AIC*) to predict the horizontal sand deformations in both training and testing modes. After GMDH 1, GMDH 5 was detected as the second-best model. Additionally, GMDH 7, GMDH 9, and GMDH 2 were the third, fourth, and fifth premium models to estimate the horizontal deformations. Based on the sensitivity analysis, GMDH 8 as the sixth-best model, GMDH 6 as the seventh-best model, GMDH 3 as the eighth-best model, and GMDH 4 as the ninth-best model were prioritized. Amongst all GMDH models, GMDH 10 had the highest level of error and the lowest level of correlation, so this model was considered the worst GMDH model to simulate horizontal deformations. The scatter plots of the GMDH 1 for the training and testing modes are depicted in Figure 4-4.

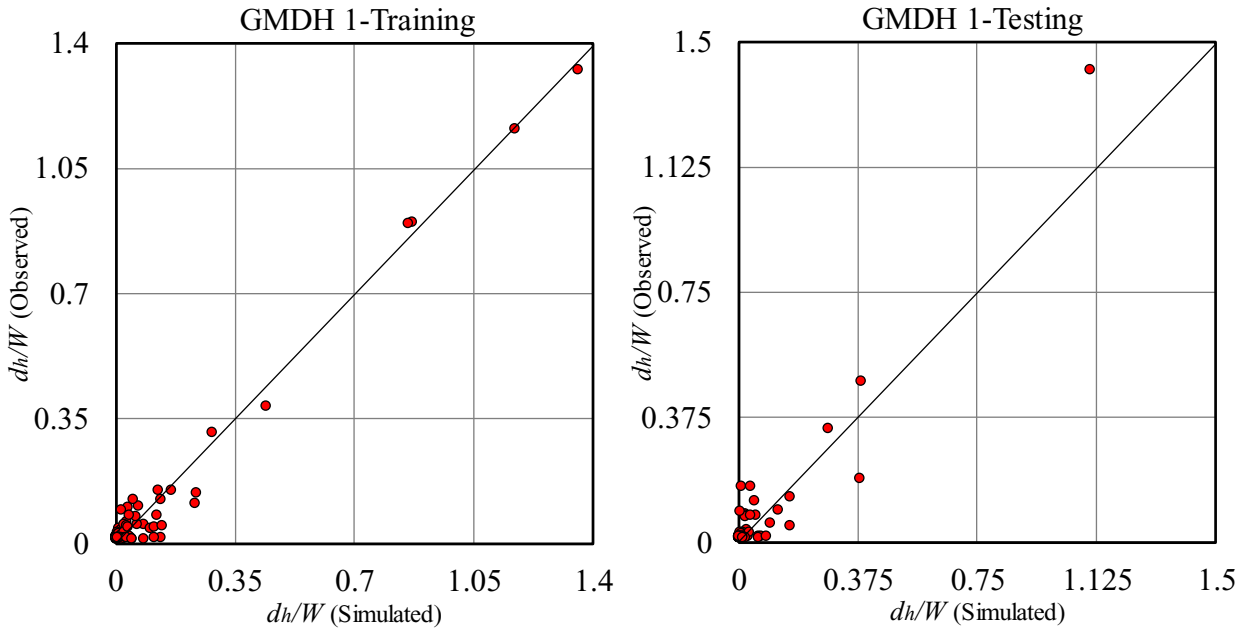


Figure 4-4. Scatter plots for simulated horizontal sand deformations by GMDH 1 in training and testing modes

According to the performed analysis, the dimensionless soil depth (y/W) had the most significant impact to model the horizontal deformations since the accuracy of the simulation had dramatically decreased by removing the effect of the y/W. After the y/W parameter, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and ω were respectively introduced as the second, the third, and the fourth influential input variables. Among all input parameters, the internal friction angle of sand (φ) had a fifth place in terms of effectiveness. Moreover, the dimensionless velocity parameter ($V^2/g.W$), the gouge depth ratio (D_s/W), and the attack angle (α) were respectively recognized as the sixth, seventh, and eighth-important inputs. The h'/W variable had insignificant importance to simulate the horizontal deformations adopting the GMDH model.

For GMDH 1, almost 87% of the simulated horizontal deformations in the training mode and 76% of these values in the testing mode had an error of less than one percent. Therefore, GMDH 1 was

known as the superior model to predict the horizontal subgouge deformations since this model had an acceptable precision and complexity compared to other GMDH models. Besides, GMDH 1 was a function of all input variables. Regarding the conducted analyses, the dimensionless soil depth (y/W) was the most significant input parameter.

4.1.3.2. Vertical soil deformations

Table 4-3 displays computed statistical indices for simulation of the vertical deformations by GMDH 1 to GMDH 10 in training, testing, and combined modes with respective ratios, i.e., 60% to 40%.

Table 4-3. Results of calculated statistical indices for modeling vertical deformations by GMDH 1 to GMDH 10 models

Indexes		Correlation			Precision			Complexity
Mode	Model	*R*	*VAF*	*NSC*	*RMSE*	*SI*	*MAE*	*AIC*
Training	GMDH 1	0.985	97.308	0.974	0.008	0.374	0.006	-83.269
	GMDH 2	0.196	-2133.5	0.044	0.038	1.873	0.023	-51.806
	GMDH 3	0.982	96.394	0.965	0.009	0.452	0.007	-79.579
	GMDH 4	0.980	95.887	0.960	0.009	0.454	0.007	-79.491
	GMDH 5	0.978	95.608	0.955	0.010	0.480	0.007	-78.399
	GMDH 6	0.985	96.849	0.969	0.008	0.407	0.006	-81.642
	GMDH 7	0.963	92.270	0.926	0.011	0.527	0.008	-76.584
	GMDH 8	0.968	93.348	0.937	0.010	0.509	0.009	-77.268
	GMDH 9	0.979	95.695	0.958	0.010	0.490	0.008	-77.975

	GMDH 10	0.319	-1114.45	0.099	0.038	1.847	0.021	-52.074
Testing	GMDH 1	0.849	57.960	0.709	0.007	0.568	0.005	-52.148
	GMDH 2	0.161	-172.808	-0.156	0.014	1.124	0.010	-43.265
	GMDH 3	0.707	30.578	0.460	0.009	0.733	0.006	-48.826
	GMDH 4	0.785	46.818	0.604	0.008	0.664	0.006	-50.120
	GMDH 5	0.768	25.094	0.583	0.009	0.705	0.006	-49.337
	GMDH 6	0.793	43.530	0.630	0.008	0.619	0.006	-51.024
	GMDH 7	0.810	47.544	0.639	0.005	0.357	0.003	-58.204
	GMDH 8	0.734	44.106	0.491	0.006	0.499	0.005	-53.849
	GMDH 9	0.707	-101.232	0.464	0.010	0.794	0.005	-47.795
	GMDH 10	-0.144	-167.531	-1.275	0.020	1.527	0.015	-39.271
Based on the training and testing ratio	GMDH 1	0.931	81.567	0.868	0.008	0.452	0.006	-70.821
	GMDH 2	0.182	-1349.223	-0.036	0.0284	1.573	0.018	-48.390
	GMDH 3	0.872	70.067	0.763	0.009	0.564	0.007	-67.278
	GMDH 4	0.902	76.259	0.818	0.009	0.538	0.007	-67.743
	GMDH 5	0.894	67.402	0.806	0.010	0.570	0.007	-66.774
	GMDH 6	0.908	75.521	0.833	0.008	0.492	0.006	-69.395
	GMDH 7	0.902	74.380	0.811	0.009	0.459	0.006	-69.232
	GMDH 8	0.874	73.651	0.759	0.008	0.505	0.007	-67.900
	GMDH 9	0.870	16.924	0.760	0.010	0.612	0.007	-65.903
	GMDH 10	0.134	-735.682	-0.451	0.031	1.719	0.019	-46.953

According to the implemented sensitivity analysis, in terms of accuracy, correlation, and complexity, GMDH 1 was the best model to simulate the vertical sand deformations, with an *R* index of 0.986 in the training mode. After GMDH 1, GMDH 7 was detected as the second-best model to simulate the vertical subgouge deformations (d_v/W), whereas GMDH 4, GMDH 6, and GMDH 5 were respectively introduced as the third, fourth, and fifth-best models. Additionally, GMDH 8 as the sixth-best model, GMDH 9 as the seventh-best model, and GMDH 3 as the eighth-best model were ranked. In addition, GMDH 2 was graded as the ninth-best GMDH model, and among all GMDH models and GMDH 10 was considered the worst model to simulate vertical deformations. In Figure 4-5, the scatter plots of GMDH 1 in both the training and testing procedure are illustrated.

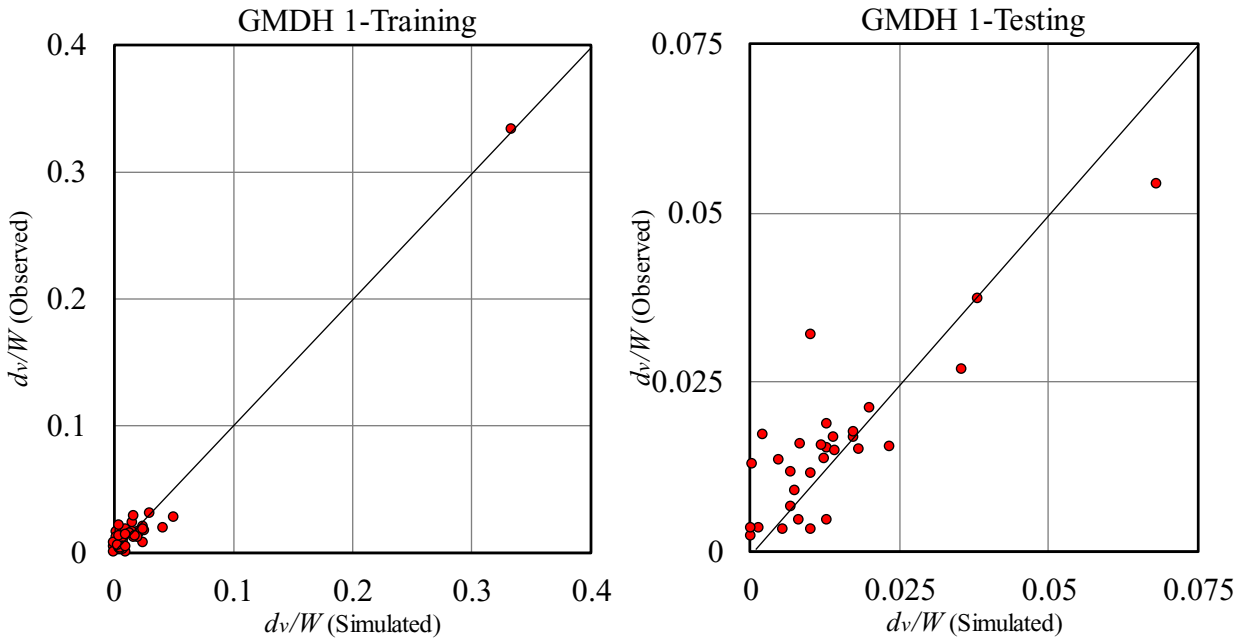


Figure 4-5. Scatter plots for simulated vertical sand deformations by GMDH 1 in training and testing modes

On the other hand, the ratio of the soil depth to the gouge width (y/W) was identified as the most effective input parameter to estimate the vertical deformations (d_v/W) through the GMDH models. Furthermore, the dimensionless velocity parameter ($^{V2}/g.W$), the dimensionless vertical load parameter ($L_v/\gamma_s.W^3$), and the gouge depth ratio (D_s/W) were prioritized as the second, third, and fourth most important input variables. Furthermore, the internal friction angle of sand (φ) was detected as the fifth significant input factor to model the vertical deformations. The h'/W as the sixth important factor, the ω as the seventh most important parameter, and the dimensionless horizontal load parameter ($L_h/\gamma_s.W^3$) as the eighth-most influencing variable was detected. Among all input factors, the attack angle has the lowest level of effectiveness in the estimation the vertical deformations.

Results of the error analysis in the training and testing procedures of GMDH 1 were quite similar, denoting that approximately 13% of the simulated vertical deformations had an error between one percent and five percent. Moreover, almost one-third of the estimated vertical deformations adopting GMDH 2 had an error from one percent to five percent, whilst one-fifth of the simulated vertical deformations by GMDH 3 had an error from one to five percent. Regarding the conducted analyses and the results in combined modes, overall, GMDH 1, was known as the best model to predict vertical deformations. Additionally, the dimensionless soil depth parameter (y/W) was regarded as the most influential input factor to model the vertical subgouge deformations.

It is worth mentioning that the best model does not necessarily mean containing all of the model parameters. Identifying the best GMDH model needs a set of criteria comprising the correlation, precision, and complexity indices in both the training and testing modes. The best model (i.e., GMDH 1) has the highest level of correlation (the closeness of the R and the NSC to one) and accuracy (closeness of the RMSE, MAE, and SI to zero) along with the lowest level of complexity

(the lowest value of AIC) overall in both training and testing modes simultaneously. Therefore, the coincidence of being the best model and containing all model parameters in GMDH1 is an accident in this study, whilst inverse cases have been observed in earlier studies (e.g., Azimi and Shiri 2021).

4.1.4. Results of Uncertainty analysis (UA)

4.1.4.1. Horizontal soil deformations

Uncertainty analysis (UA) was implemented to assess the performance of the GMDH models and describe the error predicted by each model (Azimi and Shiri 2020a). It means that the estimated error by the GMDH model (e_j) was calculated as the difference between the simulated (P_j) and the observed (O_j) subgouge sand deformations are as follows:

$$e_j = P_j - O_j \tag{4-24}$$

Subsequently, the mean value of the estimated error $(\bar{e})$ and the standard deviation (S_e) were computed as below:

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{4-25}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{4-26}$$

The negative sign of the mean value of the estimated error $(\bar{e})$ denoted an underestimated performance of the GMDH model, meaning that the predicted deformations were less than the observed values. However, the positive sign showed an overestimated performance of the numerical model, signifying that the simulated deformations were more significant than the observed values. Furthermore, by employing the $\bar{e}$ and S_e, a confidence bound was created around

an estimated error using the “Wilson score approach” without the continuity correction. The Wilson score interval, which was developed by Wilson (1927), was an improvement over the normal distribution interval in which an asymmetric normal distribution was utilized to improve the confidence interval bound (Wallis 2013). This method was used to perform the uncertainty analysis of the GMDH models. Subsequently, a ±1.96Se resulted in roughly the “95% confidence bound or 95% confidence interval (95%CB)” and then the “width of uncertainty bound (WUB)” was calculated. In Figure 4-6, a comparison between the experimental and the numerical horizontal deformations is shown.

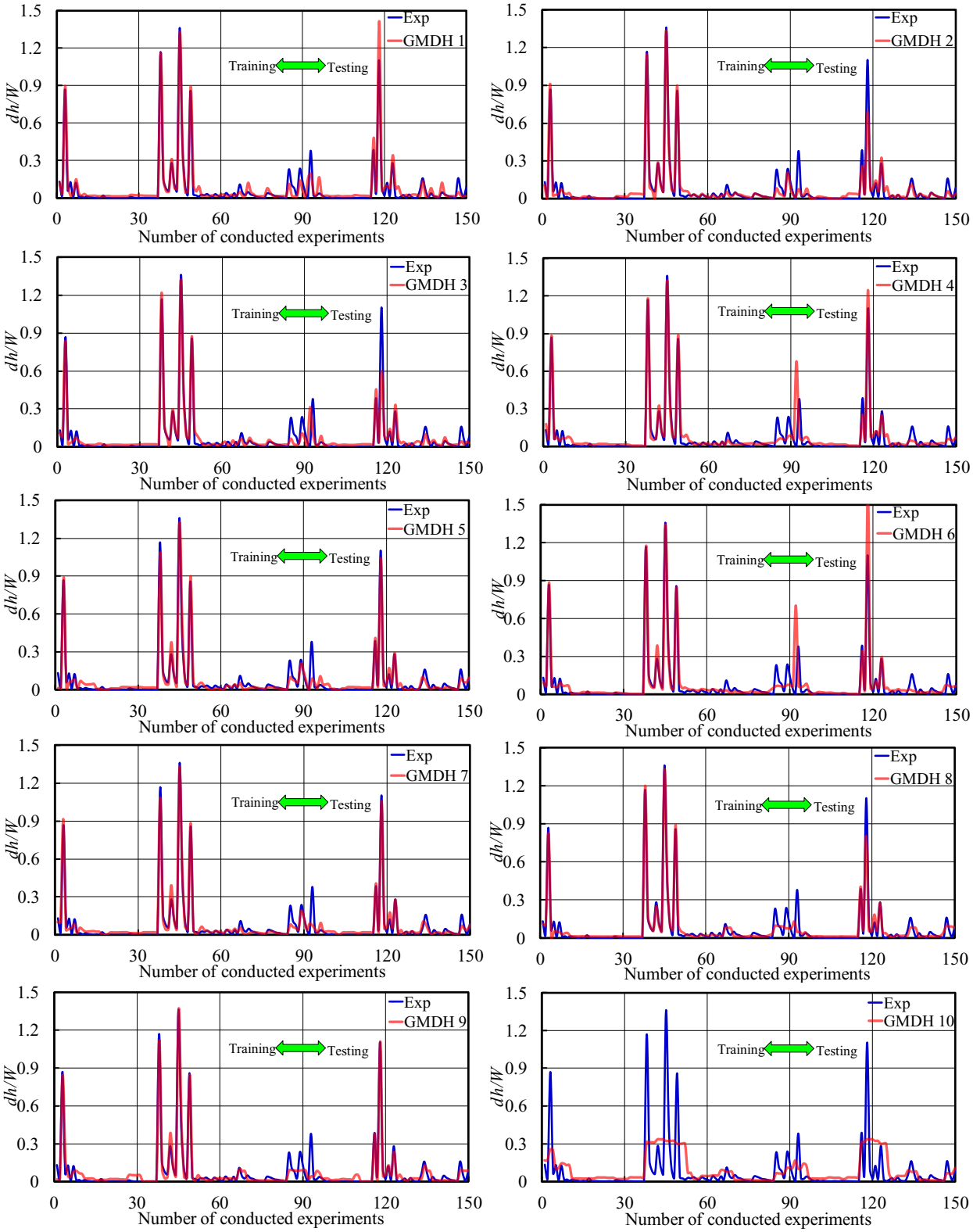


Figure 4-6. Comparison between experimental horizontal deformations and GMDH results

Regarding the conducted UA, GMDH 1, GMDH 4, GMDH 5, GMDH 6, GMDH 7, GMDH 9, and GMDH 10 underestimated the subgouge deformations, whereas GMDH 2, GMDH 3, and GMDH 8 had an overestimated performance. Additionally, the width of uncertainty bound (WUB) for GMDH 1, GMDH 2, and GMDH 8 was equal to ±0.008, while the WUB for GMDH 5 and GMDH 7 was to be ±0.006. However, the widest uncertainty bound belonged to GMDH 10, with ±0.028. Moreover, Table 4-4 displays the results of the performed UA to simulate the horizontal deformations using the GMDH models.

Table 4-4. Results of uncertainty analysis for modeling horizontal deformations by GMDH models

Model	$\bar{e}$	S_e	WUB	95% CI
GMDH 1	-0.006	0.049	±0.008	-0.014 to 0.002
GMDH 2	0.002	0.053	±0.008	-0.006 to 0.011
GMDH 3	0.0002	0.064	±0.010	-0.010 to 0.010
GMDH 4	-0.005	0.072	±0.012	-0.017 to 0.006
GMDH 5	-0.001	0.04	±0.006	-0.007 to 0.006
GMDH 6	-0.008	0.082	±0.013	-0.021 to 0.005
GMDH 7	-0.002	0.04	±0.006	-0.008 to 0.005
GMDH 8	0.001	0.049	±0.008	-0.007 to 0.008
GMDH 9	-0.001	0.046	±0.007	-0.009 to 0.006
GMDH 10	-0.017	0.175	±0.028	-0.045 to 0.011

According to the implemented UA, the 95%CB for GMDH 1, GMDH 2, and GMDH 3 were respectively estimated to be (-0.014 to 0.002), (-0.006 to 0.011), and (-0.010 to 0.010). However, for GMDH 4, GMDH 5, and GMDH 6, the 95%CB equaled (-0.017 to 0.006), (-0.007 to 0.006), and (-0.021 to 0.005), respectively. It is noteworthy that the lower and upper bound of 95%CB for GMDH 10 were -0.045 and 0.011, respectively.

4.1.4.2. Vertical soil deformations

The comparison between the experimental and numerical vertical deformations is portrayed in Figure 4-7. Based on the performed UA, except for GMDH 7, which had an overestimated performance, other GMDH models underestimated the vertical sand deformations. The WUB for GMDH 1, GMDH 3, GMDH 4, and GMDH 6 was similar, with ±0.002. In addition, for GMDH 2 and GMDH 10, the WUB equaled ±0.009. Moreover, the uncertainty bound of GMDH 5, GMDH 7, GMDH 8, and GMDH 9 had the same width, with ±0.003. It should be stated that the 95%CB for GMDH 1, GMDH 2, and GMDH 4 were respectively calculated at (-0.002 to 0.001), (-0.012 to 0.006), and (-0.003 to 0.001). The results of the UA to simulate vertical deformations by using the GMDH models are shown in Table 4-5.

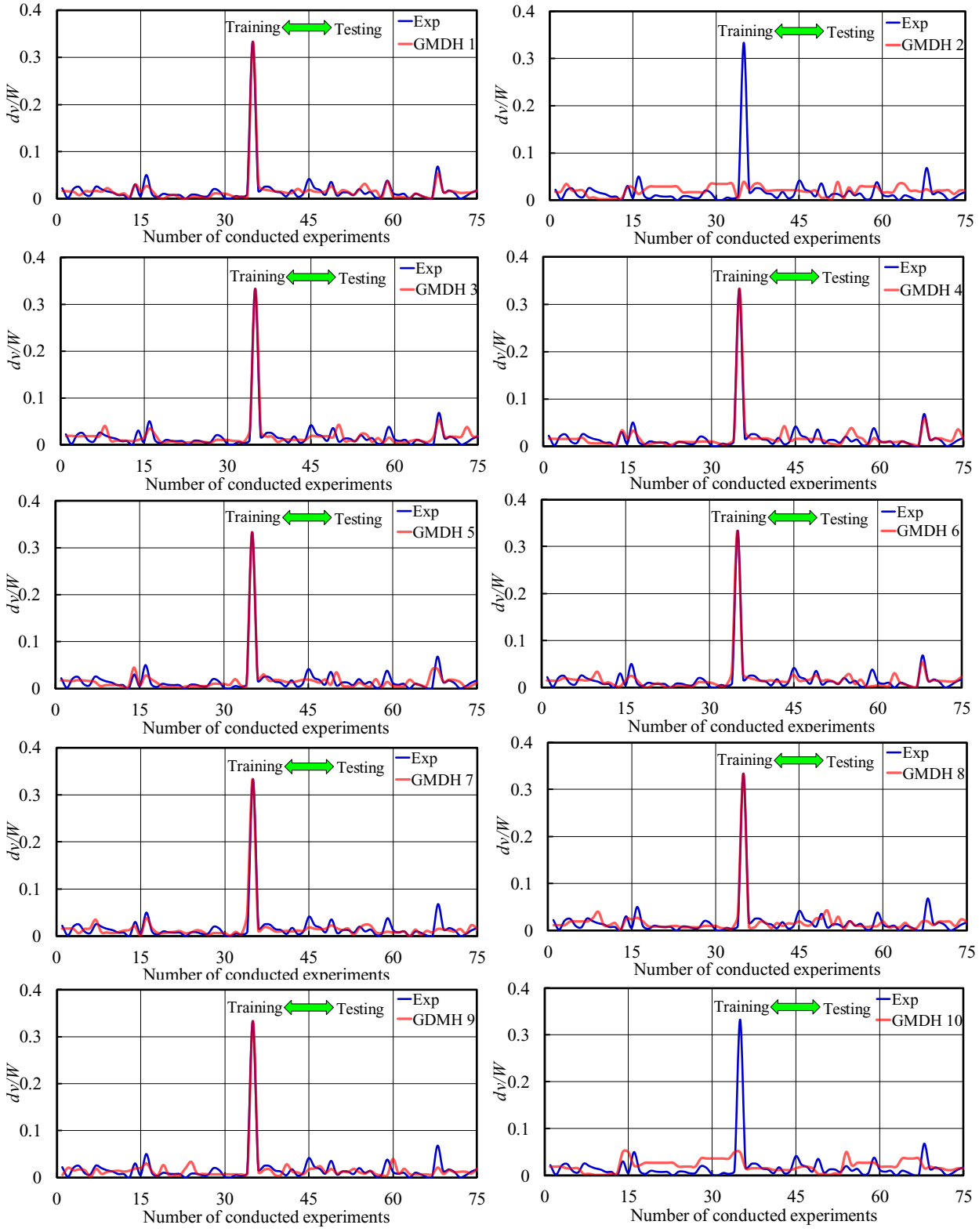


Figure 4-7. Comparison between experimental vertical deformations and GMDH results

Table 4-5. Results of uncertainty analysis for modeling vertical deformations by GMDH model

Model	$\bar{e}$	S_e.	WUB	95% CI
GMDH 1	-0.0005	0.008	±0.002	-0.002 to 0.001
GMDH 2	-0.003	0.038	±0.009	-0.012 to 0.006
GMDH 3	-0.0009	0.011	±0.002	-0.003 to 0.002
GMDH 4	-0.001	0.01	±0.002	-0.003 to 0.001
GMDH 5	-0.0004	0.011	±0.003	-0.003 to 0.002
GMDH 6	-0.0007	0.01	±0.002	-0.003 to 0.002
GMDH 7	0.0004	0.011	±0.003	-0.002 to 0.003
GMDH 8	-0.0004	0.012	±0.003	-0.003 to 0.002
GMDH 9	-0.0002	0.012	±0.003	-0.003 to 0.002
GMDH 10	-0.003	0.038	±0.009	-0.012 to 0.005

Therefore, GMDH 1 had an underestimated performance, with the WUB equaled to ±0.008 and ±0.002 to predict the horizontal and vertical sand deformations, respectively. In contrast, GMDH 10 had the wideset uncertainty bound, with an underestimated performance for modeling the deformations. It is worth mentioning that the best model does not necessarily mean containing all of the model parameters. Identifying the best GMDH model needs a set of criteria comprising the correlation, precision, and complexity indices both in the training and testing modes. The best model (i.e., GMDH 1) has the highest level of correlation (the closeness of the *R* and the *NSC* to one) and accuracy (closeness of the *RMSE, MAE*, and *SI* to zero) along with the lowest level of complexity (the lowest value of AIC) overall in both training and testing modes simultaneously. Therefore, the coincidence of being the best model and containing all model parameters in

GMDH1 is an accident in this study, whilst inverse cases have been observed in other published studies (e.g., Azimi and Shiri 2021). Also, Table 4-4 and Table 4-5 show whether the GMDH models have underestimated or overestimated the output parameters.

4.1.5. Comparison of GMDH predictions with empirical and ANN models

A comparison between the results of the best GMDH model (GMDH 1) with the ANN and the empirical (Woodworth-Lynes et al. 1998) approaches was made. It is noteworthy that Woodworth-Lynes et al. (1998) presented two empirical formulas to estimate the horizontal (d_h) and vertical (d_v) subgouge soil deformations as follow:

$$d_h = 0.6(W.D_s).exp\left(-\frac{2}{3}.\frac{y}{D_s}\right) \quad (4\text{-}27)$$

$$d_v = D_s.exp\left(-\frac{1}{3}.\frac{y}{D_s}\right) \quad (4\text{-}28)$$

where, d_h, W, D_s, d_v, and y are the horizontal subgouge deformation, the width of the gouge, the depth of the gouge, the vertical subgouge deformation, and the soil depth, respectively. The comparison between the best GMDH models to the empirical and the ANN approaches are arranged in Table 4-6. To compare the results of the NN-based models with equations (27) and (28), the horizontal and vertical subgouge deformations estimated by the empirical models become dimensionless.

Table 4-6. Comparison between results from GMDH with ANN and empirical models

Target	Model	Mode	R	VAF	RMSE	SI	MAE	NSC	AIC
	GMDH 1	Training	0.988	97.541	0.035	0.400	0.025	0.976	-120.769
		Testing	0.955	89.353	0.065	1.085	0.038	0.832	-58.032

dh/W	ANN	Training	0.971	94.394	0.052	0.677	0.023	0.938	-112.125
		Testing	0.964	92.907	0.052	0.681	0.029	0.929	-72.394
	Empirical	-	0.894	79.650	0.425	1.081	0.193	0.785	-58.533
	GMDH 1	Training	0.985	97.308	0.008	0.374	0.006	0.974	-83.269
		Testing	0.849	57.960	0.007	0.568	0.005	0.709	-52.148
dv/W	ANN	Training	0.799	63.609	0.034	1.778	0.010	0.833	-65.629
		Testing	0.739	20.722	0.010	0.680	0.006	0.984	-57.655
	Empirical	-	0.047	-2670.28	4.304	6.928	1.242	-0.006	52.709

The value of *R* to predict the horizontal deformations using the ANN in training mode was computed to be 0.971 and the value *NSC* to estimate the horizontal deformations using the empirical model was estimated as 0.785. The profiles of horizontal and vertical deformations obtained from the experimental measurements, the GMDH, the ANN, and the empirical models are depicted in Figure 4-8 and Figure 4-9.

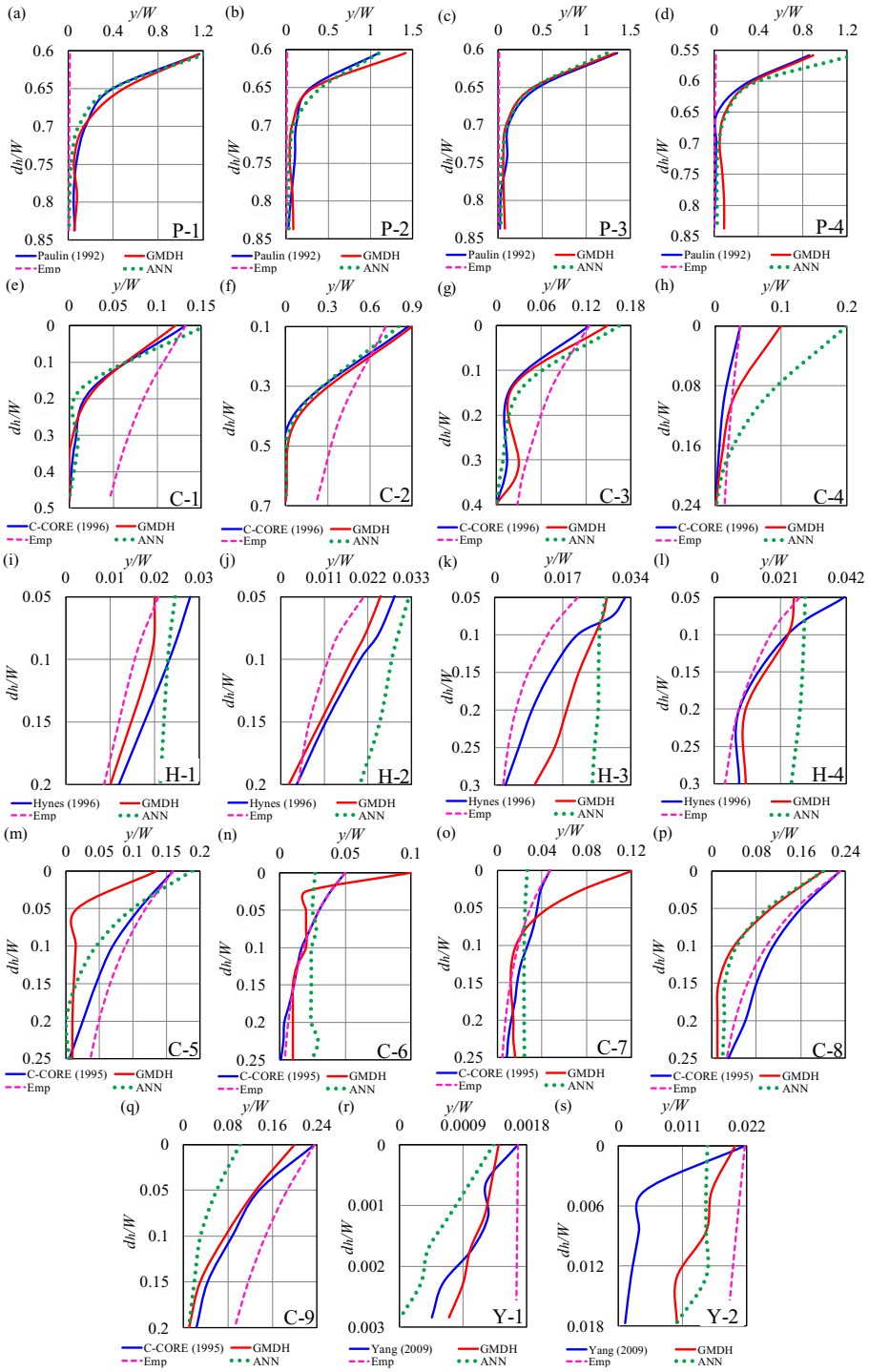


Figure 4-8. Comparison of simulated horizontal deformations by GMDH with ANN and empirical models (a-d) Paulin (1992) (e-h) C-CORE (1996) (i-l) Hynes (1996) (m-q) C-CORE (1995) (r-s) Yang (2009)

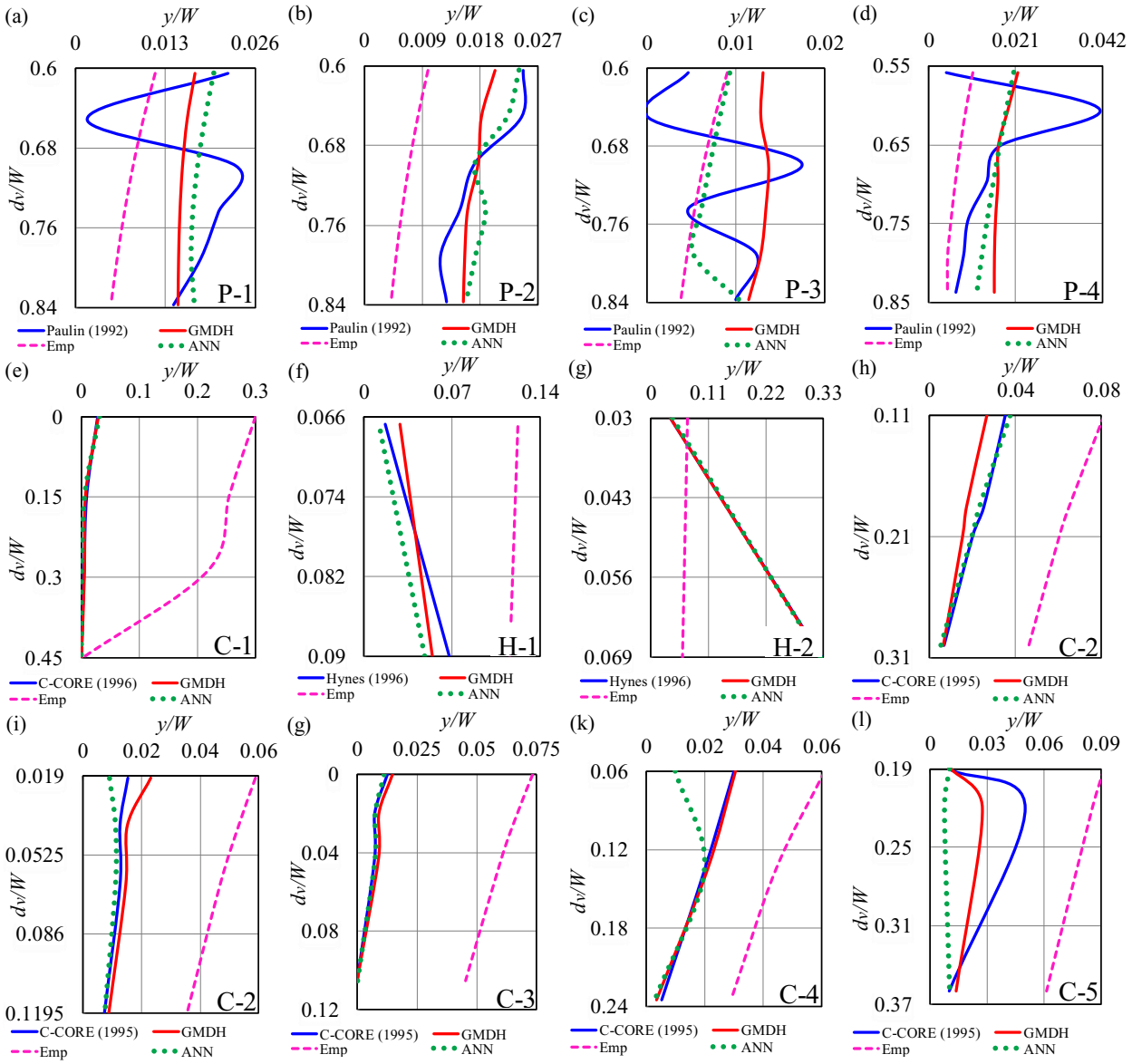


Figure 4-9. Comparison of simulated vertical deformations by GMDH with ANN and empirical models (a-d) Paulin (1992) (e) C-CORE (1996) (f-g) Hynes (1996) (h-l) C-CORE (1995)

The GMDH 1 model has simulated the subgouge deformation profiles by using all input variables although this GMDH model had different performances at dealing with various applied datasets. For instance, the GMDH model simulated the horizontal subgouge deformations reported by the P-1 to P-4 tests with a high level of accuracy since Paulin's tests included all input variables. The

correlation between the GMDH model and the experimental measurements provided by Yang (2009) was quite low (Y-1 and Y-2 tests). The constructed dataset was enough heterogeneous, with a wide range of data and this NN-based model was able to have a broad generalization capability to simulate the complex and non-linear ice-gouging problem. The GMDH model could simulate the C-1 to C-4 tests with higher accuracy in comparison with the C-5 to C-9 tests. Even though the discrepancies between the upper and lower limits of the applied datasets were high, the GMDH model estimated the subgouge deformations with acceptable precision.

Therefore, GMDH 1 had better performance in terms of accuracy, correlation, and complexity in the training mode, while the ANN model showed better performance in some criteria during the testing procedure. However, unlike the GMDH model, the ANN was not able to provide an explicit equation to estimate the subgouge deformations (d_h/W and d_v/W). In addition, during the simulation of the subgouge deformations, the computation time of the GMDH model was significantly less than the ANN model. It was observed in Figure 9, that there is a deviation between the GMDH and test results, where the reported test data exhibits an unexpected noise (probably due to equipment or human error). However, the GMDH has attempted to bypass the noise and prevented the overall trend of the horizontal subgouge deformation with a nonlinear behavior to get affected by the noise.

Furthermore, the empirical models estimated the subgouge deformations just as a function of the gouge geometry and the soil depth. The other governing parameters, such as the shear strength parameters, the attack angle, the keel velocity, the height of the berm, and the load components had not been taken into account by the existing empirical formulas. The empirical model had the worst performance to model the vertical deformations, signifying that this model showed the lowest level of correlation, and precision with experimental values (R=0.047, VAF=-2670.28,

MAE=1.242). It seems that the empirical model was not able to appropriately estimate the vertical subgouge sand deformations in a wide range of experimental values since this model was derived just for specific laboratory measurements. Therefore, the GMDH model was preferable to the ANN technique, and the empirical formulas and the GMDH were introduced as the superior model to estimate the subgouge soil deformations in the current study. Figure 4-10 displays the evolved structure of GMDH 1 to simulate both horizontal and vertical deformations.

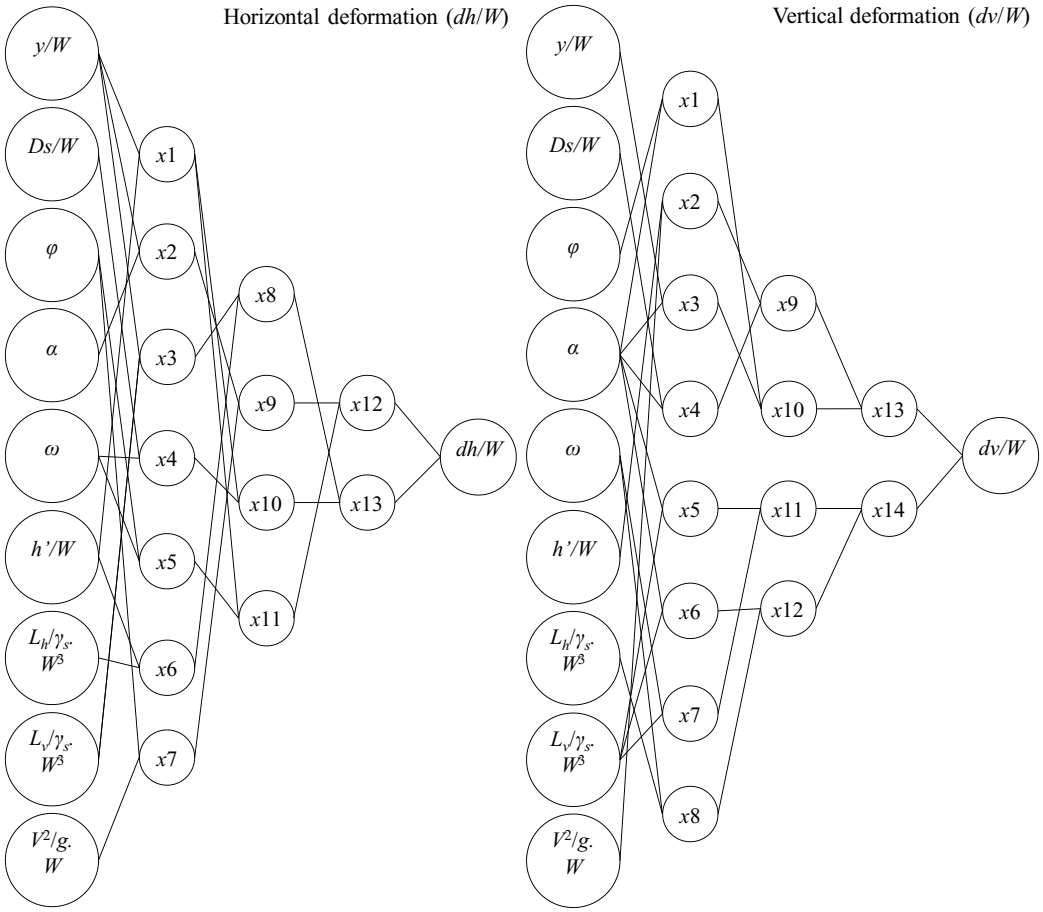


Figure 4-10. Evolved structure of the best GMDH models to predict horizontal (dh/W) and vertical (dv/W) deformations

Unlike the previous empirical models, which were just a function of the soil depth and the gouge geometry, the presented models were a function of more dimensionless subgouge parameters comprising the soil depth ratio (y/W), the gouge depth ratio (D_s/W), the shear strength of seabed soil (φ), the attack angle (α), the angle of the surcharged soil slope (ω), the berm height ratio (h'/W), the horizontal and vertical loads ($L_h/\gamma_s.W^3$ and $L_v/\gamma_s.W^3$), and the ice dynamics ($V^2/g.W$). Thus, these models may pave the way for future research seeking physical governing equations.

4.1.6. Conclusion

In this study, the horizontal and vertical deformations caused by the ice-gouging in sandy seabed were simulated by the group method of data handling (GMDH) method for the first time. Firstly, the parameters governing the subgouge deformations were identified. Then, 60% of constructed datasets were used for training the GMDH models, and 40% of the experimental measurements were employed for testing these models. By conducting a sensitivity analysis (SA) and an uncertainty analysis (UA), the premium GMDH models, and the most influencing input parameters were introduced. Moreover, the results of the GMDH simulation were compared with the ANN and empirical models. The most important results obtained from the current study are summarized as follows:

- GMDH 1 (as a function of all input parameters including y/W, D_s/W, φ, α, ω, h'/W, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$) was detected as the best GMDH model to simulate the horizontal and vertical subgouge deformations.
- The model had an acceptable accuracy in the estimation of the subgouge deformations. For instance, to simulate the horizontal deformations using GMDH 1 in the testing mode, the correlation coefficient (R), the variance accounted for (VAF), and the Nash-Sutcliff

efficiency coefficient (*NSC*) were computed at 0.955, 89.353, and 0.832, respectively. Additionally, the *RMSE*, *MAE*, and *AIC* to estimate the vertical deformations by using GMDH 1 in the testing mode were respectively approximated as 0.007, 0.005, and -54.148.

- Roughly 76% of the simulated horizontal deformations in the testing mode owned an error of less than 5%, whilst 87% of the estimated vertical deformation had an error of less than one percent.
- Regarding the conducted UA, GMDH 1 had an underestimated performance to estimate both horizontal and vertical subgouge deformations.
- In comparison to other GMDH models, GMDH 1 had the narrowest width of uncertainty bound (WUB) to approximate the vertical deformations.
- In comparison to the ANN and the empirical methods, the GMDH model outperformed.

The presented study provided invaluable insight into the simulation of subgouge problems using NN-based models. These results can facilitate proposing of new solutions to model the subgouge sand deformations. Moreover, the current study can be beneficial to the planning of costly experimental studies and time-consuming three-dimensional simulations and then deduct the hefty expenses of future investigations.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through

“Collaborative Research and Developments Grants (CRD)”. Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In Petrov, B.N.; Csáki, F. (eds.), 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, pp. 267–281.

Alba, J.L., 2015. Ice Scour and Gouging Effects With Respect to Pipeline and Wellhead Placement and Design. Bureau of Safety and Environmental Enforcement (BSEE), Wood Group Kenny, Report No. 00100.01.PL.REP.004, Houston, TX.

Azimi, H., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., Khoshbin, F., 2018. Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length. Acta Mechanica, 229(3): 1197-1214.

Azimi, H., Shiri, H., 2020a. Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021b. Modeling Subgouge Sand Deformations by Using Multi-Layer Perceptron Neural Network. 31st International Ocean and Polar Engineering Conference, ISOPE-I-21-2150.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour Displacement in Clays from Physical Model Tests. In 7th International Pipeline Conference, IPC2008-64186, pp. 239-245.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2, St. John's, NL, Canada.

C-CORE., 1996. PRISE Phase 3c: Extreme LEE Gouge Event—Modeling and Interpretation. C-CORE Report 96-C32, St. John's, NL, Canada.

Comfort, G., Graham, B., 1986. Evaluation of Sea Bottom Ice Scour Models. Environmental Studies Revolving Funds, Report No. 37, Arctic Canada Limited.

Ebtehaj, I., Bonakdari, H., Zaji, A. H., Azimi, H., Khoshbin, F., 2015. GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. Eng. Sci. Technol., an International Journal, 18(4): 746-757.

Ebtehaj, I., Bonakdari, H., Zaji, A.H., 2016. A nonlinear simulation method based on a combination of multilayer perceptron and decision trees for predicting non-deposition sediment transport. Water Science and Technology: Water Supply, 16: 1198-1206.

Ebtehaj, I., Bonakdari, H., Sharifi, A., 2014. Design criteria for sediment transport in sewers based on self-cleansing concept, Journal of Zhejiang University Science A, 15(11): 914-924.

Farlow, S.J., 1984. Self-Organizing Method in Modelling: GMDH Type Algorithm. Marcel Dekker, New York.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

ISO 19906., 2019 E. Petroleum and Natural Gas Industries - Arctic offshore structures. Geneva: ISO.

Ivakhnenko, A.G., 1971. Polynomial theory of complex systems. IEEE transactions on Systems, Man, and Cybernetics, (4): 364-378.

Ivakhnenko, A.G., 1976. The group method of data handling in prediction problems. Soviet Automatic Control, 9: 21-30.

Ivakhnenko, A.G., Ivakhnenko, G.A., 1995. The review of problems solvable by algorithms of the group method of data handling (GMDH). Pattern Recognition And Image Analysis C/C Of Raspoznavaniye Obrazov I Analiz Izobrazhenii, 5: 527-535.

Kaydani, H., Najafzadeh, M., Mohebbi, A., 2014. Wellhead choke performance in oil well pipeline systems based on genetic programming. Journal of Pipeline Systems Engineering and Practice, 5(3): 06014001.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. Thirteenth International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. Fourteenth International Offshore and Polar Engineering Conference. ISOPE-I-04-107.

Lach, P.R., 1996. Centrifuge modelling of large soil deformation due to ice scour. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Mouawad, J., 2008. Oil Survey Says Arctic Has Riches. The New York Times.

Najafzadeh. M., Saberi-Movahed, F., 2019. GMDH-GEP to predict free span expansion rates below pipelines under waves. Marine Georesources & Geotechnology, 37: 375-392.

Nazari, A., Rajeev, P., Sanjayan, J.G., 2015. Modelling of upheaval buckling of offshore pipeline buried in clay soil using genetic programming. Engineering Structures, 101, 306-317.

Nematzadeh, A, Shiri, H., 2019a. Large deformation analysis of ice scour process in dense sand. 10th Congress on Numerical Methods in Engineering, At Guimaraes, Portugal.

Nematzadeh, A, Shiri, H., 2019b. Self-correcting soil models for numerical simulation of strain rate dependent ice scour in sand. 9th ECCOMAS Thematic Conference on Smart Structures and Materials, Paris, France.

Nematzadeh, A, Shiri, H. 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Region Science and Technology, 170: 102929.

Nessim, M.A., Jordaan, I.J., 1985. Protection of Arctic Submarine Pipelines against Ice Scour. Journal of Energy Resources Technology, 107(3): 356–361.

Paulin, M.J., 1991. Preliminary Results of Physical Model Tests of ice Scour. Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering (C-CORE), St. John's, NL, Canada.

Paulin, M.J., 1992. Physical Model Analysis of Iceberg Scour in Dry and Submerged Sand, Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Shahin, M.A., Jaksa, M.B., Maier, H.R., 2001. Artificial neural network applications in geotechnical engineering. Australian Geomechanics, 36: 49-62.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip Line Field Solutions as an Approach to Understand Ice Subgouge Deformation Patterns, The Sixteenth International Offshore and Polar Engineering Conference, 628–633.

Stoica, P., Selen, Y., 2004. Model-order selection: a review of information criterion rules. IEEE Signal Processing Magazine, 21(4): 36-47.

Volterra, Vito., 1887. Above the functions that depend on other functions, III, Italy: R. Accademia dei Lincei. pp. 97–105.

Wallis, S., 2013. Binomial confidence intervals and contingency tests: mathematical fundamentals and the evaluation of alternative methods. J Journal of Quantitative Linguistics, 20: 178-208.

Walton, R., Binns, A., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., 2019. Estimating 2-year flood flows using the generalized structure of the Group Method of Data Handling. Journal of Hydrology, 575: 671-689.

Wilson, E.B., 1927. Probable inference, the law of succession, and statistical inference. Journal of the American Statistical Association, 22: 209-212.

Woodworth-Lynes, C., Nixon, D., Phillips, R., Palmer, A., 1996. Subgouge Deformations and the Security of Arctic Marine Pipelines. Offshore Technology Conference, OTC-8222-MS.

Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Section 2

Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling

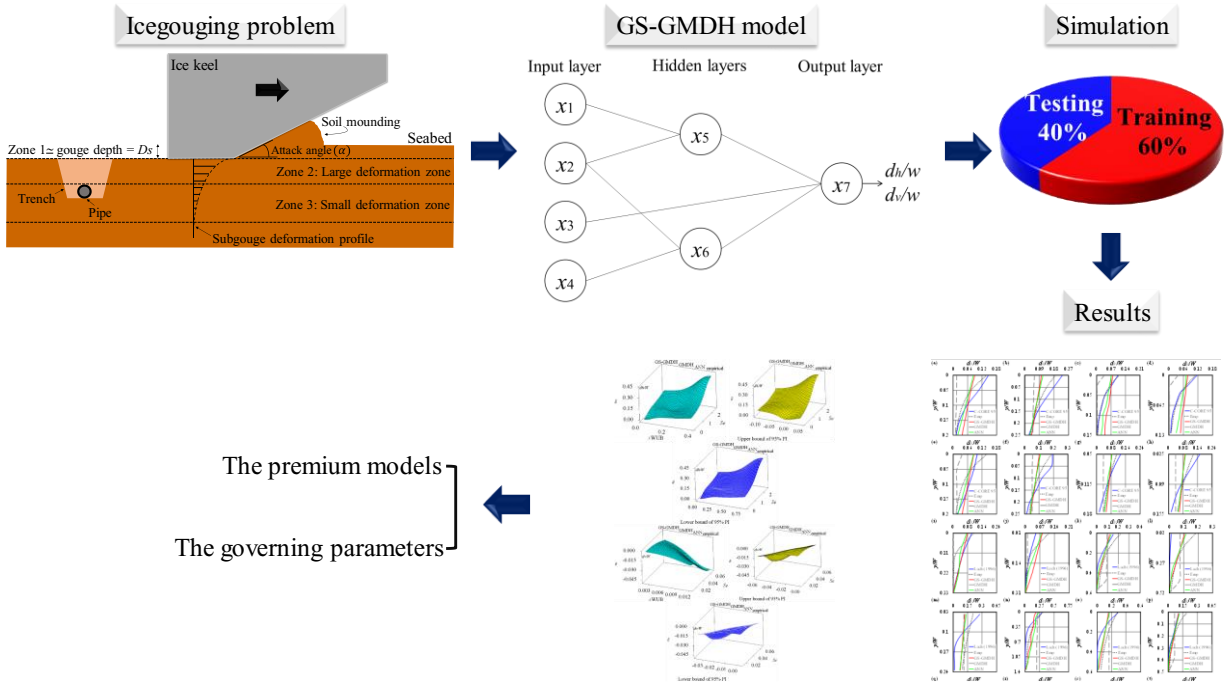
This section is a journal paper which has been published in Cold Regions Science and Technology (2022), Vol. 193, pp. 103426, ISSN: 0165-232X

Abstract

Ice-induced scour is a serious challenge for the subsea pipelines in the Arctic shallow waters. Estimation of the maximum pipeline deformation and its minimum burial depth can guarantee the operational integrity of these structures in ice-prone regions. The pipeline buried below the ice keel is still threatened by subgouge soil deformation that is extended down the ice tip due to the shear resistance of the seabed soil. Determining the subgouge soil deformations is a challenging process that needs costly experimental and numerical simulations. In this paper, an alternative and cost-effective methodology has been proposed using a robust neural network-based method titled "generalized structure of group method of data handling" (GS-GMDH) for the first time to simulate the horizontal and vertical subgouge soil deformation profiles in clay. Using the parameters governing the subgouge soil deformations, nine GS-GMDH models were defined. The premium GS-GMDH models and the most influencing input parameters comprising the soil depth and the gouge geometry were introduced by performing a sensitivity analysis. Subsequently, the results of the best GS-GMDH models were compared with the classical group method of data handling (GMDH), the artificial neural network (ANN), and the empirical approaches. An uncertainty analysis showed that the GS-GMDH slightly overestimates the horizontal and underestimates the vertical subgouge soil deformations. A partial derivative sensitivity analysis (PDSA) was also performed to assess the influence of the input parameters on the subgouge soil deformations. Lastly, a set of GS-GMDH-based equations were proposed for fast estimation of the subgouge soil deformations in clay.

Keywords: Iceberg-seabed interaction; Subgouge deformation in clay; Artificial neural network (ANN); Generalized structure of group method of data handling (GS-GMDH); Partial derivative sensitivity analysis (PDSA); Uncertainty analysis (UA)

Graphical abstract



4.1.7. Introduction

A huge amount of hydrocarbon deposits including crude oil (90 billion barrels) and natural gases (47.3 trillion cubic meters) have been located in the Arctic areas and these resources can be transferred through subsea pipelines from offshore to onshore. In the Arctic's shallow waters, the seabed might be scoured by the traveling icebergs during the warmer seasons; as a result, the subsea pipelines, communications cables, wellheads, and other infrastructures may be gouged by the traveling masses (Mouawad 2008; Azimi and Shiri 2020a). A schematic layout of the ice-induced scour considered in this study is shown in Figure 4-11.

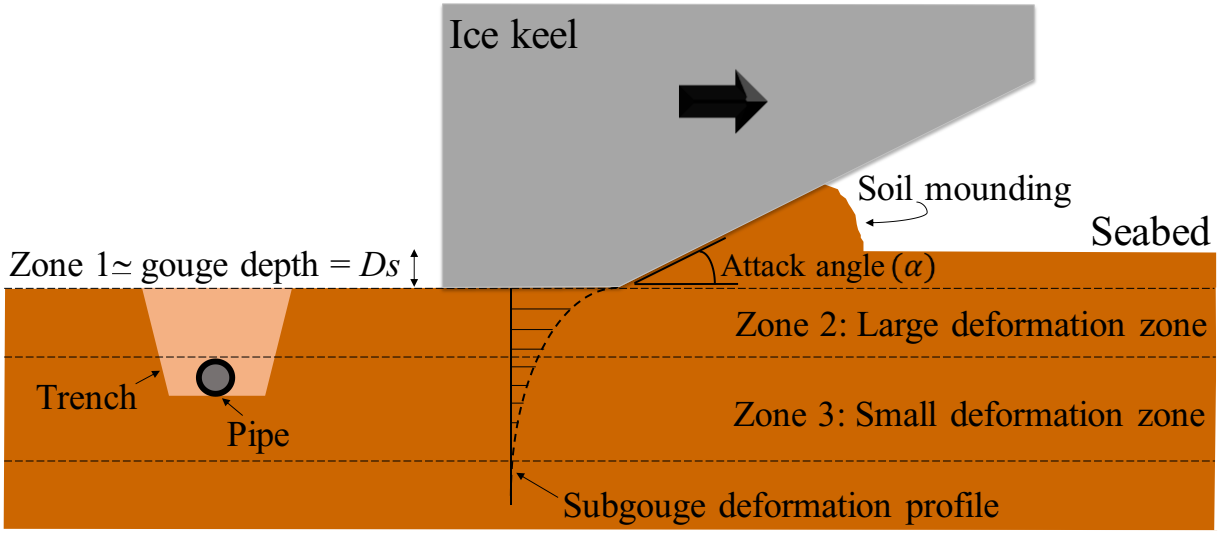


Figure 4-11. Schematic layout of an ice-scouring event

As shown, D_s is the gouge depth and the angle between the ice keel face and the contacting seabed surface is defined as the attack angle (α). The seabed attacked by the ice-gouging is divided into three distinct zones: Zone 1, where the seabed soil is displaced by the gouging keel; Zone 2, where the seabed soil undergoes very large permanent deformations; and Zone 3, where the seabed soil witnesses only elastic strain, without significant deformations. The buried pipelines may be displaced by the subgouge soil deformation which is described by an exponential distribution

through the soil depth underneath the ice tip (ISO 19906: 2019 E). An example of these exponential expressions will be provided in the coming sections. The soil deformation caused by ice-gouging is not limited to the soil in front of the ice keel being directly in contact with the ice. The shear resistance in the soil extends the soil displacement much deeper than the ice tip which is called subgouge soil deformation (see Figure 7-1). Therefore, the subgouge soil deformation may still endanger the pipeline buried deeper than the gouge depth in terms of the serviceability limit state (e.g., ovalization) or ultimate limit state (e.g., local buckling or tensile fracture) (Lach 1996, Nematzadeh and Shiri 2020). The estimation of the additional burial depth below the ice tip for pipeline protection against the subgouge soil deformation is a challenging and costly design aspect that is usually evaluated through a decoupled approach. This approach combines the accuracy of the continuum finite element (FE) analysis of the free field ice-gouging event with the simplicity of the beam-spring models for the simulation of the pipeline. The decoupled approach begins with a free-field ice-gouging analysis without the presence of a pipeline. Then the resultant subgouge soil deformations are manually transferred to the springs representing the soil in a beam-spring model to obtain the pipe response to soil movement. The beam-spring model examines the subgouge soil displacements in various depths that have been obtained from a single free-field FE analysis to come up with the best burial depth. Therefore, free-field ice-gouging is a significant part of the decoupled approach in practical pipeline design that is usually supported by costly computational and experimental studies. The current study introduces a neural network-based (NN-based) method that can be alternatively used for fast and cost-effective estimation of the subgouge soil deformations to be fed into the beam-spring models.

The proper dataset is of significant importance in the success of NN-based analysis. In this paper, the database for training and testing of the NN-based models was extracted from a set of crucial

experimental studies in clay published in the literature including C-CORE (1995), C-CORE (1996), Lash (1996), Schoonbeek et al. (2006), and Been et al. (2008). The study conducted at C-CORE (1995) included a series of free field centrifuge tests of the ice keel scour problems in different seabed conditions including loose sand, dense sand, soft clay, and medium clay. Different configurations of the gouge depth ranging from 1 to 2 meters, the gouge width ranging from 15 to 30 meters, and the attack angle ranging from 15 to 30 degrees were carried out in this study. The experimental measurements were verified with the soil displacements observed in the field and it was proved that the measurements were accurate enough. Lach (1996) performed nine centrifuge tests to study the behavior of saturated clay during ice-gouging events. The author surveyed the effects of the soil stress history, the width of the gouge, the depth of the gouge, and the attack angle on the mechanics of the ice-soil interaction process. The study concluded that the clay displacements depended on the initial state of the soil sample. Another set of experiments was conducted at C-CORE (1996) called "Pressure Ridge Ice Scour Experiment" (PRISE) tests to evaluate the soil characteristics including the soil deformations and the stress changes during the ice-induced events. Furthermore, some empirical relationships were derived from the experimental measurements to estimate the horizontal and vertical clay deformations. The study showed that the gouge geometry had a significant influence on the soil deformations and these formulas underestimated the horizontal displacements in the medium-dense silt and sand.

Schoonbeek et al. (2006) studied the clay behavior during some centrifuge testing and assessed the effects of the undrained shear stress of clay, the ice keel speed, and the depth of scour. The subgouge soil displacements in clay were measured by using the image processing of video camera data. The author asserted that the undrained shear stress was a significant parameter to form the subgouge soil deformations in clay. Been et al. (2008) carried out some centrifuge tests to measure

ice-induced clay displacements. The author also studied two failure mechanisms including steep and shallow keel angles. The investigation concluded that the soil sample was pushed up into a mound in front of the steep ice keel, whereas the scoured clay was forced under and sides of the shallow keel.

Bekker et al. (2005) simulated the effect of the traveling hummocks on the marine pipeline. Using a probabilistic model, the authors determined the bottom configuration of the subsea pipeline track. Arnau and Ivanović (2019) evaluated the impact of load rate on the scouring pattern in seabed sand in both dry and saturated circumstances. The effects of relative density and soil permeability were also assessed in 1g condition experimentally. The study showed that the rate effects grew linearly with the scour depth and the berm height. Moreover, Shin et al. (2019) simulated the ice scouring issue by means of the Coupled Eulerian Lagrangian (CEL) approach. The authors employed the contact condition and geostatic stress to model the ice keel gouging and verified the numerical results with experimental measurements. Nematzadeh and Shiri (2020) modeled the subgouge event in sand adopting the ABAQUS software and the Coupled Eulerian-Lagrangian (CEL) scheme. The authors assessed configurations of ice keel and soil properties and asserted that the higher magnitude of soil deformations stemmed from the soil with high unit weight and relative density. Azimi and Shiri (2020a) determined the parameters affecting the ice-seabed interaction problems by using Buckingham's theory for the first time. A set of linear regression (LR) relationships were provided to estimate the maximum horizontal and vertical deformations both in clay and sand. The author validated their LR models with a comprehensive dataset.

Besides the sophisticated computational methods, same as many other engineering challenges, if properly developed, artificial intelligence (AI) and soft computing (SC) algorithms can be used as potential cost-effective alternatives to analyze the ice-seabed interaction process. These techniques

can be of particular importance at the early stages of design projects, where a fast and fairly accurate estimation is required to decide on methodologies and logistics and plan the scope of detailed engineering works. However, despite many other engineering challenges, only a few AI methods and machine learning (ML) algorithms have been developed to simulate ice-induced problems to date. Kioka et al. (2003) used the artificial neural network (ANN) to estimate the ice scour depth. The authors employed the backpropagation method to train the ANN model and used five neurons in the hidden layer. Additionally, Kioka et al. (2004) integrated the ANN method and a mechanical approach for assessing the ice-gouging process. The bottom shape of the ice and the condition surrounding the ice ridge were detected as the most important variables. Azimi and Shiri (2020b) utilized gene expression programming (GEP) to estimate the horizontal subgouge displacements in the sand. The horizontal ice-induced displacements were simulated in terms of the bearing pressure, the attack angle, the sand dilation angle, and the gouge depth. The authors identified the gouge depth and the dilation index as the most decisive parameters in prediction of subgouge soil deformations and reaction forces. The sand characteristics during the gouging process were simulated using a machine learning (ML) model entitled extreme learning machine (ELM) by Azimi and Shiri (2021a). Nine ELM models were developed and the best ELM models and the most decisive input parameters were introduced by performing several analyses. The author proposed a set of ELM-based equations in order to estimate the ice-gouging features. Azimi and Shiri (2021b) estimated the horizontal subgouge soil deformations in the sand by means of a multi-layer perceptron neural network (MLPNN) model. The study demonstrated that the MLPNN model had a reasonable accuracy, signifying that the vast majority of the simulated subgouge soil deformations possessed an error of less than 5%.

Considering the high cost and challenges associated with the experimental studies and continuum FEA modeling in clay, the alternative robust neural network-based approach, proposed in this study, i.e., the "generalized structure of group method of data handling" (GS-GMDH) can mitigate the expenses of future designs and even improve the planning of the research investigations. Further details associated with the proposed method and the results compared with test data will be discussed in the coming sections.

4.1.8. Material and methods

In this section, the group method of data handling (GMDH), the generalized structure of the group method of data handling (GS-GMDH), and the artificial neural network (ANN) models were detailed. Subsequently, by using dimensional analysis, the parameters governing the ice-induced clay displacements were detected. Lastly, to verify these models, a comprehensive dataset was established.

4.1.8.1. Group method of data handling (GMDH)

The group method of data handling (GMDH) was introduced by Ivakhnenko (1976) for the first time. This approach has been broadly applied in different fields to approximate linear and non-linear problems, without having sufficient information about the system (Ivakhnenko and Ivakhnenko 1995; Ebtehaj et al. 2015; Azimi et al. 2018; Walton et al. 2019). The GMDH is considered a self-organized technique in which a set of neurons are utilized to estimate a target function like f_i with the output O_i by using multi-input $\chi = (\chi_1, \chi_2, \chi_3, \dots, \chi_k)$. This means that a binary connection between different neurons adopting quadratic polynomials results in the

development of the GMDH neural network. Thus, the O_i is expressed for n observed values in each dataset comprising m inputs and one output as follows:

$$O_i = f(\chi_{i1,}\chi_{i2,}\chi_{i3,} \cdots, \chi_{in,}) \qquad i = (1,2,3, \dots, m) \tag{4-29}$$

The target function is minimized by calculating the mean square error (*MSE*) between real subgouge soil displacements (o_i) and estimated values and also the corrected Akaike Information Criterion (AIC_c) is employed for correction of the target function to express the simple models in a noncomplex structure as below:

$$MSE = \sum_{i=1}^{m}\left(O_i\left(\chi_{i1,}\chi_{i2,}\chi_{i3,} \cdots, \chi_{in,}\right) - o_i\right)^2 \rightarrow min \qquad i = (1,2,3, \dots, m) \tag{4-30}$$

Subsequently, the corrected Akaike Information Criterion (AIC_c) is employed for the correction of the target function to express the simple models in a noncomplex structure as the following form:

$$AIC_c = n \times log\left[\sum_{i=1}^{m}\left(O_i\left(\chi_{i1,}\chi_{i2,}\chi_{i3,} \cdots, \chi_{in,}\right) - o_i\right)\right]^2 + 2\zeta + \frac{2\zeta(2\zeta+1)}{n-\zeta-1} \tag{4-31}$$

where, the ζ is the number of optimized parameters during the training procedure. Furthermore, the Volterra functional series as a polynomial function is employed so as to link between input and output parameters as follows:

$$O = \Gamma_0 + \sum_{i=1}^{n}\Gamma_i\chi_i + \sum_{i=1}^{n}\sum_{j=1}^{n}\Gamma_{ij}\chi_i\chi_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\Gamma_{ijk}\chi_i\chi_j\,\chi_k + \cdots \tag{4-32}$$

The Kolmogorov-Gabor polynomial as a quadratic and bivariate form of the Volterra functional series is used in the below equation (Ivakhnenko 1971; Farlow 1984; Nariman-Zadeh et al. 2002):

$$O = G\left(\chi_i, \chi_j\right) = \Gamma_0 + \Gamma_1\chi_i + \Gamma_2\chi_j + \Gamma_3\chi_i^2 + \Gamma_4\chi_j^2 + \Gamma_5\chi_i\chi_j \tag{4-33}$$

where, Γ_i is the unknown coefficients stemmed from regression approaches to minimize the difference between the observed ice-induced deformations (o_i) and the simulated values for each input parameters pair χ_i and χ_j (Farlow 1984). All the unknown coefficients are calculated using the least square (LS) method by solving the following matrix:

$$\Delta\Gamma = O \tag{4-34}$$

where, the Γ is the unknown coefficient vector of the Kolmogorov-Gabor polynomial $(\Gamma_0, \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5)$ and the o is considered the vector of the real subgouge soil deformations in clay, $\{o_0, o_1, \dots, o_m\}^T$, thus the Δ matrix is expressed in the form of Eq. (4-9) as below:

$$\Delta = \begin{bmatrix} 1 & \chi_{1p} & \chi_{1q} & \chi_{1p}^2 & \chi_{1q}^2 & \chi_{1p}\chi_{1q} \\ 1 & \chi_{2p} & \chi_{2q} & \chi_{2p}^2 & \chi_{2q}^2 & \chi_{2p}\chi_{2q} \\ m & m & m & m & m & m \\ 1 & \chi_{mp} & \chi_{mq} & \chi_{mp}^2 & \chi_{mq}^2 & \chi_{mp}\chi_{mq} \end{bmatrix} \tag{4-35}$$

where, the superscript T signifies the transpose of the matrix. The vector of the unknown coefficients is estimated through the least-square (LS) technique deriving from the multiple regression analysis and can be expressed as follows (Najafzadeh and Saberi-Movahed 2019):

$$\Gamma = (\Delta^T\Delta)^{-1}\Delta^T o \tag{4-36}$$

The described procedure is iterated for each particular neuron of the next hidden layer regarding the linkage topology of the GMDH network (Najafzadeh and Saberi-Movahed 2019).

4.1.8.2. Generalized structure of GMDH (GS-GMDH)

The classical GMDH has several drawbacks including the use of the quadratic polynomial, limiting the inputs of each neuron to two, and limiting the inputs of each neuron to the use of the adjacent layer neurons. These disadvantages lead to the reduction of the modeling performance of the classical GMDH model along with the development of complex structures to simulate multidimensional non-linear problems (Walton et al. 2019). To overcome these challenges, the generalized structure of the group method of data handling (GS-GMDH) method was coded in the MATLAB software environment to estimate the ice-induced clay deformations in the current study. Figure 4-12 displays the structure of the GS-GMDH topology where a neuron in a particular layer is inputted from the non-adjacent and the adjacent layer simultaneously. The GS-GMDH

model used both the quadratic and cubic polynomials, applied two and three inputs in each neuron, and took input neurons from both the adjacent and the non-adjacent layers. Lastly, the premium GS-GMDH model was selected by means of the corrected Akaike Information Criterion (AIC_c), signifying that the simplicity and the accuracy of the chosen model were simultaneously evaluated. These features enhanced the performance of the GS-GMDH model compared to the classical GMDH.

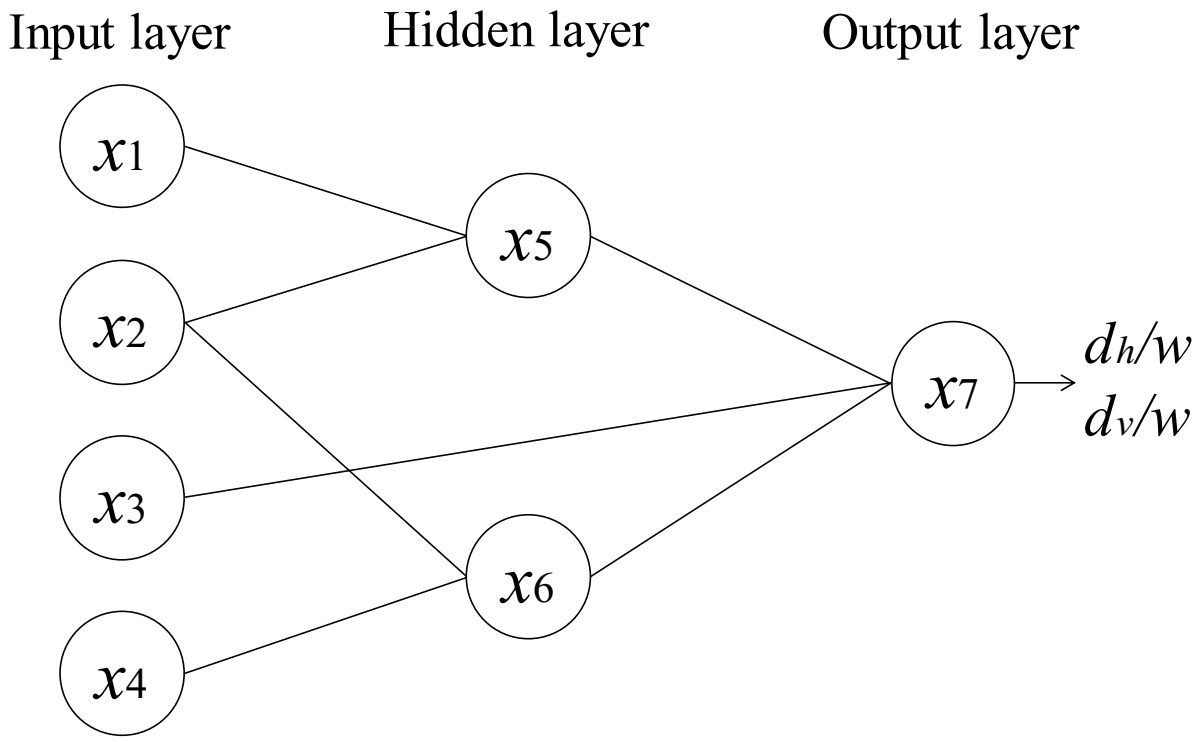


Figure 4-12. Structure of the GS-GMDH model

Figure 4-13 portrays the flowchart of the used GS-GMDH model in the current study. Firstly, the constructed dataset was divided into two sub-samples as the training and testing data. It is worth noting that four sub-samples including 50%, 60%, 70%, and 80% of the constructed dataset were respectively utilized to train the GS-GMDH models. However, the sub-sample consisting of 60%

training the models and 40% testing them was chosen in this study since the GS-GMDH models had a better performance during the modeling process. The number of inputs for each neuron, the maximum number of neurons in the layers, and the maximum number of the layers were tuned in the next step.

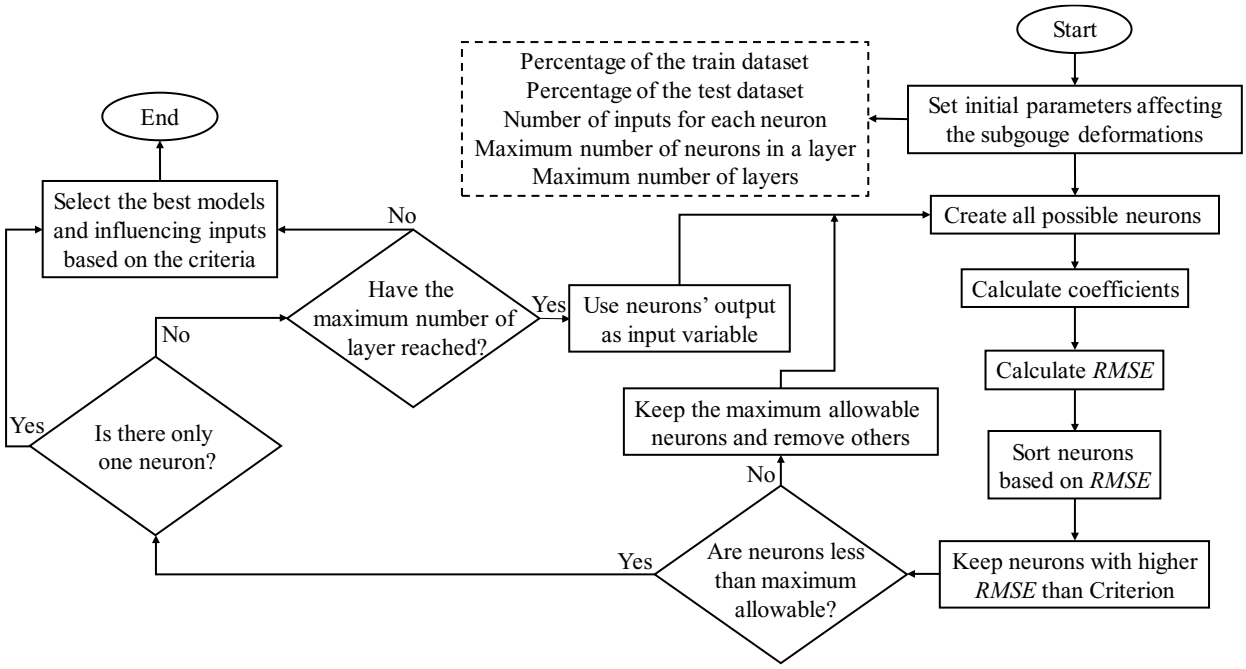


Figure 4-13. Flowchart of the applied GS-GMDH model

4.1.8.3. Artificial neural network (ANN)

Artificial neural networks (ANNs) are inspired by the biological neuron networks in which a problem is solved similar to the way the human brain works. Owing to the flexibility of the ANN model to predict complex problems, the ANN has been extensively applied to predict different high-dimensional and complicated parameters (Haykin 1994; Shahin et al. 2001; Guresen et al. 2011). It is noteworthy that a detailed description of the ANN is beyond the scope of the current study. In general, each neural network model has some layers, including an input layer, at least a hidden layer, and an output layer. Input parameters are connected to the network through the input

layer. Thus, the number of neurons in the input layer is equal to the number of input parameters of the problem. In addition, the number of neurons in the output layer equals the number of target functions, the subgouge soil deformation in clay (Shahin et al. 2001). Although there is no regulation to determine the number of neurons in the hidden layer (Azimi and Shiri 2021b), the number of neurons in the hidden layer is a function of the problem's complexity (Al-Alawi et al. 1998; Vujicic et al. 2016). In the current study, the number of neurons in the input layer equaled eight input parameters including y/W, D_s/W, $c/\gamma_s.W$, α, h'/W, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$, while one neuron in the output layer was considered as the target parameter (clay deformations or d/W). Moreover, a trial and error approach was employed to choose the number of neurons in the hidden layers (Azimi and Shiri 2021b). In other words, the number of neurons in the hidden layer was initially considered equal to one and the precision of the ANN model was evaluated by increasing this value to 12. The most optimum number of hidden layer neurons was selected to be seven since the accuracy of the ANN model was then insignificantly altered. It is worth noting that the number of neurons in the antecedent layer was gathered by using the summation of weighted neurons in the hidden and output layers and transferred to the next layer through an "activation function" (Smith 1993). In the current study, the "linear", "sigmoid", and "hyperbolic tangent" activation functions were applied in different models and then the accuracy of these activation functions was evaluated. The applied activation functions are set as below:

$$linear(x) = x \tag{4-37}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{4-38}$$

$$\tan(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{4-39}$$

Subsequently, an optimum activation function for the ANN model was selected, meaning that the hyperbolic tangent function was used for the estimation of the ice-induced clay displacements since the accuracy of this activation function was better than others. Furthermore, the values of weighted outputs were summed by means of the multilayer perceptron throughout a procedure called "model training". This means that the "Levenberg-Marquardt (LM) algorithm" was applied to training the multilayer perceptron neural network, where the biases and weights were adjusted by means of the back-propagation algorithm. The number of epochs was considered to be 5000 as the ANN model converged at some point, where the convergence was achieved at epoch number 1200.

4.1.8.4. Subgouge soil deformations in clay

Ice-induced deformation in clay seabed is a function of the scour depth (*Ds*), the undrained shear strength (c), the width of gouge (*W*), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (h'), the resultant horizontal force (L_h) and the resultant vertical force (L_v), the velocity of ice keel (V), and the specific weight of sand (γ_s) as follows (Azimi and Shiri 2020a):

$$d_{max(clay)} = f_1(D_s, c, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \quad (4\text{-}40)$$

The L_h and L_v represent the resultant horizontal and vertical driving loads comprising drag force from wind, drag force from current, buoyancy force, keel weight, friction force on the bottom of the keel, Coulomb's passive friction force acting in front and on both sides of the keel, and driving force from the surrounding floe that are all amongst the input parameters to the ice-gouging problem (Duplenskiy and Gudmestad 2013; Azimi and Shiri 2020a). It is worth mentioning that in all of these tests, the ice keel has been a rigid body moving in the horizontal direction under a

steady state condition with a restrained vertical motion and rotation. The ice-seabed interaction has been initialized and configured to properly simulate the desired ice driving force resulting from the ice dynamics. Therefore, the ice dynamics have been inherited in the dataset used for the current study and considered in the input parameters of the proposed algorithm. So, Eq. (7-12) can be written as a function of eight dimensionless groups in the following form:

$$\frac{d_{max(clay)}}{W} = f_2\left(\frac{D_s}{W}, \frac{c}{\gamma_s W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{4-41}$$

here, $d_{max(clay)}$ is considered the maximum subgouge displacement in clay formed just under the moving ice keel in the gouge centerline. It is worth noting that the ω and h' parameters were not measured by the previous experimental surveys. The maximum subgouge soil deformation occurs right under the ice keel and exponentially reduces with the soil depth (y) (ISO 19906: 2019 E). Therefore, to obtain the subgouge soil deformation (d) in a given depth, the desired soil depth (y) needs to be fed into the analysis as an input parameter. At a greater depth on the subgouge centerline, by incorporating the soil depth (y/W), the subgouge soil deformation ($d_{(clay)}/W$) can be written as follows (Azimi and Shiri 2020a):

$$\frac{d_{(clay)}}{W} = f_2\left(\frac{y}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s W}, \alpha, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{4-42}$$

Therefore, the introduced dimensionless groups in Eq. (4-42) were utilized as the input parameters in the GS-GMDH models. Furthermore, to verify the artificial intelligence models (GS-GMDH, GMDH, ANN), a comprehensive database including five experimental datasets were employed in the current study. The dataset includes the measurements reported by C-CORE (1995), Lach (1996), C-CORE (1996), Schoonbeek et al. (2006), and Been et al. (2008). The constructed dataset in the current study contains 162 cases, where 113 experiments were used for training and 49 experiments for testing the models. A large number of publications are in the literature on the

application of the GMDH for simulation of the regression problems with a similar size dataset (e.g., Ebtehaj et al. 2017; Azimi et al. 2018; Najafzadeh and Saberi-Movahed 2019, Walton et al. 2019). These studies have shown that the selected data size is sufficient for the GMDH algorithm to model different regression problems with acceptable performance. Thus, the number of experimental measurements applied in the current study was assumed to be sufficient for reliable training and testing of the ML models. In addition, the good results of a "Goodness to fit" analysis that will be discussed in the next section show the sufficiency of the size of the database.

The range of dimensionless horizontal subgouge deformations (d_h/w) for the C-CORE (1995) dataset comprising C-1 to C-8 tests was from 0.003 to 0.233 and these deformations were measured at the dimensionless soil depth up to 0.025. The keel attack angle in the C-1, C-2, C-4, C-5, C-6, and C-8 tests was reported as 15°, while this value for C-3 and C-7 was measured as 30°. The gouge depth ratio (D_s/W) in the C-1 to C-8 tests were respectively obtained at 0.053, 0.099, 0.033, 0.033, 0.054, 0.095, 0.079, and 0.080. The berm height ratio was not reported in the C-1 to C-8 tests.

The L-1 to L-8 tests were reported in Lach's datasets. The value of attack angle in L-1, L-2, L-3, L-4, L-5, L-6, and C-8 was 15° but this parameter in the L-7 test was measured as 25°. In the L-1, L-4, and L-7 tests, the value of the undrained shear strength ratio was 12.029, whilst this parameter for L-2 and L-3 was similar, with the $c/\gamma_s W$ of 9.452. The internal friction angle of sand seabed in L-5 L-6, and L-8 were respectively reported at 17.185, 24.059, and 15.037.

Moreover, the C-9 to C-12 tests have measured by C-CORE (1996) in which the value of the undrained shear strength ratio was respectively measured at 0.168, 0.167, 0.333, and 0.333. The attack angle in the C-9 test was 30°, whereas this parameter in C-10, C-11, and C-12 was observed

to be 15°. The berm height ratio in the C-9 to C-12 tests was not reported. The value of the ice velocity ratio ($V^2/g.W$) in the C-9 to C-12 tests was constant at 0.015.

In Schoonbeek's experimental setup (S-1), the gouge depth ratio, the undrained shear strength ratio, and the attach angle value were measured at 11.667, 13.637, and 14°, respectively.

The value of dimensionless horizontal subgouge deformations in Been et al.'s model (B-1 to B-5 tests) varied from 0.004 to 0.330 and the two models (B-2 and B-3) had the keel attack angle of 15°, while the attack angle in the B-1 test was equal to 45°. In the B-4 and B-5 tests, the value of the attack angle was reported at 30°. Been et al. (2008) utilized a clay seabed with an undrained shear strength ratio of 1.444.

To identify the most influencing input parameters and introduce the best GS-GMDH models, a sensitivity analysis (SA) was carried out, meaning that eight GS-GMDH models including GS-GMDH 1 to GS-GMDH 8 were defined. In other words, GS-GMDH 1 estimated the subgouge soil deformations in clay by using all input parameters, comprising y/W, D_s/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$. Figure 4-14 illustrates the applied input combinations, as well as the developed artificial intelligence models.

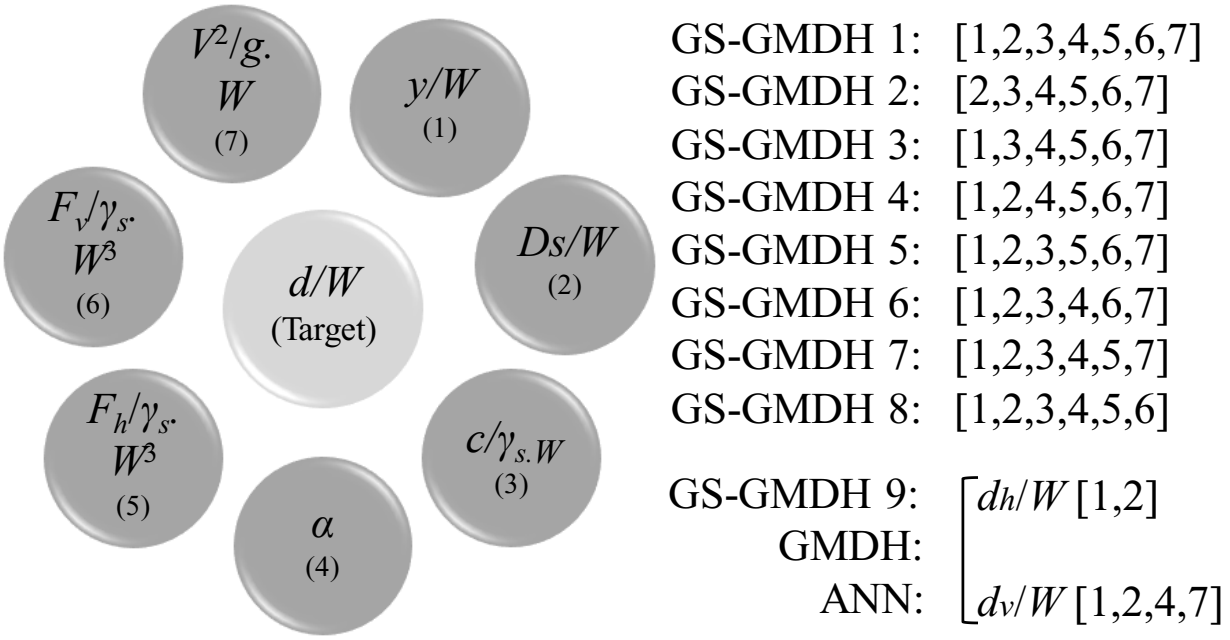


Figure 4-14. Combination of the input parameters to develop different models

Then, the effect of each input parameter was eliminated and GS-GMDH 2 to GS-GMDH 8 were developed. To estimate the horizontal clay deformations (d_h/W) using the GS-GMDH model, y/W and D_s/W were detected as the most significant input parameters, whereas y/W, D_s/W, α, and $V^2/g.W$ had the highest effects to simulate the vertical clay deformations (d_v/W). Thus, GS-GMDH 9 as the premium model was defined with a combination of the detected effective parameters. It is worth mentioning that the classical GMDH and ANN models were also developed using these influential input variables.

4.1.8.5. Goodness of fit

To assess the performance of the GS-GMDH models, some statistical indices including correlation coefficient (R), the variance accounted for (VAF), root mean square error ($RMSE$), mean absolute

error (*MAE*), scatter index (*SI*) and Nash-Sutcliff efficiency coefficient (*NSC*) were utilized as follow:

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{4-43}$$

$$VAF = \left(1 - \frac{var(P_i - O_i)}{var(P_i)}\right) \times 100 \tag{4-44}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{4-45}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{4-46}$$

$$SI = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}}{\bar{O}} \tag{4-47}$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{4-48}$$

Here, O_i, P_i, $\bar{O}$, $\bar{P}$, and n are the observed ice-induced deformations, the estimated deformations, the average observed values, the average estimated values, and the number of experimental measurements, respectively. The closeness of the correlation coefficient (*R*) and the Nash-Sutcliff efficiency coefficient (*NSC*) to one signified that the GS-GMDH model had the highest correlation with the observed values, whilst the closeness of the *RMSE*, *MAE*, and *SI* criteria to zero meant that the particular model showed the lowest level of error. Furthermore, the superior GS-GMDH model had the highest value of the *VAF* criterion though the complexity of these GS-GMDH models was not shown through the applied indices. To overcome this problem, the Akaike Information Criteria (*AIC*) was employed as follows:

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (4\text{-}49)$$

where, k is the number of estimated factors included in the GS-GMDH model (Walton et al. 2019; Sudheer and Jain 2003). The less complex GS-GMDH model had the lowest value of the *AIC* index, hence, the superior GS-GMDH model had the lowest value of the *AIC* index and error (*RMSE*, *SI*, *MAE*), and the highest level of correlation with the experimental measurements (*R* and *NSC*).

4.1.9. Results and discussion

For GS-GMDH 1 to GS-GMDH 9, a sensitivity analysis was performed and then the superior model along with the most effective input parameters was introduced. After that, the premium GS-GMDH models were compared with the GMDH, ANN, and empirical approaches. Subsequently, a set of GS-GMDH-based equations for the estimation of the horizontal and vertical deformations were provided. Lastly, a partial derivative sensitivity analysis for the best GS-GMDH models was implemented.

4.1.9.1. Sensitivity analysis (SA)

- **Horizontal soil displacement**

Figure 4-15 shows the calculated statistical indices for the horizontal clay deformation by using GS-GMDH 1 to GS-GMDH 9 in the testing mode. GS-GMDH 1 was a function of all input variables and the magnitude of computed *RMSE*, *SI*, and *MAE* of this model in testing mode was 0.162, 0.249, and 0.074, respectively. Moreover, GS-GMDH 2 simulated the target function by using D_s/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$, signifying that the effect of the soil depth

parameter (y/W) was ignored. The *AIC*, *NSC*, and *VAF* values for GS-GMDH 2 in the testing stage were respectively computed at 36.246, 0.575, and 77.081.

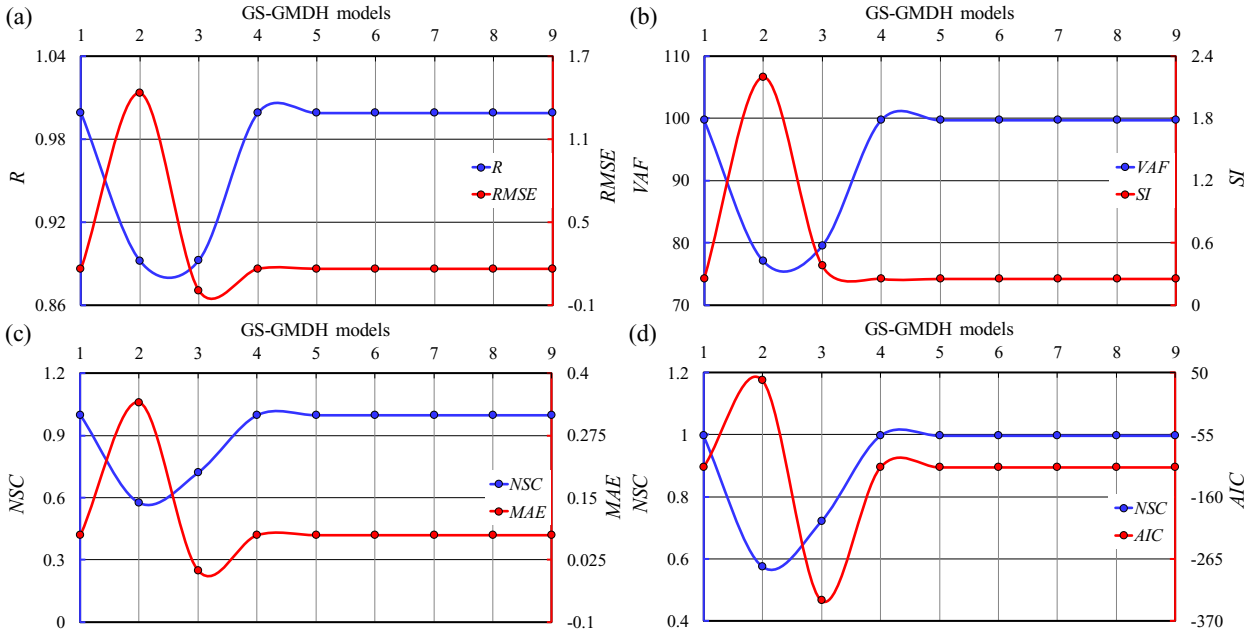


Figure 4-15. Results of statistical indices (a) *R* and *RMSE* (b) *VAF* and *SI* (c) *NSC* and *MAE* (d) *NSC* and *AIC* for the horizontal deformations simulated by GS-GMDH 1 to GSGMDH 9 in the testing mode.

Moreover, the GS-GMDH 3 with the *R*, *SI*, and *RMSE* of 0.892, 0.377, and 0.004 in the testing process predicted the horizontal displacements by means of y/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$. It is worth noting that the influence of gouge geometry (D_s/W) for GS-GMDH 3 was eliminated. Additionally, GS-GMDH 4 was a function of y/W, D_s/W, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and $V^2/g.W$, whereas the impact of shear strength parameter ($c/\gamma_s.W$) was removed for this model. The values of *AIC*, *NSC*, and *VAF* for testing GS-GMDH 4 were respectively estimated to be -110.359, 0.997, and 99.709. Moreover, GS-GMDH 5 modeled the target function in terms of y/W, Ds/W, $c/\gamma s.W$, $Lh/\gamma s.W^3$, $Lv/\gamma s.W^3$, and $V^2/g.W$, meaning that the effect of the ice keel's attack angle (α)

was ignored for GS-GMDH 5, with the *RMSE* and *MAE* of 0.162 and 0.074. The horizontal component of load ($L_h/\gamma_s.W^3$) was the removed input in GS-GMDH 6, where this model comprised the parameters y/W, Ds/W, $c/\gamma s.W$, α, $Lv/\gamma s.W^3$, and $V^2/g.W$, with a scatter index of 0.249. In GS-GMDH 7, the key input variables included y/W, Ds/W, $c/\gamma s.W$, α, $Lh/\gamma s.W^3$, and $V^2/g.W$, signifying that the vertical component of load ($Lv/\gamma s.W^3$) was eliminated. For testing the GS-GMDH 7, the values of *RMSE* and *NSC* were equal to 0.162 and 0.997. For GS-GMDH 8, the effect of velocity ratio ($V^2/g.W$) was ignored and just the parameters y/W, Ds/W, $c/\gamma s.W$, α, $Lh/\gamma s.W^3$, and $Lv/\gamma s.W^3$ were kept to simulate the target values.

Regarding the performed sensitivity analysis, the soil depth parameter (y/W) and the influence of gouge geometry (Ds/W) were detected as the most influencing input parameters to estimate the horizontal deformations. Thus, GS-GMDH 9 as a function of y/W and Ds/W was defined. The values of the correlation coefficient, Nash-Sutcliff efficiency coefficient, and scatter index for GS-GMDH 9 were respectively calculated to be 0.999, 0.997, and 0.249.

- **Vertical soil displacement**

Figure 4-16 shows the computed statistical indices for the simulated vertical deformations by the GS-GMDH models. The value of *RMSE*, *MAE*, and *SI* for testing GS-GMDH 1 was respectively estimated to be 0.017, 0.013, and 0.631. Furthermore, in the testing mode, the value of Nash-Sutcliff efficiency coefficient for GS-GMDH 2, GS-GMDH 3, and GS-GMDH 4 was equal to -1.076, -0.236, and 0.483, respectively. Regarding the performed SA, GS-GMDH 2 owned the highest level of error, with a low range of correlation with experimental values. The variance accounted for (*VAF*) values in GS-GMDH 5 and GS-GMDH 6 for modeling the vertical

displacements in the testing mode were 51.493 and 61.994. Besides, the scatter index (*SI*) for GS-GMDH 7 and GS-GMDH 8 were surmised to be 0.631 and 0.860, respectively.

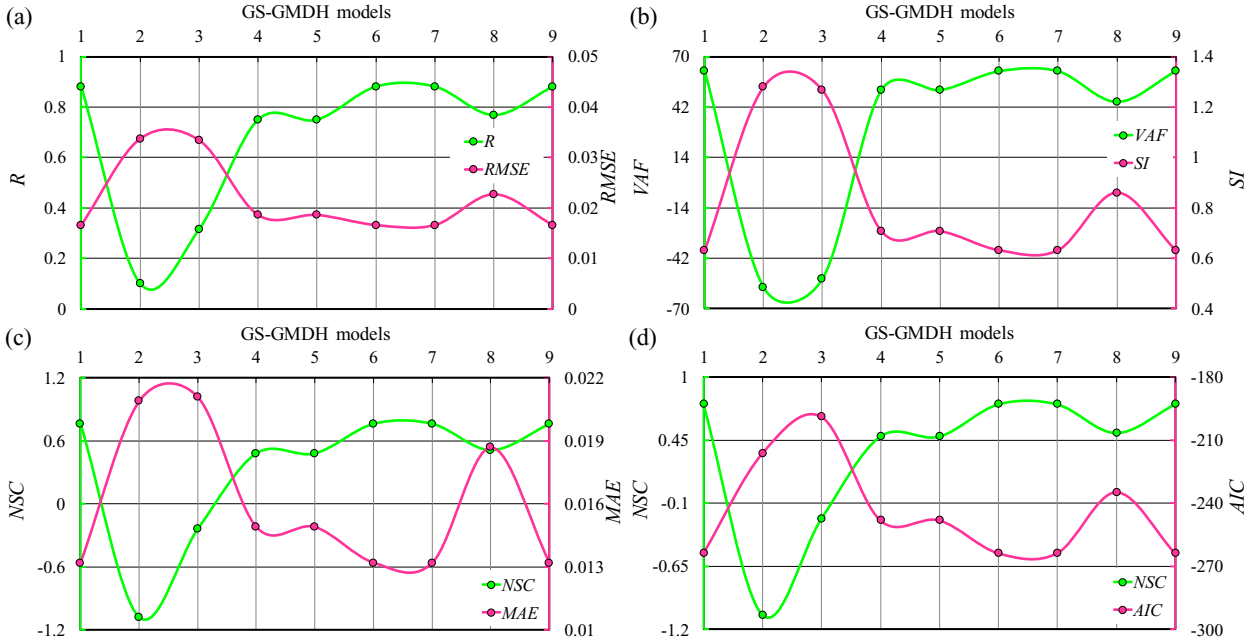


Figure 4-16. Results of statistical indices (a) *R* and *RMSE* (b) *VAF* and *SI* (c) *NSC* and *MAE* (d) *NSC* and *AIC* for the vertical deformations simulated by GS-GMDH 1 to GSGMDH 9.

Thus, GS-GMDH 9 as the best model was developed by a combination of the y/W, Ds/W, α, and $V^2/g.W$ inputs. During testing GS-GMDH 9, the values of *AIC*, *NSC*, and *R* were respectively estimated to be -263.883, 0.764, and 0.881.

4.1.9.2. Comparison of GS-GMDH with GMDH, ANN, and empirical models

- **Sub-scouring deformation profiles**

The superior GS-GMDH model (GS-GMDH 9) was compared with the GMDH, ANN, and empirical (Emp) models. It is worth noting that Woodworth-Lynas et al. (1998) suggested two

equations in order to estimate the horizontal and vertical subgouge soil deformations in clay seabed as follows:

$$d_{h(y)} = 0.6(W.D_s).exp\left(-\frac{2}{3}.\frac{y}{D_s}\right) \quad (4\text{-}50)$$

$$d_{v(y)} = D_s.exp\left(-\frac{1}{3}.\frac{y}{D_s}\right) \quad (4\text{-}51)$$

where, d_h is the maximum horizontal subgouge soil deformation, W is the width of the gouge, D_s is the depth of the gouge, d_v is the maximum vertical subgouge soil deformation, and y is the soil depth. Figure 4-17 displays the profiles of horizontal deformations estimated using the GS-GMDH, GMDH, ANN, and empirical models. The GS-GMDH simulated the horizontal ice-induced displacements with better performance, with a higher range of correlation and a lower range of error in comparison with the GMDH, ANN, and empirical approaches. For instance, the correlation coefficient (R) for the GMDH, ANN, and the empirical models was respectively computed at 0.978, 0.984, and -0.162. Furthermore, for the GMDH, ANN, and empirical models, the mean absolute error (*MAE*) was estimated to be 0.269, 0.272, and 0.469, respectively. The GS-GMDH model showed a better performance to simulate the horizontal deformations in a wide range of experimental measurements, whilst the GMDH and empirical models had an acceptable accuracy in some experimental cases. For example, the GS-GMDH model simulated the experimental measurements reported by C-CORE (1995), Lach (1996), and C-CORE (1996) with higher accuracy, while the GMDH, ANN, and the empirical models had an acceptable performance just for some particular experimental models.

Additionally, the performance of the GS-GMDH model was quite better compared to other classical NN-based models. For instance, the value of the *NSC* index for the GS-GMDH model in comparison with the GMDH and ANN models improved by approximately 10% and 29%,

respectively. The GS-GMDH model showed better performance at dealing with Schoonbeek et al.'s measurements (S-1). Although Been et al.'s experimental values were simulated using a linear behavior (B-1 to B-5), the GS-GMDH model could estimate the test results reported by C-CORE with both linear and non-linear patterns (C-1 to C12). As shown, the GS-GMDH model utilized the monotonic and non-monotonic trends to predict the experimental values measured by Lach (1996) (L-1 to L8). The profiles of simulated vertical deformations using the GS-GMDH, GMDH, ANN, and empirical models are displayed in Figure 4-17.

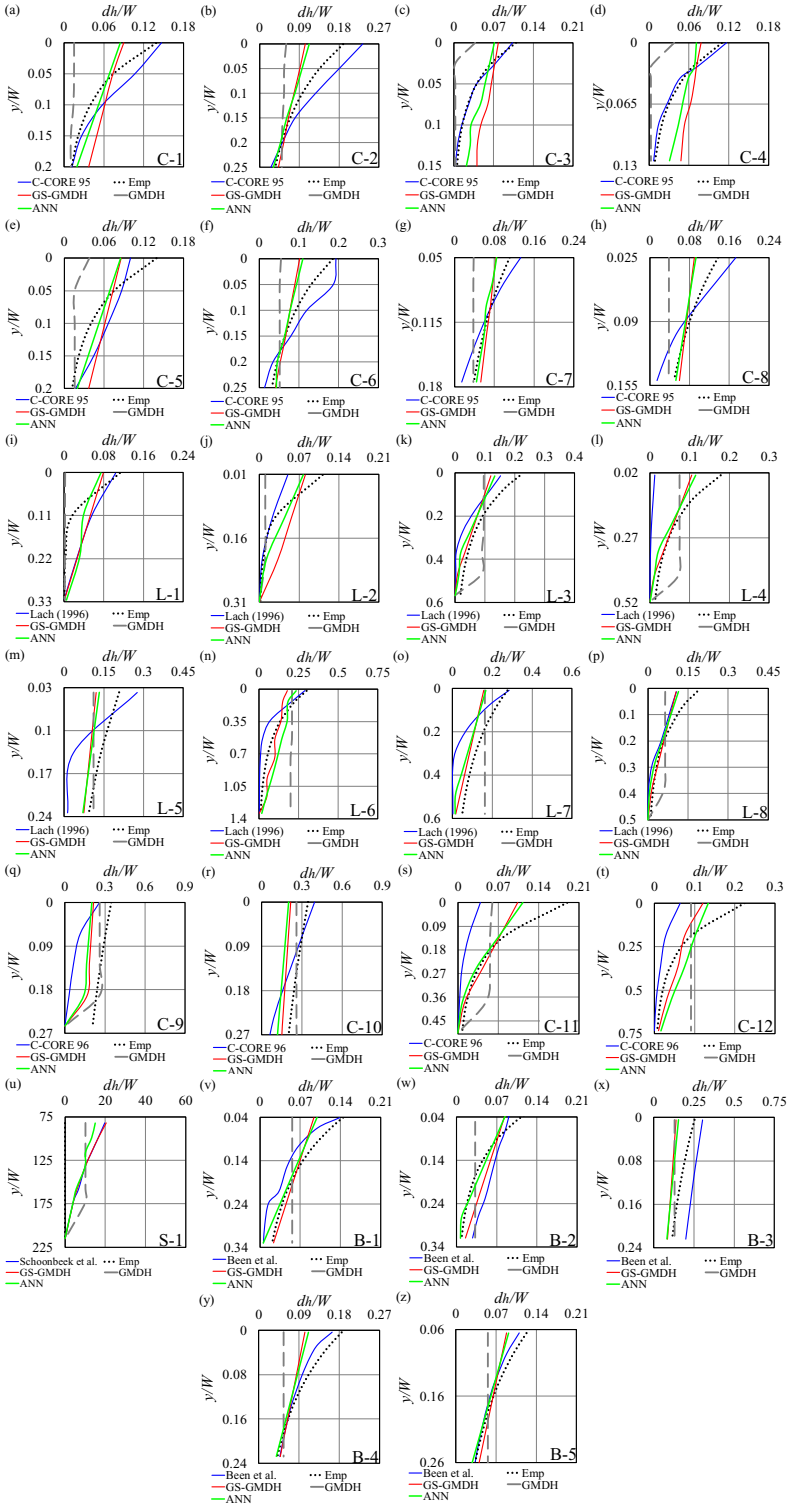


Figure 4-17. Comparison of the experimental horizontal deformations (a-h) C-CORE (1995) (i-p) Lach (1996) (q-t) C-CORE (1996) (u) Schoonbeek et al. (2006) (v-z) Been et al. (2008) with estimated values by GS-GMDH, GMDH, ANN, and empirical models

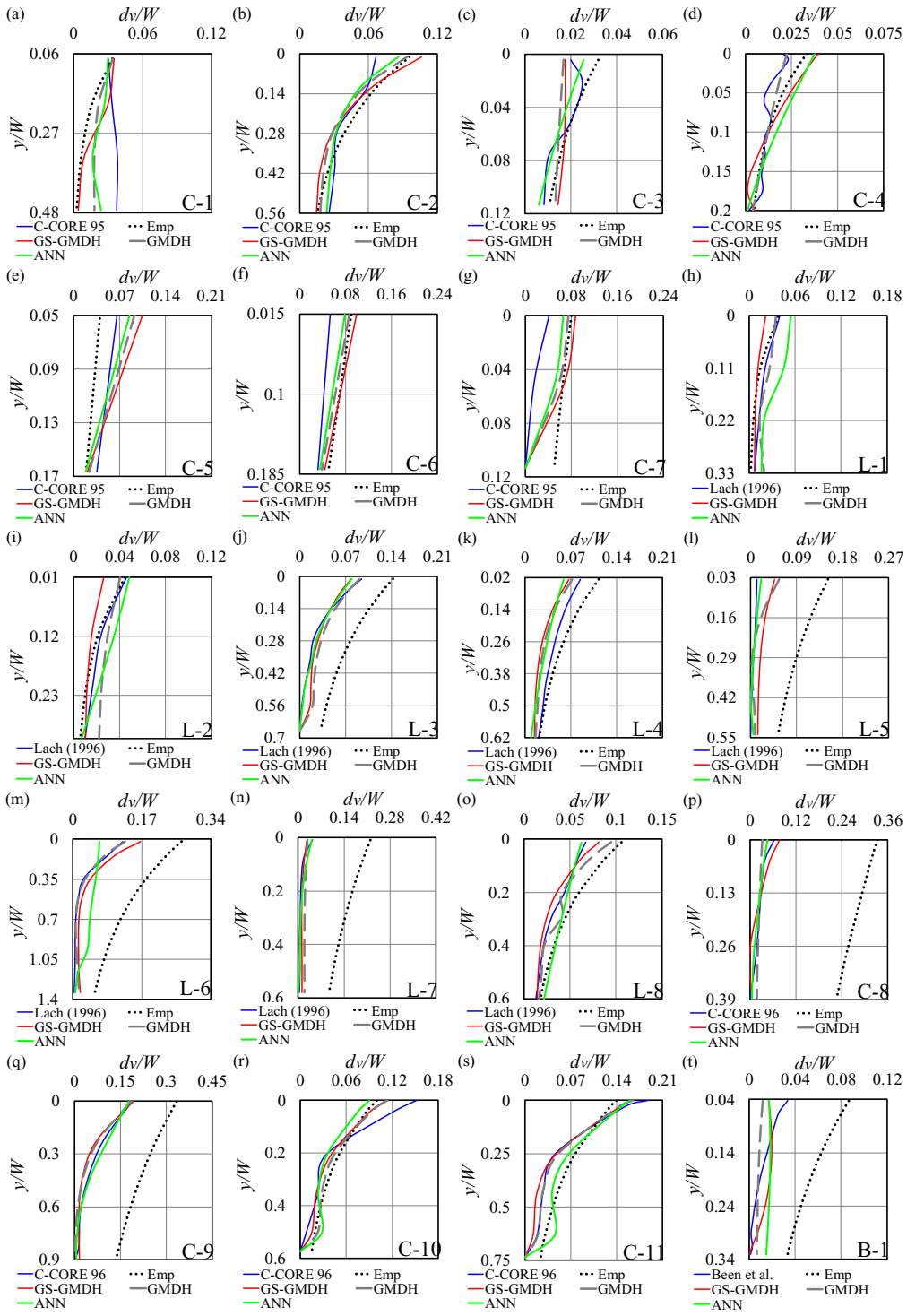


Figure 4-18. Comparison of the experimental vertical deformations (a-g) C-CORE (1995) (h-o) Lach (1996) (p-s) C-CORE (1996) (t) Been et al. (2008) with estimated values by GS-GMDH, GMDH, ANN, and empirical models

The GS-GMDH estimated the vertical displacements with better precision compared to the GMDH, ANN, and empirical models. The value of the correlation coefficient (*R*) for the GMDH, ANN, and empirical models was calculated at 0.849, 0.823, and 0.445, respectively. For the GMDH, ANN, and empirical models, the variance accounted for (*VAF*) index was respectively equal to 57.567, 39.093, and 19.632.

Unlike the GS-GMDH model, other techniques were not able to simulate the vertical displacements with acceptable accuracy in a wide range of experimental values. This means that the GS-GMDH predicted the observed vertical deformations reported by C-CORE (1995), Lach (1996), C-CORE (1996), and Been et al. (2008) with a high level of correlation along with a low level of error. However, the empirical model lacked a good performance at dealing with the simulation of the observed values reported by Lach (1996), C-CORE (1996), and Been et al. (2008). Additionally, the GMDH model could just simulate the experimental measurements reported by Lach (1996), C-CORE (1996), and Been et al. (2008) with appropriate precision.

The GS-GMDH model managed to simulate Lach's measurements (L-1 to L-8) with the highest level of precision. Several fluctuations were observed in the experimental values reported by C-CORE (1995) (C-1 to C-7), whereas this NN-based model could estimate these measurements with acceptable correlation. Moreover, the GS-GMDH model used a non-linear behavior for the estimation of the C-CORE (1996) dataset (C-8 to C-11).

From an engineering practical point of view, the GS-GMDH algorithm could estimate the ice-seabed interaction mechanism with good performance and extremely lower cost. This methodology may be a viable alternative for the initial phases of design projects since a quick and precise estimation is required to plan the scope of a detailed engineering scheme. To illustrate the superiority of the applied NN-based algorithm in comparison with the experimental and numerical

studies, the allocated time to perform the centrifuge test (e.g., Lach 1996, Test #5), FE simulation (e.g., Pike 2016, Case #4), and NN-based model (GS-GMDH) along with the accuracy of these models for simulation of the maximum horizontal subgouge deformation ($d_{max(h)}/W$) was compared.

In Table 4-7, the computation time of the FE analysis and the GS-GMDH algorithm and the accuracy of these models in comparison with the centrifuge test were tabulated. It should be stated the time needed for the centrifuge tests including the test design, model preparation, clay consolidation, instrument calibration, etc., is much greater than the reported time for the centrifuge test in the documents. Similarly, in FE analyses, additional time should be spent on the model construction and model validation. As shown in Table 4-7, the required time in GS-GMDH modeling was extremely lower than the centrifuge tests and the FE analyses. Moreover, The GS-GMDH model was able to predict the maximum horizontal deformation much more accurately than the FE analysis.

Table 4-7. The computation time of the FE analysis and the GS-GMDH algorithm and the accuracy of these models in comparison with the centrifuge test

Model	0.25Ds	1.25Ds	2.25Ds	Time required	Accuracy (0.25Ds)	Accuracy (1.25Ds)	Accuracy (3.25Ds)	Time-saving
Centrifuge test (Lach 1996, Test#5)	0.276	0.055	0.028	4.7 hrs.	-	-	-	-

FE analysis (Pike 2016, Fine mesh)	0.500	0.045	0.022	16 hrs. 8 min	18.8%	81.8	78.6	+343% more than the centrifuge test
GS-GMDH	0.239	0.039	0.018	4 min	86.6%	70.9	64.3	-7050% less than the centrifuge test

Therefore, the GS-GMDH was detected as a robust model so as to estimate both the horizontal and vertical deformation, with a better performance in terms of accuracy, correlation, complexity, and simulation time.

4.1.9.3. Error analysis (EA)

Figure 4-19 illustrates the error analysis of the GS-GMDH, GMDH, ANN, and empirical models to simulate the horizontal and vertical deformations. Approximately 75% of the horizontal deformations estimated by the empirical model produced an error of less than 6%, while this value for the GS-GMDH model was roughly 78%. Although about 99% of the results obtained from the GS-GMDH model had an error of less than 20%, this value for GMDH and ANN models was 93% and 97%, respectively.

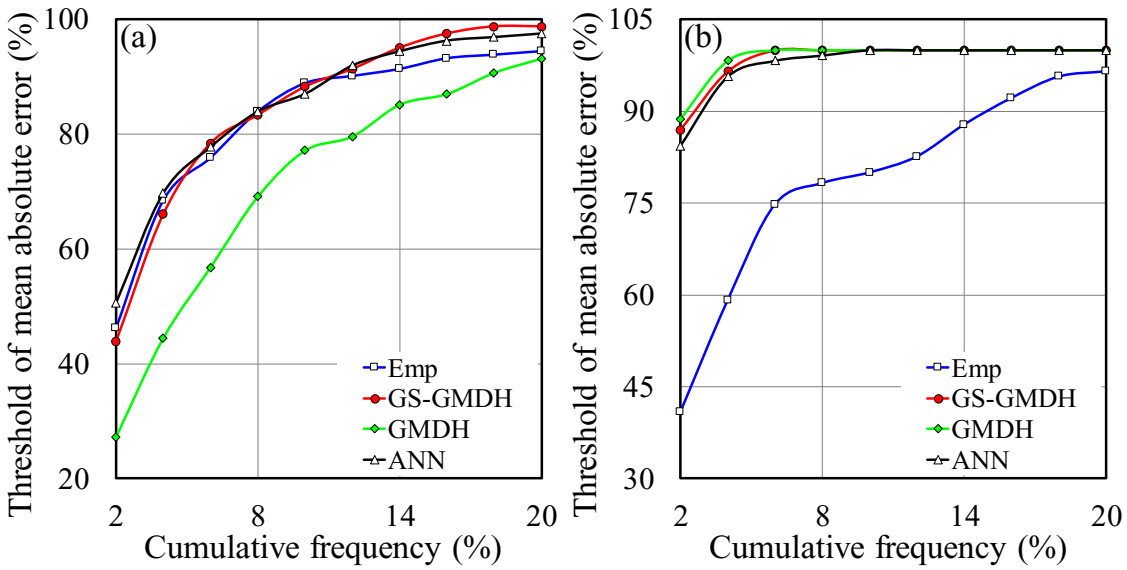


Figure 4-19. Error distribution of (a) estimated horizontal deformations (dh/W) and (b) estimated vertical deformations (dv/W) by empirical, GS-GMDH, GMDH, and ANN models

The error distribution of the GS-GMDH, GMDH, and ANN models for the estimation of the vertical deformations was almost the same. For instance, nearly all the predicted vertical displacements by the GS-GMDH and GMDH models produced an error of less than 6%. However, the error distribution showed that about 92% of the results from the empirical formula had an error of less than 16%. Thus, the lowest level of error to model the horizontal and vertical displacements was obtained for the GS-GMDH model.

4.1.9.4. Uncertainty analysis (UA)

Uncertainty analysis (UA) was performed so as to assess the performance of the GS-GMDH, GMDH, ANN, and empirical models and analyze the errors. The errors estimated by these models (e_j) were considered as the difference between the simulated (P_j) and the real (O_j) ice-induced clay displacements as follows:

$$e_j = P_j - O_j \tag{4-52}$$

Moreover, the mean value of the calculated error $(\bar{e})$ and the standard deviation (S_e) were computed as below (Azimi and Shiri 2020a):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{4-53}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{4-54}$$

The negative sign of the mean value of the calculated error $(\bar{e})$ signifies the underestimated performance of a model, the simulated values are less than the real values. However, the positive sign denotes the overestimated performance of this model, meaning that the simulated displacements are greater than the actual deformations. Additionally, a confidence bound using $\bar{e}$ and S_e was produced around an estimated error by means of the "Wilson score method" (Azimi and Shiri 2020a) without the continuity correction. The Wilson score interval is an improvement over the normal distribution interval in which an asymmetric normal distribution is employed to enhance the confidence interval bound. Then, ±1.96Se results were expressed in a 95% confidence bound which was a 95% estimated error interval (95% confidence interval). The results obtained from the performed uncertainty analysis for the GS-GMDH, GMDH, ANN, and empirical models are illustrated in Figure 4-20. To predict the horizontal displacements, the GS-GMDH, GMDH, ANN, and empirical models overestimated the target value, with the lowest level of the mean value of the calculated error $(\bar{e})$ for the GS-GMDH. To simulate the horizontal deformations, the narrowest and the widest width of the uncertainty bound were found as ±0.016 and ±0.382 for the GS-GMDH and the empirical model, respectively. In addition, the 95% confidence interval for the GS-GMDH model was calculated to be from -0.014 to 0.017.

In contrast, the mean value of the calculated error to simulate the vertical displacements adopting the GS-GMDH, GMDH, ANN, and empirical models were respectively computed at -0.001, -0.001, -0.002, and -0.047. In other words, these models had an underestimated performance to model the vertical deformations. It is worth mentioning that the highest level of $\bar{e}$ was computed for the empirical models. The width of the uncertainty bound for the GS-GMDH model was narrower than the empirical and the ANN models. Moreover, the 95% confidence interval for the GS-GMDH and GMDH models was the same, from -0.004 to 0.001.

Therefore, the GS-GMDH overestimated the horizontal deformations, with the narrowest width of uncertainty bound, whilst this model had an underestimated performance at dealing with the simulation of the vertical deformations.

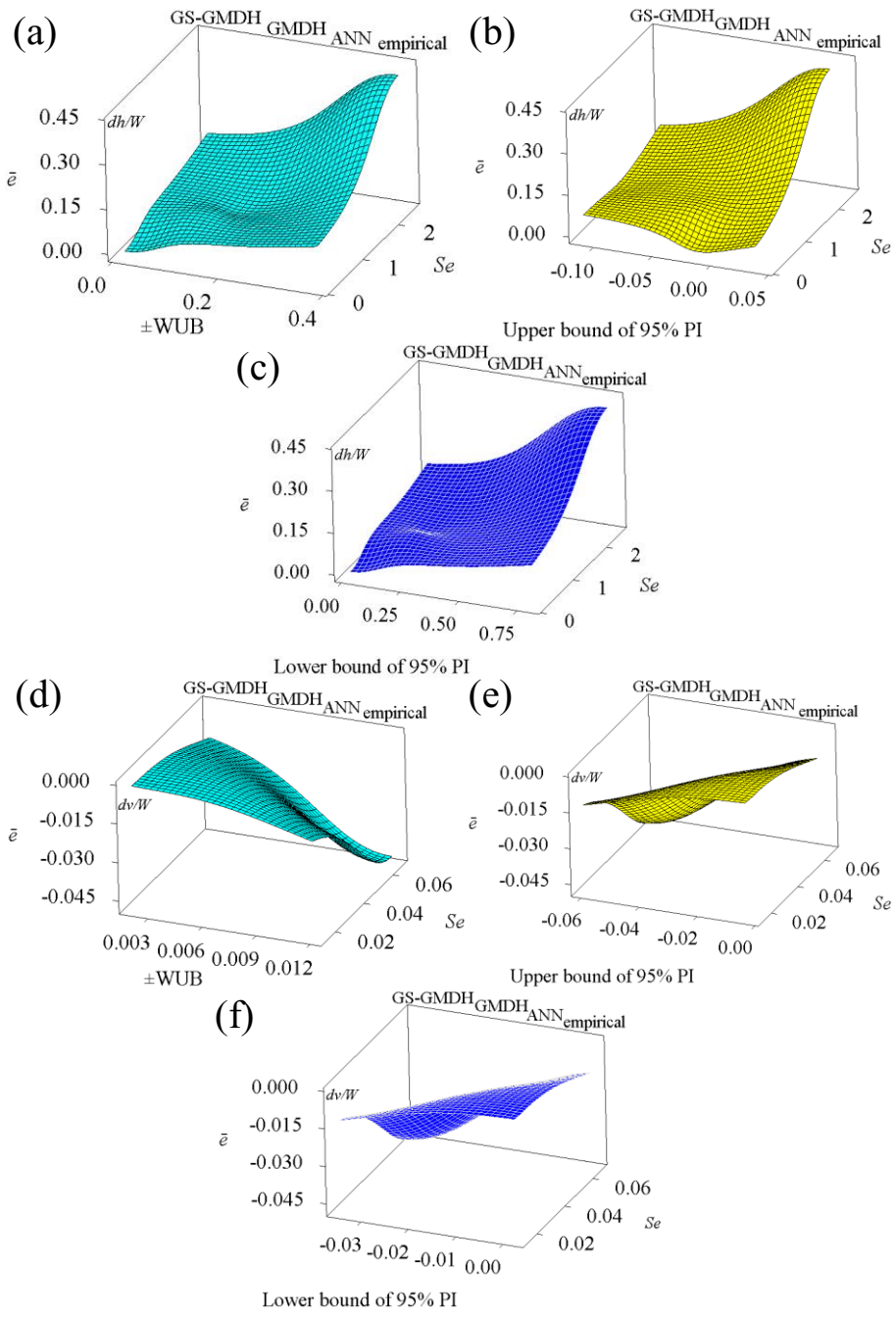


Figure 4-20. Uncertainty analysis results of empirical, GS-GMDH, GMDH, ANN models (a) e and ±WUB for the horizontal deformations (b) e and upper bound of 95%PI for the horizontal deformations (c) e and lower bound of 95%PI for the horizontal deformations (d) e and ±WUB for the vertical deformations (e) e and upper bound of 95%PI for the vertical deformations (f) e and lower bound of 95%PI for the vertical deformations

4.1.9.5. Partial derivative sensitivity analysis (PDSA)

To evaluate the influence of each input parameter as a dependent variable on subgouge soil deformations in clay as the independent parameter, a partial derivative sensitivity analysis (PDSA) for the superior GS-GMDH model (GS-GMDH 9) was carried out. The PDSA commonly shows how the dependent variables can influence the independent parameter and it is defined as follows:

Horizontal deformation: $\partial(d_h/W)/(\partial x_i)$ $x_i = \frac{y}{W}, \frac{D_s}{W}$ (4-55)

Vertical deformation: $\partial(d_v/W)/(\partial x_j)$ $x_j = \frac{y}{W}, \frac{D_s}{W}, \alpha, \frac{V^2}{g.W}$ (4-56)

A negative sign of the PDSA signifies a decreasing impact on the clay deformations, whereas a positive sign of the PDSA shows an increasing influence on the target function. The implemented PDSA for the horizontal deformations predicted by the premium GS-GMDH model is depicted in Figure 4-21.

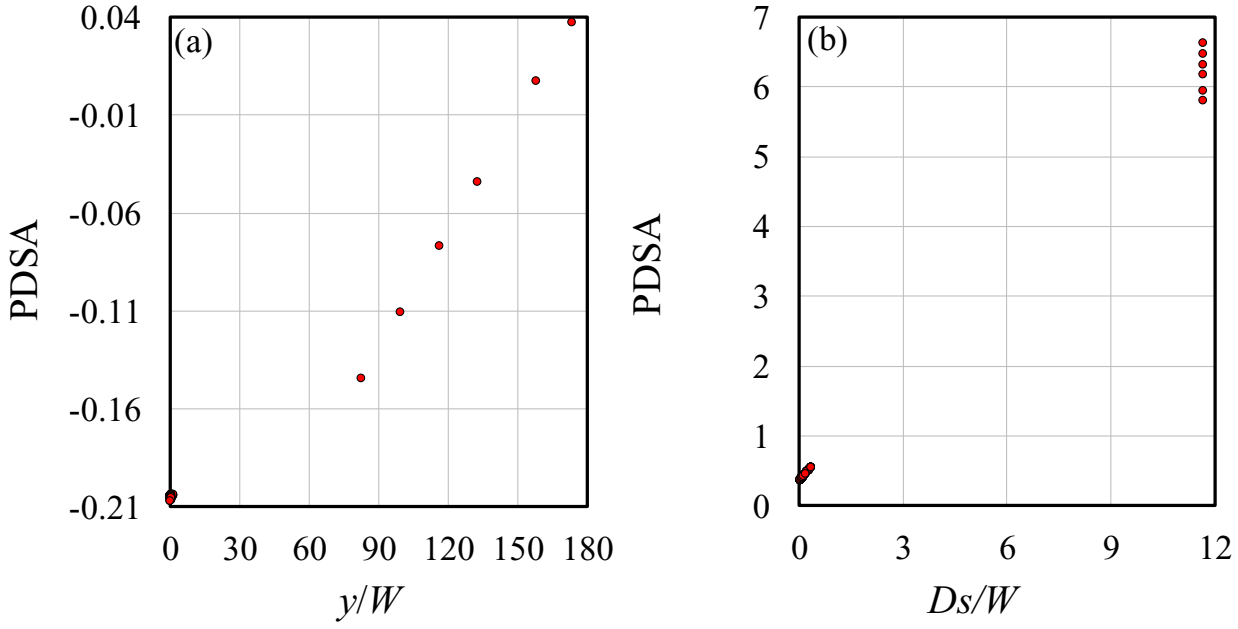


Figure 4-21. PDSA of the horizontal deformations for (a) y/W (b) Ds/W

The results of PDSA for the dimensionless soil depth parameter (y/W) were negative, signifying that the horizontal clay displacements (dh/W) decreased with increasing the dimensionless soil depth. In addition, all computed PDSA for the dimensionless gouge depth ratio (Ds/W) were positive. This means that the horizontal deformations grew with increasing the gouge depth parameter. The performed PDSA for the vertical deformations simulated by GS-GMDH 9 is illustrated in Figure 4-22.

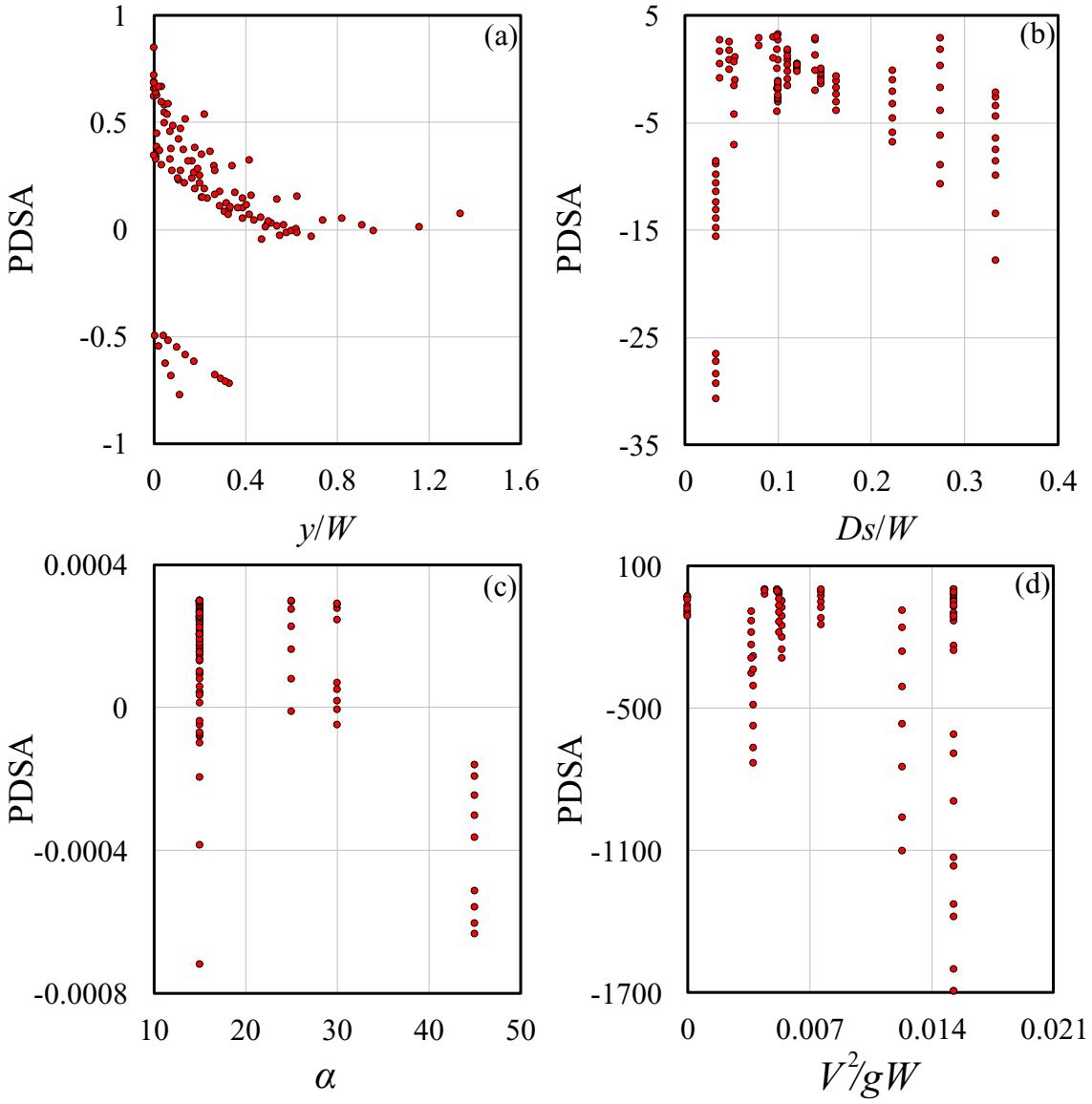


Figure 4-22. PDSA of the vertical deformations for (a) y/W (b) Ds/W (c) α (d) V^2/gW

Most obtained PDSA for the soil depth ratio (y/W) had a positive sign, with a downward pattern. Likewise, the vast majority of the taken PDSA for the attack angle (α) had a positive sign, meaning that the vertical clay deformations grew with increasing the attack angle. However, almost all computed PDSA for the gouge depth ratio (Ds/W) and the velocity parameter ($V^2/g.W$) were negative.

4.1.9.6. GS-GMDH-based equations

GS-GMDH 9 was introduced as the premium model to simulate the ice-induced clay deformations, with the lowest level of error and complexity along with the highest level of correlation with the experimental measurements. Thus, a set of GS-GMDH-based equations were presented in order to calculate the target function. The best GS-GMDH model was a function of the soil depth ratio (y/W) and the gouge geometry (D_s/W) to estimate the horizontal deformations as below:

$$d_h/W = 0.066 + 0.346(D_s/W) - 0.243(y/W) - 0.009(y/W)(D_s/W) + 0.301(D_s/W)^2 + 0.001(y/W)^2 \quad (4\text{-}57)$$

In contrast, to modeling the vertical ice-induced clay deformations, the best GS-GMDH was as a function of y/W, D_s/W, α, and $V^2/g.W$ as follows:

$$d_v/W = 0.014 - 0.109(x7) + 0.054(x6) + 31.9(x6)(x7) - 9.948(x7)(x7) - 7.897(x6)(x6)$$

where,

$$x6 = -0.001 - 0.551(V^2/g.W) + 0.689(D_s/W) - 0.108(y/W) + 46.202(D_s/W)(V^2/g.W) - 6.426(y/W)(V^2/g.W) - 0.030(y/W)(D_s/W) - 6.327(V^2/g.W)^2 - 3.142(D_s/W)^2 + 0.101(y/W)^2$$

and,

(4-58)

$$x7 = 0.111 - 0.006(\alpha) + 0.645(D_s/W) - 0.306(y/W) - 0.021(D_s/W)(\alpha) + 0.010(y/W)(\alpha) - 0.655(y/W)(D_s/W) + 0.0001(\alpha)^2 + 0.277(D_s/W)^2 + 0.184(y/W)^2$$

Therefore, the GS-GMDH was a robust model to simulate the ice-induced clay deformations, with superior performance in comparison with the classical GMDH, ANN, and the empirical models. This means that the GS-GMDH had a low level of complexity along with a high level of precision and correlation with the observed values to estimate the horizontal and vertical displacements.

4.1.10. Conclusion

In this paper, the ice gouge deformations in clay seabed were estimated by means of a robust method entitled "generalized structure of group method of data handling (GS-GMDH). Using the factors affecting the ice-induced displacements in clay, nine GS-GMDH models were developed. To verify the GS-GMDH models, a comprehensive dataset was established and then 60% of the data were employed to train the GS-GMDH models, while 40% of the rest were adopted as the testing sub-sample. By using a sensitivity analysis, the superior GS-GMDH model along with the most influencing input parameters were introduced. Subsequently, the results of the GS-GMDH, the classical GMDH, the artificial neural network (ANN), and the empirical approaches were compared, and then an uncertainty analysis was performed. Lastly, a partial derivative sensitivity analysis (PDSA) for the proposed GS-GMDH-based model was implemented. The most significant results of the current study are summarized below:

- The best GS-GMDH models simulated the subgouge soil deformations in clay with a high level of precision. For instance, the values of correlation coefficient (*R*), the variance accounted for (*VAF*), and Nash-Sutcliff efficiency coefficient (*NSC*) to model the horizontal displacements were respectively obtained to be 0.999, 99.709, and 0.997.

Moreover, for the simulation of the vertical deformations, the values of the scatter index (*SI*) and the mean absolute error (*MAE*) were calculated at 0.631 and 0.013.

- To predict the horizontal deformations, the premium GS-GMDH model was a function of y/W and D_s/W, whereas the best model estimated the vertical displacements in terms of y/W, D_s/W, α, and $V^2/g.W$.
- In comparison with the GMDH, ANN, and empirical models, the GS-GMDH showed a better performance to estimate both the horizontal and vertical displacements, with a low level of complicity and a high level of accuracy and correlation.
- The GS-GMDH model overestimated the horizontal deformations, with an underestimated performance to predict the vertical values.
- The PDSA denoted that the horizontal clay displacements were reduced by increasing the soil depth ratio. However, by increasing the attack angle, the vertical deformations grew.
- A set of GS-GMDH-based equations to estimate the subgouge clay deformation was provided for practical uses.
- The GS-GMDH demonstrated several advantages including: (1) choosing the most influential input parameters and providing an explicit relationship for the problem (2) showing better topological performance and provision of simultaneous input by a hidden neuron from the non-adjacent and adjacent layers (unlike the classical version of neural network-based models), and (3) having a simple organization so that only a few parameters in the GS-GMDH structure are set to simulate the ice-scoured deformation.
- Besides a high level of correlation/accuracy and a low degree of complexity in the simulation of the subgouge soil deformation in clay, the GS-GMDH model has some limitations as well. For instance, the architecture of the GS-GMDH algorithm is not

sufficiently flexible. This requires the algorithm to find the number of layers and neurons during the solving procedure instead of taking them as predetermined parameters.

From a practical design perspective, the study showed that the GS-GMDH algorithm could simplify the simulation of the complex and non-linear problem of ice scour with acceptable accuracy and extremely lower cost. The proposed approach can be an ideal alternative for the initial stages of design projects, where a fast and fairly accurate estimation is required to decide on methodologies and logistics as well as plan the scope of detailed engineering.

References

Al-Alawi, S.M., Al-Hinai, H.A., 1998. An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation. Renewable Energy, 14(1-4): 199-204.

Arnau, S., Ivanović, A., 2019. Rate effects during ice scour in sand. Cold Region Science Technology, 158: 182-194.

Azimi, H., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., Khoshbin, F., 2018. Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length. Acta. Mechanica, 229(3): 1197-1214.

Azimi, H., Shiri, H., 2020a. Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process. Journal of Offshore Mechanics and Arctic Engineering. 142(5), 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021b. Modeling Subgouge Sand Deformations by Using Multi-Layer Perceptron Neural Network. 31st International Ocean and Polar Engineering Conference. ISOPE-I-21-2150.

Been, K., Sancio, R. B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour Displacement in Clays From Physical Model Tests. 7th International Pipeline Conference, American Society of Mechanical Engineers Digital Collection, pp. 239–245.

Bekker, A.T., Sabodash, O.A., Truskov, P.A., 2005. Problems of Probabilistic Simulation of an Underwater Pipeline Track under Impact of Drifting Hummocks Offshore Sakhalin Island. Fifteenth International Offshore and Polar Engineering Conference. ISOPE-I-05-129.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2.

C-CORE., 1996. PRISE Phase 3c: Extreme LEE Gouge Event—Modeling and Interpretation, C-CORE Report 96-C32.

Duplenskiy, S., Gudmestad, O.T., 2013. Draft: Protection Of Subsea Pipelines Against Ice Ridge Gouging in Conditions of Substantial Surface Ice. Proceedings of the ASME 2013 32nd International Conference on Ocean, Offshore and Arctic Engineering, Paper No. 10430.

Ebtehaj, I., Bonakdari, H., Zaji, A.H., Azimi, H., Khoshbin, F., 2015. GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. Engineering Science and Technology, an International Journal, 18(4): 746-757.

Ebtehaj, I., Bonakdari, H., Khoshbin, F., Bong, C., Joo, H., Ab Ghani, A., 2017. Development of group method of data handling based on genetic algorithm to predict incipient motion in rigid rectangular storm water channel. Scientia Iranica, 24(3): 1000-1009.

Farlow, S.J., 1984. Self-Organizing Method in Modelling: GMDH Type Algorithm. Marcel Dekker, New York.

Guresen, E., Kayakutlu, G., Daim, T.U., 2011. Using artificial neural network models in stock market index prediction. Exp. Sys. Applica. 38(8): 10389-10397.

Haykin, S., 1994. Neural Networks: A Comprehensive Foundation. Prentice Hall PTR, Upper Saddle River, NJ.

ISO 19906, 2019 Petroleum and Natural Gas Industries – Arctic offshore structures. Geneva: ISO.

Ivakhnenko, A.G., 1971. Polynomial theory of complex systems. IEEE transactions on Systems, Man, and Cybernetics, 4: 364-378.

Ivakhnenko, A.G., 1976. The group method of data handling in prediction problems. Soviet Automatic Control, 9(6): 21-30.

Ivakhnenko, A.G., Ivakhnenko, G.A., 1995. The review of problems solvable by algorithms of the group method of data handling (GMDH). Pattern Recognition And Image Analysis C/C Of Raspoznavaniye Obrazov I Analiz Izobrazhenii. 5: 527-535.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. Thirteenth International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. Fourteenth International Offshore and Polar Engineering Conference. ISOPE-I-04-107.

Lach, P.R., 1996. Centrifuge Modelling of Large Soil Deformation due to Ice Scour. Doctoral dissertation, Memorial University of Newfoundland.

Mouawad Jad., 2008. Oil Survey Says Arctic Has Riches. The New York Times.

Najafzadeh, M., Saberi-Movahed, F., 2019. GMDH-GEP to predict free span expansion rates below pipelines under waves. Marine Georesources & Geotechnology, 37(3): 375-392. doi.org/10.1080/1064119X.2018.1443355

Nariman-Zadeh, N., Darvizeh, A., Felezi, M.E., Gharababei, H., 2002. Polynomial modelling of explosive compaction process of metallic powders using GMDH-type neural networks and singular value decomposition. Model. Modelling and Simulation in Materials Science and Engineering, 10: 727–744.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Pike, K.P., 2016. Physical and numerical modelling of pipe/soil interaction events for large deformation geohazards. Doctoral dissertation, Memorial University of Newfoundland, Canada.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip Line Field Solutions as an Approach to Understand Ice Subgouge Deformation Patterns. Sixteenth International Offshore and Polar Engineering Conference, ISOPE-I-06-289.

Shahin, M.A, Jaksa, M.B., Maier, H.R., 2001. Artificial neural network applications in geotechnical engineering. Australian Geomechanics, 36(1): 49-62.

Shin, M.B., Park, D.S., Seo, Y.K., 2019. Comparative Study of Ice Gouge Simulation Considering Ice Keel-Seabed Interactions. Journal of Ocean Engineering and Technology, 33(6): 556-563.

Smith, M., 1993. Neural networks for statistical modeling. Thomson Learning.

Vujicic, T., Matijevic, T., Ljucovic, J., Balota, A., Sevarac, Z., 2016. Comparative analysis of methods for determining number of hidden neurons in artificial neural network. In Central European conference on information and intelligent systems. Faculty of Organization and Informatics Varazdin, p. 219.

Walton, R., Binns, A., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., 2019. Estimating 2-year flood flows using the generalized structure of the Group Method of Data Handling. Journal of Hydrology, 575: 671-689.

Woodworth-Lynas, C.M.T., Phillips, R., Clark, J.I., Hynes, F., Xiao, X., 1998. Verification of centrifuge model results against field data: results from the Pressure Ridge Ice Scour Experiment (PRISE). 13th International Symposium on Okhotsk Sea and Sea Ice, pp. 123-138.

5. Chapter 5

Prediction of Iceberg-Seabed Interaction Process Using Extreme Learning Machine-Based Algorithms

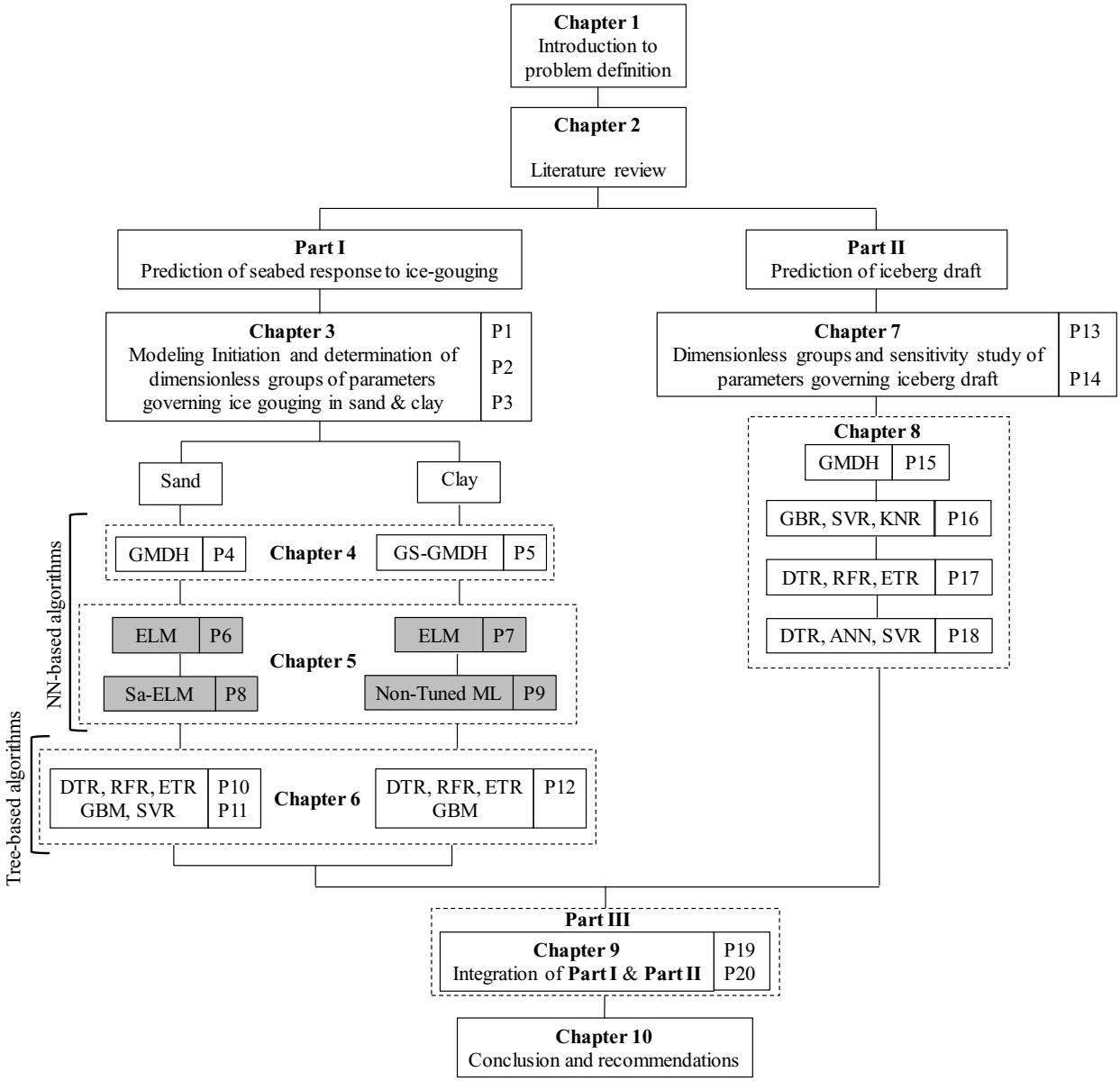
This chapter includes four sections as follows:

Section 1: Sensitivity Analysis of Parameters Influencing the Ice-Seabed Interaction in Sand by Using Extreme Learning Machine

Section 2: Assessment of Ice-Seabed Interaction Process in Clay Using Extreme Learning Machine

Section 3: Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine

Section 4: A non-tuned machine learning method to simulate ice-seabed interaction process in clay


Chapter 1
Introduction to problem definition
Chapter 2
Literature review
Part I
Prediction of seabed response to ice-gouging
Part II
Prediction of iceberg draft
Chapter 3
Modeling Initiation and determination of dimensionless groups of parameters governing ice gouging in sand & clay
P1
P2
P3
Chapter 7
Dimensionless groups and sensitivity study of parameters governing iceberg draft
P13
P14
Chapter 8
GMDH P15
GBR, SVR, KNR P16
DTR, RFR, ETR P17
DTR, ANN, SVR P18
Sand
Clay
NN-based algorithms
GMDH P4
Chapter 4
GS-GMDH P5
ELM P6
ELM P7
Chapter 5
Sa-ELM P8
Non-Tuned ML P9
Tree-based algorithms
DTR, RFR, ETR GBM, SVR P10 P11
Chapter 6
DTR, RFR, ETR GBM P12
Part III
Chapter 9
Integration of Part I & Part II
P19
P20
Chapter 10
Conclusion and recommendations

Section 1

Sensitivity Analysis of Parameters Influencing the Ice-Seabed Interaction in Sand by Using Extreme Learning Machine

This section is a journal paper which has been published in Natural Hazards (2021), Vol. 106, pp. 2307–2335, ISSN: 0921-030X

Abstract

Ice-gouging problem is a significant challenge threatening the integrity of subsea pipelines in the Arctic (e.g., Beaufort Sea) and even non-Arctic (e.g., Caspian Sea) offshore regions. Determining the seabed response to ice scour through the subgouge soil deformations and the keel reaction forces are important aspects of a safe and cost-effective design. In this study, the subgouge soil deformations and the keel reaction forces were simulated by means of the extreme learning machine (ELM) for the first time. Nine ELM models (ELM 1 to ELM 9) were developed using the key parameters governing the ice-seabed interaction. The number of neurons in the hidden layer was optimized and the best activation function for the ELM network was identified. The premium ELM model, resulting in the lowest level of inaccuracy and complexity, and the highest level of correlation with experimental values was identified by performing a sensitivity analysis. The gouge depth ratio and the shear strength of the seabed soil were found to be the most influential input parameters affecting the subgouge soil deformations and the keel reaction forces. A set of ELM-based equations were proposed to approximate these parameters. The uncertainty analysis showed that the premium ELM model slightly underestimated the subgouge soil deformation.

Keywords: Ice-seabed interaction, Sandy seabed, Extreme learning machine, Sensitivity analysis, Uncertainty analysis

5.1.1. Introduction

The Arctic areas have been discovered as one of the major hydrocarbon deposits comprising crude oil and natural gases, where marine pipelines are used to transfer these hydrocarbons from offshore to onshore (Azimi and Shiri 2020a). In the Arctic shallow waters, the ice-gouging event is universally considered a significant challenge for the bottom-founded infrastructures such as subsea pipelines, wellheads, and communication cables. The schematic layout of an ice-seabed interaction during an ice-gouging event is depicted in Figure 5-1, where a hyperbolic subgouge soil displacement occurs with the maximum magnitude right below the ice keel. Subsea trenching and backfilling are usually utilized as cost-effective solutions for the physical protection of pipelines.

In terms of pipeline response to ice-gouging, the ultimate design factors include subgouge soil deformation and the keel reaction forces that directly affect the pipeline response. These factors, in turn, are resultant from several key input parameters including the attack angle, gouge depth, seabed soil properties, ice dynamics, etc. In the current study, similar to almost all of the earlier studies (e.g., Paulin 1991 & 1992; Lach 1996; C-CORE 1995; C-CORE 1996; Hynes 1996; Nobahar et al. 2007; Abdalla et al. 2009; Liferov et al. 2014; Pike and Kenny 2016; Nematzadeh and Shiri 2019), the subgouge soil deformation and the keel reaction forces were considered as the key outputs of the ice-gouging analysis.

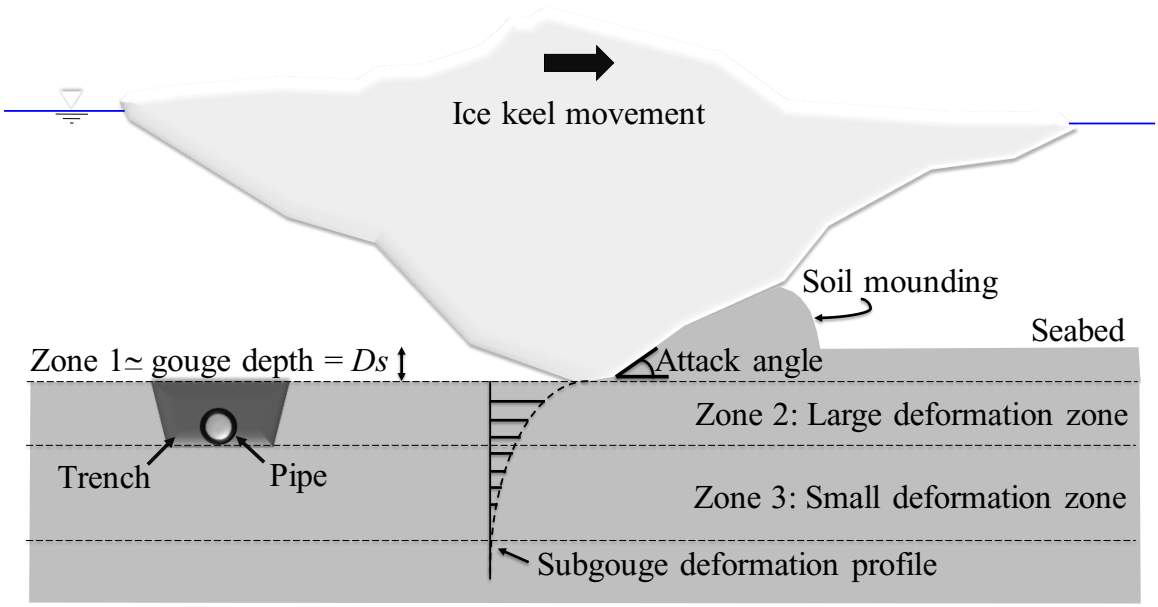


Figure 5-1. Schematic layout of ice-seabed interaction process during an ice-gouging event

Over the last decades, several experimental and numerical investigations have been done to improve the understanding of the ice-seabed interaction process. For instance, Palmer et al. (1990) performed an analytical study to approximate the horizontal and vertical subgouge forces. The authors suggested some linear models and asserted that the reaction forces were a function of iceberg weight, gouge depth, and gouge width. Paulin et al. (1991) conducted an experimental study to investigate the ice-soil interaction by 1g floor tests and 100g centrifuge tests. The authors examined clays with different saturations and concluded that the soil deformations increased by decreasing the relative density of the soil. Moreover, Paulin et al. (1992) performed a set of centrifuge tests to evaluate the subgouge deformation characteristics in the sand such as surface deformations, subscour displacements, pressure, and keel reaction forces. The study showed that measured loads and forces for the submerged models were less than the dry sand, whereas other parameters were partially similar to the dry and submerged tests. In addition, several pressure ridge ice scour experiment (PRISE) tests were performed in the sand by Hynes (1996) to measure the

horizontal and vertical keel reaction forces, gouge-induced pore pressures, stresses, and keel-soil interaction pressures. The study showed the shear dragging theory and resulted in a linear relationship between the scour loads and the scour depth. Lanan et al. (2001) reported the design and construction procedure of the Northstar offshore Arctic pipeline in Alaska. The trenching was applied for physical protection of the pipeline against the ice-gouging problem and the minimum burial depth was designed to be seven feet. The authors asserted that the ice-induced soil displacement was a function of soil depth, soil type, and gouge geometry. Nobahar et al. (2007) simulated the ice-seabed-pipe interaction mechanism by using the explicit arbitrary Lagrangian finite element method and the Winkler approach. The author displayed that the finite element model had a better performance in order to predict the ice-gouging issue. The study highlighted that the pipe movement increased by growing the magnitude of the keel gouge depth and the reaction forces were much greater for strong soil. Kenny et al. (2007) presented an overview of the numerical studies for the simulation of the ice-seabed-pipeline interaction phenomena. The authors highlighted that inputs of numerical and empirical models ought to be determined by using more statistical analysis to estimate the reaction forces and subgouge displacements. Konuk and Yu (2007) simulated the coupled ice-seabed-pipe interaction using the Arbitrary Lagrangian-Eulerian (ALE) scheme and the Winkler approach. The authors proved that the ALE FE method had better performance to model the ice-gouging problem.

Been et al. (2008) carried out an experimental investigation to measure the horizontal and vertical subgouge soil displacements in clay. The obtained results indicated that the magnitude of deformations grew by increasing the undrained shear strength since the stiffness of clay enhanced more slowly. Yang (2009) tested the ice-seabed interaction process in the sand by conducting the pipeline ice risk assessment and mitigation (PIRAM) program. The author evaluated the subgouge

sand displacements, the pressure, the keel reaction forces, the frontal berm, and the depth of the gouge. The study showed the maximum horizontal subscour deformations formed just beneath the ice keel base and concluded that the ratio of frontal berm height to gouge depth had a striking influence on the subgouge displacements and the reaction forces. Abdalla et al. (2009) modeled the horizontal ice-induced soil displacements by using the Coupled Eulerian Lagrangian (CEL) model. The study showed that the numerical models with a shallower attack angle had a better performance. The authors performed a parametric analysis and then demonstrated that the horizontal subgouge deformations were proportional to the depth of the gouge. Barrette (2011) classified ice-gouging studies into four major subgroups comprising field, laboratory, theoretical, and numerical simulation studies. The author asserted that enough knowledge of material characteristics was required to provide the guarded and economical burial depth for the subsea assets. Sancio et al. (2011) conducted a field testing program to study the ice-soil-pipeline interaction process in both clay and sand. The research concluded that the depth of the gouge had an inverse correlation with dry unit weight in the sand and undrained shear strength in clay. Phillips and Barrett (2012) reviewed the pipeline ice risk assessment and mitigation JIP (PIRAM) investigations comprising physical tests and numerical simulations. The authors demonstrated that the horizontal reaction force was a function of the submerged unit soil weight, attack angle, and gouge geometry. They also showed that the pipeline strain decreased by increasing the orientation of the ice gouge. Pike and Kenny (2012) modeled the ice-seabed-pipe interaction problem numerically through a finite element model. The study revealed that the shear stress limit had a remarkable influence on the axial pipe reaction forces. Peek et al. (2013) developed a FE approach to model the keel-soil-interaction process. The authors simulated the horizontal subgouge soil

displacement as a function of soil depth and showed that the numerical model had a reasonable performance when the size of computational elements was less than 0.1m.

Liferov et al. (2014) developed a numerical model in order to simulate the keel-soil-pipeline interaction by means of a FE model. The ice mass was assumed to be a rigid body, rather and the pipe was simulated with Lagrangian elements. The authors concluded that the gouge depth was the most important factor to design the subsea pipelines during an ice-gouging event.

Babaei and Sudom (2014) analyzed the results of eighteen numerical simulations in the field of ice-seabed interaction mechanisms. This review study showed that the pre-set gouge depth and the maximum depth of ice-scoured displacements had a linear relationship. Arnau and Ivanović (2019) conducted experiments to study the influence of the ice scour loads on a sandy seabed in saturated and dry conditions. The authors concluded that the rate effects enhanced linearly by increasing the scour width and the scour depth. Nematzadeh and Shiri (2019a) developed a CEL model for free-field ice-gouging analysis in sand using ABAQUS/Explicit. The authors incorporated the non-linear strain rate and softening effects through a user-defined subroutine. The study resulted in an improved prediction of the subgouge soil deformation and the keel reaction forces obtained from published experimental studies. Also, Nematzadeh and Shiri (2019b) simulated the ice-seabed interaction process using a self-correcting soil model in order to update the shear strength parameters during the pre-peak hardening and the post-peak softening of the sand. The authors showed that the subgouge soil deformation might be overestimated by the conventional decoupled approaches. Additionally, Nematzadeh and Shiri (2020) modeled the effect of the non-linear stress-strain behavior of dense sand in an ice-gouging problem by using a modified Mohr-Coulomb (MMC) model. The authors concluded that the size of side berms and the frontal mound were

affected by the magnitude of the attack angle, where greater subgouge deformations and reaction forces were observed for models with shallower attack angles.

Overall, the experimental and three-dimensional (3D) simulations of ice-gouging are expensive and time-consuming. Therefore, the industry is seeking potential less-costly solutions for the simulation of the ice impact on buried pipelines.

In recent years, artificial intelligence (AI) approaches and soft computing (SC) models have been successfully utilized to model the ice-seabed interaction process since these methods are accurate, quick, and inexpensive. For example, Kioka et al. (2003) modeled an ice-gouging problem using the Neural-Network (NN) approach. The NN showed a high level of accuracy and the authors asserted that this approach could be replaced with the nonlinear multiple-regression methods. Azimi and Shiri (2020a) introduced the parameters governing the ice-seabed interaction process in both clay and sandy conditions. They proposed a set of linear regression (LR) models to estimate the maximum subgouge soil displacements. The authors also concluded that the shear strength parameters of soil and the ratio of gouge depth to gouge width were the most influencing variables. Azimi and Shiri (2020b) simulated the horizontal subgouge deformations in the sand by the gene expression programming (GEP) model. The authors defined six GEP models and then the best GEP model was introduced by conducting a sensitivity analysis. The study showed that soil depth was considered the most significant input parameter.

Machine learning (ML) is a powerful, accurate, and quick tool for assessing complex problems that have never been tried for ice-gouging analysis. In this study, the extreme learning machine (ELM), as a feedforward neural network (FFNN) was used for a precise, robust, and cost-effective evaluation of the seabed response to ice, and scour. More details are presented in the coming sections.

5.1.2. Methodology

5.1.2.1. Extreme learning machine (ELM)

Extreme learning machine (ELM) as a robust neural network algorithm was introduced by Huang et al. (2004) to solve various linear and non-linear problems. The ELM is a fast training approach that has the structure of a single-layer feed-forward neural network (SLFFNN) (Huang et al. 2006). Figure 5-2 exhibits the basic structure of the ELM network. In comparison with the classical backpropagation (BP) algorithm, the ELM has several benefits as follows:

- The high training speed of feedforward neural network (FFNN) by ELM,
- Better generalization performance,
- Training with differentiable and non-differentiable activation functions,
- Overcome the gradient-based algorithm such as local minimum, improper learning rate, and overfitting problems (Huang et al. 2011; Azimi et al. 2017; Sattar et al. 2019).

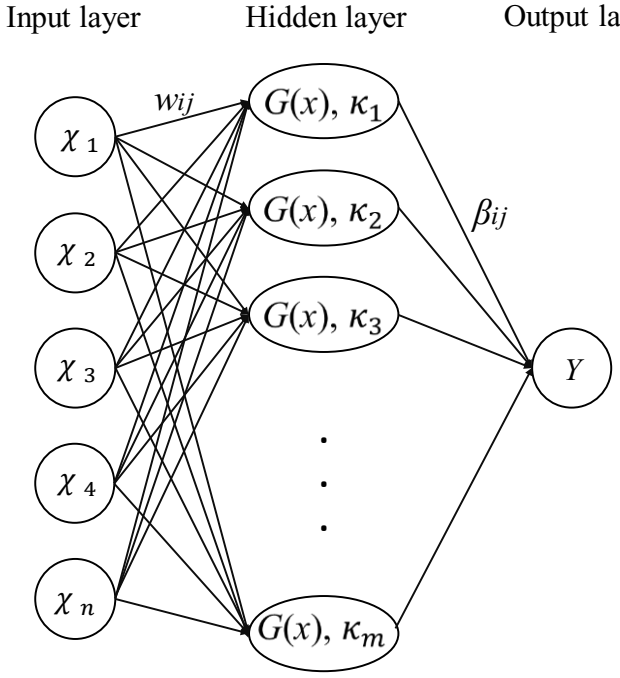


Figure 5-2. Structure of ELM network

As shown in Figure 5-2, the ELM possesses three different layers including an input layer, a hidden layer, and an output layer. The input layer contains the input parameters of the ice-gouging issue and the number of neurons in the output layer equals the number of target parameters of the problem. However, the number of neurons in the hidden layer is selected based on a trial and error procedure.

A single hidden layer forward network is formed by means of linear algebra in these three layers to solve the equations and obtain the optimum weights in the output layer (Sattar et al. 2019). The weights of the input layer are determined randomly, whereas the output weights are analytically estimated through a predefined training mechanism.

$$Y = \sum_{j=1}^{m} \beta_j G\left(\sum_{i=1}^{n} w_{i,j}\chi_i + \kappa_j\right) \tag{5-1}$$

where, Y and χ donate the output and inputs of the ELM model, respectively. Additionally, n is the number of input parameters, and m is the number of output parameters of the problem. Also, $w_{i,j}$ shows the input weight as the connection between the i^{th} neuron in the input layer and the j^{th} neuron in the hidden layer; β_j is a coefficient to link the j^{th} neuron in the hidden layer to the particular neuron in the output layer. Moreover, $G(\,)$ and κ_j are the activation function and biases of the hidden layer neurons, respectively.

Generally, the output of an SLFFNN is computed in two major stages. First, the network of the hidden layer and the number of neurons in this layer is determined by using a trial and error procedure (Azimi et al. 2017). Subsequently, to train the ELM network, an activation function (AF) is defined. Indeed, the optimum number of hidden neurons and the best activation function lead to the optimal ELM network. Using arbitrary assignments of the input layer weights $(w_{i,j})$

and the biases (κ_j), the structure of hidden layer neurons and the weight in the output layer are computed. Therefore, equation (5-1) is rewritten as below:

$$H\beta = Y \tag{5-2}$$

here, *H* is the extreme learning machine feature mapping matrix as follows:

$$H(w_{i,j}, \kappa_j, \chi_i) = \begin{bmatrix} H_{1,1} & \cdots & H_{1,m} \\ \vdots & \ddots & \vdots \\ H_{n,1} & \cdots & H_{n,m} \end{bmatrix} = \begin{bmatrix} G(w_{1,1}\chi_1 + \kappa_1) & \cdots & G(w_{1,m}\chi_m + \kappa_m) \\ \vdots & \ddots & \vdots \\ G(w_{n,1}\chi_n + \kappa_1) & \cdots & G(w_{n,m}\chi_m + \kappa_m) \end{bmatrix} \tag{5-3}$$

Also, β and *Y* are defined as below:

$$Y = [Y_1\ Y_2\ \cdots\ Y_m] \quad \text{and} \quad \beta = [\beta_1\ \beta_2\ \cdots\ \beta_n] \tag{5-4}$$

It is worth noting that β_j is obtained by minimizing error in the estimation through the Moore-Penrose generalized inverse matrix as follows (Huang et al. 2011):

$$\hat{\beta} = H^{+}Y \tag{5-5}$$

where, H^{+} represents the Moore-Penrose generalized inverse matrix of H.

5.1.2.2. Seabed interaction process

The ice-seabed interaction parameters (η) include the subgouge soil deformations (*d/W*) and reaction forces ($F/\gamma_s W^3$) that in turn are a function of the scour depth (D_s), the internal friction angle of sand φ (representing the shear strength of cohesionless seabed), the width of gouge (*W*), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (*h*'), the horizontal (L_h) and the vertical load (L_v), the velocity of ice keel (*V*), and the specific weight of sand (γ_s). These relationships are defined as follows (Lach 1996; Azimi and Shiri 2020a):

$$\eta_{(sand)} = f_1(D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \tag{5-6}$$

Besides the internal friction angle, other parameters may affect the soil response to ice-gouging such as the relative density, dilation angle, and variation of shear strength parameters with depth.

However, it is challenging to find datasets reporting these parameters altogether. Also, the strength parameters considered in this study (γs and φ) have a direct relationship with the relative density and dilation angle, which can be considered as an inherent inclusion of their potential effects in terms of sensitivity (Mitchell and Soga 2005; Cinicioglu and Abadkon 2015). Equation (5-6) can be written as a function of eight dimensionless groups as below (Azimi and Shiri 2020a):

$$\frac{d_{(sand)}}{W}, \frac{F}{\gamma_s W^3} = f_2\left(\frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{5-7}$$

The maximum subgouge deformation in the sand $\left(d_{(sand)}\right)$ occurs immediately under the ice keel in the gouge centerline. At greater depths, by incorporating the soil depth (y/W), the subgouge soil deformation ($d_{(\text{sand})}/W$) is written as follows:

$$\frac{d_{(sand)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{5-8}$$

Similarly, the keel reaction force (F) is a function of the position of the iceberg along the scour axis (x) and a set of parameters including $D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s$. Therefore, the equation (5-8) can be written as below:

$$\frac{F_{(sand)}}{\gamma_s W^3} = f_4\left(\frac{x}{W}, \frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{5-9}$$

In the current study, a series of investigations were conducted to obtain the ice-gouging output parameters (i.e., d/W and F/W). Therefore, the results from six experimental studies reported by Paulin (1991), Paulin (1992), C-CORE (1995), Hynes (1996), C-CORE (1996), and Yang (2009) were utilized to verify the ELM models. The angle of the surcharged soil slope (ω) was not reported by all applied experimental studies. Therefore, the introduced dimensionless groups in equations (5-8) and (5-9) were applied as the input parameters for the ELM model to simulate the

horizontal subgouge deformation (d_h/W), the vertical subgouge deformation (d_v/W), the horizontal reaction force (F_h/W), and the vertical reaction force (F_v/W).

There are several experimental investigations in the literature. Usually, these studies provide different input parameters depending on the objectives of the investigation. This, in turn, limits the application of datasets to evolutionary numerical models. In this study, the dataset was selected amongst the studies that provided a common set of input parameters and were widely used in the literature for validation purposes. To ensure that the selected dataset is statistically significant, a t-test was conducted and the P-values were calculated and assessed against a maximum target value of 0.05. This was to evaluate the probability of the relationship between the parameters being affected by an alternative hypothesis (Bland and Peacock 2002). The computed P-value for the applied datasets was 0.039, indicating that the correlations are statistically significant. Also, the test data from the centrifuge and 1g scale tests were non-dimensionalized to assure the consistency of the dataset and confidence of the obtained simulation results. Figure 5-3 shows the used input combinations in the current study.

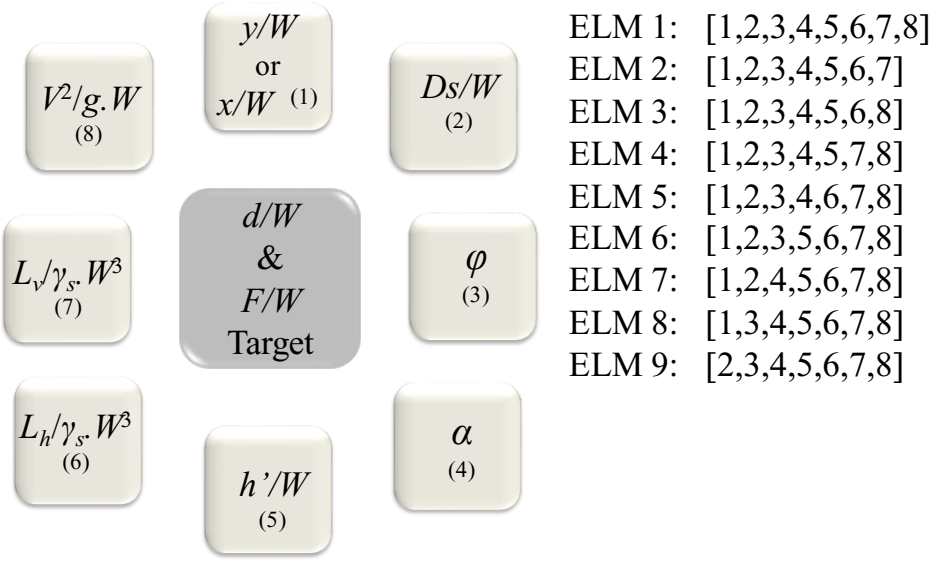


Figure 5-3. Applied input combinations in the current study

Nine ELM models (ELM 1 to ELM 9) were defined to predict the ice-induced sand parameters. It means that ELM 1 was defined by using all input parameters and then the effect of each input was eliminated (ELM 2 to ELM 9) to identify the most significant input parameters and introduce the superior ELM model. 70% of the observed values were applied to train the ELM models, and 30% of the remaining data were employed to test these models.

5.1.2.3. Goodness of fit

To examine the performance of the ELM models, several statistical indices including correlation coefficient (*R*), the variance accounted for (*VAF*), root mean square error (*RMSE*), mean absolute error (*MAE*), Nash-Sutcliff efficiency coefficient (*NSC*), and Akaike Information Criteria (*AIC*) was calculated as follows:

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{5-10}$$

$$VAF = \left(1 - \frac{var(P_i - O_i)}{var(P_i)}\right) \times 100 \tag{5-11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{5-12}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{5-13}$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{5-14}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{5-15}$$

where, Oi, Pi, $\bar{O}$, $\bar{P}$, and n are the observed ice-induced parameters, the simulated values, the average observed values, the average simulated values, and the number of experimental measurements, respectively. Moreover, the k is the number of hidden layer neurons used in the ELM network. The closeness of the correlation coefficient (R) and the Nash-Sutcliff efficiency coefficient (NSC) to unit value (i.e., one) showed that the ELM model has a high correlation with the observed values., The $RMSE$ and MAE values close to zero meant that the particular model has the lowest level of error. Furthermore, the superior ELM model had the highest value of the VAF criterion. Although, the complexity of these ELM models was not shown through the applied indices. To overcome this problem, the Akaike Information Criteria (AIC) was employed. The less complex ELM model had the lowest value of the AIC index. Hence, the superior ELM model had the lowest value of the AIC index and error ($RMSE$ and MAE) and the highest level of correlation (R and NSC) with the experimental values.

5.1.3. Results and discussion

Initially, the number of neurons in hidden layers (NHN) was optimized and then the best activation function (AF) of the extreme learning machine (ELM) was chosen. Subsequently, the subgouge sand deformations (d_h/W and d_v/W) and the subgouge reaction forces (F_h/W and F_v/W) were simulated by using ELM 1 to ELM 9. After that, a comprehensive sensitivity analysis (SA) for the ELM models was conducted. Next, an error analysis along with an uncertainty analysis (UA) was performed for the superior ELM model. Lastly, a set of ELM-based equations were provided to estimate the horizontal subgouge deformations (d_h/W), the vertical subgouge deformations (d_v/W), the horizontal subgouge reaction force (d_h/W), and the vertical subgouge reaction force (d_v/W).

5.1.3.1. Number of hidden neurons (NHN)

The optimized number of hidden neurons (NHN) for the extreme learning machine (ELM) to simulate the ice-gouging parameters was evaluated. For instance, variations of NHN versus the applied statistical indices for modeling the horizontal subgouge deformations (dh/W) are shown in Figure 5-4. In the current study, the number of neurons in the hidden layer was initially set to two and then the performance of the ELM model was assessed by increasing this value to 32. As shown, with increasing the NHN, the precision of the ELM model enhanced notably. In other words, as the NHN was chosen to be two, the values of correlation coefficient (R), mean absolute error (MAE), and Akaike Information Criteria (AIC) were respectively estimated as 0.551, 0.055, and -36.458. However, the VAF, $RMSE$, and NSC were equal to 95.771, 0.051, and 0.958 for an NHN value of 28. Therefore, the optimal NHN was chosen to be 28 since the performance of the ELM model was insignificantly altered.

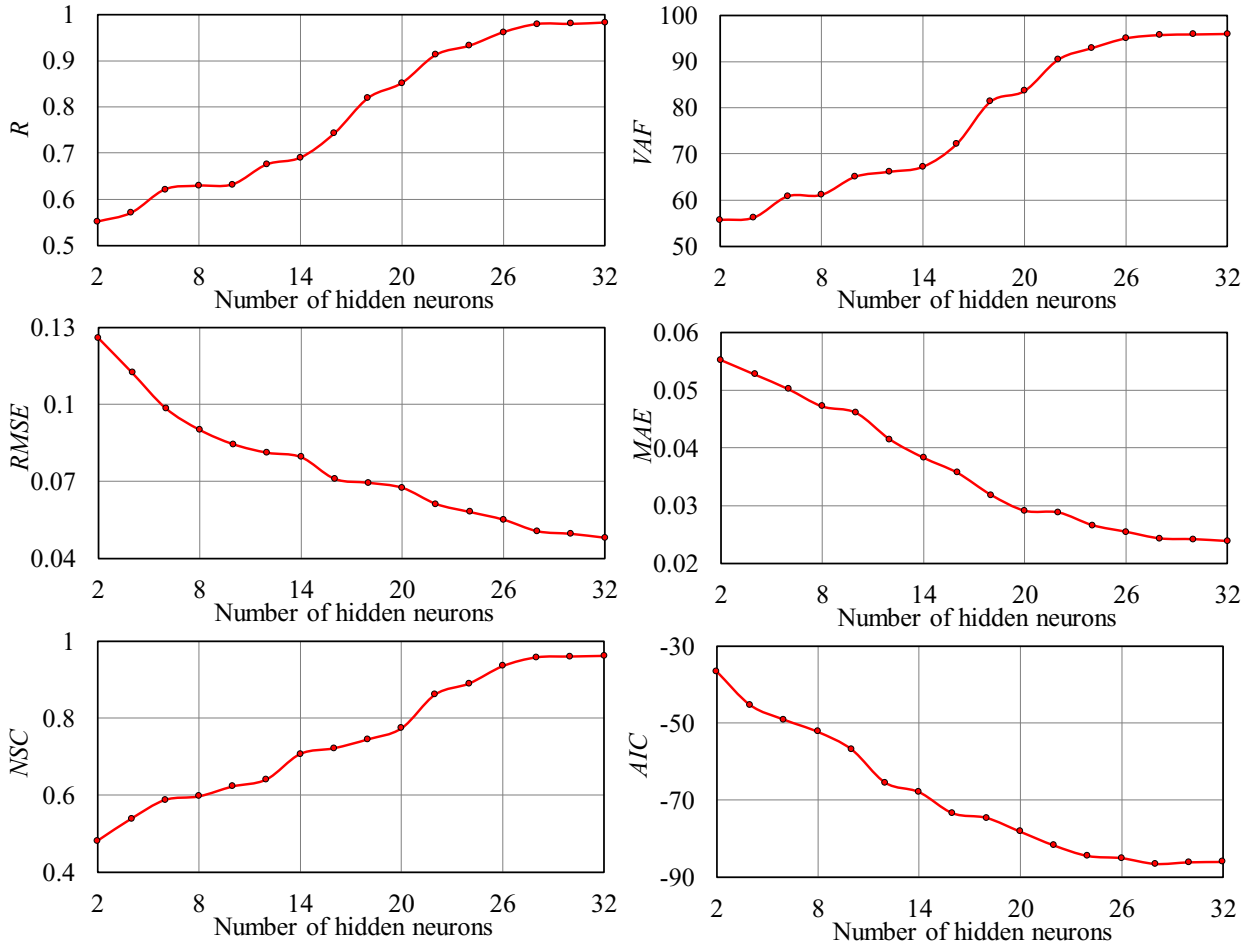


Figure 5-4. Variations of the number of hidden neurons (NHN) versus the applied statistical indices for simulation of horizontal subgouge deformations (dh/W)

5.1.3.2. Activation Function (AF)

The effect of different activation functions including sigmoidal (Sig), sinusoidal (Sin), hard limit (Hardlim), radial basis (Radbas), and triangular basis (Tribas) functions was evaluated to model the ice-gouging parameters. The performance of the ELM network for each function was assessed and then the superior activation function was introduced. Figure 5-5 illustrates the estimated statistical indices for these activation functions to model the horizontal subgouge deformations.

