Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Electrical and Computer Engineering

University of Saskatchewan

Saskatoon

By

Nima Safari

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Electrical and Computer Engineering

57 Campus Drive

University of Saskatchewan

Saskatoon, Saskatchewan (S7N 5A9)

Canada

OR

Dean

College of Graduate and Postdoctoral Studies

University of Saskatchewan

116 Thorvaldson Building, 110 Science Place

Saskatoon, Saskatchewan (S7N 5C9)

Canada

ABSTRACT

Amid growing concerns about climate change and non-renewable energy sources depletion, variable renewable energy sources (VRESs) are considered as a feasible substitute for conventional environment-polluting fossil fuel-based power plants.

Furthermore, the transition towards clean power systems requires additional transmission capacity. Dynamic thermal line rating (DTLR) is being considered as a potential solution to enhance the current transmission line capacity and omit/postpone transmission system expansion planning, while DTLR is highly dependent on weather variations. With increasing the accommodation of VRESs and application of DTLR, fluctuations and variations thereof impose severe and unprecedented challenges on power systems operation. Therefore, short-term forecasting of large-scale VERSs and DTLR play a crucial role in the electric power system operation problems. To this end, this thesis devotes on developing forecasting models for two large-scale VRESs types (i.e., wind and tidal) and DTLR.

Deterministic prediction can be employed for a variety of power system operation problems solved by deterministic optimization. Also, the outcomes of deterministic prediction can be employed for conditional probabilistic prediction, which can be used for modeling uncertainty, used in power system operation problems with robust optimization, chance-constrained optimization, etc. By virtue of the importance of deterministic prediction, deterministic prediction models are developed. Prevalently, time-frequency decomposition approaches are adapted to decompose the wind power time series (TS) into several less non-stationary and non-linear components, which can be predicted more precisely. However, in addition to non-stationarity and nonlinearity, wind power TS demonstrates chaotic characteristics, which reduces the predictability of the wind power TS. In this regard, a wind power generation prediction model based on considering the chaosity of the wind power generation TS is addressed. The model consists of a novel TS decomposition approach, named multi-scale singular spectrum analysis (MSSSA), and least squares support vector machines (LSSVMs).

Furthermore, deterministic tidal TS prediction model is developed. In the proposed prediction model, a variant of empirical mode decomposition (EMD), which alleviates the issues associated with EMD. To further improve the prediction accuracy, the impact of different components of wind power TS with different frequencies (scales) in the spatiotemporal modeling of the wind farm

is assessed. Consequently, a multiscale spatiotemporal wind power prediction is developed, using information theory-based feature selection, wavelet decomposition, and LSSVM.

Power system operation problems with robust optimization and interval optimization require prediction intervals (PIs) to model the uncertainty of renewables. The advanced PI models are mainly based on non-differentiable and non-convex cost functions, which make the use of heuristic optimization for tuning a large number of unknown parameters of the prediction models inevitable. However, heuristic optimization suffers from several issues (e.g., being trapped in local optima, irreproducibility, etc.). To this end, a new wind power PI (WPPI) model, based on a bi-level optimization structure, is put forward. In the proposed WPPI, the main unknown parameters of the prediction model are globally tuned based on optimizing a convex and differentiable cost function. In line with solving the non-differentiability and non-convexity of PI formulation, an asymmetrically adaptive quantile regression (AAQR) which benefits from a linear formulation is proposed for tidal uncertainty modeling. In the prevalent QR-based PI models, for a specified reliability level, the probabilities of the quantiles are selected symmetrically with respect the median probability. However, it is found that asymmetrical and adaptive selection of quantiles with respect to median can provide more efficient PIs. To make the formulation of AAQR linear, extreme learning machine (ELM) is adapted as the prediction engine. Prevalently, the parameters of activation functions in ELM are selected randomly; while different sets of random values might result in dissimilar prediction accuracy. To this end, a heuristic optimization is devised to tune the parameters of the activation functions.

Also, to enhance the accuracy of probabilistic DTLR, consideration of latent variables in DTLR prediction is assessed. It is observed that convective cooling rate can provide informative features for DTLR prediction. Also, to address the high dimensional feature space in DTLR, a DTR prediction based on deep learning and consideration of latent variables is put forward.

Numerical results of this thesis are provided based on realistic data. The simulations confirm the superiority of the proposed models in comparison to traditional benchmark models, as well as the state-of-the-art models.

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to my supervisor Dr. Chi Yung Chung for his continuous support throughout my Ph.D. study. I have learned priceless lessons from his expertise, vision, and personality. He has also financially supported me throughout my Ph.D. program and made it possible to only zero in on my research without any financial concerns.

Also, I would like to thank the College of Graduate and Postdoctoral Studies for providing the Dean's Scholarship which is the most prestigious scholarship in the University. Besides, I would like to thank the National Science and Engineering Research Council for providing partial financial support. Also, I would like to sincerely thank the Saskatchewan Power Corporation (SaskPower) for providing data. Special thanks go to the members of my advisory committee: Dr. Nurul A. Chowdhury, Dr. Khan Wahid, Dr. Daniel Chen, and Dr. Seok-Bum Ko for their insightful comments and suggestions throughout my Ph.D. program.

The last but not the least, I owe my deepest gratitude to my parents, Mehdi and Zahra, and my loving wife, Elnaz. Words cannot express how grateful I am to them for their continuous spiritual support.

DEDICATION

Dedicated to my Father, ***Mehdi****, my Mother,* ***Zahra****,*

and

my beloved Wife, ***Elnaz***

TABLE OF CONTENTS

LIST OF TABLES AND ILLUSTRATIONS

LIST OF FIGURES AND ILLUSTRATIONS

LIST OF ABBREVIATIONS

AAQR	Asymmetrically Adaptive Quantile Regression
AE	Autoencoder
AESO	Alberta Electric System Operator
AR	Autoregressive
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
CEEMD	Complementary Ensemble Empirical Mode Decomposition
CDF	Cumulative Distribution Function
CL	Confidence Level
CSA	Coupled Simulated Annealing
CV	Cross-Validation
CWC	Coverage Width-based Criterion
CWT	Continuous Wavelet Transform
DAE	Denoising Autoencoder
DISR	Double Input Symmetrical Relevance
DNN	Deep Neural Network
DTLR	Dynamic Thermal Line Rating
ELM	Extreme Learning Machine
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition

EPEC	Electric Power and Energy Conference
ERCOT	Electric Reliability Council of Texas
FMF	Fuzzy Membership Function
FORCE	Fundy Ocean Research Center for Energy
FS	Feature Selection
GA	Genetic Algorithm
GAM	Generalized Additive Model
GHGE	Greenhouse Gas Emissions
GWO	Gray Wolf Optimizer
HIA	Hybrid Intelligent Algorithm
ICEEMDAN	Improved Complete Ensemble Empirical Mode Decomposition Adaptive Noise
IEA	International Energy Agency
IESO	Independent Electric System Operator
KNN	K- Nearest Neighbor
LCOE	Levelized Cost of Energy
LP	Linear Programming
LSSVM	Least Square Support Vector Machine
LSTM	Long Short-Term Memory
LUBE	Lower Upper Bound Estimation
MCS	Monte Carlo Simulation
MI	Mutual Information
MLE	Maximal Lyapunov Exponent

MLP	Mathematical-Morphology-based Local Prediction
MOD	Method of Delay
MRES	Marine Renewable Energy Source
MSSSA	Multi-Scale Singular Spectrum Analysis
MTD	Mean Trend Detector
NAE	Normalized Absolute Error
NARX	Nonlinear Autoregressive Exogenous
NCWC	New Coverage Width-based Criterion
NMAE	Normalized Mean Absolute Error
NN	Neural Network
NOAA	National Oceanic and Atmospheric Administration
NPI	Nonparametric Prediction Interval
NRMSD	Normalized Root-Mean-Squared Error Deviation
NRMSE	Normalized Root-Mean-Squared Error
NWEC	National Wind Energy Center
NWP	Numerical Weather Prediction
OES	Ocean Energy System
OHL	Overhead Line
PDF	Probability Distribution Function
PI	Prediction Interval
PIMSE	Prediction Interval Mean-Squared-Error
PINAW	Prediction Interval Nominal Average Width
PM	Persistence Model

PSO	Particle Swarm Optimization
QR	Quantile Regression
QRF	Quantile Regression Forest
RBFNN	Radial Basis Function Neural Network
RES	Renewable Energy Sources
RL	Reliability Level
RMSE	Root-Mean-Squared Error
RNN	Recurrent Neural Network
SDAE	Stacked Denoising Autoencoder
SO	System Operator
SSA	Singular Spectrum Analysis
STLR	Static Thermal Line Rating
SUMD	Sacramento Municipal Utility District
SVM	Support Vector Machine
SVR	Support Vector Regression
TCD	Tidal Current Direction
TCS	Tidal Current Speed
TEP	Transmission Expansion Planning
TL	Tidal Level
TS	Time Series
VRES	Variable Renewable Energy Source
WD	Wavelet Decomposition
WT	Wavelet Transform

WPP	Wind Power Prediction
WPPI	Wind Power Prediction Interval

1- INTRODUCTION

1-1 General Context

Ever growing global electricity demand, fossil fuel depletion crisis, global warming concerns, and environmental pollution issues necessitate looking for alternative substitutes for electricity power generation. Figure 1-1 demonstrates the greenhouse gas emissions (GHGEs) per person in different provinces of Canada in 2015 [2, 3]. As it is apparent from this figure, Saskatchewan province is ranked first in producing greenhouse gas. Besides, according to the statistics presented in [4], in 2015, Canada was amongst the top-10 countries in GHGEs.

Since 2015, Canada, as a member of G7 [5], has aimed at cutting the GHGEs by terminating the use of fossil fuels by the end of the century [6]. The electricity sector is one of the major producers of greenhouse gases; therefore, as part of this plan, in 2016, Canada has announced phasing out the use of coal-fired electricity generation units by 2030 [7]. In support of this national transition to a cleaner electricity sector, provincial governments along with their electricity sectors have

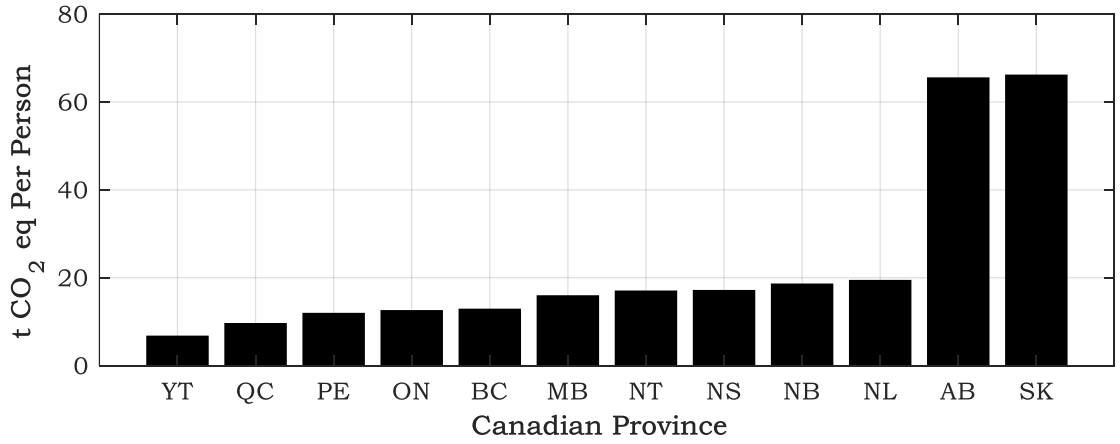

Figure 1-1: The greenhouse gas emission in Canada in 2015.

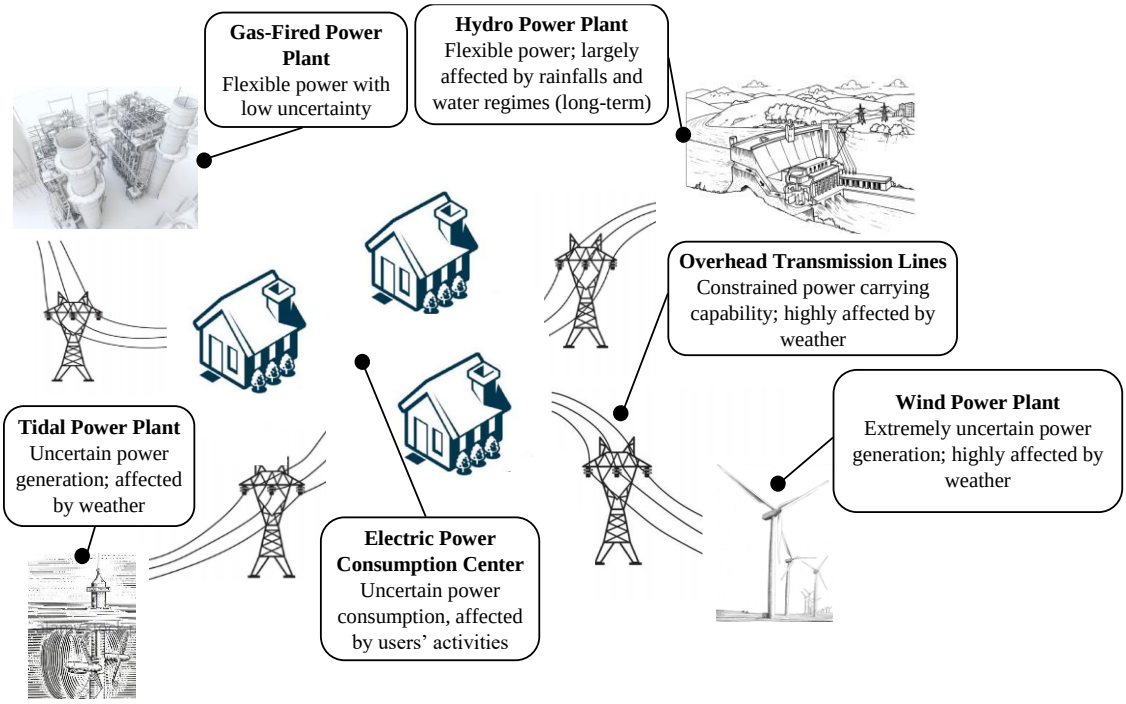

Figure 1-2: An example of the power system future scenario under high penetration of RESs.

planned to ambitiously increase the penetration of renewable energy sources (RESs), including hydro, wind, tides, etc.

Figure 1-2 displays an example of future power systems without coal-fired generation units. In such case, the electricity load, which is highly uncertain, should be predicted for the next day and/or next coming hours. Various load forecasting tools (i.e., [8-11]) have been developed in the literature to decrease the uncertainty of the electricity load values for the next hours. Based on the load forecasting output, the generation units which are usually far from load centers need to be economically scheduled such that the electricity demand is met. However, some degree of uncertainty in load forecast is inevitable. In this regard, operating reserve units need to be at disposal to meet the demand in the case that the demand is more than the forecasted values. On

the other hand, in the case that the demand is lower than the scheduled power generation, some generation units need to ramp down and decrease their generation.

Changing the generation in a short period requires the high flexibility of the generation units. As shown in Figure 1-2, in the future scenario only gas-fired generation and hydropower plants are flexible, and their generation output can be dispatched in the case operating reserve or ramp down is required; while wind power generation and tidal power generation varies based on the weather [12, 13]; and therefore, they cannot be used as an operating reserve. Besides, due to the high level of uncertainty in weather and consequently wind and tidal power generation, economic scheduling of dispatchable and flexible generation units (i.e., hydro and gas-fired generation units) for the next day and/or coming hours become hardened. Accurate forecasting of the variable RESs (VRESs), e.g., wind and tidal, can substantially facilitate the power system operation problem by alleviating the VRESs uncertainty. Based on the aforementioned discussion, precise forecasting of available generation from VRESs is a key for future power systems operation.

In generation scheduling, the capability of transmission lines in carrying the available power and delivering to load center is an imperative constraint. Generally, the large-scale VRESs are far from the load centers; therefore, transmission lines are required to connect them thereto. Transmission expansion planning (TEP) requires huge investments and time. Besides, it also brings several adverse environmental impacts. Prevalently, the transmission systems are designed conservatively. In this manner, the maximum allowable power carrying capability, known as static thermal line rating (STLR), is calculated such that the transmission line's temperature does not exceed the maximum tolerable temperature under the worst weather conditions. Unlike the conservative STLR, a dynamic thermal line rating (DTLR) estimates the transmission line's actual ampacity under the present weather conditions [14]. The additional exploitable headroom of the transmission

lines can accommodate the VRESs; and thus, the need for TEP can be postponed or even omitted. As DTLR is weather dependent, the surplus transmission line ampacity varies. To effectively consider the DTLR in generation scheduling, its prediction becomes of great importance.

To recapitulate, with the increasing need for the integration of higher amount of VRESs and curtailing fossil fuel use in the electricity industry. Uncertainty in power system operation significantly increase and therefore results in unprecedented issues. Prediction of large-scale VRESs and DTLR can greatly address the power system operation problems.

1-2 Research Motivations

Wind—one of the most distributed available energy sources—is increasingly exploited for electricity generation all around the world [15]. Due to more than 40% average reduction of the levelized cost of energy (LCOE) of renewable energy generation during the last five years, nowadays, wind provides electricity competitively comparable with fossil-fuel-fired generation units in terms of overall cost [16]. Consequently, in several countries, such as Denmark, Spain, etc., wind directly participates in the market. Furthermore, several other countries, including Canada, have already considered the wind as a valuable and clean alternative, and have substantially added wind power to their electricity supply systems. As an example, in the last five years, the average growth in wind power generation capacity in Canada has an average of 23% increase per year [17]. In 2016, Canada invested $1.5 B in the wind power industry [18]. Wind energy provides approximately 5% of Canada's total electricity and is projected to generate 20% or more of the electricity demand by 2025. Also, SaskPower set an ambitious goal of wind power penetration increment to 30% by 2030.

In addition to wind harnessing, marine RESs (MRESs) harvesting has recently grabbed remarkable attention across governments and power industries in the United States, Canada, Australia, France, Ireland, China, and South Korea [19]. Although the MRESs technology is still in developing stage, Ocean Energy Systems (OES) projects the potential of developing 748 GW installed MRESs by 2050 [20]. Among MRESs, tidal energy is of the main interest in Canada due to several potential sites [19]. The Bay of Fundy, located in Canada, has the highest tides in the world, it is assessed that more than 2,500 MW can be extracted from the 8,000 MW of the kinetic resource in the Bay of Fundy [21]. In this regard, Fundy Ocean Research Center for Energy (FORCE)—a not-for-profit corporation—is attracted $239.6 M funding from both public and private sectors for establishing research platform, site characterization and monitoring, and increasing the tidal energy capacity. In 2014, FORCE created the largest transmission capacity (64 MW) for tidal power in the world.