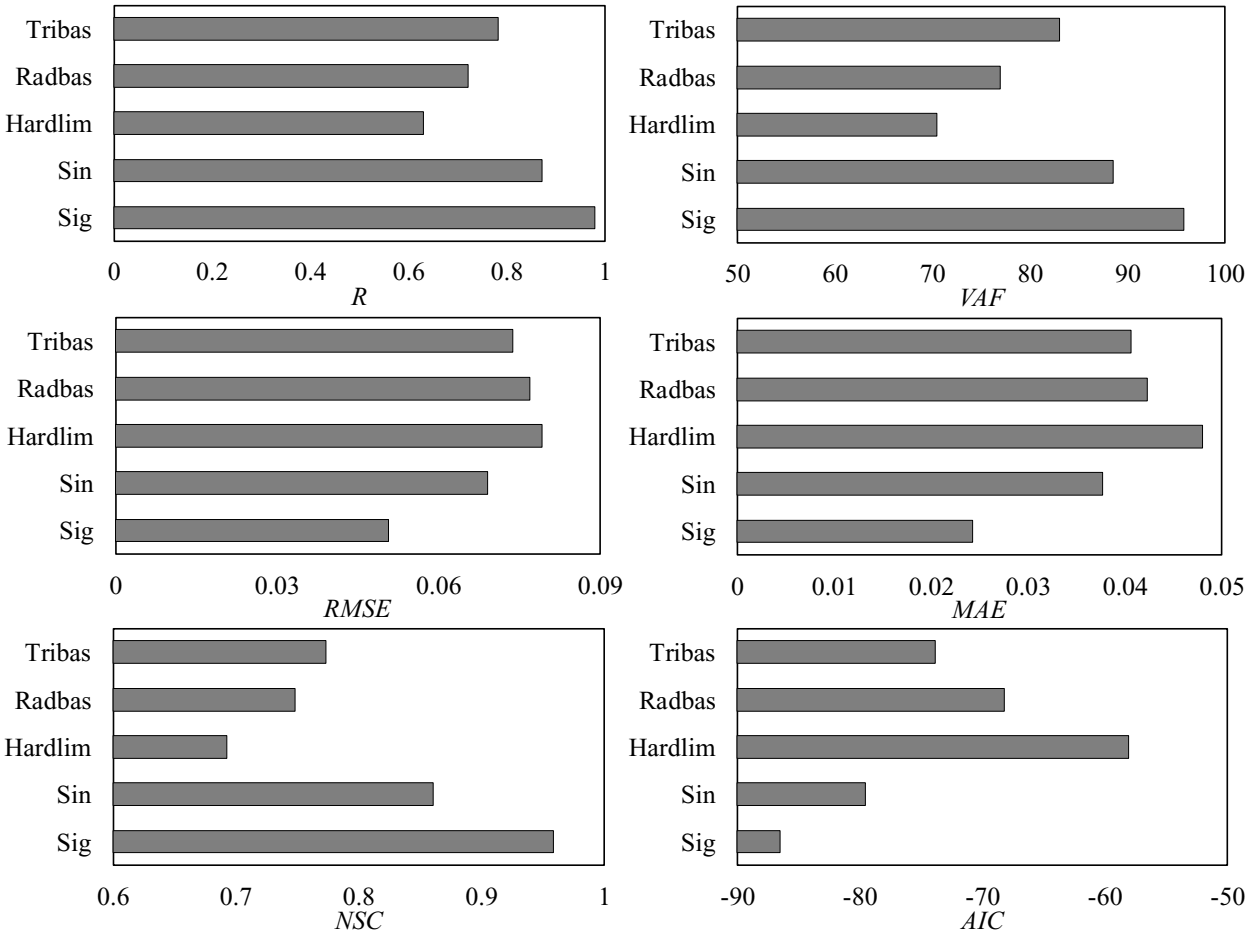


Figure 5-5. Estimated statistical indices for different activation functions to model the horizontal subgouge deformations

For example, for the Sig, Sin, and Hardlim functions, the value of the correlation coefficient was surmised as 0.980, 0.871, and 0.630, respectively. The Nash-Sutcliff efficiency coefficient for the Radbas and the Tribas functions equaled 0.748 and 0.773. The Sig activation function resulted in the highest level of accuracy and correlation with experimental values. Therefore, the Sig function was utilized as the best activation function to simulate the ice-induced process in the current study.

5.1.3.3. Sensitivity analysis (SA)

Figure 5-6 shows the calculated statistical indices of ELM 1 to ELM 9 to simulate the horizontal subgouge deformations (*dh/W*) in training and testing modes. ELM 1 was a function of all input variables including y/W, Ds/W, φ, α, h'/W, $Lh/\gamma s.W^3$, $Lv/\gamma s.W^3$, and $V^2/g.W$. The values of correlation coefficient (*R*), Nash-Sutcliff efficiency coefficient (*NSC*), and Akaike Information Criteria (*AIC*) for ELM 1 in training mode were respectively estimated as 0.980, 0.958, and -86.511. Moreover, ELM 2 predicted the horizontal deformations by means of y/W, Ds/W, φ, α, h'/W, $Lh/\gamma s.W^3$, and $Lv/\gamma s.W^3$, interpreting that the influence of the velocity parameter ($V^2/g.W$) was ignored for ELM 2. In the testing mode of ELM 2, the *VAF*, *MAE*, and *RMSE* were equal to 40.013, 0.104, and 0.163, respectively. In comparison with ELM 1, the correlation of ELM 2 in testing mode with experimental values decreased by roughly 28%, whereas its accuracy in training mode reduced by almost 127%.

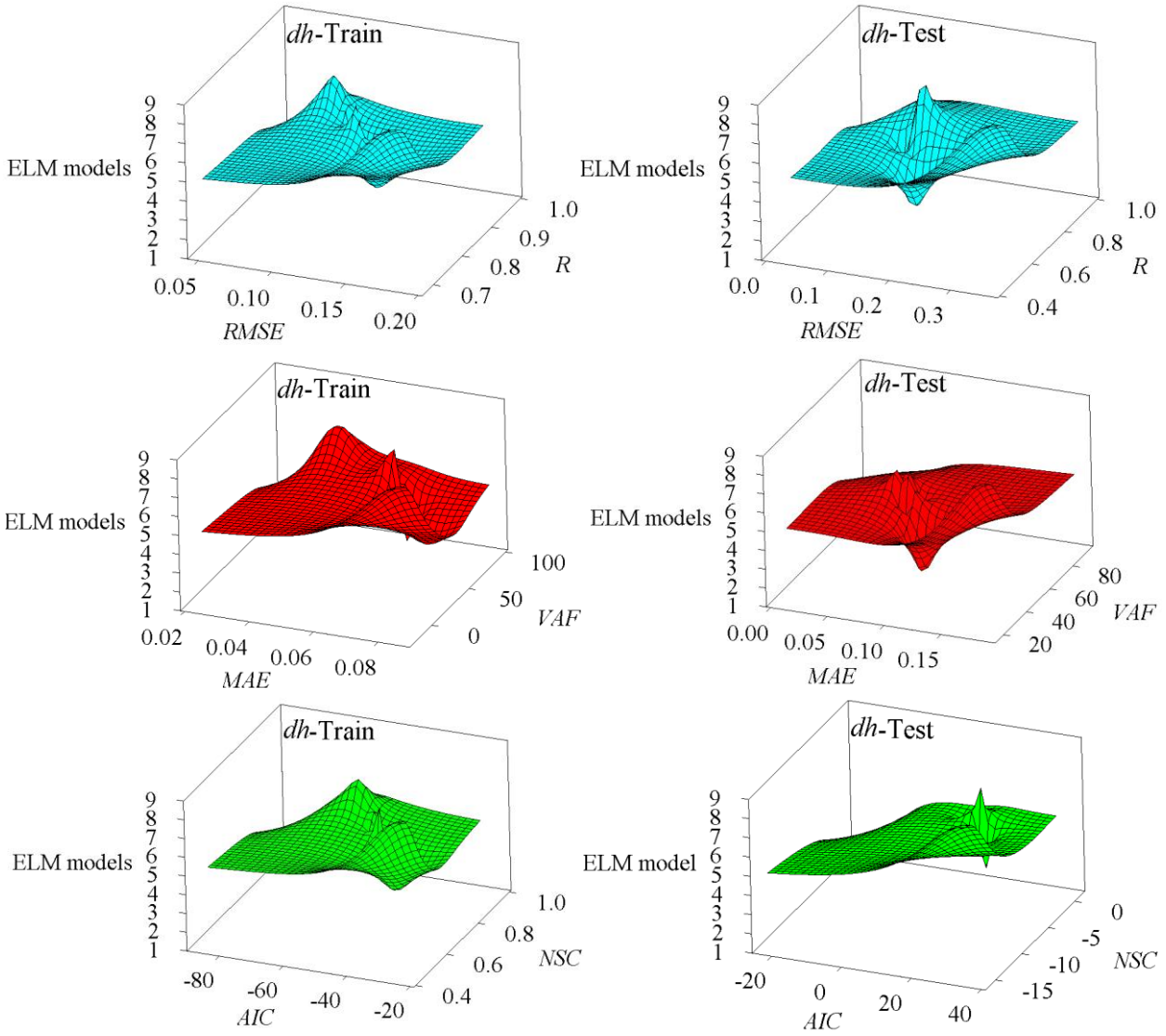


Figure 5-6. The key statistical indices for horizontal deformations simulated by ELM models in training and testing modes

The y/*W, Ds/W, φ, α, h'/W, Lh/γs.W*^{3}, and $V^2/g.W$ were used as the input variables of ELM 3 to model the horizontal deformations, meaning that the impact of dimensionless vertical load ($Lv/\gamma s.W^3$) was wiped out for this model. The complexity of ELM 3 in comparison with ELM 1 increased by about 54%. In other words, the *AIC*, *NSC*, and *VAF* values for ELM 3 in the testing mode were respectively surmised at 13.597, -1.083, and 56.308.

The effect of $L_h/\gamma_s.W^3$ was relinquished for ELM 4 and this model estimated the horizontal deformations using other input parameters. The accuracy and correlation of ELM 4 in comparison with ELM 1 decreased by approximately 105% and 23%, with *RMSE*=0.106 and *R*=0.752.

The *MAE*, *R*, and *AIC* values in ELM 5 for the simulation of the horizontal deformations in training mode were 0.060, 0.905, and -49.442, respectively. For ELM 5, the influence of *h'/W* was eliminated, meaning that ELM 5 predicted the target function by using *y/W, Ds/W,* φ*,* α*,* $Lh/\gamma s.W^3$*,* $Lv/\gamma s.W^3$, and $V^2/g.W$. The correlation of ELM 5 in testing mode compared to ELM 1 was reduced by nearly 16%. ELM 6 predicted the horizontal deformations by a combination of *y/W, Ds/W,* φ*, h'/W,* $Lh/\gamma s.W^3$*,* $Lv/\gamma s.W^3$, and $V^2/g.W$ parameters and the ice attack angle variable (α) was the removed factor in ELM 6. The *AIC* and *RMSE* values for ELM 6 in the testing phase were respectively reckoned to be 15.058 and 0.135, with a correlation coefficient of 0.870. In ELM 7, the key variables included *y/W, Ds/W,* α*, h'/W,* $Lh/\gamma s.W^3$*,* $Lv/\gamma s.W^3$, and $V^2/g.W$, signifying that the shear strength parameter of the seabed sand (φ) was ignored to model the horizontal deformations. The values of *R*, *MAE*, and *VAF* for ELM 7 in the testing situation were calculated to be 0.800, 0.074, and 52.583, respectively. Regarding the comparison between ELM 1 and ELM 7, the complexity of ELM 7 was roughly a rise of 35%, with *AIC*=-59.077 in the training mode. The influence of Ds/W was removed from the input of ELM 8 and this model simulated the target function by using *y/W,* φ*,* α*, h'/W,* $Lh/\gamma s.W^3$*,* $Lv/\gamma s.W^3$, and $V^2/g.W$. It is worth noting that the accuracy of ELM 8 compared to ELM 1 decreased by about 158%, with *MAE*=0.062 in the training stage. The ratio of soil depth to gouge width (*y*/*W*) was the eliminated factor in ELM 9, where this model comprised the input variables *Ds/W,* φ*,* α*, h'/W,* $Lh/\gamma s.W^3$*,* $Lv/\gamma s.W^3$, and $V^2/g.W$. In the testing mode of ELM 9, the *NSC, VAF*, and *RMSE* values were -15.911, 17.377, and 0.357, respectively. The accuracy and correlation of ELM 9 in comparison with ELM 1 decreased close

to 250% and 31%, whereas the complexity of this model increased to nearly 71%. Therefore, ELM 1 was identified as the best model to simulate the horizontal deformation (dh/W) since this model had a good performance in terms of accuracy, correlation, and complexity. After ELM 1 model, ELM 6, ELM 5, and ELM 8 were ranked as the second, third, and fourth-best ELM models. Moreover, ELM 7 as the fifth-best, ELM 3 as the sixth-best, ELM 4 as the seventh-best, and ELM 2 as the eighth-best ELM modes were graded. ELM 9 was detected as the worst ELM model to predict horizontal deformations.

According to the performed sensitivity analysis, the ratio of soil depth to gouge depth (y/W) was identified as the most significant input parameter to simulate the horizontal deformation. After that, the dimensionless velocity parameter ($V^2/g.W$), the dimensionless horizontal load ($L_h/\gamma_s.W^3$), the ratio of gouge depth to gouge width (Ds/W), the shear strength parameter of the seabed sand (φ), the dimensionless vertical load ($Lv/\gamma s.W^3$), the attack angle (α), and the ratio of berm height to gouge width (h'/W) were respectively prioritized in terms of effectiveness. Figure 5-7 exhibits the estimated statistical indices for the vertical deformations simulated by the ELM models in training and testing modes. In the testing mode of ELM 1, the values of R, NSC, and AIC were respectively computed to be 0.915, 0.642, and 2.339. In comparison with ELM 1, the accuracy of ELM 2, ELM 3, and ELM 4 reduced by approximately 67%, 33%, and 53%. In addition, the $RMSE$ and MAE values for ELM 5 in the testing phase were estimated at 0.007 and 0.005, where the correlation of this model decreased by nearly 30%. Additionally, the complexity of ELM 6 and ELM 7 enhanced by about 259% and 358% as compared to ELM 1, with the AIC of 8.191 and 10.725, respectively. The correlation of ELM 8 decreased by a value close to 43%, with a computed correlation coefficient of 0.519 in the testing situation. The VAF, $RMSE$, and NSC values for ELM 9 in the

testing mode were respectively equal to -68.322, 0.007, and -0.190, with a decreased accuracy of 75%.

The ELM 1 was introduced as the best ELM model to estimate the vertical deformations (dv/W), with the highest level of accuracy and correlation along with the lowest level of complexity. After ELM 1 model, ELM 2, ELM 5, ELM 4, and ELM 3 were respectively known as the second-best, third-best, fourth-best, and fifth-best ELM models. Furthermore, ELM 9, ELM 6, ELM 8, and ELM 7 were graded as the sixth-best, seventh-best, eighth-best, and ninth-best ELM models for simulating vertical deformations.

Based on the conducted sensitivity analysis, the shear strength parameter of the sandy seabed (φ) was identified as the most influencing input parameter to estimate the vertical deformations. Besides, other input variables including Ds/W, α, y/W, and $Lv/\gamma s.W^3$ were ranked as the second, third, fourth, and fifth significant input parameters, respectively. In addition, the $Lh/\gamma s.W^3$, h'/W, and $V^2/g.W$ were respectively ranked as the sixth, seventh, and eighth important inputs.

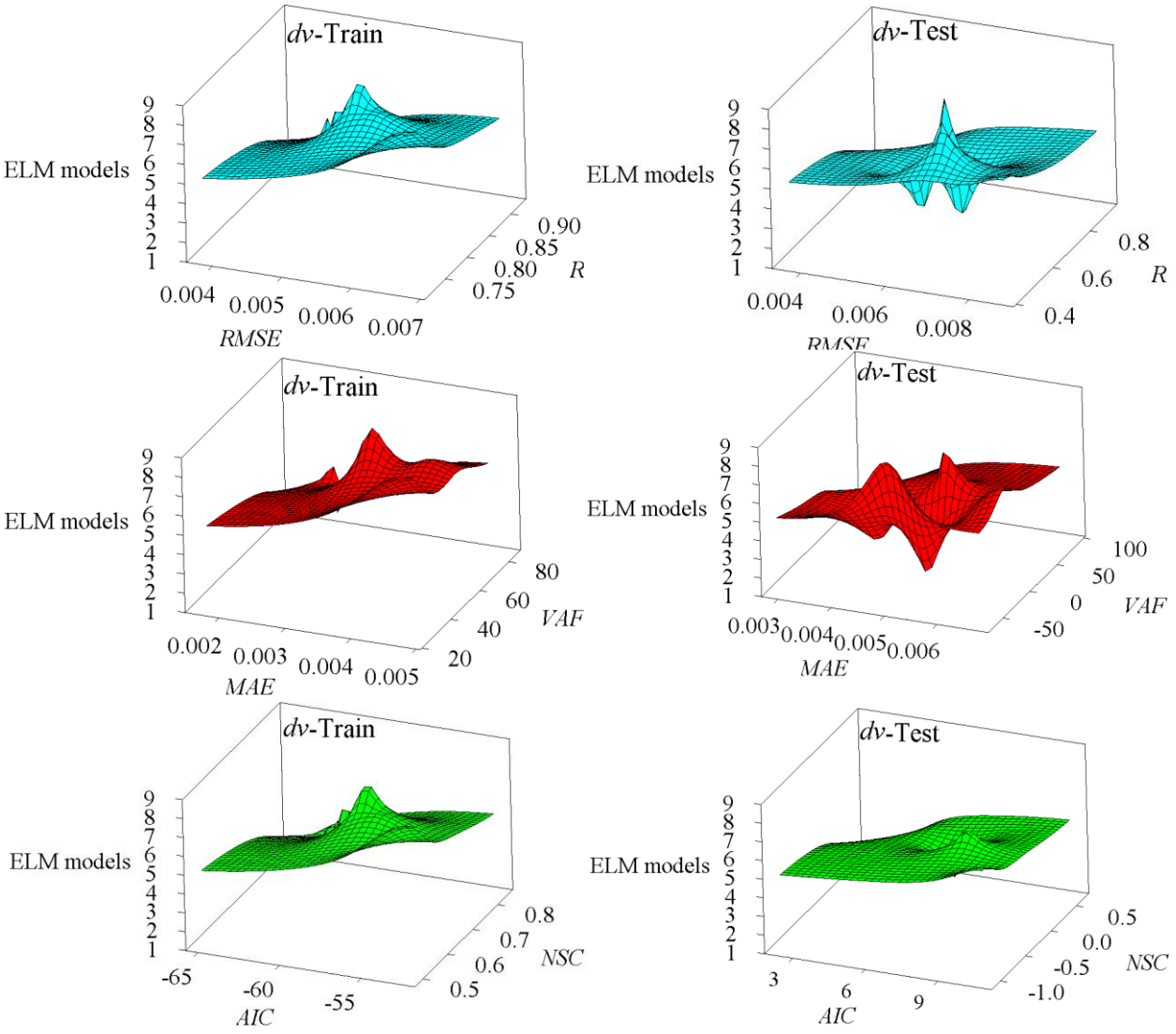


Figure 5-7. The key statistical indices for vertical deformations simulated by ELM models in training and testing modes

The calculated statistical indices for the horizontal forces simulated by the ELM models in the training and testing modes are shown in Figure 5-8. The values of *AIC*, *NSC*, and *VAF* for ELM 1 in the testing phase equaled 406.615, 0.987, and 98.665, with a computed correlation coefficient of 0.995. The *R* value for ELM 2 and ELM 3 models were the same, with 0.994. In the training mode of ELM 4, ELM 5, and ELM 6, the AIC index was respectively estimated to be 875.029,

833.302, and 842.798. Also, the *VAF* value for the testing situation of ELM 7 and ELM 8 was 98.596 and 82.855, respectively. Among all ELM models, ELM 9 had the lowest level of correlation and precision, with a computed *R* value of 0.879. Moreover, the complexity of ELM 9 was quite greater than other ELM models (*AIC*=450.833).

According to the performed analyses, the performance of ELM 1 to ELM 8 dealing with the horizontal forces was partly similar. Thus, ELM 1 was introduced as the premium ELM model since it comprised all input parameters.

Additionally, the performance of the ELM model lessened significantly by removing the soil depth parameter (*y/W*) so the *y/*W was known as the most effective input variable to predict the horizontal forces. It is worth mentioning that the ratio of gouge depth to gouge width (*Ds/W*) had also an important influence and this parameter was graded as the second significant input variable.

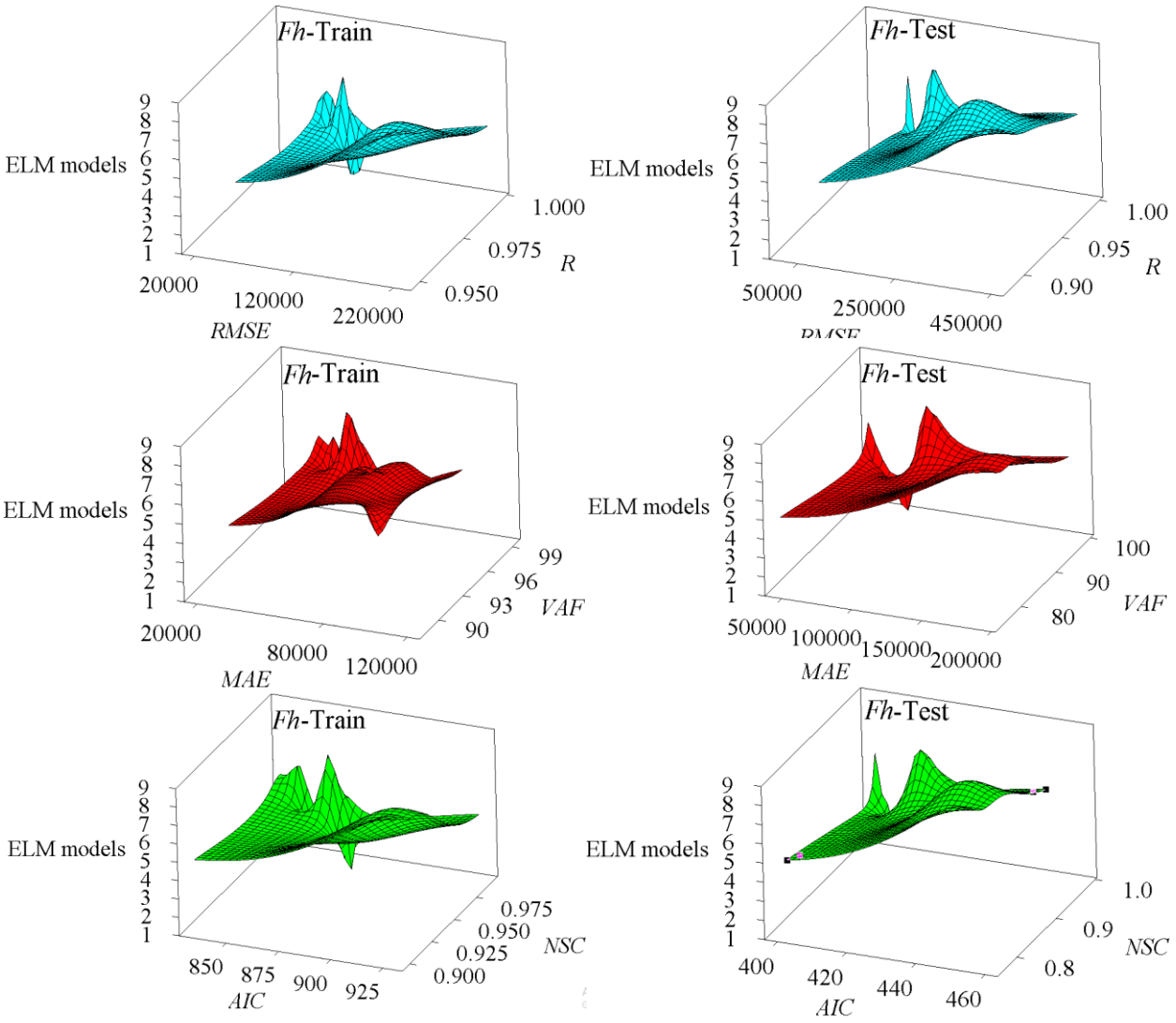


Figure 5-8. The key statistical indices for horizontal forces simulated by ELM models in training and testing modes

Figure 5-9 illustrates the estimated statistical indices for vertical forces simulated by the ELM models in the training and testing stages. The correlation coefficient (*R*), Nash-Sutcliff efficiency coefficient (*NSC*), and Akaike Information Criteria (*AIC*) values for ELM 1 in the testing mode were equal to 0.982, 0.960, and 142.206, respectively. The magnitude of computed *VAF* for ELM 2, ELM 3, and ELM 4 in the testing situation were respectively surmised to be 95.612, 95.438, and

95.231. For testing the ELM 5, ELM 6, ELM 7, and ELM 8 models, the AIC value was estimated at 142.718, 143.080, 141.565, and 142.551. Thus, ELM 1 to ELM 8 had a similar performance in terms of accuracy, correlation, and complexity so as to simulate the vertical forces.

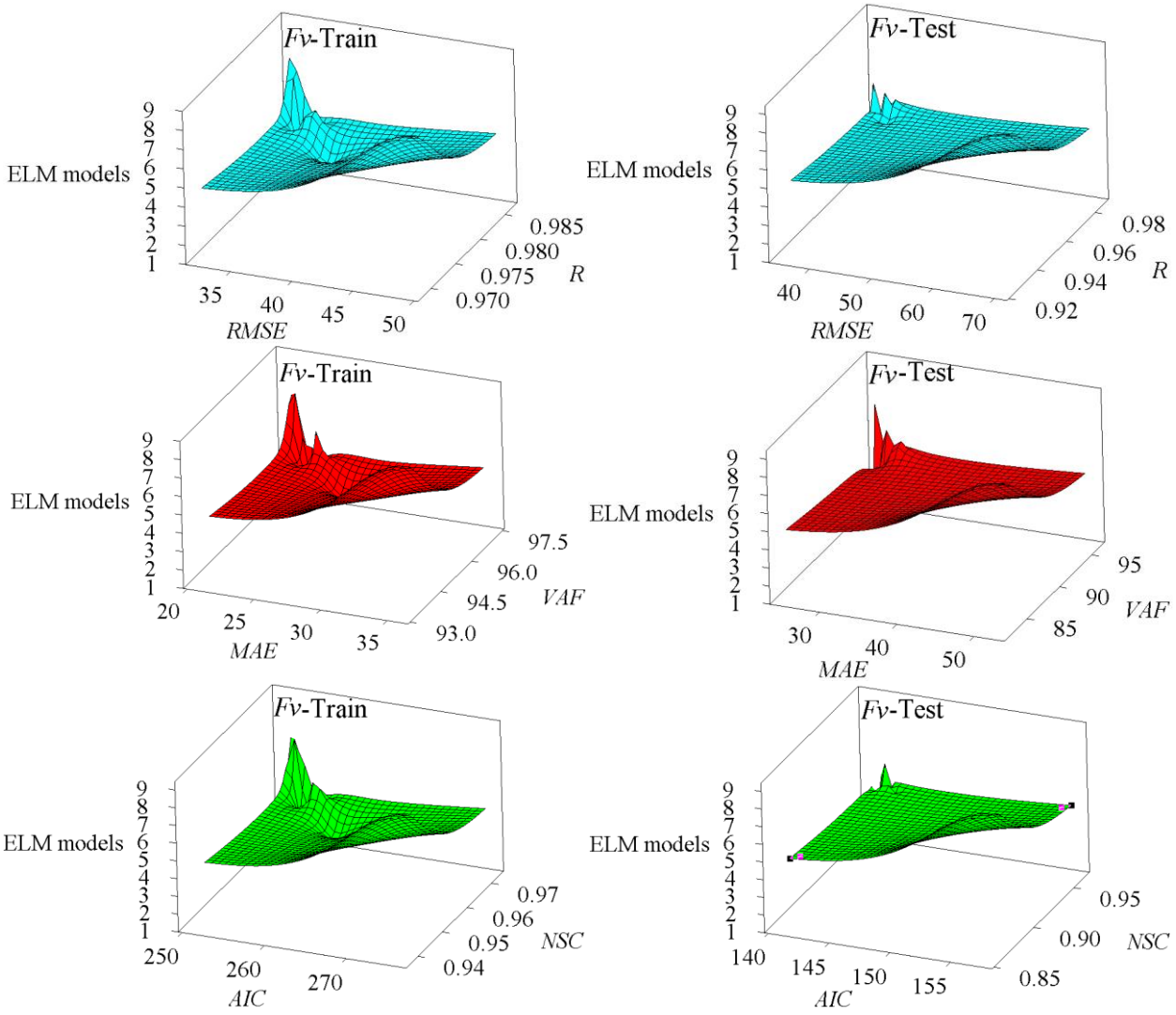


Figure 5-9. The key statistical indices for vertical forces simulated by ELM models in training and testing modes

The implemented sensitivity analysis showed that ELM 9 was the worst ELM model for the estimation of the vertical forces, with computed *R* and *NSC* indices of 0.926 and 0.857. Therefore,

ELM 1 was considered as the superior ELM model in order to predict the vertical deformations and the ratio of soil depth to gouge width (y/W) was distinguished as the most effective input parameter. Moreover, the dimensionless horizontal load factor ($L_h/\gamma_s.W^3$) was ranked as the second significant input parameter.

5.1.3.4. Error analysis (EA)

The results of Error analysis for the superior ELM model (ELM 1) to simulate the horizontal deformation, the vertical deformation, the horizontal reaction force, and the vertical reaction force are shown in Figure 5-10.

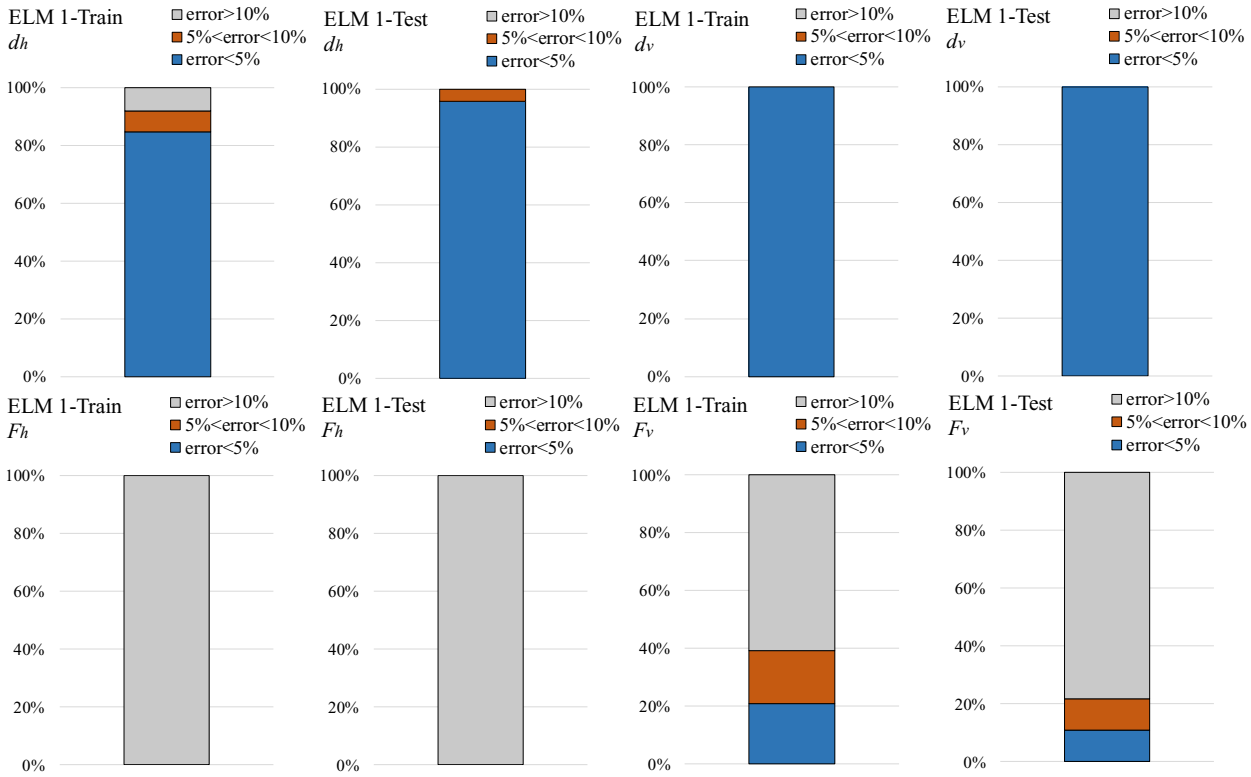


Figure 5-10. Error analysis results of the superior ELM model to simulate horizontal deformations, vertical deformations, horizontal reaction forces, vertical reaction forces

Almost 85% of the horizontal deformations estimated by the ELM model in the training mode had an error of less than 5%, whilst this percentage in the testing mode was nearly 96%. In addition, all simulated vertical deformations in both the training and testing stages had an error of less than 10%. However, all horizontal forces predicted by the ELM model in the training and testing situations were more than 10 percent. It is worth noting that one-fifth and one-tenth of the predicted vertical forces respectively in training and testing modes had an error of less than 5%, whereas approximately 78% of the simulated vertical forces in testing mode had an error of more than 10%. Therefore, the error analysis showed that the superior ELM model predicted the subgouge soil parameters with reasonable accuracy.

5.1.3.5. Uncertainty analysis (UA)

Uncertainty analysis (UA) was performed for evaluating the performance of the superior ELM model to simulate the subgouge parameters. The errors estimated by the ELM model (e_j) was considered as the difference between the simulated (P_j) and the actual (O_j) ice-induced sand parameters as below:

$$e_j = P_j - O_j \tag{5-16}$$

Moreover, the mean value of the calculated error $(\bar{e})$ and the standard deviation (S_e) were computed as below (Azimi and Shiri 2020a):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{5-17}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{5-18}$$

A positive $\bar{e}$ shows that the simulated values are greater than the observed values and the ELM model has overestimated the output parameters. By contrast, a negative $\bar{e}$ shows that the particular

ELM model has underestimated the output parameters, meaning that the simulated ice-gouging parameter is less than the experimental measurements. Moreover, a confidence bound using $\bar{e}$ and S_e was created around an estimated error by means of the "Wilson score method" without the continuity correction (Azimi and Shiri 2020a). The Wilson score interval is an improvement over the normal distribution interval in which an asymmetric normal distribution is used to enhance the confidence interval bound. Subsequently, ±1.96Se results were expressed in a 95% confidence interval (95%CI) which was a 95% estimated error interval, and then the width of uncertainty bound (WUB) was estimated. It is worth mentioning that the width of uncertainty bound (WUB) was calculated as follows:

$$WUB = \pm \frac{(Lower\ bound - upper\ bound)}{2} \tag{5-19}$$

The results of the UA for the best ELM model are tabulated in Table 5-1.

Table 5-1. Results of uncertainty analysis for the superior ELM model

Parameter	$\bar{e}$	S_e	WUB	95%CI
dh/W	-0.007	0.043	±0.007	-0.014 to -0.001
dv/W	-0.001	0.004	±0.0011	-0.002 to 0.0002
Fh/W	-26006	98424	±12704	-38710 to -13302
Fv/W	-3.010	35.120	±5.095	-8.110 to 2.080

The results show that the ELM model underestimated all ice-induced sand parameters including the *dh/W, dv/W, Fh/W*, and *Fv/W*. Furthermore, the 95% confidence interval of the ELM model to simulate the horizontal deformations was from -0.014 to -0.001, with a WUB of ±0.007. The uncertainty bound widths for the best ELM model (ELM 1) to simulate the vertical deformations

(dv/W), the horizontal reaction forces (Fh/W), and the vertical reaction forces (Fv/W) were respectively estimated as ±0.0011, ±12704, and ±5.095. Furthermore, 95% confidence intervals of the ELM model to simulate the horizontal reaction forces (Fh/W) and the vertical reaction forces (Fv/W) were approximated at "-38710 to -13302" and "-8.110 to 2.080", respectively. Therefore, the UA showed that the premium ELM model had an underestimated performance in dealing with the subgouge sand parameters.

5.1.3.6. Comparison of ELM predictions with empirical models

A set of empirical models proposed by Woodworth-Lynes et al. (1996) to estimate the subgouge deformations, and also Barker and Timco (2002) to approximate the reaction forces were used to compare the prediction results with ELM:

$$d_h = 0.6(W.D_s).exp\left(-\frac{2}{3}.\frac{y}{D_s}\right) \quad (5\text{-}20)$$

$$d_v = D_s.exp\left(-\frac{1}{3}.\frac{y}{D_s}\right) \quad (5\text{-}21)$$

$$F = 8.5V + 46.3D_s + 28.5W - 7.6 \quad (5\text{-}22)$$

Table 5-2 shows the comparison between the results of ELM 1 as the best ELM model and these empirical approaches.

Table 5-2. Comparison between the results of ELM 1 and the empirical models

Model	Parameter	R	VAF	RMSE	MAE	NSC	AIC
Eq. (5-20)	*dh*	0.894	79.65	0.425	0.193	0.785	-58.533
Eq. (5-21)	*dv*	0.047	-2670.284	4.304	1.242	-0.006	52.709

Eq. (5-22)	*Fh*	0.281	-6.083E+11	12878254.8	5910865.714	-0.267	1714.365
	dh	0.969	92.840	0.022	0.011	0.937	-22.094
	dv	0.915	81.097	0.004	0.003	0.642	2.339
ELM 1	*Fh*	0.997	99.342	86922.52	48394.323	0.991	401.739
(test)	*Fv*	0.986	95.612	35.543	25.199	0.963	141.291

The correlation coefficient (*R*) values for equations (5-20), (5-21), and (5-22) were equal to 0.894. 0.047, and 0.281, respectively. The *RMSE* values for the horizontal and vertical deformations estimated by the empirical model were surmised as 0.425 and 4.304, with an *MAE* of 1.242 for equation (5-22). Even though the empirical models had less accuracy and correlation in comparison with ELM 1, these empirical approaches were less complex than the ELM model since equations (5-20) to (5-22) possessed lower constants. Thus, the ELM model demonstrated superior performance at dealing with the simulation of the ice scouring parameters.

5.1.3.7. Comparison of ELM predictions with test results

Figures 5-11 to 5-14 depict the horizontal subgouge soil deformation (*dh/W*), the vertical subgouge soil deformation (*dv/W*), the horizontal reaction force ($Fh/\gamma_s W^3$), and the vertical reaction force ($Fv/\gamma_s W^3$) profiles simulated by the superior ELM model (i.e., ELM 1). As shown, ELM 1 showed a good performance to estimate the ice-induced soil parameters in sand since the ELM model simulated the subgouge sand deformations and reactions forces in terms of soil depth (*y/W*), gouge depth (*Ds/W*), friction angle of sand (φ), keel angle (α), berm height (*h'/W*), loads ($Lh/\gamma s.W^3$, $Lv/\gamma s.W^3$), and velocity of keel ($V^2/g.W$). By contrast, the previous investigations proposed a set of

empirical models as a function of iceberg weight, gouge geometry, keel velocity, unit soil weight, and attack angle to approximate the subgouge soil characteristics (Palmer et al. 1990; Woodworth-Lynes et al. 1996; Barker and Timco 2002; Yang 2009). It is worth mentioning that these empirical models as a function of two or three input parameters were suggested for a specific dataset, while the premium ELM model managed to simulate the ice-induced sand features in a comprehensive dataset, ranging from 1g experimental studies to centrifuge tests.

Furthermore, ELM 1 could appropriately simulate the maximum horizontal subgouge deformations at y/W=0, where just beneath the ice keel in the experimental models, and this parameter decreased at greater depth, meaning that the overall trend of horizontal subgouge deformations was correctly modeled by the ELM network.

In comparison with the centrifuge results, ELM 1 demonstrated a better performance at dealing with the 1g experimental measurements [Figures 5-11(a) to 5-11(e); Figures 5-12(a) to 5-12(d)]. There were several fluctuations in these centrifuge experimental values, rather ELM 1 predicted these measurements with an acceptable agreement [Figures 5-11(p) to Figures 5-11(t)].

Notwithstanding that the ELM network attempted to simulate the horizontal deformations with its highest performance in a wide range of experimental measurements, some discrepancies were observed with the centrifuge results because the maximum horizontal deformation parameter (dh/W) in the 1g tests varied from 0.9 to 1.4, whereas this range for the centrifuge tests was from 0.002 to 0.9. It means that the ELM network had better performance for the simulation of large deformations.

Unlike the horizontal deformation where its magnitude was decreased by a hyperbolic curve, it seems that the experimental vertical deformations lacked a meaningful trend because the vertical subgouge deformations in the 1g test (Paulin 1992) fluctuated severely [Figures 8-12(a) to Figures

8-12(d)] though ELM 1 managed to model all vertical displacements with both linear and nonlinear trends [Figures 8-12(e) to Figures 8-12(l)].

As shown in figures 5-13 and 5-14, the minimum reaction forces were recorded in the initial position of the iceberg (x/W=0) and these values increased along the scour axis. Despite the oscillation trend in the experimental reaction forces, ELM 1 used a nonlinear behavior to predict the horizontal reaction forces (Figures 5-13). Moreover, the vertical reaction forces were reported by positive and negative values but ELM 1 simulated these values linearly (Figures 5-14).

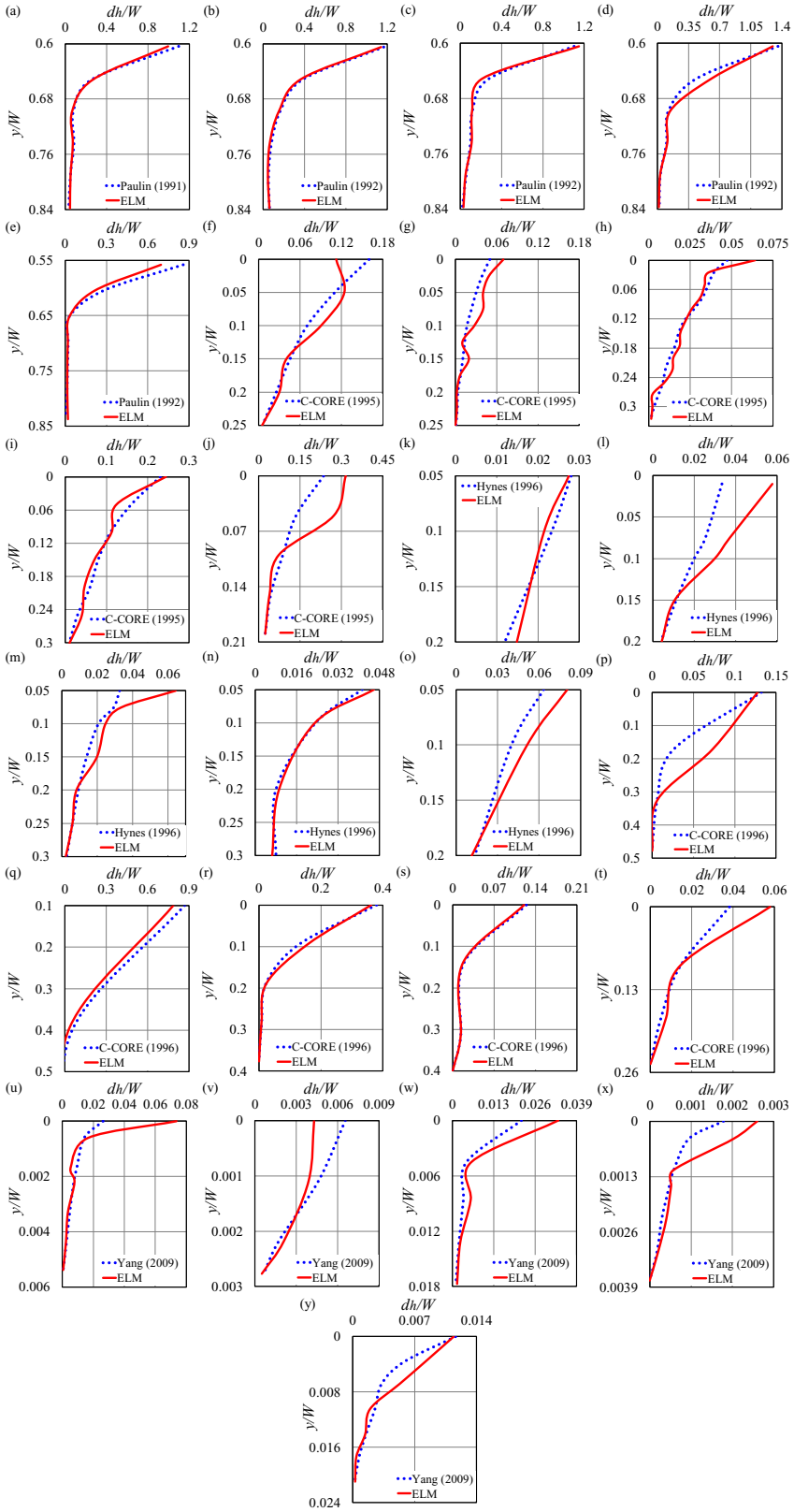


Figure 5-11. Horizontal deformation (dh/W) profiles simulated by the superior ELM model

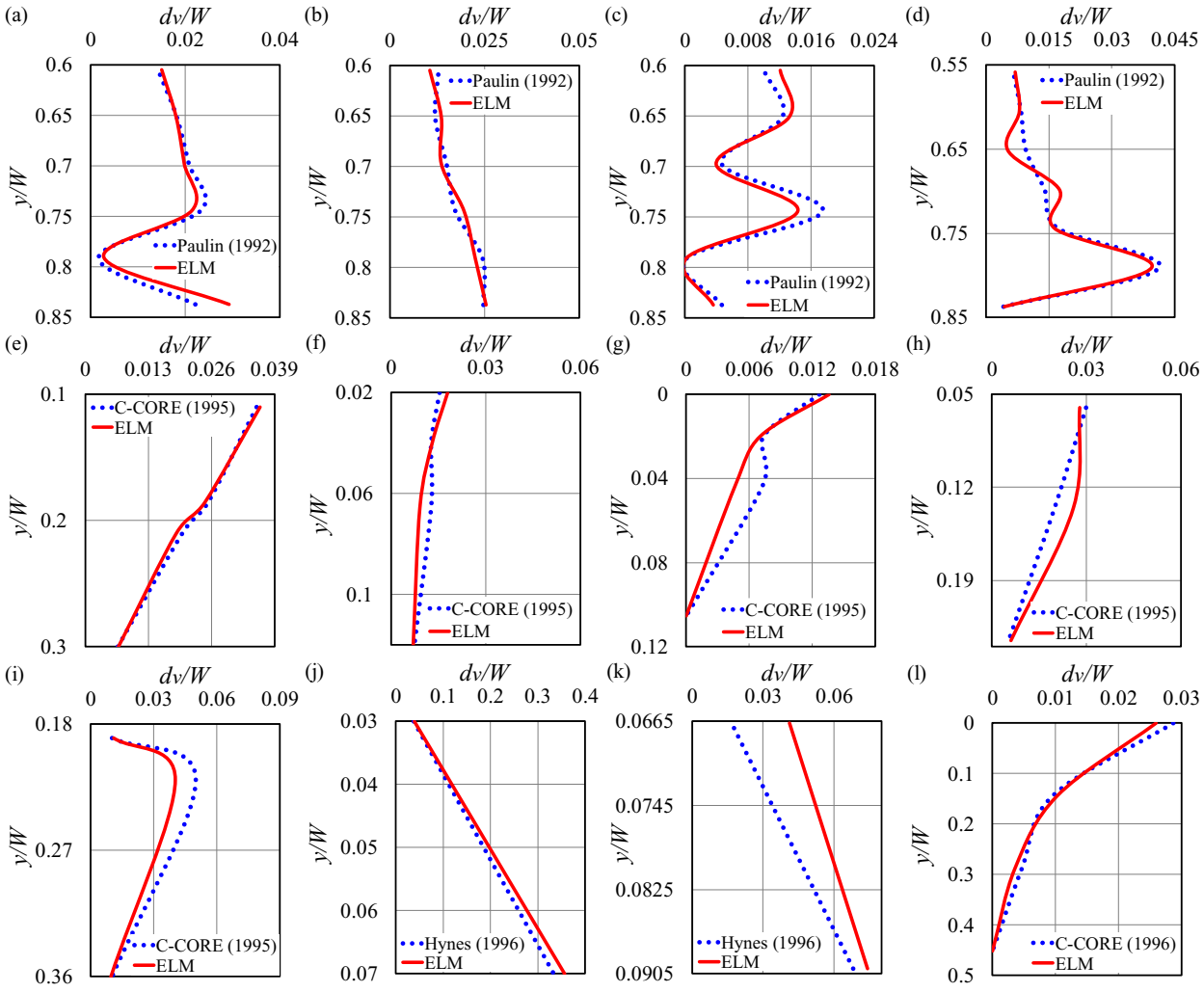


Figure 5-12. Vertical deformation (dv/W) profiles simulated by the superior ELM model

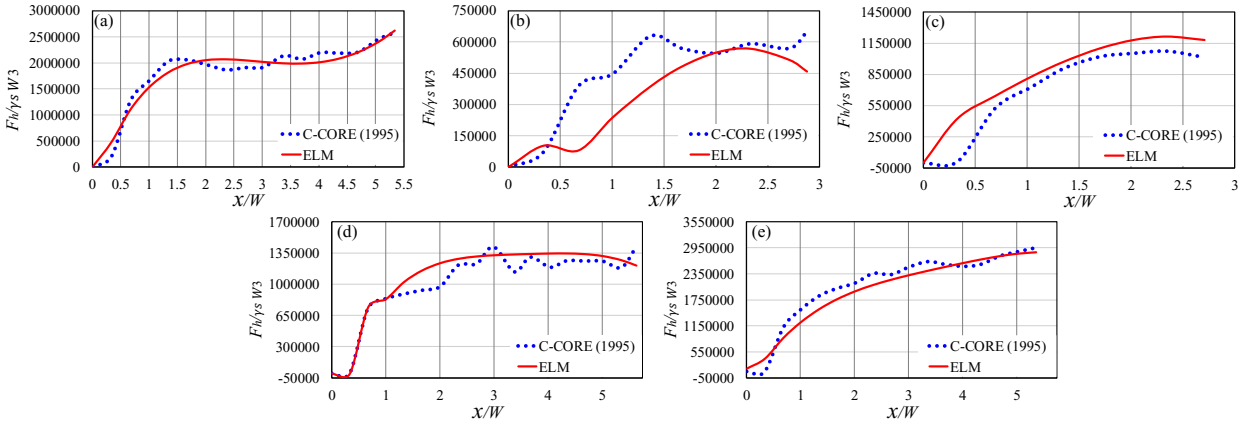


Figure 5-13. Horizontal reaction force (Fh/$\gamma_s W^3$) profiles simulated by the superior ELM model

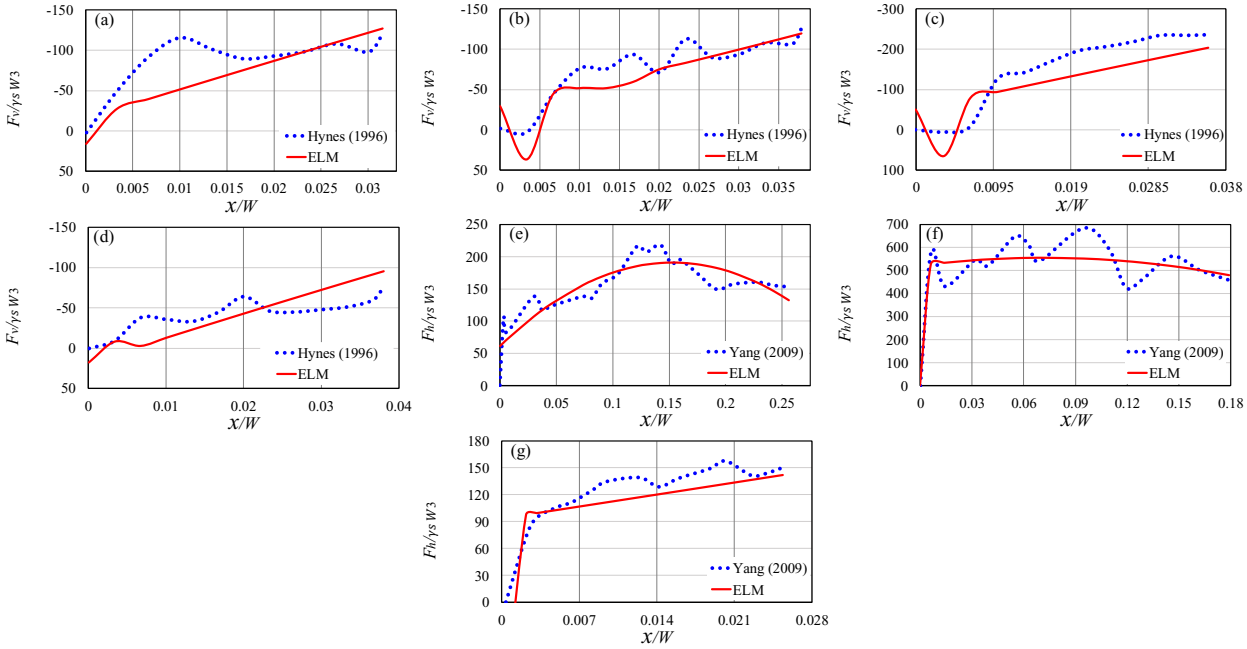


Figure 5-14. Vertical reaction force (Fv/$\gamma_s W^3$) profiles simulated by the superior ELM model

According to the obtained results in this investigation, ELM 1, as the superior ELM model, estimated the subgouge sand parameters including deformations and reaction forces with an underestimated performance along with a high level of accuracy, correlation, and low level of

complexity. Thus, a set of matrices for estimation of the subgouge deformations and the reaction forces through ELM 1 were presented in Appendix 1

5.1.4. Conclusion

The subgouge soil deformation and the keel reaction forces in the sand were simulated by the extreme learning machine (ELM). A set of ELM-based matrices were proposed to estimate the subgouge sand deformations and reaction forces. The ELM performance was evaluated through comparisons with existing empirical and experimental studies. The extreme learning machine (ELM) was found as a potential alternative tool for ice-gouging analysis. The study showed that the ELM can perform accurate, quick, and less expensive simulations compared with 3D FEA models and experimental studies.

- ELM 1 as the best model simulated the subgouge sand features in terms of soil depth, gouge depth, friction angle of sand, keel angle, berm height, loads, and velocity of the keel.
- The ELM model predicted the maximum horizontal subgouge deformations just under the ice keel in the experimental tests, whilst this parameter decreased at greater depth.
- Error analysis showed an error of less than 5% for about 85% of simulated subgouge soil deformations. As the soil depth parameter was an eliminated input, the ELM might result in less accurate predictions.
- Overall, the ELM was found to slightly underestimate the test results.
- The ratio of soil depth to gouge depth was found to be the most influencing input parameter to estimate the horizontal subgouge deformations and reaction forces, whilst

the shear strength parameter of the sandy seabed was identified as the most effective input to predict the vertical deformations.

The study provided a good insight into the applicability of the ELM method to ice-gouging in sand. More experimental and numerical investigations with managed input and output parameters would improve the performance of the developed ELM model as a potential alternative to the existing approaches.

Funding and acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

Appendix

ELM-based equations

ELM 1 was considered as the premium ELM model to simulate the subgouge sand parameters (η) comprising the subgouge deformations (*dh/W* & *dv/W*) and the subgouge forces (*Fh/W* & *Fv/W*). Among all ELM models, ELM 1 had a reasonable accuracy, correlation, and complexity. It is worth mentioning that ELM 1 estimated the subgouge sand parameters by using all input parameters and the general format of the extreme learning machine-based formula was as the following form:

$$\eta = \left[\frac{1}{1 + exp(InW \times InV \times BHI)}\right]^{T} \times OutW \tag{5-23}$$

where, InW is the matrix of input weight, InV is the matrix of input parameters, BHI is the matrix of bias of hidden layer, and $OutW$ is the matrix of output weights. As shown, ELM 1 had a good performance to estimate the horizontal ice-induced deformations (*dh/W*) in the sand. Thus, matrices of ELM 1 to estimate the horizontal deformations were presented below:

$$InV = \begin{bmatrix} \frac{y}{W} \\ \frac{D_s}{W} \\ \varphi \\ \alpha \\ \frac{h'}{W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, \; BHI = \begin{bmatrix} 0.108 \\ 0.212 \\ 0.181 \\ 0.091 \\ 0.160 \\ 0.274 \\ 0.173 \\ 0.038 \\ 0.263 \\ 0.285 \\ 0.233 \\ 0.159 \\ 0.151 \\ 0.117 \\ 0.015 \\ 0.122 \\ 0.066 \\ 0.261 \\ 0.036 \\ 0.182 \\ 0.253 \\ 0.183 \\ 0.179 \\ 0.299 \\ 0.104 \\ 0.319 \\ 0.019 \\ 0.213 \end{bmatrix}, InW = \quad (5\text{-}24)$$

$$\begin{bmatrix} 0.076 & 0.053 & 0.231 & 0.351 & 0.091 & 0.039 & 0.224 & 0.113 \\ 0.122 & 0.190 & 0.258 & 0.168 & 0.473 & 0.003 & 0.234 & 0.094 \\ 0.353 & 0.072 & 0.025 & 0.032 & 0.173 & 0.305 & 0.047 & 0.183 \\ 0.180 & 0.082 & 0.456 & 0.019 & 0.103 & 0.237 & 0.104 & 0.006 \\ 0.016 & 0.291 & 0.021 & 0.103 & 0.301 & 0.002 & 0.132 & 0.185 \\ 0.046 & 0.182 & 0.093 & 0.317 & 0.171 & 0.231 & 0.135 & 0.109 \\ 0.335 & 0.052 & 0.088 & 0.245 & 0.026 & 0.123 & 0.101 & 0.309 \\ 0.261 & 0.144 & 0.017 & 0.092 & 0.167 & 0.083 & 0.124 & 0.303 \\ 0.028 & 0.291 & 0.367 & 0.307 & 0.154 & 0.118 & 0.009 & 0.009 \\ 0.063 & 0.254 & 0.131 & 0.117 & 0.201 & 0.092 & 0.003 & 0.095 \\ 0.249 & 0.100 & 0.136 & 0.122 & 0.166 & 0.140 & 0.309 & 0.212 \\ 0.174 & 0.072 & 0.117 & 0.159 & 0.227 & 0.038 & 0.243 & 0.438 \\ 0.164 & 0.183 & 0.106 & 0.029 & 0.012 & 0.067 & 0.242 & 0.057 \\ 0.265 & 0.144 & 0.054 & 0.118 & 0.054 & 0.243 & 0.025 & 0.103 \\ 0.081 & 0.266 & 0.228 & 0.122 & 0.119 & 0.043 & 0.069 & 0.339 \\ 0.176 & 0.167 & 0.098 & 0.388 & 0.082 & 0.195 & 0.013 & 0.080 \\ 0.213 & 0.164 & 0.076 & 0.007 & 0.073 & 0.392 & 0.229 & 0.219 \\ 0.010 & 0.090 & 0.079 & 0.220 & 0.044 & 0.123 & 0.061 & 0.459 \\ 0.135 & 0.321 & 0.120 & 0.038 & 0.241 & 0.058 & 0.105 & 0.002 \\ 0.050 & 0.217 & 0.314 & 0.468 & 0.283 & 0.059 & 0.187 & 0.032 \\ 0.200 & 0.150 & 0.052 & 0.049 & 0.072 & 0.136 & 0.376 & 0.090 \\ 0.361 & 0.218 & 0.104 & 0.042 & 0.100 & 0.418 & 0.067 & 0.070 \\ 0.260 & 0.039 & 0.015 & 0.050 & 0.112 & 0.216 & 0.150 & 0.101 \\ 0.042 & 0.141 & 0.303 & 0.114 & 0.098 & 0.399 & 0.410 & 0.060 \\ 0.002 & 0.315 & 0.315 & 0.036 & 0.057 & 0.039 & 0.167 & 0.041 \\ 0.150 & 0.045 & 0.038 & 0.186 & 0.191 & 0.172 & 0.260 & 0.135 \\ 0.111 & 0.326 & 0.185 & 0.047 & 0.224 & 0.049 & 0.140 & 0.138 \\ 0.252 & 0.051 & 0.026 & 0.086 & 0.367 & 0.123 & 0.212 & 0.031 \end{bmatrix}, \; OutW = \begin{bmatrix} 29.275 \\ 72.459 \\ 165.167 \\ 2160.705 \\ 596.242 \\ 35395.487 \\ 2167.263 \\ 22.303 \\ 17132.658 \\ 75.972 \\ 62.933 \\ 4.386 \\ 866.932 \\ 311.317 \\ 594.914 \\ 42.618 \\ 703.678 \\ 129.912 \\ 775.673 \\ 17100.189 \\ 168.320 \\ 1054.218 \\ 54.598 \\ 292.875 \\ 23846.632 \\ 30.669 \\ 2671.287 \\ 130.130 \end{bmatrix}$$

Moreover, the matrices of ELM 1 to estimate the vertical sand deformations (*dv/W*) were provided as follows:

$$InV = \begin{bmatrix} \frac{y}{W} \\ \frac{D_s}{W} \\ \varphi \\ \alpha \\ \frac{h'}{W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, BHI = \begin{bmatrix} 0.007 \\ 0.081 \\ 0.077 \\ 0.169 \\ 0.217 \\ 0.218 \\ 0.130 \\ 0.112 \\ 0.298 \\ 0.113 \\ 0.060 \\ 0.166 \\ 0.155 \\ 0.015 \\ 0.285 \\ 0.298 \\ 0.212 \\ 0.030 \\ 0.140 \\ 0.162 \\ 0.277 \\ 0.143 \\ 0.048 \\ 0.175 \\ 0.292 \\ 0.291 \\ 0.167 \\ 0.289 \end{bmatrix}, \quad (5\text{-}25)$$

$$InW = \begin{bmatrix} 0.352 & 0.139 & 0.018 & 0.165 & 0.031 & 0.084 & 0.013 & 0.026 \\ 0.223 & 0.157 & 0.193 & 0.048 & 0.004 & 0.573 & 0.026 & 0.174 \\ 0.020 & 0.103 & 0.058 & 0.001 & 0.001 & 0.293 & 0.064 & 0.242 \\ 0.113 & 0.048 & 0.451 & 0.200 & 0.086 & 0.022 & 0.138 & 0.136 \\ 0.006 & 0.399 & 0.243 & 0.023 & 0.259 & 0.254 & 0.296 & 0.121 \\ 0.140 & 0.018 & 0.115 & 0.153 & 0.222 & 0.258 & 0.028 & 0.184 \\ 0.098 & 0.011 & 0.420 & 0.200 & 0.222 & 0.070 & 0.001 & 0.206 \\ 0.026 & 0.045 & 0.118 & 0.361 & 0.182 & 0.016 & 0.222 & 0.032 \\ 0.229 & 0.106 & 0.122 & 0.335 & 0.326 & 0.166 & 0.166 & 0.001 \\ 0.274 & 0.161 & 0.106 & 0.124 & 0.097 & 0.079 & 0.276 & 0.225 \\ 0.442 & 0.059 & 0.108 & 0.160 & 0.002 & 0.132 & 0.381 & 0.057 \\ 0.309 & 0.003 & 0.228 & 0.348 & 0.258 & 0.056 & 0.278 & 0.277 \\ 0.099 & 0.064 & 0.142 & 0.144 & 0.047 & 0.341 & 0.064 & 0.075 \\ 0.118 & 0.280 & 0.229 & 0.011 & 0.254 & 0.012 & 0.277 & 0.632 \\ 0.060 & 0.047 & 0.194 & 0.248 & 0.382 & 0.085 & 0.148 & 0.252 \\ 0.108 & 0.335 & 0.073 & 0.044 & 0.114 & 0.066 & 0.009 & 0.047 \\ 0.096 & 0.313 & 0.023 & 0.089 & 0.213 & 0.136 & 0.356 & 0.083 \\ 0.050 & 0.307 & 0.090 & 0.263 & 0.152 & 0.175 & 0.067 & 0.034 \\ 0.025 & 0.286 & 0.088 & 0.001 & 0.144 & 0.037 & 0.158 & 0.036 \\ 0.107 & 0.024 & 0.101 & 0.087 & 0.188 & 0.092 & 0.098 & 0.039 \\ 0.141 & 0.098 & 0.006 & 0.259 & 0.241 & 0.048 & 0.001 & 0.045 \\ 0.126 & 0.134 & 0.244 & 0.085 & 0.183 & 0.159 & 0.082 & 0.012 \\ 0.227 & 0.117 & 0.015 & 0.146 & 0.102 & 0.036 & 0.286 & 0.360 \\ 0.034 & 0.010 & 0.129 & 0.435 & 0.154 & 0.061 & 0.213 & 0.090 \\ 0.330 & 0.212 & 0.126 & 0.038 & 0.147 & 0.111 & 0.047 & 0.046 \\ 0.135 & 0.247 & 0.306 & 0.056 & 0.189 & 0.127 & 0.220 & 0.097 \\ 0.278 & 0.264 & 0.263 & 0.015 & 0.124 & 0.194 & 0.198 & 0.178 \\ 0.066 & 0.218 & 0.037 & 0.003 & 0.237 & 0.313 & 0.123 & 0.004 \end{bmatrix}, \; OutW =$$

$$\begin{bmatrix} 6.367 \\ 129.078 \\ 13.435 \\ 734.438 \\ 1073.040 \\ 1914.414 \\ 3822.172 \\ 137.341 \\ 28.803 \\ 24.231 \\ 412.769 \\ 736.786 \\ 11.185 \\ 3327.398 \\ 385.798 \\ 4.120 \\ 9.616 \\ 43.026 \\ 300.883 \\ 74.807 \\ 342.412 \\ 750.359 \\ 18.367 \\ 124.443 \\ 336.378 \\ 305.840 \\ 78.977 \\ 31.751 \end{bmatrix}$$

Besides, the matrices of ELM 1 to estimate the ice-induced horizontal forces ($Fh/\gamma sW^3$) in the sand were written as below:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_s}{W} \\ \varphi \\ \alpha \\ \frac{h'}{W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, BHI = \begin{bmatrix} -0.173 \\ -0.218 \\ -0.337 \\ 0.036 \\ -0.026 \\ 0.113 \\ -0.077 \\ -0.133 \\ -0.026 \\ 0.295 \\ -0.052 \\ 0.321 \\ 0.331 \\ 0.233 \\ -0.014 \\ -0.101 \\ -0.023 \\ -0.117 \\ -0.020 \\ -0.281 \\ -0.244 \\ -0.235 \\ -0.023 \\ 0.083 \\ -0.162 \\ -0.342 \\ -0.095 \\ 0.177 \end{bmatrix}, \quad (5\text{-}26)$$

$$InW = \begin{bmatrix} -0.158 & 0.154 & -0.294 & -0.299 & 0.018 & 0.032 & -0.197 & 0.258 \\ 0.340 & -0.035 & -0.082 & -0.268 & -0.013 & -0.158 & 0.098 & -0.434 \\ -0.050 & 0.200 & 0.160 & 0.287 & -0.250 & -0.034 & 0.046 & -0.161 \\ 0.177 & 0.119 & 0.061 & 0.252 & -0.122 & 0.222 & 0.264 & -0.023 \\ 0.278 & -0.153 & 0.073 & -0.149 & -0.090 & 0.211 & -0.003 & 0.327 \\ 0.002 & 0.197 & -0.269 & -0.134 & 0.214 & 0.254 & -0.262 & -0.179 \\ 0.264 & -0.012 & -0.045 & -0.233 & -0.232 & 0.128 & -0.222 & -0.042 \\ 0.033 & 0.445 & 0.018 & -0.089 & 0.029 & -0.082 & 0.203 & 0.282 \\ -0.095 & -0.260 & -0.093 & -0.128 & -0.269 & -0.142 & 0.003 & -0.011 \\ -0.163 & -0.220 & -0.006 & 0.121 & -0.364 & 0.232 & -0.033 & 0.312 \\ 0.233 & -0.101 & 0.260 & 0.091 & 0.022 & -0.085 & -0.311 & -0.044 \\ -0.202 & -0.113 & 0.019 & -0.056 & -0.172 & -0.240 & -0.058 & -0.023 \\ -0.279 & -0.097 & 0.178 & -0.188 & -0.197 & 0.190 & -0.082 & -0.102 \\ -0.216 & -0.019 & -0.038 & -0.043 & -0.052 & 0.115 & -0.413 & -0.229 \\ -0.103 & -0.033 & -0.292 & 0.342 & -0.192 & -0.200 & -0.232 & 0.131 \\ -0.101 & 0.023 & 0.174 & 0.198 & 0.212 & 0.184 & -0.217 & -0.052 \\ 0.046 & -0.171 & -0.367 & 0.320 & -0.136 & -0.024 & 0.147 & -0.194 \\ 0.098 & 0.346 & -0.134 & 0.056 & -0.109 & 0.214 & 0.048 & -0.053 \\ 0.075 & -0.021 & -0.234 & -0.340 & -0.267 & 0.038 & 0.171 & 0.053 \\ 0.098 & -0.097 & -0.320 & 0.256 & 0.328 & -0.147 & 0.024 & 0.217 \\ -0.071 & 0.385 & 0.190 & -0.024 & -0.163 & -0.059 & -0.033 & 0.243 \\ 0.382 & -0.213 & -0.030 & 0.045 & 0.037 & 0.043 & -0.263 & 0.186 \\ 0.002 & -0.151 & -0.116 & -0.106 & 0.320 & 0.289 & 0.172 & 0.055 \\ -0.292 & -0.088 & 0.079 & 0.023 & 0.131 & 0.356 & 0.171 & -0.202 \\ 0.064 & -0.281 & 0.192 & -0.035 & -0.061 & 0.064 & 0.339 & 0.033 \\ 0.356 & 0.114 & 0.152 & 0.173 & -0.172 & 0.229 & -0.175 & -0.084 \\ -0.018 & 0.006 & -0.302 & 0.140 & -0.151 & 0.418 & 0.0431 & 0.036 \\ -0.051 & -0.177 & 0.224 & -0.001 & 0.196 & 0.099 & -0.084 & 0.272 \end{bmatrix},$$

$$OutW = \begin{bmatrix} -579043842.543 \\ 11327617709.333 \\ 25804356551.036 \\ 23880116986.524 \\ 49235561.598 \\ -2107351719.547 \\ -6133664925.287 \\ -324830407.099 \\ 14514446831.585 \\ -691139063.538 \\ 69679699824.426 \\ -270014640.236 \\ -180938532.185 \\ 520312750 \\ 16177672.834 \\ -45918830515.734 \\ 86028151.093 \\ -641785582.707 \\ 914415067.981 \\ 444471868.539 \\ 23259646121.250 \\ -118218046.915 \\ 4266202972.363 \\ 109167304.838 \\ 7941707982.082 \\ 9530629767.052 \\ 2019320579.769 \\ -113281225148.813 \end{bmatrix}$$

Furthermore, the matrices of this ELM model to surmise the ice-induced vertical reaction forces ($Fv/\gamma sW^3$) in a sandy seabed were suggested as follows:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_s}{W} \\ \varphi \\ \alpha \\ \frac{h'}{W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, BHI = \begin{bmatrix} -0.256 \\ 0.079 \\ 0.091 \\ -0.278 \\ 0.127 \\ -0.222 \\ -0.338 \\ 0.336 \\ -0.215 \\ -0.119 \\ 0.311 \\ 0.097 \\ 0.326 \\ 0.152 \\ -0.042 \\ -0.159 \\ -0.083 \\ 0.003 \\ -0.140 \\ -0.106 \\ -0.045 \\ 0.135 \\ 0.138 \\ -0.104 \\ 0.037 \\ 0.129 \\ 0.029 \\ 0.341 \end{bmatrix}, \quad (5\text{-}27)$$

$$InW = \begin{bmatrix} -0.045 & -0.358 & 0.040 & 0.217 & -0.267 & 0.029 & 0.051 & -0.338 \\ 0.020 & 0.120 & 0.061 & 0.146 & -0.123 & -0.137 & 0.232 & 0.254 \\ 0.009 & 0.371 & 0.027 & 0.181 & 0.115 & -0.013 & -0.055 & -0.284 \\ 0.300 & -0.042 & -0.015 & 0.086 & -0.045 & 0.032 & -0.185 & -0.208 \\ -0.293 & -0.060 & -0.118 & 0.273 & 0.029 & -0.231 & -0.196 & 0.283 \\ -0.107 & 0.386 & 0.208 & -0.028 & -0.351 & 0.136 & 0.236 & -0.132 \\ -0.259 & -0.235 & -0.155 & 0.220 & -0.044 & -0.001 & -0.018 & 0.167 \\ -0.324 & 0.014 & 0.166 & -0.165 & -0.013 & -0.336 & 0.347 & 0.253 \\ 0.317 & 0.153 & 0.149 & -0.209 & 0.012 & -0.338 & -0.056 & -0.043 \\ 0.139 & 0.155 & -0.012 & 0.223 & 0.454 & 0.389 & -0.001 & 0.269 \\ -0.059 & 0.146 & -0.012 & -0.122 & -0.077 & 0.163 & -0.101 & 0.109 \\ -0.278 & 0.050 & -0.093 & -0.032 & 0.198 & -0.106 & 0.218 & -0.326 \\ -0.107 & 0.401 & 0.031 & 0.125 & -0.210 & 0.018 & -0.084 & -0.015 \\ 0.070 & 0.113 & 0.259 & -0.188 & 0.110 & -0.071 & -0.161 & 0.055 \\ -0.084 & -0.079 & 0.095 & -0.296 & -0.046 & 0.160 & 0.028 & -0.046 \\ 0.251 & -0.106 & 0.191 & 0.115 & -0.183 & 0.111 & 0.194 & 0.169 \\ -0.272 & -0.154 & 0.267 & -0.277 & 0.241 & 0.182 & 0.225 & 0.050 \\ 0.091 & -0.156 & -0.031 & 0.138 & 0.234 & -0.167 & 0.107 & -0.129 \\ -0.186 & -0.020 & 0.057 & 0.180 & -0.425 & 0.014 & -0.239 & 0.043 \\ 0.257 & -0.036 & 0.141 & 0.077 & -0.183 & 0.064 & 0.086 & 0.440 \\ 0.211 & -0.088 & -0.285 & -0.164 & -0.033 & -0.052 & 0.224 & 0.046 \\ 0.001 & 0.248 & 0.064 & 0.033 & 0.056 & 0.162 & 0.079 & 0.006 \\ -0.055 & -0.079 & -0.233 & -0.254 & -0.203 & 0.307 & 0.024 & -0.096 \\ -0.146 & 0.175 & -0.149 & -0.257 & 0.039 & -0.271 & -0.437 & 0.127 \\ -0.191 & 0.066 & 0.320 & 0.237 & 0.068 & 0.010 & 0.084 & -0.138 \\ 0.061 & 0.158 & -0.431 & -0.206 & -0.193 & -0.057 & 0.247 & 0.048 \\ 0.013 & 0.229 & -0.387 & 0.281 & 0.087 & -0.087 & 0.311 & -0.005 \\ -0.233 & 0.072 & -0.203 & -0.064 & 0.039 & 0.413 & -0.125 & 0.114 \end{bmatrix},$$

$$OutW = \begin{bmatrix} -9846993.568 \\ -3754470.686 \\ -57306.747 \\ 739540.221 \\ 567279.640 \\ 4111811.376 \\ 115947.041 \\ 776649.753 \\ -101460.754 \\ -3108267.921 \\ -412567.206 \\ 528094.443 \\ -4728862.345 \\ 896003.104 \\ 592571.805 \\ 6710766.618 \\ -571166.680 \\ -287955.147 \\ -5324488.298 \\ -1195984.731 \\ 37129.165 \\ 6906839.302 \\ 61221.738 \\ 625421.204 \\ 8430674.410 \\ 70.211 \\ 1328898.924 \\ 555030.486 \end{bmatrix}$$

References

Abdalla, B., Pike, K., Eltaher, A., Jukes, P., Duron, B., 2009. Development and validation of a coupled Eulerian Lagrangian finite element ice scour model. In International Conference on Offshore Mechanics and Arctic Engineering, 43451: 87-95.

Alba, J. L., 2015, Ice Scour and Gouging Effects With Respect to Pipeline and Wellhead Placement and Design, Bureau of Safety and Environmental Enforcement (BSEE), Wood Group Kenny, Report No. 100100.01.PL.REP.004.

Arnau, S., Ivanović, A., 2019. Rate effects during ice scour in sand. Cold Regions Science and Technology, 158: 182-194.

Azimi, H., Bonakdari, H., Ebtehaj, I., 2017. Sensitivity analysis of the factors affecting the discharge capacity of side weirs in trapezoidal channels using extreme learning machines. Flow Measurement and Instrumentation, 54: 216-223.

Azimi, H., Shiri, H., 2020a. Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b, Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Babaei, M.H., Sudom, D., 2014. Ice-seabed gouging database: review and analysis of available numerical models. Offshore Technology Conference, OTC-24603-MS.

Barker, A., Timco, G., 2002. Laboratory experiments of ice scour processes: rigid ice indentor. Cold regions science and technology, 35(3): 195-206.

Barrette, P. 2011., Offshore pipeline protection against seabed gouging by ice: An overview. Cold regions science and technology, 69(1): 3-20.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour displacement in clays from physical model tests. International Pipeline Conference, 48609: 239-245.

Bland, J.M., Peacock, J.L., 2002. Interpreting Statistics with Confidence,

Journal of General Internal Medicine, 4(3): 176–180.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2.

C-CORE., 1996. PRISE Phase 3c: Extreme lee Gouge Event - Modeling and Interpretation, C-CORE Report 96-C32.

Cinicioglu, O., Abadkon, A., 2015. Dilatancy and friction angles based on in situ soil conditions. Journal of Geotechnical and Geoenvironmental Engineering, 141(4): 06014019.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In 2004 IEEE international joint conference on neural networks, IEEE Cat. No. 04CH37541, Vol. 2, pp. 985-990.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing, 70(1-3): 489-501.

Huang, G.B., Zhou, H., Ding, X., Zhang, R., 2011. Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(2): 513-529.

Hynes, F. Centrifuge modelling of ice scour in sand. Doctoral dissertation, Memorial University of Newfoundland. (1996).

Lach, P.R., 1996. Centrifuge modelling of large soil deformation due to ice scour. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Kenny, S., Barrett, J., Phillips, R., Popescu, R., 2007. Integrating geohazard demand and structural capacity modelling within a probabilistic design framework for offshore arctic pipelines. The Seventeenth International Offshore and Polar Engineering Conference, ISOPE-I-07-492.

Kioka, S. D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. Thirteenth International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Konuk, I., Yu, S., 2007. A pipeline case study for ice scour design. International Conference on Offshore Mechanics and Arctic Engineering, Vol. 42711, pp. 163-169.

Lanan, G.A., Ennis, J.O., Egger, P.S., Yockey, K.E., 2001. Northstar offshore Arctic pipeline design and construction. Offshore Technology Conference, OTC-13133-MS.

Liferov, P., Nes, H., Asklund, J., Shkhinek, K., Jilenkov, A., 2014. Ice-gouging and its effect on pipelines. Offshore Technology Conference, OTC-24605-MS.

Mitchell, J.K., Soga, K., 2005. Fundamentals of soil behavior, Vol. 3. New York: John Wiley & Sons.

Nematzadeh, A, Shiri, H., 2019a. Large deformation analysis of ice scour process in dense sand. 10th Congress on Numerical Methods in Engineering, At Guimaraes, Portugal.

Nematzadeh, A, Shiri, H., 2019b. Self-correcting soil models for numerical simulation of strain rate dependent ice scour in sand. 9th ECCOMAS Thematic Conference on Smart Structures and Materials, Paris, France.

Nematzadeh, A, Shiri, H. 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Region Science and Technology, 170: 102929.

Nobahar, A., Kenny, S., Phillips, R., 2007. Buried pipelines subject to subgouge deformations. International Journal of Geomechanics, 7(3): 206-216.

Palmer, A.C., Konuk, I., Comfort, G., Been, K., 1990. Ice-gouging and the safety of marine pipelines. Offshore Technology Conference, OTC-6371-MS.

Paulin, M.J., 1991. Preliminary results of physical model tests of ice scour. Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering.

Paulin, M.J. 1991. Physical model analysis of iceberg scour in dry and submerged sand. Doctoral dissertation, Memorial University of Newfoundland.

Peek, R., Been, K., Bouwman, V., Nobahar, A., Sancio, R., Schalkwijk, R.V., 2013. Buried pipeline response to ice-gouging on a clay seabed large scale tests and finite element

analysis. 22nd International Conference on Port and Ocean Engineering under Arctic Conditions (POAC), pp. 9-13.

Phillips, R., Barrett, J., 2012. PIRAM: pipeline response to ice-gouging. Offshore Technology Conference, OTC-23790-MS.

Pike, K., Kenny, S., 2012. Numerical pipe/soil interaction modelling: Sensitivity study and extension to ice-gouging. Offshore Technology Conference, OTC-23731-MS.

Pike, K., Kenny, S., 2016. Offshore pipelines and ice gouge geohazards: comparative performance assessment of decoupled structural and coupled continuum models. Canadian Geotechnical Journal, 53(11): 1866-1881.

Sattar, A.M., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. Neural Computing and Applications, 31(1): 157-169.

Sancio, R., Been, K., Lopez, J., 2011. Large scale indenter test program to measure sub gouge displacements. Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions, No. POAC11-096.

Woodworth-Lynes, C., Nixon, D., Phillips, R., Palmer, A., 1996. Subgouge deformations and the security of Arctic marine pipelines. Offshore Technology Conference, OTC-8222-MS.

Yang, W., 2009. Physical modeling of subgouge deformations in sand. Doctoral dissertation, Memorial University of Newfoundland.

Section 2

Assessment of Ice-Seabed Interaction Process in Clay Using Extreme Learning Machine

This section is a journal paper which has been published in the International Journal of Offshore and Polar Engineering (2021), Vol. 31, pp. 411–420, ISSN: 1053-5381

Abstract

The buried subsea pipelines in the Arctic offshore territories are commonly threatened by drifting icebergs scouring the sea bottom (ice-gouging). In this study, the subgouge soil deformations in clay and the ice keel-seabed reaction forces were simulated using Extreme Learning Machine (ELM) algorithm. Using the effective input parameters, eight ELM models were developed to predict the seabed response to ice-gouging. The superior ELM models and the most significant input parameters were identified by performing a sensitivity analysis. The developed ELM models well predicted the ice-seabed interaction parameters with the soil depth being the most influencing input factor affecting subgouge soil deformations, and the vertical load and the attack angle having the highest level of effectiveness for simulation of the ice keel-seabed reaction forces.

Keywords: Extreme learning machine, Ice-seabed interaction, Clay seabed, Simulation, Uncertainty analysis

5.1.5. Introduction

Discovered crude oil and natural gases in the Arctic areas are respectively estimated as 3 billion bbl. and 4 trillion cubic ft. The overwhelming majority of the future development projects for these oil and gas reserves in the areas comprise the deployment of subsea pipelines, manifolds, wellheads, and communication cables (Alba 2015). However, these bottom-founded structures are at risk of damage and potential failure from moving pressure ridges and icebergs since the seabed may be gouged by these traveling masses in the ice-prone regions during warmer seasons. Figure 5-15 shows the layout of the ice-seabed interaction process during an ice-gouging phenomenon. As shown, the maximum subgouge deformation occurs just beneath the iceberg keel base (zone 1) and the magnitude of the soil displacements decreases by extending into Zone 2 and Zone 3.

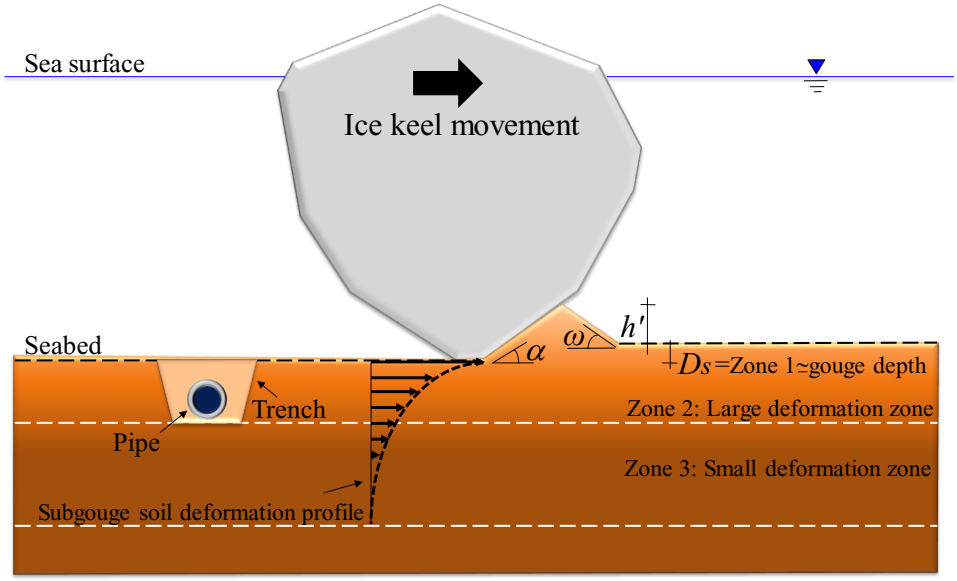


Figure 5-15. The layout of the ice-seabed interaction process during an ice-gouging phenomenon

Moreover, the gouge depth equals the maximum depth of ice intrusion. On the other hand, subsea trenching and backfilling are commonly utilized to bury these structures for physical protection

against ice scour events (Nematzadeh and Shiri 2020; Azimi and Shiri 2020a) and the ice-gouging parameters ought to be estimated to avoid the collision (Croasdale 2000).

Thus, thanks to the significant importance of this issue, so many expensive fields and laboratory studies have been performed on the ice-gouging problem (Barnes et al. 1984; Woodworth-Lynas et al. 1990; Chouinard 1995). Lach (1996) conducted nine centrifuge tests to measure the ice-induced parameters comprising the reaction forces and the subgouge displacements. The author assumed that the ice block moved horizontally on a uniform clay seabed at a constant speed. The author developed a two-dimensional finite element model to simulate the ice-scoured problem and showed that the numerical model could estimate the subgouge behavior appropriately. Woodworth-Lynes et al. (1996) performed a set of the Pressure Ridge Ice Scour Experiment (PRISE) joint-industry projects to evaluate the scoured soil deformations. The investigation proved that the pipeline burial depth should be beneath the maximum depth of ice intrusion. The authors asserted that the subgouge soil deformations were a function of the soil depth and the gouge geometry. Schoonbeek et al. (2006) reported some centrifuge experimental values of ice-induced clay displacements. Effects of the gouge depth, the undrained shear strength of clay, and the ice keel velocity on the deformed subgouge clay were assessed. The survey confirmed that the large plastic strains took place under the moving ice keel base. Been et al. (2008) investigated the impact of ice keel shape including the steep and shallow keels on the clay failure mechanism in the ice-gouging events. The authors concluded that the soil failure pattern depended upon the keel geometry and the impact of the keel length parameter should be evaluated in the future. Arnau and Ivanović (2019) studied experimentally the ice-induced loads on a sandy seabed in both saturated and dry circumstances. The effects of the relative density, soil permeability, gouge geometry, and

drifting velocities were assessed. The author concluded that the drifting velocity possessed a remarkable influence on the ice-gouging loads.

Furthermore, to provide a good understanding of the local stresses and strains in the ice-seabed interaction process, several time-consuming 2D or 3D finite element simulations have been performed in this field (Lele et al. 2011; Pike and Kenny2012; Sabodash and Bekker 2019; Nematzadeh and Shiri 2019a). Nematzadeh and Shiri (2019b) modeled an ice-gouging problem through a self-correcting soil method to update the shear strength parameter of soil. The authors showed that the ice-induced deformations were overestimated by the conventional decoupled models. Nematzadeh and Shiri (2020) simulated the impact of the non-linear stress-strain behavior of an ice-gouging event in a sandy seabed by means of a modified Mohr-Coulomb (MMC) model. The authors asserted that the side berm size was affected by the magnitude of the attack angle, while the frontal berm was not significantly altered.

It is worth noting that artificial intelligence (AI) methods and machine learning (ML) technology as precise, fast, and cost-effective tools have been recently utilized to simulate the ice-gouging problem. Kioka et al. (2003) modeled the scour depth as a function of the ice keel by using a Layered Neural-Network (NN) method. The authors showed that the NN algorithm managed to simulate the target value with good precision. Azimi and Shiri (2020a) conducted a dimensional analysis to introduce factors governing the ice-seabed interaction event. The authors suggested several linear regression (LR) models so as to calculate the maximum ice-scoured deformations. The shear strength of seabed soil and the gouge geometry were considered the most influencing parameters. Azimi and Shiri (2020b) modeled the ice-induced sand displacements through a gene expression programming-based (GEP) method. The GEP model outperformed the artificial neural network (ANN) and empirical approaches. The author concluded that the soil depth was a

significant factor to simulate the subgouge displacements. Azimi and Shiri (2021) estimated the subgouge sand features by means of a machine learning model. The author showed that the gouge depth ratio and the angle of internal friction of the seabed sand had a significant impact on the simulation results.

The extreme learning machine (ELM) algorithm was used by Azimi and Shiri (2021) to model the ice-gouging event. The current study added several novelties to improve the limitations of the earlier study. For instance, the architecture of the ELM algorithm was significantly improved in the current study with a reduced number of hidden layer neurons (20 vs. 28 in the earlier study), and a hyperbolic-tangent (Hypertan) activation function (sigmoid in the earlier study). The Hypertan function is monotonic (with a non-monotonic derivative) and has an S-shaped curve with a range of -1 to 1. This helped well negative mapping of the negative input and near-zero mapping of the zero input. The sigmoid activation function in the earlier study (Azimi and Shiri 2021) was monotonic (with a bell-shaped derivative) and had a range of 0 to 1. The sigmoid function was concave for values greater than zero, and convex for values less than zero. In addition, the earlier study conducted by Azimi and Shiri (2021) was limited to the cohesionless sand seabed, which follows an entirely different geomechanical response compared with clay. The current study investigated a cohesive clay seabed using a completely different set of experimental data (i.e., five sets of experiments in clay including C-CORE 1995; Lach 1996; C-CORE 1996; Schoonbeek et al. 2006; and Been et al. 2008). This was the first application of the ELM algorithm in the literature on modeling the subgouge soil deformation in clay. The parameters governing the subgouge soil characteristics in sand and clay are significantly different. The input parameters in the earlier study (i.e., Azimi and Shiri 2021) was including the ice keel geometry, the angle of internal friction in the sand, the berm height, the environmental loads, and the ice dynamics. However, the input

parameters in the current study included a different set of parameters comprising the gouge depth ratio (Ds/W), the clay cohesion ($c/\gamma s.W$), the attack angle (α), the horizontal and vertical loads ($Lh/\gamma s.W^3$, $Lv/\gamma s.W^3$), and the ice keel velocity ($V^2/g.W$).

Therefore, the ice-seabed interaction process in clay was simulated by using the ELM algorithm in this study for the first time.

The literature tried to review the important studies in the field of ice-gouging and provide a good insight into the parameters governing the problem. Although, owing to the high level of uncertainty of the ice-gouging issue, the industry still prefers to implement costly experimental investigations and long-running 3D simulations, the companies are looking for a reliable, cost-effective, and quick alternative to simulate the behavior of subgouge soil to alleviate the collision risk of the moving icebergs with the sea bottom founded structures (Banke and Smith 1984; Azimi and Shiri 2020b). Thus, to fill this knowledge gap, the ability of an accurate, quick, and powerful ML approach called Extreme Learning Machine (ELM) was evaluated to simulate the ice-seabed interaction features in this study. Further information is detailed in the upcoming sections.

5.1.6. Methodology

5.1.6.1. Extreme Learning Machine (ELM)

An extreme learning machine (ELM) is a powerful neural network algorithm introduced by Huang et al. (2004) to solve different linear and non-linear problems (Azimi et al. 2017). The ELM is a quick training approach possessing the structure of a single-layer feed-forward neural network (SLFFNN) (Huang et al. 2006). In comparison with the classical backpropagation (BP) algorithm, the ELM has several benefits as follows:

- The high training speed of feedforward neural network (FFNN) by ELM,

- Better generalization performance,
- Training with differentiable and non-differentiable activation functions,
- Overcome the gradient-based algorithm such as the local optimum, improper learning rate, and overfitting problem (Huang et al. 2011; Azimi et al. 2017; Sattar et al. 2019).

The ELM has three distinct layers comprising an input layer, a hidden layer, and an output layer. The input layer contains the input parameters of the ice-gouging event and the number of neurons in the output layer equals the number of target parameters of the problem. However, the number of neurons in the hidden layer is chosen based on a trial and error procedure (Azimi et al. 2017). A single hidden layer forward network is formed by means of linear algebra in these three layers so as to solve the equations for obtaining the optimum weights in the output layer (Sattar et al. 2019). It is worth noting that the weights of the input layer are determined randomly, whereas the output weights are analytically estimated through a predefined training mechanism.

The input weights and the hidden layer biases are randomly assigned in the ELM algorithm. However, different random assignments were used in a simulation of each ice-gouging event to ensure a proper representation of the results.

Tuning the value of hidden layer hyperparameters, e.g., hidden layer biases, in the ELM network was not necessary since the hidden layer output matrix remained unchanged once the arbitrary value of input weights and hidden layer biases were assigned at the beginning of the learning process. Thus, the initialization of these random numbers has no impact on the accuracy of the prediction (Huang et al. 2004).

Thus, the training procedure in the ELM model is extremely fast and its generalization capability is quite high. Commonly, the output of an SLFFNN is solved by using the below equation:

$$Y = \sum_{j=1}^{m} \beta_j G\left(\sum_{i=1}^{n} w_{i,j}\chi_i + \kappa_j\right) \tag{5-28}$$

where, Y and χ donate the output and inputs of the ELM model, respectively. Additionally, n equals the number of input parameters and m is the number of output parameters of the problem. Moreover, $w_{i,j}$ indicates input weight that is the connection between the i^{th} neuron in the input layer to the j^{th} neuron in the hidden layer, β_j is a coefficient that links the j^{th} neuron in the hidden layer to the particular neuron in the output layer. Besides, $G()$ and κ_j are respectively the activation function and biases of the hidden layer neurons.

Generally, the output of an SLFFNN is computed in two major steps. First, the network of the hidden layer, meaning that the number of neurons in this layer is determined by using a trial and error procedure (Azimi et al. 2017). Subsequently, in order to train the ELM network, an activation function (AF) is defined. Indeed, the optimum number of hidden neurons and the best activation function lead to the optimal ELM network. After that, using arbitrary assignments of the input layer weights $(w_{i,j})$ and the biases (κ_j), the structure of hidden layer neurons and the weight in the output layer are calculated. Therefore, equation (5-28) is rewritten as below:

$$H\beta = Y \tag{5-29}$$

here, H is the extreme learning machine feature mapping matrix as follows:

$$H(w_{i,j}, \kappa_j, \chi_i) = \begin{bmatrix} H_{1,1} & \cdots & H_{1,m} \\ \vdots & \ddots & \vdots \\ H_{n,1} & \cdots & H_{n,m} \end{bmatrix} = \begin{bmatrix} G(w_{1,1}\chi_1 + \kappa_1) & \cdots & G(w_{1,m}\chi_m + \kappa_m) \\ \vdots & \ddots & \vdots \\ G(w_{n,1}\chi_n + \kappa_1) & \cdots & G(w_{n,m}\chi_m + \kappa_m) \end{bmatrix} \tag{5-30}$$

In addition, β and Y are defined as below:

$$Y = [Y_1\ Y_2\ \cdots\ Y_m] \quad \text{and} \quad \beta = [\beta_1\ \beta_2\ \cdots\ \beta_n] \tag{5-31}$$

It is worth noting that β_j is obtained by minimizing error in an estimation through the Moore-Penrose generalized inverse matrix as follows (Huang et al. 2011):

$$\hat{\beta} = H^{+}Y \tag{5-32}$$

here, H^{+} represents the Moore-Penrose generalized inverse matrix of H. Figure 5-16 depicts the ELM flowchart to simulate the subgouge clay parameters in the current study.

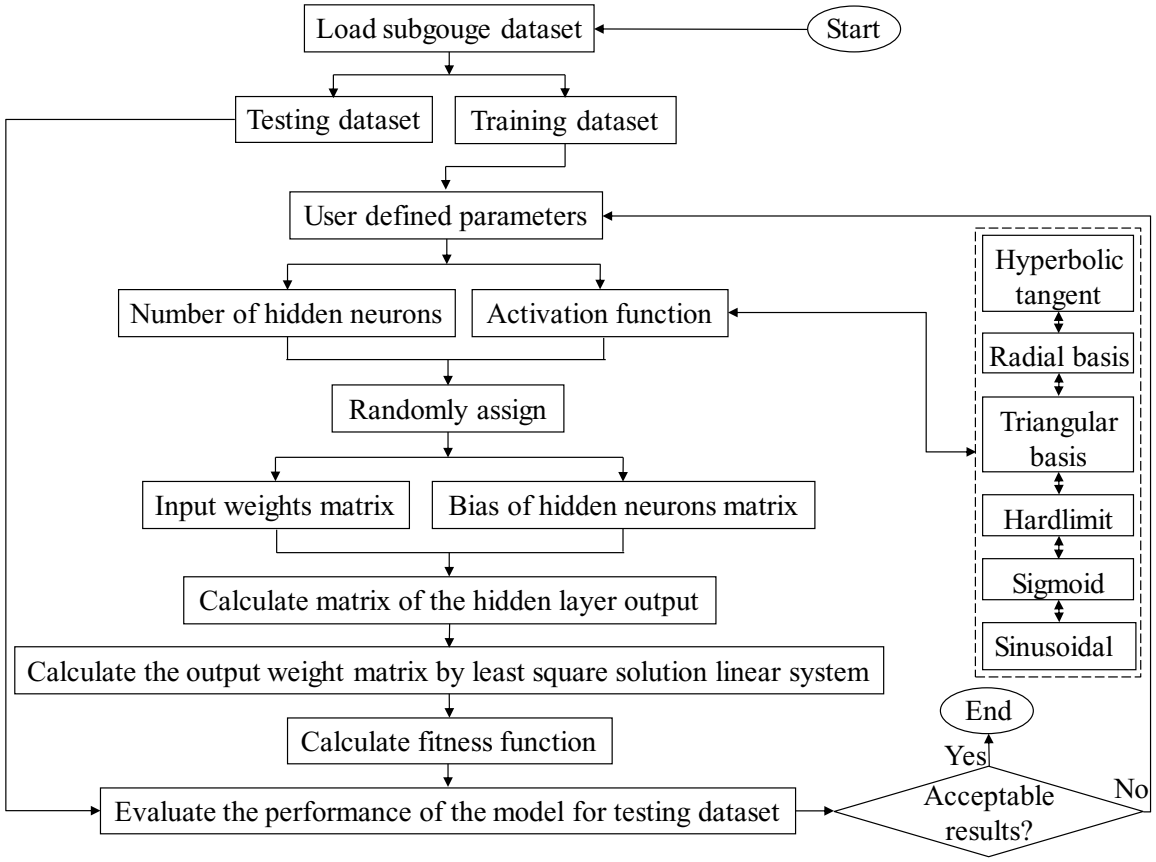


Figure 5-16. ELM flowchart to simulate the subgouge clay parameters

Initially, the constructed dataset including the training and testing sub-samples was loaded. For the training dataset, the user parameters such as the number of iterations, the number of hidden neurons, and the activation function were defined. In the next step, the input weight matrix and the bias of the hidden neuron matrix were randomly assigned which resulted in the calculation of the hidden layer output matrix. Subsequently, by using the least square solution, the output weight matrix was determined. Next, the fitness function was computed and the performance of the ELM

model was assessed for the testing dataset. Lastly, if the level of accuracy, correlation, and complexity of the ELM model was acceptable, the simulation was ended, otherwise, this procedure was iterated.

5.1.6.2. Seabed Interaction Process in Clay

Seabed interaction parameters (η) in a clay mass comprising the soil deformations (d/W) and the reaction forces ($F/\gamma_s W^3$) are a function of the scour depth (Ds), the shear strength parameter of the clay (c), the width of gouge (W), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (h'), the horizontal load (L_h) and the vertical load (L_v), the velocity of ice keel (V), and the specific weight of clay (γ_s) as follows (Lach 1996; Azimi and Shiri 2020a):

$$\eta_{(clay)} = f_1(D_s, c, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \tag{5-33}$$

Equation (5-33) can be written as a function of eight dimensionless groups as below (Azimi and Shiri 2020a):

$$\frac{\eta_{(clay)}}{W} = f_2\left(\frac{D_s}{W}, \frac{c}{\gamma_s . W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{5-34}$$

It is worth noticing that the maximum subgouge deformation in the clay $\left(d_{(clay)}\right)$ is formed just under the moving ice keel in the gouge centerline. However, at greater depth on the subgouge centerline, by incorporating the soil depth (y/W), the subgouge clay displacement ($d_{(clay)}/W$) is written as follows:

$$\frac{d_{(clay)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s . W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{5-35}$$

Or

$$\frac{d_{(h)}}{W} = f_4\left(\frac{y}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s . W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{5-36}$$

$$\frac{d_{(v)}}{W} = f_5\left(\frac{y}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}37)$$

Similarly, the ice-induced reaction force (F) is a function of the position of the iceberg along the scour axis (x) and $D_s, c, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s$, so equation (5-37) can be written as below:

$$\frac{F_{(clay)}}{\gamma_s W^3} = f_6\left(\frac{x}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}35)$$

Or

$$\frac{F_{(h)}}{\gamma_s W^3} = f_7\left(\frac{x}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}36)$$

$$\frac{F_{(y)}}{\gamma_s W^3} = f_8\left(\frac{x}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}37)$$

In the current study, different laboratory investigations were applied so as to simulate the ice-gouging clay parameters (d/W and $F/\gamma_s W^3$). In other words, the key measurements of five experimental investigations reported by C-CORE (1995), Lach (1996), C-CORE (1996), Schoonbeek et al. (2006), and Been et al. (2008) were utilized to verify the ELM models. It should be noted that the angle of the surcharged soil slope (ω) and the height of the berm (h') were not reported by all applied laboratory studies. Therefore, other introduced dimensionless groups in equations (5-36) and (5-37) were utilized as the input parameters for the ELM model to simulate the horizontal subgouge deformation (d_h/W), the vertical subgouge deformation (d_v/W), the horizontal reaction force ($F_h/\gamma_s W^3$), and the vertical reaction force ($F_v/\gamma_s W^3$). Figure 5-17 portrays the defined input combinations in the current study.

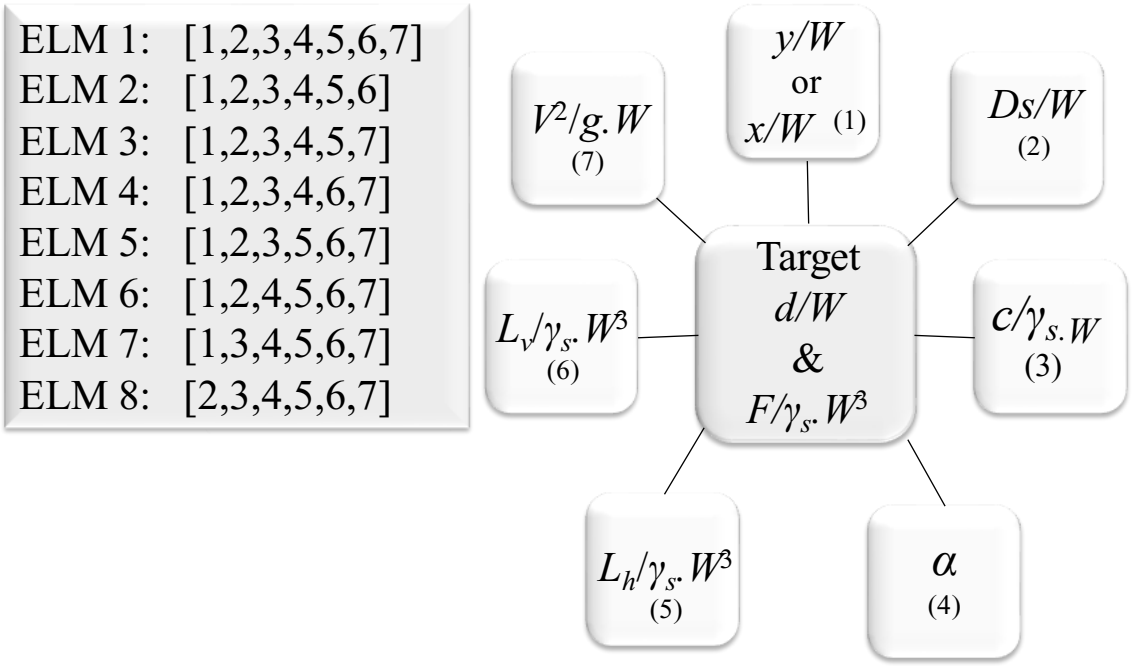


Figure 5-17. Used input combinations to develop ELM models

The ELM algorithm was fed with x/W, Ds/W, $c/\gamma s.W$, α, $Lh/\gamma s.W^3$, $Lh/\gamma s.W^3$, and $V^2/g.W$ as input parameters to simulate the d_v/W and d_h/W as output. Moreover, the y/W, Ds/W, $c/\gamma s.W$, α, $Lh/\gamma s.W^3$, $Lh/\gamma s.W^3$, and $V^2/g.W$ were regarded as inputs of the ELM model to predict the $F_h/\gamma_s W^3$ and $F_v/\gamma_s W^3$ as output.

Eight ELM models (ELM 1 to ELM 8) were developed to simulate the ice-induced clay parameters. This means that ELM 1 was defined by using all input parameters and then the effect of each input was eliminated (ELM 2 to ELM 8) to identify the most significant input parameters. It is noteworthy that 70% of the observed values were applied to train the ELM models, whereas 30% of the remaining were utilized to test them.

5.1.6.3. Goodness of Fit

To examine the performance of the ELM models, a set of statistical indices including correlation coefficient (*R*), the variance accounted for (*VAF*), root mean square error (*RMSE*), mean absolute error (*MAE*), Nash-Sutcliff efficiency coefficient (*NSC*), and Akaike Information Criteria (*AIC*) were applied as below:

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{5-38}$$

$$VAF = \left(1 - \frac{var(P_i - O_i)}{var(P_i)}\right) \times 100 \tag{5-39}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{5-40}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{5-41}$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{5-42}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{5-43}$$

here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, and *n* are the observed ice-induced seabed parameters, the simulated values, the average observed values, the average simulated values, and the number of experimental measurements, respectively. Moreover, the *k* is the number of hidden layer neurons used in the ELM network. The closeness of the correlation coefficient (*R*) and the Nash-Sutcliff efficiency coefficient (*NSC*) to one signified that the ELM model owned the highest correlation with the observed values, whilst the closeness of the *RMSE* and *MAE* criteria to zero meant that the particular model had the lowest level of error. Furthermore, the superior ELM model possessed the highest value of the *VAF* criterion though the complexity of these ELM models was not shown

through the applied indices. To overcome this drawback, the Akaike Information Criteria (*AIC*) was employed. The less complex ELM model had the lowest value of the *AIC* index, hence, the superior ELM model owned the lowest value of the *AIC* index and error (*RMSE* and *MAE*) and the highest level of correlation (*R* and *NSC*) with the experimental measurements.

5.1.7. Results and discussion

In the upcoming sections, the optimal number of hidden neurons (NHN) and the best activation function (AF) were chosen. The ELM 1 model was used to select the optimized number of hidden layer neurons (NHN) and the best activation function (AF).

5.1.7.1. Number of Hidden Neurons (NHN)

Determination of the optimal NHN can play a significant role in the precision of the ELM structure, meaning that by increasing the NHN, the generalization ability of the ELM model is improved (Feng et al. 2009). The NHN for the ELM network was initially chosen as 2 and then was increased to 22. The ELM network could simulate the ice-gouging parameters (d_h/W, d_v/W, $F_h/\gamma s.W^3$, $F_v/\gamma s.W^3$) with its highest performance once the NHN was adopted as 20. Figure 5-18 portrays the variations of NHN versus the calculated statistical indices for the simulation of ice-induced parameters (e.g., the horizontal reaction forces).

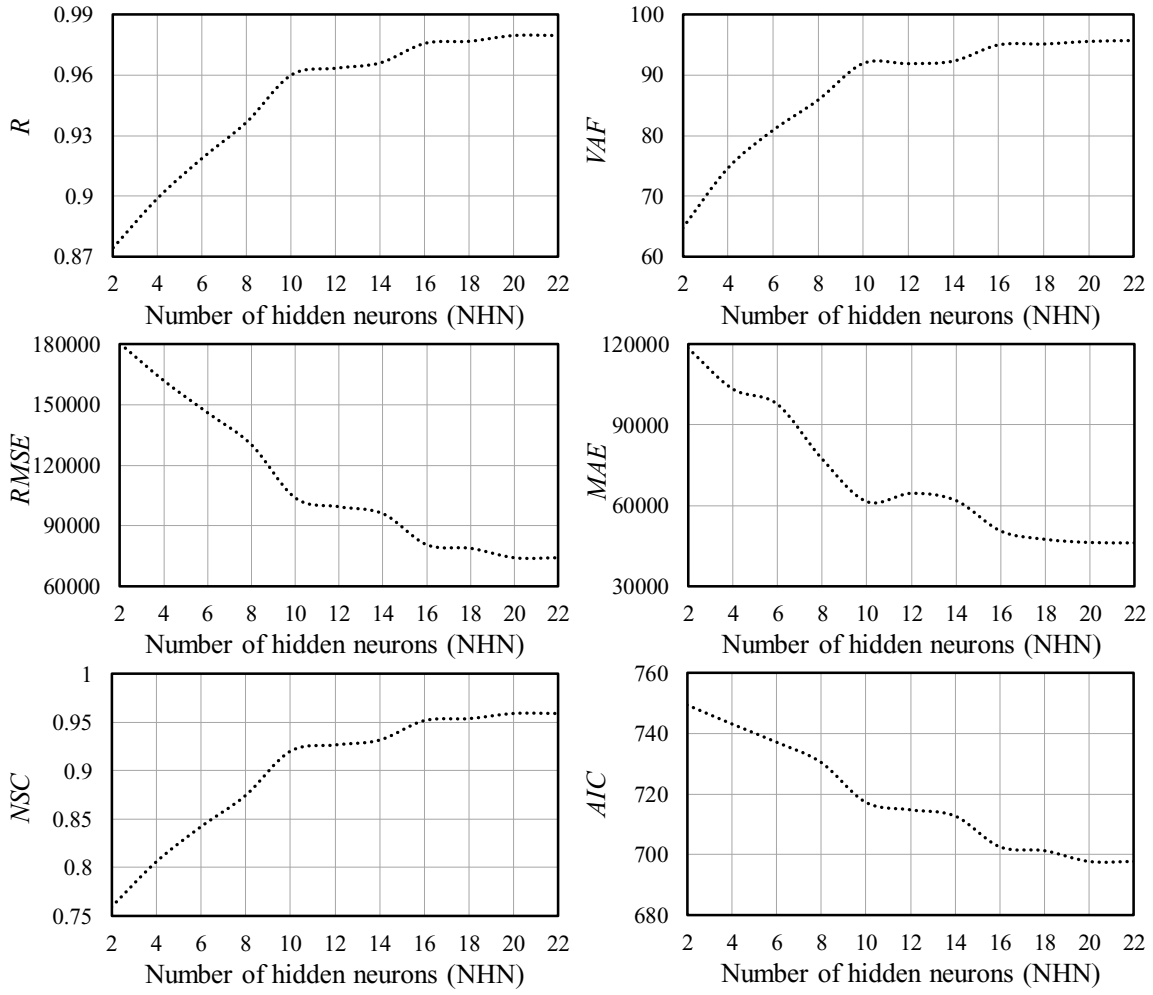


Figure 5-18. Variations of the number of hidden neurons (NHN) versus the used statistical indices for simulation of horizontal reaction forces ($F_h/\gamma_s W^3$)

As shown, the NHN for the ELM network was initially chosen as two, with a low level of accuracy and correlation (*MAE*=118770.564 and *R*=0.874). However, the *R*, *MAE*, and *NSC* values for the ELM network were respectively estimated to be 0.980, 46193.161, and 0.959 when the NHN was considered at 20. Additionally, once the NHN was equal to 20, the value of the *NSC* criterion for the ELM algorithm to predict the d_h/W, d_v/W, and $F_v/\gamma_s.W^3$ was surmised at 0.999, 0.841, and 0.993, respectively. Regarding the simulation results, the performance of the ELM model was

insignificantly changed when the NHN was greater than 20. Thus, the optimal number of NHN for the ELM network was selected to be 20 in the current study.

5.1.7.2. Activation Function (AF)

Before the evaluation of the activation functions in the ELM network, an optimal NHN of 20 was set for this algorithm. Generally, the weighted input parameters and biases in the ELM structure are nonlinearly transformed through an activation function (AF) (Hertz 2018). In the current study, six different AFs comprising the hyperbolic tangent (Hypertan), the triangular basis (Tribas), the radial basis (Radbas), the hard limit (Hardlim), the sigmoid (Sig), and the sinusoidal (Sin) were assessed to simulate the ice-scoured seabed interaction features. Figure 5-19 shows the calculated statistical indices for the applied activation functions to predict the subgouge parameters (e.g., the horizontal reaction forces). The Hypertan activation function with the *NSC*, *AIC*, and *R* of 0.959, 697.554, and 0.980, resulted in the highest correlation and accuracy along with the lowest level of complexity among the applied AFs. The variance accounted for (VAF) and the correlation coefficient (*R*) values in the Tribas function for modeling the horizontal reaction forces were 84.515 and 0.931. The calculated values of *NSC*, *MAE*, and *AIC* for the Radbas function were respectively surmised as 0.938, 59343.908, and 709.939. The Hardlim was detected as the worst activation function in the ELM network so as to simulate the ice-seabed interaction process, with the *R* and *AIC* values equal to 0.899 and 743.350. For the simulation of the subgouge clay parameters by using the Sig activation function, the values of *AIC*, *NSC*, and *R* criteria were respectively estimated at 699.367, 0.957, and 0.978. The magnitude of computed *NSC*, *RMSE*, and *VAF* values for the Sin function was calculated as 0.829, 151877.675, and 80.276, respectively.

Moreover, the value of the correlation coefficient (R) for the simulation of vertical reaction forces, horizontal deformations, and vertical deformations by using the Hypertan function was estimated as 0.997, 0.999, and 0.935, respectively. Therefore, regarding the simulation results for various activation functions in the ELM structure, the hyperbolic tangent was chosen as the premium activation function to model the ice-seabed interaction event in the current investigation.

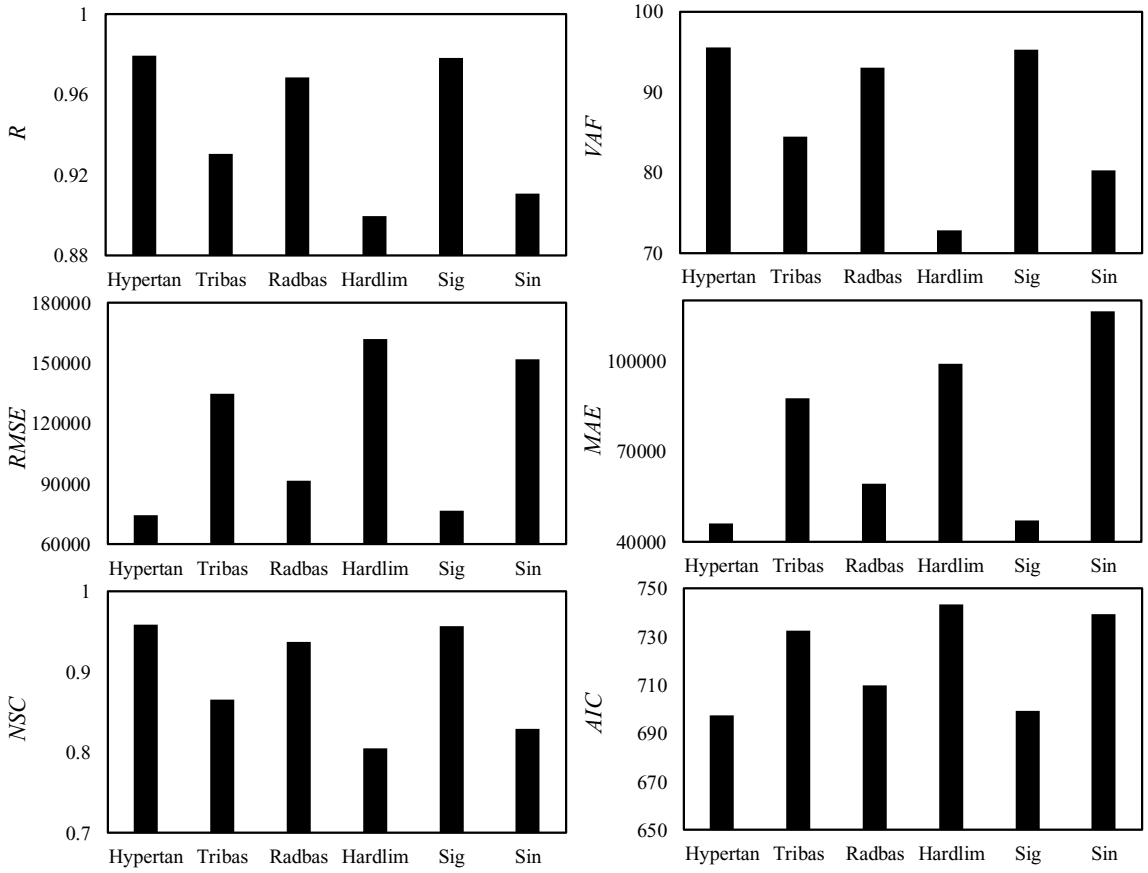


Figure 5-19. Calculated statistical indices for various activation functions to estimate the horizontal reaction forces

Hence, the performance of the ELM algorithm to simulate the subgouge characteristics was evaluated for the different number of neurons in the hidden layer (NHN) and various activation

functions (AFs), meaning that the selected NHN and AF were the optimized values to predict the reaction forces and subgouge deformations.

5.1.7.3. Sensitivity Analysis

In the current study, eight ELM models (ELM 1 to ELM 8) were defined to simulate the horizontal reaction forces ($Fh/\gamma_s W^3$), the vertical reaction forces ($Fv/\gamma_s W^3$), the horizontal subgouge deformations (dh/W), and the vertical subgouge deformations (dv/W) in the clay seabed. Table 5-3 depicts the calculated statistical indices for the horizontal reaction forces simulated by these ELM models in both training and testing modes. ELM 1 simulated the $Fh/\gamma_s W^3$ by using all input factors including x/w, D_s/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW. Subsequently, these input parameters that were removed one at a time in ELM 2 to ELM 8 were respectively V^2/gW, $L_v/\gamma_s.W^3$, $L_h/\gamma_s.W^3$, α, $c/\gamma_s.W$, D_s/W, and x/w. According to the simulation results, the values of variance accounted for (*VAF*), Akaike Information Criteria (*AIC*), and correlation coefficient (*R*) for ELM 1 in the testing phase were respectively reckoned to be 95.594, 697.554, and 0.980. ELM 2 predicted the Fh/$\gamma_s W^3$ parameter in terms of x/w, D_s/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, with a Nash-Sutcliff efficiency coefficient (*NSC*) of 0.955 in the testing mode. To predict the horizontal reaction forces by using ELM 3 in the testing mode, the *RMSE*, *NSC*, and *AIC* criteria were respectively obtained as 87624.302, 0.943, and 707.254. ELM 3 estimated the target function in terms of x/w, D_s/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, V^2/gW. The $L_h/\gamma_s.W^3$ was an eliminated input factor for the ELM 4, signifying that this model prognosticated the horizontal reaction forces using x/w, D_s/W, $c/\gamma_s.W$, α, $L_v/\gamma_s.W^3$, V^2/gW parameters, with the computed *R* and *AIC* values of 0.982 and 694.015. The VAF, R, and AIC values for ELM 5 in the testing mode were 96.618,

0.985, and 689.900, respectively. The ELM 5 was a function of x/w, D_s/W, $c/\gamma_s.W$, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW, meaning that the attack angle (α) was a deleted factor for this model. ELM 6 simulated the horizontal reaction forces adopting the x/w, D_s/W, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW variables, signifying that the effect of the shear strength of clay ($c/\gamma_s.W$) was ignored. In the testing mode of ELM 6, the values of MAE, NSC, and VAF indices were respectively computed at 50942.981, 0.951, and 94.759. The AIC and R criteria for ELM 7 were equal to 695.920, and 0.981, with the RMSE index of 72221.886. In ELM 7, the key input parameters included x/w, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW.

The x/w input was the eliminated factor in ELM 8, where this model comprised the input parameters D_s/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and V^2/gW. For ELM 8 in the testing situation, the values of *RMSE*, *AIC*, and *MAE* were equal to 227442.199, 763.178, and 125394.709. Thus, ELM 5 possessed the highest level of precision and correlation with the laboratory measurements in order to estimate the horizontal reaction forces, with the lowest level of complexity. After ELM 5, other models including ELM 4, ELM 7, ELM 1, ELM 2, ELM 6, and ELM 3 were ranked as the second-best to seventh-best ELM models to simulate the $Fh/\gamma_s W^3$. It is worth noticing that the ELM 8 was detected as the worst model.

On the other hand, regarding the performed sensitivity analysis, the x/w parameter was selected as the most influencing input factor to simulate the horizontal reaction forces. Furthermore, the $L_v/\gamma_s.W^3$, $c/\gamma_s.W$, V^2/gW, D_s/W, $L_h/\gamma_s.W^3$ inputs were identified as the second-important to sixth-important variables to predict the horizontal reaction forces. The attack angle (α) had an insignificant impact on modeling the horizontal reaction forces.

Table 5-3. Calculated statistical indices for horizontal reaction forces simulated by ELM models in training and testing modes

Model	Mode	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
ELM 1	Train	0.98	95.8	72505.47	43385.9	0.96	1571.
	Test	0.98	95.5	74262.76	46193.1	0.95	697.5
ELM 2	Train	0.98	97.6	54848.88	37968.5	0.97	1532.
	Test	0.97	95.0	78129.78	50640.9	0.95	700.5
ELM 3	Train	0.98	95.9	71938.39	44841.8	0.96	1569.
	Test	0.97	93.6	87624.30	57080.9	0.94	707.2
ELM 4	Train	0.98	97.0	61803.44	35025.6	0.97	1549.
	Test	0.98	96.1	69912.11	40528.3	0.96	694.0
ELM 5	Train	0.98	97.2	59946.94	36168.0	0.97	1544.
	Test	0.98	96.6	65173.66	37288.7	0.96	689.9
ELM 6	Train	0.98	97.1	60282.87	38490.5	0.97	1545.
	Test	0.97	94.7	81221.26	50942.9	0.95	702.8
ELM 7	Train	0.98	96.9	63046.37	36952.2	0.97	1551.
	Test	0.98	95.8	72221.88	42691.5	0.96	695.9
ELM 8	Train	0.87	68.2	179088.3	101250.	0.75	1694.
	Test	0.79	53.36	227442.1	125394.7	0.617	763.17

Table 5-4 illustrates the computed statistical indices for vertical reaction forces simulated by ELM models in training and testing modes. Among the ELM models, ELM 3 owned the highest level of precision and correlation with the observed values, with the MAE and R of 271.175 and 0.997.

The values of *AIC*, *VAF*, and *R* for ELM 1 in the testing stage were respectively approximated to be 252.291, 98.507, and 0.994. Moreover, the magnitude of *RMSE* and *NSC* for ELM 2 in the testing phase were calculated as 434.985 and 0.989, with a correlation coefficient of 0.996. In the testing mode of ELM 4, the variance accounted for (*VAF*), correlation coefficient (*R*), and Akaike Information Criteria (*AIC*) values were 98.811, 0.994, and 253.065, respectively.

The ELM 5 was the seventh-best model to estimate the $Fv/\gamma_s W^3$ parameter, where the RMSE and AIC values for this model in the testing situation were estimated as 573.596 and 257.930. For ELM 6 during the testing mode, the *R*, *AIC*, and *VAF* statistical indices were respectively surmised to be 0.994, 255.574, and 98.520. In addition, the values of *MAE* and *NSC* indices for the ELM 7 in the testing situation were estimated as 359.620 and 0.987, with a calculated *VAF* criterion of 98.797. ELM 8 with an *R*, *RMSE*, and *VAF* criteria of 0.967, 1389.374, and 81.532, resulting in the worst ELM model to predict the vertical reaction forces. After the ELM 3, ELM 2, ELM 7, ELM 1, ELM 4, ELM 6, and ELM 5 were identified as the third-best to seventh-best ELM models to simulate the $Fv/\gamma_s W^3$ parameter. According to the performed sensitivity analysis, the position of the iceberg along the scour axis (*x/W*), the attack angle (α), and the shear strength of clay seabed ($c/\gamma_s.W$) were considered the most significant input factors. Moreover, the $L_h/\gamma_s.W^3$, D_s/W, V^2/gW, and $L_v/\gamma_s.W^3$ variables were prioritized as the fourth-important to seventh-important input parameters.

Table 5-4. Calculated statistical indices for vertical reaction forces simulated by ELM models in training and testing modes

Model	Mode	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
ELM 1	Train	0.993	98.56	535.2	405.4	0.986	536.5

	Test	0.994	98.50	486.6	362.8	0.987	252.2
ELM 2	Train	0.993	98.59	528.6	399.3	0.986	535.6
	Test	0.996	98.91	434.9	330.7	0.989	248.4
ELM 3	Train	0.997	99.49	320.1	244.5	0.995	495.9
	Test	0.997	99.31	343.5	271.1	0.993	240.3
ELM 4	Train	0.992	98.41	561.0	428.6	0.984	540.3
	Test	0.994	98.81	497.7	371.6	0.986	253.0
ELM 5	Train	0.991	98.27	585.6	476.5	0.983	543.7
	Test	0.993	98.18	573.5	469.1	0.982	257.9
ELM 6	Train	0.990	97.87	648.4	467.0	0.979	551.7
	Test	0.994	98.52	535.5	421.0	0.984	255.5
ELM 7	Train	0.995	99.05	435.7	354.2	0.991	520.3
	Test	0.994	98.79	465.2	359.6	0.987	250.7
ELM 8	Train	0.966	92.83	1161.	800.8	0.933	597.8
	Test	0.967	81.53	1389.	923.8	0.893	288.2

The calculated statistical indices for horizontal subgouge deformations predicted by the ELM models in both training and testing phases are shown in Table 5-5. ELM 6 was the superior model to simulate the *dh/W* parameter, where the values of the correlation coefficient and Akaike Information Criteria in the testing mode of this model were equal to 0.999 and -28.275. Besides, ELM 4 was the second-best model, with RMSE and MAE values of 0.044 and 0.034. The *VAF* and NSC indices for the third-best model (ELM 5) in the testing mode were computed to be 99.939 and 0.999. The magnitude of *AIC*, *RMSE*, and *MAE* criteria for ELM 2 as the fourth-premium

model were respectively reckoned at -24.626, 0.048, and 0.036. For ELM 1 (the fifth-best ELM model), the *NSC* and *VAF* values were approximated to be 0.999 and 99.940. The computed *AIC* index for the sixth-best model (ELM 3) and the seventh-best model (ELM 7) in the testing phase was respectively calculated as -23.284 and -22.417. ELM 8 with the highest level of complexity and the lowest level of accuracy was the worst ELM model to estimate the horizontal deformations, where the values of *AIC* and *RMSE* equaled 26.430 and 0.528.

The sensitivity analysis showed that the soil depth parameter (*y/W*), the gouge dimension factor (*Ds*/*W*), and the vertical load factor $(L_v/\gamma_s.W^3)$ were respectively detected as the most influencing input parameters to predict the horizontal subgouge deformations. Furthermore, other variables including the velocity factor (V^2/gW), attack angle (α), horizontal load factor $(L_h/\gamma_s.W^3)$, and the clay shear strength $(c/\gamma_s.W)$ were respectively ranked as the fourth-significant to seventh-significant input parameters to model the horizontal subgouge displacements.

Table 5-5. Calculated statistical indices for horizontal deformations simulated by ELM models in training and testing modes

Model	Mode	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
ELM 1	Train	0.991	98.11	0.360	0.102	0.981	-10.12
	Test	0.999	99.94	0.049	0.038	0.999	-24.37
ELM 2	Train	0.999	99.81	0.113	0.052	0.998	-67.21
	Test	0.999	99.93	0.048	0.036	0.999	-24.62
ELM 3	Train	0.998	99.69	0.146	0.063	0.997	-54.33
	Test	0.999	99.92	0.051	0.043	0.999	-23.28

ELM 4	Train	0.999	99.89	0.085	0.053	0.999	-81.12
	Test	0.999	99.95	0.044	0.034	0.999	-26.62
ELM 5	Train	0.987	97.36	0.424	0.101	0.974	-2.057
	Test	0.999	99.93	0.047	0.037	0.999	-24.90
ELM 6	Train	0.999	99.90	0.084	0.047	0.999	-81.70
	Test	0.999	99.95	0.040	0.032	0.999	-28.27
ELM 7	Train	0.982	96.37	0.495	0.139	0.965	5.483
	Test	0.999	99.92	0.053	0.041	0.999	-22.41
ELM 8	Train	0.800	43.84	1.587	0.361	0.640	62.67
	Test	0.999	84.63	0.528	0.116	0.919	26.43

Table 5-6 portrays the computed statistical indices for the vertical deformations (dv/W) simulated by the ELM models in training and testing modes. The highest level of precision and correlation along with the lowest level of complexity were obtained for ELM 3, with the calculated *MAE*, *R*, and *AIC* of 0.010, 0.935, and -25.294 in testing mode. ELM 7 was the second-best model to simulate the vertical displacements, where the values of *RMSE* and *AIC* for this model were approximated to be 0.015 and -24.281. The *VAF*, *NSC*, and *R* indices for ELM 6 as the third-best model were respectively equal to 78.669, 0.812, and 0.915. In the testing mode of the fourth-best model (ELM 1), the *MAE*, *AIC*, and *RMSE* values were respectively computed as 0.012, -22.934, and 0.016. ELM 2 was the fifth-best ELM model, with a correlation coefficient of 0.914. The estimated *MAE* and *AIC* values for ELM 4 as the sixth-premium model equaled 0.013 and -22.580. ELM 5 was known as the seventh-best model, with the calculated *RMSE* and *VAF* of 0.017 and

77.201. ELM 8 with the *R*, *MAE*, and *AIC* of 0.485, 0.021, and -12.787, leading to the lowest level of precision and the highest level of complexity among all ELM models.

The applied sensitivity analysis proved that the soil depth variable (y/w) was identified as the most influencing input parameter to simulate the vertical deformations. Besides, the α, $L_h/\gamma_s.W^3$, V^2/gW, $c/\gamma_s.W$, D_s/W, and $L_v/\gamma_s.W^3$ were ranked as the second-important to seventh-important input factors.

Table 5-6. Calculated statistical indices for vertical horizontal deformations simulated by ELM models in training and testing modes

Model	Mode	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
ELM 1	Train	0.843	59.24	0.020	0.013	0.710	-95.82
	Test	0.917	81.63	0.016	0.012	0.783	-22.93
ELM 2	Train	0.877	70.09	0.018	0.013	0.770	-99.80
	Test	0.914	78.20	0.016	0.013	0.777	-22.72
ELM 3	Train	0.934	85.31	0.013	0.010	0.872	-109.9
	Test	0.935	84.25	0.014	0.010	0.841	-25.29
ELM 4	Train	0.933	85.35	0.013	0.009	0.872	-110.0
	Test	0.899	74.33	0.016	0.013	0.772	-22.58
ELM 5	Train	0.890	73.84	0.017	0.012	0.793	-101.6
	Test	0.900	77.20	0.017	0.013	0.764	-22.29
ELM 6	Train	0.924	82.78	0.014	0.011	0.853	-107.6
	Test	0.915	78.66	0.015	0.011	0.812	-24.09
ELM 7	Train	0.953	89.84	0.011	0.008	0.908	-115.7

	Test	0.946	89.38	0.015	0.011	0.818	-24.28
ELM 8	Train	0.404	-413.3	0.034	0.024	0.163	-77.38
	Test	0.485	-236.7	0.031	0.021	0.175	-12.78

5.1.7.4. Superior ELM Models

Regarding the performed sensitivity analysis, ELM 5, ELM 3, ELM 6, and ELM 3 were known as the premium models so as to simulate the horizontal reaction forces ($Fh/\gamma_s W^3$), the vertical reaction forces ($Fv/\gamma_s W^3$), the horizontal deformations (dh/W), and the vertical deformations (dv/W), respectively. Almost all horizontal and vertical reaction forces simulated by the superior models possessed an error of less than 10%. Even though approximately one-fourth of the horizontal deformations predicted by ELM 6 owned an error of between 5% and 10%, about 73% of these results had an error of less than 5%. Additionally, the overwhelming majority of vertical deformations predicted by using ELM 3 had an error of less than 5%.

The accuracy of laboratory measurements has originated from two main sources comprising systematic error (equipment error) and random error (human error). The source of the systematic error can be detected and the value of this error is measurable, whereas identification of a random error source is quite laborious. To prevent the random error, the experimental measurements had been repeated and their average values were reported as the laboratory records. It seems that the experimental values had some noises but the ELM algorithm could properly simulate the subgouge parameters. The discrepancy ratio (DR) of the superior ELM models was computed as follows:

$$DR = \frac{\eta_{(Simulated)}}{\eta_{(Observed)}} \tag{5-44}$$

where, $\eta_{(Simulated)}$ and $\eta_{(Observed)}$ are respectively the simulated subgouge clay parameters and the observed subgouge clay parameters. The magnitude of the *DR* index approaching the unity signified the higher performance of the ELM model (Azimi and Shiri 2020b). Figure 5-20 depicts the changes in discrepancy ratio (*DR*) versus the subgouge clay parameters for the superior ELM models. The average discrepancy ratio (DR_{ave}) for the horizontal and vertical reaction forces simulated by using the ELM 5 and ELM 3 models were respectively estimated at 1.367 and 1.004. Furthermore, for the *dh/W* and *dv/W* predicted through the superior ELM models, the DR_{ave} was equal to 2.481 and 1.578, respectively. As shown, the premium ELM models owned a reasonable performance in terms of the discrepancy ratio for the simulation of the subgouge clay parameters.

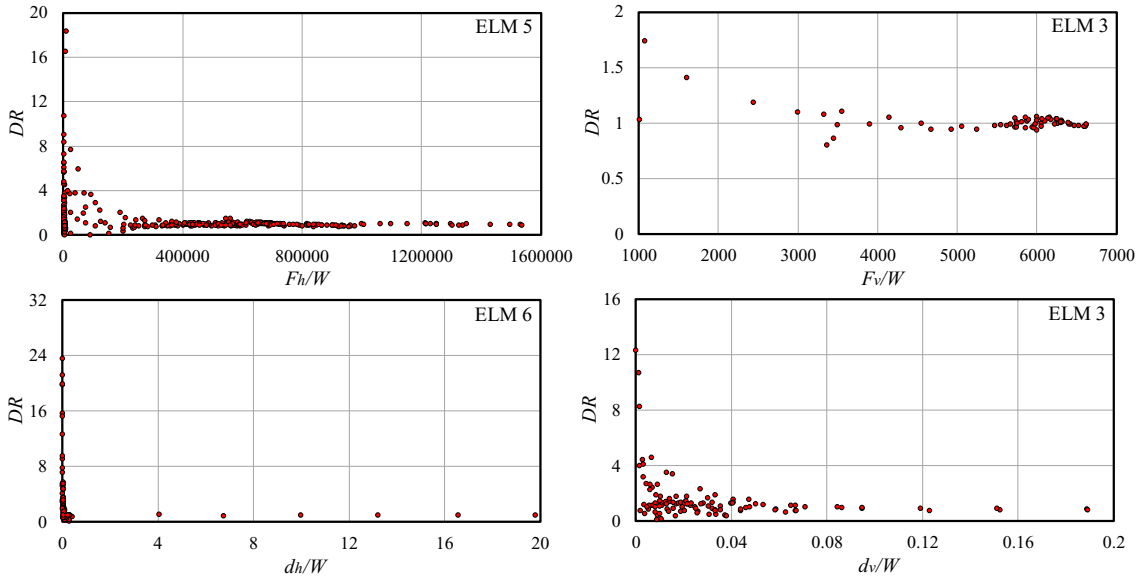


Figure 5-20. Changes of discrepancy ratio (*DR*) versus the subgouge clay parameters for the superior ELM models

To assess the performance of the superior ELM models, an uncertainty analysis (UA) was implemented. The results of the UA for the premium ELM models are arranged in Table 5-7. The

errors computed by the ELM model (e_j) was calculated as the difference between the simulated ice-gouging parameters (P_j) and the laboratory values (O_j) as follows:

$$e_j = P_j - O_j \tag{5-45}$$

In addition, the mean value of the estimated error $(\bar{e})$ and the standard deviation (S_e) were obtained by using the following equations (Azimi and Shiri 2020a):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{5-46}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{5-47}$$

If the mean value of the predicted error $(\bar{e})$ had a negative sign, the ELM model had an underestimated performance, meaning that the simulated ice-induced parameter was less than the real values. However, the positive sign of $\bar{e}$ signifies that the ELM model overestimated the subgouge parameters. It means that the simulated values were greater than the actual ones.

Subsequently, using the $\bar{e}$ and S_e, confidence bound was assumed around the computed error through the "Wilson score method" without the continuity correction (Azimi and Shiri 2020a).

It is worth mentioning that the Wilson score interval is an improvement over the normal distribution interval, where an asymmetric normal distribution is employed to improve the confidence interval bound. After that, a ±1.96Se led to a 95% confidence interval (95%CI) and then the width of uncertainty bound (WUB) was calculated. Regarding the performed UA, ELM 5 overestimated the horizontal reaction forces, whilst ELM 6 had an underestimated performance to model the horizontal subgouge deformations. Furthermore, ELM 3 simulated the vertical reaction forces with an overestimated performance, however, this model underestimated the vertical subgouge deformations. As shown in Table 9-5, the WUB for the ELM models to simulate

the $Fh/\gamma_s W^3$, $Fv/\gamma_s W^3$, dh/W, and dv/W parameters were respectively estimated as ±5531, ±39.950, ±0.012, and ±0.002. Additionally, in order to estimate the subgouge parameters including the $Fh/\gamma_s W^3$, $Fv/\gamma_s W^3$, dh/W, and dv/W, the 95%CI for the superior ELM models were calculated to be -3402 to 7660, -29 to 50.900, -0.014 to 0.010, and -0.004 to 0.0002, respectively.

Table 5-7. Uncertainty analysis results of the superior ELM models to simulate the subgouge parameters

Model	$\bar{e}$	*Se*	WUB	95% CI
ELM 5	2129	59366	±5531	-3402 to 7660
ELM 3	10.900	327.900	±39.950	-29 to 50.900
ELM 6	-0.002	0.073	±0.012	-0.014 to 0.010
ELM 3	-0.002	0.012	±0.002	-0.004 to 0.0002

A detailed comparison was made between the performance of the ELM and the multilayer perceptron neural networks (MLPNN), as a classical feedforward artificial neural network algorithm. Table 5-8 shows the results of a comparison between the MLPNN and the best ELM models in the simulation of the subgouge soil response in clay. It was observed that the training procedure of the ELM algorithm was faster than the MLPNN model in the same simulation circumstances, with the training time of ELM and MLPNN at 13.4 and 27.4 seconds, respectively. Also, the ELM showed a higher precision and correlation along with lower complexity. For instance, the value of MAE, R, and AIC indices for the MLPNN model to simulate the vertical deformations was calculated as 0.019, 0.823, and 226.508, respectively.

Table 5-8. Comparison between MLPNN and ELM models for simulation of subgouge clay parameters

Model	Parameter	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
	$Fh/\gamma s.W^3$	0.77	60.6	299490	70801	0.35	779.3
	$Fv/\gamma s.W^3$	0.98	97.69	691.34	451.93	0.99	264.3
MLPNN	*dh/W*	0.98	27.67	1.42	0.272	0.77	43.59
	dv/W	0.82	39.09	0.026	0.019	0.07	-226.5
ELM 5	$Fh/\gamma s.W^3$	0.98	96.61	65173	37288.	0.96	689.9
ELM 3	$Fv/\gamma s.W^3$	0.99	99.31	343.53	271.2	0.99	240.3
ELM 6	*dh/W*	0.99	99.95	0.04	0.032	0.99	-28.3
ELM 3	*dv/W*	0.93	84.2	0.014	0.01	0.84	-25.3

As shown, the performed analyses for the ELM models demonstrated that this methodology could appropriately simulate the subgouge clay parameters. In other words, the premium ELM models had reasonable accuracy and acceptable performance. Figure 5-21 illustrates the profiles of horizontal reaction forces ($Fh/\gamma_s W^3$) simulated by ELM 5. This model in comparison with other ELM models possessed a high correlation, good precision along with a low level of complexity and an underestimated performance to prognosticate the horizontal reaction forces.

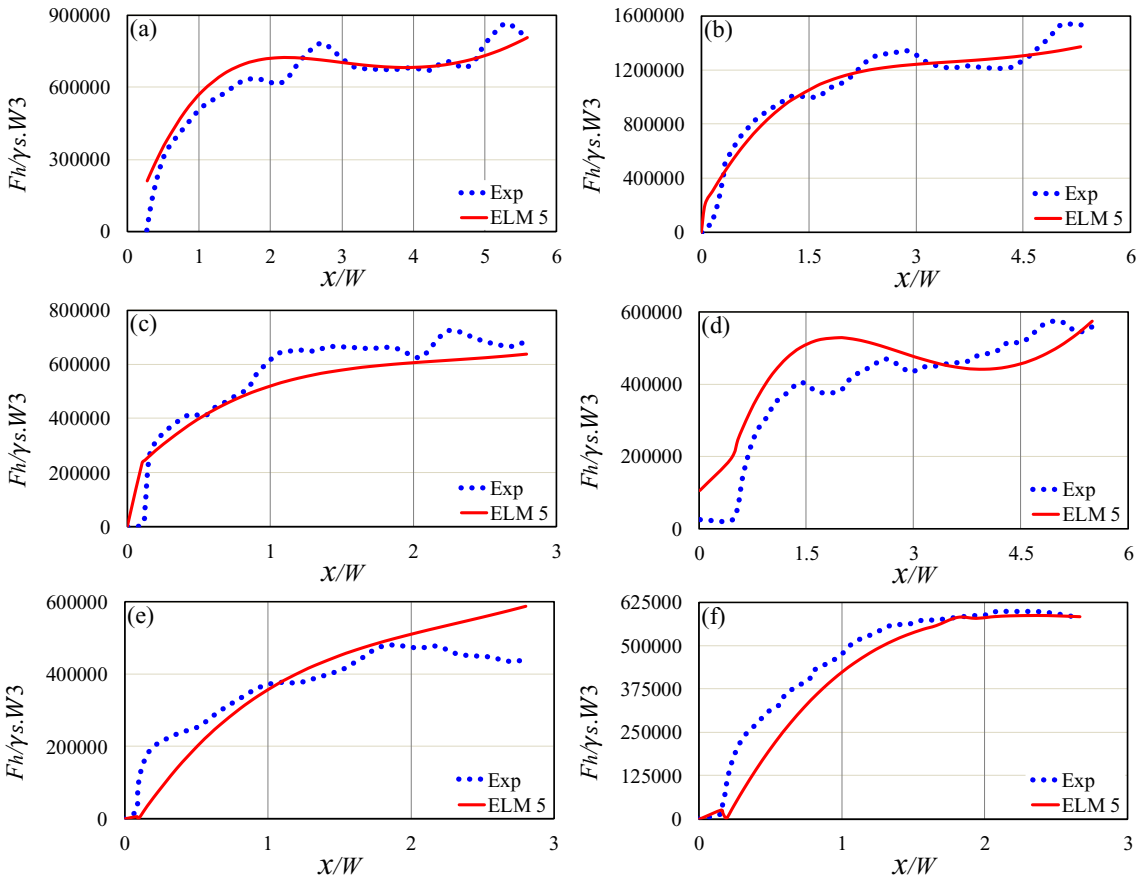


Figure 5-21. Horizontal reaction force ($Fh/\gamma_s W^3$) profiles simulated by ELM 5

ELM 3 was identified as the best model to predict the vertical reaction forces. The profiles of vertical reaction forces simulated by this model are depicted in Figure 5-22. According to simulation results, ELM 3 managed to estimate the vertical reaction forces meticulously, with a high index of precision and correlation as well as a low level of complicacy.

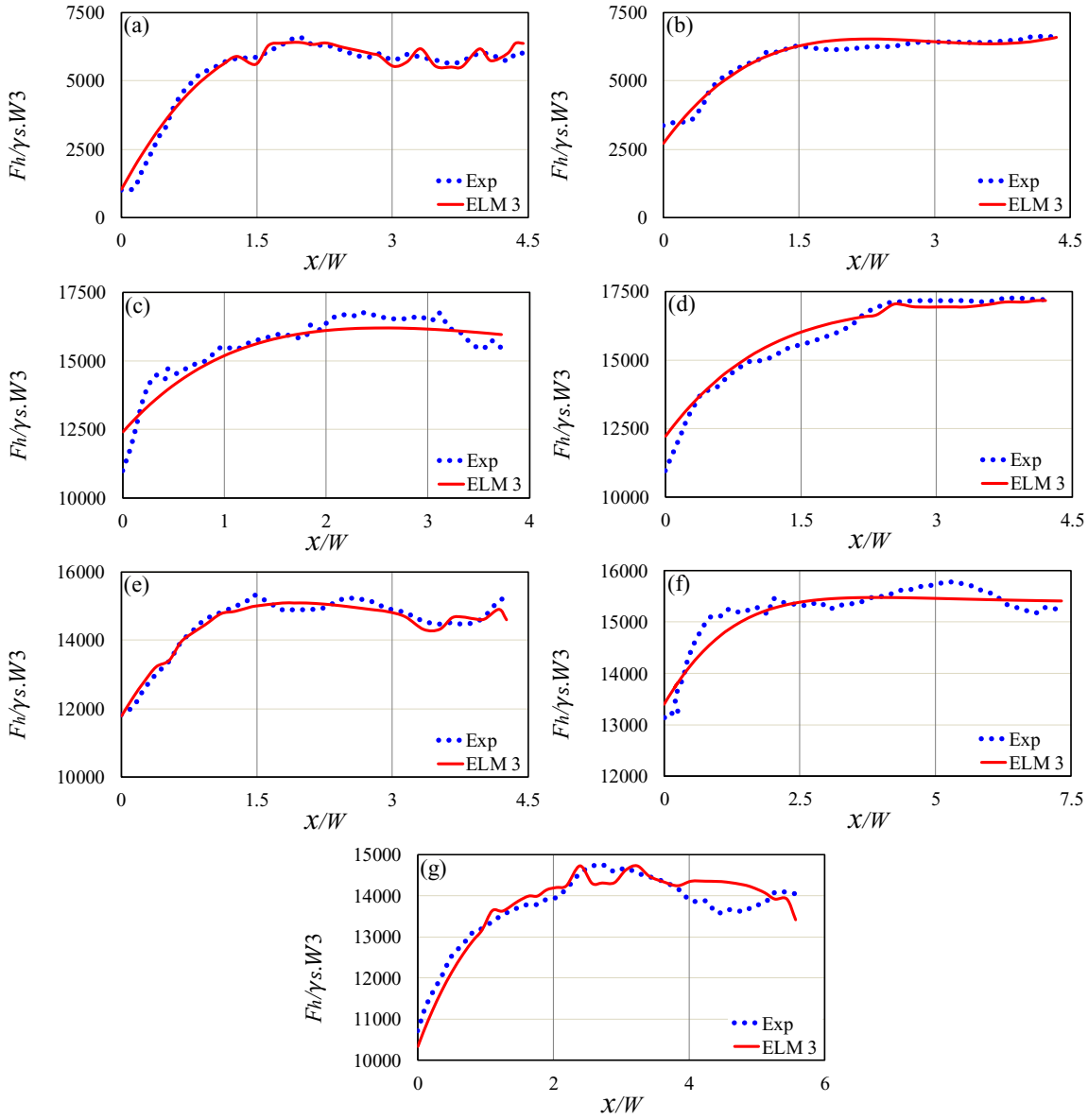


Figure 5-22. Vertical reaction force ($Fv/\gamma_s W^3$) profiles simulated by ELM 3

The profiles of horizontal subgouge deformation simulated by the best ELM model (ELM 6) are displayed in Figure 5-23. As shown, the ice-induced horizontal displacements were predicted accurately and efficiently along with a narrow bound of uncertainty by the ELM 6 model. It is worth noting that the highest level of correlation and the lowest level of complexity were obtained for this ELM model.

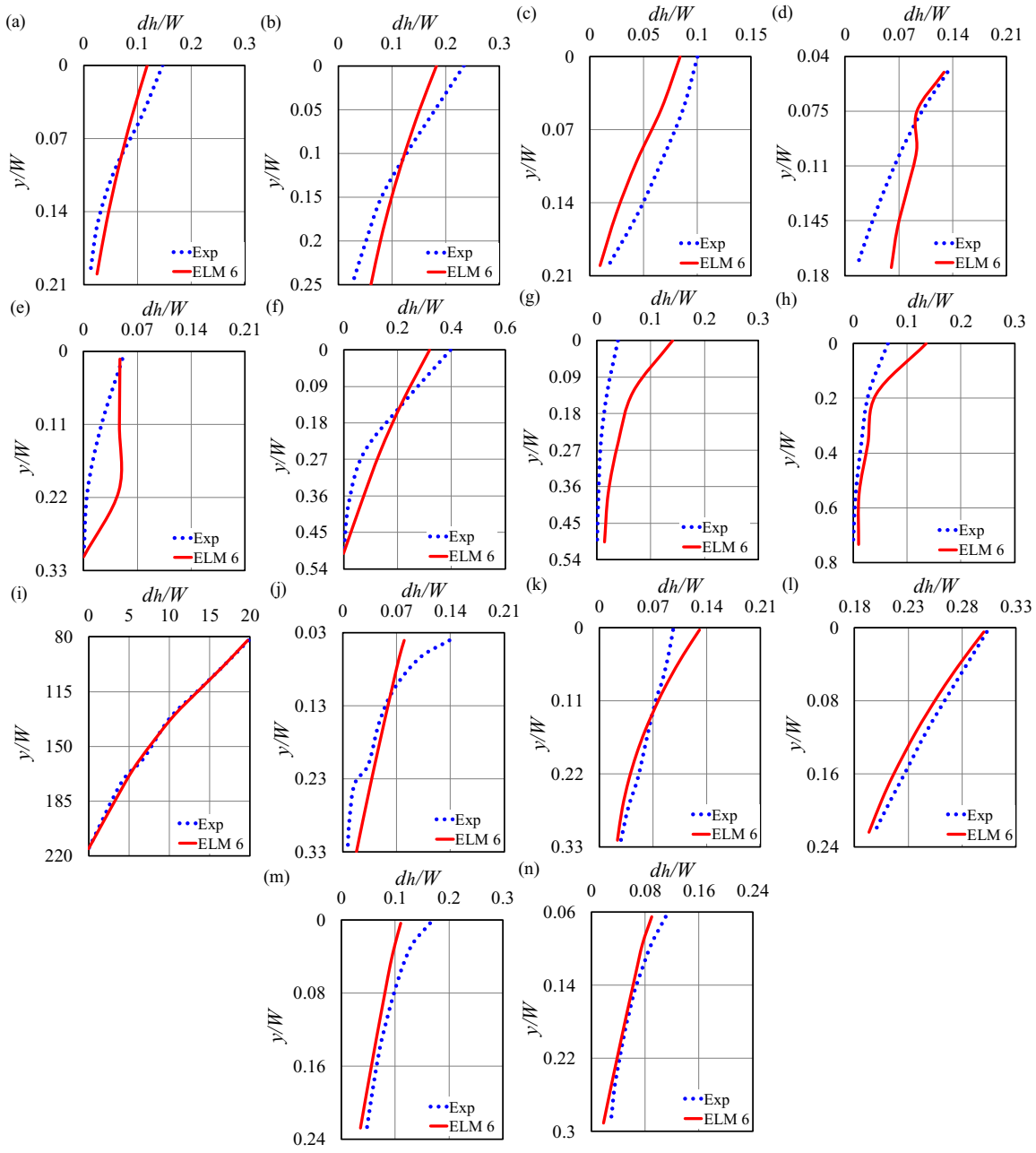


Figure 5-23. Horizontal deformation (*dh/W*) profiles simulated by ELM 6

Figure 5-24 demonstrates the profiles of vertical ice-scoured deformations estimated by ELM 3, as the best ELM model. Based on the obtained results, ELM 3 possessed the narrowest uncertainty bound at dealing with the simulation of the vertical subgouge deformations, with an

underestimated performance. This model was quite precise, and its complicacy was entirely low compared to other ELM models.

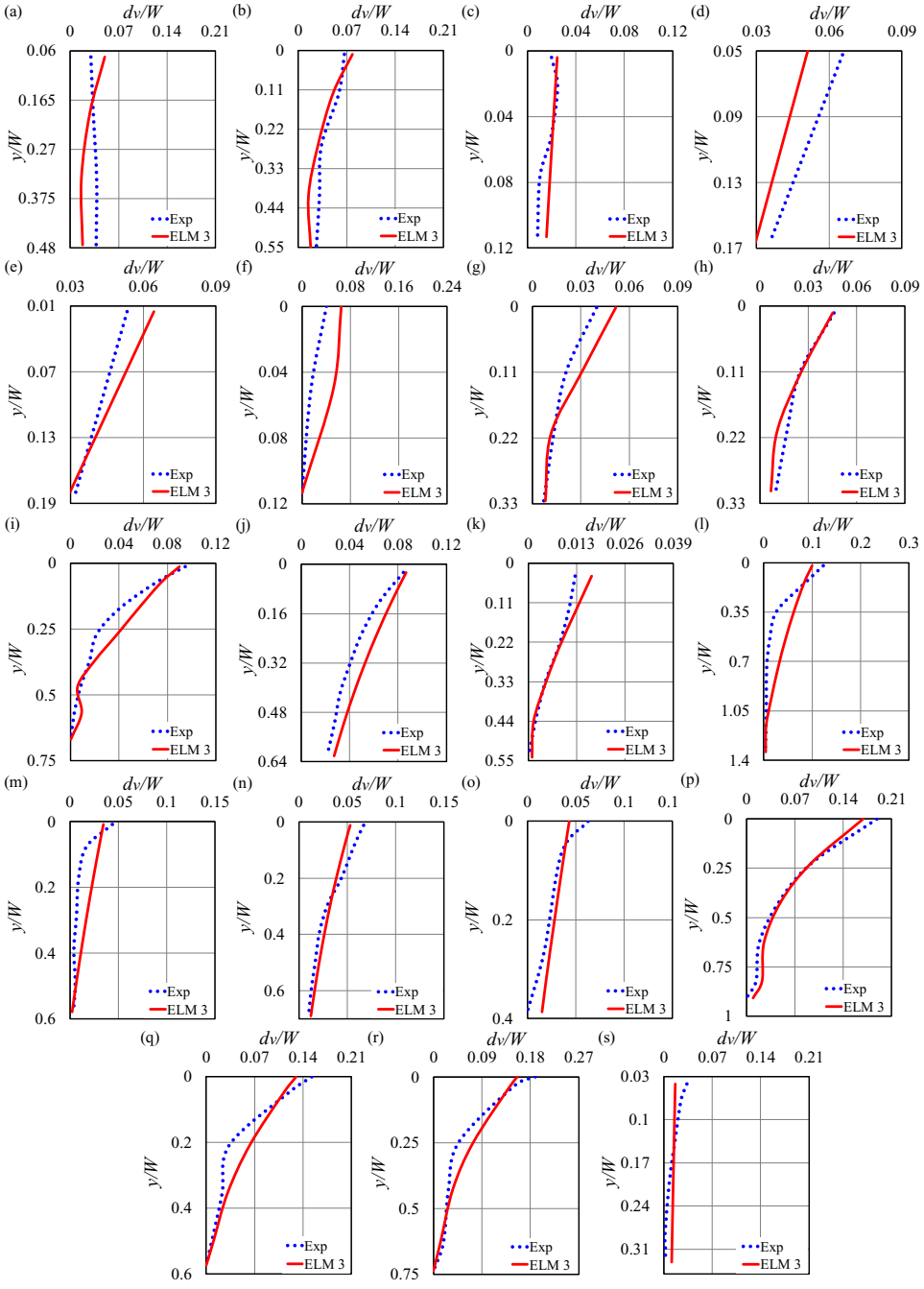


Figure 5-24. Vertical deformation (*dv/W*) profiles simulated by ELM 3

Therefore, the developed ELM models could estimate the ice-scoured clay parameters including reaction forces and deformations with reasonable accuracy and correlation along with a low level of complexity. Furthermore, the most important variables governing the subgouge displacements and reaction forces were identified in this study.

5.1.8. Conclusion

In the current study, the extreme learning machine (ELM) as a robust machine learning approach was utilized to simulate the subgouge clay parameters comprising the reaction forces and deformations. Firstly, by using the input parameters, eight ELM models were developed. It is worth mentioning that a comprehensive dataset was built to verify the ELM models and then 70% of these data were considered as the training dataset and 30% of the remaining as the testing dataset. The most significant results were summarized below:

- The optimal number of hidden neurons in the ELM network was selected to be 20, with the correlation coefficient and Nash-Sutcliff efficiency coefficient of 0.980 and 0.959.
- The hyperbolic tangent (Hypertan) was chosen as the optimized activation function in the ELM structure, with an Akaike Information Criteria of 697.554.
- The superior ELM models were introduced by conducting a sensitivity analysis. For instance, to predict the horizontal subgouge deformations, ELM 6 as a function of y/w , D_s/W , α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW had a high level of precision and low level of complexity compared to other ELM models, with an *MAE* and *R* of 0.032 and 0.999.

- The vertical load factor ($L_v/\gamma_s.W^3$), the attack angle (α), and the soil depth parameter (y/w) were identified as the most influencing input factors to estimate the ice-scoured seabed features.
- Detailed analyses proved that the ELM models had acceptable performance, for example, right 2% of the horizontal subgouge deformations simulated by ELM 6 had an error of greater than 10%.
- The implemented uncertainty analysis demonstrated that the premium ELM models possessed an overestimation performance at dealing with the simulation of reaction forces, whilst they underestimated the subgouge deformations.

The study showed that machine learning advancement as a cost-effective, robust, and fast alternative could be employed to deduct the expenses of costly experimental investigations and long-running numerical simulations in the field of the ice-seabed interaction process. Though the results presented an appropriate understanding of this complicated problem, more analyses should be done to gain a profound knowledge associated with this issue.

Funding and acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Alba, J.L., 2015. Ice Scour and Gouging Effects with Respect to Pipeline and Wellhead Placement and Design. Bureau of Safety and Environmental Enforcement (BSEE).

Arnau, S., Ivanović, A., 2019. Rate effects during ice scour in sand. Cold Regions Science and Technology, 158: 182-194.

Azimi, H., Bonakdari, H., Ebtehaj, I., 2017. Sensitivity analysis of the factors affecting the discharge capacity of side weirs in trapezoidal channels using extreme learning machines. Flow Measurement and Instrumentation, 54: 216-223.

Azimi, H., Shiri, H., 2020a. Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Banke, E.G., Smith, S.D., 1984. A Hindcast Study of Iceberg Drift on the Labrador Coast, Canadian Technical Report of Hydrography and Ocean Sciences. Canada Department of Fisheries and Oceans, No. 49.

Barnes, P.W., Rearic, D.M., Reimnitz, E., 1984. Ice-gouging characteristics and processes. In The Alaskan Beaufort Sea, Academic Press, pp. 185-212.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour Displacement in Clays from Physical Model Tests. In the 7th International

Pipeline Conference, American Society of Mechanical Engineers Digital Collection, pp. 239–245.

C-CORE., 1996. PRISE Phase 3c: Extreme LEE Gouge Event-Modeling and Interpretation. C-CORE Report 96-C32.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour. C-CORE Report 95-Cl2.

Croasdale, K.R., 2000. Study of iceberg scour and risk in the grand banks region. PERD/CHC Report, pp. 31-26.

Chouinard, L.E., 1995. Estimation of burial depths for pipelines in arctic regions. Journal of cold regions engineering, 9(4): 167-182.

Feng, G., Huang, G.B., Lin, Q., Gay, R., 2009. Error minimized extreme learning machine with growth of hidden nodes and incremental learning. IEEE Transactions on Neural Networks, 20(8): 1352-1357.

Hertz, J.A., 2018. Introduction to the theory of neural computation. CRC Press.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. IEEE international joint conference on neural networks, 2: 985-990.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing, 70(1-3): 489-501.

Huang, G.B., Zhou, H., Ding, X., Zhang, R., 2011. Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(2): 513-529.

Hynes. F., 1996. Centrifuge modelling of ice scour in sand. Doctoral dissertation, Memorial University of Newfoundland. Canada.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. In the 13th International Offshore and Polar Engineering Conference, 1: 539-544.

Lach, P.R., 1996. Centrifuge Modelling of Large Soil Deformation due to Ice Scour. Doctoral dissertation, Memorial University of Newfoundland.

Lele, S.P., Hamilton, J.M., Panico, M., Arslan, H., 2011. Advanced continuum modeling to determine pipeline strain demand due to ice-gouging. 21st International Offshore and Polar Engineering Conference, 4: 615-621.

Nematzadeh, A., Shiri, H., 2019a. Large deformation analysis of ice scour process in dense sand, 10th Congress on Numerical Methods in Engineering, At Guimaraes, Portugal.

Nematzadeh, A., Shiri, H., 2019b. Self-correcting soil models for numerical simulation of strain rate dependent ice scour in sand. 9th ECCOMAS Thematic Conference on Smart Structures and Materials, Paris, France.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Paulin, M.J., 1991. Preliminary results of physical model tests of ice scour. Master thesis, Memorial University of Newfoundland, Canada.

Paulin, M.J., 1992. Physical model analysis of iceberg scour in dry and submerged sand. Doctoral dissertation, Memorial University of Newfoundland, Canada.

Pike, K.P, Kenny, S.P., 2012. Advanced continuum modeling of the ice gouge process: Assessment of keel shape effect and geotechnical data. In the 22nd International Offshore and Polar Engineering Conference, 1: 1281-1287.

Sabodash, O.A., Bekker, A.T., 2019. Numerical Modelling of Interaction of the Hummock with a Seabed in Abaqus Software. In the 29th International Ocean and Polar Engineering Conference, ISOPE-I-19-702.

Sattar, A.M., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. Neural Computing and Applications, 31(1): 157-169.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip Line Field Solutions as an Approach to Understand Ice Subgouge Deformation Patterns. In the 16th International Offshore and Polar Engineering Conference, 1: 628-633.

Woodworth-Lynas, C.M.T., Guigné, J.Y., 1990. Iceberg scours in the geological record: examples from glacial Lake Agassiz. Geological Society, London, Special Publications, 53(1): 217-233.

Woodworth-Lynes, C., Nixon, D., Phillips, R., Palmer, A., 1996. Subgouge deformations and the security of Arctic marine pipelines. Offshore Technology Conference, OTC-8222-MS.

Yang, W., 2009. Physical modeling of subgouge deformations in sand. Doctoral dissertation, Memorial University of Newfoundland. Canada

Section 3

Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine

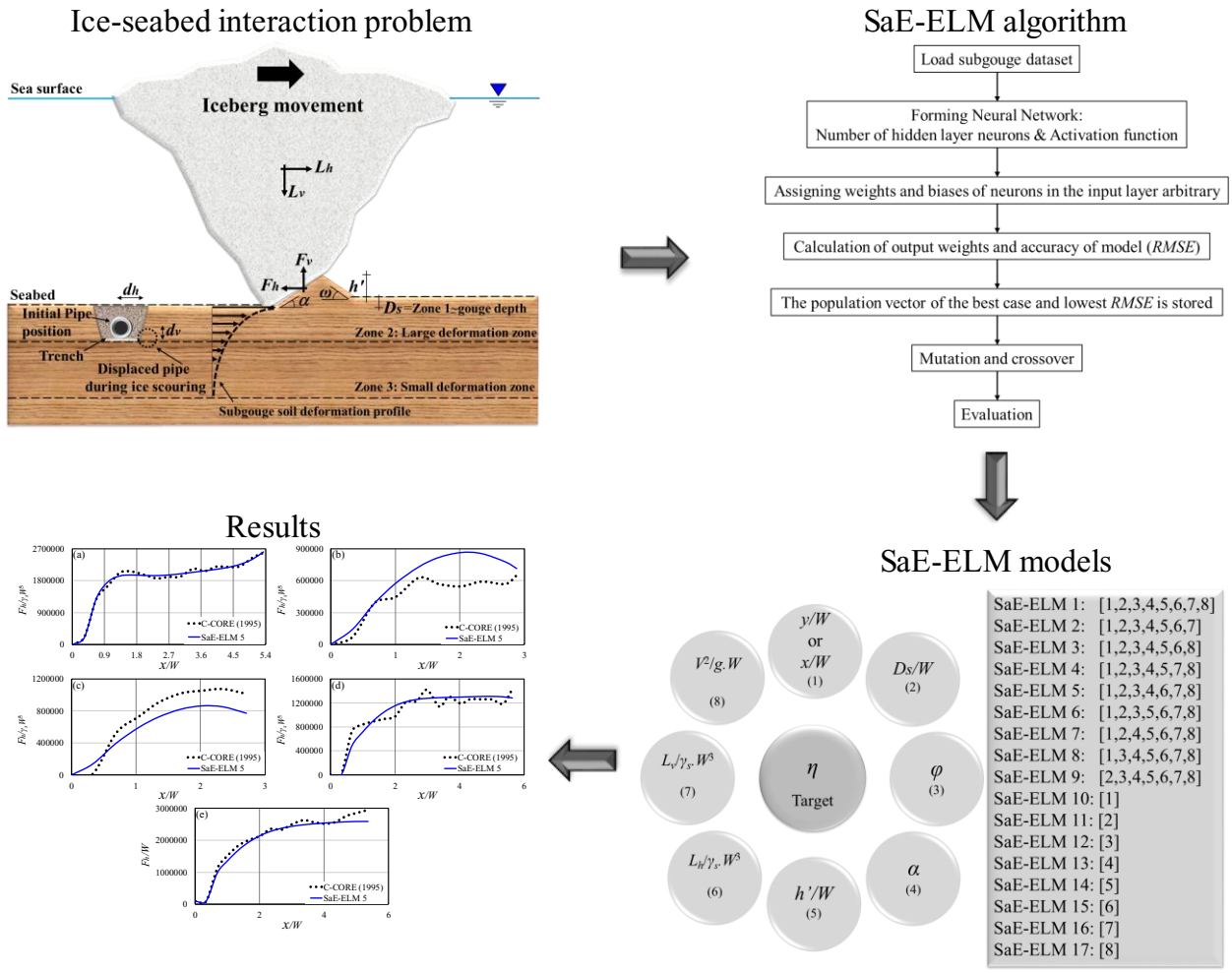
This section is a journal paper which has been published in Ocean Engineering (2021), Vol. 239, pp. 109795, ISSN: 0029-8018

Abstract

Recently discovered oil and gases in the Arctic area have heightened the need for more attention to ice-seabed interaction during an ice-scouring event. The seabed is gouged by these drifting icebergs in warmer months threatening the subsea pipelines and transferring the hydrocarbons from offshore to onshore. The simulation of ice scouring the seabed needs costly large deformation finite element analysis for the guaranteed operational integrity of the subsea pipelines. In this paper, a cost-effective alternative approach using the self-adaptive evolutionary extreme learning machine (SaE-ELM) algorithm was taken to model the ice-induced seabed scour. Initially, using parameters governing the ice-gouging process, 17 SaE-ELM models were developed. Then, a comprehensive dataset was established and properly allocated for the training and testing of the developed models. The optimal number of hidden layer neurons and the best activation function opted for the SaE-ELM network. The premium SaE-ELM models and the most influencing inputs were recognized by conducting a sensitivity analysis. The vertical component of load showed a significant impact on the reaction forces, rather the soil depth and berm height possessed a striking effect for modeling the soil displacements. Ultimately, a set of SaE-ELM-based equations were presented to estimate the subgouge soil parameters.

Keywords: Ice-gouging, Sandy seabed, Self-adaptive evolutionary extreme learning machine, Error analysis, Sensitivity analysis, Uncertainty analysis

Graphical abstract



5.1.9. Introduction

It is worthwhile noting that approximately 30% and 15% of the globe's undiscovered natural gases and crude oil are stored in the Arctic territories (Gautier et al. 2009) and the vast majority of these hydrocarbons should be transferred from offshore to onshore by using the subsea pipelines as an economical and safe way (Arnau Almirall 2017). Seabed scouring by icebergs is recognized as being one of the most important features in the Arctic shallow waters where traveling icebergs and pressure ridges would gouge the sea floor during warmer months. During an ice-gouging phenomenon, the maximum subgouge soil deformation occurs right under the iceberg keel base, and the magnitude of the soil displacements decreases by extending to the greater depths (Azimi and Shiri 2020a). Figure 5-25 displays the schematic layout of an iceberg-seabed interaction problem.

The gouge depth (*Ds*) is defined as the scour incision depth and the subsea pipeline is displaced in both horizontal (*dh*) and vertical (*dv*) directions during an ice-gouging problem (see Figure 5-25). The *Lh* and *Lv* represent the resultant horizontal and vertical loads comprising drag force from wind, drag force from current, buoyancy force, keel weight, friction force on the bottom of the keel, Coulomb's passive friction force acting in front and on both sides of the keel, and driving force from the surrounding floe that are all amongst the input parameters to the problem. It is worthwhile noting that the resultant horizontal (L_h) and vertical loads (L_v) were reported in the applied experimental investigations to feed the used machine learning (ML) model in the current study (Azimi and Shiri 2020a). In terms of pipeline response to the ice-gouging, the ultimate design factors include subgouge soil displacements (*d*) and the keel reaction forces (*F*). These factors are resultant of several key input variables including the attack angle (α), gouge depth (*Ds*), gouge width (*W*), seabed soil properties (φ), ice dynamics (*V*), environmental loads (*L*), etc. In the current

investigation, similar to all of the previous studies (e.g., Paulin 1991 & 1992; Lach 1996; C-CORE 1995; C-CORE 1996; Hynes 1996; Nematzadeh and Shiri 2019a; Azimi and Shiri 2021), the subgouge soil deformations (d) and the keel reaction forces (F) were considered as the key outputs of the ice-gouging analysis.

Currently, subsea trenching and backfilling are commonly utilized to bury these structures for physical protection against ice scour events (Nematzadeh and Shiri 2020). Thus, the estimation of subgouge soil characteristics comprising the reaction forces and soil displacements are significant design factors playing with the project's budget redlines to minimize the required burial depth for physical protection against operational and environmental loads (Azimi and Shiri 2020b).

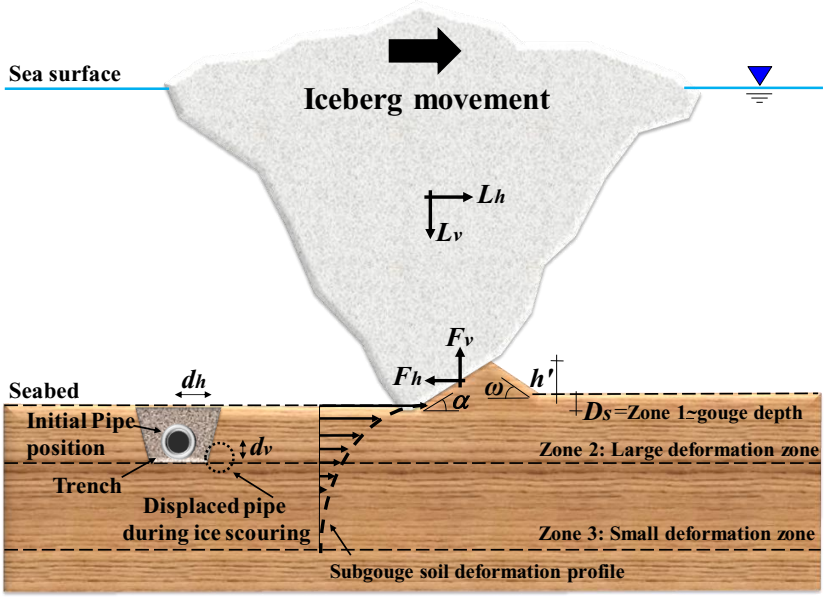


Figure 5-25. Schematic layout of an iceberg-seabed interaction problem

Some preliminary field and experimental investigations associated with sea ice scouring were carried out in the early 1980s (Chari 1980; Comfort and Graham 1986; Machemehl and Jo1989), rather ice-gouging issue has received much attention over the last three decades and several costly experimental studies have been implemented by many researchers and scholars. For instance,

Paulin (1991) performed ten experimental ice-gouging tests for the sandy and clay seabed under different circumstances. The author highlighted that the vertical reaction forces were significant for models with a smaller attack angle, whereas the horizontal reaction forces were quite dominant for a greater attack angle. Paulin (1992) carried out several ice-gouging tests experimentally for both dry and submerged sand seabed in order to measure the subgouge displacements and reaction forces. The author maintained that the ice-gouging process was similar for both submerged and dry seabed, whereas reaction forces and loads for the submerged models were much less than for the dry tests. Subsequently, a set of the Pressure Ridge Ice Scour Experiment (PRISE) tests was implemented at the Center for Cold Ocean Resources Engineering (C-CORE) to study the subscour deformations and reaction forces for different clay and sandy seabed (C-CORE 1995 and C-CORE 1996). These research programs showed that the horizontal reaction forces enhanced by increasing the scour width, whilst this parameter decreased with the steeper attack angle of the ice keel. Hynes (1996) investigated the ice keel scour problem for a sandy seabed in a centrifuge study. The author reported the values of scour-induced reaction force, displacement, and pore pressure. They concluded that the relation between the scour loads and scour depths could be described linearly. Yang (2009) conducted some Pipeline Ice Risk Assessment and Mitigation (PIRAM) tests for the sand seabed. The author focused on ice-induced displacements and evaluated the effects of gouge depth, frontal berm height, and gouge rate. The study showed that the reaction force was a function of the gouge geometry and attack angle.

Although there has been considerable interest in the large deformation finite element simulations of the ice-gouging problem in recent years (Nematzadeh and Shiri 2019a; Nematzadeh and Shiri 2019b), these numerical analyses entail time-consuming calculations (Azimi and Shiri 2020a). For example, Shin et al. (2019) modeled the ice scouring issue by means of the Coupled Eulerian

Lagrangian (CEL) method. These authors applied the contact condition and geostatic stress to predict the ice keel gouging and verified the simulation results with experimental values.

Arnau and Ivanović (2019) performed a set of small-scale 1g laboratory ice scour tests for dry and saturated sand. The investigation showed that the drifting speed had a significant impact on the scouring loads. The authors concluded that the rate effects grew linearly with the berm height and the scour depth.

Moreover, Nematzadeh and Shiri (2020) simulated the influence of the non-linear stress-strain behavior of dense sand in an ice-gouging event through a modified Mohr-Coulomb (MMC) scheme. The study proved that the side berm size was affected by the magnitude of the attack angle, rather the frontal berm was not significantly changed.

On the other hand, the last decades have witnessed a huge growth in the use of artificial intelligence (AI) and machine learning (ML) advancement since they are considered precise, fast, and cost-effective tools. The first investigations into the application of AI technology in the simulation of the ice-gouging phenomenon were reported by Kioka et al. (2003) and Kioka et al. (2004). The authors utilized the neural network (NN) algorithm to simulate the ice scouring problem and validated these models with a mechanical approach. Azimi and Shiri (2020a) introduced the parameters affecting the ice-gouging process regarding Buckingham's theory for both clay and sandy seabed. They suggested a set of linear regression (LR) models to estimate the maximum ice-induced displacements. Additionally, Gene Expression Programming (GEP) was applied to predict the subgouge sand deformations for the first time by Azimi and Shiri (2020b). They pointed out that the soil depth and dilation index had a significant effect to simulate the horizontal subgouge deformations by the GEP model.

Azimi and Shiri (2021) simulated the ice-seabed interaction parameters by means of a machine learning (ML) algorithm. The authors asserted that the sand shear strength parameter and gouge depth ratio possessed a significant effect on the subgouge reaction forces and displacements. The study also showed that the ML algorithm outperformed the empirical models to estimate the ice-induced sand deformations and the keel reaction forces.

As shown, simulation of the subgouge soil response is a key design requirement for subsea facilities, so the industry keeps looking for more reliable and cost-effective alternatives to model the behavior of these sea bottom-founded structures to lessen the collision risk of icebergs with them. Therefore, to fill this knowledge gap, the capability of a robust self-adaptive machine learning called a self-adaptive evolutionary extreme learning machine (SaE-ELM) to model the subgouge sand characteristics is assessed in the current investigation for the first time. More details on this will be given in the next sections.

5.1.10. Methodology

The extreme learning machine (ELM), differential evolution (DE) algorithm, and self-adaptive evolutionary extreme learning machine (SaE-ELM) were detailed and then the parameters affecting the ice-induced sand seabed were presented. After that, using these inputs, seventeen SaE-ELM models were defined. Ultimately, several criteria were introduced for the appraisal of these self-adaptive machine learning models.

5.1.10.1. Extreme learning machine (ELM)

An extreme learning machine (ELM) is a powerful neural network algorithm introduced by Huang et al. (2004) to solve different linear and non-linear problems (Azimi and Shiri 2021). The ELM

is a quick training machine learning owning the structure of a single-layer feed-forward neural network (SLFFNN) (Huang et al. 2006). Figure 5-26 exhibits the basic structure of the ELM network. In comparison with the classical backpropagation (BP) algorithm, the ELM possesses several benefits as follows:

- The high training speed of feedforward neural network (FFNN) by ELM,
- Better generalization performance,
- Training with differentiable and non-differentiable activation functions,
- Overcome the gradient-based algorithm such as the local optimum, improper learning rate, and overfitting problem (Huang et al. 2011; Sattar et al. 2019, Azimi and Shiri 2021).

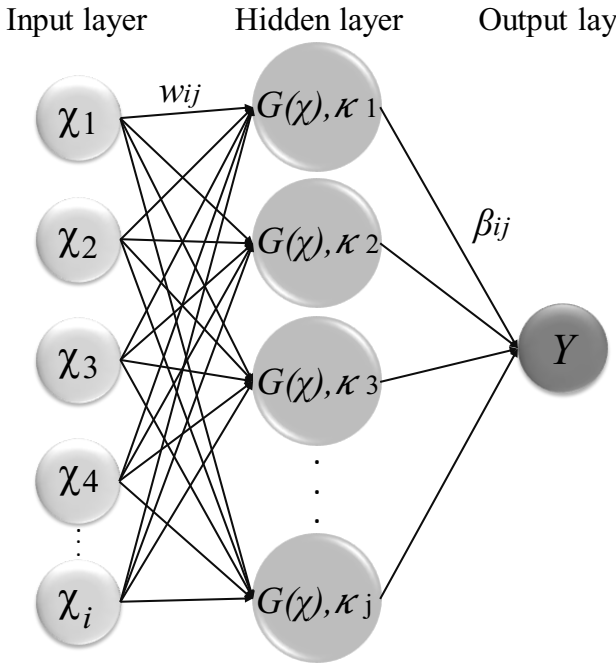


Figure 5-26. Structure of ELM network

As shown in Figure 5-26, the ELM has three distinct layers comprising an input layer, a hidden layer, and an output layer. The input layer contains the input parameters of the ice-gouging problem and the number of neurons in the output layer equals the number of target parameters of

the issue. However, the number of neurons in the hidden layer is chosen based on a trial and error procedure (Azimi and Shiri 2021).

A single hidden layer forward network is formed by means of linear algebra in these three layers so as to solve the equations for obtaining the optimum weights in the output layer (Sattar et al. 2019). It is worth noting that the weights of the input layer are determined randomly, rather the output weights are analytically estimated through a predefined training mechanism. Therefore, the training procedure in the ELM model is extremely fast and consequently, its generalization capability is quite significant. The output of an SLFFNN is solved as below:

$$Y = \sum_{j=1}^{m} \beta_j G\left(\sum_{i=1}^{n} w_{i,j}\chi_i + \kappa_j\right) \tag{5-48}$$

where, Y and χ denote the output and input of the ELM algorithm, respectively. Additionally, n and m are the number of input and output parameters, respectively. Moreover, $w_{i,j}$ indicates input weight that is the connection between the i^{th} neuron in the input layer to the j^{th} neuron in the hidden layer; β_j is a coefficient that links the j^{th} neuron in the hidden layer to the particular neuron within the output layer. Besides, $G(\,)$ and κ_j are the activation function and biases of the hidden layer neurons, respectively.