With the enormous investments in wind and tidal energy generation, unprecedented challenges are faced by power systems. These are mainly due to uncertainty and variability of the output generation of these VRESs, such as wind and tidal. However, employing forecasting tools and using their outputs as the input of the generation scheduling optimization problem can substantially alleviate the emerged issues [22].

With the electricity demand growth as well as the extensive VERSs exploitations, transmission lines, which are operated near their STLR constraints, may not be able to transfer the electric power due to transmission line congestions. In such a case, there are two main solutions including transmission system expansion and DTLR. The former requires vast investments and time; also, it causes several environmental concerns, while the latter can lead to an increment of the current transmission lines' ampacity and therefore delaying the need for transmission expansion with

minimum costs. However, DTLR highly varies based on meteorological variables, and therefore, the maximum available transmission line capacity is uncertain.

Increasing the penetration of VRESs, including wind and tidal, surges the power systems uncertainty. On the other hand, employing the DTLR for meeting the ever-growing demand through exploiting VRESs intensifies the uncertainty in power systems. The progressively growing uncertainty can lead to several challenges in both power systems operation and planning problems. Forecasting tools can provide estimations of the available generation from VRESs and available ampacity of transmission lines. Employing the forecasting tools outcome as the input of the operation optimization problems can markedly decrease the uncertainty in VRESs and DTLR. The cost of VRESs error in forecasting manifests depending on power system operation structures. In a study conducted by Xcel Energy[1] [23], it is concluded that reduction of the normalized mean absolute error (NMAE) by one percent would have saved over $1 million in Public Service of Colorado in 2008 [24]. Besides, the prediction model which is preferred by power producers and system operators (ISOs) are widely different from each other. For instance, some system operators like the Electric Reliability Council of Texas (ERCOT) use historical data of prediction error statistics to determine the non-spinning reserve[2] requirements. While some others like Sacramento Municipal Utility District (SUMD) in CA, USA requires spatial and temporal variability of VRESs for energy market and dispatch decisions [16]. Therefore, the need for more accurate prediction models and different types of prediction models are the main motivations of this thesis.

[1] Xcel Energy currently manages about 7000 megawatts of installed wind power.

[2] The non-spinning reserve, which is also known as supplemental reserve, is the ancillary generating capacity that is not currently offline and disconnected from the electric power grid but can be brought online within a short period of time. In an isolated electric power grid, non-spinning reserve is supplied by fast-start generators. While, in an interconnected electric system, the non-spinning reserve might be the additional power which can be imported from other electric power grids.

Based on the forecast horizon, prediction models can be categorized into four groups, comprising long-, medium-, short-, and very short-term prediction [12]. Long- and medium-term prediction models with horizons more than 6 hours ahead are employed for maintenance scheduling, reserve planning, unit commitment, etc., while short- and very short-term prediction models are vital for economic dispatch, electricity market clearing, regulation and control actions, etc. Physical models are preferred for prediction horizons longer than 6 hours ahead; however, statistical models are utilized for shorter horizons due to their low computational complexity and high accuracy. Due to the high importance of short-term prediction horizons, this work zeros in on developing a statistical prediction for short horizons.

1-3 Research Objectives and Scope

The main objective of this research is to provide proficient short-term prediction models required for optimal operation of electric power system in high penetration of large-scale VRESs, specifically wind and tidal. The objective of this Ph.D. study can be apportioned to six sub-objectives.

The **first** is to develop accurate deterministic prediction models which are capable of predicting the wind power and tidal energy. This deterministic prediction can be used for a range of power system problems, such as economic dispatch [25], energy storage sizing and planning [26], etc. The **second** is to develop wind power prediction (WPP) considering the spatiotemporal correlation among wind farms. Usually, several wind farms may be situated in a region; and therefore, the reflection of the correlation of wind farms power generation during the last samples with the target wind farm output in the next samples in the prediction model can enhance the prediction accuracy. The **third** is to address probabilistic prediction tools for robust [27] or interval optimization-based

[28] power system problems. The **fourth** is to develop DTLR forecasting model which facilitates efficient integration of large-scale VRESs to the electric power systems.

In line with the **first** sub-objective, in Chapters 0 and 3-, two deterministic prediction models are developed for wind and tidal prediction, respectively. According to the literature, wind power time series (TS) possesses nonlinearity and nonstationarity, which are the main culprits for inaccurate prediction of wind power. Signal processing approaches have been recently used to decompose the wind power TS into several less non-stationary and non-linear components which can be predicted more efficiently. Some recent studies reveal that wind power TS contains chaotic components. chaos is the property of some types of nonlinear system and results in wild and random looking patterns in TS. The existence of strong chaos in wind power TS can lead to inaccurate prediction. In Chapter 4-, the importance of considering different components with different frequency ranges (scales) in spatiotemporal wind power prediction modeling is analyzed, and a multiscale spatiotemporal prediction model is developed towards realizing the **second** sub-objective. To facilitate the power system operators with PI models which are of great importance for robust or interval-based power system optimization, prediction models for wind and tidal are respectively developed in Chapters 5-6- and 6 as the **third** sub-objective. In Chapter 7-, a DTLR probabilistic forecasting is proposed towards realizing the **fourth** sub-objective.

In this thesis, real-world data are used to develop and validate prediction models. The data are mainly available for the public. For wind power generation prediction, the data from AESO, Sotavento wind farm (located in Spain), and Saskatchewan wind farms, are used. While for tidal prediction due to unavailability of historical tidal power data, the influential factor to tidal power generation, including tidal level (TL), tidal current speed (TCS), and tidal current direction (TCD), are predicted. In the literature (e.g., [29, 30]), models are developed to estimate the tidal power

based on the mentioned factors. To develop tidal prediction models, the TL data from Port Dover (located in Ontario, Canada) and TCS and TCD data from Bay of Fundy (located in Nova Scotia, Canada) and Shark River Entrance (located in New Jersey, US) are used. Also, the DTLR model is developed based on the data related to M2 met tower in National Wind Energy Center (NWEC) [31].

1-4 Thesis Contributions

To achieve the objectives stated in Section 1-3, several contributions have been made to the related literature. A summary of these contributions is illustrated in the following. However, in each chapter of this manuscript-based thesis, these contributions are further elaborated.

The first contribution of this thesis is to develop a multi-step deterministic prediction model for wind power forecasting. This multi-step prediction model can be used for a range of power system problems, e.g., economic dispatch [32], energy storage planning and coordination [33], etc. The proposed prediction model consists of two main stages, including decomposition, and prediction. In the decomposition stage, a TS decomposition framework, based on considering the non-linearity, non-stationarity, and chaosity of the wind power TS. Prevalently, wind power TS is decomposed into several components using the decomposition tools, such as wavelet decomposition (WD), empirical mode decomposition (EMD), etc. Afterward, the decomposed components, which are less non-linear and non-stationary, are used to develop the prediction models. The existence of strong chaotic components can deteriorate the predictability of a TS. In recent literature, the presence of chaotic components in the decomposed wind power TS components are shown; however, no decomposition framework has been proposed to deal with these chaotic components, thus far. To this end, In Chapter 0 [12], I propose a decomposition

framework based on ensemble EMD (EEMD), chaotic TS analysis, and singular spectrum analysis (SSA). To realize the multi-step prediction, in the prediction stage, iterative prediction strategy is devised for non-chaotic components. While the chaotic components are predicted using direct prediction strategy, which prevents from accumulating the prediction error in predicting the chaotic components.

As the second contribution, a tidal energy prediction is proposed which considers the non-stationarity and non-linearity of the TCS and TCD TS. In the recent literature, it is shown that the TCS and TCD TS cannot be represented by a set of periodic components, and therefore, the TCD and TCS are non-stationary [34]. Besides, it is shown that TCS and TCD are nonlinear, and thus, linear prediction models, such as autoregressive moving average (ARMA), are not apposite for TCS and TCD prediction [34]. To address these issues, first, a decomposition approach named improved complete EEMD adaptive noise (ICEEMDAN), is proposed. ICEEMDAN is employed to decompose the TCS/ TCD TS into several less non-stationary components. The main reason behind proposing ICEEMDAN is to address the issues related to EEMD. Traditionally, to perform EEMD different realizations of the white noise are added to the original TS (i.e., TCS and TCD) [13]. However, using various realizations of the noise with the original TS can result in various extracted components for each noise associated TS; thus, the final components, obtained from EEMD, might vary for different sets of noise. Secondly, adding various noises can result in a different number of EMD components; therefore, the final averaging process for finding EEMD components becomes problematic. To further improve the prediction accuracy, a localized prediction model is developed. In the proposed prediction model, unlike the thus far TCD and TCS prediction, the prediction model is trained and developed based on the feature input vectors and their corresponding target, which are nearest

to the test feature input vector. To this end, *K*-nearest neighbor (KNN) is employed to identify the training set. Least square support vector machine (LSSVM), which has several advantages compared to the other prediction engines (e.g., neural networks, support vector machine, etc.) is employed to the TCS and TCD for the first time. To modify the prediction error, an error correction stage, consisting of an ensemble of extreme learning machines (ELMs), is developed.

As the third contribution, a multiscale spatiotemporal wind power prediction model is developed. In this prediction model, the correlations of wind farms located in the vicinity of the target wind farms are considered, using wavelet decomposition and a mutual information-based feature selection (FS). In this prediction model, first, the wind power TS of different wind farms are decomposed into different frequency ranges (scales). Then, using FS, the most relevant scales of different wind farms in different time lags are considered as the input of the prediction engine. Afterward, localized prediction models are developed to predict the next value of each component, and the wind power generation in the next sample is predicted by aggregation of predicted values of all components.

As the fourth contribution, a WPP interval (WPPI) for short-term prediction is proposed to provide information about the wind power generation uncertainty. This WPPI can be employed as the input of different power system regulation tasks. Tuning the prediction engine is one of the most imperative steps in developing a WPPI. The prediction engine parameters are tuned using a cost function, which needs to be optimized. Most of the thus far proposed WPPI models are based on non-differentiable and non-convex cost functions, which make the use of heuristic optimization unescapable. However, heuristic optimizations are prone to be trapped in local solutions. Besides, with increasing the number of prediction parameters, which need to be tuned, finding the optimal prediction engine parameters by heuristic optimizations become

more troublesome. To this end, I propose a bi-level WPPI, in which the upper-level controls the quality of the WPPI based on the power system operator preference using few hyperparameters. While the lower-level tunes the ELM parameters via a differentiable and convex optimization using global optimization techniques.

As the fifth contribution, a tidal prediction interval (PI) for short-term prediction is put forward. This PI model is developed for TL, TCS, and TCD prediction, which are key factors in tidal power generation by means of different technologies. The proposed model is based on a bi-level optimization formulation, in which ELM prediction engine and quantile regression (QR) are employed. The quantile probabilities are asymmetrically and adaptively chosen in the upper-level optimization, which assesses the quality of the PIs. Additionally, the training process of ELM is amended by adaptively selecting ELM's hidden neurons via upper-level optimization. The lower-level problem attains ELM's output weighting coefficients by linear programming of QR. The heuristic optimization, consisting of gray wolf optimizer and simplex method, is designed to equip the NPI with high exploration and exploitation capabilities in the upper-level optimization.

As the sixth and the last contribution, a probabilistic DTLR is introduced for short-term power system operation problems. The proposed DTLR PI model benefits from new predictors which can provide valuable information about DTLR by representing the highly nonlinear and complex relation of DTLR values with meteorological variables. Besides, in the DTLR PI, for the first time, a deep neural network (DNN) is adapted as the prediction engine. Stacked denoising autoencoder (SDAE) is employed for feature extraction and dimension reduction. The DNN is composed of recurrent neural networks (RNNs) which are more suitable in TS

forecasting. A novel cost function, defined to facilitate the training of the prediction model for probabilistic forecasting, is another valuable feature of this work.

1-5 Thesis Organizations

The rest of this manuscript-based thesis is as follows.

Chapter 0 is titled “A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis.” In this Chapter, the proposed multi-step deterministic wind power prediction, which is published in *IEEE Transactions on Power Systems*, is described. Mr. G. C. D. Price is a co-author of this paper, and he provided me with invaluable technical comments. Also, he helped me to evaluate the proposed model using the Centennial wind power generation. I developed and implemented the model, performed the simulation, and analyzed the results, along with the paper write-up

Chapter 3- is titled “An advanced multistage multi-step tidal current speed and direction prediction model.” This chapter presents the proposed deterministic prediction for TCS and TCD, which has been published as an article in Proceedings of the Electric Power and Energy Conference (EPEC), Saskatoon, Saskatchewan, CA, October 2017. Mr. Khorramdel and Mr. Zare assisted me in preparing the presentation and programming the forecasting model in MATLAB software. I developed and implemented the model, performed the simulation, and analyzed the results, along with the paper write-up

Chapter 4- titled “A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction.” This chapter demonstrates the multiscale spatiotemporal wind power prediction modeling, which has been published as an article in Proceedings of the EPEC, Saskatoon, Saskatchewan, CA, October 2017. Ms. Y. Chen, Mr. B.

Khorramdel, and Mr. L. P. Mao assisted me in developing the model in MATLAB and preparing the presentation. I developed and implemented the model, performed the simulation, and analyzed the results, along with the paper write-up.

Chapter 5- titled "Very short-term wind power prediction interval framework via bi-level optimization and novel convex cost function." This chapter describes the proposed wind power PI, which has been submitted to *IEEE Transactions on Power systems* for the third round of revision. Dr. Mazhari as the co-author provided invaluable comments and suggestions during the development of the model. I developed and implemented the model, performed the simulation, and analyzed the results, along with the paper write-up.

Chapter 6- is titled "Tidal current and level uncertainty prediction via adaptive linear programming." This chapter demonstrates the proposed tidal PI, which has been published in *IEEE Transactions on Sustainable Energy*. Dr. Mazhari as the co-author provided invaluable comments and suggestions during the development of the model. I developed and implemented the model, performed the simulation, and analyzed the results, along with the paper write-up.

Chapter 7- is titled "A secure deep probabilistic dynamic thermal line rating prediction." This chapter demonstrates the proposed DTLR probabilistic prediction, which is going to be submitted to *IEEE Transactions on Power Systems.* Dr. Mazhari as the co-author provided invaluable comments and suggestions during the development of the model. Dr. S. B. Ko provided priceless comments on presentation of the proposed model. I developed and implemented the model, performed the simulation, and analyzed the results, along with the paper write-up. The conclusions of this thesis are provided in Chapter 8-.

2- A NOVEL MULTI-STEP SHORT-TERM WIND POWER PREDICTION FRAMEWORK BASED ON CHAOTIC TIME SERIES ANALYSIS AND SINGULAR SPECTRUM ANALYSIS[3]

Decomposition methods are widely applied as a prestage of wind power prediction (WPP) to reduce the prediction errors caused by the nonstationarity and nonlinearity of wind power time series (TS); however, they cannot address the issues posed by the chaotic behavior of wind power TS. This paper, therefore, proposes a novel decomposition approach to take the chaotic nature of wind power TS into account and to improve WPP accuracy. In this decomposition approach, as a primary step, the wind power TS is separated into several components with different time-frequency characteristics (scales) by means of ensemble empirical mode decomposition. Chaotic TS analysis is then applied to determine which components are chaotic, and then singular spectrum analysis (SSA) is applied thereto. This multi-scale SSA (MSSSA) can maintain the general trend of chaotic components, which become smoother by eliminating extremely rapid changes with low amplitudes, and thus several steps ahead WPP with higher accuracy can be realized. Following the proposed decomposition, a novel short-term WPP method comprised of localized direct and iterative prediction is proposed to perform multi-step prediction for the chaotic and nonchaotic components of MSSSA, respectively. The proposed framework is finally validated using historical data related to overall wind power generation for Alberta (Canada), the Sotavento wind farm (Spain), and Centennial wind farm in Saskatchewan (Canada).

[3] [12] N. Safari, C. Chung, and G. Price, "A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis," IEEE Trans. Power Syst., vol. 33, no. 1, pp. 590-601, Jan. 2018.

2-1 Introduction

Duo to high intermittency and non-dispatchability, large-scale wind power integration creates new challenges in power system operation, but these can be significantly alleviated by short-term wind power prediction (WPP) [35]. Existing WPP methods are mainly based on three types of prediction models: physical models [36], statistical models [37, 38], and hybrid models (combining aspects of physical and statistical models) [39, 40]. Physical models are based on detailed physical descriptions of terrain and wind farm layout as well as simulation results obtained from numerical weather prediction (NWP), while statistical models only require time series (TS) of historical data such as wind speed and wind power [38]. Due to the high computational cost of NWP, physical models based on NWP are not preferred for short-term WPP ($\leq$ 6-h ahead) [35]. Because hybrid models also require results of NWP, their application to short-term WPP is also limited. In contrast, statistical WPP models have lower cost and complexity and therefore can be used for shorter-term WPP [37, 38].

Generally, statistical WPP can be classified into two types: indirect and direct prediction. In indirect prediction, wind speed values are first predicted and then mapped into wind power by a parametric or non-parametric wind power curve (e.g., [38, 39, 41, 42]). An accurate wind power curve requires data for several environmental and meteorological variables [43], but such variables have not been considered in indirect prediction [44]. Consequently, the indirect prediction has limited accuracy. In direct prediction, wind power is directly predicted from historical data (e.g., [37, 45, 46]), resulting in greater accuracy [44, 47].

Most recent direct statistical WPP methods [37, 46, 48-50] decompose the TS into several components with different characteristics [51] as a pre-stage of prediction; this remedies the

inaccuracy originating from the non-linearity and non-stationarity of the original TS. In [46], a morphological-based decomposition approach for separating wind power TS into a mean trend component and a stochastic component is proposed. In [37, 49], the effectiveness of the wavelet transform (WT) decomposition method is assessed. Recently, a few studies in WPP have employed empirical mode decomposition (EMD) variants [48, 50]. Generally, WT decomposition requires some prior knowledge and assumptions about the TS to find an appropriate mother wavelet, while EMD is a data-driven heuristic method that does not need initial assumptions about TS shape prior to decomposition [52]. A comprehensive comparison of the performance of EMD vs. WT variants shows the superiority of the former [50]. Ensemble EMD (EEMD) has the best performance in terms of improvement in WPP accuracy compared to other EMD variants and WT [50]. Additionally, the low computational complexity of EEMD has been theoretically demonstrated [53], confirming its suitability in terms of a low computational burden for short-term WPP. Although decomposition methods can deal with the non-linearity of the components to a large extent, the components obtained from wind power TS can still be of a chaotic nature, which can lead to prediction errors [54]. In a chaotic component, some subcomponents highly fluctuate in unpredictable patterns with low amplitudes [55, 56]. Singular spectrum analysis (SSA) can remove highly fluctuating and low amplitude variations from the chaotic TS, and therefore significantly improve the accuracy of the prediction [56]; however, this has not yet been applied to WPP as a supplementary tool for decomposition processes to handle the chaotic behavior of wind power TS.