Generally, the output of an SLFFNN is computed in two major steps. Initially, the network of the hidden layer is constructed, meaning that the number of neurons in this layer is determined by using a trial and error procedure (Azimi and Shiri 2021). Subsequently, in order to train the ELM network, an activation function (AF) is defined. Indeed, the optimum number of hidden neurons and the best activation function lead to the optimal ELM network. Then, using arbitrary assignments of the input layer weights $(w_{i,j})$ and the biases (κ_j), the structure of hidden layer

neurons and the weight in the output layer are calculated. Therefore, equation (5-48) is rewritten as follows:

$$H\beta = Y \tag{5-49}$$

here, H is the extreme learning machine feature mapping matrix as below:

$$H\left(w_{i,j}, \kappa_j, \chi_i\right) = \begin{bmatrix} H_{1,1} & \cdots & H_{1,m} \\ \vdots & \ddots & \vdots \\ H_{n,1} & \cdots & H_{n,m} \end{bmatrix} = \begin{bmatrix} G\left(w_{1,1}\chi_1 + \kappa_1\right) & \cdots & G\left(w_{1,m}\chi_m + \kappa_m\right) \\ \vdots & \ddots & \vdots \\ G\left(w_{n,1}\chi_n + \kappa_1\right) & \cdots & G\left(w_{n,m}\chi_m + \kappa_m\right) \end{bmatrix} \tag{5-50}$$

In addition, β and Y are defined in the following form:

$$Y = [Y_1\ Y_2\ \cdots\ Y_m] \quad \text{and} \quad \beta = [\beta_1\ \beta_2\ \cdots\ \beta_n] \tag{5-51}$$

The β_j is obtained by minimizing error in an estimation through the Moore-Penrose generalized inverse matrix as follows (Huang et al. 2011):

$$\hat{\beta} = H^{+}Y \tag{5-52}$$

here, H^{+} represents the Moore-Penrose generalized inverse matrix of H.

5.1.10.2. Differential evolution (DE)

Differential evolution (DE) as a robust optimizer algorithm was first proposed by Storn and Price (1997). The automatic detection-derivation mechanism and the high convergence speed are the most significant benefits of DE over other optimizer tools (Bonakdari et al. 2020). Commonly, minimizing the below equation is the major aim of the DE algorithm:

$$\min f(\vartheta) \quad \text{where} \quad \vartheta \in \boldsymbol{R}^{\mathrm{D}} \tag{5-53}$$

here, ϑ is a vector, where the DE algorithm generally employs a population of I individuals so as to calculate the global optimal of the ice-seabed interaction parameters. Therefore, the i^{th} vector is surmised as below:

$$\vartheta_{i,\varpi} = \vartheta_{i,\varpi}^{1}, \vartheta_{i,\varpi}^{2}, \vartheta_{i,\varpi}^{3}, \cdots, \vartheta_{i,\varpi}^{\mathrm{D}} \qquad \text{where } i = 1,2,3, \cdots I \tag{5-54}$$

where, ϖ is defined as the generation number. In the current study, the DE algorithm utilizes four major stages to solve the ice-seabed interaction problem as follows:

(I) Initialization: *I* vectors are randomly produced by using the below equation:

$$\vartheta_{i,\varpi} = \vartheta_{min} + rand(0,1) \tag{5-55}$$

$$\times (\vartheta_{max} - \vartheta_{min}) \qquad \text{where} \begin{cases} \vartheta_{min} = [\vartheta_{min}^{1}, \vartheta_{min}^{2}, \cdots, \vartheta_{min}^{\mathrm{D}}] \\ \vartheta_{max} = [\vartheta_{max}^{1}, \vartheta_{max}^{2}, \cdots, \vartheta_{max}^{\mathrm{D}}] \end{cases}$$

here, the ϑ_{max} and ϑ_{min} are defined as the boundaries considered for the ice-gouging problem.

(II) Mutation: The mutant vectors ($\Psi_{i,\varpi}$) is created by means of the generated vector ($\vartheta_{i,\varpi}$) in the mutation phase. Various mutation strategies have been introduced by Storn and Price (1997); however, four of them are applied in the current investigation comprised as follows:

Strategy 1:

$$\Psi_{i,\varpi} = \vartheta_{\sigma_{1,\varpi}^{i}} + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) \tag{5-56}$$

Strategy 2:

$$\Psi_{i,\varpi} = \vartheta_{\sigma_{1,\varpi}^{i}} + \Lambda \times \left(\vartheta_{best,\varpi} - \vartheta_{\sigma_{1,\varpi}^{i}}\right) + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) + \Lambda \times \left(\vartheta_{\sigma_{4,\varpi}^{i}} - \vartheta_{\sigma_{5,\varpi}^{i}}\right) \tag{5-57}$$

Strategy 3:

$$\Psi_{i,\varpi} = \vartheta_{\sigma_{1,\varpi}^{i}} + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) + \Lambda \times \left(\vartheta_{\sigma_{4,\varpi}^{i}} - \vartheta_{\sigma_{5,\varpi}^{i}}\right) \tag{5-58}$$

Strategy 4:

$$\Psi_{i,\varpi} = \vartheta_{i,\varpi} + \Lambda \times \left(\vartheta_{\sigma_{1,\varpi}^{i}} - \vartheta_{i,\varpi}\right) + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) \tag{5-59}$$

here, σ_{k}^{i} is randomly calculated from the $[1,2,3,\cdots,I]$ interval and *k* as the control parameter is randomly adjusted between 0 and 1. The Λ as a scaling factor is selected between 0 and 2 (Storn and Price 1997).

(III) Crossover: The crossover phase is performed on the mutated vectors to enhance the model's versatility. During this stage, the trial vector $\left(\Delta_{i,\varpi} = \left\lfloor \Delta_{i,\varpi}^{1}, \Delta_{i,\varpi}^{2}, \cdots, \Delta_{i,\varpi}^{\mathrm{D}}, \right\rfloor\right)$ is produced for the mutated vectors $\left(\Psi_{i,\varpi} = \left\lfloor \Psi_{i,\varpi}^{1}, \Psi_{i,\varpi}^{2}, \cdots, \Psi_{i,\varpi}^{\mathrm{D}}, \right\rfloor\right)$ using the equation below:

$$\Delta_{i,\varpi}^{j} = \begin{cases} \Psi_{i,\varpi}^{j} & \text{if}\left(\mathrm{rand}_j \le CC\right) \text{ or } \left(j = j_{\mathrm{rand}}\right) \\ \vartheta_{i,\varpi}^{j} & \text{otherwise} \end{cases} \tag{5-60}$$

here, *CC* is the coefficient of crossover and the rand_j variable is randomly taken into account as a constant value. The j_{rand} parameter is adjusted between 1 and Ɗ, since one of the $\Delta_{i,\varpi}$ variables ought to be different from the $\vartheta_{i,\varpi}$ values.

(IV) Selection: The best individual with the minimum target value is chosen as the solution to the ice-seabed interaction problem.

Phases (II) to (IV) are iterated till either the desired accuracy or maximum predefined iterations are achieved.

5.1.10.3. Self-adaptive evolutionary extreme learning machine (SaE-ELM)

Even though the ELM algorithm has an acceptable performance to simulate multifarious linear or nonlinear issues, selecting the optimized value of the parameters in the hidden layer manually plays a crucial role to result in simulating a problem appropriately. Moreover, in practical applications, the input parameters should be dimensionless which is another drawback of the ELM network (Cao et al 2012). In order to overcome these limitations, the self-adaptive evolutionary extreme learning machine (SaE-ELM), as a hybrid model of the ELM network and DE algorithm, was developed for the first time by Cao et al. (2012).

The DE algorithm has been widely utilized in the literature to optimize the parameters of SLFFNN models (e.g., Cao et al. 2012). In the current study, the parameters of neurons in the hidden layer

were optimized by using the DE algorithm, while the output weights were analytically determined through the ELM model. The developed SaE-ELM network has a better generalization ability and also a further compact network structure in comparison with the classical ELM algorithm (Cao et al. 2012).

The developed SaE-ELM model in this study has several advantages compared with the classical ELM models (e.g., Azimi and Shiri 2021):

- SaE-ELM algorithm has a better generalization capability owing to optimization of the parameters of neurons in the hidden layer by using the Differential evolution (DE). However, the parameters of the hidden layer neurons in the ELM algorithm were randomly determined (e.g., Cao et al. 2012).
- The classical ELM algorithm was sensitive to the number of hidden layer neurons and the determination of the optimal number of hidden layer neurons was quite laborious (Azimi and Shiri 2021) but these parameters in the SaE-ELM network were analytically optimized by means of the DE algorithm.
- The input parameters in the ELM network had to be dimensionless which prevented the ELM usage in practical applications (Azimi and Shiri 2021) whereas the input parameters in the SaE-ELM model could be either dimensional or dimensionless.
- In the current study, 17 SaE-ELM models had been defined in which the effect of each single input parameter on the target function was assessed, whilst the published works in the field of ice-gouging lacked this comprehensive analysis.

Generally, SaE-ELM was a novel version of the ELM algorithm that overcame the drawbacks of the classical form. Although the classical ELM algorithm had a good performance at dealing with the simulation of the ice-gouging characteristics (Azimi and Shiri 2021), the ELM trapped in local

optima and the algorithm had overfitting problems in some cases, while the SaE-ELM algorithm overcome these disadvantages. Figure 5-27 details the flowchart of the SaE-ELM model in the current study.

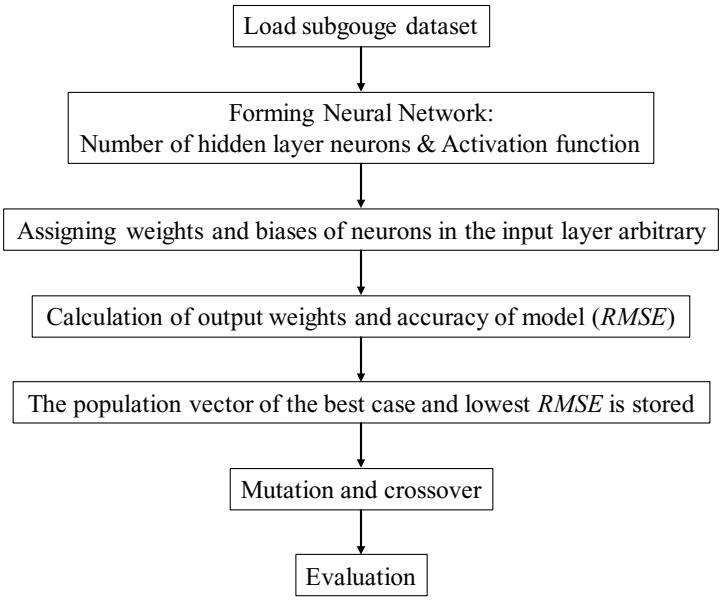


Figure 5-27. Flowchart of the SaE-ELM in the current study

In the SaE-ELM algorithm, the number of neurons in the hidden layer is optimized by using the self-adaptive differential evolution algorithm. During the training procedure, the control coefficients in the DE algorithm and the strategies for trial vector generation are determined as a self-adaptive mechanism. In other words, during the simulation process, the output weights are estimated by means of the ELM network, whereas the input weights and biases of the hidden layer neurons are determined through the DE algorithm. To model the ice-seabed interaction process, the main steps comprising initialization, output weights, mutation and crossover, and evaluation are applied by using the SaE-ELM algorithm as follows:

Initialization: the *I* vector is randomly created in the first DE generation and the following vector is computed:

$$\vartheta_{k,\varpi} = \left\lfloor a^{Y}_{1,(k,\varpi)}, \cdots, a^{Y}_{l,(k,\varpi)}, b_{1,(k,\varpi)}, \cdots, b_{l,(k,\varpi)} \right\rfloor \tag{5-61}$$

where, a and b are the learning parameters and l is the size of the hidden layers.

Output weights: the output weights are defined by using the following equation:

$$\beta_{k,\varpi} = H_{k,\varpi} Y \tag{5-62}$$

where, the $H_{k,\varpi}$ is determined by using the equation below:

$$H_{k,\varpi} = \begin{bmatrix} G\left(a_{1,(k,\varpi)}, b_{1,(k,\varpi)}, \chi_1\right) & \cdots & G\left(a_{n,(k,\varpi)}, b_{n,(k,\varpi)}, \chi_1\right) \\ \vdots & \ddots & \vdots \\ G\left(a_{1,(k,\varpi)}, b_{1,(k,\varpi)}, \chi_m\right) & \cdots & G\left(a_{L,(k,\varpi)}, b_{L,(k,\varpi)}, \chi_m\right) \end{bmatrix} \tag{5-63}$$

The Root Mean Squared Error (*RMSE*) of the generated individuals is computed as follows:

$$RMSE_{k,\varpi} = \sqrt{\frac{\sum_{i=1}^{m} \left\| \sum_{j=1}^{L} \beta_j . G\left(a_{j,(k,\varpi)}, b_{j,(k,\varpi)}, \chi_i\right) - Y_i \right\|}{n \times m}} \tag{5-64}$$

Then, the next generation of individuals is calculated using the following equation:

$$\vartheta_{k,\varpi+1} = \begin{cases} \Delta_{k,\varpi+1} & if \quad RMSE_{\vartheta_{k,\varpi}} - RMSE_{\vartheta_{k,\varpi}} > \pi . RMSE_{\vartheta_{k,\varpi}} \\ \Delta_{k,\varpi+1} & if \quad \left| RMSE_{\vartheta_{k,\varpi}} - RMSE_{\vartheta_{k,\varpi}} \right| < \pi . RMSE_{\vartheta_{k,\varpi}} \quad and \quad \left\| \beta_{\Delta_{k,\varpi+1}} \right\| < \left\| \beta_{k,\varpi} \right\| \\ \vartheta_{k,\varpi} & else \end{cases} \tag{5-65}$$

where, π is defined as a positive constant value to hinder the zero improvement rate.

Mutation and Crossover: using the introduced strategies in equations (5-56) to (5-59), the trial vectors of the DE algorithm are defined. Furthermore, the best strategy is chosen regarding the probability procedure $P_{l,\varpi}$ where the l^{th} strategy is considered for the ϖ^{th} and l is 1, 2, 3, or 4. Generally, $P_{l,\varpi}$ is updated in each step, meaning that if the population of generated vectors (I) is higher than ϖ, the four applied strategies possess an identical probability ($P_{l,\varpi} = 0.25$). If ϖ is greater than the I, $P_{l,\varpi}$ is calculated as follows:

$$P_{l,\varpi} = \frac{\varrho_{l,\varpi}}{\sum_{l=1}^{4} \varrho_{l,\varpi}} \qquad \text{where} \qquad \varrho_{l,\varpi} = \frac{\sum_{q=\varpi-I}^{\varpi-1} n\varrho_{l,\varpi}}{\sum_{q=\varpi-I}^{\varpi-1} n\varrho_{l,\varpi} + \sum_{q=\varpi-I}^{\varpi-1} nf_{l,\varpi}} + \pi \tag{5-66}$$

where, $nf_{l,\varpi}$ and $n\varrho_{l,\varpi}$ are trial vectors. The *CC* and Λ values are determined for a target vector through the selection phase from the normal distribution function.

Evaluation: the next generation of trial vectors are created by equation (5-65). The mutation, crossover, and evaluation steps are iterated until either the predefined iteration number or required accuracy is reached during simulating the ice-induced seabed parameters using the SaE-ELM algorithm.

Although the generalization ability of SaE-ELM network was improved by using the DE algorithm and this hybrid ML model could overcome the sensitivity to the number of neurons in the hidden layer, in comparison with the classical ELM model, the SaE-ELM algorithm was slightly long-running during the simulation of target function.

5.1.10.4. Ice-seabed interaction process

The ice-induced interaction parameters (η) in a sand mass including soil deformations (d/W) and reaction forces ($F/\gamma_s W^3$) are a function of a set of parameters including the scour depth (Ds), the internal friction angle of sand (φ), the width of gouge (W), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (h'), the horizontal load (L_h), the vertical load (L_v), the velocity of ice keel (V), and the specific weight of sand (γ_s) (Lach 1996; Azimi and Shiri 2020a):

$$\eta_{(sand)} = f_1(D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \tag{5-67}$$

Equation (5-67) can be written as a function of eight dimensionless groups as below (Azimi and Shiri 2020a):

$$\frac{\eta_{(sand)}}{W} = f_2\left(\frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}68)$$

It is worth noting that the maximum subgouge deformation in the seabed sand $\left(d_{(sand)}\right)$ is formed just under the moving ice keel in the gouge centerline. However, at greater depth on the subgouge centerline, by incorporating the soil depth (y/W), the subgouge sand deformation ($d_{(sand)}/W$) is written as follows:

$$\frac{d_{(sand)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}69)$$

Similarly, the ice-induced reaction force (F) is a function of the position of the iceberg along the scour axis (x) and $D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s$ so equation (5-69) can be written as below:

$$\frac{F_{(sand)}}{\gamma_s W^3} = f_4\left(\frac{x}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}70)$$

In this study, a set of experimental investigations were used to simulate the ice-induced seabed parameters (d/W and $F/\gamma_s W^3$). This means that the key measurements of six experimental datasets reported by Paulin (1991), Paulin (1992), C-CORE (1995), Hynes (1996), C-CORE (1996), and Yang (2009) were applied to validate the SaE-ELM models. The values of the surcharged soil slope (ω) were not reported by the aforementioned research works. Therefore, the introduced dimensionless groups in equations (5-69) and (5-70) were applied as the input parameters for the SaE-ELM models to simulate the subgouge deformations and the reaction forces. The applied input combinations in this study are displayed in Figure 5-28.

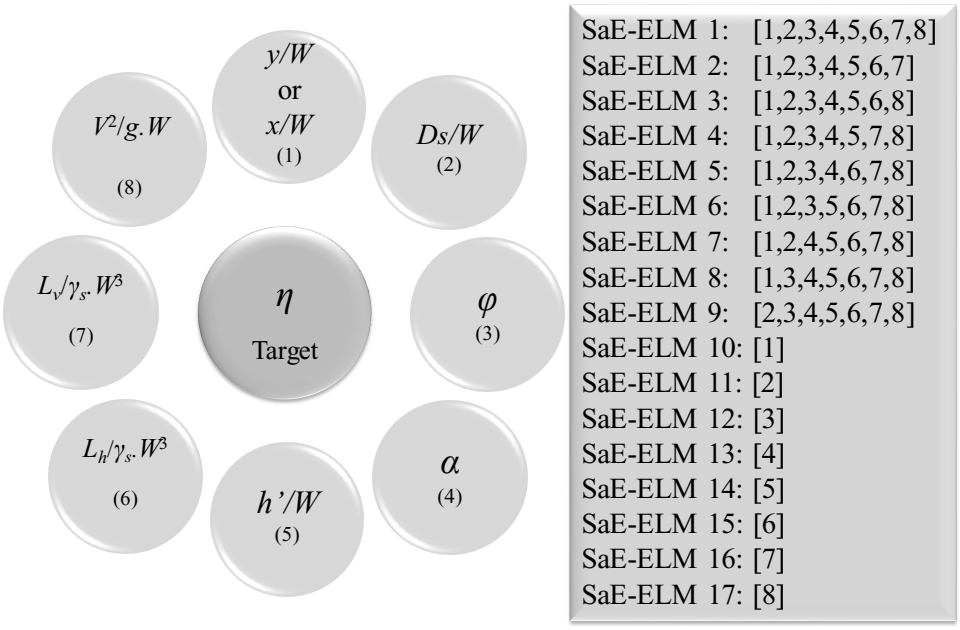


Figure 5-28. Applied input combinations in this study

Regarding the performed sensitivity analysis for the ice-gouging problem, there were eight dimensionless groups of parameters governing the subgouge sand characteristics. The SaE-ELM 1 model was a function of all input parameters, whereas the input parameters were eliminated one at a time in SaE-ELM 2 to SaE-ELM 9 (8 models) and eight models (SaE-ELM 10 to SaE-ELM 17) were developed by means of solely one input parameter

Therefore, seventeen SaE-ELM models were defined and the best models along with the most significant input parameters were identified by performing a sensitivity analysis. The superior SaE-ELM model was recognized by evaluation of the SaE-ELM 1 to SaE-ELM 9 models, while the most important input factors were respectively ranked by analyzing the SaE-ELM 10 to SaE-ELM 17 models. Subsequently, the applied dataset were split into two sub-samples comprising training (70% of the entire data) and testing (30% of the remaining) dataset.

5.1.10.5. Goodness of fit

A set of criteria such as correlation coefficient (*R*), variance accounted for (*VAF*), root mean square error (*RMSE*), mean absolute error (*MAE*), Nash-Sutcliff efficiency coefficient (*NSC*), and Akaike Information Criteria (*AIC*) were utilized to evaluate the accuracy, correctness, correlation, and complexity of the SaE-ELM models. The correlation of the SaE-ELM models was examined by the *R*, *NSC* indices, while the precision of these models was evaluated by the *RMSE* and *MAE* criteria. The correctness of the models was verified through the *VAF* index but the *AIC* indicator was applied in order for evaluation of the complexity of the defined models.

Closeness of the *R* and *NSC* indices to one signified that the SaE-ELM model owned the highest correlation with the experimental values, rather the closeness of the *RMSE* and *MAE* values to zero meant that a particular model had the lowest level of error. Moreover, the best SaE-ELM model possessed the highest value of the *VAF* index though the complexity of these SaE-ELM models was not assessed using the defined criteria. To overcome this limitation, the Akaike Information Criteria (*AIC*) was used, meaning that the less complex SaE-ELM model owned the lowest value of *AIC* value; hence, the premium SaE-ELM model had the lowest value of the *AIC* index and error (*RMSE* and *MAE*), with the highest level of correlation (*R* and *NSC*).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{5-71}$$

$$VAF = \left(1 - \frac{var(P_i - O_i)}{var(P_i)}\right) \times 100 \tag{5-72}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{5-73}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{5-74}$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (5\text{-}75)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (5\text{-}76)$$

here, Oi, Pi, $\bar{O}$, $\bar{P}$, and n are respectively the experimental measurements, the simulated values, the average experimental values, the average simulated values, and the number of experimental measurements. In addition, the k is the number of hidden layer neurons used in the SaE-ELM network.

5.1.11. Results and discussion

First, the best value for the number of hidden layer neurons and the premium activation function were opted. Subsequently, a comprehensive sensitivity analysis for SaE-ELM 1 to SaE-ELM 17 was performed and then the superior SaE-ELM models along with the most significant inputs were introduced. More analyses were conducted including the error analysis and uncertainty analysis for the best models. Ultimately, a set of SaE-ELM-based equations were proposed to estimate the subgouge sand parameters. It should be mentioned that only the simulation results in the testing mode are presented in this paper.

5.1.11.1. Hidden layer neurons

Determination of the optimal number of hidden layer neurons (NHN) is considered as a striking feature of the SaE-ELM structure since the accuracy and complexity of the SaE-ELM model is increased by growing the NHN (Sattar et al. 2019). Thus, the optimized value of NHN was selected based on a trial and error procedure in the current investigation. Figure 5-29 shows the variation of NHN against the key statistical indices for simulation of the horizontal reaction forces.

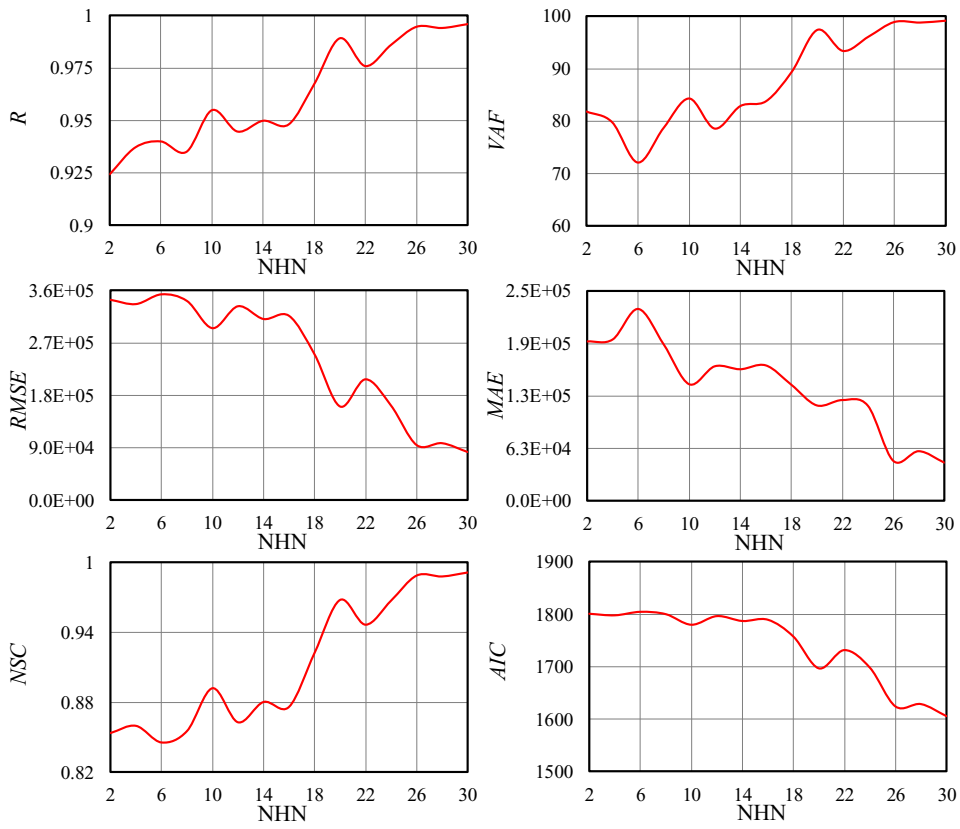


Figure 5-29. Variation of the number of hidden layer neurons against the key statistical criteria for simulation of reaction forces

The initial value of NHN was chosen as 2, where the computed variance accounted for (*VAF*) and correlation coefficient (*R*) were estimated as 81.844 and 0.924, respectively. The NHN was considered to be 28, the key criteria such as the *R*, *AIC*, and *NSC* were respectively surmised at 0.994, 1628.419, and 0.988. As shown in Figure 5, by increasing the value of NHN greater than 28, the performance of SaE-ELM model was negligibly altered. Therefore, the optimal value of NHN for the SaE-ELM network was selected to be 28.

5.1.11.2. Activation function (AF)

The weighted input parameters and biases in the applied machine learning are nonlinearly transformed by using an activation function (AF) (Hertz 2018). Six different AFs in the SaE-ELM network comprising the hyperbolic tangent (Hypertan), the triangular basis (Tribas), the radial basis (Radbas), the hard limit (Hardlim), the sigmoid (Sig), and the sinusoidal (Sin) functions were applied to simulate the ice-gouging parameters in this study. The Hypertan and Sig activation functions have widely applied in feedforward neural networks (FFNN). The Hypertan has an S-shaped curve in which the range of the function is from -1 to 1, whereas the Sig function varies from zero to 1. The Hypertan is monotonic function and its derivative is non-monotonic curve but the sigmoid is monotonic function and its derivative has a bell-shaped curve. Using the Hypertan activation function, the negative input is mapped quite negative and the zero input is mapped near zero. The Sig function is concave for values greater than zero, while it is convex for values less than zero. The Tribas is a linear Gaussian-based activation functions but it is simpler and quicker. In the Radbas activation function, the output ranges from $-\infty$ to $+\infty$ and this value is estimated regarding the distance from the origin point. By using the Hardlim activation function, the output of neuron is 1 as its input approaches a threshold, otherwise zero. The Sin is a sinusoidal activation function and it goes up and down but it outputs zero for large negative and positive input values. These activation functions have been extensively utilized in the field of machine learning (ML) for regression, classification, and clustering problems. The schematic layout of the applied activation functions is presented in Figure 5-30.

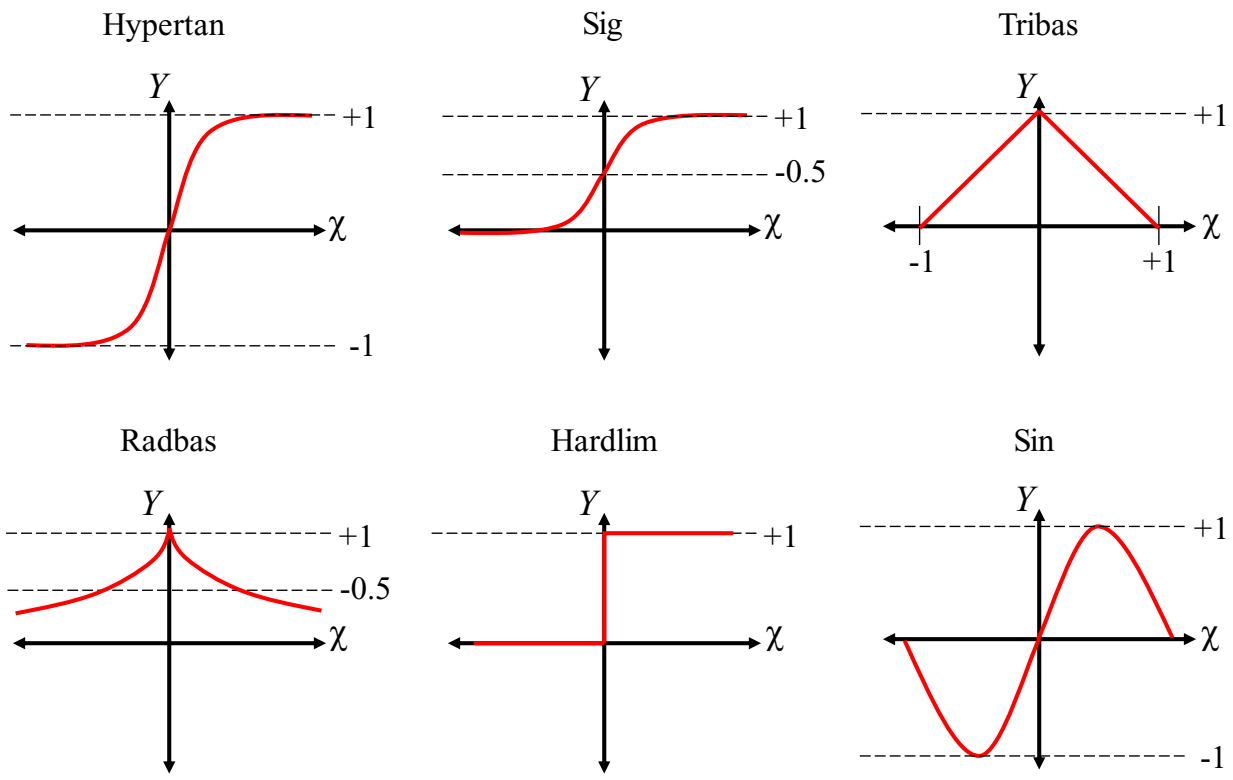


Figure 5-30. The schematic layout of the applied activation functions in the current study

Figure 5-31 portrays the comparison between the performances of various activation functions in the SaE-ELM network for simulation of the horizontal reaction forces.

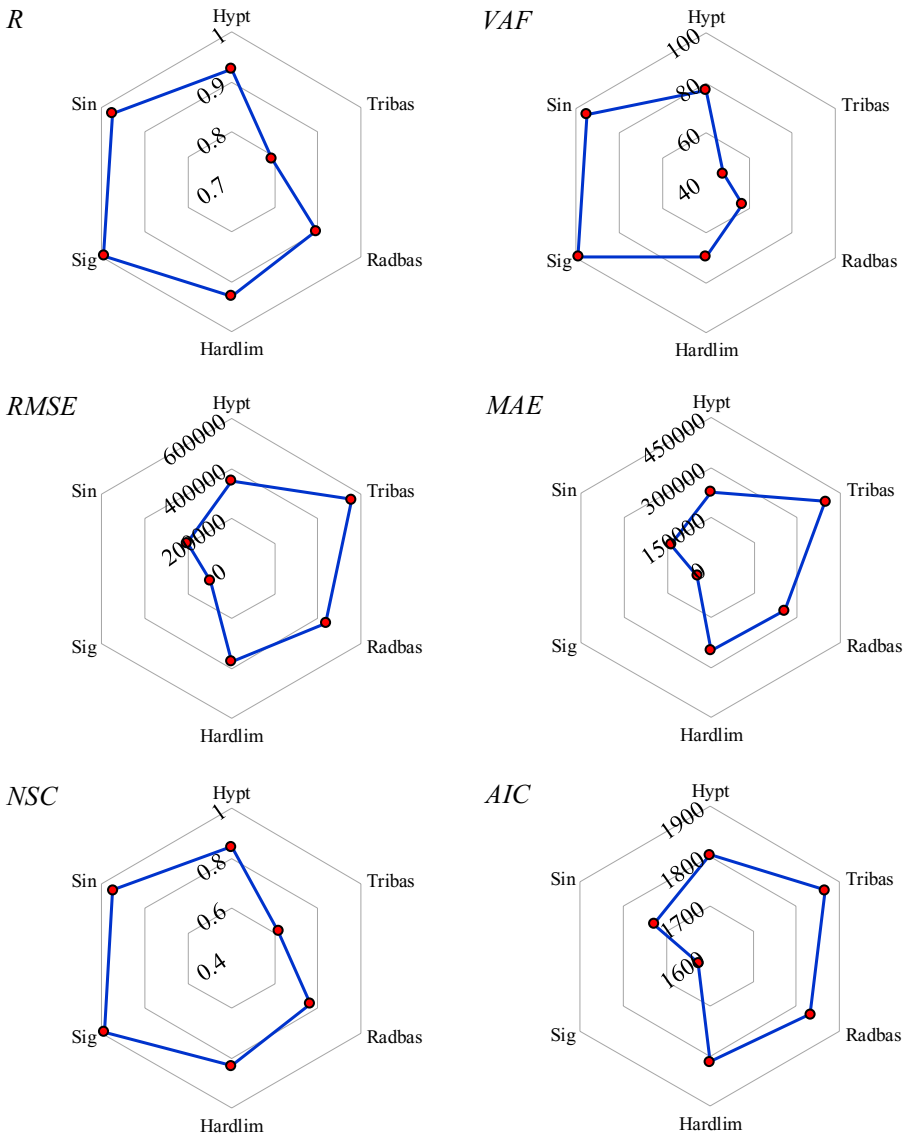


Figure 5-31. Comparison between the performances of various activation functions in the SaE-ELM network for simulation of the horizontal reaction forces

Among the utilized AFs, the Tribas function demonstrated the worst performance, where the values of AIC and R for this activation function were equal to 1865.751 and 0.793. It was found that the Sig function had the best performance to model the target value in comparison with other

AFs, with the VAF, NSC, and AIC indices of 98.929, 0.988, and 1626.670, respectively. Thus, the sigmoid function was opted as the best activation function for the SaE-ELM network in this paper.

5.1.11.3. SaE-ELM models

To introduce the superior SaE-ELM model, SaE-ELM 1 to SaE-ELM 9 were developed where SaE-ELM 1 was a function of all input factors, whereas the effect of each input parameter was removed one at a time in the SaE-ELM 2 to SaE-ELM 9 models. On the other hand, in order to detect the most influencing input factors, the SaE-ELM 10 to SaE-ELM 17 were defined. Figure 5-32 displays the calculated statistical criteria of the SaE-ELM models for simulation of the horizontal reaction forces. SaE-ELM 1 predicted the target value through x/W, Ds/W, φ, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W. The value of R for SaE-ELM 1 had been found to be 0.993. The influence of the velocity parameter (V^2/g.W) for SaE-ELM 2 was ignored and this model simulated the horizontal reaction forces by using other inputs (x/W, Ds/W, φ, α, h'/W, Lh/γs.W3, and Lv/γs.W3) where the values of NSC for the SaE-ELM 2 model equaled to 0.983. SaE-ELM 3 was a function of x/W, Ds/W, φ, α, h'/W, Lh/γs.W3, and V2/g.W, signifying that the Lv/γs.W3 input was a removed parameter for this model. The values of VAF for SaE-ELM 3 were respectively computed to be 98.636. SaE-ELM 4 predicted the Fh/γs.W3 in terms of x/W, Ds/W, φ, α, h'/W, Lv/γs.W3, and V2/g.W, with a correlation coefficient of 0.995. However, the effect of h'/W was removed for SaE-ELM 5, with a low value of complexity and high value of correlation (AIC=1628.924 and R= 0.995). The ice keel attack angle (α) was an eliminated input for SaE-ELM 6, indicating that this model simulated the horizontal reaction forces by means of x/W, Ds/W, φ, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W inputs, with the computed NSC value of 0.986. To model the horizontal reaction forces using SaE-ELM 7, the x/W, Ds/W, α, h'/W, Lh/γs.W3,

Lv/γs.W3, and V2/g.W were as input factors where the shear strength parameter of the sand seabed (φ) was an eliminated parameter. The values of AIC for SaE-ELM 7 was reckoned at 1671.649. SaE-ELM 8 predicted the target values using the x/W, φ, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W inputs and the ratio of gouge depth to gouge width (Ds/W) was excluded, with a correlation coefficient of 0.936. Amongst the SaE-ELM models with eight input factors, SaE-ELM 9 scored as the worst model, with the lowest correlation (R=0.893), the highest error (RMSE=408259.6), and the highest complexity (AIC=1823.445), since the effect of soil depth parameter (y/W) was deleted for the SaE-ELM 9 model, revealing that the y/W had a striking impact to simulate the Fh/γs.W3.

As shown in figure 10-4, the SaE-ELM 10 to SaE-ELM 17 models were just a function of one input factor comprising x/W, Ds/W, φ, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W, respectively. Regarding the performed sensitivity analysis, Lv/γs.W3, Lh/γs.W3, and x/W were respectively identified as the most significant inputs to model the horizontal reaction forces. Moreover, h'/W, V2/g.W, φ, and Ds/W were the fourth, fifth, sixth, and seventh influencing input factors to predict the Fh/γs.W3, rather the effect of attack angle parameter was partially insignificant.

The SaE-ELM 5 was identified as the superior model to estimate the horizontal reaction forces as a function of x/W, Ds/W, φ, α, Lh/γs.W3, Lv/γs.W3, and V2/g.W missing h'/W that was the fourth significant input factor in the sensitivity analysis. This emphasizes that the relationship between the input factors and output parameter may not be linear explicit. In other words, the best model may still exclude some of the secondary significant input parameters, even if a nonlinear machine learning (ML) algorithm is used for the simulation.

Therefore, SaE-ELM 5 was regarded as the superior SaE-ELM model to simulate the horizontal reaction forces, nevertheless, the components of vertical and horizontal loads possessed an overriding influence.

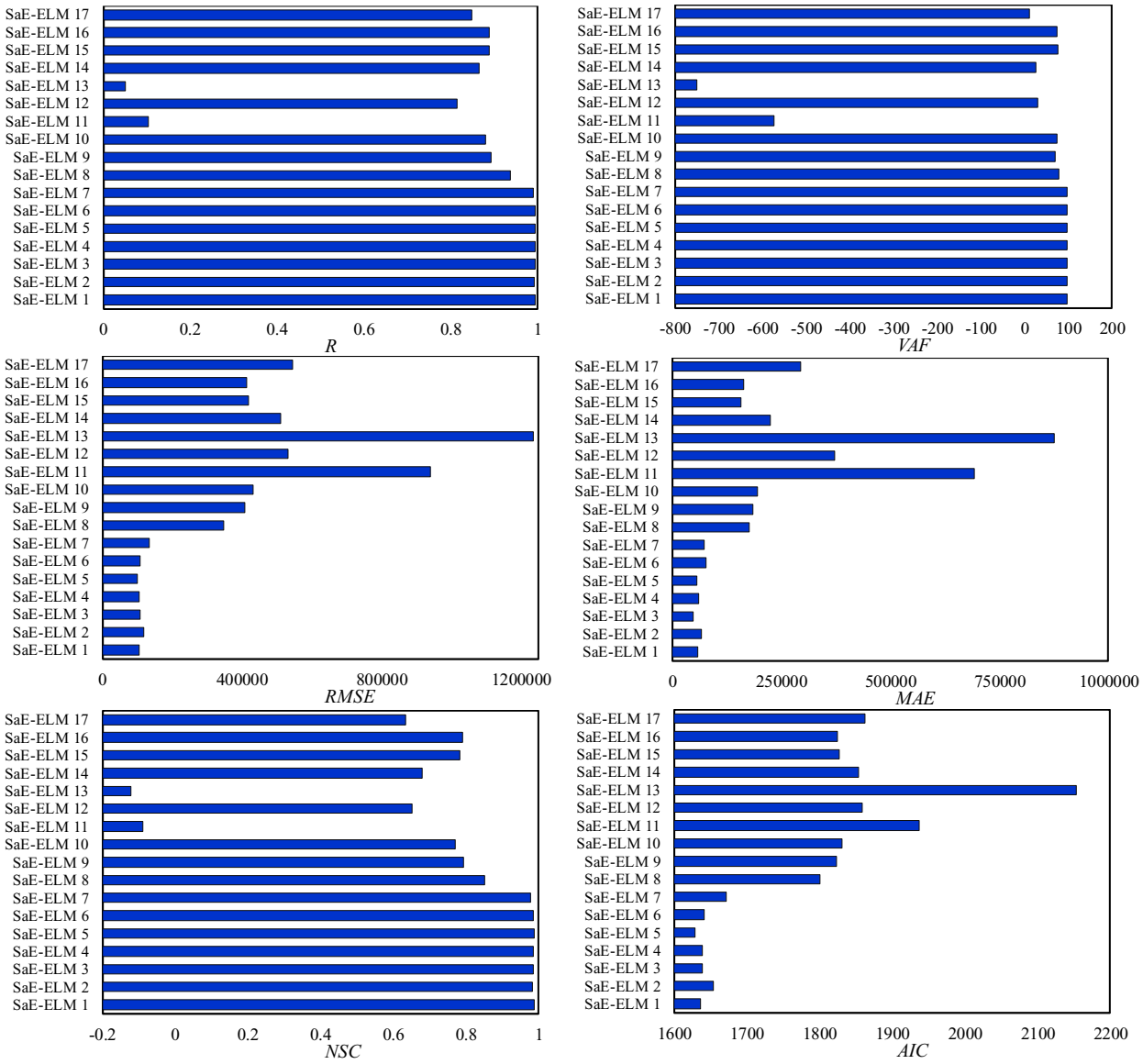


Figure 5-32. Comparison between the performances of SaE-ELM 1 to SaE-ELM 17 to simulate the horizontal reaction forces

Figure 5-33 shows the comparison between the performances of SaE-ELM models to simulate the vertical reaction forces. According to the simulation results, SaE-ELM 1 was distinguished as the third-best model to predict the vertical reaction forces. Furthermore, SaE-ELM 2 and Sa-ELM 3 were ranked as the fourth and seventh-premium models, respectively. The values of AIC for SaE-ELM 4 as the fifth-superior model was equal to 143.373. Additionally, SaE-ELM 5 and SaE-ELM 6 with the AIC index of 141.606 and 143.303 resulting in the second-best and sixth-best models. Among the SaE-ELM models to simulate the vertical reaction forces, SaE-ELM 7 was preferred as the best model. The AIC index for the eighth-best model (SaE-ELM 8) equaled to 143.490. Amongst the SaE-ELM models with seven input parameters, SaE-ELM 9 possessed the lowest level of precision and highest level of complexity with RMSE=64.196 and AIC=155.413.

On the other hand, the Lv/γs.W3, Lh/γs.W3, and h'/W parameters were recognized as the most effective input parameters to prognosticate the vertical reaction forces. Furthermore, V2/g.W, Ds/W, φ, and x/W were prioritized as the fourth, fifth and sixth-influencing input variables, whilst the implemented sensitivity analysis proved that the impact of ice keel attack angle was quite trivial.

Therefore, to approximate the vertical reaction forces, SaE-ELM 7 outperformed other SaE-ELM models and the Lv/γs.W3, Lh/γs.W3 inputs represented the highest level of effectiveness.

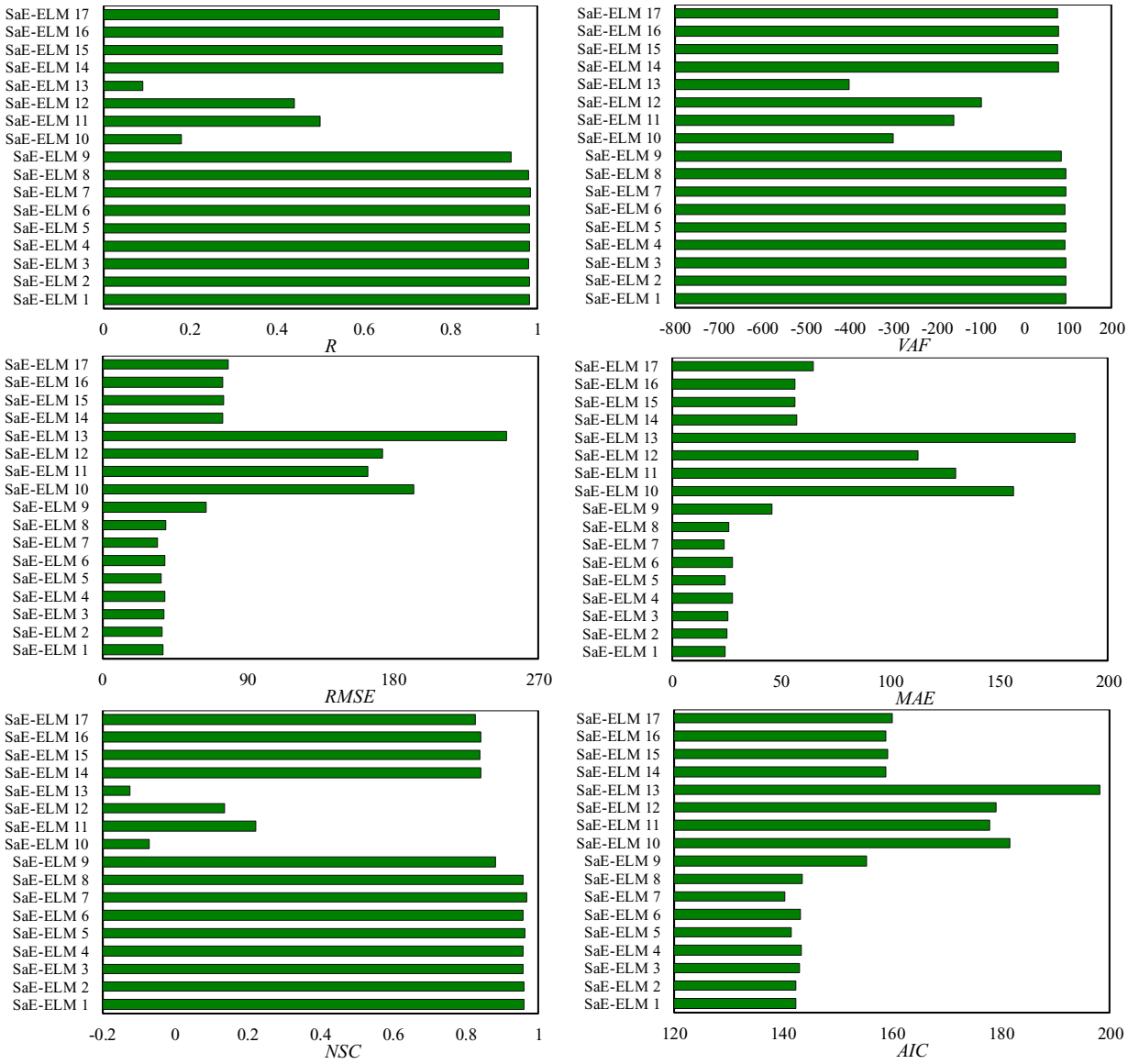


Figure 5-33. Comparison between the performances of SaE-ELM 1 to SaE-ELM 17 to simulate the vertical reaction forces

Figure 5-34 compares the performances of SaE-ELM models to estimate the horizontal ice-intrusion displacements. Regarding the carried out analyses, SaE-ELM 1 scored the lowest level of complexity (AIC=-2.860) and inaccuracy (RMSE=0.056) along with the highest level of correlation (R=0.992). After the SaE-ELM 1 model, SaE-ELM 7 and SaE-ELM 8, and SaE-ELM

5 were preferred as the second (AIC=15.456), third (AIC=16.006), and fourth-best (AIC=17.496) models to estimate the values of horizontal deformations. In addition, SaE-ELM 3 was regarded as the fifth-superior model. The R criterion for the sixth-superior model (SaE-ELM 4) were respectively surmised at 0.817. The computed Akaike Information Criteria (AIC) for the seventh-best (SaE-ELM 2) and eighth-best (SaE-ELM 6) models were estimated to be 17.742 and 17.698. The lowest level of precision (MAE=0.080) and correlation (R=0.491) as well as the highest level of complication (AIC=25.771) to model the horizontal deformations were obtained for SAE-ELM 9. Among the SaE-ELM models with one input parameter, SaE-ELM 10 was recognized as the premium model; as a result, the soil depth parameter (y/W) possessed the highest level of effectiveness to predict the horizontal subgouge displacements. After the soil depth variable, the velocity parameter (V2/g.W), the shear strength of sand seabed (φ), and the horizontal component of load (Lh/γs.W3) were ranked as the second, third, and fourth-effective input factors. In terms of significance, the h'/W, α, and Ds/W were put in the fifth, sixth, and seventh ranks, whereas the influence of the Lv/γs.W3 parameter was utterly negligible.

Thus, SaE-ELM 1 and soil depth (y/W) were detected as the best model and the most influencing input factor to estimate the horizontal ice-scoured displacements.

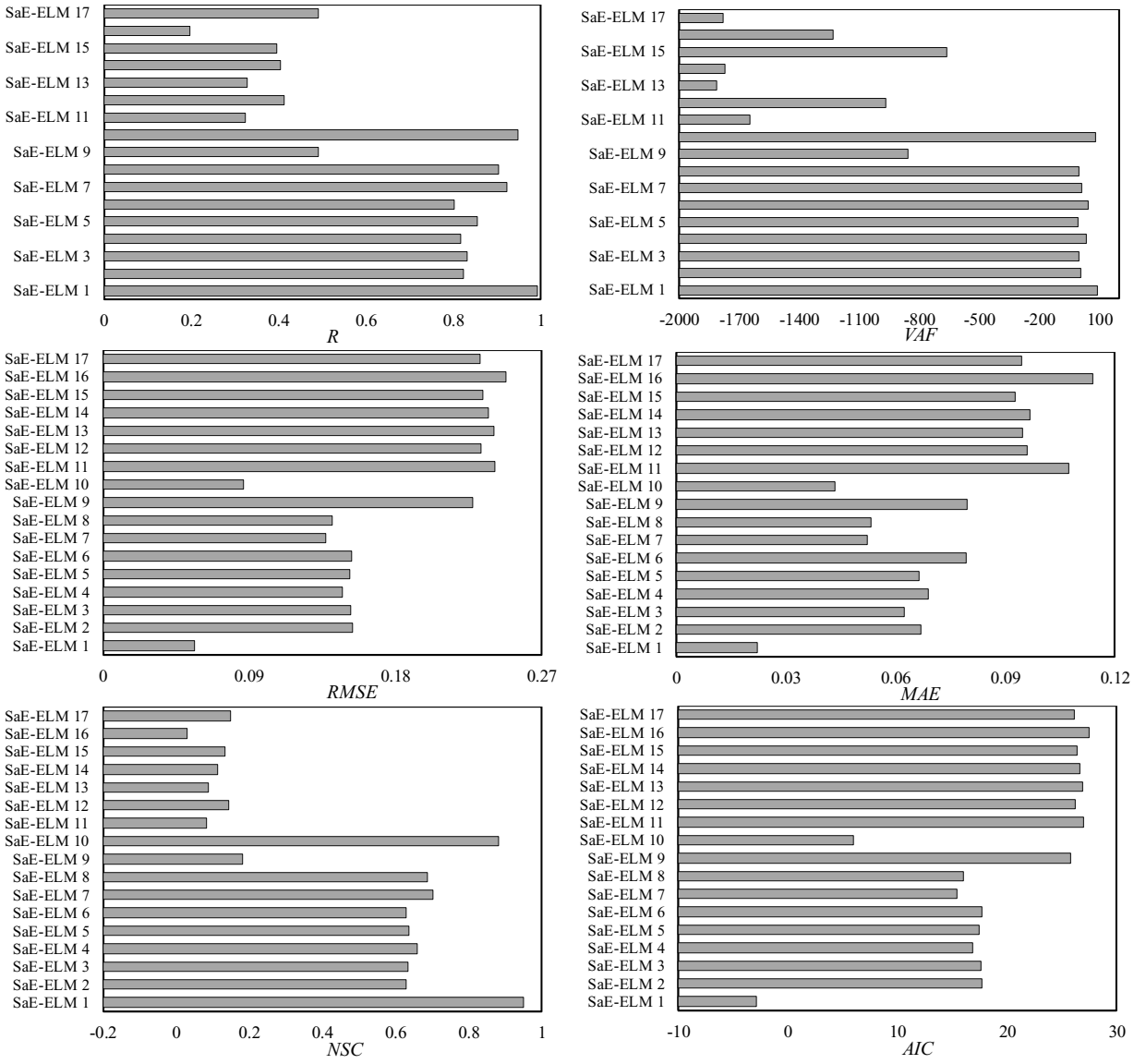


Figure 5-34. Comparison between the performances of SaE-ELM 1 to SaE-ELM 17 to simulate the horizontal displacements

Figure 5-35 shows the comparison between the SaE-ELM models to predict the vertical subgouge deformations. The performed sensitivity analysis indicated that SaE-ELM 1 owned an excellent performance to simulate the vertical deformations among these SaE-ELM models. The estimated value of AIC carrion for SaE-ELM 2 (the second-best model), SaE-ELM 3 (the fifth-best model),

and SaE-ELM 4 (the third-best model) were respectively approximated as 14.595, 15.403, 15.242. The SaE-ELM 5 performed as the sixth-premium model to foresee the vertical deformation, with a correlation coefficient of 0.376. The value of Nash-Sutcliff efficiency coefficient (NSC) index for the seventh-best (SaE-ELM 6) and eighth-best (SaE-ELM 9) models equaled to 0.068 and 16.789. The VAF index associated with the worst SaE-ELM model (SaE-ELM 7) was calculated to be -525.217. The made sensitivity analysis for SaE-ELM 10 to SaE-ELM 17 indicated that the h'/W parameter was the most important factor affecting to simulate the vertical displacements. After the h'/W factor, the shear strength parameter (φ), the horizontal load component (Lh/γs.W3), and the vertical load component (Lv/γs.W3) were graded as the second, third, and fourth-important input variables. Furthermore, the V2/g.W, y/W, and Ds/W were recognized as the fifth, sixth, and seventh-governing parameters. The ice keel attack angle (α) was proved to be as the less important input variable. Therefore, SaE-ELM 1 was introduced as the superior SaE-ELM model to forecast the vertical ice-scoured deformations, while the ratio of berm height to gouge width (h'/W) was a dominating input factor.

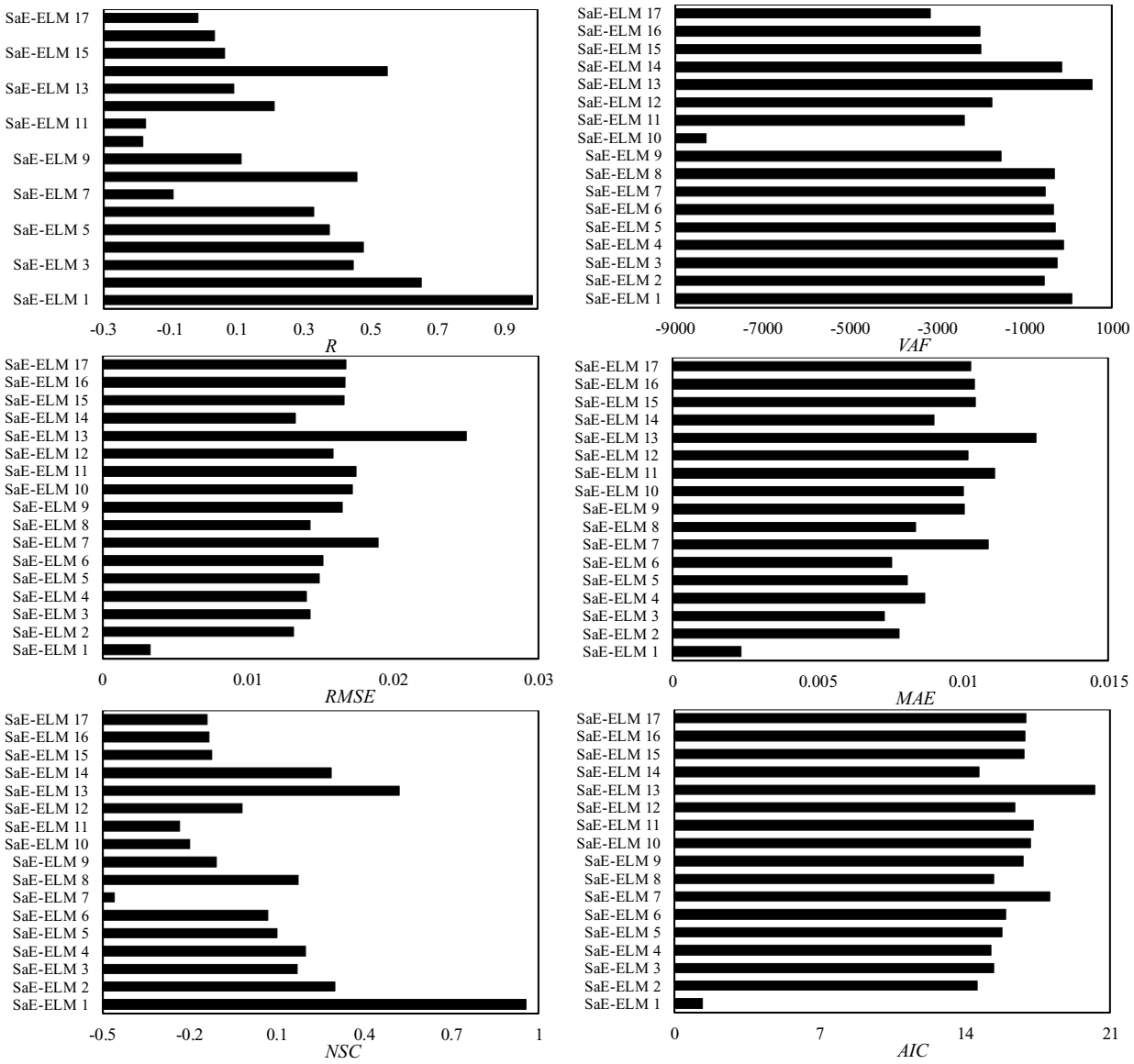


Figure 5-35. Comparison between the performances of SaE-ELM 1 to SaE-ELM 17 to simulate the vertical displacements

5.1.11.4. Premium SaE-ELM models

To estimate the subgouge soil parameters, the premium SaE-ELM models were detected and further evaluations associated with the performance of the best SaE-ELM models ought to be conducted. Figure 5-36 demonstrates the results of error analysis for the superior SaE-ELM

models. As shown, the overwhelming majority of the horizontal reaction forces predicted by SaE-ELM 5 possessed an error of greater than 20%, rather a small minority of the horizontal reaction forces had an error between 10% and 20%. Over half of the vertical reaction forces simulated by SaE-ELM 7 owned an error of less than 10% and approximately a quarter of the predicted vertical reaction forces had an error of more than 20%. Although about a third of the horizontal deformations modeled by SaE-ELM 1 indicated an error of less than 10%, this figure for the vertical deformation was almost 42%. The performed error analysis showed that roughly 27% and 28% of the simulated horizontal and vertical deformations had an error between 10% and 20%. Thus, the conducted error analysis proved that the superior SaE-ELM models predicted the ice-induced seabed parameters with a high level of precision.

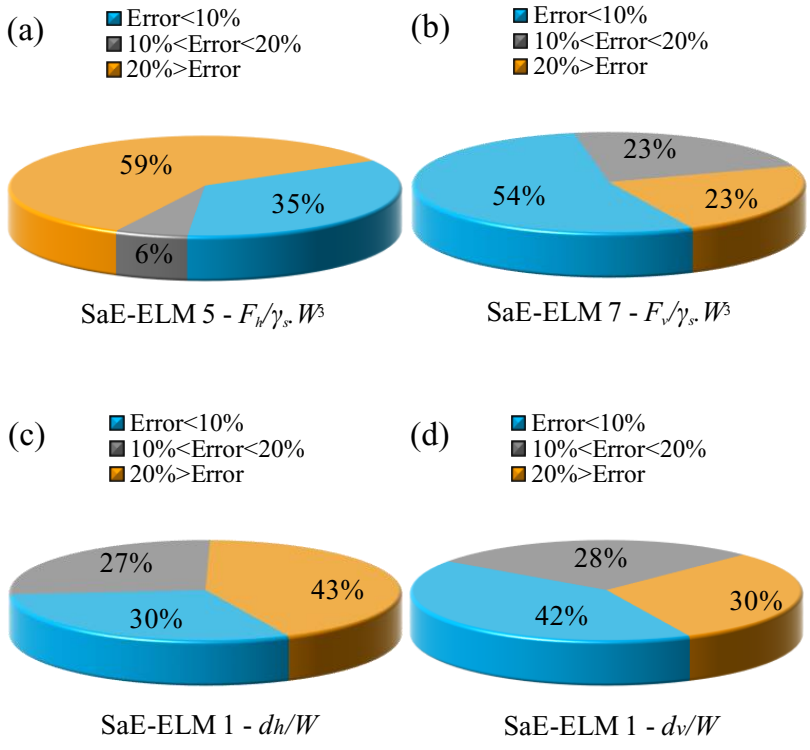


Figure 5-36. Results of error analysis (a) SaE-ELM 5 for horizontal reaction force (b) SaE-ELM 7 for vertical reaction force (c) SaE-ELM 1 for horizontal deformation (d) SaE-ELM 1 for vertical deformation

After that, the discrepancy ratio (*DR*) of the best SaE-ELM models was also calculated using the following equation:

$$DR = \frac{\eta_{(Simulated)}}{\eta_{(Observed)}} \tag{5-77}$$

here, $\eta_{(Simulated)}$ and $\eta_{(Observed)}$ are respectively the simulated and observed subgouge values. The *DR* compares the distance between the simulated and observed values, which is a ratio of the simulated ice-gouging parameter to the real data. Also, the accuracy of the SaE-ELM model can be determined by *DR*, where the magnitude of *DR* parameter approaching the unity indicates a better performance of the SaE-ELM model (Azimi and Shiri 2020b). The result of calculated *DR* for the premium SaE-ELM models are depicted in Figure 5-37. The minimum discrepancy ratio for SaE-ELM 5 and SaE-ELM 1 to model the horizontal reaction forces and vertical deformations were respectively estimated as 0.335 and 0.640, whereas the averaged DR for the horizontal and vertical subgouge displacements simulated by the SaE-ELM 1 model equaled to 1.669 and 1.309. So, regarding the discrepancy ratio analysis, the superior SaE-ELM models exhibited an acceptable performance to forecast the ice-scoured seabed characteristics.

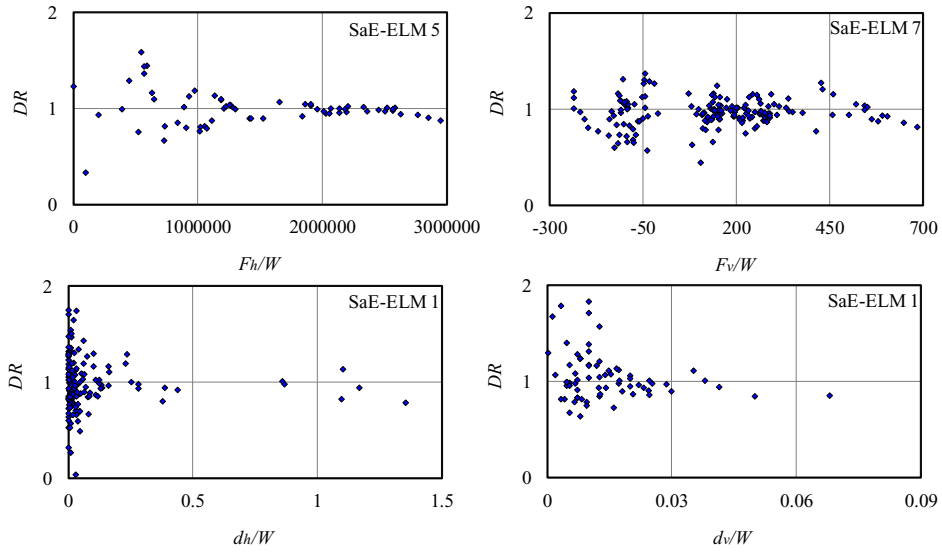


Figure 5-37. Results of discrepancy ratio (DR) for the premium SaE-ELM models

Subsequently, the uncertainty analysis (UA) was conducted for precise examination of the SaE-ELM model performance. Figure 5-38 details the results from the performed UA. The errors approximated by the SaE-ELM model (e_j) were computed as the difference between the predicted ice-gouging values (P_j) and the observed values (O_j) as below:

$$e_j = P_j - O_j \tag{5-78}$$

The averaged calculated errors $(\bar{e})$ and the standard deviation (S_e) of errors were estimated as follows (Azimi and Shiri 2020a):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{5-79}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{5-80}$$

A negative $\bar{e}$ means that the SaE-ELM model has underestimated the output parameter, showing that the simulated ice-induced parameter is less than the measured value. By contrast, a positive $\bar{e}$ proves that the predicted value is greater than the real value and the SaE-ELM model has

overestimated the target value. In the next step, regarding the "Wilson score method" without the continuity correction, the confidence bound was considered around the estimated error by using the $\overline{e}$ and S_e values (Azimi and Shiri 2020a). Note that the Wilson score interval is an improvement over the normal distribution interval where an asymmetric normal distribution is utilized so as to rectify the confidence interval bound. Moreover, a ±1.96Se brought about a 95% confidence interval (95%CI), and then the width of uncertainty bound (WUB) was calculated as follows:

$$\mathrm{WUB} = \pm \frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{5-81}$$

Even though the SaE-ELM 5 model underestimated the horizontal reaction forces, SaE-ELM 7 overestimated the vertical reaction forces. Besides, the SaE-ELM 1 model represented an overestimated and underestimated performance to simulate the horizontal and vertical deformations, respectively. The values of WUB to foresee the Fh/γs.W3, Fv/γs.W3, dh/W, and dv/W parameters by the superior SaE-ELM models equaled to -12214.5, -4.745, -0.006, and -0.001. As it could be seen, the narrowest width of uncertainty bound was obtained for SaE-ELM 1, rather SaE-ELM 5 was recognized as a model with the widest uncertainty bound.

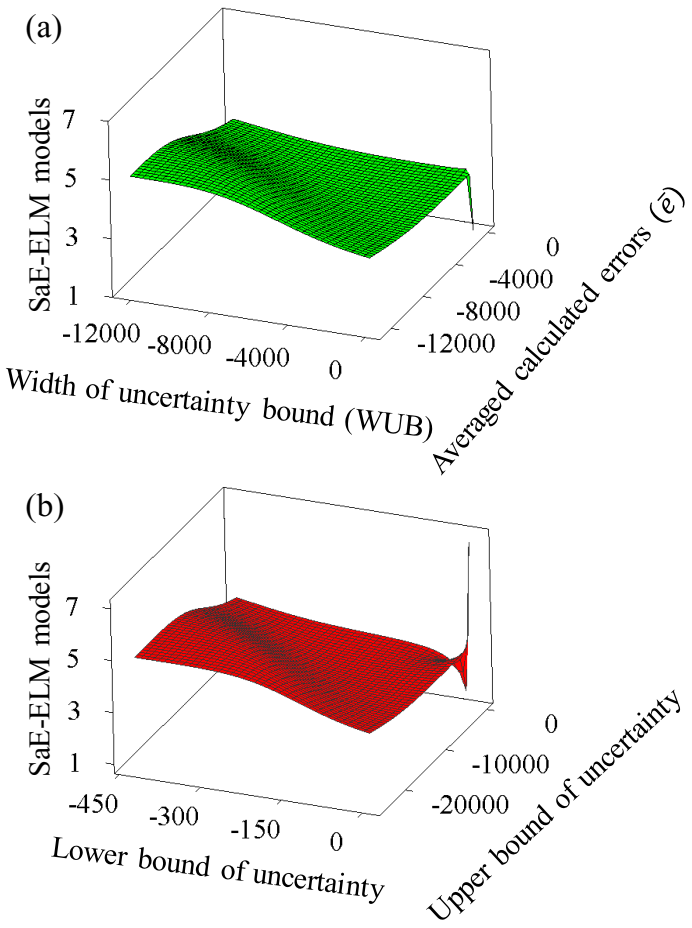


Figure 5-38. Results of uncertainty analysis (a) with of uncertainty bound (WUB) and averaged calculated error (b) lower and upper bound of uncertainty for the premium SaE-ELM models

Figure 5-39 illustrates the profiles of horizontal reaction forces simulated by SaE-ELM 5. As shown, a significant correlation was found between the experimental measurements and the results obtained from the SaE-ELM 5 model. This model was a function of x/W, Ds/W, φ, α, Lh/γs.W3, Lv/γs.W3, and V2/g.W, with a high level of precision and an underestimated performance. Regarding the Figures 15, the measured horizontal reaction forces had a conspicuous pattern, signifying that the minimum value of horizontal reaction forces were measured in initial position of the ice keel (x/W=0) and this parameter grew along the scour axis. Notwithstanding the fluctuant

trend in the experimental measurements, the SaE-ELM 5 model could predict the horizontal reaction forces by means of a nonlinear behavior. It means that this SaE-ELM model simulated appropriately the overall trend of horizontal reaction forces.

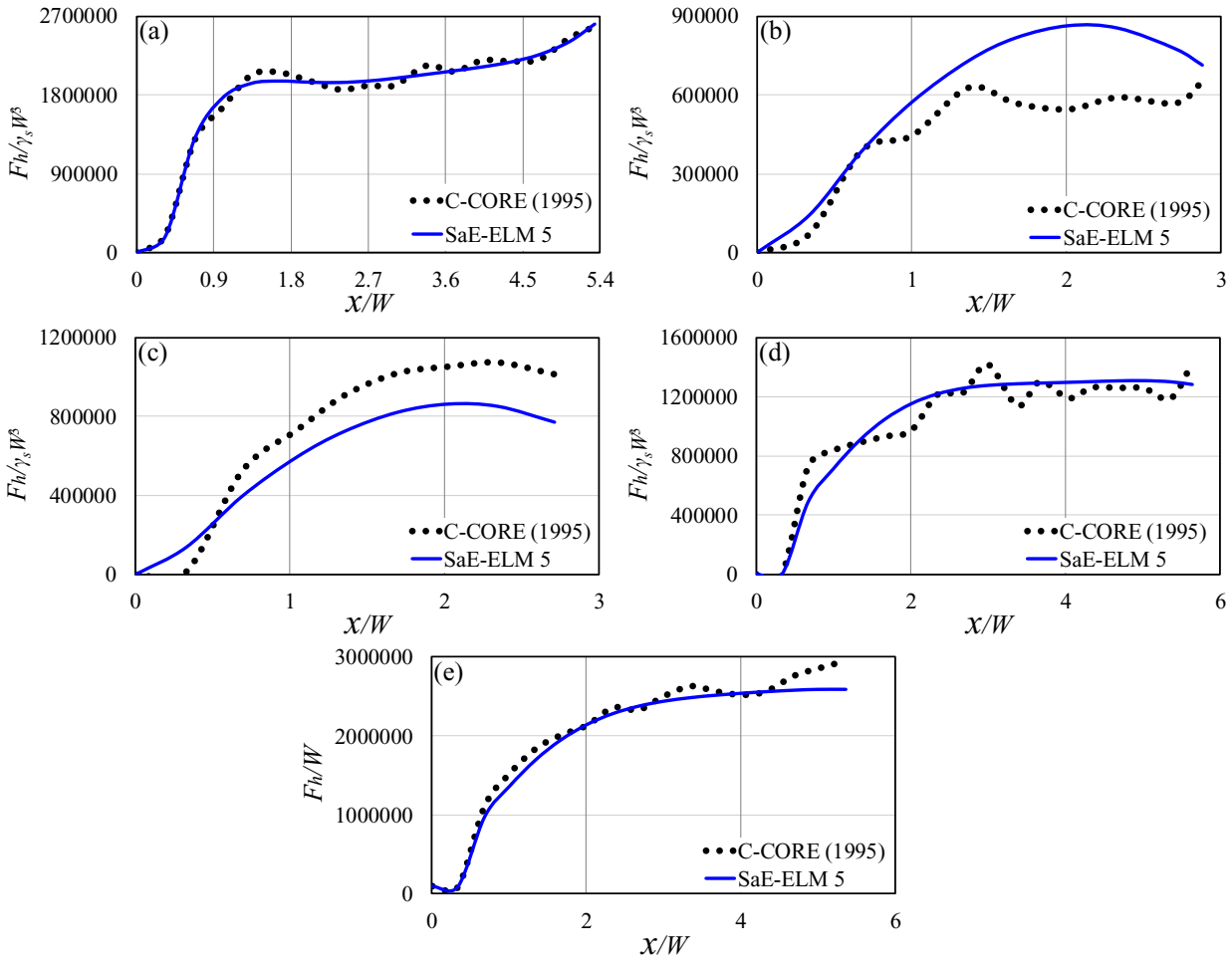


Figure 5-39. Comparison between the simulated horizontal reaction forces by SaE-ELM 5 with experimental values (a-e) C-CORE (1995)

Figure 5-40 compared the simulated vertical reaction forces by SaE-ELM 7 with experimental values. SaE-ELM 7 predicted the vertical reaction forces through the x/W, Ds/W, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W inputs, with a low level of inaccuracy and complexity. Regarding Figure 10-16, the magnitude of vertical reaction forces were reported as zero in initial

position of iceberg and this value changed along the scour axis. Although the vertical reaction forces witnessed severe fluctuations where the experimental measurements were recorded by positive and negative values, the SaE-ELM 7 model managed to simulate the vertical reaction forces with both monotonic linearity [Figures 5-40(c, f, g, i, j)] and nonlinearity patterns [Figures 10-16(a, b, d, e, h, k, l)].

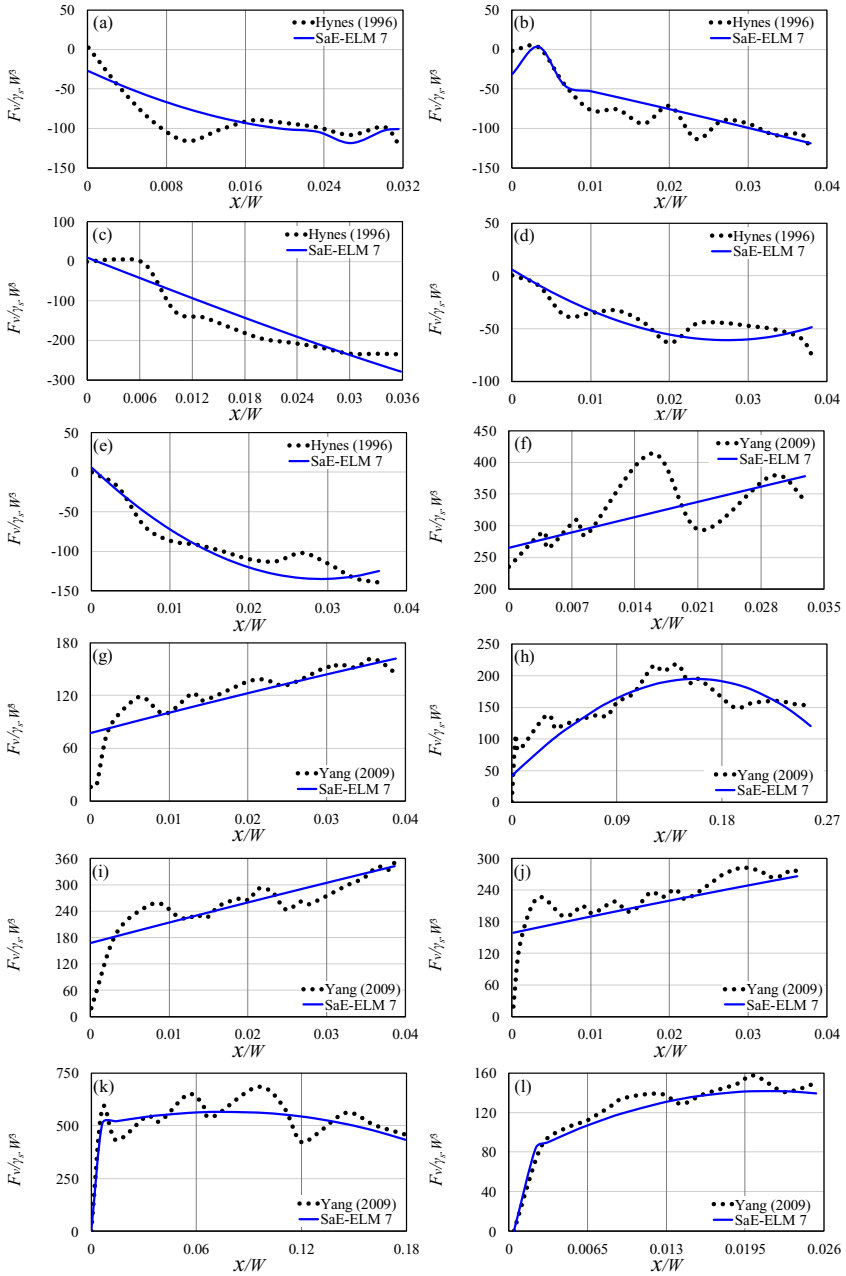


Figure 5-40. Comparison between the simulated vertical reaction forces by SaE-ELM 7 with experimental values (a-e) Hynes (1996) (f-l) Yang (2009)

The simulated horizontal deformations by the SaE-ELM 1 model with experimental measurements were compared in Figure 5-41. SaE-ELM 1 modeled the target values in terms of all inputs (y/W, Ds/W, φ, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W). This model overestimated the subgouge horizontal displacements, with a high level of correlation and accuracy along with a low level of complexity.

The horizontal subgouge deformations were measured as a hyperbolic curve where the maximum magnitude was observed below the keel iceberg and the horizontal soil displacements reduced at greater depth on the subgouge centerline. Despite some discrepancies between the experimental and numerical results, the SaE-ELM 1 model was able to predict the horizontal soil deformations with an appropriate performance in different experimental conditions, ranging from 1g laboratory circumstance [Figures 5-41(a-e)] to centrifuge tests [Figures 5-41(f-x)].

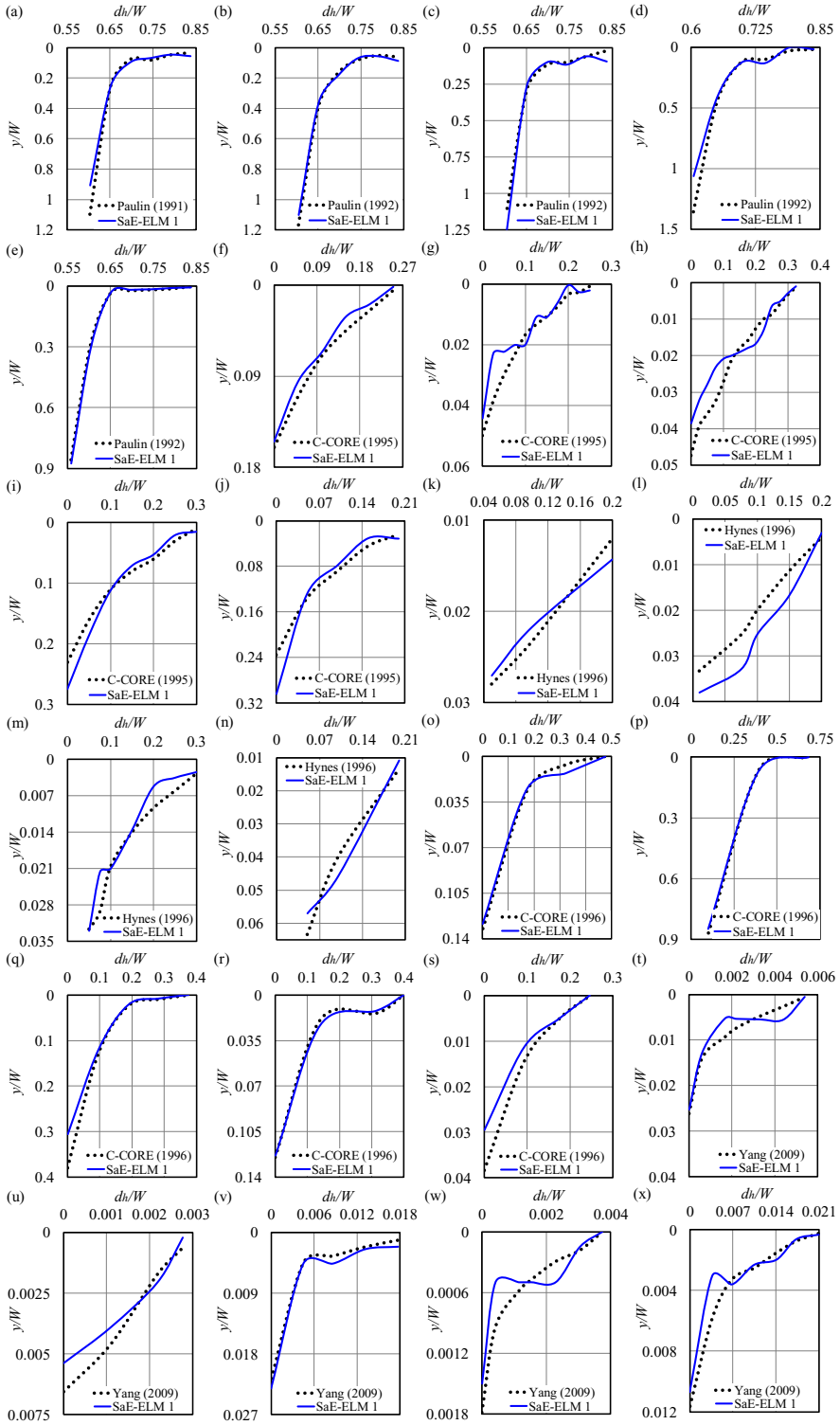


Figure 5-41. Comparison between the simulated horizontal deformations by SaE-ELM 1 with experimental values (a) Paulin (1991) (b-e) Paulin (1992) (f-j) C-CORE (1995) (k-n) Hynes (1996) (o-s) C-CORE (1996) (t-x) Yang (2009)

In Figure 5-42, the comparison between the modeled vertical deformations by SaE-ELM 1 with observed values is highlighted. As aforementioned, the key inputs comprising the y/W, Ds/W, φ, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W factors were applied in order for anticipation of the vertical deformations by using SaE-ELM 1. This model possessed an underestimated performance and demonstrated an acceptable precision with a low level of complicacy.

The accuracy of laboratory measurements has originated from two main sources comprising systematic error (equipment error) and random error (human error). The source of the systematic error can be detected and value of this error is measurable, whereas identification of random error source is quite laborious. In order to prevent the random error, the experimental measurements had been repeated and their average values were reported as the laboratory records. It seems that the experimental vertical displacements lacked a meaningful trend and several oscillations were reported for this parameter but the SaE-ELM 1 model could properly simulate the vertical subgouge deformations with both linear [Figures 5-42(b, f)] and nonlinear [Figures 5-42(a, c, d, e)] behaviors.

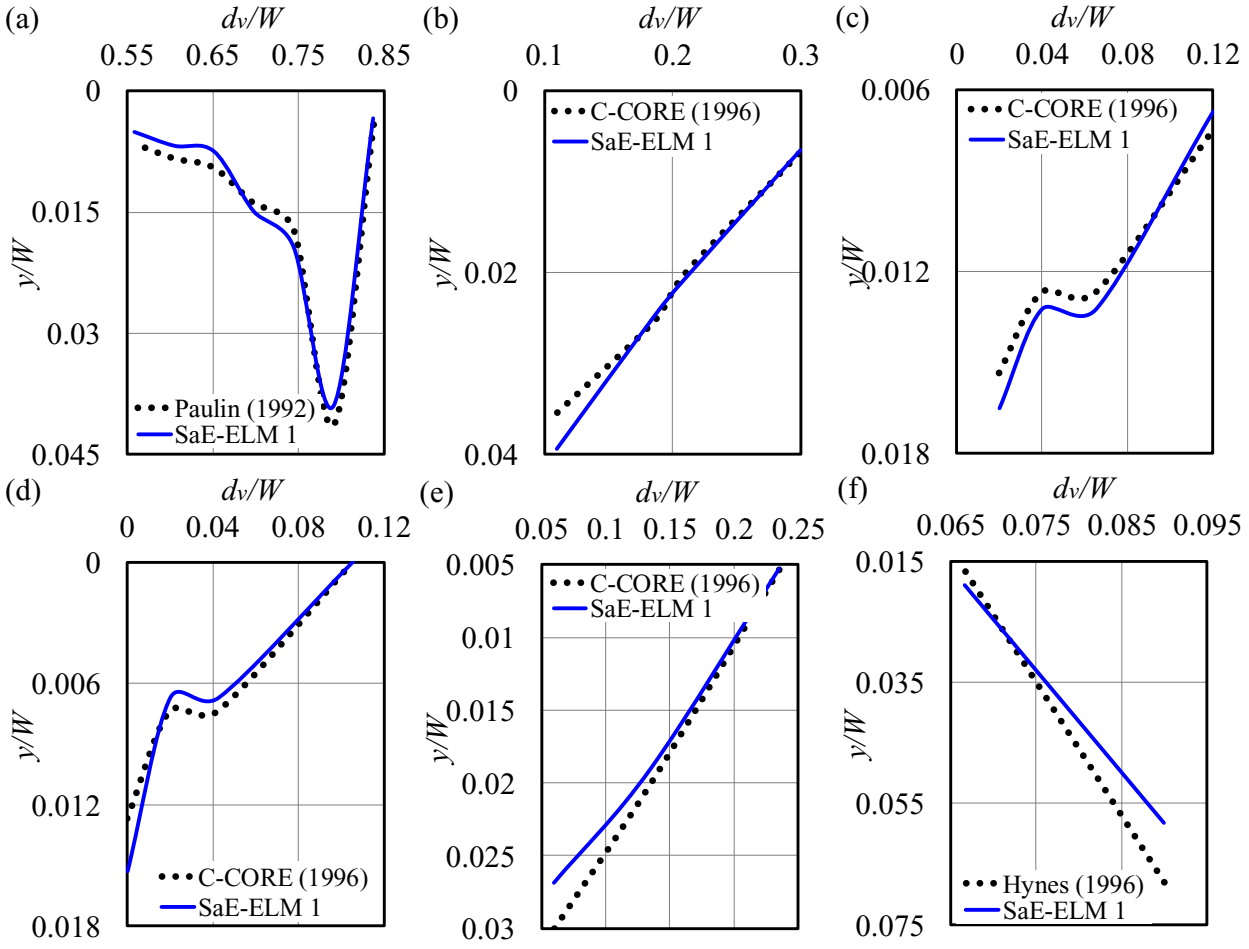


Figure 5-42. Comparison between the simulated vertical deformations by SaE-ELM 1 with experimental values (a) Paulin (1992) (b-e) C-CORE (1996) (f) Hynes (1996)

Therefore, the SaE-ELM methodology could simulate the ice-induced seabed parameters in sand with an acceptable precision. The best SaE-ELM models represented a low level of complexity and a high level of correlation with observed values.

5.1.12. Conclusion

In this study, for the first time, a self-adaptive evolutionary machine learning algorithm entitled the self-adaptive evolutionary extreme learning machine (SaE-ELM) was applied to simulate the ice-induced scour in the sandy seabed. Initially, the factors affecting the ice-seabed interaction phenomenon were introduced and 17 SaE-ELM models were defined by using these parameters. After that, the number of neurons in the hidden layer (NHN) was optimized and then the most accurate activation function was selected for the SaE-ELM network. Subsequently, sensitivity analysis, error analysis, and uncertainty analysis were conducted for the SaE-ELM models. The most important obtained results are listed below:

- The SaE-ELM was found to be a cost-effective and robust alternative for simulation of the ice-seabed interaction, particularly for the early stages of the pipeline design projects where fast estimations are required with an acceptable level of accuracies.
- A set of the SaE-ELM-based equations were proposed to approximate the ice-scoured seabed parameters in practical design projects.
- The implemented sensitivity analysis demonstrated that SaE-ELM 1 as a function of all input factor was the premium model to predict the ice-scoured seabed deformations, whilst SaE-ELM 5 and SaE-ELM 7 had the lowest level of complexity along with highest level of correlation with the experimental values for simulation of the reaction forces.
- The optimum NHN was adopted at 28, while the sigmoid function was the best activation function to model the subgouge seabed parameters using the SaE-ELM network.
- Though the vertical component of the load ($L_v/\gamma_s.W^3$) possessed the highest level of effectiveness to estimate the reaction forces, the soil depth (y/W) and berm height (h'/W) parameters were distinguished to be the most important input factors to model the subgouge displacements.

- The performed error analysis showed that about 54% of the vertical reaction forces predicted by SaE-ELM 7 had an error of less than 10% and this figure for SaE-ELM 5 was almost 35%.
- SaE-ELM 1 showed an overestimated performance dealing with the horizontal deformations; however, this model underestimated the vertical subgouge displacements.

This study was the first step towards evaluating the capability of a self-adaptive machine learning to enhance our understanding of the ice-gouging phenomenon. The SaE-ELM algorithm showed a better generalization performance at dealing with simulation of target parameter overcoming the sensitivity to the number of neurons in the hidden layer, but the computation time of this model was slightly longer than the classical version of ELM algorithm. The present findings can support providing fast and cost-effective alternatives for simulation of ice-seabed interaction but more investigations still need to be conducted using other optimization tools and machine learning algorithms to identify the most efficient and robust approach.

Appendix I:

SaE-ELM-based matrices

SaE-ELM 5, SaE-ELM 7, and SaE-ELM 1 were respectively selected as the superior models so as to estimate the horizontal reaction forces ($Fh/\gamma s.W^3$), the vertical reaction forces ($Fv/\gamma s.W^3$), the horizontal deformations (dh/W), and the vertical deformations (dv/W). It is worth noting that the general format of the SaE-ELM-based equation is as below:

$$\eta = \left[\frac{1}{1 + exp(InW \times InV \times BHI)}\right]^T \times OutW \tag{5-82}$$

where, η is the ice-seabed interaction parameter ($F/\gamma s W^3$ *or* d/W), InW is the matrix of input weight, InV is the matrix of input variables, BHI is the matrix of bias of hidden layer, and $OutW$

is the matrix of output weights. Matrices of the SaE-ELM 5 model to calculate the horizontal reaction forces were provided below:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_S}{W} \\ \varphi \\ \alpha \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, \; BHI = \begin{bmatrix} 0.201 \\ 0.466 \\ 0.095 \\ 0.962 \\ 0.735 \\ 0.058 \\ -0.107 \\ 0.411 \\ 0.885 \\ 0.355 \\ 0.330 \\ 0.845 \\ 0.253 \\ 0.696 \\ 0.015 \\ 0.988 \\ 0.161 \\ 0.089 \\ 0.095 \\ 0.637 \\ 0.352 \\ 0.787 \\ 0.582 \\ -0.197 \\ -0.107 \\ 0.491 \\ 0.134 \\ 0.564 \end{bmatrix}, InW = \begin{bmatrix} -0.521 & 0.838 & 0.130 & -0.035 & 0.450 & 0.877 & 0.419 \\ 0.663 & 0.668 & 0.355 & 0.003 & -0.781 & 0.523 & 0.036 \\ 0.935 & 0.111 & 0.073 & 0.484 & 0.294 & 0.841 & 0.589 \\ -0.687 & 0.771 & 0.484 & 0.368 & 0.478 & 0.027 & 0.326 \\ 0.149 & 0.397 & 0.401 & 0.090 & 0.789 & -0.212 & 0.525 \\ 0.361 & 0.410 & 0.536 & -0.519 & 0.385 & 0.747 & 0.787 \\ 0.538 & 0.733 & 0.336 & 0.433 & 0.164 & 1.031 & 0.584 \\ 0.219 & 0.222 & 0.178 & 0.697 & 0.532 & 0.108 & 0.587 \\ -0.868 & 0.762 & -0.419 & -0.039 & -0.616 & 0.835 & 0.924 \\ 0.052 & 0.796 & 0.105 & 0.813 & -0.680 & 0.689 & 0.500 \\ 0.583 & -0.471 & 0.933 & 0.122 & -0.768 & -0.306 & 0.448 \\ 0.159 & -0.284 & 0.666 & 0.410 & 0.025 & 0.475 & 0.270 \\ 0.814 & 0.089 & 0.865 & 0.789 & 0.284 & -0.454 & -0.106 \\ -0.645 & 0.503 & 0.146 & 0.482 & 0.841 & 0.828 & 0.385 \\ 0.918 & 0.146 & 0.270 & 0.325 & 0.025 & 0.163 & 0.613 \\ 0.594 & 0.148 & 0.044 & 0.789 & 0.588 & 0.983 & 0.651 \\ 0.235 & 0.188 & 0.356 & -0.687 & 0.565 & -0.367 & 0.607 \\ 0.276 & 0.373 & -0.905 & 0.610 & -0.159 & 0.475 & 0.811 \\ 0.837 & 0.470 & -0.003 & -0.596 & 0.690 & -0.034 & -0.096 \\ 0.434 & 0.632 & 0.877 & 0.358 & 0.388 & 0.181 & 1.272 \\ 0.529 & -0.040 & 0.257 & 0.566 & -0.586 & 0.410 & 0.514 \\ 0.803 & 0.921 & 0.416 & 0.835 & 0.573 & 0.842 & -0.044 \\ 0.662 & -0.407 & -0.068 & -0.377 & -0.494 & 0.423 & -0.902 \\ 0.310 & -0.599 & 0.399 & 0.549 & 0.718 & 0.282 & 0.098 \\ -0.310 & 0.735 & 0.640 & -0.445 & -0.987 & 0.750 & 0.751 \\ -0.206 & 0.001 & 0.241 & 0.295 & -0.545 & 0.301 & 0.415 \\ 0.240 & 0.426 & 0.134 & 0.673 & 0.585 & 0.348 & -0.084 \\ 0.621 & -0.592 & 0.108 & -0.976 & 0.523 & -0.295 & 0.146 \end{bmatrix} \quad (5\text{-}83)$$

$$OutW = \begin{bmatrix} 28215.598 \\ 49118158019.649 \\ -625826814413.669 \\ -6353322556826090 \\ -14225295030822.5 \\ 5040932.1014631 \\ -167095253419015 \\ 2503010096988850 \\ -56910687024.4821 \\ 839441093095939 \\ -2926907161308550 \\ -2862808413664180 \\ -2932341606220150 \\ 2026662830725.44 \\ 24263621253068.7 \\ 138999255982638 \\ 13985926.7575799 \\ 2535924543.34386 \\ 2453850378.24901 \\ -2932300664626640 \\ -1336509355744550 \\ -2914764047124080 \\ -5920820452.55523 \\ 19492217685272700 \\ 243284955.180934 \\ -5893042007554.72 \\ -553214559215756 \\ 227307328116.784 \end{bmatrix}$$

For SaE-ELM 7, the matrices to predict the vertical reaction forces were suggested as follows:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_s}{W} \\ \alpha \\ \frac{h'}{W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, \; BHI = \begin{bmatrix} 0.889 \\ -0.167 \\ 0.871 \\ 0.996 \\ 0.938 \\ -0.179 \\ -0.208 \\ -0.361 \\ -0.579 \\ -0.946 \\ 0.626 \\ -0.166 \\ 0.598 \\ 0.898 \\ -0.099 \\ -0.423 \\ -0.805 \\ -0.649 \\ 0.369 \\ 0.753 \\ -0.958 \\ 0.993 \\ 0.449 \\ -0.313 \\ 0.334 \\ 0.276 \\ 0.780 \\ 0.116 \end{bmatrix}, InW = \quad (5\text{-}84)$$

$$\begin{bmatrix}
-0.386 & 0.270 & 0.822 & -0.478 & -0.556 & 0.653 & -0.553 \\
-0.916 & -0.863 & 0.716 & 0.862 & 0.549 & -0.564 & 0.733 \\
0.816 & 0.278 & -0.068 & -0.349 & 0.458 & 0.705 & -0.292 \\
-0.028 & 0.899 & -0.631 & 0.367 & 0.275 & -0.335 & 0.298 \\
0.983 & -0.694 & 0.428 & 0.039 & -0.325 & 0.145 & 0.793 \\
0.434 & -0.352 & -0.967 & 0.622 & -0.790 & -0.792 & -0.394 \\
-0.675 & 0.126 & 0.779 & -0.896 & -0.849 & -0.315 & -0.584 \\
-0.118 & 0.171 & -0.288 & 0.564 & 0.082 & 0.916 & -0.424 \\
-0.830 & -0.398 & 0.688 & 0.764 & 0.703 & -0.488 & -0.321 \\
-0.226 & -0.288 & -0.366 & -0.060 & 0.210 & 0.254 & 0.215 \\
0.749 & -0.503 & 0.203 & 0.103 & -0.228 & 0.194 & -0.291 \\
-0.079 & 0.149 & 0.728 & -0.227 & 0.947 & -0.084 & 0.258 \\
0.966 & -0.191 & -0.443 & -0.088 & 0.995 & 0.421 & -0.141 \\
-0.623 & 0.919 & 0.585 & -0.162 & 0.234 & -0.639 & -0.796 \\
0.417 & 0.022 & 0.595 & -0.858 & -0.459 & -0.842 & -0.467 \\
0.490 & 0.981 & -0.189 & 0.792 & -0.505 & 0.400 & -0.668 \\
-0.137 & -0.546 & -0.933 & 0.793 & -0.300 & 0.549 & -0.841 \\
0.664 & -0.742 & 0.767 & -0.648 & -0.488 & 0.113 & 0.097 \\
0.612 & -0.410 & 0.622 & 0.505 & -0.908 & -0.548 & 0.516 \\
-0.154 & -0.346 & -0.936 & -0.800 & 0.268 & -0.945 & -0.360 \\
-0.715 & 0.676 & 0.365 & -0.215 & 0.294 & -0.579 & 0.256 \\
0.850 & -0.134 & 0.416 & -0.874 & -0.807 & 0.793 & -0.180 \\
0.494 & 0.726 & 0.087 & 0.988 & -0.970 & 0.254 & -0.665 \\
0.121 & -0.887 & 0.085 & 0.053 & 0.323 & 0.599 & -0.898 \\
0.056 & -0.634 & -0.453 & 0.876 & 0.573 & -0.968 & 0.299 \\
0.863 & -0.950 & 0.107 & -0.353 & 0.419 & 0.222 & -0.171 \\
-0.643 & 0.306 & 0.017 & -0.876 & 0.140 & -0.268 & 0.712 \\
0.426 & 0.634 & -0.696 & -0.669 & 0.796 & -0.018 & -0.522
\end{bmatrix}$$

$$OutW = \begin{bmatrix} -1872648442408.43 \\ 443032259289.844 \\ 14700.3467612937 \\ 16866269825.812 \\ -10461575827.599 \\ -25023424110740.5 \\ 380359024714.12 \\ 25166709.063 \\ -267990850064.857 \\ -2481091752.490 \\ 68785463.057 \\ 947274762390.157 \\ -205171386.997 \\ 115713376933.871 \\ 66506525782.960 \\ 3427836.565 \\ 3235981911222.91 \\ 316660180663.925 \\ -118168125102.645 \\ -6189763479086.53 \\ -308279882.112 \\ -36784665.632 \\ -388448.092 \\ 624276.579 \\ 20425622883.656 \\ -1321923.470 \\ 333103.014 \\ 363439040980.949 \end{bmatrix}$$

To compute the horizontal subgouge deformations by using the SaE-ELM 1 model, the proposed matrices were as below:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_S}{W} \\ \varphi \\ \alpha \\ \frac{h'}{W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, \; BHI = \begin{bmatrix} 0.334 \\ 0.706 \\ 0.079 \\ 0.244 \\ 0.680 \\ 0.476 \\ 0.365 \\ 0.414 \\ 0.712 \\ 0.622 \\ 0.728 \\ 0.745 \\ 0.537 \\ 0.330 \\ 0.244 \\ 0.345 \\ 0.693 \\ 0.687 \\ 0.605 \\ 0.399 \\ 0.376 \\ 0.470 \\ 0.497 \\ 0.091 \\ 0.168 \\ 0.402 \\ 0.469 \\ 0.517 \end{bmatrix}, InW = \quad (5\text{-}85)$$

$$\begin{bmatrix}
0.199 & 0.704 & 0.350 & 0.362 & 0.493 & 0.454 & 0.835 & 0.390 \\
0.021 & 0.197 & 0.427 & 0.526 & 0.365 & 0.655 & 0.349 & 0.638 \\
0.533 & 0.212 & 0.433 & 0.941 & 0.675 & 0.163 & 0.539 & 0.729 \\
0.918 & 0.313 & 0.092 & 0.444 & 0.102 & 0.593 & 0.304 & 0.670 \\
0.213 & 0.035 & 0.361 & 0.722 & 0.027 & 0.066 & 0.080 & 0.658 \\
0.348 & 0.557 & 0.791 & 1.052 & 0.566 & 0.535 & 0.290 & 0.638 \\
0.277 & 0.462 & 0.136 & 0.758 & 0.617 & 0.612 & 0.036 & 0.462 \\
0.959 & 0.262 & 0.483 & 0.148 & 0.105 & 0.617 & 1.013 & 0.226 \\
0.933 & 0.571 & 0.193 & 1.049 & 1.117 & 0.813 & 0.285 & 0.257 \\
0.563 & 0.161 & 0.372 & 0.759 & 0.187 & 0.181 & 0.074 & 0.704 \\
0.668 & 0.797 & 0.569 & 0.436 & 0.429 & 0.801 & 0.295 & 0.373 \\
0.784 & 0.792 & 0.458 & 0.770 & 0.171 & 0.697 & 0.470 & 0.801 \\
0.982 & 0.224 & 1.058 & 0.738 & 0.863 & 0.004 & 0.497 & 0.173 \\
0.206 & 0.897 & 0.126 & 0.600 & 0.712 & 0.577 & 0.528 & 0.979 \\
0.078 & 0.136 & 0.346 & 0.351 & 0.402 & 0.047 & 0.788 & 0.840 \\
0.622 & 0.508 & 0.655 & 0.824 & 0.815 & 0.619 & 0.581 & 0.831 \\
0.421 & 1.162 & 0.829 & 0.852 & 0.649 & 0.123 & 1.089 & 0.009 \\
0.112 & 0.0001 & 0.404 & 0.638 & 0.490 & 0.616 & 0.034 & 0.076 \\
0.503 & 0.502 & 0.104 & 0.650 & 0.876 & 0.664 & 1.146 & 0.964 \\
0.663 & 0.949 & 0.426 & 0.716 & 0.221 & 0.404 & 0.655 & 0.587 \\
0.280 & 0.137 & 0.178 & 0.954 & 0.185 & 0.367 & 0.744 & 0.002 \\
0.185 & 0.537 & 0.276 & 0.612 & 0.397 & 0.155 & 0.694 & 0.480 \\
0.029 & 0.552 & 0.179 & 0.856 & 0.726 & 0.983 & 0.917 & 0.260 \\
0.003 & 0.274 & 0.898 & 0.622 & 0.340 & 0.727 & 0.086 & 0.553 \\
0.342 & 0.733 & 0.247 & 0.425 & 0.272 & 0.043 & 0.345 & 0.671 \\
0.782 & 0.961 & 0.915 & 0.046 & 0.347 & 0.613 & 0.593 & 0.684 \\
0.414 & 0.205 & 0.416 & 1.127 & 0.119 & 0.556 & 0.374 & 0.230 \\
0.665 & 0.171 & 1.022 & 0.739 & 0.354 & 0.560 & 0.548 & 0.828
\end{bmatrix}$$

$$OutW = \begin{bmatrix} 150861655.981 \\ 731354216.280 \\ 4.934 \\ 116444.748 \\ 56.468 \\ 10296137831.450 \\ 203183265.333 \\ 1693.236 \\ 54006255729.879 \\ 60.008 \\ 4.250 \\ 1.338 \\ 10295957606.547 \\ 116.042 \\ 133651705.737 \\ 10245905286.115 \\ 10295911748.938 \\ 13898001832.974 \\ 6599893.098 \\ 2977907325.429 \\ 8010461965.166 \\ 13141833.297 \\ 581717767.903 \\ 10279887032.661 \\ 10659119.639 \\ 5126627993.304 \\ 2.437 \\ 10297342423.233 \end{bmatrix}$$

And for estimation of the vertical ice-induced deformations, the SaE-ELM-based matrices are written as follows:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_S}{W} \\ \varphi \\ \alpha \\ \frac{h'}{W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, \; BHI = \begin{bmatrix} 0.886 \\ 0.374 \\ 0.267 \\ 0.028 \\ 0.514 \\ 0.021 \\ 0.269 \\ 0.242 \\ 0.424 \\ 0.591 \\ 0.119 \\ 0.522 \\ 0.016 \\ 0.749 \\ 0.669 \\ 0.080 \\ 0.349 \\ 0.031 \\ 0.094 \\ 0.791 \\ 0.545 \\ 0.902 \\ 0.469 \\ 0.009 \\ 0.061 \\ 0.775 \\ 0.712 \\ 0.744 \end{bmatrix}, InW = \quad (5\text{-}86)$$

$$\begin{bmatrix}
0.008 & 0.047 & 0.395 & 0.215 & 0.200 & 0.321 & 0.889 & 0.014 \\
0.343 & 0.999 & 0.160 & 0.127 & 0.042 & 0.746 & 0.574 & 0.028 \\
0.002 & 0.009 & 0.588 & 0.294 & 0.216 & 0.508 & 0.744 & 0.417 \\
0.014 & 0.987 & 0.204 & 0.015 & 0.559 & 0.738 & 0.952 & 0.525 \\
0.574 & 0.014 & 0.356 & 0.858 & 0.610 & 0.885 & 0.598 & 0.004 \\
0.563 & 0.301 & 0.247 & 0.751 & 0.209 & 0.810 & 0.085 & 0.303 \\
0.503 & 0.003 & 0.386 & 0.329 & 0.490 & 0.038 & 0.659 & 0.430 \\
0.807 & 0.427 & 0.461 & 0.281 & 0.864 & 0.412 & 0.364 & 0.686 \\
0.828 & 0.126 & 0.321 & 0.546 & 0.912 & 0.554 & 0.022 & 0.935 \\
0.268 & 0.001 & 0.123 & 0.180 & 0.516 & 0.180 & 0.008 & 0.155 \\
0.381 & 0.271 & 0.389 & 0.788 & 0.976 & 0.428 & 0.384 & 0.013 \\
0.667 & 0.939 & 0.022 & 0.879 & 0.635 & 0.473 & 0.033 & 0.079 \\
0.747 & 0.984 & 0.341 & 0.607 & 0.327 & 0.406 & 0.699 & 0.033 \\
0.864 & 0.530 & 0.219 & 0.498 & 0.017 & 0.034 & 0.942 & 0.507 \\
0.003 & 0.811 & 0.892 & 0.802 & 0.292 & 0.988 & 0.774 & 0.476 \\
0.022 & 0.651 & 0.861 & 0.238 & 0.958 & 0.889 & 0.149 & 0.821 \\
0.014 & 0.625 & 0.176 & 0.144 & 0.644 & 0.0003 & 0.155 & 0.722 \\
0.002 & 0.072 & 0.249 & 0.330 & 0.139 & 0.637 & 0.754 & 0.033 \\
0.588 & 0.477 & 0.018 & 0.639 & 0.434 & 0.886 & 0.781 & 0.238 \\
0.871 & 0.430 & 0.316 & 0.486 & 0.039 & 0.166 & 0.335 & 0.988 \\
0.031 & 0.830 & 0.206 & 0.919 & 0.313 & 0.436 & 0.640 & 0.565 \\
0.033 & 0.563 & 0.536 & 0.008 & 0.740 & 0.705 & 0.794 & 0.936 \\
0.653 & 0.313 & 0.019 & 0.641 & 0.778 & 0.505 & 0.215 & 0.909 \\
0.234 & 0.023 & 0.897 & 0.837 & 0.878 & 0.173 & 0.025 & 0.058 \\
0.176 & 0.595 & 0.195 & 0.983 & 0.257 & 0.403 & 0.265 & 0.801 \\
0.621 & 0.878 & 0.642 & 0.591 & 0.034 & 0.085 & 0.455 & 0.900 \\
0.757 & 0.571 & 0.942 & 0.309 & 0.474 & 0.260 & 0.727 & 0.592 \\
0.041 & 0.255 & 0.486 & 0.859 & 0.234 & 0.697 & 0.888 & 0.231
\end{bmatrix}$$

$$OutW = \begin{bmatrix} 4733749.051 \\ 1.880 \\ 255412246.148 \\ 98.642 \\ 4.426 \\ 166098349.443 \\ 49864560.027 \\ 219901900.279 \\ 133652612.513 \\ 4427.562 \\ 8151325219.923 \\ 487905.145 \\ 14.712 \\ 7.038 \\ 133.347 \\ 103389656.302 \\ 14304.200 \\ 4.032 \\ 4203.011 \\ 19519300.856 \\ 111.251 \\ 2106938.603 \\ 29048.120 \\ 3.326 \\ 8179037224.417 \\ 1.098 \\ 37176.757 \\ 29.504 \end{bmatrix}$$

References

Arnau Almirall, S., 2017. Ice-gouging in sand and the associated rate effects. Doctoral dissertation, University of Aberdeen, UK.

Arnau, S., Ivanović, A., 2019. Rate effects during ice scour in sand. Cold Regions Science and Technology, 158: 182-194.

Azimi, H., Shiri, H., 2020a. Dimensionless Groups of Parameters Governing the Ice-Seabed Interaction Process. Journal of Offshore Mechanics and Arctic Engineering, 142(5), 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Bonakdari, H., Qasem, S.N., Ebtehaj, I., Zaji, A.H., Gharabaghi, B., Moazamnia, M., 2020. An expert system for predicting the velocity field in narrow open channel flows using self-adaptive extreme learning machines. Measurement, 151: 107202.

Cao, J., Lin, Z., Huang, G.B., 2012. Self-adaptive evolutionary extreme learning machine. Neural processing letters, 36(3): 285-305.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2, St. John's, NL, Canada.

C-CORE., 1996. PRISE Phase 3c: Extreme LEE Gouge Event—Modeling and Interpretation, C-CORE Report 96-C32, St. John's, NL, Canada.

Chari, T.R., 1980. A model study of iceberg scouring in the north Atlantic. Journal of Petroleum Technology, 32(12): 2-247.

Comfort, G., Graham, B., 1986. Evaluation of Sea Bottom Ice Scour Models (No. 37). Environmental Studies Revolving Funds, Report No. 37, Arctic Canada Limited, Calgary, Alberta, Canada.

Gautier, D.L., Bird, K.J., Charpentier, R.R., Grantz, A., Houseknecht, D.W., Klett, T.R., Moore, T.E., Pitman, J.K., Schenk, C.J., Schuenemeyer, J.H., Sorensen, K., Tennyson, M.E., Valin, Z.C., Wandrey, C.J., 2009. Assessment of undiscovered oil and gas in the Arctic, Science, 324: 1175-1179.

Hertz, J.A., 2018. Introduction to the theory of neural computation. CRC Press.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In 2004 IEEE international joint conference on neural networks, IEEE Cat. No. 04CH37541, 2: 985-990.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2006. Extreme learning machine: theory and applications. Neurocomputing, 70(1-3): 489-501.

Huang, G.B., Zhou, H., Ding, X., Zhang, R., 2011. Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(2): 513-529.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand, Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Lach, P.R., 1996. Centrifuge modelling of large soil deformation due to ice scour. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. In The Thirteenth International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. In The Fourteenth International Offshore and Polar Engineering Conference. ISOPE-I-04-107.

Nematzadeh, A., Shiri, H., 2019a. Large deformation analysis of ice scour process in dense sand. In the 10th Congress on Numerical Methods in Engineering, At Guimaraes, Portugal.

Nematzadeh, A., Shiri, H., 2019b. Self-correcting soil models for numerical simulation of strain rate dependent ice scour in sand. In the 9th ECCOMAS Thematic Conference on Smart Structures and Materials, Paris, France.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Machemehl, J.L., Jo, C.H., 1989. Note on new ice gouge events in Alaskan Beaufort Sea. Journal of cold regions engineering, 3(3): 145-149.

Paulin, M.J., 1991. Preliminary Results of Physical Model Tests of ice Scour, Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering, St. John's, NL, Canada.

Paulin, M.J., 1992. Physical Model Analysis of Iceberg Scour in Dry and Submerged Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Sattar, A.M., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. Neural Computing and Applications, 31(1): 157-169.

Shin, M.B., Park, D.S., Seo, Y.K., 2019. Comparative Study of Ice Gouge Simulation Considering Ice Keel-Seabed Interactions. Journal of Ocean Engineering and Technology, 33(6): 556-563.

Storn, R., Price, K., 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 11(4): 341-359.

Wang, G.G., Lu, M., Dong, Y.Q., Zhao, X.J., 2016. Self-adaptive extreme learning machine. Neural Computing and Applications, 27(2): 291-303.

Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Section 4

A non-tuned machine learning method to simulate ice-seabed interaction process in clay

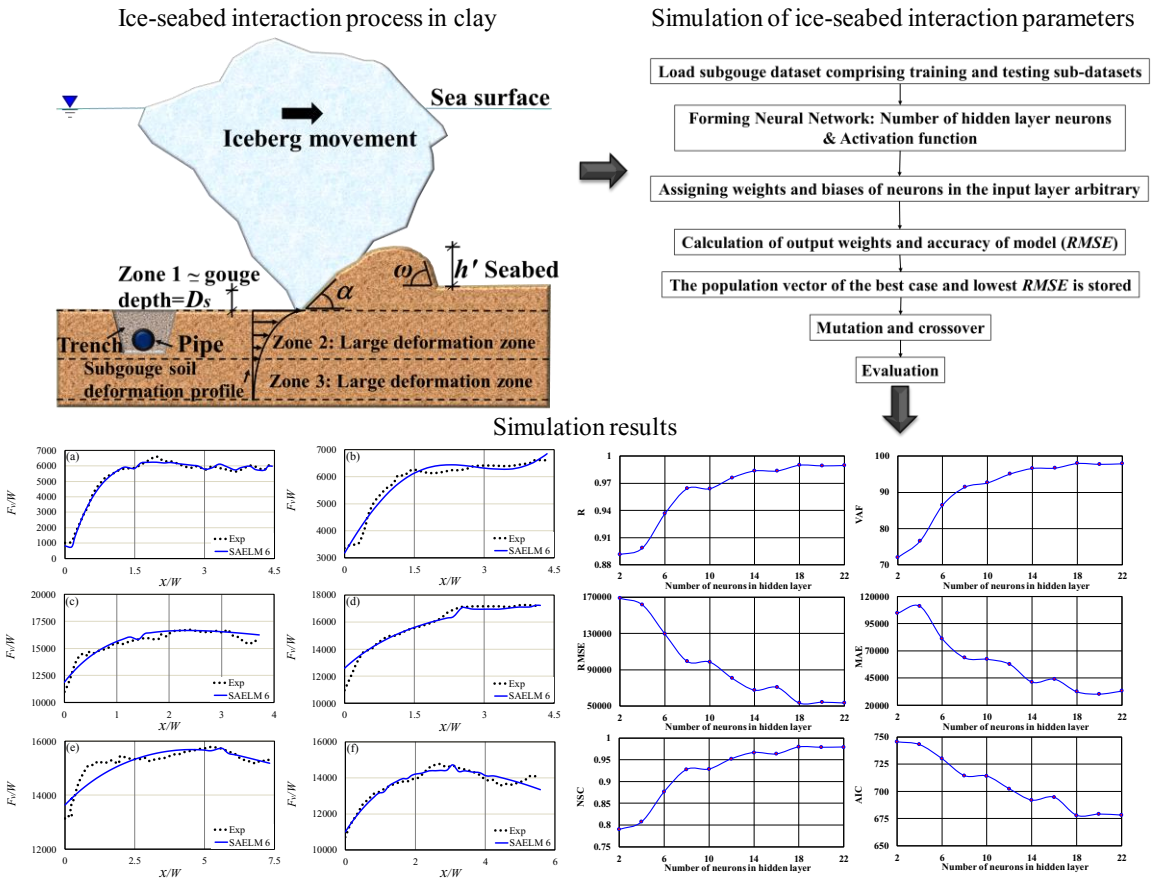
This section is a journal paper which has been published in the Journal of Pipeline Science and Engineering (2021), Vol. 01, pp. 379-394, ISSN: 2667-1433

Abstract

The exploitation of oil and gas in the Arctic area is expected to expand in the coming years. These hydrocarbons are transferred through subsea pipelines from offshore to onshore; however, the marine pipelines are threatened by traveling icebergs where the seabed may be gouged by the moving masses in warmer months. Subsea trenching and backfilling are usually utilized to bury the subsea pipelines for physical protection against ice scouring. Regarding the stress-based design methods, deformations and forces are generally the controlling design factors for the subsea assets. In this study, the subgouge clay displacements and the reaction forces were simulated using a non-tuned self-adaptive machine learning (ML) entitled "self-adaptive extreme learning machine" (SAELM). Initially, fifteen SAELM models were defined by means of the parameters affecting the ice-scoured features. Subsequently, 70% and 30% of the constructed dataset were respectively applied to train and test the machine learning models. After that, the optimum number of hidden layer neurons and the best activation function was selected for the SAELM network. By conducting a comprehensive sensitivity analysis, the premium SAELM models and the most influencing input parameters in the estimation of the subgouge clay characteristics were introduced. Regarding the performed analyses, the horizontal load factor and the gouge depth ratio were identified as the most influential parameters to model the reaction forces, whereas the soil depth had a significant impact on the simulation of the ice-induced clay deformations. Finally, a set of SAELM-based equations were presented to estimate the subgouge clay parameters.

Keywords: Machine learning, Self-adaptive model, Ice-gouging, Clay seabed, Parametrical analysis, Simulation

Graphical abstract



5.1.13. Introduction

Recent discovered crude oil and natural gases in the Arctic area have led to an increased interest in developing subsea pipelines for transferring hydrocarbons from offshore to onshore. Despite being safe and cost-effective, these pipelines are subjected to the hazards inherent to the arctic environment such as ice-scouring. This particular phenomenon occurs when a traveling iceberg keel plows the seafloor, presenting a threat to the subsea assets. This is more likely to happen during the warmer seasons (Azimi and Shiri, 2021).

Backfilling and trenching are generally used to protect these sea bottom-founded infrastructures against the ice-gouging problem, so estimation of the subgouge soil response is considered an important design factor for the subsea infrastructures since minimizing the required burial depth to physical protection is crucial for the project budget. The maximum subgouge deformation occurs just beneath the ice keel and the magnitude of the soil deformation decreases with greater depths (Nematzadeh and Shiri, 2020). The schematic layout of the ice-seabed interaction process is displayed in Figure 5-43. Regarding the figure, D_s is the gouge depth, α is the attack angle, ω is the angle of the surcharged soil slope, and h' is the height of the berm.

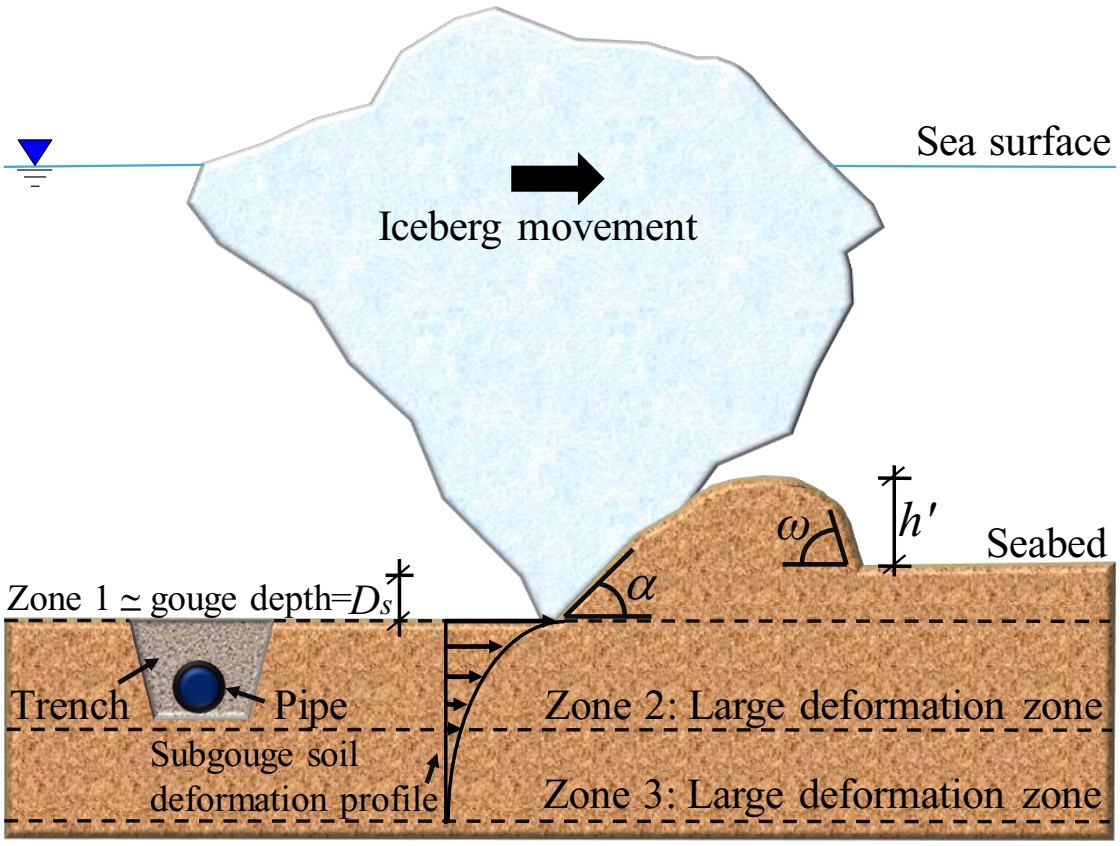


Figure 5-43. Schematic layout of the ice-seabed interaction process

Hefty field and experimental investigations as well as time-consuming numerical simulations are currently utilized to evaluate the ice-seabed interaction process (Comfort and Graham, 1986; Machemehl and Jo, 1989; Kioka et al., 2000; Abdalla et al., 2009; Bailey Dudley et al., 2017; Nematzadeh and Shiri, 2019); however, the industry is constantly looking for alternative approaches to predict the behavior of the subsea pipelines to lessen the collision risk of icebergs with the subsea assets (Azimi and Shiri, 2020). In the past three decades, a number of researchers have sought to study the parameters affecting the ice-seabed interaction process. For instance, Lach (1996) investigated the ice scouring problem experimentally and numerically. The author performed a set of centrifuge tests to determine the ice-clay interaction parameters. The study also

showed that the scouring force was a function of undrained shear strength, keel geometry, and subgouge deformation. A series of the Pressure Ridge Ice Scour Experiment (PRISE) joint-industry tests were conducted by Woodworth-Lynes et al. (1996). The authors demonstrated that the subgouge deformations could be estimated in terms of the soil depth, gouge depth, and gouge width. At the center for Cold Ocean Resources Engineering (C-CORE), several centrifuge tests were conducted to measure the ice-induced deformations and reaction forces in both sand and clay seabed (C-CORE, 1995, 1996). The experimental results showed that the magnitude of scour force increased with increasing the soil shear strength, scour width, and scour depth, while this parameter reduced when the keel attack angle increased. Schoonbeek et al. (2006) measured the subgouge clay deformations through centrifuge tests. The authors presented an empirical equation in terms of gouge geometry and soil parameters to estimate the subgouge depth. Been et al. (2008) studied the mechanism of clay failure in ice scouring problems. The authors concluded that greater deformations occurred in a clay mass with higher undrained shear strength, meaning that the scour depth was a function of the soil properties. Arnau Almirall (2017) conducted several 1g laboratory tests to perform the subgouge sand features in saturated and dry circumstances. The author studied the effect of velocity change, scour geometry, and soil conditions on the subgouge parameters. The study demonstrated that the ice-induced sand deformations in the 1g test are less than those resulting from the centrifuge tests.

Liu et al. (2015) simulated the ice keel-seabed interaction by means of a discrete element method (DEM) and the cohesive frictional model (CFM) was applied to model the freeze-bonds between blocks of ice. The numerical results were validated by PIRAM tests and the authors asserted that the numerical results improved by considering the water drag. Shin et al. (2019) developed a three-dimensional model to simulate the ice-gouging phenomenon. The authors showed that the

numerical model had a better performance by considering the geostatic stress and contact condition. Using the modified Mohr-Coulomb (MMC) scheme, the ice-gouging problem in a sandy seabed was modeled by Nematzadeh and Shiri (2020). The authors concluded that the magnitude of subgouge sand displacements and reaction forces enhanced with the growing value of the unit weight and relative density.

Although several expensive and time-consuming experimental and numerical research has been carried out on the ice-seabed interaction process, there have been few artificial intelligence (AI) investigations into studying this phenomenon. Different linear and non-linear problems have been modeled by using AI and machine learning (ML) algorithms since these tools are sufficiently precise, quick, and cost-effective. For example, Kioka et al. (2002, 2003) estimated the scour depth through the artificial neural network (ANN). The optimized number of neurons in the hidden layer was considered to be 5, and then the scour depth was approximated by four inputs: ice drift velocity, sand property, sea-bottom gradient, and attack angle. Azimi and Shiri (2020a) performed a dimensional analysis to identify the factors affecting the ice-seabed interaction process by using the Buckingham theorem. The authors proposed a series of linear regression (LR) models to estimate the maximum subgouge soil displacements. Azimi and Shiri (2020b) simulated the horizontal subgouge sand deformations by means of gene expression programming (GEP). Using the input parameters, five GEP models were defined and then the best model was selected by performing a sensitivity analysis. In the end, a GEP-based equation was suggested to compute the subgouge sand displacements. Azimi and Shiri (2021a) estimated the subgouge sand features by using the extreme learning machine (ELM). The authors performed several analyses in order to identify the superior model and the influencing input variables. The study concluded that the gouge depth ratio had a considerable impact on the estimation of the sand displacements. A multi-layer

perceptron neural network (MLPNN) was applied to predict the horizontal subgouge sand deformation by Azimi and Shiri (2021b). The authors asserted that the soil depth and bearing pressure were the most important input parameters to simulate the objective function.

By analyzing the literature, it is possible to notice the considerable amount of effort dedicated to the study of the ice seabed interaction through experimental or 3D numerical models. Even though in small numbers, these novel alternatives are still very much sought by the industry given their cost-effectiveness and high potential for design purposes. Hence, in the current study, the ability of a non-tuned self-adaptive machine learning (ML) called "self-adaptive extreme learning machine" (SAELM) to model the ice-clay interaction process is scrutinized for the first time. More information about this will be provided in the next sections.

5.1.14. Methodology

Initially, the extreme learning machine (ELM), differential evolution (DE) algorithm, and self-adaptive extreme learning machine (SAELM) were introduced and then the parameters affecting the ice-induced clay seabed were detailed. Subsequently, fifteen SAELM models were defined by means of these input parameters. Lastly, several statistical indices were also presented for the appraisal of these SAELM models. Through that analysis, the best models were selected and the most influential parameters were determined.

5.1.14.1. Extreme learning machine (ELM)

Extreme learning machine (ELM) as a single-layer feed-forward neural network (SLFFNN) was first developed by Huang et al. (2004). Using the ELM, the ice-gouging features including the reaction forces ($F/\gamma_s W^3$) and the subgouge deformations (d/W) as the output of the problem (Q)

were simulated by *m* as the number of input parameters and *L* as the size of hidden layers as follows (Ebtehaj et al., 2016):

$$\Omega = \sum_{j=1}^{L} \beta_j . G(a_j, b_j, \chi_i) \tag{5-87}$$

here, $i = 1, 2, 3, m$ and $j = 1, 2, 3, L$. Moreover, a_j and b_j are training parameters, while β_j is defined as the connection between the i^{th} input parameter and the j^{th} neuron within the hidden layer. Regarding the χ_i as an input parameter, the output of the j^{th} neuron in the hidden layer is $G(a_j, b_j, \chi_i)$ and it is connected to the output neuron via β_j. The cost function of the ELM model is estimated below (Ebtehaj et al., 2016):

$$\mathrm{C} = \sum_{j=1}^{m} \|\Omega_i - Y_i\| \tag{5-88}$$

where, *Y* is the value of the target parameter. In order to minimize the cost function of the ELM model, the network parameter (Ŋ) is a set of a_j, b_j, and β_j that is repetitively updated through the conventional gradient-based algorithm as follows (Ebtehaj et al., 2016)::

$$\text{Ŋ}_{k+1} = \text{Ŋ}_k - \mathrm{l}\frac{\partial \mathrm{C}(\text{Ŋ}_k)}{\partial(\text{Ŋ}_k)} \tag{5-89}$$

here, l is the learning rate. To estimate the outputs with zero error ($\sum_{j=1}^{m}\|\Omega_i - Y_i\| = 0$), $G(.)$ is applied as the activation function, therefore (Ebtehaj et al., 2017):

$$H\beta = Y \tag{5-90}$$

where, *H* is the hidden layer matrix, β is the matrix of output weight, Y is the matrix of the target. The SLFFNN functions as a common estimator by randomly selecting the neurons in the hidden layer and calculating the output weights regarding the target parameter. Thus, the output weights are computed by solving the least square problem below (Ebtehaj et al., 2017):

$$\hat{\beta} = H^{+}Y \tag{5-91}$$

here, H^+ represents the Moore-Penrose generalized inverse matrix of H.

5.1.14.2. Differential evolution (DE)

Differential evolution (DE), as a robust optimizer algorithm, was first presented by Storn and Price (1997). The automatic detection-derivation mechanism and the high convergence speed are the most significant benefits of DE over other optimizer tools (Bonakdari et al., 2020). Commonly, minimizing the following equation is the main goal of a DE algorithm:

$$\min f(\vartheta) \quad \text{where} \quad \vartheta \in \boldsymbol{R}^{\mathrm{D}} \tag{5-92}$$

here, ϑ is a vector where the DE algorithm generally employs a population of I individuals so as to calculate the global optimal of the ice-seabed interaction parameters. So, the i^{th} vector is surmised as below:

$$\vartheta_{i,\varpi} = \vartheta_{i,\varpi}^{1}, \vartheta_{i,\varpi}^{2}, \vartheta_{i,\varpi}^{3}, \cdots, \vartheta_{i,\varpi}^{\mathrm{D}} \qquad \text{where } i = 1,2,3, \cdots I \tag{5-93}$$

where, ϖ is defined as the generation number. In the current study, the DE algorithm utilizes four major stages to solve the ice-seabed interaction problem as follows:

(I) Initialization: I vectors are randomly produced by using the below equation:

$$\vartheta_{i,\varpi} = \vartheta_{min} + rand(0,1) \times (\vartheta_{max} - \vartheta_{min}) \qquad \text{where} \begin{cases} \vartheta_{min} = \left[\vartheta_{min}^{1}, \vartheta_{min}^{2}, \cdots, \vartheta_{min}^{\mathrm{D}}\right] \\ \vartheta_{max} = [\vartheta_{max}^{1}, \vartheta_{max}^{2}, \cdots, \vartheta_{max}^{\mathrm{D}}] \end{cases} \tag{5-94}$$

here, the ϑ_{max} and ϑ_{min} are defined as the boundaries considered for the ice-gouging problem.

(II) Mutation: The mutant vectors ($\Psi_{i,\varpi}$) are created by means of the generated vector ($\vartheta_{i,\varpi}$) in the mutation phase. Various mutation strategies have been introduced by Storn and Price (1997); however, four of them are applied in the current investigation:

Strategy 1:

$$\Psi_{i,\varpi} = \vartheta_{\sigma_{1,\varpi}^{i}} + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) \tag{5-95}$$

Strategy 2:

$$\Psi_{i,\varpi} = \vartheta_{\sigma_{1,\varpi}^{i}} + \Lambda \times \left(\vartheta_{best,\varpi} - \vartheta_{\sigma_{1,\varpi}^{i}}\right) + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) + \Lambda \times \left(\vartheta_{\sigma_{4,\varpi}^{i}} - \vartheta_{\sigma_{5,\varpi}^{i}}\right) \tag{5-96}$$

Strategy 3:

$$\Psi_{i,\varpi} = \vartheta_{\sigma_{1,\varpi}^{i}} + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) + \Lambda \times \left(\vartheta_{\sigma_{4,\varpi}^{i}} - \vartheta_{\sigma_{5,\varpi}^{i}}\right) \tag{5-97}$$

Strategy 4:

$$\Psi_{i,\varpi} = \vartheta_{i,\varpi} + \Lambda \times \left(\vartheta_{\sigma_{1,\varpi}^{i}} - \vartheta_{i,\varpi}\right) + \Lambda \times \left(\vartheta_{\sigma_{2,\varpi}^{i}} - \vartheta_{\sigma_{3,\varpi}^{i}}\right) \tag{5-98}$$

here, σ_k^i is randomly calculated from the $[1,2,3,\cdots,I]$ interval and k as the control parameter is randomly adjusted between 0 and 1. The Λ is a scaling factor selected between 0 and 2 (Storn and Price, 1997).

(III) Crossover: The crossover phase is performed on the mutated vectors to enhance the model's versatility. During this stage, the trial vector $\left(\Delta_{i,\varpi} = \left\lfloor \Delta_{i,\varpi}^{1}, \Delta_{i,\varpi}^{2}, \cdots, \Delta_{i,\varpi}^{\mathrm{D}}, \right\rfloor\right)$ is produced for the mutated vectors $\left(\Psi_{i,\varpi} = \left\lfloor \Psi_{i,\varpi}^{1}, \Psi_{i,\varpi}^{2}, \cdots, \Psi_{i,\varpi}^{\mathrm{D}}, \right\rfloor\right)$ using the equation below:

$$\Delta_{i,\varpi}^{j} = \begin{cases} \Psi_{i,\varpi}^{j} & \text{if}\left(\mathrm{rand}_j \le CC\right) \text{ or } \left(j = j_{\mathrm{rand}}\right) \\ \vartheta_{i,\varpi}^{j} & \text{otherwise} \end{cases} \tag{5-99}$$

here, CC is the coefficient of crossover and the rand_j variable is randomly taken into account as a constant value. It is worth mentioning that the j_{rand} parameter is adjusted between 1 and D since one of the $\Delta_{i,\varpi}$ variables ought to be different from the $\vartheta_{i,\varpi}$ values.

(IV) Selection: The best individual with the minimum target value is chosen as the solution to the ice-seabed interaction problem.

Phases (II) to (IV) are iterated till either the desired accuracy or maximum iterations predefined are achieved.

5.1.14.3. Self-adaptive extreme learning machine (SAELM)

Even though the ELM algorithm has an acceptable performance while simulating multifarious linear or nonlinear problems, the manual selection of optimized hidden layer neurons plays a crucial role in the results. Moreover, in practical applications, the input parameters should be dimensionless, and that is another drawback of the ELM network (Cao et al., 2012). In order to overcome these limitations, the self-adaptive evolutionary extreme learning machine (SAELM), as a hybrid model of the ELM network and DE algorithm, was developed for the first time by Cao et al. (2012). Figure 5-44 details the flowchart of the SAELM model in the current study. In the SAELM algorithm, the number of neurons in the hidden layer is optimized by using the self-adaptive differential evolution algorithm. During the training procedure, the control coefficients in the DE algorithm and the strategies for trial vector generation are determined through a self-adaptive mechanism. In other words, during the simulation process, the output weights are estimated by means of the ELM network, whereas the input weights and biases of the hidden layer neurons are determined through the DE algorithm. To model the ice-seabed interaction process, the main steps comprising initialization, output weights, mutation and crossover, and evaluation are applied by using the SAELM algorithm as follows:

Initialization: The *I* vector is randomly created in the first DE generation and the following vector is computed:

$$\vartheta_{k,\varpi} = \left\lfloor a^{Y}_{1,(k,\varpi)}, \cdots, a^{Y}_{l,(k,\varpi)}, b_{1,(k,\varpi)}, \cdots, b_{l,(k,\varpi)} \right\rfloor \tag{5-100}$$

where *a* and *b* are the learning parameters. Additionally, *l* is the size of hidden layers.

Output weights: The output weights are defined by using the following equation:

$$\beta_{k,\varpi} = H_{k,\varpi} Y \tag{5-101}$$

where, $H_{k,\varpi}$ is determined by using the below equation:

$$H_{k,\varpi} = \begin{bmatrix} G\left(a_{1,(k,\varpi)}, b_{1,(k,\varpi)}, \chi_1\right) & \cdots & G\left(a_{n,(k,\varpi)}, b_{n,(k,\varpi)}, \chi_1\right) \\ \vdots & \ddots & \vdots \\ G\left(a_{1,(k,\varpi)}, b_{1,(k,\varpi)}, \chi_m\right) & \cdots & G\left(a_{L,(k,\varpi)}, b_{L,(k,\varpi)}, \chi_m\right) \end{bmatrix} \quad (5\text{-}102)$$

Additionally, the Root Mean Squared Error (*RMSE*) of the generated individuals is computed as follows:

$$RMSE_{k,\varpi} = \sqrt{\frac{\sum_{i=1}^{m} \left\| \sum_{j=1}^{L} \beta_j . G\left(a_{j,(k,\varpi)}, b_{j,(k,\varpi)}, \chi_i\right) - Y_i \right\|}{n \times m}} \quad (5\text{-}103)$$

After that, the next generation of individuals is calculated using the following equation:

$$\vartheta_{k,\varpi+1} = \begin{cases} \Delta_{k,\varpi+1} & if \quad RMSE_{\vartheta_{k,\varpi}} - RMSE_{\vartheta_{k,\varpi}} > \pi . RMSE_{\vartheta_{k,\varpi}} \\ \Delta_{k,\varpi+1} & if \quad \left|RMSE_{\vartheta_{k,\varpi}} - RMSE_{\vartheta_{k,\varpi}}\right| < \pi . RMSE_{\vartheta_{k,\varpi}} \quad and \quad \left\|\beta_{\Delta_{k,\varpi+1}}\right\| < \left\|\beta_{k,\varpi}\right\| \\ \vartheta_{k,\varpi} & else \end{cases} \quad (5\text{-}104)$$

where π is defined as a positive constant value so as to hinder the zero improvement rate.

Mutation and Crossover: Using the introduced strategies in equations (5-95) to (5-98), the trial vectors of the DE algorithm are defined. Furthermore, the best strategy is chosen regarding the probability procedure $P_{l,\varpi}$ where the l^{th} strategy is considered for the ϖ^{th} generation and *l* is 1, 2, 3, or 4. Generally, $P_{l,\varpi}$ is updated in each step, meaning that if the population of generated vectors (I) is higher than ϖ, the four applied strategies possess an identical probability ($P_{l,\varpi} = 0.25$). If ϖ is greater than the I, $P_{l,\varpi}$ is calculated as follows:

$$P_{l,\varpi} = \frac{\varrho_{l,\varpi}}{\sum_{l=1}^{4} \varrho_{l,\varpi}} \qquad \text{where} \qquad \varrho_{l,\varpi} = \frac{\sum_{q=\varpi-I}^{\varpi-1} n\varrho_{l,\varpi}}{\sum_{q=\varpi-I}^{\varpi-1} n\varrho_{l,\varpi} + \sum_{q=\varpi-I}^{\varpi-1} nf_{l,\varpi}} + \pi \quad (5\text{-}105)$$

where, $nf_{l,\varpi}$ and $n\varrho_{l,\varpi}$ are trial vectors. It is worth mentioning that the *CC* and Λ values are determined for a target vector through the selection phase from the normal distribution function.

Evaluation: The next generation of trial vectors is created by means of equation (5-104). The mutation, crossover, and evaluation steps are iterated until either the predefined iteration number or required accuracy is reached while simulating the ice-induced seabed deformation using the SAELM algorithm.

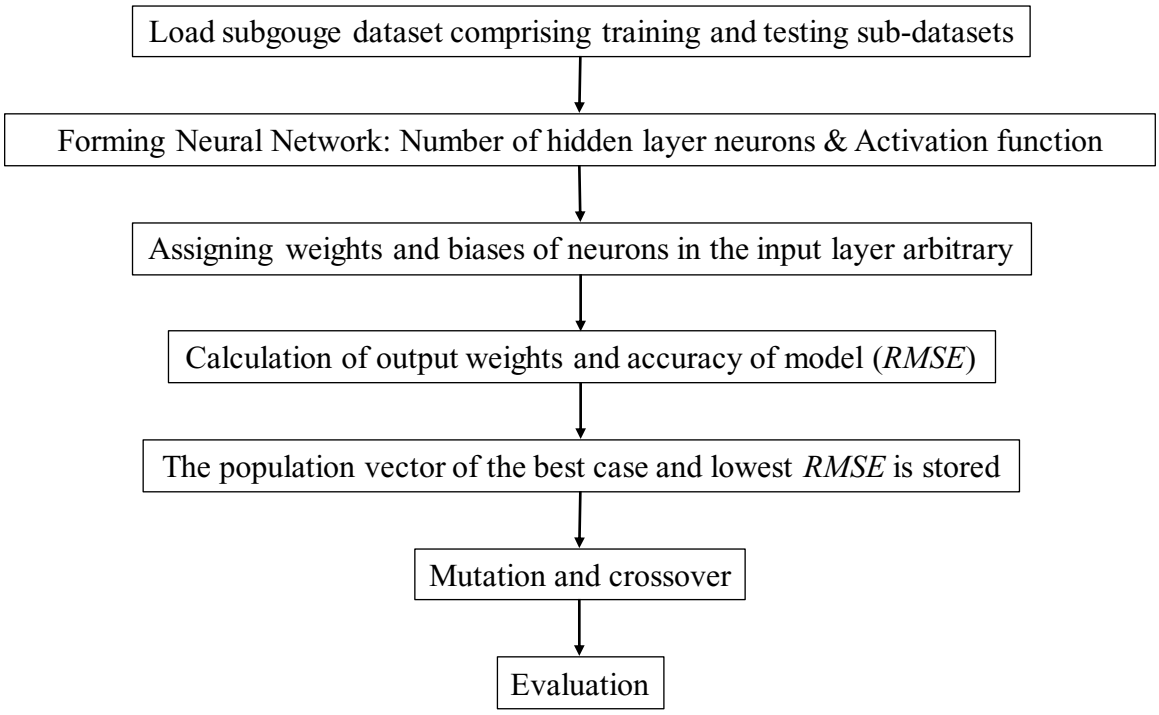


Figure 5-44. Flowchart of the SAELM in the current study

5.1.14.4. Seabed interaction process in clay

Seabed interaction parameters (η) in a clay mass comprising the soil deformations (d/W) and the reaction forces ($F/\gamma_s W^3$) are a function of the scour depth (Ds), the shear strength parameter of the clay (c), the width of gouge (W), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (h'), the horizontal load (Lh), the vertical load (Lv), the velocity of ice keel (V), and the specific weight of clay (γ_s) as follows (Azimi and Shiri, 2020; Lach, 1996):

$$\eta_{(clay)} = f_1(D_s, c, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \quad (5\text{-}105)$$

Equation (5-105) can be written as a function of eight dimensionless groups as below (Azimi and Shiri, 2020):

$$\frac{\eta_{(clay)}}{W} = f_2\left(\frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}106)$$

It is worth noticing that the maximum subgouge deformation in the clay $\left(d_{(clay)}\right)$ is formed just under the moving ice keel in the gouge centerline. However, at a greater depth on the subgouge centerline, by incorporating the soil depth (*y/W*), the subgouge clay displacement ($d_{(clay)}/W$) is written as follows:

$$\frac{d_{(clay)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}107)$$

Similarly, the ice-induced reaction force (*F*) is a function of the position of the iceberg along the scour axis (*x*) and $D_s, c, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s$, so equation (5-107) can be written as below:

$$\frac{F_{(clay)}}{\gamma_s W^3} = f_4\left(\frac{x}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (5\text{-}108)$$

In the current study, different laboratory investigations were used in order to simulate the ice-gouging clay parameters (*d/W* and $F/\gamma_s W^3$). In other words, the key measurements of five experimental investigations reported by C-CORE (1995), Lach (1996), C-CORE (1996), Schoonbeek et al. (2006), and Been et al. (2008) were utilized to verify the SAELM models. It should be noted that the angle of the surcharged soil slope (ω) and the height of the berm (*h*') were not reported by all applied laboratory studies. Therefore, other introduced dimensionless groups in equations (5-107) and (5-108) were utilized as the input parameters for the SAELM model to simulate the horizontal subgouge deformation (*dh/W*), the vertical subgouge deformation (*dv/W*), the horizontal reaction force ($Fh/\gamma_s W^3$), and the vertical reaction force ($Fv/\gamma_s W^3$). Figure 5-45 portrays the defined input combinations in the current study. Fifteen SAELM models (SAELM 1

to SAELM 15) were developed to simulate the ice-induced clay parameters. This means that SAELM 1 was defined by using all input parameters and then the effect of each input was eliminated (SAELM 2 to SAELM 8) and then SAELM 9 to SAELM 15 were developed so as to detect the most significant input parameters. It is noteworthy that 70% of the experimental measurements were utilized to train the SAELM network, whilst 30% of the rest were used to test it.

SAELM 1: [1,2,3,4,5,6,7]
SAELM 2: [1,2,3,4,5,6]
SAELM 3: [1,2,3,4,5,7]
SAELM 4: [1,2,3,4,6,7]
SAELM 5: [1,2,3,5,6,7]
SAELM 6: [1,2,4,5,6,7]
SAELM 7: [1,3,4,5,6,7]
SAELM 8: [2,3,4,5,6,7]
SAELM 9: [1]
SAELM 10: [2]
SAELM 11: [3]
SAELM 12: [4]
SAELM 13: [5]
SAELM 14: [6]
SAELM 15: [7]

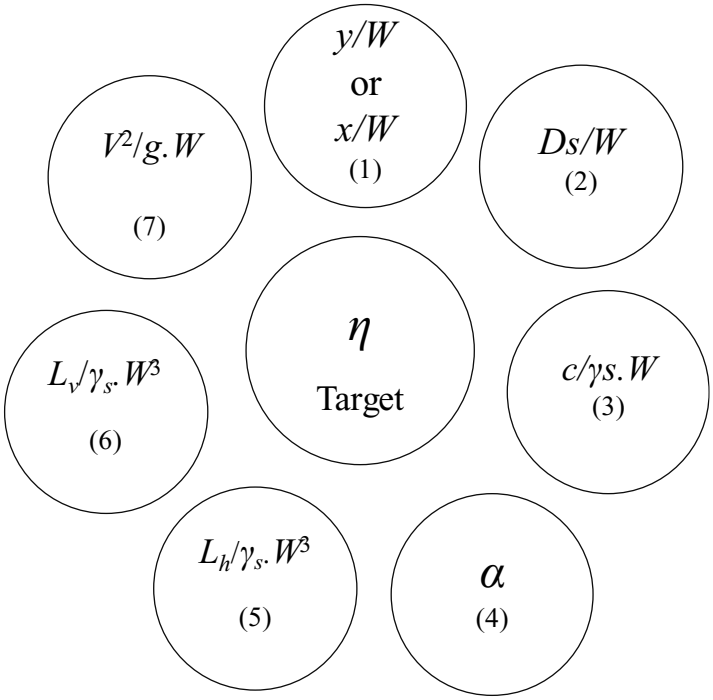


Figure 5-45. Combination of input parameters to develop SAELM models

5.1.14.5. Quality of fit

A set of criteria such as correlation coefficient (*R*), variance accounted for (*VAF*), root mean square error (*RMSE*), mean absolute error (*MAE*), Nash-Sutcliff efficiency coefficient (*NSC*), and Akaike Information Criteria (*AIC*) were utilized to evaluate the accuracy, correlation, and complexity of the SAELM models. These criteria were utilized to assess the correlation, precision, correctness,

and complexity of the SAELM models. The correlation of the models was examined by the R, NSC indices, while the precision of these models was evaluated by the *RMSE* and *MAE* criteria. The correctness of the models was verified through the *VAF* index, but the *AIC* indicator was applied in order for evaluation of the complexity of the defined models.

Closeness of the *R* and *NSC* indices to one signified that the SAELM model had the best correlation with the experimental values, on the other hand, the closeness of the *RMSE* and *MAE* values to zero meant that a particular model had the lowest level of error. Moreover, the best SAELM model possessed the highest value of the *VAF* index though the complexity of these SAELM models was not assessed using the defined criteria. To overcome this limitation, the Akaike Information Criteria (*AIC*) was used, meaning that the less complex SAELM model owned the lowest value of *AIC*; hence, the premium SAELM model had the lowest values of *AIC* index and error (*RMSE* and *MAE*), with the highest level of correlation (*R* and *NSC*).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{5-109}$$

$$VAF = \left(1 - \frac{var(P_i - O_i)}{var(P_i)}\right) \times 100 \tag{5-110}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{5-111}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - O_i|}{n} \tag{5-112}$$

$$NSC = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \tag{5-113}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{5-114}$$

here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, and *n* are respectively the experimental measurements, the simulated values, the average experimental values, the average simulated values, and the number of experimental measurements. In addition, *k* is the number of hidden layer neurons used in the SAELM network.

5.1.15. Results and discussion

In the first instance, the optimized number of hidden layer neurons and the premium activation function were determined. After that, a sensitivity analysis for SAELM 1 to SAELM 15 was implemented and then the best SAELM models, along with the most significant input variables, were introduced. Subsequently, further analyses comprising the error analysis and the uncertainty analysis were performed for the premium SAELM models. Lastly, a set of SAELM-based equations were suggested to approximate the subgouge clay characteristics. The simulation results, obtained during the test stage, will be presented in the upcoming sections.

5.1.15.1. Number of hidden neurons (NHN)

Although the precision of SAELM network is improved with increasing the number of hidden neurons (NHN), the computation time and complexity of the SAELM model grow (Sattar et al., 2019). It means that the determination of the optimal NHN can play a significant role to improve the performance of the SAELM network. The alterations of NHN against different computed indices to predict the horizontal reaction forces are shown in Figure 11-4. In the current study, the SAELM 1 model was used for evaluation of the NHN and the optimum number of NHN was selected by means of a trial and error approach, meaning that the initial number of neurons in the hidden layer was assumed to be 2 where the SAELM model demonstrated the worst performance, with the *R*, *RMSE*, and *AIC* values of 0.891, 168566.038, and 745.614. As shown in Figure 5-46,

the performance of SAELM model was slightly changed as the NHN was greater than 20 where the values of *VAF*, *AIC*, and *R* were respectively approximated at 97.759, 678.923, and 0.989. It seems that the SAELM model with NHN=18 had a similar performance to the model with NHN=20, with a *VAF*, *AIC*, and *R* of 96.719, 694.474, and 0.984. However, the model with NHN=20 outperformed the model with NHN=18. Therefore, to simulate the subgouge clay characteristics in the current investigation, the optimized number of neurons in the hidden layer was selected as 20.

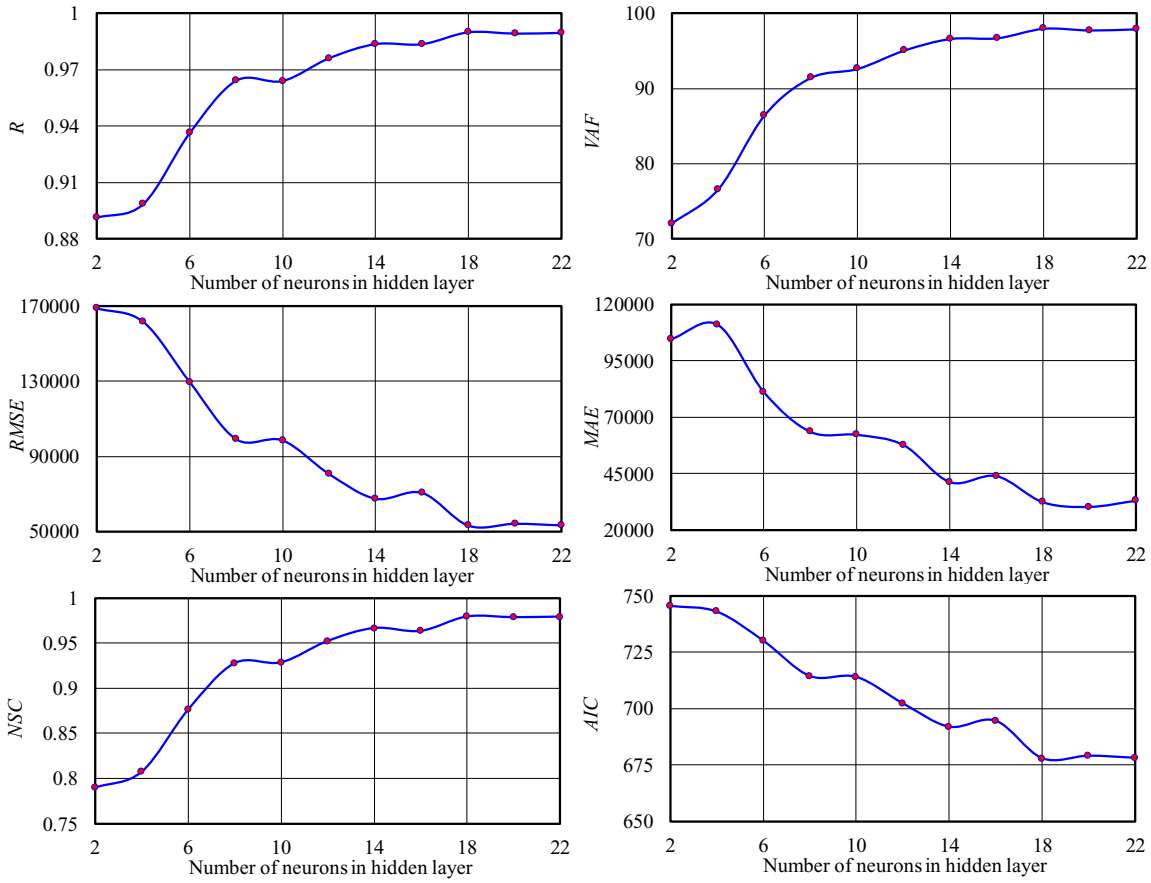


Figure 5-46. Alterations of NHN against computed indices for modeling the horizontal reaction forces

5.1.15.2. Activation Function (AF)

In the machine learning algorithms, the weighted input parameters and biases are nonlinearly transformed through a transfer function or activation function (AF) (Hertz, 2018). In this study, various AF for the SAELM network comprising the hyperbolic tangent (Hypertan), the triangular basis (Tribas), the radial basis (Radbas), the hard limit (Hardlim), the sigmoid (Sig), and the sinusoidal (Sin) functions were utilized to simulate the ice-seabed interaction process in clay. The performance of different AFs for simulation of the horizontal reaction forces is illustrated in Figure 11-5. It is worth noting that the SAELM 1 model was used to assess the effect of the AFs.

According to the simulation results, the value of correlation coefficient for the Hypertan, Tribas, and Radbas functions was respectively estimated as 0.977, 0.893, and 0.969. For the Hardlim function, the computed *MAE* and *NSC* indices were equal to 99083.311 and 0.829.

The Sin function showed the worst performance among the applied activation functions, with an *AIC* value of 746.767. However, the Sig function may be considered the most accurate activation function to model the ice-induced features in clay seabed where the values of *AIC* and *VAF* for this function were computed at 674.927 and 98.236. Thus, the sigmoid activation function was selected for the simulation of the ice-seabed interaction event in this investigation.

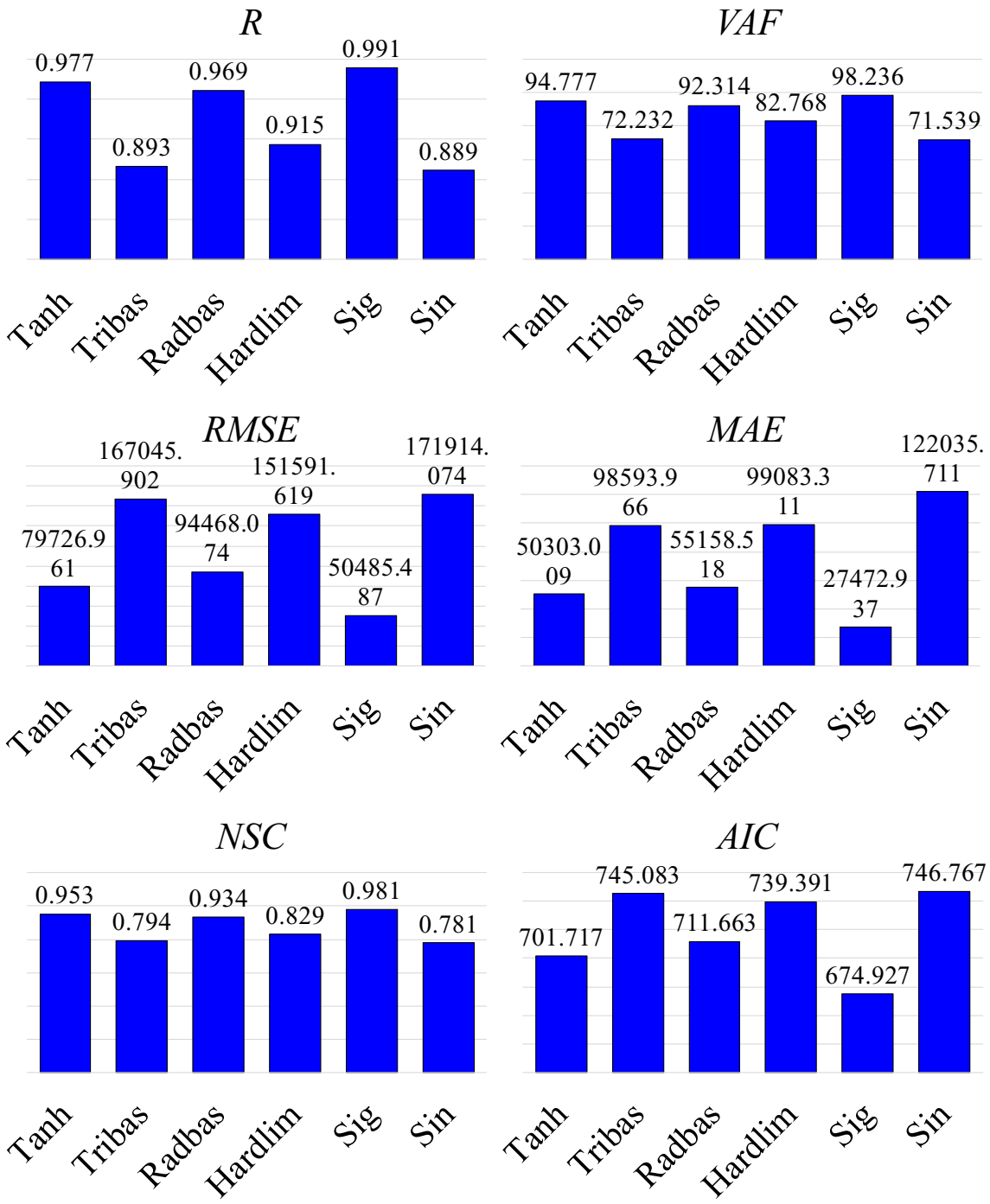


Figure 5-47. Comparison between the performances of various activation functions in the SAELM network for the simulation of the horizontal reaction forces

5.1.15.3. Evaluation of SAELM models

The performance of the defined SAELM models was evaluated by conducting a sensitivity analysis. This means that the capability of SAELM 1 to SAELM 8 was assessed in order to introduce the best SAELM model. The impact of each input was eliminated one at a time in the SAELM 2 to SAELM 8 models, while SEALM 1 was a function of all input parameters. Furthermore, to detect the most influencing input parameters, the ability of the SAELM 9 to SAELM 15 models for the simulation of the ice-induced parameters was investigated. Table 5-9 shows the comparison between the performances of the SAELM models to simulate the horizontal reaction forces. For SAELM 1 a function of x/w , D_s/W , $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW inputs, the correlation coefficient (R), and Akaike Information Criteria (*AIC*) were calculated to be 0.987 and 684.047. It is worth noticing that SAELM 1 was ranked as the sixth-best model to predict horizontal reaction forces. SAELM 2 as the third-best model simulated the horizontal reaction forces by using x/w , D_s/W , $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, with a VAF value of 97.768. The effect of $L_v/\gamma_s.W^3$ for SAELM 3 (the forth-best model) was removed, signifying that this model was a function of x/w , D_s/W , $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, V^2/gW. The *NSC* and R values for SAELM 3 were respectively estimated at 0.975 and 0.987. The $L_h/\gamma_s.W^3$ parameter was the eliminated factor for SAELM 4 (the seventh-best model) and other inputs including x/w , D_s/W , $c/\gamma_s.W$, α, $L_v/\gamma_s.W^3$, V^2/gW were used to simulate the horizontal reaction forces where the *VAF*, *MAE*, and R criteria for this model equaled 97.428, 35066.664, and 0.987, respectively. The attack angle (α) was the removed parameter for SAELM 5 (the best SAELM model) and this model predicted the $Fh/\gamma_s W^3$ in terms of x/w , D_s/W , $c/\gamma_s.W$, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and V^2/gW, with an *RMSE* of 49662.134. SAELM 6 was fed by using the x/w , D_s/W , α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW inputs so as to

model the horizontal reaction forces, whereas the clay shear strength ($c/\gamma_s.W$) was the eliminated factor. For SAELM 6 (the fifth-best model), the values of *AIC* and *NSC* were equal to 683.531 and 0.975. The second-best model (SAELM 7) forecasted the target function (F_h/W) through x/w, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$ and V^2/gW input parameters and the ratio of gouge depth to gouge width (D_s/W) was ignored for this SAELM model, with a correlation coefficient of 0.989. The *MAE*, *VAF*, and *AIC* values for SAELM 8 (the worst model) were respectively estimated to be 125104.727, 53.391, and 763.182. The x/w was a removed parameter for SAELM 8 where this model simulated the horizontal reaction forces in terms of D_s/W, $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and V^2/gW parameters. Among the SAELM models with one input (SAELM 9 to SAELM 16), SAELM 13 a function of the horizontal load parameter ($L_h/\gamma_s.W^3$) had the highest level of correlation and precision along with the lowest level of complexity (*R*=0.801, *MAE*=121211.917, and *AIC*=761.914), so the $L_h/\gamma_s.W^3$ was detected as the most significant input parameter. Sensitivity analysis also showed that the vertical load parameter ($L_v/\gamma_s.W^3$), the velocity parameter (V^2/gW), the clay shear strength ($c/\gamma_s.W$), the ratio of gouge depth to gouge width (D_s/W), and the position of iceberg along the scour axis (*x/W*) were recognized as the second important to sixth important input parameters. However, the attack angle (α) demonstrated a negligible influence to simulate the horizontal reaction forces.

Thus, SAELM 5 was introduced as the best model to predict the horizontal reaction forces (Fh/$\gamma_s W^3$), while the horizontal load parameter ($L_h/\gamma_s.W^3$) possessed the highest level of influence for simulation of the target values.

Table 5-9. Comparison between the performances of the SAELM models to simulate the horizontal reaction forces

Model	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
SAELM 1	0.987	97.488	58981.48	33873.16	0.974	684.047
SAELM 2	0.989	97.768	55163.91	33828.04	0.977	680.123
SAELM 3	0.987	97.477	58456.26	37083.99	0.975	683.522
SAELM 4	0.987	97.428	58598.64	35066.66	0.974	683.665
SAELM 5	0.991	98.13	49662.13	27756.89	0.982	673.963
SAELM 6	0.988	97.498	58464.64	37982	0.975	683.531
SAELM 7	0.989	97.866	55541.67	32207.49	0.977	680.524
SAELM 8	0.794	53.391	227462.5	125104.7	0.617	763.182
SAELM 9	0.387	-689.185	345849.6	283069.2	0.115	787.750
SAELM 10	0.594	-117.046	296853.1	231393.8	0.348	778.793
SAELM 11	0.764	43.737	240648.3	141417.2	0.571	766.487
SAELM 12	0.182	-2193.92	364218.6	304225.2	0.018	790.784
SAELM 13	0.801	54.080	222593.8	121211.9	0.633	761.914
SAELM 14	0.793	52.001	226642.6	131446.7	0.620	762.971
SAELM 15	0.764	22.961	237717.2	145973.8	0.582	765.768

The comparison of the computed key statistical indices for SAELM models in order to predict the vertical reaction forces is tabulated in Table 5-10. Based on the simulation results, SAELM 6 showed the best performance to model the vertical reaction forces ($Fv/\gamma_s W^3$), where the computed *AIC*, *R*, and *VAF* values for this model were respectively calculated at 238.217, 0.997, *and* 99.423. The correlation coefficient for SAELM 3 (the second-best model) equaled 0.996. The AIC index for the third-best model (SAELM 4), the fourth-best model (SAELM 1), and the fifth-best model

(SAELM 2) were respectively approximated to be 252.771, 254.259, and 258.008. The values of RMSE and NSC for SAELM 8 (the sixth-best) were about 1388.961 and 0.893. SAELM 5 was known as the fifth-best model in which the variance accounted for (*VAF*) and correlation coefficient (*R*) values were at 98.029 and 0.990. SAELM 7 represented the poorest performance at dealing with the simulation of the vertical reaction forces, with an NSC of 0.978. Amongst SAELM 9 to SAELM 15 models, SAELM 10 (a function of *Ds/W*) simulated the target values with the highest level of accuracy and correlation, so the ratio of gouge depth to gouge width was recognized as the most influencing input parameter to model the vertical reaction forces. By contrast, the position of the iceberg along the scour axis (*x/W*) showed up to be the least effective parameter to model the vertical reaction forces. SAELM 9 (a function of *x/W*) showed the worst results among the models with one input parameter. Moreover, the input parameters $L_v/\gamma_s.W^3$, $L_h/\gamma_s.W^3$, V^2/gW, $c/\gamma_s.W$, and α were classified in a range from second to sixth respectively, in terms of their influence to model the vertical reaction forces. Therefore, SAELM 6 was the most accurate and showed the highest level of correlation in order to estimation of the vertical reaction forces. Still regarding SAELM 6, the ratio of gouge depth to gouge width was the most influential input parameter.

Table 5-10. Comparison between the performances of SAELM models to simulate the vertical reaction forces

Model	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
SAELM 1	0.994	98.811	515.390	379.113	0.985	254.259
SAELM 2	0.992	98.441	574.906	419.319	0.982	258.008
SAELM 3	0.996	99.152	396.392	270.272	0.991	245.252

SAELM 4	0.995	99.067	493.520	349.677	0.986	252.771
SAELM 5	0.991	98.029	580.714	431.109	0.981	258.353
SAELM 6	0.997	99.423	322.908	260.777	0.994	238.217
SAELM 7	0.990	97.809	633.188	430.489	0.978	261.321
SAELM 8	0.966	81.614	1388.961	928.155	0.893	288.273
SAELM 9	-0.034	-2021.88	5074.151	4522.248	-0.431	332.724
SAELM 10	0.966	81.723	1387.301	924.233	0.893	288.232
SAELM 11	-0.002	-598.027	4980.19	4005.37	-0.379	332.082
SAELM 12	-0.002	-623.025	5012.235	4235.255	-0.211	345.258
SAELM 13	0.951	78.749	1512.42	1094.644	0.873	291.194
SAELM 14	0.951	78.777	1511.814	1094.532	0.873	291.180
SAELM 15	0.319	-242.73	4542.99	3680.212	-0.147	328.93

The calculated statistical indices for SAELM models to predict the horizontal subgouge displacements are arranged in Table 5-11. SAELM 5 demonstrated a better performance than other SAELM models, meaning that the *RMSE*, *VAF*, and *AIC* values for this model were 0.047, 99.936, and -25.085. SAELM 3 was the second-best model for the simulation of the horizontal deformations, with the *MAE* and *NSC* of 0.054 and 0.999. The performed sensitivity analysis proved that SAELM 1 was the third-best model where the value of *AIC* was at -22.025. The *VAF* and *MAE* values associated with the fourth-best model (SAELM 7) were respectively reckoned to be 99.901 and 0.043. The computed statistical indices showed that SAELM 2 was the fifth-best model (*RMSE*= 0.064 and *AIC*=-18.332). The estimated *NSC* and *MAE* criteria for the sixth-best model (SAELM 6) equaled 0.999 and 0.042. The values of variance accounted for (*VAF*) and

Akaike Information Criteria (*AIC*) for the worst SAELM model (SAELM 8) were respectively approximated as 84.588 and 26.450. According to the implemented sensitivity analysis for SAELM 9 to SAELM 15, the soil depth parameter (*y/W*) and the ratio of gouge depth to gouge width (*Ds/W*) were detected as the most effective inputs in estimating the horizontal subgouge deformations. SAELM 15 in terms of the velocity parameter ($V^2/g.W$) had the lowest level of correlation and precision, indicating that the $V^2/g.W$ input possessed the lowest level of effectiveness to model the horizontal displacements. Additionally, $L_v/\gamma_s.W^3$, $L_h/\gamma_s.W^3$, α, and $c/\gamma_s.W$ were classified in a range between third and sixth in terms of their relevance as input variables for this particular simulation. Thus, to estimate the horizontal subgouge deformations, SAELM 5 and the soil depth (*y/W*) were respectively identified as the premium model and the most substantial input variable.

Table 5-11. Comparison between the performances of SAELM models to simulate the horizontal subgouge deformations

Model	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
SAELM 1	0.999	99.917	0.054	0.036	0.999	-22.025
SAELM 2	0.999	99.898	0.064	0.044	0.999	-18.332
SAELM 3	0.999	99.916	0.054	0.040	0.999	-22.272
SAELM 4	0.999	99.434	0.154	0.061	0.993	0.177
SAELM 5	0.999	99.936	0.047	0.034	0.999	-25.085
SAELM 6	0.999	99.849	0.070	0.042	0.999	-16.462
SAELM 7	0.999	99.901	0.058	0.043	0.999	-20.537
SAELM 8	0.999	84.588	0.529	0.118	0.919	26.450

SAELM 9	0.999	99.915	0.059	0.040	0.999	-20.377
SAELM 10	0.999	84.634	0.529	0.118	0.919	26.444
SAELM 11	0.999	84.634	0.529	0.120	0.919	26.458
SAELM 12	0.999	84.403	0.530	0.131	0.919	26.494
SAELM 13	0.999	84.669	0.530	0.123	0.919	26.473
SAELM 14	0.999	84.598	0.529	0.122	0.919	26.467
SAELM 15	0.295	-760.043	1.778	0.557	0.086	52.245

Table 5-12 presents the calculated key criteria for SAELM models in order to estimate the vertical subgouge displacements. As shown in Figure 9, SAELM 3 was able to predict the vertical deformations with the highest correlation (R=0.952) and lowest complexity (AIC=-28.896). After SAELM 3, SAELM 5 was preferred as the second-best model, where the values of *VAF* and *MAE* equaled 77.781 and 0.010 respectively. Although SAELM 7 exhibited an acceptable performance, it was the third-best model to simulate vertical deformation ($RMSE$=0.015 and R=0.920). The *AIC* and *NSC* values of SAELM 2 were respectively equal to -23.595 and 0.801 making SAELM 2 the fourth-best model. The obtained *VAF* and *MAE* values for SAELM 6 (the fifth-best model) were about 77.393 and 0.012, while these indices for the sixth-best model (SAELM 4) were estimated to be 74.986 and 0.011. SAELM 1 in terms of all input factors was classified as the seventh-best model, with an Akaike Information Criteria of -22.963. SAELM 8 had an unsatisfactory performance to model the vertical ice-induced deformations since the values of *AIC*, *R*, and *NSC* for this model were respectively computed at -13.184, 0.530, and 0.217. Amongst SAELM 9 to SAELM 15, the lowest level of error and complexity was attributed to SAELM 9, so the soil parameter had also a significant influence in order to estimate the vertical deformations. In

addition, different components of load factor ($L_h/\gamma_s.W^3$ and $L_v/\gamma_s.W^3$) were found to be respectively the second and the third most influential input parameters. Based on the simulation results, the clay shear strength, the velocity factor, and the attack angle were ranked as the fourth, fifth and sixth most significant input variables. The ratio of gouge depth to gouge width exhibited a poor correlation with the vertical ice scouring displacements. Hence, SAELM 3 was introduced as the best model, and the soil depth parameter had a considerable impact on predicting the vertical subgouge deformations.

Table 5-12. Comparison between the performances of SAELM models to simulate the vertical subgouge deformations

Model	*R*	*VAF*	*RMSE*	*MAE*	*NSC*	*AIC*
SAELM 1	0.911	76.338	0.0159	0.012	0.784	-22.963
SAELM 2	0.939	86.418	0.015	0.012	0.801	-23.595
SAELM 3	0.952	90.322	0.011	0.008	0.901	-28.896
SAELM 4	0.911	74.986	0.015	0.011	0.795	-23.386
SAELM 5	0.915	77.781	0.014283	0.010	0.825	-24.581
SAELM 6	0.910	77.393	0.015	0.012	0.800	-23.569
SAELM 7	0.920	82.104	0.015	0.012	0.805	-23.743
SAELM 8	0.530	-190.847	0.030	0.020	0.217	-13.184
SAELM 9	0.619	-261.429	0.028	0.019	0.324	-14.307
SAELM 10	0.305	-525.673	0.034	0.023	-0.011	-11.245
SAELM 11	0.474	-335.679	0.031	0.021	0.189	-12.921
SAELM 12	0.351	-1820.84	0.033	0.021	0.082	-11.979

SAELM 13	0.548	-582.127	0.030	0.021	0.216	-13.182
SAELM 14	0.515	-350.095	0.030	0.021	0.204	-13.060
SAELM 15	0.432	-563.028	0.031	0.019	0.156	-12.614

5.1.15.4. Superior SAELM models

The performed sensitivity analysis showed that SAELM 5, SAELM 6, SAELM 5, and SAELM 3 were selected as the premium models in predicting the horizontal reaction forces, the vertical reaction forces, the horizontal deformations, and the vertical deformations, respectively. Thus, these models were carried on for further evaluation. The discrepancy ratio (*DR*) of the premium SAELM models was computed by using the equation below:

$$DR = \frac{\eta_{(Simulated)}}{\eta_{(Observed)}} \tag{5-115}$$

where, $\eta_{(Simulated)}$ and $\eta_{(Observed)}$ are respectively the simulated and observed subgouge parameters. The closer *DR* is to the unity, the better is the SEALM model l (Azimi and Shiri, 2020b). Figure 5-48 shows the results of discrepancy ratio (*DR*) for these SAELM models. For instance, the average discrepancy ratio (DR_{ave}) for the SAELM 5 and SAELM 6 models for simulation of the horizontal and vertical reaction forces were computed as 1.031 and 0.994 respectively. Additionally, the minimum discrepancy ratio (DR_{min}) for SAELM 5 to predict the horizontal subgouge deformations was 0.067 and the maximum discrepancy ratio (DR_{max}) for SAELM 3 in modeling the vertical ice-induced displacements equaled to 0.403. Therefore, the superior SAELM models could estimate the subgouge clay characteristics with a reasonable performance.

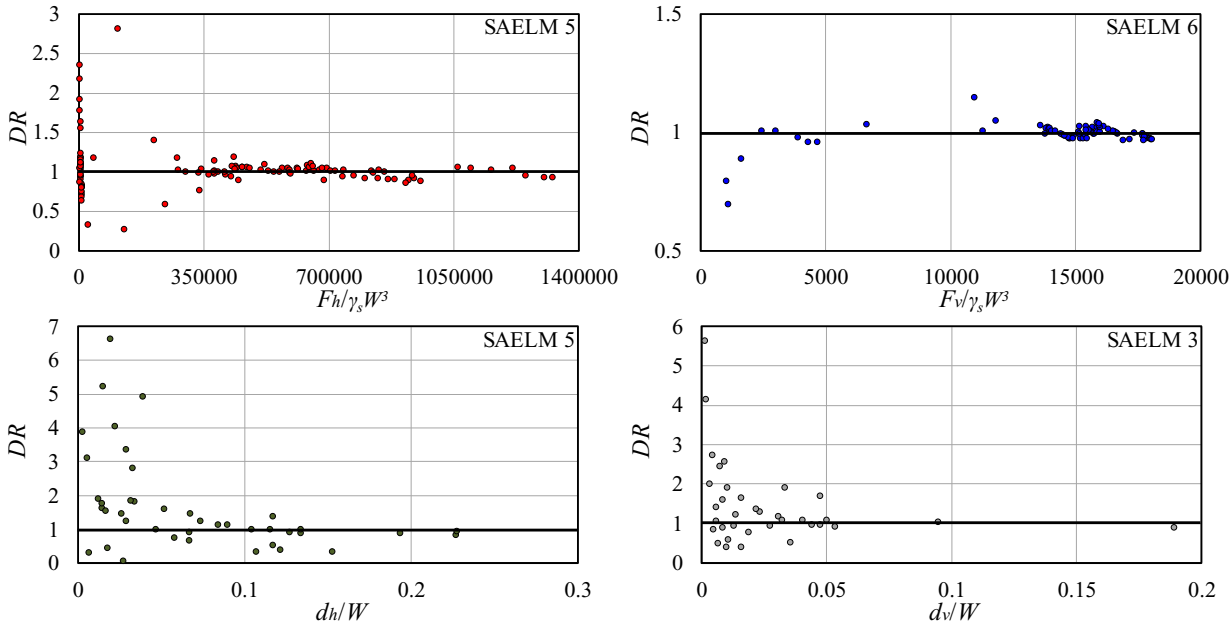


Figure 5-48. Results of discrepancy ratio (*DR*) for the best SAELM models

Additionally, the implemented error analysis for the SAELM models is depicted in Figure 5-49. For example, about 57% of the horizontal reaction forces obtained from SAELM 5 had an error of less than 10%, whereas this value for the vertical reaction forces was almost 95%. A minority of the vertical reaction forces predicted by SAELM 6 (roughly 2.5%) possessed an error greater than 20%. As shown, approximately one-fifth of the simulated horizontal displacements had an error of less than 10% (though this figure for the vertical deformations was nearly 34%). Moreover, a small part of the vertical ice-scoured displacements modeled by SAELM 3 (virtually 6%) indicated an error between 10% and 20%. Thus, the superior SAELM models showed a reasonable level of accuracy in order to simulate the subgouge clay features.