After decomposition, various prediction methods can be used for WPP. Machine learning tools including neural networks (NNs) and support vector machines (SVMs) have recently been employed for this purpose due to their efficiency in predicting non-linear TS (e.g., [46, 57]). In [57], a comparative study of the conventional statistical approach known as autoregressive

integrated moving average (ARIMA) and NNs shows the superiority of the latter approach. Recently, a modified SVM, known as least squares-SVM (LSSVM), was successfully employed in WPP [46, 58]. LSSVM demonstrates better performance than SVM and NNs in terms of computational burden, simplicity, and the probability of convergence to global minima [58].

For all of the methods mentioned above, the prediction model should first be trained. Global [58] and localized [46] training are two possible training procedures that can be utilized. Localized prediction with an on-line training process, which makes use of historical data points that are very similar to the most recent data points, has superior performance in terms of prediction accuracy compared to global fitting approaches [46, 59]. Localized prediction focuses only on predicting the upcoming output from the current data point; hence, the model should be kept updated by historical data points with the smallest Euclidean distance to the most recent data points.

According to prediction output, statistical WPP models proposed in the literature can be categorized into probabilistic prediction [60], prediction interval [45], and deterministic prediction [46-52]. Probabilistic prediction and prediction interval are the extension of deterministic WPP to model the uncertainty in prediction [61]. Having information about WPP uncertainty is very useful. However, system operators need probabilistic or interval WPP results accompanied by deterministic WPP to make an optimal decision [62]. Deterministic WPP can be interpreted as the most likely wind power [27]. The application of deterministic WPP has been reported for short-term power system security [35], various unit commitment strategies [63], energy storage sizing [26], etc. In addition, a deeper understanding of the chaotic behavior of TS, such as wind power, can significantly reduce the prediction uncertainty and make the predicted value closer to the actual value [59]. For the aforementioned reasons, this paper aims at developing deterministic WPP

framework and improving prediction accuracy by proposing a deterministic chaotic time series analysis-based WPP.

Multi-step WPP is crucial for multi-step optimization problems such as unit commitment [64]. Several machine learning algorithms can be used in multi-step prediction; however, only a limited number of studies have been carried out for short-term multi-step WPP [46]. Multi-step prediction can be divided into direct prediction and iterative prediction. In the latter approach, the predicted values from previous steps are used to predict the next step; in the former approach, the predicted values at different steps are only calculated based on historical data. Iterative prediction is appropriate for predicting multi-step non-chaotic TS while multi-step direct prediction is suitable for chaotic TS [59, 65]. Indeed, iterative prediction can lead to significant accumulative error in predicting chaotic components [59]; however, hitherto no WPP has employed direct and iterative prediction methods to predict chaotic and non-chaotic components, respectively.

From the above discussion, the main purpose of this paper is to consider both the chaotic and non-chaotic natures of wind power TS. Considering the distinguishing characteristics of chaotic and non-chaotic components in both decomposition and prediction stages can result in a significant increase in WPP accuracy. In this regard, a novel decomposition stage based on the proposed multi-scale singular spectrum analysis (MSSSA) is developed. The MSSSA is used to decompose the wind power TS. Utilizing MSSSA alleviates the issues related to wild patterns in wind power TS. MSSSA consists of three main building blocks: EEMD, chaotic TS analysis, and SSA. EEMD is first applied to wind power TS, resulting in decomposition of TS into several components (scales). Then, by employing chaotic TS analysis, chaotic components (scales) are detected. To increase the predictability of the chaotic components, the wildest and most unpredictable subcomponents in each scale (component) are recognized based on SSA. Thereafter, the chaotic

components are processed to make them smoother and more predictable. Hence, this MSSSA framework is different from the decomposition-based WPP models presented thus far. After the decomposition stage, a localized LSSVM-based framework is employed and a multi-step short-term WPP method is proposed that comprises localized direct and iterative prediction of chaotic and non-chaotic components, respectively. Using direct multi-step prediction helps to avoid the accumulated error in the prediction of chaotic components, while iterative multi-step increases the accuracy of the prediction for non-chaotic components. The proposed framework that comprises MSSSA and LSSVM is herein named MSSSA-LSSVM.

To see the effectiveness of the proposed WPP in the prediction of aggregated wind power generation, as well prediction of wind farm generation, data from Alberta Electric System Operator (AESO), in Canada, Sotavento wind farm located in Spain, and Centennial wind farm, in Saskatchewan, Canada, are reported. Several well-established WPP models have been employed to confirm the superiority of the proposed MSSSA-LSSVM.

The remainder of this paper is organized as follows. Section 2-2 presents the proposed decomposition method. Section 2-3 introduces the proposed multi-step WPP. Data, evaluation indices, and simulation results of the proposed MSSSA-LSSVM framework are discussed in Section 2-4. Section 2-5 concludes the paper.

2-2 Decomposition Method

This section first provides a brief overview of EEMD, chaotic TS analysis, and SSA, and then introduces the proposed decomposition method in detail.

2-2-1 EEMD

EEMD is obtained from an ensemble of *NE* (number of generated trials from combinations of the original TS and white noises) EMDs. Hereafter, we assume $x(nT_S)$ is the TS of wind power, *n* is the sample number with T_S sampling time, and *N* is the total number of available samples. Decomposing the wind power TS with EEMD can be summarized according to the following steps. A detailed explication can be found in [66].

1) Add white noise ($w_j(nT_S), j = 1, \dots, NE$) to the original TS ($x(nT_S)$) to construct NE trials of TS ($x(nT_S)$) as follows:

$$x_j(nT_S) = x(nT_S) + w_j(nT_S) \tag{2-1}$$

then perform Steps 2-6 for every trial (j) to find corresponding EMD components.

2) Find the local maxima and local minima of $x_j(nT_S)$.

3) Interpolate the identified local maxima and minima points to find the upper envelope ($UE_j(nT_S)$) and lower envelope ($LE_j(nT_S)$) of $x_j(nT_S)$ respectively.

4) Calculate the difference between $x_j(nT_S)$ and the mean of $UE_j(nT_S)$ and $LE_j(nT_S)$ as follows:

$$m_j(nT_S) = x_j(nT_S) - \frac{UE_j(nT_S) + LE_j(nT_S)}{2} \tag{2-2}$$

5) Repeat Steps 2-4 with $m_j(nT_S)$ instead of $x_j(nT_S)$, until $\frac{UE_j(nT_S)+LE_j(nT_S)}{2} \leq \varepsilon$ (where ε is the acceptable error and should be close to zero), then keep $c_j^1(nT_S) = m_j(nT_S)$ as the first EMD component of$x_j(nT_S)$. At this stage, the residue $r_j^1(nT_S) = x_j(nT_S) - c_j^1(nT_S)$ is then calculated.

6) Given $c_j^i(nT_S)$ and r_j^{i-1}, for $i > 1$ then $c_j^{i+1}(nT_S)$ can be calculated by setting $r_j^i(nT_S) = r_j^{i-1}(nT_S) - c_j^i(nT_S)$ and repeating Steps 2-5 with $x_j(nT_S)$ replaced by $r_j^i(nT_S)$. Theoretically, the above iteration should continue until the residue becomes monotonic; however, different stoppage criteria are required to find the appropriate number of necessary components [51, 66]. In this paper, the number of required components (M) is selected based on a tradeoff between computational burden and monotonicity. Based on this process, $x_j(nT_S)$ can be decomposed into different EMD components as follows:

$$x_j(nT_S) = \sum_{i=1}^{M} c_j^i(nT_S) + r_j^M(nT_S) \tag{2-3}$$

7) Calculate the components of EEMD ($c^i(nT_S)$) by averag-ing the respective components ($c^i(nT_S)$) of all trials.

Finally, $x(nT_S)$ can be reconstructed as follows:

$$x(nT_S) = \sum_{i=1}^{M} c^i(nT_S) + r^M(nT_S) \tag{2-4}$$

where $r^M(nT_S)$ is the residue and calculated by averaging $r_j^M(nT_S)$, $j = 1, \dots, NE$. Notably, the amplitude of the residue is very small and hence is negligible in WPP.

2-2-2 Chaotic TS Analysis

Chaotic TS has wild and non-periodic behaviors [59] and its chaosity can be determined by its corresponding maximal Lyapunov exponent (MLE) (Appendix A). To find the MLE, the TS must first be embedded in a multidimensional space to form the trajectory matrix. This can be done with the method of delays (MOD) [67]. In such an embedding, each point in multidimensional space is

a vector whose components are the delayed version of the TS. In this multidimensional space, Lyapunov exponents determine the exponential convergence or divergence of nearby points, and MLE is the largest Lyapunov exponent that shows the largest convergence or divergence. If the sign of MLE is positive, then the TS is recognized as a chaotic TS; otherwise, it is non-chaotic. A detailed discussion of the MLE, as well as the program for calculating it, is available in [67].

2-2-3 SSA

A TS is comprised of several intrinsic components with different time-frequency characteristics [55]. SSA is a non-parametric spectral extraction method that can be used to separate the general trend, fluctuations, and noise components in a TS. First, to find components of the TS by means of SSA, the TS is mapped into a multidimensional space by the construction of the trajectory matrix. Then, eigenvalues and corresponding eigenvectors of the trajectory matrix are calculated by means of singular value decomposition. Using eigenvectors, the components in the trajectory matrix corresponding to every eigenvalue can be extracted. Next, in a grouping stage, the components which share some similarity are grouped together. In the last step, using diagonal averaging [55], the grouped components should be mapped from reconstructed multidimensional space to form a TS.

The smallest eigenvalues in SSA correspond to components with the smallest amplitudes, highest fluctuations, or noise, which are called non-informative (as they do not provide useful information and distort the main trend of the TS); the largest eigenvalues (dominant eigenvalues) are related to the highest amplitudes and smallest variations (i.e., the general trend of the TS) [55].

Using the advantages of SSA, the components obtained from SSA can be grouped into informative and non-informative groups. In other words, SSA can reconstruct the TS with

informative components (large eigenvalues), thereby increasing the prediction accuracy. Further information on SSA can be found in [56].

2-2-4 Proposed MSSSA Decomposition Method

Figure 2-1 shows the flowchart of the proposed MSSSA decomposition method for obtaining its *i*th component. The detailed procedure of the proposed method is also provided as follows:

1) Apply EEMD to find components related to wind power TS by following the procedure described in Section 2-2-1.
2) Identify the chaotic and non-chaotic components based on chaotic TS analysis as explained in Section 2-2-2.
3) Carry out SSA as described in Section 2-2-3 for any chaotic components, while non-chaotic components remain unchanged. To perform SSA for a chaotic component, the corresponding trajectory matrix by means of MOD is constructed. If $c^i(nT_S)$ in Eq. ((2-4) is a chaotic component, the corresponding trajectory matrix can be defined as

$$\boldsymbol{C}^i = [C_1^i \quad C_2^i \quad \cdots \quad C_l^i \quad \cdots \quad C_m^i]_{d\times m}^T \tag{2-5}$$

where d is the number of lags (or embedding dimension) and its appropriate value [54] can be found based on Cao's method [68]. Cao's method is a computationally efficient way to find the minimum number of time-delay elements (i.e., the number of lags) required to represent the TS in a multidimensional space if every row of the trajectory matrix in Eq. (2-5) is considered as a point in that space (Appendix B). In Eq. (2-5), $m = N - d + 1$and C_l^i is

$$C_l^i = \left[c^i(lT_S) \quad c^i\big((l+1)T_S\big) \quad \cdots \quad c^i\big((l+d-1)T_S\big)\right]_{1\times d}^T \tag{2-6}$$

4) Find eigenvalues and the corresponding eigenvectors for trajectory matrix $\boldsymbol{C}^i$. Here, the auto-correlation of $\boldsymbol{C}^i$ should be first constructed as follows [55]:

$$\boldsymbol{A}^i = \left(\boldsymbol{C}^i_{m\times d}\right)^T \boldsymbol{C}^i_{m\times d} \tag{2-7}$$

where $\boldsymbol{A}^i$ is a $(d \times d)$ matrix. Afterward, eigenvalues of $\boldsymbol{A}^i$, $\left(\lambda^i_1, \dots, \lambda^i_d\right)$, and its corresponding eigenvectors $\left(U^i_1, \dots, U^i_d\right)$ can be found and are equal to those of $\boldsymbol{C}^i$.

5) Discard the non-informative and wild subcomponents from the chaotic components. The eigenvalues are first ordered from largest to smallest. According to Section 2-2-3, non-informative and high-frequency variations with low amplitudes result from subcomponents related to small eigenvalues. An analysis of historical data for different case studies conducted for this paper showed those eigenvalues that are smaller than one-tenth of the largest eigenvalue can be removed so that the non-informative subcomponents of the chaotic component are discarded.

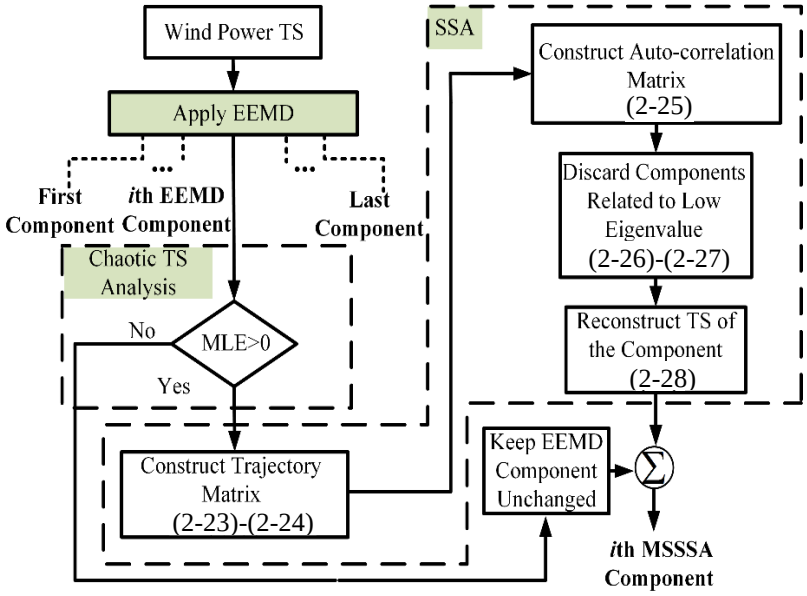

Figure 2-1: Flowchart of the MSSSA method for obtaining the *i*th component.

6) Find the trajectory matrix corresponding to dominant eigenvalues. For r dominant eigenvalues, the trajectory matrix related to U_j^i $(j = 1, \dots, r)$ can be obtained as follows:

$$\boldsymbol{C}_j^i = \boldsymbol{C}_{m\times d}^i U_{j(d\times 1)}^i \left(U_{j(d\times 1)}^i\right)^T \tag{2-8}$$

7) Construct the refined trajectory matrix. By adding all the reconstructed matrices corresponding to dominant eigenvalues in Eq. (2-8), the original trajectory matrix $\boldsymbol{C}^i$ can be refined by removing information related to small eigenvalues:

$$\boldsymbol{C}_{\mathrm{MSSSA}}^i = \boldsymbol{C}_1^i + \boldsymbol{C}_2^i + \cdots + \boldsymbol{C}_r^i \tag{2-9}$$

8) Convert $\boldsymbol{C}_{\mathrm{MSSSA}}^i$ to the MSSSA component (c_{MSSSA}^i) corresponding to $c^i(nT_S)$ by means of diagonal averaging in Eq. (2-10) [55].

$$\boldsymbol{c}_{\mathbf{MSSSA}}^{\boldsymbol{i}}(nT_S) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} \boldsymbol{C}_{\mathrm{MSSSA}}^i(n-k+1,k) & 1 \le n \le d \\ \dfrac{1}{d}\displaystyle\sum_{k=1}^{n} \boldsymbol{C}_{\mathrm{MSSSA}}^i(n-k+1,k) & d < n < m \\ \dfrac{1}{N-n+1}\displaystyle\sum_{k=n-m+1}^{d} \boldsymbol{C}_{\mathrm{MSSSA}}^i(n-k+1,k) & m < n < N \end{cases} \tag{2-10}$$

After all MSSSA components are obtained, the original wind power TS can be then reconstructed as follows:

$$x(nT_S) = \sum_{i=1}^{M} \boldsymbol{c}_{\mathbf{MSSSA}}^{\boldsymbol{i}}(nT_S) + r_{MSSSA}(nT_S) \tag{2-11}$$

2-3 Prediction Method

LSSVM is a powerful machine learning approach that can be used as a main integrant in WPP [46, 58]. A detailed description of LSSVM can be found in [69]. This section briefly explains the LSSVM and then introduces the proposed WPP.

2-3-1 LSSVM

In the LSSVM, the predicted value (y) corresponding to the input vector, also called the test vector (X_{Test}), can be calculated as follows [69]:

$$y = \boldsymbol{w}^T \varphi(X_{Test}) + b \tag{2-12}$$

where $\boldsymbol{w}$ and b are the weight vector and bias, respectively. φ is a feature vector and depends on the kernel function and its parameters [70]. There are various kernel functions, such as linear kernel, polynomial kernel, Gaussian kernel, and so on. Among all kernel functions, the Gaussian kernel has the best performance for wind speed prediction [70] and so is employed herein. The width of the Gaussian kernel (σ) must be determined. In addition, $\boldsymbol{w}$ and b should also be found through optimization.

The values of $\boldsymbol{w}$ and b can be found by solving the following optimization problem with a training dataset comprised of K training input vectors X_i $(i = 1,2, \dots, K)$ and their corresponding training output y_i:

$$C = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{i=1}^{K}\varepsilon_i^2 \tag{2-13}$$

$$\text{s.t. } \boldsymbol{w}^T\varphi(X_i) + b = y_i - \varepsilon_i$$

where C is the cost function (to be minimized) and K and γ are the number of training vectors and a regularization parameter, respectively. ε_i is the error of the model related to the ith training input vector and its output. Here, γ, which controls the tradeoff between bias and variance of the model, should be determined [70].

To solve the optimization problem in Eq. (2-13), γ and σ, which are called hyperparameters, should be selected through optimization before training the LSSVM. In this regard, a coupled simulated annealing (CSA) [71] is first performed to find a starting point for the simplex algorithm, and then simplex is performed as a fine-tuning step. Employing CSA-simplex in tuning regularization and kernel function parameters in LSSVM has been successfully achieved (e.g., [72]). Notably, to make optimization of hyper-parameters robust, cross-validation (CV) [69] is employed herein. CV is a statistical technique to prevent over-fitting as well as the dependency on datasets for finding parameters. In this paper, *k*-fold CV is used and the original training set is randomly divided into *k* equal-sized subsets. A subset is chosen as a validation set and the remainder are used as the training set. The LSSVM is then trained, and $\boldsymbol{w}$ and *b*, which suitably fit the LSSVM to the training set, are calculated by using different combinations of hyperparameters based on the CSA. For each chosen combination of hyper-parameters, the trained LSSVM is evaluated with the validation set by means of an evaluation index (such as mean square error or mean absolute error). The process is repeated *k* times, with each of the subsets used once as the validation set. The average value of the evaluation index is calculated. The CSA process is repeated until the predetermined acceptable evaluation index value is satisfied. The optimized hyperparameters are then used as the starting point for simplex for fine-tuning the hyper-parameters. After the hyper-parameters are determined, Eq. (2-13) can be simply solved by a Karush-Kuhn-Tucker approach to find $\boldsymbol{w}$ and *b*.

2-3-2 Proposed Multi-Step WPP

In this proposed WPP, the localized direct multi-step prediction is employed for chaotic components while iterative multi-step prediction is used for non-chaotic components.

To predict multi-step wind power values for, say, *p*-steps ahead and related to a chaotic component, the direct multi-step prediction model comprises *p* LSSVMs. Each LSSVM is trained to predict a specific step, but the predicted value in each step is not used to predict other steps. Assume that, at time NT_s, the procedure for predicting the next *p* values for the *i*th component of MSSSA ($c^i_{\mathrm{MSSSA}}(nT_S)$) in (2-10) is summarized as follows:

1) Find the test vector and training datasets. The test vector is used as the input for the prediction model. The test vector is identical for all of the *p* LSSVMs. To find the required test vector, the trajectory matrix corresponding to the *i*th chaotic component ($c^i_{MSSSA}(nT_S)$) should be found. To find the trajectory matrix, a similar approach to Step 3 in Section 2-2-4 should be employed by using the MSSSA components instead of the original components of wind power TS in Eq. (2-5). The last row of trajectory matrix (m), which contains the most recent data points, is selected as the test vector, as follows:

$$X_{Test} = \left[c^i_{\mathrm{MSSSA}}\big((m)T_S\big) \quad \cdots \quad c^i_{\mathrm{MSSSA}}\big((m+d-1)T_S\big)\right] \tag{2-14}$$

where *m* is defined as in Section 2-2-4. For each LSSVM, the training dataset consists of input vectors and their corresponding outputs. These training input vectors and outputs are used to tune LSSVM parameters by solving (2-12). Based on the trajectory matrix related to $c^i_{\mathrm{MSSSA}}(nT_S)$, the matrix $\boldsymbol{X}$ in (2-15) can be found.

$$\boldsymbol{X} = \begin{bmatrix} c^i_{\mathrm{MSSSA}}(1T_S) & c^i_{\mathrm{MSSSA}}(2T_S) & \cdots & c^i_{\mathrm{MSSSA}}(dT_S) \\ c^i_{\mathrm{MSSSA}}(2T_S) & c^i_{\mathrm{MSSSA}}(3T_S) & \cdots & c^i_{\mathrm{MSSSA}}\big((d+1)T_S\big) \\ \vdots & \vdots & \vdots & \vdots \\ c^i_{\mathrm{MSSSA}}(rT_S) & \cdots & \cdots & c^i_{\mathrm{MSSSA}}\big((r+d-1)T_S\big) \end{bmatrix} \tag{2-15}$$

Each row of matrix **X** is a candidate for training input. For prediction step *l*, where *l*=1,…*p*, the number of rows of matrix **X** is $r = m - l$; this is because training output does not exist for

rows in the trajectory matrix that are greater than $m - l$. The column vector with all of the training outputs corresponding to matrix $\boldsymbol{X}$ is as follows:

$$Y_l = \left[c^i_{\mathrm{MSSSA}}\big((d+l)T_S\big) \quad c^i_{\mathrm{MSSSA}}\big((d+l+1)T_S\big) \quad \cdots \right.$$
$$\left. c^i_{\mathrm{MSSSA}}\big((d+l+i)T_S\big) \quad \cdots \quad c^i_{\mathrm{MSSSA}}\big((r+d+l-1)T_S\big)\right]^T \tag{2-16}$$

where the *i*th row constitutes the training output vector corresponding to the *i*th training row vector in matrix $\boldsymbol{X}$. However, a localized prediction model only requires the training vectors that have smaller Euclidean distances to the test vector. Therefore, by means of the k-nearest neighbors algorithm (KNN), and using Eqs. (2-15) and (2-16), for the *l*th LSSVM the training input matrix $\boldsymbol{X}_{KNN(K\times d)}$ contains *K* row vectors of matrix $\boldsymbol{X}$ that have the closest Euclidean distance to the test vector in Eq. (2-14). Consequently, the corresponding training output to $\boldsymbol{X}_{KNN(K\times d)}$ can be represented by $Y_{lKNN(k\times 1)}$.

2) Train the LSSVM model. Using the training input vectors and training output, and based on Section 2-3-1, kernel function parameters, the weight vector, and the bias corresponding to each LSSVM can be found.

3) Predict the value of the next *p*-step using the *p* trained LSSVMs and the obtained test vector. The LSSVM in step *l* is as follows:

$$\hat{c}^i_{\mathrm{MSSSA}}\big((N+l)T\big) = \boldsymbol{w}^T\varphi(X_{Test}) + b \tag{2-17}$$

As shown above, to predict *p*-steps ahead in direct multi-step prediction, *p* LSSVMs are trained separately by means of historical data and are independent of predicted values in previous steps. Unlike the direct approach, however, the predicted values at previous steps in iterative prediction are used as test and training datasets for the next steps.

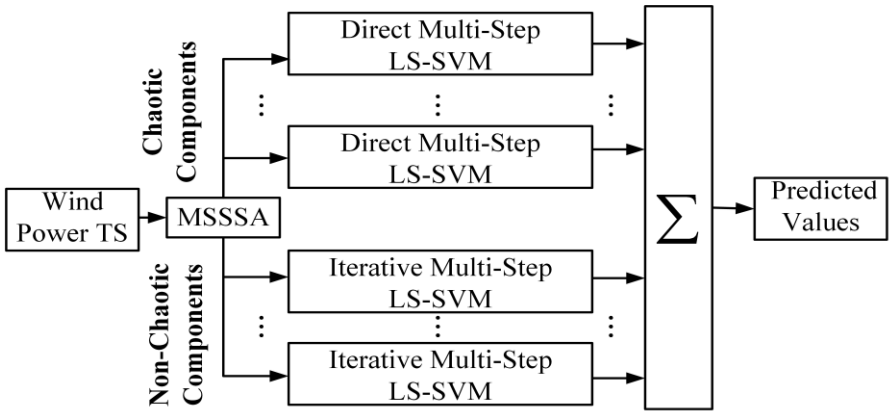

Figure 2-2: Flowchart of the proposed MSSSA-LSSVM framework.

Values predicted for each step for all chaotic and non-chaotic components are then respectively added to find the predicted values (i.e., predicted wind power) for the next p-step in the original wind power TS. The flowchart of the WPP framework (MSSSA-LSSVM) is provided in Figure 2-2.

2-4 Case Studies

2-4-1 Description of Datasets

Simulation studies are carried out to evaluate the performance of the proposed MSSSA-LSSVM. For this purpose, three sets are used: Set 1 is wind power generation data related to AESO [62], located in Canada; Set 2 is the historical wind power generation data for a wind farm in Sotavento [73], located in Spain; and Set 3 is the historical wind power generation of wind power in Centennial wind farm, located in Saskatchewan, Canada. The main reason for choosing these three different sets is to validate the performance of the MSSSA-LSSVM for different types of wind power generation profiles.

As case studies, WPP for Set 1 is carried out from February to March 2012, for Set 2 it is carried out from Aug. 2011 to Oct. 2011, and for Set 3 is carried out for October 2016.

The historical wind power generation data in all Sets are sampled every 10 minutes. In this study, the multi-step WPP was conducted with the step length of 10-min. Note that wind power generation data is the only data used for developing the WPP. More studies can be carried out to find other effective exogenous variables that can be employed for predicting both chaotic and non-chaotic components.

2-4-2 Benchmark Models for Numerical Comparison

For comparison, the proposed MSSSA-LSSVM framework is compared with a persistence model (PM) commonly used as a benchmark model for WPP. As LSSVM and EEMD are employed as building blocks in the proposed framework, the performance of LSSVM [58] and EEMD-LSSVM [51] benchmarks are also compared to the proposed WPP model. A radial basis function neural network (RBFNN) is a mature ANN that is widely used in WPP. Hence, RBFNN is used here as another benchmark model. Similar to the WPP model in [46], the MSSSA-LSSVM is based on localized prediction. Hence, for further comparison, the performance of the proposed framework is compared with a recently developed localized WPP based on a mean trend detector (MTD) and a mathematical-morphology-based wind local prediction (MLP) (MTD/MLP) [46]. It is also useful to compare the performance of the proposed method against a standard nonlinear regression benchmark model such as generalized additive model (GAM) [74]. Thus, a well-trained GAM has also been used as the last benchmark model.

It is noteworthy to mention that for a fair comparison, LSSVM and EEMD-LSSVM are locally trained and their parameters and coefficients have been kept updated by performing training for

each prediction step, like MSSSA-LSSVM. In fact, without this consideration, the improvement in prediction accuracy of MSSSA-LSSVM in comparison to LSSVM and EEMD-LSSVM is more significant than what is reported in this section.

The results of the simulation are limited to prediction horizons up to 6 hours ahead. Prediction horizons up to 6 hours ahead are of great importance for electricity market and power system security. While, hybrid and physical WPP models are preferred for longer prediction horizons [35], which are beyond the scope of the current study.

2-4-3 Evaluation Indices

A variety of well-defined and widely accepted evaluation indices are used. These include the normalized mean absolute error (NMAE), the normalized root-mean-squared error (NRMSE), and the normalized absolute error (NAE). NMAE and NAE_i measure the accuracy of the predicted values as compared to actual values, and NRMSE estimates the variation and degree of bias in the predicted values. NRMSE also gives a measure of large errors. NMAE, NRMSE, and NAE_i can be defined, respectively, as

$$NMAE = \frac{1}{L}\sum_{i=1}^{L}\frac{|x_i - \hat{x}_i|}{P_{Inst.}} \times 100\%, \tag{2-18}$$

$$NRMSE = \frac{1}{P_{Inst.}} \times \sqrt{\frac{1}{L}\sum_{i=1}^{L}(x_i - \hat{x}_i)^2} \times 100\%, \tag{2-19}$$

$$NAE_i = \frac{|x_i - \hat{x}_i|}{P_{Inst.}} \times 100, \tag{2-20}$$

where L is the number of test points, $P_{Inst.}$ is the installed wind power capacity of the case study, x_i is the actual value, and $\hat{x}_i$ is the corresponding predicted value. For Sets 1-3, $P_{Inst.}$ values are 967 MW, 17.5 MW, and 150 MW, respectively.

2-4-4 Numerical Results and Analysis

2-4-4-1 Length of Training Data and Computation Time

The simulation was done on a Windows 7 PC with a 3.4 GHz Intel Core i7 CPU and 16 GB of memory. The WPP was run on MATLAB 2014b. A range of lengths for the training dataset was evaluated to find a suitable length of the training dataset with 6-hour ahead WPP carried out for Set 1.

The NRMSE values, as well as computation time for different training dataset lengths, are presented in Figure 2-3. Increasing the length of the training dataset can enhance WPP accuracy and increase the computation time. In Figure 2-3, the trend of improvement in NRMSE by increasing the training dataset length to more than 10 days is very gentle, while the slope of

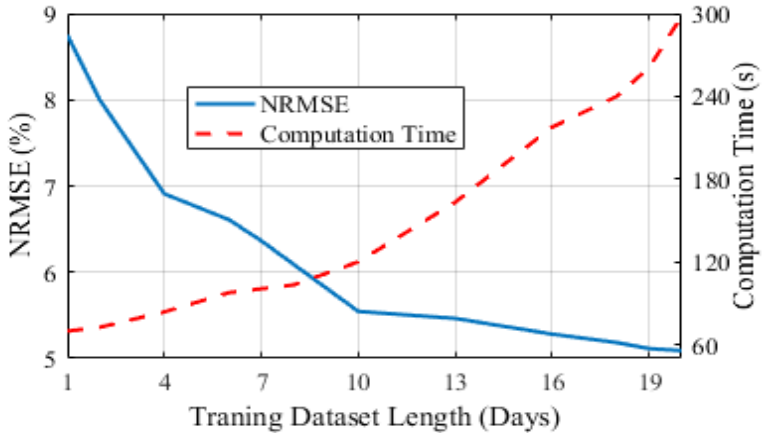

Figure 2-3 NRMSE and computation time variation for different training dataset lengths.

incrementally increasing computation time is severe. To obtain an appropriate WPP accuracy with an affordable computation burden, 10 days of historical data prior to the predicting moments was used as the training dataset in this study.

In 6-hour ahead WPP, with 10-day training dataset, the training and prediction for the PM, RBFNN, LSSVM, EEMD-LSSVM, and MSSSA-LSSVM took 0.0050, 34.9201, 30.2298, 110.8001, and 120.0108 s, respectively. Although the computation time of the proposed prediction model is higher than the benchmark prediction models, this computation time is acceptable considering the 6-hour WPP horizon.

2-4-4-2 Numerical Comparisons of WPP Models

WPP for 1-hour (6-step), 2-hour (12-step), 4-hour (24-step), and 6-hour (36-step) ahead prediction were carried out for Sets 1-3 and the results summarized in Tables 2-1-2-3, respectively.

From the NRMSE values in Tables 2-1-2-3, it can be easily observed that the proposed MSSSA-LSSVM outperforms all of the benchmark models; hence, it can be concluded that MSSSA-LSSVM successfully eliminates large prediction errors. It is clear from the NRMSE related to EEMD-LSSVM, which has the best performance among benchmark models, that increasing the number of prediction steps significantly increases the NRMSE until the performance is even worse than PM. This observation is an additional confirmation of the importance of considering the chaotic behavior of wind power TS and how the existence of large errors results from accumulated prediction error as the number of prediction steps increases in EEMD-LSSVM. To depict the impact of improvement in NRMSE more clearly, the NAEs related to different WPP models for 6-h ahead (36 samples) are presented in Figures 2-4 -2-6. The improvement in error reduction by means of the proposed WPP model can be observed in these figures. Benchmark models lead to

large errors (high NRMSE) in prediction, while, the MSSSA-LSSVM is more robust with respect to prediction and limits prediction errors.

The NMAE for different prediction models are also shown in Tables 2-1-2-3. The NMAE for MSSSA-LSSVM is less than all prediction models for all observations except for three cases in which the EEMD-LSSVM has about 2% improvement compared to the MSSSA-LSSVM. However, by increasing the number of prediction steps, the MSSSA-LSSVM performs significantly better in terms of NMAE. Moreover, for those very few cases in which the EEMD-LSSVM provides a slightly lower NMAE compared to the MSSSA-LSSVM, its NRMSE is significantly worse than the MSSSA-LSSVM. Considering both NMAE and NRMSE for all case studies as well as the importance of reducing large errors, MSSSA-LSSVM can be concluded to have the best overall performance.

As a comparison, the histograms of the NAE for the proposed model and benchmark models are depicted in Figures 2-7-2-9. The percentage of NAE is divided into 10 bins and the figures highlight the reduction in prediction error. The largest proportion of prediction errors related to MSSSA- LSSVM lies in the first bin (NAE<10%). This result confirms the effectiveness of MSSSA-LSSVM in reducing large prediction errors. Table 2-4 shows the percentage of NRMSE and NMAE reduction compared with the PM for Sets 1-3 for one month. The significant improvement even with increasing prediction horizon indicates the need to consider the chaotic behavior of wind power TS in WPP.

Table 2-1: Performance evaluation of different prediction models for set 1 (%).

Method	1-h ahead (6-step)		2-h ahead (12-step)		4-h ahead (24-step)		6-h ahead (36-step)	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
PM	3.7136	2.3275	5.8915	3.7519	9.0445	6.0778	11.1534	7.6267
RBFNN	4.0971	2.6343	6.3300	4.1005	9.3550	6.2947	12.0060	8.0824
LSSVM	4.4329	2.8627	7.2480	4.7615	11.1493	7.6406	13.7940	9.5471
EEMD-LSSVM	2.9793	1.3419	3.9113	**1.6431**	8.5352	3.7225	13.5456	6.9815
MSSSA-LSSVM	**2.2375**	**1.2957**	**2.8880**	1.6543	**5.5206**	**3.3645**	**5.8514**	**4.0267**

Table 2-2: Performance evaluation of different prediction models for set 2 (%).

Method	Month	1-h ahead (6-step)		2-h ahead (12-step)		4-h ahead (24-step)		6-h ahead (36-step)	
		NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
PM	Aug.	5.3631	2.8342	6.8941	3.6823	8.8727	4.9713	9.9232	5.6647
	Sep.	4.7240	2.7712	5.8742	3.5647	7.4720	4.6145	8.4314	5.3941
	Oct.	7.6521	4.2648	9.1524	5.3317	11.3025	6.8641	12.5321	7.5326
RBFNN	Aug.	5.4490	2.9184	7.3508	4.0046	9.2790	5.3593	9.6805	6.2036
	Sep.	4.7302	2.9189	6.1443	3.8753	7.6890	4.8738	9.2236	5.9151
	Oct.	7.4801	4.4303	9.0022	5.6201	11.5695	7.2954	12.7602	8.2512
LSSVM	Aug.	5.7523	3.0421	7.2517	4.0448	9.4601	5.5634	9.8612	5.9732
	Sep.	5.1635	3.0345	6.5730	3.9823	8.1921	5.1236	9.0512	5.6501
	Oct.	8.7276	4.9148	9.8712	6.1302	12.3914	7.8742	13.3123	8.7325
EEMD-LSSVM	Aug.	3.2138	1.4686	5.3532	2.2612	9.7347	3.9248	7.1332	3.9887
	Sep.	2.5778	**1.3245**	5.8314	2.3247	11.9721	4.5415	12.1813	5.4832
	Oct.	4.9112	2.5197	7.8311	3.9236	10.6845	5.5132	15.6403	8.1245
MSSSA-LSSVM	Aug.	**2.7345**	**1.2813**	**3.7002**	**1.4221**	**5.5096**	**2.7797**	**5.9671**	**3.4262**
	Sep.	**2.5321**	1.3613	**2.7834**	**1.4132**	**5.5714**	**3.1215**	**6.3932**	**3.6113**
	Oct.	**4.2331**	**2.3912**	**6.5693**	**3.2452**	**8.4302**	**4.6278**	**9.4322**	**5.4665**

Table 2-3: Performance evaluation of different prediction models for set 3 (%).