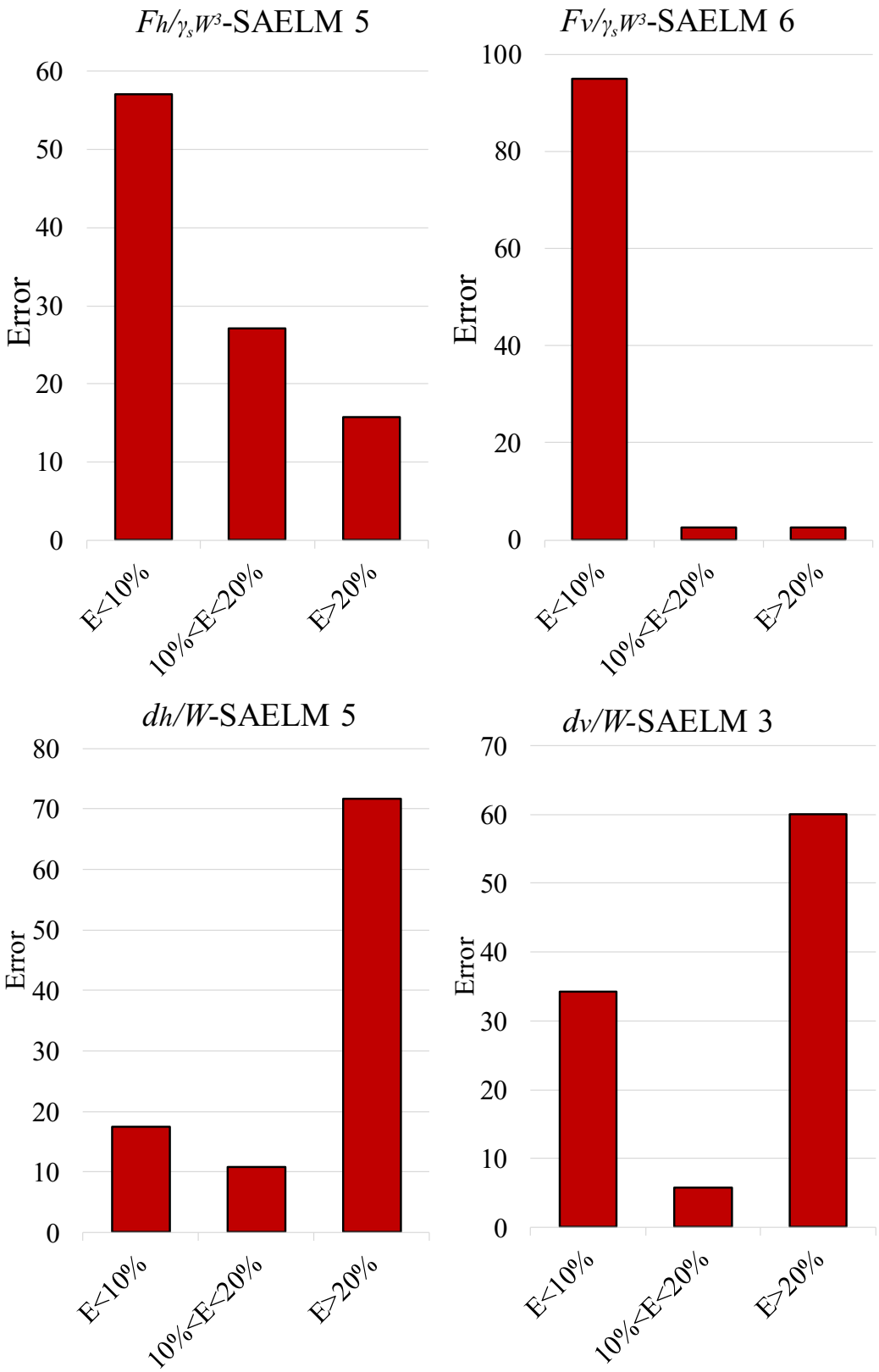


Figure 5-49. Results of error analysis for the best SAELM models

The uncertainty analysis (UA) was carried out to determine the performance of the best SAELM models. Results of the UA for the selected SAELM models are presented in Table 5-13. The errors obtained from the SAELM model (e_j) were commonly approximated as the difference between the predicted ice-scoured characteristics (P_j) and the experimental values (O_j) as follows:

$$e_j = P_j - O_j \tag{5-116}$$

The averaged estimated errors $(\bar{e})$ and the standard deviation (S_e) of errors were calculated using the equations below (Azimi and Shiri, 2020a):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{5-117}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{5-118}$$

A positive averaged estimated error $(\bar{e})$ means that the SAELM models overestimated the subgouge sand parameters, whilst a negative estimated error means that the SAELM models underestimated the experimental results. In the next step, regarding the "Wilson score approach" without the continuity correction, a confidence interval was built around the estimated error by using the $\bar{e}$ and S_e values l (Azimi and Shiri, 2020b). It is worthwhile noting that the Wilson score interval is an improvement over the normal distribution interval where an asymmetric normal distribution is utilized so as to rectify the confidence interval bounds. Moreover, a ±1.96Se brought about a 95% confidence interval (95%CI), and then the width of uncertainty bound (WUB) was calculated as follows:

$$\text{WUB} = \pm \frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{5-119}$$

Figure 5-50 shows the normal distribution graphs for the uncertainty analysis of the best SAELM models. Regarding the performed UA, SAELM 5 underestimated the horizontal reaction forces ($\bar{e} = -1572$), while SAELM 6 overestimated the vertical reaction forces ($\bar{e} = 2.9$). Furthermore, SAELM 5 ($\bar{e} = -0.0032$) and SAELM 3 ($\bar{e} = -0.0009$) slightly underestimated the values of horizontal and vertical displacements. The values of standard deviation of predicted error for SAELM 5, SAELM 6, SAELM 5, and SAELM 3 were respectively equal to 49800, 400.8, 0.356, and 0.011. Additionally, the widest uncertainty bound was obtained for SAELM 5 while modeling the horizontal reaction forces, with a WUB of ±4613.5, whereas SAELM 3 had the narrowest uncertainty bound (WUB=±0.002).

Table 5-13. Results of the uncertainty analysis for SAELM models

Parameter	*Fh*	*Fv*	*dh*	*dv*
Model	SAELM 5	SAELM 6	SAELM 5	SAELM 3
Number of observations	450	261	162	115
Mean	-1572	2.9	-0.0032	-0.0009
St.Dev	49800	400.8	0.3555	0.01145
SE Mean	2348	24.8	0.0279	0.00107
WUB	±4613.5	±48.85	±0.055	±0.002
95%CI	(-6185 to 3042)	(-46.0, 51.7)	(-0.0584, 0.0519)	(-0.00302, 0.00121)

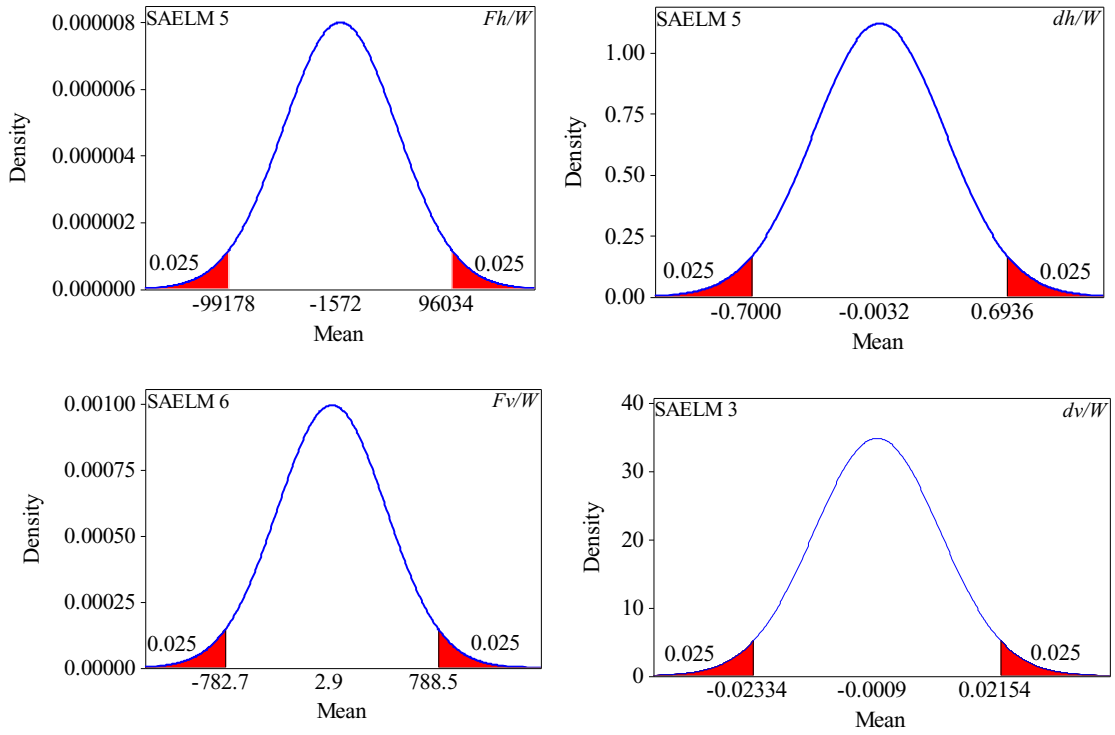


Figure 5-50. Normal distribution graphs for the performed uncertainty analysis of the best SAELM models

In Figures 5-51 to 5-54, the comparisons between the simulated subgouge parameters by the best SAELM models with experimental measurements are displayed. As shown, SAELM 5 in terms of x/w, D_s/W, $c/\gamma_s.W$, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW could model the horizontal reaction forces with the best performance. The minimum values of reaction forces were measured at the initial position of the keel (*x/W*=0), whereas the magnitude of the forces increased along the scour axis. Notwithstanding the oscillatory behavior in the experimental measurements, SAELM 5 utilized a monotonic trend to simulate the horizontal reaction forces (Figures 11-9). Likewise, the vertical reaction forces were almost zero at the beginning of scour axis and this parameter reached its highest value along the scour axis, after several fluctuations. In other words, SAELM 6, a function

of x/w , D_s/W , α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, V^2/gW, managed to model the vertical reaction forces by using a nonlinear trend with reasonable precision (Figures 5-51).

SAELM 5 simulated the horizontal ice-induced deformations in terms of the soil depth, ratio of gouge depth to gouge width, clay shear strength, horizontal and vertical loads, and velocity parameters. Furthermore, SAELM 5 had the lowest level of complexity and the highest level of accuracy. According to the simulation results, the maximum horizontal subgouge deformations were predicted under the position of the iceberg keel, and the magnitude of this parameter was reduced with increasing the soil depth (Figure 5-52). Even though some oscillations were reported in the applied experimental measurements, SAELM 3 was able to estimate the vertical subgouge displacements with good performance [Figure 5-54(b, c, o, and p)]. It means that SAELM 3 modeled the target function (*dv/W*) using the y/w , D_s/W , $c/\gamma_s.W$, α, $L_h/\gamma_s.W^3$, V^2/gW parameters, with a low level of complexity along with a high level of accuracy. Like the horizontal deformations profiles, the maximum vertical displacements were observed at the soil surface and then the magnitude of this parameter decreased at greater depths.

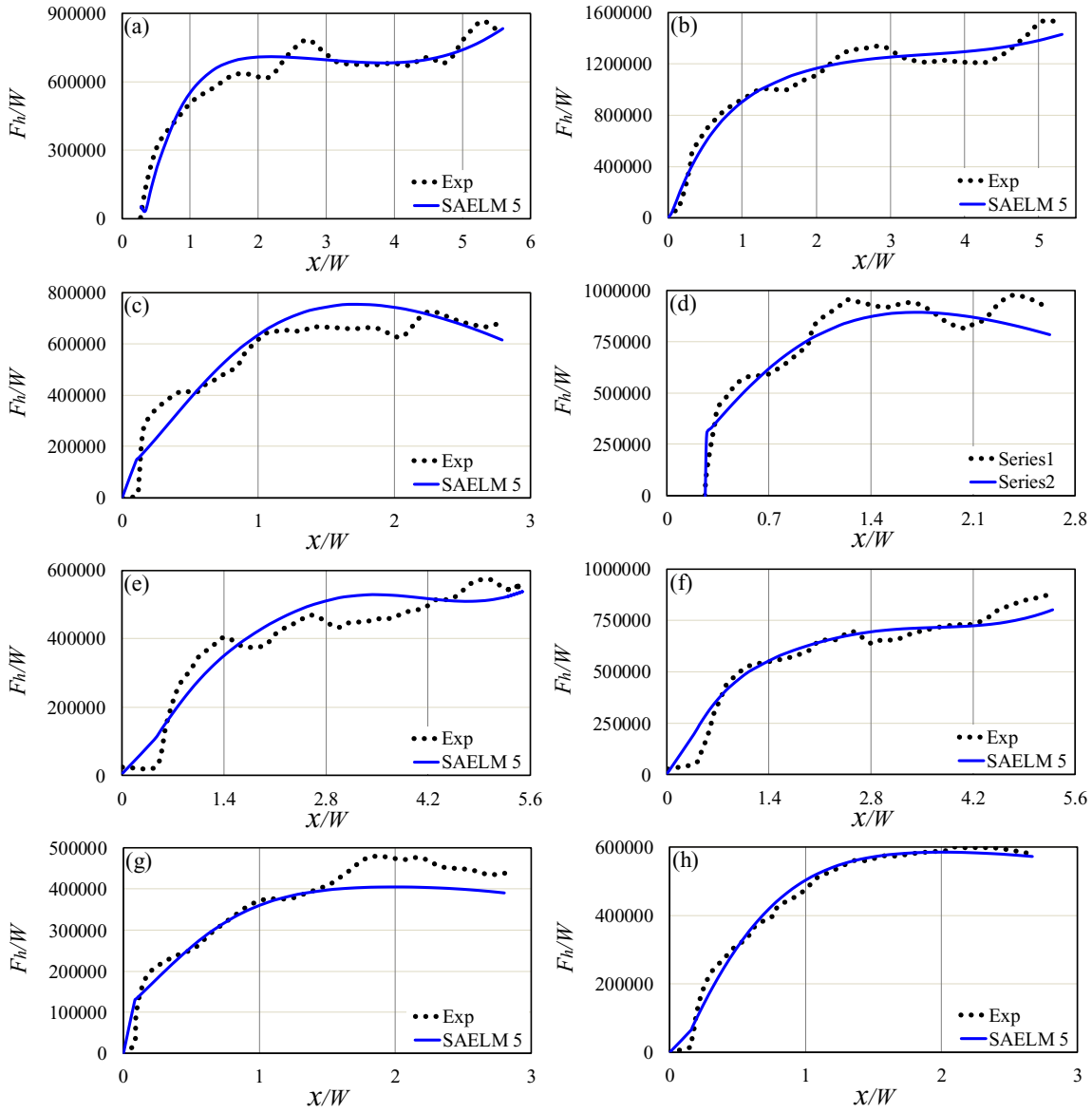


Figure 5-51. Comparison between the simulated horizontal reaction forces by SAELM 5 with experimental measurements

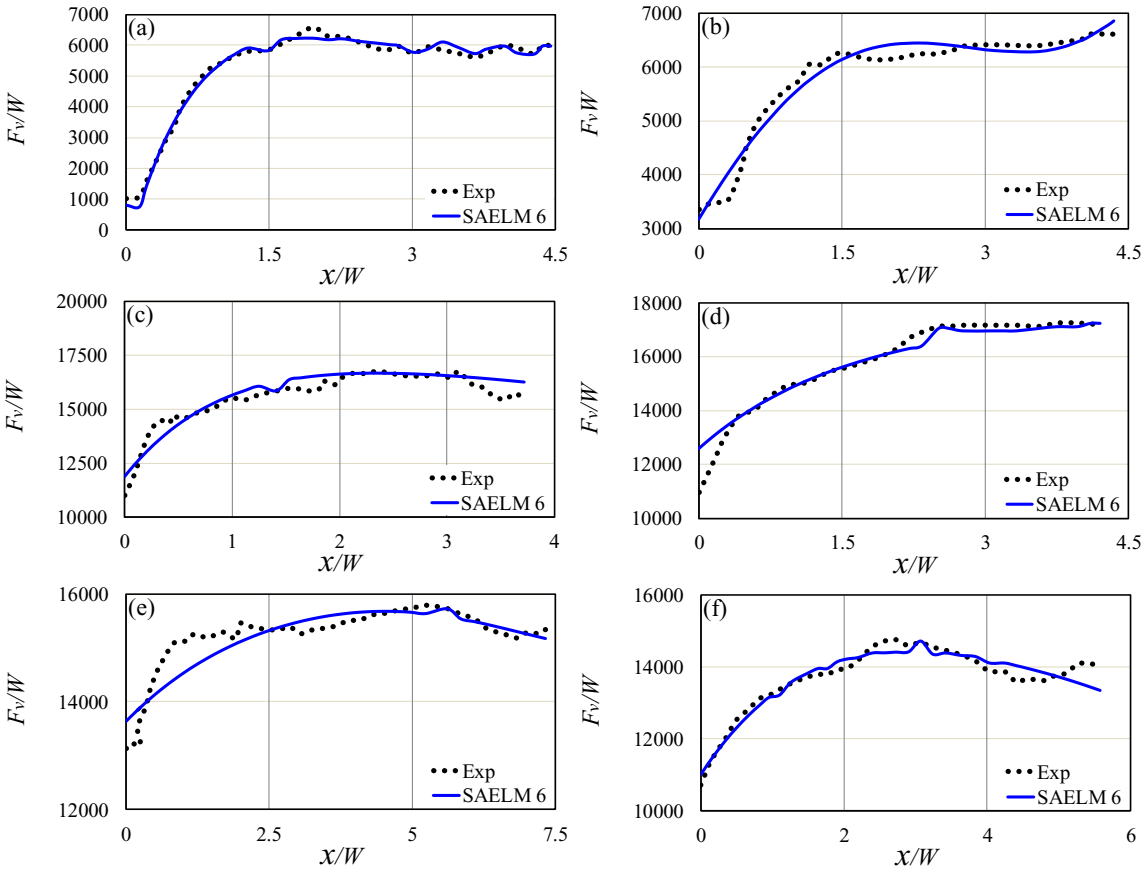


Figure 5-52. Comparison between the simulated vertical reaction forces by SAELM 6 with experimental measurements

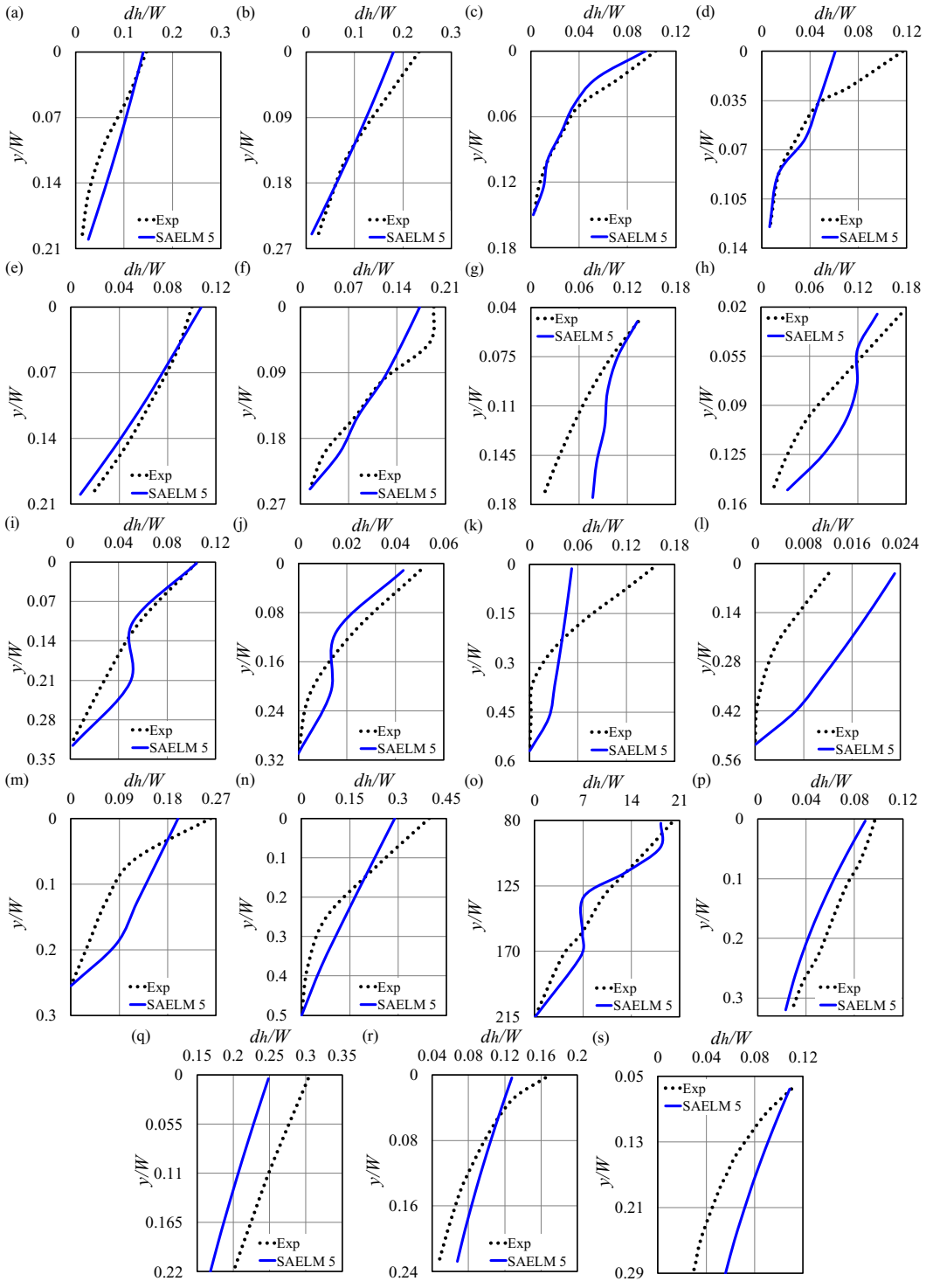


Figure 5-53. Comparison between the simulated horizontal deformations by SAELM 5 with experimental measurements

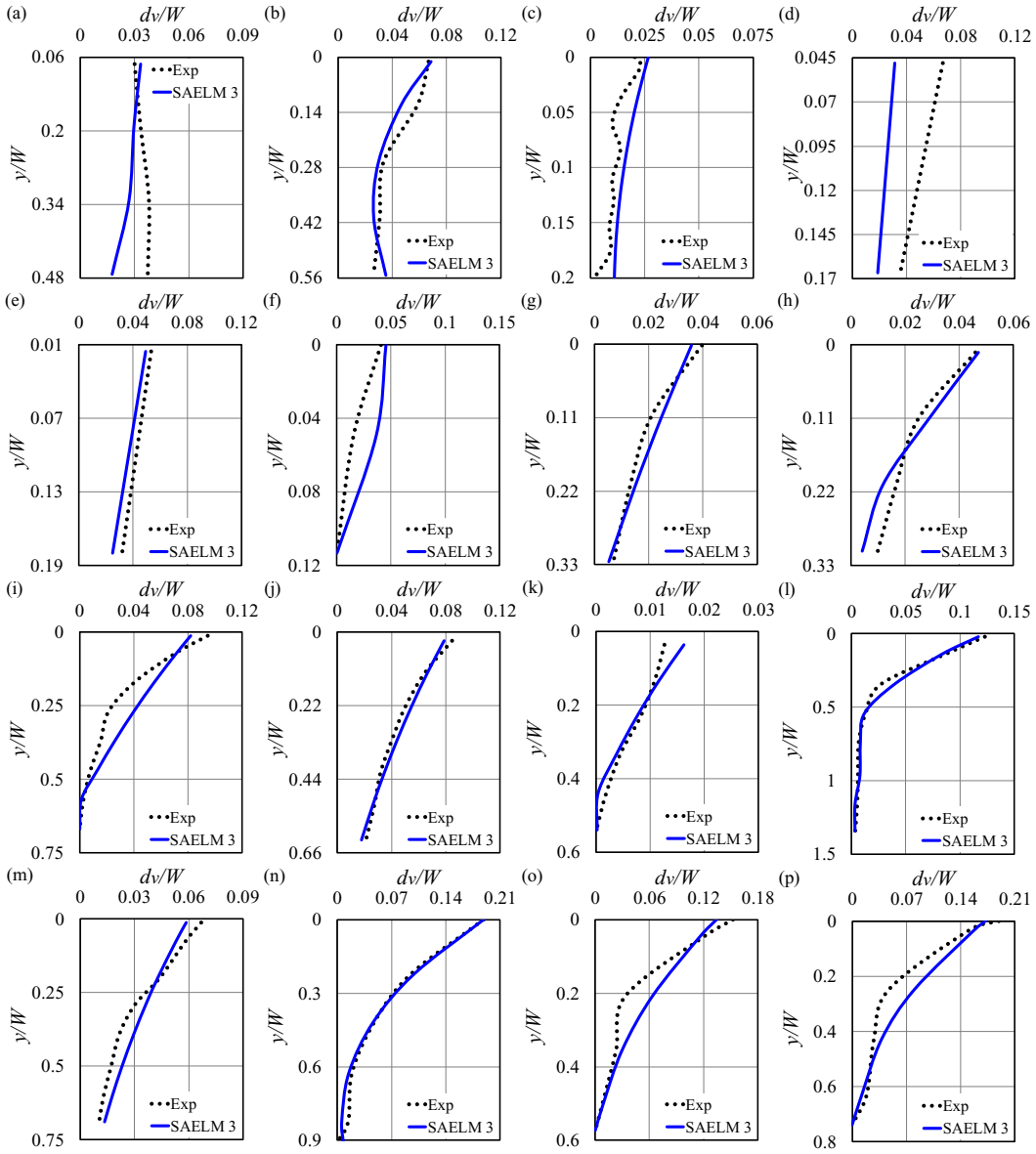


Figure 5-54. Comparison between the simulated vertical deformations by SAELM 3 with experimental measurements

The conducted analyses showed that the premium SAELM models can simulate the ice-induced clay parameters with a good level of accuracy, a low level of complexity along with a high level of correlation. The most significant input parameters were also introduced by performing a comprehensive sensitivity analysis.

5.1.16. Conclusion

In this study, the subgouge clay characteristics comprising the reaction forces and soil displacements were predicted by means of a non-tuned self-adaptive machine learning model called "self-adaptive extreme learning machine" (SAELM). Initially, using the parameters affecting the ice-induced clay seabed, fifteen SAELM models were defined. After that, by making use of published experimental results, a dataset was constructed and divided into two major sub-samples including the training sub-sample (70% of the dataset) and the testing sub-sample (30% of the rest). Subsequently, different analyses for the SAELM models were implemented. The most important results are summarized as follows:

- The most optimized number of neurons in the hidden layer was determined to be 20, with the correlation coefficient (*R*) and the Nash-Sutcliff efficiency coefficient (*NSC*) of 0.989 and 0.978.
- The sigmoid was selected as the best activation function for the SAELM network where the *AIC* and *VAF* indices were respectively equal to 674.927 and 98.236.
- SAELM 5, SAELM 6, SAELM 5, and SAELM 3 were respectively selected as the superior models to simulate the horizontal reaction forces, the vertical reaction forces, the horizontal subgouge deformations, and the vertical subgouge deformation.
- The horizontal component of load parameter and the ratio of gouge depth to gouge width were the most influencing inputs to model the reaction forces, whereas the soil depth variable had a significant impact in estimating the ice-induced deformation.

- The error analysis demonstrated that more than 90% of the vertical reaction forces predicted by SAELM 6 possessed an error of less than 10%, whilst this value for SAELM 3 was 34%.
- Regarding the uncertainty analysis, SAELM 6 overestimated the vertical reaction forces; however, SAELM 5 and SAELM 3 underestimated the other subgouge parameters.
- Finally, a set of SAELM-based equations were presented to approximate the ice-induced clay features.

Though several expensive experimental studies and long-running numerical simulations have been conducted in the field of ice scouring, the capability of self-adaptive machine learning for modeling this problem has not been utilized so much. The SAELM network demonstrated a superior generalization performance at dealing with simulation of the objective function overcoming the sensitivity to the number of neurons in the hidden layer; however, the computation time of this model was slightly longer than the classical form of ELM model. The obtained results proved that the SAELM algorithm is a precise, cost-effective, and quick alternative and could be used to simulate the subgouge clay parameters. The current study was the first footstep of the non-tuned self-adaptive machine learning application in the field of ice-gouging, further research on this theme is to be expected in the future. The present findings can support providing fast and cost-effective alternatives for simulation of ice-seabed interaction but more investigations still need to be conducted using other optimization algorithms and machine learning models to identify the most efficient and robust approach.

Appendix

SAELM-based equations

The common format of the self-adaptive extreme learning machine-based (SAELM) formula is:

$$\eta = \left[\frac{1}{1 + exp(InW \times InV \times BHI)}\right]^{T} \times OutW \tag{5-120}$$

where InW is the matrix of input weight, InV is the matrix of input parameters, BHI is the matrix of bias of hidden layer, and $OutW$ is the matrix of output weights. SAELM 5, SAELM 6, SAELM 5, and SAELM 3 were respectively selected as the premium SAELM models for the simulation of the horizontal reaction forces, the vertical reaction forces, the horizontal deformations, and the vertical deformations. The matrices of SAELM 5 to predict the horizontal reaction forces are presented below:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_s}{W} \\ \frac{c}{\gamma_s . W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, BHI = \begin{bmatrix} 0.137 \\ 0.672 \\ 0.429 \\ 0.367 \\ 0.759 \\ 0.276 \\ 0.887 \\ 0.125 \\ 0.249 \\ 0.209 \\ 0.034 \\ 0.465 \\ 0.600 \\ 0.433 \\ 0.201 \\ 0.120 \\ 0.823 \\ 0.665 \\ 0.588 \\ 0.341 \end{bmatrix}, InW = \begin{bmatrix} 0.804 & 1.368 & 0.283 & 0.219 & 0.452 & 0.338 \\ 0.709 & 0.288 & 0.087 & 1.131 & 0.106 & 0.508 \\ 0.106 & 0.848 & 0.172 & 0.396 & 0.377 & -0.014 \\ 0.439 & 0.254 & 0.549 & 0.565 & 0.603 & 0.208 \\ 0.382 & 0.786 & 0.394 & 1.110 & 0.775 & 0.843 \\ 0.362 & 0.967 & 0.785 & 0.113 & 0.094 & 0.537 \\ 0.685 & 0.303 & 0.905 & 0.436 & 0.982 & 1.632 \\ 0.919 & 0.596 & 0.570 & 0.382 & 0.387 & 0.285 \\ 0.346 & 0.804 & -0.539 & 0.041 & 0.379 & 0.889 \\ 0.427 & 0.268 & 0.414 & 0.137 & 0.464 & 0.426 \\ 0.421 & 0.996 & 0.456 & 0.440 & 0.044 & 0.037 \\ 0.222 & 1.472 & 0.693 & 1.476 & 0.281 & 0.737 \\ 0.546 & 1.079 & 0.551 & 0.968 & 1.546 & 0.224 \\ 0.424 & 0.659 & 0.872 & -0.535 & 0.854 & 0.090 \\ 0.430 & 0.363 & 0.156 & 0.660 & 0.633 & 1.801 \\ 0.373 & 0.278 & 0.709 & 1.115 & 0.465 & 0.955 \\ 0.547 & 0.724 & 0.632 & 0.230 & 0.015 & 0.861 \\ 0.750 & 0.885 & 1.268 & 1.254 & 0.473 & 0.907 \\ 0.470 & 0.081 & 0.038 & 0.636 & 1.624 & 0.081 \\ 0.928 & 0.141 & 0.887 & 0.014 & 0.659 & 0.439 \end{bmatrix} OutW \quad (5\text{-}121)$$

$$= \begin{bmatrix} 81454574.880 \\ 9858947.424 \\ -43398676.342 \\ 226126922.942 \\ 76594229.866 \\ 69992267.559 \\ -196450851.039 \\ -212768962.673 \\ 339950.056 \\ 634508844.353 \\ -393599219.977 \\ 166528636.712 \\ 249733803.590 \\ -239205854.575 \\ 82795238.535 \\ -305218224.322 \\ -68505659.283 \\ -36440008.877 \\ -330160337.170 \\ 227818062.419 \end{bmatrix}$$

Furthermore, the matrices of SAELM 6 to approximate the vertical reaction forces are:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_s}{W} \\ \alpha \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, BHI = \begin{bmatrix} -0.231 \\ 0.278 \\ 0.372 \\ 0.476 \\ -0.545 \\ -0.821 \\ -0.160 \\ -0.401 \\ 0.440 \\ -0.171 \\ -0.552 \\ 0.730 \\ 0.950 \\ 0.255 \\ -0.057 \\ -0.915 \\ -0.290 \\ -0.611 \\ 0.962 \\ 0.073 \end{bmatrix}, InW = \begin{bmatrix} 0.328 & -0.111 & -0.607 & -0.427 & -0.154 & 0.770 \\ 0.885 & -0.260 & -0.980 & 0.918 & 0.603 & -0.867 \\ -0.148 & -0.0005 & -0.953 & 0.962 & 0.009 & -0.858 \\ 0.376 & 0.944 & -0.745 & 0.583 & -0.547 & 0.813 \\ 0.247 & -0.375 & 0.325 & -0.591 & -0.832 & 0.975 \\ -0.078 & 0.046 & 0.995 & 0.382 & 0.678 & 0.939 \\ -0.102 & 0.894 & 0.445 & 0.910 & 0.687 & 0.965 \\ 0.533 & -0.748 & -0.423 & -0.795 & -0.340 & -0.756 \\ -0.621 & 0.721 & 0.997 & -0.566 & 0.041 & 0.431 \\ 0.615 & -0.572 & 0.237 & 0.375 & -0.335 & -0.745 \\ -0.341 & 0.635 & -0.451 & -0.455 & 0.452 & 0.811 \\ 0.603 & -0.117 & 0.208 & -0.828 & -0.294 & -0.888 \\ -0.055 & 0.254 & 0.637 & -0.413 & 0.970 & 0.127 \\ 0.957 & -0.431 & 0.863 & -0.040 & 0.032 & -0.144 \\ -0.154 & -0.907 & -0.762 & -0.580 & 0.223 & -0.966 \\ 0.323 & -0.711 & 0.286 & 0.006 & 0.015 & 0.291 \\ -0.262 & -0.884 & -0.860 & -0.947 & 0.061 & -0.889 \\ -0.812 & -0.671 & -0.383 & 0.370 & -0.445 & -0.815 \\ -0.505 & -0.693 & -0.114 & -0.139 & 0.487 & -0.450 \\ -0.625 & 0.155 & 0.088 & -0.472 & -0.986 & -0.560 \end{bmatrix} OutW \quad (5\text{-}122)$$

$$= \begin{bmatrix} 4710292681462.980 \\ -23130960.575 \\ -2712.673 \\ 811259256833.264 \\ -117105903408286 \\ -31950438000.231 \\ -15811860133.841 \\ -216018097792274 \\ 374.977 \\ -131721.752 \\ -20328.146 \\ 1161778663.675 \\ 36342192406.932 \\ 11056026674.790 \\ -841071.755 \\ -250395.332 \\ -10336542922678.700 \\ -2239866434.206 \\ 387498565.804 \\ -7457874661382.450 \end{bmatrix}$$

Likewise, the matrices of SAELM 5 to calculate the ice-induced horizontal deformations are:

$$InV = \begin{bmatrix} \frac{y}{W} \\ \frac{D_S}{W} \\ \frac{c}{\gamma_s.W} \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix}, \; BHI = \begin{bmatrix} -0.676 \\ -0.042 \\ -0.136 \\ -0.347 \\ 0.595 \\ 0.733 \\ 0.885 \\ 0.736 \\ 0.571 \\ -0.346 \\ -0.609 \\ 0.989 \\ 0.994 \\ -0.497 \\ -0.345 \\ 0.670 \\ -0.581 \\ -0.581 \\ -0.139 \\ 0.440 \end{bmatrix}, InW = \tag{5-123}$$

$$\begin{bmatrix} 0.240 & 0.282 & -0.064 & -0.775 & -0.605 & 0.035 \\ -0.515 & 0.899 & -0.174 & -0.749 & -0.479 & -0.669 \\ -0.232 & -0.479 & 0.227 & -0.502 & -0.464 & 0.186 \\ 0.458 & -0.734 & -0.143 & 0.757 & -0.637 & -0.549 \\ 0.927 & -0.666 & -0.739 & 0.598 & -0.285 & 0.817 \\ -0.080 & -0.631 & -0.177 & 0.737 & 0.874 & 0.101 \\ 0.695 & -0.627 & 0.568 & 0.781 & 0.751 & 0.292 \\ 0.326 & 0.314 & -0.879 & 0.182 & 0.134 & -0.577 \\ -0.536 & -0.298 & -0.272 & 0.591 & -0.379 & 0.746 \\ -0.139 & -0.878 & 0.778 & 0.312 & -0.482 & -0.871 \\ 0.483 & 0.604 & 0.401 & 0.980 & 0.770 & 0.783 \\ -0.490 & 0.377 & -0.521 & 0.611 & -0.917 & -0.450 \\ -0.191 & -0.454 & -0.527 & 0.621 & -0.712 & 0.339 \\ -0.569 & -0.305 & 0.990 & 0.323 & -0.157 & -0.419 \\ -0.071 & -0.815 & 0.903 & -0.684 & 0.003 & -0.790 \\ -0.330 & 0.203 & -0.737 & 0.535 & -0.653 & 0.557 \\ 0.375 & 0.631 & 0.510 & -0.880 & 0.812 & 0.218 \\ -0.918 & -0.968 & -0.356 & -0.041 & -0.087 & 0.007 \\ -0.046 & 0.647 & -0.582 & 0.311 & 0.363 & -0.618 \\ -0.139 & 0.602 & 0.850 & 0.520 & -0.145 & -0.709 \end{bmatrix} \; OutW = \begin{bmatrix} -0.676 \\ -0.042 \\ -0.136 \\ -0.347 \\ 0.595 \\ 0.733 \\ 0.885 \\ 0.736 \\ 0.571 \\ -0.346 \\ -0.609 \\ 0.989 \\ 0.994 \\ -0.497 \\ -0.345 \\ 0.670 \\ -0.581 \\ -0.581 \\ -0.139 \\ 0.440 \end{bmatrix}$$

And, the matrices of SAELM 3 to predict the subgouge vertical displacements in a clay seabed are:

$$InV = \begin{bmatrix} \frac{x}{W} \\ \frac{D_s}{W} \\ \alpha \\ \frac{L_h}{\gamma_s W^3} \\ \frac{L_v}{\gamma_s W^3} \\ \frac{V^2}{gW} \end{bmatrix},\ BHI = \begin{bmatrix} 0.331 \\ 0.348 \\ 0.139 \\ 0.832 \\ 0.763 \\ 0.109 \\ 0.708 \\ 0.823 \\ 0.508 \\ 0.176 \\ 0.252 \\ 0.832 \\ 0.977 \\ 0.115 \\ 0.426 \\ 0.047 \\ 0.094 \\ 0.204 \\ 0.957 \\ 0.909 \end{bmatrix}, InW = \begin{bmatrix} 0.683 & 0.387 & 0.061 & 0.912 & 0.131 & 0.087 \\ 0.182 & 0.630 & 0.100 & 0.713 & 0.090 & 0.573 \\ 0.251 & 0.660 & 0.125 & 0.995 & 0.053 & 0.027 \\ 0.894 & 0.720 & 0.502 & 0.178 & 0.453 & 0.167 \\ 0.133 & 0.012 & 0.899 & 0.224 & 0.037 & 0.403 \\ 0.131 & 0.446 & 0.360 & 0.095 & 0.942 & 0.030 \\ 0.851 & 0.703 & 0.952 & 0.396 & 0.114 & 0.151 \\ 0.360 & 0.883 & 0.404 & 0.593 & 0.118 & 0.034 \\ 0.639 & 0.687 & 0.741 & 0.062 & 0.026 & 0.363 \\ 0.499 & 0.001 & 0.461 & 0.214 & 0.485 & 0.396 \\ 0.807 & 0.543 & 0.040 & 0.002 & 0.065 & 0.324 \\ 0.809 & 0.278 & 0.910 & 0.177 & 0.069 & 0.336 \\ 0.154 & 0.421 & 0.019 & 0.124 & 0.024 & 0.029 \\ 0.244 & 0.511 & 0.042 & 0.933 & 0.103 & 0.962 \\ 0.029 & 0.228 & 0.060 & 0.143 & 0.552 & 0.165 \\ 0.962 & 0.097 & 0.036 & 0.287 & 0.958 & 0.007 \\ 0.349 & 0.771 & 0.362 & 0.112 & 0.187 & 0.127 \\ 0.222 & 0.009 & 0.173 & 0.849 & 0.108 & 0.803 \\ 0.633 & 0.116 & 0.646 & 0.008 & 0.482 & 0.206 \\ 0.753 & 0.225 & 0.189 & 0.856 & 0.579 & 0.910 \end{bmatrix} OutW = \begin{bmatrix} 158656.805 \\ 749.405 \\ 76.735 \\ 7.806 \\ 56.786 \\ 3.743 \\ 122.843 \\ 0.057 \\ 3.900 \\ 2.053 \\ 0.439 \\ 14.120 \\ 0.558 \\ 0.246 \\ 11.314 \\ 23.141 \\ 2.825 \\ 778237.724 \\ 0.864 \\ 158766.616 \end{bmatrix} \quad (5\text{-}124)$$

References

Abdalla, B., Pike, K., Eltaher, A., Jukes, P., Duron, B., 2009. Development and validation of a coupled Eulerian Lagrangian finite element ice scour model. In: 28th International Conference on Ocean, Offshore and Arctic Engineering, 43451, pp. 87–95.

Arnau Almirall, S., 2017. Ice-gouging in sand and the associated rate effects. University of Aberdeen, UK.

Azimi, H., Shiri, H., 2020a. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142: 051601.

Azimi, H., Shiri, H., 2020b. Ice-seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice-seabed interaction in sand by using extreme learning machine. Natural Hazards, 106 (3): 2307–2335.

Azimi, H., Shiri, H., 2021b. Modeling subgouge sand deformations by using multi-layer perceptron neural network. 31st International Ocean and Polar Engineering Conference, ISOPE-I-21-2150.

Bailey Dudley, E., Liu, L., Sarracino, R., Taylor, R., 2017. Using discrete element model to simulate Keel-Gouging: a sensitivity analysis. In the 36th International Conference on Ocean, Offshore and Arctic Engineering, V008T07A022: OMAE2017-62479.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour displacement in clays from physical model tests. In the 7th International Pipeline Conference, 48609: 239–245.

Bonakdari, H., Qasem, S.N., Ebtehaj, I., Zaji, A.H., Gharabaghi, B., Moazamnia, M., 2020. An expert system for predicting the velocity field in narrow open channel flows using self-adaptive extreme learning machines. Measurement, 151: 107202.

Cao, J., Lin, Z., Huang, G.B., 2012. Self-adaptive evolutionary extreme learning machine. Neural processing letters, 36: 285–305.

C-CORE., 1995. Phase 3: centrifuge modelling of ice Keel Scour, C-CORE Report 95-Cl2.

C-CORE., 1996. PRISE Phase 3c: extreme LEE Gouge event—modeling and interpretation, C-CORE Report 96-C32.

Comfort, G., Graham, B., 1986. Evaluation of sea bottom ice scour models (No. 37). Environ. Stud. Revolving Funds.

Ebtehaj, I., Bonakdari, H., Shamshirband, S., 2016. Extreme learning machine assessment for estimating sediment transport in open channels. Engineering with Computers, 32(4): 691–704.

Ebtehaj, I., Sattar, A.M, Bonakdari, H., Zaji, A.H., 2017. Prediction of scour depth around bridge piers using self-adaptive extreme learning machine. Journal of Hydroinformatics, 19(2): 207–224.

Hertz, J.A., 2018. Introduction to the theory of neural computation. CRC Press.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In: 2004 IEEE International Joint Conference on Neural Networks, IEEE Cat. No. 04CH37541, 2. IEEE, pp. 985–990.

Kioka, S., Yasunaga, Y., Watanabe, Y., Saeki, H., 2000. Evaluation of ice forces acting on sea bed due to ice scouring. 10th International Offshore and Polar Engineering Conference, ISOPE-I-00-111.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and generalization of experimental values of ice Scour event by a neural-network. 13th International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S.D., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of Scour depths. 14th International Offshore and Polar Engineering Conference, ISOPE-I-04-107.

Lach, P.R., 1996. Centrifuge modelling of large soil deformation due to ice scour. Memorial University of Newfoundland. St. John's. NL, Canada.

Liu, L., Bailey, E., Sarracino, R., Taylor, R., Power, C., Stanbridge, C., 2015. Numerical simulation of ice ridge gouging. 34th International Conference on Ocean, Offshore and Arctic Engineering, 56567: V008T07A022

Machemehl, J.L., Jo, C.H., 1989. Note on nearshore ice gouge depths in Alaskan Beaufort Sea. Journal of cold regions engineering, 3: 150–153.

Nematzadeh, A., Shiri, H., 2019. Large deformation analysis of ice scour process in dense sand. 10th Congress on Numerical Methods in Engineering.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Sattar, A.M., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. Neural Computing and Applications, 31: 157–169.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip line field solutions as an approach to understand ice subgouge deformation patterns. In the16th International Offshore and Polar Engineering Conference, ISOPE-I-06-289.

Shin, M.B., Park, D.S., Seo, Y.K., 2019. Comparative study of ice gouge simulation considering ice keel-seabed interactions. Journal of Ocean Engineering and Technology, 33: 556–563.

Storn, R., Price, K., 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 11: 341–359.

Woodworth-Lynes, C., Nixon, D., Phillips, P., Palmer, A., 1996. Subgouge deformations and the security of Arctic marine pipelines. Offshore Technology Conference, OTC-8222-MS.

6. Chapter 6

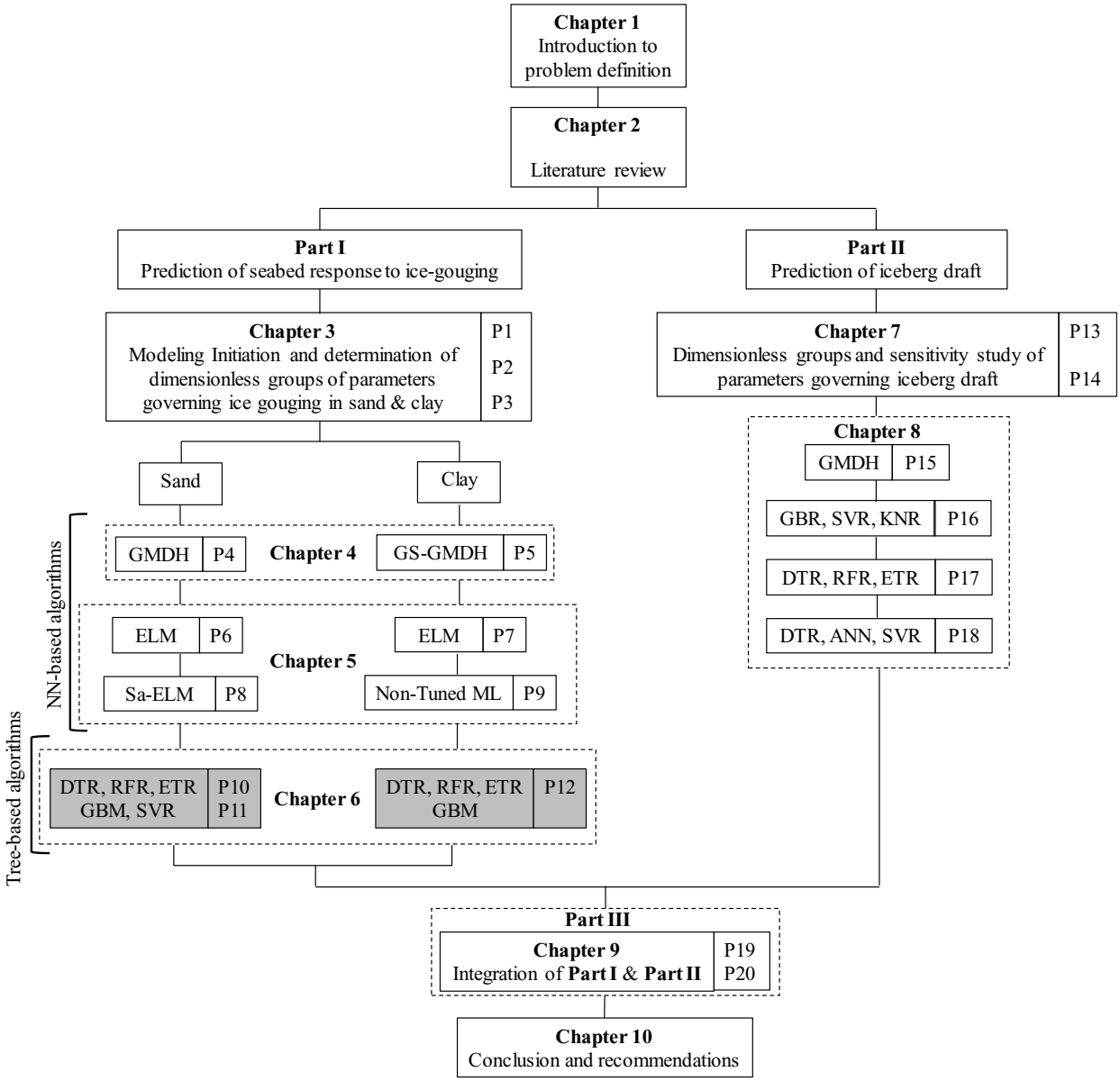
Prediction of Iceberg-Seabed Interaction Process Using Tree-Based Machine Learning Algorithms

This chapter includes three sections as follows:

Section 1: Simulation of subgouge sand deformations using robust machine learning algorithms

Section 2: Iceberg-seabed interaction analysis in sand by a random forest algorithm

Section 3: Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms


Chapter 1
Introduction to problem definition
Chapter 2
Literature review
Part I
Prediction of seabed response to ice-gouging
Part II
Prediction of iceberg draft
Chapter 3
Modeling Initiation and determination of dimensionless groups of parameters governing ice gouging in sand & clay
P1
P2
P3
Chapter 7
Dimensionless groups and sensitivity study of parameters governing iceberg draft
P13
P14
Chapter 8
GMDH P15
GBR, SVR, KNR P16
DTR, RFR, ETR P17
DTR, ANN, SVR P18
Sand
Clay
NN-based algorithms
GMDH P4
Chapter 4
GS-GMDH P5
ELM P6
ELM P7
Chapter 5
Sa-ELM P8
Non-Tuned ML P9
Tree-based algorithms
DTR, RFR, ETR GBM, SVR P10 P11
Chapter 6
DTR, RFR, ETR GBM P12
Part III
Chapter 9
Integration of Part I & Part II
P19
P20
Chapter 10
Conclusion and recommendations

Section 1

Simulation of subgouge sand deformations using robust machine learning algorithms

This section was published at the Offshore Technology Conference (OTC), 2022, Houston, Texas, US

Abstract

Ice-gouging is one of the major menaces to the subsea pipelines crossing the Arctic (e.g., Beaufort Sea) or the non-Arctic (e.g., Caspian Sea) shallow waters. Burial of the sea-bottom-founded infrastructures is regarded as a feasible method for protection of the subsea assets against the ice-gouging threat. These pipelines are commonly embedded underneath the deepest ice-scoured records in the area, whereas the pipeline system is still threatened by the ice-induced soil displacement developed into the ice tip owing to the shear resistance of the seabed soil.

Determination of the subgouge soil displacements is a governing design factor for the subsea structures in the Arctic offshore that commonly need costly laboratory studies and long-running finite element (FE) analyses to guarantee the operational integrity of the subsea pipeline against the ice-gouging event. Thus, the industry is still seeking more cost-effective, reliable, and faster alternative approaches for simulation of the iceberg-seabed-pipeline interaction process to minimize the collision risk of ice keels with the subsea structures. Recently, the application of machine learning (ML) in different fields has witnessed impressive growth since the ML technology is sufficiently precise, quick, reliable, and cost-effective to model various linear and non-linear problems.

In this study, three robust ML algorithms comprising the Decision Tree Regression (DTR), Random Forest Regression (RFR), and Extra Tree Regression (ETR) models were used for the first time to simulate the iceberg-seabed interaction process in the sandy seabed. Using the parameters governing the ice-seabed interaction mechanism, a set of the DTR, RFR, and ETR models were developed. To verify the ML models, a comprehensive dataset was constructed and the data was divided into two sub-samples including the training (70% of data) and testing (the remaining 30% of the data) datasets. Subsequently, for the DTR, RFR, and ETR models, several

analyses such as sensitivity analysis, error analysis, and uncertainty analysis were performed. The conducted analyses demonstrated that the ETR algorithm had a reasonable performance to simulate both horizontal and vertical subgouge soil deformations in the sand. The soil depth ratio (y/W) and the horizontal load factor ($Lh/\gamma s.W^3$) had substantial significance to model the horizontal and vertical deformations in the present study.

The presented results provided a good notion of modeling the ice-gouging problem through the ETR algorithm. The outcomes may facilitate proposing new solutions to estimate the subgouge soil deformations in the sandy seabed. The present work can also be used for the planning of expensive field, laboratory, and FE simulations and to reduce expenditures on future studies.

Keywords: Iceberg-seabed interaction process; Seabed sand; Decision Tree Regression (DTR), Random Forest Regression (RFR), Extra Tree Regression (ETR); Uncertainty analysis (UA)

6.1.1. Introduction

The past decades have witnessed the rapid exploitation of crude oil and natural gases in the Arctic territories, where a huge amount of these hydrocarbons is transferred from offshore to onshore through the subsea pipelines safely and economically. However, ice scouring is one of the primary hazards to the subsea pipelines crossing the Arctic area, when the traveling icebergs may gouge the seabed and collide with the sea bottom-founded infrastructures in shallow waters during warmer months. Figure 6-1 demonstrates the layout of the ice-seabed interaction mechanism.

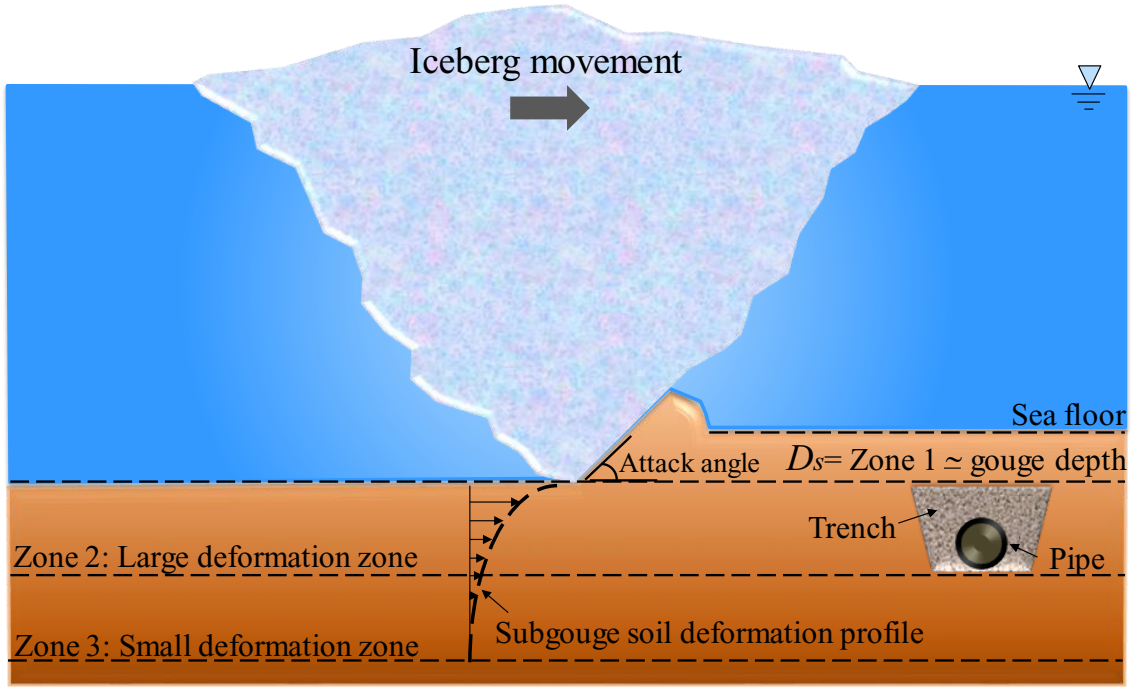


Figure 6-1. Layout of the ice-seabed interaction mechanism

To safeguard the subsea pipelines against ice scouring, these marine pipelines are embedded below the maximum estimated gouge depth. It is worth noting that the soil displacements are not limited to the soil in front of the ice keel, rather the soil shear resistance displaces the subgouge soil much deeper than the ice tip. This subgouge soil deformation has an exponential distribution through the soil (Figure 12-1) underneath the ice tip (ISO 19906: 2019 E) and can still threaten buried

pipelines. The estimation of the depth of the subgouge deformation zone underneath the ice tip is a burdensome and costly design aspect that is generally performed by a decoupled simulation.

In the decoupled simulation, initially, free-field iceberg-seabed interaction analysis is modeled without the presence of the pipeline. Subsequently, the deformed sub-gauge soil is manually transferred to the beam-spring model to simulate the pipe response to soil displacements. As shown, the free field iceberg-seabed interaction process is a significant phase of the decoupled simulation in practical pipeline design and relies on costly laboratory and long-running finite element (FE) analyses.

The past years have seen increasingly rapid advances in the application of machine learning (ML) in different fields although the ML models have seen limited use in the analysis of the ice-seabed interaction phenomenon. As an example, Kioka et al. (2003, 2004) used a neural network-based (NN) model to estimate the ice-gouging problem. In the NN model, five hidden neurons were embedded and the ice drift velocity, sand property, sea-bottom gradient, and attack angle were assumed as the input parameters. The authors validated this model with a mechanical method and showed that it was able to approximate the ice-seabed interaction features with good accuracy. Azimi and Shiri (2020a) detected the dimensionless group governing the iceberg-seabed interaction parameters for both sandy and clay seabeds. The authors proposed a set of relationships to calculate the maximum horizontal and vertical subgouge deformations. A gene expression programming-based approach was applied to simulate the horizontal subgouge soil deformations in the sand by Azimi and Shiri (2020b). The study highlighted that the gouge depth and the dilation angle had a significant effect on the predicted horizontal subgouge soil displacements. Azimi and Shiri (2021a) simulated the subgouge soil characteristics in sandy seabed through the extreme learning machine (ELM) algorithm. The authors showed that the gouge depth ratio and the shear

strength of the seabed soil were the most important input parameters to model the objective parameter. The horizontal subgouge deformations in the sand were modeled using a multi-layer perceptron neural network algorithm by Azimi and Shiri (2021b). The authors employed six neurons in the hidden layer and used the sigmoid as the activation function. Azimi and Shiri (2021c) combined the extreme learning machine (ELM) and the differential evolution (DE) algorithm to simulate the ice-seabed interaction process. The study concluded that roughly 30% of the horizontal subgouge deformations estimated by the SaE-ELM algorithm had an error of less than 10%, and almost 43% of the results showed an error between 10% and 20%.

The present investigation utilized extra tree regression (ETR) as a robust ML algorithm and compared the results with the random forest regression (RFR) and decision tree regression (DTR) as other alternatives for estimation of the horizontal and vertical subgouge soil deformations to be fed into the beam-spring models.

The dataset for training and testing of the models was constructed by using six well-recognized experimental studies in sand comprising Paulin (1991, 1992), C-CORE (1995, 1996), Hynes (1996), and Yang (2009). Paulin (1991, 1992) performed different experimental tests for submerged and dry sand to measure the reaction forces and deformations during the ice-scouring phenomenon. The author asserted that the loads and reaction forces for the submerged sand were smaller than in the dry seabed circumstances. Several Pressure Ridge Ice Scour Experiment (PRISE) tests at the Center for Cold Ocean Resources Engineering (C-CORE) were implemented to quantify the subgouge deformations and keel reaction forces in both clay and sand seabed (C-CORE 1995, 1996). These studies showed that the horizontal reaction forces increased by growing the width of scouring, rather the horizontal reaction forces decreased with the steeper keel attack angle. Hynes (1996) investigated the ice-seabed interaction process in a centrifuge study. The

author reported that the scour depth and scour load had a linear correlation. Yang (2009) conducted various Pipeline Ice Risk Assessment and Mitigation (PIRAM) tests on the sandy seabed. The author evaluated the influence of gouge depth, frontal berm height, and gouge rate and concluded that the attack angle and gouge geometry significantly affected the reaction forces.

The ETR algorithm presented in the current study provides fast estimations of the horizontal and vertical subgouge soil deformations, which can be used in the initial stages of engineering design, for planning the construction methodologies, logistics, and the scope of detailed engineering. This method is fast and cost-effective as compared to centrifuge testing or continuum FE analyses. More details about the proposed methodology will be presented in the upcoming sections.

6.1.2. Methodology

The three types of regression used in this study are described in the following sections.

6.1.2.1. Decision Tree Regression (DTR)

A tree data structure comprising a set of nodes and branches in every single node is considered a decision tree (DT). The DT can be applied for the estimation of the classification and regression problems. The DT consists of several sections including a root node, some leaf nodes, internal nodes, and branches. The uppermost node in the tree is defined as the root node and the terminal nodes (leaf nodes) end with the labels of classes, whereas the non-leaf nodes are regarded as the internal nodes. These nodes connect to each other by using the branches. In the current investigation, the mean squared error (*MSE*) is applied to control the fitness function in the DT algorithm (Pekel 2020). The value of DTR parameters including the *max_depth*, *max_features*, *max_leaf_nodes*, *min_samples_leaf*, *min_weight_fraction_leaf*, and *splitter* was respectively

adjusted as 8, *'auto'*, 12, 1, 0.01, and *'random'*. It should be asserted that these parameters were found using a trial-and-error procedure.

6.1.2.2. Random Forest Regression (RFR)

The random forest (RF) algorithm has been developed by Breiman (2001) to solve classification and regression problems. The RF algorithm, as an ensemble learning method, is the extension of the Classification and Regression Trees (CART) algorithm where the tree-based CART models have the overfitting problem and the RF algorithm overcomes this limitation (Breiman 2001). In the RF algorithm, several decision trees are created and decisions with the highest number of votes are chosen as the simulation results (Sahani and Ghosh 2021). In the random forest regression, the number of trees for a random vector is increased and the tree predictor, input parameters, and output parameters are considered as numerical values. Regarding the random vector distribution, the training dataset is independently utilized. The RFR model reduces the average error of the trees utilized by means of the weighted correlation between the residuals and the applied randomization ought to be employed at the low level of correlation. In the investigation, the parameters applied in the RFR algorithm comprising the *max_depth*, *random_state*, *number of estimators*, and *verbose* were tuned by a trial and error method as 5000, 500, 1000, and 1, respectively.

6.1.2.3. Extra Tree Regression (ETR)

ETR was originally driven from the RF algorithm by Geurts et al. (2006). The RF algorithm utilizes the bootstrapping, generation of the decision trees, and bagging, division of the decision tree leaves, in order to solve a regression problem. The ETR commonly generates a set of unexpected regression trees. By choosing the best subset, the decision-making procedure is implemented. The

ETR and RF algorithms have two main differences: (1) In the ETR model, the leaves are split using the random selection from the cutting points (2) The ETR algorithm utilizes the complete training dataset in order to generate the trees to minimize the value of bias. The ETR algorithm (i) applies the number of features that are selected randomly in the leave (ii) minimizes the sample size for the separation of leaves. The feature selection strength is determined using (i) and (ii) mechanisms; rather the strength of the averaged outcome noise is defined through (ii) procedure. These mechanisms can lessen the overfitting problem in the ETR algorithm and enhance the performance of this model (Hammed et al. 2021). It should be stated that the ETR parameters were tuned through a trial-and-error approach in the current study. This means that the value of *n_estimators*, *random_state*, *criterion*, *max_depth*, *min_sample_split*, and *min_sample_leaf* was regulated to be 500, 12, *'mse'*, 2000, 4, and 2 respectively.

6.1.2.4. Ice-seabed interaction process

The maximum subgouge soil deformation in the sand ($d_{(sand\_max)}$) is a function of a set of parameters including the scour depth (D_s), the internal friction angle of sand (φ), the width of gouge (W), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (h'), the horizontal load (L_h), the vertical load (L_v), the velocity of ice keel (V), and the specific weight of sand (γ_s) (Azimi and Shiri 2020a):

$$d_{(sand\_max)} = f_1(D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \quad (6\text{-}1)$$

Equation (6-1) can be written as a function of eight dimensionless groups as below (Azimi and Shiri 2020a):

$$\frac{d_{(sand\_max)}}{W} = f_2\left(\frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}2)$$

It is worth noting that the maximum subgouge deformation in the seabed sand $\left(d_{(sand\_max)}\right)$ is formed just under the moving ice keel in the gouge centerline. However, on the subgouge centerline, by incorporating the soil depth (y/W) at greater depth, the subgouge sand deformation ($d_{(\text{sand-}y)}/W$) is written as follows:

$$\frac{d_{(sand-y)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}3)$$

In this study, a set of experimental investigations were used to simulate the ice-induced seabed deformations (d/W). This means that the key measurements of six experimental datasets reported by Paulin (1991, 1992), C-CORE (1995, 1996), Hynes (1996), and Yang (2009) were applied to validate the ML models. Table 6-1 shows the range of experimental parameters used in the present study.

Table 6-1. Range of experimental parameters used in the present study

Model	Test	Ds/W	φ (Deg)	α (Deg)	h'/W	$Lh/\gamma s.W^3$	$Lv/\gamma s.W^3$	$V^2/g.W$
Paulin (1991, 1992)	P-1	0.930	18	15	-	1.051	1.524	0.001
	P-2	0.095	35	15	1.163	1.271	1.642	0.008
	P-3	0.091	35	15	0.119	1.054	1.528	0.008
	P-4	0.088	35	15	0.109	0.721	0.910	0.008
	P-5	0.093	35	15	0.102	0.751	0.777	0.008
	C-1	0.113	29.022	15	-	0.001	0.001	6.8E-05
	C-2	0.065	28.912	30	-	0.0002	0.0002	6.8E-05
	C-3	0.073	28.912	30	-	0.0002	0.0003	6.8E-05

	C-4	0.079	34.542	15	-	0.0004	0.0003	6.8E-05
	C-5	0.143	34.542	15	-	0.001	0.001	6.8E-05
	C-6	0.300	41.056	30	-	1.065	1.397	-
C-CORE	C-7	0.333	43.540	15	-	2.235	2.598	-
(1995, 1996)	C-8	0.140	33.880	15	-	0.731	0.646	-
	C-9	0.177	41.332	15	-	0.845	0.764	-
	C-10	0.167	37.882	15	-	0.764	0.723	-
	H-1	0.113	38.2	15	0.233	2E-05	2.27E-05	0.015
	H-2	0.065	36.6	30	0.207	2E-05	6.46E-06	0.015
	H-3	0.073	37	15	0.207	1E-05	1.18E-05	0.015
Hynes (1996)	H-4	0.080	39.3	15	0.240	1.17E-05	1.28E-05	0.015
	H-5	0.143	38.4	15	0.240	2.17E-05	2.44E-05	0.015
	Y-1	0.130	44	30	0.338	0.512	1.620	2.55E-09
	Y-2	0.14	32	30	0.215	0.539	0.693	2.55E-09
	Y-3	0.082	34	30	0.295	0.614	0.905	5.61E-08
Yang (2009)	Y-4	0.160	32	30	0.260	1.103	1.530	8.57E-09
	Y-5	0.167	29	30	0.280	1.091	1.261	2.14E-07
	Y-6	0.086	37	30	0.336	2.033	2.859	0.0001

The aforementioned research works did not report the value of the surcharged soil slope (ω), so equation (6-3) is written as below:

$$\frac{d_{(sand\_y)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}4)$$

Therefore, the dimensionless groups introduced in equation (6-4) were applied as the input parameters for the ML models to simulate the horizontal and vertical subgouge deformations. The applied input combinations in the present study are drawn in Figure 6-2.

Model	*y/W*	*Ds/W*	*φ*	*α*	*h'/W*	*Lh/γs.W3*	*Lv/γs.W3*	*V2/g.W*
ETR 1	■	■	■	■	■	■	■	■
ETR 2	■	■	■	■	■	■	■	
ETR 3	■	■	■	■	■	■		■
ETR 4	■	■	■	■	■		■	■
ETR 5	■	■	■	■		■	■	■
ETR 6	■	■	■		■	■	■	■
ETR 7	■	■		■	■	■	■	■
ETR 8	■		■	■	■	■	■	■
ETR 9		■	■	■	■	■	■	■

Figure 6-2. Applied input combinations in the current study

Regarding the sensitivity analysis performed for the ice-gouging problem, there were eight dimensionless groups of parameters governing the subgouge soil deformations in the sand. The ETR 1 was a function of all input parameters, whereas the input parameters were removed one at a time in ETR 2 to ETR 9. Thus, nine ETR models were developed and the best models along with the governing input parameters were identified by performing several analyses. It should be noted that the applied dataset was split into two main sub-samples comprising training (70% of the entire data) and testing (the remaining 30% of the data) dataset.

6.1.2.5. Quality of fit

A set of criteria, such as correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike

Information Criteria (*AIC*) were utilized to evaluate the accuracy, correlation, and complexity of the ML models. The closeness of the *R* and *WI* indices to one signified that the model had a high correlation with the experimental values. On the other hand, the closeness of the *RMSE*, *MAPE*, and *CRM* values to zero meant that a particular model had the lowest level of error though the complexity of the ML models was not assessed using the aforementioned criteria. To overcome this limitation, the Akaike Information Criteria (*AIC*) was used, meaning that the less complex ML model owned the lowest value of *AIC*; hence, the best model had the lowest values of *AIC* index and error (*RMSE*, *MAPE*, and CRM*),* with the highest level of correlation (*R* and *WI*).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{6-5}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{6-6}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{6-7}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{6-8}$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \tag{6-9}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{6-10}$$

here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, and *n* are the experimental measurements, the simulated values, the average experimental values, the average simulated values, and the number of experimental measurements, respectively. In addition, k is the number of input parameters in the ML models.

6.1.3. Results and Discussion

6.1.3.1. Sensitivity analysis

By performing a sensitivity analysis, the ETR model's performance (ETR 1 to ETR 9) was assessed. In Figure 6-3, the results of key criteria computed for ETR 1 to ETR 9 to model the horizontal subgouge deformations are illustrated. In this figure, the shading and the size of the rectangles do not signify the significance of the ML models. Among the ETR models, ETR 4 was the best model to simulate the horizontal displacements since it had the lowest degree of complexity along with the highest level of correlation and accuracy, with the AIC, R, and RMSE of -38.368, 0.978, and 0.077, respectively. After ETR 4, the ETR 3, ETR 1, and ETR 7 models were recognized as the second-best to fourth-best ML models to predict horizontal deformations. The sensitivity analysis showed that ETR 9 had the worst performance for the simulation of the objective parameter. Regarding the obtained results, ETR 2, ETR 8, ETR 6, and ETR 5 were respectively ranked as the fifth-best to eighth-best ETR models so as to estimate the horizontal subgouge deformations.

Moreover, the performed sensitivity analysis showed that the soil depth ratio (y/W) had the most significant influence to model the objective parameter (dh/W). However, the berm height ratio (h'/W), the attack angle (α), the gouge depth ratio (Ds/W), the ice dynamic factor (V2/g.W), and the shear strength parameter of the sand seabed (φ) were distinguished as the second-effective to the seventh-effective input parameters to model the horizontal displacements, respectively.

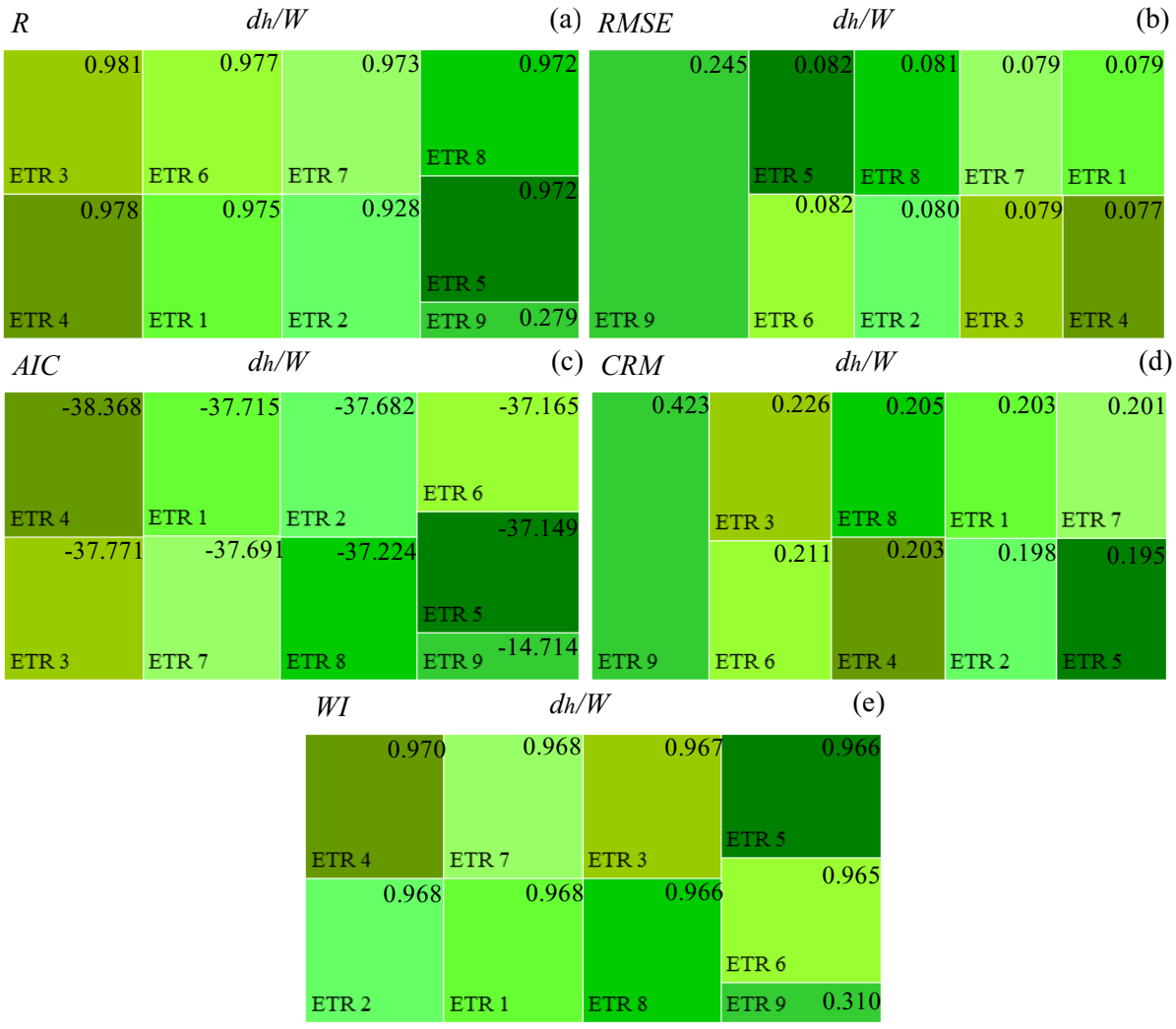


Figure 6-3. Results of key criteria computed for ETR 1 to ETR 9 to model the horizontal subgouge deformations (a) R (b) RMSE (c) AIC (d) CRM (e) WI

Figure 6-4 displays the calculated key criteria for ETR 1 to ETR 9 to model the vertical subgouge deformations. The shading and the size of the rectangles do not signify the significance of the ML models in this figure. According to the simulation results, ETR 1 as a function of all input parameters was the best ETR model to approximate the vertical subgouge deformation, where the value of AIC, CRM, and WI for ETR 1 was equal to -41.783, -0.093, and 0.975. However, ETR 4 demonstrated the worst performance to simulate the vertical displacements, with the R and AIC of

0.695 and -24.395. Based on the conducted sensitivity analysis, ETR 7, ETR 5, ETR 2, ETR 3, ETR 8, ETR 9, and ETR 6 were known as the second-best to eighth-best models to predict the target parameter. To estimate the vertical subgouge soil deformations in the sand, the vertical load factor (Lv/γs.W3) possessed the highest level of effectiveness, while the influence of sand shear strength (φ) was slightly insignificant. In terms of effectiveness, the attack angle (α), the soil depth ratio (y/W), the gouge depth ratio (Ds/W), the vertical load factor (Lv/γs.W3), ice velocity factor (V2/g.W), and the berm height ratio (h'/W) were respectively rated as the second-important to seventh-important input variables to model the vertical displacements.

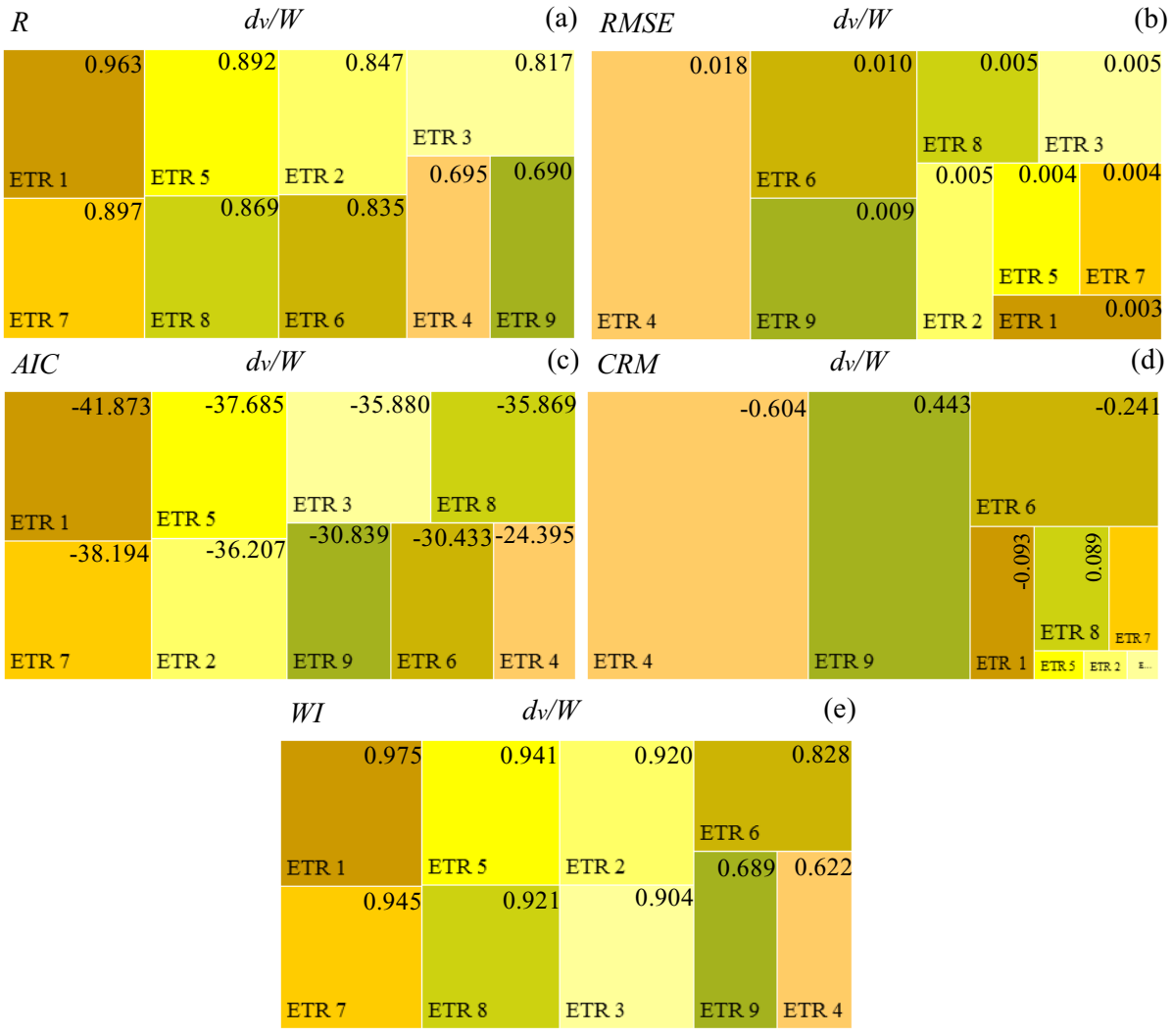


Figure 6-4. Results of key criteria computed for ETR 1 to ETR 9 to model the vertical subgouge deformations (a) R (b) RMSE (c) AIC (d) CRM (e) WI

6.1.3.2. Comparison with other ML algorithms

The ETR 4 and ETR 1 models gave good predictions of the horizontal and vertical subgouge soil deformations in the sandy seabed, respectively. The performance of these superior ETR models was compared with DTR and RFR algorithms. In Figure 6-5, the comparison of ETR 4, DTR, and RFR models' performance is presented. As shown, the ETR 4 models simulated the horizontal subgouge displacements with a better level of accuracy, correlation, and simplicity in comparison with DTR and RFR models. For instance, the AIC value for the ETR 4, DTR, and RFR algorithms was respectively calculated as -38.368, -20.720, and -37.368.

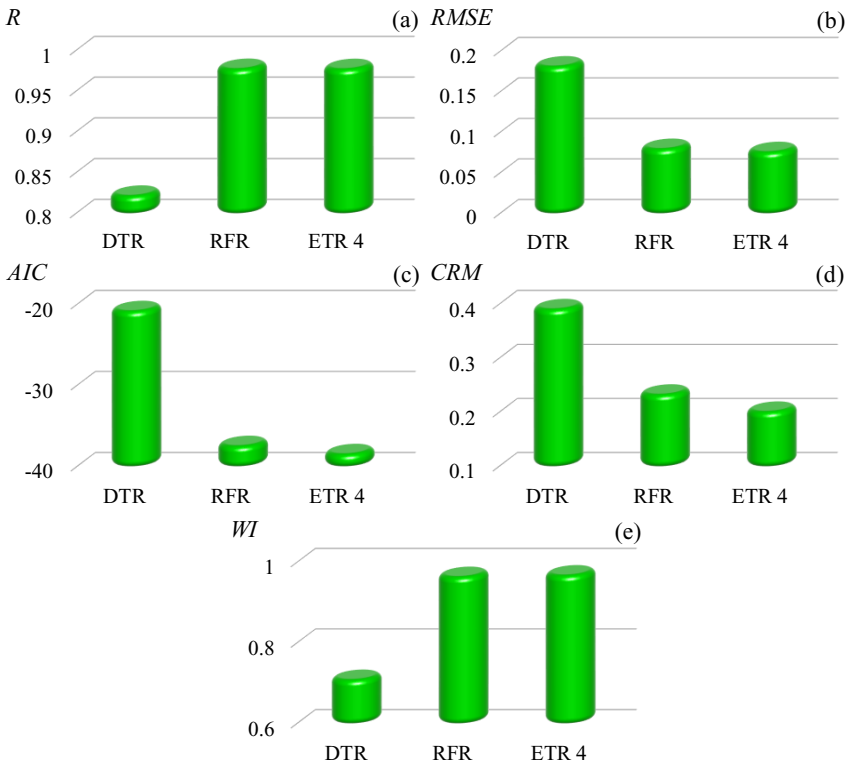


Figure 6-5. Comparison of ETR 4, DTR, and RFR models performance (a) R (b) RMSE (c) AIC (d) CRM (e) WI

Figure 6-6 displays the compression between the performance of ETR 1, DTR, and RFR models to estimate the vertical subgouge deformations. The computed RMSE value for the ETR 1, DTR, and RFR models was surmised as 0.003, 0.006, and 0.005, respectively. As can be seen, the value of all indices (R, RMSE, AIC, CRM, and WI) showed that ETR 1 is a better model than DTR and RFR to simulate vertical deformations.

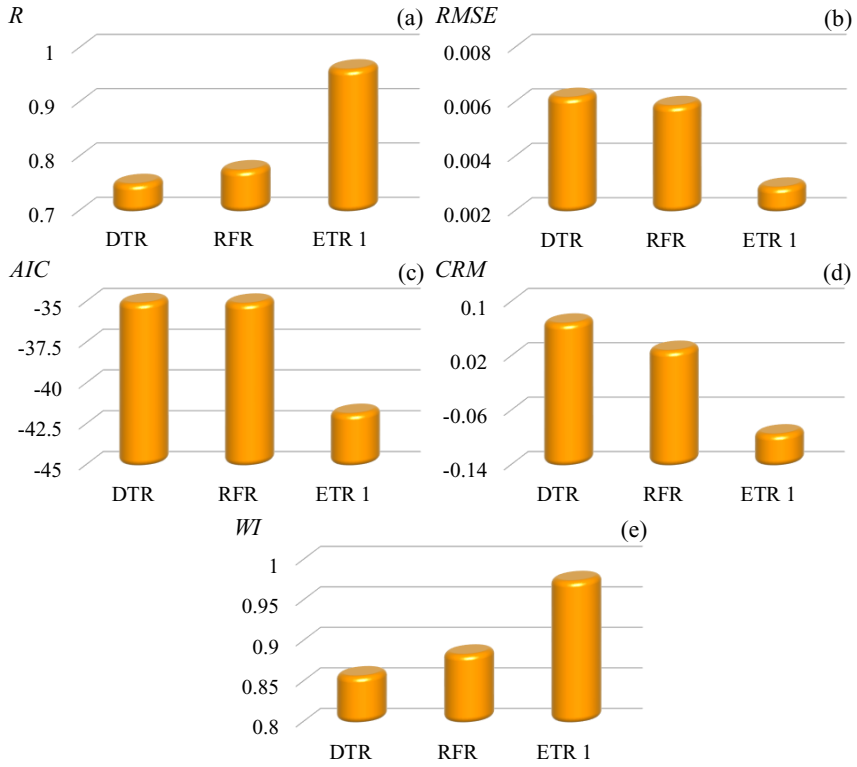


Figure 6-6. Comparison of ETR 1, DTR, and RFR models performance (a) R (b) RMSE (c) AIC (d) CRM (e) WI

6.1.3.3. Error analysis

The error analysis of the ETR, DTR, and RFR models for simulation of the horizontal and vertical deformations was carried out. The results of the error analysis performed for the horizontal displacements estimated by the ETR 4, DTR, and RFR models are drawn in Figure 6-7. The

conducted error analysis demonstrated that approximately 83% of the horizontal deformations predicted by the DTR model had an error of greater than 20%, whereas this value for RFR and ETR 4 models equaled roughly 72% and 51%. Additionally, almost one-third of the ETR 4 results possessed an error of less than 10% but a tiny minority of the horizontal displacements predicted by the DTR model had an error of smaller than 10%.

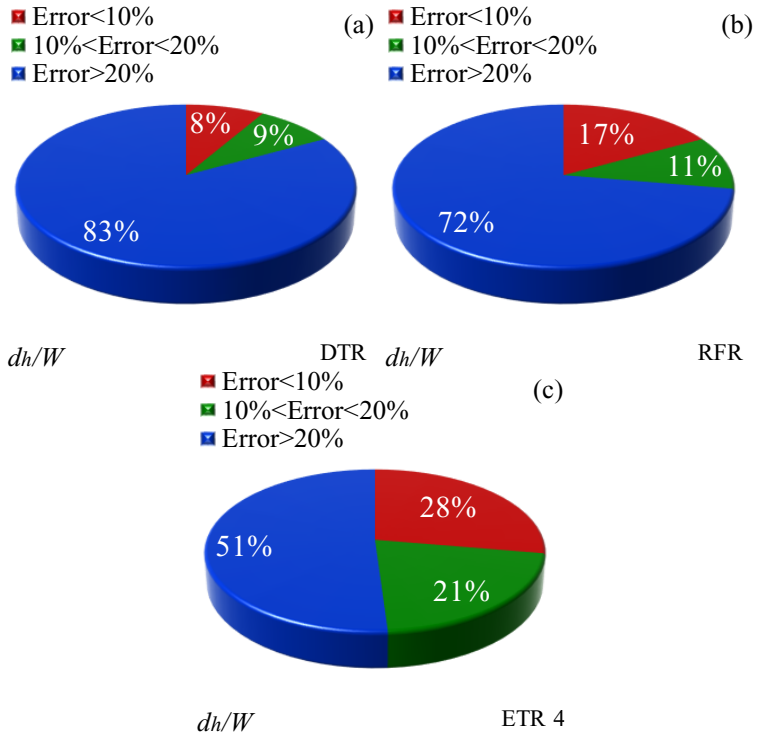


Figure 6-7. Results of the performed error analysis for the horizontal displacements estimated by (a) DTR (b) RFR (c) ETR 4

Figure 6-8 presents the results of the performed error analysis for the vertical displacements estimated by DTR, RFR, and ETR 1. Almost 72% of the vertical deformations simulated by the DTR and RFR algorithms showed an error of more than 20% though this amount for the ETR 1 model was 36%.

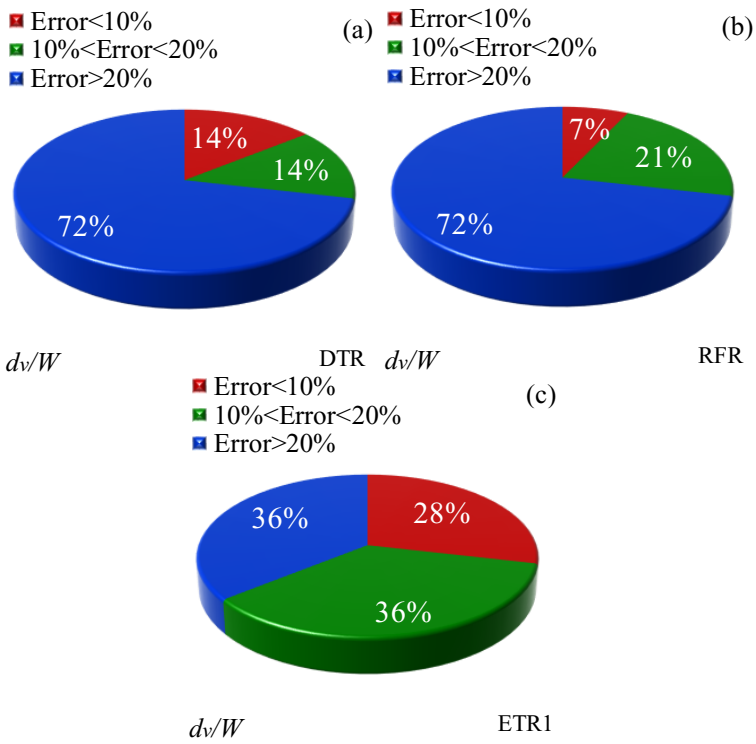


Figure 6-8. Results of the performed error analysis for the vertical displacements estimated by (a) DTR (b) RFR (c) ETR 1

Therefore, the performed error analysis proved that the ETR algorithms had better performance in comparison with the DTR and RFR methods to simulate the subgouge soil displacements in the sandy seabed.

6.1.3.4. Uncertainty analysis

To appraise the performance of the ML models used in this study, an uncertainty analysis was carried out. To perform the uncertainty analysis, errors from these ML models (e_j) were calculated as the difference between the simulated ice-induced deformation (P_j) and the observed values (O_j) as follows:

$$e_j = P_j - O_j \tag{6-11}$$

The mean calculated errors (Mean) and the standard deviation (StDev) of errors were obtained by the following equations:

$$\text{Mean} = \frac{1}{n}\sum_{j=1}^{n} e_j \tag{6-12}$$

$$\text{StDev} = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2/(n-1)} \tag{6-13}$$

The ML model had an underestimated performance if the sign of Mean was negative, while the positive Mean reprinted that the ML model overestimated the deformations. Using the "Wilson score method" and disregarding the continuity correction, a confidence interval (CI) was generated near the computed error by using the Mean and StDev values. A modified normal distribution interval as an asymmetric normal distribution, entitled the Wilson score interval, was employed to rectify the CI bounds. After that, a ±1.96Se created a 95%CI. It is worth noting that the width of uncertainty bound (WUB) was calculated as follows (Azimi and Shiri 2021c):

$$\text{WUB} = \pm\frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{6-14}$$

The results of uncertainty analysis for the horizontal and vertical deformations predicted by the DTR, RFR, and ETR algorithms are exhibited in Figure 6-9. Regarding the performed uncertainty analysis, the DTR, RFR, and ETR models indicated an overestimated performance in order to simulate the horizontal deformations, signifying that the Mean value for these ML models was respectively obtained at 0.011, 0.008, and 0.006. In addition, the ETR algorithm had the narrowest width of uncertainty bound to model the horizontal ice-induced displacements, with a WUB of 0.007. Although the DTR and ETR models overestimated the vertical displacements, these deformations were underestimated by the RFR method, where the Mean value for DTR, ETR, and

RFR was equal to 0.0007, -0.00002, and 0.0004. To simulate the vertical subgouge deformations, the widest uncertainty bound belonged to the DTR model (WUB=0.006), whereas the narrowest one was for the ETR algorithm (WUB=0.001).

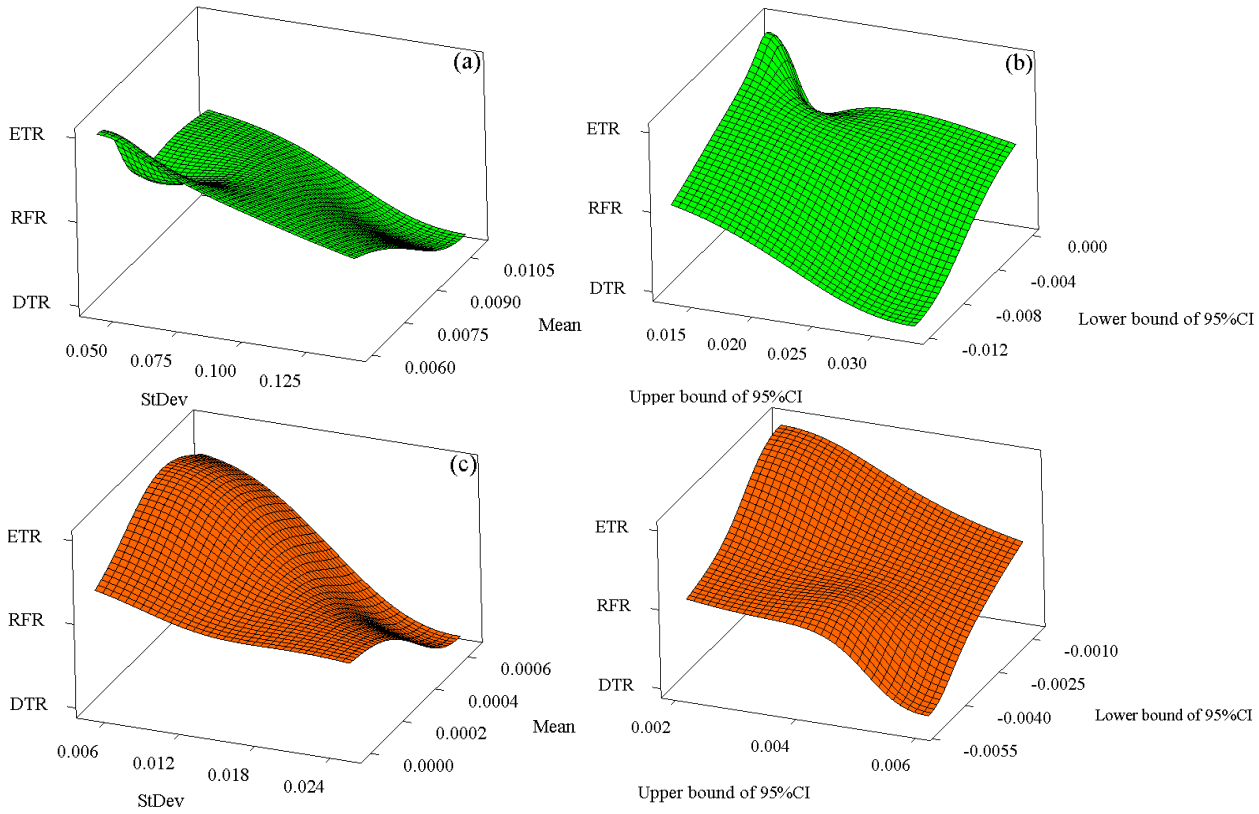


Figure 6-9. Results of uncertainty analysis for the DTR, RFR, and ETR algorithms (a) Mean and StDev for horizontal deformations (b) Lower bound and Upper bound of 95% for horizontal deformations (c) Mean and StDev for vertical deformations (d) Lower bound and Upper bound of 95% for vertical deformations

6.1.3.5. Comparison with test results

ETR was identified as the premium ML model to simulate the subgouge soil deformations in the sand in the present study. Figure 6-10 compares the horizontal ice-induced soil deformations estimated by the ETR 4 and the test results. As shown, the maximum value of horizontal subgouge

deformations was predicted just under the ice keel, and the magnitude of this parameter was reduced at greater soil depth. ETR 4 managed to model the P-1, P-2, P-3, P-5, H-2, H-5, C-6, Y-5, and Y-6 with a high level of correlation and precision. Even though several discrepancies were reported between the simulation results and the experimental values (P-1, C-1, C-2, C-4, C-5, H-1, H-3, C-8, C-10, Y-1, Y-3, and Y-4), ETR 4 was able to simulate the objective function with its highest performance.

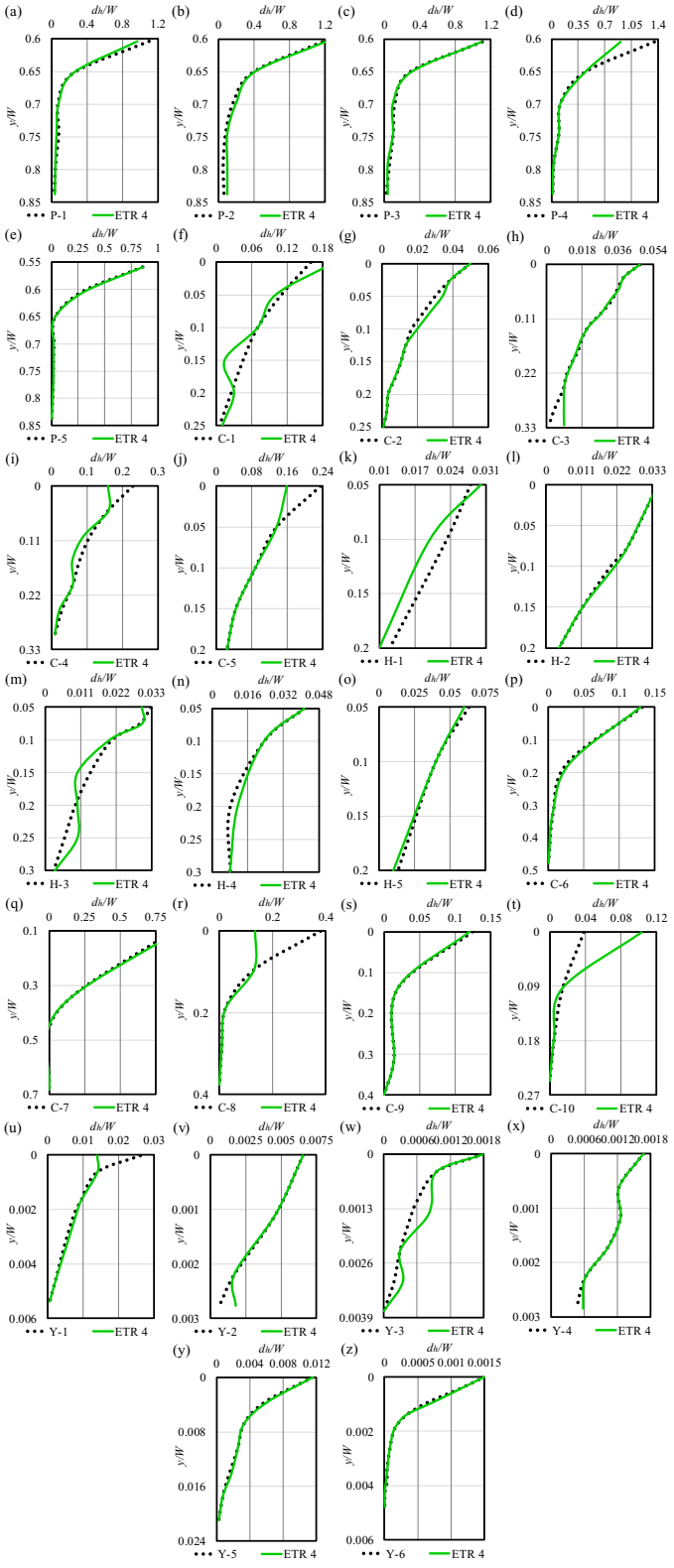


Figure 6-10. Comparison between the simulated horizontal deformations by ETR 4 with test results

The comparison between the simulated vertical displacements by ETR 1 with experimental results is illustrated in Figure 6-11. As can be seen, there were some fluctuations in the experimental measurements (P-1, P-2, P-3, and P-4); rather ETR 1 could model the vertical subgouge deformations with reasonable accuracy. The C-2, H-1, H-2, and H-3 tests were well simulated by ETR 1; however, several differences were observed between the ETR 1 results and C-1, C-3, C-4, and C-5 tests. The ETR 1 model using the linear (C-3, H-2, and H-3) and non-linear (C-2, C-4, H-1, C-5) trends commonly predicted the vertical subgouge deformations.

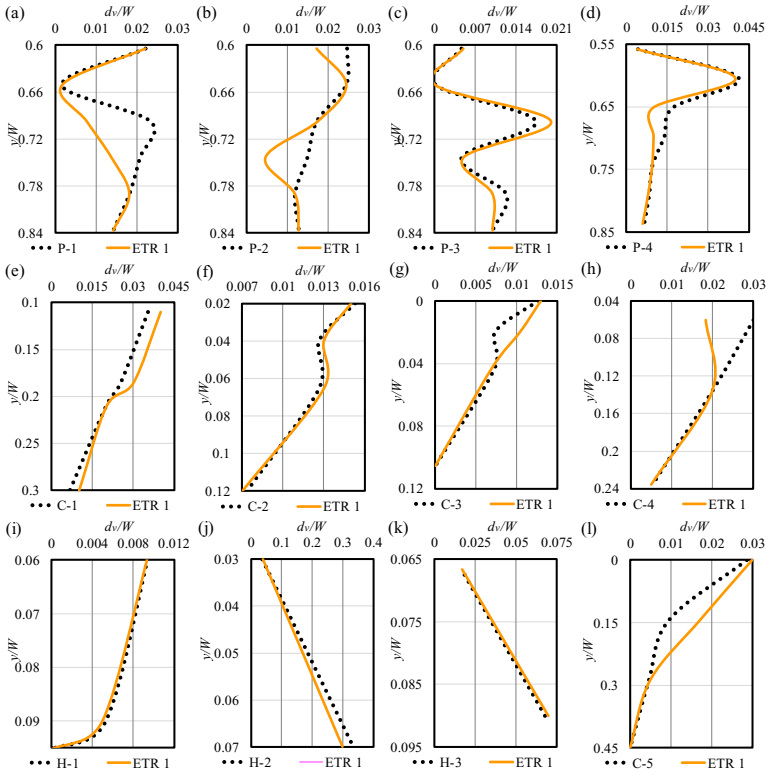


Figure 6-11. Comparison between the simulated vertical deformations by ETR 1 with test results

Therefore, the ETR 4 a function of the y/W, Ds/W, φ, α, h'/W, $Lv/\gamma s.W^3$, and $V^2/g.W$ input parameters was determined to be the best ML model to simulate the horizontal subgouge soil

deformations in the sand, meaning that the horizontal load factor ($Lh/\gamma s.W^3$) was a removed variable for this model. ETR 4 could model the horizontal displacements with a good level of correlation, precision, and simplicity. Furthermore, ETR 1 was determined to be the best model to simulate the vertical deformations, and was fed with all inputs including y/W, Ds/W, φ, α, h'/W, $Lh/\gamma s.W^3$, $Lv/\gamma s.W^3$, and $V^2/g.W$. The ETR 1 was able to estimate the objective parameter with the highest degree of correlation and accuracy along with the lowest level of complexity.

6.1.4. Conclusion

In this study, the extra tree regression (ETR) algorithm was applied to simulate the horizontal and vertical subgouge soil deformations in the sandy seabed. Nine ETR models, comprising ETR 1 to ETR 9, were developed to compute deformations from the parameters governing the iceberg-seabed interaction process. A comprehensive dataset was constructed and it was divided into two major sub-samples comprising the training sub-sample (70% of the dataset) and the testing sub-sample (the remaining 30% of the data). The superior ETR model and the governing input parameters were determined by performing a sensitivity analysis. Furthermore, the performance of the best ETR models was compared with the decision tree regression (DTR) and random forest regression (RFR) algorithms. The most important results obtained from the current study are summarized as follows:

- The ETR 4 and ETR 1 models were identified as the superior models to simulate the horizontal and vertical subgouge soil deformations in the sand, respectively. ETR 4 modeled the horizontal deformations using the y/W, Ds/W, φ, α, h'/W, Lv/γs.W3, and V2/g.W inputs, while the ETR 1 model was a function of y/W, Ds/W, φ, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W.

- The superior ETR models had the highest level of accuracy and correlation along with the lowest degree of complexity.
- The soil depth ratio (y/W) and the berm height ratio (h'/W) governed the prediction of the horizontal deformations, whereas the horizontal load factor (Lh/γs.W3) and the attack angle (α) were identified as the most significant inputs to simulate the vertical deformations.
- The ETR algorithm outperformed the DTR and RFR models for the estimation of both the horizontal and vertical displacements.
- Almost one-third of the horizontal and vertical deformations simulated by the ETR model had an error of less than 10%.
- Uncertainty analysis showed that the ETR model was biased towards overestimating both horizontal and vertical deformations.
- The narrowest width of uncertainty bound belonged to the ETR algorithm.

These results proved that the ETR algorithm was well suited to the simulation of the subgouge soil deformations in the sandy seabed. This investigation can pave the way to develop fast and cost-effective alternatives to centrifuge tests or continuum FE modeling of the ice-seabed interaction problem. The ETR algorithm is considered especially useful in the early phases of engineering design projects, where a quick and precise evaluation is needed to plan the construction methodology, corresponding logistics, and the future scope of detailed schemes.

Funding and Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University

of Newfoundland, the “Natural Science and Engineering Research Council of Canada (NSERC)”, and the “Newfoundland Research and Development Corporation (RDC) (now TCII)” through “Collaborative Research and Developments Grants (CRD)”. Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Azimi, H., Shiri, H., 2020a. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021b. Modeling subgouge sand deformations by using multi-layer perceptron neural network. In The 31st International Ocean and Polar Engineering Conference, ISOPE-I-21-2150.

Azimi, H., Shiri, H., 2021c. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Breiman, L., 2001. Random forests, Machine learning, 45(1): 5-32.

C-CORE., 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2, St. John’s, NL, Canada.

C-CORE., 1996. PRISE Phase 3c: Extreme LEE Gouge Event—Modeling and Interpretation, C-CORE Report 96-C32, St. John's, NL, Canada.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning 63: 3-42.

Hammed, M.M., AlOmar, M.K., Khaleel, F., Al-Ansari, N., 2021. An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. Mathematical Problems in Engineering, 2021: 1-19.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand, Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

ISO. ISO 19906, 2019. Petroleum and Natural Gas Industries – Arctic Offshore Structures. ISO, Geneva.

Kioka, S.D, Kubouchi, A, Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. 13th International Offshore and Polar Engineering Conference. International Society of Offshore and Polar Engineers, ISOPE-I-03-081.

Kioka, S, Kubouchi, A, Ishikawa, R, Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. 14th International Offshore and Polar Engineering Conference. International Society of Offshore and Polar Engineers, ISOPE-I-04-107.

Paulin, M.J., 1991. Preliminary Results of Physical Model Tests of ice Scour, Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering, St. John's, NL, Canada.

Paulin, M.J., 1992. Physical model analysis of iceberg scour in dry and submerged sand. Master thesis, Memorial University of Newfoundland, Canada.

Pekel, E., 2020. Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3): 1111-1119.

Sahani, N., Ghosh, T., 2021. GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model, Ecological Informatics, 64: 101352.

Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Section 2

Iceberg-seabed interaction analysis in sand by a random forest algorithm

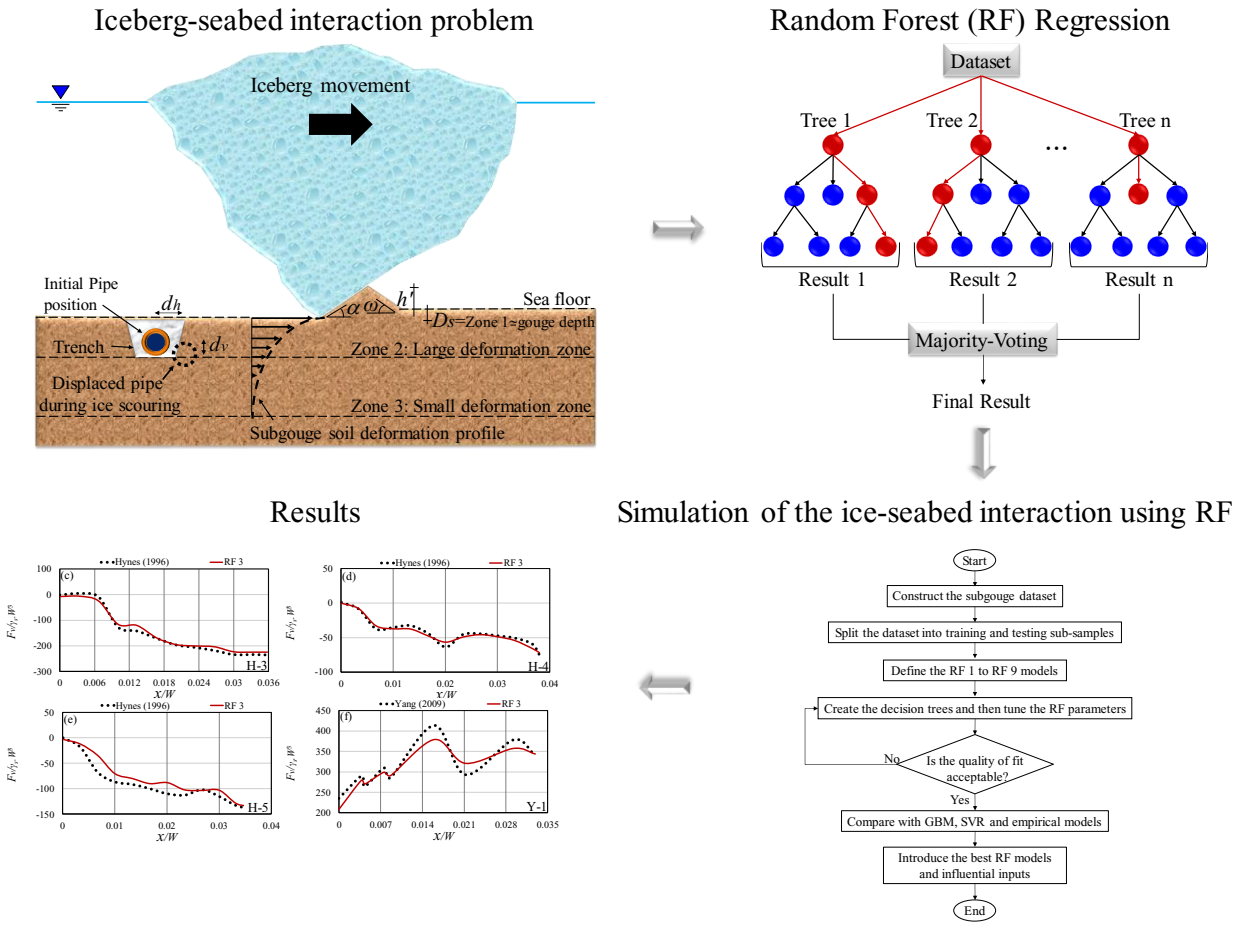
This section is a journal paper which has been published in Polar Science (2022), Vol. 34, pp. 100902, ISSN: 1873-9652

Abstract

Iceberg-seabed interaction that threatens subsea pipelines and structures is a challenging and costly engineering design aspect of Arctic offshore infrastructures. In this study, the subgouge soil deformation in the sand along with the keel reaction forces was simulated using Random Forest (RF) as a strong machine learning (ML) model and compared with the Gradient Boosting Model (GBM), and Support Vector Regression (SVR) as other alternatives. Nine RF models were built based on the most influential parameters and the best model was identified by performing a sensitivity analysis. The study showed that the proposed RF model outperformed its counterparts and proved to be a cost-effective and reliable alternative to assess the iceberg-seabed interaction in the sand, particularly at the early stages of the projects, where a fast and accurate estimation is required for planning the construction methodologies, logistics, and the scope of detailed engineering.

Keywords: Iceberg-seabed interaction, Sandy seabed, Random forest, Gradient boosting model Support vector regression,

Graphical abstract



6.1.5. Introduction

Recently discovered crude oil (90 billion barrels) and natural gases (47.3 trillion cubic meters) in the Arctic regions have led to a growing interest in these regions, where the overwhelming majority of these resources are transferred through the subsea pipelines from offshore to onshore. The seabed may be gouged into the shallower waters by drifting icebergs, threatening the integrity of the subsea pipelines and other bottom-founded structures (Alba, 2015). Figure 6-12 illustrates the layout of the iceberg-seabed interaction process schematically.

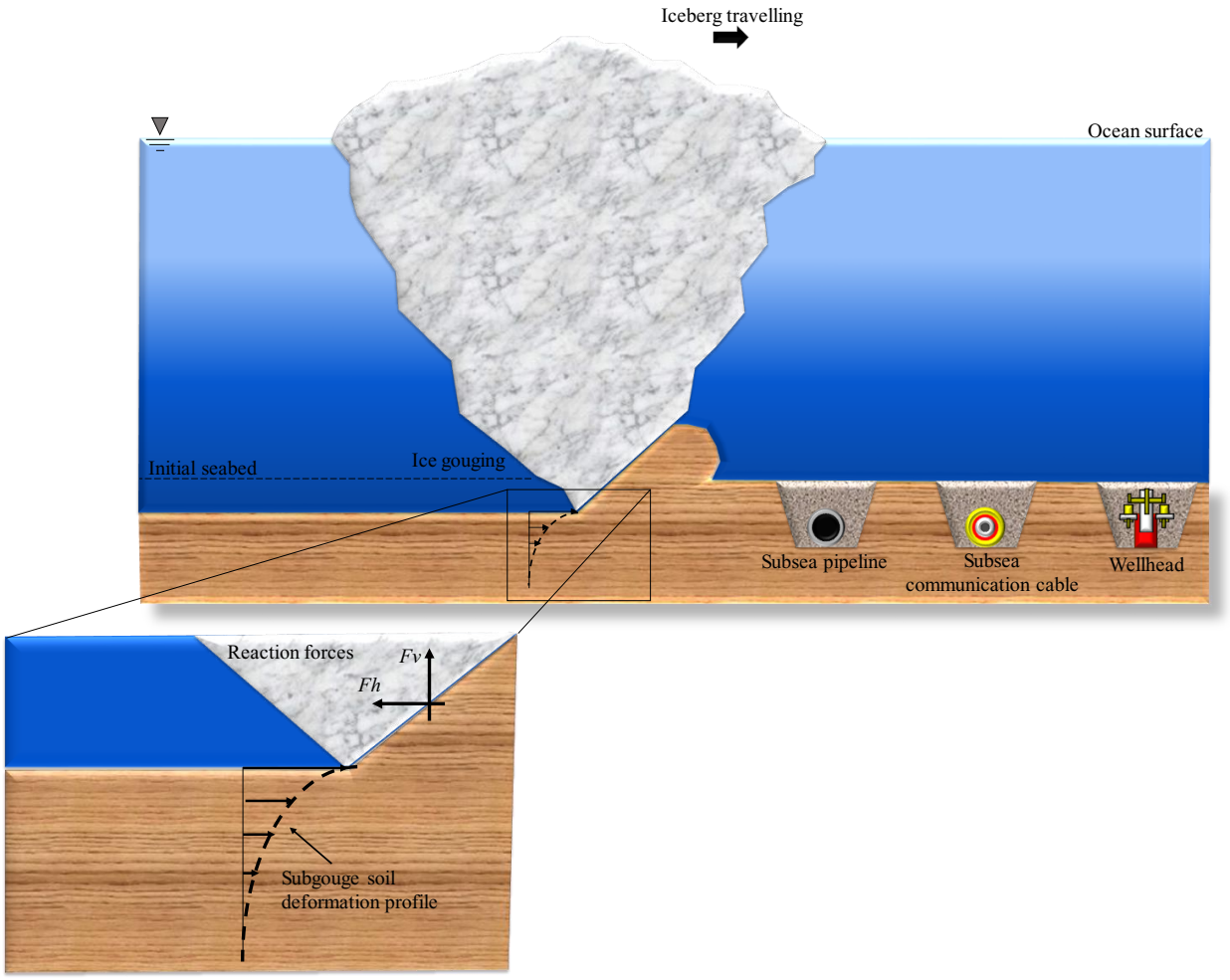


Figure 6-12. Schematic layout of the iceberg-seabed interaction problem

Papusha et al. (2013) described several subsea systems damaged by icebergs in the shallow Arctic waters on the bottom of Lake Erie, Great Slave Lake, and the Labrador Sea. Grand Banks Scour Catalogue (GBSC) GeoDatabase", one of the complete investigations of the ice-gouging characteristics identified on the Grand Banks of Canada, reported that several icebergs drifted and gouged the ocean floor across the shallow water area of the Grand Banks, where scour marks observed in depths of up to 400 m (Campbell 2014). Moreover, Minerals Management Service (2002) recorded that 529 out of 836 gouges (roughly 63%) surveyed from 1995 to 1998 were observed in the shallower waters.

For pipeline protection against the ice-gouging, the pipeline is buried below the maximum predicted ice gouge. However, the soil deformation is not limited to the soil in front of the iceberg keel. The soil shear resistance causes the soil displacement to extend much deeper than the iceberg tip and still threatens the pipeline which is called subgouge soil deformation (see Figure 1). The estimation of the additional burial depth below the iceberg tip is a challenging and costly design aspect that is usually conducted by a decoupled approach. The decoupled approach combines the accuracy of the continuum finite element analysis (FEA) of the free field ice-gouging event with the simplicity of the beam-spring models for the simulation of the pipeline. In this approach, first, a free-field ice-gouging analysis is conducted without the presence of the pipeline. Then, the resultant subgouge soil deformations are manually transferred to the beam-spring model to obtain the pipe response to soil movement. Therefore, free-field ice-gouging is a significant part of the decoupled approach in practical pipeline design and is usually supported by costly numerical and experimental studies.

Artificial intelligence (AI) and machine learning (ML) models have been limitedly used for the analysis of iceberg-seabed interaction. Kioka et al. (2003, 2004) used a Neural Network (NN)

model for the simulation of the ice-gouging problem in the sand. The NN model was validated using a mechanical approach and the author concluded that the NN results had a strong correlation with this mechanical approach. Azimi and Shiri (2020a) introduced the dimensionless parameters affecting the iceberg-seabed interaction in the sand by using Buckingham's theory for the first time. The authors proposed some linear regression (LR) models to estimate the maximum subgouge soil displacements. Azimi and Shiri (2020b) simulated the horizontal subgouge soil displacements in the sand through Gene Expression Programming (GEP). The gouge depth and the dilation angle were found as the most significant input variables to predict the objective function. The extreme learning machine (ELM) was utilized for modeling the iceberg-scoured parameters by Azimi and Shiri (2021a). The authors suggested a set of ELM-based formulas for estimating the subgouge deformations and reaction forces. The shear strength of the seabed soil and the gouge depth ratio was the most influential input to predict the reaction forces and the subgouge deformations, respectively. Multi-layer perceptron neural network (MLPNN) was applied to simulate the horizontal ice-intrusion displacements by Azimi and Shiri (2021b). The authors concluded that the soil depth and the bearing pressure had a remarkable impact on the estimation of the target parameter. Moreover, Azimi and Shiri (2021c) optimized the ELM model using the differential evolution (DE) algorithm to simulate the iceberg-seabed interaction mechanism in the sand. The study highlighted that the berm height ratio, gouge depth ratio, and vertical component of the load had a significant effect on the simulation of the subgouge soil parameters.

Azimi et al. (2021) applied a non-tuned machine learning algorithm to model the sub-gauge soil displacements in clay seabed along with the reaction forces. The authors concluded that the horizontal load factor and the gouge depth ratio were recognized as the most influential input

parameters and a set of equations was provided to estimate the sub-gauge soil features. Azimi and Shiri (2021d) assessed the iceberg-seabed interaction process in clay mass through the ELM model. The soil depth was the most significant input factor governing the subgouge soil deformations. Azimi et al (2022a) utilized an evolutionary design of the generalized group method of data handling (GS-GMDH) to model the iceberg-seabed interaction mechanism. The comparison of the best GS-GMDH model with the artificial neural network (ANN) and the GMDH algorithm showed the better performance of the GS-GMDH model. Azimi et al (2022b) simulated the horizontal and vertical subgouge soil deformations in clay by using the Decision Tree Regression (DTR), Random Forest Regression (RFR), and Extra Tree Regression (ETR) models. The simulation results demonstrated that the ETR model possessed the highest degree of precision and correlation with the experimental values. Azimi et al (2022c) estimated the iceberg-seabed interaction characteristics in clay seabed through decision tree regression (DTR), random forest regression (RFR), and gradient boosting regression (GBR) algorithms. The authors stated that the ETR algorithm was able to predict the target parameters.

The current study used Random Forest (RF) as a strong machine-learning model and compared the results with the Gradient Boosting Model (GBM), and Support Vector Regression (SVR) as other alternatives for a fast and cost-effective estimation of the subgouge soil deformations to be fed into the beam-spring models. This can be of significant importance at the fair stages of design projects, where a fast and fairly accurate estimation of burial depth is required for planning the construction methodologies, logistics, and the scope of detailed engineering.

The dataset is of significant importance in the success of ML models. In this paper, the database for training and testing of the models was extracted from six series of well-recognized experimental studies in sand including Paulin (1991), Paulin (1992), C-CORE (1995), Hynes

(1996), C-CORE (1996), and Yang (2009). The last five decades have witnessed a huge growth in the costly field and experimental studies of the iceberg-seabed interaction process (Chari 1980; Prasad 1986; Machemehl 1986). For instance, several experimental ice-gouging tests in sand and clay seabed were performed by Paulin (1991). The author showed that the vertical reaction forces were more influential for experimental models with a smaller angle of attack, and the horizontal reaction force was the dominant factor for a greater attack angle. Additionally, Paulin (1992) implemented some laboratory tests for submerged and dry sand to quantify the reaction forces and deformations during the ice scouring event. The author concluded that the ice scouring mechanism in submerged and dry circumstances was almost identical; however, the loads and reaction forces for the submerged sand were smaller than in the dry seabed condition. At the Center for Cold Ocean Resources Engineering (C-CORE), some Pressure Ridge Ice Scour Experiment (PRISE) tests were done to the measurement of the subgouge displacements and keel reaction forces in both clay and sand seabed (C-CORE 1995; C-CORE 1996). The investigations proved that the horizontal component of the reaction forces grew by increasing the width of the scouring, whereas the horizontal reaction forces reduced with the steeper attack angle. A set of centrifuge tests in seabed sand was implemented by Hynes (1996). They observed a linear correlation between the scour load and scour depth. Several Pipeline Ice Risk Assessment and Mitigation (PIRAM) tests in the sand condition were conducted by Yang (2009). The effects of gouge depth, frontal berm height, and gouge rate were assessed in this investigation. The author asserted that the gouge geometry and attack angle had a remarkable effect on the reaction forces.

Hashemi and Shiri (2022) simulated the iceberg-seabed interaction in clay by incorporating the strain rate and strain-softening effects. The authors conducted a parametric study to assess the impact of different model parameters on the seabed response to iceberg intrusion. This study

concluded that strain softening had a significant influence on a larger amount of plastic shear strain below and in front of the iceberg tip. Hashemi et al. (2022) modeled the impact of the layered seabed, e.g., soft over stiff clay, in the ice-gouging problem. The keel reaction force was reduced by growing the attack angle.

Considering the high cost and challenges associated with the experimental studies and continuum FEA modeling, the Random Forest (RF) model proposed in this study can mitigate the expenses of future designs and even improve the planning of the research investigations.

In this paper, first, the applied machine learning (ML) algorithms, the iceberg-seabed interaction process, and the used criteria to assess the ML models are discussed in the materials and methods section. Then the performance of the ML models is evaluated in the results and discussion section. Eventually, the best ML models along with the most important input parameters are introduced. Further details associated with the proposed method and the results compared with test data will be discussed in the coming sections.

6.1.6. Materials and methods

In this section, RF, GBM, and SVR are described and then the parameters governing the iceberg-seabed interaction mechanism in the sand will be introduced. Regarding the input parameters, nine RF models were developed in this study. Ultimately, a set of criteria were also provided for the appraisal of the ML models. The flowchart of the current investigation is shown in Figure 6-13.

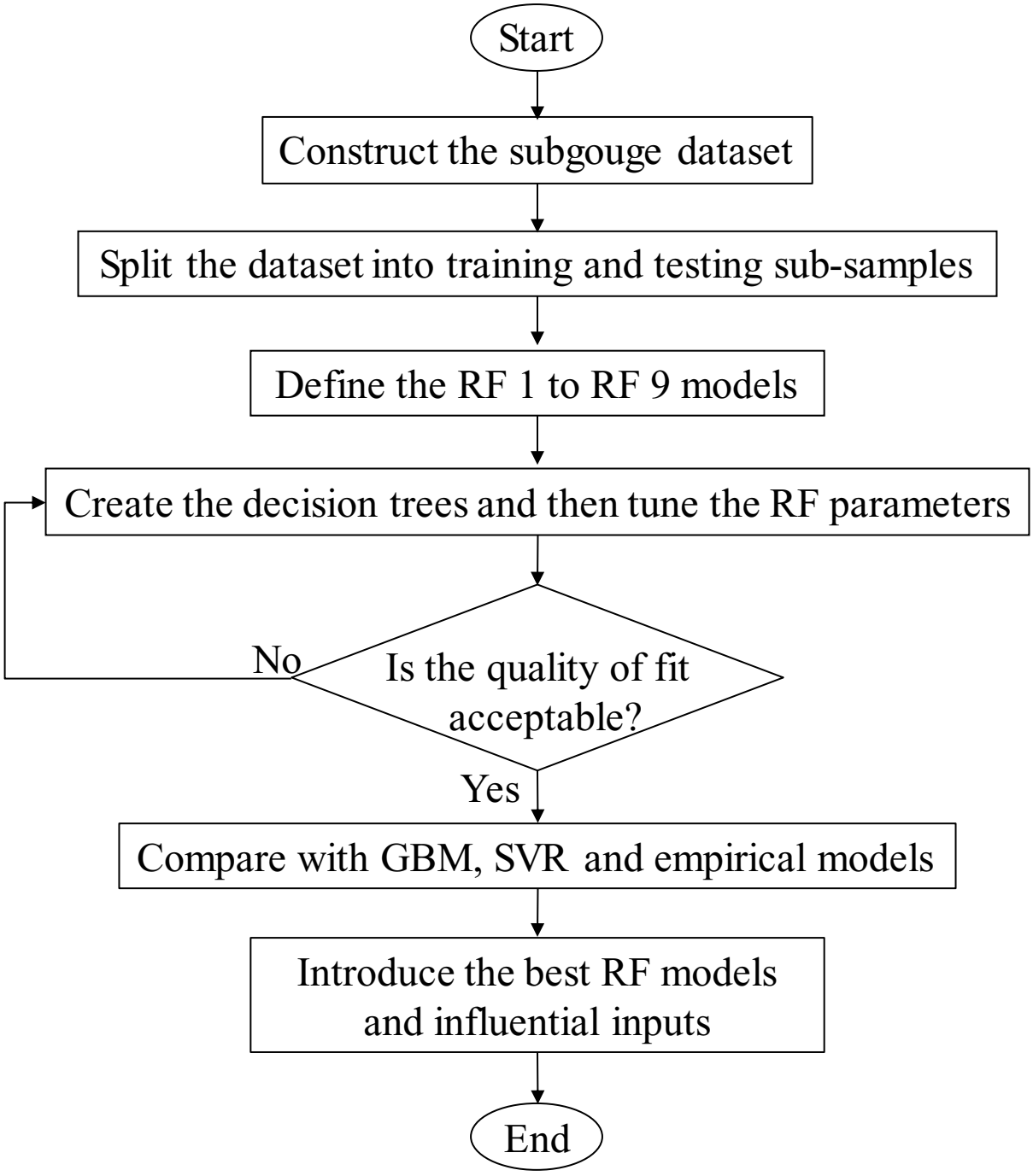


Figure 6-13. Flowchart of the current study

6.1.7. Random forest (RF)

RF algorithm has been developed by Breiman (2001) to solve classification and regression problems. The RF algorithm, as an ensemble learning method, is the extension of the Classification and Regression Trees (CART) algorithm where the tree-based CART models have the overfitting problem and the RF algorithm overcomes this limitation (Breiman 2001). Figure 6-14 shows the diagram of the RF regression schematically.

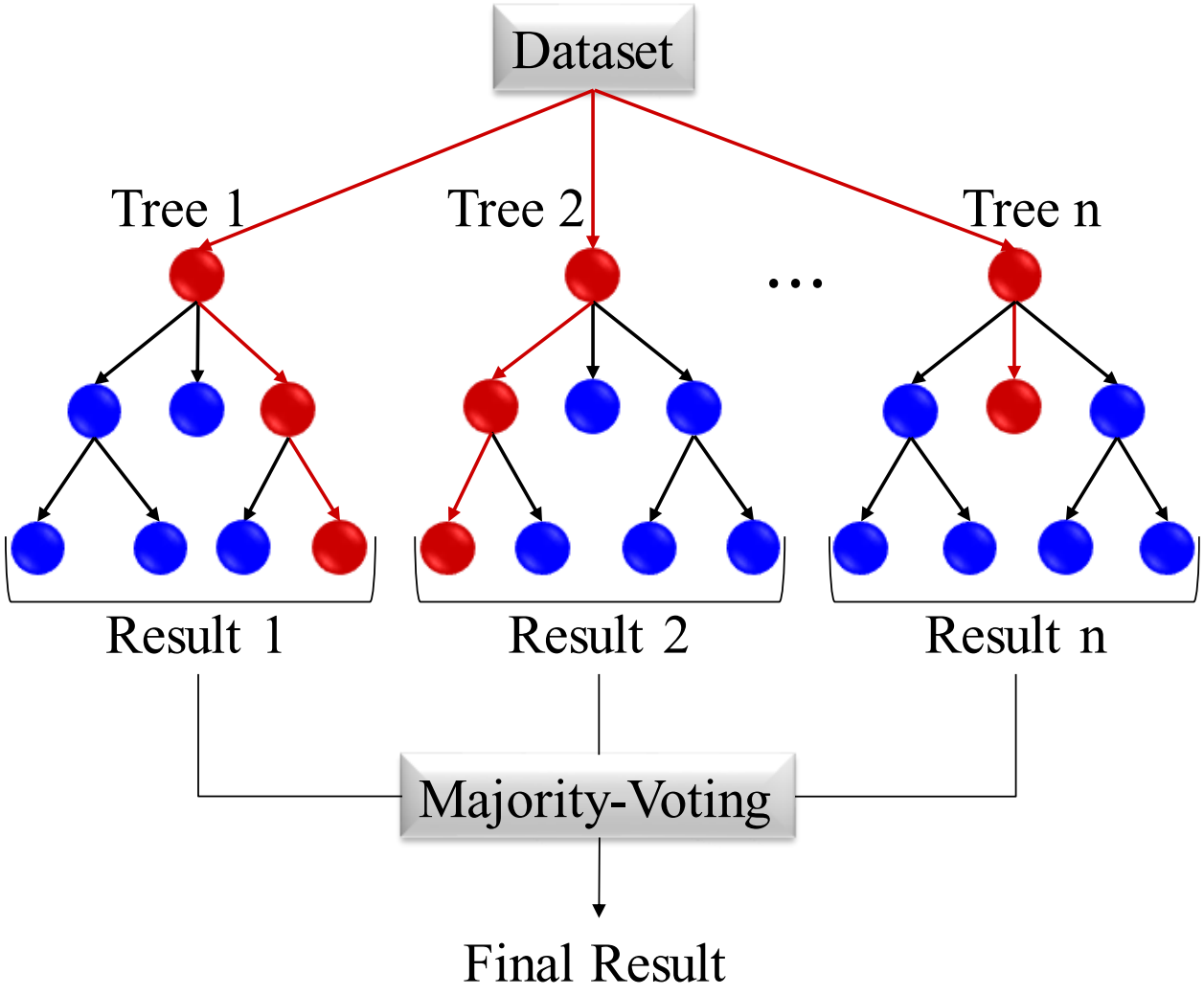


Figure 6-14. Diagram of the Random Forest (RF) regression

In the RF algorithm, several decision trees are created and decisions with the highest number of votes are chosen as the simulation results (Sahani and Ghosh 2021). In the random forest regression, the number of trees for a random vector is increased and the tree predictor, input parameters, and output parameters are considered numerical values. Regarding the random vector distribution, the training dataset is independently utilized. Therefore, the mean-squared generalization error for the tree predictor is calculated (Sahani and Ghosh 2021). The RF model reduces the average error of the trees utilized by utilizing the weighted correlation between the residuals and the applied randomization ought to be employed at a low level of correlation. In the current study, the parameters applied in the RF algorithm comprising the *max_samples*, *bootstrap*, *max_depth*, *random_state*, *number of estimators*, and *verbose* were tuned by a trial-and-error method as *None*, *True* 5000, 500, 1000, 1, respectively.

6.1.8. Gradient boosting model (GBM)

GBM creates an ensemble estimator, regarding the decision trees, to overcome the drawbacks of weak ML tools. In this model, a single decision tree may perform poorly; however, the ensemble algorithm, entitled gradient boosted trees, usually outperforms other decision-based models (Hastie et al. 2009). The GBM estimates the latent function "*f*"' that maps the input parameters to the objective function in an additive expansion form (Li and Bai 2016).

The regression tree is fitted within every single leave in each leave tree for an entire tree (Friedman 2001). It is worth noting that the used parameters of the GBM algorithm, including the *number of estimators*, *learning rate*, *random_state*, *loss function*, *max_depth*, and *verbose* were adjusted using a trial and error approach at 1000, 0.1, 2000, *Huber*, 5, and 1, respectively.

6.1.9. Support vector regression (SVR)

Support vector machine (SVM) is based on Vapnik-Chervonenkis (VC) theory, and this algorithm has been proposed by Vapnik (2004). To simulate the regression problems, SVR is applied in which the training data is mapped from the input variables (input space) into the objective parameter (feature space) through a function (Ϙ). In the feature space, a separating hyperplane with the highest margin is produced. In a regression problem, a nonlinear transformation from the input space to high-dimensional space is made by using the Ϙ function. Regardless of the transformation function (Ϙ), the kernel function can implement the dot product in the multidimensional feature space through the low-dimensional space input variables. In practical applications, several kernel functions comprising the linear, polynomial, and radial basis functions (RBF) are utilized in the SVR algorithm. Moreover, the ε-insensitive loss function is employed as a cost function in this model. To simulate the subgouge soil characteristics in this study, the parameters of the SVM algorithm, such as the *penalty parameter* (*C*), the *kernel coefficient* (*gamma*), *epsilon*, *verbose*, and *kernel* were respectively tuned as 1000, 1.5, 0.001, 1, and RBF. The applied parameters of the SVM model in the current study were chosen based on a trial and error method.

6.1.10. Modeling Process of Iceberg-seabed Interaction

The ice-induced interaction parameters (η) in a sand mass including soil deformations (*d/W*) and reaction forces (*F*/γsW3) are a function of a set of parameters including the scour depth (*Ds*), the internal friction angle of sand (φ), the width of gouge (*W*), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (*h*'), the horizontal load (*Lh*), the vertical load (*Lv*), the velocity of iceberg keel (*V*), and the specific weight of sand (γ_s) (Lach 2006, Azimi and Shiri 2020a):

$$\eta_{(sand)} = f_1(D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \quad (6\text{-}15)$$

Equation (6-15) can be written as a function of eight dimensionless groups as below (Azimi and Shiri 2020a):

$$\frac{\eta_{(sand)}}{W} = f_2\left(\frac{D_s}{W}, \varphi, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}16)$$

It is worth noting that the maximum subgouge deformation in sand $\left(d_{(sand)}\right)$ is formed just under the moving iceberg keel in the gouge centerline. However, at greater depth on the subgouge centerline, by incorporating the soil depth (y/W), the subgouge soil deformation in the sand ($d_{(sand)}/W$) is written as follows:

$$\frac{d_{(sand)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}17)$$

Similarly, the ice-induced reaction force (F) is a function of the position of the iceberg along the scour axis (x) and $D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s$ so equation (6-17) can be written as below:

$$\frac{F_{(sand)}}{\gamma_s W^3} = f_4\left(\frac{x}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}18)$$

The Lh and Lv indicate the consequence of horizontal and vertical loads incorporating drag force from wind, drag force from current, buoyancy force, keel weight, friction force on the base of the keel, Coulomb's passive friction force acting in front and on both sides of the keel, and driving force from the surrounding floe that all are among the input parameters to the problem. However, the Fh and Fv are the seabed reaction forces resultant from the iceberg collision with the seabed.

6.1.11. The physical process of Iceberg-seabed interaction

The past decades have seen the physical modeling application of the iceberg-seabed interaction processes under highly controlled circumstances in the laboratory environment. The ice-gouging

modeling in the laboratory has been carried out adopting two different methods, e.g., the Earth gravity simulation or 1g condition and centrifuge test. Simulation of the iceberg-seabed interaction process in the laboratory may deal with several challenges such as the expensive lab equipment, heavily regulated lab safety, time spent, scaling issues, etc. For instance, the time in-flight for centrifuge simulation excluding the days of preparation for Test #5 in Lach (1996) was 4.7 hours. Additionally, scaling the vertical stress distribution in the 1g condition is impossible. Hence, to overcome these challenges, the iceberg-seabed process in the sand is simulated through different machine learning (ML) algorithms in the current study.

This means that a set of experimental investigations were used to simulate the ice-induced seabed parameters (d/W and $F/\gamma s W^3$) using ML technology. This means that the key measurements of six experimental datasets reported by Paulin (1991), Paulin (1992), C-CORE (1995), Hynes (1996), C-CORE (1996), and Yang (2009) were applied to validate the RF models. The values of the surcharged soil slope (ω) were not reported by the aforementioned research works. The value of the internal friction angle of sand (φ), the keel attack angle (α), the gouge depth ratio (Ds/W), and the velocity ratio (V^2/gW) in Paulin's (1991) experimental model (P-1) were 18^o, 15^o, 0.093, and 0.00054, respectively. In the P-1 model, the value of the berm height ratio (h'/W) was not reported. The range of dimensionless horizontal subgouge deformations (dh/w) for the Paulin's (1992) dataset comprising P-2 to P-5 tests was from 0.002 to 1.356 and these deformations were measured at the dimensionless soil depth up to 0.837. The keel attack angle and the internal friction angle of sand in the P-2 to P-5 tests were reported as 15^o and 35^o, respectively. The value of the surcharged soil slope angle (ω) in the P-1 to P-4 tests was 39.1^o, 30.5^o, 29.2^o, and 26.3^o, respectively. The gouge depth ratio (Ds/W) in P-1 to P-4 were respectively obtained at 0.059, 0.091, 0.088, and 0.093. The C-1 to C-10 tests were reported in C-CORE's (1995) and C-CORE's (1996) datasets.

The value of attack angle in C-1, C-4, C-5, C-7, C-8, C-9, and C-10 was 15° but this parameter in the C-2, C-3, and C-6 tests was measured as 30°. In the C-1, C-4, and C-5 tests, the value of the internal friction angle of sand was 29.022°, 34.542°, and 34.542°, whilst this parameter for C-2 and C-3 was similar, with the φ of 28.912°. The internal friction angle of sand seabed in C-6 to C-10 were respectively reported at 41.056°, 43.540°, 33.880°, 41.332°, and 37.882°. The berm height ratio was not reported in the C-1 to C-10 tests.

Moreover, the H-1 to H-5 tests have measured by Hynes (1996) in which the value of internal friction angle was respectively measured at 38.2°, 36.6°, 37°, 39.3°, and 38.4°. The attack angle in the H-2 test was 30°, whereas this parameter in H-1, H-3, H-4, and H-5 was observed to be 15°. The angle of a surcharged soil slope in the H-1 to H-4 tests was not reported and the frontal berm ratio (h'/W) in these tests was at 0.233, 0.207, 0.207, 0.240, and 0.240, respectively. The value of the iceberg velocity ratio ($V^2/g.W$) in the H-1 to H-5 tests was constant at 0.015. The gouge depth ratio (Ds/W) in the H-1 to H-5 models were equal to 0.113, 0.065, 0.073, 0.080, and 0.143, respectively.

The value of dimensionless horizontal subgouge deformations in Yang's model (Y-1 to Y-7 tests) varied from zero to 0.026. Although the value of the keel attack angle in Y-1 to Y-6 models was 30°, this parameter in Y-7 was measured as 15°. Additionally, the internal friction angle in Y-1 to Y-7 tests was at 44°, 32°, 34°, 32°, 29°, 37°, and 29°, respectively. Therefore, the introduced dimensionless groups in equations (6-17) and (6-18) were applied as the input parameters for the RF models to simulate the subgouge deformations and the reaction forces. The applied input combinations in this study are displayed in Figure 6-15. The number of experimental measurements in the training and testing modes to simulate the horizontal reaction forces ($Fh/\gamma s.W^3$) was 163 and 70, respectively. However, 130 and 55 experiments were respectively

utilized for training and testing to model the vertical reaction forces ($Fv/\gamma s.W^3$). Additionally, 110 and 47 experimental cases were respectively used to train and test the ML model for the prediction of the horizontal sub-gauge soil deformation (dh/W). In addition, the ML model was fed with 50 and 22 laboratory measurements to model the vertical subgouge soil deformations (dv/W) in the training and testing modes

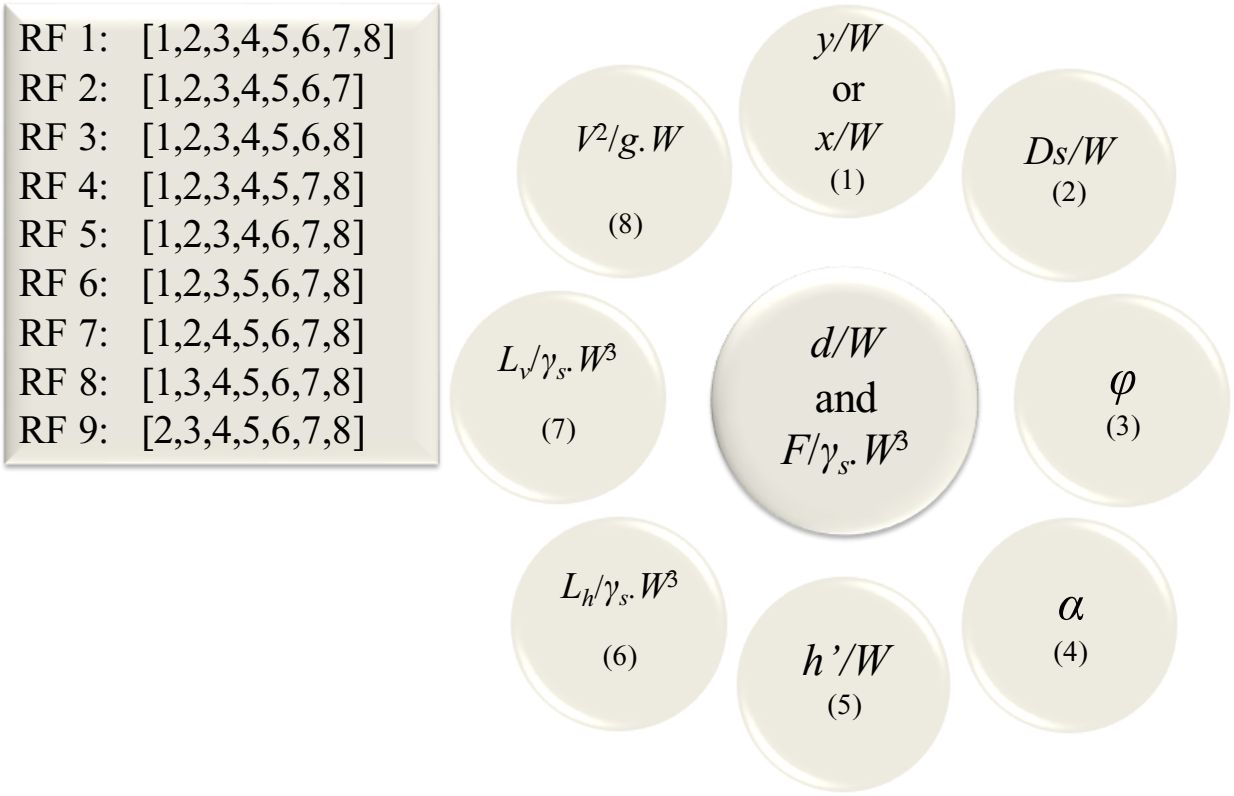


Figure 6-15. Applied input combinations in the current study

Regarding the performed sensitivity analysis for the ice-gouging problem, there were eight dimensionless groups of parameters governing the subgouge soil characteristics. RF 1 was a function of all input parameters, whereas the input parameters were eliminated one at a time in RF 2 to RF 9. Therefore, nine RF models were defined and the best models along with the most significant input parameters were identified by performing a sensitivity analysis. The constructed

dataset was divided into two main sub-samples including the training and testing sub-samples with four different ratios, e.g., (i) 50% for training against 50% for testing, (ii) 60% for training against 40% for testing, (iii) 70% for training against 30% for testing, and (iv) 80% for training against 20% for testing. However, the ML model demonstrated a good performance when 70% and 30% of data were used in the training and testing modes. Thus, the applied dataset was split into two sub-samples comprising training (70% of the entire data) and testing (30% of the remaining) dataset.

6.1.12. Quality of fit

A set of criteria, such as correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were utilized to evaluate the accuracy, correlation, and complexity of the ML models. The closeness of the *R* and *WI* indices to one signified that the model had a high correlation with the experimental values. On the other hand, the closeness of the *RMSE*, *MAPE*, and *CRM* values to zero meant that a particular model had the lowest level of error though the complexity of these models was not assessed using the defined criteria. To overcome this limitation, the Akaike Information Criteria (*AIC*) was used, meaning that the less complex ML model owned the lowest value of *AIC*; hence, the premium model had the lowest values of *AIC* index and error (*RMSE*, *MAPE*, and *CRM*), with the highest level of correlation (*R* and *WI*) (Ahmed et al. 2021).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \quad (6\text{-}19)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (6\text{-}20)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \quad (6\text{-}21)$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (6\text{-}22)$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \quad (6\text{-}23)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (6\text{-}24)$$

here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, and *n* are the experimental measurements, the simulated values, the average experimental values, the average simulated values, and the number of experimental measurements, respectively. In addition, *k* is the number of input parameters in the ML models.

6.1.13. Results and Discussion

The performance of the defined RF models was assessed in the first step and then the best RF models along with the most effective input parameters were identified. Subsequently, the premium RF models were compared with the GBM, SVR, and empirical models, and several analyses, including error analysis and uncertainty analysis, were performed for these models.

6.1.13.1. Evaluation of Random Forest (RF) Models

Figure 6-15 demonstrates the results of the computed key statistical indices for the RF 1 to RF 9 models for the simulation of the horizontal reaction forces. As shown, RF 1 to RF 8 had a similar performance in dealing with modeling the horizontal reaction forces. It is worth noting that RF 1

was a function of all input variables and the effect of each input was disregarded one at a time in RF 2 to RF 9. The value of the correlation coefficient and Willmott Index for the RF 1 model was surmised as 0.995 and 0.998, respectively. The impact of iceberg dynamic ($V^2/g.W$) was removed for the RF 2 model. For the RF 2 model, the value of *CRM* and *AIC* were equal to 0.011 and 360.321. The influence of the vertical load factor ($Lv/\gamma s.W^3$) for RF 3 was removed, with an Akaike Information Criteria of 360.274. The value of the coefficient of residual mass for the RF 4 model was 0.012. The h'/W was a removed input for the RF 5 model where the *AIC* criterion for this model equaled 360.531. The attack angle (α) was the removed input for RF 6, with a *CRM* of 0.012. The shear strength parameter of the sand seabed (φ) was eliminated for RF 7 and the value of *WI* for RF 7 was 0.998. The gouge depth ratio (Ds/W) was excluded from the simulation of $Fh/\gamma s.W^3$ in the RF 8 model. For the RF 8 model, the values of *RMSE* and *WI* criteria were respectively reckoned at 88748.668 and 0.998. The position of the iceberg along the scour axis (x/W) was disregarded for the RF 9 model, with an *AIC* index of 409.091.

The performed sensitivity analysis indicated that RF 4 was the premium model to simulate the horizontal reaction forces, whereas RF 9 had the worst performance. Moreover, RF 7, RF 1, RF 3, RF 2, RF 8, RF 6, and RF 5 were ranked as the second-best to the seventh-best model to predict the horizontal reaction forces. Regarding the simulation results, the position of the iceberg along the scour axis (x/W) had a significant effect on the estimation of the target function since the performance of the RF model worsened by removing this input parameter. Besides, the berm height ratio (h'/W), the attack angle (α), the gouge depth ratio (Ds/W), the vertical load factor ($Lv/\gamma s.W^3$), the shear strength parameter of the sand seabed (φ), the iceberg dynamic factor ($V^2/g.W$), and the horizontal load factor ($Lh/\gamma s.W^3$) were recognized as the second-effective to the eighth-effective input parameters to model the horizontal reaction forces. Even though the

sensitivity analysis demonstrated that RF 4 was the premium model to simulate the horizontal reaction forces in the current study, other combinations of input parameters, e.g., model 1 a function of all inputs, were introduced as the superior model to simulate the subgouge characteristics (Azimi et al. 2022a; Azimi et al. 2022b; Azimi et al. 2022c). Hence, conducting the sensitivity analysis does not always result in introducing model 1 as the best model to estimate the target parameter since a particular ML algorithm may show that a specific input parameter has a remarkable influence but other inputs lack this significance.

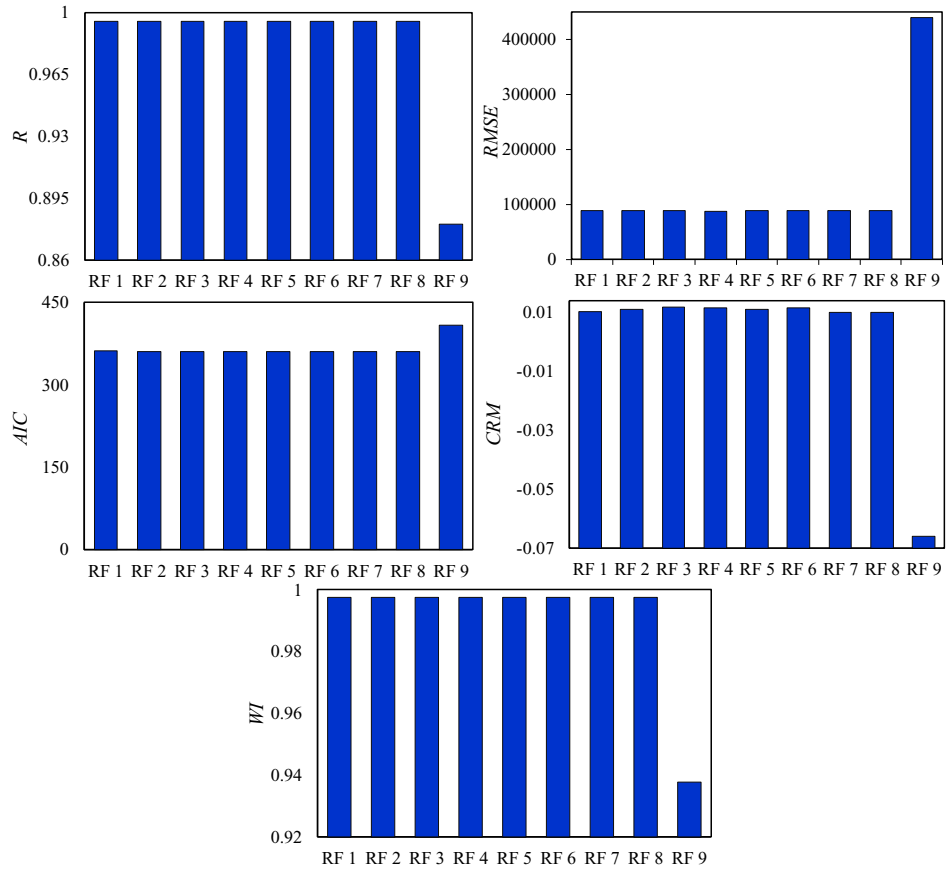


Figure 6-16. Key statistical indices for horizontal reaction forces estimated by RF models

The calculated statistical indices for vertical reaction forces estimated by RF models are compared in Figure 6-17. To simulate the vertical reaction forces, RF 3 was distinguished as the premium

random forest model, with a correlation coefficient of 0.980. Additionally, RF 1, RF 2, and RF 7 were ranked in the second to fourth places and the value of *AIC* for them was respectively approximated to be 103.708, 101.736, and 101.765. For the fifth-best model (RF 8), the value of *RMSE* and *WI* criteria was equal to 39.466 and 0.988, respectively. RF 6 and RF 5, and RF 4 with the AIC index of 101.803, 101.914, and 102.759 resulting in the sixth-best to eight-best RF model. Amongst the RF models, RF 9 was known as the worst model for the simulation of the vertical reaction forces where the value of the Willmott Index was equal to 0.960. The position of the iceberg along the scour axis was also found to as the most influential input factor to model the vertical reaction forces. The horizontal load factor, berm height ratio, attack angle, gouge depth ratio, shear strength parameter of the sandy seabed, iceberg dynamic factor, and vertical load factor were sorted as the second-important to the eighth-important input variables.

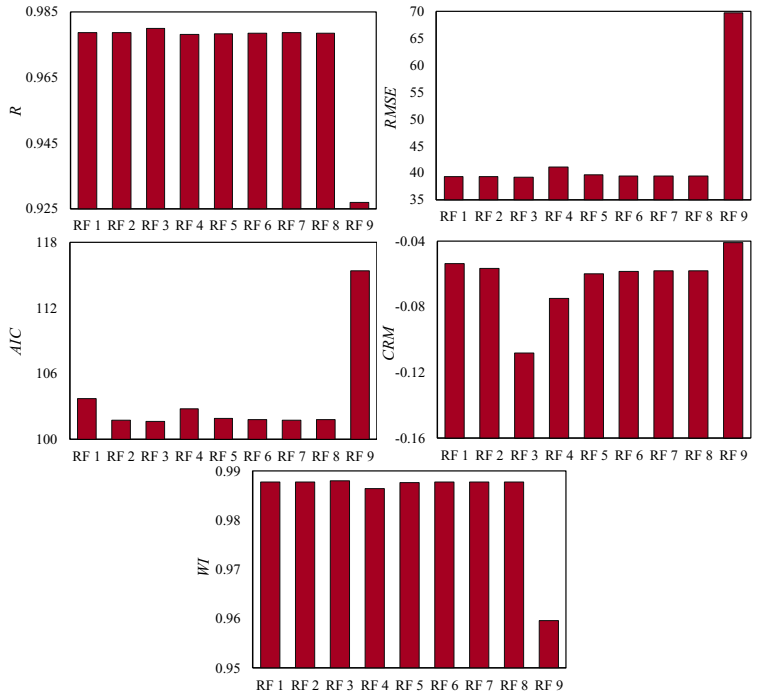


Figure 6-17. Key statistical indices for vertical reaction forces estimated by RF models

In Figure 6-18, the comparison between the performances of RF models to predict horizontal deformations is drawn. RF 2 had the lowest level of complexity (*AIC*=-37.708) along with appropriate precision and correlation (*RMSE*= 0.079 and *R*= 0.978). After the RF 2 model, RF 3, RF 8, and RF 7 were preferred as the second-superior to fourth-superior RF models. The value of *CRM* for the fifth-best RF model (RF 4) was computed as 0.219, whilst this index for the sixth-best model (RF 6) equaled 0.222. For RF 5 and RF 1 as the seventh and eighth-premium random forest models, the value of Akaike Information Criteria was respectively calculated at -37.303 and -35.695. RF 9 exhibited the worst performance to simulate the horizontal deformations among all RF models. The soil depth ratio (y/W) was the most influential input factor in predicting the target function, whereas the berm height ratio, the attack angle, the horizontal load factor, the shear strength parameter of the sandy seabed, the gouge depth ratio, and the vertical load factor were placed at the second to seventh-significant input parameters. The iceberg dynamic parameter ($V^2/g.W$) also had an insignificant effect to model the horizontal subgouge displacements.

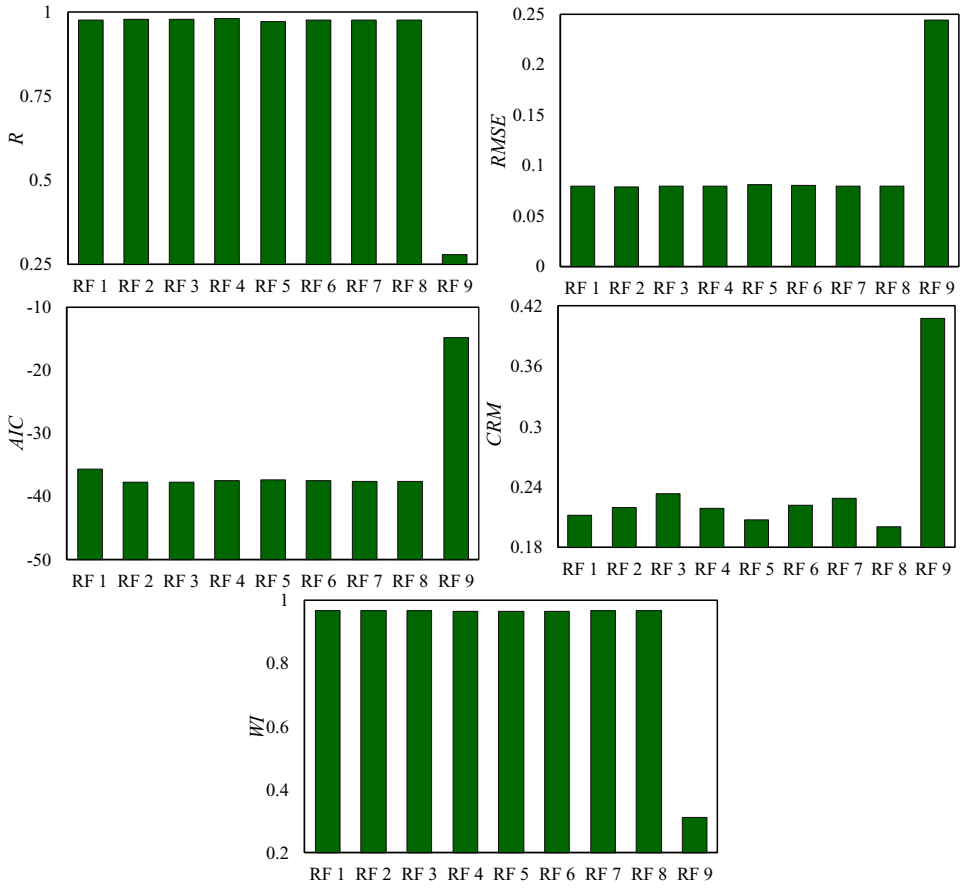


Figure 6-18. Key statistical indices for horizontal deformation estimated by RF models

The comparison between the calculated statistical indices for the vertical ice-induced displacements using the RF models is illustrated in Figure 6-19. As shown, RF 7 demonstrated the best performance for the simulation of the vertical deformation, with the *RMSE* and *AIC* indices of 0.014 and -26.568. The value of the Willmott Index (*WI*) for RF 8 (the second-best model) and RF 6 (the third-best model) was respectively approximated to be 0.325 and 0.307. RF 1, RF 2, and RF 4 were respectively the fourth, the fifth, and the sixth-premium RF models where the value of correlation coefficient (*R*) for these RF models equaled 0.500, 0.494, and 0.477. The computed WI criteria for the seventh-best model (RF 3) was 0.290, whilst this index for the RF 9 (the eighth-superior model) was at 0.248. RF 5 had the lowest level of correlation (R=0.211) and precision

(*RMSE*=0.016) along with the highest level of complexity (*AIC*=-25.452). The berm height ratio (*h'*/*W*) possessed the highest level of effectiveness for estimation of the vertical deformations, rather the influence of sand shear strength (φ) was quite insignificant. Furthermore, the soil depth ratio (*y*/*W*), the vertical load factor ($Lv/\gamma s.W^3$), and the horizontal load factor ($Lh/\gamma s.W^3$) were identified as the second to fourth-important input factors. The iceberg velocity factor ($V^2/g.W$), the attack angle (α), and the gouge depth ratio (*Ds*/*W*) were ranked at six to eight places in the estimation of the target parameter by the RF models.

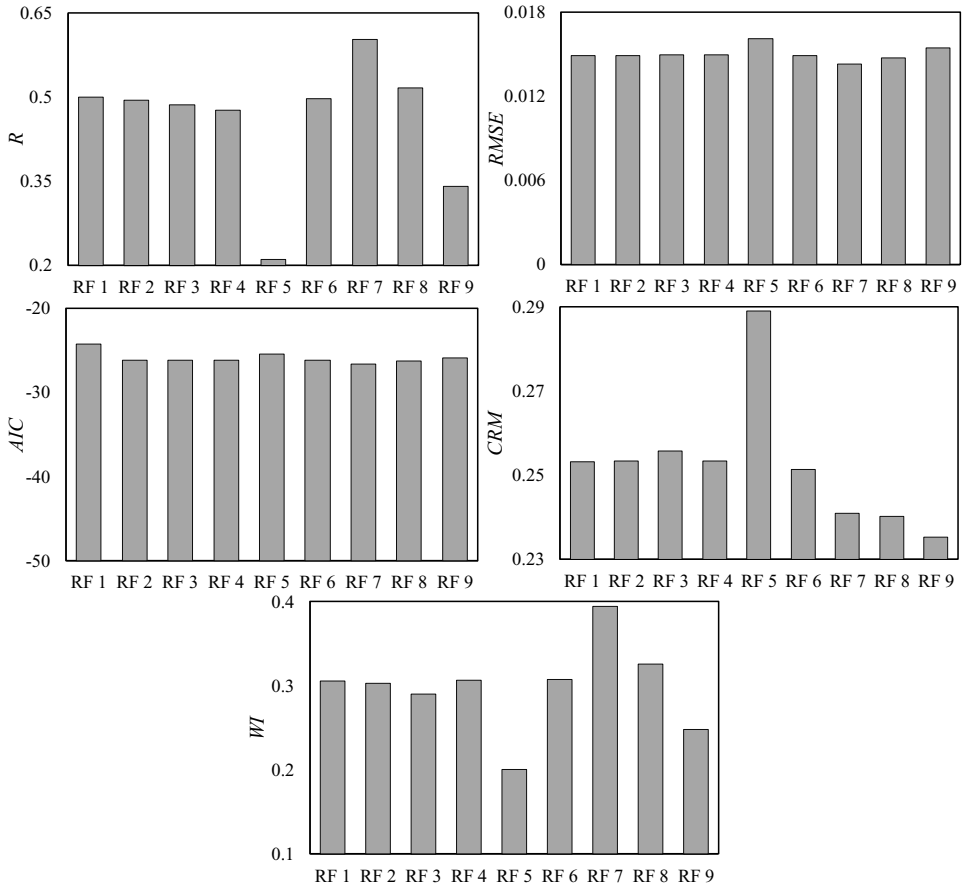


Figure 6-19. Key statistical indices for vertical deformation estimated by RF models

The performed sensitivity analysis for the RF models indicated that the RF 4, RF 3, RF 2, and RF 7 models well simulated the ice-seabed interaction characteristics, with a good level of accuracy and correlation with experimental measurements.

6.1.13.2. Comparison with GBM, SVR, and empirical models

The performance of the superior RF models (RF 4, RF 3, RF 2, and RF 7) was compared with GBM, SVR, and empirical (Emp) models. It is worth noting that some empirical models were suggested by Woodworth-Lynes et al. (1996) to compute the ice-induced soil displacements (equations 6-25 and 6-26), and Barker and Timco (2002) to estimate the subgouge reaction forces as follows (equation 6-27):

$$d_h = 0.6(W.D_s).exp\left(-\frac{2}{3}.\frac{y}{D_s}\right) \quad (6\text{-}25)$$

$$d_v = D_s.exp\left(-\frac{1}{3}.\frac{y}{D_s}\right) \quad (6\text{-}26)$$

$$F = 8.5V + 46.3D_s + 28.5W - 7.6 \quad (6\text{-}27)$$

Figure 6-20 displays the key criteria computed for RF 4, GBM, SVR, and empirical (Barker and Timco 2002) models to simulate horizontal reaction forces. The value of the correlation coefficient for GBM, SVR, and Emp was equal to 0.973, 0.887, and 0.281, with an *AIC* index of 389.663, 410.046, and 1714.365, respectively.

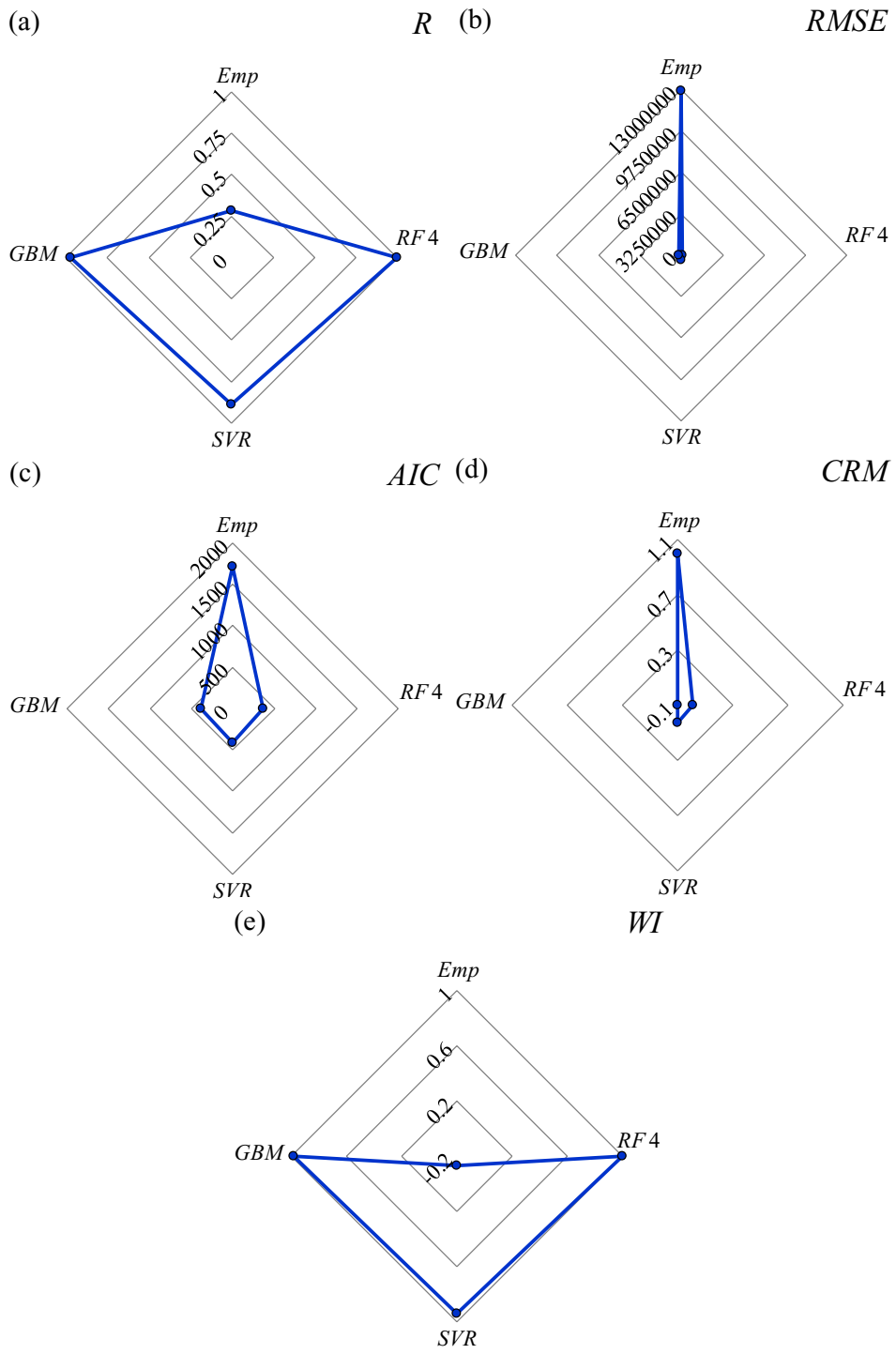


Figure 6-20. Key indices computed for RF 4, GBM, SVR, and empirical models to simulate horizontal reaction forces (a) correlation coefficient (b) RMSE (c) AIC (d) CRM (e) WI

The conducted error analysis for the RF 4, GBM, SVR, and Emp models is demonstrated in Figure 6-21. For instance, half of the simulated reaction forces predicted by RF 4 had an error of less than 10%, whereas this figure for GBM, SVR, and Emp models were 14%, 3%, and zero, respectively. Thus, to simulate the horizontal reaction forces, RF 4 outperformed the GBM, SVR, and Emp models.

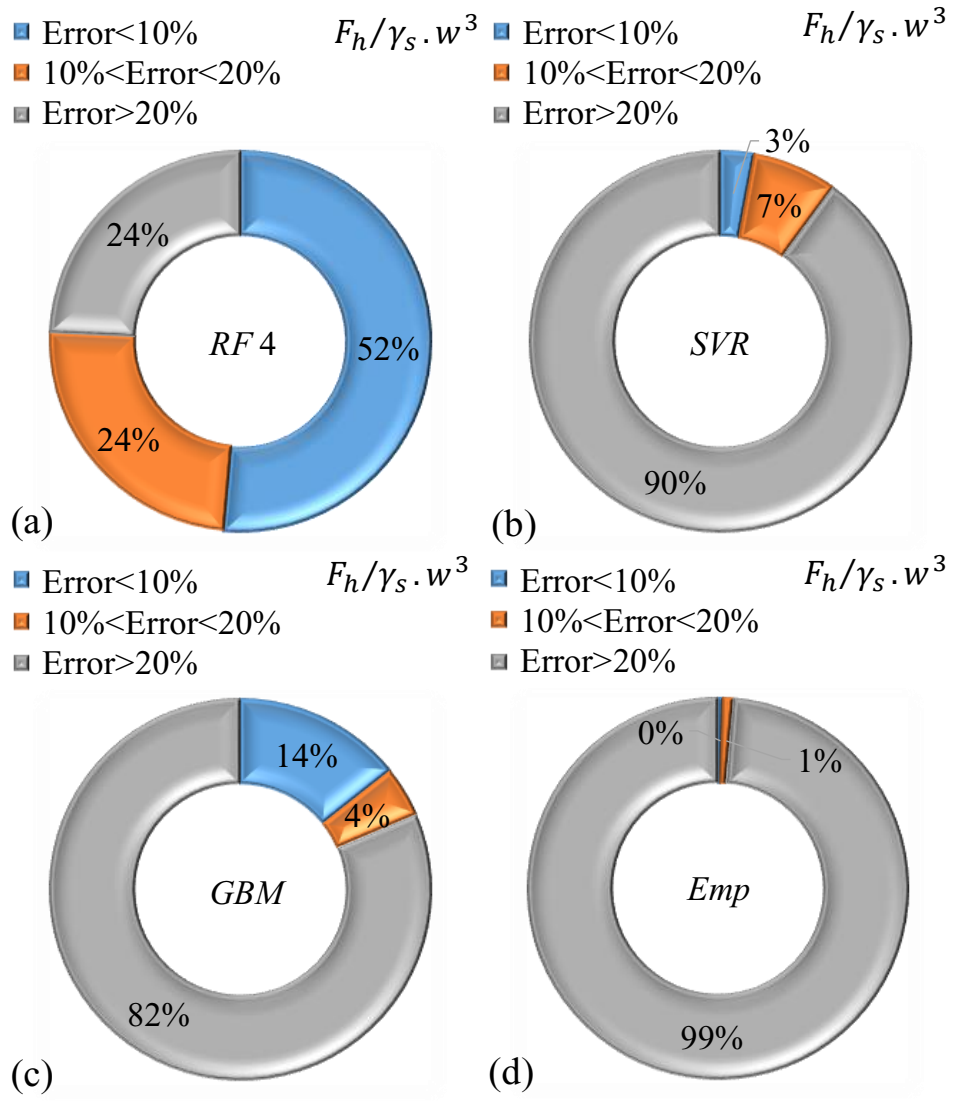


Figure 6-21. Result of MAPE index for (a) RF 4, (b) SVR, (c) GBM, and (d) empirical models to estimate the horizontal reaction forces

Figure 6-22 illustrates the comparison between the RF 3, GBM, and SVR, algorithms to estimate the vertical reaction forces. Similarly, RF 3 had a good performance at dealing with the simulation of the objective function, with a high level of correlation, low level of inaccuracy, and complexity. For GBM and SVR models, the value of WI criteria was computed as 0.957 and 0.954, whilst this index for RF 3 was at 0.988.

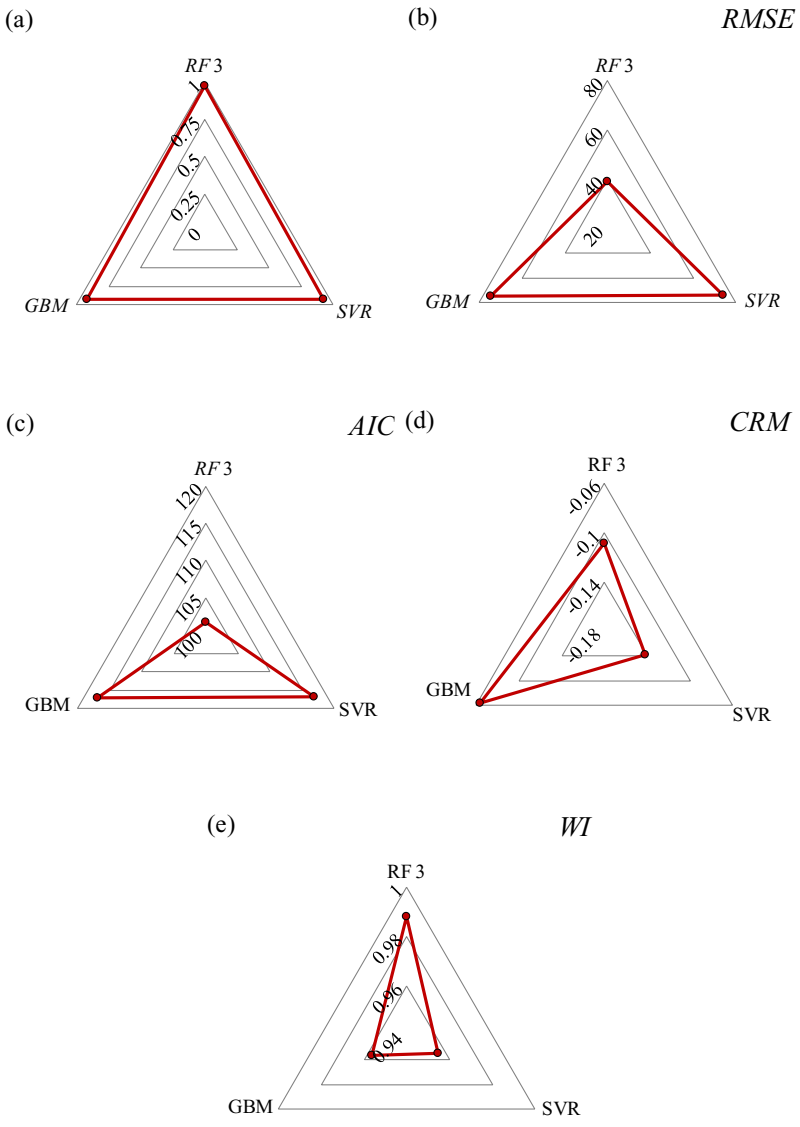


Figure 6-22. Key indices computed for RF 3, GBM, SVR, and empirical models to simulate vertical reaction forces (a) correlation coefficient (b) RMSE (c) AIC (d) CRM (e) WI

In Figure 6-23, the results of error analysis for RF 3, GBM, and SVR algorithms for modeling the vertical reaction forces are presented. The vast majority of the predicted vertical reaction forces by RF 3 (94%) possessed an error of less than 10%; however, this value for GBM and SVR models was 93%.

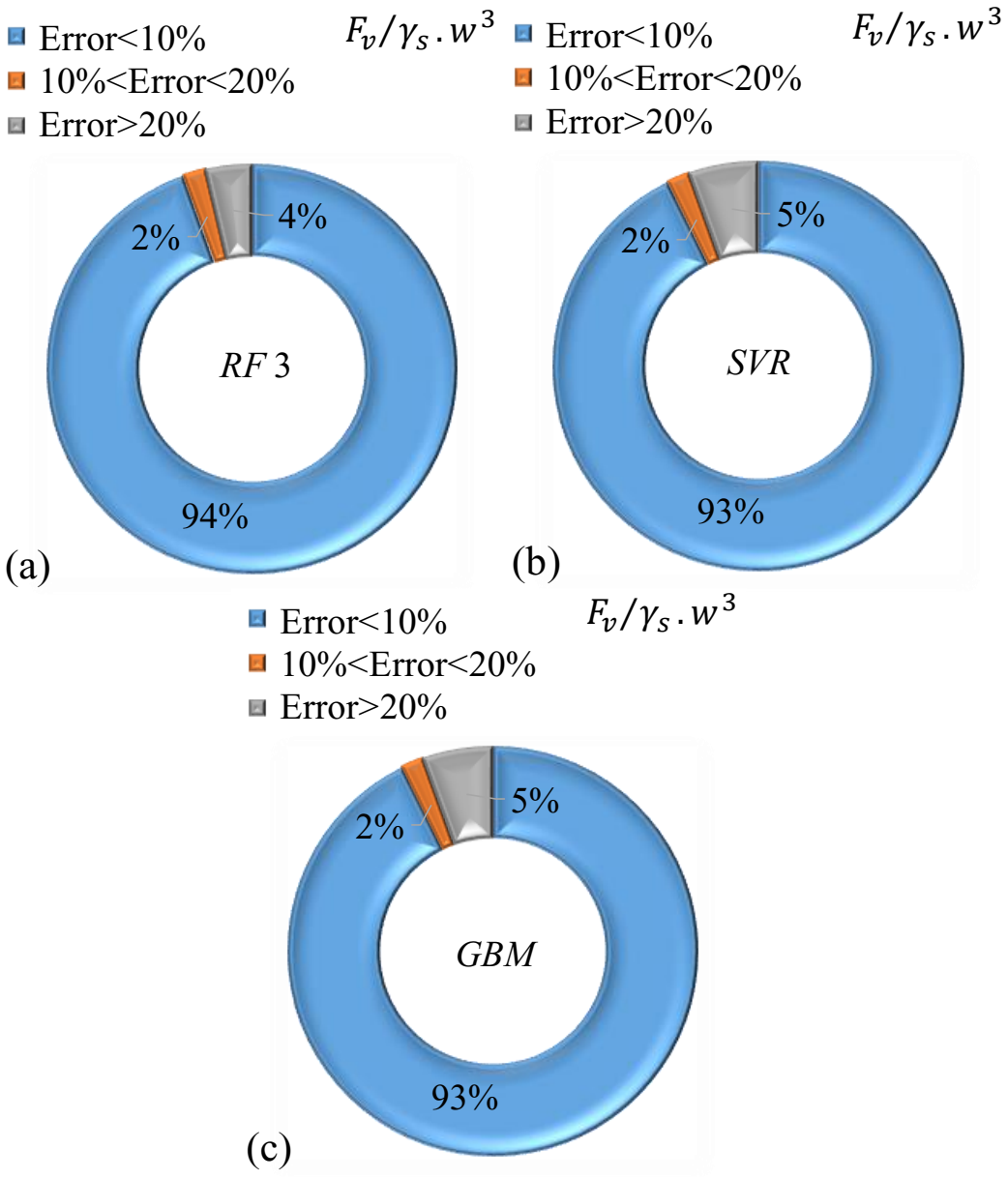


Figure 6-23. Result of MAPE index for (a) RF 3, (b) SVR, (c) GBM, and (d) empirical models to estimate the vertical reaction forces

The performance of the RF 2, GBM, SVR, and empirical models for simulation of the horizontal deformations are compared in Figure 6-24. The RMSE Index of the GBM, SVR, and Emp models was calculated to be 0.066, 0.151, and 0.425. According to the simulation results, the RF 2 model estimated the target function better than its counterparts.