Method	1-h ahead (6-step)		2-h ahead (12-step)		4-h ahead (24-step)		6-h ahead (36-step)	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
PM	9.7862	5.1654	13.3897	7.4448	17.6991	10.5855	23.5440	14.4854
RBFNN	11.0354	6.4792	14.2940	8.9314	17.9384	11.8037	22.1108	13.5694
LSSVM	11.3892	6.3665	13.7598	8.3654	18.3379	11.6879	21.8332	13.8207
EEMD-LSSVM	6.1853	**3.2891**	8.5135	4.2330	14.8693	7.3922	19.7464	10.7848
MSSSA-LSSVM	**5.9457**	3.3657	**6.9232**	**4.0859**	**8.2468**	**5.2468**	**9.5223**	**6.1952**

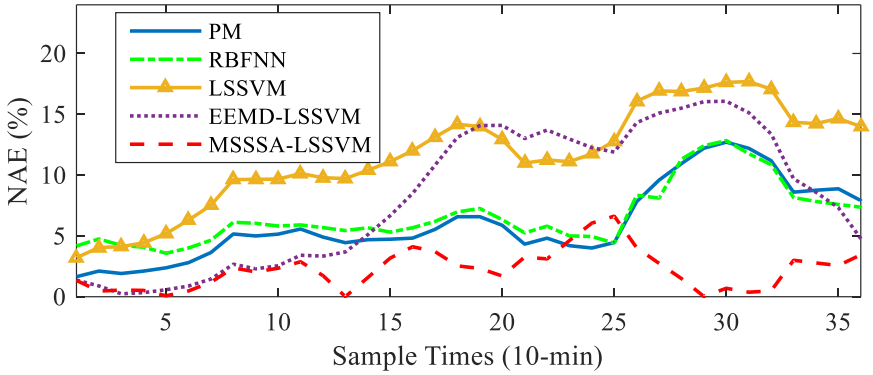

Figure 2-4: The NAE error corresponding to different 6-h ahead WPP for Set 1.

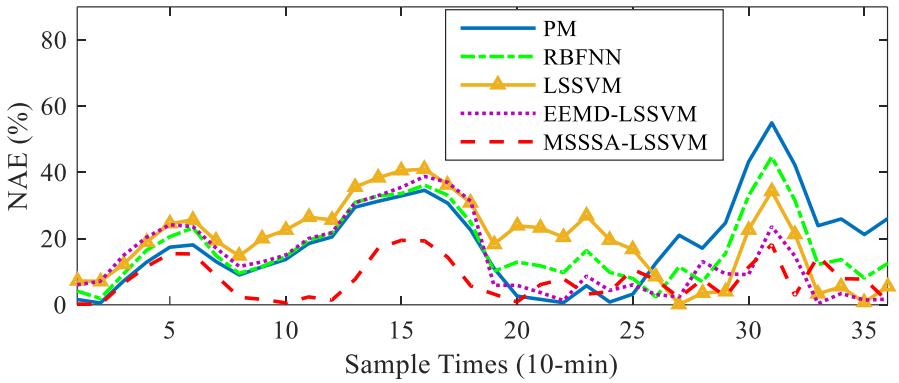

Figure 2-5: The NAE error corresponding to different 6-h ahead WPP for Set 2.

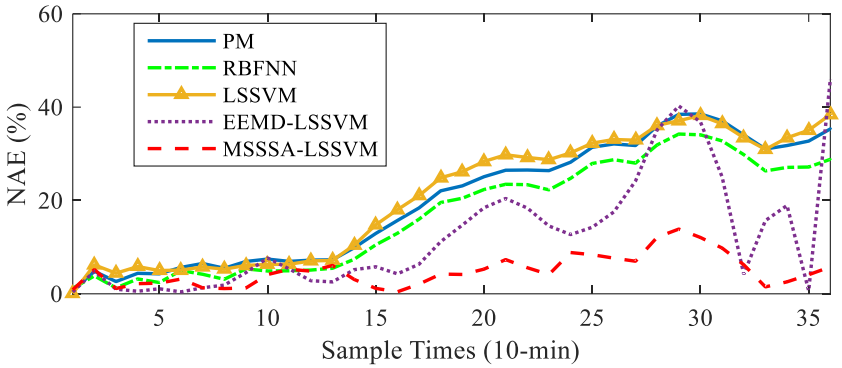

Figure 2-6: The NAE error corresponding to different 6-h ahead WPP for Set 3.

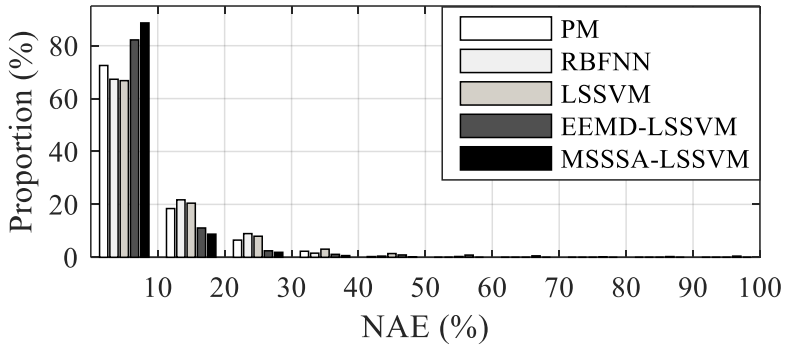

Figure 2-7: NAE distribution for 6-h ahead WPP in Set 1, Feb 2012.

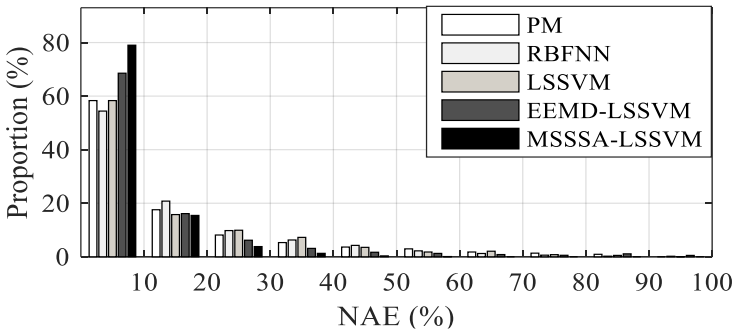

Figure 2-8: NAE distribution for 6-h ahead WPP in Set 2, Oct. 2011.

To closely compare the effectiveness of the proposed MSSSA against EEMD, the prediction

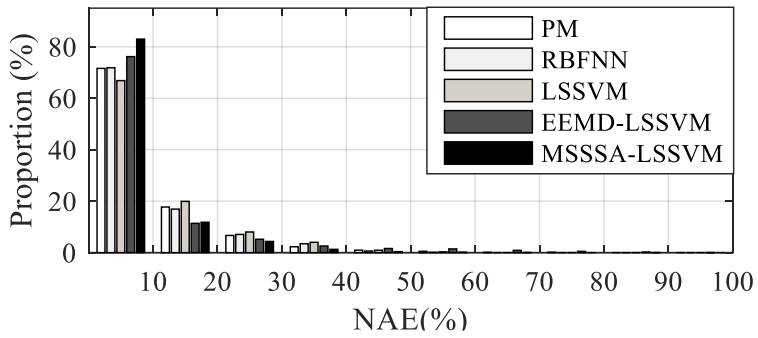

Figure 2-9: NAE distribution for 6-h ahead WPP in Set 3, Oct. 2016.

results for 36-step ahead prediction of one of the chaotic components in Set 3 for 72 samples are shown in Figure 2-10. This figure clearly indicates that employing MSSSA makes the training data smoother and hence, the suitable *K* nearest neighbors can be obtained. In contrast, due to wild changes in training data in EEMD-based prediction, the *K* nearest neighbors cannot properly train the prediction engine. In this regard, using EEMD-LSSVM for multi-step prediction results in large errors and consequently, a large NRMSE. In this case study, to further examine the

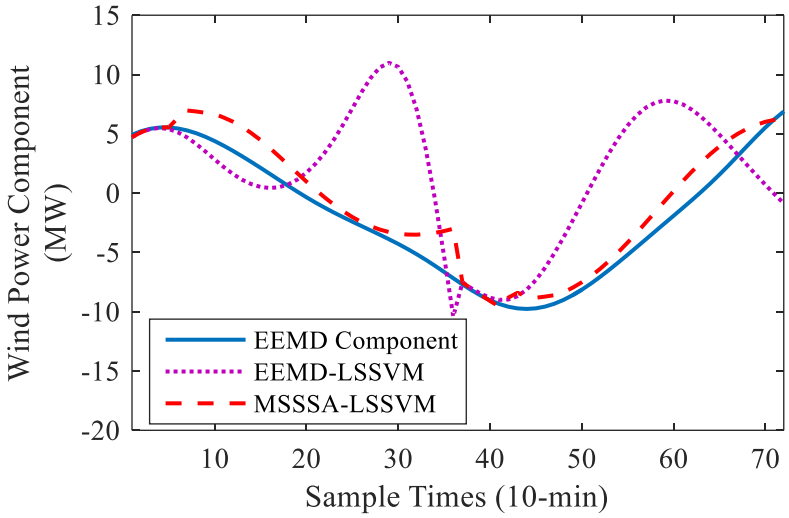

Figure 2-10: Comparing the effectiveness of MSSSA and EEMD in WPP.

Table 2-4: Percentage of NRMSE and NMAE reduction compared to persistence model

Set	6-step ahead		12-step ahead		24-step ahead		36-step ahead	
	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE	NRMSE	NMAE
1	44.33	44.33	51.12	55.91	38.96	44.64	47.54	47.20
2	49.01	54.79	46.33	61.38	37.90	44.85	39.87	65.33
3	39.24	34.84	48.29	45.14	53.40	50.43	59.55	57.23

effectiveness of the proposed MSSSA-LSSVM in comparison to EEMD-LSSVM, the performance of the MSSS-LSSVM in predicting the aggregated chaotic components, as well as the aggregated non-chaotic components have been separately evaluated. To conduct a meaningful comparison, in calculating NRMSE of the chaotic and non-chaotic components, the $P_{Inst.}$ in Eq. (2-19) is replaced with the maximum values of the aggregated actual chaotic components and aggregated actual non-chaotic components, respectively. The NRMSE values which are resulted from EEMD-LSSVM and MSSSA-LSSVM in predicting chaotic components are 17.2670% and 6.0478%, respectively. While the NRMSE in the prediction of the non-chaotic components for both EEMD-LSSVM and MSSSA-LSSVM is 5.4803%. Thus, due to a significant reduction in prediction error of chaotic component in MSSSA-LSSVM, the overall wind power prediction using MSSSA-LSSVM leads to considerably lower prediction error.

2-4-4-3 Further Numerical Comparisons with Localized WPP Models

As another comparison, a 5-h ahead WPP for Set 1 in March 2012, which has been used in [46], has been conducted. The NRMSE and NMAE of the MSSSA-LSSVM, are compared with the results of MTD/MLP [46]. To make the proposed WPP comparable with MTD/MLP WPP results reported in [46], for this case study similar to [46], 2 days of historical data prior to the prediction point is used as the training dataset. Quoting the results of MTD/MLP WPP from [46], the NRMSE and NMAE for 5-h ahead WPP using MTD/MLP are 12.974% and 9.2940%, respectively. While,

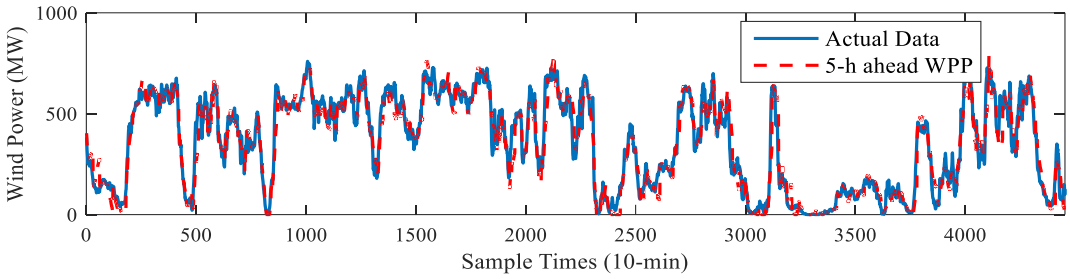

Figure 2-11: 5-h ahead WPP results of MSSSA-LSSVM in Set 1 for March 2012.

for MSSSA-LSSVM the NRMSE, with the same length of training data, are 7.8144% and 5.4683%, respectively. The result of WPP, using MSSSA-LSSVM, for Set 1, March 2012 is depicted in Figure 2-11. From this figure, it can be concluded that in the whole month the proposed WPP can effectively track the actual data. To closely investigate the performance of the proposed model, the close-up view of the WPP result for one day has been shown in Figure 2-12. This figure can validate the significant accuracy of WPP model is prediction wind power for in day which the wind power has a large range of variations, from approximately 100 MW to 600 MW. This WPP model can be a very helpful tool for power system operators in short-term decision-making applications.

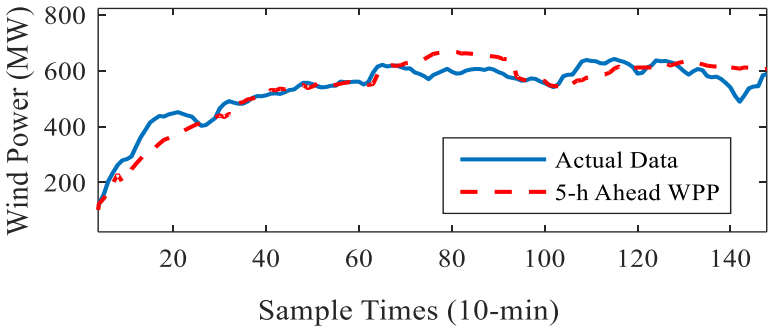

Figure 2-12: 5-h ahead WPP results of MSSSA-LSSVM in Set 1 for March 2012 for 25 hours

Table 2-5: Performance comparison of GAM and MSSSA-LSSVM

Prediction horizon	GAM		MSSSA-LSSVM	
	NRMSE	NMAE	NRMSE	NMAE
6-step ahead	5.7182	3.1378	2.7345	1.2813
12-step ahead	6.7947	4.1540	3.7002	1.4221
24-step ahead	9.7345	5.8139	5.5096	2.7797
36-step ahead	12.6565	7.0902	5.9671	3.4262

2-4-4-4 Further Numerical Comparisons with GAM

A well-trained GAM based WPP model has been developed and the results of prediction for Set 2 in August 2011 are reported in Table 2-5. Both NMAE and NRMSE caused by GAM are significantly larger compared to the proposed MSSSA-LSSVM. Furthermore, with increasing the prediction horizon the NRMSE and NMAE of GAM markedly increase, while the proposed MSSA-LSSVM provides a more robust prediction accuracy with increasing the prediction horizons.

2-5 Conclusion

A novel framework (MSSSA-LSSVM) for short-term multi-step WPP that can consider the chaotic nature of the components obtained from wind power TS in both decomposition and prediction stages is proposed. In the decomposition stage, the chaotic TS analysis identifies the chaotic components obtained from EEMD, and then the chaotic components become more predictable by eliminating high-frequency variations that are small in amplitude based on SSA theory. In the prediction stage, the proposed WPP employs localized direct and iterative prediction for chaotic and non-chaotic components, respectively. Simulation results validate the efficacy of the MSSSA-LSSVM for predicting aggregated wind power generation of a region, as well as wind

power generation in different locations for single wind farm power generation. The model significantly outperforms PM (a widely-used benchmark model) as well as other comparable models including RBFNN, LSSVM, EEMD-LSSVM, MTD/MLP, and GAM.

3- AN ADVANCED MULTISTAGE MULTI-STEP TIDAL CURRENT SPEED AND DIRECTION PREDICTION MODEL[4]

Non-stationarity and non-linearity of the tidal current speed (TCS) and tidal current direction (TCD) time series are among the main barriers for enhancing the TCS and TCD prediction accuracy. In this regard, this paper proposes an improved complete ensemble empirical mode decomposition adaptive noise (ICEEMDAN) which is employed to decompose the non-stationary TCS and TCD time series into several components (modes) with unique characteristics. Then, to capture the nonlinear pattern of TCS and TCD in different modes, several prediction engines based on least squares support vector machine (LSSVM) are developed. To modify the prediction error which occurs in predicting different components, a prediction modification stage based on a combination of extreme learning machines (ELMs) is utilized to reconstruct the final prediction values. The proposed TCS and TCD prediction model, named ICEEMDAN-LSSVM-ELM, has been evaluated using the data recorded from Shark river entrance, NJ. Performance of the proposed prediction model is compared with various well-developed benchmark models.

3-1 Introduction

Hydrokinetic tidal power resulting from extracting kinetic energy of tides has a huge potential to be an alternative for fossil fuel-fired generation [76]. As a result, the International Energy Agency (IEA) expects increase in tidal energy generation in near future [77].

[4] © 2018 IEEE. Reprinted, with permission from [75] N. Safari, B. Khorramdel, A. Zare, and C. Y. Chung, "An advanced multistage multi-step tidal current speed and direction prediction model," in *Proc. IEEE Electrical Power and Energy Conference (EPEC)*, 2017, pp. 1-6.

As the penetration of tidal power generation units will be increased, the uncertainty and variability of tidal power can cause several unprecedented power system operation challenges [78]. Accurate short-term prediction and uncertainty models of tidal power generation can be used as a key tool in alleviating the adverse impacts of tidal generation units on power system operation [79].

To predict tidal power, first, the effective tidal current speed (TCS) which flows into the area swept by tidal turbine blades needs to be predicted. Then, using the tidal turbine characteristics, the tidal power can be predicted [78]. The effective TCS can be calculated by TCS accompanied by tidal current direction (TCD). For this reason, several prediction models have been proposed for both TCS and TCD [13, 34, 79, 80]. As TCS and TCD time series are non-stationary and non-linear employing stationary and linear framework can result in large prediction errors [80].

Wavelet decomposition (WD) and empirical mode decomposition (EMD) variants have been considered as effective tools in decomposing non-stationary time series into several components with considerably lower non-stationarity [81, 82]. In [79], WD is utilized to decompose the TCS and TCD time series into several components; then, each component is individually predicted, and all the predicted components are reconstructed to find the predicted TCS and TCD in the next sample time. In [13], ensemble EMD (EEMD) is used to decompose the TCS and TCD into different components; then, the prediction is conducted on the extracted components. By adding some noise to the time series and constructing an ensemble of the original time series associated with noise, EEMD addresses the mode mixing issue of traditional EMD. In EEMD, first, by applying EMD on each noise added time series, EMD components are obtained. Then, averaging the EMD components related to noise added times series results in the find the EEMD components.

Despite promising advantages of EEMD in TCS and TCD, some aspects of EEMD need further considerations. Firstly, associating different realizations of the noise with the original time series can result in various extracted components for each noise associated time series; thus, the final components, obtained from EEMD, might vary for different sets of noise. Secondly, adding various noises can result in dissimilar number of EMD components; therefore, the final averaging process for finding EEMD components becomes problematic. In [83], a complementary EEMD (CEEMD) is proposed to circumvent the above-mentioned issues. In CEEMD, both positive and negative realizations of white noise are used to construct ensembles of the original time series. This approach can substantially eliminate the first issue; while the second issue remains unaddressed. In [84, 85], variants of EEMD, known as complete EEMD with adaptive noise (CEEMDAN) is proposed to alleviate the second issue. The proposed approach can deal with both issues, but there is no guarantee that the noises added to the time series eliminate each other in the final CEEMDAN components. However, due to employing various noises for constructing ensemble of noise associated time series, the final components, obtained from CEEMDAN, might vary every time that CEEMDAN is conducted. In this paper, to benefit from the advantages of both CEEMD and CEEMDAN in TCS and TCD prediction and address the drawbacks of CEEMDAN, an improved version of CEEMDAN, named ICEEMDAN is proposed.

Following the decomposition stage, prediction engines are employed to predict the corresponding value of different components in the next sample time. To address the nonlinearity of the time series, in [34] artificial neural network (ANN)-based prediction engine is used; however, ANN is prone to overfitting and being trapped in local minima during the process of parameters tuning [80]. In this regard, in [80] a TCS and TCD prediction model based on support vector machine (SVM) is proposed. SVM training involves quadratic programming which requires

a high computational burden [69]. Hence, an SVM variant, known as least squares support vector machine (LSSVM) has been proposed in [69]. In this powerful machine learning tool, the training process is done by solving linear equations, instead of quadratic programming. Among LSSVM kernels, the Gaussian kernel has superior performance in modeling nonlinear time series [70]. LSSVM has been successfully employed in renewable energy generation prediction [12, 13]. For this reason, LSSVM is selected as the prediction engine of the proposed model.

After Predicting TCS and TCD components through ICEEMDAN-LSSVM, the overall TCS and TCD predicted values need to be obtained. Conventionally, in decomposition-based prediction models, the sum of all the predicted components are considered as the final predicted value [51, 80]. However, aggregation of the predicted values of different components can result in erroneous prediction result. Besides, the prediction of each component is accompanied by some degrees of uncertainty. In this regard, we propose a prediction aggregation and modification stage which compensates the prediction error exists in the ICEEMDAN-LSSVM. To realize this stage, an extreme learning machine (ELM)-based stage is hybridized with ICEEMDAN-LSSVM. As the ELM is greatly impacted by the initialization, a combination of ELMs has been developed, and the well-trained ELMs are selected.

From the above discussion, proposing a hybrid TCS and TCD prediction model, named ICEEMDAN-LSSVM-ELM, is the main contribution of this paper. This prediction model benefits from ICEEMDAN in decomposition stage, LSSVM-based prediction engines in prediction stage, and an ELM-based prediction aggregation and modification stage. The proposed ICEEMDAN-LSSVM-ELM is tested, using the historical TCS and TCD datasets which are recorded in Shark River Entrance [86], one of the potential sites for tidal energy generation [87].

The remainder of this paper is organized as follows. Section 3-2 briefly illustrates the main building blocks of the proposed model. The proposed prediction model is expounded in Section 3-3. Data, evaluation, and simulation results of the proposed ICEEMDAN-LSSVM-ELM are discussed in Section 3-4. Section 3-5 concludes the paper.

3-2 Building blocks of the proposed prediction model

The general scheme of the proposed prediction model is depicted in Figure 3-1. As can be seen in this figure, ICEEMDAN, LSSVM, and ELM are three main building blocks of the proposed prediction model, and they are briefly elucidated in this section.

3-2-1 Proposed ICEEMDAN

The procedure of the applying the proposed ICEEMDAN to the TCS and TCD time series, $\{x(nT_s)\}_{n=1}^{N}$ where T_s is the sample time and N is the number of samples, is as follows:

1) Generate M different realizations of white noise with zero mean, $w_i(nT_s)$ where $i = 1, \dots, M$. Then, perform Steps 2 to 6 to find the first component of ICEEMDAN, $c^1(nT_s)$.
2) Find the M complimentary white noises $w_i(nT_s)$, where $i = M+1, \dots, 2M$, to $w_i(nT_s)$, where $i = 1, \dots, M$ as follows:

$$w_{\mathrm{i}}(nT_S) = -w_{i-M}(nT_S) \tag{3-1}$$

3) Find the $2M$ ensembles of $\{x(nT_s)\}_{n=1}^{N}$ aggregated with $w_i(nT_s)$, as follows:

$$x_i(nT_S) = x(nT_S) + w_i(nT_s) \tag{3-2}$$

4) Find the first EMD component of $x_i(nT_s)$, where $i = 1, \dots, 2M$.
5) Find the jth ICEEMDAN component, $c^j(nT_s)$, by averaging all the first EMD components obtained from the previous stage.

6) Find the residual of the $x(nT_s)$ after subtracting $c^j(nT_s)$, then replace it with $x(nT_s)$ as follows:

$$x(nT_s) = x(nT_s) - c^j(nT_s) \quad (3\text{-}3)$$

7) Check the stopping criteria. If the stopping criteria are not satisfied, repeat Steps 2-6 to find all the components, otherwise, terminate the decompostion. Note: there are several stopping criteria which exist in the literature. Based on the application, the most appropriate criteria can be selected. Here, the maximum number of components and the number of local extrema points have been employed as the stopping criteria.

3-2-2 LSSVM

In LSSVM, the relationship between the input vector (X) and the output y is as follows:

$$y = w\phi(X) + b \quad (3\text{-}4)$$

where *w* and *b* are the weight vector and bias, respectively. The input vector, X, can be formed based on different feature selection approaches. However, in this paper, the features are determined based on Cao's embedding dimension [12]. In (3-4), ϕ is the feature vector and depends on kernel function. Here, due to the advantages of the Gaussian kernel function in prediction [70], we use this kernel function.

To find *w* and *b*, the following constrained optimization problem should be solved [69]:

$$\begin{aligned} &Min\ C(w,b) = \frac{1}{2}\|w^2\| + \sigma \sum_{k=1}^{K} \quad \alpha_k^2 \\ &s.t.\ |y_k - \langle w, x_k\rangle - b| \le \alpha_{\mathrm{k}} \quad k \\ &\qquad = 1,2,\dots,K \end{aligned} \quad (3\text{-}5)$$

where σ and α are user defined parameters which make a trade-off between the empirical risk and the model flatness. These parameters have been chosen by hybrid optimization based on coupled simulated annealing (CSA) and simplex.

3-2-3 Single Layer ELM

Single layer ELM is simply a single-hidden layer feedforward ANN. Thus, in an ELM with *J* neurons in hidden layer, the relationship between i^{th} input vector, $X_i = [X_{i1}, \dots, X_{iN}]$ and i^{th} output y_i can be written as follows [88]:

$$y_i = \sum_{j=1}^{J} \beta_j f_j(a_j . X_i + b_j) \tag{3-6}$$

where β_j is the weighting coefficient which connects the output of j^{th} hidden neurons to the output and f_j is the activation function of the j^{th} hidden neuron. In (3-6), $a_j = \left[a_{j1}, \dots, a_{jN}\right]^T$ is the weighting vector, including the weights of different elements of input vector, and b_j is the threshold of the activation function.

For the training data with the length of *L*, (3-6) can be written in a compact form as follows:

$$Y = \boldsymbol{H}\boldsymbol{\beta} \tag{3-7}$$

where

$$\boldsymbol{H} = \begin{bmatrix} f_1(a_1 . X_1 + b_1) & \cdots & f_J(a_J . X_1 + b_J) \\ \vdots & \ddots & \vdots \\ f_1(a_1 . X_N + b_1) & \cdots & f_J(a_J . X_N + b_J) \end{bmatrix}_{N \times J} \tag{3-8}$$

$$\boldsymbol{\beta} = \left[\beta_1, \dots, \beta_J\right]^T \tag{3-9}$$

where $\boldsymbol{\beta}$ can be calculated as follows:

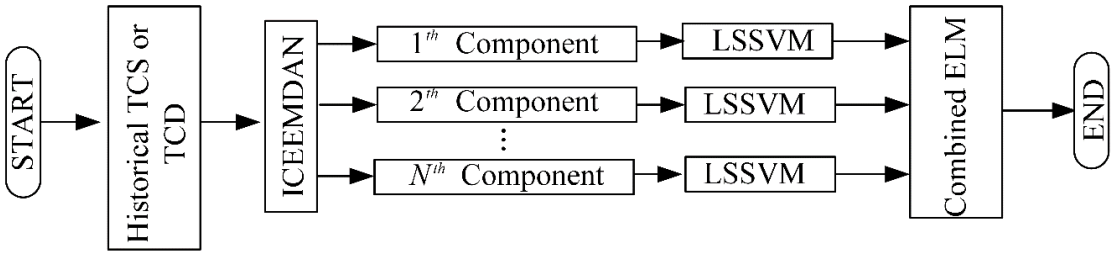

Figure 3-1: The flowchart of the proposed prediction model.

$$\boldsymbol{\beta} = \boldsymbol{H}^{+}Y \tag{3-10}$$

where $\boldsymbol{H}^{+}$ is the Moore-Penrose generalized inverse of $\boldsymbol{H}$. As it is clear, to find $\boldsymbol{H}^{+}$, the hidden neurons weight vectors, $\{a_j\ j = 1, \dots, J\}$, and activation thresholds, $\{b_j\ j = 1, \dots, J\}$, should be known a priori. Usually, the mentioned parameters are randomly generated. Hence, the initial random numbers highly affect the performance of the ELM. This issue will be further discussed and addressed in the Section 3-3.

3-3 The Proposed Prediction Model Description

Now based on the above-explained building blocks the proposed prediction model can be constructed as shown in Figure 3-1. In this prediction model, first, TCS and TCD time series are decomposed into several components by following the proposed ICEEMDAN procedure in Section 3-2-1. Then, the corresponding predicted values of each component are obtained using the LSSVM. In this paper, in training the LSSVM, a localized and online training strategy has been employed. The localized training strategy for predicting the value of j^{th} component, c^j, in the next sample time, $(N+1)T_S$, is as follows:

1) Find the training input vector candidates, their corresponding output, and the test input vector. For this purpose, construct the Hankel matrix, $\boldsymbol{C}^j$, for N available historical sample time of j^{th} component, c^j, as follows:

$$\boldsymbol{C}^j = \begin{bmatrix} C_1^j & \cdots & C_k^j & \cdots & C_r^j \end{bmatrix}^T \quad (3\text{-}11)$$

where $r = N - d + 1$ and d is the embedding dimension, obtained from Cao's embedding dimension method. C_k^j in (3-11) can be obtained as follows:

$$C_k^j = \begin{bmatrix} c^j(kT_S) & c^j\big((k+1)T_S\big) & \ldots & c^j(mT_S) \end{bmatrix} \quad (3\text{-}12)$$

The $r-1$ row vector constructing the first $r-1$ rows of $\boldsymbol{C}^j$ can be considered as the candidates of the training input vectors, X_{Train}, and hereafter the matrix, including all the training input vectors, is denoted by $\boldsymbol{X}_{Train}$. Outputs $\boldsymbol{Y}_{Train}$ can be simply obtained from the last column of the last $r-1$ rows in (3-11). The vector including all elements in the last row of $\boldsymbol{C}^j$ is the test vector, X_{Test}.

2) Find the K most relevant training sets $\{X_{Train}^*, y_{train}^*\}$ based on K-nearest neighbor (KNN) approach. In other words, by means of KNN, the candidate training input vectors which are nearest to X_{Test} are selected.
3) Train LSSVM, using $\{X_{Train}^*, y_{train}^*\}$ and following the procedure explained in Section 3-2-2.
4) Use the trained LSSVM and input vector, X_{Test}, to find the corresponding predicted value of j^{th} component.
5) To perform a multi-step-ahead prediction, the predicted value at each step is recursively used, and Steps 1-4 are repeated.

After finding the predicted values of all the components, in the aggregation and modification stage, the final TCS and TCD in the next sample time are obtained by using a combined ELM. The initial random number generation highly influences the accuracy of ELM. For instance, in two ELMs with the same number of hidden neurons and similar training sets, the accuracy of might be significantly different. In this paper, a combined ELM module, including several ELMs with various initializations, has been used. Then, based on the accuracy of ELMs in training datasets, they are ranked, and the first K ELMs are selected to be utilized for testing input. Finally, using a weighted averaging among the K ELMs, the TCS and TCD in the next sample time is predicted.

3-4 Experimental Results and Comparisons

3-4-1 Data Description

Simulation studies are conducted to assess the efficacy of the proposed ICEEMDAN-LSSVM-ELM prediction model. For this purpose, TCS and TCD historical data of Shark River Entrance, recorded from Oct. 1999 to Nov. 1999, has been used. The data resolution is 6-min, and the prediction horizon is set 1-h with 6-min resolution. In other words, in this study, a 10-step ahead prediction is carried out.

3-4-2 Training Length and Prediction Horizon

To make a trade-off between computation time and the prediction accuracy, the training data with the length of 20 days has been chosen. In other words, 20 days prior to the prediction point are used as the training data.

3-4-3 Evaluation Indices

For comparing the proposed prediction model with other models, three widely-used evaluation indices including the absolute error (AE), normalized mean absolute error (NMAE) and normalized root-mean-squared error deviation (NRMSD) have been used [13]. AE measures the deviation of the predicted values from the real value at each sample point. NMAE indicates the average absolute deviation of the predicted values from actual values, and NRMSD is more sensitive to large errors and the lowers the importance of the small prediction error. NMAE and NRMSD can be calculated as follows:

$$\mathrm{NMAE}(\%) = \frac{1}{(y_{max} - y_{min})N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \times 100 \tag{3-13}$$

$$\mathrm{NRMSD}(\%) = \frac{1}{(y_{max} - y_{min})} \sqrt{\frac{1}{N} |y_i - \hat{y}_i|} \times 100 \tag{3-14}$$

3-4-4 Benchmark Models

To compare the performance of ICEEMDAN-LSSVM-ELM with other prediction approaches. Two well-developed benchmark models have been utilized. As LSSVM is used in TCS and TCD prediction [13], and it is one the main building blocks of the proposed approach, LSSVM-based prediction model is developed as the first benchmark model. Proposing a novel decomposition method, named ICEEMDAN, is one of the contributions of this paper. For this reason, as the second benchmark model, EEMD-LSSVM, which has been recently developed in [13], is used to compare the performance of ICEEMADAN with EEMD in enhancing prediction accuracy.

3-4-5 Numerical Results and Analysis

The simulations are implemented in two parts. In the first part, the performance of the proposed ICEEMDAN is studied. Figure 3-2 shows a 20-day historical TCS which is used for training the prediction model, its extracted components, resulting from ICEEMDAN, and the residual

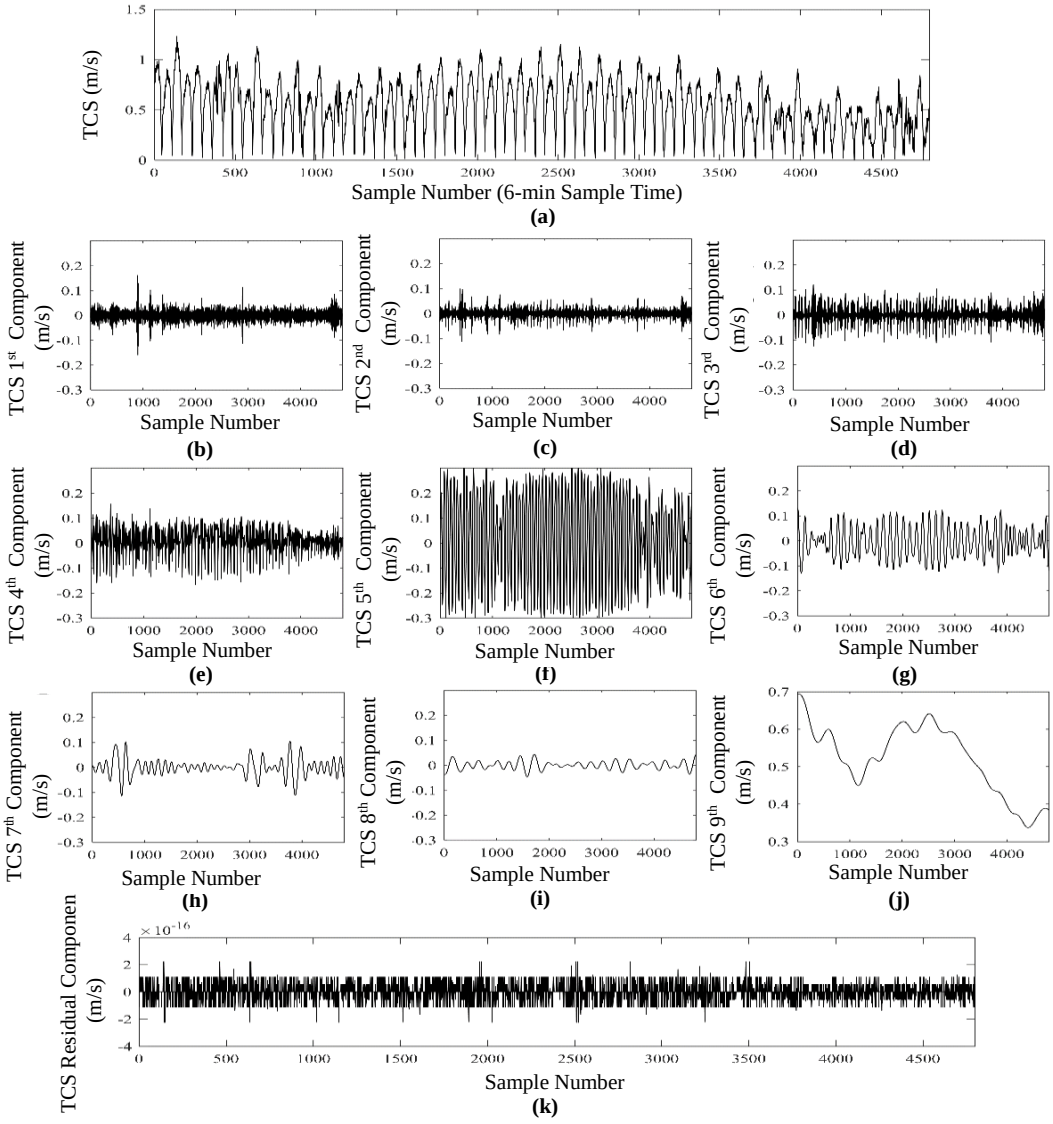

Figure 3-2: Historical TCD and its ICEEMDAN components for 20 days, (a) Original time series, (b)-(j) ICEEMDAN components, (k) the residual component.