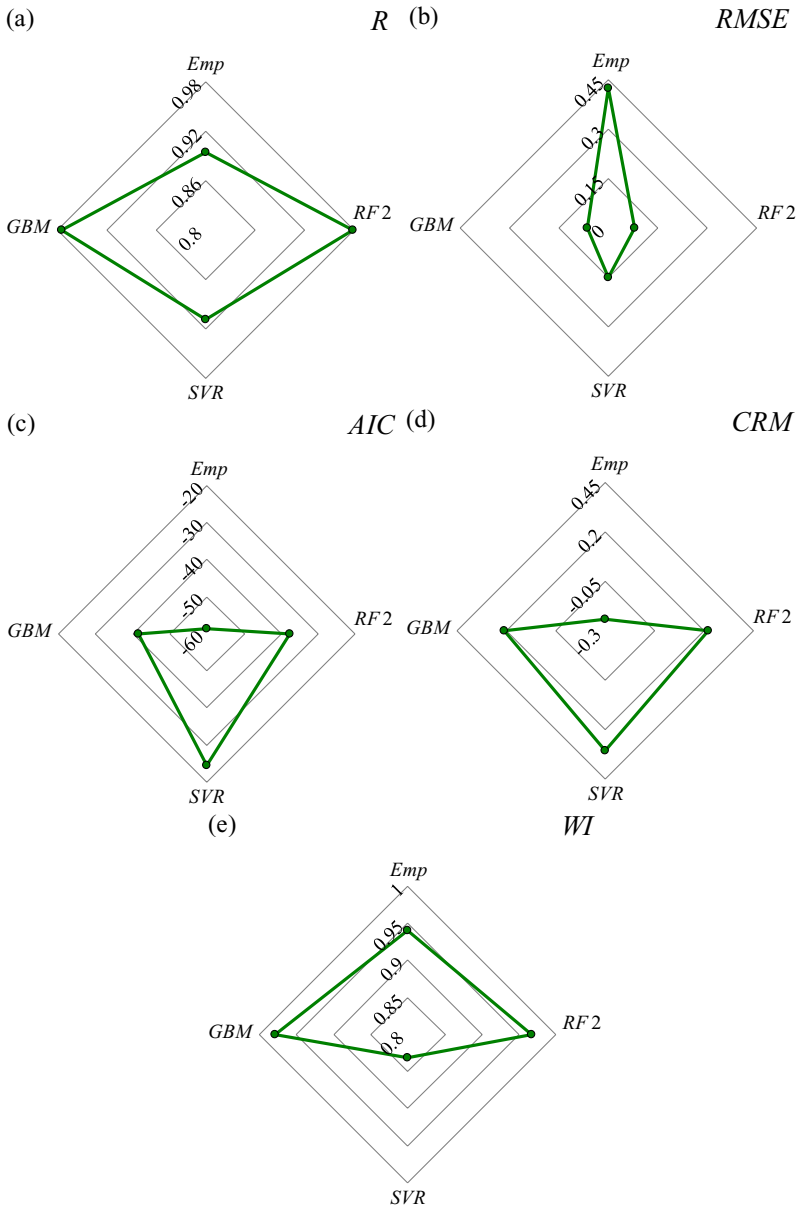


Figure 6-24. Key indices computed for RF 2, GBM, SVR, and empirical models to simulate the horizontal deformations (a) correlation coefficient (b) RMSE (c) AIC (d) CRM (e) WI

Error analysis results for the predicted horizontal deformations by the RF 2, GBM, SVR, and Emp models are depicted in Figure 6-25. Almost 15% of the modeled horizontal displacements by RF 2 represented an error of between 10% and 20%; rather this value for the SVR model was roughly 24%. Although about 72% of the obtained result of RF 2 had an error of more than 20%, this amount for the GBM and Emp models was approximately 85% and 76%. Therefore, amongst the four models, the RF 2 performed superior to the SVR, GBM, and Emp models in modeling the horizontal deformations.

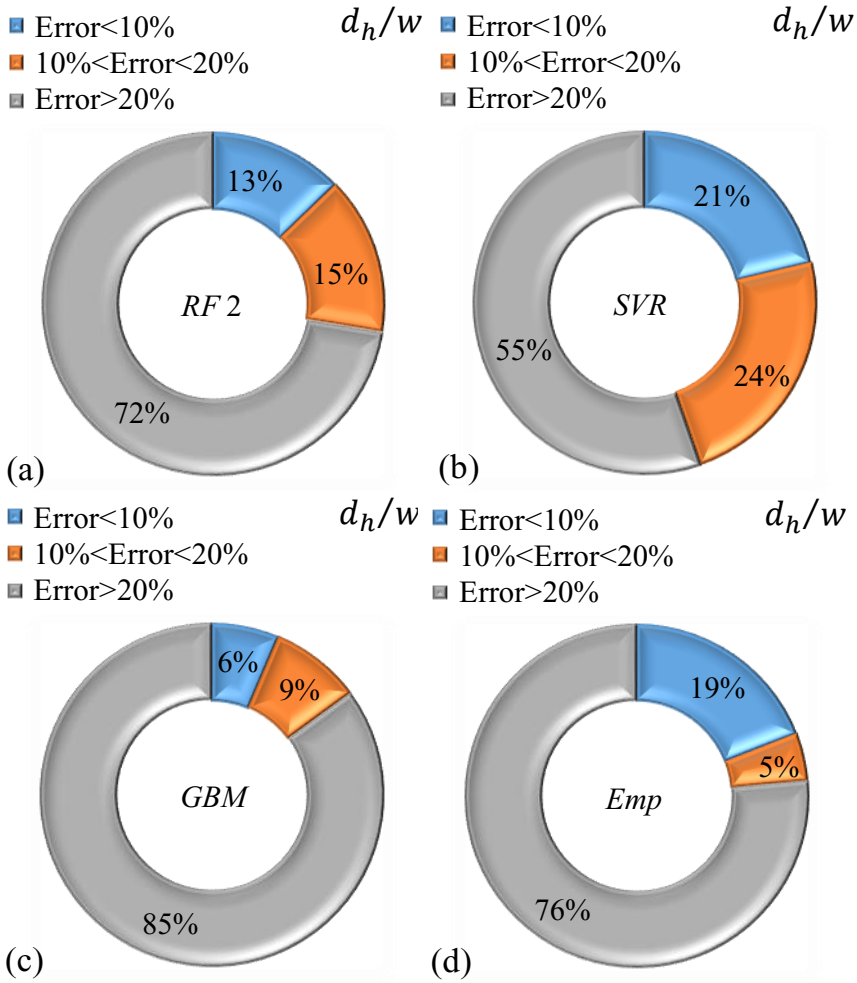


Figure 6-25. Result of MAPE index for (a) RF 2, (b) SVR, (c) GBM, and (d) empirical models to estimate the horizontal deformations

Figure 6-26 demonstrates the calculated indices for the simulated vertical deformation by RF 7, GBM, SVR, and Emp models. SVR and GBM provided a correlation coefficient of 0.156 and 0.153, respectively. The comparison showed that the empirical model (Emp) gave the worst estimation since its RMSE value was the highest (4.304) and its correlation was close to zero (R=0.047).

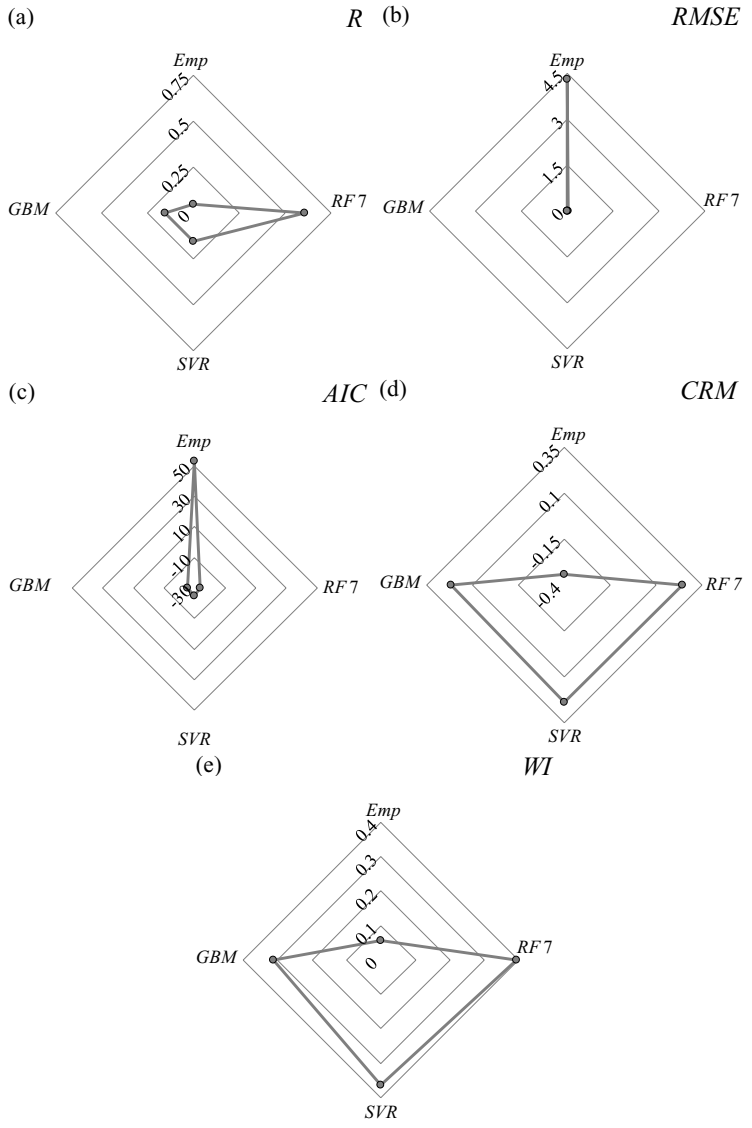


Figure 6-26. Key indices computed for RF 7, GBM, SVR, and empirical models to simulate the horizontal deformations (a) correlation coefficient (b) RMSE (c) AIC (d) CRM (e) WI

The error analysis for the vertical deformation predicted by the RF 7, GBM, SVR, and Emp models is presented in Figure 6-27. Though just about a quarter of the vertical deformations modeled by RF 7 possessed an error of between 10% and 20%, this value for the SVR model was obtained as about 27%. Almost all vertical deformation simulated by the empirical model had an error of greater than 20% and this amount for GBM was nearly 68%. As compared with the other models, the RF 7 model provided better performance in terms of accuracy, correlation, and complexity.

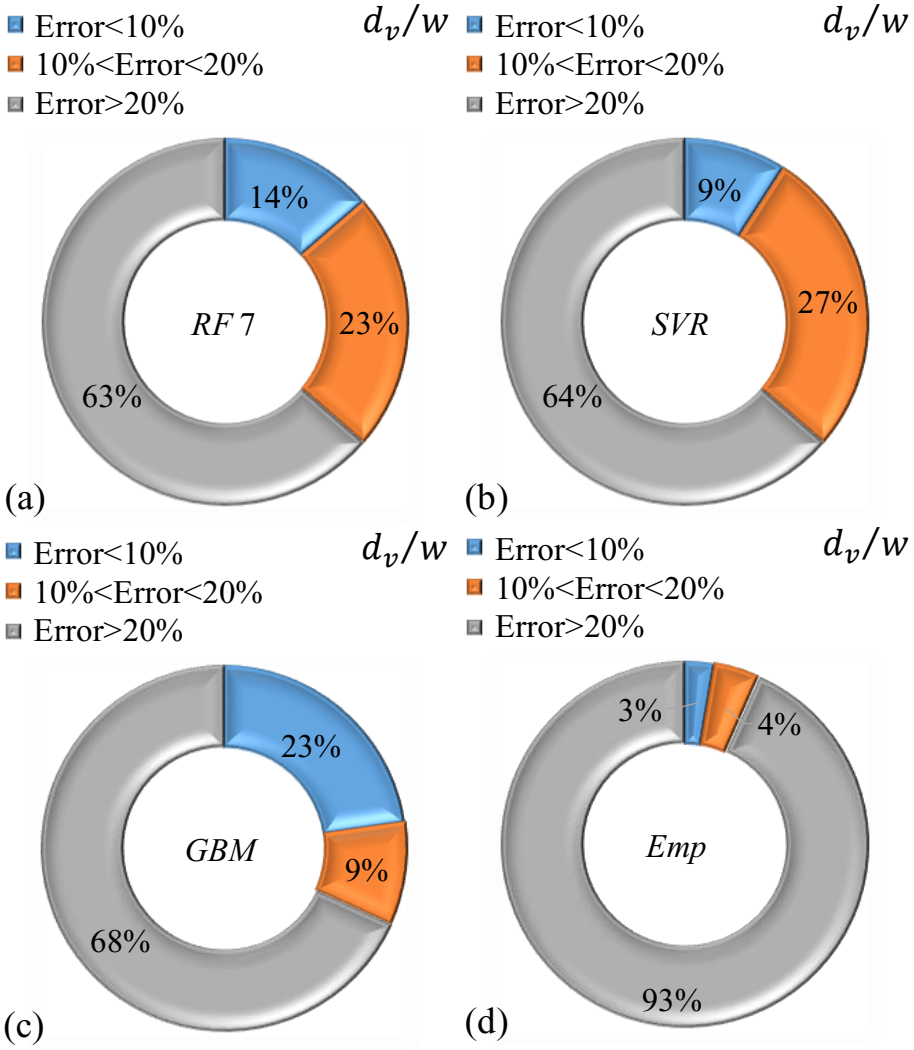


Figure 6-27. Result of MAPE index for (a) RF 7, (b) SVR, (c) GBM, and (d) empirical models to estimate the vertical deformations

To determine the performance of the RF, SVR, GBM, and empirical models, uncertainty analysis (UA) was conducted. To perform this UA, the errors obtained from each model (e_j) were calculated as the difference between the predicted ice-scoured parameters (P_j) and the experimental values (O_j) as:

$$e_j = P_j - O_j \tag{6-28}$$

The averaged computed errors $(\bar{e})$ and the standard deviation (S_e) of errors were obtained using the equations below (Azimi and Shiri 2021a):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{6-29}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{6-30}$$

A positive averaged estimated error $(\bar{e})$ means that the models overestimated the subgouge sand parameters, whilst a negative estimated error means that the models underestimated the experimental results. In the next step, regarding the "Wilson score approach" without the continuity correction, a confidence interval was built around the estimated error by using the $\bar{e}$ and S_e values (Azimi and Shiri 2021a). It is worthwhile noting that the Wilson score interval is an improvement over the normal distribution interval where an asymmetric normal distribution is utilized to rectify the confidence interval bounds and then a $\pm 1.96 S_e$ brought about a 95% confidence interval (95%CI). In Figure 6-28, the normal distribution graphs for the performed UA of the horizontal reaction forces predicted by RF 4, GBM, SVR, and Emp models are presented.

According to the conducted UA, the RF 4, SVR, and Emp models showed an overestimated performance, although the GBM model underestimated the target parameter. The 95%CI for the RF 4, SVR, GBM, and Emp models was obtained as (-15432 and 26778), (-89603 and 114840), (-98837 and 3223), and (4452548 and 7368490), respectively.

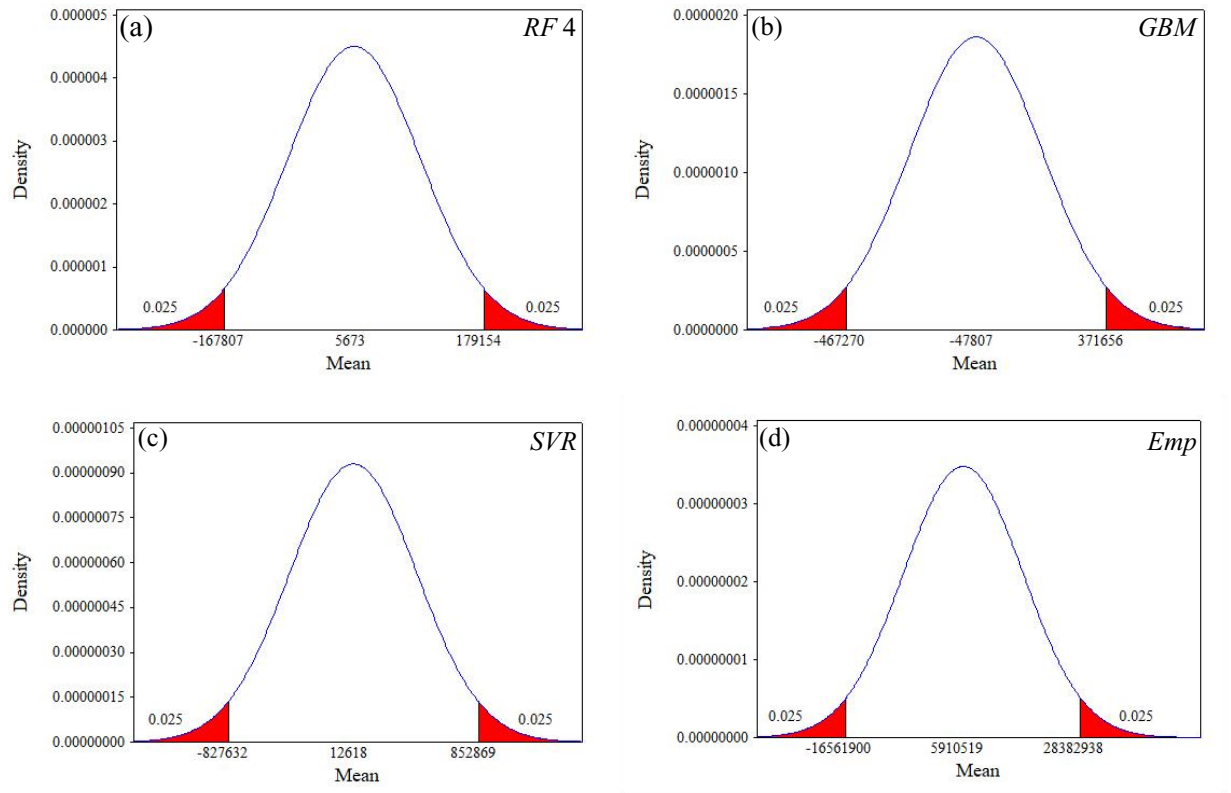


Figure 6-28. Normal distribution graphs for the performed uncertainty analysis of the horizontal reaction forces predicted by (a) RF 4, (b) GBM, (c) SVR, and (d) empirical models

Figure 6-29 depicts the normal distribution graphs for the UA of the vertical reaction forces estimated by RF 3, GBM, and SVR. The performed UA showed that the RF 3, GBM, and SVR models underestimated the vertical reaction forces, with average computed errors ($\bar{e}$) of -12.10, -15.91, and -7.1, respectively. The computed 95%CI for RF 3, GBM, and SVR were equal to (-22.28 and -1.93), (-35.65 and 3.84), and (-27.3 and 13.2).

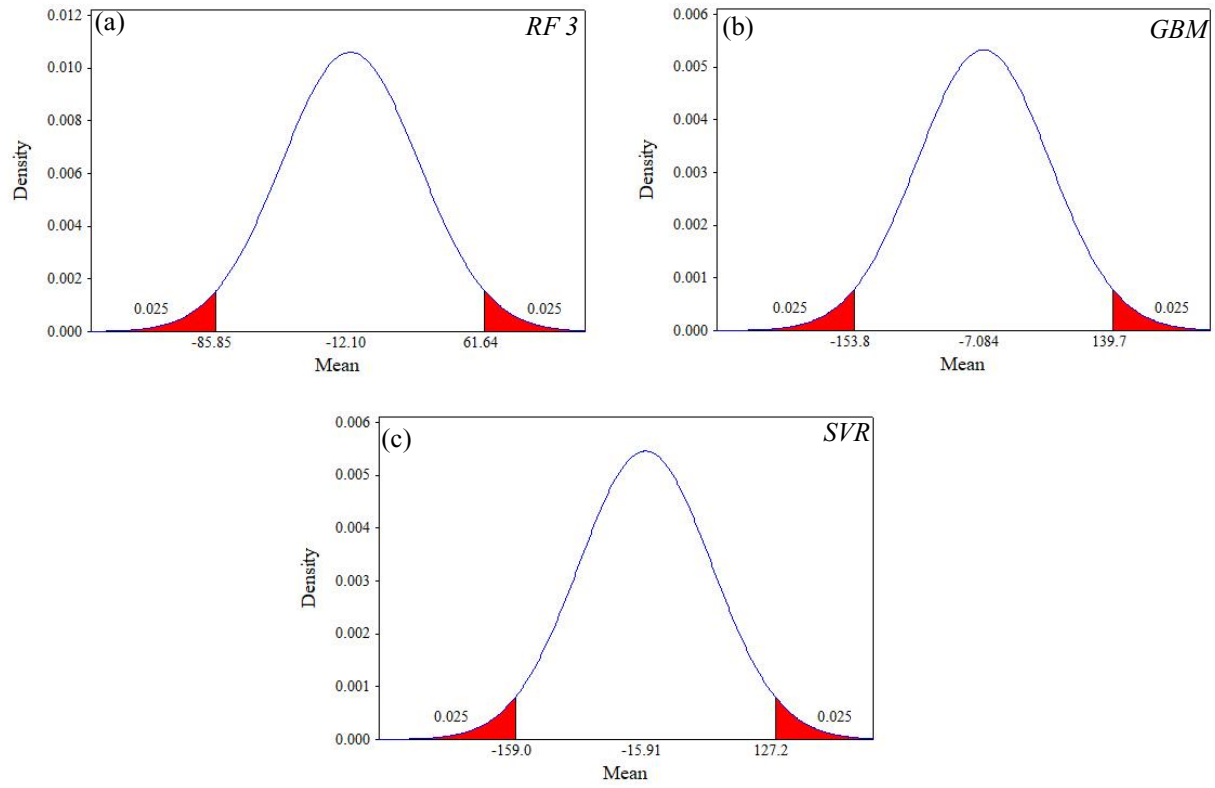


Figure 6-29. Normal distribution graphs for the performed uncertainty analysis of the vertical reaction forces predicted by (a) RF 3, (b) GBM, and (c) SVR

The normal distribution graphs for the UA of the horizontal deformations simulated by RF 2, GBM, SVR, and Emp models are shown in Figure 6-30. The RF 2 model had an overestimated error, with a 95%CI of (-0.003, 0.042). The SVR and GBM models similarly overestimated the target parameter; however, the empirical model underestimated the horizontal subgouge displacements. The value of 95%CI for the SVR, GBM, and Emp models was respectively approximated at (-0.017 and 0.071), (0.0001 and 0.038), and (-0.159 and -0.031).

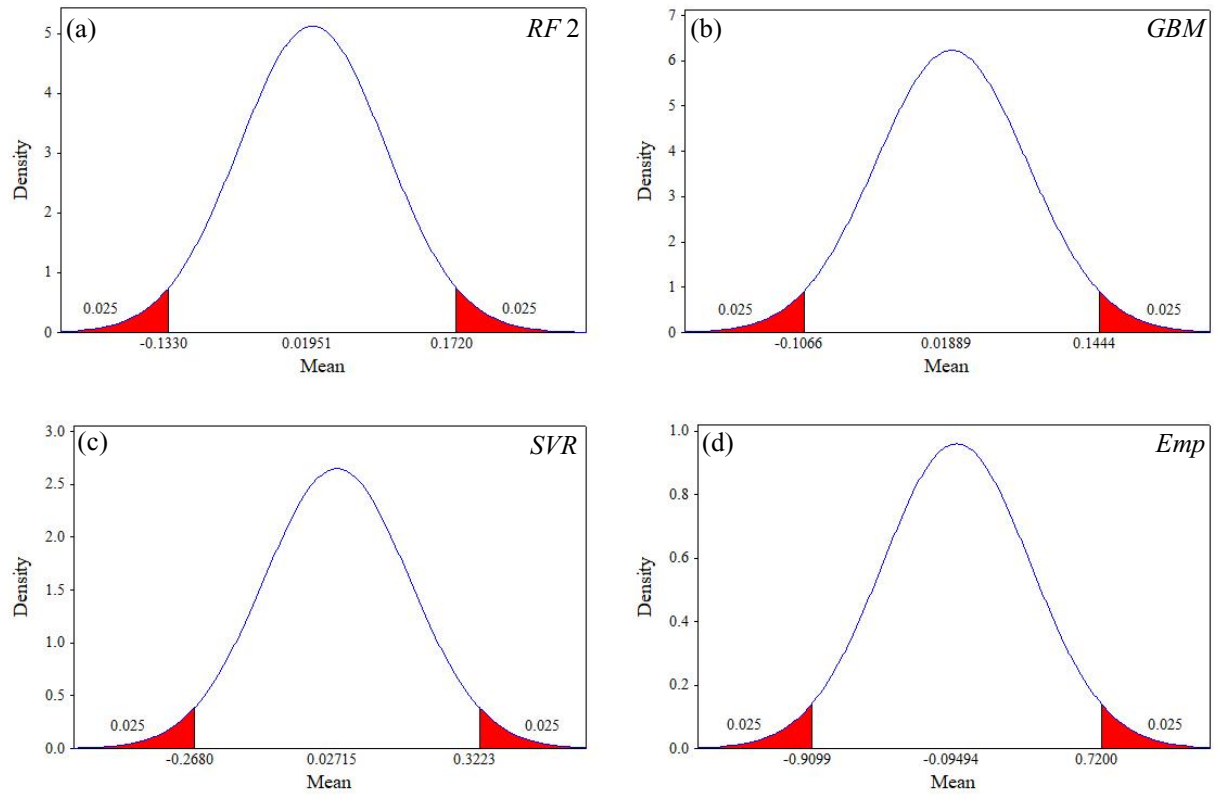


Figure 6-30. Normal distribution graphs for the performed uncertainty analysis of the horizontal deformations predicted by (a) RF 2, (b) GBM, (c) SVR, and (d) empirical models

Figure 6-31 displays the normal distribution graphs for the UA of the vertical deformations estimated by RF 7, GBM, SVR, and Emp models. The ML-based models (RF 7, GBM, and SVR) overestimated the vertical deformations, whilst the empirical model had an underestimated performance. The 95%CI for RF 7 was (-0.002 and 0.010) and this value for GBM, SVR, and Emp models equaled (-0.003 and 0.011), (-0.003 and 0.012), and (-1.176 and 0.749).

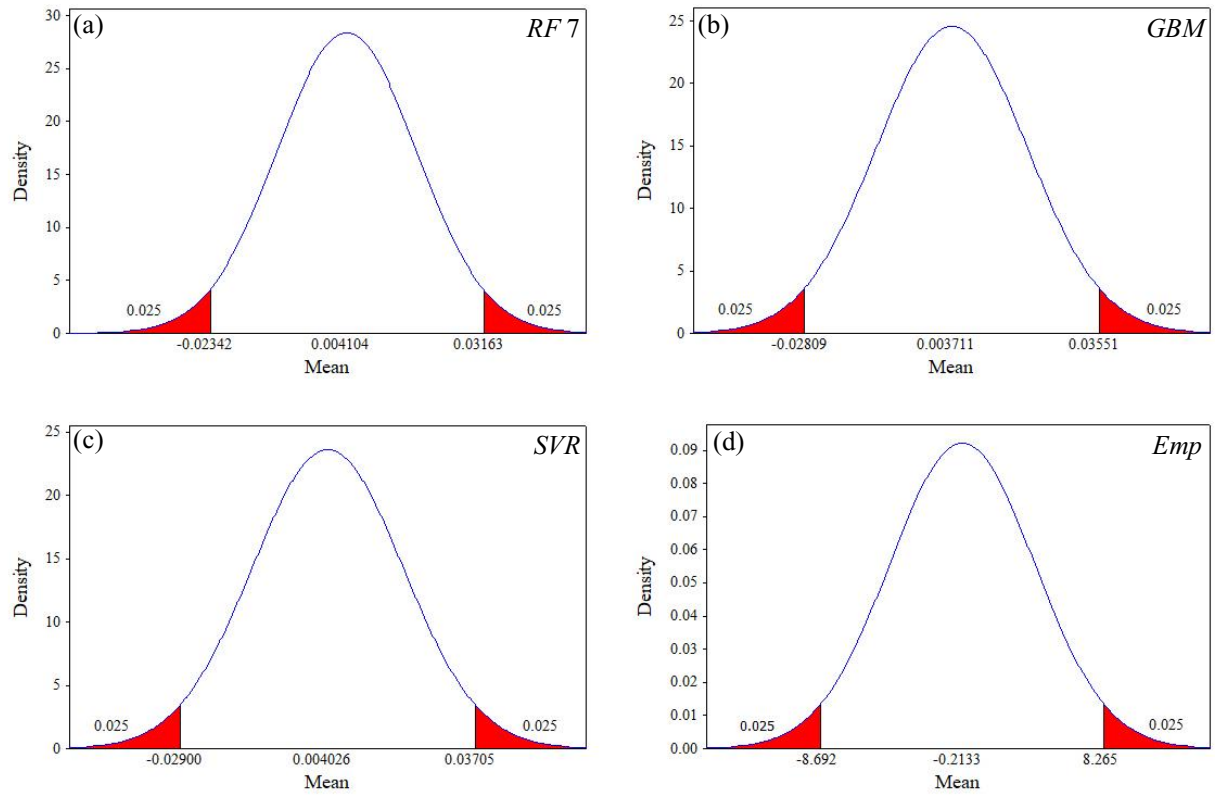


Figure 6-31. Normal distribution graphs for the performed uncertainty analysis of the vertical deformations predicted by (a) RF 7, (b) GBM, (c) SVR, and (d) empirical models

Therefore, the RF model overestimated the ice-induced sand displacements, while the vertical reaction forces and horizontal reaction forces were simulated by an underestimated and overestimated performance by the RF model, respectively.

The performance of the superior RF models was compared with other ML models (GBM & SVR) and empirical approaches to exemplify how accurate and flexible are the RF models. The empirical models that have been proposed based on a specific dataset showed an appropriate performance in a limited range of variables. In contrast, the RF models were able to estimate the subgouge soil features in a wide domain of datasets (six applied datasets), with the lowest degree of inaccuracy and complexity along with the highest level of correlation with the experimental values.

Furthermore, the RF algorithm was a meta-estimator that fitted several classifying decision trees on different sub-samples of the dataset and applied averaging to enhance the predictive precision and handled the over-fitting issue. The size of the sub-sample was also regulated with the max_samples hyperparameter when the bootstrap parameter was set as True; otherwise, the entire dataset was going to be utilized to construct each tree. The GBM and SVR models lacked these features. Therefore, the RF algorithm was found as the best method to simulate the subgouge soil characteristics in the current study.

6.1.13.3. Comparison of RF predictions with test results

RF was the premium model to simulate the ice-induced characteristics in the current investigation. Figure 6-32 compares the modeled horizontal reaction forces by RF 4 with experimental measurements. As shown, RF 4 had an acceptable performance to estimate the horizontal reaction forces, with a high level of accuracy and correlation along with a low level of complexity. The minimum amount of the horizontal reaction forces occurred just at the beginning of the iceberg position and this parameter increased along the scour axis. Although several oscillatory behaviors were observed in the applied experimental values (C-2, C-4, H-1, Y-2, and Y-3), the RF 4 model attempted to well predict the horizontal reaction forces. The RF 4 model could simulate the horizontal reaction forces with a non-linear trend, and this model showed an overestimated performance in some cases (C-2, H-1, and H-2) but the C-4 and Y-3 tests were modeled by using an underestimated performance.

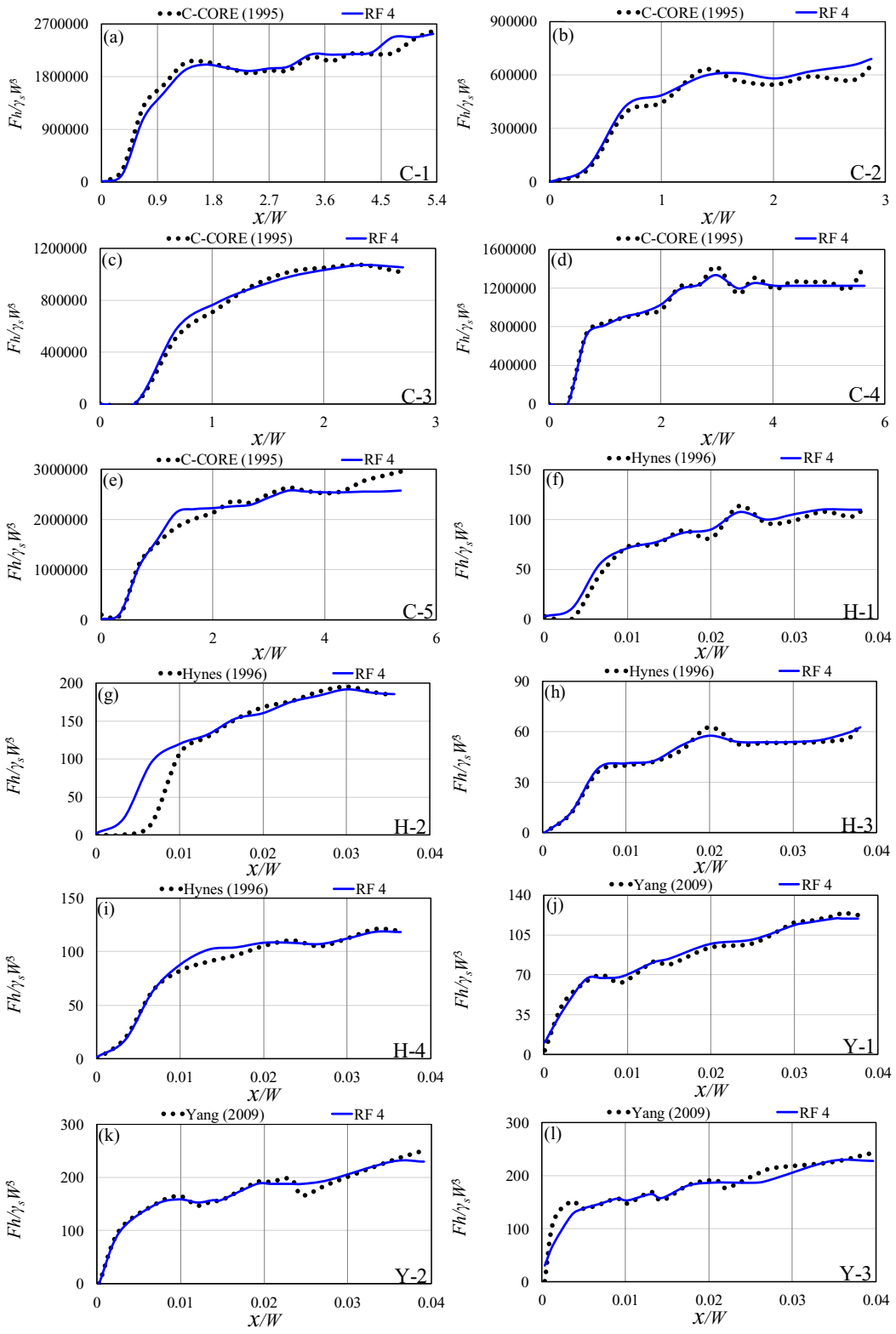


Figure 6-32. Comparison between the simulated horizontal reaction forces by RF 4 with experimental values

The predicted vertical reaction forces by RF 3 were compared with the used experimental values in Figure 6-32. The magnitude of the vertical reaction forces was near zero at the beginning of the scouring axis and then, after some oscillations, the vertical reaction forces reached their highest amount along the ice-gouging axis. However, the RF 3 model could simulate the target parameter using a nonlinear trend with acceptable accuracy. Even though the RF 3 model overestimated the H-1, H-5, and Y-3 tests, this ML model was able to well simulate the H-2, H-3, H-4, and Y-2 tests.

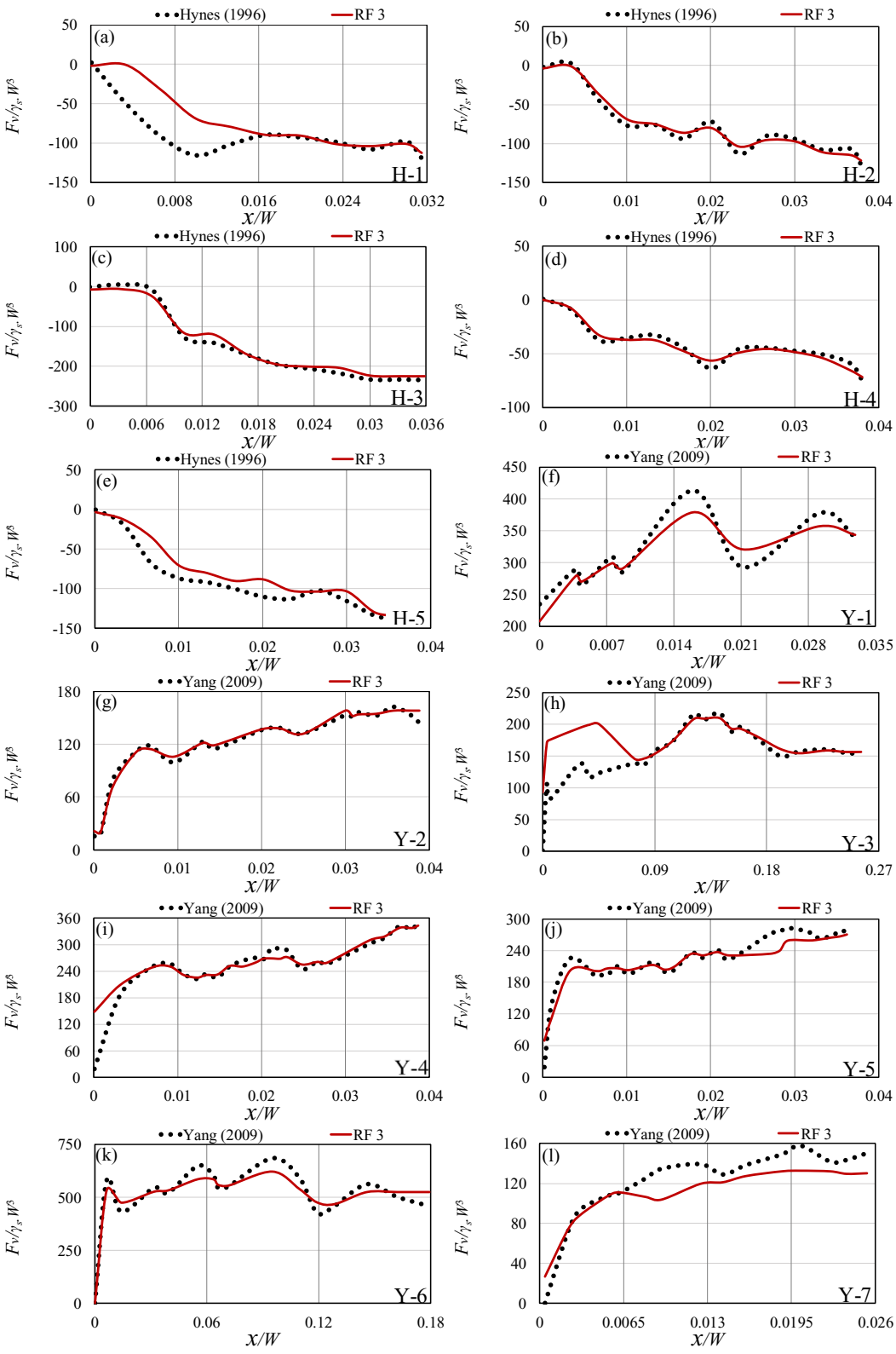

Hynes (1996)
RF 3
(a)
H-1
(b)
H-2
(c)
H-3
(d)
H-4
(e)
H-5
Yang (2009)
(f)
Y-1
(g)
Y-2
(h)
Y-3
(i)
Y-4
(j)
Y-5
(k)
Y-6
(l)
Y-7
$F_v/\gamma_s W^3$
x/W

Figure 6-33. Comparison between the simulated vertical reaction forces by RF 3 with experimental values

In Figure 6-34, the simulated horizontal subgouge deformations by RF 2 are compared with the experimental values. The maximum horizontal subgouge displacements took place beneath the iceberg keel and the enlargement of this parameter decreased with increasing the soil depth. Though some discrepancies between the experimental values and simulation results were observed (C-2, C-3, C-4, C-5, H-1, C-10, Y-1, Y-4, and Y-5), RF 2 managed to model the horizontal deformation with its highest performance. The RF 2 model applied both linear (H-1, H-4, C-7) and non-linear (P-1 to P-5) behavior to estimate the horizontal subgouge deformations. Although there were several discrepancies between the simulated values and experimental measurements (C-2 to C-4, C-10, Y-1, Y-4, and Y-5), the P-1 to P-5, C-1, H-3, H-4, C-6, and C-9 tests were simulated with a high level of correlation.

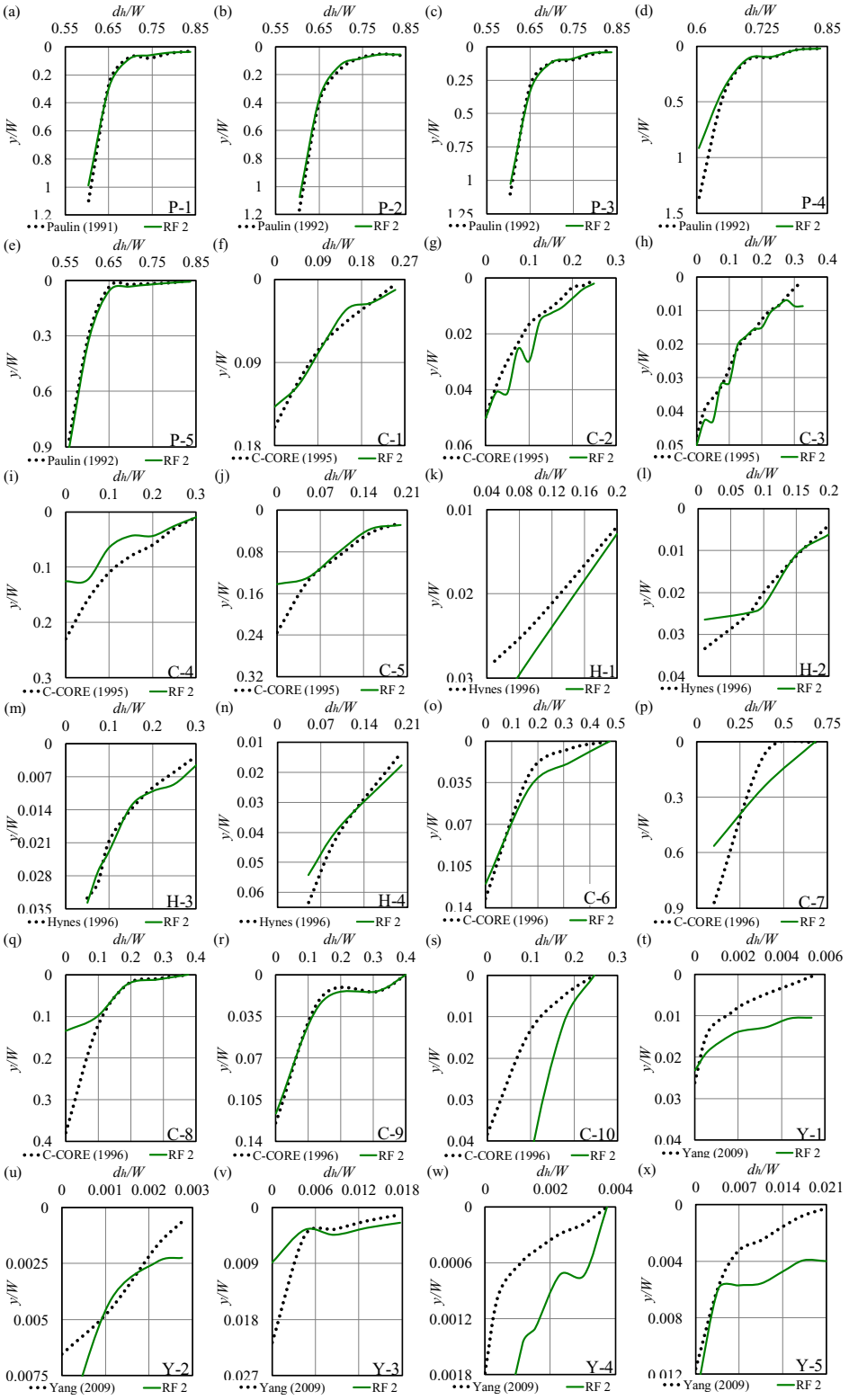


Figure 6-34. Comparison between the simulated horizontal deformations by RF 2 with experimental values

Figure 6-35 shows the comparison between the simulated vertical deformations by RF 7 with experimental measurements. The RF 7 model estimated the objective function by using y/W, Ds/W, α, h'/W, Lh/γs.W3, Lv/γs.W3, and V2/g.W. There were several fluctuations in experimental values (P-2 to P-4); however, the RF 7 model was able to simulate the vertical subgouge displacements with both linear (H-1 to H-4) and nonlinear patterns (C-1 and C-6).

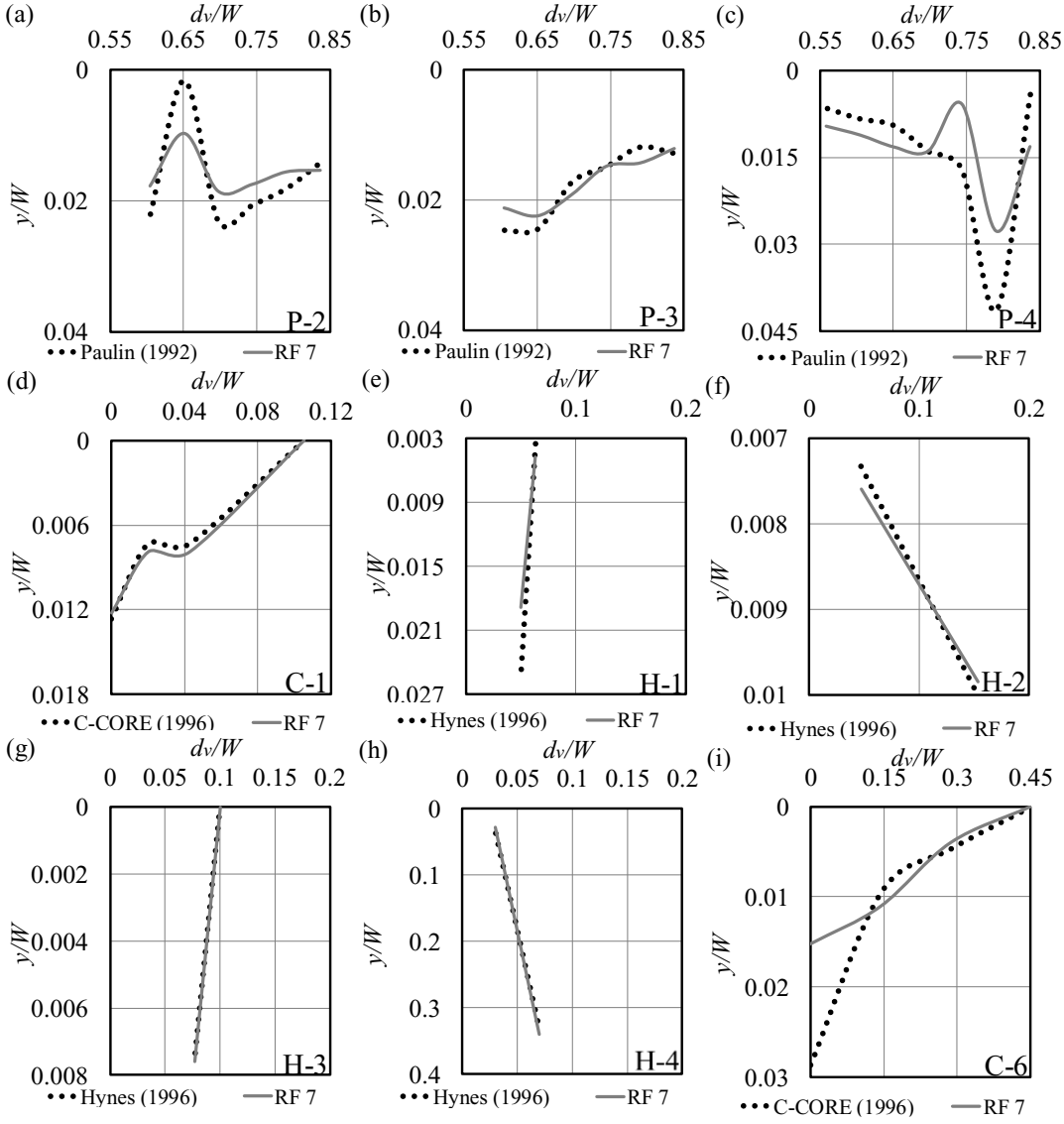


Figure 6-35. Comparison between the simulated vertical deformations by RF 7 with experimental values

The RF model as an ensemble ML model had a reasonable performance to predict the ice-scoured sand parameters comprising the reaction forces and sand displacements in the current study. The RF indicated a low level of complexity and a high level of precision along with a high level of correlation in comparison with the GBM, SVR, and empirical models.

6.1.14. Conclusion

In this study, the subgouge reaction forces and sand displacements were simulated by the RF model. Using the parameters affecting the iceberg-seabed interaction mechanism, nine RF models were built. Subsequently, 70% of the constructed dataset was used for training the RF model, whereas 30% of the remaining were employed for testing these models. By conducting a sensitivity analysis, the superior RF models and the most effective input variables were observed. Furthermore, the results of the RF, GBM, and SVR models, as well as empirical methods were compared. The most important outcomes of the current investigation are summarized as follows:

- RF 4, RF 3, RF 2, and RF 7 were known as superior models for the simulation of the horizontal reaction forces, vertical reaction forces, horizontal deformations, and vertical deformations, respectively.
- The best RF models demonstrated a good capability to predict the iceberg-gouging parameters. For instance, the value of correlation coefficient (R), Willmott Index (WI), and coefficient of residual mass (CRM) for RF 4 were estimated to be 0.995, 0.998, and 0.012, respectively.
- Regarding the sensitivity analysis, the position of the iceberg along the scour axis and the berm height ratio were regarded as the most influential inputs to model the reaction

forces; however, the soil depth ratio and the berm height ratio had a significant effect on the estimation of the subgouge displacements.

- The best RF models, with the highest level of precision, correlation, and lowest level of complexity, outperformed the GBM, SVR, and empirical models.
- RF 4, RF 2, and RF 7 overestimated the horizontal reaction forces, horizontal deformations, and vertical deformations, whereas the RF 3 model showed an underestimated performance to simulate the vertical reaction forces.

This research was the first step toward assessing the ability of the RF as an ensemble ML algorithm to improve our understanding of the iceberg-seabed interaction mechanism. The RF model demonstrated an appropriate performance in dealing with the simulation of the ice-gouging parameters. The present results may pave the way to develop some fast and cost-effective approaches for modeling the iceberg-seabed interaction problem, particularly at the early stages of design projects; whereas a fast and fairly accurate assessment is required to plan the construction methodology, corresponding logistics, and the future scope of detailed engineering. In addition, the results can benefit planning future numerical and experimental studies.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Ahmed, A.N., Van Lam, T., Hung, N.D., Van Thieu, N., Kisi, O., El-Shafie, A., 2021. A comprehensive comparison of recent developed meta-heuristic algorithms for streamflow time series forecasting problem. Applied Soft Computing, 105: 107282.

Alba, J.L., 2015. Ice Scour and Gouging Effects with Respect to Pipeline and Wellhead Placement and Design, Report, Bureau of Safety and Environmental Enforcement (BSEE), Washington, DC, USA, 100100.01.PL.REP.04.

Azimi, H., Shiri, H., 2020a. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021b. Modeling Subgouge Sand Deformations by Using Multi-Layer Perceptron Neural Network. 31st International Ocean and Polar Engineering Conference. Rhodes, Greece, ISOPE-I-21-2150.

Azimi, H., Shiri, H., 2021c. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., Malta, E.R., 2021. A non-tuned machine learning method to simulate ice-seabed interaction process in clay. Journal of Pipeline Science and Engineering, 1(4): 379-394.

Azimi, H., Shiri, H., 2021d. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04): 411-420.

Azimi, H., Shiri, H., Zendehboudi, S., 2022a. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology, 193: 103426.https://doi.org/10.1016/j.coldregions.2021.103426

Azimi, H., Shiri, H., Mahdianpari, M., 2022b. Simulation of Subgouge Sand Deformations Using Robust Machine Learning Algorithms. Offshore Technology Conference, OTC-31937-MS.

Azimi, H., Shiri, H., Mahdianpari, M., 2022c. Iceberg-Seabed Interaction Evaluation in Clay Seabed using Tree-Based Machine Learning Algorithms. Journal of Pipeline Science and Engineering, 2(4): 100075.

Barker, A., Timco, G., 2002. Laboratory experiments of ice scour processes: rigid ice indentor. Cold Regions Science and Technology, 35(3): 195-206.

Breiman, L., 2001. Random forests. Machine learning, 45(1): 5-32.

Campbell, P.J., 2014. Grand Banks Scour Catalogue (GBSC) GeoDatabase. Geological Survey of Canada, Atlantic.

Chari, T.R., 1980. A model study of iceberg scouring in the north Atlantic. Journal of Petroleum Technology, 32(12): 2247-2252.

C-CORE., 1995.Phase 3: Centrifuge Modelling of Ice Keel Scour. C-CORE Report 95-Cl2. St. John's, NL, Canada.

C-CORE., 1996. PRISE Phase 3c: Extreme LEE Gouge Event—Modeling and Interpretation. C-CORE Report 96-C32. St. John's, NL, Canada.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, 29(5): 1189-1232.

https://www.jstor.org/stable/2699986Hastie, T., Tibshirani, R., Friedman, J., 2009. Boosting and additive trees. In The elements of statistical learning, Springer., New York, pp. 337-387.

Hashemi, S., Shiri, H., 2022. Numerical Modeling of Ice–Seabed Interaction in Clay by Incorporation of the Strain Rate and Strain-Softening Effects. Journal of Offshore Mechanics and Arctic Engineering, 144(4): 042101.

Hashemi, S., Shiri, H., Dong, X., 2022. The influence of layered soil on ice-seabed interaction: Soft over stiff clay. Applied Ocean Research, 120: 103033.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand, Doctoral dissertation, Memorial University of Newfoundland, St. John’s, NL, Canada.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. In the 13th International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. 14th International Offshore and Polar Engineering Conference, ISOPE-I-04-107.

Lach, P.R., 1996. Centrifuge Modelling of Large Soil Deformation due to Ice Scour. Doctoral dissertation, Memorial University of Newfoundland. St. John’s. NL. Canada.

Li, X., Bai, R., 2016. Freight vehicle travel time prediction using gradient boosting regression tree. In the 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1010-1015.

Machemehl, J.L., Jo, C.H., 1989. Note on nearshore ice gouge depths in Alaskan Beaufort Sea. Journal of cold regions engineering, 3(3): 150-153.

Minerals Management Service., 2002. Evaluation of Sub–Sea Physical Environmental Data for the Beaufort Sea OCS and Incorporation into a Geographic Information System (GIS) Database. OCS Study MMS 2002–017.

Papusha, A., Kazunin, D., Gontarev, D., Vasilevich, V., Kyalbieva, S., 2013. Stability and strength of the subsea pipeline under iceberg load in Arctic. SPE Arctic and Extreme Environments Technical Conference and Exhibition, Moscow, Russia. SPE-166941-MS.

Paulin, M.J., 1991. Preliminary Results of Physical Model Tests of ice Scour, Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering, St. John's, NL, Canada.

Paulin, M.J., 1992. Physical model analysis of iceberg scour in dry and submerged sand. Master thesis, Memorial University of Newfoundland, Canada.

Prasad, K.S.R., Chari, T.R., 1986. Some factors influencing iceberg scour estimates. Journal of Energy Resources Technology, 108(3): 234-239.

Sahani, N., Ghosh, T., 2021. GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. Ecological Informatics, 64: 101352.https://doi.org/10.1016/j.ecoinf.2021.101352

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory, Springer, New York.

Woodworth-Lynes, C., Nixon, D., Phillips, R., Palmer, A., 1996. Subgouge Deformations and the Security of Arctic Marine Pipelines. Offshore Technology Conference, OTC-8222-MS.

Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Section 3

Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms

This section is a journal paper which has been published in the Journal of Pipeline Science and Engineering (2022), Vol. 2, pp. 100075, ISSN: 2667-1433

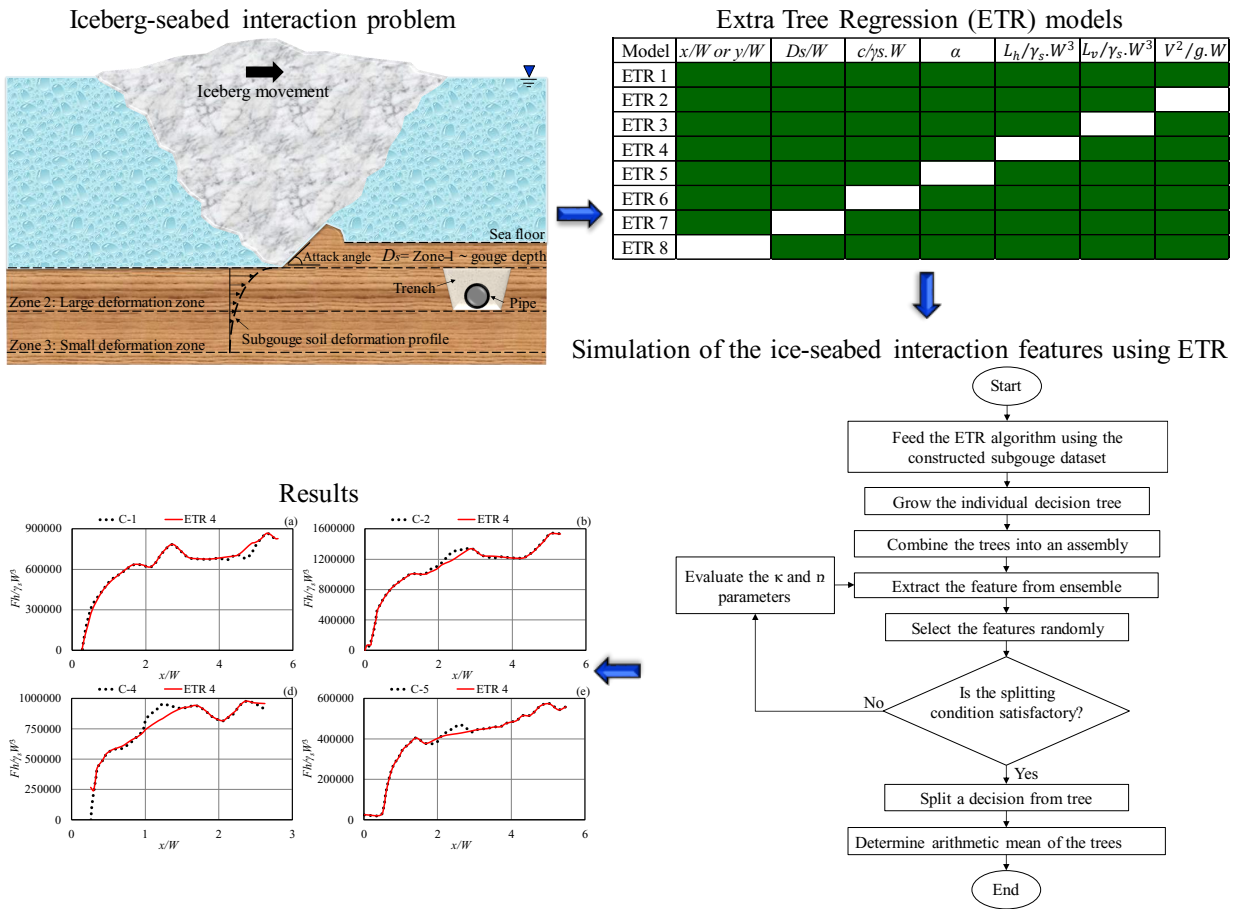
Abstract

The upcoming years are likely to witness a considerable rise in oil and gas exploitation in the Arctic regions, where the vast majority of these hydrocarbons can be transferred through the subsea pipelines from offshore to onshore. However, the operational integrity of the subsea pipeline may be at risk of collision with traveling icebergs, which gouge the seabed in the Arctic shallow waters. Even though these sea bottom-founded structures are buried at a secure depth below the seafloor, the pipeline is still threatened by the ice scouring event extended much deeper than the ice tip due to the shear resistance of the seabed soil. Modeling the subgouge soil characteristics is a challenging problem that requires costly experimental and long-running finite element (FE) simulations. To overcome these challenges, in this paper, the reaction forces and subgouge soil deformations in clay seabed were modeled using the advanced extra tree regression (ETR) algorithm, as a quick and cost-effective alternative in the early design phases of pipeline engineering projects. Eight ETR models, ETR-1 to ETR-8, were developed by using the input parameters governing the iceberg-seabed interaction problem.

The collected data were randomly split into 70% for training the machine learning (ML) models and 30% for validation purposes. The most accurate ETR models and the most significant input parameters were identified by performing a sensitivity analysis. The comparison of the most accurate ETR models and decision tree regression (DTR), random forest regression (RFR), and gradient boosting regression (GBR) algorithms proved that the ETR models had better performance to simulate the ice-induced soil features.

Keywords: Ice-seabed interaction process, Clay seabed, Extra tree regression (ETR), Decision tree regression (DTR), Random forest regression (RFR), Gradient boosting regression (GBR)

Graphical abstract



6.1.15. Introduction

Recent developments of oil and gases in Arctic areas have led to scientific interest in iceberg-seabed interaction processes where the moving icebergs can scour the seafloor and may collide with the subsea pipelines conveying these hydrocarbons exploited from offshore to onshore (Azimi and Shiri 2021a). The layout of the ice-seabed interaction process is schematically drawn in Figure 6-36.

The gouge depth (*Ds*) is specified as the scour incision deepness and the submarine pipeline is replaced in both horizontal (*dh*) and vertical (*dv*) directions during the ice-gouging occurrence.

The *Lh* and *Lv* indicate the consequence of horizontal and vertical loads incorporating drag force from wind, drag force from current, buoyancy force, keel weight, friction force on the base of the keel, Coulomb's passive friction force acting in front and on both sides of the keel, and driving force from the surrounding floe that are all among the input parameters to the problem (Azimi and Shiri 2021a).

The subsea pipelines and even other sea bottom-funded infrastructures, such as wellheads or communication cables, are commonly backfilled below the maximum estimated ice-gouging depth.

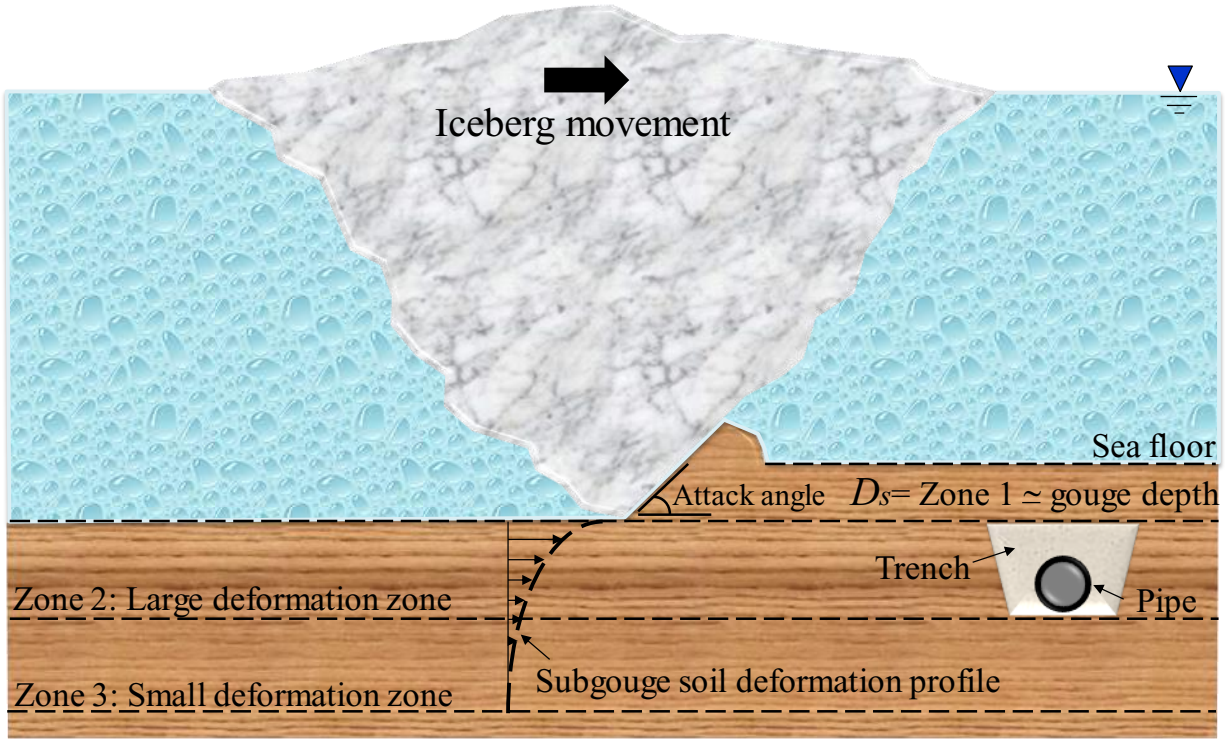


Figure 6-36. Layout of the ice-seabed interaction process schematically (Azimi et al 2021a)

As shown in Figure 6-36, the soil in front of the ice keel is displaced and the subgouge soil is also deformed by an exponential distribution at greater soil depth (ISO 19906: 2019 E). However, the structural integrity of the buried pipeline may be threatened by the traveling-icebergs, so the estimation of the required burial depth and subgouge soil characteristics play a crucial role to guarantee the pipeline operations safety involving costly centrifuge tests and long-running finite element (FE) analyses that are commonly conducted by a decoupled approach.

The decoupled approach combines the accuracy of the continuum FE analysis of the free field ice-gouging problem with the simplicity of the beam-spring models for pipeline simulation. In this case, first, a free-field ice-gouging analysis is performed without the presence of the pipeline. Then, the resultant subgouge soil deformations are manually transferred to the beam-spring model to obtain the pipe response to soil movement. Thus, the free field ice-gouging is an important part

of the decoupled approach in practical pipeline design and is usually supported by high-cost centrifuge and FE modeling.

The two past decades have witnessed the rapid development of machine learning (ML) algorithms in different fields, particularly the iceberg-seabed interaction process. For instance, a simple neural network (NN) model was applied for the simulation of the ice-scouring problem by Kioka et al. (2003, 2004). In the developed NN architecture, five neurons in the hidden layer were used and the ice drift velocity, sand property, sea-bottom gradient, and attack angle were put in the input layer. They verified the NN results with a mechanical approach and asserted that this methodology could predict the ice-seabed interaction parameters with an acceptable correlation. Barrette et al. (2015) discussed several standards addressing the subsea pipeline protection against the ice-scouring problem. The authors stated that the estimation of the required gouge depth, clearance underneath the required gouge depth, and the subsea infrastructure response were key phases to design the pipeline. Zvyagin and Heinonen (2017) presented a stochastic method to determine the parameters affecting the keel depths. The authors analyzed the ice-intrusion depths ranging from 6.5 m to 10 m and discussed both lognormality and an exponential distribution of gouge depth. The study concluded that the exponential distribution resulted in wide confidence intervals. Azimi and Shiri (2020) introduced the dimensionless groups of parameters affecting the iceberg-seabed interaction mechanism in both clay and sandy seabed. The authors estimated the maximum horizontal subgouge soil displacements regarding the provided dimensionless variables by using the linear regression (LR) models. Azimi et al. (2021a) utilized a non-tuned ML model to approximate the reaction forces and the ice-induced soil deformations in clay seabed. In that study, the optimized number of neurons in the hidden layer and the best activation function were obtained as 20 and the sigmoid function, respectively. Azimi et al. (2021b) also simulated the horizontal

and vertical ice-scoured soil displacements in clay seabed utilizing an NN-based model entitled the generalized structure of group method of data handling (GS-GMDH). A set of GS-GMDH-based equations were proposed in order to calculate the horizontal and vertical deformations by the authors.

To study the ice keel-seabed interaction precisely, some centrifuge tests were done at the Center for Cold Ocean Resources Engineering (C-CORE). In these investigations, the reaction forces and the subgouge soil displacements in clay and sand seabed were measured. It was concluded that the value of the scouring force was enhanced, while the magnitude of other parameters such as the soil shear strength, scour width, and scour depth increased; however, the scouring force decreased as the attack angle grew (C-CORE 1995 & 1996). Lach (1996) conducted a centrifuge work to study the iceberg-seabed interaction mechanism. This author implemented nine centrifuge tests in different geometric and physical circumstances for clay seabed. It was highlighted that the scouring force could be estimated using ice-induced displacement, keel geometry, and undrained shear strength. The ice-induced soil displacements in clay seabed were studied by Schoonbeek et al. (2006). The authors evaluated the influence of keel velocity, gouge depth, and undrained shear strength of clay seabed and reported that the plastic deformations of clay seabed occurred beneath the traveling keel bottom. Been et al. (2008) utilized an idealized sub-scour model in an experimental investigation disregarding the presence of a pipe, stratified seabed, backfill characteristics, and trench geometry to measure the reaction forces and subgouge soil deformations in clay seabed. The authors demonstrated that there was a strong correlation between the scour depth and the clay properties.

In the current study, the extra tree regression (ETR) algorithm, as a quick and cost-effective alternative for modeling the ice-induced soil features, was applied to be fed into the beam-spring

models. Furthermore, the robustness of ETR was compared with decision tree regression (DTR), random forest regression (RFR), and gradient boosting regression (GBR) models. The results obtained may be applied at the primary phases of pipeline design projects where a rapid and accurate approximation of subgouge soil behavior is required for planning the construction methodologies, logistics, and correct engineering perspective. In this study, a dataset comprising C-CORE (1995, 1996), Lach (1996), Schoonbeek et al. (2006), and Been et al. (2008) experimental measurements was built.

Regarding the high expenses and difficulties related to the experimental and FE studies, the ETR model used in the current work may deduct the expenditures of future projects and progress the planning of the research investigations. More details about the applied algorithm and the obtained results are presented in the upcoming parts.

6.1.16. Methodology

6.1.16.1. Decision Tree Regression (DTR)

A tree data structure comprising a set of nodes and branches wherein every single node is considered as a decision tree (DT). The DT can be applied to solve both classification and regression problems. The DT consists of several sections, including a root node, some leaf nodes, internal nodes, and branches. The uppermost node in the tree is defined as the root node and the terminal nodes (leaf nodes) end with the labels of classes, whereas the non-leaf nodes are regarded as the internal nodes. These nodes connect each other by using the branches (Pekel 2020). In the current investigation, the mean squared error (*MSE*) is applied to control the fitness function in the DT algorithm.

In the current study, the hyperparameters of the DTR algorithms were selected based on a trial and error procedure, meaning that the value of *max_depth*, *max_leaf_nodes*, and *min_weight_fraction_leaf* was initially set as 2, 4, and 0.1, respectively, and the performance of the DTR model was evaluated. The value of these hyperparameters increased in each step until the DTR's performance reached a satisfactory level. The DTR algorithm simulated the ice-seabed interaction features with its highest degree of accuracy and correlation along with its lowest level of complexity when the value of hyperparameters including the max_depth, max_features, max_leaf_nodes, *min_samples_leaf*, *min_weight_fraction_leaf*, and splitter was respectively adjusted as 8, 'auto', 12, 1, 0.01, and 'random'.

6.1.16.2. Random Forest Regression (RFR)

The random forest (RF) algorithm has been developed by Breiman (2001) to solve classification and regression problems. The RF algorithm, as an ensemble learning method, is the extension of the Classification and Regression Trees (CART) algorithm. In contrast to the tree-based CART models, RF has a great capability to deal with a high-dimensional, multi-source dataset while not being over-fitted to the data (Breiman 2001). More specifically, several decision trees are created in the RF model, and decisions with the highest number of votes are chosen as the simulation results (Sahani and Ghosh 2021). The RFR model reduces the average error of the trees utilized through the weighted correlation between the residuals and the applied randomization.

Similar to the DTR algorithm, the RFR's hyperparameters were also chosen using a trial-and-error approach. This means that the initial value of max_depth, random_state, and the *number of estimators* was respectively regulated at 500, 50, and 10. The RFR algorithm showed its best

performance as the *max_depth*, *random_state*, *number of estimators*, and *verbose* values were tuned at 8000, 500, 100, and 1, respectively.

6.1.16.3. Extra Tree Regression (ETR)

ETR was initially driven from the RF algorithm by Geurts et al. (2006). The RF algorithm utilizes the bootstrapping, generation of the decision trees, and bagging, division of the decision tree leaves, to solve a regression problem. The ETR commonly generates a set of unexpected regression trees. By choosing the best subset, the decision-making procedure is implemented. The ETR and RF algorithms have two main differences: (1) In the ETR model, the leaves are split using the random selection from the cutting points (2) The ETR algorithm utilizes the entire training dataset to generate the trees to minimize the value of bias. The ETR algorithm (i) applies the number of features that are selected randomly in the leave using κ parameter (ii) minimizes the sample size for the separation of leaves utilizing the ɳ parameter. The feature selection strength is determined by the κ parameter; rather the strength of the averaged outcome noise is defined using the ɳ parameter. This means that the κ and ɳ parameters will lessen the overfitting problem in the ETR algorithm and enhance the performance of this model (Hammed et al. 2021). The flowchart of the ETR algorithm is illustrated in Figure 6-37.

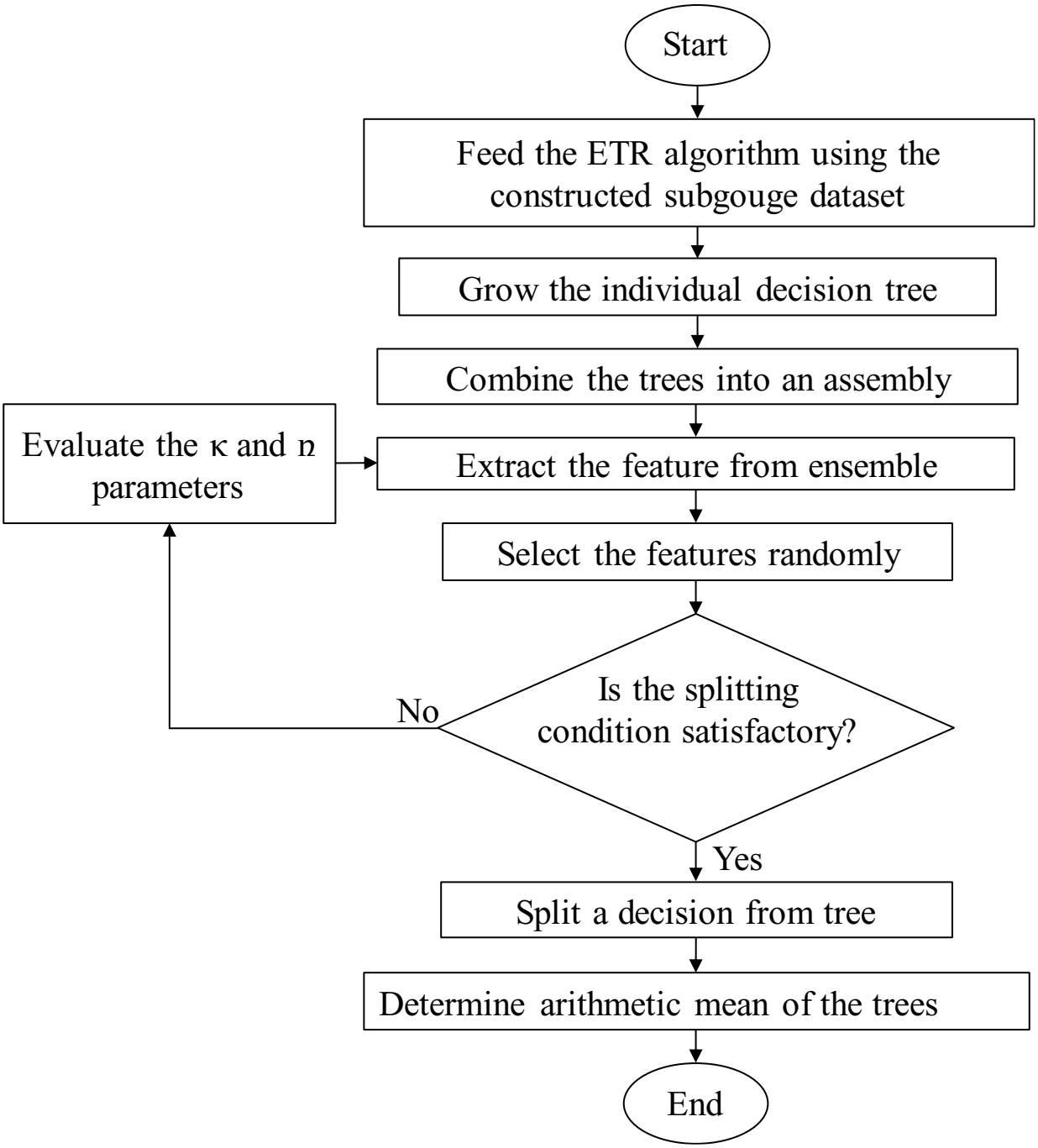


Figure 6-37. Flowchart of ETR algorithm

It should be stated that the ETR parameters were tuned through a trial-and-error approach in the current study. This means that the value of *n_estimators*, *random_state*, *criterion*, *max_depth*,

min_sample_split, and *min_sample_leaf* was regulated to be 100, 1, *'mse'*, 2000, 4, and 2, respectively.

6.1.16.4. Gradient Boosting Regression (GBR)

GBR is considered an ML technique to model various regression and classification problems. GBR creates an ensemble estimator, regarding the decision trees, to overcome the drawbacks of weak ML tools. In this model, a single decision tree may perform poorly; however, the ensemble algorithm, entitled gradient boosted trees, usually outperforms other decision-based models (Hastie et al. 2009). In the current study, the parameters applied for the GBR algorithm, including the *number of estimators*, *learning rate*, *random_state*, *loss function*, *max_depth*, and *verbose* were adjusted using a trial and error approach at 1000, 0.1, 2000, *Huber*, 5, and 1, respectively.

6.1.16.5. Seabed interaction process in clay

Seabed interaction parameters (η) in a clay mass comprising the soil deformations (*d/W*) and the reaction forces (*F/W*) are a function of the scour depth (*Ds*), the shear strength parameter of the clay (c), the width of gouge (*W*), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (*h'*), the horizontal load (*Lh*), the vertical load (*Lv*), the velocity of ice keel (*V*), and the specific weight of clay (γ_s) as follows (Azimi and Shiri 2020):

$$\eta_{(clay)} = f_1(D_s, c, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \quad (6\text{-}31)$$

Equation (6-31) can be written as a function of eight dimensionless groups as below (Azimi and Shiri 2020):

$$\frac{\eta_{(clay)}}{W} = f_2\left(\frac{D_s}{W}, \frac{c}{\gamma_s . W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}32)$$

It is worth noticing that the maximum subgouge deformation in the clay $\left(d_{max(clay)}\right)$ is formed just under the moving ice keel in the gouge centerline. However, at a greater depth on the subgouge centerline, by incorporating the soil depth (y/W), the subgouge clay displacement ($d_{(clay)}/W$) is written as follows:

$$\frac{d_{(clay)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}33)$$

Similarly, the ice-induced reaction force (F) is a function of the position of the iceberg along the scour axis (x) and $D_s, c, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s$, so equation (6-33) can be written as below:

$$\frac{F_{(clay)}}{\gamma_s W^3} = f_4\left(\frac{x}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s.W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (6\text{-}34)$$

The performance of ML models on simulation of both the soil deformations and the reaction forces was independently evaluated. This means that the subgouge deformations and the reaction forces were separately simulated using the ML algorithms. In other words, the ML models were fed with x/W, Ds/W, $c/\gamma s.W$, α, $Lh/\gamma s.W^3$, $Lh/\gamma s.W^3$, and $V^2/g.W$ as input parameters to simulate the dh/W and dv/W as output (Eq. 6-33). In addition, the y/W, Ds/W, $c/\gamma s.W$, α, $Lh/\gamma s.W^3$, $Lh/\gamma s.W^3$, and $V^2/g.W$ were employed as inputs of the ML models to estimate the $Fh/\gamma s W^3$ and $Fv/\gamma s W^3$ as output (Eq. 6-34).

In the current study, different laboratory investigations were used in order to simulate the ice-gouging clay parameters (d/W and $F/\gamma_s W^3$). In other words, the key measurements of five experimental investigations reported by C-CORE (1995), Lach (1996), C-CORE (1996), Schoonbeek et al. (2006), and Been et al. (2008) were utilized to verify the ML models.

The range of dimensionless horizontal subgouge deformations (dh/w) for the C-CORE (1995) dataset comprising C-1 to C-8 tests was from 0.003 to 0.233 and these deformations were measured at the dimensionless soil depth up to 0.025. The keel attack angle in the C-1, C-2, C-4, C-5, C-6,

and C-8 tests was reported as 15o, while this value for C-3 and C-7 was measured as 30o. The gouge depth ratio (Ds/W) in the C-1 to C-8 tests were respectively obtained at 0.053, 0.099, 0.033, 0.033, 0.054, 0.095, 0.079, and 0.080. The berm height ratio was not reported in the C-1 to C-8 tests.

The L-1 to L-8 tests were reported in Lach's datasets. The value of attack angle in L-1, L-2, L-3, L-4, L-5, L-6, and C-8 was 15^o but this parameter in the L-7 test was measured as 25^o. In the L-1, L-4, and L-7 tests, the value of the undrained shear strength ratio was 12.029, whilst this parameter for L-2 and L-3 was similar, with the $c/\gamma_s W$ of 9.452. The internal friction angle of sand seabed in L-5, L-6, and L-8 was respectively reported at 17.185, 24.059, and 15.037.

Moreover, the C-9 to C-12 tests have measured by C-CORE (1996) in which the value of the undrained shear strength ratio was respectively measured at 0.168, 0.167, 0.333, and 0.333. The attack angle in the C-9 test was 30o, whereas this parameter in C-10, C-11, and C-12 was observed to be 15^o. The berm height ratio in the C-9 to C-12 tests was not reported. The value of the ice velocity ratio ($V^2/g.W$) in the C-9 to C-12 tests was constant at 0.015.

In Schoonbeek et al.'s experimental setup (S-1), the gouge depth ratio, the undrained shear strength ratio, and the attach angle values were measured at 11.667, 13.637, and 14o, respectively.

The value of dimensionless horizontal subgouge deformations in Been et al.'s model (B-1 to B-5 tests) varied from 0.004 to 0.330 and the two models (B-2 and B-3) had the keel attack angle of 15o, while the attack angle in the B-1 test was equal to 45o. In the B-4 and B-5 tests, the value of attack angle was reported at 30o. Been et al. (2008) utilized a clay seabed with an undrained shear strength ratio of 1.444.

It should be noted that the angle of the surcharged soil slope (ω) and the height of the berm (h') were not reported by all applied laboratory studies. Therefore, other introduced dimensionless

groups in equations (14-3) and (14-4) were utilized as the input parameters for the ML model to simulate the horizontal subgouge deformation (dh/W), the vertical subgouge deformation (dv/W), the horizontal reaction force ($Fh/\gamma_s W^3$), and the vertical reaction force ($Fv/\gamma_s W^3$). Figure 6-38 shows the defined input combinations to develop ETR models in the current study.

In order to assess the pipeline behavior during the ice scouring event, the ultimate design aspects comprise the sub-gauge soil displacements (d) and the reaction forces of the iceberg keel (F) are predicted. These aspects are consequential of different key input parameters such as the attack angle (α), gouge depth (Ds), gouge width (W), seabed soil properties (φ), ice dynamics (V), environmental loads (L), etc. In the present research, similar to the previous investigations (e.g., Lach 1996; C-CORE 1995; C-CORE 1996; Azimi and Shiri 2021a), the ice scouring soil displacements (d) and the reaction forces (F) were regarded as the essential outcomes of the ice-gouging simulation.

Model	x/W or y/W	Ds/W	$c/\gamma s.W$	α	$L_h/\gamma_s.W^3$	$L_v/\gamma_s.W^3$	$V^2/g.W$
ETR 1	■	■	■	■	■	■	■
ETR 2	■	■	■	■	■	■	
ETR 3	■	■	■	■	■		■
ETR 4	■	■	■	■		■	■
ETR 5	■	■	■		■	■	■
ETR 6	■	■		■	■	■	■
ETR 7	■		■	■	■	■	■
ETR 8		■	■	■	■	■	■

Figure 6-38. The combination of input parameters to develop the ETR 1 to ETR 8

Eight ETR models (ETR 1 to ETR 8) were defined to simulate the ice-induced clay parameters. This means that ETR 1 was defined by using all input parameters and then the effect of each input

was ignored in the ETR 2 to ETR 8 models. To train the ML models, 70% of the constructed data was applied, whereas 30% of the rest was used to test the trained models.

6.1.16.6. Quality of fit

A set of criteria, such as correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were utilized to evaluate the accuracy, correlation, and complexity of the ML models. The closeness of the *R* and *WI* indices to one signified that the model had a high correlation with the experimental values. On the other hand, the closeness of the *RMSE*, *MAPE*, and *CRM* values to zero meant that a particular model had the lowest level of error though the complexity of these models was not assessed using the defined criteria. To overcome this limitation, the Akaike Information Criteria (*AIC*) was used, meaning that the less complex ML model owned the lowest value of *AIC*; hence, the premium model had the lowest values of *AIC* index and error (*RMSE*, *MAPE*, and *CRM*), with the highest level of correlation (*R* and *WI*) (Azimi et al. 2021a).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{6-35}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{6-36}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{6-37}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{6-38}$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \quad (6\text{-}39)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (6\text{-}40)$$

here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, and *n* are the experimental measurements, the simulated values, the average experimental values, the average simulated values, and the number of experimental measurements, respectively. In addition, k is the number of input parameters in the ML models.

6.1.17. Results and discussion

The sensitivity analysis for the ETR 1 to ETR 8 models was performed. Then, the robustness of the best ETR models was compared to the DTR, RFR, and GBR algorithms. Finally, the results of the superior ETR models were compared to the experimental measurements.

6.1.17.1. Sensitivity analysis of ETR models

Figure 6-39 demonstrates the calculated statistical indices for the horizontal reaction forces simulated by ETR 1 to ETR 8. The ETR 1 model adopted all input parameters so as to predict the horizontal component of the reaction forces, with a *WI* value of 0.994. Regarding the performed sensitivity analysis, ETR 1 was the six-best model for the estimation of the horizontal forces. Although the effect of the ice dynamics parameter ($V^2/g.W$) was excluded for ETR 2, this model was known as the fifth-best ETR model, where the value of *R* and *AIC* criteria for it was surmised as 0.988 and 656.577. The sensitivity analysis showed that ETR 3 was the fourth-best model since the influence of the vertical component of loads ($L_v/\gamma_s.W^3$) was disregarded for that, with the *CRM* and *AIC* indices of 0.006 and 655.956. ETR 4 as a function of x/w, D_s/W, $c/\gamma_s.W$, α, $L_v/\gamma_s.W^3$, and V^2/gW was the premium ETR model for the

simulation of the horizontal reaction forces, with the highest degree of correlation (*R*=0.989), the lowest level of complexity (*AIC*=653.553), and inaccuracy (*RMSE*=54630.154). The keel's attack angle (α) was an eliminated input factor for ETR 5 (the seventh-best model), while the value of *CRM*, *WI*, and *AIC* was respectively equal to 0.011, 0.994, and 657.811. For the third-best model (ETR 6), the impact of the clay shear strength ($c/\gamma_s.W$) was ignored where the *RMSE* and *WI* criteria for this model were reckoned as 56624.436 and 0.994. ETR 7 was ranked as the second-best model to estimate the horizontal forces once the gouge depth ratio (*Ds/W*) was the removed input parameter for this model, with a correlation coefficient value of 0.989. Amongst ETR models for simulation of the horizontal reaction forces, ETR 8 had the worst performance (*R*=0.794, *AIC*=737.183, and *WI*=0.898) because the effectiveness of the x/w input factor was eliminated. According to the implemented sensitivity analysis, the x/w had a striking effect to model the horizontal reaction forces, rather the attack angle (α), the ice dynamics parameter (V^2/gW), the vertical component of the load factor ($L_v/\gamma_s.W^3$), the clay shear strength parameter ($c/\gamma_s.W$), the gouge depth ratio (D_s/W), and the horizontal component of the load factor ($L_h/\gamma_s.W^3$) were ranked as the second important to seventh important input parameters.

Hence, ETR 4 and the x/w along with α variables were respectively recognized as the superior ETR model and the most significant input parameters for simulation of the horizontal reaction forces.

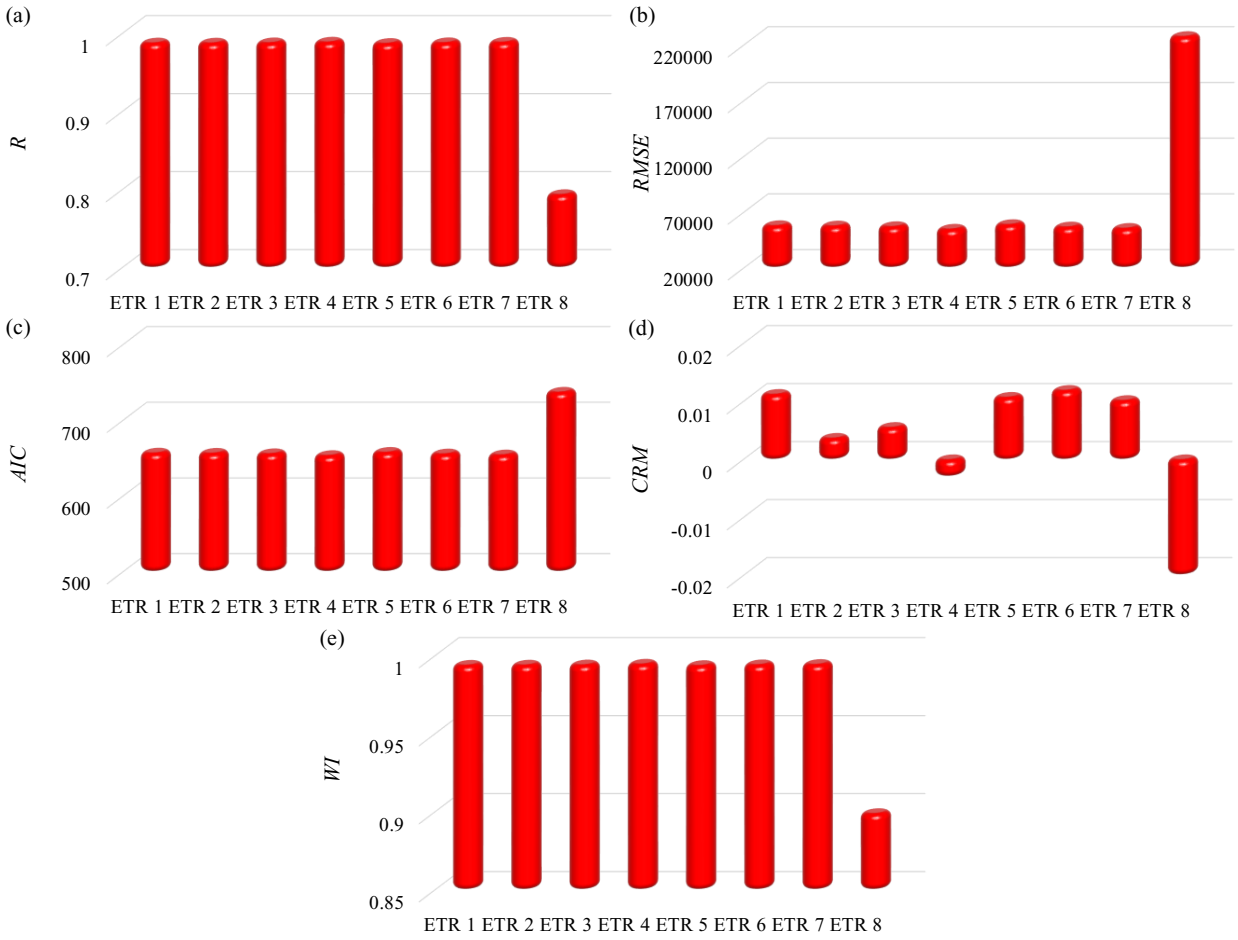


Figure 6-39. Accuracy assessment indices (a) R, (b) RMSE, (c) AIC, (d) CRM, and (e) WI for the horizontal reaction forces simulated by ETR 1 to ETR 8

The key statistical indices computed for the vertical reaction forces predicted by ETR 1 to ETR 8 are compared in Figure 6-40. The value of *CRM* and *WI* for the premium ETR model to estimate the vertical reaction forces, the ETR 6 model, were approximated as -0.006 and 0.997. This sensitivity analysis proved that ETR 2, ETR 3, and ETR 5 were respectively detected as the second-best to fourth-best ETR model to forecast the vertical component of reaction forces where the value of *AIC* criterion for these models were 225.696, 225.680, and 225.538, respectively. The ETR 1 model represented the fifth-best model for the estimation of the vertical forces, with an

RMSE value of 475.567. For ETR 4 as the sixth-best ETR model, the *R* and *AIC* values were calculated at 0.996 and 225.367. ETR 7 was determined as the seventh-best model and the value of *CRM* and *RMSE* indices for this model equaled to -0.006 and 473.307. ETR 8 possessed the poorest performance to model the vertical reaction forces, indicating that the *AIC* value for this model was estimated to be 265.194.

According to the sensitivity analysis, the position of the iceberg along the scour axis (x/W) showed up to be the most influential input parameter for modeling the vertical component of reaction forces. The second significant input parameter was recognized to be the gouge depth ratio (D_s/W), whereas the horizontal component of the load factor ($L_h/\gamma_s.W^3$) and the attack angle (α) were the third important and the fourth important input factors, respectively. The vertical load factor ($L_v/\gamma_s.W^3$) and the ice velocity parameter (V^2/gW) were ranked as the fifth important and the sixth significant input parameters for the simulation of the vertical reaction forces. The degree of effectiveness of the clay shear strength ($c/\gamma_s.W$) was insignificant to model the vertical reaction forces.

Thus, the ETR 6 model and the position of the iceberg along the scour axis (x/W) were introduced as the superior model and the most important input variable to estimate the vertical reaction forces, respectively.

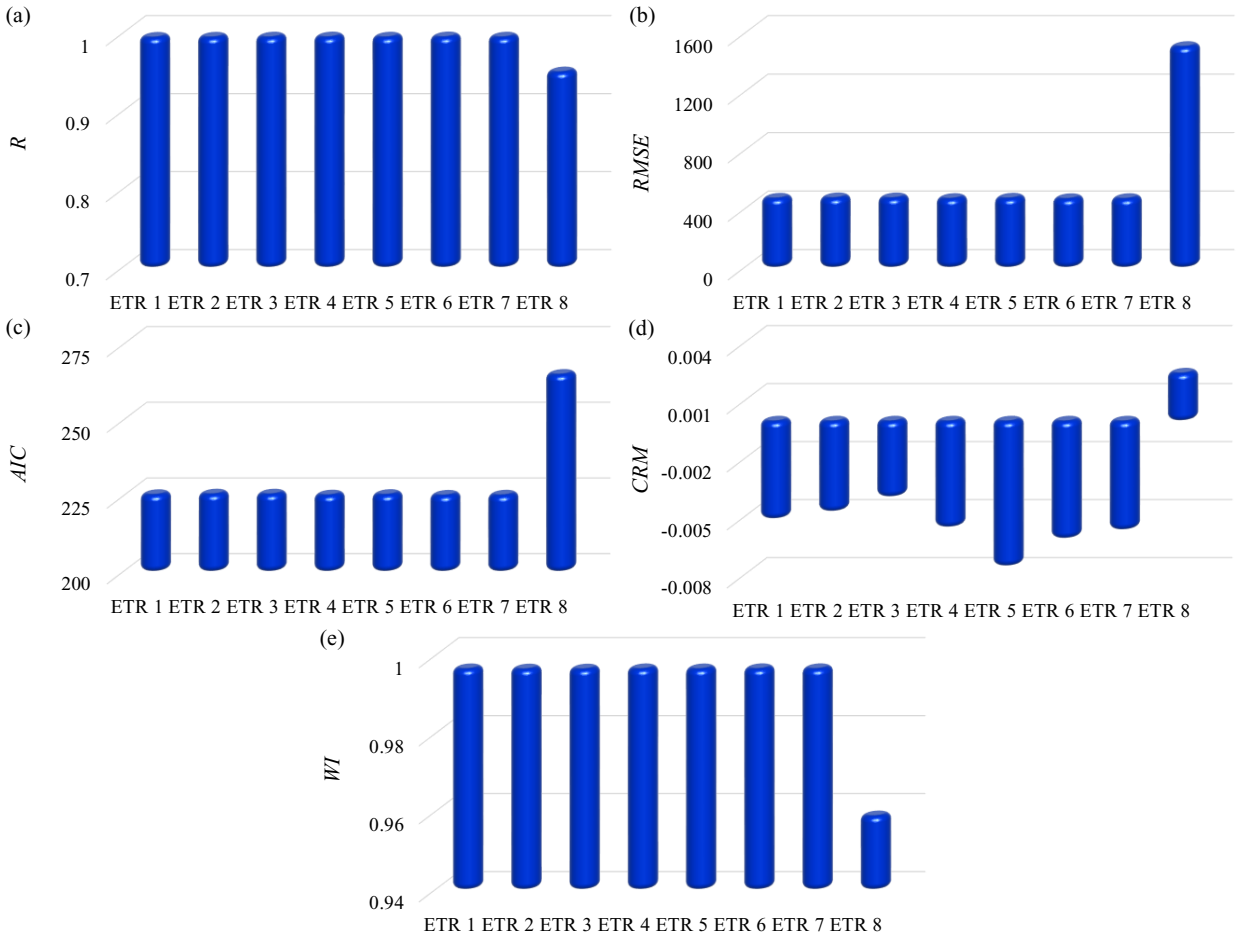


Figure 6-40. Accuracy assessment indices (a) R, (b) RMSE, (c) AIC, (d) CRM, and (e) WI for the vertical reaction forces simulated by ETR 1 to ETR 8

In Figure 6-41, the statistical indices computed for the horizontal subgouge deformations modeled by ETR 1 to ETR 8 are compared. As shown, ETR 6 was the excellent model for approximation of the horizontal subgouge deformations, with the highest level of correlation, precision, and lowest degree of complexity (R=0.999, $RMSE$=0.037, AIC=-56.264). However, ETR 8 demonstrated the worst performance to model the horizontal displacements, signifying that the value of AIC and $RMSE$ criteria for ETR 8 was obtained to be 0.463 and 0.529. With respect to the performed sensitivity analysis, ETR 5, ETR 2, and ETR 3 were preferred to be the second-best to

the fourth-best ETR model at dealing with the simulation of horizontal deformations. This means that the value of the *AIC* index for the ETR 5, ETR 2, and ETR 3 models was equal to -55.679, -53.137, and -51.624, respectively. For the fifth-best ETR model (ETR 7), the *RMSE* and CRM values were surmised at 0.057 and -0.044, while the indices for the sixth-best ETR model (ETR 4) were about 0.066 and -0.057, respectively. The RMSE and AIC criteria of ETR 1 were obtained as 0.091 and -36.939, making ETR 1 the seventh-best model.

The soil depth ratio (y/W) was found to be the most substantial input parameter to model the horizontal displacements, whilst the horizontal load factor ($L_h/\gamma_s.W^3$), the gouge depth ratio (Ds/W), the vertical load factor ($L_v/\gamma_s.W^3$), the ice keel velocity (V2/g.W), and the attack angle (α) were ranked as the second to sixth most significant input parameters. However, the clay shear strength ($c/\gamma_s.W$) had the lowest degree of effectiveness to model the horizontal deformations using the ETR algorithm.

Therefore, ETR 6 was preferred as the premium ETR model and the soil depth ratio possessed the highest impact so as to predict the horizontal subgouge deformations.

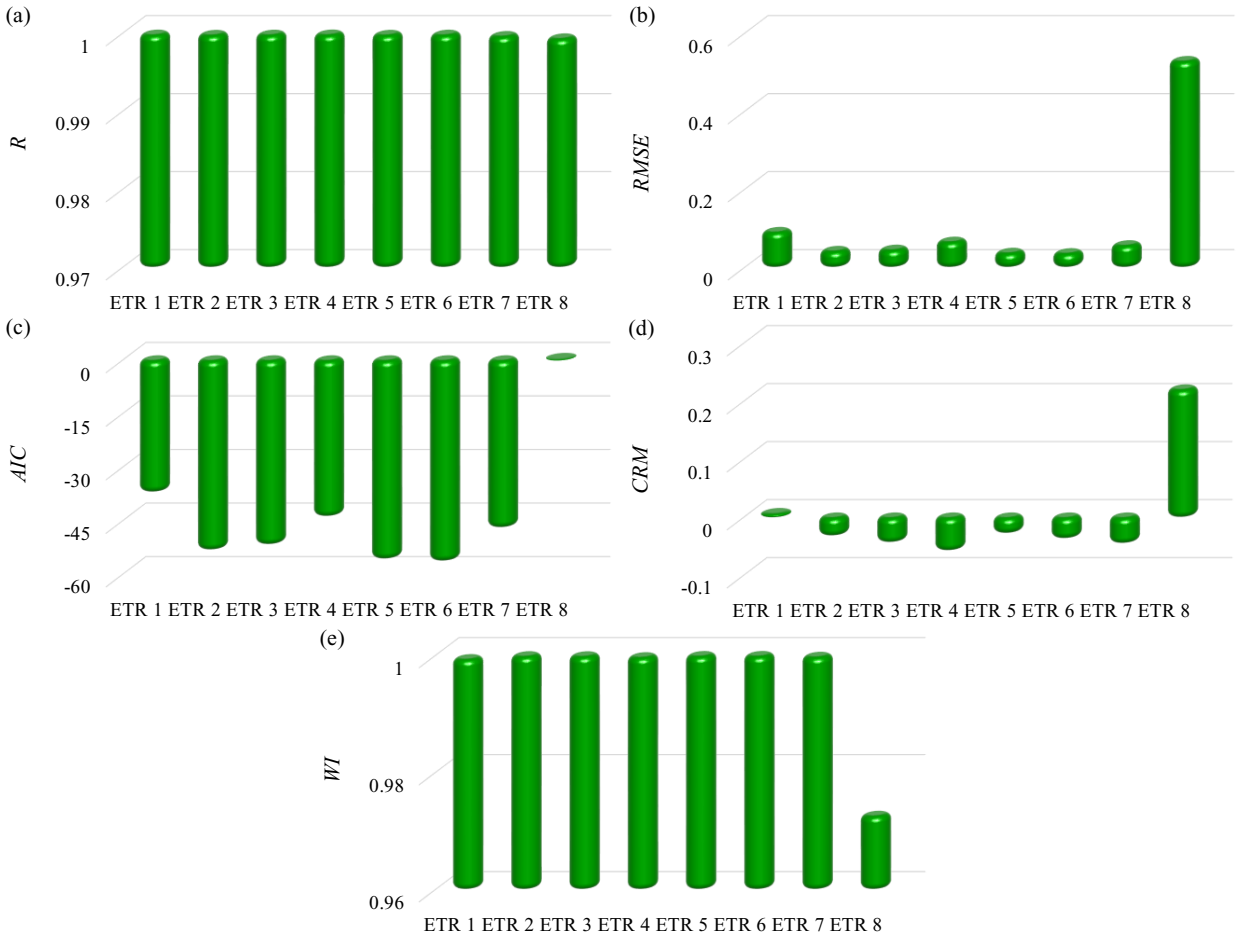


Figure 6-41. Accuracy assessment indices (a) R, (b) RMSE, (c) AIC, (d) CRM, and (e) WI for the horizontal subgouge deformations simulated by ETR 1 to ETR 8

The computed key criteria for the vertical ice-induced displacements estimated by ETR 1 to ETR 8 are illustrated in Figure 6-42. For the ETR 1 model as the fifth-best model, the *R* and *AIC* indices were obtained to be 0.963 and -53.633, whereas ETR 2 was ranked in the first position in terms of accuracy, correlation, and simplicity, with the *RMSE* and *WI* value of 0.009 and 0.977. ETR 3 was identified as the third-best model where the value of the correlation coefficient and Willmott index for the ETR 3 model were equal to 0.974 and 0.968. ETR 4, ETR 5, and ETR 6 were the fourth, seventh, and second-best models for simulation of the vertical deformations where the value of

correlation coefficient for these models was respectively calculated to be 0.971, 0.920, and 0.980. The simulation results showed that ETR 7 was the sixth-best ETR model, with an *AIC* criterion of -51.091. ETR 8 represented the weakest performance to model the vertical ice-induced deformation in clay, and the value of *RMSE* and *CRM* indices for ETR 8 was reckoned at 0.033 and -0.417. For the simulation of the vertical subgouge soil deformations in clay using the ETR algorithm, ETR 2 was distinguished as the superior model. Furthermore, the soil depth ratio and the attack angle possessed a remarkable influence in predicting the target parameter.

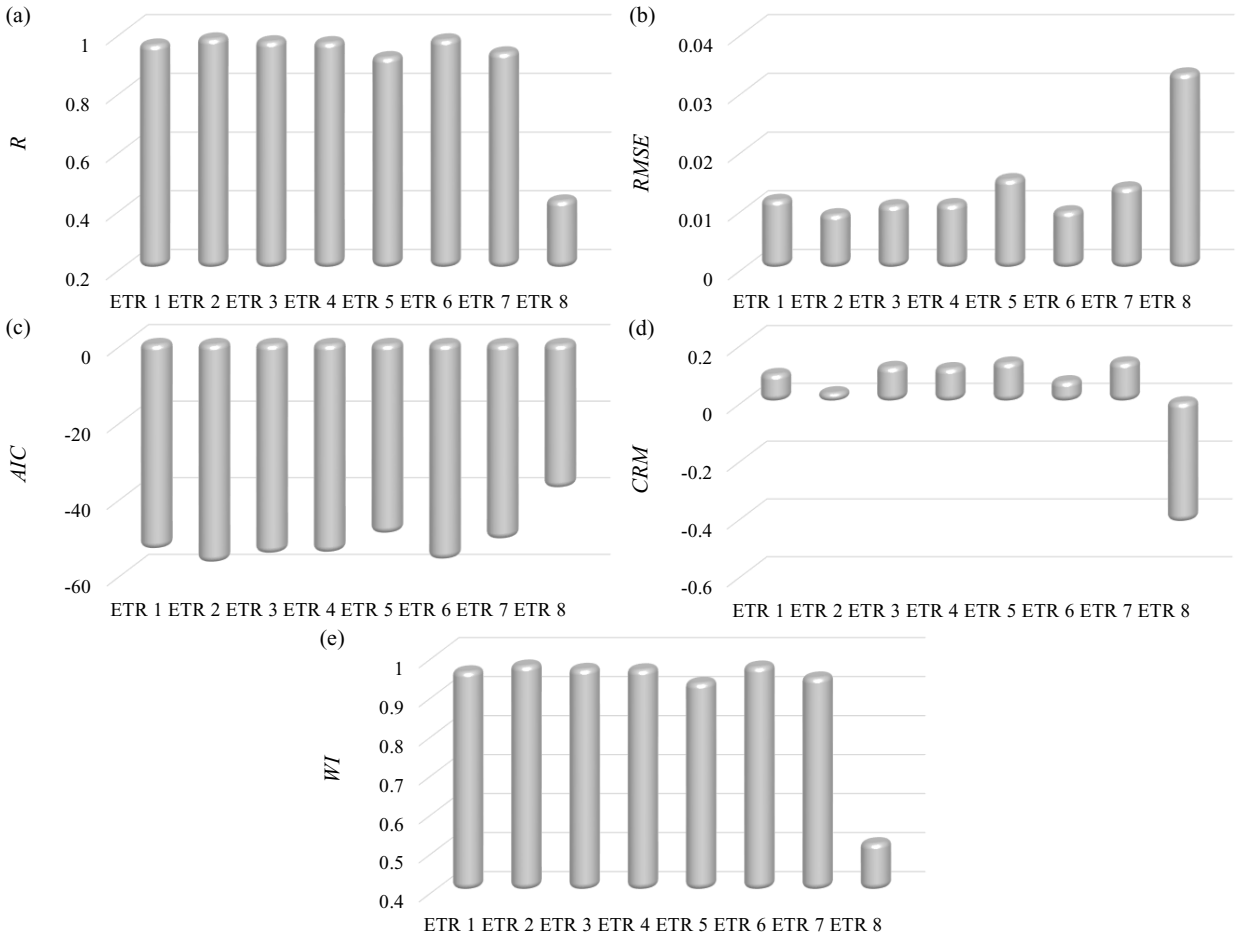


Figure 6-42. Accuracy assessment indices (a) R, (b) RMSE, (c) AIC, (d) CRM, and (e) WI for the vertical subgouge deformations simulated by ETR 1 to ETR 8

6.1.17.2. Comparison with DTR, RFR, and GBR models

The performed sensitivity analysis demonstrated that the ETR 4, ETR 6, ETR 6, and ETR 2 models were respectively the superior models to estimate the horizontal reaction forces, the vertical reaction forces, the horizontal subgouge deformations, and the vertical subgouge deformations. The performance of these ETR models was compared to the DTR, RFR, and GBR algorithms. In Figure 6-43, the capability of the ETR 4 model in the simulation of the horizontal reaction forces is compared to the DTR, RFR, and GBR methods. Even though the RFR algorithm represented the worst outcome among these approaches (R=0.927 and AIC=708.288), ETR 2 still had the highest amount of precision and correlation ($RMSE$=54630.154 and R=0.989). Based on the simulation results, the DTR and GBR algorithms were ranked as the second-best and the third-best decision-based ML methodology for modeling the horizontal reaction forces. Then, the ETR 2 managed to outperform its counterparts in the estimation of the target parameters.

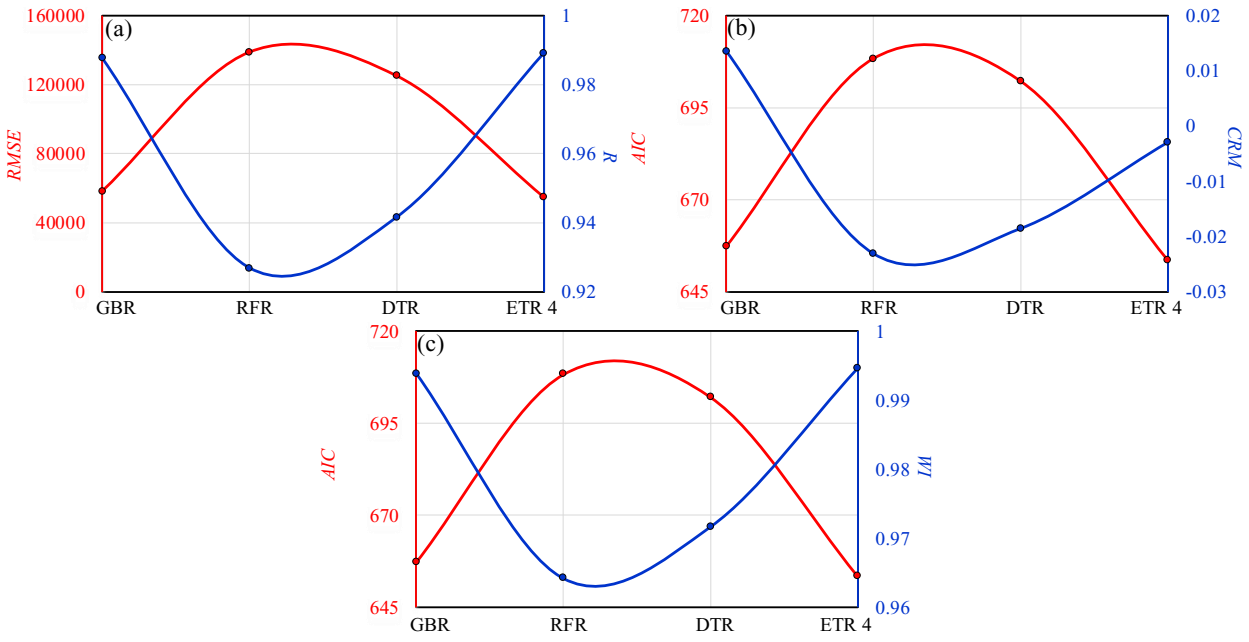


Figure 6-43. Comparison of the ETR 4 model in simulation of the horizontal reaction forces with the DTR, RFR, and GBR methods (a) RMSE and R (b) AIC and CRM (c) AIC and WI

Figure 6-44 shows the comparison of the ETR 6 model with other decision-based machine learning algorithms for modeling the vertical reaction forces. The best performance was obtained for ETR 6; however, the DTR algorithm had the lowest degree of accuracy and simplicity. In other words, the value of the RMSE index for the ETR 6, DTR, RFR, and GBR models was respectively calculated as 472.658, 1255.463, 529.059, and 506.399. Therefore, the most intriguing performance for the prediction of the vertical reaction forces was with the ETR 6 algorithm as a function of x/w, D_s/W, α, $L_h/\gamma_s.W^3$, $L_v/\gamma_s.W^3$, and V^2/gW.

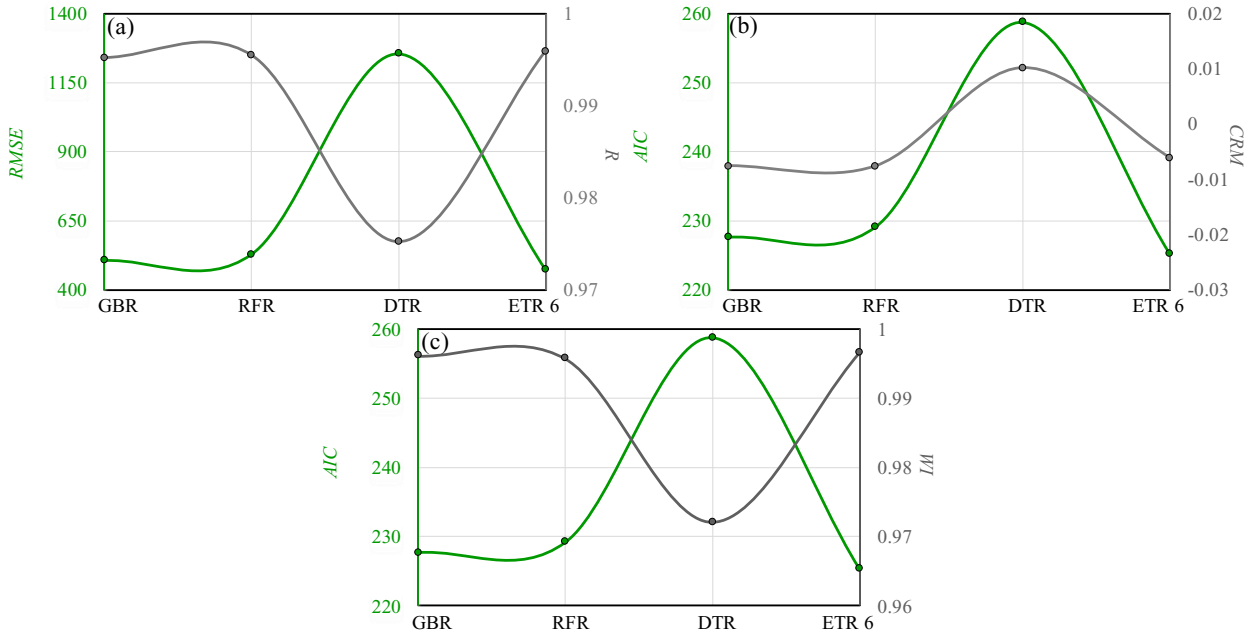


Figure 6-44. Comparison of the ETR 6 model in simulation of the vertical reaction forces with the DTR, RFR, and GBR methods (a) RMSE and R (b) AIC and CRM (c) AIC and WI

The key statistical indices computed for the horizontal subgouge deformations simulated by the ETR 6, DTR, RFR, and GBR models are illustrated in Figure 6-45. The made comparison proved that the ETR 6 had the highest level of agreement with the experimental values though the DTR algorithm produced the highest amount of inaccuracy, indicating that the value of the *AIC* index for the ETR 6, DTR, RFR, and GBR models was respectively equal to -56.264, 6.822, -43.925, and -2.385. Hence, the study found that the ETR 6 model surpassed other ML algorithms to forecast the horizontal subgouge deformations in terms of accuracy, correlation, and complexity.

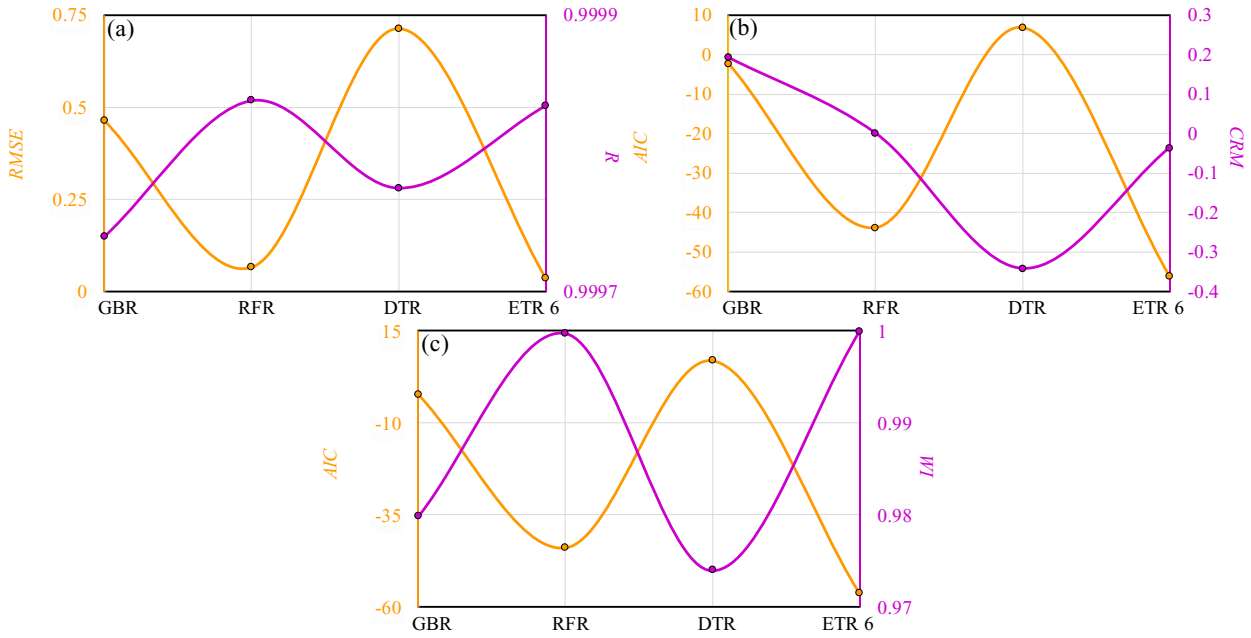


Figure 6-45. Comparison of the ETR 6 model in simulation of the horizontal subgouge deformations with the DTR, RFR, and GBR methods (a) RMSE and R (b) AIC and CRM (c) AIC and WI

The made comparison between the ETR 2, RFR, DTR, and GBR algorithms to estimate the vertical subgouge deformations is shown in Figure 6-46. The simulation results demonstrated that the ETR 2, GBR, RFR, and DTR models were ranked as the best to worst decision-based ML methods for

dealing with the simulation of the vertical deformations where the value of the AIC index for these models was computed to be -57.192, -49.221, -45.838, and -45.021, respectively. Thus, the findings indicate that the performance of the ETR 2 model exceeded other ML algorithms.

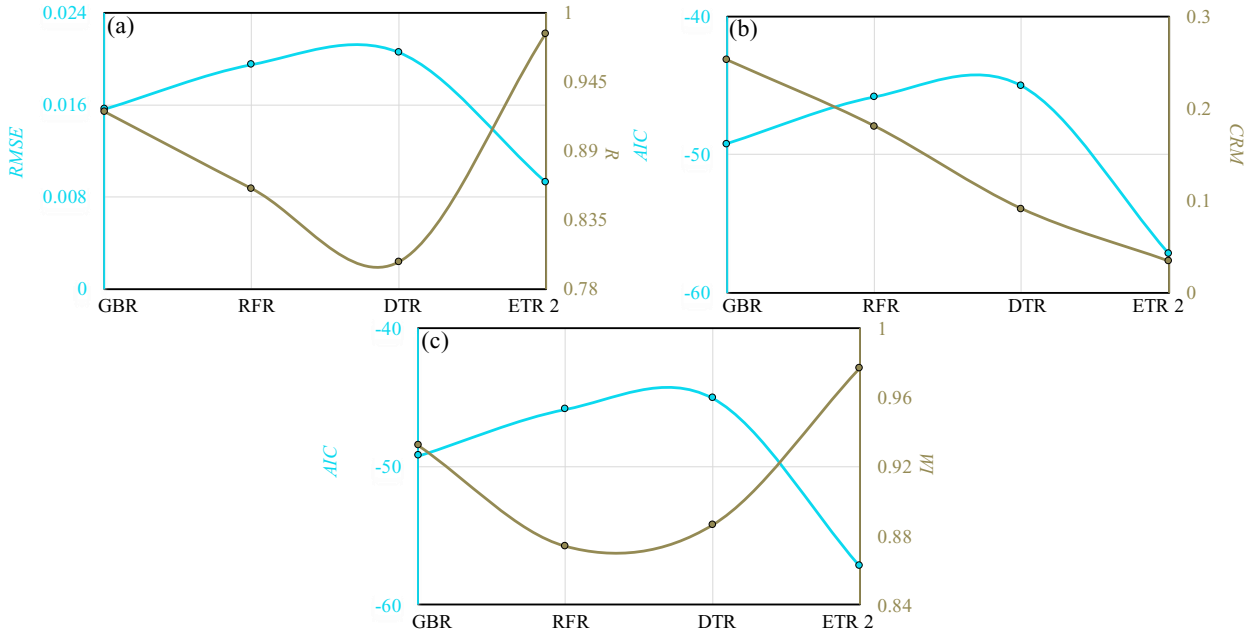


Figure 6-46. Comparison of the ETR 2 model in simulation of the vertical subgouge deformations with the DTR, RFR, and GBR methods (a) RMSE and R (b) AIC and CRM (c) AIC and WI

The error analysis for the superior ETR models along with the DTR, RFR, and GBR algorithms is presented in Figure 6-47. For instance, roughly half of the horizontal reaction forces simulated by the GBR algorithm had an error of less than 8%, whilst this amount for RFR, DTR, and ETR 4 was approximately 63%, 30%, and 66%. To predict the vertical reaction forces, almost 92% of the results obtained from the ETR 6 model possessed an error of less than 4%, although this value for the DTR, RFR, and GBR approaches was about 52%, 85%, and 88%. Nearly a third of the horizontal subgouge displacements modeled by the ETR 6 algorithm had an error of smaller than

8%, but close to 10% of the results from the GBR and DTR methodologies showed an error of less than 10%. The performed error analysis for simulation of the vertical deformations indicated that virtually one-fourth of the ETR 2 results held an error of slighter than 14%; however, for the RFR, GBR, and DTR methods, this number was respectively equal to 20%, 22%, and 6%. Hence, the error analysis presented that the premium ETR models could estimate the subgouge soil characteristics in clay seabed with acceptable accuracy.

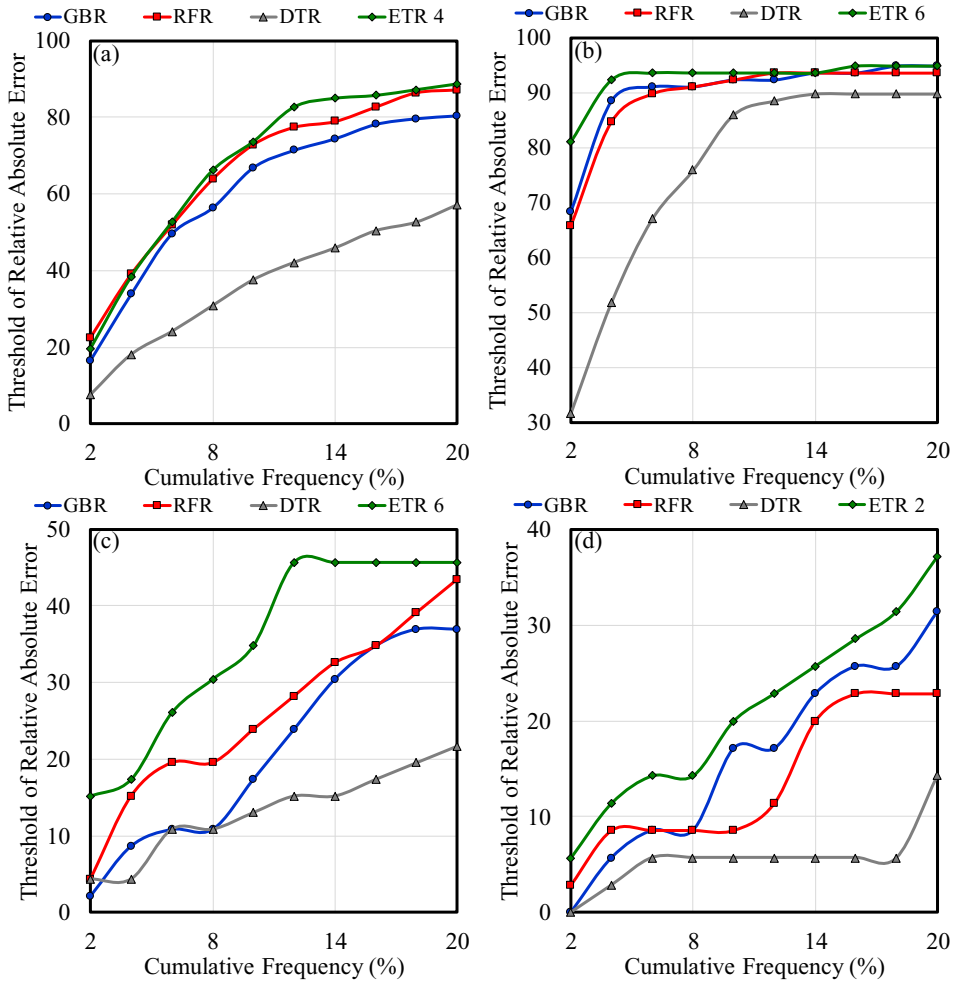


Figure 6-47. Error analysis for the superior ETR models along with the DTR, RFR, and GBR algorithms (a) horizontal reaction forces (b) vertical reaction forces (c) horizontal subgouge deformations (d) vertical subgouge deformations

The performance of the superior ETR models along with the GBR, RFR, and DTR algorithms was evaluated using the discrepancy ratio (*DR*) as below (Azimi and Shiri 2021a):

$$DR = \frac{P_i}{O_i} \tag{6-41}$$

here, P_i and O_i are the predicted and observed subgouge soil characteristics, respectively. Commonly, the closer the value of *DR* is to the unity, the better performance represents the ML algorithm. The predicted and observed ice-gouging parameters are compared using the DR index, which is the estimated value to the actual amount ratio. Additionally, the precision of the ML model is evaluated by the *DR*, where the significance of the DR criterion approaching the unity shows a better performance of the ML model (Azimi and Shiri 2021a).

Figure 6-48 depicts the *DR* computed for ice-seabed interaction parameters predicted by the ML models. As shown, the overwhelming majority of the simulated ice-induced parameters had a *DR* near to one. For example, the average discrepancy ratio (DR_{ave}) for the horizontal reaction forces estimated by the ETR 4 model was 1.063; rather this value for the GBR algorithm was appraised at 1.095. The maximum *DR* (DR_{max}) for the vertical reaction forces estimated by the GBR, RFR, DTR, and ETR 6 models were equal to 3.269, 3.349, 5.643, and 3.320, respectively. The minimum discrepancy ratio (DR_{min}) for the horizontal displacements predicted by ETR 6 and GBR models was 0.522 and 0.350. The DRmin for ETR 2 in modeling the vertical ice-induced deformations equaled 0.650. Therefore, the ETR algorithm showed good efficiency in predicting the ice-seabed interaction parameters.

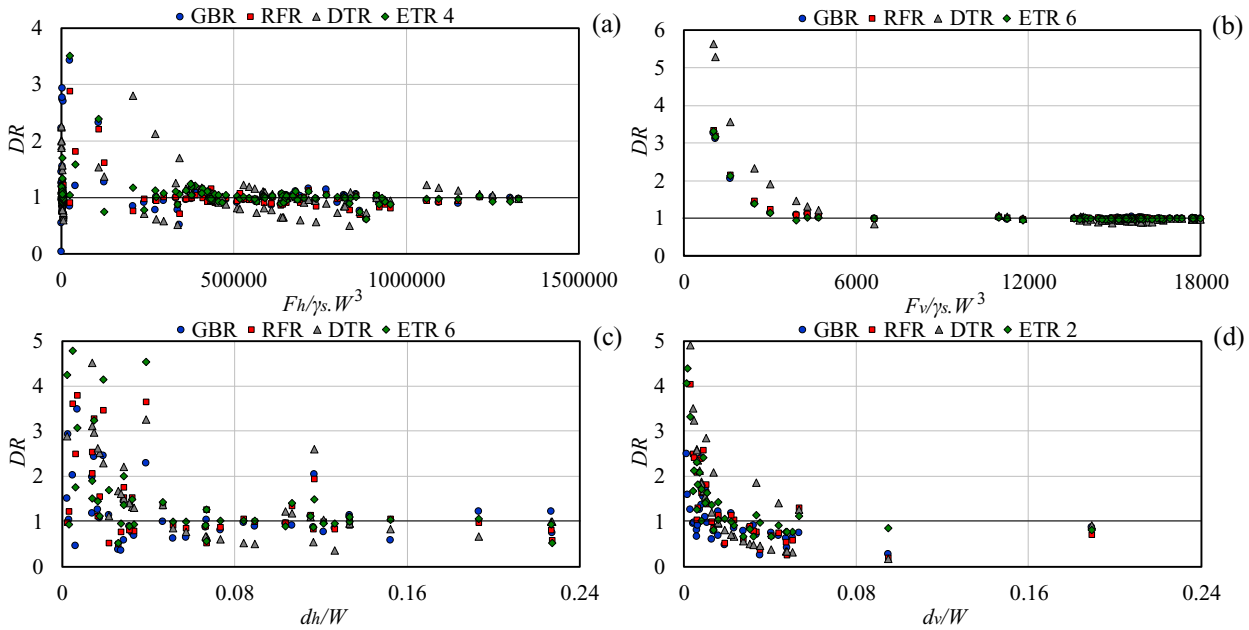


Figure 6-48. *DR* computed for ice-seabed interaction parameters (a) horizontal reaction forces (b) vertical reaction forces (c) horizontal subgouge deformations (d) vertical subgouge deformations predicted by the ML models

To assess the performance of the applied ML algorithms, an uncertainty analysis was conducted. To perform the uncertainty analysis, errors obtained from the ML models (e_j) were computed as the difference between the simulated ice-induced characteristics (P_j) and the observed values (O_j) as below (Azimi and Shiri 2021a):

$$e_j = P_j - O_j \tag{6-62}$$

The averaged calculated errors $(\bar{e})$ and the standard deviation (S_e) of errors were obtained by the following equations (Azimi and Shiri 2021a):

$$\bar{e} = \sum_{j=1}^{n} e_j \tag{6-63}$$

$$S_e = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{6-64}$$

The ML model had an underestimated performance if the sign of averaged obtained error ($\bar{e}$) was negative; however, the positive averaged calculated error indicated that the ML model overestimated the target parameters. Concerning the "Wilson score method" and disregarding the continuity correction, a confidence interval was generated near the computed error by means of the $\bar{e}$ and S_e values. A modified normal distribution interval as an asymmetric normal distribution, entitled the Wilson score interval, was applied to amend the confidence interval bounds. Then, a ±1.96Se created a 95% confidence interval (95%CI), and the width of uncertainty bound (WUB) was obtained as below (Azimi and Shiri 2021b):

$$\text{WUB} = \pm \frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{6-65}$$

The uncertainty analysis results from the ML algorithms are tabulated in Table 6-2. According to the uncertainty analysis, the GBR model overestimated the horizontal reaction forces; by contrast, the RFR, DTR, and ETR 4 models underestimated this parameter. The uncertainty analysis showed that the narrowest WUB was built for ETR models, as an example, the value of WUB for the vertical reaction forces predicted by the ETR 6 model equaled ±104.75. The vertical reaction forces were underestimated by using the GBR and ETR 6 models although the RFR and DTR algorithms overestimated this parameter. For the simulation of the horizontal deformations, the GBR and RFR models demonstrated an overestimated performance, but this parameter was underestimated through the DTR and ETR 6 models. All applied ML models in the current study performed an overestimated behavior.

Table 6-2. Uncertainty analysis parameters for ML algorithms

Parameter	Model	Mean ($\bar{e}$)	St.Dev (S_e)	WUB	95%CI
Fh	GBR	5532	58508	±10035.5	-4504 to 15567
	RFR	-7751	140164	±24041.5	-31793 to 16290
	DTR	-1125	114846	±19699	-20824 to 18574
	ETR 4	-373	54930	±9422	-9795 to 9049
Fv	GBR	-107.6	498	±111.55	-219.1 to 4
	RFR	108.1	521.2	±116.75	-224.8 to 8.7
	DTR	144	1255	±281	-137 to 425
	ETR 6	-86.4	467.7	±104.75	-191.1 to 18.4
dh	GBR	0.068	0.478	±0.142	-0.075 to 0.210
	RFR	0.001	0.069	±0.020	-0.020 to 0.021
	DTR	-0.119	0.735	±0.218	-0.337 to 0.099
	ETR 6	-0.012	0.036	±0.011	-0.023 to -0.001
dv	GBR	0.007	0.014	±0.005	0.002 to 0.012
	RFR	0.005	0.019	±0.007	-0.002 to 0.011
	DTR	0.002	0.021	±0.008	-0.005 to 0.010
	ETR 2	0.001	0.009	±0.003	-0.002 to 0.004

The uncertainty analysis showed that the narrowest WUB was built for ETR models, as an example, the value of WUB for the vertical reaction forces predicted by the ETR 6 model equaled ±104.75, similarly, the WUB for the ETR 2 model was ±0.003.

6.1.17.3. Comparison with experimental measurements

Figure 6-49 displays the comparison made between the horizontal reaction forces simulated by the ETR 4 model and the experimental measurements. As shown, the superior ETR model managed to approximate the reaction forces with its highest level of accuracy, correlation, and lowest degree of complexity. Regarding the simulation results, the minimum value of the horizontal reaction forces was modeled at the initial position of scouring; after that, the magnitude of this parameter grew along the scour axis. The ETR 4 model utilized a non-linear trend to predict this ice-gouging parameter (C-1, C-2, C-3, and C-8). Though there were several fluctuations in the experimental tests, the ETR 4 model was able to simulate the horizontal reaction forces with acceptable performance. Moreover, some discrepancies between the simulation outcomes and test results were observed (C-7, L-1, L-2, L-4, and L-5).

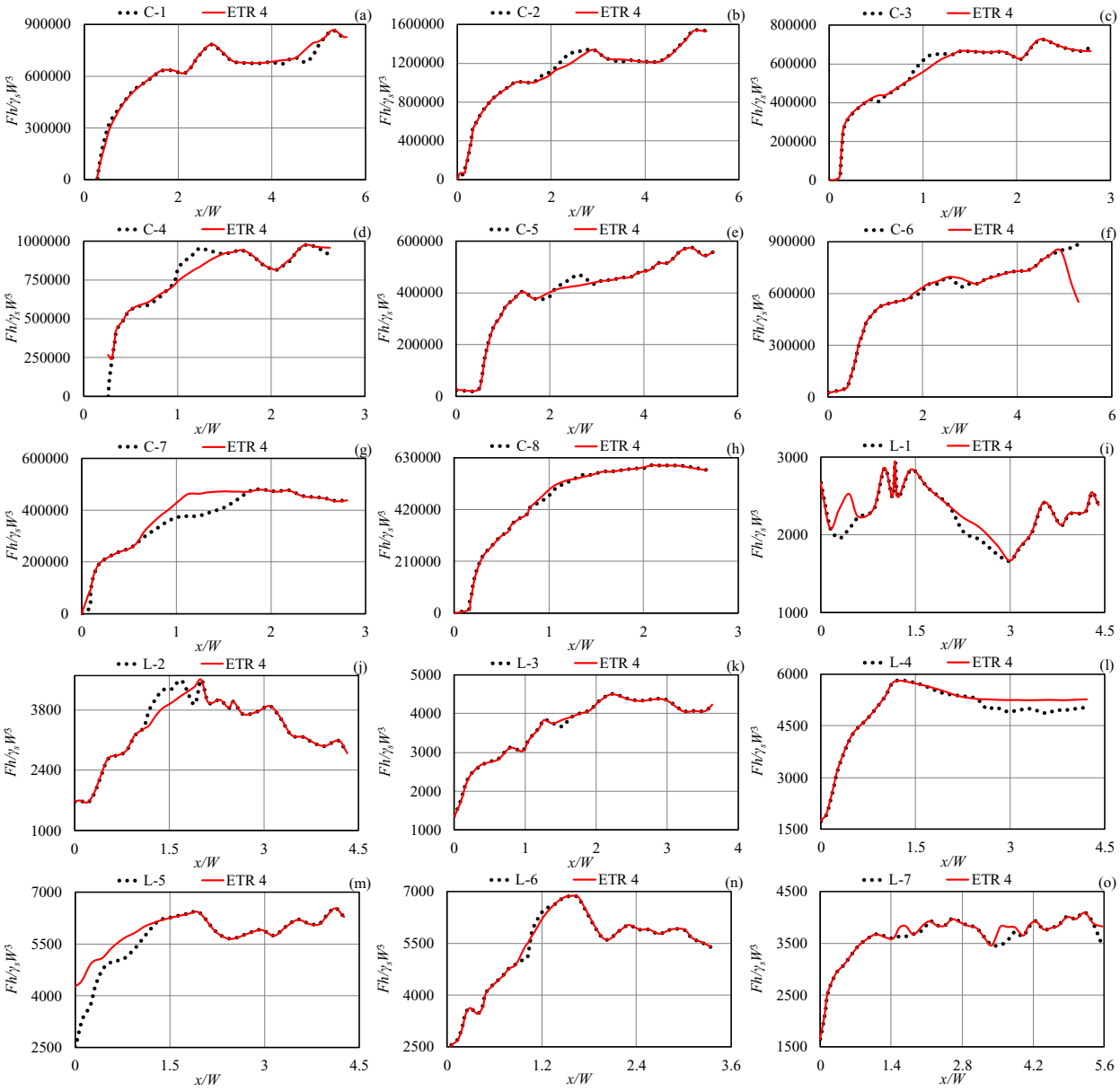


Figure 6-49. Comparison between the horizontal reaction forces simulated by the ETR 4 model and the experimental measurements (a-h) C-1 to C-8 (i-o) L-1 to L-7

The results of the ETR 6 model and the vertical reaction forces are compared in Figure 6-50. Similar to the ETR 4 model, the smallest amount of the vertical reaction forces was estimated in the initial location of the scouring and it reached the maximum amount by developing along the scour axis. ETR 6 modeled the L-2, L-4, and L-5 cases with a strong correlation; however, in the

L-1, L-3, and L-6 tests, several disagreements were reported between the test and the numerical results. Despite the oscillatory trend in the L-7 and L-8 cases, the ETR 6 model could appropriately predict the alteration pattern of the vertical reaction forces.

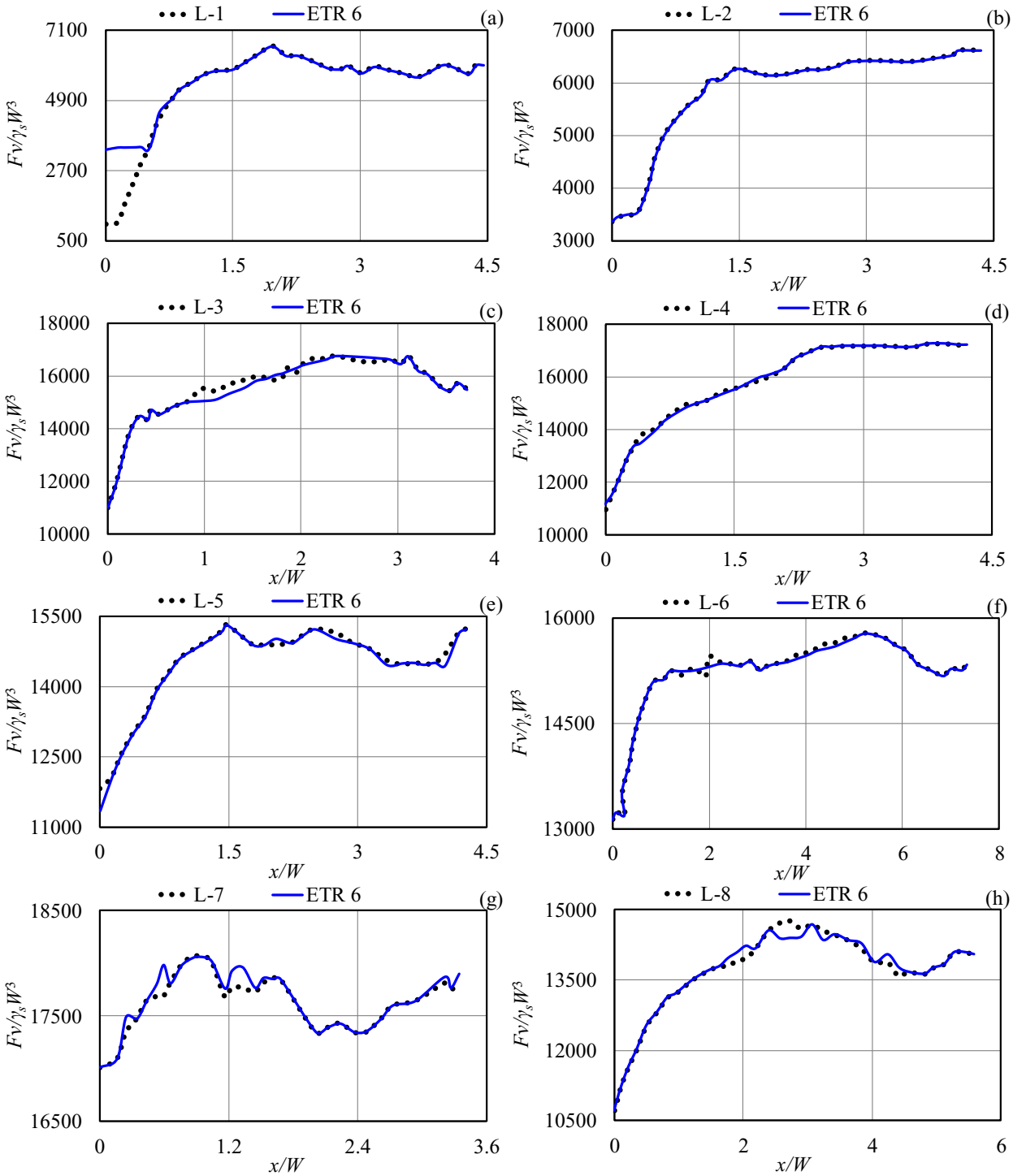


Figure 6-50. Comparison between the vertical reaction forces simulated by the ETR 6 model and the experimental measurements (a-h) L-1 to L-8

Figure 6-51 demonstrates the comparison between the horizontal ice-induced soil deformations estimated by ETR 6 and the test results. As can be seen, the maximum subgouge displacement was recorded beneath the ice keel body, rather the magnitude of this parameter was reduced by an exponential distribution at greater soil depth. Apart from some slight discordances between the numerical and experimental results (C-4, C-8, L-1, L-2, L-3, and L-8 to L-12), the ETR 6 model approximated the horizontal displacements with reasonable exactness. This comparison indicated that the ETR 6 model estimated the C-1, C-2, C-3, C-7, L-5, L-7, S-1, and B-2 tests with a strong correlation.

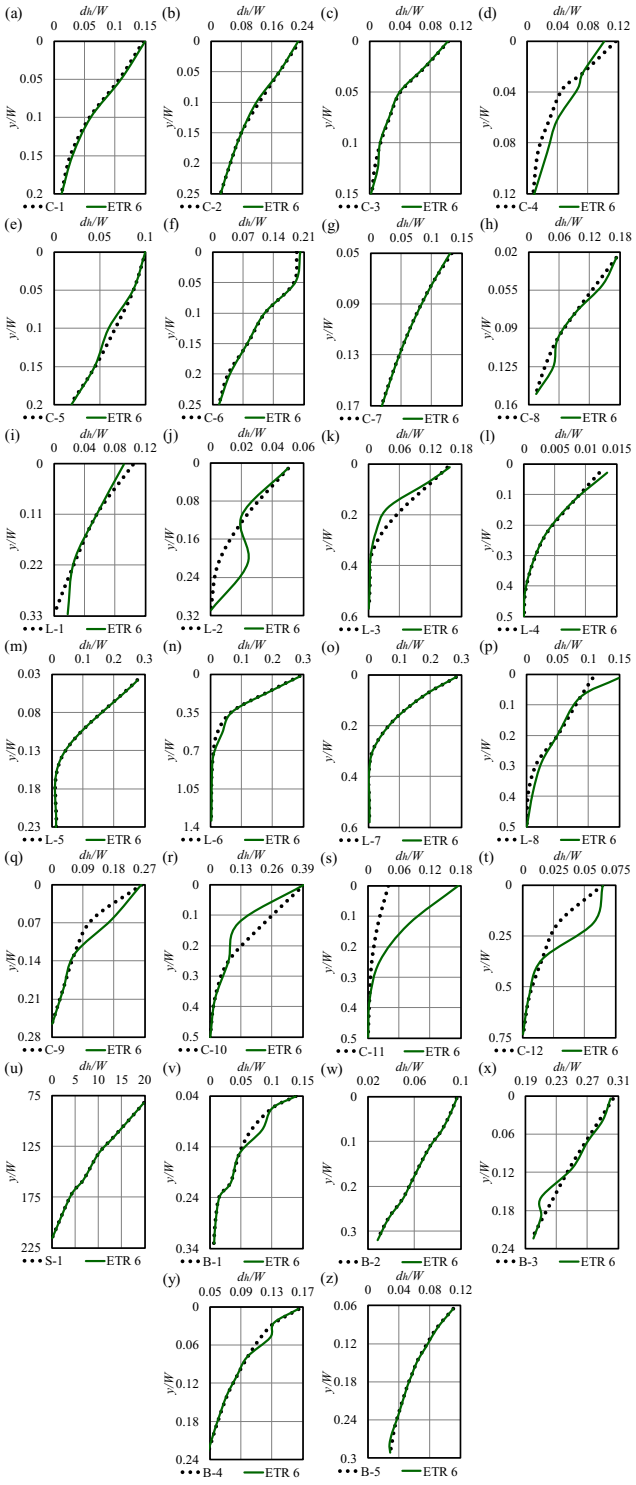


Figure 6-51. Comparison between the horizontal subgouge deformations simulated by the ETR 6 model and the experimental measurements (a-h) C-1 to C-8 (i-p) L-1 to L-8 (q-t) C-9 to C-12 (u) S-1 (v-z) B-1 to B-5

The comparison between the vertical subgouge deformations simulated by the ETR 2 model and the test measurements is drawn in Figure 6-52. A remarkable correlation was found between the test and simulation results (C-1 to C-3, C-7, L-1, L-3, and C-8), but several differences occurred in the C-4, L-4 to L-8, and B-1 tests. The ETR 2 model could simulate the C-5 and C-6 tests by using a linear trend; whilst the C-8 to C-11 tests were modeled regarding a non-linear pattern.

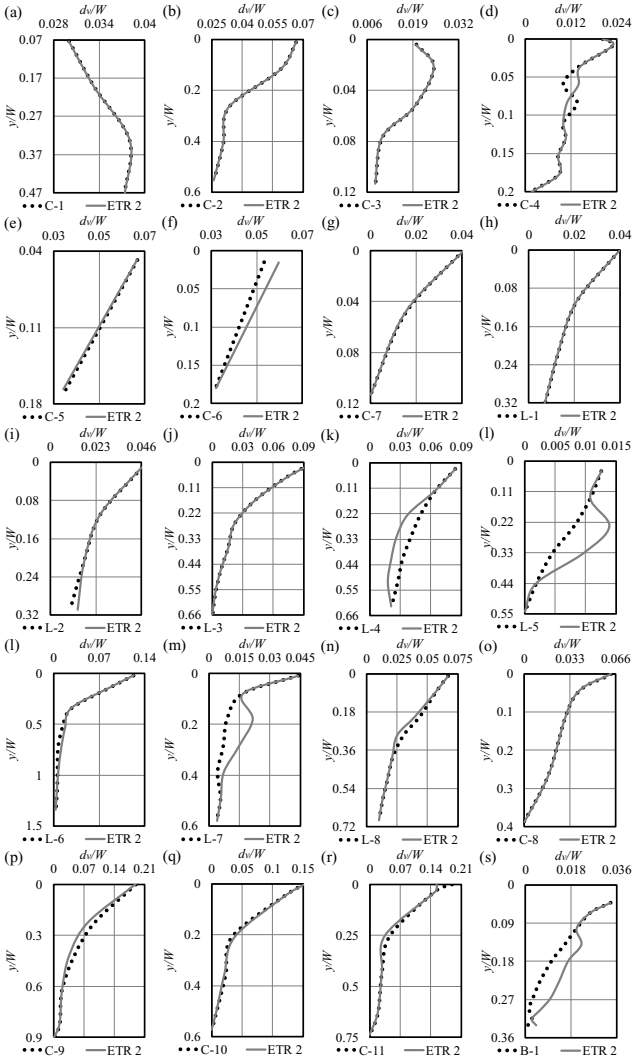


Figure 6-52. Comparison between the vertical subgouge deformations simulated by the ETR 2 model and the experimental measurements (a-g) C-1 to C-7 (h-n) L-1 to L-8 (o-r) C-8 to C-11 (s) B-1

The precision of experimental values has originated from two major sources containing systematic error (equipment error) and random error (human error). The source of the systematic error may be identified and the value of this error is measurable, while identification of a random error source is quite arduous. To hinder the random error, the experimental measurements had been repeated and their average values were reported as the laboratory records. It seems that the experimental values lacked a meaningful trend and there were several oscillations in the experimental measurements (Figure 14-14-i, Figure 14-17-c, and 14-17-d); however, the superior ETR model was able to appropriately estimate the vertical sub-gauge displacements with the linear and the nonlinear trends.

Therefore, the ETR model estimated the ice-induced soil characteristics, including the reaction forces and the subgouge soil deformations in clay seabed with sufficient precision, a high level of correlation, a low degree of discrepancy, and a high amount of simplicity.

6.1.18. Conclusion

In this study, the ETR algorithm was used for the simulation of the iceberg-seabed interaction parameters comprising the reaction forces and the subgouge soil deformations in clay. Regarding the variables governing the ice-seabed interaction, eight ETR models (ETR 1 to ETR 8) were defined. Using the test studies published in the literature, a comprehensive dataset was established, where 70% of the data was allocated to train the ML models, and the trained models were tested using the remaining 30%. The superior ETR models and the most influencing input parameters were introduced by performing a sensitivity analysis. Subsequently, the performance of premium ETR models was compared with three ML algorithms, including the DTR, RFR, and GBR models.

A set of analyses, e.g., error analysis, discrepancy analysis, and uncertainty analysis, were implemented for a detailed evaluation of the applied ML models. The most important outcomes of the current investigation are summarized below:

- The simulation results proved that the ETR model could predict the ice-seabed interaction features with the highest level of accuracy and correlation along with the lowest level of complexity.
- ETR 4, a function of the x/W, Ds/W, φ, α, h'/W, $Lv/\gamma s.W^3$, and $V^2/g.W$ inputs, was identified as the best model to approximate the horizontal reaction forces, while ETR 6 a function of x/W, Ds/W, φ, h'/W, $Lh/\gamma s.W^3$, $Lv/\gamma s.W^3$, and $V^2/g.W$ was the best ML model for estimation of the vertical reaction forces.
- To model the horizontal and vertical ice-induced soil displacements in clay, ETR 6 and ETR 2 were respectively recognized as the best ML models.
- The attack angle (α) and the gouge depth ratio (Ds/W) had a striking effect on the simulation of the subgouge reaction forces; however, the soil depth ratio (y/W), the horizontal component of the load ($Lh/\gamma s.W^3$), and the attack angle (α) were detected as the most significant input factors to predict the ice-induced soil deformations in clay seabed.
- The comparison between the ETR model with the DTR, RFR, and DTR algorithms represented that the ETR method outperformed the counterparts.
- Error analysis showed that almost 82% of the results from the ETR 4 model possessed an error of less than 12%, but this amount for GBR, RFR, and DTR were about 71%, 77%, and 42%, respectively.

From a practical design point of view, this investigation demonstrated that the ETR algorithm could simply simulate the complex and non-linear ice-scouring problem with acceptable precision

and low expenses. The proposed model may be a perfect alternative to the early stages of pipeline design projects in the Arctic area, even though a quick and accurate estimate is still needed to decide on the methodology, logistics, and planning of the accurate engineering area. Further studies with more focus on advanced ML methods, Neuro-Fuzzy models, and optimization algorithms are therefore suggested in future works.

Funding and Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Azimi, H., Shiri, H., 2020. Dimensionless groups of parameters governing the ice-seabed interaction process, Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2021a. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., Malta, E.R., 2021a. A Non-Tuned Machine Learning Method to Simulate Ice-Seabed Interaction Process in Clay. Journal of Pipeline Science and Engineering, 1(4): 379-394.

Azimi, H., Shiri, H., Zendehboudi, S., 2021b. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology: 103426.

Azimi, H., Shiri, H., 2021b. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Barrette, P., Sudom, D., Babaei, H., 2015. Protecting offshore pipelines against drifting ice: A discussion on standards and guidelines. International Conference on Offshore Mechanics and Arctic Engineering. American Society of Mechanical Engineers, Vol. 56567, p. V008T07A028.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A., 2008. Subscour displacement in clays from physical model tests. In International Pipeline Conference, Vol. 48609: pp. 239-245.

Breiman, L., 2001, Random forests, Machine learning, 45(1): 5-32.

C-CORE, 1995, Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2, St. John’s, NL, Canada.

C-CORE, 1996, PRISE Phase 3c: Extreme LEE Gouge Event—Modeling and Interpretation, C-CORE Report 96-C32, St. John's, NL, Canada.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning, 63(1): 3-42.

Hammed, M.M., AlOmar, M.K., Khaleel, F., Al-Ansari, N., 2021. An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. Mathematical Problems in Engineering.

Hastie, T, Tibshirani, R, Friedman, J., 2009. Boosting and additive trees. In The elements of statistical learning, Springer, New York, pp. 337-387.

ISO. ISO 19906, 2019. Petroleum and Natural Gas Industries – Arctic Offshore Structures. ISO, Geneva.

Kioka, S.D, Kubouchi, A, Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. In the 13th International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S, Kubouchi, A, Ishikawa, R, Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. In the 14th International Offshore and Polar Engineering Conference, ISOPE-I-04-107.

Lach, P.R., 1996. Centrifuge modelling of large soil deformation due to ice scour. Doctoral dissertation, Memorial University of Newfoundland. St. John's. NL. Canada.

Pekel, E., 2020. Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3): 1111-1119.

Sahani, N. Ghosh, T., 2021. GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. Ecological Informatics, 64: 101352.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip line field solutions as an approach to understand ice subgouge deformation patterns. In the 16th International Offshore and Polar Engineering Conference, ISOPE-I-06-289.

Zvyagin, P., Heinonen, J., 2017. In the 36th International Conference on Ocean, Offshore and Arctic Engineering. International Conference on Offshore Mechanics and Arctic Engineering. Vol. 57762: p. V008T07A010.

Part II

Prediction of Iceberg Draft

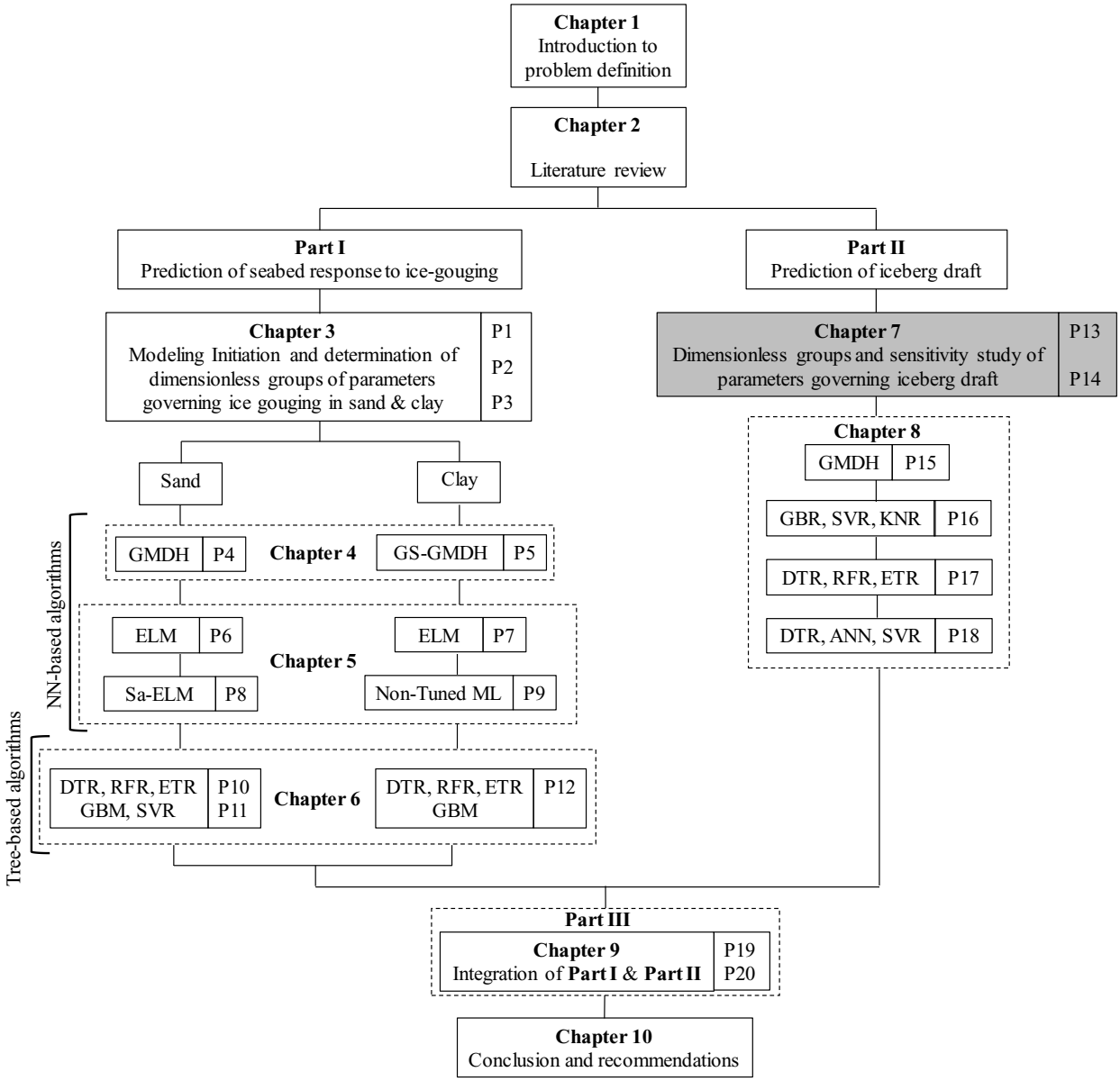
7. Chapter 7

Dimensionless Groups and Sensitivity Study of Parameters Governing Iceberg Draft

This chapter includes two sections as follows:

Section 1: Determination of Parameters Affecting the Estimation of Iceberg Draft

Section 1: Sensitivity Analysis of Parameters Governing the Iceberg Draft through Neural Network-Based Models


Chapter 1
Introduction to problem definition
Chapter 2
Literature review
Part I
Prediction of seabed response to ice-gouging
Part II
Prediction of iceberg draft
Chapter 3
Modeling Initiation and determination of dimensionless groups of parameters governing ice gouging in sand & clay
P1
P2
P3
Chapter 7
Dimensionless groups and sensitivity study of parameters governing iceberg draft
P13
P14
Chapter 8
GMDH P15
GBR, SVR, KNR P16
DTR, RFR, ETR P17
DTR, ANN, SVR P18
Sand
Clay
NN-based algorithms
GMDH P4
Chapter 4
GS-GMDH P5
ELM P6
ELM P7
Chapter 5
Sa-ELM P8
Non-Tuned ML P9
Tree-based algorithms
DTR, RFR, ETR GBM, SVR P10 P11
Chapter 6
DTR, RFR, ETR GBM P12
Part III
Chapter 9
Integration of Part I & Part II
P19
P20
Chapter 10
Conclusion and recommendations

Section 1

Determination of Parameters Affecting the Estimation of Iceberg Draft

This section is a journal paper which has been published in China Ocean Engineering (2023),

Vol. 37, pp. 62–72, ISSN: 0890-5487

Abstract

Recent offshore oil and gas loading facilities developed in the Arctic area have led to a considerable awareness of the iceberg draft approximation, where deep keel icebergs may gouge the ocean floor, and these submarine infrastructures would be damaged in the shallower waters. Developing reliable solutions to estimate the iceberg draft requires a profound understanding of the problem's dominant parameters. As such, the dimensionless groups of the parameters affecting the iceberg draft estimation were determined for the first time in the present study. Using the dimensionless groups recognized and the linear regression (LR) analysis, nine LR models (i.e., LR 1 to LR 9) were developed and then validated using a comprehensive dataset, which has been constructed in this study. A sensitivity analysis distinguished the premium LR models and important dimensionless groups. The best LR model, as a function of all dimensionless parameters, was able to estimate the iceberg draft with the highest level of precision and correlation along with the lowest degree of complexity. The ratio of iceberg length to iceberg height as the "iceberg length ratio" and the ratio of iceberg width to iceberg height as the "iceberg width ratio" was detected as the important dimensionless groups in the estimation of the iceberg draft. An uncertainty analysis demonstrated that the best LR model was biased towards underestimating the iceberg drafts. The premium LR model outperformed the previous empirical models. Ultimately, a set of LR-based relationships were derived for estimating the iceberg drafts for practical engineering applications, e.g., the early stages of the iceberg management projects.

Keywords: Dimensional analysis, Iceberg draft, Subsea assets, Linear regression (LR), Sensitivity analysi

7.1.1. Introduction

Deep draft icebergs may scour the seafloor in the Arctic and shallow subarctic waters, posing a threat to the operational integrity of the submarine assets, such as subsea pipelines, wellheads, offshore oil loading equipment, and communication cables. As shown in Figure 7-1, if the iceberg draft is greater than the water depth, the seabed may be gouged; otherwise, the iceberg keel is considered in the free-floating condition. The estimation of the iceberg draft is a challenging issue to warn of the collision risk of the iceberg keel with the subsea infrastructures. Therefore, determining the parameters affecting the iceberg draft estimation may be crucial in protecting the subsea structures.

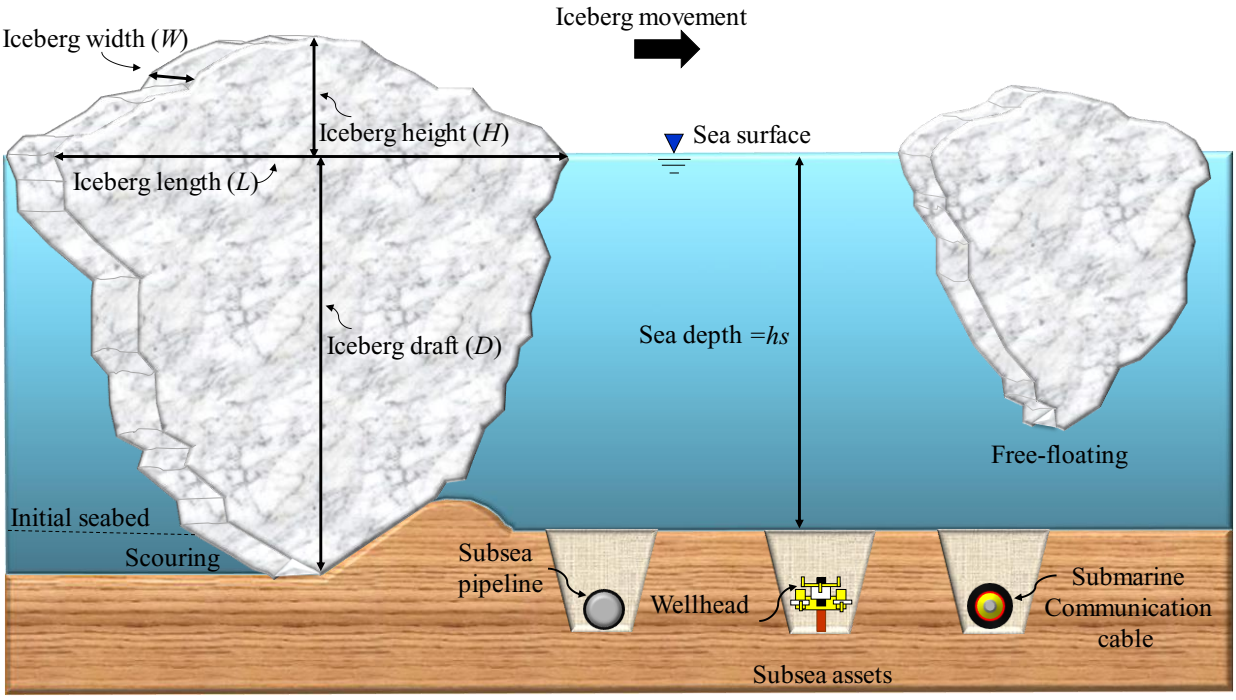


Figure 7-1. Schematic layout of the iceberg free-floating and iceberg scouring in the Arctic waters

The precise iceberg draft estimation is essential for improving the ice management projects, diminishing the collision risk of the iceberg keel with the submarine assets, and deducting the expenses and downtime.

Papusha et al. (2013) reported many impaired subsea infrastructures by traveling icebergs in the Arctic shallow waters in the Labrador Sea, Great Slave Lake, and Lake Erie. "Grand Banks Scour Catalogue (GBSC) GeoDatabase," as one of the comprehensive investigations of the ice collision events on the Grand Banks of Canada, documented that multi icebergs drifted and gouged the ocean floor across the shallow water area of the Grand Banks, where scour marks reported in the water depth of up to 400 m (Campbell 2014). Furthermore, Minerals Management Service (2002) recorded 529 out of 836 gouges (roughly 63%) monitored from 1995 to 1998 were marked in the shallower waters. Additionally, several studies have sought to determine the iceberg draft, where preceding iceberg draft models have correlated the height of the underwater iceberg portion (D) with the iceberg length (L) or iceberg mass (M). For instance, Robe and Farmer (1976) reported the features of tabular, broken tabular, pinnacle, dry dock, and domed icebergs. They asserted that the iceberg draft (D) correlated with the iceberg height (H). In particular, the authors proposed a relationship for the estimation of the iceberg draft as follows:

$$\frac{D}{H} = 49.4H^{-0.8} \tag{7-1}$$

Mognor and Zorn (1979) investigated the iceberg dimensions and showed that iceberg stability was a function of its draft and width (w). The authors suggested an equation to surmise the iceberg draft:

$$D=1.2w+50 \tag{7-2}$$

El-Tahan and El-Tahan (1982) used the iceberg characteristics, including the height and draft of different bergs. The study highlighted that the domed and tabular icebergs had more remarkable

drafts than the dry dock and pinnacle icebergs. The authors introduced the following relationships to calculate the iceberg drafts:

$$\frac{D}{H} = 111.04\left(H\right)^{-1.111} \quad \text{For domed icebergs} \tag{7-3}$$

$$\frac{D}{H} = 198\left(H\right)^{-1.235} \quad \text{For Tabular icebergs} \tag{7-4}$$

Hotzel and Miller (1983) analyzed the iceberg dimensions and assumed that the planimetric section of icebergs was circular, while the length of berg (L) was twice the circle radius. The authors derived a formula for the approximation of iceberg draft through the power curve method as below:

$$D = 3.781\left(L^{0.63}\right) \tag{7-5}$$

C-CORE (2001a) and C-CORE (2001b) reported several iceberg management programs to measure iceberg characteristics. In these studies, a set of relationships were extracted from the field measurements analyzed for approximation of the iceberg draft as follows:

$$D = 2.1\left(M^{0.28}\right) \tag{7-6}$$

$$D = 3.239\left(L^{0.68}\right) \tag{7-7}$$

Where M is the iceberg mass. Barker, Sayed, and Carrieres (2004) assessed the geometry of iceberg sails and keels. The authors calculated the cross-sectional areas of the iceberg at various water depth intervals from a particular waterline length. They presented some empirical equations for the estimation of the iceberg draft in terms of the iceberg length by using the power curve and regression analysis:

$$D = 2.91\left(L^{0.71}\right) \tag{7-8}$$

$$D = 0.7\left(L\right) \tag{7-9}$$

Sacchetti et al. (2012) studied the berg characteristics and ice scouring in Labrador and Hibernia regions. The features of different icebergs, such as wedged, domed, tabular, and pinnacle bergs, were evaluated in this study. The authors offered a set of equations in terms of the iceberg length to appraise the iceberg draft in the following forms:

$$D = 3.9\left(L^{0.63}\right) \quad \text{for Labrador region} \tag{7-10}$$

$$D = 1.95\left(L^{0.79}\right) \quad \text{for Hibernia region} \tag{7-11}$$

King et al. (2016) conducted a field investigation to estimate the rolling rate of the iceberg. The iceberg draft was modeled through a calving analysis, with a calculated standard deviation of draft variations from 19% to 34%. The iceberg draft has corresponded with the iceberg mass as follows:

$$D = 2.05\left(M^{0.276}\right) \tag{7-12}$$

Most recently, Stuckey et al. (2021) simulated the three-dimensional iceberg shapes by adopting field measurements. The investigation demonstrated that the iceberg draft was computed in terms of iceberg length by utilizing the power curve. The authors derived two empirical models regarding the data collected in 2016 and the post-2000 data as below:

$$D = 4.78\left(L^{0.58}\right) \quad \text{data collected in 2016} \tag{7-13}$$

$$D = 1.87\left(L^{0.72}\right) \quad \text{post-2000 data} \tag{7-14}$$

Although several empirical models for estimating the iceberg draft have been proposed, determining a dimensionless group of parameters for rapid simulation of the iceberg draft can be used in the early stages of engineering projects to plan construction methodologies, logistics, and the scope of detailed engineering.

This study recognized the dimensionless group of parameters affecting the iceberg draft estimation through the Buckingham-π theorem for the first time. A set of LR models were developed using

the dimensionless groups and linear regression (LR) approach. The best LR models and the most influencing dimensionless groups were ascertained by conducting several analyses.

7.1.2. Methodology

The flowchart of the methodology applied in the current study is drawn in Figure 7-2. In the proposed method, dimensional analysis of the iceberg draft is performed, and the dimensionless groups affecting the iceberg draft estimation are introduced through the Buckingham-π theorem. Several linear regression (LR) models are proposed to estimate the iceberg draft. Using the characteristics of the real icebergs reported in the literature, a comprehensive dataset was created, which has been used for validating the accuracy, correlation, and complexity of the LR models. The study compared the performance of the LR models with the previous models in the literature. The premium LR models and the most influencing input parameters were recognized by performing several analyses.

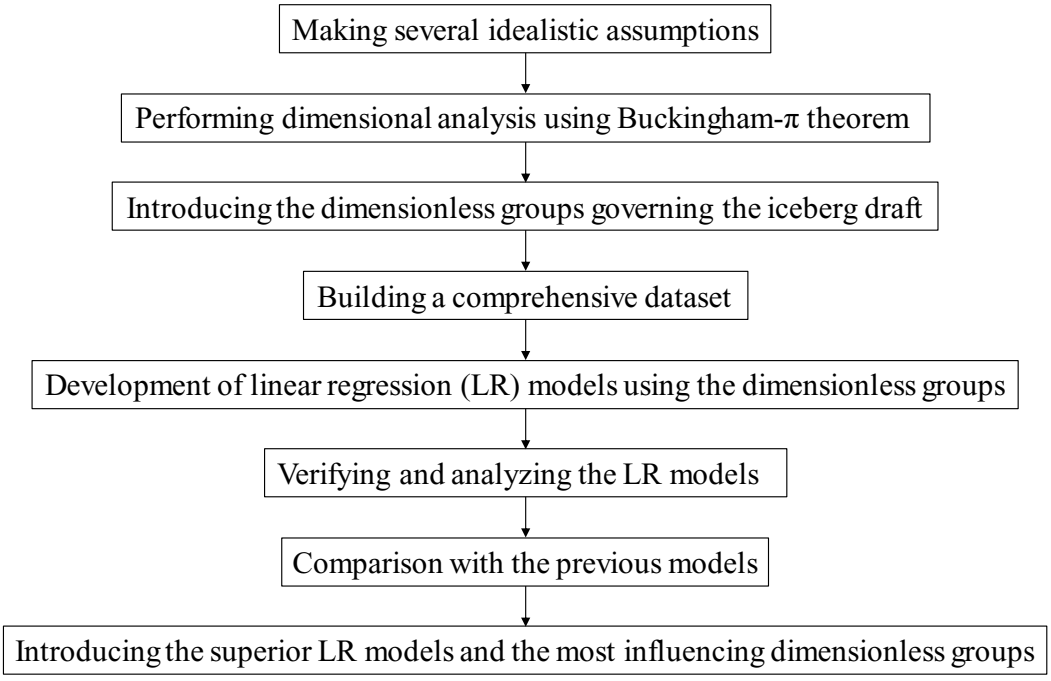


Figure 7-2. Flowchart of the proposed method employed in this study

7.1.2.1. Dimensional analysis

Dimensional analysis is a type of analysis in which correlations between significant quantities, such as mass, time, and length, are assessed in terms of their dimensions. Dimensional analysis is a globally valid method used in various engineering designs, applied mathematics, and physical challenges. This analysis represents that quantities involving a specific equation are dimensionally homogeneous. By using the analysis, scholars and engineers can gain insight into measuring and scaling different quantitative parameters to provide an elaborate design. The dimensional analysis is utilized as a relation that can be written between the target function as an output parameter (δ_1) and input parameters (δ_2, δ_3,…, δ_n) , as below:

$$\delta_1 = f\left(\delta_2, \delta_3, \ldots, \delta_n\right) \tag{7-15}$$

Accordingly,

$$f\left(\delta_1, \delta_2, \delta_3, \ldots, \delta_n\right) = 0 \tag{7-16}$$

Where *f* is "some function of"' and *n* is defined as the number of input parameters describing the physical problem (Massey 1971; Simon, Weigand, and Gomaa 2017). Thus, the appropriate primary output and input parameters should be chosen, meaning that just one relation between input variables and output parameters may exist. Subsequently, several dimensional groups $(\pi_1, \pi_2, \pi_3, \ldots, \pi_{n-m'})$ are defined regarding the primary parameters; hence, the new form of the relationship is written as follows:

$$f(\pi_1, \pi_2, \pi_3, \ldots, \pi_{n-m'}) = 0 \tag{7-17}$$

Where "m'" is the number of the principal quantities, and "n-m'" is the number of dimensionless groups. In other words, if a physical problem involves a certain number "n" of independent input parameters, which have "m'" main quantities, then "n-m'" dimensional groups can be developed.

The procedure mentioned is universally considered the Buckingham-π theorem to introduce the dimensionless groups in sophisticated and complex physical problems (Massey 1971; Simon Simon, Weigand, and Gomaa 2017).

To conduct a robust dimensional analysis of the iceberg draft estimation, several simplifying assumptions were made below to assist the implementation of the methodology developed in the practical applications:

- The iceberg is in free-floating condition;
- The density of seawater is constant;
- The temperature of seawater is constant, and;
- The viscosity of seawater is constant.

Even though determining the parameters to consider in the dimensional analysis is one of the most critical aspects, there is no specific regulation or rule for selecting those parameters. Notwithstanding this, thoroughly comprehending the problem can significantly help the issue (Stuckey 2008). Hence, the parameters involved in the dimensional analysis of the iceberg draft estimation can be categorized into three distinct types:

(i) The geometric characteristics of the iceberg, e.g., the above-water parameters and shape of the iceberg;

(ii) The physical characteristics of the iceberg, e.g., mass and density of the iceberg;

(iii) The physical characteristics of the ocean water, e.g., the density and viscosity of ocean water.

Although some of the parameters mentioned above might not be completely independent, the parameters related to the geometric characteristics of the iceberg, physical characteristics of the iceberg, and physical characteristics of the ocean water could be used to determine the

dimensionless groups of iceberg properties, e.g., iceberg drift speed estimation (Stuckey 2008). Furthermore, in dimensional analysis, a dependent variable is generally estimated in terms of independent variables. A dimensionless expression of the acquired relations then involves a dependent parameter, including the dependent variable and other independent parameters comprising the independent variables. Several independent variables happen in both the dependent and independent parameters. The existence of standard variables in the dependent and independent parameters may produce a statistical correlation that cannot be seen between the dependent and independent variables of the problem (Mahmood and Siddiqui 1980).

The iceberg draft (D) was considered as a function of the physical characteristics of the iceberg, including the iceberg length (L), iceberg height (H), iceberg width (w), and iceberg mass (M) in several fields, analytical, and numerical studies as follows (Barker, Sayed, and Carrieres 2004; McKenna et al. 2019; and Stuckey et al. 2021):

$$D = f_1(L, H, w, M) \tag{7-18}$$

Moreover, the density of an iceberg (ρ_i), the density of seawater (ρ_{sw}), seawater viscosity (μ_{sw}), and gravitational acceleration (g) can affect the iceberg draft as below:

$$D = f_2(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g) \tag{7-19}$$

The iceberg shape factor (S_f) represents the global shape of icebergs, which can affect the magnitude of the iceberg draft (Turnbull, King, and Ralph 2018). The shape factor of the traveling icebergs is placed into six categories, as demonstrated in Figure 7-3 (Rudkin 2005). Thus, equations (7-19) can be written in the following form:

$$D = f_3\left(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g, S_f\right) \tag{7-20}$$

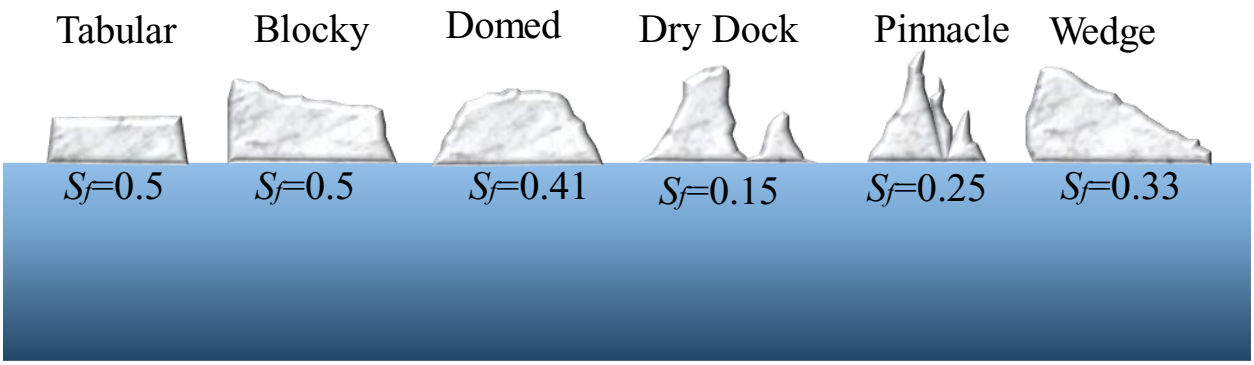


Figure 7-3. Shape factors for different icebergs

Assuming that the density, viscosity of the seawater, and gravitational acceleration are constant parameters, hence equation (7-20) is rewritten as below:

$$D = f_4\left(L, H, w, M, \rho_i, S_f\right) \tag{7-21}$$

The dimensional form of equation (7-21) is formed as follows:

$$D = f_5\left(\Pi_1, \Pi_2, \Pi_3, \Pi_4\right) \tag{7-22}$$

where Π_1, Π_2,…, Π_4 are dimensionless groups and f_5 is a functional symbol based on the Buckingham-π theorem. Hence, the dimensionless groups below are written:

$$\Pi_1 = \frac{L}{H} \tag{7-23}$$

$$\Pi_2 = \frac{w}{H} \tag{7-24}$$

$$\Pi_3 = \frac{M}{\rho_i . H^3} \tag{7-25}$$

$$\Pi_4 = S_f \tag{7-26}$$

Equation (7-21) is then formulated as a function of four dimensionless groups as follows:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \tag{7-27}$$

Therefore, the iceberg draft ratio, *D/H*, is a function of the length ratio (*L/H*), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor (S_f).

7.1.2.2. Construction of Dataset

The study adopted several field observations to analyze the iceberg draft. The key values of 12 field studies reported by El-Tahan et al. (1985) (38 cases), Woodworth-Lynas et al. (1985) (one case), Løset and Carstens (1996) (52 cases), Barker, Sayed, and Carrieres (2004) (14 cases), McKenna (2004) (two cases), Sonnichsen et al. (2006) (nine cases), Turnbull et al. (2015) (two cases), McGuire et al. (2016) (eight cases), Younan et al. (2016) (29 cases), Talimi et al. (2016) (one case), Zhou (2017) (three cases), and Turnbull, King, and Ralph (2018) (two cases) were used. The total number of field observations in the present investigation was 161 cases. Table 7-1 summarizes the current study's key parameters, comprising the field observations' maximum, minimum, average, variance, and standard deviation.

To demonstrate the significant difference between the means of the individual datasets, the T-test and the P-value for the datasets were calculated, presuming that the P-value of 0.05 or less is statistically significant (Azimi and Shiri 2020), where an alternative hypothesis influences the likelihood of the relationship between the observed values. This P-value for the constructed dataset was estimated as 0.008, representing that the correlations were statistically significant.

Table 7-1. Summary of the key parameters employed in the present study

Value	*D*(m)	*H*(m)	*w*(m)	*M*(kg)	*L*(m)	S_f
Max	240	94	408	1.13E+13	499	0.58
Min	18	3.9	9.672	8317400	34	0.11

Ave	88.988	30.956	101.605	1.68E+11	144.522	0.298
Var.	1696.272	399.389	3733.368	1.14E+24	6546.271	0.011
Std.	41.186	19.985	61.101	1.07E+12	80.909	0.107

Hence, four dimensionless groups, including length ratio (*L/D*), width ratio (w/H), the mass ratio ($M/\rho_i.H^3$), and the iceberg shape factor (S_f) were applied to estimate the iceberg draft ratio (*D*/*H*) through the LR approach in the present investigation. Figure 7-4 shows the combinations of four dimensionless groups introduced to develop the LR models.

LR is a linear method for the estimation of the relationship between a dependent parameter (*D*/*H*) and independent parameters (L/H , w/H , $M/\rho_i.H^3$, and S_f). The unknown parameters of the LR relationship were extracted from the constructed dataset using the LR analysis, and the relationship between the dependent and independent variables was determined utilizing a linear predictor function. As shown in Figure 7-4, nine LR models, including LR 1 to LR 9, were defined to introduce the premium LR models and the most influencing dimensionless groups. The LR 1 model included all dimensionless groups, while these dimensionless groups were disregarded one at a time in LR 2 to LR 5 models. It is worth noting that the LR 6 to LR 7 models function solely as one dimensionless group.

Model	L/H	W/H	$M/\rho_i H^3$	S_f
LR 1	■	■	■	■
LR 2	■	■	■	
LR 3	■	■		■
LR 4	■		■	■
LR 5		■	■	■
LR 6	■			
LR 7		■		
LR 8			■	
LR 9				■

Figure 7-4. Combinations of four dimensionless groups introduced to develop the LR models

7.1.2.3. Quality of fit

Several criteria like correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were employed to assess the precision, correlation, and complexity of the LR models. The *R* and *WI* criteria evaluated the correlation of these LR models. However, the *RMSE* and *MAPE* indices assessed the precision of these models. Besides, the performance and the complexity of the LR models have been evaluated using the *WI* and *AIC* value, respectively. The proximity of the *R* and *WI* criteria to one representing the LR model tended to have a high degree of correlation with the values observed. The nearness of the *RMSE*, *MAPE*, and *CRM* indices to zero signify the LR model possessed the lowest degree of impreciseness, but the complexity of the LR models was not examined through the indices introduced. The Akaike Information Criteria (*AIC*) was utilized to overcome this restriction. It means that the less complex LR model had the lowest amount of *AIC* so that the best LR model might have the lowest degree of *AIC* index and error (*RMSE*, *MAPE*, and *CRM*), with the highest amount of correlation (*R* and *WI*) (Ahmed et al. 2021):

$$R = \frac{\sum_{i=1}^{n}\left(P_i - \bar{P}\right)\left(O_i - \bar{O}\right)}{\sqrt{\sum_{i=1}^{n}\left(P_i - \bar{P}\right)^2 \sum_{i=1}^{n}\left(O_i - \bar{O}\right)^2}} \tag{7-28}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_i - O_i\right)^2} \tag{7-29}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{7-30}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}\left(O_i - P_i\right)^2}{\sum_{i=1}^{n}\left(\left|P_i - \bar{O}\right| + \left|O_i - \bar{O}\right|\right)^2} \quad (7\text{-}31)$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \quad (7\text{-}32)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_i - O_i\right)^2}\right) + 2k \quad (7\text{-}33)$$

Where *Oi, Pi,* $\bar{O}_i$*,* $\bar{P}_i$*, n*, and *k* are the observational value, the predicted amount, the average observational values, the average predicted amount, the number of observations, and the number of independent variables in the LR models.

7.1.3. Results and discussion

The results obtained from the LR models were evaluated by performing a sensitivity analysis. The superior LR models and the most influential dimensionless groups were also identified. Error analysis and uncertainty analysis were conducted for LR 1 to LR 9. The performance of the premium LR model was compared with the previous empirical models in the literature. Ultimately, a set of LR relationships was proposed to estimate the iceberg draft for practical applications.

7.1.3.1. Sensitivity analysis

Figure 7-5 demonstrates the key criteria obtained from the LR 1 to 9 models. Regarding the simulation results, the *R*, *AIC*, and *WI* values for the LR 1 model were respectively surmised as 0.882, 26.701, and 0.946. The LR 1 model estimated the iceberg draft ratio (*D/H*) by using all dimensionless groups, including L/H, w/H, $M/\rho_i.H^3$, and S_f. For the LR 2 model, the shape factor (S_f) was a dimensionless group dropped, indicating that this LR model utilized

dimensionless groups, L/H , w/H , and $M/\rho_i.H^3$, to approximate the iceberg draft ratio, with the *RMSE* and *R* values of 1.338 and 28.373. The magnitude of calculated *AIC*, *RMSE*, and *R* criteria for the LR 3 model was equaled at 26.922, 1.311, and 0.881, respectively. The LR 3 model was a function of L/H , w/H , and S_f, whereas the mass ratio ($M/\rho_i.H^3$) was eliminated to model the iceberg draft. The amount of *WI* and *R* indices for the LR 4 model was calculated to be 0.941 and 0.875, respectively. The LR 4 model predicted the iceberg draft in terms of L/H , $M/\rho_i.H^3$, and S_f, while the *w*/*H* was removed from the model. Even though the effect of the iceberg length ratio (*L*/*H*) was excluded in the LR 5 model, the iceberg draft ratio was modeled utilizing other factors, such as w/H , $M/\rho_i.H^3$, and S_f by this linear regression model. In the current investigation, four LR models (LR 6 to LR 9) were developed using just one dimensionless group. For instance, the LR 6 model was a function of the iceberg length ratio (*L*/*H*), with an *AIC* and *RMSE* criteria of 30.994 and 1.389. In the LR 7 model, the key input dimensionless group included the iceberg width ratio (*w*/*H*). The value of *CRM*, *WI*, and *R* for the LR 7 model was respectively estimated at 0.0004, 0.930, and 0.858. The iceberg mass ratio ($M/\rho_i.H^3$) was used to simulate the iceberg draft by the LR 8 model, with the *RMSE*, *R*, and *AIC* of 2.774, 0.040, and 79.328. The LR 9 was merely fed with the iceberg shape factor (S_f), while the magnitude of the *WI* and *CRM* indices was reckoned as 0.670 and -0.002, respectively.

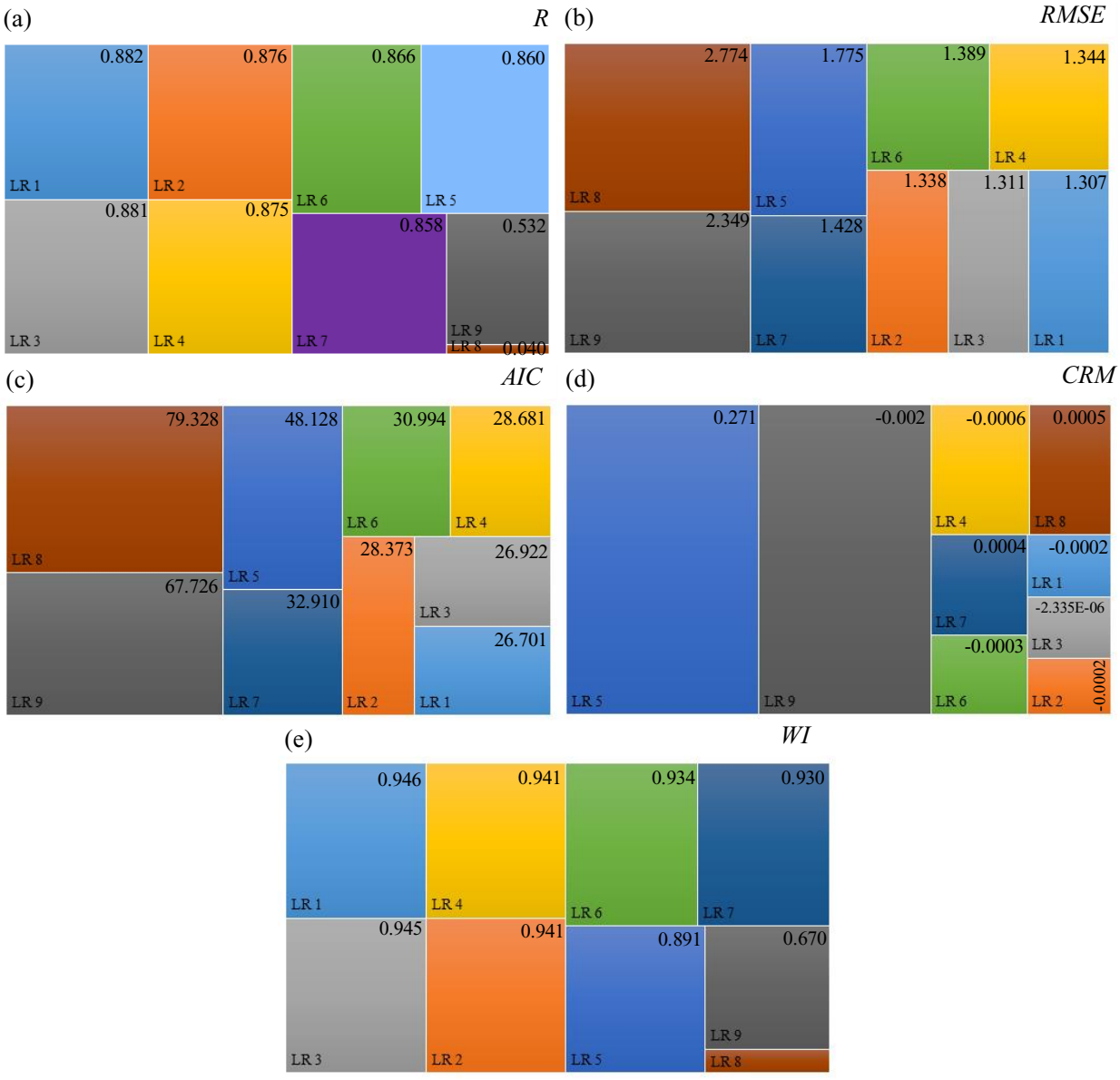


Figure 7-5. Key criteria obtained from the LR 1 to LR 9 models (a) *R* (b) *RMSE* (c) *AIC* (d) *CRM* (e) *WI*

The sensitivity analysis showed that the LR 1 model best estimates the iceberg draft ratio. The comparison between the iceberg drafts estimated by the LR models with the observational values is presented in Figure 7-6. The LR 1 model had the highest correlation, accuracy, and lowest level of complexity in predicting the target parameter.

The LR 3, LR 2, and LR 4 models were the second, third, and fourth best LR models for estimating iceberg draft ratio, followed by LR 6, LR 7, LR 5, LR 9, and LR 8 models, respectively. On the other hand, the iceberg length ratio (*L/H*) was identified as the most significant dimensionless group to model the iceberg draft, but the iceberg width ratio (*w/H*) was situated in second place in terms of effectiveness. Furthermore, the third-significant dimensionless group was allocated to the iceberg shape factor (S_f) for the simulation of the iceberg draft ratio. The influence of the iceberg mass ratio, $M/\rho_i.H^3$, was relatively trivial to estimate the iceberg draft.

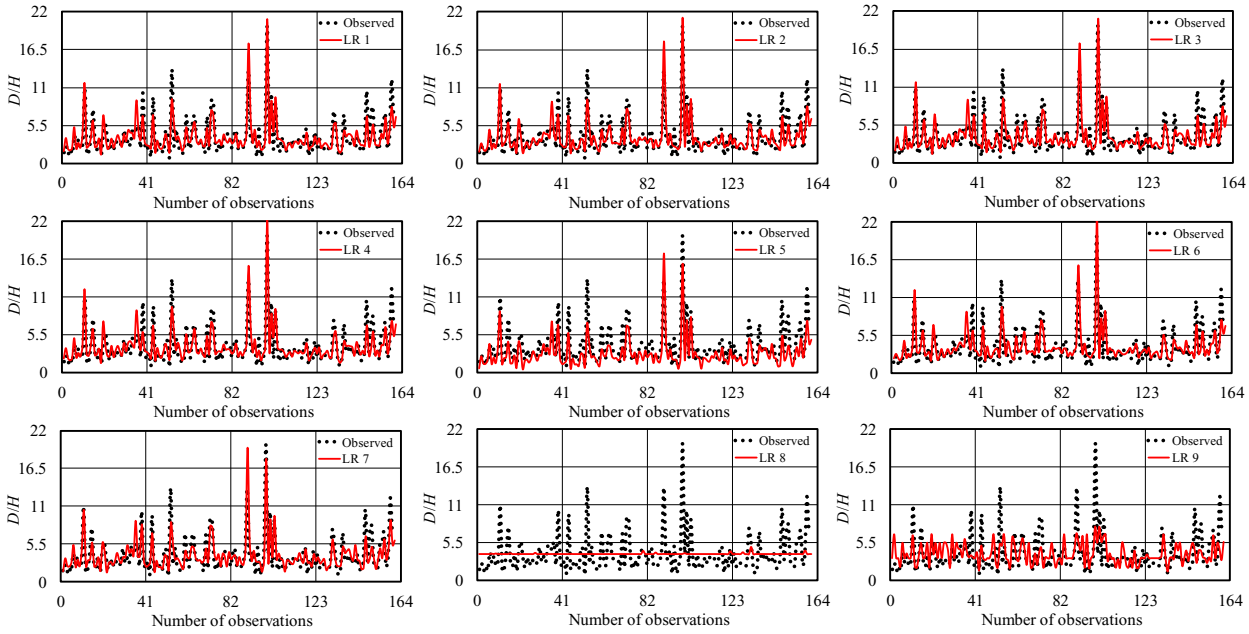


Figure 7-6. Comparison between the iceberg drafts estimated by the LR models with the observational values (a) LR 1 (b) LR 2 (c) LR 3 (d) LR 4 (e) LR 5 (f) LR 6 (g) LR 7 (h) LR 8 (i) LR 9

Although there were several discrepancies between the results simulated by the LR 1 model and the observational values, this model managed to approximate the iceberg draft. As shown in Figure

15-6, the LR 1 model had the highest level of correlation with the field data, but the LR 8 model showed the worst performance in estimating the target value.

7.1.3.2. Error analysis

Figure 7-7 demonstrates the error analysis results obtained from the LR models. For example, roughly one-quarter of the iceberg draft estimated by the LR 1 model showed an error of less than 10% though this amount for the LR 2, LR 3, and LR 4 models was about 20%, 25%, and 22%, respectively. Regarding the error analysis, a tiny minority of the iceberg draft simulated by the LR 5 model possessed an error of between 10% and 20%, and this value for the LR 1 model was approximately 24%. Nearly half of the outcomes of the LR 6 and LR 7 models had an error of greater than 20% (55%); however, a vast majority of the iceberg drafts predicted by the LR 8 model contained an error of more than 20% (77%). In addition, the error analysis indicated that about 85% of the LR 9 model's results had more than 20% error, while this value for the LR 1 model was just about 52%.

In summary, the conducted error analysis proved that the LR 1 model performed the iceberg draft well compared with the other LR models; instead, the LR 8 and LR 9 models showed the worst performance in terms of error analysis.

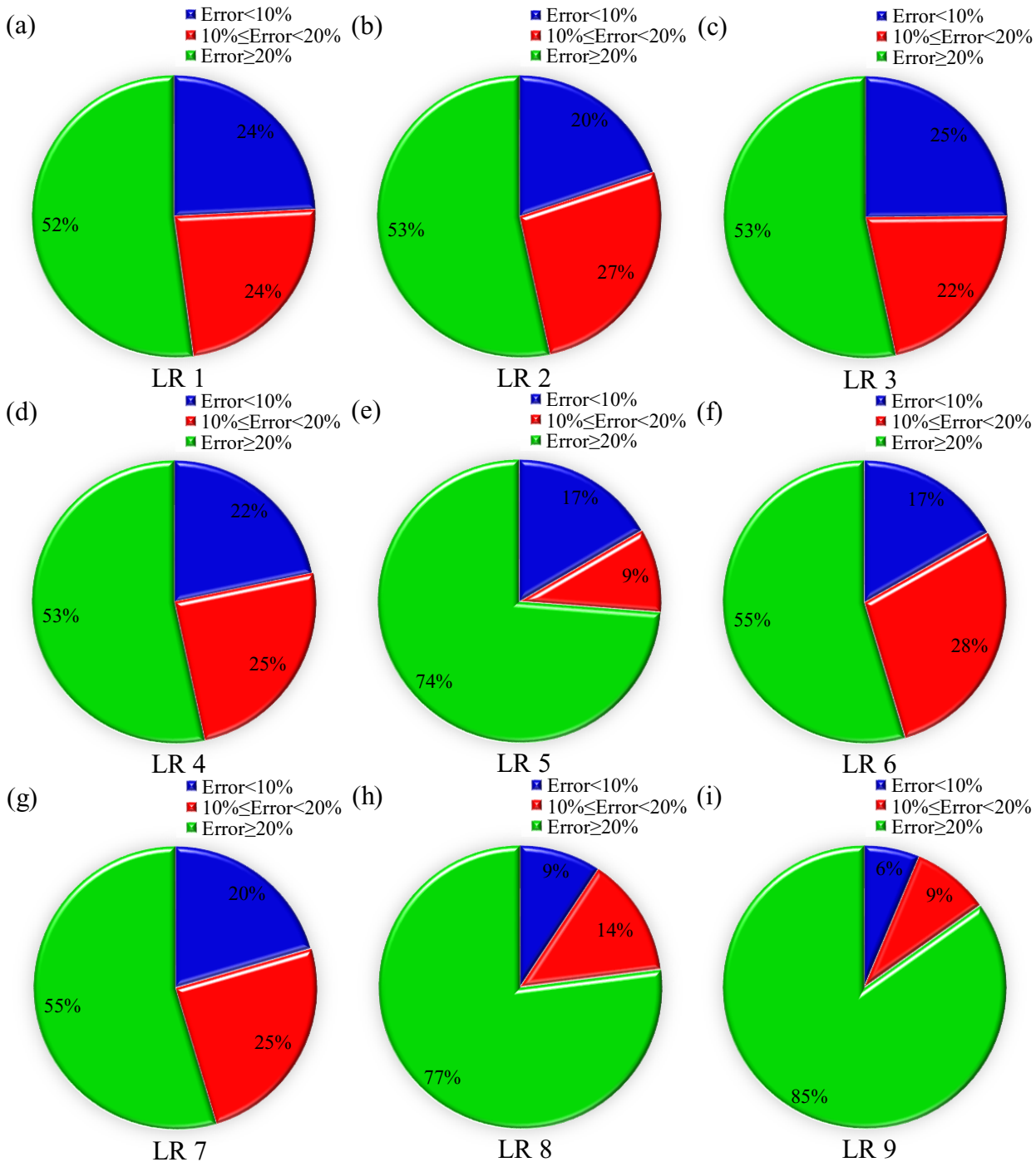


Figure 7-7. Results of error analysis obtained from (a) LR 1 (b) LR 2 (c) LR 3 (d) LR 4 (e) LR 5 (f) LR 6 (g) LR 7 (h) LR 8 (i) LR 9 models

The error analysis conducted for the LR models demonstrated that the LR 1 model could approximate the iceberg draft ratio with a reasonable degree of precision.

7.1.3.3. Uncertainty analysis

An uncertainty analysis was conducted to appraise the performance of the LR models developed in the present investigation. To implement the uncertainty analysis, errors from the LR models (e_j) were obtained as the difference between the iceberg drafts estimated (P_j) and the observational amounts (O_j), as below:

$$e_j = P_j - O_j \tag{7-34}$$

The mean (Mean) and the standard deviation (StDev) of these errors were computed through the equations below:

$$\text{Mean} = \frac{1}{n}\sum_{j=1}^{n} e_j \tag{7-35}$$

$$\text{StDev} = \sqrt{\sum_{j=1}^{n}\left(e_j - \bar{e}\right)^2 / \left(n-1\right)} \tag{7-36}$$

A particular LR model underestimated the iceberg draft if the sign of the Mean value was negative, whereas the positive sign of the Mean reprinted that the LR model had an overestimated performance on the iceberg draft. Subsequently, a confidence interval (CI) was generated around the error estimated using the Mean, StDev values, and the "Wilson score approach" by disregarding the continuity correction. A regular distribution interval rectified as an asymmetric normal distribution called the Wilson score interval, was utilized to modify the CI bounds. After that, a ±1.96Se generated a 95%CI. It should be stated that the width of uncertainty bound (WUB) of each single LR model was obtained as below (Azimi, Shiri, and Zendehboudi 2022):

$$\mathrm{WUB} = \pm \frac{\left(\text{Lower bound} - \text{upper bound}\right)}{2} \tag{7-37}$$

The normal distribution of the CI for the iceberg drafts simulated by the LR models is illustrated in Figure 7-8. The LR 1, LR 2, LR 3, and LR 4 models estimated the CI value as (-0.205 to 0.203), (-0.208 to 0.210), (-0.205 to 0.205), and (-0.212 to 0.208), respectively. According to the uncertainty analysis, the LR 1, LR 3, LR 4, LR 6, and LR 9 models underestimated the iceberg draft, with a Mean value of -0.001, -9.046E-06, -0.002, -0.001, and -0.009, respectively. However, the LR 2, LR 5, LR 7, and LR 8 models simulated the iceberg draft with an overestimated performance. The LR 1 model had the narrowest width of uncertainty bound (WUB=±0.204), whereas the widest uncertainty bound belonged to the LR 8 model, where the value of WUB was equal to ±0.433. The CI estimated for the LR 5, LR 6, LR 7, LR 8, and LR 9 models were at (0.827 to 1.274), (-0.218 to 0.216), (-0.221 to 0.225), (-0.431 to 0.435), and (-0.376 to 0.357), respectively.

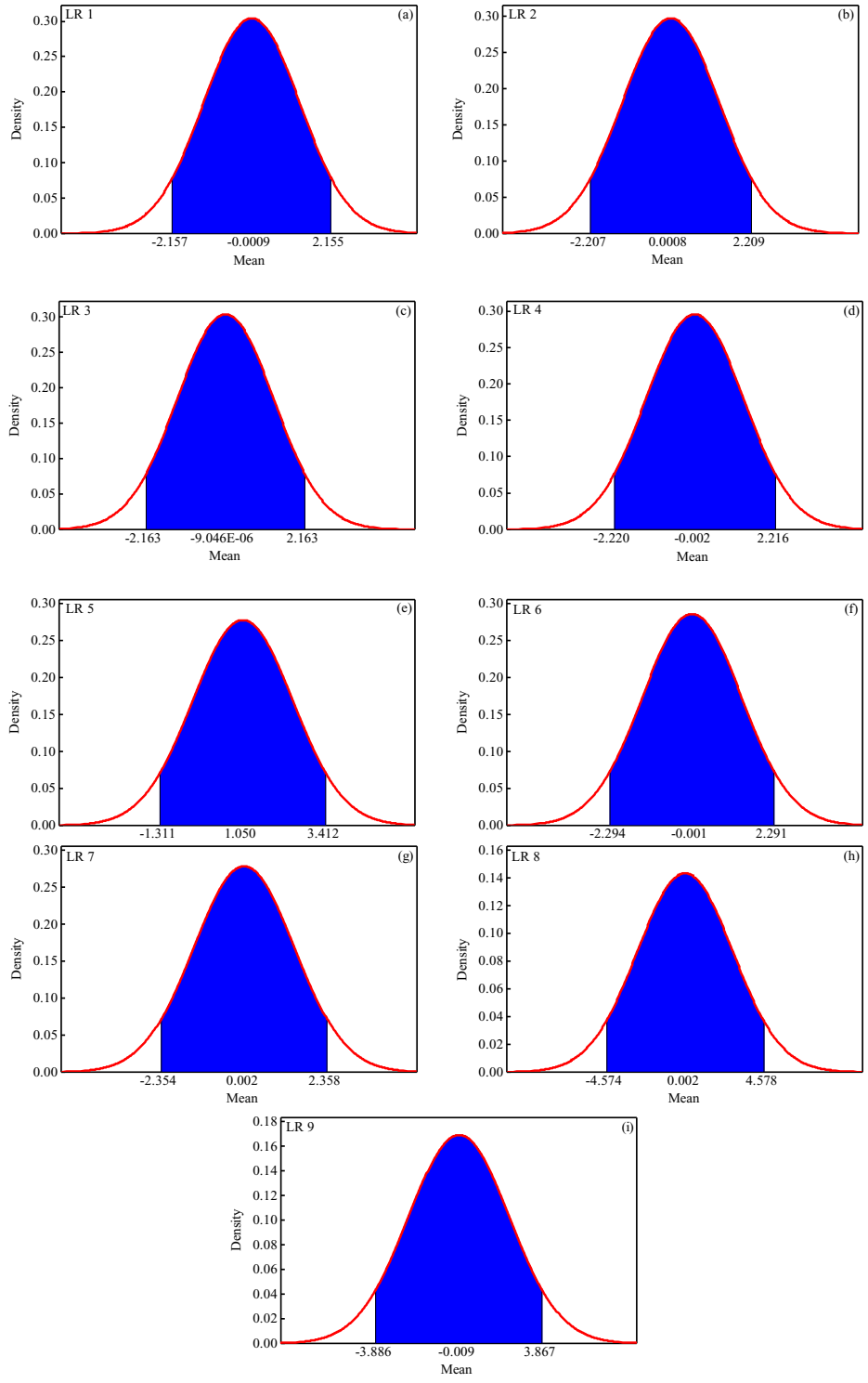


Figure 7-8. Normal distribution of the confidence interval for the iceberg drafts simulated by the LR models

Thus, the uncertainty analysis done for the LR models proved that the premium LR model (LR 1) possessed the narrowest width of uncertainty bound, and this model underestimated the iceberg draft, with a CI of (-0.205 to 0.203).

7.1.3.4. Comparison with the previous models

Several empirical models were developed to estimate the iceberg draft by Robe and Farmer (1976) (Eq. 7-1), Mognor and Zorn (1979) (Eq. 7-2), El-Tahan and El-Tahan (1982) (Eq. 7-3 and Eq. 15-4), Hotzel and Miller (1983) (Eq. 7-5), C-CORE (2001) (Eq. 7-6 and Eq. 7-7), Barker, Sayed, and Carrieres (2004) (Eq. 7-8 and Eq. 7-9), Sacchetti et al. (2012) (Eq. 7-10 and Eq. 7-11), King et al. (2016) (Eq. 7-12), and Stuckey et al. (2021) (Eq. 7-13 and Eq. 7-14). The performance of the premium LR model (LR 1) was compared with the empirical models proposed for the constructed dataset. The results of key criteria obtained from these empirical models are presented in Table 7-2. For example, the value of the *R* index for Eq. (7-1), Eq. (7-2), and Eq. (7-3) were respectively equaled 0.526, 0.767, and -0.500. Additionally, for Eq. (7-4), Eq. (7-5), and Eq. (7-6), the *WI* criterion was estimated at 0.859, 0.850, and 0.958, respectively. It is important to note that the empirical models were functions of solely one feature of the iceberg, e.g., the iceberg length, iceberg height, iceberg width, and iceberg mass; however, the LR 1 model as the superior model introduced in the current study managed to simulate the iceberg draft in terms of all iceberg features. Amongst the empirical models, Eq. (7-7) had the lowest level of precision and correlation, but Eq. (7-10) and Eq. (7-13) showed the highest degree of correlation to predict the iceberg draft. It is noteworthy that Eq. (7-5), Eq. (7-6), Eq. (7-8), Eq. (7-9), Eq. (7-10), Eq. (7-11), Eq. (7-13),

and Eq. (7-14) surmised the iceberg draft with almost identical performance, while Eq. (7-7) and Eq. (7-12) had a similar capability in the estimation of the target parameter.

Table 7-2. Results of key criteria obtained from the empirical models

Model	*R*	*RMSE*	*AIC*	*CRM*	*WI*
LR 1	0.882	1.307	26.701	0.0002	0.946
Eq. (7-1)	0.526	36.589	253.699	-0.066	0.505
Eq. (7-2)	0.767	96.518	321.522	-0.932	0.351
Eq. (7-3)	-0.500	51.201	277.194	-0.060	0.775
Eq. (7-4)	-0.509	45.921	269.584	0.124	0.859
Eq. (7-5)	0.769	26.964	232.355	0.057	0.850
Eq. (7-6)	0.769	26.684	231.628	-0.039	0.958
Eq. (7-7)	0.330	185.770	367.305	-0.565	0.222
Eq. (7-8)	0.769	38.301	256.896	-0.137	0.867
Eq. (7-9)	0.769	28.011	235.020	-0.085	0.895
Eq. (7-10)	0.770	29.389	238.378	-0.089	0.896
Eq. (7-11)	0.769	26.517	231.189	0.027	0.863
Eq. (7-12)	0.333	163.520	358.385	-0.429	0.257
Eq. (7-13)	0.770	35.917	252.404	0.267	0.674
Eq. (7-14)	0.768	27.684	234.199	0.071	0.824

As shown, the LR 1 model outperformed the empirical models, meaning that the superior linear regression model could simulate the iceberg draft with the highest level of correlation and accuracy

and the lowest degree of complexity. Furthermore, the LR 1 model had more generalization capability since this model estimated the iceberg draft in a wide range of measurements; instead, the empirical models were derived from solely particular circumstances or a specific study area. Even though the previous empirical equations were a function of just one parameter, e.g., the length, width, height, or mass of icebergs, the influence of all dimensionless groups was considered in the LR 1 model.

It is worth mentioning that remote sensing technology has been widely used to assess different aspects of iceberg properties comprising iceberg detection and iceberg trajectory (Lopez-Lopez et al. 2021; Braakmann-Folgmann et al. 2022), iceberg drift (Monteban et al. 2020), and iceberg draft (Liu et al. 2021). The geometrical properties of the exposed part of the icebergs are obtained from remote sensing studies. However, the iceberg draft, which is the underwater portion, cannot be estimated directly through satellite images. The current study is to explicitly predict the iceberg draft with reasonable accuracy using the above-water properties of the iceberg.

The significant advantages of the current study, which can facilitate practical solutions to protect the subsea infrastructures in actual iceberg grounding events, were summarized as follows:

(i) The literature review revealed that the dimensionless groups affecting the iceberg draft estimation had not been ascertained, which were determined for the first time in this work. These dimensionless groups affecting the iceberg draft estimation may be utilized to present additional cost-effective and quick solutions in practical applications.

(ii) Even though several empirical models were suggested in the past by other scholars, these models were able to estimate the iceberg draft using merely one parameter, e.g., iceberg length, iceberg mass, iceberg height, or iceberg width, and influence of other parameters affecting the iceberg draft estimation was disregarded. However, the LR models proposed in the current study

predicted the iceberg draft through the above-water parameters of the iceberg, including iceberg length, iceberg width, iceberg mass, and iceberg shape factor simultaneously.

(iii) The analyses performed in this investigation represented the better performance of the explicit LR models than prior empirical methods. It means that the premium LR model had the lowest level of inaccuracy and complexity along with the highest degree of correlation with the observational values compared with the previous methods.

(iv) The methodology presented in the current investigation may pave the way to achieve robust and versatile answers for the iceberg draft prediction. This methodology may preserve a significant amount of time and expenses, particularly in the initial phases of the iceberg management projects. Therefore, the current study was the first step to identifying the most influencing parameter for the iceberg draft estimation, and the obtained results might inspire further investigations in this domain.

7.1.4. Conclusion

In the present investigation, the dimensionless groups affecting the iceberg draft were distinguished using the Buckingham-π theorem for the first time. Regarding the dimensionless groups detected and using the linear regression (LR) method, nine LR models (LR 1 to LR 9) were defined. A comprehensive dataset was constructed to validate the LR model results. For the LR models developed, several analyses, including sensitivity analysis, error analysis, and uncertainty analysis, were implemented. The performance of the LR models was compared with the previous empirical models. The most important findings of this study can be summarized as follows:

- Four dimensionless groups comprising the length ratio (*L/H*), width ratio (*w/H*), the mass ratio ($M/\rho_i.H^3$), and iceberg shape factor (S_f) were recognized to predict the iceberg draft.
- The best LR model (LR 1) could approximate the iceberg draft by using all dimensionless groups introduced.
- The LR 1 model estimated the iceberg draft with the highest level of correlation (*R*=0.882), the highest degree of precision (*RMSE*=1.307), and the lowest level of complexity (*AIC*=26.701).
- The performed sensitivity analysis demonstrated that the iceberg length ratio (*L*/*H*) and the iceberg width ratio (*w/H*) had the highest level of effectiveness in modeling the iceberg draft.
- The error analysis showed that almost a quarter of the iceberg draft estimated had an error of less than 10%.
- According to the uncertainty analysis, the superior LR model was biased towards underestimating the iceberg draft ratio.
- The comparison between the LR 1 and the previous models unfolded that the LR 1 model outperformed its counterparts.
- A set of LR models were proposed to simulate the iceberg drafts.

The proposed dimensionless groups may provide better insight into the key parameter affecting the iceberg draft. The outcome of this study can build a way to foster quick and cost-effective alternatives to use in the early phases of engineering design projects, where an intuition and detailed assessment is required for planning the construction methodology, corresponding logistics, and the future scope of precise schemes.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)," and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)." Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

Appendix

In the current investigation, nine LR models (LR 1 to LR 9) were proposed to estimate the iceberg draft ratio. The LR-based relationships derived in this study are shown in Table 7-3. It is worth noting that the coefficients of the proposed LR models were derived from the Minitab statistical software.

Table 7-3. The LR-based relationship proposed to estimate the iceberg draft ratio

Model	Relationship	Equation
LR 1	$D/H=-0.159+0.295(L/H)+0.330(w/H)+0.000006(M/\rho_i.H^3)+3.16(S_f)$	(7-38)
LR 2	$D/H=0.542+0.309(L/H)+0.367(w/H)+0.000008(M/\rho_i.H^3)$	(7-39)
LR 3	$D/H=-0.165+0.298(L/H)+0.322(w/H)+3.30(S_f)$	(7-40)
LR 4	$D/H=-0.084+0.488(L/H)+0.000005(M/\rho_i.H^3)+3.58(S_f)$	(7-41)
LR 5	$D/H=-0.274+0.761(w/H)+0.000007(M/\rho_i.H^3)+3.52(S_f)$	(7-42)
LR 6	$D/H=0.759+0.529(L/H)$	(7-43)

LR 7	$D/H=0.546+0.824(w/H)$	(7-44)
LR 8	$D/H=3.85+0.000006(M/\rho_i.H^3)$	(7-45)
LR 9	$D/H=-0.260+13.9(S_f)$	(7-46)

References

Ahmed, A. N., Van Lam, T., Hung, N. D., Van Thieu, N., Kisi, O., & El-Shafie, A., 2021. A comprehensive comparison of recent developed meta-heuristic algorithms for streamflow time series forecasting problem. Applied Soft Computing, 105: 107282.

Azimi, H., Shiri, H., 2020. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142 (5): 051601.

Azimi, H., Shiri, H., Zendehboudi, S., 2022. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology, 193: 103426.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In the 14th international offshore and polar engineering conference, ISOPE-I-04-116.

Braakmann-Folgmann, A., Shepherd, A., Gerrish, L., Izzard, J, & Ridout, A., 2022. Observing the disintegration of the A68A iceberg from space. Remote Sensing of Environment, 270: 112855.

Campbell, P.J., 2014. Grand Banks Scour Catalogue (GBSC) GeoDatabase. Geological Survey of Canada, Atlantic.

C-CORE., 2001a. Integrated Ice Management R&D Initiative - Year 2000, C-CORE Report 00-C36.

C-CORE., 2001b. Documentation of iceberg grounding events from the 2000 season. C-CORE publication 01-C10 (Revision 0). Report submitted to the Geological Survey of Canada, Atlantic.

El-Tahan, M., El-Tahan, H., 1982. Estimation of iceberg draft. OCEANS 82, IEEE, Washington, DC, USA, 689-695.

El-Tahan, M., El-Tahan, H., Courage, D., Mitten, P., 1985. Documentation of Iceberg Groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

Enderlin, E.M., Carrigan, C.J., Kochtitzky, W.H., Cuadros, A., Moon, T, Hamilton, G. S., 2018. Greenland iceberg melt variability from high-resolution satellite observations. The Cryosphere, 12(2): 565-575.

Hotzel, I.S., Miller, J.D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Arctic Technology Conference, OTC-27358-MS.

Liu, X., Cheng, X., Liang, Q., Li, T., Peng, F., Chi, Z, He, J., 2021. Grounding Event of Iceberg D28 and Its Interactions with Seabed Topography. Remote Sensing, 14(1): 154.

Lopez-Lopez, L., Parmiggiani, F., Moctezuma-Flores, M, Guerrieri, L., 2021. On the detection and long-term path visualisation of a-68 iceberg. Remote Sensing, 13(3): 460.

Løset, S, T. Carstens, T., 1996. Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Regions Science and Technology, 24 (4): 323-340.

Luckman, A., Padman, L, Jansen, D., 2010. Persistent iceberg groundings in the western Weddell Sea, Antarctica. Remote Sensing of Environment, 114(2): 385-391.

Mahmood, K, Siddiqui, M.M., 1980. Spurious Correlation in Dimensional Analysis. Journal of the Engineering Mechanics Division, 106(1): 93-109.

Massey, B.S., 1971. Units, Dimensional Analysis and Physical Similarity. University College London, Van Nostrand Reinhold Company, London.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. In the 25th International Conference on Port and Ocean Engineering under Arctic Conditions, pp. 9-19.

McGuire, P., Younan, A., Wang, Y., Bruce, J., Gandi, M., King, T., Keeping, K., Regular, K., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

Minerals Management Service., 2002. Evaluation of Sub–Sea Physical Environmental Data for the Beaufort Sea OCS and Incorporation into a Geographic Information System (GIS) Database. OCS Study MMS 2002–017.

Mognor, K, Zorn, R., 1979. Discussion of Another Hypothesis About Iceberg Draft, 3rd International Conference on Port and Ocean Engineering Under Arctic Conditions, Fairbanks.

Monteban, D., Lubbad, R., Samardzija, I, Løset, S., 2020. Enhanced iceberg drift modelling in the Barents Sea with estimates of the release rates and size characteristics at the major

glacial sources using Sentinel-1 and Sentinel-2. Cold Regions Science and Technology, 175: 103084.

Papusha, A., Kazunin, D., Gontarev, D., Vasilevich, V., Kyalbieva, S., 2013. Stability and strength of the subsea pipeline under iceberg load in Arctic. SPE Arctic and Extreme Environments Technical Conference and Exhibition, Moscow, Russia. SPE-166941-MS.

Robe, R.Q., Farmer, L.D., 1976. Physical Properties of Icebergs. Part I. Height to Draft Ratios of Icebergs. Part II. Mass Estimation of Arctic Icebergs. Coast Guard Research and Development Center Groton Conn.

Rudkin, P., 2005. Comprehensive Iceberg Management Database Report 2005 update. PERD/CHC Report 20-72. National Research Council of Canada (NRC) and Panel on Energy Research and Development (PERD).

Sacchetti, F., Benetti, S., Cofaigh, C.Ó., Georgiopoulou, A., 2012. Geophysical evidence of deep-keeled icebergs on the Rockall Bank, Northeast Atlantic Ocean. Geomorphology, 159: 63-72.

Simon, V., Weigand, B., Gomaa, H., 2017. Dimensional analysis for engineers. Berlin: Springer.

Sonnichsen, G., Hundert, T., Myers, R., Pocklington, P., 2006. Documentation of Recent Iceberg Grounding Evens and a Comparison with Older Events of Know Age. Environmental Studies Research Funds. Report ESRF Vol. 157.

Stuckey, P.D., 2008. Drift speed distribution of icebergs on the grand banks and influence on design loads. PhD dissertation, Memorial University of Newfoundland, NL, Canada.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., & King, T., 2021. Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. In the 40th

International Conference on Ocean, Offshore and Arctic Engineering, Vol. 85178, p. V007T07A018.

Talimi, V., Ni, S., Qiu, W., Fuglem, M., MacNeill, A., Younan, A., 2016. Investigation of Iceberg Hydrodynamics. Arctic Technology Conference, OTC-27493-MS.

Turnbull, I. D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Regions Science and Technology, 110: 1-18.

Turnbull, I. D., King, T., Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Woodworth-Lynas, C.M.T., Simms, A., Rendell, C.M., 1985. Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Regions Science and Technology, 10 (2): 163-186.

Younan, A., Ralph, F., Ralph, T., Bruce, J., 2016. Overview of the 2012 iceberg profiling program. Arctic Technology Conference, OTC-27469-MS.

Zhou, M., 2017. Underwater iceberg profiling and motion estimation using autonomous underwater vehicles. PhD dissertation, Memorial University of Newfoundland, NL, Canada.

Section 2

Sensitivity Analysis of Parameters Governing the Iceberg Draft through Neural Network-Based Models

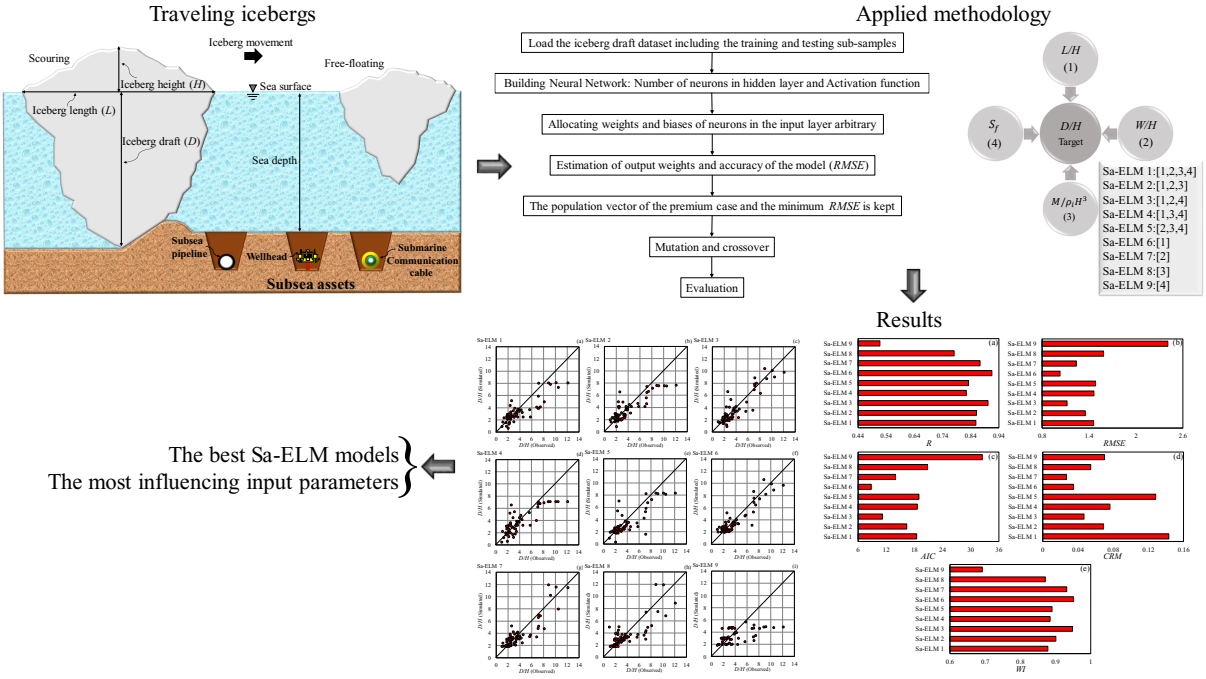
This section is a journal paper which has been published in Journal of Ocean Engineering and Marine Energy (2023), Vol. 9, ISSN: 2198-6444

Abstract

Precise estimation of the iceberg draft may significantly reduce the collision risk of deep keel icebergs with the offshore facilities comprising the submarine pipelines, wellheads, communication cables, and hydrocarbon loading equipment crossing the Arctic shallow waters. As such, in this study, the iceberg drafts were simulated using a self-adaptive machine learning (ML) algorithm entitled self-adaptive extreme learning machine (Sa-ELM) for the first time, to the best of our knowledge. Initially, the parameters governing the iceberg drafts were specified, and then nine Sa-ELM models were defined using these parameters. To test and train the Sa-ELM models, a comprehensive dataset was constructed, where 60% of the dataset was utilized for model training and 40% for model validation. In addition, several hyper-parameters have been optimized during the training procedure to obtain the most accurate results. The superior Sa-ELM model and the most influencing input parameters were determined by conducting a sensitivity analysis. The comparison of the premium Sa-ELM model with the artificial neural network (ANN) and extreme learning machine (ELM) models demonstrated that the Sa-ELM model had the highest level of accuracy and correlation as well as the lowest degree of complexity. Ultimately, a Sa-ELM-based equation was presented to estimate the iceberg draft in practical applications.

Keywords: Artificial neural network, Machine learning, Iceberg Draft, Subsea assets, Sensitivity analysis, Self-adaptive evolutionary extreme learning machine, Extreme learning machine

Graphical abstract



7.1.5. Introduction

Determining the iceberg draft in the Arctic and subarctic shallower waters have led to a large number of investigations, where the operational and structural integrity of subsea pipelines, wellheads, offshore oil loading equipment, and communication cables are likely to threaten by the deep iceberg keels. The schematic layout of the iceberg free-floating and iceberg scouring in cold waters is demonstrated in Figure 7-9. As seen, the iceberg is in a free-floating situation if the ocean depth is greater than the iceberg draft; otherwise, the seafloor is scoured and the seabed soil shear resistance causes the soil displacement to extend deeper than the iceberg tip threatening the subsea assets, which is called the ice-gouging event.

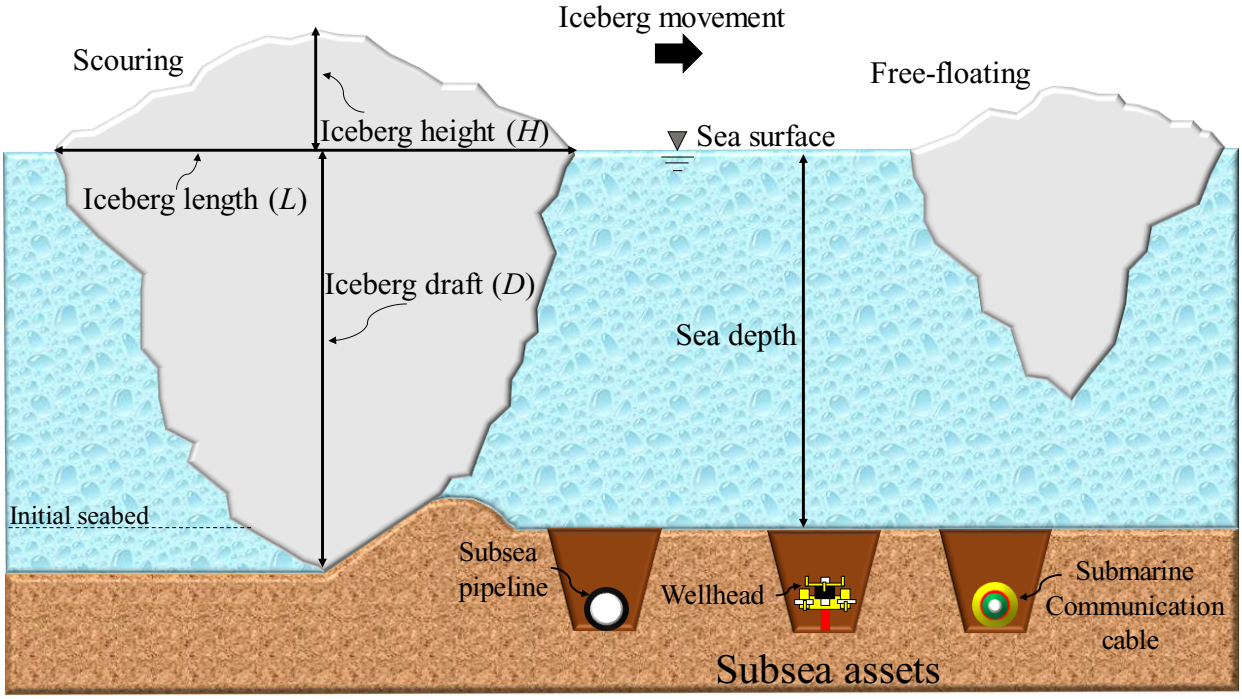


Figure 7-9. The iceberg free-floating and iceberg scouring schematically

The efficient iceberg management designs and the guaranteed operational integrity of the sea bottom-founded infrastructure against the berg attacks in the ice-prone areas demand the

appropriate iceberg draft appraisal, which may lead to a potential decrease in operating expenses and downtime. The side scan sonar mounted on autonomous underwater vehicles (AUVs), and remote operating vehicles (ROVs) is currently used for measuring iceberg drafts. However, these methods are costly because of needing vessels, experienced crew, expensive logistics, and advanced pieces of equipment. The industry is still looking for cost-effective solutions that can assess the iceberg draft using the easily measurable exposed parameters. The current study has tried to address this need by performing a sensitivity analysis to identify the most influential parameters that can be used in ML-based algorithms for draft estimation. Several investigations have tended to focus on iceberg characteristics and behavior. For instance, iceberg stability, iceberg calving evaluation, iceberg drift, and iceberg draft were assessed in analytical studies, e.g., Allaire (1972) analyzed the stability status of various icebergs regarding the sail features. They demonstrated that the minimum stable ratio of iceberg width to sail height for tabular, dry dock, and dome bergs were 6:1, 4:1, and 1.8:1, respectively. Bass (1980) analytically evaluated the stability of icebergs in different configurations. The study highlighted that the volume of the underwater section was almost seven times greater than the volume of the above-water section. In addition, Brooks (1980) analyzed the stability of theoretical icebergs and provided a hypothesis to estimate the iceberg draft. The investigation demonstrated that the draft of icebergs was smaller than the length of icebergs. In another study, Hotzel and Miller (1983) assessed the iceberg dimensions and supposed that the planimetric section of icebergs was circular, while the length of a berg was twice the circle radius. The iceberg draft was estimated in terms of iceberg length by using the power-law method. Liang (2001) presented a model to determine the relationship between the dynamics and stability of icebergs using the size and shape of bergs. The author asserted that the draft and hydrostatic force distribution affected the iceberg's stability. In another

study, Barker et al. (2004) evaluated the geometry of iceberg sails and keels. They estimated the cross-sectional areas of the berg at different water depth intervals from a particular waterline length. To approximate the iceberg draft, a set of models in terms of the iceberg length were suggested using the power curve and regression analysis. McKenna (2004) studied the spatial correlation between the above-water and below-water sections of the iceberg. The relationships between the length, width, height, and draft of icebergs were determined, and the three-dimensional shape of icebergs was simulated. Stuckey (2008) modeled the iceberg drift speed by using the probabilistic approach. The author stated that the above-water dimension, below-water dimension, and shape factor of the icebergs impacted the environmental driving forces. McKenna and King (2009) simulated the deterioration mechanisms of various icebergs by considering the incremental changes in draft, mass, and shape of bergs. The study highlighted that the draft and length of icebergs were reduced by decreasing the iceberg mass. In addition, Sacchetti et al. (2012) analyzed the features of different icebergs, such as wedged, domed, tabular, and pinnacle and ice scouring in the Northeast Atlantic Ocean. The investigation reported that the bimodal distribution of the scour depth was observed across the study area.

Furthermore, iceberg characteristics were documented in the field investigation, e.g., Robe and Farmer (1976) measured the drafts of tabular, broken tabular, pinnacle, dry dock, and domed by the sonar technology. They correlated the iceberg draft with the iceberg height and proposed a regression model for the estimation of the iceberg draft. The International Ice Patrol (IIP) has been annually monitoring the characteristics of traveling icebergs in the Atlantic and Arctic Oceans. In 1987, the IIP reported that the iceberg draft was surmised as 3.95 times the height of a sail. Several investigations were performed by the Centre for Cold Ocean Resources Engineering (C-CORE) to study the shape, behavior, and dynamics of icebergs over the years. The above-water shape of

bergs comprising the blocky, domed, dry dock, tabular, pinnacle, and wedge icebergs were reported regarding the standard categories. The iceberg dimensions were recorded using the scan sonar technology and marine sextant. The study showed that the iceberg draft was correlated with iceberg length (C-CORE, 2001). Dowdeswell and Bamber (2007) examined the keel depths of traveling icebergs in Antarctic waters. The authors estimated the depth of the keel utilizing the ice thickness and surface elevation. The research concluded that a tiny minority of icebergs in the Antarctica and Greenland waters had drafts of greater than 650 m.

Different numerical models were used to simulate the iceberg features, e.g., King (2012) modeled the iceberg's characteristics including the draft, length, and mass through the Monte Carlo simulation. The author outlined that the survey of iceberg drafts greater than 150 m was quite limited. Turnbull et al. (2015) forecasted the drift trajectory of the traveling icebergs in Northwest Greenland using the hindcast simulation. It was revealed that the trajectory of icebergs was sensitized to the iceberg drafts. King et al. (2016) performed a field investigation to calculate the iceberg rolling rate. The iceberg drafts were estimated through a calving analysis, with a calculated standard deviation of draft variations from 19% to 34%. The iceberg drafts corresponded with the mass of the icebergs. Talimi et al. (2016) simulated the iceberg-structure interaction using computational fluid dynamics (CFD). The modeling results demonstrated that the iceberg was shifted upward by almost 10% of the iceberg draft. Turnbull et al. (2018) proposed a model for the drift estimation of moving icebergs on the Grand Banks of Newfoundland. This model approximated the draft of icebergs roughly 1.3 times more than the real values.

McKenna et al. (2019) have recently simulated ice scouring on the Grand Banks of Canada by using the Monte Carlo method. The iceberg draft alterations were utilized to lessen the size of draft variations in this modeling. Most recently, Stuckey et al. (2021) simulated the three-dimensional

iceberg shapes by adopting field measurements. The investigation demonstrated that the draft and mass of the iceberg were estimated in terms of the iceberg length by utilizing the power curve.

The determination of objective function and parameters affecting a specific problem using machine learning (ML) advancement were widely investigated in prior studies, e.g., Guyon and Elisseeff (2003), Ma et al. (2020) Azimi et al. (2023) since the selection of objective functions and parameters affecting a problem can play a significant role in improving the simulation performance of ML algorithms, e.g., providing a reliable solution, presenting the more cost-effective methodology, and giving a better comprehension for generating datasets.

Despite the wide application of ML technology in the estimation of linear and nonlinear problems (Azimi and Shiri 2020a, Mahdianpari et al. 2021, Azimi et al. 2022, 2023), no previous study has estimated the iceberg drafts through the ML algorithm so far. To fill this knowledge gap, the iceberg drafts were simulated using the artificial neural network (ANN), extreme learning machine (ELM), and self-adaptive extreme learning machine (Sa-ELM) in the current study, for the first time. Several analyses were performed to ascertain the superior ML models and the most significant input parameters to estimate the iceberg draft.

7.1.6. Methodology

To simulate the iceberg drafts, three ML algorithms comprising the ANN, ELM, and Sa-ELM models were employed. More details associated with these ML approaches are provided in the next sections. Subsequently, the parameters affecting the iceberg draft are introduced. A comprehensive dataset was then constructed to train and validate the ML algorithms. Ultimately, a set of statistical criteria were utilized to assess the performance of the ML algorithms.

7.1.6.1. Artificial neural networks (ANN)

ANN is one of the most universally supervised ML algorithms, which is widely utilized to simulate linear and non-linear relationships, such as regression and classification problems since this method is quite flexible and simple. The main structure of an ANN algorithm consists of at least three distinct layers comprising an input layer, a hidden layer, and an output layer. The input parameters are embedded within the input layer, while the target parameter, e.g., the iceberg draft, is considered in the output layer. The hidden neurons are situated within the hidden layer, where the size of this layer is determined by the problem's complexity and desired accuracy. It should be noted that there is no particular formula to adjust the number of neurons within the hidden layer, thus a trial and error procedure was employed to determine the optimum number of hidden layer neurons in the current study (Shahin et al. 2001). The number of hidden layer neurons was initially set as one and the magnitude of this hyperparameter was increased to 15, where the optimum number of the hidden layer neurons was chosen at 12 for the reason that the proficiency of the ANN algorithm was negligibly altered after this amount. In each hidden neuron, both the input parameters and their weights are calculated using mathematical operations, and the outcome is passed through a transfer function entitled the activation function. The sigmoid function was applied for the current architecture because it had better performance in comparison with other activation functions. The output value was then computed by randomly assigning the input parameters and weights. Subsequently, the performance of the ANN algorithm was evaluated by the Mean squared error (*MSE*) as the loss function in the present study to measure the difference between the computed outputs and the target outputs. The weights updating in each epoch using a procedure called “model training” which has been defined as the backpropagation algorithm,

where the Levenberg-Marquardt (LM) algorithm was selected as the model training approach and the number of epochs was set as 5000 in this study.

7.1.6.2. Extreme learning machine (ELM)

ELM as a single-layer feed-forward neural network (SLFFNN) has been introduced by Huang et al. (2004) to estimate various non-linear and linear problems. An ELM algorithm consists of three different layers, including the input layer, the hidden layer, and the output layer. The input parameters governing the iceberg draft were assigned to the input layer, whereas the iceberg draft value as the output parameter was considered as the output layer. The optimal number of the hidden layer neurons was selected based on a trial and error process in this study. It means that the initial number of neurons within the hidden layer was regarded at one and this hyperparameter was incremented one at a time until 35 although the optimum value of the hidden layer neurons was set as 30 once the ELM simulation demonstrated the best performance. The input weights were determined randomly though the output weights were analytically obtained by using a predefined training procedure; as a result, the training process speed of the ELM algorithm was fast. The weights and biases of the input parameters in the ELM model were nonlinearly transformed along with an activation function, where the sigmoid was chosen for the ELM structure since this activation function had a better capability in comparison with its counterparts. It should be noted that the coefficients connecting the neurons in the hidden layer to the neurons in the output layer were calculated by minimizing the error using the Moore-Penrose generalized inverse matrix.

7.1.6.3. Self-Adaptive Extreme Learning Machine (Sa-ELM)

Although the ELM model possessed a reasonable performance; however, the manual selection of the optimal number of hidden neurons might affect the simulation results. As another downside of the ELM model, the input parameters ought to be dimensionless in practice (Azimi and Shiri 2021a). To solve these issues, the ELM and differential evolution (DE) algorithms were combined, and then the Sa-ELM algorithm was developed to simulate the iceberg draft in the current investigation. The DE, as a powerful optimization algorithm, had a high convergence speed and an automatic detection-derivation mechanism. The DE algorithm uses four main steps comprising initialization, mutation, crossover, and selection to simulate the iceberg drafts (Storn and Price 1997). The Sa-ELM flowchart applied in the present study is shown in Figure 7-10. In the Sa-ELM approach, the number of hidden neurons was optimized by adopting the self-adaptive differential evolution algorithm. In the training process, the control coefficients in the DE algorithm and the strategies for trial vector generation were adjusted by a self-adaptive method, meaning that the output weights were determined using the ELM algorithm but the input weights and biases of the hidden neurons were tuned utilizing the DE tool. The Sa-ELM algorithm simulated the iceberg drafts using the major phases, including the initialization, output weights, mutation and crossover, and evaluation (Azimi and Shiri 2021a).

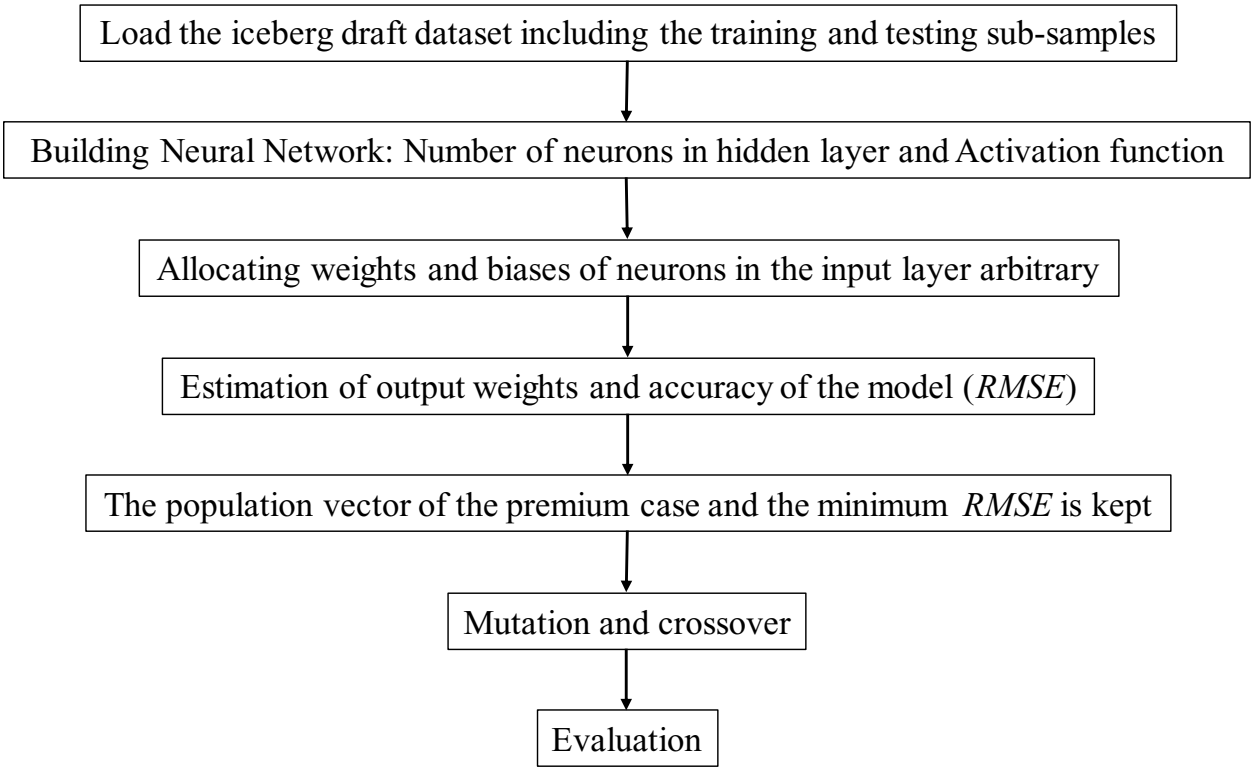


Figure 7-10. Sa-ELM flowchart applied in the present study

7.1.6.4. Iceberg draft

The iceberg draft (D) was assumed as a function of the physical characteristics of the iceberg, comprising the iceberg length (L), iceberg height (H), iceberg width (w), iceberg mass (M) in several fields, analytical, and numerical studies in the form below (Barker et al. 2004; McKenna et al. 2019; and Stuckey et al. 2021):

$$D = f_1(L, H, w, M). \tag{7-47}$$

Furthermore, the density of an iceberg (ρ_i), the density of seawater (ρ_{sw}), seawater viscosity (μ_{sw}), and gravitational acceleration (g) may influence the iceberg draft as follows:

$$D = f_2(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g) \tag{7-48}$$

The iceberg shape factor (S_f) signifies the global shape of icebergs, which can affect the magnitude of the iceberg draft (Turnbull et al. 2018). The shape factor of the traveling icebergs is considered universally into six categories as illustrated in Figure 7-11 (Rudkin 2014). Hence, equation (7-48) can be stated as below:

$$D = f_3\left(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g, S_f\right) \tag{7-49}$$

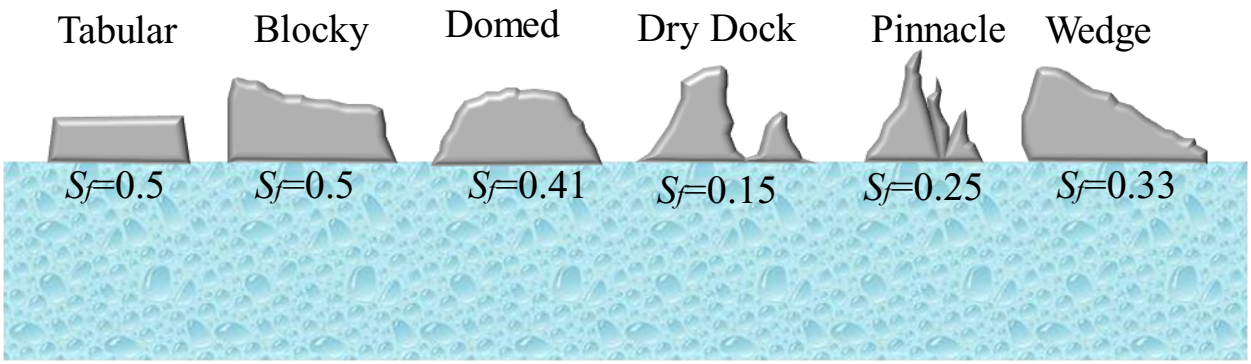


Figure 7-11. Shape factors for different icebergs

It is assumed that the density along with the viscosity of the seawater is constant and the value of gravitational acceleration can be regarded as a constant value; as a result, equation (7-49) can be rewritten as follows:

$$D = f_4\left(L, H, w, M, \rho_i, S_f\right). \tag{7-50}$$

The dimensional form of equation (7-50) is as below:

$$D = f_5(\Pi_1, \Pi_2, \Pi_3, \Pi_4) \tag{7-51}$$

here, Π_1, Π_2,…, Π_4 are dimensionless groups and f_5 is a functional symbol based on the Buckingham-π theorem. Thus, the dimensionless groups below can be calculated as:

$$\Pi_1 = \frac{L}{H} \tag{7-52}$$

$$\Pi_2 = \frac{w}{H} \tag{7-53}$$

$$\Pi_3 = \frac{M}{\rho_i . H^3} \tag{7-54}$$

$$\Pi_4 = S_f \tag{7-55}$$

Equation (7-51) is then formulated as a function of four dimensionless groups as follows:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \tag{7-52}$$

Therefore, D/H as the iceberg draft ratio is a function of the length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$. Subsequently, the ML models applied in the current investigations were fed with the input parameters in equation (7-52).

7.1.6.5. Construction of Dataset

Several field observations were adopted to analyze the iceberg draft. The key values of 12 field studies applied in the present work are tabulated in Table 7-4.

Table 7-4. The studies applied in the present work

No.	Study	Number of reported cases	Type of Study
1	El-Tahan et al. (1985)	38	Documentation
2	Woodworth-Lynas et al. (1985)	one	Analytical
3	Løset and Carstens (1996)	52	Field
4	Barker et al. (2004)	14	Analytical
5	McKenna (2004)	2	Analytical

6	Sonnichsen et al. (2006)	9	Documentation & analytical
7	Turnbull et al. (2015)	2	Analytical
8	McGuire et al. (2016)	8	Field
9	Younan et al. (2016)	29	Field
10	Talimi et al. (2016)	1	CFD
11	Zhou (2017)	3	Autonomous Underwater Vehicle (AUV)
12	Turnbull et al. (2018)	2	Field & numerical

Table 7-5 also tabulates a summary of the key parameters employed in the present study comprising the maximum, minimum, average, variance, and standard deviation of the field observations. To demonstrate whether a population is statistically significant, the t-test is performed and then the P-value is calculated. The t-test has a probability called P-value stemming from the sample of the population that happened by chance. The P-values range from 0% to 100% and the low P-values signify the applied dataset has not happened by chance. The T-test and the P-value for the dataset were calculated, presuming that the P-value of 0.05 or less is statistically significant (Azimi and Shiri 2020b), where the likelihood of the relationship between the observed values is influenced by an alternative hypothesis. This P-value for the constructed dataset was estimated as 0.008, representing that the correlations were statistically significant.

Table 7-5. Summary of the key parameters employed in the present study

Value	D(m)	H(m)	w(m)	M(kg)	L(m)	Sf

Max	240	94	408	1.13E+13	499	0.58
Min	18	3.9	9.672	8317400	34	0.11
Ave	88.988	30.956	101.605	1.68E+11	144.522	0.298
Var.	1696.272	399.389	3733.368	1.14E+24	6546.271	0.011
Std.	41.186	19.985	61.101	1.07E+12	80.909	0.107

Hence, the dimensionless parameters in Eq. (16-10) were applied to estimate the iceberg draft ratio through the ML models. Figure 7-12 demonstrates the combinations of four dimensionless groups introduced to define the ML models. As seen, to introduce the premium Sa-ELM models and the most influencing input parameters, nine Sa-ELM models, including Sa-ELM 1 to Sa-ELM 9 were defined. The Sa-ELM 1 model included all dimensionless groups, whilst these dimensionless groups were disregarded one at a time in Sa-ELM 2 to Sa-ELM 5 models. It is worth noting that the Sa-ELM 6 to Sa-ELM 9 models were function of solely one input parameter.

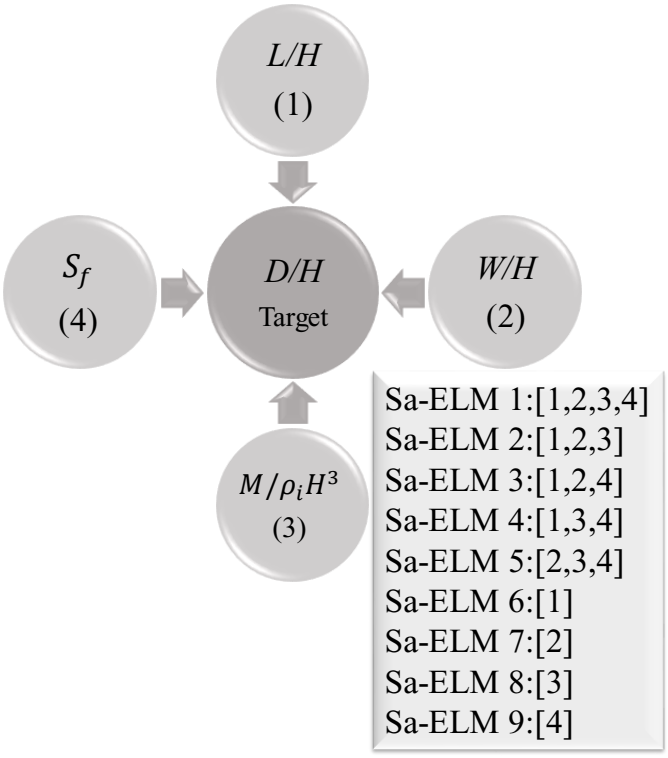


Figure 7-12. Combination of input parameters to develop the Sa-ELM models

It is worth noting that the constructed database contains 161 case studies, where 97 cases (i.e., 60%) were randomly separated for the training and 64 cases (i.e., 40%) were randomly selected for testing of the models. A similar size dataset has been successfully used in the literature to simulate the regression problems with sufficient reliability for training and testing the ML models (e.g., Ebtehaj et al., 2015; Azimi et al., 2018; Walton et al., 2019). Also, the results of a "Goodness to fit" analysis that will be further discussed in the next section have shown the sufficiency of the size of the constructed database.

7.1.6.6. Goodness of fit

In this study, a set of criteria such as correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*),

and Akaike Information Criteria (*AIC*) were utilized to evaluate the precision, correlation, and complexity of the Sa-ELM models.

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{7-53}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{7-54}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{7-55}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{7-56}$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \tag{7-57}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{7-58}$$

where, *Oi, Pi*, $\bar{O}$, $\bar{P}$, *n*, and *k* are respectively the observational value, the predicted amount, the average observational values, the average predicted amount, the number of observations, and the number of independent variables in the Sa-ELM models. The proximity of the *R* and *WI* criteria to one showing the Sa-ELM model tended to have a high degree of correlation with the observed values. The closeness of the *RMSE*, *MAPE*, and *CRM* indices to zero representing the Sa-ELM model possessed a high degree of accuracy; however, the complexity of the Sa-ELM models was not examined through the indices applied. To overcome this restriction, the Akaike Information Criteria (*AIC*) was used. In other words, the less complex Sa-ELM model had the lowest amount of *AIC*, so the best Sa-ELM model might have the lowest degree of *AIC* index and error (*RMSE*, *MAPE*, and *CRM*), with the highest amount of correlation (*R* and *WI*).

7.1.7. Results and discussion

Through the implementation of several defined scenarios, the optimal number of hidden layer neurons and the best activation function within the Sa-ELM architecture has been selected in this study. In the next step, several analyses, such as sensitivity analysis, error analysis, and uncertainty analysis are carried out to detect the superior Sa-ELM model and the most effective input parameters. For this purpose, the results of the NN-based models in the testing mode are analyzed. The best Sa-ELM model is compared with the ELM and ANN algorithms. Eventually, a Sa-ELM-based equation is presented to estimate the iceberg draft.

7.1.7.1. Number of hidden neurons (NHN)

The performance of the neural network-based models alters by changing the number of neurons within the hidden layer, signifying that finding the optimum number of hidden layer neurons (NHN) plays an important role in the precision and complexity of the applied neural-network-based algorithm (Sattar et al. 2019 and Azimi and Shiri 2021a). The variation of the Sa-ELM algorithm performance regarding the NHN parameter against the determined accuracy assessment indices is depicted in Figure 7-13. In the present study, the NHN was initially taken as one and this parameter was increased to eight in each step.

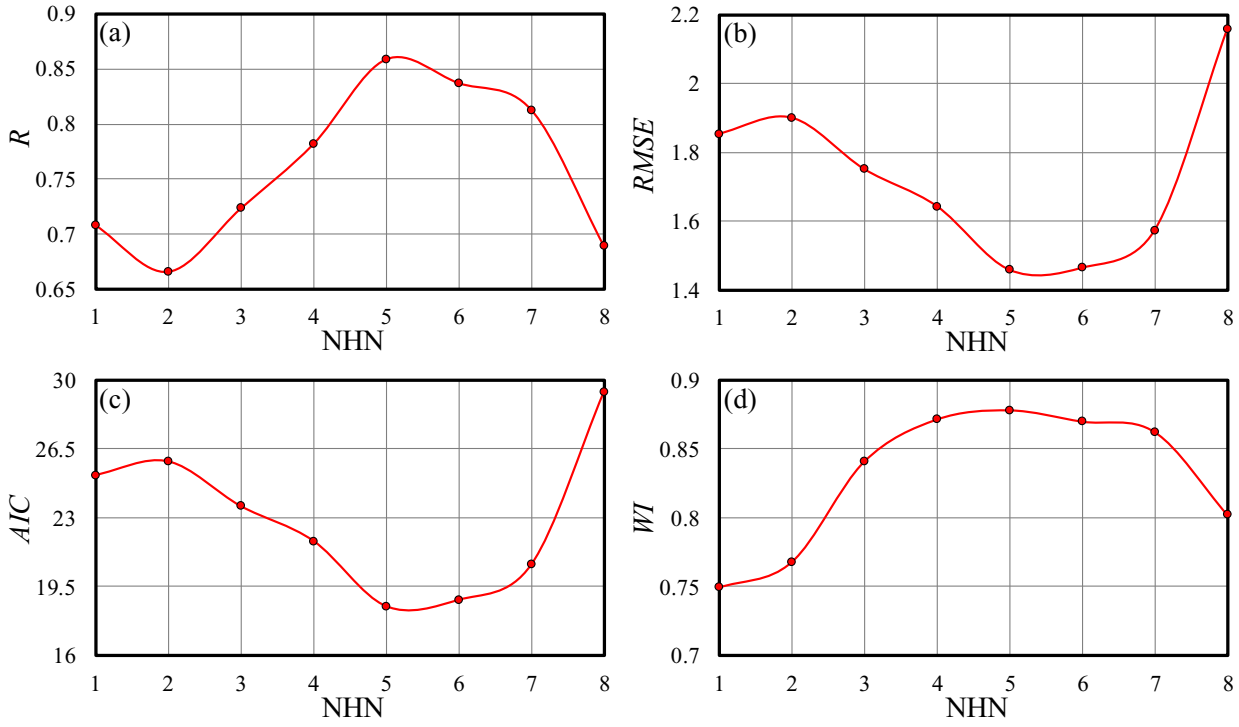


Figure 7-13. Variation of the Sa-ELM algorithm performance regarding the NHN parameter against the statistical criteria calculated (a) R (b) RMSE (c) AIC (d) WI

As shown in Figure 7-13, when the NHN was assumed as one, the *R*, *WI*, and *AIC* indices were obtained at 0.708, 0.749, and 25.159, respectively. The best performance of the Sa-ELM model was achieved once the NHN value was set to five, with the *AIC* and *R* criteria of 18.488 and 0.858. By increasing the NHN value, the accuracy and correlation of the Sa-ELM model have been significantly decreased. For instance, the value of *RMSE* and *WI* indices were computed to be 2.159 and 0.802 for the Sa-ELM model with eight neurons in its hidden layer. As such, the optimal NHN in the Sa-ELM structure for the simulation of the iceberg draft was set to 5 in this study.

7.1.7.2. Activation functions

Using a nonlinear transformation, both the input variables (e.g., weights and biases) are conveyed through an activation function in the neural network-based architecture (Hertz 2018). In the Sa-ELM structure applied, six activation functions, including the hyperbolic tangent (Tanh), the triangular basis (Tribas), the radial basis (Radbas), the hard limit (Hardlim), the sigmoid (Sig), and the sinusoidal (Sin) were examined for the simulation of the iceberg draft. Figure 7-14 displays the comparison between the applied activation functions in this study.

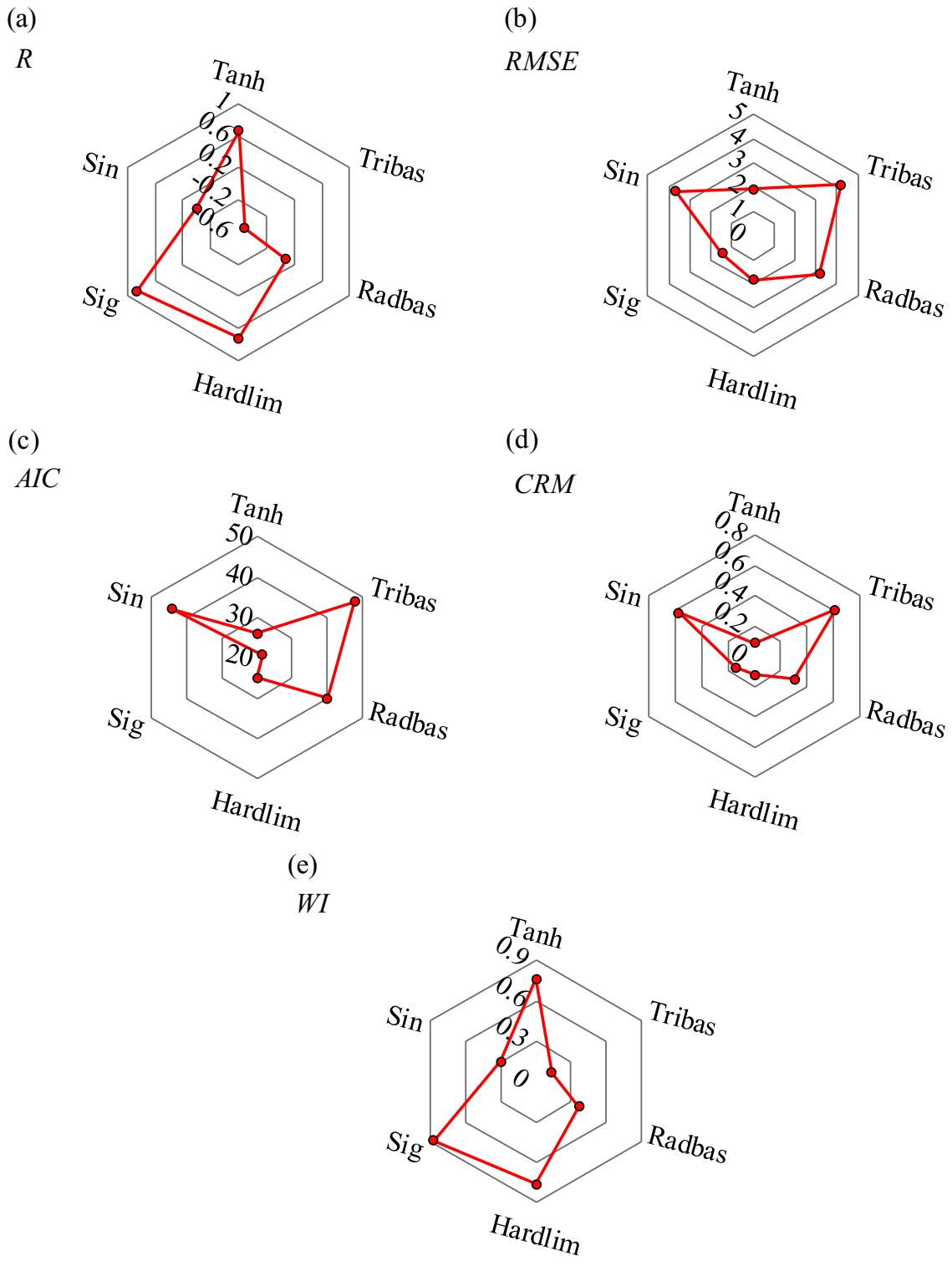


Figure 7-14. Comparison between the applied activation functions (a) R (b) RMSE (c) AIC (d) CRM (e) WI

The value of the *AIC* index for the Tanh, Sin, and Radbas functions was estimated at 26.037, 44.187, and 40.119, respectively. As seen in Figure 7-14, the Sig activation function presented the best performance for modeling the iceberg draft (R=0.858, WI=0.878, and AIC=18.488). Regarding the simulation results, the Tribas function represented the worst performance in order to estimate the iceberg draft, with the *RMSE* calculated at 4.200. Hence, the Sig was selected as the best activation function to predict the target parameter in the current work.

7.1.7.3. Sensitivity analysis

Figure 7-15 depicts the comparison between the statistical indices calculated for Sa-ELM 1 to Sa-ELM 9. The Sa-ELM 1 model was a function of all input parameters, and the influence of each input was disregarded in the Sa-ELM 2 to Sa-ELM 5 models one at a time. The superior Sa-ELM model in the estimation of the iceberg draft was identified by the evaluation of the Sa-ELM 2 to Sa-ELM 5 models. In order to determine the effectiveness level of each input parameter in the prediction of the iceberg draft (D/H), sensitivity analysis was conducted by feeding only one input parameter into the Sa-ELM 6, Sa-ELM 7, Sa-ELM 8, and Sa-ELM 9 models, i.e., L/H, w/H, $M/\rho_i.H^3$, and S_f, respectively.

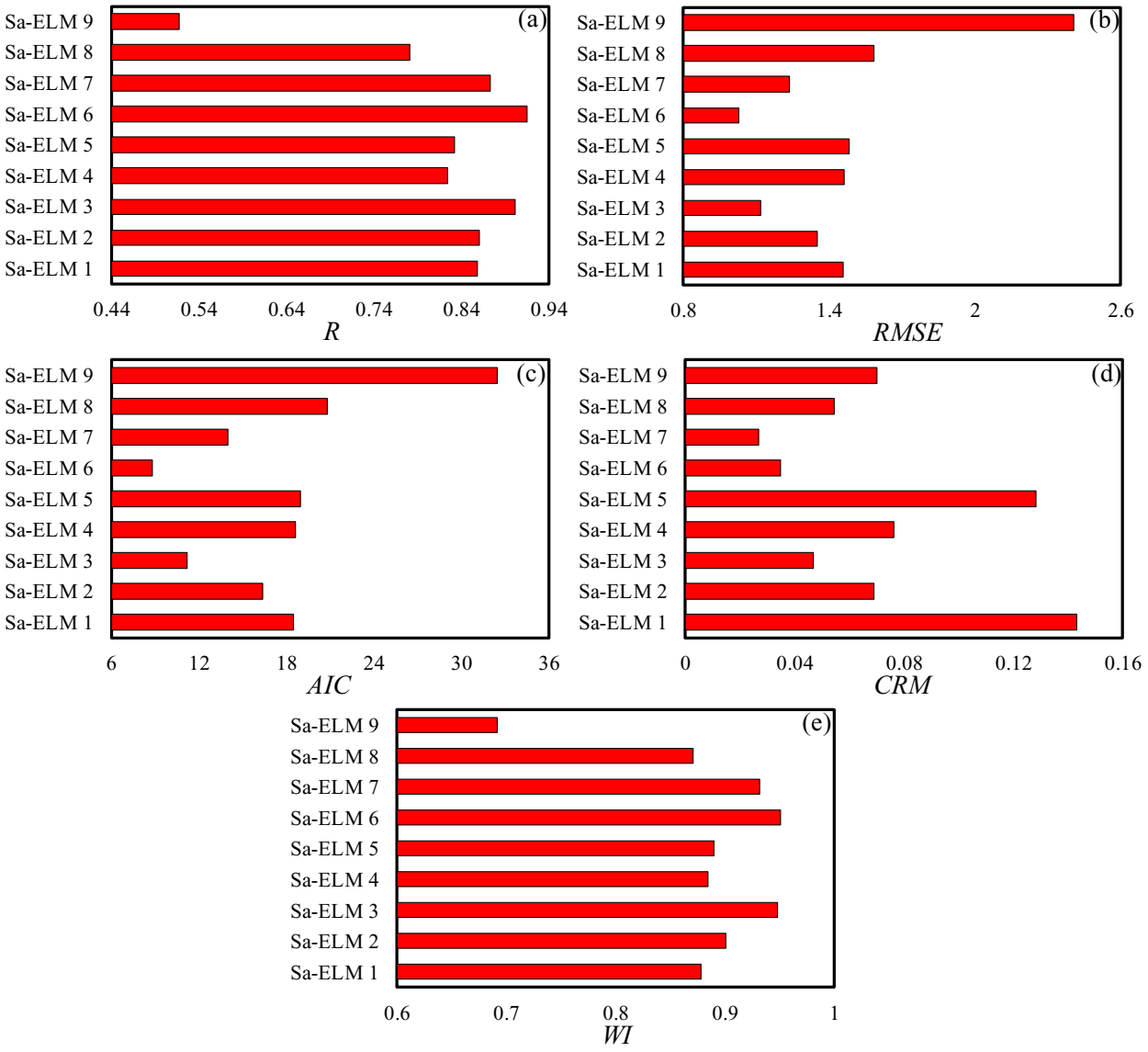


Figure 7-15. Comparison between the statistical indices calculated for the Sa-ELM 1 to Sa-ELM 9 (a) R (b) RMSE (c) AIC (d) CRM (e) WI

The scatter plots of the Sa-ELM 1 to Sa-ELM 9 are displayed in Figure 7-16. According to the simulation results, the Sa-ELM 3 model estimated the iceberg draft with a low degree of complexity along with a high level of correlation and accuracy, meaning that the value *AIC*, *R*, and *RMSE* indices for the Sa-ELM 3 were obtained as 11.155, 0.902, and 1.120, respectively. In

addition, the iceberg draft values were predicted using the L/H , w/H , and S_f inputs by the Sa-ELM 3 model, and the impact of iceberg mass ratio ($M/\rho_i.H^3$) was eliminated for this model. The sensitivity analysis demonstrated that Sa-ELM 2 was the second-best model for the simulation of iceberg draft, with an *RMSE* of 1.351. For Sa-ELM 2, the effect of the iceberg shape factor (S_f) was ignored and this model predicted the target parameter by means of L/H , w/H , and $M/\rho_i.H^3$. The Sa-ELM 1 model, the third-best model, forecasted the iceberg draft by using all inputs. For the Sa-ELM 1 model, the *WI* and *CRM* values were equal to 0.878 and 0.143, respectively. The influence of the iceberg width ratio (w/H) was dropped for the fourth-best model (Sa-ELM 4), with an *AIC* computed of 18.601. Although Sa-ELM 5 modeled the iceberg draft through w/H , $M/\rho_i.H^3$, and S_f, the effectiveness of the iceberg length ratio (L/H) was removed, where the value of RMSE and R criteria for this model was respectively calculated at 1.482 and 0.832.

Amongst the Sa-ELM 6 to Sa-ELM 9 models, the Sa-ELM 6 model, which is a function of L/H, has a better performance compared to other models. As a result, the iceberg length ratio was the most important input parameter to simulate the iceberg draft. It is worth noting that the overwhelming majority of studies estimating iceberg drafts through monitoring the above-water features of icebergs have used the iceberg length as the first influential parameter to predict the draft (Hotzel and Miller 1983; C-CORE 2001; Barker et al. 2004; Sacchetti et al. 2012; Stuckey et al. 2021). In the current study, the iceberg width ratio (w/H) and the iceberg mass ratio ($M/\rho_i.H^3$) were identified as the second and third most influential parameters for determining the iceberg draft. This is in agreement with several investigations in the literature that have revealed that the iceberg width and iceberg mass are amongst the key governing parameters in the iceberg draft estimation (Mognor and Zorn 1979; King et al. 2016).

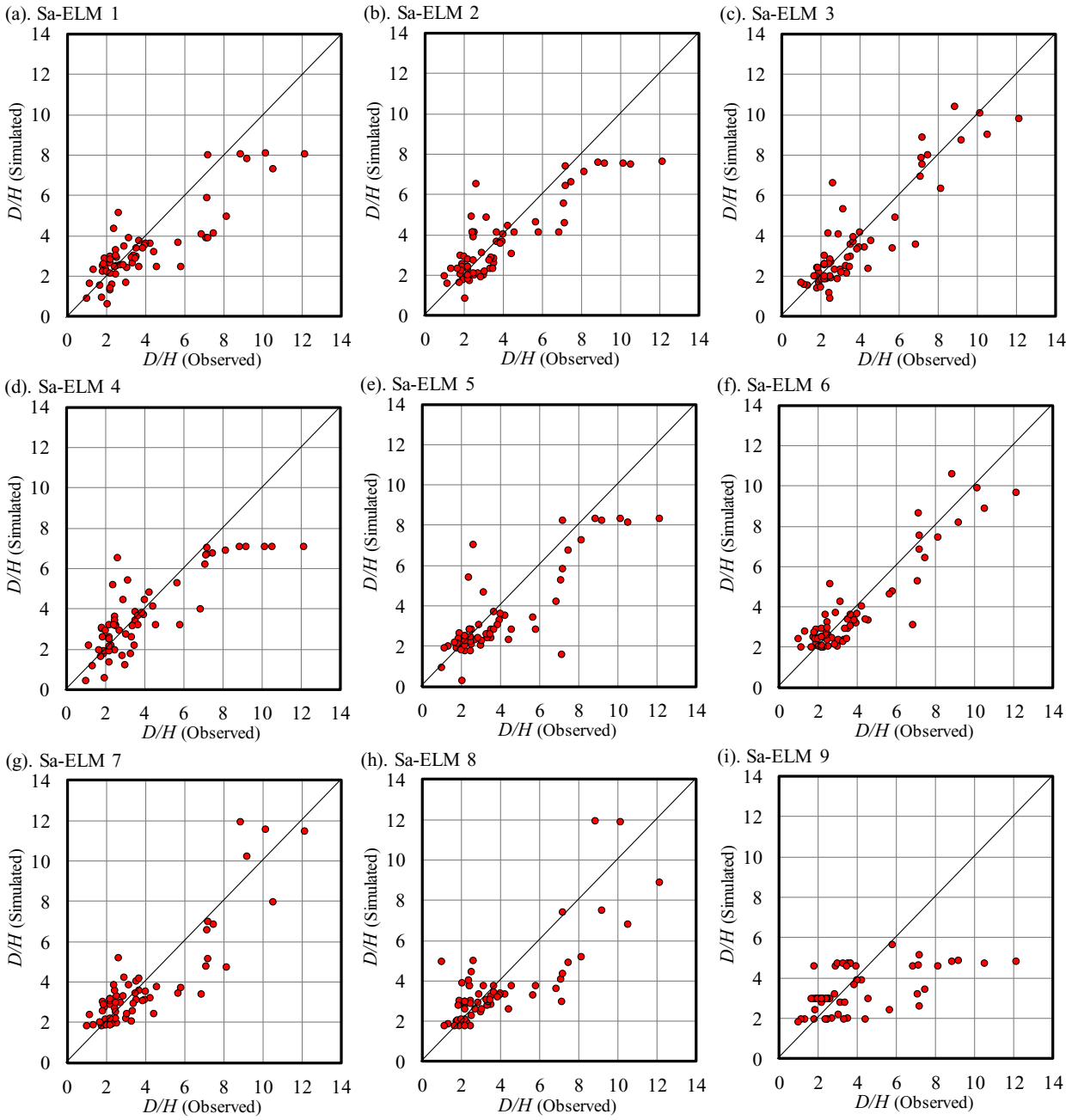


Figure 7-16. Scatter Plots of Sa-ELM 1 to Sa-ELM-9

Ultimately, the performed sensitivity analysis highlighted that the influence of the iceberg shape factor (*Sf*) in the prediction of the draft was insignificant. These findings are in agreement with physical fundamentals. The length and width are the dimensions both contributing to the iceberg's

volume, mass, weight, and draft, consequently. These dimensions can be used for draft estimation by aircraft, vessels, and satellites. The shape factor is a dimensionless parameter, which depends on the simplified shape of the iceberg and the governing convective heat transfer law applicable across the thin boundary layer over the keel, which is more suitable for predominantly flat or axisymmetric icebergs (Shipilova et al. 2022). The shape factor for irregular icebergs that are widely found in the hired database carries large uncertainties that make it insignificant in the determination of the draft. As the shape factors are further developed in the literature and incorporate more complicated characteristics such as flow separation effects, the current type of studies is suggested to be re-conducted to re-assess the significance of this parameter.

7.1.7.4. Error analysis

Table 7-6 shows the error analysis results for the Sa-ELM models. In this table, the percentage cumulative errors from 2% to 20% for the Sa-ELM models are tabulated. Regarding this error analysis, almost a quarter of iceberg drafts estimated by the Sa-ELM 3 model possessed an error of less than 12%, while about half of the results from this model showed an error of less than 18%. On the other hand, more than 60% of the Sa-ELM 3 model's outcomes comprised an error of smaller than 20%. As shown, the Sa-ELM 2 and Sa-ELM 4 models demonstrated a similar performance for the estimation of the iceberg draft. For instance, approximately 46% of iceberg drafts predicted by these models had an error of less than 20%. Roughly one-third of the Sa-ELM 1 and Sa-ELM 7 models' results had an error of less than 14%. About 40% of results simulated by the Sa-ELM 5 and Sa-ELM 6 models included an error of smaller than 16%. Based on the error analysis, Sa-ELM 9 displayed the worst performance for modeling the iceberg drafts, where just about one-fifth of the results provided by the Sa-ELM 9 model had an error of less than 20%.

Table 7-6. Results of the percentage cumulative errors from 2% to 20% for the Sa-ELM models (%)

Cumulative error (%)	2	4	6	8	10	12	14	16	18	20
Sa-ELM 1	4.69	7.81	10.94	10.94	15.63	23.44	28.13	34.38	40.63	43.75
Sa-ELM 2	3.13	4.69	6.25	12.50	18.75	28.13	31.25	34.38	42.19	46.88
Sa-ELM 3	4.69	7.81	10.94	17.19	23.44	25.00	29.69	45.31	51.56	60.94
Sa-ELM 4	3.13	6.25	9.38	20.31	26.56	31.25	43.75	45.31	46.88	46.88
Sa-ELM 5	3.13	6.25	9.38	14.06	17.19	26.56	35.94	40.63	46.88	53.13
Sa-ELM 6	3.13	6.25	15.63	18.75	25.00	28.13	37.50	42.19	50.00	57.81
Sa-ELM 7	4.69	10.94	12.50	15.63	20.31	26.56	32.81	40.63	43.75	48.44
Sa-ELM 8	0.00	7.81	10.94	14.06	17.19	20.31	26.56	34.38	43.75	50.00
Sa-ELM 9	0.00	3.13	4.69	4.69	7.81	7.81	12.50	12.50	17.19	21.88

7.1.7.5. Uncertainty analysis

Uncertainty is inherent in machine learning for predictive modeling. There can be many sources of uncertainty in an ML project, including variance in the specific data values, the sample of data collected from the domain, and the imperfect nature of any models developed from such data. Therefore, to examine the efficiency of the Sa-ELM models, an uncertainty analysis was carried out. To do this, errors from the Sa-ELM model (e_j) were calculated as the difference between the iceberg drafts predicted (P_j) and the observed amounts (O_j) as follows:

$$e_j = P_j - O_j \tag{7-59}$$

The mean errors (Mean) and the standard deviation (StDev) values of the errors were obtained by the following equations:

$$\text{Mean} = \frac{1}{n}\sum_{j=1}^{n} e_j \tag{7-60}$$

$$\text{StDev} = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{7-61}$$

A certain Sa-ELM model underestimated the iceberg draft if the sign of the Mean value was negative; however, the positive sign of the Mean reprinted that the Sa-ELM model possessed an overestimated performance on the iceberg draft. In addition, a prediction interval (PI) was generated around the error values using the Mean, StDev values, and the "Wilson score approach" by disregarding the continuity correction. A normal distribution interval rectified as an asymmetric normal distribution, entitled the Wilson score interval, was employed to modify the PI bounds; as a result, a ±1.96Se produced a 95% of PI (95%PI). It is worth mentioning that the width of uncertainty bound (WUB) of the Sa-ELM model was computed as below (Azimi and Shiri 2021b):

$$\text{WUB} = \pm\frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{7-62}$$

Table 7-7 and Figure 7-17 show the uncertainty analysis results from the Sa-ELM 1 to Sa-ELM models. According to the uncertainty analysis performed, all Sa-ELM models overestimated the iceberg drafts. Among the Sa-ELM 1 to Sa-ELM 5 models, Sa-ELM 3 created the narrowest uncertainty bound, whereas the widest uncertainty bound was calculated for the Sa-ELM 4 model.

Table 7-7. Results of uncertainty analysis for Sa-ELM models

Model	Mean	StDev	WUB	95%PI
Sa-ELM 1	0.553	1.360	±0.340	0.213 to 0.893

Sa-ELM 2	0.266	1.335	±0.334	-0.068 to 0.599
Sa-ELM 3	0.181	1.114	±0.279	-0.098 to 0.459
Sa-ELM 4	0.295	1.446	±0.361	-0.066 to 0.656
Sa-ELM 5	0.496	1.408	±0.352	0.144 to 0.847
Sa-ELM 6	0.135	1.027	±0.257	-0.121 to 0.392
Sa-ELM 7	0.104	1.244	±0.311	-0.207 to 0.414
Sa-ELM 8	0.21	1.583	±0.396	-0.186 to 0.605
Sa-ELM 9	0.271	2.413	±0.603	-0.332 to 0.874

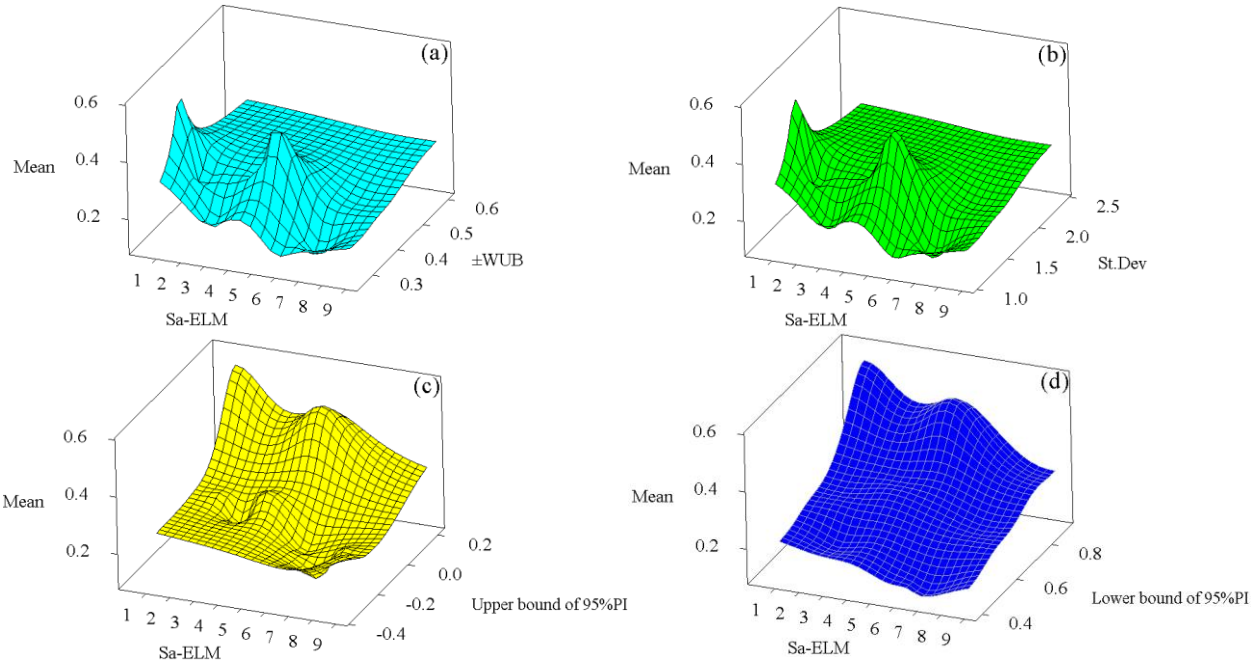


Figure 7-17. Uncertainty analysis results from the Sa-ELM 1 to Sa-ELM 9 models variations of Mean against estimates of WUB (a), St Dev (b), Upper and Lower bounds of 95% PI (c, d)

7.1.7.6. Comparison with ELM and ANN

The capability of Sa-ELM 3, as the best model in the present study, to simulate the iceberg draft has been compared to the proficiency of other neural network-based models comprising the ELM and ANN algorithms. Figure 7-18 compares the performance of the Sa-ELM, ELM, and ANN models in the estimation of the iceberg draft. The value of the correlation coefficient for the Sa-ELM, ELM, and ANN models was calculated at 0.902, 0.896, and 0.746, whilst the *AIC* index for these models equaled 11.116, 13.082, and 25.346, respectively. As a result, the Sa-ELM model had higher precision and correlation along with a lower complexity to predict the iceberg draft. This means that the Sa-ELM method outperformed other neural network-based algorithms for the simulation of the iceberg draft.

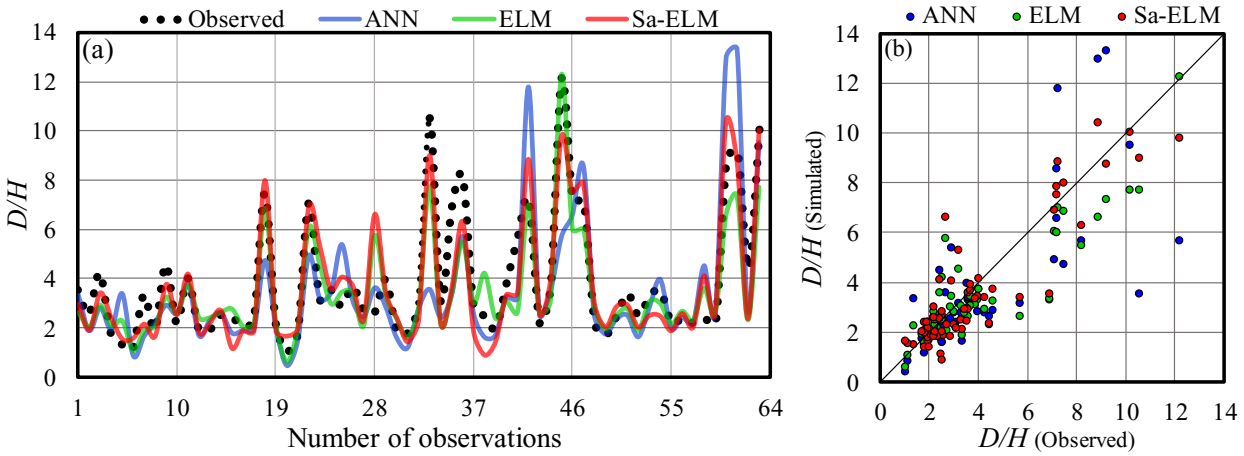


Figure 7-18. (a) Performance comparison of Sa-ELM, ELM, and ANN (b) scatter plot of the Sa-ELM, ELM, and ANN

The general form of the Sa-ELM formula to estimate the target parameter is written as:

$$\eta = \left[\frac{1}{1 + exp(InW \times InV \times BHI)}\right]^{T} \times OutW \tag{7-63}$$

where, the InW, InV, BHI, and $OutW$ are the matrix of input weight, the matrix of input parameters, the matrix of bias of hidden layer, and the matrix of output weights, respectively. In

this study, the Sa-ELM 3, which was a function of the iceberg length ratio (L/H), the iceberg width ratio (w/H), and the iceberg shape factor (Sf) has been recognized as the premium model to predict the iceberg draft. The matrices of Sa-ELM 3 to approximate the iceberg draft is formulated as follows:

$$InV = \begin{bmatrix} L/H \\ w/H \\ Sf \end{bmatrix}, BHI = \begin{bmatrix} 0.260 \\ -0.071 \\ -0.116 \\ 0.483 \\ -0.247 \end{bmatrix}, InW = \begin{bmatrix} 0.444 & -0.492 & 0.694 \\ 0.080 & 0.568 & -0.467 \\ -0.554 & -0.265 & 0.767 \\ -0.465 & 0.216 & 0.574 \\ -0.029 & 0.748 & 0.841 \end{bmatrix}, OutW = \begin{bmatrix} -6.297 \\ -31.396 \\ 40.934 \\ -26.879 \\ 46.468 \end{bmatrix} \quad (7\text{-}64)$$

Therefore, Sa-ELM 3 is better able to simulate the iceberg draft with reasonable proficiency and the highest degree of correlation accuracy compared to other neural network-based methods. The iceberg length ratio (L/H) and iceberg width ratio (w/H) were identified as the most important input parameters to simulate the iceberg draft using the Sa-ELM algorithm.

7.1.8. Conclusion

In the present investigation, the iceberg drafts were modeled by a self-adaptive machine learning algorithm entitled "self-adaptive extreme learning machine" (Sa-ELM). In the first step, the parameters governing the iceberg draft were recognized and nine Sa-ELM models were developed by using these parameters. To train and test these ML models, a comprehensive dataset was constructed, where 60% of the data were employed for the training of the ML models and 40% for validation. An optimized structure for the Sa-ELM algorithm was obtained, meaning that the optimal number of hidden layer neurons (NHN) and the best activation function have been determined for this algorithm. By defining different scenarios, the premium Sa-ELM model and the most important input parameters affecting the iceberg draft were introduced. The proficiency of the Sa-ELM was compared to the ELM and ANN approaches. A Sa-ELM-based equation was

finally presented for the estimation of the iceberg draft for practical applications. The most important findings of this study are summarized as:

- The optimum NHN for the Sa-ELM architecture was chosen at 5 since the applied algorithm demonstrated the best performance.
- The sigmoid function showed an excellent outcome to model the iceberg draft in comparison with other activation functions.
- The Sa-ELM 3 is a function of w/H , $M/\rho_i.H^3$, and S_f input parameters were identified as the best ML model to predict the iceberg draft.
- The sensitivity analysis proved that the iceberg length ratio (*L/H*) and iceberg width ratio (*w/H*) had a remarkable influence to model the iceberg drafts using the Sa-ELM algorithm.
- The performed error analysis indicated that almost 60% of the iceberg draft estimated by the Sa-ELM 3 model had an error of less than 20%.
- Regarding the uncertainty analysis, the Sa-ELM 3 model overestimated the iceberg drafts.
- The Sa-ELM 3 model simulated the iceberg drafts with better performance in comparison with the ELM and ANN methods, signifying that the Sa-ELM 3 had the highest degree of precision, correlation, and simplicity compared to its counterparts.

The superior Sa-ELM model demonstrated a reasonable generalization ability in the iceberg draft estimation. The model was trained and validated through a comprehensive dataset, comprising the most important field studies in this area. Therefore, the results obtained from this model are sufficiently general to apply to other independent draft prediction scenarios.

Although the current work was the first application of the ML advancement for the simulation of the iceberg draft, the provided results exhibited that the ML algorithms can pave the way to present alternative approaches to preserve the time and expenses of iceberg management projects. For future studies, it is recommended to collect more field measurements using the underwater survey facilities to improve the model performance and validate against a wider dataset.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Allaire, P.E., 1972. Stability of simply shaped icebergs. Journal of Canadian Petroleum Technology, 11:04.

Azimi, H., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., Khoshbin, F., 2018. Evolutionary design of generalized group method of data handling-type neural network for estimating the hydraulic jump roller length. Acta Mechanica, 229(3): 1197–1214.

Azimi, H., Shiri, H., 2021a. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., 2021b. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04): 411-420.

Azimi, H., Shiri, H., 2020a. Ice-seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2020b. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., Zendehboudi, S., 2022. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Region Science and Technology, 193: 103426.

Azimi, H., Shiri, H., Zendehboudi, S., 2023. Prediction of ice-induced subgouge soil deformation in sand using group method of data handling–based neural network. Journal of Cold Regions Engineering, 37(2): 04023006.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In The fourteenth international offshore and polar engineering conference, ISOPE-I-04-116.

Bass, D.W., 1980. Stability of icebergs. Annals of Glaciology, 1: 43-47.

Bilhan, O., Emiroglu, M.E., Kisi, O., 2010, Application of two different neural network techniques to lateral outflow over rectangular side weirs located on a straight channel. Advanced Engineering Software, 41(6): 831-837.

Brooks, L.D., 1980. Another hypothesis about iceberg draft. In POAC 79, Proceedings of the 5th International Conference on Port and Ocean Engineering under Arctic Conditions, Vol. 1, pp. 241-252.

C-CORE., 2001. Documentation of iceberg grounding events from the 2000 season. C-CORE publication 01-C10 (Revision 0). Report submitted to the Geological Survey of Canada, Atlantic.

Coast Guard, 1991 Report of the International Ice Patrol in the North Atlantic. Bulletin, (77).

Dowdeswell, J.A., Bamber, J.L., 2007. Keel depths of modern Antarctic icebergs and implications for sea-floor scouring in the geological record. Marine Geology, 243(1-4): 120-131.

Ebtehaj, I., Bonakdari, H, Zaji AH, Azimi H, Khoshbin F., 2015. GMDH-type neural network approach for modeling the discharge coefficient of rectangular sharp-crested side weirs. Engineering Science and Technology, an International Journal, 18(4): 746–757.

El-Tahan, M., El-Tahan, H., Courage, D., Mitten, P., 1985. Documentation of iceberg groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of machine learning research, 3: 1157-1182.

Hertz, J.A., 2018. Introduction to the Theory of Neural Computation. CRC Press.

Hotzel, I.S., Miller, J.D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

Huang, G.B., Zhu, Q.Y., Siew, C.K., 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. In 2004 IEEE international joint conference on neural networks, IEEE Cat. No. 04CH37541, Vol. 2, pp. 985-990.

King, T., 2012. Iceberg interaction frequency model for subsea structures. Ocean Technology Conference, OTC-23787-MS.

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Ocean Technology Conference, OTC-27358-MS.

Liang, B., 2001. Iceberg Stability and Deterioration. Doctoral dissertation, Memorial University of Newfoundland, NL, Canada.

Løset. S., Carstens, T., 1996. Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Region Science and Technology, 24(4): 323-340.

Ma, L., Resvanis, T.L., Vandiver, J.K., 2020. Using machine learning to identify important parameters for flow-induced vibration. In the 39th International Conference on Ocean, Offshore and Arctic Engineering, 84355:V004T04A011.

Mahdianpari, M., Homayouni, S., Foucher, S., 2021. CJRS' special issue on deep learning for environmental applications of remote sensing data. Canadian Journal of Remote Sensing, 47(2): 159-161.

McGuire. P., Younan, A., Wang, Y., Bruce, J., Gandi, M., King, T., Regular, K., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., 2009. Modelling iceberg shape, mass and draft changes. In the 20th International Conference on Port and Ocean Engineering under Arctic Conditions (POAC'09), No. POAC09-107.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. In Proceedings of the 25th International Conference on Port and Ocean Engineering under Arctic Conditions, pp. 9-19.

Mognor, K., Zorn, R., 1979. Discussion of another hypothesis about iceberg draft. In the 3rd International Conference on Port and Ocean Engineering Under Arctic Conditions, POAC, Fairbanks, Alaska, USA.

Shipilova, O., Olsson, A., Thomson, M., 2022. On the shape factor in iceberg deterioration by forced convection. Cold Region Science and Technology, 195: 103472.

Robe, R.Q., Farmer, L.D., 1976. Physical properties of icebergs. Part I. height to draft ratios of icebergs. Part II. mass estimation of Arctic icebergs. Coast Guard Research and Development Center Groton Conn.

Rudkin, P., 2014. Comprehensive iceberg management database report 2005 update. PERD/CHC Report 20-72. National Research Council of Canada (NRC) and Panel on Energy Research and Development (PERD), 2005, (March 2005).

Sacchetti, F., Benetti, S., Cofaigh, C.Ó., Georgiopoulou, A., 2012. Geophysical evidence of deep-keeled icebergs on the Rockall Bank, Northeast Atlantic Ocean. Geomorphology, 159:63-72.

Sattar, A., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. Neural Computing and Applications, 31(1): 157-169.

Schulman, P.R., 2022. Reliability, uncertainty and the management of error: New perspectives in the COVID-19 era. Journal of Contingencies and Crisis Management, 30(1): 92-101.

Sen, D., 2014. The uncertainty relations in quantum mechanics. Current Science, 203-218.

Shahin, M.A., Jaksa, M.B., Maier, H.R., 2001, Artificial neural network applications in geotechnical engineering. Australian Geomechanics, 36(1): 49-62.

Sonnichsen, G., Hundert, T., Myers, R., Pocklington, P., 2006. Documentation of recent iceberg grounding evens and a comparison with older events of know age. Environmental Studies Research Funds. Report ESRF Vol. 157.

Storn, R., Price, K., 1997. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 11(4): 341-359.

Stuckey, P.D., 2008. Drift Speed Distribution of Icebergs on the Grand Banks and Influence on Design Loads. Doctoral dissertation, Memorial University of Newfoundland, NL, Canada.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., King, T., 2021. Iceberg load software update using 2019 iceberg profile dataset. 40th International Conference on Ocean, Offshore and Arctic Engineering, 85178: V007T07A018.

Talimi, V., Ni, S., Qiu, W., Fuglem, M., MacNeill, A., Younan, A., 2016. Investigation of iceberg hydrodynamics. Arctic Technology Conference, OTC-27493-MS.

Turnbull, I.D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Region Science Technology, 110: 1-18.

Turnbull, I.D., King, T., Ralph, F., 2018. Development of a new operational iceberg drift forecast model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Walton, R., Binns, A., Bonakdari, H., Ebtehaj, I., Gharabaghi, B., 2019. Estimating 2-year flood flows using the generalized structure of the group method of data handling. Journal of Hydrology, 575: 671–689.

Woodworth-Lynas, C.M.T., Simms, A., Rendell, C.M., 1985. Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Region Science and Technology, 10(2): 163-186.

Younan, A., Ralph, F., Ralph, T., Bruce, J., 2016. Overview of the 2012 iceberg profiling program. Arctic Technology Conference, OTC-27469-MS.

Zhou, M., 2017. Underwater Iceberg Profiling and Motion Estimation using Autonomous Underwater Vehicles, Doctoral dissertation, Memorial University of Newfoundland, NL, Canada.

8. Chapter 8

Prediction of Iceberg Draft using ML algorithms

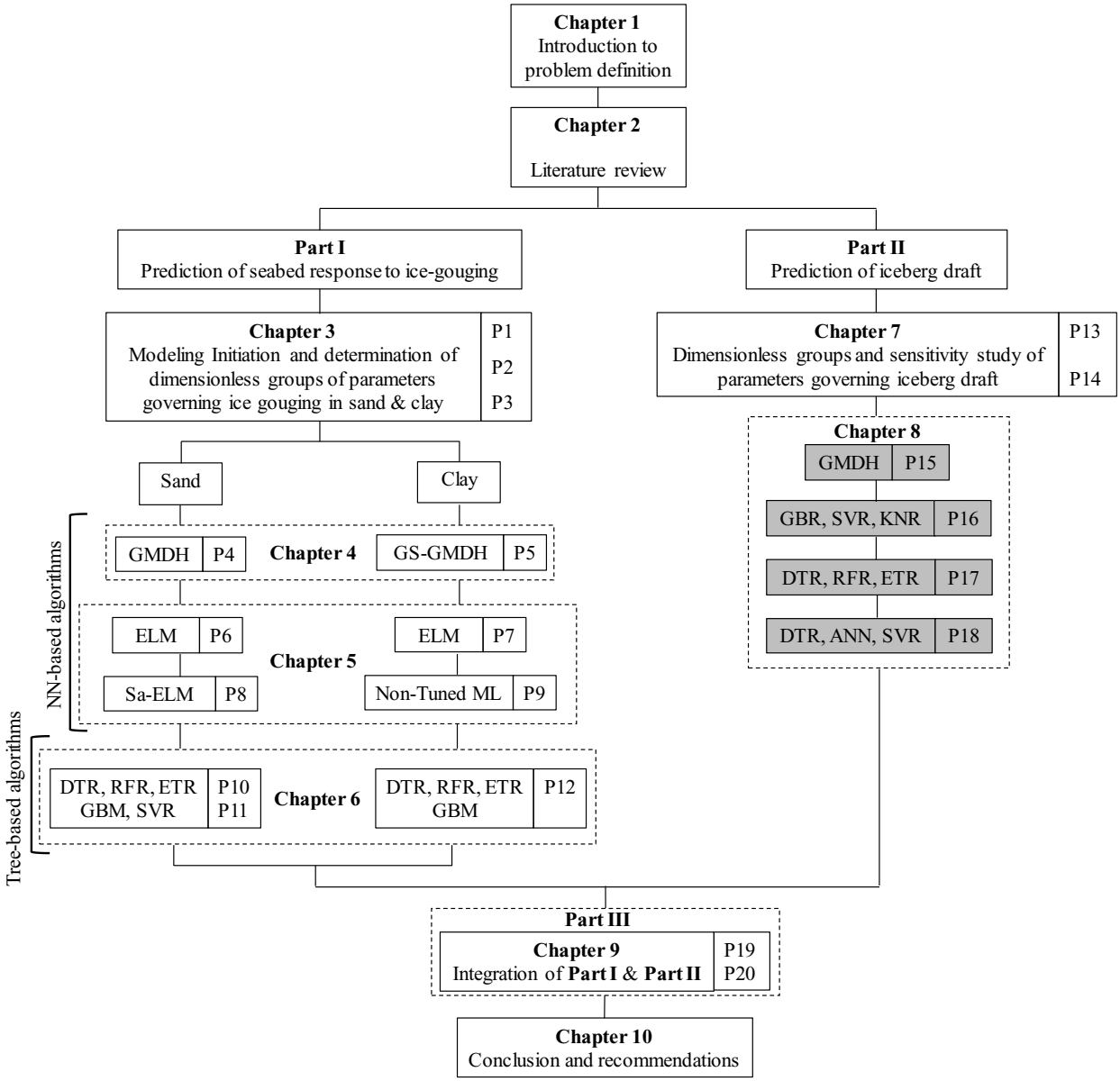
This chapter includes two sections as follows:

Section 1: Generalized Structure of the Group Method of Data Handling for Modeling Iceberg Drafts

Section 2: Iceberg Draft Prediction using Gradient Boosting Regression Algorithm

Section 3: Iceberg Draft Evaluation through Robust Tree-Based Machine Learning Algorithms

Section 4: Iceberg Drafts Assessment using Decision Tree Regression (DTR), Artificial Neural Network (ANN), and Support Vector Regression (SVR) algorithms


Chapter 1
Introduction to problem definition
Chapter 2
Literature review
Part I
Prediction of seabed response to ice-gouging
Part II
Prediction of iceberg draft
Chapter 3
Modeling Initiation and determination of dimensionless groups of parameters governing ice gouging in sand & clay
P1
P2
P3
Chapter 7
Dimensionless groups and sensitivity study of parameters governing iceberg draft
P13
P14
Chapter 8
GMDH P15
GBR, SVR, KNR P16
DTR, RFR, ETR P17
DTR, ANN, SVR P18
Sand
Clay
NN-based algorithms
GMDH P4
Chapter 4
GS-GMDH P5
ELM P6
ELM P7
Chapter 5
Sa-ELM P8
Non-Tuned ML P9
Tree-based algorithms
DTR, RFR, ETR GBM, SVR P10 P11
Chapter 6
DTR, RFR, ETR GBM P12
Part III
Chapter 9
Integration of Part I & Part II
P19
P20
Chapter 10
Conclusion and recommendations

Section 1

Generalized Structure of the Group Method of Data Handling for Modeling Iceberg Drafts

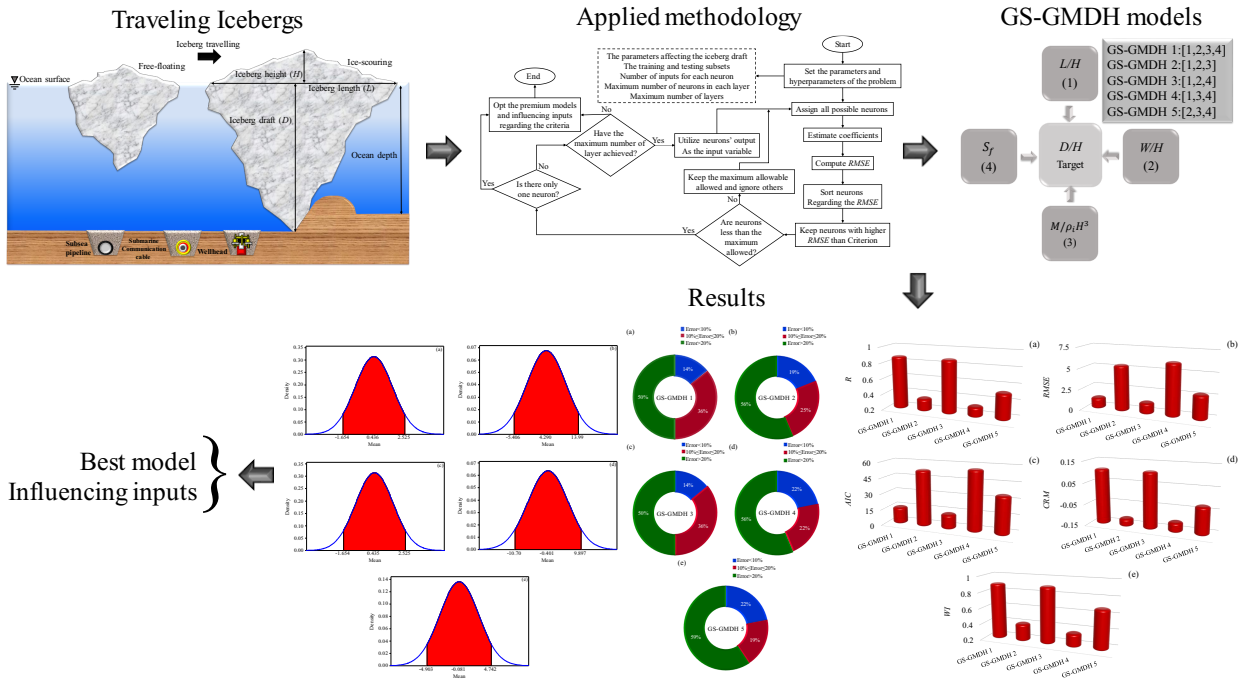
This section was submitted as a journal manuscript

Abstract

The iceberg draft prediction is vital to mitigate the collision risk of deep keel icebergs with the seafloor-founded infrastructures, including the subsea pipelines, wellheads, hydrocarbon loading equipment, and communication cables crossing the Arctic areas (e.g., the Grand Bank of Canada, Beaufort Sea, or the Caspian Sea), since the traveling icebergs may gouge the ocean floor and the physical and operational integrity of the submarine structures would be threatened. In this study, the iceberg drafts were simulated using the generalized structure of the group method of data handling (GS-GMDH) algorithm for the first time. The parameters affecting the iceberg drafts were determined, and five GS-GMDH models comprising GS-GMDH 1 to GS-GMDH 5 were developed utilizing those parameters governing. A dataset was generated, and the GS-GMDH models were trained through 60% of the data, the rest of the data (i.e., 40%) were considered for the GS-GMDH models' validation. By defining different scenarios, the most accurate GS-GMDH model and the most important input parameters were identified. The sensitivity analysis demonstrated that the iceberg width ratio (w/H) and the iceberg shape factor (S_f) were identified as the most influencing input parameters. The comparison between the performance of the premium GS-GMDH model and the group method of data handling (GMDH), artificial neural network (ANN) algorithms, and the empirical models proved that the GS-GMDH model simulated the iceberg drafts with the highest level of precision and correlation along with the lowest degree of complexity. Based on the partial derivative sensitivity analysis (PDSA), the magnitude of iceberg drafts grew by increasing the value of the iceberg width and iceberg length. Ultimately, a GS-GMDH-based equation was presented to estimate the iceberg drafts for practical applications, particularly in the early stages of iceberg management projects and engineering designs.

Keywords: Iceberg Draft, Generalized structure of group method of data handling (GS-GMDH), Subsea assets, Machine learning, Artificial neural network (ANN), Partial derivative sensitivity analysis (PDSA)

Graphical abstract



8.1.1. Introduction

Although the Arctic region has provided approximately 25% of the natural gas and 10% of the crude oil across the world, about 22% of the planet's undiscovered hydrocarbons are located in this ice-prone area. (WWF Arctic Programme). The recent offshore oil and gas loading facilities (e.g., subsea pipelines, wellheads, and communication cables) developed in the Arctic area have led to a considerable awareness of the iceberg draft approximation, where deep keel icebergs may plow the ocean floor; as a result, the operational integrity of these submarine infrastructures may be threatened in the shallower waters (Bruce et al. 2016; Nematzadeh and Shiri 2020). The iceberg free-floating and iceberg scouring in the Arctic waters are schematically displayed in Figure 8-1. As seen, if the ocean depth is deeper than the iceberg draft, the traveling iceberg is in free-floating condition; however, the iceberg tip would gouge the seafloor, which is called the ice-gouging problem.

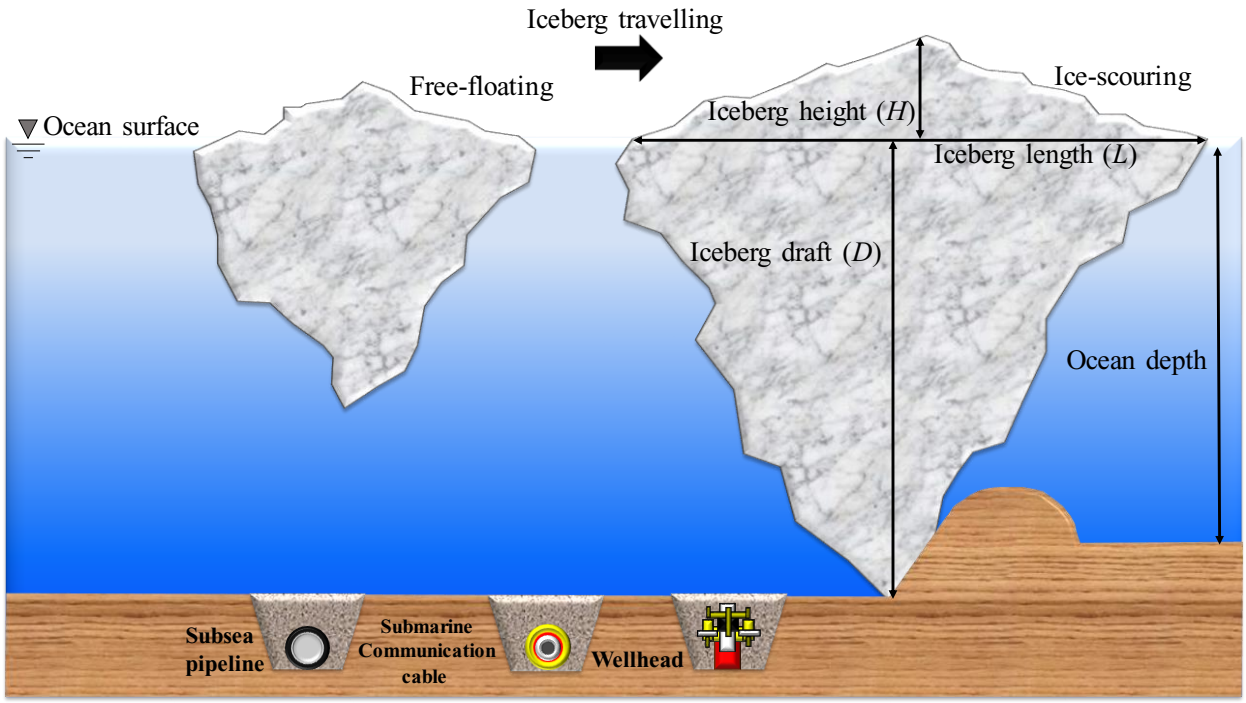


Figure 8-1. Schematic layout of the iceberg in free-floating and seabed gouging

The precise iceberg draft prediction may reduce expenditures and provide an effective iceberg management system to guarantee the physical and operational integrity of the subsea structures against the deep-keel iceberg attacks in the Arctic region. The studies conducted up to date have tended to focus on the estimation of the iceberg drafts regarding the iceberg length or mass.

Allaire (1972) studied the stability of icebergs with different shapes. The author demonstrated that the above-water features had a significant impact on the iceberg stability, rather the effect of the below-water section was insignificant. Robe and Farmer (1976) measured the characteristics of various icebergs in order to estimate the iceberg draft. The investigation concluded that the iceberg with massive height had the greatest height-to-draft ratios. They proposed a set of relationships to surmise the iceberg drafts in terms of iceberg height. Hotzel and Miller (1983) assessed the iceberg dimensions and supposed that the planimetric section of icebergs was circular, while the length of a berg was twice the circle radius. The iceberg draft was estimated in terms of iceberg length by using the power-law method. Furthermore, the International Ice Patrol (IIP) has been annually monitoring the characteristics of traveling icebergs in the Atlantic and Arctic Oceans. McKena (2000) assessed the risk of ice-gouging to the offshore oil and gas facilities operating in the Grand Banks area. The author asserted that the iceberg dimension was a function of the iceberg length. Barker et al. (2004) determined the iceberg sails and keels using the measurements observed in the field. The research exhibited that the iceberg draft could be calculated using the iceberg waterline length. McKenna (2004) evaluated the influence of iceberg shapes on the above-water and below-water portions. The author simulated the three-dimensional iceberg configuration using a probabilistic-based approach, where the iceberg shape was modeled regarding the principles of spatial statistics.

Timco (2007) described different iceberg management systems applied on the Grand Banks of Newfoundland. The iceberg draft estimation chart was evaluated in this work, and it was shown that the precision of this chart was nearly 25% for the iceberg draft prediction. Ralph et al. (2008) conducted a field experiment to measure the above-water and below-water characteristics of several icebergs moving on the Grand Bank of Newfoundland. They calculated the iceberg draft by using the multi-beam sonar system, whilst the iceberg mass was surmised through the photographs of the profile. Stuckey (2008) presented a probabilistic method for the estimation of the drift velocity distribution of the icebergs moving on the Grand Bank of Canada. The author introduced the dimensionless groups affecting the drift speed of icebergs. McKenna and King (2009) focused on the mass variation of icebergs due to different deterioration factors, including the size and shape of icebergs. The study demonstrated that the gradual relationship between mass fluxes and draft variations for single icebergs was potentially different from existing iceberg dimensions and volumes. King (2012) computed the iceberg impact risk for the sea bottom-founded infrastructures using the Monte Carlo simulation technique. This model predicted the collision risk of iceberg tips with submarine equipment regarding the iceberg draft height. Sacchetti et al. (2012) analyzed the features of different icebergs, such as wedged, domed, tabular, as well as pinnacle, and ice scouring in the Northeast Atlantic Ocean. The investigation reported that the bimodal distribution of the scour depth was observed across the study area. Turnbull et al. (2015) used a numerical model through a MATLAB environment so as to forecast the iceberg drift, where the iceberg drafts were approximated using the available empirical equations. The obtained results demonstrated that the iceberg trajectory was quite sensitive to iceberg drafts in some cases. Stuckey et al. (2016) utilized the Iceberg Load Software (ILS) for the simulation of load posing on structures such as gravity-based structures (GBS) and floating, production, storage,

and offloading (FPSO) operating in the Arctic and sub-Arctic areas. The equations correlating the iceberg draft, mass, and length were updated in this investigation. Turnbull et al. (2018) proposed a model for iceberg drift prediction for up to 24 hours. The authors highlighted that the iceberg mass and underwater portion were the primary sources of uncertainty in forecasting the iceberg drift. The study discussed that the empirical equation as a function of the iceberg length overestimated the iceberg drafts. Stuckey et al. (2021) most recently modeled the three-dimensional iceberg configuration using the adopting field measurements. The power curve, draft, and mass of icebergs were calculated in terms of the iceberg length in this research.

Several fields, analytical, and numerical studies have been carried out over the years to estimate the under-water section and draft of icebergs to ensure the operational integrity of the subsea assets. However, despite the accuracy, reliability, and cost-effectiveness of machine learning (ML) models, ML algorithms have not been applied in this field. Hence, developing an advanced ML-based solution to predict iceberg drafts is a real need in this field. To fill this knowledge gap, the use of a generalized structure of group method of data handling (GS-GMDH) algorithm as a neural network-based ML method was proposed in this study to model the iceberg drafts for the first time.

8.1.2. Methodology

In this section, a short overview of the artificial neural network (ANN), group method of data handling (GMDH), and generalized structure of group method of data handling (GS-GMDH) are presented. Subsequently, the parameter governing the iceberg drafts were defined and five GS-GMDH models were generated. A dataset was constructed using the field observations reported in the literature. A set of criteria were then utilized for the evaluation of the ML models.

8.1.2.1. Artificial neural network (ANN)

ANN is assumed as one of the universally supervised ML methods that are extensively applied to model different linear and nonlinear issues, e.g., the regression to classification problems. The major architecture of an ANN algorithm comprises at least three different layers such as an input layer, a hidden layer, and an output layer. The input parameters are situated in the input layer, whereas the objective function, e.g., the iceberg draft, is embedded in the output layer. The hidden neurons are located in the hidden layer, while the dimension of the hidden layer is ascertained regarding the problem's intricacy and desirable precision. It is worth noting that there is no certain regulation to determine the number of hidden layer neurons. The optimized number of hidden layer neurons was adjusted using a trial-and-error procedure in the present study. In other words, the number of hidden layer neurons was enhanced from one to 15 and the optimal number of this hyperparameter was 12 when the performance of the ANN algorithm was negligibly changed after this amount. In all neurons, the inputs and their weights are computed by mathematical operations, and the result is gone through a transfer function called the activation function. The sigmoid activation function was utilized since it showed better performance in comparison with the other activation functions evaluated in the present work. The capability of the applied ANN model was then assessed by the Mean squared error (*MSE*) when the loss function was to calculate the difference between the simulated values and the actual amounts.

8.1.2.2. Group method of data handling (GMDH)

GMDH was initially developed by Ivakhnenko (1976). The GMDH has been widely used in various areas to model linear and nonlinear issues, despite knowing insufficient knowledge of the problem (Ivakhnenko and Ivakhnenko 1995; Azimi et al. 2022). This method is a self-organized

approach in that different neurons are applied to simulate an objective parameter employing the input parameters. A binary relationship between the neurons using a quadratic polynomial leads to the construction of the GMDH algorithm. Hence, the objective function is predicted for the number of real values in a specific dataset including the output and input parameters.

The objective parameter is optimized by controlling the amount of the *MSE* between real iceberg drafts and the iceberg drafts simulated. The corrected Akaike Information Criterion (AIC_c) is applied for the amendment of the objective parameter to demonstrate the output in an uncomplicated form.

A polynomial function entitled the Volterra functional series is used to connect the input parameters and output one. Indeed, a quadratic and bivariate form of the Volterra functional series named the Kolmogorov-Gabor polynomial is employed in the GMDH solution (Ivakhnenko 1971). The vectors of unresolved factors of the Kolmogorov-Gabor polynomial arose from the regression methods computed by the least square (LS) approach. These vectors are surmised using the LS method stemming from the multiple regression approach. The explained process is retreaded for all neurons of the subsequent hidden layer based on the nexus topology of the GMDH algorithm (Azimi et al. 2022).

8.1.2.3. Generalized structure of group method of data handling (GS-GMDH)

The traditional version of the GMDH algorithm has a few drawbacks, comprising the application of the quadratic polynomial, restricting the contributions of each neuron to two, and restricting the contribution of each neuron to the use of neurons from the adjacent layer. These drawbacks result in the decline of the simulation capability of the traditional version of the GMDH algorithm, as well as the progress of complicated structures to model nonlinear and multidimensional issues. To

solve these problems, the generalized structure of the group method of data handling (GS-GMDH) approach was developed to simulate the iceberg drafts in the present work. The GS-GMDH approach applied the quadratic and cubic polynomials at the same time, employed two and three input parameters in all neurons, and inputted neurons from the non-adjacent and adjacent layers simultaneously. Ultimately, the superior GS-GMDH model was chosen by using the *AICc*, meaning that the plainness and precision of the selected model were assessed at the same time. These characteristics increased the ability of the GS-GMDH algorithm in comparison with the traditional version of the GMDH model.

Figure 8-2 illustrates the flowchart of the GS-GMDH approach used in the current study. Initially, the built dataset was divided into the training and testing sub-samples, e.g., the sub-sample consists of 60% for training and 40% for testing the GS-GMDH model. The number of inputs for all neurons, the highest number of neurons in each layer, and the highest number of layers were adjusted in the next stages.

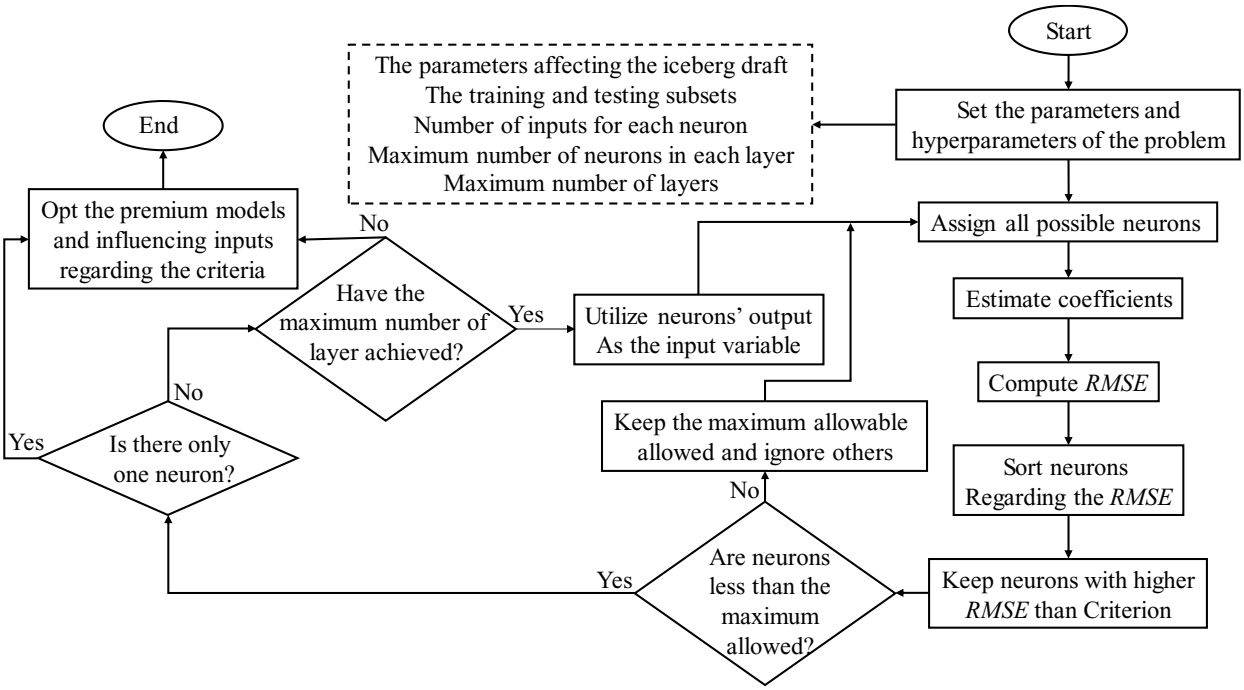


Figure 8-2. Flowchart of the GS-GMDH approach used in the current study

8.1.2.4. Iceberg draft

The iceberg draft (D) was assumed as a function of the physical characteristics of the iceberg, comprising the iceberg length (L), iceberg height (H), iceberg width (w), iceberg mass (M) in several fields, analytical, and numerical studies in the form below (Barker et al. 2004; McKenna et al. 2019; and Stuckey et al. 2021):

$$D = f_1(L, H, w, M) \tag{8-1}$$

Furthermore, the density of an iceberg (ρ_i), the density of seawater (ρ_{sw}), seawater viscosity (μ_{sw}), and gravitational acceleration (g) may influence the iceberg draft as follows:

$$D = f_2(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g) \tag{8-2}$$

The iceberg shape factor (Sf) signifies the global shape of icebergs, which can affect the magnitude of the iceberg draft (Turnbull et al. 2018). The shape factor of the traveling icebergs is considered universally into six categories as illustrated in Figure 8-3 (Rudkin 2005). Hence, equation (8-2) can be summarized below:

$$D = f_3\left(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g, S_f\right) \tag{8-3}$$

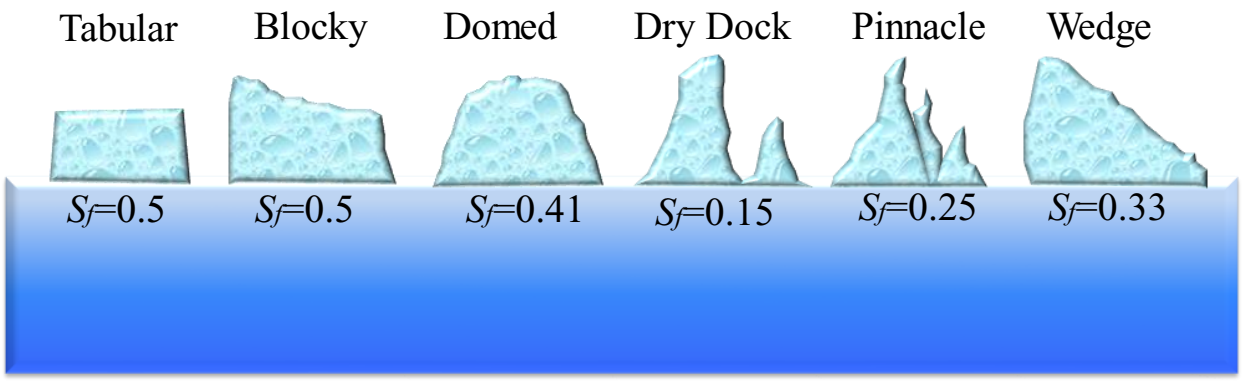


Figure 8-3. Shape factor values for different icebergs

It is assumed that the density along with viscosity of the seawater is constant and the value of gravitational acceleration can be regarded as a constant value; as a result, equation (8-3) is rewritten as follows:

$$D = f_4\left(L, H, w, M, \rho_i, S_f\right). \tag{8-4}$$

The dimensional form of equation (8-4) is written as:

$$D = f_5(\Pi_1, \Pi_2, \Pi_3, \Pi_4) \tag{8-5}$$

here, Π_1, Π_2,…, Π_4 are dimensionless groups and f_5 is a functional symbol based on the Buckingham-π theorem. Thus, the dimensionless groups below are written:

$$\Pi_1 = \frac{L}{H} \tag{8-6}$$

$$\Pi_2 = \frac{w}{H} \tag{8-7}$$

$$\Pi_3 = \frac{M}{\rho_i . H^3} \tag{8-8}$$

$$\Pi_4 = S_f \tag{8-9}$$

Equation (8-5) is then formulated as a function of four dimensionless groups as follows:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \tag{8-10}$$

Therefore, D/H as the iceberg draft ratio is a function of the length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$. Subsequently, the ML models applied in the current investigations were fed with the input parameters in equation (8-10).

Hence, four dimensionless groups, including length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$ were applied to estimate the iceberg draft ratio (D/H) through the ML models in the present work. Figure 8-4 shows the combinations of four dimensionless groups introduced to define the GS-GMDH models.

As it can be seen, to introduce the premium GS-GMDH models and the most influencing input parameters, five GS-GMDH models, including GS-GMDH 1 to GS-GMDH 5 were defined. The GS-GMDH 1 model included all dimensionless groups, whilst these dimensionless groups were disregarded one at a time in GS-GMDH 2 to GS-GMDH 5 models.

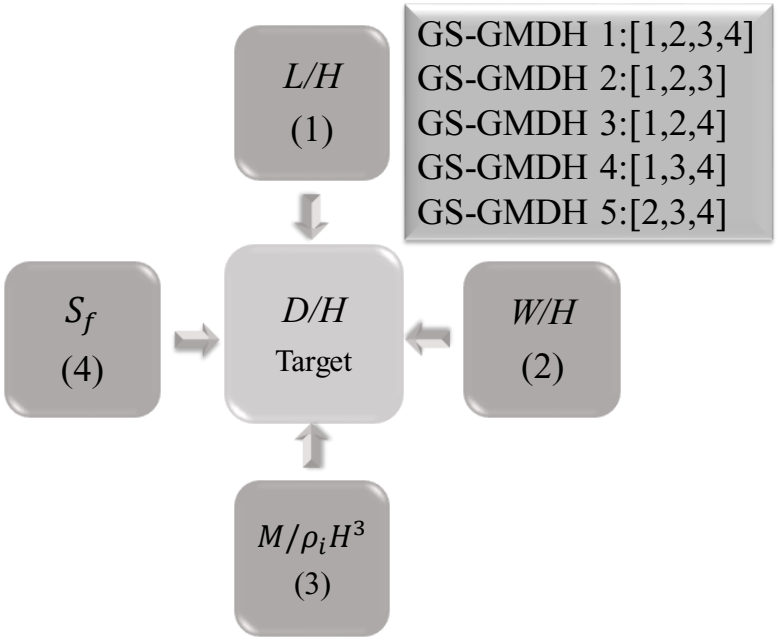


Figure 8-4. The combinations of the input parameters to introduce the GS-GMDH models

8.1.2.5. Construction of Dataset

Several field observations were adopted to analyze the iceberg draft. The key values of 12 field studies reported by El-Tahan et al. (1985) (38 cases), Woodworth-Lynas et al. (1985) (one case), Løset and Carstens (1996) (52 cases), Barker et al. (2004) (14 cases), McKenna (2004) (two cases), Sonnichsen et al. (2006) (nine cases), Turnbull et al. (2015) (two cases), McGuire et al. (2016) (eight cases), Younan et al. (2016) (29 cases), Talimi et al. (2016) (one case), Zhou (2017) (three cases), Turnbull et al. (2018) (two cases) were used. Table 8-1 tabulates a summary of the key

parameters employed in the present study comprising the maximum, minimum, average, variance, and standard deviation of the field observations.

The T-test and the P-value for the dataset were calculated, presuming that the P-value of 0.05 or less is statistically significant (Azimi and Shiri 2020), where the likelihood of the relationship between the observed values is influenced by an alternative hypothesis. This P-value for the constructed dataset was estimated as 0.008, representing that the correlations were statistically significant.

Table 8-1. Summary of the key parameters employed in the present study

Value	*D(m)*	*H(m)*	*w(m)*	*M(kg)*	*L(m)*	*Sf*
Max	240	94	408	1.13E+13	499	0.58
Min	18	3.9	9.672	8317400	34	0.11
Ave	88.988	30.956	101.605	1.68E+11	144.522	0.298
Var.	1696.272	399.389	3733.368	1.14E+24	6546.271	0.011
Std.	41.186	19.985	61.101	1.07E+12	80.909	0.107

8.1.2.6. Goodness of fit

In the current study, the correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were employed to examine the precision, correlation, and complexity of the GS-GMDH models. The closeness of the *R* and *WI* indices to one demonstrating the GS-GMDH model possessed a high level of correlation with the observed reports. The low *RMSE*,

MAPE, and *CRM* criteria showed the GS-GMDH model had the highest level of accuracy, while the complexity of the GS-GMDH models was not assessed using these indices. To overcome this limitation, the Akaike Information Criteria (*AIC*) was introduced in this study. This means that the less complex GS-GMDH model had the lowest magnitude of *AIC*, so the best GS-GMDH model may have the lowest degree of AIC index and error (*RMSE*, *MAPE*, and *CRM*), with the highest level of correlation (*R* and *WI*).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{8-11}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{8-12}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{8-13}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{8-14}$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \tag{8-15}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{8-16}$$

Here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, *n* and *k* are respectively the observational value, the predicted amount, the average observational values, the average predicted amount, the number of observations, and the number of independent variables in the GS-GMDH models.

8.1.3. Results and discussion

The GS-GMDH models' performance was assessed by conducting several analyses. The premium GS-GMDH model was also compared to GMDH, ANN, and empirical methods. Next, a partial derivative sensitivity analysis (PDSA) was performed for the best GS-GMDH model. A GS-GMDH-based equation was ultimately proposed for the estimation of the iceberg drafts.

8.1.3.1. Sensitivity analysis

The statistical indices estimated for the GS-GMDH models are depicted in Figure 8-5. The results demonstrated that GS-GMDH 3 was the premium model since it had the lowest degree of inaccuracy (*RMSE*=1.333) and complexity (*AIC*=13.996) along with the highest degree of correlation (*R*=0.877) with the iceberg drafts observed. The GS-GMDH 3 predicted the iceberg draft values utilizing the L/H, w/H, and S_f inputs; however, the effect of iceberg mass ratio ($M/\rho_i.H^3$) was disregarded in this model. The GS-GMDH 1 model as a function of all input parameters (L/H, w/H, $M/\rho_i.H^3$, and S_f) was known as the second-best model to estimate the iceberg draft, with the *AIC* and *WI* of 15.995 and 0.902, respectively. The value of *R*, *CRM*, and *RMSE* for the third-best model (i.e., GS-GMDH 5) was respectively calculated at 0.531, -0.021, and 2.910. It is worth noting that GS-GMDH 5 used the w/H, $M/\rho_i.H^3$, and S_f inputs to model the iceberg draft but the iceberg length ratio (L/H) was a removed factor in the GS-GMDH 5 model. Using the L/H, w/H, $M/\rho_i.H^3$ input factors, the iceberg drafts were simulated through the GS-GMDH 2 model, meaning that the impact of the iceberg shape factor (S_f) was neglected for this model, with an *RMSE* and *AIC* estimated of 5.484 and 53.302. Regarding the sensitivity analysis conducted, the GS-GMDH 4 model as a function of L/H, $M/\rho_i.H^3$, and S_f was detected as the worst GS-GMDH model to forecast the iceberg draft, where the effectiveness of

the iceberg width ratio (w/H) was eliminated for GS-GMDH 4. The computed *WI*, *R*, and *RMSE* values for the GS-GMDH 4 model were equal to 0.367, 0.331, and 6.225, respectively.

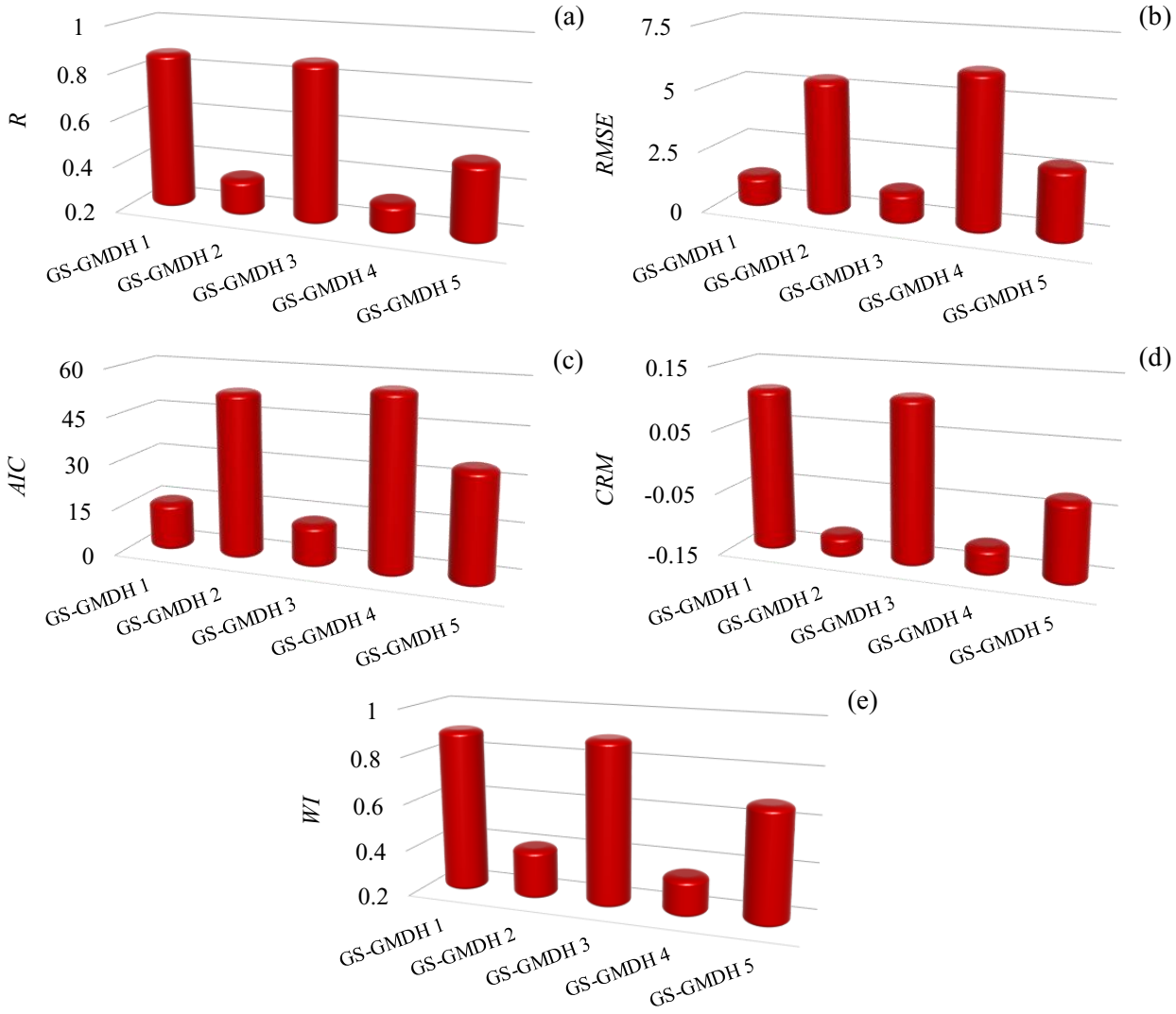


Figure 8-5. Statistical indices estimated for the GS-GMDH models (a) R (b) RMSE (c) AIC (d) CRM (e) WI

Figure 8-6 displays the scatter plots of the GS-GMDH 1 to GS-GMDH 5 models. As shown, the GS-GMDH 3 model had the highest level of correlation with the value of iceberg drafts observed; hence, this model was introduced as the superior GS-GMDH model for the estimation of the target

parameter. The sensitivity analysis proved that the iceberg width ratio (w/H) was the most significant input parameter to model the iceberg drafts. Furthermore, the iceberg shape factor (S_f) and the iceberg length ratio (L/H) were ranked as the second-important and the third-important input parameter in the present study. The results analyzed showed that the iceberg mass ratio ($M/\rho_i.H^3$) had the lowest level of influence to model the iceberg draft using the GS-GMDH algorithm.

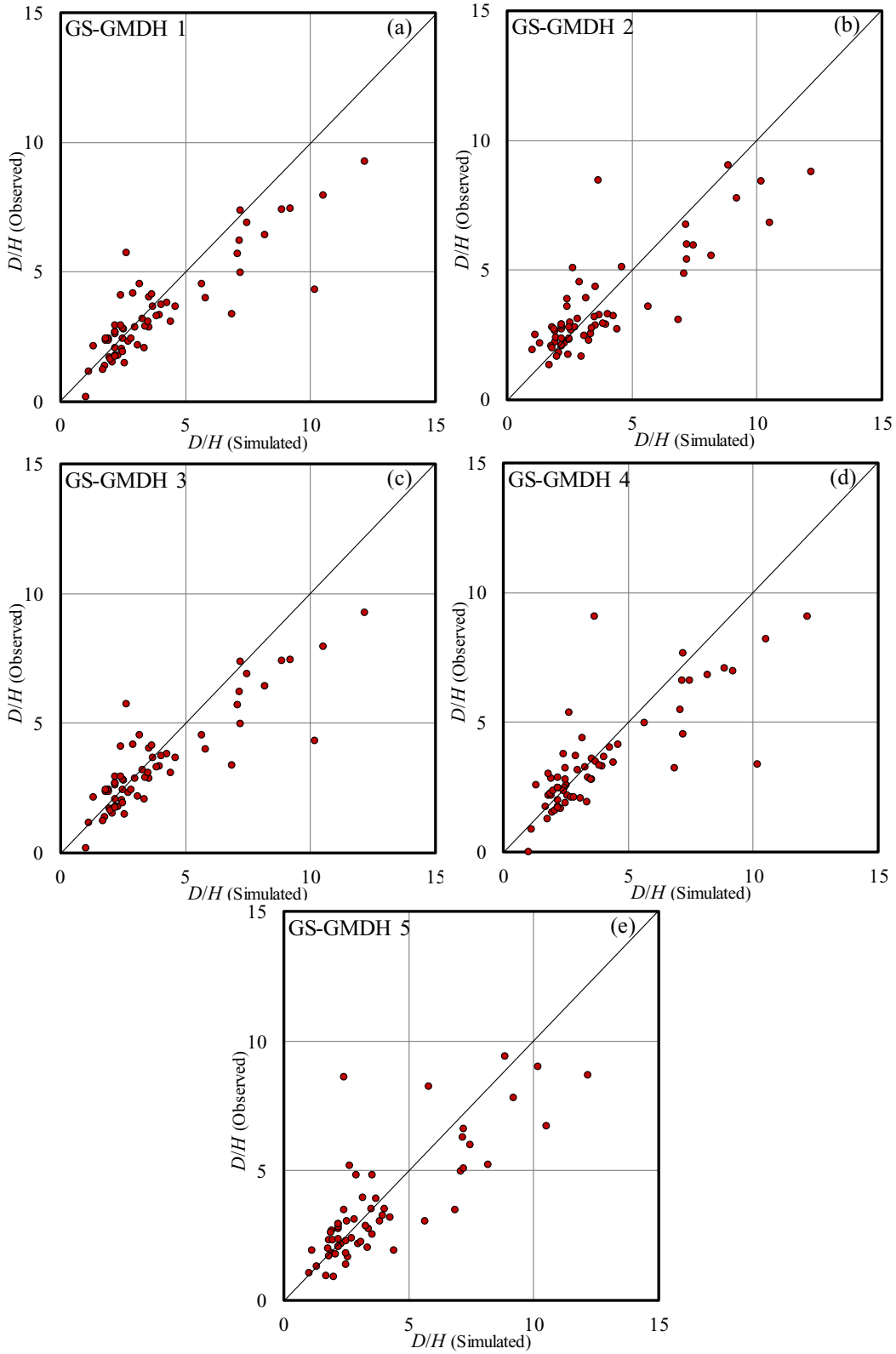

GS-GMDH 1
(a)
GS-GMDH 2
(b)
GS-GMDH 3
(c)
GS-GMDH 4
(d)
GS-GMDH 5
(e)
D/H (Observed)
D/H (Simulated)
0
5
10
15

Figure 8-6. Scatter plots of (a) GS-GMDH 1 (b) GS-GMDH 2 (c) GS-GMDH 3 (d) GS-GMDH 4 (e) GS-GMDH 5 models

8.1.3.2. Error analysis

The error analysis performed for the GS-GMDH models is shown in Figure 8-7. The GS-GMDH 1 and GS-GMDH 3 models had a similar performance in terms of the implemented error analysis, signifying that almost half of the iceberg drafts simulated by those models possessed an error of greater than 20%, while approximately one-third of results obtained from them had an error of between 10% and 20%. About one-fifth of the iceberg drafts estimated by the GS-GMDH 2, GS-GMDH 4, and GS-GMDH 5 models provided an error of smaller than 10%, whereas this value for the GS-GMDH 3 was surmised as roughly 14%. Moreover, a 10% error arose from 22% of the GS-GMDH 4 and GS-GMDH 5 models' outcomes, while this amount for the GS-GMDH 3 model was roughly 14%. In terms of the error analysis conducted, the GS-GMDH 5 model demonstrated unsatisfactory performance; however, the GS-GMDH 3 simulated the iceberg drafts reasonably.

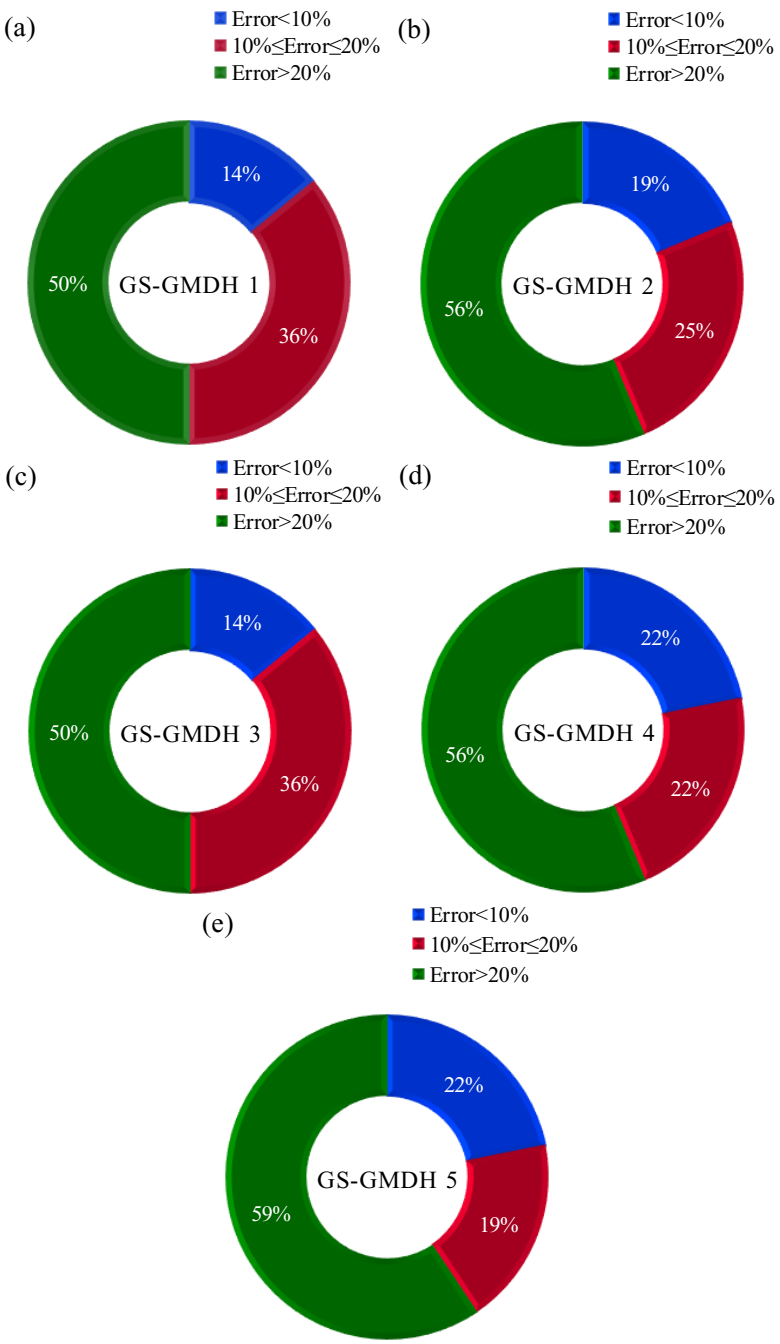


Figure 8-7. Error analysis performed for (a) GS-GMDH 1 (b) GS-GMDH 2 (c) GS-GMDH 3 (d) GS-GMDH 4 (e) GS-GMDH 5 models

8.1.3.3. Uncertainty analysis

To examine the efficiency of the GS-GMDH models, an uncertainty analysis was also conducted. To do so, the errors arising from the GS-GMDH model (e_j) were calculated as the difference between the iceberg drafts predicted (P_j) and the observed amounts (O_j) as follows:

$$e_j = P_j - O_j \tag{8-17}$$

The mean (Mean) and the standard deviation (StDev) values of the error values were obtained by the equations below:

$$\mathrm{Mean} = \frac{1}{n}\sum_{j=1}^{n} e_j \tag{8-18}$$

$$\mathrm{StDev} = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2/(n-1)} \tag{8-19}$$

A particular GS-GMDH model underestimated the iceberg draft if the sign of the Mean value was negative; however, the positive sign of the Mean reprinted that the GS-GMDH model possessed an overestimated performance on the iceberg draft. In addition, a confidence interval (CI) was generated around the error values using the Mean, StDev values, and the "Wilson score approach" by disregarding the continuity correction. A normal distribution interval rectified as an asymmetric normal distribution, entitled the Wilson score interval, was employed to modify the CI bounds; as a result, a ±1.96Se produced a 95% of the confidence interval as 95%CI. It is worth mentioning that the width of uncertainty bound (WUB) of the GS-GMDH model was computed as below (Azimi and Shiri 2021):

$$\mathrm{WUB} = \pm\frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{8-20}$$

Figure 8-8 presents the uncertainty analysis outcomes of the GS-GMDH 1 to GS-GMDH 5 models. Regarding the uncertainty analysis results, the GS-GMDH 1 and GS-GMDH 3 models showed an overestimated performance, with a Mean value of 0.436 and 0.435. In contrast, the GS-GMDH 2,

GS-GMDH 4, and GS-GMDH 5 models underestimated the iceberg drafts, where the Mean values were -0.434, -0.401, and -0.081, respectively. The uncertainty analysis demonstrated that the narrowest bound of uncertainty belonged to the GS-GMDH 3 model (WUB=0.317), rather the GS-GMDH 4 had the widest uncertainty bound, with a WUB of 1.564. Additionally, 95%CI for the GS-GMDH 1 to GS-GMDH 5 models was approximated at (0.118 to 0.753), (-1.810 to 0.943), (0.118 to 0.753), (-1.965 to 1.163), and (-0.813 to 0.652), respectively.

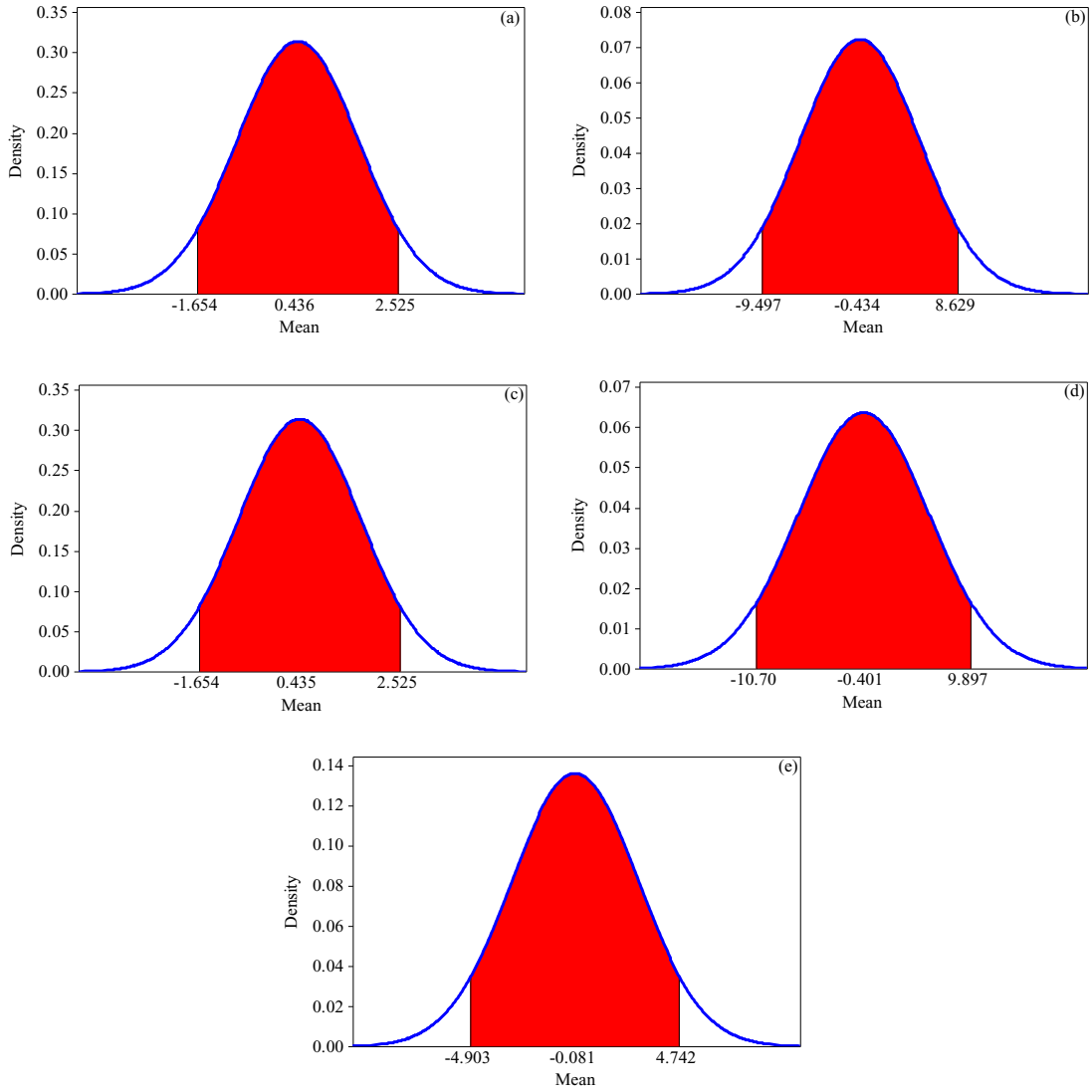


Figure 8-8. Normal distribution of the error calculated for (a) GS-GMDH 1 (b) GS-GMDH 2 (c) GS-GMDH 3 (d) GS-GMDH 4 (e) GS-GMDH 5 models

8.1.3.4. Comparison with GMDH, ANN, and empirical models

The performance of the premium GS-GMDH model was compared to the GMDH, ANN, and a set of empirical models. Barker et al. (2004) presented a relationship for estimation of the iceberg draft in terms of the iceberg length by using the power curve and regression analysis as below:

$$D = 2.91(L^{0.71}) \quad (8\text{-}21)$$

Sacchetti et al. (2012) offered an equation in terms of the iceberg length to appraise the iceberg draft in the following form:

$$D = 3.9(L^{0.63}) \quad (8\text{-}22)$$

King et al. (2016) proposed an empirical model that corresponded with the iceberg mass as follows:

$$D = 2.05(M^{0.276}) \quad (8\text{-}23)$$

Stuckey et al. (2021) derived an empirical model regarding the data collected in 2016 as below:

$$D = 4.78(L^{0.58}) \quad (8\text{-}24)$$

Figure 8-9 displays the comparison of the premium GS-GMDH model with GMDH, ANN, and empirical approaches. As shown, the value of the *R* index for the GS-GMDH, GMDH, ANN, equations (8-21), (8-22), (8-23), and (8-24) was computed to be 0.877, 0.876, 0.746, 0.769, 0.770, 0.333, and 0.770, respectively. The GS-GMDH algorithm was able to predict the iceberg draft with the lowest degree of complexity, where the *AIC* criteria for the ANN model, equations (8-21), (8-22), (8-23), and (8-24) equaled 25.346, 256.896, 238.378, 358.385, and 252.404. Thus, the premium GS-GMDH model outperformed the GMDH, ANN, and empirical models.

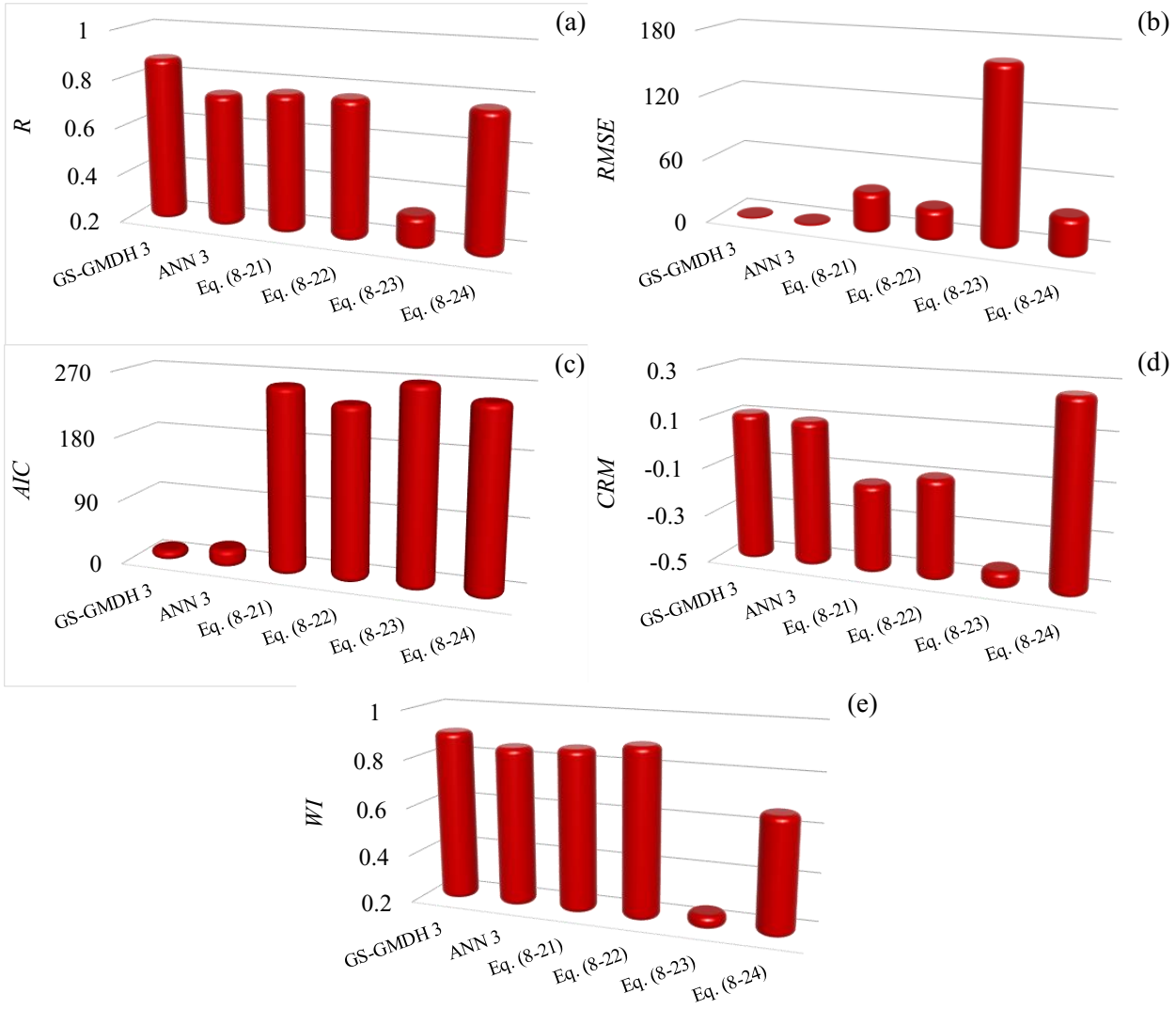


Figure 8-9. Comparison of the premium GS-GMDH model with GMDH, ANN, and empirical approaches

8.1.3.5. Partial derivative sensitivity analysis (PDSA)

To assess the impact of the input parameters (dependent variables) on the iceberg drafts (the independent variable), a partial derivative sensitivity analysis (PDSA) for the premium GS-GMDH model (GS-GMDH 3) was conducted. The PDSA generally demonstrates how the dependent variables may affect the independent variable and it is determined as:

$$\partial(d/H)/(\partial x_i) \qquad x_i = L/H\,,\ w/H\,, \text{and } S_f \tag{8-25}$$

A negative sign of the PDSA represents a decreasing influence on the independent variable, rather a positive sign of the PDSA means an increasing effect on the iceberg draft (Azimi et al. 2022). The PDSA performed for the iceberg drafts simulated by the GS-GMDH 3 model is shown in Figure 8-10.

The overwhelming majority of the PDSA taken for the iceberg length ratio (L/H) had the positive sign, with an upward pattern. In other words, the iceberg drafts grew with increasing the iceberg length ratio (L/H). Furthermore, almost all the PDSA calculated for the iceberg width ratio (w/H) had the positive sign, meaning that the magnitude of the iceberg drafts incremented when the value of the iceberg width ratio (w/H) was enhanced. Although most PDSA estimated for the iceberg shape factor (*Sf*) were positive, a minority of the obtained PDSA had a negative sign, with a downward trend.

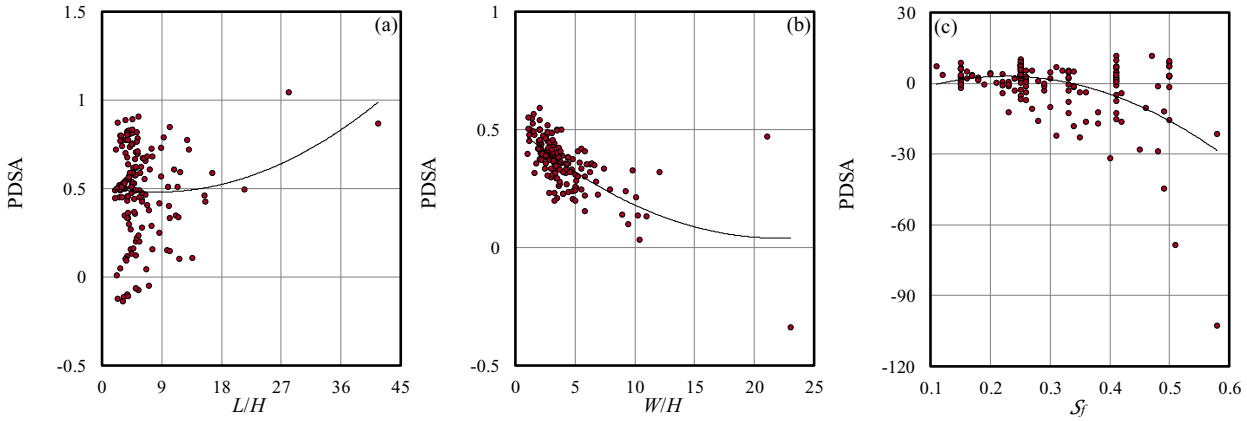


Figure 8-10. Result of PDSA for the premium GS-GMDH model (a) *L*/*H* (b) w/H (c) S_f

8.1.3.6. Superior model

The conducted analyses demonstrated that the GS-GMDH 3 model simulated the iceberg draft with reasonable performance. This model estimated the target parameter in terms of L/H, w/H,

and S_f; hence, an explicit GS-GMDH-based equation was extracted to model the iceberg drafts for practical applications as below:

$$D/H = -4.205 + 13.281(S_f) + 0.579(w/H) + 1.049(L/H) - 0.499(w/H).(S_f) - 2.706(L/H).(S_f) + 0.050(L/H).(w/H) + 5.963(S_f)^2 - 0.045(w/H)^2 + 0.004(L/H)^2 \quad (8\text{-}26)$$

The proposed model biased slightly toward overestimation; however, it was able to calculate the iceberg drafts with acceptable performance. This model had a significant superiority compared to the GMDH, ANN algorithms, and empirical models.