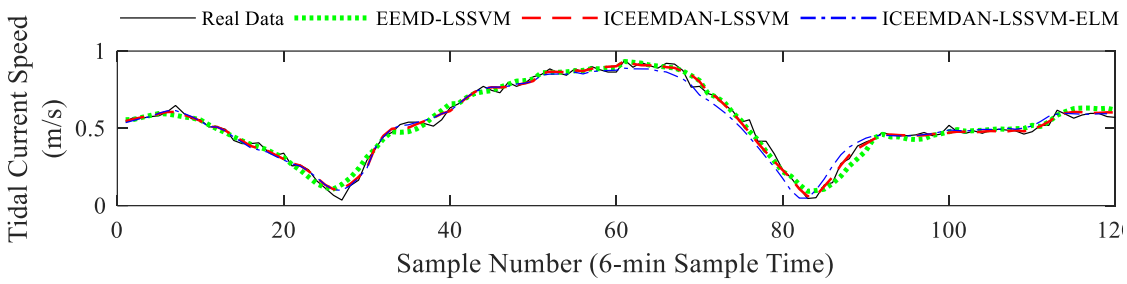

Figure 3-3: The results of 1-h ahead TCS prediction obtained from different decomposition-based

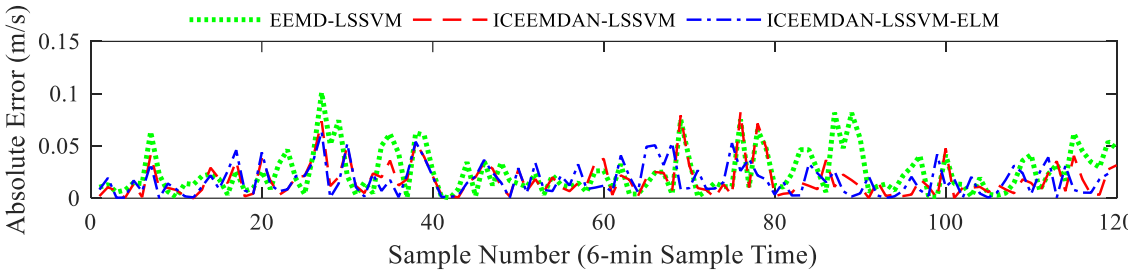

Figure 3-4: The results of 1-h ahead TCS prediction obtained from different decomposition-based prediction models.

component. As it can be observed, the proposed decomposition approach is efficient in decomposing the TCS components into several components, ranging from high-frequency variations to low-frequency variations, with the negligible residual component.

In the second part, the effectiveness of the proposed ICEEMDAN in TCS and TCD is evaluated and compared against different benchmark models. Tables 3-1-3-2 summarize the prediction error,

Table 3-1: Performance evaluation of different prediction models for 1-h (10-step) ahead TCS prediction

Method	NMAE (%)	NRMSD (%)
LSSVM	6.56	9.30
EEMD-LSSVM	1.96	2.93
ICEEMDAN-LSSVM	1.65	2.45
ICEEMDAN-LSSVM-ELM	1.54	2.18

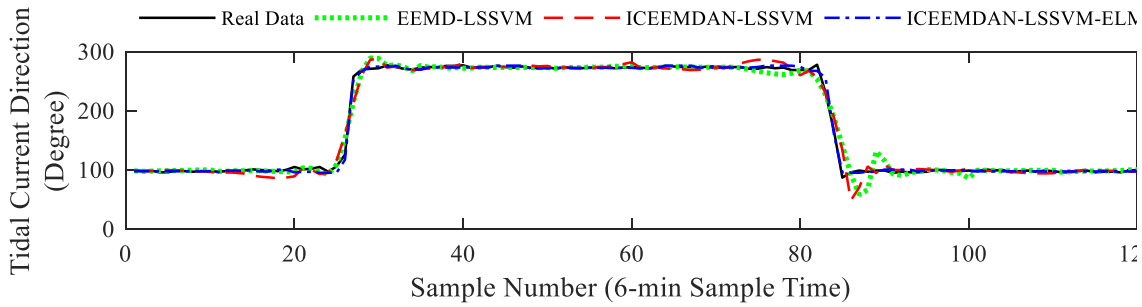

Figure 3-5: The results of 1-h ahead TCD prediction obtained from different decomposition-based prediction

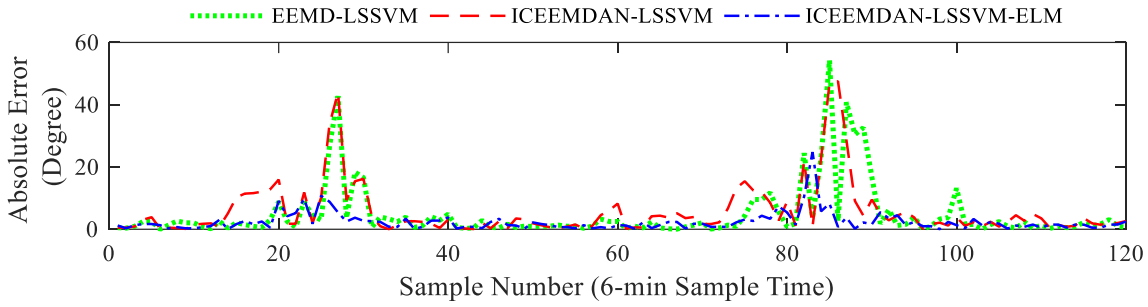

Figure 3-6: The results of 1-h ahead TCS prediction obtained from different decomposition-based prediction

in terms of NMAE and NRMSD, resulted from different prediction models for 1-h ahead prediction. As it can be observed, compared to EEMD, the proposed ICEEMDAN decomposition is more effective in decreasing the prediction error. Besides, in ICEEMDAN-LSSVM-ELM, the prediction accuracy is even higher.

To more closely investigate the superiority of the proposed ICEEMDAN-LSSVM-ELM in comparison to other decomposition-based models, a close view of TCS and TCD 1-h prediction and their corresponding absolute error for a 12-h interval are shown in Figures 3-3-3-6. It can be observed that the proposed ICEEMDAN-LSVVM-ELM possesses a high level of robustness in predicting sudden changes in both TCS and TCD prediction.

Table 3-2: Performance evaluation of different prediction models for 1-h (10-step) ahead TCS prediction

Method	NMAE (%)	NRMSD (%)
LSSVM	6.56	9.30
EEMD-LSSVM	1.96	2.93
ICEEMDAN-LSSVM	1.65	2.45
ICEEMDAN-LSSVM-ELM	1.54	2.18

3-5 Conclusion

A novel deterministic prediction model, named ICEEMDAN-LSSVM, has been addressed for tidal energy application. The proposed approach benefits from a novel ICEEMDAN decomposition approach, which is a modified version of the prevalent CEEMDAN, which avoids EEMD drawbacks. Besides, to further enhance the prediction accuracy an additional stage including a combination of ELMs is considered to ICEEMDAN-LSSVM. The proposed ICEEMDAN-LSSVM-ELM outperforms various well-developed benchmark for 1-h ahead prediction with 6-min resolution. This short-term prediction model can be utilized a wide range of power system applications, such as optimal power flow, real-time unit commitment, and so forth.

4- A SPATIOTEMPORAL WIND POWER PREDICTION BASED ON WAVELET DECOMPOSITION, FEATURE SELECTION, AND LOCALIZED PREDICTION[5]

Wind power possesses a high level of non-linearity and non-stationarity which are the main barriers to developing an accurate wind power prediction (WPP). In this regard, a multiresolution wavelet decomposition (WD), based on discrete wavelet transform, is employed to decompose the wind power time series (TS) into several components. Afterward, in a feature selection (FS) stage, which benefits from the spatiotemporal relation among the wind farms, the double input symmetrical relevance (DISR) has been adopted to find the most suitable features in predicting each component. Then, to have a high-accuracy prediction with an affordable computation time, localized prediction engines have been used to predict each component. The final WPP value is obtained by superposition of all the predicted values corresponding to components. The proposed spatiotemporal WPP is evaluated using the wind power generation historical data in Saskatchewan, Canada. The performance of the proposed WPP is compared with other well-developed and widely-used WPP models. Various evaluation indices have been utilized for conducting the performance evaluation.

[5] © 2018 IEEE. Reprinted, with permission from [89] N. Safari, Y. Chen, B. Khorramdel, L. Mao, and C. Chung, "A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction," in *Proc. IEEE Electrical Power and Energy Conference (EPEC)*, 2017, pp. 1-6.

4-1 Introduction

Wind power generation is one of the fastest-increasing types of clean and renewable energy generations. Due to the intermittent and variable nature of wind power generation, high penetration of wind power generation can cause several unprecedented challenges in power system operation [90, 91]. Wind power prediction (WPP) models can provide useful information about the upcoming wind power generation profile. In this regard, WPP models have been proposed in the literature (e.g. [12, 45, 92-94]).

Based on the nature of WPP output, the WPP literature can be divided into two main categories of deterministic WPP and probabilistic WPP. In deterministic WPP, the most probable power generation value for the next sample time is predicted, while in probabilistic WPP, upper and lower bounds for wind power in the next sample time are calculated. The application of deterministic or probabilistic WPP depends on the decision-making problems in power systems. For example, in [95] probabilistic WPP has been used in economic dispatch and unit commitment, and in [26] energy storage sizing and coordination strategies have been proposed based on deterministic WPP. Besides, in some probabilistic WPP approaches, deterministic WPP models need to be employed since probabilistic WPP results conditionally depend on the deterministic WPP values [96]. Hence, due to the importance of deterministic WPP [12], it is the focus of this paper.

Taking the spatiotemporal correlations among the meteorological variables into account, the accuracy of wind power generation prediction can be improved [97]; thus, the spatiotemporal based WPP models have been proposed in [98, 99]. In [98], using differential equations of wind turbines located in a wind farm with wake effect considerations, the spatial correlation matrix of the wind turbines speed has been constructed. Although this approach proposes an effective way

to model the correlation among the wind turbines in a wind farm, the spatiotemporal correlations among the wind farms in a wide region are not studied. In [99], the spatiotemporal correlations among the generation of wind farms have been considered, and a WPP model has been developed based on sparse vector auto-regression.

An efficient feature selection (FS) can effectively enhance the spatial WPP model accuracy. Among feature selection approaches, information theory-based FS approaches have gained the attention of many researchers in dealing with wide range of problems, ranging from load forecasting [100] to WPP [92, 101]. Unlike traditional correlation coefficient-based FS approaches, in this type of FS, the non-linear relations between the feature candidates and the target variable can be measured [92]. In [102], an FS named double input symmetrical relevance (DISR) has been proposed for cancer classification based on an objective function defined using mutual information (MI). Due to the effectiveness of double input symmetrical relevance (DISR) in different problems [103, 104], we adopt the DISR for spatiotemporal wind power prediction.

The non-stationary wind power time series (TS), obtained from historical data of wind farms, can be decomposed into several components with less non-stationarity [12]. These components are more predictable; hence, the prediction accuracy can be significantly enhanced. Among decomposition approaches, wavelet decomposition (WD) [105] and empirical mode decomposition (EMD) [12, 13] have been effectively applied to non-stationary TS prediction. Wavelet transformation (WT)-based decomposition approaches, e.g., discrete wavelet transform (DWT), can decompose the TS into several components with a predefined frequency range. Thus, adopting WT-based decomposition on all the wind farms TS using DISR can lead to detecting the most suitable feature from various wind farms in predicting every component. For this reason, in

this paper, first, the wind power TS of all wind farms are decomposed, then the appropriate features for each component are identified through an FS stage.

As the last stage of the WPP, a prediction engine for predicting the value of each component should be employed and trained. Due to the non-linearity of wind power TS, nonlinear prediction engines can be employed. Artificial neural networks (ANNs) variants and support vector machine (SVM) variants are widely used for non-linear TS prediction [11-13]. Least squares support vector machine (LSSVM), which is a variant of SVM has shown promising advantages in wind power and wind speed prediction [12, 70]. Hence, this paper employs the LSSVM for spatiotemporal WPP.

In this paper, based on the above discussion, a spatiotemporal WPP model which is comprised of the wavelet transform, DISR, and LSSVM is proposed. The proposed WPP, named WD-FS-LSSVM, has been validated using the wind power generation historical data of the Saskatchewan, Canada. The WPP accuracy of WD-FS-LSSVM has been compared against well-developed WPP benchmark models, using various evaluation metrics. The remainder of this paper is as follows. Section 4-2, presents the building blocks of the WPP. The proposed approach is introduced in Section 4-3. Data sets, evaluation indices, and simulation results of the proposed WD-FD-LSSVM are discussed in Section 4-4. Section 4-5 concludes the paper.

4-2 Building Blocks of the Wind Power Prediction

4-2-1 Wavelet Decomposition

To obtain more accurate results from WPP, WD is employed. In the following, a summary of WT is presented.

WD can be categorized into decomposition based on continuous wavelet transform (CWT) and DWT. Both CWT and DWT functions are obtained by continuously scaling and translating the mother wavelet using discretely sampled. However, increased computational time and memory required to calculate the wavelet coefficients are needed in CWT. Thus, DWT is used because of its efficiency to present the similarity between the original TS and the translated mother wavelet. The DWT is calculated as follows [105]:

$$W(a,b) = \frac{1}{\sqrt{a}} \sum_{n=0}^{N} x(nT_S) g\big((n-b)/a\big) \tag{4-1}$$

where $x\ (nT_S\ )$ is the wind power TS with the sample time $T_{S,}$ which is kept constant and equal to 30-min, and g is the mother wavelet function. In (4-1), a and b, which can be obtained from (4-2), are scaling and translation functions [105].

$$\begin{aligned} a &= 2^m \\ b &= na \end{aligned} \tag{4-2}$$

where *m* and *n* are integer variables.

This paper utilizes Mallat's decomposition algorithm. As an example, a three-level Mallat's decomposition algorithm is depicted in Figure 4-1. The "*approximations*" and "*details*" of the wind power TS are extracted from the multiresolution procedure via Mallat's pyramidal algorithm.

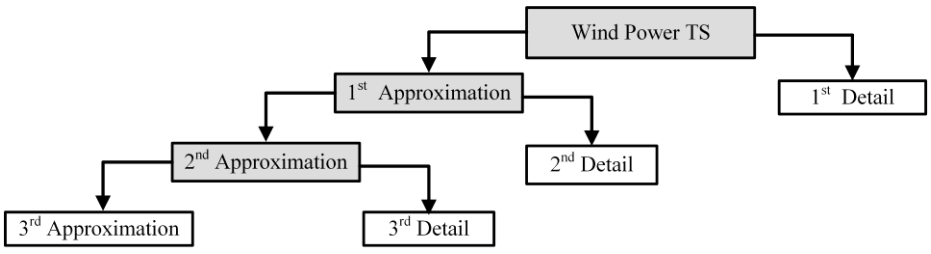

Figure 4-1: The process of multilevel decomposition based on Mallat's algorithm.

An *approximation* is a low-frequency depiction of the TS, holding general trend of the original signal; and *detail* illustrates the difference between two successive *approximations*, representing high-frequency components of the original signal. In every level of decomposition, the approximation components can be further decomposed into approximation and detail. Decomposing the approximation components is repeated until obtaining a desirable number of decomposition levels.

4-2-2 Feature Selection

In this paper, in the feature selection stage, the appropriate features are selected using DIRS [102]. To adopt the DIRS, in WPP application in which the TS can have elements with any positive values, the TS ($z^i(nT_S)$), which represents i^{th} feature, should be discretized. For this purpose, first, the data is normalized as follows:

$$z_n^i(nT_S) = \frac{z^i(nT_S) - \underline{z}^i}{\overline{z}^i - \underline{z}^i} \tag{4-3}$$

where $\overline{z}^i$ and $\underline{z}^i$ represents the maximum and minimum values of $z^i(nT_S)$. Afterward, $z^i(nT_S)$ is divided into 10 bins which are labeled by 10 different values from 1 to 10. The objective of the DIRS, like other feature selection approaches, is to find a set of *k* features, $FS_{Best} = \{f_{S1}, \dots, f_{Sk}\}$, and can be formulated as follows:

$$\max I(\boldsymbol{FS}; Y) \tag{4-4}$$

where *Y* denotes the wind power TS labels for the wind farm which its WPP is our interest. In (4-4), $\boldsymbol{FS}$ is the joint variable which is obtained from multiple features. As an illustrative example, the joint variable of two vectors $a = [1,0,1,0,2]$ and $b = [1,1,0,1,3]$ is $c = [1,2,3,2,4]$ ($c = J(a,b)$). In (4-4), operator $I(\cdot)$ calculates the MI. The MI is a measure of mutual relevancy of

variables. Hence, the higher of $I(\boldsymbol{FS}; Y)$ means the more information about *Y* can be obtained from FS_{Best} [102]. The MI between potential feature *X* and a target can be calculated using (4-5).

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log\left(\frac{p(x).p(y)}{p(x,y)}\right) \tag{4-5}$$

here $p(.)$ is the joint probability distribution function of *x* and *y*.

Now, let's assume we want to select the *k* most appropriate features from a feature pool, the feature selection stage can be performed by the following steps.