8.1.4. Conclusion

The generalized structure of the group method of data handling (GS-GMDH) simulated the iceberg drafts in this study for the first time. First, the parameters governing the iceberg drafts were detected and five GS-GMDH models comprising the GS-GMDH 1 to GS-GMDH 5 models were then produced. Using the field measurements published in the literature, a comprehensive dataset was constructed, and the GS-GMDH models were trained and tested through 60% and 40% of the dataset, respectively. The superior GS-GMDH model and the most effective input parameters were identified using a sensitivity analysis. The error analysis, uncertainty analysis, and partial derivative sensitivity analysis were conducted for the GS-GMDH model. The best GS-GMDH model was also compared to the group method of data handling (GMDH), artificial neural network (ANN), and empirical models. The most important results obtained from the present study are summarized as follows:

- The GS-GMDH 3 model as a function of L/H, w/H, and S_f was identified as the premium model since it predicted the iceberg drafts meticulously, with the *R*, *RMSE*, and *AIC* of 0.877, 1.333, and 13.996, respectively.

- The iceberg width ratio (w/H) and the iceberg shape factor (S_f) had a remarkable impact on estimating the iceberg drafts using the GS-GMDH algorithm.
- The accuracy of the GS-GMDH 3 model was relatively high, signifying that approximately 36% of the iceberg drafts simulated by this model had an error of between 10% and 20%.
- Regarding the uncertainty analysis implemented, the narrowest bound of uncertainty belonged to the GS-GMDH 3 model, where this model overestimated the iceberg drafts.
- The comparison between the performance of the GS-GMDH 3 model with GMDH, ANN, and empirical models demonstrated that the GS-GMDH 3 model forecasted the iceberg drafts with the highest degree of accuracy and correlation along with the lowest level of complexity.
- The performed partial derivative sensitivity analysis for the GS-GMDH 3 model indicated that by growing the iceberg lengths and the iceberg widths, the magnitude of the iceberg drafts incremented.
- A GS-GMDH-based equation was proposed for the estimation of the iceberg drafts in the practical applications of iceberg management operations and engineering design projects.

Although the current study was the first implementation of the GS-GMDH algorithm for the iceberg drafts modeling to guarantee the integrity of offshore oil and gas equipment operating in the Arctic shallow waters, the results obtained may facilitate the cost-effective, quick, and accurate alternatives to estimate the iceberg drafts.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial

University of Newfoundland, the “Natural Science and Engineering Research Council of Canada (NSERC)”, and the “Newfoundland Research and Development Corporation (RDC) (now TCII)” through “Collaborative Research and Developments Grants (CRD)”. Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Allaire, P.E., 1972. Stability of simply shaped icebergs. Journal of Canadian Petroleum Technology, 11: 4.

Azimi, H., Shiri, H., 2020. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2021. Assessment of Ice-Seabed Interaction Process in Clay Using Extreme Learning Machine. International Journal of Offshore and Polar Engineering, 31(04): 411-420.

Azimi, H., Shiri, H., Zendehboudi, S., 2022. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Region Science and Technology, 193: 103426.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In the fourteenth international offshore and polar engineering conference. OnePetro, ISOPE-I-04-116.

Bruce, J., Younan, A., MacNeill, A., 2016. Applications of iceberg profiling data to improve iceberg management success. Arctic Technology Conference, OTC-27470-MS.

El-Tahan, M., El-Tahan, H., Courage, D., Mitten, P., 1985. Documentation of Iceberg Groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

Hotzel, I.S., Miller, J. D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

https://www.arcticwwf.org/threats/oil-and-gas/

McGuire, P., Younan, A., Wang, Y., Bruce, J., Gandi, M., King, T., Regular, K., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

McKena, R., 2000. Study of iceberg scour & risk in the Grand Banks region. KR Croasdale & Associates Ltd., Ballicater Consulting Ltd., Canadian Seabed Research Ltd., C-CORE, and Ian Jordaan & Associates Inc., PERD/CHC Report, 31-26.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., 2009. Modelling Iceberg Shape, Mass and Draft Changes. In Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions, No. POAC09-107.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. In Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, POAC, Delft, The Netherlands (pp. 9-19).

Ivakhnenko, A.G., 1971. Polynomial theory of complex systems. IEEE Transactions on Systems, Man, and Cybernetics, 4: 364-378.

Ivakhnenko, A.G., 1976. The group method of data handling in prediction problems. Soviet Automatic Control, 9(6): 21-30.

Ivakhnenko, A.G., Ivakhnenko, G.A., 1995. The review of problems solvable by algorithms of the group method of data handling (GMDH). Pattern Recognition And Image Analysis C/C Of Raspoznavaniye Obrazov I Analiz Izobrazhenii. 5: 527-535.

King, T., 2012. Iceberg Interaction Frequency Model for Subsea Structures. Arctic Technology Conference. OTC-23787-MS

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Arctic Technology Conference, OTC-27358-MS.

Løset, S., Carstens, T., 1996. Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Regions Science and Technology, 24(4): 323-340.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Regions Science and Technology, 170: 102929.

Ralph, F., McKenna, R., Gagnon, R., 2008. Iceberg characterization for the bergy bit impact study. Cold Regions Science and Technology, 52(1): 7-28.

Robe, R. Q., Farmer, L. D., 1976. Physical Properties of Icebergs. Part I. Height to Draft Ratios of Icebergs. Part II. Mass Estimation of Arctic Icebergs. Coast Guard Research and Development Center Groton Conn.

Rudkin, P., 2005. Comprehensive Iceberg Management Database Report 2005 update. PERD/CHC Report 20-72. National Research Council of Canada (NRC) and Panel on Energy Research and Development (PERD).

Sacchetti, F., Benetti, S., Cofaigh, C.Ó., Georgiopoulou, A., 2012. Geophysical evidence of deep-keeled icebergs on the Rockall Bank, Northeast Atlantic Ocean. Geomorphology. 159: 63-72.

Sonnichsen, G., Hundert, T., Myers, R., Pocklington, P., 2006. Documentation of Recent Iceberg Grounding Evens and a Comparison with Older Events of Know Age. Environmental Studies Research Funds. Report ESRF Vol. 157.

Stuckey, P.D., 2008. Drift speed distribution of icebergs on the grand banks and influence on design loads. Doctoral dissertation, Memorial University of Newfoundland, NL, Canada.

Stuckey, P., Younan, A., Parr, G., Fuglem, M., Habib, K., 2016. Updating the Iceberg Load Software Using High Resolution Iceberg Profiles. Arctic Technology Conference. OnePetro. OTC-27492-MS.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., King, T., 2021. Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. International Conference on Offshore Mechanics and Arctic Engineering. 85178: V007T07A018.

Talimi, V., Ni, S., Qiu, W., Fuglem, M., MacNeill, A., Younan, A., 2016. Investigation of Iceberg Hydrodynamics. Arctic Technology Conference, OTC-27493-MS.

Timco, G., 2007. Grand Banks Iceberg Management. PERD/CHC Report 20-84.

Turnbull, I. D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Regions Science and Technology, 110: 1-18.

Turnbull, I.D., King, T., Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Woodworth-Lynas, C.M.T., Simms, A., Rendell, C. M., 1985. Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Regions Science and Technology, 10(2): 163-186.

Younan, A., Ralph, F., Ralph, T., Bruce, J., 2016. Overview of the 2012 iceberg profiling program. In Arctic Technology Conference, OTC-27469-MS.

Zhou, M., 2017. Underwater iceberg profiling and motion estimation using autonomous underwater vehicles, Doctoral dissertation, Memorial University of Newfoundland, NL, Canada.

Section 2

Iceberg Draft Prediction using Gradient Boosting Regression Algorithm

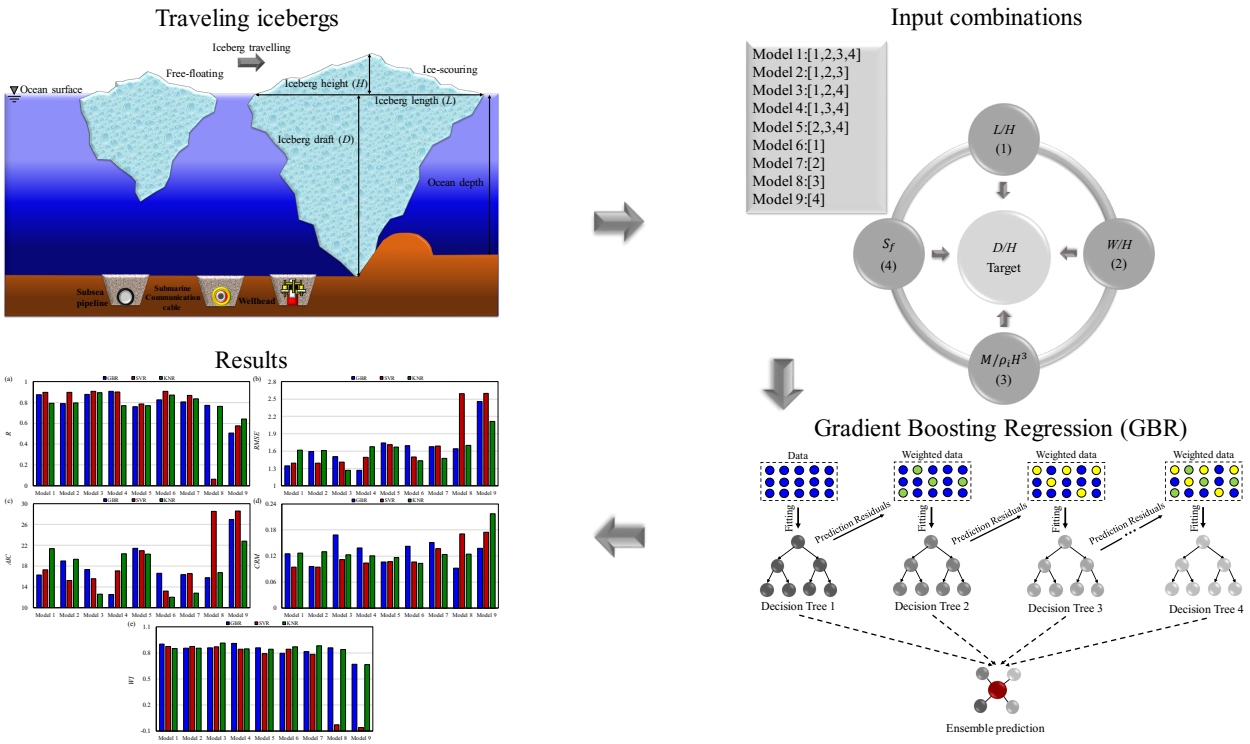
This section is submitted as a journal manuscript

Abstract

The Arctic area is one of the best destinations for the development of oil and gas loading equipment. However, the recent development of oil and gas facilities, including the submarine pipelines and wellheads crossing the Arctic area, has elevated the need for more attention to iceberg draft (under-water height of icebergs) estimation during an ice scouring event. This means if the draft of an iceberg is more than the ocean deepness, the iceberg tip can gouge the ocean floor and collide with the subsea assets; as a result, the operational integrity of the submarine infrastructures is threatened. Hence, the estimation of the iceberg drafts is vital for the oil and gas operators in the Arctic waters. In the present study, the use of the Gradient Boosting Regression (GBR) algorithm was proposed to simulate the iceberg drafts for the first time. Initially, the parameters governing the iceberg drafts were recognized, and nine GBR models were then developed. The premium GBR model along with the most important inputs was known by conducting a sensitivity analysis. The best GBR model was compared with Support Vector Regression (SVR) and K-Nearest Neighbors Regression (KNR) algorithms, rather the GBR algorithm had the highest degree of accuracy and correlation as well as the lowest amount of complexity. Lastly, the performed uncertainty analysis proved that the superior GBR model possessed the narrowest bound of uncertainty, with an overestimated performance in the iceberg draft simulation.

Keywords: Iceberg Draft; Gradient Boosting Regression; Support Vector Regression; K-Nearest Neighbors Regression; Subsea assets, Iceberg management system

Graphical abstract



8.1.5. Introduction

Roughly 22% of the Earth's undiscovered hydrocarbons are held in the Arctic area (WWF Arctic Programme), where crude oil, natural gases, and natural gas liquids discovered on the Grand Banks of Newfoundland are estimated at 2 billion barrels, 5.4 trillion cubic feet, and 313 million barrels, respectively (CNLOPB, 2005). However, current offshore hydrocarbon transferring equipment (e.g., subsea pipelines, wellheads, and communication cables) developed on the Grand Banks of Canada has led to a significant awareness of the iceberg draft prediction and the iceberg-seabed interaction process. The iceberg scouring and iceberg free-floating in the Arctic waters are schematically displayed in Figure 8-11. As shown, if the ocean depth is deeper than the iceberg draft, the traveling berg is in a free-floating circumstance, rather the iceberg tip would gouge the seafloor, which is called the ice-gouging problem. The profound keel bergs can plow the seabed and the subgauge soil deformations would extend down the ice keel owing to the shear resistance of the sea bottom soil; as a result, the operational integrity of such subsea assets may be threatened in the shallow areas (Nematzadeh and Shiri 2020).

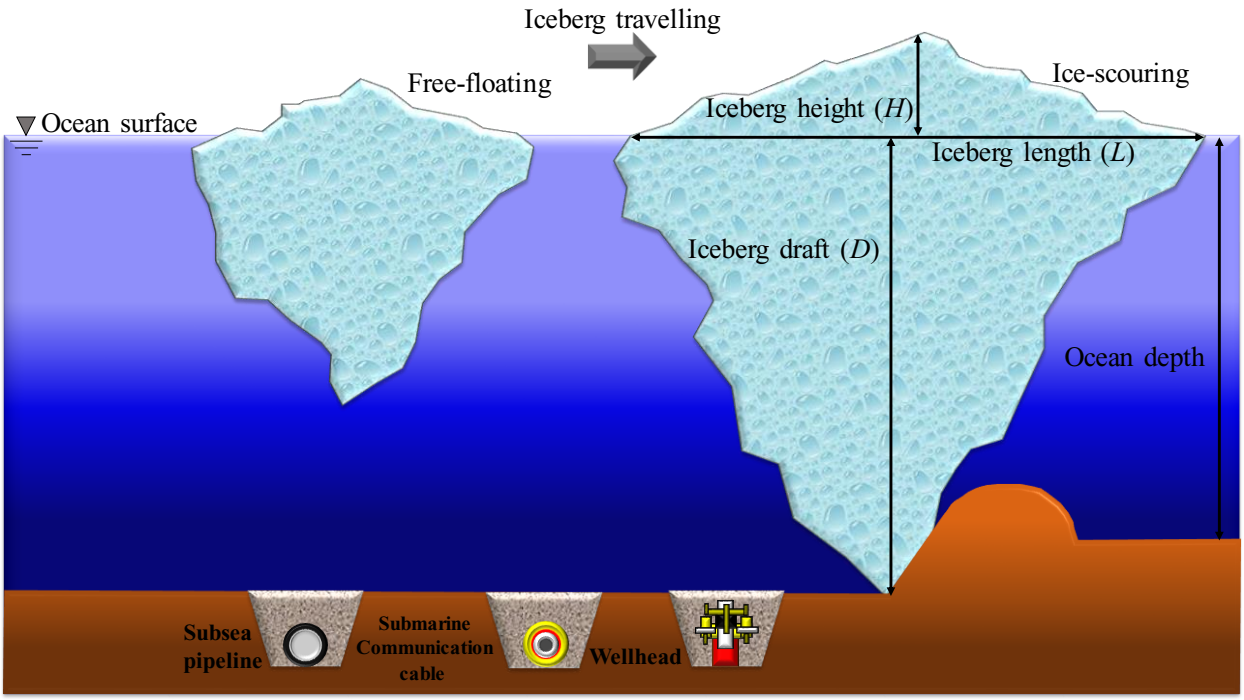


Figure 8-11. Iceberg in scouring and free-floating circumstances

Although extensive research comprising field, theoretical, numerical, and field explorations have been conducted on the iceberg draft estimation (Robe and Farmer 1976; Hotzel and Miller 1983; Barker et al. 2004; King et al. 2016; Stuckey et al. 2021), ML-based study has not been conducted to estimate the iceberg drafts.

Allaire (1972) analyzed the stability of icebergs under various conditions. The author explained that the above-water segments had a considerable influence on the iceberg stability, instead, the impact of the below-water part was negligible. Robe and Farmer (1976) measured different iceberg characteristics to estimate the iceberg drafts. The study stated that the iceberg with massive height possessed the largest height-to-draft ratios. A couple of relationships as a function of iceberg height to predict iceberg drafts were proposed. Hotzel and Miller (1983) considered the ice proportions and presumed that the planimetric part of bergs was circular, but the length of an iceberg was

double the circle radius. The draft was counted using iceberg length through the power-law technique. McKena (2000) evaluated the threat of ice-gouging to the offshore petroleum installations serving the Grand Banks region. The author showed that the iceberg proportions were a function of the iceberg length. Sonnichsen et al. (2003) reported the seafloor surveys and ice-gouging on the Grand Bank of Canada in the 2000 iceberg season. The drafts of the iceberg were recorded through a lateral scan sonar tool mounted on the tracking boat. The study demonstrated that there was serious concern about the precision of lateral scans of iceberg draft estimation. McKenna (2004) analyzed the impact of iceberg configurations on the above-water and below-water parts. The author modeled the three-dimensional iceberg formatting by adopting a probabilistic-based system, where the iceberg profile was estimated regarding the principles of spatial statistics. Barker et al. (2004) specified the iceberg sails and keels utilizing the dimensions marked in the field. The investigation revealed that the iceberg draft could be summed by employing the iceberg waterline length. Dowdeswell and Bamber (2007) scrutinized the keel deepness of crossing icebergs in the Antarctic waters. The authors calculated the depth of the keel through the ice thickness and surface elevation. The study ended that a small minority of icebergs in the Antarctica and Greenland waters had drafts deeper than 650 m. Sacchetti et al. (2012) explored the iceberg features and ice-gouging in Labrador and Hibernia territories. The characteristics of various bergs like wedged, domed, tabular, and pinnacle icebergs were considered in the examination. They proposed some relationships as a function of the berg length to predict the draft. King et al. (2016) completed field experimentation to count the rolling iceberg rate. The draft was surmised by a calving study, with a computed standard deviation of draft changes from 19% to 34%. The draft was approximated using the mass of the iceberg. Turnbull et al. (2018) suggested an instance of the drift mensuration of shifting bergs in Newfoundland's

waters. The study stated that the iceberg draft estimated was roughly 1.3 times deeper than the real iceberg draft. McKenna et al. (2019) modeled the ice scouring on the Grand Banks of Canada adopting the Monte Carlo simulation (MCs). The iceberg draft variations were also employed to reduce the dimension of draft changes in the applied methodology. Stuckey et al. (2021) modeled the 3D iceberg forms through a field survey. The examination exhibited that the iceberg drafts were summed in terms of iceberg length through the power method. They provided two practical instances concerning the information gathered in 2016 and the post-2000 report.

As shown in the literature, developing a reliable and cost-effective solution to predict iceberg drafts is necessary to guarantee the operational and physical integrity of the submarine assets. Despite the quick progress of machine learning (ML) algorithms for modeling linear and non-linear problems, ML models have not been applied in the estimation of iceberg draughts. To fill this knowledge gap, in this research, several ML algorithms comprising gradient boosting regression (GBR), support vector regression (SVR), and k-nearest neighbors regression (KNR) were applied to predict the iceberg drafts for the first time.

8.1.6. Methodology

In this section, a summary of the ML algorithms (i.e., GBR, SVR, and KNR) applied in the current study, was presented. Next, the parameters governing the iceberg drafts were introduced and nine GBR models were developed by using these input parameters. Moreover, a comprehensive dataset for the training and testing of the ML models was constructed. Lastly, the used statistical criteria for the evaluation of the ML models were discussed.

8.1.6.1. Gradient Boosting Regression (GBR)

GBR is an ML model to simulate different regression and classification issues. GBR forms an ensemble estimator, concerning the decision trees, to overwhelm the weaknesses of powerless ML models. In this algorithm, a single decision tree may act inadequately; nonetheless, the ensemble algorithm, entitled gradient boosted trees, generally exceeds other sole decision-based methods (Hastie et al. 2009). In the present investigation, the hyperparameters of the GBR model were adjusted utilizing the trial-and-error procedure, e.g., *loss='huber'*, *n_estimators*=25, *subsample*=1.0, *learning_rate*=0.14, *min_samples_split*=4, *criterion='friedman_mse'*, *min_samples_leaf*=1, *max_depth*=3, *min_weight_fraction_leaf*=0.0, *min_impurity_decrease*=0.0, *random_state=None*, *max_features=None*, *init=None*, *alpha*=0.9, *warm_start=False*, *verbose*=0, *max_leaf_nodes=None*, *n_iter_no_change=None*, *tol*=0.0001, *ccp_alpha*=0.0, *validation_fraction*=0.1. The structure of the GBR algorithm is depicted in Figure 8-12.

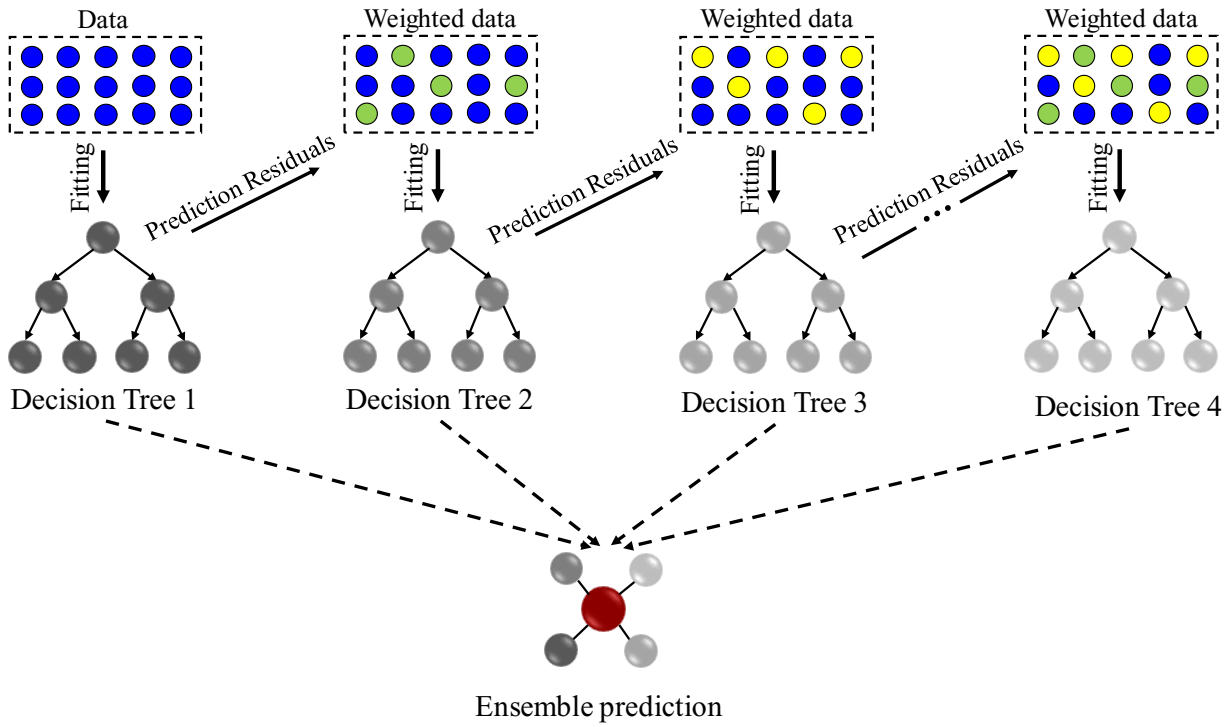


Figure 8-12. Structure of GBR algorithm

8.1.6.2. Support Vector Regression (SVR)

A support vector machine (SVM) is a supervised ML algorithm to decode classification and regression issues. The SVM is established on Vapnik-Chervonenkis (VC) hypothesis, and this algorithm was suggested by Vapnik (1995). To approach the regression problems, SVR is involved in which the training dataset is mapped from the inputs (input space) into the target function (feature space) via a function (Q). A separating hyperplane with the highest margin is constructed in the feature space. In a regression problem, a nonlinear transformation from the input space to high-dimensional space is created employing the Q function. Regardless of the transformation function (Q), the kernel function may perform the dot product in the multidimensional feature space via the low-dimensional space inputs. In realistic problems, several kernel functions containing the linear, polynomial, and radial basis functions (RBF) are employed in the SVR method. Additionally, the ε-insensitive loss function is engaged as a cost function in this algorithm. To predict the iceberg drafts in the present work, the hyperparameters of the SVM model, e.g., the penalty parameter (*C*), the kernel coefficient (*gamma*), *epsilon*, *verbose*, and kernel were adjusted as 0.01, 1, 0.5, 1, and linear, respectively. The involved hyperparameters of the SVM algorithm in the present investigation were selected according to a trial and error strategy.

8.1.6.3. K-Nearest Neighbors Regression (KNR)

The K-Nearest Neighbors algorithm is also applied in classification and regression studies. The KNR is a non-parametric machine learning approach that intuitively estimates the correlation between independent and dependent features by averaging the observed values in the same neighborhood (Altman 1992). In the first step, the distance between the new value and each training point is computed using the most commonly known approaches, e.g., the Euclidean

distance, Manhattan distance, or Hamming distance. Subsequently, according to the calculated distance, the closest K value is chosen regarding the error calculation for the training and testing datasets. Lastly, the average of these data is estimated to predict the new value. In the current study, the hyperparameters of the KNR algorithm for estimation of the iceberg drafts were set utilizing a trial and error approach, wherein *n_neighbors*=5, *weights='uniform'*, *algorithm='auto'*, *leaf_size*=1, *p*=1, *metric='minkowski'*, *metric_params=None*, *n_jobs=None*.

8.1.6.4. Iceberg draft

The iceberg draft (D) was considered in terms of the physical characteristics of the iceberg, comprising the iceberg length (L), iceberg height (H), iceberg width (w), and iceberg mass (M). Azimi et al. (2023) introduced the dimensionless groups governing the iceberg draft in the following form:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \tag{8-27}$$

Hence, the iceberg draft ratio (D/H) is a function of the length ratio (L/H), width ratio (w/H), the mass ratio $(M/\rho_i . H^3)$, and iceberg shape factor $\left(S_f\right)$ (Azimi et al. 2023). In the present research, the ETR algorithm was fed with the dimensionless groups in equation (8-1) as input parameters to simulate the iceberg draft ratio.

Figure 8-13 illustrates the composition of four dimensionless groups to create the ML models. As seen, to identify the premium ML models, five ML models, e.g., Model 1 to Model 5 were developed, while Model 6 to Model 9 were defined to recognize the most influencing input parameters. Model 1 included all input factors, whilst those dimensionless groups were ignored one at a time in Model 2 to Model 5.

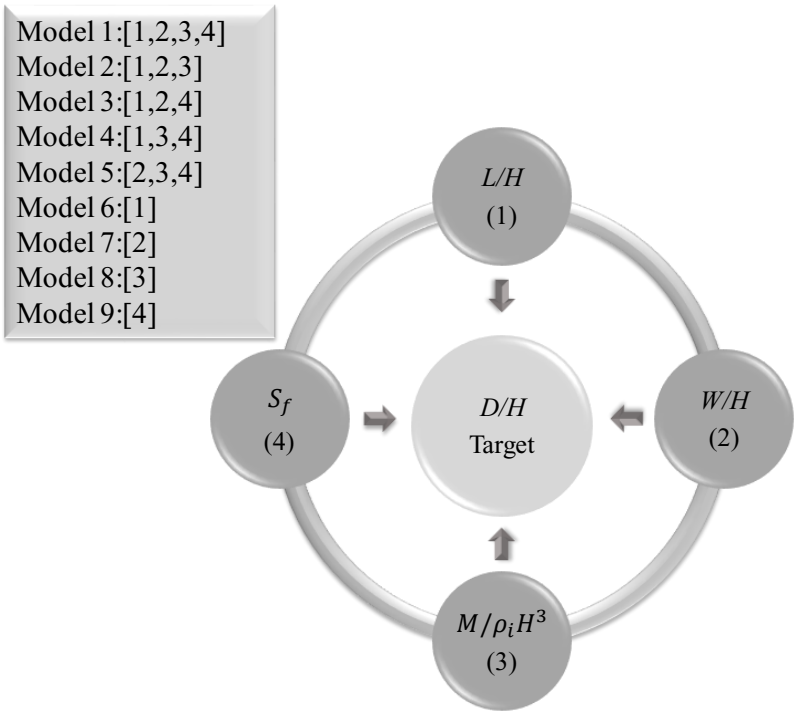


Figure 8-13. Input combination applied for developing the ML models

8.1.6.5. Construction of Dataset

The observational values of 12 field investigations documented by El-Tahan et al. (1985) (38 cases), Woodworth-Lynas et al. (1985) (one case), Løset and Carstens (1996) (52 cases), Barker et al. (2004) (14 cases), McKenna (2004) (two cases), Sonnichsen et al. (2006) (nine cases), Turnbull et al. (2015) (two cases), McGuire et al. (2016) (eight cases), Younan et al. (2016) (29 cases), Talimi et al. (2016) (one case), Zhou (2017) (three cases), Turnbull et al. (2018) (two cases) were applied in the current study to build the dataset (Azimi et al. 2023). Table 8-2 presents the maximum, minimum, average, variance, and standard deviation of the field observations. To show the significance of the constructed dataset, the T-test and the P-value were calculated, assuming that the P-value of 0.05 or less is statistically significant (Azimi and Shiri 2020), where the likelihood of the relationship between the observational amounts is influenced by an alternative

hypothesis. The P-value for the built database was calculated as 0.008, expressing that the correlations were statistically significant.

Table 8-2. Maximum, minimum, average, variance, and standard deviation of the field observations

Value	*D(m)*	*H(m)*	*w(m)*	*M(kg)*	*L(m)*	S_f
Max	240	94	408	1.13E+13	499	0.58
Min	18	3.9	9.672	8317400	34	0.11
Ave	88.988	30.956	101.605	1.68E+11	144.522	0.298
Var.	1696.272	399.389	3733.368	1.14E+24	6546.271	0.011
Std.	41.186	19.985	61.101	1.07E+12	80.909	0.107

8.1.6.6. Goodness of fit

A set of benchmarks like correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were employed to assess the accurateness, correlation, and complexness of the ML models. The nearness of the *R* and *WI* indices to one indicating the ML model had a high level of correlation with the observational measurements. The proximity of the *RMSE*, *MAPE*, and *CRM* criteria to zero denoting the ML model included the lowest grade of impreciseness, whereas the complexity of the ML models was not scrutinized via these indices. To solve this problem, the Akaike Information Criteria (*AIC*) was operated. It means that the less complex ML model had the lowest value of *AIC*, so the best ML model might have the lowest

degree of *AIC* index and error (*RMSE*, *MAPE*, and *CRM*), with the highest amount of correlation (*R* and *WI*) (Azimi et al. 2022a).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \quad (8\text{-}28)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (8\text{-}29)$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \quad (8\text{-}30)$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (8\text{-}31)$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \quad (8\text{-}32)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (8\text{-}33)$$

Here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, *n* and *k* are respectively the field measurements, the simulated values, the average field values, the average simulated amount, the number of experiments, and the number of independent variables in the ML models.

8.1.7. Results and discussion

The results of the GBR models were assessed and the best GBR models along with the most influencing input parameters were detected in the sensitivity analysis. The performance of the premium GBR model was compared to the SVR and KNR algorithms by conducting several statistical analyses.

8.1.7.1. Sensitivity analysis

Figure 8-14 shows the comparison of the calculated statistical indices for the GBR, SVR, and KNR models. The GBR 1, SVR 1, and KNR 1 models predicted the iceberg drafts adopting L/H , w/H , $M/\rho_i.H^3$, and S_f. The value of the *RMSE* index for the GBR 1, SVR 1, and KNR 1 models was obtained at 1.348, 1.395, and 1.617, respectively. For the GBR 2, SVR 2, and KNR 2 models, the influence of the iceberg shape factor (S_f) was ignored for the estimation of the iceberg drafts, signifying that these models were a function of L/H , w/H , and $M/\rho_i.H^3$, with an AIC criterion of 18.964, 15.250, and 19.299, respectively. The GBR 3, SVR 3, and KNR 3 models utilized the L/H , w/H , and S_f inputs to forecast the iceberg drafts, whereas the effect of iceberg mass ratio ($M/\rho_i.H^3$) was eliminated, where the WI value for these models were 0.860, 0.870, and 0.915. Using the L/H , $M/\rho_i.H^3$, and S_f input parameters, the iceberg drafts were simulated through the GBR 4, SVR 4, and KNR 4 models, rather than the w/H the factor was eliminated. The *CRM* statistical index for GBR 4, SVR 4, and KNR 4 were respectively equal to 0.138, 0.104, and 0.120. The iceberg length ratio (L/H) was an eliminated input factor in the GBR 5, SVR 5, and KNR 5 models although these models were fed with w/H , $M/\rho_i.H^3$, and S_f input parameters, where the *AIC* value for GBR 5, SVR 5, and KNR 5 was computed as 21.407, 20.967, and 20.295. It is worth noting that models 6 to 9 were individually a function of the L/H , w/H , $M/\rho_i.H^3$, and S_f input factors. For example, the value of *RMSE* for the GBR 6, SVR 6, and KNR 6 models were 1.692, 1.497, and 1.433.

The performed sensitivity analysis demonstrated that the GBR 4 model was identified as the superior model among all GBR models, while SVR 2 and KNR 3 were detected as the premium models amongst the SVR and KNR models. Moreover, the iceberg length ratio (*L/H*) and iceberg

width ratio (w/H) had the highest degree of effectiveness to model the iceberg drafts; however, the iceberg shape factor (S_f) showed an insignificant influence.

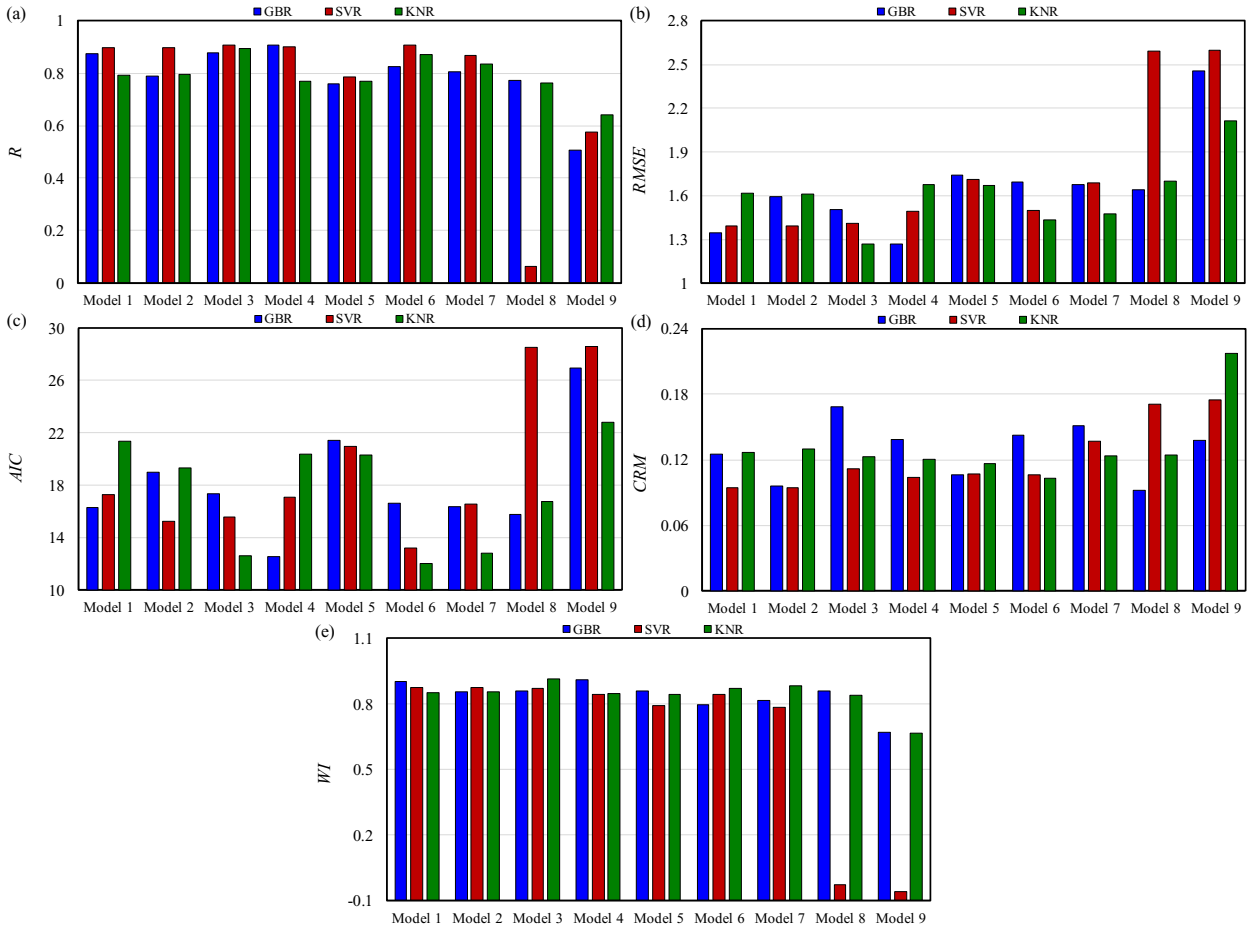


Figure 8-14. Comparison of the GBR, SVR, and KNR models' performance (a) R (b) RMSE (c) AIC (d) CRM (e) WI

8.1.7.2. Superior models

Among all GBR, SVR, and KNR models, the GBR 4, SVR 2, and KNR 3 models were recognized as the superior ML models to simulate the iceberg drafts. The scatter plots for these superior models are depicted in Figure 8-15. According to simulation results, the values of the correlation coefficient (R) for the GBR 4, SVR 2, and KNR 3 models were surmised to be 0.907, 0.896, and

0.893, respectively. As shown, the superior model possessed a high level of correlation in comparison with other ML models. Furthermore, the lowest degree of complexity was obtained for the GBR 4, SVR 2, and KNR 3 models, where the *AIC* index for these models was to be 12.571, 15.250, and 12.633, respectively. Thus, the GBR 4 model showed the highest degree of correlation and simplicity among the superior ML models.

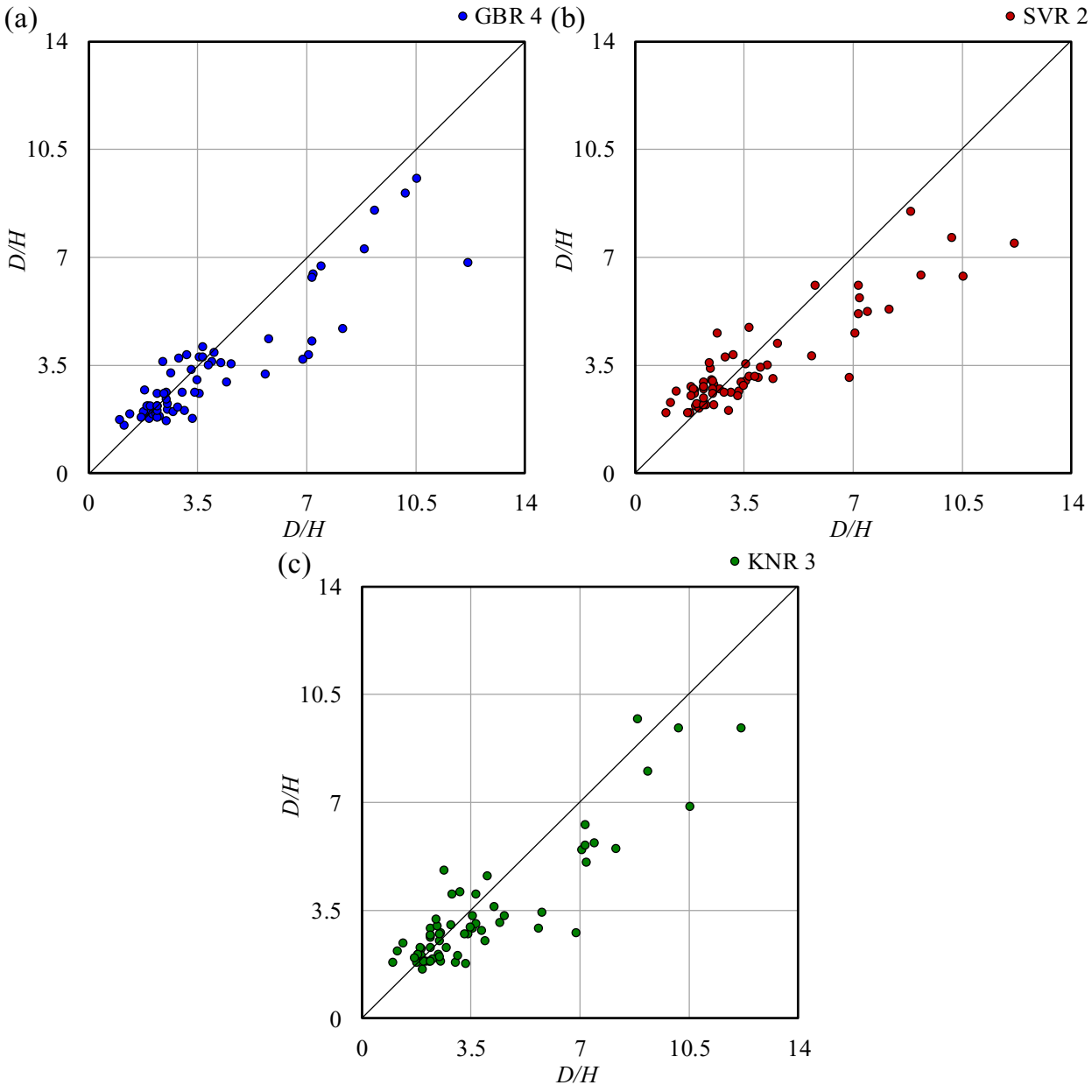


Figure 8-15. Scatter plots for (a) GBR 4 (b) SVR 2 (c) KNR 3 as the superior models

The iceberg drafts simulated by the GBR 4, SVR 2, and KNR 3 models are compared with the iceberg drafts observed in Figure 8-16. As can be seen, even though some discrepancies were witnessed between the results simulated with the iceberg drafts observed, the superior ML models attempted to forecast the target parameter with their highest degree of performance. However, the GBR 4 model predicted the iceberg drafts with higher accuracy (*RMSE*=1.267) in comparison with the SVR 2 (*RMSE*=1.395) and KNR 3 (*RMSE*=1.270). Hence, the simulation outcomes proved that the GBR 4 demonstrated the highest level of precision among the superior ML models.

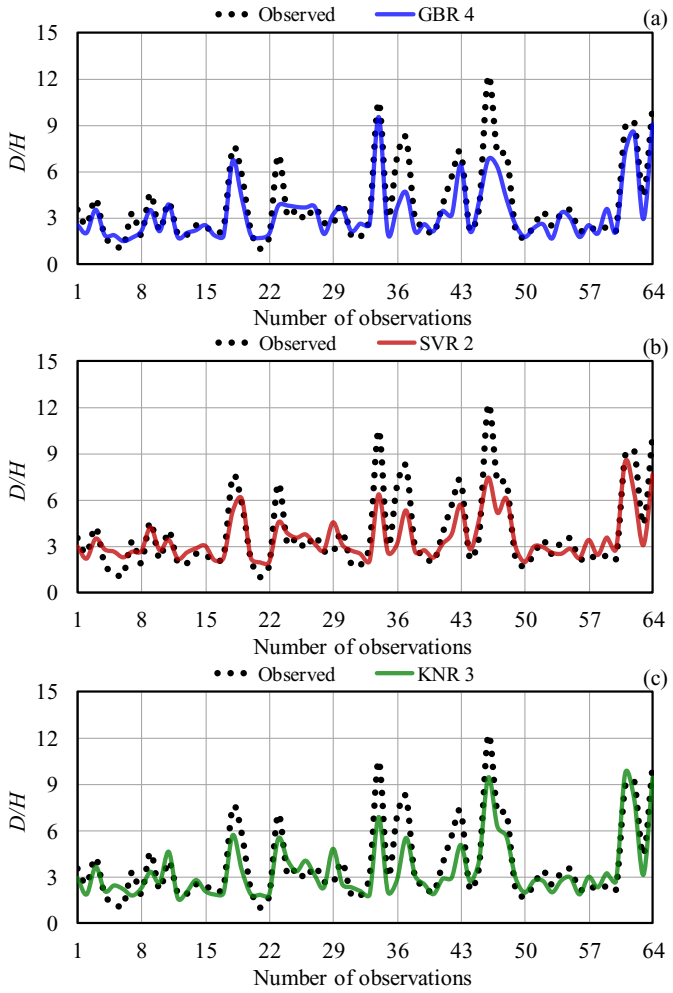


Figure 8-16. Comparison between the observed iceberg drafts and (a) GBR 4 (b) SVR 2 (c) KNR 3 models

Figure 8-17 illustrates the results of the error analysis for the GBR 4, SVR 2, and KNR 3 models. Regarding the error analysis conducted, approximately one-third of the iceberg drafts simulated by GBR 4 possessed an error of less than 10%; rather, nearly one-fifth of the results obtained from the SVR 2 and KNR 3 models had an error of smaller than 10%. Almost 28% of the outcomes achieved from the GBR 4 model demonstrated an error between 10% and 20%, whilst this amount for SVR 2 and KNR 3 was estimated as 26% and 31%. Roughly half of the iceberg drafts modeled using KNR 3 and SVR 2 showed an error of greater than 20% though this figure for GBR 4 was about 38%. Therefore, the GBR 4 model managed to simulate the iceberg draft values with the lowest level of incorrectness.

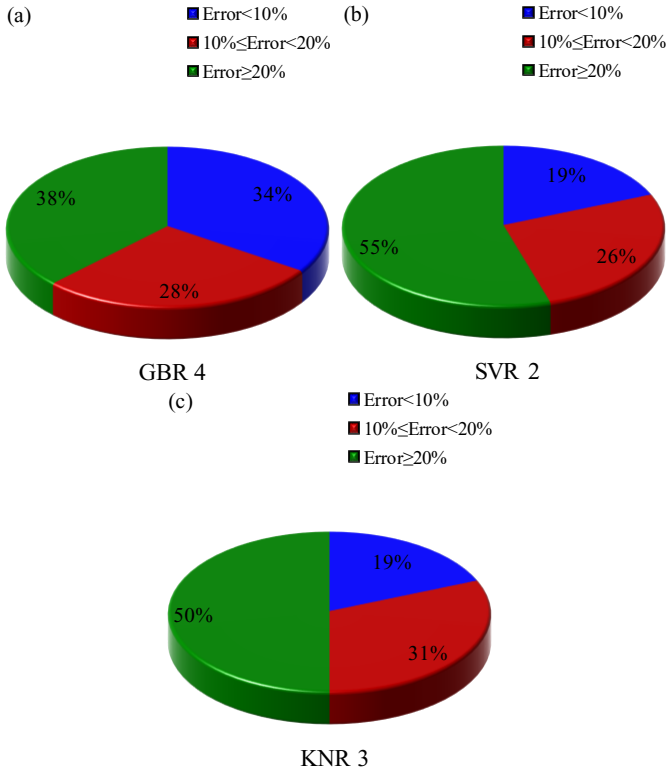


Figure 8-17. Results of the error analysis for (a) GBR 4 (b) SVR 2 (c) KNR 3 models

The efficiency of the premium ML models was assessed through the discrepancy ratio (*DR*) as follows (Azimi et al. 2022b):

$$DR = \frac{P_i}{O_i} \tag{8-34}$$

where, P_i and O_i are respectively the predicted and observed iceberg drafts. Generally, the closer the magnitude of *DR* is to the unity, the higher performance demonstrates the ML model. The value of the maximum ($DR_{(\max)}$), minimum ($DR_{(\min)}$), and average ($DR_{(\text{ave})}$) discrepancy ratio for the superior models are drawn in Figure 8-18. For the GBR 4, SVR 2, and KNR 3 models, the maximum discrepancy ratio ($DR_{(\max)}$) was reckoned at 1.704, 1.998, and 1.916; however, the minimum discrepancy ratio ($DR_{(min)}$) for these models equaled as 0.526, 0.453, and 0.402, respectively. It is worth mentioning that the average discrepancy ratio ($DR_{(ave)}$) for GBR 4, SVR 2, and KNR 3 were at 0.943, 1.049, and 0.971. As seen, the lowest value of $DR_{(max)}$ and the highest amount of $DR_{(min)}$ were acquired for the GBR 4 model.

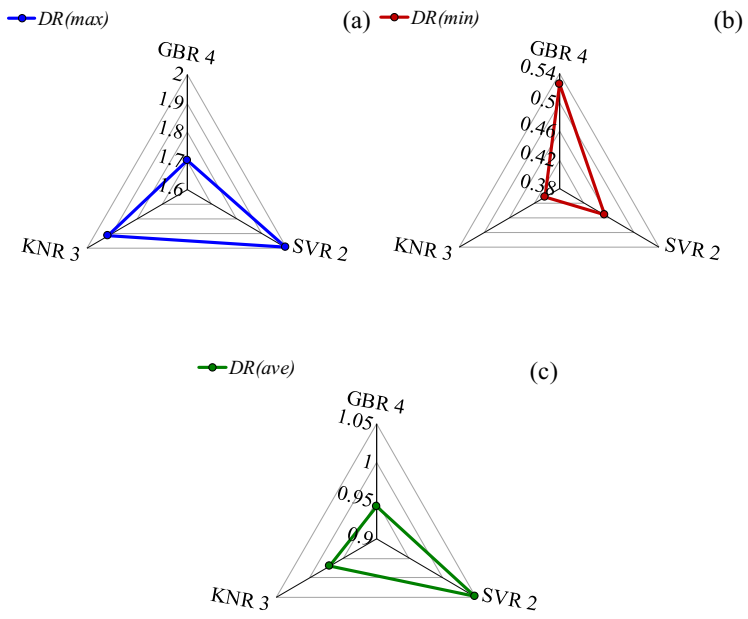


Figure 8-18. Value of (a) maximum discrepancy ratio (DR(max)) (b) minimum discrepancy ratio (DR(min)) (c) average discrepancy ratio (DR(ave)) for the superior models

In order to evaluate the performance of the superior ML models, an uncertainty analysis was also implemented. In other words, the difference between the predicted iceberg drafts (P_j) and the observed values (O_j) is defined as the error (e_j) as below:

$$e_i = P_i - O_i \tag{8-35}$$

The mean (Mean) and the standard deviation (StDev) of the error values were calculated using the equations below:

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} e_i \tag{8-36}$$

$$\text{StDev} = \sqrt{\sum_{i=1}^{n} (e_i - \text{Mean})^2/(n-1)} \tag{8-37}$$

Assuming that the sign of the Mean error was positive, the ML model overestimated the iceberg drafts; otherwise, if the Mean error had a negative sign, the ML model tended to underestimate the target values. By disregarding the continuity correction, a confidence interval (CI) was created around the values of error utilizing the Mean, StDev values, and the "Wilson score approach". It is worth mentioning that an asymmetric normal distribution interval entitled the Wilson score interval was utilized to rectify the CI bounds; hence, a ±1.96Se brought about a 95% of confidence interval (95%CI). Moreover, the width of uncertainty bound (WUB) of the superior ML models was calculated as follows (Azimi et al. 2022b):

$$\mathrm{WUB} = \pm\frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{8-38}$$

Figure 8-19 illustrates the error normal distribution of the GBR 4, SVR 2, and KNR 3 models. Additionally, the results of uncertainty analysis for the superior ML models are exhibited in Figure 8-20. Regarding the uncertainty analysis, the value of Mean error for the GBR 4, SVR 2, and KNR 3 models was approximated at 0.533, 0.363, and 0.474, so the superior models biased toward overestimation, with a StDev of 1.158, 1.358, and, 1.187, respectively. The GBR 4 had the narrowest width of uncertainty bound (WUB=±0.290), whereas the widest WUB was for SVR 2 (WUB=±0.339).

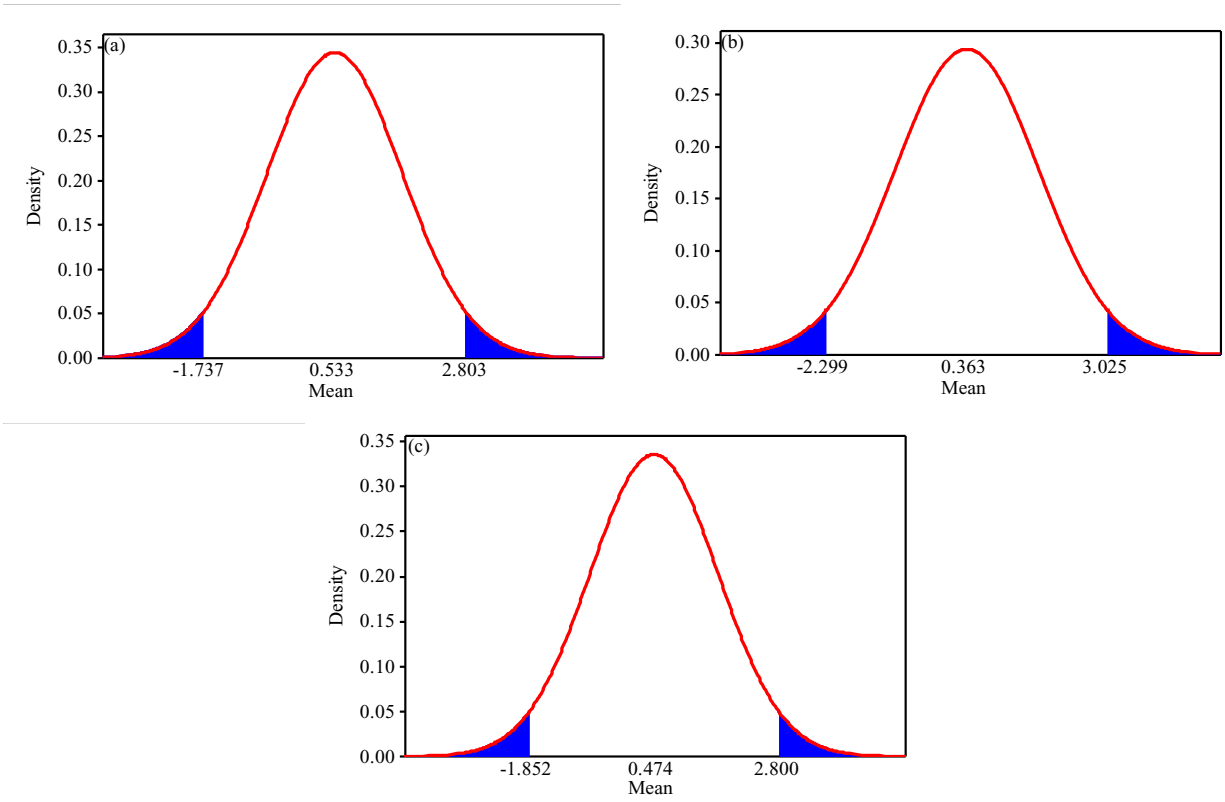


Figure 8-19. Error normal distribution of (a) GBR 4 (b) SVR 2 (c) KNR 3 models

The lower and upper bound of uncertainty for the KNR 3 model was 0.178 and 0.771, with a WUB of 0.297. It should be stated that the lower and upper uncertainty bound for the GBR 4 was computed to 0.243 and 0.822 but these values for the SVR 2 were at 0.024 and 0.702.

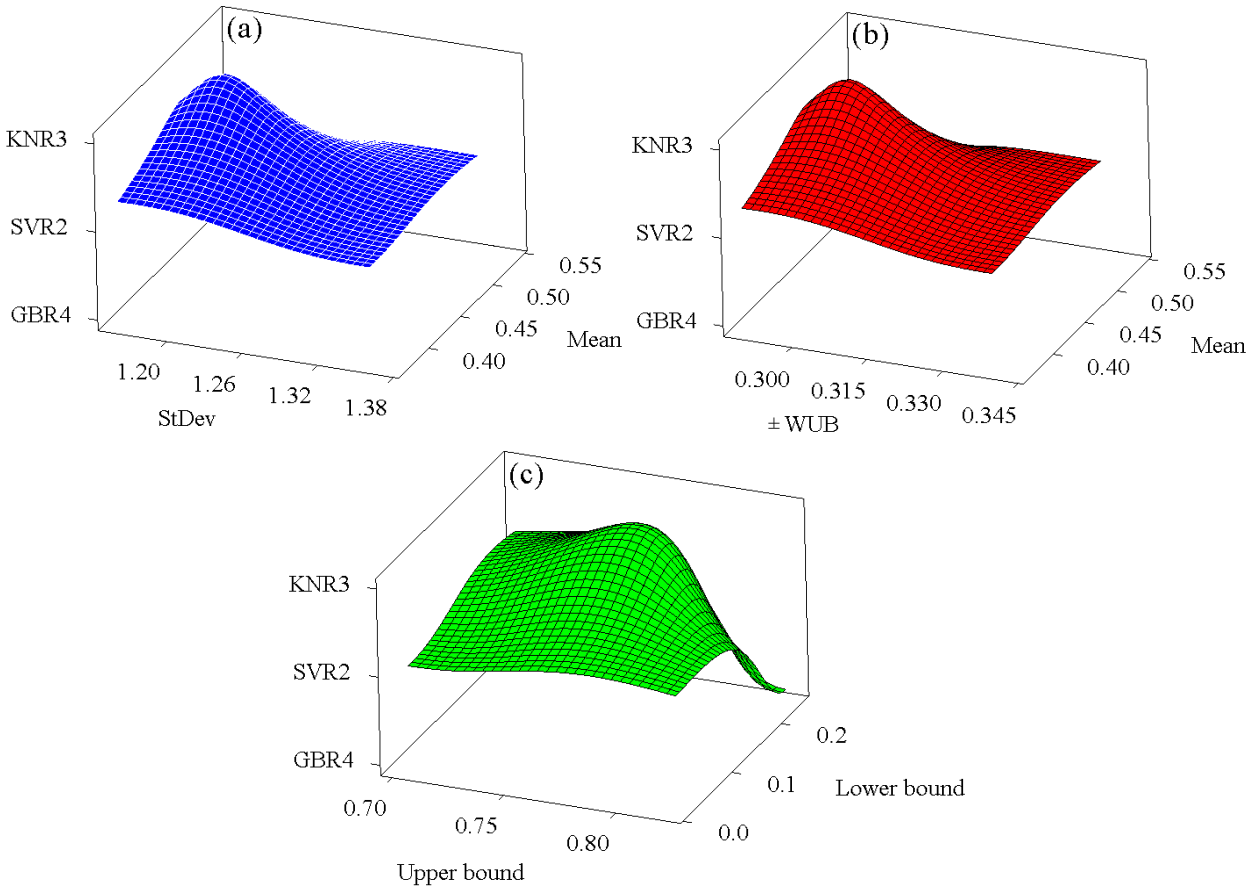


Figure 8-20. The results of uncertainty analysis for the superior ML models (a) Mean and StDev (b) Mean and WUB (c) Lower bound and Upper bound

As shown, the GBR 4 model managed to simulate the iceberg drafts with better performance in terms of accuracy, correlation, complexity, and discrepancy. Furthermore, the iceberg length ratio (L/H) and iceberg width ratio (w/H) input parameters had a remarkable influence on the iceberg drafts prediction by the GBR model.

8.1.8. Conclusion

In this investigation, the Gradient Boosting Regression (GBR) algorithm was utilized to model the iceberg drafts for the first time. In the first step, the dimensionless groups affecting the iceberg

drafts were found and nine GBR models, e.g., GBR 1 to GBR 9, were defined by employing these dimensionless groups as input parameters. The performance of the GBR models was assessed by conducting sensitivity analysis and then the premium GBR model as well as the most influencing input parameters were distinguished. Subsequently, the ability of the best GBR model was compared with Support Vector Regression (SVR) and K-Nearest Neighbors Regression (KNR) models. Several statistical analyses, such as error analysis, discrepancy analysis, and uncertainty analysis were performed to present a robust and reliable model for the iceberg draft prediction. The most important outcomes of the current study are summarized below:

- Amongst the GBR models, the GBR 4 model as a function of L/H , $M/\rho_i.H^3$, and S_f was able to predict the iceberg drafts with the best performance.
- The value of *R*, *RMSE*, and *AIC* indices for the GBR 4 model was respectively reckoned to be 0.907, 1.267, and 12.571.
- The sensitivity analysis showed that the iceberg length ratio (*L/H*) and iceberg width ratio (w/H) possessed the highest level of importance in approximating the iceberg drafts.
- The GBR 4 model had the highest degree of accuracy, correlation, and simplicity in comparison with the SVR and KNR methods.
- Roughly 34% of the iceberg draft predicted by the GBR 4 model had an error of less than 10%, whereas this amount for SVR and KNR models was 19%.
- The maximum discrepancy ratio (DR_{max}) for the GBR 4, SVR, and KNR models was obtained as 1.704, 1.998, and 1.916, respectively.
- Regarding the sensitivity analysis implemented, the GBR 4 model was biased towards overestimating, with the narrowest bound of uncertainty.

The presented results demonstrated that the ML-based algorithms are able to predict the iceberg drafts with acceptable performance. These models can significantly reduce the expenditures and downtime in the iceberg management system to guarantee the physical and operational integrity of the subsea structures against the deep-keel iceberg attacks in the Arctic region.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Allaire, P.E., 1972. Stability of simply shaped icebergs. Journal of Canadian Petroleum Technology 11: 04.

Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46: 175-185.

Azimi, H., Shiri, H., 2020. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142: 051601.

Azimi, H., Shiri, H., Mahdianpari, M., 2020a. Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. Journal of Pipeline Science and Engineering, 2: 100075.

Azimi, H., Shiri, H., Zendehboudi, S., 2022b. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Region Science and Technology, 193: 103426.

Azimi, H., Mahdianpari, M., Shiri, H., 2023. Determination of parameters affecting the estimation of iceberg draft. China Ocean Engineering, 37: 62-72.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In the Fourteenth international offshore and polar engineering conference, ISOPE-I-04-116.

CNLOPB, 2005. Canada-Newfoundland and Labrador Offshore Petroleum Board. St. John's, NL.

Dowdeswell, J.A., Bamber, J.L., 2007. Keel depths of modern Antarctic icebergs and implications for sea-floor scouring in the geological record. Marine Geology, 243: 120-131.

El-Tahan, M., El-Tahan, H., Courage, D., Mitten, P., 1985. Documentation of Iceberg Groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Boosting and additive trees. In The elements of statistical learning, Springer, New York, pp. 337-387.

Hotzel, I.S., Miller, J.D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

https://www.arcticwwf.org/threats/oil-and-gas/

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Arctic Technology Conference, OTC-27358-MS.

Løset, S., Carstens, T., Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Regions Science and Technology, 24: 323-340.

McGuire, P., Younan, A., Wang, Y., Bruce, J., Gandi, M., King, T., Keeping, K. and Regular, K., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

McKena, R., 2000. Study of iceberg scour & risk in the Grand Banks region. KR Croasdale & Associates Ltd., Ballicater Consulting Ltd., Canadian Seabed Research Ltd., C-CORE, and Ian Jordaan & Associates Inc., PERD/CHC Report, pp. 31-26.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., Crocker, G., Bruneau, S. and German, P., 2019. Modelling iceberg grounding on the grand banks. Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, POAC, pp. 9-19.

Nematzadeh, A., Shiri, H., 2020. The influence of non-linear stress-strain behavior of dense sand on seabed response to ice-gouging. Cold Region Science and Technology, 170, 102929.

Robe, R.Q., Farmer, L.D., 1976. Physical Properties of Icebergs. Part I. Height to Draft Ratios of Icebergs. Part II. Mass Estimation of Arctic Icebergs. Coast Guard Research and Development Center Groton Conn.

Rudkin, P., 2014. Comprehensive Iceberg Management Database Report 2005 update. PERD/CHC Report 20-72. National Research Council of Canada (NRC) and Panel on Energy Research and Development (PERD).

Sacchetti, F., Benetti, S., Cofaigh, CÓ., Georgiopoulou, A., 2012. Geophysical evidence of deep-keeled icebergs on the Rockall Bank, Northeast Atlantic Ocean. Geomorphology, 159, 63-72.

Sonnichsen, G., R. Myers, K. Simpson, Rob Brown, J. McClintock, T. Bullock, P. Rudkin., 2003. Seabed Surveys to Document Groundings from the 2000 Grand Bank Iceberg Season. Proceedings of the 17th International Conference on Port and Ocean Engineering under Arctic Conditions. POAC 03.

Sonnichsen, G., Hundert, T., Myers, R., Pocklington, P., 2006. Documentation of Recent Iceberg Grounding Evens and a Comparison with Older Events of Know Age. Environmental Studies Research Funds. Report ESRF, 157.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L. and King, T., Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. In International Conference on Offshore Mechanics and Arctic Engineering, 85178, pp. V007T07A018.

Talimi, V., Ni, S., Qiu, W., Fuglem, M., MacNeill, A. Younan, A., 2016. Investigation of Iceberg Hydrodynamics. Arctic Technology Conference, OTC-27493-MS.

Turnbull, I.D., Fournier, N., Stolwijk, M., Fosnaes, T. McGonigal, D., Operational iceberg drift forecasting in Northwest Greenland. Cold Region Science and Technology, 110: 1-18.

Turnbull, I.D., King, T. Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Woodworth-Lynas, C.M.T., Simms, A. and Rendell, C.M., 1985. Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Region Science Technology. 10, 163-186.

Younan, A., Ralph, F., Ralph, T. and Bruce, J., 2016. Overview of the 2012 iceberg profiling program. Arctic Technology Conference, OTC-27469-MS.

Zhou, M., 2017. Underwater iceberg profiling and motion estimation using autonomous underwater vehicles, Doctoral dissertation, Memorial University of Newfoundland, NL, Canada.

Section 3

Iceberg Draft Evaluation through Robust Tree-Based Machine Learning Algorithms

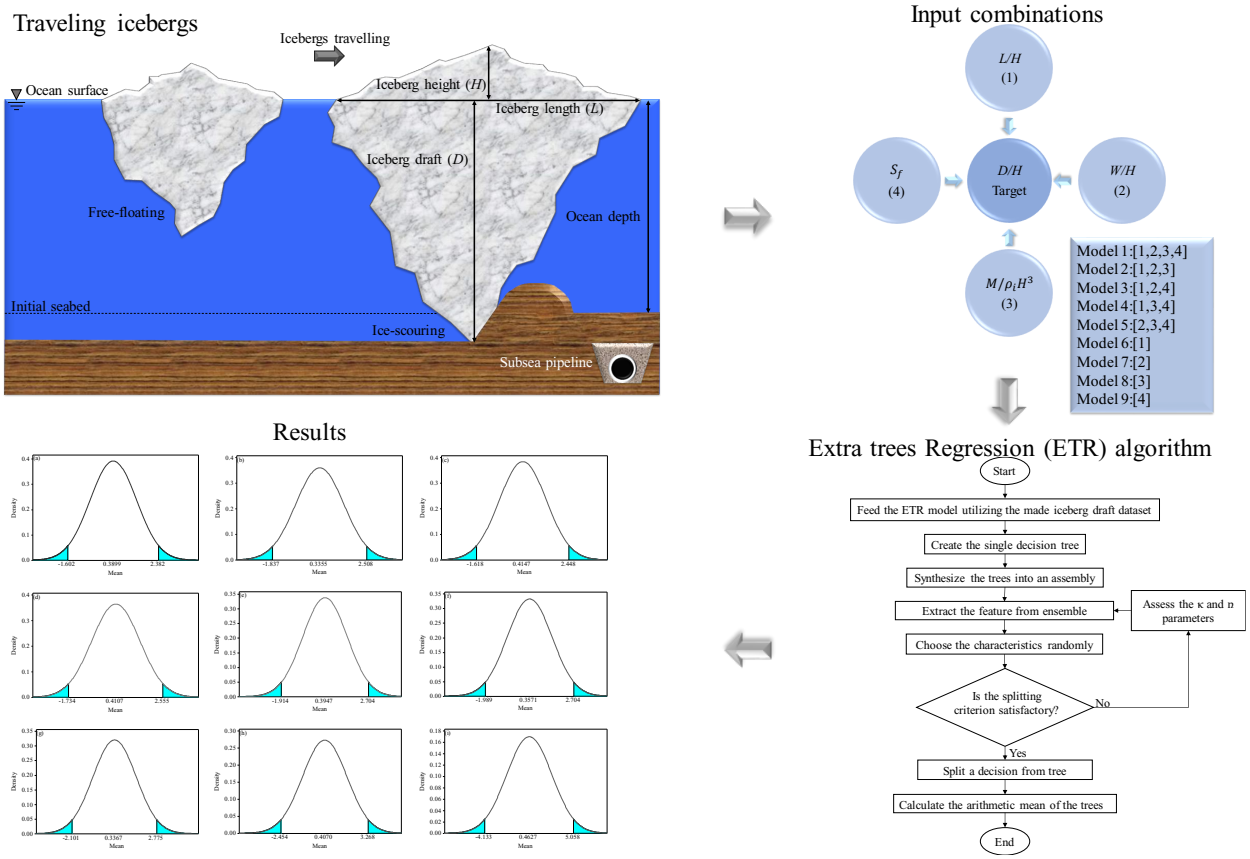
This section was submitted as a journal manuscript

Abstract

Climate change is causing the polar ice caps and glaciers in the Arctic to melt faster, leading to an increase in the number of icebergs that break off each year. These traveling icebergs are a potential threat and one of the governing design factors of subsea assets crossing the ocean floor in the shallower waters when the iceberg's tip scratches the seabed, and the integrity of subsea structures may be impaired. Hence, iceberg management operations with significant downtime and expenses, such as iceberg towing and re-routing, are currently performed to protect the subsea infrastructures, offshore structures, and ships. In this study, Extra Tree Regression (ETR) was used as a cost-effective and efficient method to estimate the draft of icebergs in the early stages of decision-making projects for managing icebergs. To end this, nine ETR models were defined using the parameters affecting the iceberg drafts, then the superior ETR model alongside the most influential inputs was introduced by performing several analyses. The uncertainty analysis demonstrated that the superior ETR model was biased toward the overestimation of the iceberg drafts, and the best ETR model simulated the iceberg draft with the highest level of precision, correlation, and simplicity. The comparison between the best ETR model with decision tree regression (DTR), random forest regression (RFR), and empirical approaches demonstrated the outperformance of the ETR algorithm to predict the iceberg drafts.

Keywords: Subsea assets, Offshore structures, Iceberg draft, Decision tree regression (DTR), Random forest regression (RFR), Extra tree regression (ETR)

Graphical abstract



8.1.9. Introduction

Roughly one-fifth of the Earth's undiscovered hydrocarbons are situated in the Arctic area (WWF Arctic Programme), and the Arctic offshore regions with rich wind culture also have a high potential for developing offshore wind farms (CNLOPB, 2005). However, thousands of icebergs are born every year out of glaciers in the Arctic zone and carried away by the currents into the North Atlantic. These icebergs may touch the ocean bottom in shallow waters and scratch the seabed, causing so-called "ice-gouging" that may endanger the physical and operational integrity of subsea pipelines and power cables or even directly collide with offshore structures like ships, platforms, wind turbines, and subsea manifolds. Currently, so-called "Ice Management," i.e., iceberg towing and re-routing, is the most reliable approach to protect the subsea and offshore infrastructures, where the threatening icebergs are hooked and towed in a safe direction. This operation was conducted under a collaborative program between the North Atlantic and Europe called the International Ice Patrol (IIP) program, which was established after the well-known Titanic sank in 1912. Thus, ice management operations are costly and require standby marine spread with a range of advanced tools and equipment, such as subsea survey facilities, vessels, and expert crews, to investigate the iceberg draft and determine if it is a threat to infrastructures.

In addition to using underwater survey equipment such as side scan sonars and multi-beam echo sounders, there has been a growing need for data-based methods to predict the draft of icebergs, which has led to numerous studies and investigations. Despite the offshore structures such as ships, wind turbines, and floating or fixed platforms that icebergs can directly attack, the threat to the buried pipelines and power cables is much more complicated involving the ice-soil-structure interactions. During the ice-gouging, the subgouge soil displacement is vastly extended down the

seabed much deeper than the gouge depth, and this may cause large deformations in the pipelines buried even below the gouge depth.

The schematic layout of the iceberg free-floating and iceberg scouring in cold waters is demonstrated in Figure 8-21. As shown, the iceberg is in a free-floating situation if the ocean depth is greater than the iceberg draft; otherwise, the seafloor is scoured, and the seabed soil shear resistance causes the soil displacement to extend more profound than the iceberg tip threatening the buried subsea assets.

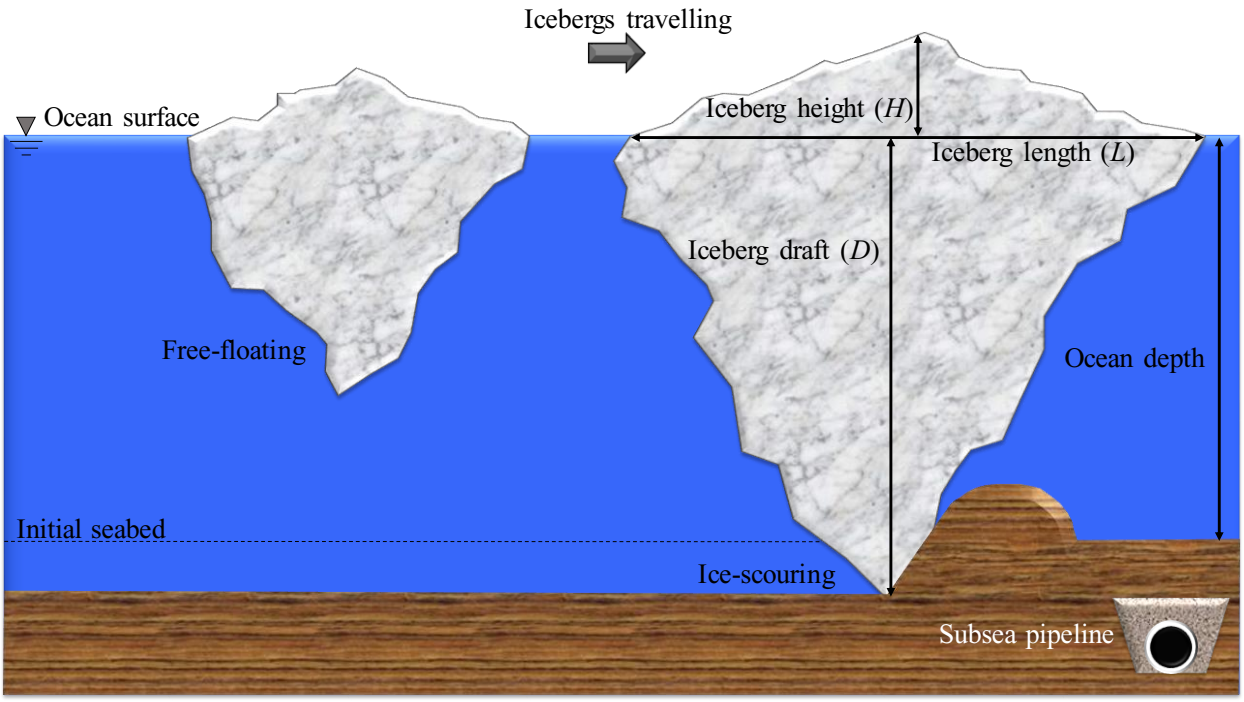


Figure 8-21. Icebergs in free-floating and scouring conditions schematically

The efficient iceberg management designs and the guaranteed operational integrity of the sea bottom-funded infrastructure against the iceberg attacks in the ice-prone areas demand the appropriate iceberg draft appraisal, which may lead to a potential decrease in operating expenses and downtime. Earlier investigations have focused on modeling the iceberg draft by using the

iceberg length or mass. For instance, Allaire (1972) analyzed the stability status of various icebergs regarding the sail features. They demonstrated that the minimum stable ratio of iceberg width to sail height for tabular, dry dock, and dome bergs were 6:1, 4:1, and 1.8:1, respectively. Similarly, Robe and Farmer (1976) measured the drafts of tabular, broken tabular, pinnacle, dry dock, and domed by the sonar technology. They correlated the iceberg draft with the iceberg height and proposed a regression model for estimating the iceberg draft. Bass (1980) analytically evaluated the stability of icebergs in different configurations. The study highlighted that the volume of the underwater section was almost seven times greater than the volume of the above-water area. Brooks (1980) evaluated iceberg lengths and drafts to decide whether an iceberg's length might be regarded as its utmost potential draft. The maximum iceberg length and draft were reported to be 599m and 219m in this survey. In another investigation, Hotzel and Miller (1983) assessed the iceberg dimensions and supposed that the planimetric section of icebergs was circular, while the length of a berg was twice the circle radius. The study estimated the iceberg drafts in terms of length using the power-law approach.

Furthermore, the IIP has been annually monitoring the characteristics of traveling icebergs in the Atlantic and Arctic Oceans. In 1987, this organization IIP reported that the iceberg draft was 3.95 times the height of a sail. PERD (1999) distinguished the data collected from traveling icebergs observed in the Labrador Sea and Grand Banks and then provided a relationship for the iceberg draft prediction in terms of length. Liang (2001) presented a model to determine the relationship between the dynamics and stability of icebergs using the size and shape of bergs. The author asserted that the draft and hydrostatic force distribution affected the iceberg's stability.

The Centre for Cold Ocean Resources Engineering (C-CORE) performed several investigations to study icebergs' shape, behavior, and dynamics over the years. The above-water form of bergs

comprising the blocky, domed, dry dock, tabular, pinnacle, and wedge icebergs were reported regarding the standard categories. The iceberg dimensions were recorded using the scan sonar technology and marine sextant. The study showed that the iceberg draft was correlated with iceberg length (C-CORE, 2001). In another study, Barker et al. (2004) evaluated the geometry of iceberg sails and keels. They estimated the cross-sectional areas of the berg at different water depth intervals from a particular waterline length. The authors provided a set of models in terms of the length using the power curve and regression analysis to approximate the iceberg draft. Dowdeswell and Bamber (2007) examined the keel depths of traveling icebergs in Antarctic waters. The authors estimated the depth of the keel through the ice thickness and surface elevation. The research concluded that a tiny minority of icebergs in the Antarctica and Greenland waters had drafts greater than 650m. Moreover, Stuckey (2008) modeled the iceberg drift speed using the probabilistic approach. The author stated that the above-water dimension, below-water dimension, and shape factor of the icebergs impacted the environmental driving forces. McKenna and King (2009) simulated the deterioration mechanisms of various icebergs by considering the incremental changes in draft, mass, and shape of bergs. The study highlighted that decreasing the iceberg mass reduced the draft and length of icebergs. In another study, Sacchetti et al. (2012) analyzed the features of different icebergs, such as wedged, domed, tabular, and pinnacle and ice scouring in the Northeast Atlantic Ocean. The investigation reported that the bimodal distribution of the scour depth was observed across the study area. Turnbull et al. (2015) forecasted the drift trajectory of the traveling icebergs in Northwest Greenland using the hindcast simulation. It was revealed that the trajectory of icebergs was sensitized to the iceberg drafts. King et al. (2016) performed a field investigation to calculate the rolling iceberg rate. The iceberg drafts were estimated utilizing a calving analysis, with a calculated standard deviation of draft variations from 19% to 34%. The

iceberg drafts corresponded with the mass of the icebergs. Turnbull et al. (2018) proposed a model for the drift estimation of moving icebergs on the Grand Banks of Newfoundland. This model approximated the draft of icebergs roughly 1.3 times more than the actual values. McKenna et al. (2019) have recently simulated ice scouring on the Grand Banks of Canada using the Monte Carlo method. The iceberg draft alterations were utilized to lessen the size of draft variations in this modeling. Most recently, Stuckey et al. (2021) simulated the three-dimensional iceberg shapes by adopting field measurements. The investigation demonstrated that the power curve estimated the draft and mass in terms of the iceberg length.

Machine learning (ML) technology has been widely applied to solve different linear and nonlinear problems in various fields such as environmental monitoring, remote sensing, and geotechnical engineering (Azimi and Shiri 2021; Azimi et al. 2021; Mahdianpari 2021; Azimi et al. 2022), due to its reliability, precision, cost-effectiveness, and speed. However, despite its potential, the literature reveals that ML has not yet been investigated in simulating iceberg drafts. Hence, to fill this knowledge gap, iceberg drafts were modeled through three robust tree-based ML algorithms comprising decision tree regression (DTR), random forest regression (RFR), and extra tree regression (ETR). More descriptions are provided in the upcoming sections.

8.1.10. Methodology

The applied ML algorithms in the present study, including the DTR, RFR, and ETR, are initially introduced. Subsequently, the parameters affecting the iceberg drafts were determined, and then nine ETR models were developed using the presented parameters. A comprehensive dataset was constructed using the reported iceberg characteristics in the literature to train and test the ML models. Finally, different statistical criteria were employed to assess the ML models' performance.

8.1.10.1. Decision Tree Regression (DTR)

A tree data system contains a series of leaves and branches in which each node is regarded as a decision tree (DT). The DT may be used to solve both regression and classification problems. The DT comprises many components, such as a root node, several leaf nodes, internal nodes, and branches. The topmost node in this tree is considered the root node, and the leaf nodes (terminal nodes) end with the titles of types, while the non-leaf nodes are assumed as the internal nodes. Such nodes link to each other through the branches (Pekel 2020). In the present work, the mean squared error (*MSE*) is involved in maintaining the fitness function in the DT algorithm.

In this investigation, the hyperparameters of the DTR algorithms were determined using a trial-and-error strategy. It means that the value of *max_depth*, *max_leaf_nodes*, and *min_weight_fraction_leaf* was primary at 10, 2, and 0.01, and the performance of the DTR algorithm was assessed. The number of hyperparameters was raised in the following stages until the DTR's results reached an adequate level. The DTR model estimated the iceberg drafts with its highest level of precision and correlation as well as its lowest degree of complicatedness until the number of hyperparameters comprising the *max_depth*, *max_features*, *max_leaf_nodes*, *min_samples_leaf*, *min_weight_fraction_leaf*, and splitter was, in turn, adapted as 150, *'auto'*, 2, 2, 0.001, and *'random'*.

8.1.10.2. Random Forest Regression (RFR)

Random forest (RF) has been proposed by Breiman (2001) to model different problems like classification and regression data. The RF is an ensemble algorithm, an extended version of the Classification and Regression Trees (CART) learning method. However, the RF can deal with

multi-source and high-dimensional data without over-fitted problems (Breiman 2001). Multiple decision trees are constructed in the RF algorithm, and decisions with the highest value of votes are selected as the modeling outcomes (Sahani and Ghosh 2021). The RFR algorithm decreases the mean error of the trees employed using the weighted correlation between the applied randomization and the residuals.

The hyperparameters of the RFR algorithm were also chosen using a trial and error method, similar to the DTR model. In particular, the initial number of *max_depth*, *random_state*, and the *number of estimators* were set to be 2, 2, and 2, respectively. The simulation results demonstrated that the RFR algorithm had the best performance when the *max_depth*, *random_state*, *number of estimators*, and *verbose* values were adjusted as 12, 6, 8, and 1, respectively.

8.1.10.3. Extra Tree Regression (ETR)

The ETR algorithm is an advanced version of the RF algorithm initially proposed by Geurts et al. (2006). To solve a regression problem, the RF algorithm operates bootstrapping, creating the decision trees, bagging, and separating the decision tree leaves. The ETR often builds several random regression trees. By selecting the most suitable subset, the decision-making process is performed. The ETR and RF models possess two major dissimilarities: (1) in the ETR model; these leaves are divided utilizing the random choice from the cutting points and (2) the ETR algorithm uses the whole training data to create the trees to minimize the value of bias. The ETR model (i) involves the number of features chosen randomly in the leave utilizing κ variable (ii) minimizes the specimen size for the splitting of leaves utilizing the ɲ variable. The κ value specifies the characteristic selection strength; however, the strength of the averaged result noise is expressed utilizing the ɲ variable. In other words, the κ and ɲ values will reduce the overfitting issue in the

ETR model and improve the efficiency of this algorithm. Figure 8-22 demonstrates the flowchart of the ETR algorithm. It is worth noting that the hyperparameters of the ETR model were adjusted using a trial-and-error strategy in the present work, as *max_depth=None*, *min_samples_split*=8, *min_samples_leaf*=1, *min_weight_fraction_leaf*=0, *max_features*=1.0, *random_state=None*, *min_impurity_decrease*=0.0, *max_leaf_nodes=None*, and *ccp_alpha*=0.0.

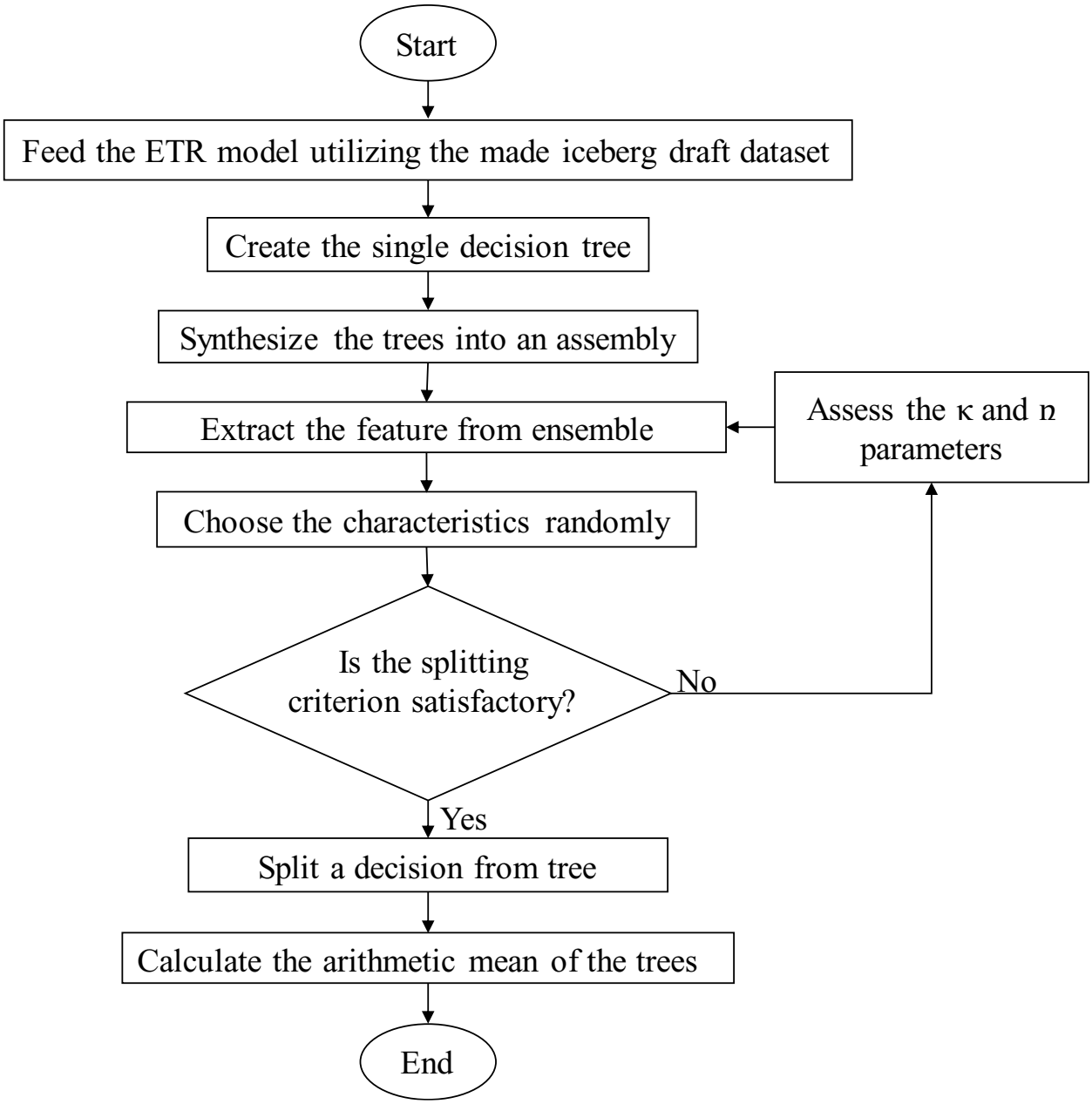


Figure 8-22. Flowchart of the ETR algorithm applied in the current study

8.1.10.4. Iceberg draft

The iceberg draft (D) was considered as a function of the physical characteristics of the iceberg, comprising the iceberg length (L), iceberg height (H), iceberg width (w), iceberg mass (M) in so many fields, analytical, and numerical investigations in the form below (Barker et al. 2004; McKenna et al. 2019; and Stuckey et al. 2021):

$$D = f_1(L, H, w, M). \tag{8-39}$$

Furthermore, the density of an iceberg (ρ_i), the density of seawater (ρ_{sw}), seawater viscosity (μ_{sw}), and gravitational acceleration (g) may influence the iceberg draft as follows:

$$D = f_2(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g) \tag{8-40}$$

The iceberg shape factor (S_f) signifies the global shape of icebergs, which can affect the magnitude of the iceberg draft (Turnbull et al. 2018). The shape factor of the traveling icebergs is considered universally into six categories, as illustrated in Figure 19-3 (Rudkin 2005). Hence, equation (8-40) can be summarized below:

$$D = f_3\left(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g, S_f\right) \tag{8-41}$$

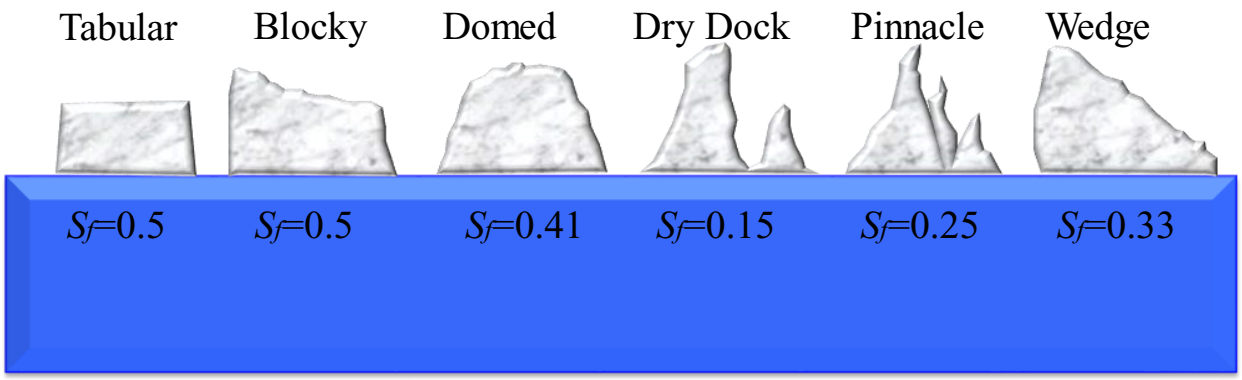


Figure 8-23. Applied shape factor values for different icebergs in the present study

Presumably, the density along with the viscosity of the seawater is constant, and the value of gravitational acceleration may be regarded as a constant value; as a result, equation (8-41) is rewritten as follows:

$$D = f_4\left(L, H, w, M, \rho_i, S_f\right). \tag{8-42}$$

The dimensional form of equation (8-42) is written below:

$$D = f_5(\Pi_1, \Pi_2, \Pi_3, \Pi_4) \tag{8-43}$$

here, Π_1, Π_2,…, and Π_4 are dimensionless groups and f_5 is a functional symbol based on the Buckingham-π theorem. Thus, the dimensionless groups below are written:

$$\Pi_1 = \frac{L}{H} \tag{8-44}$$

$$\Pi_2 = \frac{w}{H} \tag{8-45}$$

$$\Pi_3 = \frac{M}{\rho_i . H^3} \tag{8-46}$$

$$\Pi_4 = S_f \tag{8-47}$$

Equation (8-47) is then formulated as a function of four dimensionless groups as follows:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \tag{8-48}$$

Therefore, D/H is the iceberg draft ratio and is a function of the length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$. Subsequently, the ML models applied in the current investigations were fed with the input parameters in equation (8-48).

Hence, four dimensionless groups, including length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$ were applied to estimate the iceberg draft ratio (D/H) through the ML models in the present work. Figure 8-24 illustrates the combinations of four dimensionless groups introduced to develop the ML models. As seen, to identify the premium ML

models, five ML models, e.g., Model 1 to Model 5, were acquired, while Model 6 to Model 9 were defined to recognize the most influencing input parameters. Model 1 included all input factors, whilst these dimensionless groups were disregarded one at a time in Model 2 to Model 5.

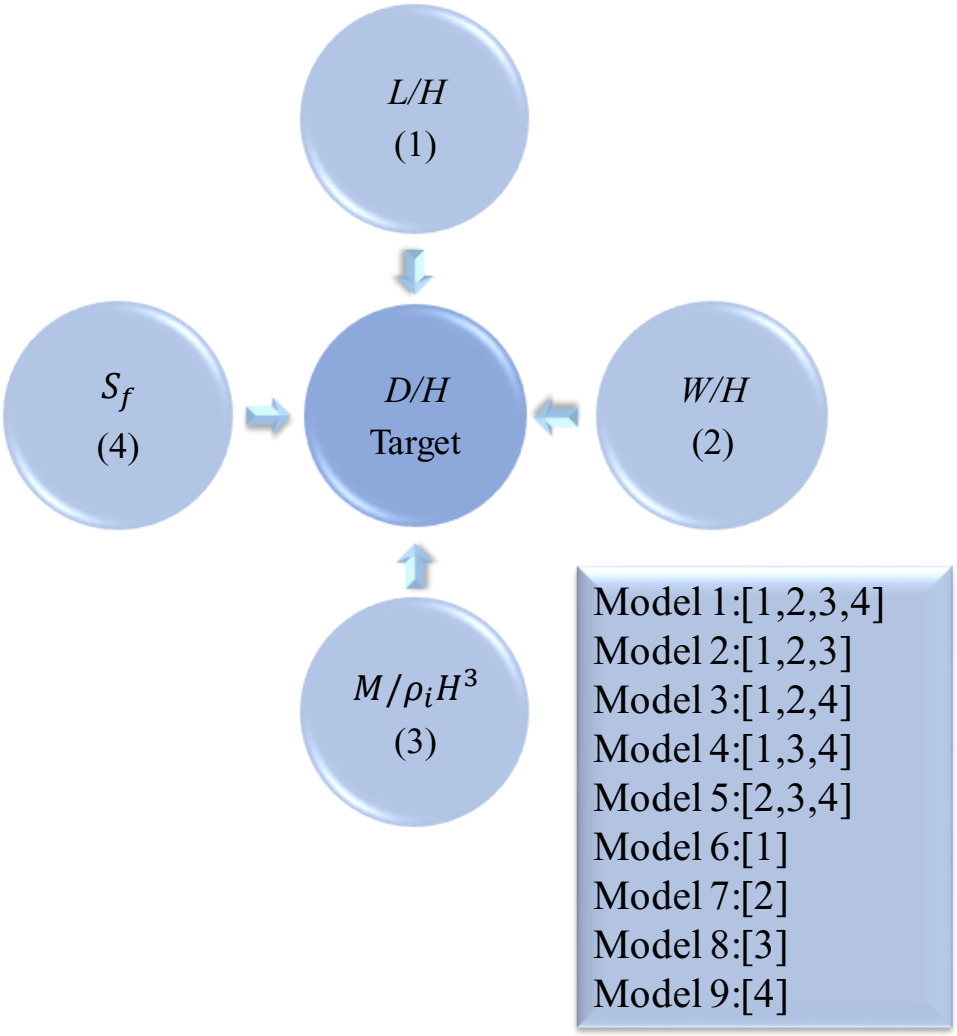


Figure 8-24. Input combination utilized for developing the ML models in the current study

8.1.10.5. Dataset construction

The current work utilized many field studies to analyze the iceberg draft. The key values of 12 field studies reported by El-Tahan et al. (1985) (38 cases), Woodworth-Lynas et al. (1985) (one

case), Løset and Carstens (1996) (52 cases), Barker et al. (2004) (14 cases), McKenna (2004) (two cases), Sonnichsen et al. (2006) (nine cases), Turnbull et al. (2015) (two cases), McGuire et al. (2016) (eight cases), Younan et al. (2016) (29 cases), Talimi et al. (2016) (one case), Zhou (2017) (three cases), Turnbull et al. (2018) (two cases) were used. Table 8-3 summarizes the key parameters employed in the present study, comprising the field observations' maximum, minimum, average, variance, and standard deviation.

The T-test and the P-value for the dataset were calculated, presuming that the P-value of 0.05 or less is statistically significant (Azimi and Shiri 2020), where an alternative hypothesis influences the likelihood of the relationship between the observed values. This P-value for the constructed dataset was estimated as 0.008, representing that the correlations were statistically significant. It is worth mentioning that 60% of the constructed dataset was used to train the ML algorithms, whereas 40% of the rest data was applied to test these models.

Table 8-3. Summary of the key parameters employed in the present study

Value	*D*(m)	*H*(m)	*w*(m)	*M*(kg)	*L*(m)	*Sf*
Max	240	94	408	1.13E+13	499	0.58
Min	18	3.9	9.672	8317400	34	0.11
Ave	88.988	30.956	101.605	1.68E+11	144.522	0.298
Var.	1696.272	399.389	3733.368	1.14E+24	6546.271	0.011
Std.	41.186	19.985	61.101	1.07E+12	80.909	0.107

8.1.10.6. Goodness of fit

A set of criteria such as correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were utilized to evaluate the precision, correlation, and complexity of the ML models. The proximity of the *R* and *WI* criteria to one showing the ML model tended to have a high degree of correlation with the values observed. The nearness of the *RMSE*, *MAPE*, and *CRM* indices to zero representing the ML model possessed the lowest degree of impreciseness; however, the complexity of the ML models was not examined through the indices applied. Hence, the Akaike Information Criteria (*AIC*) was used to overcome this restriction. In other words, the less complex ML model had the lowest amount of *AIC*, so the best ML model might have the lowest degree of *AIC* index and error (*RMSE*, *MAPE*, and *CRM*), with the highest amount of correlation (*R* and *WI*) (Azimi et al. 2022).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{8-49}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{8-50}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{8-51}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{8-52}$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \tag{8-53}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{8-54}$$

Here, Oi, Pi, $\bar{O}$, $\bar{P}$, n, and k are the observational value, the predicted amount, the average observational values, the average predicted amount, the number of observations, and the number of independent variables in the ML models.

8.1.11. Results and discussion

The ETR models' performance was evaluated by performing several analyses, then the superior ETR model and the most significant input parameters were identified. Ultimately, the capability of the superior ETR model was compared with DTR, RFR, and empirical models.

8.1.11.1. Evaluation of ETR models

Figure 8-25 displays the results of key statistical indices obtained from the ETR 1 to ETR 9 models. The slices' size in this figure denotes the ETR models' significance. For instance, the highest level of precision was for the ETR 1 model, with an *RMSE* index of 1.081. The ETR 1 model approximated the iceberg draft through all inputs, including L/H, w/H, $M/\rho_i.H^3$, and S_f, where the value of *WI* and *AIC* criteria for this model equaled 0.943 and 10.165. The highest degree of accuracy was also obtained for the ETR 2 model (*RMSE*=1.081), while the iceberg shape factor (S_f) was an eliminated input for this model. The influence of mass ratio ($M/\rho_i.H^3$) was ignored for the ETR 3 model, and the *CRM* and R indices for this model were calculated to be 0.108 and 0.917. The ETR 4 model estimated the iceberg drafts using L/H, $M/\rho_i.H^3$, and S_f inputs and the effect of w/H was disregarded for this ML model, with the *WI* and *AIC* values of 0.926 and 10.138. Among the ETR models with three inputs, the ETR 5 model had the lowest amount of precision (*RMSE*=1.234) and the highest level of complexity (*AIC*=11.837). For the ETR 5 model, the iceberg length ratio (L/H) was removed from its input combination.

Four ETR models with one input, e.g., ETR 6 to ETR 9, were defined to recognize the most significant input parameters to approximate the iceberg drafts. The simulation results proved that the iceberg length ratio (L/H) was detected as the most influential input since the ETR 6 model as a function of L/H, possessed the highest level of precision, correlation, and simplicity to predict iceberg drafts. Furthermore, the width ratio (w/H), the mass ratio ($M/\rho_i.H^3$), and shape factor (S_f) were ranked in terms of effectiveness in the second, third, and fourth places.

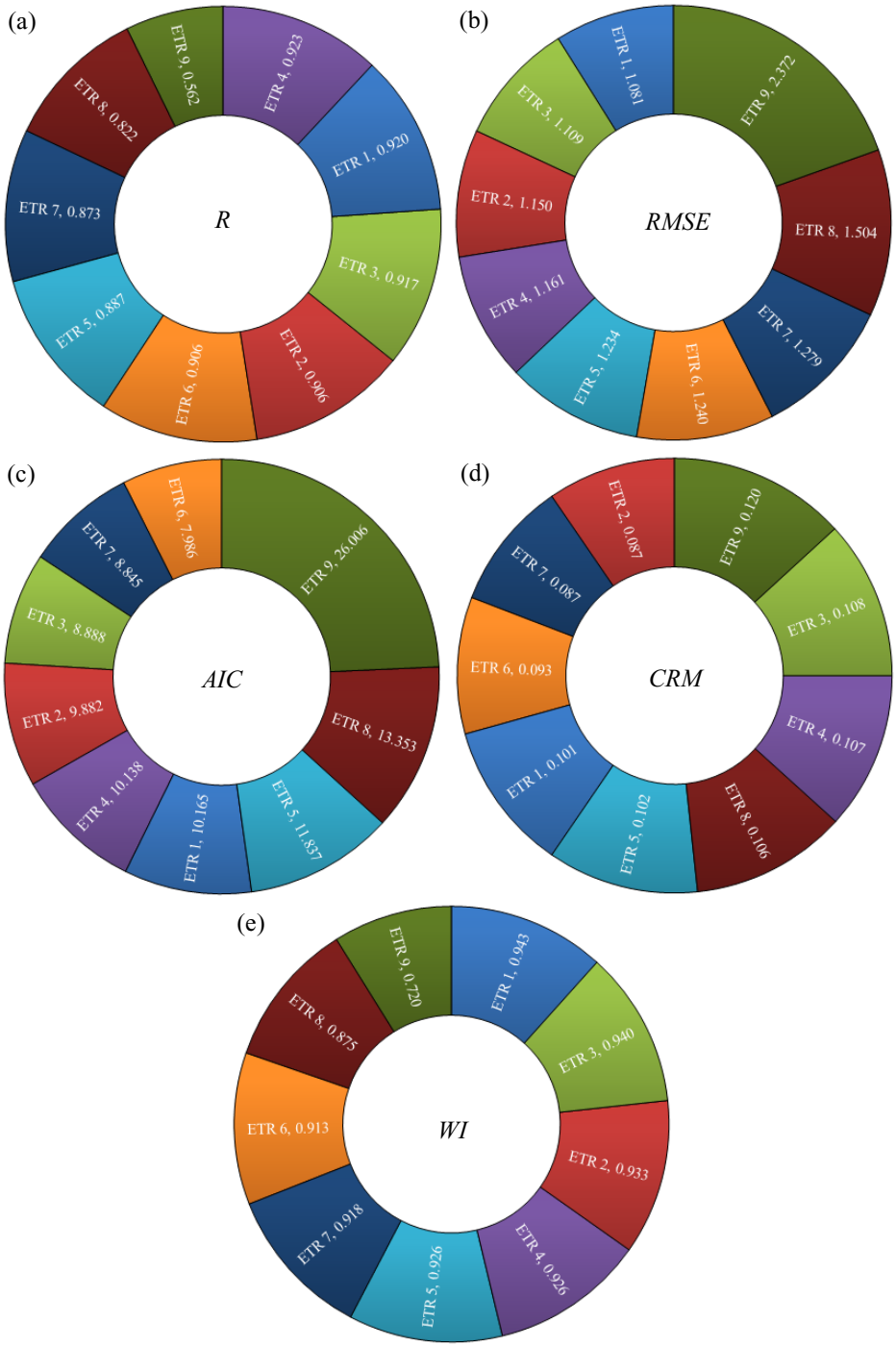


Figure 8-25. Results of key statistical indices obtained from the ETR 1 to ETR 9 models (a) R (b) RMSE (c) AIC (d) CRM (e) WI

Figure 8-26 illustrates the scatter plots for the ETR 1 to ETR 9 models. Regarding this figure, the highest degree of correlation belonged to the ETR 4 model, with an *R* criterion of 0.924, whereas, amongst the ETR models with three input parameters, the ETR 5 model's correlation with the field data was relatively low. Moreover, the *R* index for the ETR 1, ETR 2, ETR 3, and ETR 5 models was surmised as 0.920, 0.906, 0.917, and 0.887, respectively. Among the ETR models with just one input parameter, the ETR 6 model had the uppermost correlation with the observed values, while the lowest level of correlation was achieved for the ETR 9 model.

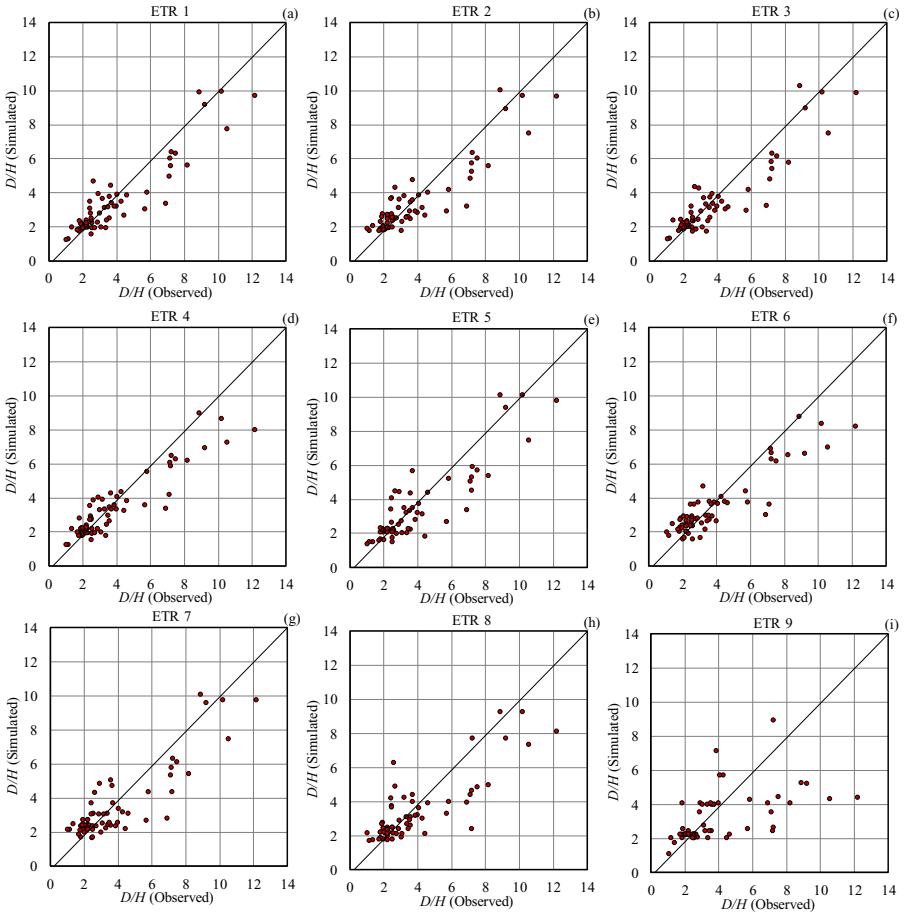


Figure 8-26. scatter plots for the ETR 1 to ETR 9 models (a) ETR 1 (b) ETR 2 (c) ETR 3 (d) ETR 4 (e) ETR 5 (f) ETR 6 (g) ETR 7 (h) ETR 8 (i) ETR 9

The error analysis for the ETR models was performed to evaluate these ML models' performance further. In Figure 8-27, the results of the error analysis for the ETR 1 to ETR 9 models are exhibited. Almost 36% of the iceberg drafts simulated by the ETR 1 and ETR 4 models showed an error of more than 20%. However, more than one-third of the iceberg drafts modeled by ETR 1 had an error of less than 10%, and this amount for the ETR 4 model was about 30%. Approximately half of the iceberg drafts simulated by the ETR 2 model possessed an error of greater than 20%, whilst this value for the ETR 3 and ETR 5 model was nearly 38% and 45%. The error analysis for the ETR models with three and four inputs, e.g., ETR 1 to ETR 5, demonstrated that the ETR 1 model had the best performance in estimating the iceberg drafts.

Among the ETR models with one input parameter, the ETR 6 model exhibited better performance in terms of error analysis, signifying that virtually 44% of the ETR 6 model's results provided an error of bigger than 20%. In contrast, this value for ETR 7, ETR 8, and ETR 9 was just about 56%, 52%, and 53%, respectively. The ETR 7 and ETR 9 models showed the worst error analysis amount than the ML models with just one input parameter.

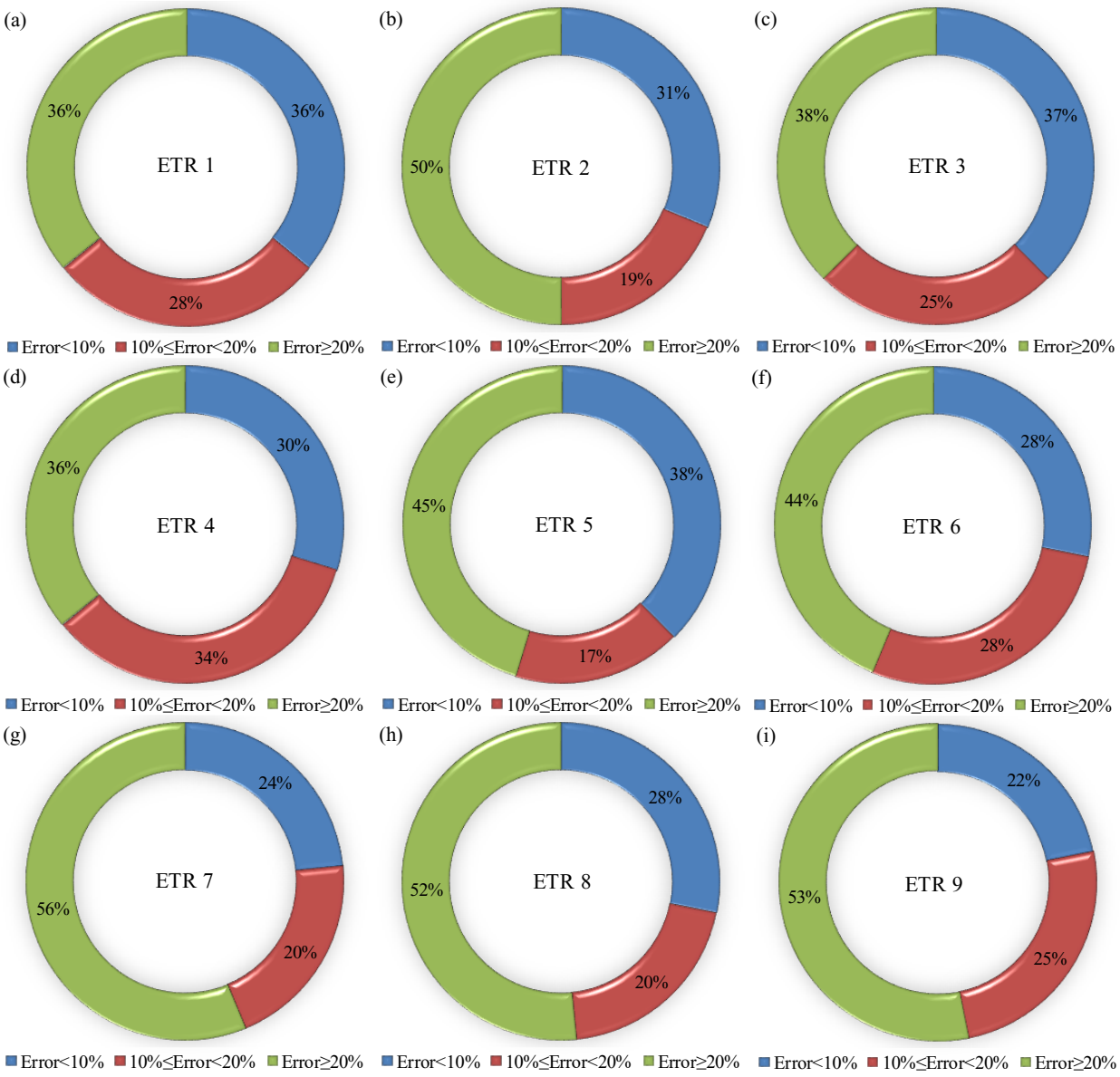


Figure 8-27. Error analysis of the ETR 1 to ETR 9 models (a) ETR 1 (b) ETR 2 (c) ETR 3 (d) ETR 4 (e) ETR 5 (f) ETR 6 (g) ETR 7 (h) ETR 8 (i) ETR 9

The results of the discrepancy analysis for the ETR models are depicted in Figure 8-28. The efficiency of the ETR models was examined using the discrepancy ratio (*DR*) as below:

$$DR = \frac{P_i}{O_i} \tag{8-55}$$

where, P_i and O_i are the simulated and observed iceberg drafts, respectively. The closer the magnitude of *DR* is to the unity, the higher performance shows the ETR model. The value of the maximum ($DR_{(max)}$), minimum ($DR_{(min)}$), and average ($DR_{(ave)}$) discrepancy ratio for these models was calculated. For example, the $DR_{(min)}$ value for ETR 1, ETR 2, and ETR 3 models equaled, in turn, 0.493, 0.474, and 0.480. The nearest $DR_{(ave)}$ to one belonged to ETR 2, with a $DR_{(ave)}$ value of 1.005, whereas this index for ETR 4 and ETR 5 was 0.973 and 0.965.

Among the ETR models with solely one input, e.g., ETR 6 to ETR 9, the highest value of $DR_{(max)}$ was obtained for the ETR 8 model ($DR_{(max)}$=2.470), and this criterion for the ETR 6, ETR 7, and ETR 9 models were respectively computed to be 1.984, 2.139, and 2.293.

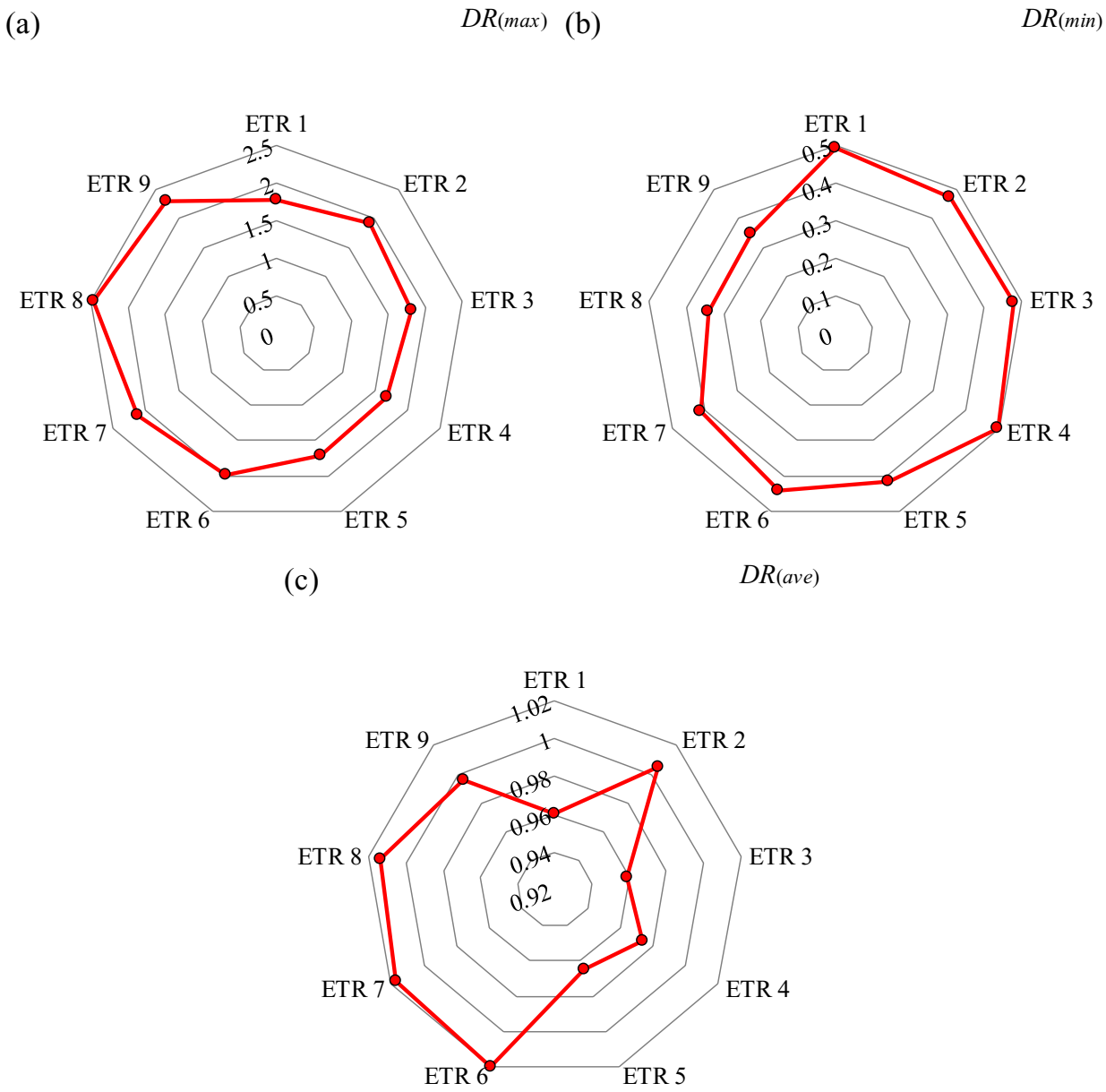


Figure 8-28. Results of the discrepancy analysis for the ETR models (a) DR(max) (b) DR(min) (c) DR(ave)

The study performed an uncertainty analysis (UA) to further assessment of the ETR models' performance. To do so, the ETR model's error (e_j) were calculated as the difference between the iceberg drafts predicted (P_j) through this model and the actual iceberg drafts (O_j), as follows:

$$e_j = P_j - O_j \tag{8-56}$$

The mean (Mean) and the standard deviation (StDev) of such error values were obtained by the equations as follows:

$$\text{Mean} = \frac{1}{n}\sum_{j=1}^{n} e_j \tag{8-57}$$

$$\text{StDev} = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2 / (n-1)} \tag{8-58}$$

An individual ETR model underestimated the iceberg draft if the sign of the Mean value was negative, while the positive sign of the Mean meant that the ETR model overestimated the iceberg drafts. Thereon, a confidence interval (CI) was produced near the error counted using the Mean, StDev values, and the "Wilson score technic" by omitting the continuity correction. A normal distribution interval corrected as an asymmetric normal distribution, named the Wilson score interval, was employed to adjust the CI bounds. Subsequently, a ±1.96Se yielded a 95%CI. It should be remarked that the width of uncertainty bound (WUB) of each ETR model was acquired below (Azimi et al. 2022):

$$\text{WUB} = \pm\frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{8-59}$$

Figure 8-29 exhibits the normal distribution of the CI for the iceberg drafts simulated by the ETR models. According to the performed UA, all ETR models overestimated the iceberg drafts since the value of "Mean" for these models had positive signs, e.g., for the ETR 1 to ETR 9 models, the "Mean" values were surmised at 0.390, 0.335, 0.415, 0.411, 0.395, 0.357, 0.337, 0.407, and 0.463, respectively.

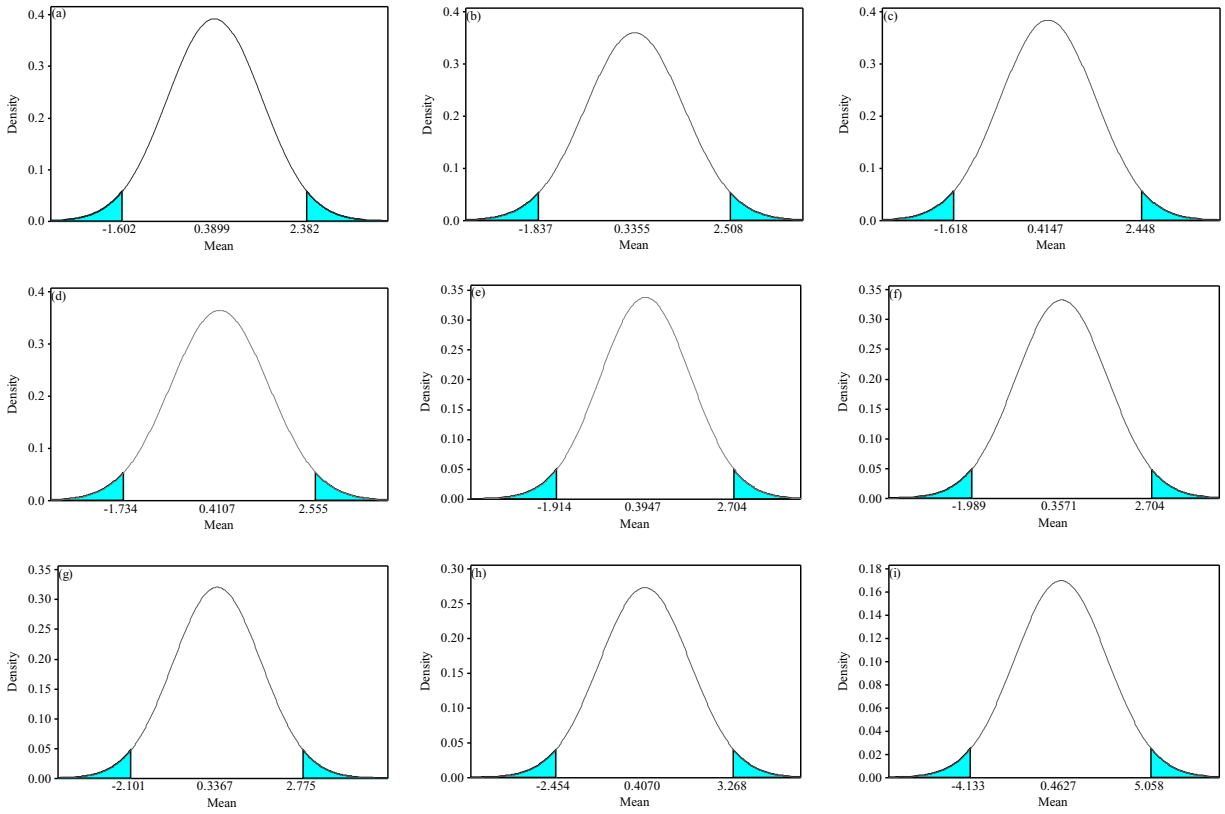


Figure 8-29. Error normal distribution of (a) ETR 1 (b) ETR 2 (c) ETR 3 (d) ETR 4 (e) ETR 5 (f) ETR 6 (g) ETR 7 (h) ETR 8 (i) ETR 9

The smallest StDev value alongside the narrowest WUB was for the ETR 1 model, with a StDev and WUB of 1.016 and ±0.254. Amid the ETR 1 to ETR 5 models, the highest StDev and the widest WUB belonged to the ETR 5 model (StDev=1.178 & WUB=±0.295). The 95%CI value for the ETR 2, ETR 3, and ETR 4 models was at (0.059 to 0.612), (0.156 to 0.674), and (0.137 to 0.684), respectively. Amongst the ETR models with one input parameter, the lowest and highest StDev and WUB amounts referred to the ETR 6 and ETR 9 models.

Thus, several analyses were performed, including sensitivity, error, discrepancy, and uncertainty analysis, to identify the superior ETR model. The ETR 1 model, as a function of all inputs, was recognized as the superior ETR model to estimate the iceberg drafts in the present study.

8.1.11.2. Comparison between the superior ETR model with DTR, RFR, and empirical models

The ETR 1 model was introduced as the best ETR model to predict the iceberg drafts in the previous sections. Hence, the capability of the ETR 1 model was compared with several tree-based ML models and empirical methods. In other words, the performance of DTR and RFR algorithms, along with four empirical models provided by Barker et al. (2004) (Eq. 8-59), Sacchetti et al. (2012) (Eq. 8-60), King et al. (2016) (Eq. 8-61), and Stuckey et al. (2021) (Eq. 8-62) was compared with the ETR 1 model.

$$D = 0.7(L) \tag{8-59}$$

$$D = 3.9(L^{0.63}) \tag{8-60}$$

$$D = 2.05(M^{0.276}) \tag{8-61}$$

$$D = 4.78(L^{0.58}) \tag{8-62}$$

Figure 8-30 compares the key statistical criteria between the ETR 1 model, ML algorithms, and empirical approaches. The correlation coefficient for DTR, RFR, Eq. (8-59), Eq. (8-60), Eq. (8-61), and Eq. (8-62) equaled 0.848, 0.900, 0.769, 0.770, 0.333, and 0.769. The ETR 1 model's correlation with the observed values was 2.2% and 19.5% higher than the RFR algorithm and Eq. (8-60). Additionally, the ETR 1 model was almost 19.3% and 96.1% more accurate than the DTR and Eq. (8-59), where its complexity level was 97.2% and 96% less than Eq. (8-61) and Eq. (8-62). The *CRM* index for DTR and RFR models was estimated to be 0.005 and 0.115, while the *WI* criterion for Eq. (8-59) to Eq. (8-62) equaled 0.895, 0.896, 0.257, and 0.674, respectively.

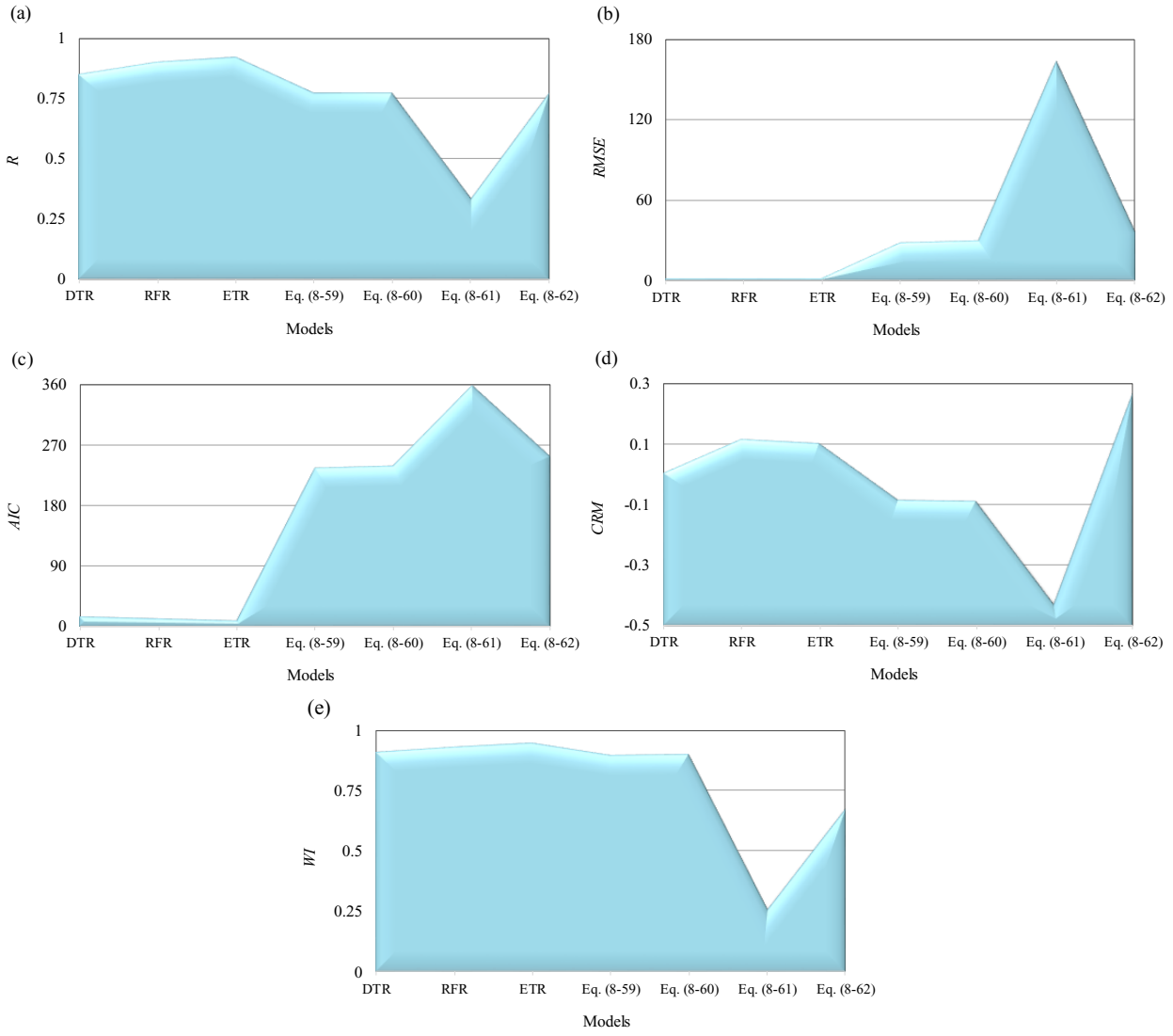


Figure 8-30. Comparison between the ETR 1 model, ML algorithms, and empirical approaches

As shown, the ETR 1 model could simulate the iceberg draft values more efficiently than its counterparts, with more accuracy, better correlation, and lower complexity. It is worth mentioning that the generalization capability of the ETR 1 model was more significant than the empirical models since the ability of ETR 1 to model the iceberg draft was examined through 12 different field datasets. In contrast, the empirical approaches were derived from merely specific geographical locations and datasets. Furthermore, the empirical models suggested in the literature

were able to estimate the iceberg drafts just using one input parameter, e.g., length or mass, etc., while the best ETR model managed to simulate the draft values through the length, width, mass, and shape factor of the iceberg. Moreover, the iceberg length and width ratio were identified as the most influential input parameters to model the iceberg drafts using the ETR algorithm.

8.1.12. Conclusion

This study proposed the use of extra tree regression (ETR) to model the iceberg drafts for the first time. Initially, nine ETR models were developed utilizing the parameters affecting the iceberg draft; then, a dataset was built using the observational measurements in the literature, and 60% and 40% of the data were used for the training and testing of the ML models, respectively. The superior ETR model and the most significant input parameters were recognized by performing sensitivity, error, discrepancy, and uncertainty analyses. Lastly, the performance of the superior ETR model was compared with the decision tree regression (DTR), random forest regression (RFR), and empirical methods. The most important outcomes of this study are summarized as follows:

- The superior ETR model, ETR 1, could simulate the iceberg drafts with the highest level of precision and correlation alongside the lowest degree of complexity.
- The ETR 1 model predicted the iceberg drafts as a function of all input parameters, including L/H , w/H , $M/\rho_i.H^3$, and S_f,
- The value of *RMSE*, *R*, and *AIC* indices for the superior ETR model was obtained at 1.081, 0.920, and 10.165, respectively.
- Approximately 36% of the iceberg drafts estimated by the ETR 1 model possessed an error of less than 10%.

- The sensitivity analysis showed that the iceberg length ratio (*L/H*) and iceberg width ratio (*w/H*) was the most influential input parameters.
- Regarding the performed uncertainty analysis, the ETR 1 model overestimated the iceberg drafts, with the smallest value of the standard deviation of computed error (StDev) and narrowest bound of uncertainty analysis (WUB).
- Comparing the ETR 1 model's performance with the DTR, RFR, and empirical approaches demonstrated that the ETR 1 model outperformed its counterparts.

The presented investigation was the first application of the tree-based ML algorithms in estimating iceberg drafts, and the results gave an excellent understanding of this domain. The achieved outcomes can preserve significant downtime and expenditures, particularly in the initial phases of the iceberg management projects, to enhance the safety and integrity of offshore structures and subsea assets.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)," and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)." Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Allaire, P.E., 1972. Stability of simply shaped icebergs. Journal of Canadian Petroleum Technology, 11: 04.

Azimi, H., Shiri, H., 2020. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2021, Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., Malta, E.R., 2021. A non-tuned machine learning method to simulate ice-seabed interaction process in clay. Journal of Pipeline Science and Engineering, 1(4): 379-394.

Azimi, H., Shiri, H., Mahdianpari, M., 2022. Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. Journal of Pipeline Science and Engineering, 2(4): 100075.

Barker, A., Sayed. M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In the14th international offshore and polar engineering conference, ISOPE-I-04-116.

Bass, D.W., 1980. Stability of icebergs. Annals of Glaciology, 1: 43-47.

Blažauskas, N., Włodarski, M., Paulauskas, S., 2013. Perspectives for offshore wind energy development in the South-East Baltics. Klaip eda University: Klaipeda, Lithuania, 1-54.

Breiman, L., 2001. Random forests, Machine learning, 45(1): 5-32.

Brooks, L., 1982. West Greenland iceberg drift and ocean current investigations. 64-65.

C-CORE, 2001. Documentation of iceberg grounding events from the 2000 season. C-CORE publication 01-C10 (Revision 0). Report submitted to the Geological Survey of Canada, Atlantic.

Dowdeswell, J.A., Bamber, J.L., 2007. Keel depths of modern Antarctic icebergs and implications for sea-floor scouring in the geological record. Marine Geology, 243(1-4): 120-131.

El-Tahan, M., El-Tahan, H., Courage, D., Mitten, P., 1985. Documentation of Iceberg Groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning, 63(1): 3-42.

Hammed, M.M., AlOmar, M.K., Khaleel. F., Al-Ansari, N., 2021. An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. Mathematical Problems in Engineering.

Hotzel, I.S., Miller, J.D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Arctic Technology Conference, OTC-27358-MS.

Liang, B., 2001. Iceberg stability and deterioration. Doctoral dissertation, Memorial University of Newfoundland. St. John's. NL. Canada.

Løset, S., Carstens, T., 1996. Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Region Science and Technology, 24(4): 323-340.

Mahdianpari, M., Homayouni, S., Foucher, S., 2021. CJRS'special issue on deep learning for environmental applications of remote sensing data. Canadian Journal of Remote Sensing, 47(2): 159-161.

McGuire, P., Younan, A., Wang, Y., Bruce, J., Gandi, M., King, T., Regular, K., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., 2009. Modelling Iceberg Shape, Mass and Draft Changes. In Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions, POAC09-107.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. In Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, POAC, pp. 9-19.

Pekel, E., 2020. Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3): 1111-1119.

PERD, 1999. Compilation of Iceberg Shape and Geometry Data for the Grand Banks Region. Prepared by Canatec, ICL Isometrics, Coretec & Westmar for the National Research Council of Canada. PERDICHC Report 9-80.

Robe, R.Q., Farmer, L.D., 1976. Physical Properties of Icebergs. Part I. Height to Draft Ratios of Icebergs. Part II. Mass Estimation of Arctic Icebergs. Coast Guard Research and Development Center Groton Conn.

Rudkin, P., 2005. Comprehensive Iceberg Management Database Report 2005 update. PERD/CHC Report 20-72. National Research Council of Canada (NRC) and Panel on Energy Research and Development (PERD).

Sacchetti, F., Benetti, S., Cofaigh, C.Ó., Georgiopoulou, A., 2012. Geophysical evidence of deep-keeled icebergs on the Rockall Bank, Northeast Atlantic Ocean. Geomorphology, 159: 63-72.

Sahani, N., Ghosh, T., 2021. GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. Ecological Informatics, 64: 101352.

Sonnichsen, G., Hundert, T., Myers, R., Pocklington, P., 2006. Documentation of Recent Iceberg Grounding Evens and a Comparison with Older Events of Know Age." Environmental Studies Research Funds. Report ESRF Vol. 157.

Stuckey, P.D., 2008. Drift speed distribution of icebergs on the grand banks and influence on design loads. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., King, T., 2021. Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. In International Conference on Offshore Mechanics and Arctic Engineering, 85178: V007T07A018.

Talimi, V., Ni, S., Qiu, W., Fuglem, M., MacNeill, A., Younan, A., 2016. Investigation of Iceberg Hydrodynamics. Arctic Technology Conference, OTC-27493-MS.

Turnbull, I.D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Region Science Technology, 110: 1-18.

Turnbull, I.D., King, T., Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Woodworth-Lynas, C.M.T., Simms, A., Rendell, C.M., 1985. Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Region Science and Technology, 10(2): 163-186.

WWF Arctic Programme; https://www.arcticwwf.org/threats/oil-and-gas/

Younan, A., Ralph, F., Ralph, T., Bruce, J., 2016. Overview of the 2012 iceberg profiling program. Arctic Technology Conference, OTC-27469-MS.

Zhou, M., 2017. Underwater iceberg profiling and motion estimation using autonomous underwater vehicles. PhD dissertation. Memorial University of Newfoundland, NL, Canada.

Section 4

Iceberg Drafts Assessment using Decision Tree Regression (DTR), Artificial Neural Network (ANN), and Support Vector Regression (SVR) algorithms

This section was published in the 27th International Conference on Port and Ocean Engineering under Arctic Conditions (POAC), 2023, Glasgow, UK

Abstract

Nearly one-fifth of the Earth's undiscovered hydrocarbons are reserved in the Arctic area whereas, the recent offshore oil and gas loading equipment, e.g., subsea pipelines, wellheads, and communication cables, developed in the Arctic waters has led to a considerable awareness of the iceberg draft prediction. The iceberg tip would gouge the ocean floor and the operational integrity of the sea bottom-founded infrastructures may be threatened in the Arctic shallower waters if the ocean depth is smaller than the traveling iceberg draft. Hence, developing an intelligent and cost-effective solution to predict iceberg drafts is necessary to guarantee the operational integrity of the subsea assets. In this study, the iceberg drafts were simulated using three machine learning (ML) algorithms comprising decision tree regression (DTR), artificial neural network (ANN), and support vector regression (SVR). Initially, using the parameters governing the iceberg draft simulation, a set of ML models was defined. By performing several analyses including sensitivity analysis, error analysis, and uncertainty analysis, the premium ML model along with the most significant input parameters was introduced. The obtained outcome can smooth the path to offer alternative techniques to maintain the time and expenditures of the iceberg management projects and subsea structure design, specifically in the primary phases of the construction methodology, corresponding logistics, and the prospective scope of engineering design projects.

Keywords: Iceberg draft simulation; Subsea infrastructures; decision tree regression (DTR); artificial neural network (ANN); support vector regression (SVR).

8.1.13. Introduction

Every year many icebergs are born out of glaciers in the Arctic area and carried away by the currents and into the North Atlantic. These traveling masses may touch the sea bottom in shallow waters and scratch the seabed, causing “ice-gouging” that can endanger the integrity of subsea pipelines and power cables or even directly collide with offshore structures such as ships, platforms, wind turbines, subsea manifolds, etc. Currently, ice management such as iceberg towing and re-routing is the most reliable approach to protect the subsea and offshore infrastructures, where the threatening icebergs are hooked and towed in a safe direction. Ice management is generally a costly operation and requires standby marine spread with a range of advanced tools, vessels, and equipment, like subsea survey facilities, to investigate the iceberg draft and determine if it is a threat to infrastructures. The Arctic region is one of the best resources to grow oil and gas loading equipment. Moreover, the Arctic offshore regions with rich wind culture have a high potential for the development of offshore wind farms (Blažauskas et al. 2013). The schematic layout of the iceberg free-floating and iceberg scouring in cold waters is displayed in Figure 8-31. As shown, the iceberg is in a free-floating circumstance if the ocean depth is greater than the iceberg draft; otherwise, the seafloor is scoured, and the seabed soil shear resistance causes the soil displacement to extend deeper than the iceberg tip threatening the buried subsea assets.

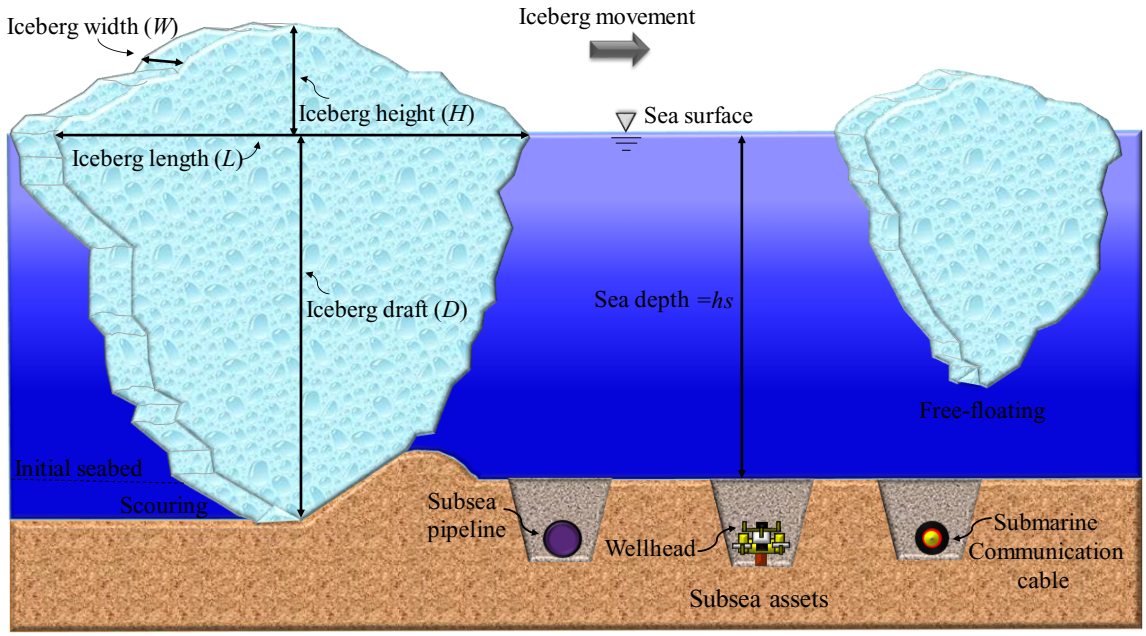


Figure 8-31. Icebergs in scouring and free-floating circumstances

Owing to the significance of the iceberg draft estimation, extensive experimental, analytical, and numerical studies have been conducted in the domain. For example, McKena (2000) evaluated the threat of ice-gouging to offshore petroleum installations serving in the Grand Banks region. The author showed that the iceberg proportions were a function of the iceberg length. Sonnichsen et al. (2003) reported the seafloor surveys and ice-gouging on the Grand Bank of Canada in the 2000 iceberg season. The drafts of the iceberg were recorded through a lateral scan sonar tool mounted on the tracking boat. The study demonstrated that there was serious concern about the precision of lateral scans of iceberg draft estimation. Barker et al. (2004) specified the iceberg sails and drafts utilizing the dimensions marked in the field. The investigation revealed that the iceberg draft could be approximated regarding the iceberg waterline length. Dowdeswell and Bamber (2007) scrutinized the draft deepness of traveling icebergs in the Antarctic waters. The authors calculated the draft through the ice thickness, keel depth at the grounding line, and surface elevation. The

study ended that a small minority of icebergs in the Antarctica and Greenland waters had drafts deeper than 650 m. Sacchetti et al. (2012) explored the iceberg features and ice scouring in Labrador and Hibernia territories. The characteristics of various icebergs such as wedged, domed, tabular, and pinnacle bergs were considered in this examination. The authors proposed a set of relationships in terms of the iceberg length to predict the iceberg draft. King et al. (2016) completed field experimentation to count the rolling iceberg rate. The iceberg draft was surmised by a calving study, with a computed standard deviation of draft changes from 19% to 34%. The iceberg draft has been approximated using the iceberg mass. Turnbull et al. (2018) suggested an instance of the drift mensuration of shifting icebergs on the Grand Banks of Newfoundland. The study stated that the iceberg draft estimated was roughly 1.3 times deeper than the real iceberg draft. McKenna et al. (2019) modeled the ice-gouging on the Grand Banks of Canada adopting the Monte Carlo simulation (MCs). The iceberg draft variations were also employed to reduce the dimension of draft changes in the applied methodology. Stuckey et al. (2021) modeled the 3D iceberg forms through a field survey. The examination exhibited that the iceberg drafts were summed in terms of iceberg length through the power method. They provided two practical instances concerning the information gathered in 2016 and the post-2000 report. Despite the wide applications of machine learning (ML) technology in multifarious fields to simulate different linear and nonlinear problems (Azimi and Shiri 2021a, Azimi et al. 2022, Azimi et al. 2023), the literature indicated that the iceberg draft has not been predicted using ML algorithms so far. Hence, to fill this knowledge gap, the iceberg drafts were simulated through three robust ML algorithms, e.g., artificial neural network (ANN), decision tree regression (DTR), and support vector regression (SVR) models. More information will be presented in the forthcoming sections.

8.1.14. Methodology

8.1.14.1. Artificial neural networks (ANN)

ANN is one of the most universally supervised machine learning (ML) algorithms. The main structure of an ANN algorithm consists of at least three distinct layers comprising an input layer, a hidden layer, and an output layer. The input parameters are embedded within the input layer, while the target parameter, e.g., the iceberg draft, is considered in the output layer. The hidden neurons are situated within the hidden layer, where the size of this layer is determined by the problem's complexity and desired accuracy (Azimi and Shiri 2021b). The number of hidden layer neurons was initially set as one and the magnitude of this hyperparameter was increased to 15, where the optimum number of the hidden layer neurons was chosen at 12 for the reason that the proficiency of the ANN algorithm was negligibly altered after this amount. In each hidden neuron, both the input parameters and their weights are calculated using mathematical operations, and the outcome is passed through a transfer function entitled the activation function. The sigmoid function was applied for the current architecture because it had better performance in comparison with other activation functions. Subsequently, the performance of the ANN algorithm was evaluated by the Mean squared error (*MSE*) as the loss function in the present study to measure the difference between the computed outputs and the target outputs.

8.1.14.2. Decision tree regression (DTR)

A tree data system contains a series of leaves and branches in which each node is regarded as a decision tree (DT). The DT may be used to solve both regression and classification problems. The DT comprises many components, such as a root node, several leaf nodes, internal nodes, and branches. The topmost node in this tree is considered the root node, and the leaf nodes (terminal

nodes) end with the titles of types, while the non-leaf nodes are assumed as the internal nodes. Such nodes link to each other through the branches (Pekel 2020). In the present work, the mean squared error (MSE) is involved in maintaining the fitness function in the DT algorithm. In this investigation, the hyperparameters of the DTR algorithms were determined using a trial-and-error strategy. It means that the value of *max_depth*, *max_leaf_nodes*, and *min_weight_fraction_leaf* was primary at 10, 2, and 0.01, and the performance of the DTR algorithm was assessed. The number of hyperparameters was raised in the following stages until the DTR's results reached an adequate level. The DTR model estimated the iceberg drafts with its highest level of precision and correlation as well as its lowest degree of complicatedness until the number of hyperparameters comprising the *max_depth*, *max_features*, *max_leaf_nodes*, *min_samples_leaf*, *min_weight_fraction_leaf*, and *splitter* was, in turn, adapted as 150, *'auto'*, 2, 2, 0.001, and *'random'*.

8.1.14.3. Support vector regression (SVR)

A support vector machine (SVM) is known as a supervised learning ML algorithm to solve both classification and regression problems. The SVM is based on Vapnik-Chervonenkis (VC) theory, and this algorithm was proposed by Vapnik (1995). To simulate the regression problems, SVR is applied in which the training data is mapped from the input variables (input space) into the objective parameter (feature space) through a function (Ϙ). In the feature space, a separating hyperplane with the highest margin is produced. In a regression problem, a nonlinear transformation from the input space to high-dimensional space is made by using the Ϙ function. Regardless of the transformation function (Ϙ), the kernel function can implement the dot product in the multidimensional feature space through the low-dimensional space input variables. In

practical applications, several kernel functions comprising the linear, polynomial, and radial basis functions (RBF) are utilized in the SVR algorithm. Moreover, the ε-insensitive loss function is employed as a cost function in this model. To simulate the iceberg drafts in this study, the parameters of the SVM algorithm, such as the penalty parameter (*C*), the kernel coefficient (*gamma*), *epsilon*, *verbose*, and kernel were respectively tuned as 0.01, 1, 0.5, 1, and *linear*. The applied parameters of the SVM model in the current study were chosen based on a trial and error method. The Flowchart of the SVR algorithm applied in the current study is shown in Figure 8-32. Regarding the flowchart, the constructed dataset was initially loaded and it was divided into the training and testing sub-samples. Subsequently, the iceberg drafts were simulated by using the parameters affecting the SVR model. If the performance of the SVR was acceptable, the results were compared with the ANN and DTR algorithms; otherwise, the SVR's hyperparameters were tuned. It is worth mentioning that the definition of acceptable performance for the ML algorithm is determined by the applied indices in equations (8-8) to (8-13). These statistical indices assess the accuracy, correlation, and complexity of the ML model at the same time.

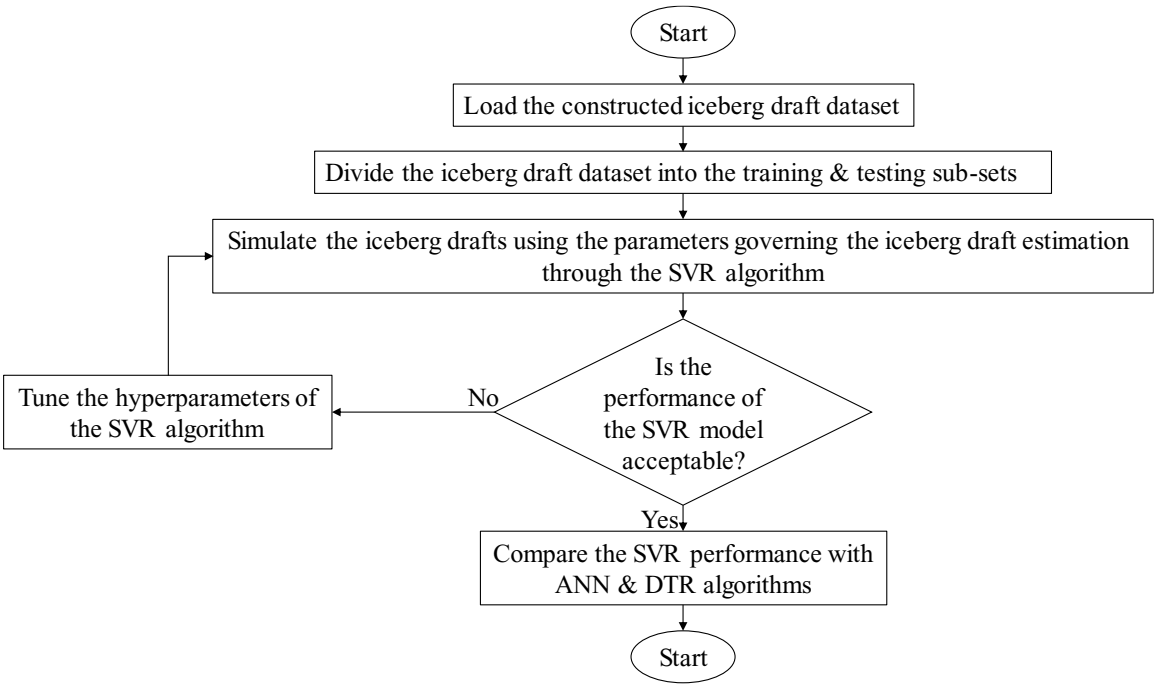


Figure 8-32. Flowchart of the SVR algorithm applied in the current study

8.1.14.4. Iceberg drafts

The iceberg draft (D) was assumed as a function of the physical characteristics of the iceberg, comprising the iceberg length (L), iceberg height (H), iceberg width (w), iceberg mass (M) in several fields, analytical, and numerical studies in the form below (Barker et al. 2004; McKenna et al. 2019; and Stuckey et al. 2021):

$$D = f_1(L, H, W, M). \tag{8-63}$$

Furthermore, the density of an iceberg (ρ_i), the density of seawater (ρ_{sw}), seawater viscosity (μ_{sw}), and gravitational acceleration (g) may influence the iceberg draft as follows:

$$D = f_2(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g) \tag{8-64}$$

The iceberg shape factor (S_f) signifies the global shape of icebergs, which can affect the magnitude of the iceberg draft (Turnbull et al. 2018). The iceberg shape factor has been already defined. The

shape factor of an iceberg describes the estimated fraction filled by the iceberg sail of a rectangle whose dimensions are the length by the height (Turnbull et al. 2015). The shape factor of the traveling icebergs is considered universally into six categories including Tabular (S_f=0.5), Blocky (S_f=0.5), Domed (S_f=0.41), Dry Dock (S_f=0.15), Pinnacle (S_f=0.25), and Wedge (S_f=0.5) (Rudkin 2005). Hence, equation (2) can be summarized below:

$$D = f_3\left(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g, S_f\right) \tag{8-65}$$

It is assumed that the density along with viscosity of the seawater is constant and the value of gravitational acceleration can be regarded as a constant value; as a result, equation (8-65) is rewritten as follows:

$$D = f_4\left(L, H, w, M, \rho_i, S_f\right). \tag{8-66}$$

The dimensional form of equation (20-4) is written below:

$$D = f_5(\Pi_1, \Pi_2, \Pi_3, \Pi_4) \tag{8-67}$$

here, Π_1, Π_2,…, and Π_4 are dimensionless groups and f_5 is a functional symbol based on the Buckingham-π theorem. Thus, the dimensionless groups below are written:

$$\Pi_1 = \frac{L}{H}, \Pi_2 = \frac{w}{H}, \Pi_3 = \frac{M}{\rho_i . H^3}, \Pi_4 = S_f \tag{8-68}$$

Equation (8-67) is then formulated as a function of four dimensionless groups as follows:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \tag{8-69}$$

Therefore, D/H as the iceberg draft ratio is a function of the length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$. Subsequently, the ML models applied in the current investigations were fed with the input parameters in equation (8-69). Hence, four dimensionless groups, including length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$ were applied to estimate the iceberg draft ratio (D/H) through the

ML models in the present work. Figure 8-33 illustrates the combinations of four dimensionless groups introduced to develop the ML models. As seen, to identify the premium ML models, five ML models, e.g., Model 1 to Model 5 were developed, while Model 6 to Model 9 were defined to recognize the most influencing input parameters. Model 1 included all input factors, whilst these dimensionless groups were disregarded one at a time in Model 2 to Model 5. Additionally, models 6 to 9 predicted the iceberg drafts solely one input parameter.

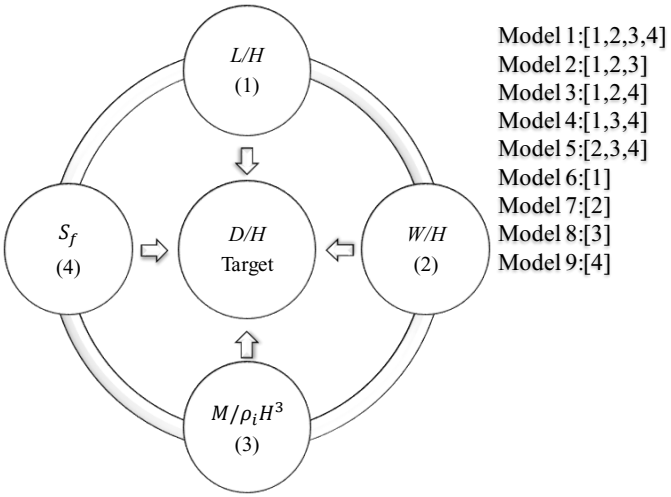


Figure 8-33. Input combination applied for developing the ML models

8.1.14.5. Construction of database

Several field observations were adopted to analyze the iceberg draft. The key values of 12 field studies reported by El-Tahan et al. (1985) (38 cases), Woodworth-Lynas et al. (1985) (one case), Løset and Carstens (1996) (52 cases), Barker et al. (2004) (14 cases), McKenna (2004) (two cases), Sonnichsen et al. (2006) (nine cases), Turnbull et al. (2015) (two cases), McGuire et al. (2016) (eight cases), Younan et al. (2016) (29 cases), Talimi et al. (2016) (one case), Zhou (2017) (three cases), Turnbull et al. (2018) (two cases) were used. It is worth noting that the iceberg drafts in the El-Tahan et al. (1985) study were measured using a set of technices such as submarine cables,

exploration vessels, shore-based radar, drill-rig radar, etc. Woodworth-Lynas et al. (1985) utilized acoustic profilers and sextants to measure the iceberg draft. Løset and Carstens (1996) calculated the iceberg draft through the real dimension above the water surface. Barker et al. (2004) employed the sonar profiles of the iceberg draft to document the iceberg draft. McKenna (2004) reported the magnitude of the iceberg draft by analyzing the iceberg profiles obtained from sonar technology. Sonnichsen et al. (2006) asserted that the side scan sonar equipment had been applied to approximate the iceberg draft. Turnbull et al. (2015) estimated iceberg draft regarding the relationships between above-water height and underwater depth. McGuire et al. (2016) applied a multi-beam sonar system for the iceberg draft measurement. Younan et al. (2016) scanned the iceberg drafts by means of a sideway-oriented multi-beam mounted on a remotely operated vehicle (ROV). Talimi et al. (2016) reported that the dimensions of the applied iceberg in this study were obtained from multi-beam sonar profiling. Zhou (2017) used the digital iceberg which was profiled by the National Research Council Canada (NRC). Turnbull et al. (2018) quantified the iceberg draft using a multi-beam profiling system. The T-test and the P-value for the dataset were calculated, presuming that the P-value of 0.05 or less is statistically significant (Azimi and Shiri 2020), where the likelihood of the relationship between the observed values is influenced by an alternative hypothesis. This P-value for the constructed dataset was estimated as 0.008, representing that the correlations were statistically significant. It is worth mentioning that 60% of the constructed database was utilized for training the ANN, DTR, and SVR models, whereas 40% of the remaining dataset was employed to test these models.

8.1.14.6. Goodness of fit

To examine the precision, correlation, and complexity of the ML models, several criteria such as correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were utilized. The proximity of the R and WI criteria to one showing the ML model tended to have a high degree of correlation with the values observed. The nearness of the RMSE, *MAPE*, and *CRM* indices to zero representing the ML model possessed the lowest degree of impreciseness; however, the complexity of the ML models was not examined through the indices applied. To overcome this restriction, the Akaike Information Criteria (*AIC*) was used. (Azimi et al. 2022).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{8-70}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{8-71}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{8-72}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{8-73}$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \tag{8-74}$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \tag{8-75}$$

Here, *Oi, Pi*, $\bar{O}$, $\bar{P}$, *n* and *k* are respectively the observational value, the predicted amount, the average observational values, the average predicted amount, the number of observations, and the number of independent variables in the ML models.

8.1.15. Results and discussion

8.1.15.1. Sensitivity analysis

Figure 8-34 demonstrates the statistical indices calculated for the ANN, DTR, and SVR models. Model 1 estimated the iceberg drafts using L/H , w/H , $M/\rho_i.H^3$, and S_f parameter. The *RMSE* index for the ANN 1, DTR 1, and SVR 1 models was equal to 0.698, 0.848, and 0.896. The effect of the iceberg shape factor (S_f) was disregarded for model 2, where the value of the *AIC* criterion for the ANN 2, DTR 2, and SVR 2 models were respectively computed to be 132.007, 18.512, and 15.250. The iceberg mass ratio ($M/\rho_i.H^3$) was an eliminated factor for Model 3, e.g., ANN 3, DTR 3, and SVR 3, with a *WI* value of 0.455, 0.785, and 0.870, in turn. The value of the *CRM* statistical index for ANN 4, DTR 4, and SVR 4 was -1.016, 0.025, and 0.104 as the influence of the iceberg width ratio (w/H) was removed for these ML models. The iceberg length ratio (L/H) was ignored for the iceberg draft estimation using the ANN 5, DTR 5, and SVR 5 models when the *RMSE* amount for such models was surmised as 1.598, 2.059, and 1.713. Models 6 to 9 were the function of solely one input parameter, e.g., L/H , w/H , $M/\rho_i.H^3$, and S_f, respectively. The simulation outcomes demonstrated that the ANN 5, DTR 1, and SVR 3 models were detected as the superior models among the ANN, DTR, and SVR models. According to the performed sensitivity analysis, the iceberg length ratio (L/H) and the iceberg width ratio (w/H) had the highest degree of effectiveness to simulate the iceberg drafts, while the iceberg mass ratio ($M/\rho_i.H^3$) possessed an insignificant influence.

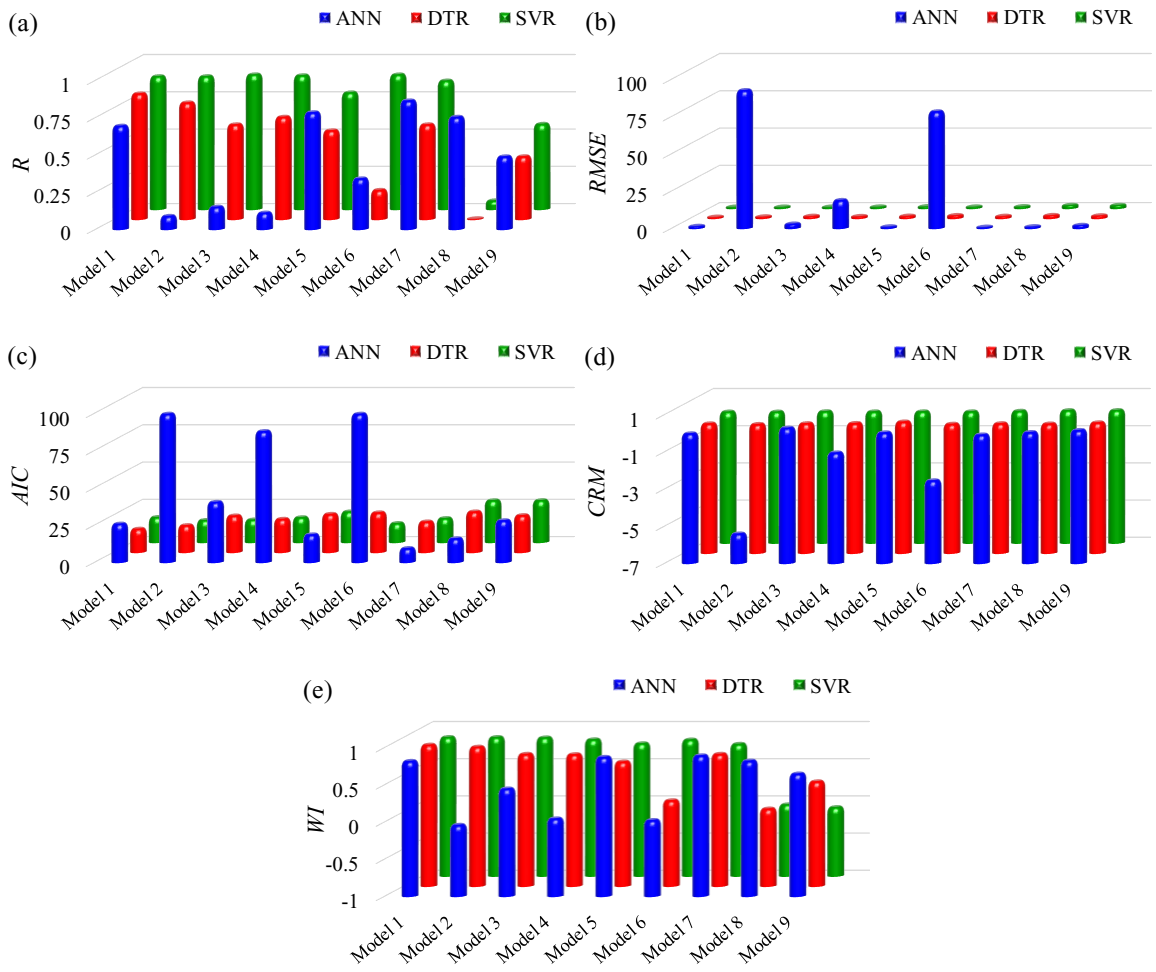


Figure 8-34. Comparison between the statistical indices calculated for the ANN, DTR, and SVR models

The error analysis showed that roughly one-fourth of iceberg drafts predicted by the ANN 1 model had an error of less than 10%, whereas this amount for the ANN 2, ANN 3, ANN 4, and ANN 5 models was approximately 17%, 23%, 31, and 33%, in turn. Almost 53% of the ANN 5 model's results possessed an error of smaller than 20% but this value for the ANN 6, ANN 7, ANN 8, and ANN 9 models was nearly 40%, 34%, 39%, and 50%, respectively. About one-third of the DTR 1, DTR 2, DTR 3, DTR 4, and DTR 5 models' results had an error of less than 18%; however, this

value for the DRE 6, DTR 7, DTR 8, and DTR 9 models was near 21%, 26%, 18%, and 25%. Just about half of the iceberg drafts predicted by the SVR 3 model showed an error of less than 20% and this figure for the SVR 4 and SVR 5 was 42% and 40%. The performed error analysis demonstrated that ANN 5, DTR 1, and SVR 3 had the highest level of accuracy among the ANN, DTR, and SVR models, respectively.

8.1.15.2. Comparison between the superior models

The performed analysis in the previous section demonstrated that the ANN 5, DTR 1, and SVR 3 models were the superior models to estimate the iceberg drafts. Figure 8-35 exhibits the scatter plots for these superior models. The value of correlation coefficient (R) for the ANN 5, DTR 1, and SVR 3 models was respectively obtained at 0.789, 0.848, and 0.908. Hence, the SVR 3 predicted the iceberg drafts with the highest level of correlation.

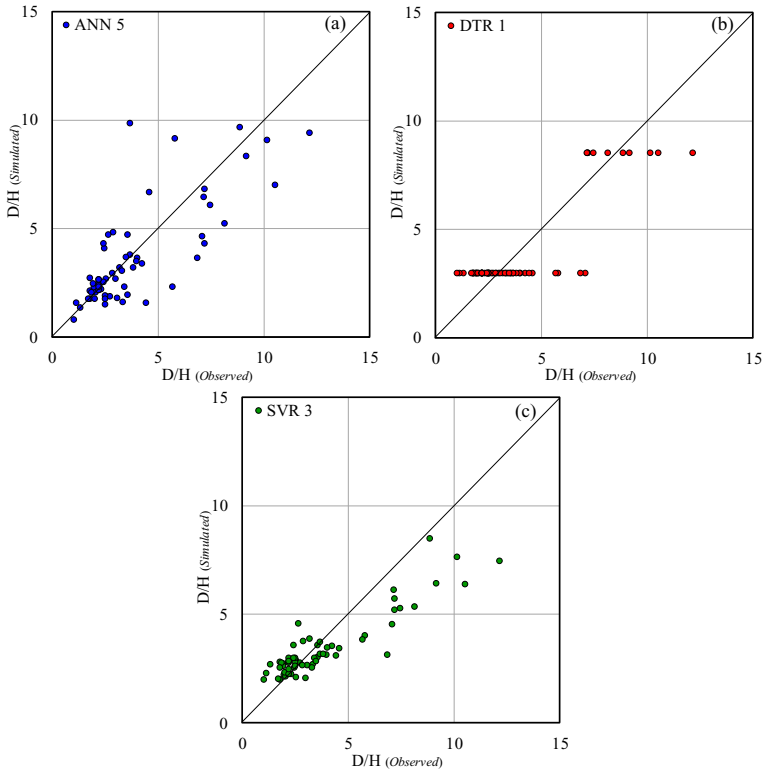


Figure 8-35. Scatter plots for the superior models (a) ANN 5 (b) DTR 1 (c) SVR 3

The study performed an uncertainty analysis (UA) to further assessment of the ETR models' performance. To do so, the ETR model's error was calculated as the difference between the iceberg drafts predicted through this model and the actual iceberg drafts. The mean (Mean) and the standard deviation (StDev) of such error, values were obtained. An individual ETR model underestimated the iceberg draft if the sign of the Mean value was negative, while the positive sign of the Mean meant that the ML models overestimated the iceberg drafts. Thereon, a confidence interval (CI) was produced near the error counted using the Mean, StDev values, and the "Wilson score technic" by omitting the continuity correction. A normal distribution interval corrected as an asymmetric normal distribution, named the Wilson score interval, was employed to adjust the CI bounds. Subsequently, a $\pm 1.96 S_e$ yielded a 95%CI. It should be remarked that the width of the uncertainty bound (WUB) is half of the difference between the lower and upper bound (Azimi et al. 2023). Figure 8-36 illustrates the binomial and normal error distribution of the ANN 5, DTR 1, and SVR 3 models. Regarding the performed uncertainty analysis, the ANN 5, DTR 1, and SVR 3 models overestimated the iceberg drafts, with a Mean value of 0.252, 0.019, and 0.432. Moreover, the lowest value of StDev was known for the SVR 3 model. Although the widest uncertainty bound was calculated for the ANN 5 (WUB=±0.398), the width of the uncertainty bound for the DTR 1 and SVR 3 was equal to ±0.338. Therefore, the SVR 3 model as a function of L/H , w/H , and S_f was recognized to be the superior ML model to simulate the iceberg drafts in the present study. This model showed the lowest degree of complexity alongside the highest degree of accuracy and correlation with the observational values, where it biased toward the overestimation to predict the target parameter. It is worth noting that the iceberg length ratio (L/H)

and the iceberg width ratio (w/H) were known as the most influential inputs in order to model the iceberg drafts.

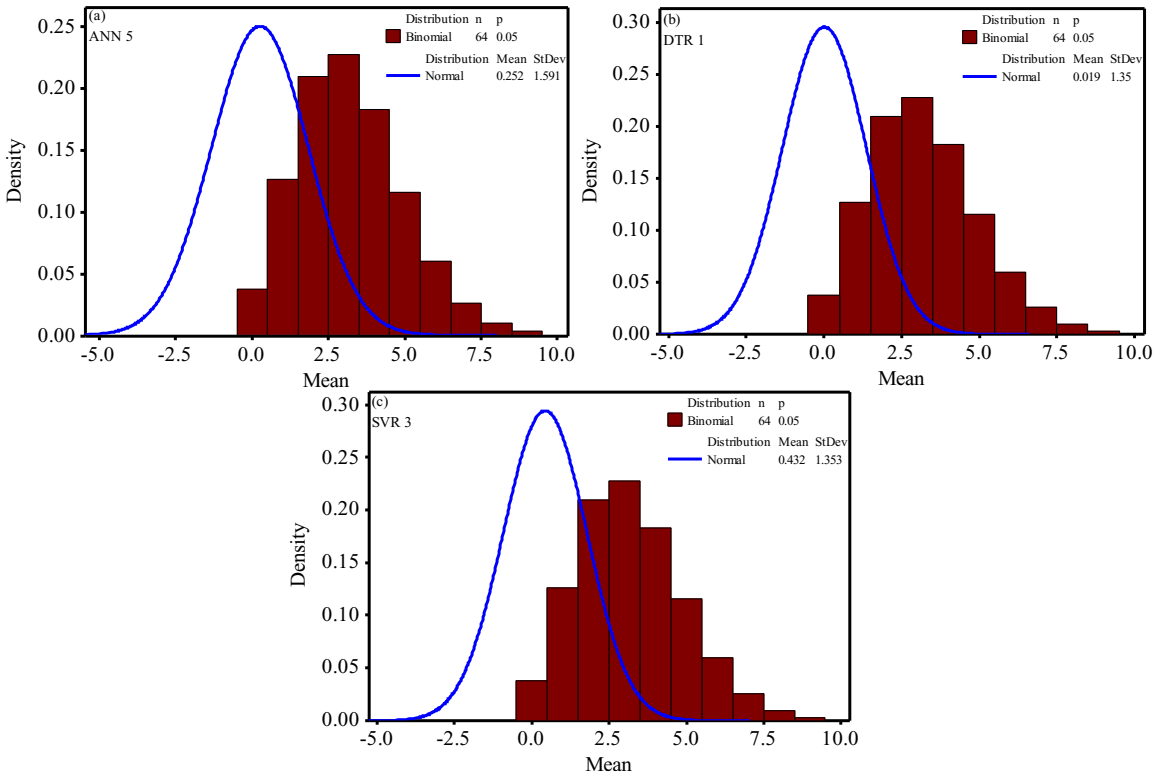


Figure 8-36. Binomial and normal error distribution of the superior models (a) ANN 5 (b) DTR 1 (c) SVR 3

8.1.16. Conclusion

In the present study, three ML algorithms, e.g., ANN, DTR, and SVR, were used to model the iceberg drafts. By performing several analyses including, the sensitivity, error, and uncertainty analyses, the superior ML model alongside the most influencing input parameters was distinguished. The SVR algorithm outperformed the ANN and DTR methods in the estimation of the iceberg drafts. The conducted sensitivity analysis proved that the SVR 3 model as a function of L/H , w/H , and S_f was the superior ML model for the simulation of iceberg drafts. The SVR

3 model possessed the highest degree of precision, correlation, and simplicity, meaning that the value of *RMSE*, *R*, and *AIC* for this model was reckoned to be 1.411, 0.908, and 15.562. Regarding the performed error analysis, almost one-fifth of the simulation results of the SVR 3 model had an error of smaller than 12%. The performed uncertainty analysis revealed that the SVR 3 model overestimated the iceberg drafts, with the lowest value of the error standard deviation (StDev) at 1.353. The iceberg length ratio (L/H) and the iceberg width ratio (w/H) were found to be the most effective input parameters to model the iceberg drafts using the SVR algorithm. The obtained results gave a good understanding of the modeling of the iceberg drafts through ML algorithms in order to protect the offshore structures and subsea assets in the Arctic and subarctic areas. These outcomes can facilitate proposing of cost-effective and quick alternatives in the early stages of iceberg management projects and subsea structure designs.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Azimi, H., Shiri, H., 2020. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): pp.051601.

Azimi, H., Shiri, H., 2021a. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04): pp.411-420.

Azimi, H., Shiri, H., 2021b. Modeling subgouge sand deformations by using multi-layer perceptron neural network. In the 31st International Ocean and Polar Engineering Conference. Rhodes, Greece, ISOPE-I-21-2150.

Azimi, H., Shiri, H., Mahdianpari, M., 2022. Simulation of Subgouge Sand Deformations Using Robust Machine Learning Algorithms. Offshore Technology Conference, OTC-31937-MS.

Azimi, H., Shiri, H., Zendehboudi, S., 2023. Prediction of Ice-Induced Subgouge Soil Deformation in Sand Using Group Method of Data Handling–Based Neural Network. Journal of Cold Regions Engineering, 37(2), pp.04023006.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In the 14th international offshore and polar engineering conference, ISOPE-I-04-116.

Blažauskas, N., Włodarski, M., Paulauskas, S., 2013. Perspectives for offshore wind energy development in the South-East Baltics. Klaipėda University: Klaipėda, Lithuania. 1-54.

Dowdeswell, J.A., Bamber, J.L., 2007. Keel depths of modern Antarctic icebergs and implications for sea-floor scouring in the geological record. Marine Geology, 243(1-4): pp.120-131.

El-Tahan, M., El-Tahan, H., Courage, D., Mitten, P., 1985. Documentation of Iceberg Groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Arctic Technology Conference, OTC-27358-MS.

Løset, S., Carstens, T., 1996. Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Regions Science and Technology, 24(4): 323-340.

McGuire, P., Younan, A., Wang, Y., Bruce, J., Gandi, M., King, T., Regular, K., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

McKena, R., 2000. Study of iceberg scour & risk in the Grand Banks region. KR Croasdale & Associates Ltd., Ballicater Consulting Ltd., Canadian Seabed Research Ltd., C-CORE, and Ian Jordaan & Associates Inc., PERD/CHC Report, 31-26.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, pp. 9-19.

Pekel, E., 2020. Estimation of soil moisture using decision tree regression. Theoretical and Applied Climatology, 139(3): 1111-1119.

Rudkin, P., 2005. Comprehensive Iceberg Management Database Report 2005 update. PERD/CHC Report 20-72. National Research Council of Canada (NRC) and Panel on Energy Research and Development (PERD).

Sacchetti, F., Benetti, S., Cofaigh, CÓ., & Georgiopoulou, A., 2012. Geophysical evidence of deep-keeled icebergs on the Rockall Bank, Northeast Atlantic Ocean. Geomorphology 159: 63-72.

Sonnichsen, G., Myers, R., Simpson, K., Brown, R., McClintock, J., Bullock, T., Rudkin, P., 2003. Seabed Surveys to Document Groundings from the 2000 Grand Bank Iceberg Season. Proceedings of the 17th International Conference on Port and Ocean Engineering under Arctic Conditions. POAC 03.

Sonnichsen, G., Hundert, T., Myers, R., Pocklington, P., 2006. Documentation of Recent Iceberg Grounding Evens and a Comparison with Older Events of Know Age. Environmental Studies Research Funds. Report ESRF Vol. 157.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., King, T., 2021. Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. In International Conference on Offshore Mechanics and Arctic Engineering, 85178, pp.V007T07A018.

Talimi, V., Ni, S., Qiu, W., Fuglem, M., MacNeill, A., Younan, A., 2016, Investigation of Iceberg Hydrodynamics. Arctic Technology Conference, OTC-27493-MS.

Turnbull, I.D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Regions Science and Technology, 110: 1-18.

Turnbull, I.D., King, T., Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory, Springer, New York.

Woodworth-Lynas, C.M.T., Simms, A., Rendell, C.M., 1985, Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Regions Science and Technology, 10(2): 163-186.

Younan, A., Ralph, F., Ralph, T., Bruce, J., 2016. Overview of the 2012 iceberg profiling program. Arctic Technology Conference, OTC-27469-MS.

Zhou, M., 2017. Underwater iceberg profiling and motion estimation using autonomous underwater vehicles. PhD, Memorial University of Newfoundland, NL, Canada.

Part III

Prediction of Iceberg Draft and iceberg Response to Ice-gouging Using Machine Learning Algorithms: Integration of Part I and Part II

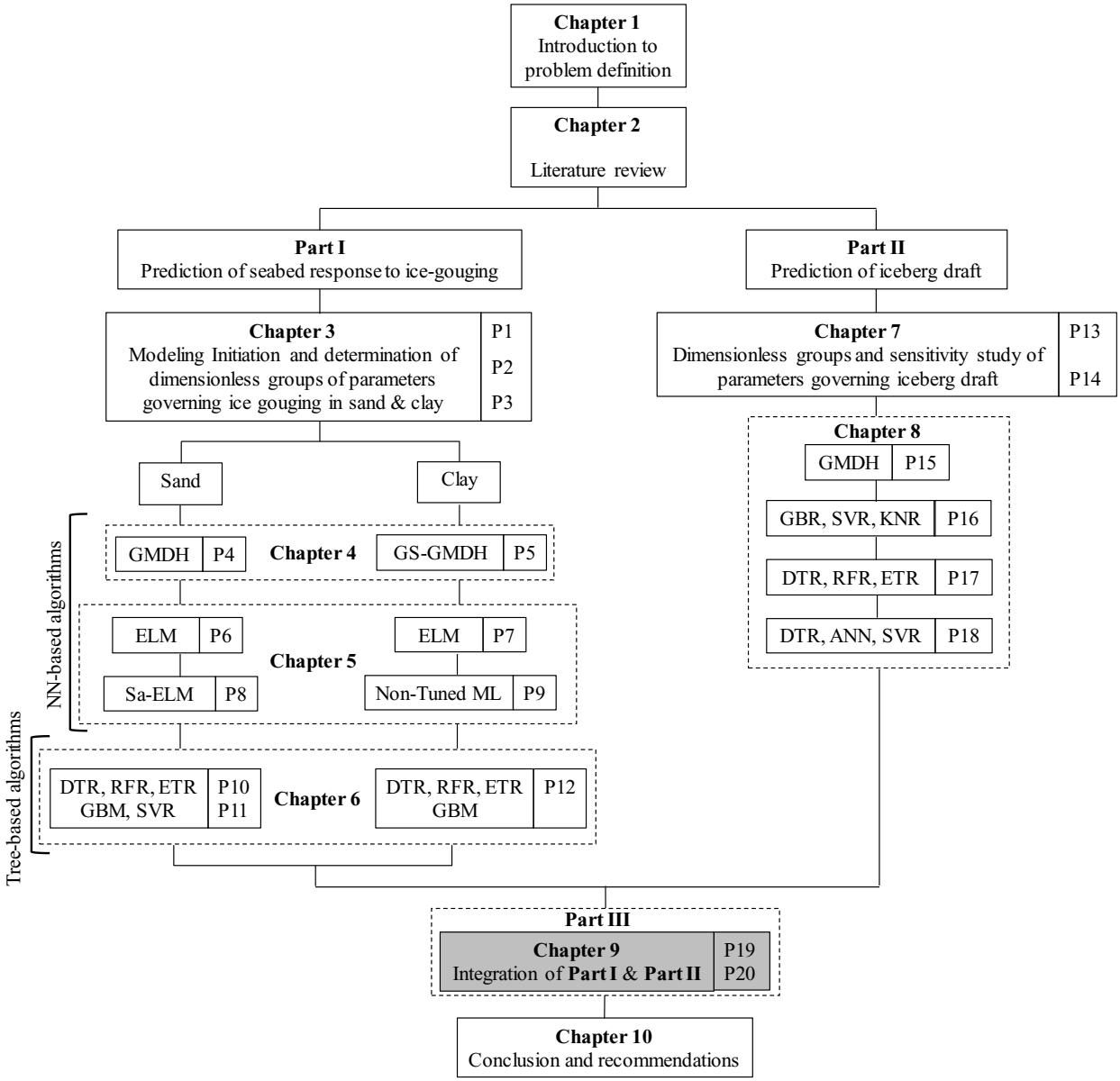
9. Chapter 9

Prediction of Iceberg Draft and iceberg Response to Ice-gouging Using Machine Learning Algorithms

This chapter includes two sections as follows:

Section 1: Evaluation of Iceberg Draft and Iceberg-Seabed Interaction Using Random Forest Regression Algorithm

Section 2: Simulation of Iceberg Draft and Subgouge Soil Characteristics Using Extra Tree Regression Algorithm


Chapter 1
Introduction to problem definition
Chapter 2
Literature review
Part I
Prediction of seabed response to ice-gouging
Part II
Prediction of iceberg draft
Chapter 3
Modeling Initiation and determination of dimensionless groups of parameters governing ice gouging in sand & clay
P1
P2
P3
Chapter 7
Dimensionless groups and sensitivity study of parameters governing iceberg draft
P13
P14
Chapter 8
GMDH P15
GBR, SVR, KNR P16
DTR, RFR, ETR P17
DTR, ANN, SVR P18
Sand
Clay
NN-based algorithms
GMDH P4
Chapter 4
GS-GMDH P5
ELM P6
ELM P7
Chapter 5
Sa-ELM P8
Non-Tuned ML P9
Tree-based algorithms
DTR, RFR, ETR GBM, SVR P10 P11
Chapter 6
DTR, RFR, ETR GBM P12
Part III
Chapter 9
Integration of Part I & Part II
P19
P20
Chapter 10
Conclusion and recommendations

Section 1

Evaluation of Iceberg Draft and Iceberg-Seabed Interaction Using Random Forest Regression Algorithm

This section was published in the 33rd International Ocean and Polar Engineering Conference (ISOPE), 2023, Ottawa, Canada

Abstract

In the present study, the iceberg drafts and iceberg-seabed interaction process were simulated using the random forest regression (RFR) algorithm for the first time. Initially, utilizing the parameters governing the iceberg drafts and the iceberg-seabed interaction process in the sandy seabed, a set of RFR models were developed. To train and test the RFR models, a comprehensive dataset was subsequently constructed by using the field and experimental values reported in the published literature. By performing a sensitivity analysis, the premium RFR model and the most significant input parameters were introduced.

Keywords: Iceberg draft; Iceberg-seabed interaction; Random forest regression (RFR); Subsea assets; Iceberg management system; Arctic area.

9.1.1. Introduction

Approximately 22% of the Earth's undiscovered hydrocarbons are stored in the Arctic region. Therefore, the Arctic area is one of the best resources to grow oil and gas loading equipment. Moreover, the Arctic offshore regions with rich wind culture have a high potential for the development of offshore wind farms. This combined with climate change and global warming means an increased number of ice management operations to protect the subsea pipelines, power cables, and offshore and subsea structures. The schematic layout of the iceberg free-floating and iceberg scouring in cold waters is shown in Figure 9-1.

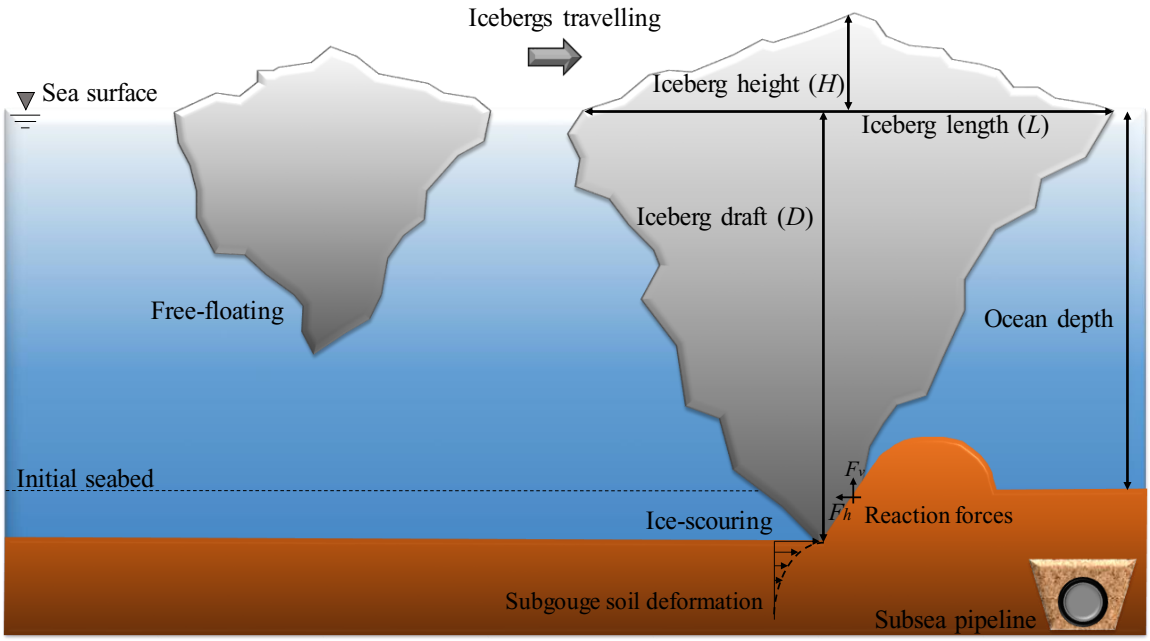


Figure 9-1. Schematic layout of the iceberg free-floating and iceberg scouring in cold waters

As seen, the iceberg is in a free-floating situation if the ocean depth is greater than the iceberg draft; otherwise, the seafloor is scoured, and the seabed soil shear resistance causes the soil displacement to extend deeper than the iceberg tip threatening the buried subsea assets.

The efficient iceberg management designs and the guaranteed operational integrity of the sea bottom-funded infrastructure against the iceberg attacks in the ice-prone areas demand the appropriate iceberg draft appraisal, which may lead to a potential decrease in operating expenses and downtime. Earlier investigations have tended to focus on modeling the iceberg draft by using the iceberg length or iceberg mass (Robe and Farmer 1976; Hotzel and Miller 1983; Barker et al. 2004). King et al. (2016) performed a field investigation to calculate the iceberg rolling rate. The iceberg drafts were estimated utilizing a calving analysis, with a calculated standard deviation of draft variations from 19% to 34%. The iceberg drafts corresponded with the mass of the icebergs. In another investigation, Turnbull et al. (2018) proposed a model for the drift estimation of moving icebergs on the Grand Banks of Newfoundland. This model approximated the draft of icebergs roughly 1.3 times more than the real values.

McKenna et al. (2019) have recently simulated ice scouring on the Grand Banks of Canada using the Monte Carlo method. The iceberg draft alterations were utilized to lessen the size of draft variations in this modeling. Most recently, Stuckey et al. (2021) simulated the three-dimensional iceberg shapes by adopting field measurements. The investigation demonstrated that the draft and mass of the iceberg were estimated in terms of the iceberg length by utilizing the power curve.

Despite the offshore structures such as ships, wind turbines, and floating or fixed platforms that can be directly attacked by icebergs, the threat that goes to the buried pipelines and power cables is much more complicated involving the ice-soil-structure interactions. During the ice-gouging, the subgouge soil displacement is largely vastly extended down the seabed much deeper than the gouge depth. This may cause large deformations in the pipelines buried even below the gouge depth. Investigation of the subgouge soil deformation and the best trench depth to bury the pipelines and power cables have been a hot topic in the literature over the past few decades (Paulin

1992; Hynes 1996; Yang 2009). Arnau Almirall (2017) conducted several 1g laboratory tests to perform the subgouge sand features in saturated and dry circumstances. The study demonstrated that the ice-induced sand deformations in the 1g test are less than those resulting from the centrifuge tests. Hashemi and Shiri (2022) simulated the iceberg-seabed interaction in clay by incorporating the strain rate and strain-softening effects. This study concluded that strain softening had a significant influence on a larger amount of plastic shear strain below and in front of the iceberg tip.

Moreover, artificial intelligence (AI) models and machine learning (ML) algorithms have been recently employed for the analysis of iceberg-seabed interaction (Kioka et al. 2003, 2004; Azimi and Shiri 2020a, 2020b; Azimi and Shiri 2021a, 2021b, 2021c; 2021d, Azimi et al. 2021) since these tools are cost-effective, fast, and accurate. Azimi et al. (2022a) utilized an evolutionary design of the generalized group method of data handling (GS-GMDH) to model the iceberg-seabed interaction mechanism. The comparison of the best GS-GMDH model with the artificial neural network (ANN) and the GMDH algorithm showed the better performance of the GS-GMDH model. Azimi et al. (2022b) simulated the horizontal and vertical subgouge soil deformations in clay by using the Decision Tree Regression (DTR), Random Forest Regression (RFR), and Extra Tree Regression (ETR) models. The simulation results demonstrated that the ETR model possessed the highest degree of precision and correlation with the experimental values. Azimi et al. (2022c) estimated the iceberg-seabed interaction characteristics in clay seabed through decision tree regression (DTR), random forest regression (RFR), and gradient boosting regression (GBR) algorithms. The authors stated that the ETR algorithm was able to predict the target parameters with a better performance. Literature shows that there is no study to predict iceberg drafts and subgouge soil features at the same time. Hence, to fill this knowledge gap, a tree-based ML

algorithm is used in the current study to simulate the iceberg drafts and subgouge soil characteristics simultaneously. More details are presented in the next sections.

Every year thousands of icebergs are born out of glaciers in the Arctic zone and carried away by the currents and into the North Atlantic. These icebergs may touch the sea bottom in shallow waters and scratch the seabed, causing so-called "ice-gouging" that may endanger the integrity of subsea pipelines and power cables. Currently, so-called "ice management," i.e., iceberg towing and re-routing, is the most reliable approach to protect the subsea and offshore infrastructures, where the threatening icebergs are hooked and towed in a safe direction. Ice management is a costly operation and needs standby marine spread with a range of advanced tools and equipment, such as subsea survey facilities, to investigate the iceberg draft and determine if it is a threat to infrastructures.

On the other hand, expensive numerical simulations and experimental testing programs are conducted to explore the soil displacement under the tip of the iceberg, scouring the seabed or so-called "subgouge soil displacement" that, in turn, governs the structural response of the buried subsea pipelines and power cables. In this research program, using advanced Machine Learning (ML) algorithms, the iceberg draft and subgouge soil parameters were simulated.

Firstly, the parameters affecting the iceberg draft estimation along with the parameters governing the subgouge soil feature modeling were identified utilizing Buckingham's theorem. Subsequently, regarding the identified parameters affecting the problems, two distinct datasets for iceberg draft and subgouge soil feature estimation were constructed by using the published papers and reports that have been cited in this work. Each dataset was divided into two subsets comprising the training and testing subsets. The ML algorithm was fed with these parameters affecting to estimate the iceberg drafts and subgouge soil characteristics in the training mode. After that, the ML model was validated by employing the testing datasets. The results of ML models were

analyzed to introduce the best combination of the parameters affecting, the superior ML model, alongside the most significant parameters. The applied ML algorithm in the current study as a fast and cost-effective alternative can be used to simulate the iceberg drafts and iceberg-seabed interaction process, particularly in the early stages of iceberg management systems and subsea design projects.

9.1.2. Methodology

9.1.2.1. Random forest regression (RFR)

RF algorithm has been developed by Breiman (2001) to solve classification and regression problems. The RF algorithm, as an ensemble learning method, is the extension of the Classification and Regression Trees (CART) algorithm where the tree-based CART models have the overfitting problem and the RF algorithm overcomes this limitation (Breiman 2001). Figure 9-2 demonstrates the RFR model's flowchart in the present study.

In the RF algorithm, several decision trees are created and decisions with the highest number of votes are chosen as the simulation results. In the random forest regression, the number of trees for a random vector is increased and the tree predictor, input parameters, and output parameters are considered numerical values. Regarding the random vector distribution, the training dataset is independently utilized. Therefore, the mean-squared generalization error for the tree predictor is calculated (Sahani and Ghosh 2021). The RF model reduces the average error of the trees utilized by utilizing the weighted correlation between the residuals and the applied randomization ought to be employed at a low level of correlation. In this study, the hyperparameters tuned in the RFR algorithm including the *max_samples*, *bootstrap*, *max_depth*, *random_state*, *number of estimators*, and *verbose* were adjusted by a trial and error procedure at *None*, *True*, 12, 6, 8, 1, respectively.

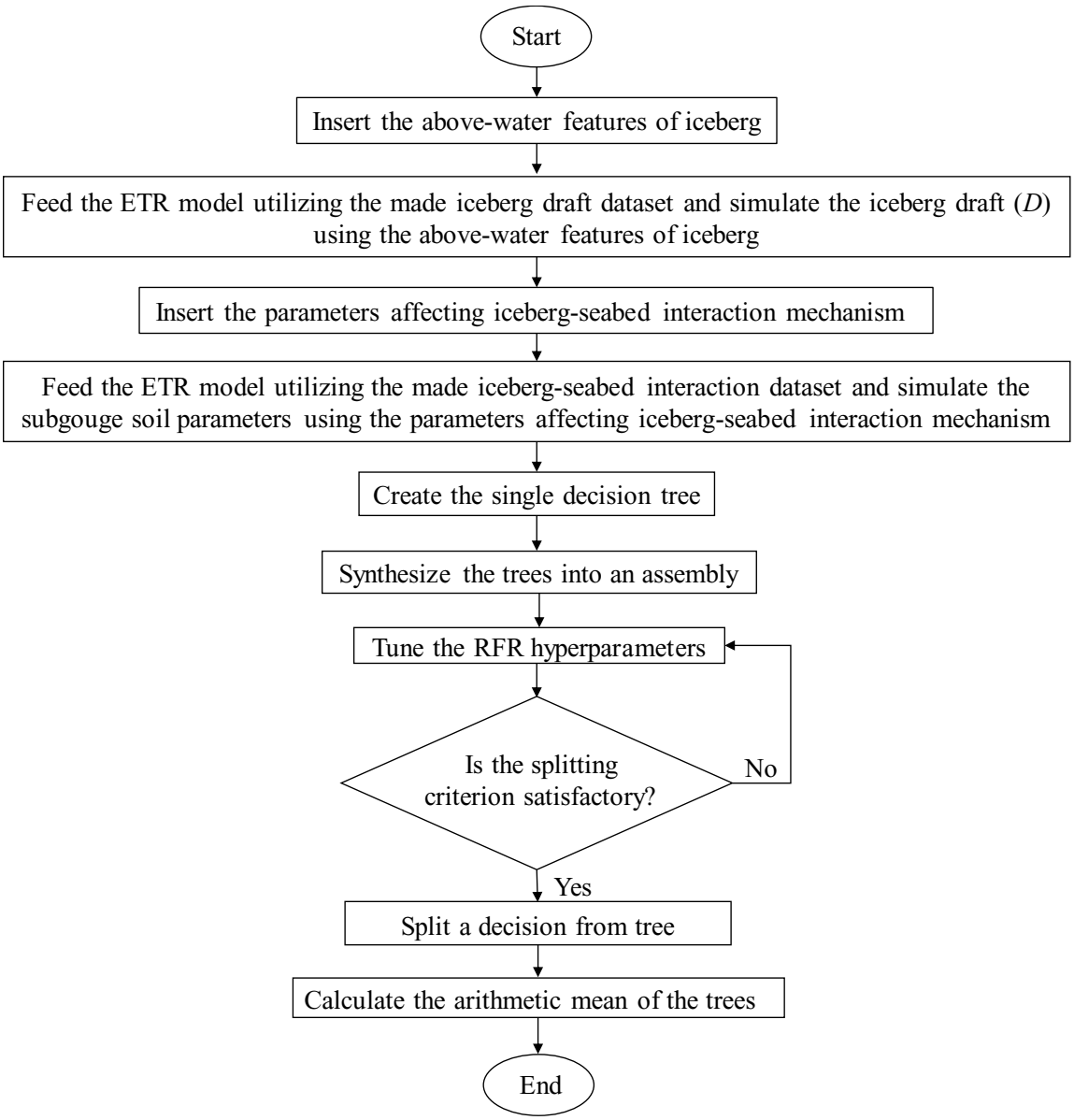


Figure 9-2. RFR model's flowchart in the present study

9.1.2.2. Iceberg draft

The iceberg draft (D) was considered as a function of the physical characteristics of the iceberg, comprising the iceberg length (L), iceberg height (H), iceberg width (w), iceberg mass (M) in so

many fields, analytical, and numerical investigations in the form below (Barker et al. 2004; McKenna et al. 2019; and Stuckey et al. 2021):

$$D = f_1(L, H, w, M) \tag{9-1}$$

Furthermore, the density of an iceberg (ρ_i), the density of seawater (ρ_{sw}), seawater viscosity (μ_{sw}), and gravitational acceleration (g) may influence the iceberg draft as follows:

$$D = f_2(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g) \tag{9-2}$$

The iceberg shape factor (S_f) signifies the global shape of icebergs, which can affect the magnitude of the iceberg draft (Turnbull et al. 2018). The shape factor of the traveling icebergs is considered universally into six categories, e.g., the shape factor for tabular, blocky, domed, dry dock, pinnacle, and wedge icebergs are 0.5, 0.5, 0.41, 0.15, 0.25, 0.33, respectively (Rudkin 2005). Hence, equation (9-2) can be summarized below:

$$D = f_3\left(L, H, w, M, \rho_i, \rho_{sw}, \mu_{sw}, g, S_f\right) \tag{9-3}$$

Presumably, the density along with the viscosity of the seawater is constant, and the value of gravitational acceleration may be regarded as a constant value. The dimensional form of equation (9-3) is written below:

$$D = f_5(\Pi_1, \Pi_2, \Pi_3, \Pi_4) \tag{9-4}$$

here, Π_1, Π_2,…, and Π_4 are dimensionless groups and f_5 is a functional symbol based on the Buckingham-π theorem. Thus, equation (9-5) is then formulated as a function of four dimensionless groups as follows:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \tag{9-5}$$

Therefore, D/H is the iceberg draft ratio and is a function of the length ratio (L/H), width ratio (w/H), mass ratio ($M/\rho_i . H^3$), and iceberg shape factor $\left(S_f\right)$.

9.1.2.3. Iceberg-seabed interaction process

The ice-induced interaction parameters (η) in a sand mass including soil deformations (d/W) and reaction forces ($F/\gamma sW3$) are a function of a set of parameters including the scour depth (Ds), the internal friction angle of sand (φ), the width of gouge (W), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (h'), the horizontal load (L_h), the vertical load (L_v), the velocity of iceberg keel (V), and the specific weight of sand (γ_s) (Azimi and Shiri 2020a). The maximum subgouge deformation in the sand $\left(d_{(sand)}\right)$ is formed just under the moving iceberg keel in the gouge centerline. However, at greater depth on the subgouge centerline, by incorporating the soil depth (y/W), the subgouge soil deformation in the sand ($d_{(sand)}/W$) is written as follows:

$$\frac{d_{(sand)}}{W} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{9-6}$$

Similarly, the ice-induced reaction force (F) is a function of the position of the iceberg along the scour axis (x) and $D_s, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s$ so equation (9-6) can be written as below:

$$\frac{F_{(sand)}}{\gamma_s W^3} = f_4\left(\frac{x}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \tag{9-7}$$

Here, Lh and Lv indicate the consequence of horizontal and vertical loads. Figure 9-3 illustrates the combination of input parameters for the development of the RFR model to simulate iceberg draft and subgouge soil characteristics.

(a)

	L/H	w/H	M/ρi.H3	Sf
RFR 1				
RFR 2				
RFR 3				
RFR 4				
RFR 5				

(b)

Model	x/W	Ds/Ws	φ (Deg)	α (Deg)	h'/W	Lh/γs.W3	Lv/γs.W3	V2/gW
RFR 6								
RFR 7								
RFR 8								
RFR 9								
RFR 10								
RFR 11								
RFR 12								
RFR 13								
RFR 14								

Figure 9-3. Combination of input parameters for the development of RFR model to simulate (a) iceberg draft (b) subgouge soil characteristics

9.1.2.4. Construction of dataset

The measured values of 12 field studies reported by El-Tahan et al. (1985) (38 cases), Woodworth-Lynas et al. (1985) (one case), Løset and Carstens (1996) (52 cases), Barker et al. (2004) (14 cases), McKenna (2004) (two cases), Sonnichsen et al. (2006) (nine cases), Turnbull et al. (2015) (two cases), McGuire et al. (2016) (eight cases), Younan et al. (2016) (29 cases), Talimi et al. (2016) (one case), Zhou (2017) (three cases), Turnbull et al. (2018) (two cases) were used.

The key measurements of six experimental datasets reported by Paulin (1991, 1992) (P-1 to P-5), C-CORE (1995, 1996) (C-1 to C-10), Hynes (1996) (H-1 to H-5), and Yang (2009) (Y-1 to Y-7) were applied to validate the subgouge soil features simulated by RFR algorithm.

To assess the quality of the datasets, the t-test and the P-value for the datasets were calculated, presuming that the P-value of 0.05 or less is statistically significant, where an alternative hypothesis influences the likelihood of the relationship between the observed values. This P-value

for the iceberg draft dataset and iceberg-seabed interaction dataset was respectively estimated as 0.008 and 0.039 representing that the correlations were statistically significant. It should be stated that 70% of the constructed dataset was utilized for training the RFR models, whilst 30% of the rest was employed for testing these ML models.

9.1.2.5. Goodness of fit

A set of criteria, such as correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were utilized to evaluate the accuracy, correlation, and complexity of the RFR models. The closeness of the *R* and *WI* indices to one signified that the model had a high correlation with the experimental values. On the other hand, the closeness of the *RMSE*, *MAPE*, and *CRM* values to zero meant that a particular model had the lowest level of error though the complexity of these models was not assessed using the defined criteria. To overcome this limitation, the Akaike Information Criteria (*AIC*) was used, meaning that the less complex RFR model owned the lowest value of *AIC*; hence, the premium model had the lowest values of *AIC* index and error (*RMSE*, *MAPE*, and *CRM*), with the highest level of correlation (*R* and *WI*).

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{9-8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{9-9}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{9-10}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (9\text{-}11)$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \quad (9\text{-}12)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (9\text{-}13)$$

here, Oi, Pi, $\bar{O}$, $\bar{P}$, and n are the experimental measurements, the simulated values, the average experimental values, the average simulated values, and the number of experimental measurements, respectively. In addition, k is the number of input parameters in the RFR models.

9.1.3. Results and discussion

Firstly, the performance of the RFR models was examined through a sensitivity analysis. Then, the superior RFR models in the estimation of the iceberg drafts along with the subgouge soil parameters are introduced. Finally, the subgouge soil profiles and reaction forces simulated by the superior models are depicted.

9.1.3.1. RFR models

Figure 9-4 exhibits the computed statistical indices for the RFR 1 to RFR 5 models to predict the iceberg draft. Regarding the simulation results, the RFR 3 model showed the best performance so as to estimate the iceberg drafts, where this model was a function of L/H, w/H, and S_f, and the value of R, $RMSE$, and AIC criteria for it was 0.915, 1.186, and 10.734, respectively. The RFR 1 as a function of all inputs, e.g., L/H, w/H, $M/\rho_i.H^3$, and S_f, was ranked as the second-best model in order to approximate the iceberg drafts, with an AIC of 13.181. The impact of the iceberg

shape factor (S_f) was removed for the RFR 2 model, as the *RMSE* and *WI* values for this model were calculated to be 1.309 and 0.909. The RFR 4 model was recognized as the third-best model for the iceberg draft simulation, while the iceberg width ratio (*w/H*) was an eliminated input and the *R* index for this model was obtained at 0.871. Among the RFR models for the iceberg draft estimation, the RFR 5 model had the worst act to forecast the iceberg draft values since the *RMSE* and *R* indices for such a model were 1.345 and 0.856. According to the implemented sensitivity analysis, the iceberg length ratio (L/H) had the highest level of effectiveness to model the iceberg draft using the RFR algorithm although the iceberg mass ratio ($M/\rho_i.H^3$) had an insignificant impact in this aspect.

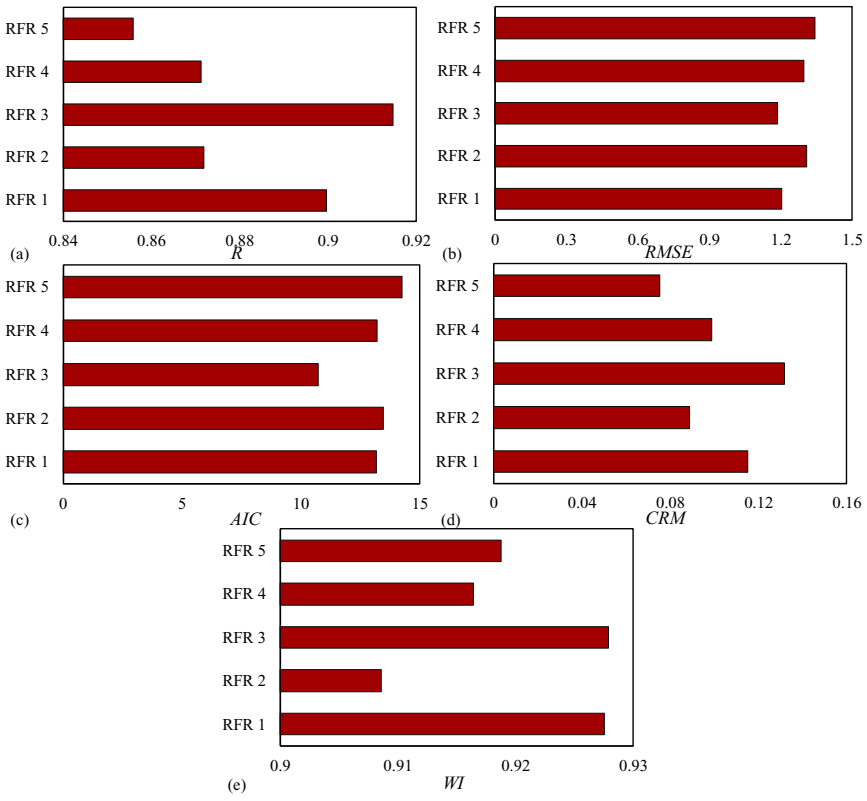


Figure 9-4. Key statistical indices calculated for the RFR 1 to RFR 5 models to simulate the iceberg drafts

Figure 9-5 illustrates the criteria computed for the RFR 6 to RFR 14 models to simulate the horizontal subgouge deformations. In this situation, RFR 7 had the highest amount of precision and correlation with experimental values (*RMSE*=0.079 & *WI*=0.967), where the influence of iceberg dynamic ($V^2/g.W$) was ignored for the RFR 2 model. After the RFR 2 model, the RFR 8, RFR 13, RFR 12, RFR 9, RFR 11, RFR 10, and RFR 6 were respectively introduced as the second-best to the eighth-best models. The inferior performance belonged to the RFR 14 because the value of *AIC* and *CRM* criteria for these models was -14.737 and 0.407. The simulation results proved that the soil depth ratio (y/W) and the berm height ratio (h'/W) were the most influential input parameters to model the horizontal subgouge soil deformations through the RFR algorithm.

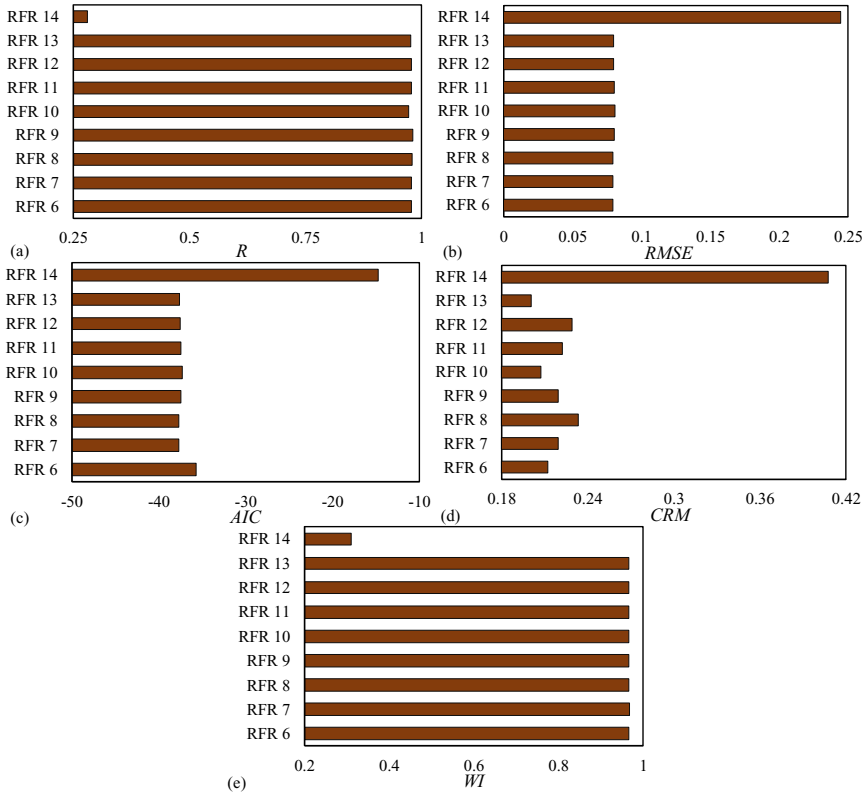


Figure 9-5. Key statistical indices calculated for the RFR 6 to RFR 14 models to simulate the horizontal subgouge deformations

Figure 9-6 displays the obtained statistical criteria for the RFR 6 to RFR 14 models to estimate the vertical subgouge soil deformations. As seen, RFR 12 was identified as the superior model to predict the vertical subgouge deformations. The RFR 12 simulated the vertical deformations in terms of $y/W, D_s/W, \alpha, h'/W, L_h/\gamma_s W^3, L_v/\gamma_s W^3,$ and V^2/gW, whereas the effect of the sand shear strength (φ) was extinguished, and the value of *AIC* and *CRM* indices for this model was -26.568 and 0.241. After RFR 12, the RFR 13 model was in second place in terms of performance to estimate the vertical deformations, with an *RMSE* of 0.015. The performed analysis demonstrated that the RFR 11, RFR 6, RFR 7, RFR 9, RFR 8, and RFR 14 models located third to eight places when it came to accuracy, correlation, and complexity. According to the sensitivity analysis, the berm height and soil depth ratios were known as the most significant input parameters to simulate the vertical displacements using the RFR algorithm.

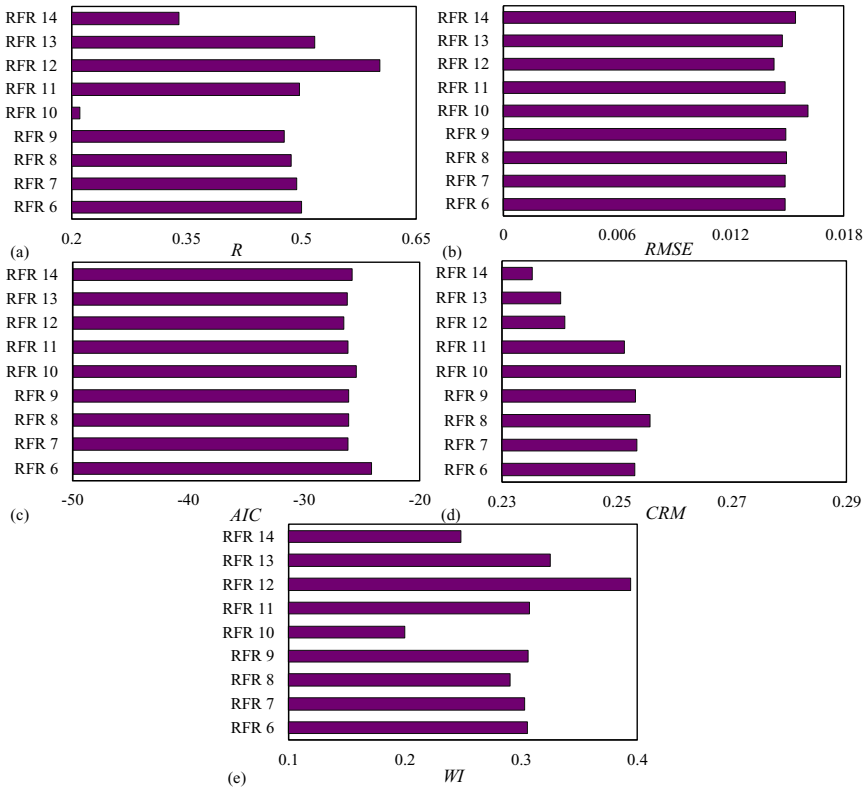


Figure 9-6. Key statistical indices calculated for the RFR 6 to RFR 14 models to simulate the vertical subgouge deformations

In Figure 9-7, the results of the statistical indices obtained for the ETR 6 to ETR 14 model to simulate the horizontal reaction forces are depicted. The RFR 9 as a function of $y/W, D_s/W, \varphi, \alpha, h'/W, L_v/\gamma_s W^3$, and V^2/gW was the best model to predict the horizontal reaction forces (R=0.995 & AIC=360.135), where $L_h/\gamma_s W^3$ was an eliminated factor. It is worth noting that the RFR 14 was identified as the worst model to approximate the horizontal reaction forces, with the *CRM* and *R* of -0.066 and 0.880. The simulation outcomes showed that the RFR 12, RFR 6, RFR 8, RFR 7, RFR 13, RFR 11, and RFR 10 models were located in the second to eighth places. Among the input parameters to model the horizontal reaction forces by the RFR algorithm, the position of the iceberg along the scour axis (*x/W*) and the berm height ratio (*h'/W*) were realized to be the most important inputs.

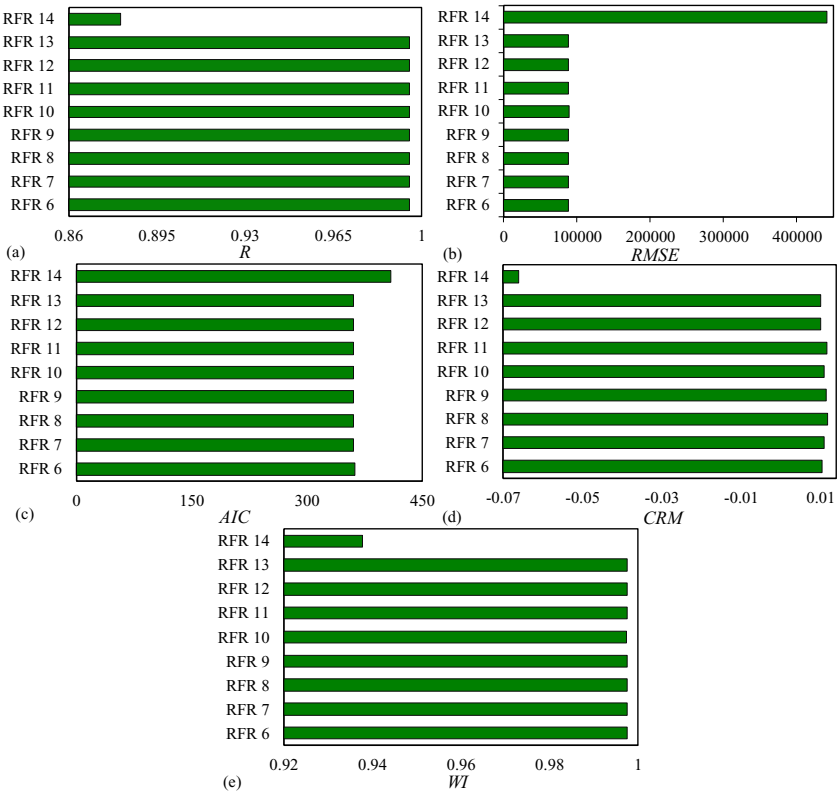


Figure 9-7. Key statistical indices calculated for the RFR 6 to RFR 14 models to simulate the horizontal reaction forces

Figure 9-8 shows the key statistical indices calculated for the RFR 6 to RFR 14 models to simulate the vertical reaction forces. Based on the obtained results, RFR 8 possessed the highest degree of precision (*RMSE*=39.198), correlation (*R*=0.980), and simplicity (*AIC*=101.629) to model the vertical reaction forces, as this model predicted the vertical reaction forces utilizing $y/W, D_s/W, \varphi, \alpha, h'/W, L_h/\gamma_s W^3,$ and V^2/gW. RFR 6 as a function of all inputs was also the second-best model to estimate the target parameter, where the *AIC* and *RMSE* criteria for this model were achieved to be 103.708 and 39.327. Additionally, the RFR 7, RFR 12, RFR 13, RFR 11, RFR 10, and RFR 9 models were placed in the third to eighth ranks in terms of efficiency, respectively. The conducted sensitivity analysis proved that the iceberg along the scour axis (x/W)

and the horizontal load factor ($Lh/\gamma s.W^3$) were identified as the most impactful input parameters to model the vertical reaction forces by the RFR algorithm.

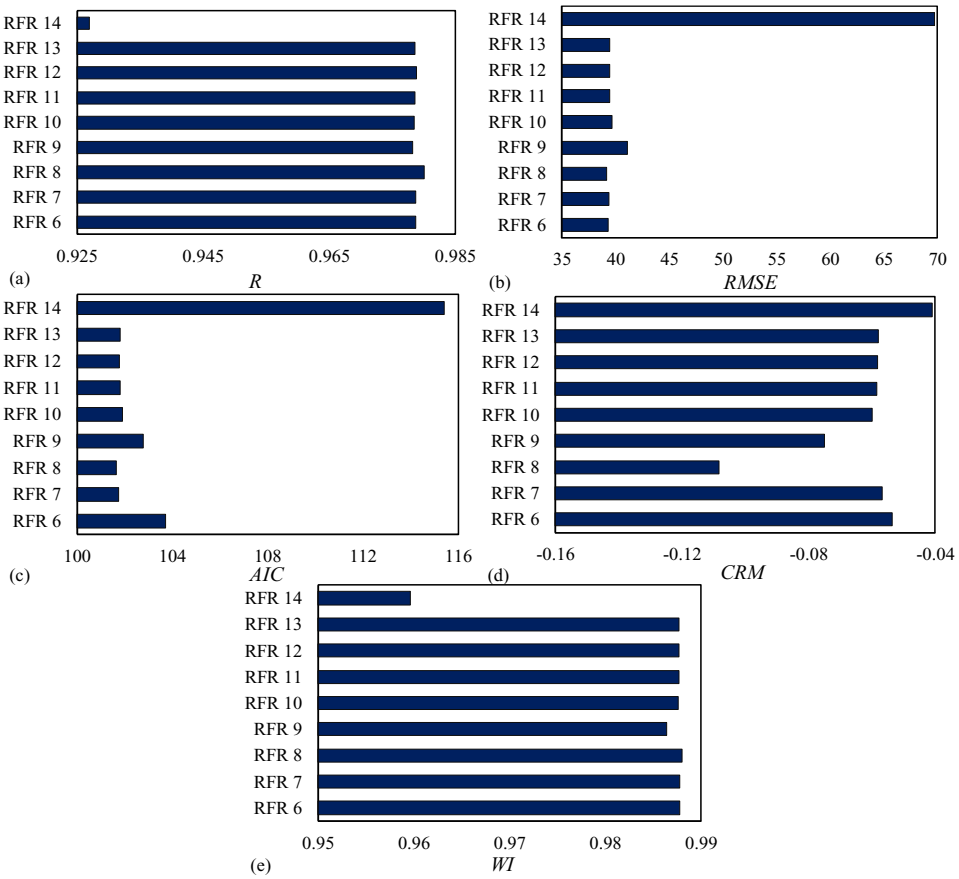


Figure 9-8. Key statistical indices calculated for the RFR 6 to RFR 14 models to simulate the vertical reaction forces

9.1.3.2. Superior RFR model

Figure 9-9 presents the comparison between the horizontal subgouge deformations predicted by the RFR 7 model and the experimental measurements. The maximum horizontal subgouge deformations occurred at the bottom of the iceberg keel and the magnitude of this parameter was reduced at greater depth. Despite a few discrepancies between the laboratory amounts and simulation outcomes (C-2, C-3, C-4, C-5, H-1, C-10, Y-1, Y-4, and Y-5), the RFR 7 model could

estimate the horizontal subgouge deformations with a reasonable agreement. The RFR 7 model employed both linear (H-1, H-4, C-7) and non-linear (P-1 to P-5) systems in estimating the horizontal displacements. Several discrepancies were observed between the modeled horizontal deformations and reported values (C-2 to C-4, C-10, Y-1, Y-4, and Y-5); however, the P-1 to P-5, C-1, H-3, H-4, C-6, and C-9 experimental cases were surmised with a strong correlation.

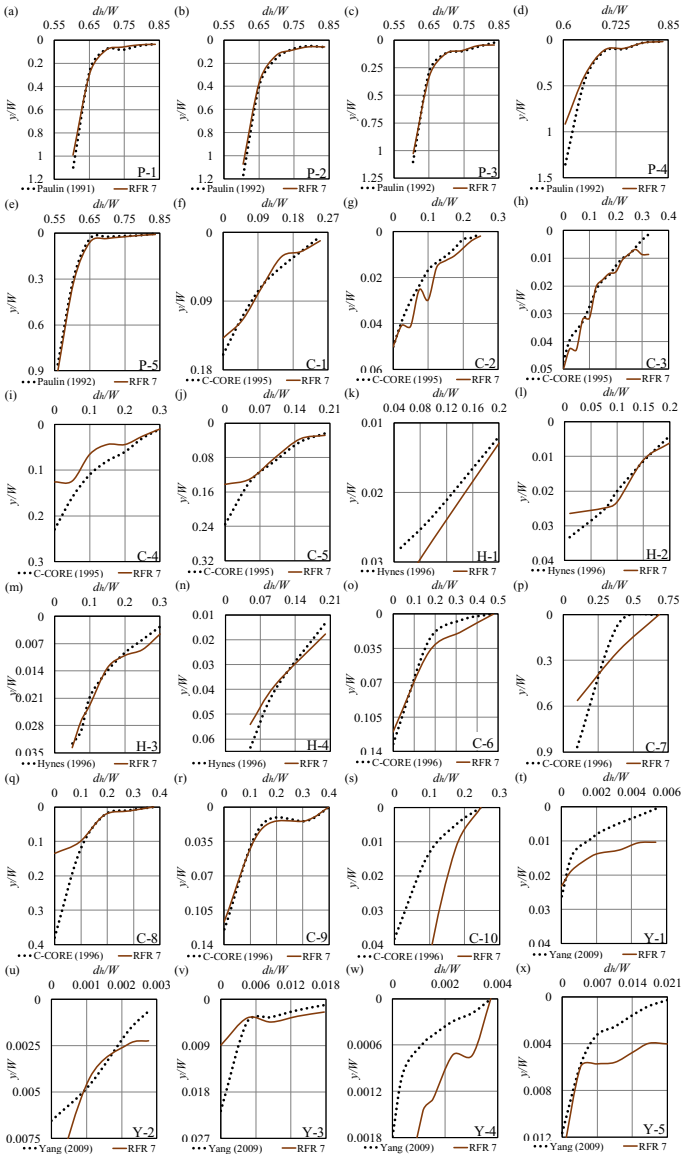


Figure 9-9. Horizontal subgouge deformations profiles simulated by the RFR 7 model

Figure 9-10 draws the comparison between the modeled vertical subgouge deformations through RFR 12 with laboratory values. RFR 7 predicted the vertical deformations employing y/W, Ds/W, α, h'/W, $Lh/\gamma s.W^3$, $Lv/\gamma s.W^3$, and $V^2/g.W$. There was some oscillatory behaviors in laboratory reports (P-2 to P-4), but the RFR 12 model attempted to estimate the vertical deformations with the linear (H-1 to H-4) and also nonlinear patterns (C-1 and C-6).

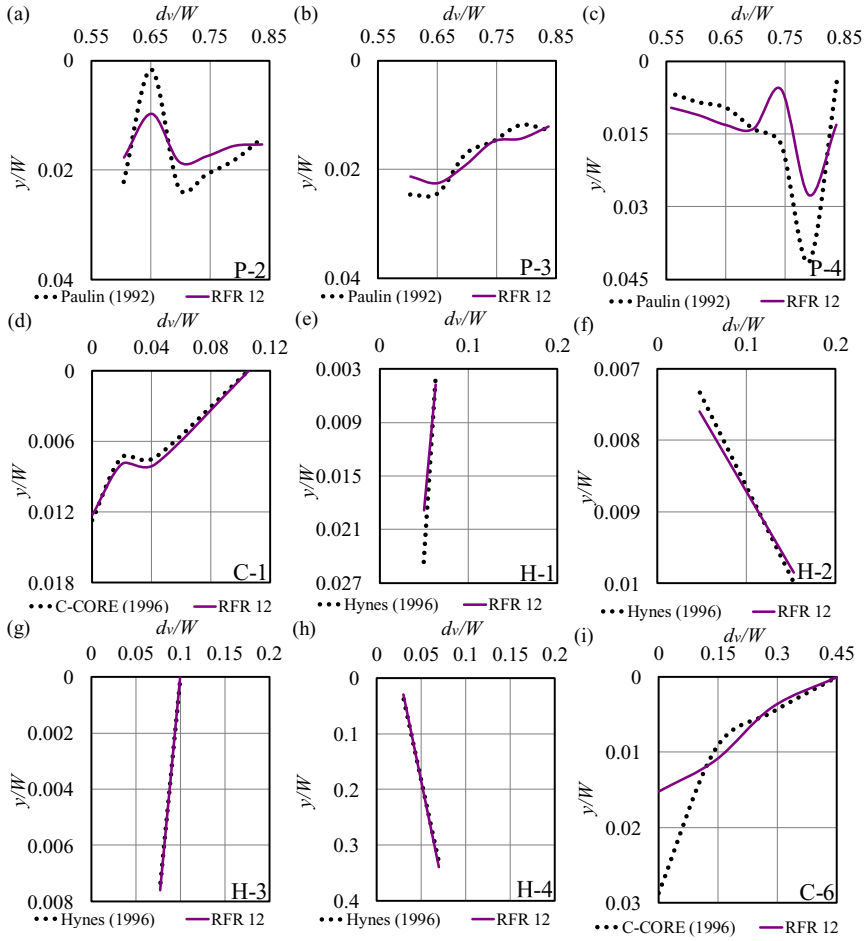


Figure 9-10. Vertical subgouge deformations profiles simulated by the RFR 7 model

Figure 9-11 compares the predicted horizontal reaction forces by RFR 9 with laboratory values. As seen, the RFR 9 well simulated the horizontal reaction forces, with a high degree of precision and correlation alongside a low amount of complexity. The minimum amount of the horizontal

reaction forces took place exactly at the beginning of the keel position and this parameter enhanced along the scour axis. Some discrepancies were reported in the used experimental measurements (C-2, C-4, H-1, Y-2, and Y-3); however, the RFR 9 model attempted to simulate the target parameter with a reasonable performance. RFR 9 was able to model the horizontal reaction forces with a nonlinear pattern, and it overestimated the target values (C-2, H-1, and H-2), whereas the C-4 and Y-3 cases were underestimated by such a model.

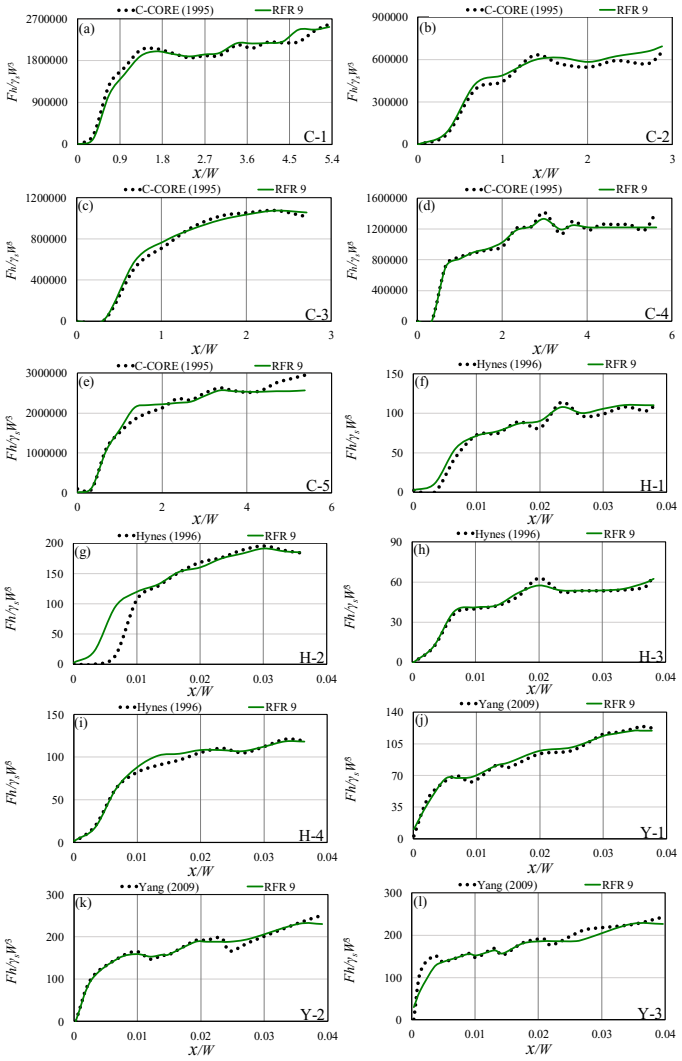


Figure 9-11. Horizontal reaction force profiles simulated by the RFR 9 model

Figure 9-12 compares the modeled vertical reaction forces by RFR 8 with the experimental measurements. The amount of vertical reaction forces was almost zero at the beginning of the scouring axis, but after some fluctuations, this parameter reached its highest value along the ice-gouging axis. The RF 3 model managed to predict the vertical reaction forces utilizing a nonlinear behavior with good efficiency. Although the RFR 8 model had an overestimation performance in the H-1, H-5, and Y-3 cases, this model could well simulate the H-2, H-3, H-4, and Y-2 experimental cases. The RFR algorithm was able to simulate the iceberg drafts and the subgouge soil characteristics in the sandy seabed with reasonable performance, meaning that the superior RFR models had a high level of accuracy and correlation alongside the low level of complexity in order to model the iceberg drafts, subgouge soil deformations, and keel reaction forces simultaneously.

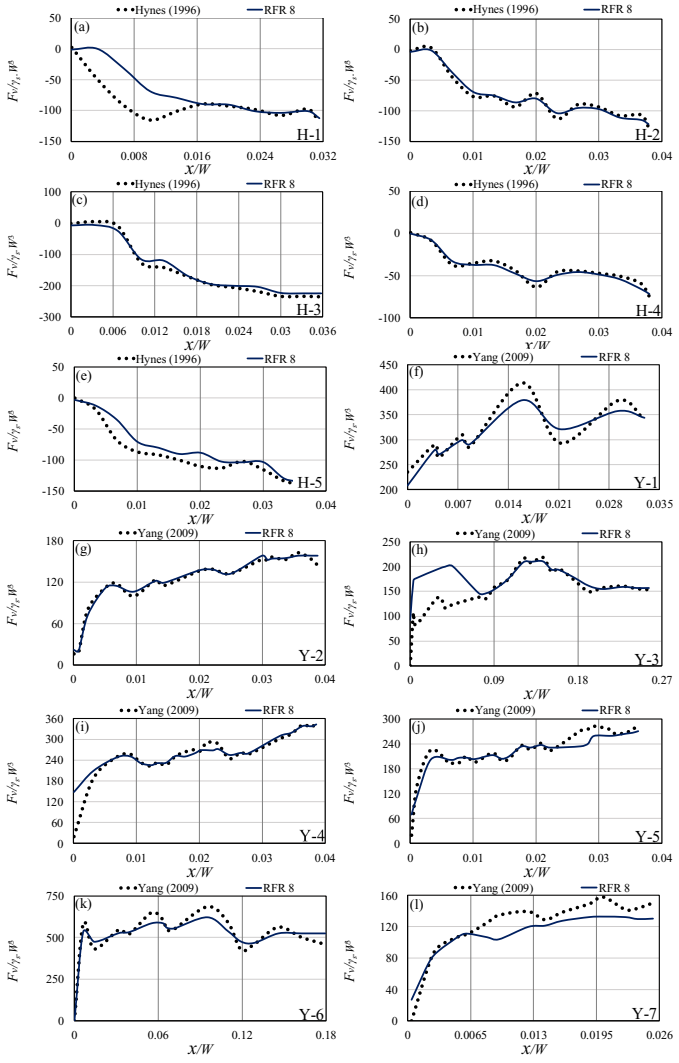


Figure 9-12. Vertical reaction force profiles simulated by the RFR 8 model

9.1.3.3. Comparison with previous works

Different ML methods comprising neural network (NN)-based, tree-based, and ensemble algorithms to estimate iceberg drafts and subgouge soil characteristics were utilized by authors. The NN-based algorithms are a series of models that try to realize underlying connections in specific data using a procedure that simulates how the human brain performs. An NN-based algorithm includes layers of corresponding neurons. Each neuron is defined as a perceptron and is

identical to the multiple linear regression. The perceptrons feed the signal made by a multiple linear regression into an activation function that can be linear or nonlinear. Tree-based algorithms are powerful predictors and valuable in data exploration with high precision, stability, and ease of understanding. Unlike linear algorithms, these models can map nonlinear connections entirely well. They utilize the mean for continuous attributes or mode for definite features to make predictions on training data. Ensemble algorithms are models that integrate the predictions from two or more algorithms, in which several simple models are combined to create one optimum predictive algorithm. Ensembles deliver better outcomes when there is considerable diversity between the models. The ML model used in this study (RFR), as a tree-based algorithm, does not provide an explicit formula for the prediction of ice-seabed interaction parameters and iceberg drafts. Several ML algorithms were used to simulate the iceberg-seabed interaction process, e.g., Azimi and Shiri (2021a), (2021b), (2021c), and Azimi et al. (2022b) modeled the subgouge soil parameters in the sandy seabed, while Azimi and Shiri (2021d), Azimi et al. (2021), Azimi et al. (2022a), (2022c) predicted the ice-induced soil characteristics in the clay seabed. For the horizontal reaction forces simulated by the extreme learning machine (ELM) in Azimi and Shiri (2021a) and (Azimi 2021d) studies, the value of *AIC* was obtained to be -22.094 and -28.275. Additionally, Azimi and Shiri (2021c) and Azimi et al. (2021) employed the self-adaptive extreme learning machine (Sa-ELM) for the horizontal reaction forces in sandy and clay seabed, where the correlation coefficient (*R*) for these studies equaled 0.995 and 0.990, respectively. For the multilayer perceptron neural network (MLPNN) utilized by Azimi and Shiri (2021b) to simulate the horizontal subgouge soil deformations, the *RMSE*, *R*, and *AIC* indices were at 0.344, 0.897, and 4.727. As shown, the RFR model applied in the current study had better performance in terms of accuracy, correlation, and complexity compared with the mentioned studies. The superior RFR

model demonstrated a reasonable generalization ability in the iceberg draft estimation and subgouge soil simulation. The model was trained and validated through a comprehensive dataset, comprising the most important studies in this area. Therefore, the results obtained from this model are sufficiently general to apply to other independent draft prediction scenarios. For future studies, it is recommended to collect more field measurements using the underwater survey facilities to improve the model performance and validate against a wider dataset.

9.1.4. Conclusion

In this study, the iceberg drafts and the subgouge soil features in a sandy seabed were simulated using a robust tree-based machine learning algorithm entitled random forest regression (RFR). To do so, several RFR models, e.g., RFR 1 to RFR 14, were defined using the parameters affecting the iceberg drafts and subgouge soil characteristics. Subsequently, a comprehensive dataset was constructed to train and test the RFR models. Indeed, 70% of the built dataset was applied to train the RFR models, and 30% of the remaining data was used to test these models. The results of the RFR models were evaluated and the most important outcomes are summarized below:

- RFR 3 as a function of L/H , w/H , and S_f could simulate the iceberg drafts with a good performance, e.g., the value of *RMSE* and *R* for this model was reckoned as 1.186 and 0.915.
- RFR 7, RFR 12, RFR 9, and RFR 8 were respectively known as the superior models to simulate the horizontal subgouge deformations, vertical subgouge deformations, horizontal reaction forces, and vertical reaction forces.
- The superior RFR models demonstrated the highest degree of precision, correlation, and low level of complexity in predicting the target parameters, e.g., the correlation

coefficient (R) for the RFR 7, RFR 12, RFR 9, and RFR 8 models equaled 0.978, 0.603, 0.995, and 0.980, respectively.

- The analyses showed that the iceberg length ratio was the most influential input parameter to estimate the iceberg drafts, while the soil depth ratio, berm height ratio, and the position of the iceberg along the scour axis had the highest level of effectiveness so as to simulate the subgouge soil parameters using the RFR algorithm.
- The RFR algorithm was able to predict the profiles of subgouge soil deformation and reaction force with acceptable performance.

The current study gave a good understanding of the application of the tree-based ML algorithm to model the iceberg drafts and subgouge soil characteristics. The obtained results demonstrated that ML models, as cost-effective, fast, and accurate alternatives, can be used in the early stages of iceberg management operations and subsea structure designs to preserve downtime and expenses of the offshore companies operating in the Arctic waters.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)", and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)". Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Arnau Almirall., S, 2017. Ice-gouging in sand and the associated rate effects. Doctoral dissertation, University of Aberdeen, UK.

Azimi, H., Shiri, H., 2020a. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards: 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021b. Modeling subgouge sand deformations by using multi-layer perceptron neural network. In the 31st International Ocean and Polar Engineering Conference, ISOPE-I-21-2150.

Azimi, H., Shiri, H., 2021c. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., 2021d. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04): 411-420.

Azimi, H., Shiri, H., Malta, E.R., 2021. A non-tuned machine learning method to simulate ice-seabed interaction process in clay. Journal of Pipeline Science and Engineering, 1(4): 379-394.

Azimi, H., Shiri, H., Zendehboudi, S., 2022a. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology, 193: 103426.

Azimi, H., Shiri, H., Mahdianpari, M., 2022. Simulation of Subgouge Sand Deformations Using Robust Machine Learning Algorithms. Offshore Technology Conference, OTC-31937-MS.

Azimi, H., Shiri, H., Mahdianpari, M., 2022c. Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. Journal of Pipeline Science and Engineering, 2(4): 100075.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In the 14th International Ocean and Polar Engineering Conference. 2, pp. 899-904

Breiman, L., 2001. Random forests. Machine learning, 45(1): 5-32.

C-CORE, 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour. C-CORE Report 95-Cl2. St. John's, NL, Canada.

C-CORE, 1996. PRISE Phase 3c: Extreme LEE Gouge Event—Modeling and Interpretation. C-CORE Report 96-C32. St. John's, NL, Canada.

El-Tahan, M., El-Tahan, H., Courage, Mitten, D., 1985. Documentation of Iceberg Groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

Hashemi, S., Shiri, H., 2022. The response of layered seabed to ice-gouging: Sand over clay. Ocean Engineering, 266: 113134.

Hotzel, I.S., Miller, J.D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. Arctic Technology Conference, OTC-27358-MS.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. In the 13th International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. In the 14th International Offshore and Polar Engineering Conference, ISOPE-I-04-107.

Løset, S., Carstens, T., 1996. Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Regions Science and Technology, 24 (4): 323-340.

McGuire, P., Younan, A., Wang, Y., Bruce, J., Gandi, M., King, T., Keeping, K., Regular, K., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. In Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, pp. 9-1.

Paulin, M.J., 1991. Preliminary Results of Physical Model Tests of ice Scour. Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering, St. John's, NL, Canada.

Paulin, M.J., 1992. Physical model analysis of iceberg scour in dry and submerged sand. Master thesis, Memorial University of Newfoundland, Canada.

Robe, R.Q., Farmer, L.D., 1976. Physical Properties of Icebergs. Part I. Height to Draft Ratios of Icebergs. Part II. Mass Estimation of Arctic Icebergs. Coast Guard Research and Development Center Groton Conn.

Rudkin, P., 2005. Comprehensive Iceberg Management Database Report 2005 update. PERD/CHC Report 20-72. National Research Council of Canada (NRC) and Panel on Energy Research and Development (PERD).

Sahani, N., Ghosh, T., 2021. GIS-based spatial prediction of recreational trail susceptibility in protected area of Sikkim Himalaya using logistic regression, decision tree and random forest model. Ecological Informatics, 64: 101352.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., King, T., 2021. Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. International Conference on Offshore Mechanics and Arctic Engineering, 85178: V007T07A018.

Sonnichsen, G., Hundert, T., Myers, R., Pocklington, P., 2006. Documentation of Recent Iceberg Grounding Evens and a Comparison with Older Events of Know Age. Environmental Studies Research Funds. Report ESRF Vol. 157.

Turnbull, I.D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Regions Science and Technology, 110: 1-18.

Turnbull, I.D., King, T., Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Woodworth-Lynas, C.M.T., Simms, A., Rendell, C.M., 1985. Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Regions Science and Technology, 10 (2): 163-186.

Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Younan, A., Ralph, F., Ralph, T., Bruce, J., 2016. Overview of the 2012 iceberg profiling program. Arctic Technology Conference, OTC-27469-MS.

Zhou, M., 2017. Underwater iceberg profiling and motion estimation using autonomous underwater vehicles. PhD dissertation, Memorial University of Newfoundland, NL, Canada.

Section 2

Simulation of Iceberg Draft and Subgouge Soil Characteristics Using Extra Tree Regression Algorithm

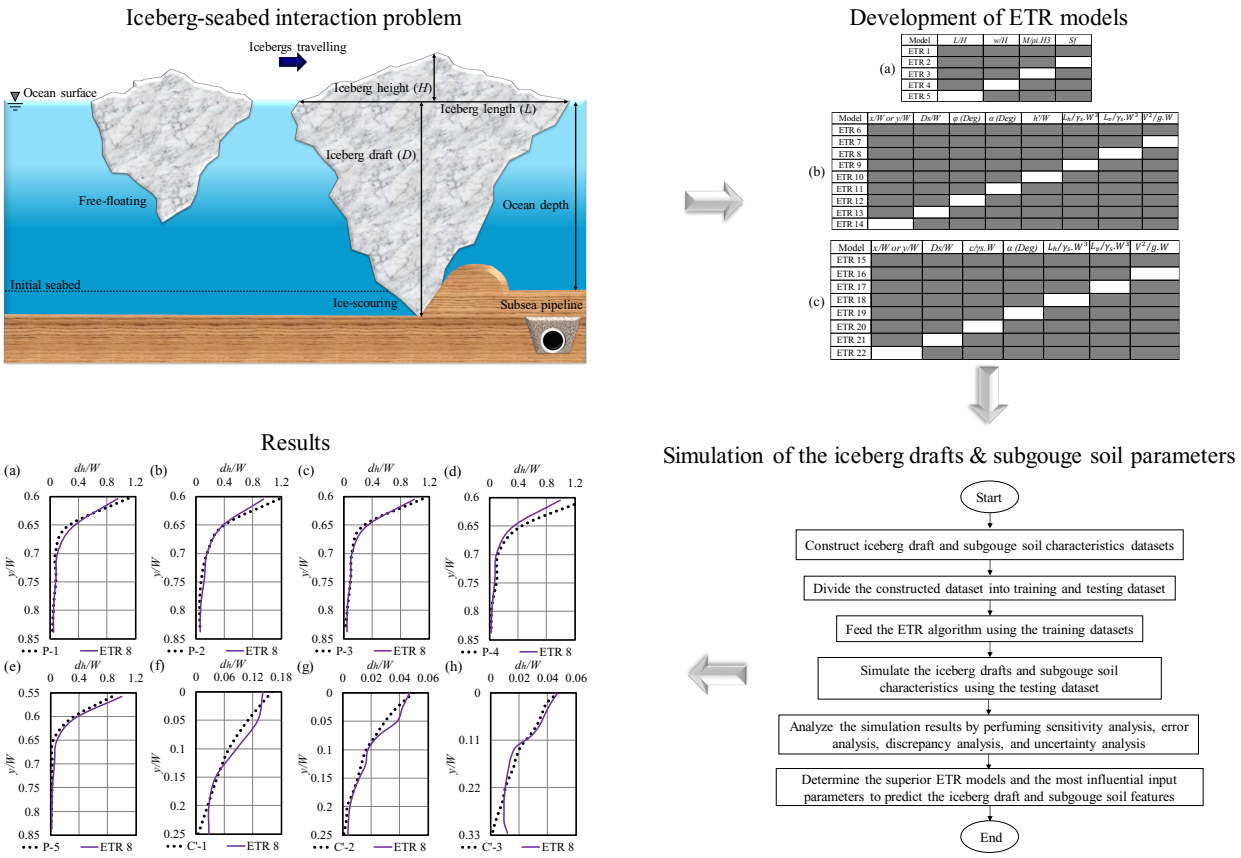
This section was submitted as a journal manuscript

Abstract

Along with this growth in offshore and subsea infrastructures in the Arctic and subarctic areas; however, owing to climate change and global warming, there is increasing concern over the collision risk of traveling icebergs with those structures in the cold waters. Besides using costly and time-consuming underwater survey apparatus and centrifuge tests to surmise the underwater height of icebergs and subgouge soil characteristics, the industry is always pursuing fast and cost-effective alternatives in this realm. Hence, in the present investigation, the extra tree regression (ETR) algorithm was applied to model simultaneously the iceberg drafts and subgouge soil features in sandy and clay seabed for the first time. The ETR algorithm predicted the iceberg draft in the first step and then modeled the subgouge soil reaction forces and deformations. To do so, 22 ETR models were defined utilizing the parameters affecting the iceberg draft estimation and subgouge soil characteristics simulation. The premium ETR models alongside the most influential inputs were recognized by implementing several analyses including sensitivity analysis, error analysis, discrepancy analysis, and uncertainty analysis. The superior ETR model predicted the iceberg draft using all input parameters. The error analysis revealed that roughly 70% of the vertical reaction forces in the sand seabed had an error of less than 16%. The proposed methodology can be applied in the initial phases of iceberg management plans and offshore structure designs.

Keywords: Iceberg draft, Subgouge soil parameters, Extra tree regression (ETR), Offshore structure design, Machine Learning

Graphical abstract



9.1.5. Introduction

Almost 22% of the Earth's undiscovered hydrocarbons are stored in the Arctic region; hence, the Arctic area is one of the best resources to grow oil and gas loading equipment. The Arctic offshore regions with rich wind culture have also a high potential for the expansion of offshore wind farms (Blažauskas et al. 2013). This combined with climate change and global warming means an increased number of ice management operations to protect the subsea pipelines, power cables, and offshore and subsea structures. Every year thousands of icebergs are born out of glaciers in the Arctic zone and carried away by the currents and into the North Atlantic. These icebergs may touch the sea bottom in shallow waters and scratch the seabed, causing “ice-gouging” that may endanger the integrity of subsea pipelines and power cables or even directly collide with offshore structures such as ships, platforms, wind turbines, subsea manifolds, etc. Currently, “ice management,” i.e., iceberg towing and re-routing, is the most reliable approach to protect the subsea and offshore infrastructures, where the threatening icebergs are hooked and towed in a safe direction (Tsvetkova 2020). Ice management is a costly operation and needs standby marine spread with a range of advanced tools and equipment, such as subsea survey facilities, to investigate the iceberg draft and determine if it is a threat to infrastructures. Figure 9-13 illustrates the scheme of the iceberg in free-floating and ice-gouging conditions

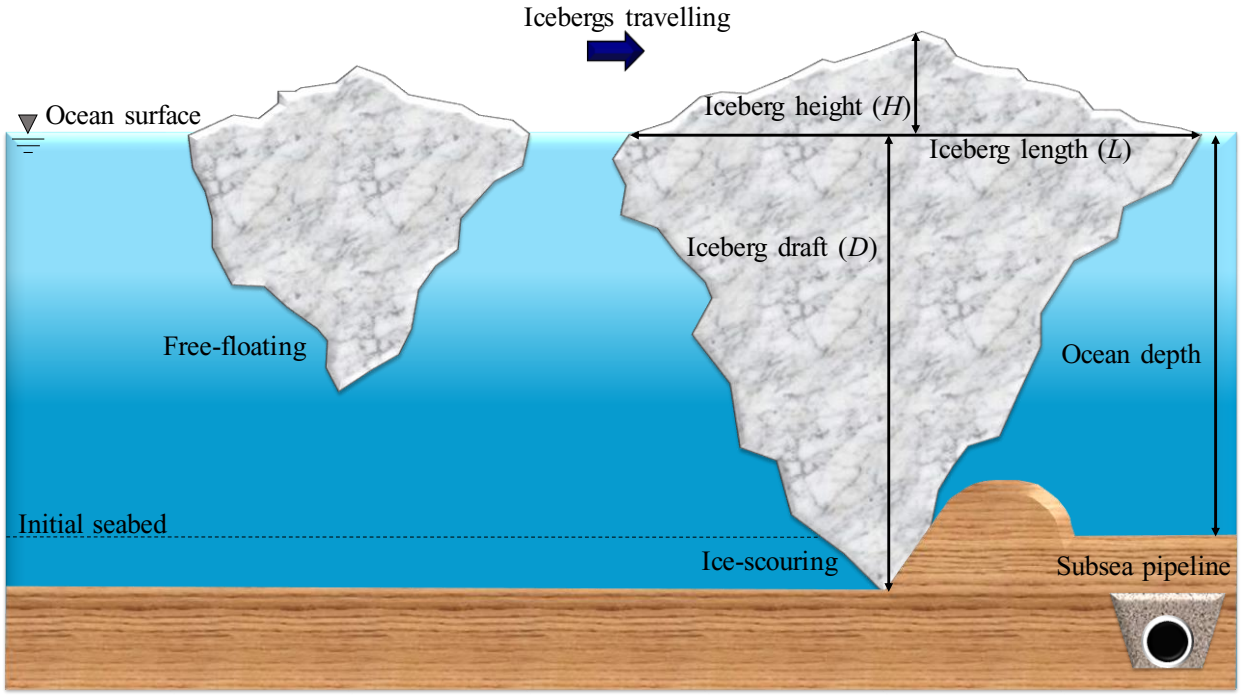


Figure 9-13. Schematic layout of the iceberg in free-floating and ice-gouging conditions

The well-planned iceberg management systems and the guaranteed operational integrity of the sea bottom-funded structures against iceberg attacks in the ice-prone regions demand an adequate iceberg draft estimation, which may lead to a possible reduction in operating costs and downtime. Earlier investigations have focused on predicting the iceberg draft using the iceberg length or iceberg mass. As an example, Allaire (1972), Robe and Farmer (1976), Bass (1980), Hotzel and Miller (1983), and C-CORE (2001) evaluated different aspects of iceberg geometry in the field, analytical, and numerical investigations.

Additionally, Barker et al. (2004) considered the iceberg dimensions including keel and sail. The authors evaluated the cross-sectional areas of the iceberg at diverse sea depth intervals from a certain waterline length. To reckon the draft of the iceberg, some equations as the function of the length were proposed utilizing the regression technique and power curve. Dowdeswell and Bamber

(2007) scrutinized the keel height of voyaging icebergs in the Antarctic Ocean. The scholars assessed the profoundness of the keel via the ice height and surface size. The study stated that a small number of icebergs in the Greenland waters and Antarctica possessed drafts of more significant than 650 m.

Stuckey (2008) estimated the iceberg drift velocity through the probabilistic procedure. He remarked that the above-water extent, below-water size, and iceberg shape factors affected the environmental driving loads. In another study, McKenna and King (2009) predicted the deterioration process of diverse icebergs by viewing the accumulative evolutions in draft, mass, and formation of bergs. The analysis underlined that the draft and length of icebergs were diminished by reducing the iceberg mass. King (2012) simulated the iceberg's features comprising the draft, length, and mass via the Monte Carlo modeling. This author summarized that the evaluation of iceberg drafts deeper than 150 m was fairly restricted. Turnbull et al. (2015) indicated the drift course of the transient icebergs in Northwest Greenland utilizing the hindcast simulation. It was disclosed that the trajectory of bergs might be sensitized to the draft of the iceberg. King et al. (2016) completed a field examination to estimate the iceberg rolling rate. The iceberg drafts were calculated through a calving method, with a computed standard deviation of draft changes from 19% to 34%. The iceberg drafts were correlated with the mass of the icebergs. Turnbull et al. (2018) suggested a relationship in the estimation of drift for traveling icebergs on the Grand Banks of Canada. This equation estimated the iceberg drafts of around 1.3 times more additional than the observed amounts. McKenna et al. (2019) modeled ice-gouging on the Grand Banks of Canada by the Monte Carlo simulation. The iceberg draft changes were employed to reduce the dimension of draft alterations in such a simulation. Stuckey et al. (2021) modeled the 3D iceberg formations using field data. The mass and draft of the iceberg were considered in terms of the

iceberg length regarding the power law assumptions. Most recently, Azimi et al. (2023) determined the dimensionless groups of the variables governing the iceberg draft approximation for the first time. The Authors developed a set of linear regression (LR) equations through the identified dimensionless groups and recognized the most influencing parameters affecting along with the premium LR models. They presented several LR-based models to approximate the iceberg drafts in daily engineering practices. The best LR model had better performance rather than the prior empirical relationships.

In the past three decades, numerous scholars have attempted to investigate the parameters governing subgouge soil characteristics. For example, Paulin (1991), Paulin (1992), Lach (1996), Woodworth-Lynes et al. (1996), C-CORE (1995, 1996), Hynes (1996), Schoonbeek et al. (2006), Been et al. (2008), and Yang (2009) assessed the iceberg-seabed interaction mechanism for a sandy or clay seabed in the laboratory scales. Arnau Almirall (2017) has recently done a couple of 1g experimental studies to measure the sub-gouge sand features in saturated and dry cases. The author examined the influence of velocity modification, scour geometry, and soil characteristics on the subgouge features. The analysis revealed that the ice-scoured sand displacements in the 1g condition are smaller than the outcomes from the centrifuge circumstance.

The application of machine learning (ML) advancement is growing in ubiquity, ranging from engineering to medicine since this technology is fast, precise, and cost-effective.

ML algorithms have been limitedly applied for the simulation of the iceberg-seabed interaction process. Kioka et al. (2003, 2004) utilized a Neural Network (NN) method to model the ice-gouging issue in the sandy seabed. The NN approach was verified by a mechanical method, and the scholars asserted that the NN results possessed a high correlation with the mechanical model. Azimi and Shiri (2020a) ascertained the dimensionless groups impacting the iceberg-seabed

process in both clay and sandy seabed using Buckingham's theory for the first time. They provided a set of LR-based relationships to calculate the maximum subgouge soil deformations. In another research, Azimi and Shiri (2020b) modeled the horizontal subgouge soil deformations in the sandy seabed via gene expression programming (GEP). The dilation index, gouge depth, and attack angle were recognized as the most influential inputs to simulate the target parameter. Azimi and Shiri (2021a), Azimi and Shiri (2021b), Azimi and Shiri (2021c), Azimi et al. (2021), Azimi and Shiri (2021d), Azimi et al. (2022a) applied several neural network-based (NN-based) algorithms such as artificial neural network (ANN), extreme learning machine (ELM), self-adaptive extreme learning machine (Sa-ELM), group method of data handling (GMDH), and generalized structure of group method of data handling (GS-GMDH) to model the subgouge soil characteristics in clay and sand seabed.

To overcome the challenges in the applied NN-based algorithms including overfitting problems, limited data, and hyperparameters tuning, Azimi et al. (2022b) and Azimi et al. (2022c) estimated the iceberg-seabed interaction characteristics in clay and sandy seabed through the tree-based algorithms comprising decision tree regression (DTR), random forest regression (RFR), and gradient boosting regression (GBR) algorithms. They concluded that the ETR model could estimate the objective parameters with a better performance.

The literature revealed that the iceberg draft and the subgouge soil features in clay and sandy seabed have not been simultaneously simulated. Therefore, to fill this knowledge gap, the iceberg drafts along with the subgouge soil parameters, including reaction forces and deformations in both sandy and clay seabed, were modeled using the extra tree (ETR) regression algorithm in the current investigation. The main objectives of the current work are framed as follows:

- Establishing an ETR model to simulate the iceberg draft and subgouge soil features simultaneously.
- Analyzing the simulation results to identify the superior ETR models.
- Determination of the most influential input parameters to predict the iceberg draft and subgouge soil characteristics in clay and sandy seabed.

More details about the applied methodology and simulation results will be provided in the forthcoming sections.

9.1.6. Methodology

Figure 9-14 exhibits the flowchart of the current investigation. In the present research, the iceberg draft and subgouge soil characteristics datasets were constructed and the datasets were divided into the training and testing datasets. The ETR algorithm was fed with the training dataset and the efficiency of the ETR models was validated through the testing dataset. By performing several analyses, the best ETR models alongside the most significant inputs were ascertained.

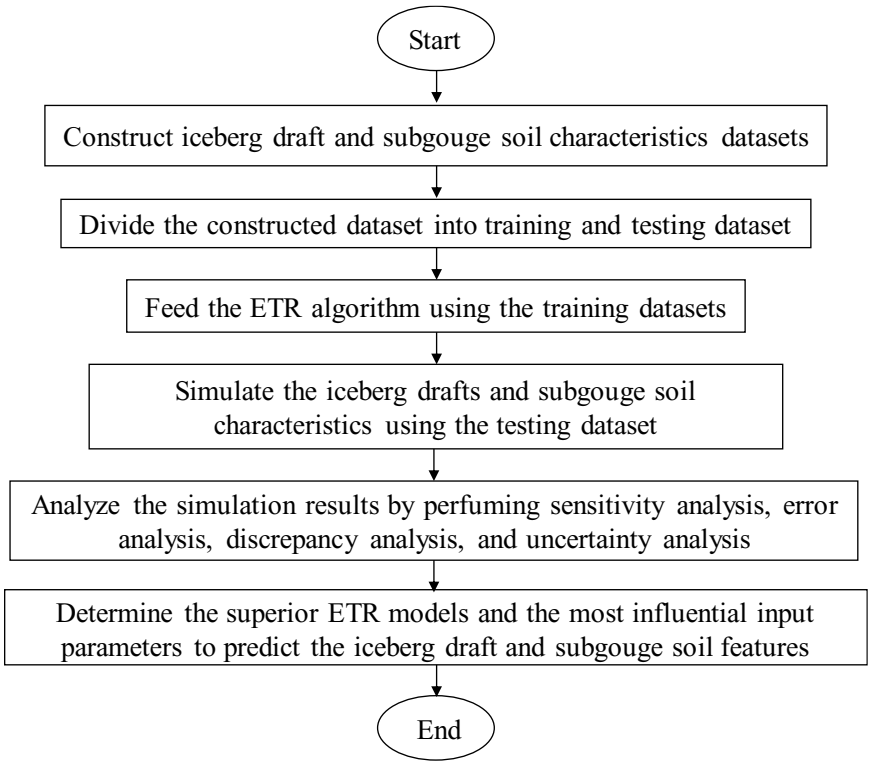


Figure 9-14. Flowchart of the present research

9.1.6.1. Extra Tree Regression (ETR)

The ETR algorithm is a developed version of the RF algorithm originally offered by Geurts et al. (2006). To model a regression issue, the RF algorithm functions by bootstrapping, forming the decision trees, bagging, and splitting the decision tree leaves. The ETR frequently constructs different random regression trees. By choosing the most appropriate subset, the decision-making process is conducted. The ETR and RF models have two primary distinctions: (1) in the ETR method; these leaves are split employing the random selection from the cutting points and (2) the ETR method employs the whole training data to build the trees to minimize the amount of bias. The ETR method (i) utilizes the number of features selected randomly in the leave employing κ variable and (ii) minimizes the sample size for the splitting of leaves operating the n variable. The

κ value determines the characteristic sampling strength; nonetheless, the strength of the averaged outcome noise is described by the ɲ variable. It means that the κ and ɲ variables will decrease the overfitting issue in the ETR methodology and improve the efficiency of the simulation (Hammed et al. 2021). It is worth remarking that the hyperparameters tuning of the ETR algorithm was accommodated through a trial-and-error process in the present task, which is presented in Table 9-1.

Table 9-1. Results of the hyperparameters tuning of the ETR algorithm in the current research

Hyperparameter	Value
min_samples_leaf	1
min_samples_split	8
max_features	1
random_state	*None*
max_depth	*None*
min_impurity_decrease	0
max_leaf_nodes	*None*
min_weight_fraction_leaf	0
ccp_alpha	0

9.1.6.2. Iceberg draft

The iceberg draft (D) was considered in terms of the physical features of the iceberg, such as the iceberg length (L), iceberg height (H), iceberg width (W), and iceberg mass (M). Azimi et al. (2023) introduced the dimensionless groups governing the iceberg draft in the following form:

$$\frac{D}{H} = f_6\left(\frac{L}{H}, \frac{w}{H}, \frac{M}{\rho_i . H^3}, S_f\right) \quad (9\text{-}14)$$

Hence, the iceberg draft ratio (D/H) is a function of the length ratio (L/H), width ratio (w/H), the mass ratio ($M/\rho_i.H^3$), and iceberg shape factor $\left(S_f\right)$. In the present research, the ETR algorithm was fed with the dimensionless groups in equation (22-1) as input parameters to simulate the iceberg draft ratio.

The iceberg-seabed interaction parameters (η) in a soil mass such as soil deformations (d/W) and reaction forces ($F/\gamma s W^3$) are a function of several parameters like the scour depth (Ds), the shear strength parameter of the soil (c), the internal friction angle of soil (φ), the width of gouge (W), the attack angle (α), the angle of the surcharged soil slope (ω), the height of the berm (h'), the horizontal load (Lh), the vertical load (Lv), the velocity of iceberg keel (V), and the specific weight of sand (γ_s) (Lach 1996, Azimi and Shiri 2020a, b):

$$\eta_{(soil)} = f_1(D_s, c, \varphi, W, \alpha, \omega, h', L_h, L_v, V, \gamma_s) \quad (9\text{-}15)$$

Equation (9-15) can be written in the form of Equations (9-16) and (9-17) for the sandy and clay seabed as below (Azimi and Shiri 2020a):

$$\frac{d_{(sand)}}{W}, \frac{F_{(sand)}}{\gamma_s W^3} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \varphi, \alpha, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (9\text{-}16)$$

$$\frac{d_{(clay)}}{W}, \frac{F_{(clay)}}{\gamma_s W^3} = f_3\left(\frac{y}{W}, \frac{D_s}{W}, \frac{c}{\gamma_s . W}, \alpha, \omega, \frac{h'}{W}, \frac{L_h}{\gamma_s W^3}, \frac{L_v}{\gamma_s W^3}, \frac{V^2}{gW}\right) \quad (9\text{-}17)$$

9.1.6.3. Constructed dataset

The observational measurements of various field explorations related to the iceberg draft were utilized to create a comprehensive dataset for the training and testing of the ETR algorithm. The key values of 12 field investigations documented were used (Azimi et al. 2023). The summary of

these field works to construct the iceberg draft dataset is provided in Table 9-2. To model the iceberg drafts utilizing the ETR algorithm, 60% of the applied data was employed to train the ETR model, while 40% of the rest data was allocated to test it.

Table 9-2. Summary of the applied field works in the present research to construct the iceberg draft dataset

Research	Number of case studies
El-Tahan et al. (1985)	38 cases
Woodworth-Lynas et al. (1985)	One case
Løset and Carstens (1996)	52 cases
Barker et al. (2004)	14 cases
McKenna (2004)	Two cases
Sonnichsen et al. (2006)	Nine cases
Turnbull et al. (2015)	Two cases
McGuire et al. (2016)	Eight cases
Younan et al. (2016)	29 cases
Talimi et al. (2016)	One case
Zhou (2017)	Three cases
Turnbull et al. (2018)	Two cases

The experimental values of six laboratory studies recorded by Paulin (1991, 1992) (P-1 to P-5), C-CORE (1995, 1996) (C'-1 to C'-10), Hynes (1996) (H-1 to H-5), and Yang (2009) (Y-1 to Y-7) were employed to verify the ETR algorithm. The amount of the surcharged soil slope (ω) was not

mentioned in the previous works. The internal friction angle of sand (φ), the keel attack angle (α), the gouge depth ratio (D_s/W), and the velocity ratio (V^2/gW) in Paulin's (1991) study (P-1) were reported at 18o, 15o, 0.093, and 0.00054, in turn. However, the berm height ratio (h'/W) was not provided in the P-1 case (Azimi et al. 2022a).

A set of laboratory studies were also applied for the simulation of the iceberg-seabed interaction mechanism in clay seabed. This means that the key measurements of five experimental works documented by C-CORE (1995, 1996) (C-1 to C-12), Lach (1996) (L-1 to L-8), Schoonbeek et al. (2006) (S-1), and Been et al. (2008) (B-1 to B-5) were adopted to validate the ETR algorithm (Azimi et al. 2022b). To simulate the subgouge soil parameters, 70% of the built dataset was for the training of the ETR models and 30% of the remaining data was considered in the testing mode. The applied input combinations to develop different ETR models in this work are depicted in Figure 9-15. It is worth mentioning that the results from the testing phase were analyzed in the present research.

(a)

Model	*L/H*	*w/H*	*M/ρi.H3*	*Sf*
ETR 1	■	■	■	■
ETR 2	■	■	■	
ETR 3	■	■		■
ETR 4	■		■	■
ETR 5		■	■	■

(b)

Model	*x/W or y/W*	*Ds/W*	*φ (Deg)*	*α (Deg)*	*h'/W*	$L_h/\gamma_s.W^3$	$L_v/\gamma_s.W^3$	$V^2/g.W$
ETR 6	■	■	■	■	■	■	■	■
ETR 7	■	■	■	■	■	■	■	
ETR 8	■	■	■	■	■	■		■
ETR 9	■	■	■	■	■		■	■
ETR 10	■	■	■	■		■	■	■
ETR 11	■	■	■		■	■	■	■
ETR 12	■	■		■	■	■	■	■
ETR 13	■		■	■	■	■	■	■
ETR 14		■	■	■	■	■	■	■

(c)

Model	*x/W or y/W*	*Ds/W*	*c/γs.W*	*α (Deg)*	$L_h/\gamma_s.W^3$	$L_v/\gamma_s.W^3$	$V^2/g.W$
ETR 15	■	■	■	■	■	■	■
ETR 16	■	■	■	■	■	■	
ETR 17	■	■	■	■	■		■
ETR 18	■	■	■	■		■	■
ETR 19	■	■	■		■	■	■
ETR 20	■	■		■	■	■	■
ETR 21	■		■	■	■	■	■
ETR 22		■	■	■	■	■	■

Figure 9-15. Input combinations for development of ETR models (a) iceberg draft (b) iceberg-seabed interaction in the sand (c) iceberg-seabed interaction in clay

9.1.6.4. Goodness of fit

Different statistical indices like correlation coefficient (*R*), root mean square error (*RMSE*), mean absolute percentage error (*MAPE*), Willmott Index (*WI*), coefficient of residual mass (*CRM*), and Akaike Information Criteria (*AIC*) were employed to assess the accuracy, correlation, and complexity of the ETR models. The closeness of the *R* and *WI* criteria to one representing the ETR model had a high level of correlation with the observational values. The nearness of the *RMSE*, *MAPE*, and *CRM* criteria to zero indicates the ETR model had the lowest level of impreciseness, while the complexity of the ETR models was not evaluated using the indices introduced. Thus, the Akaike Information Criteria (*AIC*) was considered to cope with this challenge. This means that the less complex ETR model showed the lowest degree of *AIC*, hence the best ETR model could have the lowest amount of *AIC* criterion and error (*RMSE*, *MAPE*, and *CRM*), with the highest level of correlation (*R* and *WI*) (Azimi et al. 2021):

$$R = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \tag{9-18}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{9-19}$$

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{P_i - O_i}{O_i}\right| \tag{9-20}$$

$$WI = 1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(|P_i - \bar{O}| + |O_i - \bar{O}|)^2} \tag{9-21}$$

$$CRM = \frac{\sum_{i=1}^{n} O_i - \sum_{i=1}^{n} P_i}{\sum_{i=1}^{n} O_i} \quad (9\text{-}22)$$

$$AIC = n \times \log\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}\right) + 2k \quad (9\text{-}23)$$

Here, *Oi*, *Pi*, $\bar{O}$, $\bar{P}$, *n*, and *k* are the observational value, the predicted amount, the average observational values, the average predicted amount, the number of observations, and the number of independent variables in the ETR models.

9.1.7. Results and discussion

The ETR models' performance is evaluated through several analyses including sensitivity analysis, error analysis, and uncertainty analysis. Then, the superior ETR models in the estimation of the iceberg drafts and the subgouge soil parameters along wide the most influential inputs are distinguished, and additional analyses are performed for the best ETR models in the end.

9.1.7.1. Sensitivity analysis

The key statistical indices obtained from the developed ETR models are arranged in Table 9-3. Among the ETR models in the iceberg draft estimation, the ETR 1 model as a function of all input parameters comprising L/H , w/H , $M/\rho_i.H^3$, and S_f was known as the model with the highest degree of accuracy, signifying that the RMSE value was achieved as 1.081. The effect of the iceberg shape factor was ignored in ETR 2, where the lowest value of *CRM* was computed for this model (*CRM*=0.087). To predict the iceberg draft using the ETR 3, the mass ratio ($M/\rho_i.H^3$) was an eliminated input and the value of *AIC* for this model was 8.888. The correlation coefficient (*R*) and Willmott Index (*WI*) criteria for the ETR 4 were estimated as 0.923 and 0.926, where the

influence of the iceberg width ratio (w/H) was omitted. The ETR 5 model showed the worst performance to approximate the iceberg draft as the impact of L/H was removed from the ETR 5 model's inputs, with the *RMSE*, *R*, and *AIC* indices of 1.234, 0.887, and 11.837.

Regarding the conducted sensitivity analysis, the iceberg length ratio (L/H) was distinguished as the most influencing input to simulate the iceberg draft using the ETR algorithm; whereas, the iceberg width ratio (w/H), the iceberg mass ratio ($M/\rho_i.H^3$), and iceberg shape factor (S_f) were ranked as the second-effective to the fourth-effective input parameters.

The ETR 6 to ETR 14 models were defined to approximate the subgouge soil features in the sandy seabed. Among the ETR models in the estimation of the horizontal reaction force, the ETR 10 had the highest degree of precision, correlation, and simplicity, meaning that the value of *RMSE*, *R*, and *AIC* for this model was appraised to be 99805.670, 0.994, and 363.941, respectively. The ETR 10 simulated the horizontal reaction forces using x/W , D_s/W , $\varphi, \alpha,$ $L_h/\gamma_s W^3$, $L_v/\gamma_s W^3$, V^2/gW , while the influence of h'/W was removed from this model. The performed sensitivity analysis proved that the position of the iceberg along the scour axis (*x/W*), the attack angle (α), and the shear strength parameter of the sand seabed (φ) were recognized to be the most significant inputs.

The ETR 8 model showed up to be the best ETR model in the estimation of the vertical reaction forces in the sandy seabed when the *AIC* and *WI* values for such model were obtained as 104.364 and 0.985. The ETR 8 model simulated the vertical reaction forces regarding x/W , D_s/W , $\varphi, \alpha, h'/W$, $L_h/\gamma_s W^3$, V^2/gW inputs and the $L_v/\gamma_s W^3$ was dropped from the model's inputs. Additionally, the $x/W, \alpha$, and $L_h/\gamma_s W^3$ were detected to be the most effective input parameters to model the vertical reaction forces. The value of *RMSE* and *AIC* for the superior ETR model to predict the subgouge horizontal deformation in the sand, the ETR 8 model, was

equal to 0.079 and -39.618. These horizontal deformations approximated in terms of y/W , D_s/W , $\varphi, \alpha, h'/W$, $L_h/\gamma_s W^3$, V^2/gW using the ETR 8 model, where the influence of $L_v/\gamma_s W^3$ was taken away from the input combinations. The soil depth ratio (*y/W*) was found as the most important input to simulate the horizontal deformations in the sand, whilst $h'/W, L_h/\gamma_s W^3$, α, D_s/W , $\varphi, V^2/gW$, and $L_v/\gamma_s W^3$ were identified in the second-significant to eighth-significant inputs in terms of effectiveness. The ETR 9 model surmised the vertical deformation in sand using y/W , D_s/W , $\varphi, \alpha, h'/W$, $L_v/\gamma_s W^3$, V^2/gW inputs and the impact of $L_h/\gamma_s W^3$ was disregarded for this model, with the R and WI indices of 0.905 and 0.916. The soil depth ratio (*y/W*) and the berm height ratio (*h'/W*) were the most important input parameters in the vertical deformation estimation in the sand. In order to estimate the horizontal reaction forces in clay seabed, the ETR 18 model as a function of x/W , D_s/W , $c/\gamma_s W, \alpha$, $L_v/\gamma_s W^3$, V^2/gW had the best performance, with *WI* and *AIC* values of 0.995 and 653.553. The effectiveness degree of the iceberg along the scour axis (*x/W*), attack angle (α), and the iceberg dynamic factor ($V^2/g.W$) was much higher than other input parameters to model the horizontal forces in clay. It is worth mentioning that $L_v/\gamma_s W^3, c/\gamma_s W, D_s/W$, $L_h/\gamma_s W^3$ were placed in terms of significance in the fourth to seventh positions. The *AIC* and *RMSE* criteria for ETR 20, as the superior model for the vertical reaction force simulation in clay, were surmised as 225.289 and 472.658. The ETR 20 model was inputted with x/W , D_s/W , α, $L_h/\gamma_s W^3$, $L_v/\gamma_s W^3$, V^2/gW parameters but the clay shear strength parameter ($c/\gamma_s.W$) was a removed input in this model. According to the presented sensitivity analysis, *x/W*, D_s/W, and $L_h/\gamma_s W^3$ possessed the highest degree of effectiveness and other inputs including α, $L_v/\gamma_s W^3$, V^2/gW , and $c/\gamma_s W$ were ranked in the fourth to seventh places, respectively. The ETR 20 model represented the premium model for the estimation of the horizontal deformations

in the clay seabed, with the *WI* and *AIC* value of 0.999 and -56.264. The y/W, $L_h/\gamma_s W^3$, D_s/W, $L_v/\gamma_s W^3$, V^2/gW , α, and $c/\gamma_s W$ were identified as the most important to the less important inputs to forecast the horizontal deformations in clay. ETR 16 was the excellent model for the simulation of the vertical subgouge deformations in clay seabed, with the highest level of correlation, precision, and lowest degree of complexity (*R*=0.983, *RMSE*=0.009, *AIC*=-57.192). The ETR 16 model was fed with y/W , D_s/W , $c/\gamma_s W$, α, $L_h/\gamma_s W^3$, $L_v/\gamma_s W^3$ inputs to approximate the vertical deformations and the iceberg dynamic factor ($V^2/g.W$) was an eliminated factor. The sensitivity analysis exhibited that the soil depth ratio (y/W), the keel attack angle (α), and the gouge depth ratio (Ds/W) were the most remarkable input parameters in the prediction of clay vertical deformations.

Table 9-3. Key statistical indices calculated for the ETR models

Parameter		Model	*R*	*RMSE*	*AIC*	*CRM*	*WI*
Draft		ETR 1	0.920	1.081	10.165	0.101	0.943
		ETR 2	0.906	1.150	9.882	0.087	0.933
		ETR 3	0.917	1.109	8.888	0.108	0.940
		ETR 4	0.923	1.161	10.138	0.107	0.926
		ETR 5	0.887	1.234	11.837	0.102	0.926
Sand-Horizontal	reaction	ETR 6	0.985	154280.556	379.182	-0.030	0.992
		ETR 7	0.983	167992.998	379.770	-0.035	0.991
		ETR 8	0.982	168842.614	379.924	-0.033	0.991
		ETR 9	0.983	165178.002	379.257	-0.035	0.991

	ETR 10	0.994	99805.670	363.941	0.004	0.997
	ETR 11	0.982	171349.321	380.372	-0.040	0.991
	ETR 12	0.982	169286.105	380.003	-0.038	0.991
	ETR 13	0.986	150379.199	376.403	-0.033	0.993
	ETR 14	0.880	439271.280	408.991	-0.058	0.938
Sand-Vertical reaction force	ETR 6	0.970	46.348	107.632	-0.059	0.983
	ETR 7	0.973	44.818	104.830	-0.066	0.984
	ETR 8	0.974	43.953	104.364	-0.079	0.985
	ETR 9	0.972	45.123	104.992	-0.059	0.984
	ETR 10	0.973	44.460	104.638	-0.065	0.984
	ETR 11	0.972	45.189	105.027	-0.062	0.984
	ETR 12	0.972	44.911	104.880	-0.064	0.984
	ETR 13	0.972	45.027	104.941	-0.061	0.984
	ETR 14	0.926	70.017	115.486	-0.044	0.959
Sand-Horizontal deformation	ETR 6	0.739	0.187	-18.256	-0.094	0.849
	ETR 7	0.980	0.073	-39.551	0.180	0.974
	ETR 8	0.984	0.072	-39.618	0.193	0.973
	ETR 9	0.982	0.079	-37.901	0.197	0.968
	ETR 10	0.981	0.079	-37.872	0.211	0.968
	ETR 11	0.981	0.078	-38.078	0.199	0.968
	ETR 12	0.982	0.074	-39.172	0.204	0.972
	ETR 13	0.981	0.074	-39.015	0.179	0.972

	ETR 14	0.278	0.245	-14.719	0.416	0.307
Sand-Vertical deformation	ETR 6	0.828	0.009	-28.827	-0.001	0.913
	ETR 7	0.826	0.009	-30.733	-0.012	0.912
	ETR 8	0.818	0.009	-30.560	0.147	0.879
	ETR 9	0.905	0.011	-29.462	-0.227	0.916
	ETR 10	0.818	0.011	-29.495	-0.313	0.849
	ETR 11	0.828	0.010	-30.027	-0.230	0.897
	ETR 12	0.828	0.009	-30.699	-0.089	0.911
	ETR 13	0.828	0.009	-30.776	0.056	0.912
	ETR 14	0.277	0.016	-25.467	0.329	0.207
Clay-Horizontal reaction force	ETR 15	0.988	57795.569	656.856	0.011	0.994
	ETR 16	0.988	57521.716	656.577	0.004	0.994
	ETR 17	0.988	56914.968	655.956	0.006	0.994
	ETR 18	0.989	54630.154	653.553	-0.003	0.995
	ETR 19	0.988	58744.465	657.811	0.011	0.994
	ETR 20	0.988	56624.436	655.656	0.012	0.994
	ETR 21	0.989	55408.261	654.383	0.010	0.994
	ETR 22	0.794	227462.840	737.183	-0.020	0.899
Clay-Vertical reaction force	ETR 15	0.995	475.567	225.500	-0.005	0.997
	ETR 16	0.996	478.300	225.696	-0.005	0.997
	ETR 17	0.996	478.067	225.680	-0.004	0.997
	ETR 18	0.996	473.733	225.367	-0.006	0.997

	ETR 19	0.996	476.092	225.538	-0.008	0.997
	ETR 20	0.996	472.658	225.289	-0.006	0.997
	ETR 21	0.996	473.307	225.336	-0.006	0.997
	ETR 22	0.951	1512.420	265.194	0.002	0.959
Clay-Horizontal deformation	ETR 15	0.999	0.091	-36.939	0.007	0.999
	ETR 16	0.999	0.043	-53.137	-0.032	0.999
	ETR 17	0.999	0.046	-51.624	-0.043	0.999
	ETR 18	0.999	0.066	-43.745	-0.057	0.999
	ETR 19	0.999	0.038	-55.679	-0.027	0.999
	ETR 20	0.999	0.037	-56.264	-0.036	0.999
	ETR 21	0.999	0.057	-46.891	-0.045	0.999
	ETR 22	0.999	0.529	0.463	0.221	0.973
Clay-Vertical deformation	ETR 15	0.963	0.012	-53.633	0.098	0.962
	ETR 16	0.983	0.009	-57.192	0.035	0.977
	ETR 17	0.974	0.011	-54.854	0.122	0.968
	ETR 18	0.971	0.011	-54.597	0.118	0.967
	ETR 19	0.920	0.015	-49.616	0.137	0.933
	ETR 20	0.980	0.010	-56.382	0.073	0.974
	ETR 21	0.936	0.014	-51.091	0.138	0.946
	ETR 22	0.432	0.033	-37.753	-0.417	0.521

Figure 9-16 depicts the error analysis results for the ETR models to estimate the iceberg drafts and subgouge soil characteristics. Approximately one-third of the iceberg draft simulated by the ETR

1 model showed an error of less than 8%, whilst this amount for the ETR 2, ETR 3, ETR 4, and ETR 5 models equaled almost 26%, 32%, 21, and 29%, respectively. Roughly half of the horizontal reaction forces in the sandy seabed predicted using the ETR 10 model possessed an error of smaller than 12%, whereas about 70% of the vertical reaction forces in the sand modeled by ETR 8 demonstrated an error of less than 16%. Moreover, nearly one-fourth of the horizontal deformations and one-third of the vertical deformations in the sandy seabed estimated by the ETR 8 and ETR 9 models had an error of smaller than 12%. In the clay seabed, just about 85% of the horizontal reaction forces reckoned by the ETR 18 model contained an error of lower than 18% but it was around 95% for the ETR 20 model in the estimation of the vertical reaction forces. Virtually 45% of the horizontal deformations and 37% of the vertical deformations modeled by ETR 20 and ETR 16 provided an error of smaller than 20.

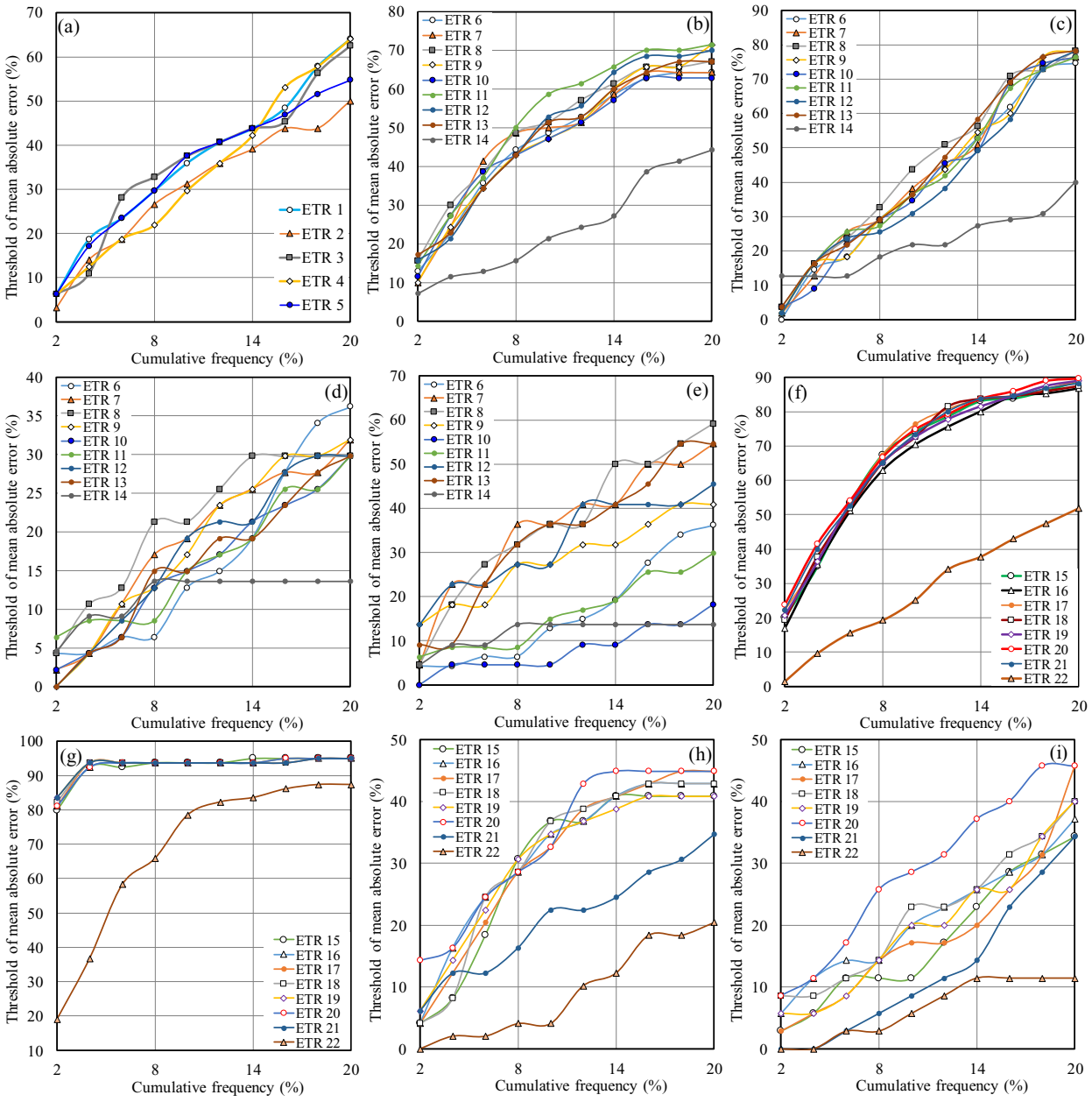


Figure 9-16. Results of error analysis for the ETR models to simulate (a) iceberg draft (b) horizontal reaction force in the sand (c) vertical reaction force in the sand (d) horizontal deformation in the sand (e) vertical deformation in the sand (f) horizontal reaction force in clay (g) vertical reaction force in clay (h) horizontal deformation in clay (i) vertical deformation in clay

9.1.7.2. Uncertainty analysis

The study performed an uncertainty analysis (UA) to further evaluation of the ETR models' performance. In other words, the ETR model's error (e_j) were computed as the difference between the predicted value (P_j) through this model and the observed value (O_j), as follows:

$$e_j = P_j - O_j \tag{9-24}$$

The mean (Mean) and the standard deviation (StDev) of such error values were calculated by the equations below:

$$\text{Mean} = \frac{1}{n}\sum_{j=1}^{n} e_j \tag{9-25}$$

$$\text{StDev} = \sqrt{\sum_{j=1}^{n} \left(e_j - \bar{e}\right)^2/(n-1)} \tag{9-26}$$

An individual ETR model underestimated the subgouge soil parameters or iceberg draft if the sign of the Mean value was negative; however, the positive sign of the Mean meant that the ETR model overestimated the target parameter. Hence, a confidence interval (CI) was produced near the error counted using the Mean, StDev values, and the "Wilson score technic" by omitting the continuity correction. A normal distribution interval corrected as an asymmetric normal distribution, named the Wilson score interval, was employed to adjust the CI bounds. Next, a ±1.96Se yielded a 95%CI. It should be remarked that the width of uncertainty bound (WUB) of each ETR model was achieved below (Azimi et al. 2022c):

$$\text{WUB} = \pm\frac{(\text{Lower bound} - \text{upper bound})}{2} \tag{9-27}$$

In Table 9-4, the results of the performed uncertainty analysis for the ETR models are tabulated. Regarding the conducted uncertainty analysis, all ETR models developed to simulate the iceberg

drafts, e.g., ETR 1 to ETR 5, represented an overestimated performance, where the ETR 1 model had the narrowest bound of uncertainty (WUB=±0.254). Likewise, the horizontal reaction forces in the sandy seabed were overestimated by the ETR 6 to ETR 14 models and the narrowest WUB belonged to the superior model (ETR 10) in this case. The best ETR model for the vertical reaction forces estimation in the sandy seabed had an underestimation performance, with a Mean and WUB of -2.060 and ±8.700. The horizontal deformations in the sandy seabed were overestimated by the ETR 6 to ETR 14 models. To model the horizontal reaction forces in clay, the smallest value of Mean and the narrowest WUB was obtained for ETR 18, with an overestimation performance; however, the best ETR model in the prediction of vertical reaction forces (ETR 20) underestimated the target parameter. The uncertainty analysis revealed that the ETR 20 and ETR 16 models, the premium models, were biased towards underestimation and overestimating the horizontal and vertical subgouge deformations in the clay seabed.

Table 9-4. Results of uncertainty analysis for the ETR models

Parameter		Model	Mean	StDev	95%CI		WUB
Draft		ETR 1	0.39	1.016	0.136	0.644	±0.254
		ETR 2	0.335	1.109	0.059	0.612	±0.277
		ETR 3	0.415	1.037	0.156	0.674	±0.259
		ETR 4	0.411	1.094	0.137	0.684	±0.274
		ETR 5	0.395	1.178	0.1	0.689	±0.295
Sand-	Horizontal	ETR 6	7010.948	94635.970	-16823	30845	±23834
		ETR 7	6948.011	98479.210	-17854	31750	±24802

	ETR 8	8940.125	92871.530	-14449	32330	±23389.500
	ETR 9	7094.299	95202.510	-16882	31071	±23976.500
	ETR 10	12830.91	85894.21	-8801	34436	±21618.500
	ETR 11	5328.068	97348.02	-19189	29845	±24517
	ETR 12	6038.568	97902.03	-18618	30695	±24656.500
	ETR 13	4120.492	98780.39	-20757	28998	±24877.500
	ETR 14	38833.970	238386.700	-21203	98871	±60037
Sand-Vertical reaction force	ETR 6	0.453	33.726	-8.940	9.840	±9.390
	ETR 7	-0.586	32.718	-9.690	8.520	±9.105
	ETR 8	-2.060	31.235	-10.760	6.640	±8.700
	ETR 9	0.195	32.835	-8.950	9.340	±9.145
	ETR 10	-0.472	32.171	-9.430	8.480	±8.955
	ETR 11	-0.167	32.783	-9.290	8.960	±9.125
	ETR 12	-0.278	32.495	-9.330	8.770	±9.050

	ETR 13	0.087	32.607	-8.990	9.160	±9.075
	ETR 14	5.049	56.605	-10.710	20.810	±15.760
Sand-Horizontal deformation	ETR 6	0.019	0.084	-0.007	0.045	±0.026
	ETR 7	0.016	0.074	-0.007	0.040	±0.024
	ETR 8	0.018	0.074	-0.006	0.041	±0.024
	ETR 9	0.018	0.081	-0.007	0.043	±0.025
	ETR 10	0.019	0.081	-0.006	0.045	±0.026
	ETR 11	0.018	0.080	-0.007	0.043	±0.025
	ETR 12	0.018	0.075	-0.005	0.042	±0.024
	ETR 13	0.016	0.077	-0.008	0.040	±0.024
	ETR 14	0.038	0.258	-0.042	0.119	±0.081
Sand-vertical deformation	ETR 6	0.001	0.008	-0.002	0.005	±0.004
	ETR 7	0.001	0.008	-0.003	0.005	±0.004
	ETR 8	0.003	0.008	-0.0003	0.007	±0.004
	ETR 9	-0.004	0.010	-0.009	0.0007	±0.005

	ETR 10	-0.004	0.008	-0.008	-0.0006	±0.004
	ETR 11	-0.003	0.008	-0.006	0.0008	±0.003
	ETR 12	-0.0003	0.008	-0.004	0.003	±0.004
	ETR 13	0.002	0.008	-0.001	0.006	±0.004
	ETR 14	0.006	0.015	-0.001	0.013	±0.007
Clay-Horizontal reaction force	ETR 15	6936.124	53473.540	-2271	16143	±9207
	ETR 16	4131.597	53311.010	-5048	13311	±9179.500
	ETR 17	4786.232	52530.400	-4259	13831	±9045
	ETR 18	1655.058	50082.490	-6968	10278	±8623
	ETR 19	6782.686	54190.590	-2548	16113	±9330.500
	ETR 20	7050.163	52443.450	-1980	16080	±9030

		ETR 21	6556.398	50604.960	-2157	15270	±8713.500
		ETR 22	8330.371	202623.600	-26558	43219	±34888.500
Clay-Vertical reaction force		ETR 15	-71.925	473.101	-177.900	34	±105.950
		ETR 16	-66.530	476.6775	-173.300	42.200	±107.750
		ETR 17	-55.842	477.829	-162.900	51.200	±107.050
		ETR 18	-78.119	470.233	-183.400	27.200	±105.300
		ETR 19	-106.604	466.960	-211.200	-2	±104.600
		ETR 20	-86.388	467.666	-191.100	18.400	±104.750
		ETR 21	-80.068	469.467	-185.200	25.100	±105.150
		ETR 22	35.192	1521.672	-306	376	±341
Clay-	Horizontal	ETR 15	0.007	0.098	-0.023	0.038	±0.031

	ETR 16	-0.007	0.040	-0.019	0.006	±0.012
	ETR 17	-0.012	0.045	-0.026	0.002	±0.014
	ETR 18	-0.017	0.068	-0.039	0.004	±0.022
	ETR 19	-0.006	0.037	-0.017	0.005	±0.011
	ETR 20	-0.009	0.034	-0.019	0.002	±0.011
	ETR 21	-0.011	0.056	-0.028	0.006	±0.017
	ETR 22	0.092	0.570	-0.085	0.270	±0.177
Clay-Vertical deformation	ETR 15	0.003	0.012	-0.001	0.007	±0.004
	ETR 16	0.001	0.009	-0.002	0.004	±0.003
	ETR 17	0.003	0.010	-0.0003	0.007	±0.004
	ETR 18	0.003	0.011	-0.0005	0.007	±0.004

ETR 19	0.004	0.015	-0.001	0.009	±0.005
ETR 20	0.002	0.010	-0.001	0.006	±0.004
ETR 21	0.004	0.013	-0.0009	0.008	±0.004
ETR 22	-0.011	0.032	-0.022	-0.0003	±0.011

9.1.7.3. Superior ETR models

The performed analysis disclosed that ETR 1, ETR 10, ETR 8, ETR 8, ETR 9, ETR 18, ETR 20, ETR 20, and ETR 16 were the excellent models to simulate the iceberg draft, horizontal reaction forces in clay, vertical reaction forces in clay, horizontal deformations in clay, vertical deformations in clay, horizontal reaction forces in the sand, vertical reaction forces in the sand, horizontal deformations in the sand, and vertical deformations in the sand, respectively. To assess the performance of these superior models, a discrepancy analysis was conducted for them. The results of the discrepancy analysis for the ETR models are depicted in Figure 9-17. The efficiency of the ETR models was examined using the discrepancy ratio (*DR*) as below:

$$DR = \frac{P_i}{O_i} \tag{9-28}$$

where, P_i and O_i are the predicted and observed values. The closer the magnitude of *DR* is to the unity, the higher performance shows the ETR model. The value of the maximum ($DR_{(max)}$), minimum ($DR_{(min)}$), and average ($DR_{(ave)}$) discrepancy ratio for these models was computed. For

instance, the $DR_{(max)}$, $DR_{(min)}$, and $DR_{(ave)}$ values for the ETR 1 model equaled 1.778, 0.493, and 0.960. Moreover, the *DR*(ave) criterion for the subgouge soil features in sandy seabed predicted by the ETR 10, ETR 8, ETR 8, and ETR 9 models was acquired to be 1.273, 0.897, 1.463, and 1.206, respectively, and this index for ETR 18, ETR 20, ETR 20, and ETR 16 models to estimate the subgouge soil parameters in clay seabed was at 1.062, 1.078, 1.455, and 1.458.

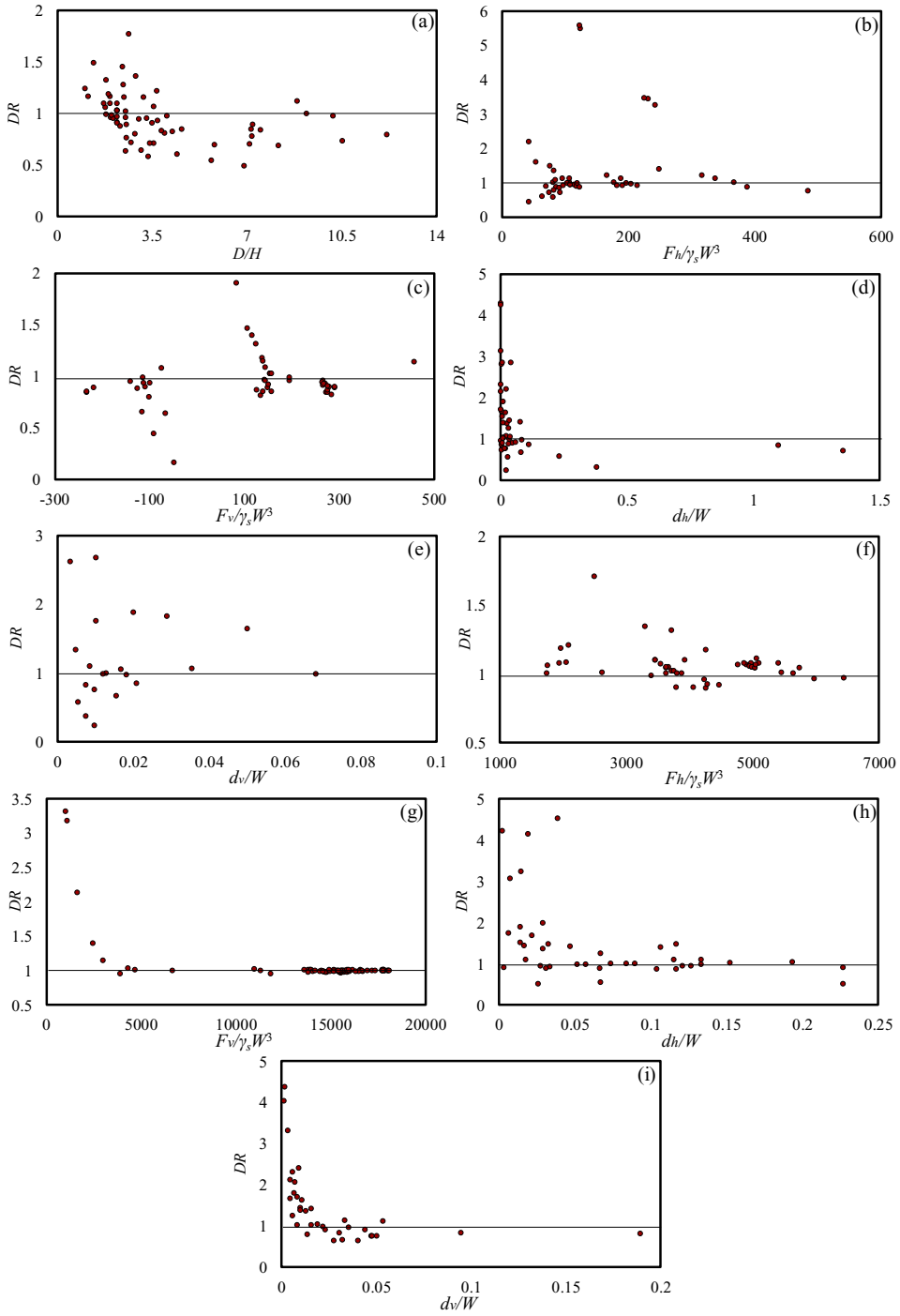


Figure 9-17. Results of discrepancy analysis for the simulation of (a) iceberg draft by ETR 1 (b) horizontal reaction force in the sand by ETR 10 (c) vertical reaction force in the sand by ETR 8 (d) horizontal deformation in sand by ETR 8 (e) vertical deformation in sand by ETR 9 (f)

horizontal reaction force in clay by ETR 18 (g) vertical reaction force in clay by ETR 20 (h) horizontal deformation in clay by ETR 20 (i) vertical deformation in clay by ETR 16

ETR 10 was the premium model in the estimation of the horizontal reaction forces in the sandy seabed. Figure 9-18 compares the horizontal reaction forces predicted by ETR 10 with the laboratory reports. The lowest amount of horizontal reaction force was recorded at the commencement of the scouring process and then it grew along the scour axis. Despite some oscillatory behavior in the laboratory measurements (C'-2, C'-4, and H-2), ETR 10 attempted to model this parameter with its high level of performance. ETR 4 overestimated the horizontal reaction forces utilizing a nonlinear pattern (e.g., C'-3, C'-5, and H-4).

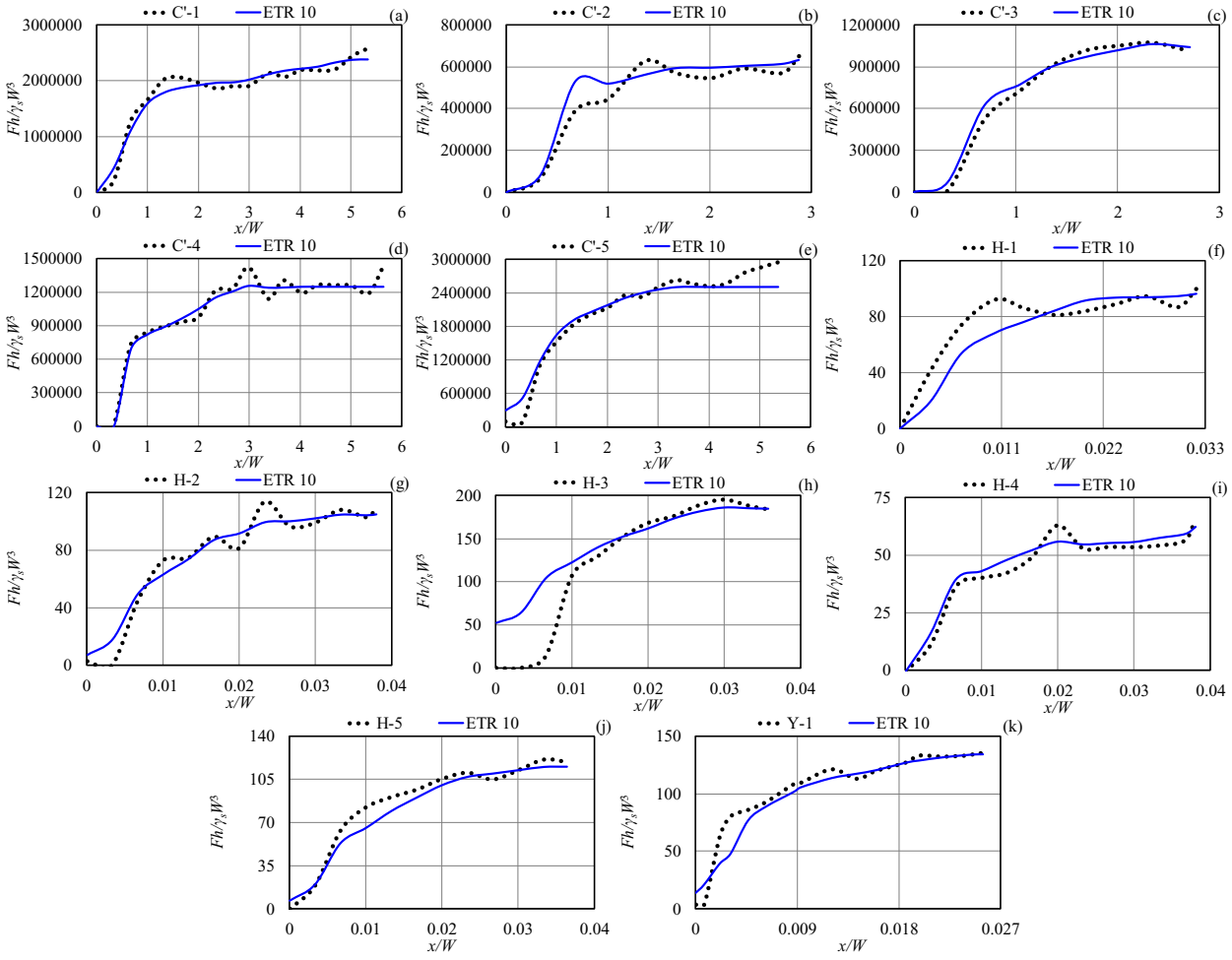


Figure 9-18. Horizontal reaction force profiles in sandy seabed simulated by ETR 10 (a-e) C'-1 to C'-5 (f-j) H-1 to H-5 (k) Y-1

Figure 9-19 compares the vertical reaction forces estimated by ETR 8 and the laboratory values. The simulation results showed that these reaction forces were almost zero at the starting point of the ice-gouging process and the amount of this parameter enhanced along the scouring axis. There were several fluctuations in the experimental value but ETR 8 could predict the vertical reaction forces utilizing a nonlinear pattern with acceptable performance.

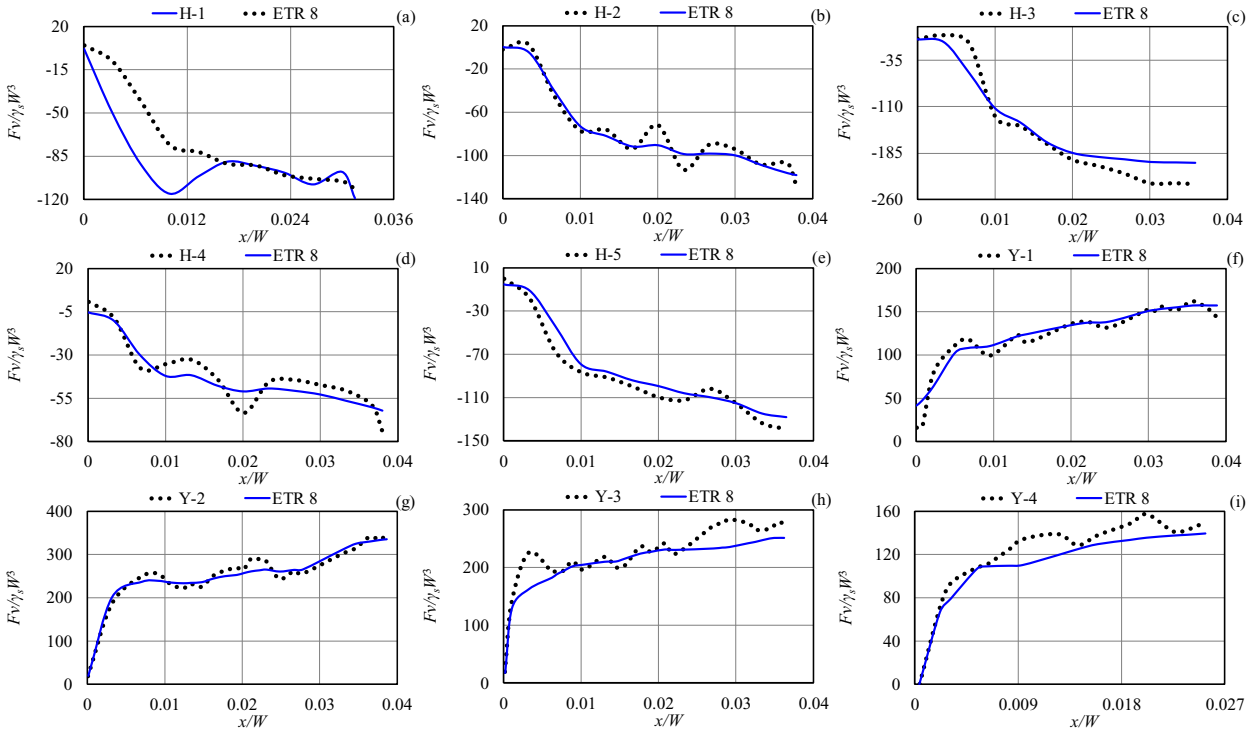


Figure 9-19. Vertical reaction force profiles in sandy seabed simulated by ETR 8 (a-e) H-1 to H-5 (f-i) Y-1 to H-4

Figure 9-20 depicts the horizontal deformation profiles in sandy seabed simulated by ETR 8. The maximum horizontal deformation was observed solely underneath the iceberg keel and it decreased in deeper soil depth. The ETR 8 model simulated the horizontal deformations as an exponential distribution through the soil depth beneath the iceberg contact point and the shear resistance in the seabed soil extended the soil displacement much deeper than the iceberg tip (P-1, P-2, P-3, P-4, P-5, H-3, and Y-1). The ML model overestimated the target parameter with the nonlinear trends (C'-5, H-5, and Y-1).

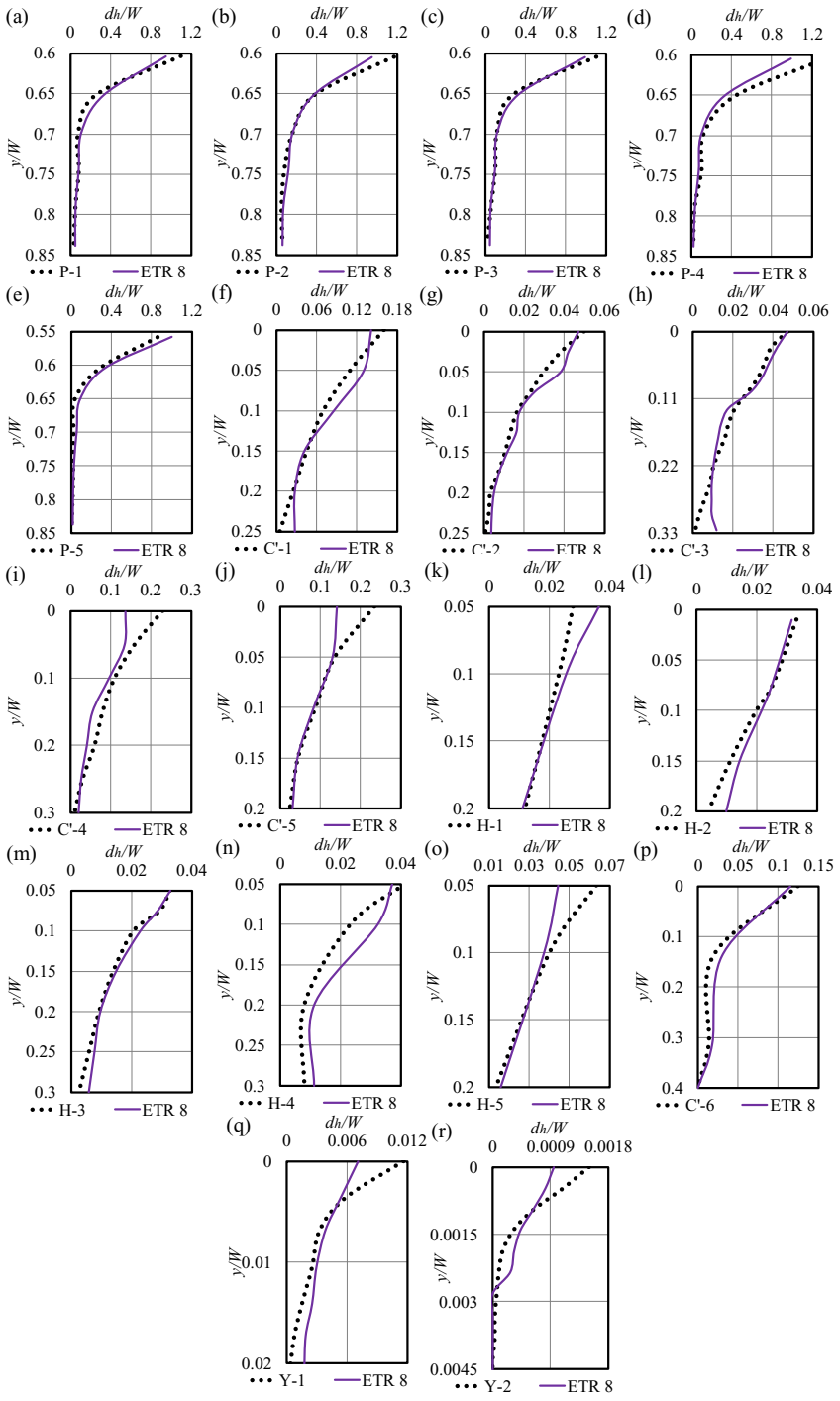


Figure 9-20. Horizontal deformation profiles in sandy seabed simulated by ETR 8 (a-e) P-1 to P-5 (f-j) C'-1 to C'-5 (k-o) H-1 to H-5 (p) C'-6 (q-r) Y-1 to Y 2

The vertical deformation profiles in sandy seabed simulated by ETR 9 are demonstrated in Figure 9-21. This ML model simulated the vertical deformation through both linear (H-2) and nonlinear (P-1, P-3, C'-1, and C'-2) patterns although many fluctuations have been seen in the experimental reports (P-2, P-3, and H-1). A few discrepancies between the observational values and simulation results were documented (P-1, P-2, P-4, C'-1, and C'-2) but ETR 9 tried to model the vertical deformations with the best performance.

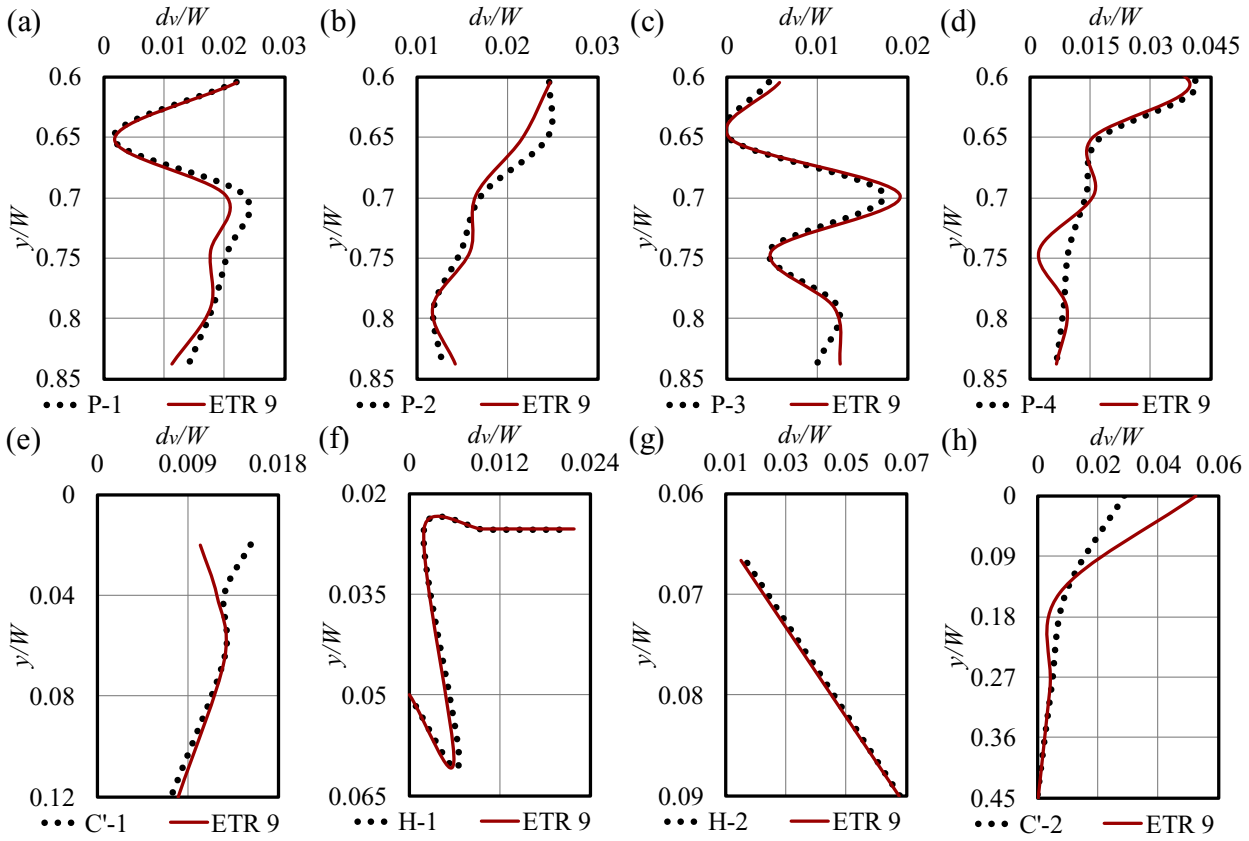


Figure 9-21. Vertical deformation profiles in sandy seabed simulated by ETR 9 (a-d) P-1 to P-4 (e) C'-1 (f-g) H-1 to H-2 (h) C'-2

Figure 9-22 compares the horizontal reaction forces in clay seabed modeled by ETR 18 with laboratory values. Similar to the horizontal reaction forces in the sand, the lowest amount of horizontal reaction force in clay seabed was predicted at the primary location of the scouring mechanism (x=0.0), where this parameter was augmented along the scouring axis. Even though some oscillatory behavior were in the experimental reports (L-1, L-2, and L-7), ETR 18 simulated the horizontal reaction forces in clay seabed with the maximum precision, correlation, and simplicity (C-5 and L-4).

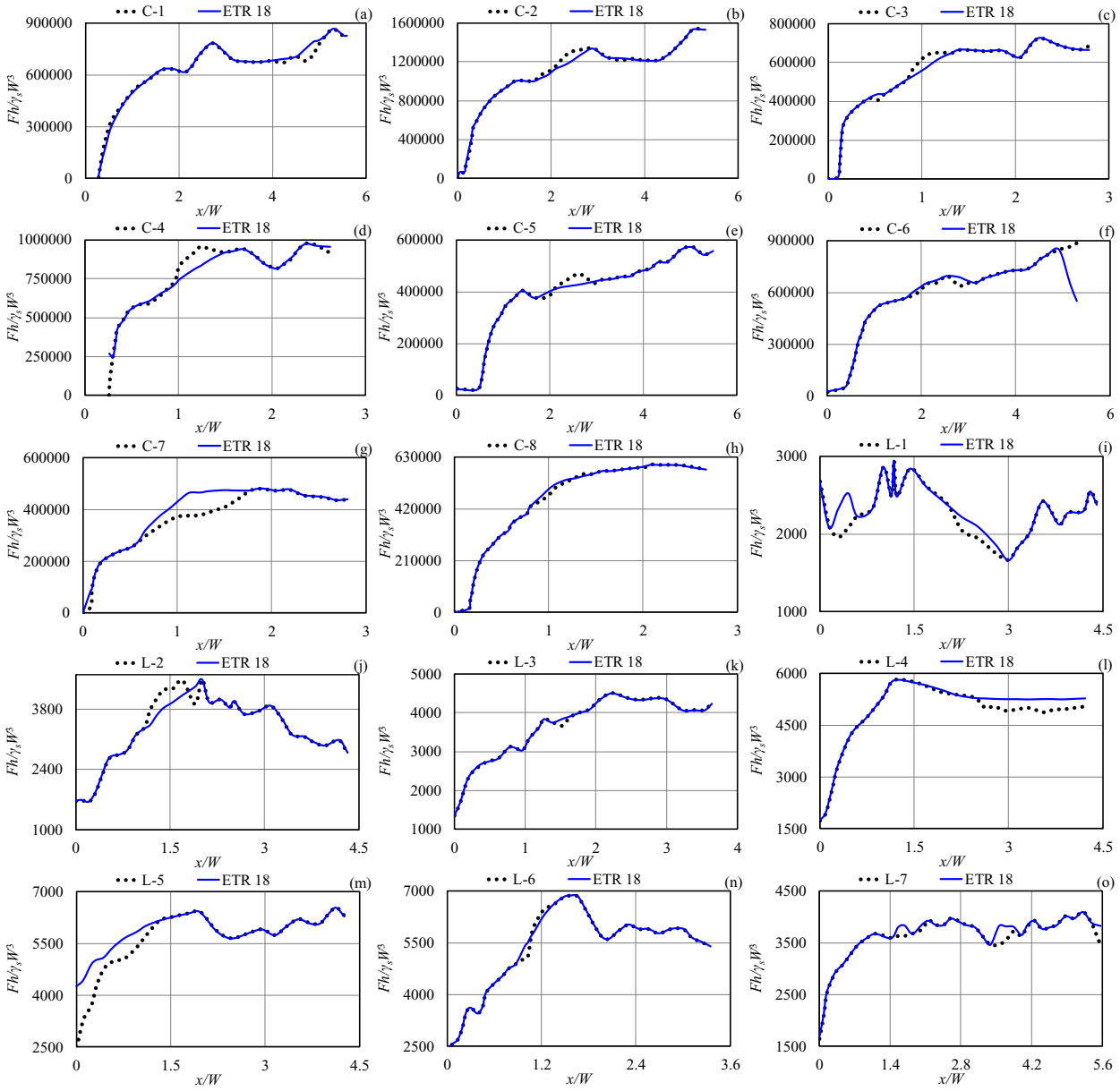


Figure 9-22. Horizontal reaction force profiles in clay seabed simulated by ETR 18 (a-h) C-1 to C-8 (i-o) L-1 to L-7

The vertical reaction force profiles in clay seabed simulated by ETR 20 are compared with the experimental measurements in Figure 9-23. The lowest value of vertical reaction forces was predicted in the primary location of the scouring process and this parameter increased to the highest

amount by developing along the scouring axis. The ETR 20 model underestimated the objective function using a nonlinear pattern (L-3 and L-4). This ML model strived to approximate the vertical reaction forces with its best performance despite several alterations observed in the laboratory reports.

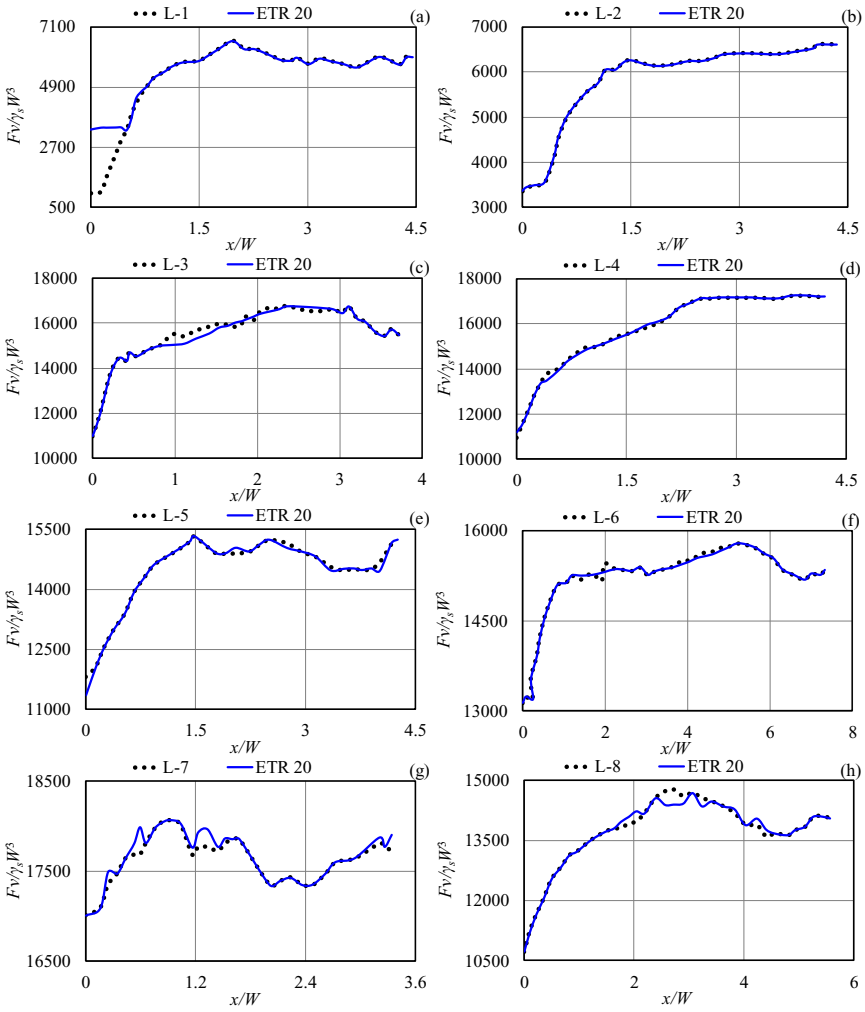


Figure 9-23. Vertical force profiles in clay seabed simulated by ETR 20 (a-h) L-1 to L-8

Figure 9-24 presents the comparison between the test results and the horizontal deformation in clay seabed modeled by ETR 20. The obtained results signified that the maximum horizontal deformations were observed in the interaction plane of the clay seabed and iceberg bottom, where

it reduced at greater soil depth. The ETR 20 model demonstrated excellent performance in the prediction of the horizontal displacements (C-1, C-2, C-3, C-7, L-4, L-5, and L-7), rather a small number of differences were seen during the simulation process (C-4, L-2, C-9, C-10, C-11, and C-12).

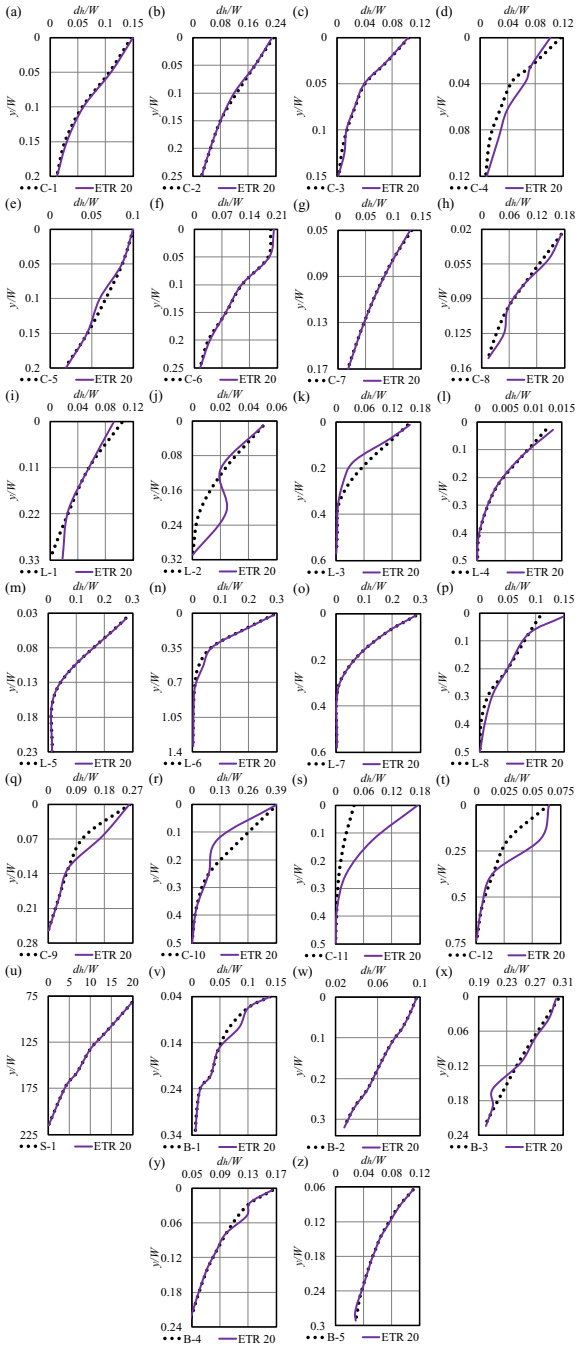


Figure 9-24. Horizontal deformation profiles in clay seabed simulated by ETR 20 (a-h) C-1 to C-8 (i-p) L-1 to L8 (q-t) C-9 to C-12 (u) S-1 (v-z) B-1 to B-5

The comparison between the vertical ice-scoured displacements in the clay seabed simulated by the ETR 16 model and the test results is depicted in Figure 9-25. A high correlation was recorded between the experimental values and the modeling outcomes (C-1, C-2, C-3, C-7, L-1, L-3, and C-8), rather some discrepancies took place in C-4, C-6, L-4, L-5, L-7, L-8, and C-9, and B-1 cases. The ETR 16 model forecasted the C-1, C-2, C-3, C-7, L-1, L-2, L-3, L-6, C-8, C-9, C-10, and C-11 using the nonlinear trend, while the C-5 and C-6 cases were modeled by the linear behavior.

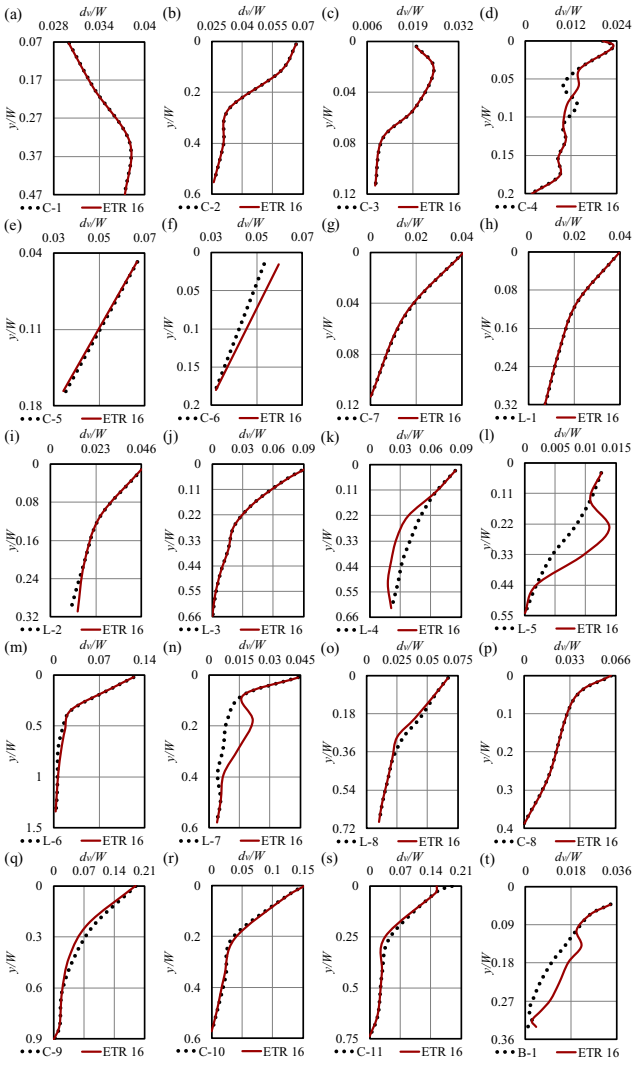


Figure 9-25. Vertical deformation profiles in clay seabed simulated by ETR 16 (a-g) C-1 to C-7 (h-o) L-1 to L-8 (p-s) C-8 to C-11 (t) B-1

The ETR algorithm modeled the iceberg draft with acceptable performance and the subgouge soil parameters in the sandy and clay seabed including the reaction forces and deformations predicted by this method when the iceberg tip collided with the ocean floor. The best ETR models alongside the most significant input parameters were identified. The best ETR models had a good capability

to estimate the objective function appropriately, with the highest degree of accuracy and correlation as well as the lowest level of complexity.

9.1.8. Conclusion

The development of an ETR model to simulate the iceberg draft and subgouge soil features at the same time was the major goal of the present investigation. The study modeled simultaneously the iceberg draft and subgouge soil parameters in sand and clay seabed using the extra tree regression (ETR) algorithm for the first time. Using the parameters affecting the iceberg draft modeling and subgouge soil features estimation, 22 ETR models were defined. Two comprehensive datasets were built to train and test the ETR models. By conducting different analyses, the best ETR models and the most significant inputs were known. The most important outcomes of this investigation are summarized as follows:

- ETR 1 predicted the iceberg drafts utilizing all inputs and demonstrated the best performance in the iceberg draft estimation.
- The iceberg length ratio (*L/H*) was the most influential input to simulate the iceberg draft through the ETR algorithm.
- The highest degree of accuracy, correlation, and simplicity was obtained for the best ETR models to predict the subgouge soil parameters in both clay and sandy seabed.
- The best ETR model was biased toward underestimation with the narrowest bound of uncertainty to approximate the horizontal deformations in the clay seabed.
- The performed error analysis showed roughly one-third of the vertical displacements in clay seabed simulated by the superior ETR had an error of less than 16%.

The present outcomes, which were the first application of the ML technology to model simultaneously the iceberg drafts and subgouge soil parameters, revealed that the ML advancement can pave to develop of more cost-effective and quick solutions to protect offshore structures against iceberg attacks, especially in the primary steps of the iceberg management systems and offshore infrastructures designs. The achieved results in the present research were promising as the first step toward the development of an alternative methodology to predict the iceberg drafts and subgouge soil characteristics at the same time. However, there were several challenges in the current investigation, e.g., the need for more field and experimental data to enhance the performance and versatility of the ML algorithms and the ETR algorithm was not able to give an explicit model for the estimation of the objective parameters in the daily engineering practices. Hence, these challenges can be tackled in upcoming studies.

Acknowledgment

The authors gratefully acknowledge the financial support of "Wood Group," which established a Research Chair program in the Arctic and Harsh Environment Engineering at the Memorial University of Newfoundland, the "Natural Science and Engineering Research Council of Canada (NSERC)," and the "Newfoundland Research and Development Corporation (RDC) (now TCII)" through "Collaborative Research and Developments Grants (CRD)." Special thanks are extended to Memorial University for providing excellent resources to conduct this research.

References

Allaire, P.E., 1972. Stability of simply shaped icebergs. Journal of Canadian Petroleum Technology, 11(04).

Arnau Almirall, S., 2017. Ice-gouging in sand and the associated rate effects. Doctoral dissertation, University of Aberdeen.

Azimi, H., Mahdianpari, M., Shiri, H., 2023. Determination of Parameters Affecting the Estimation of Iceberg Draft. China Ocean Engineering, 37(1): 62-72.

Azimi, H., Shiri, H., 2020a. Dimensionless groups of parameters governing the ice-seabed interaction process. Journal of Offshore Mechanics and Arctic Engineering, 142(5): 051601.

Azimi, H., Shiri, H., 2020b. Ice-Seabed interaction analysis in sand using a gene expression programming-based approach. Applied Ocean Research, 98: 102120.

Azimi, H., Shiri, H., 2021a. Sensitivity analysis of parameters influencing the ice–seabed interaction in sand by using extreme learning machine. Natural Hazards, 106(3): 2307-2335.

Azimi, H., Shiri, H., 2021b. Modeling subgouge sand deformations by using multi-layer perceptron neural network. In the 31st International Ocean and Polar Engineering Conference, ISOPE-I-21-2150.

Azimi, H., Shiri, H., 2021c. Evaluation of ice-seabed interaction mechanism in sand by using self-adaptive evolutionary extreme learning machine. Ocean Engineering, 239: 109795.

Azimi, H., Shiri, H., Malta, E.R., 2021. A non-tuned machine learning method to simulate ice-seabed interaction process in clay. Journal of Pipeline Science and Engineering, 1(4): 379-394.

Azimi, H., Shiri, H., 2021d. Assessment of ice-seabed interaction process in clay using extreme learning machine. International Journal of Offshore and Polar Engineering, 31(04): 411-420.

Azimi, H., Shiri, H., Zendehboudi, S., 2022a. Ice-seabed interaction modeling in clay by using evolutionary design of generalized group method of data handling. Cold Regions Science and Technology, 193: 103426.

Azimi, H., Shiri, H., Mahdianpari, M., 2022b. Simulation of Subgouge Sand Deformations Using Robust Machine Learning Algorithms. Offshore Technology Conference, OTC-31937-MS.

Azimi, H., Shiri, H., Mahdianpari, M., 2022c. Iceberg-seabed interaction evaluation in clay seabed using tree-based machine learning algorithms. Journal of Pipeline Science and Engineering, 2(4): 100075.

Barker, A., Sayed, M., Carrieres, T., 2004. Determination of iceberg draft, mass and cross-sectional areas. In The fourteenth international offshore and polar engineering conference. ISOPE-I-04-116.

Bass, D.W., 1980. Stability of icebergs. Annals of Glaciology, 1: 43-47.

Blažauskas, N., Włodarski, M., Paulauskas, S., 2013. Perspectives for offshore wind energy development in the South-East Baltics. Klaip eda University: Klaipeda, Lithuania, pp. 1-54.

Been, K., Sancio, R.B., Ahrabian, D., van Kesteren, W., Croasdale, K., Palmer, A. 2008. Subscour displacement in clays from physical model tests. In International Pipeline Conference, 48609: 239-245.

C-CORE, 1995. Phase 3: Centrifuge Modelling of Ice Keel Scour, C-CORE Report 95-Cl2.

C-CORE, 1996. PRISE Phase 3c: Extreme LEE Gouge Event-Modeling and Interpretation, C-CORE Report 96-C32.

C-CORE, 2001. Documentation of iceberg grounding events from the 2000 season. C-CORE publication 01-C10 (Revision 0). Report submitted to the Geological Survey of Canada, Atlantic.

Dowdeswell, J.A., Bamber, J.L., 2007. Keel depths of modern Antarctic icebergs and implications for sea-floor scouring in the geological record. Marine Geology, 243(1-4): 120-131.

El-Tahan, M., El-Tahan, H., Courage, D., Mitten, P., 1985. Documentation of Iceberg Groundings. Environmental Studies Research Funds. Report ESRF Vol. 7.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine learning, 63(1): 3-42.

Hammed, M.M., AlOmar, M.K., Khaleel, F., Al-Ansari, N., 2021. An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. Mathematical Problems in Engineering.

Hotzel, I.S., Miller, J.D., 1983. Icebergs: their physical dimensions and the presentation and application of measured data. Annals of Glaciology, 4: 116-123.

Hynes, F., 1996. Centrifuge Modelling of Ice Scour in Sand, Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

King, T., 2012. Iceberg Interaction Frequency Model for Subsea Structures. Arctic Technology Conference, OTC-23787-MS.

King, T., Younan, A., Richard, M., Bruce, J., Fuglem, M., Phillips, R., 2016. Subsea risk update using high resolution iceberg profiles. In Arctic Technology Conference, OTC-27358-MS.

Kioka, S.D., Kubouchi, A., Saeki, H., 2003. Training and Generalization of Experimental Values of Ice Scour Event by a Neural-Network. In the 13th International Offshore and Polar Engineering Conference, ISOPE-I-03-081.

Kioka, S., Kubouchi, A., Ishikawa, R., Saeki, H., 2004. Application of the mechanical model for ice scour to a field site and simulation method of scour depths. In the 14th International Offshore and Polar Engineering Conference, ISOPE-I-04-107.

Lach, P.R., 1996. Centrifuge Modelling of Large Soil Deformation due to Ice Scour. Doctoral dissertation, Memorial University of Newfoundland. St. John's. NL. Canada.

Løset, S., Carstens, T., 1996. Sea ice and iceberg observations in the western Barents Sea in 1987. Cold Regions Science and Technology, 24(4): 323-340.

McGuire, P., A. Younan, Y. Wang, J. Bruce, M. Gandi, T. King, K. Regular., 2016. Smart iceberg management system–rapid iceberg profiling system. Arctic Technology Conference, OTC-27473-MS.

McKenna, R., 2004. Development of iceberg shape characterization for risk to Grand Banks installations. PERD/CHC Report, 20473.

McKenna, R., King, T., 2009. Modelling Iceberg Shape, Mass and Draft Changes. In Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions, No. POAC09-107.

McKenna, R., King, T., Crocker, G., Bruneau, S., German, P., 2019. Modelling iceberg grounding on the grand banks. In Proceedings of the International Conference on Port and Ocean Engineering under Arctic Conditions, pp. 9-19.

Paulin, M.J., 1991. Preliminary Results of Physical Model Tests of ice Scour, Memorial University of Newfoundland, Centre for Cold Ocean Resources Engineering, St. John's, NL, Canada.

Paulin, M.J., 1992. Physical model analysis of iceberg scour in dry and submerged sand. Master thesis, Memorial University of Newfoundland, St. John's, NL, Canada.

Robe, R.Q., Farmer, L.D., 1976. Physical Properties of Icebergs. Part I. Height to Draft Ratios of Icebergs. Part II. Mass Estimation of Arctic Icebergs. Coast Guard Research and Development Center Groton Conn.

Schoonbeek, I.S., van Kesteren, W.G., Xin, M.X., Been, K., 2006. Slip Line Field Solutions as an Approach to Understand Ice Subgouge Deformation Patterns. In the 16th International Offshore and Polar Engineering Conference, ISOPE-I-06-289.

Stuckey, P.D, 2008. Drift speed distribution of icebergs on the grand banks and influence on design loads. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Sonnichsen, G., T. Hundert, R. Myers, P. Pocklington., 2006. Documentation of Recent Iceberg Grounding Evens and a Comparison with Older Events of Know Age. Environmental Studies Research Funds. Report ESRF Vol. 157.

Stuckey, P., Fuglem, M., Younan, A., Shayanfar, H., Huang, Y., Liu, L., King, T., 2021. Iceberg Load Software Update Using 2019 Iceberg Profile Dataset. In International Conference on Offshore Mechanics and Arctic Engineering, 85178, V007T07A018.

Turnbull, I. D., Fournier, N., Stolwijk, M., Fosnaes, T., McGonigal, D., 2015. Operational iceberg drift forecasting in Northwest Greenland. Cold Regions Science and Technology, 110: 1-18.

Turnbull, I. D., King, T., Ralph, F., 2018. Development of a New Operational Iceberg Drift Forecast Model for the Grand Banks of Newfoundland. Arctic Technology Conference, OTC-29109-MS.

Woodworth-Lynas, C.M.T., A. Simms, C.M. Rendell., 1985. Iceberg grounding and scouring on the Labrador Continental Shelf. Cold Regions Science and Technology, 10 (2): 163-186.

Woodworth-Lynes, C., Nixon, D., Phillips, R., Palmer, A., 1996. Subgouge deformations and the security of Arctic marine pipelines. Offshore Technology Conference, OTC-8222-MS.

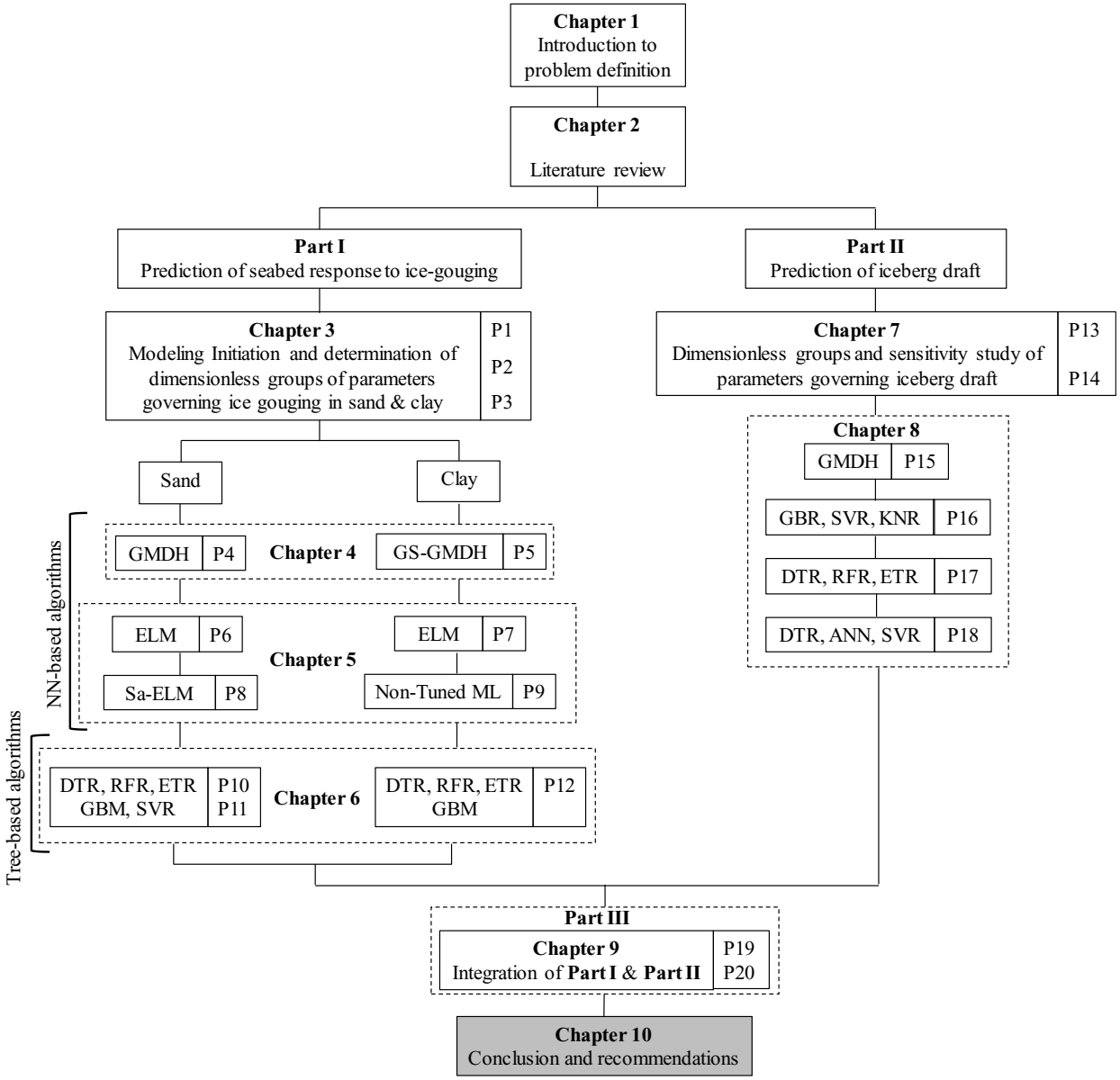
Yang, W., 2009. Physical Modeling of Subgouge Deformations in Sand. Doctoral dissertation, Memorial University of Newfoundland, St. John's, NL, Canada.

Younan, A., F. Ralph, T. Ralph, J. Bruce., 2016. Overview of the 2012 iceberg profiling program. Arctic Technology Conference. St. John's, NL, Canada, OTC-27469-MS.

Zhou, M., 2017. Underwater iceberg profiling and motion estimation using autonomous underwater vehicles. PhD dissertation. Memorial University of Newfoundland, NL, Canada.

10. Chapter 10

Conclusions and Recommendations



10.1.1. Conclusions

Developing a reliable and cost-effective solution to predict iceberg drafts and the subgouge soil displacement resulting from ice-gouging is a crucial task to assess the serviceability and structural integrity of subsea assets and offshore structures since the estimation of iceberg drafts can assist to evaluate the risk of ice-gouging events, which can threaten the operational and physical integrity of subsea structures and other infrastructures in the Arctic and non-Arctic offshore regions.

Moreover, the prediction of iceberg-seabed interaction parameters helps to estimate the seabed response to ice scour, which can affect the stress-strain behavior of the soil, the plastic shear strains, and the developed dead wedge underneath the ice keel. In other words, the subgouge soil characteristics determine the optimal design of subsea assets, such as the required burial depth, the physical protection, and the cost-effectiveness of the project.

Artificial intelligence (AI) and machine learning (ML) algorithms are undoubtedly changing the trends of the modern world and have wide applications in a range of fields. The current research project applied ML advancements for the first time to address the challenges associated with the cost-effective prediction of seabed response to ice-gouging. A series of investigations were conducted to address the key objectives of the study. Using the π-Buckingham theory and the sensitivity analysis, the most influencing dimensionless groups of parameters have been identified to model the iceberg drafts alongside subgouge soil characteristics in sand and clay. The iceberg draft, subgouge soil displacements, and reaction forces have been predicted using 14 ML algorithms comprising LR, ANN, MLPNN, GEP, ELM, SAELM, GMDH, GS-GMDH, DTR, RFR, ETR, GBR, SVM, and KNN.

This research has provided a fast and accurate ML-based solution for the iceberg draft prediction, which is vital for assessing the collision risk of deep keel icebergs with offshore facilities and subsea assets in ice-prone areas.

The applied methodology assists to overcome the limitations and uncertainties of the existing methods for measuring iceberg drafts, such as side scan sonar, remote sensing, and empirical formulas. It may be used as an additional source of information for operational iceberg management and drift forecasting, complementing the conventional dynamic models.

The ML-based solutions can design guidelines for iceberg management programs and subsea design projects. These ML models are able to predict the iceberg draft using iceberg length, width, mass, and shape factor; hence, it is useful for iceberg management programs that aim to monitor and track the movement and behavior of icebergs in navigable waters. For example, the Canadian Ice Service (CIS) is the leading authority for ice and iceberg information in Canada's navigable waters. This organization provides accurate and timely information on ice and icebergs to protect Canadians, their property, and their environment, by warning of dangerous ice conditions.

Hence, this intelligent tool provided in the current research potentially enhances the accuracy and timeliness of the CIS's information by predicting the iceberg draft based on available data.

Additionally, iceberg-seabed interaction modeling using ML technology can aid to evaluate the seabed response to ice scour, which is an important aspect of a safe and cost-effective design of subsea pipelines and other infrastructures in the ice-prone regions.

The proposed methodology in the current research may help to determine the required burial depth for the physical protection of the subsea assets against ice-gouging, which can significantly reduce operating expenses and downtime.

In other words, the prediction of subgouge soil features, e.g., deformations and reaction forces using ML advancement can improve the understanding of the stress-strain behavior of soil under an ice keel, which can affect the simple direct shear and the plastic shear strains.

The ML-based methods can be valuable for subsea design projects that involve the installation, operation, and recovery of subsea pipelines, structures, and facilities for companies that provide subsea project solutions for deep-water subsea installations and field developments. It may potentially improve the safety and efficiency of subsea project solutions by predicting the subgouge soil deformations and reaction forces based on relevant parameters.

The superior ML models for the estimation of the iceberg drafts as well as subgouge soil displacements and iceberg keel reaction forces have been introduced by performing different analyses such as sensitivity analysis, error analysis, discrepancy analysis, uncertainty analysis, and partial derivative sensitivity analysis.

The performed sensitivity analysis using different ML algorithms has provided a profound understanding of the influence of various input parameters such as iceberg length, iceberg width, iceberg mass, and iceberg shape factor on the iceberg draft prediction as well as attack angle, gouge depth, seabed soil properties, ice dynamics, etc., on the iceberg-seabed interaction mechanisms.

Several key observations were made throughout the study that can be summarized as follows:

- The GEP algorithm could automatically generate mathematical expressions from data to predict the horizontal subgouge soil deformations in the sandy seabed, which are important for assessing the potential damage to subsea pipelines and structures.
- The best MLPNN model was able to accurately predict the horizontal subgouge deformations in the sandy seabed, which were linearly related to the gouge depth and inversely related to the soil friction angle in this case.

- The MLPNN model was an applicable and reliable algorithm for simulating the horizontal subgouge sand deformations, which could help engineers and designers to optimize the burial depth and physical protection of subsea pipelines and structures.
- A new approach was applied to model the iceberg-seabed interaction process using dimensional analysis and Buckingham's π theorem, which was a powerful and simple technique that could reduce the complexity and dimensionality of the problem.
- A systematic and rigorous methodology was utilized to identify the dimensionless groups of parameters governing the iceberg-seabed interaction process, which was able to facilitate proposing potential new solutions and improving existing ones.
- The research demonstrated the applicability and reliability of the LR models for simulating the subgouge soil deformations, which could assist engineers and designers to optimize the burial depth and physical protection of submarine infrastructures.
- A new dimensionless parameter, the frontal berm height ratio, was recognized as a significant input parameter on horizontal and vertical subgouge deformations in the sandy seabed predicted by the GMDH algorithm.
- Several explicit GMDH and GS-GMDH equations were provided to estimate the horizontal and vertical soil deformations, which could support the decision-making and planning of subsea pipeline installation, operation, and maintenance in the Arctic and non-Arctic offshore regions.
- A set of ELM-based matrices, as simple and reliable tools, were derived to approximate the iceberg-seabed interaction parameters, e.g., subgouge deformations and reaction forces, in sandy and clay seabed for engineering applications.

- The ELM algorithm found that the soil depth was the most influential input factor affecting the subgouge soil deformations, while the vertical component of load alongside the attack angle parameter was the most effective for the simulation of the iceberg keel-seabed reaction forces.
- SaE-ELM algorithm identified the sand shear strength parameter and gouge depth ratio as influential input parameters in the prediction of the subgouge reaction forces and displacements.
- A non-tuned extreme learning machine algorithm could detect the soil depth as the most effective input factor affecting the subgouge soil deformations, and the vertical component of load and attack angle parameter were the most important factors for the simulation of the iceberg-seabed reaction forces.
- Tree-based ML algorithms were able to provide accurate and reliable predictions of the subgouge soil deformations with less computational cost and time than other solutions. These algorithms could capture the nonlinear and complex behavior of the soil-ice interaction process.
- The RFR method showed better performance than other ML-based models such as the gradient boosting model (GBM) and support vector regression (SVR).
- The RFR model revealed that the most important parameters influencing the iceberg-seabed interaction process in the sandy seabed were the position of the iceberg along the scour axis and the berm height ratio to model the reaction forces; however, the soil depth ratio and the berm height ratio had a significant effect on the estimation of the subgouge displacements.

- A comparison of the performance of the ETR algorithm with other tree-based ML methods, e.g., DTR, RFR, and GBR, proved that the ETR method achieved the highest accuracy and correlation among the tree-based ML methods for both the keel reaction forces and the subgouge soil deformations.
- A novel methodology for predicting the iceberg draft was developed using LR analysis based on four dimensionless groups of parameters governing the iceberg draft estimation.
- The LR models could provide simple and reliable predictions of the iceberg draft with lower computational time and expenses than existing numerical and analytical methods.
- A robust and systematic approach for conducting a sensitivity analysis of the iceberg draft estimation was implemented using neural network-based models, which could support identifying the most meaningful parameters and improve the accuracy and reliability of the predictions.
- The SVR algorithm demonstrated higher precision, correlation, and lower complexity to approximate the iceberg draft than ANN and DTR models.

It is worth noting that the iceberg-seabed interaction process and iceberg draft prediction are very complex and nonlinear phenomena involving a high level of uncertainty. Estimating the subgouge soil deformations and reaction forces as well as the iceberg drafts is essential for the design and management of the sea bottom-founded infrastructures. However, existing methods for estimating these quantities are either computationally expensive, data-intensive, or empirically based. To overcome these challenges, this novel ML-based methodology has been proposed in this dissertation that could predict the subgouge soil deformations and reaction forces using a set of dimensionless groups of parameters in clay and sandy seabed. Moreover, this intelligent model was able to predict the iceberg drafts by utilizing the length, width, mass, and shape factors of the

iceberg. According to the obtained results, this research has several advantages and benefits over existing methods, such as high accuracy, low complexity, fast computation, and wide applicability. The models have been validated using experimental and field data and compared their performance with existing empirical models. It was demonstrated that the model could provide reliable and robust predictions of the iceberg-seabed interaction process in different regions and conditions. For example:

- The offered methodology could seize the nonlinear and complex behavior of the iceberg-seabed interaction mechanism very well. It was applied to different soil types, attack angle geometries, and loading conditions.
- This investigation contributed to the advancement of knowledge and understanding of the iceberg-seabed interaction process and iceberg draft prediction, which might improve the design and management of subsea assets in cold waters.
- By evaluating the performance and applicability of different ML algorithms, the best ML models were introduced for the estimation of the iceberg draft and iceberg-seabed interaction process.
- The most influential parameters for estimating the iceberg drafts and iceberg-seabed interaction mechanism were determined, which might reduce the uncertainty and variability of the predictions.
- The comparison of the accuracy and reliability of the ML algorithms and existing empirical models might validate and improve the performance of the empirical methods.
- Overall, the extra tree regression (ETR) algorithm was found to be the superior ML algorithm to predict the iceberg draft and subgouge soil characteristics.

The study has shown that the usage of ML algorithms for the prediction of iceberg drafts and the subgouge soil characteristics can be an attractive and cost-effective alternative to the current iceberg management systems and subsea structure designs. However, the accuracy and sufficiency of these approaches may be affected by the following aspects:

- The MLPNN algorithm was not able to provide an explicit relationship to estimate the horizontal subgouge soil deformations.
- Relationships between the dimensionless groups and the subgouge soil deformations were linearly considered. Although considering the nonlinearity between these dimensionless groups and subgouge soil deformations might increase the complexity of the solution, it could help to have more realistic conditions.
- The seabed was assumed as homogeneous and isotropic soil to perform the dimensionless analysis of the iceberg-seabed interaction mechanism, and the effects of soil heterogeneity, strain rate effects, pore water pressure, and anisotropy on the subgouge soil deformations were not considered.
- The iceberg keel shape was presumed rectangular and trapezoidal, and the effects of different ice keel shapes, such as triangular and irregular, were not assessed on the subgouge soil deformations.
- It was presumed that the ice block moved horizontally on a uniform seabed at a constant velocity to analyze the iceberg-seabed interaction mechanism in the present study, whereas a realistic iceberg-seabed interaction process might be on a non-uniform seabed with non-constant velocity in nature.

- The iceberg-seabed interaction process was regarded quasi-static mechanism, and the influence of dynamic and transient factors like iceberg acceleration and iceberg inertia was not considered on the subgouge soil deformations.
- The ML methods required a large amount of training data to achieve optimal results.
- The uncertainty and variability of the environmental and operational factors, e.g., salinity, temperature, current, wind, ice crushing, and waves were not considered in the present investigation that might influence the iceberg draft estimation.

10.1.2. Recommendations for future studies

In this thesis, a novel ML-based methodology was developed to predict the subgouge soil deformations, reaction forces, and iceberg drafts. The model has shown high accuracy, low complexity, fast computation, and wide applicability in estimating the iceberg-seabed interaction process. However, there are still some possible ways to expand our work in future studies as below:

- To improve the performance of ML algorithms, having a comprehensive and high-quality dataset is crucial. A combination of field measurements and surficial data including satellite images, aerial photography, etc., are recommended to be conducted to construct a strong dataset for training and testing of the ML algorithms.
- More reliable predictions of the subgouge soil displacements need accurate identification of the advanced parameters that govern the iceberg-seabed interaction mechanisms including the rate effects, ice dynamics, etc., the better modeling of the physics of the problem, the more accurate and reliable outcome.
- The study is recommended to be extended by the inclusion of the pipeline in the analysis. This will need coupled analysis and testing of the iceberg-pipeline-seabed interaction,

particularly the trenching and backfilling effects that may significantly affect the pipe response to the ice-gouging.

- The effects of the mechanical properties of ice and iceberg crushing can be considered in the iceberg-seabed interaction process.
- Other aspects of the iceberg-seabed interaction process, e.g., soil failure mechanisms can be evaluated in future studies.
- Incorporating more parameters, e.g., the iceberg orientation, trajectory, and drifting to improve the precision and reliability of the simulation procedure, may affect the prediction of the iceberg-seabed interaction process and iceberg draft estimation.
- The effect of other types of soils in different seabed conditions, such as silt, gravel, or mixed soils, can be evaluated on the performance and applicability of ML-based solutions.
- The influence of other parameters, e.g., soil anisotropy, strain rate, pore water pressure, iceberg density, and water salinity, may be considered in the future works.