1) Find the relevant features through the feature relevancy stage using (4-5), and discard the irrelevant features from feature pool.
2) Find the most informative feature using (4-5), and store that as the first best feature in FS_{Best}.
3) Build the joint variable of FS_{Best} with every feature from feature pool.
4) Calculate the MI for each constructed joint variable and Y using (4-5).
5) Find the highest MI obtained in Step 4, and store the feature which was jointed with the feature(s) in FS_{Best}.
6) Check the stopping criteria. If the stopping criteria have not been satisfied, repeat Steps 3-5. It is noteworthy to mention that in this work the number of preferred features, determined by evaluating the performance of the prediction model on an evaluation set, is used as the stopping criterion.

4-2-3 Least Squares Support Vector Machine

In the LSSVM, the predicted value (y) corresponding to the input vector, also called the test vector (X_{Test}), can be estimated as follows [69]:

$$\hat{y} = \boldsymbol{w}^T \varphi(X_{Test}) + b \tag{4-6}$$

where $\hat{y}$, $\boldsymbol{w}$, and b are the estimation of y, the weight vector, and bias, respectively. φ is a feature vector and depends on the kernel function and its parameters [69]. Among the kernel functions, the Gaussian kernel has shown promising advantages in WPP [91]. In LSSVM with Gaussian kernel, the width of the Gaussian kernel (σ) must be determined. Moreover, the weight vector and bias should also be found through an optimization procedure.

The values of weight vector and bias can be calculated by solving the optimization problem in (4-7), with a training dataset comprised of K training input vectors X_i $(i = 1,2, \dots, K)$ and their corresponding training output y_i.

$$C = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{\gamma}{2}\sum_{i=1}^{K} e_i^2 \tag{4-7}$$

s.t. $\boldsymbol{w}^T \varphi(X_i) + b = y_i - e_i$

where C is the cost function (to be minimized) and K and γ are the number of training vectors and a regularization parameter, respectively.e_i is the error of the model related to i^{th} training input vector and its output. Here, γ, which controls the tradeoff between bias and variance of the model, should be determined [69].

To solve the optimization problem (4-7), first, γ and σ, which are called hyperparameters, should be selected through an optimization before training the LSSVM. In this regard, a coupled simulated annealing (CSA) [69] is first performed to find a starting point for the simplex algorithm, which is one of the most popular linear programming approaches, and then simplex is performed as a fine-tuning step for non-linear optimization problems. CSA-simplex in tuning regulation and kernel function parameters in LSSVM has been successfully employed (e.g., [12]). After the

hyper-parameters are determined, (4-7) can be simply solved by a linear optimization to find $\boldsymbol{w}$ and b.

4-3 Proposed Wind Power Prediction Model Procedure

To conduct the spatiotemporal WPP of L wind farms at $(N+1)T_S$, where N is the number of available data points in the historical data, the following steps should be conducted.

1) Decompose the TS constructed by wind power historical data of all wind farms, based on what is discussed in Section 4-2-1. Repeat Steps 2-10 to perform the prediction of each component of the target wind farm.
2) Hankelize the j^{th} component of wind power historical data for wind farm i ($i=1,\ldots, L$), and construct $\mathbf{W}^i$, with (N-48+1)-by-(48) dimension, as shown in (4-8) for all the wind farms. It is noteworthy to mention that the 48 elements in each row of $\mathbf{W}^i$ represent the wind power generation in one day with the sample time of 30-min.

$$\boldsymbol{W}^i = \begin{bmatrix} w^i(1T_S) & w^i(2T_S) & \ldots & w^i(48T_S) \\ w^i(2T_S) & \vdots & \vdots & w^i(49T_S) \\ \vdots & \vdots & \vdots & \vdots \\ w^i(RT_S) & \ldots & \ldots & w^i(NT_S) \end{bmatrix} \tag{4-8}$$

where $R = N - 48 + 1$.

3) Construct the feature candidate vectors, using (4-8), as follows:

$$\boldsymbol{FS} = \left[FS_1^1, \ldots, FS_{48}^1, \ldots, FS_i^l, \ldots, FS_1^L, \ldots, FS_{48}^L\right] \tag{4-9}$$

where FS_i^l corresponds to all elements in i^{th} column (excluding the last row) of the Henkel matrix of l^{th} wind farm.

4) Construct the target vector. If WPP is related to i^{th} wind farm, the target vector, Y, is a column vector consisting of the elements of the last columns of $\boldsymbol{W}^i$ in (4-8), except the first row.
5) Perform FS, illustrated in Section 4-2-2. Find the best features and store in $\boldsymbol{FS}_{Best}$.
6) Build the input of training candidate and input for the test, using $\boldsymbol{FS}_{Best}$.
7) Find the most suitable training input vectors corresponding to the test vector. It is worth noting that the test vector is the last row of $\boldsymbol{FS}_{Best}$, and by means of K-nearest neighbor (KNN) approach, the closest rows of $\boldsymbol{FS}_{Best}$ to test vector is selected as the suitable training input vectors.
8) Find the target vector for the most suitable candidates. If an input vector is constructed from r^{th} row of Hankel matrix in (4-8), and we want to construct the WPP for i^{th} wind farm, the target value corresponding to this input vector is the last element of $(r+1)^{th}$ row of $\boldsymbol{W}^i$.
9) Train the LSSVM, introduced in Section 4-2-3, for predicting j^{th} component,
10) Apply the test vector and find the estimation of the component for the next sample time.
11) Aggregate all the predicted values and find the final WPP for i^{th} wind farm, considering the spatiotemporal correlation among the wind farms.

4-4 Case Studies and Analysis

4-4-1 Description of Data

To evaluate the performance of the proposed WPP, the historical wind power generation data of wind farms, located in Saskatchewan province, is used. The location of wind farms in

Saskatchewan has been shown in Figure 4-2. As it is clear from this figure, the wind farms are installed over a wide area; thus, considering the spatiotemporal correlation among the wind farms can enhance WPP accuracy.

The data are recorded with 1-min resolution. In this paper, 30-min ahead WPP has been carried out, and the one-minute data has been averaged to construct time series with 30-min resolution. Hence, in one step ahead prediction, the average wind power in the next 30-min is predicted. Wind power generation data of wind farms is the only data which have been utilized in this paper. However, other meteorological variables might help in improving WPP accuracy.

The historical data recorded from June 2016 to July 2016 has been used for training and testing the proposed WPP. The last 30 days prior to prediction has been used for constructing the training vectors. In this paper, spatiotemporal WPP of Red Lily and Sunbridge wind farms, using the wind power generation data of all the five existing wind farms, have been developed.

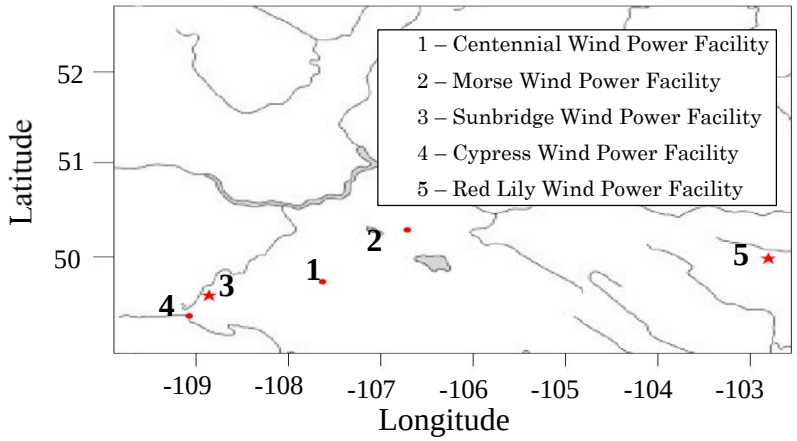

Figure 4-2: Location of wind farms in Saskatchewan, Canada 2016.

4-4-2 Evaluation Indices

Three well-defined and widely used evaluation indices including the normalized absolute error (NAE), normalized root-mean-squared error (NRMSE), and normalized mean absolute error (NMAE) have been employed to assess the accuracy of the WPP. Using (4-10)-(4-12), NAE, NRMSE, and NMAE can be calculated, respectively.

$$NAE(nT_S) = \left| \frac{\hat{P}(nT_S) - P(nT_S)}{P_{Inst.}} \right| \tag{4-10}$$

$$NRMSE = \sqrt{\frac{\sum_{n=1}^{N} \left(NAE(NT_S) \right)^2}{N}} \tag{4-11}$$

$$NMAE = \frac{\sum_{n=1}^{N} \left(NAE(NT_S) \right)}{N} \tag{4-12}$$

where $\hat{P}(nT_S)$ is the n^{th} predicted wind power; and N is the number of test points.

4-4-3 Benchmark Models

To compare the prediction accuracy of the proposed spatiotemporal WD-FS-LSSVM, the performance of the proposed WPP is compared with four well-developed benchmark models, including persistence model (PM), FS-LSSVM without considering the spatiotemporal correlation, FS-LSSVM with considering the spatiotemporal correlation, and WD-FS-LSSVM without considering the spatiotemporal correlation.

The PM, which is widely used as the main benchmark, can provide useful information about the percentage of improvement in reducing the WPP error [12]. As WD, FS, and LSSVM are the three main building blocks of the proposed spatiotemporal WD-FS-LSSVM WPP, comparing the performance of the proposed WPP with benchmark models comprised of these building blocks can

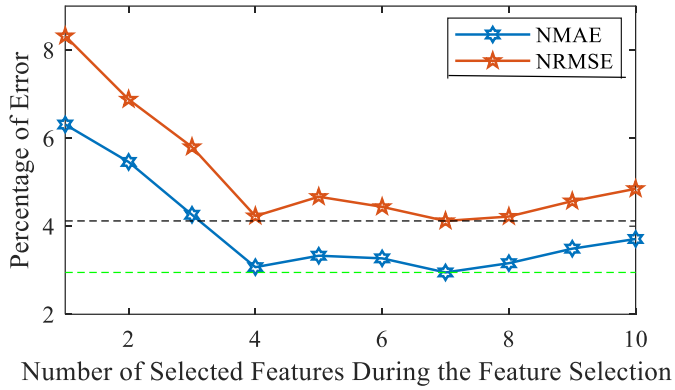

Figure 4-3: The relation between number of features and WPP.

provide a fair comparison. Besides, the impact of considering spatiotemporal can be investigated by developing both spatiotemporal and non-spatiotemporal WPP models.

4-4-4 Numerical Results and Comparisons

To find the optimal number of features in the FS stage, the different number of features have been used and the resultant NRMSE and NMAE have been analyzed. The NRMSE and NMAE in 30-min ahead prediction of Sunbridge for the different number of features is shown in Figure 4-3. From this figure, it can be found that by selecting 7 features from the feature pool, DISR provides the most informative and useful features.

The performance of the proposed spatiotemporal WD-FS-LSSVM WPP and benchmark models in 30-min ahead WPP of Sunbridge and Red Lily have been summarized in Tables 4-1-4-24-1. From these tables, it can be observed that compared to benchmark models, the proposed WPP has the lowest NRMSE and NMAE. Furthermore, it can be also seen that considering the

Table 4-1: Performance comparison of different WPP models for Sunbridge wind power facility in July 2016.

Prediction Models	NRMSE (%)	NMAE (%)
PM	12.9901	10.0091
FS-LSSVM (Non-Spatiotemporal)	13.7150	10.0722
FS-LSSVM (Spatiotemporal)	13.8276	10.7934
WD-FS-LSSVM (Non-Spatiotemporel)	5.7034	3.8945
WD-FS-LSSVM (Spatiotemporal)	4.1200	2.9567

Table 4-2: Performance Comparison of Different WPP Models for Red Lily Wind Power Facility in July 2016.

Prediction Models	NRMSE (%)	NMAE (%)
PM	16.88	12.42
FS-LSSVM (Non-Spatiotemporal)	16.46	12.20
FS-LSSVM (Spatiotemporal)	17.69	13.01
WD-FS-LSSVM (Non-Spatiotemporel)	5.51	3.77
WD-FS-LSSVM (Spatiotemporal)	4.41	3.27

spatiotemporal correlation in WPP without decomposition stage might deteriorate the WPP accuracy. From these tables, it can be concluded that in different frequency ranges, different wind farms can provide useful information. Hence, decomposition prior to FS, plays a crucial role in identifying the informative feature which enhance the accuracy of predicting each component, and can consequently improve the WPP accuracy.

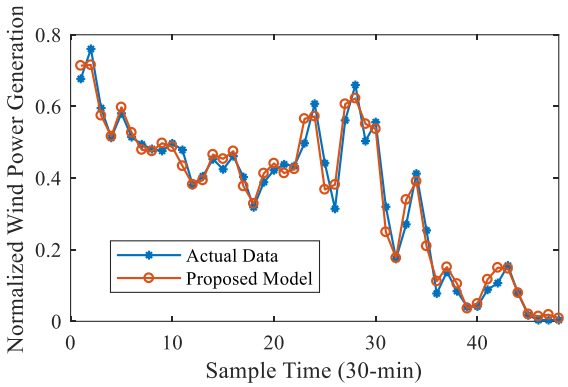

Figure 4-4: WPP for 30-min ahead in Sunbridge wind farm.

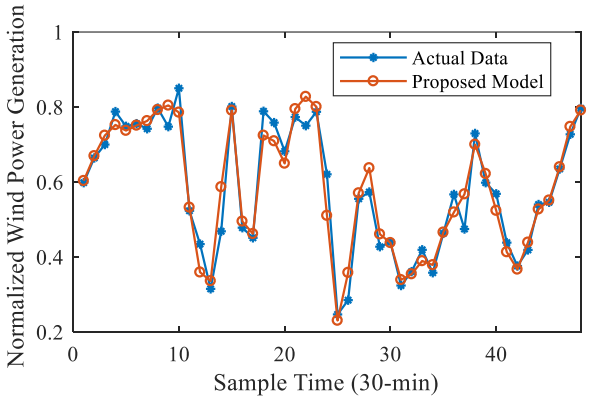

Figure 4-5: WPP for 30-min ahead in Red Lily wind farm.

To more closely investigate the performance of the proposed WPP model, the 30-min WPP in Sunbridge and Red Lily wind farms for one day have been depicted in Figures 4-4- 4-5, respectively. Besides, the kernel density estimation of NAE distribution of various prediction models for Sunbridge has been shown in Figure 4-6. This figure clarifies the effectiveness of the proposed WPP model in limiting the NAE to lower values.

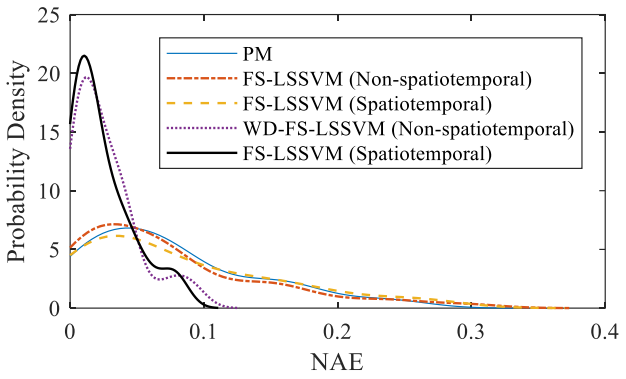

Figure 4-6: Kernel density estimation of NAE distribution of various WPP models in Sunbridge wind farm.

4-5 Conclusion

In this paper, a novel spatiotemporal WPP based on WD and DISR has been addressed. The proposed WPP benefits from various wind farms information in different frequency range, thus the WPP accuracy is significantly enhanced. Besides, it has been observed that, the spatiotemporal might deteriorate the WPP accuracy, unless the decomposition approaches are employed. The proposed WPP is compared with other well-developed and widely used benchmark models. This WPP tool can be employed in short-term power system operation.

5- VERY SHORT-TERM WIND POWER PREDICTION INTERVAL FRAMEWORK VIA BI-LEVEL OPTIMIZATION AND NOVEL CONVEX COST FUNCTION[6]

Substantial challenges in power systems operation and control as a result of the intermittent and stochastic nature of wind power generation can be significantly alleviated by proficient short-term wind power prediction interval (WPPI) models. In WPPI models, minimization of cost functions is conducted to train prediction engines and consequently tune their parameters. The prevalent cost functions of prediction engines in WPPI models are mainly non-differentiable and non-convex, and therefore the training process becomes problematic. To transcend such a crucial barrier, this paper addresses a new short-term WPPI framework based on a bi-level formulation and benefiting from a differentiable and convex cost function. The prediction engine is trained by classical global optimization of the cost function in the lower-level problem, while hyperparameters that control the quality of the WPPIs are injected thereto from the upper-level problem. The hyperparameters can be tuned such that the most useful WPPIs are constructed from the lower-level problem depending on the power system operator's preferences. Lessening the need to heuristically tune a large number of prediction engine parameters is the foremost contribution of this work to the WPPI literature. The superior performance of the proposed WPPI is verified in the multistep ahead prediction of real wind power generation data in comparison to well-tailored benchmark models.

[6] © 2018 IEEE. Reprinted, with permission from [106] N. Safari, S. M. Mazhari, and C. Y. Chung, "Very Short-Term Wind Power Prediction Interval Framework via Bi-level Optimization and Novel Convex Cost Function," *IEEE Trans. Power Syst., Early Access*.

5-1 Nomenclature

$\{y_n\}_{n=1}^{N}$	Wind power time series with *N* points.
T_S	Sample time.
L	Number of sample points in the datasets for training, validation, or testing.
$\overline{y}_n$, $\underline{y}_n$	Upper and lower bounds for the nth wind power observation.
M	Number of hidden neurons.
$h_m(\cdot), \boldsymbol{w}_m, b_m$	Activation function, weight vector, and bias for the mth hidden neuron.
$\boldsymbol{X}_n$	Input vector corresponding to the nth point.
$\overline{\beta}_m$, $\underline{\beta}_m$	Output weights related to the mth hidden neurons of the upper and lower bounds.
τ	Embedding delay time.
D	Embedding dimension.
δ_n	Binary variable equal to one when the lower and upper bounds encircle the nth observation, and zero otherwise.
RL, RL^*	Calculated and preferred reliability level.
$\overline{\Delta y}_n$, $\underline{\Delta y}_n$	Deviation of the upper and lower bounds from actual wind power.
$C(\cdot)$	The proposed cost function.
$\mathcal{E}(\cdot)$, $R(\cdot)$	Elements of $C(\cdot)$.
θ	Indicator of $\overline{\Delta y}_n$ and $\underline{\Delta y}_n$.

$\omega, \boldsymbol{a}, c$	Hyperparameters of the proposed cost function.
$F_{up}(\cdot), F_{lo}(\cdot)$	Lower-level function of upper and lower bounds.
u, U	Set of all upper and lower bounds hyperparameters and its domain.
$\mathscr{k}, \mathcal{K}$	Set of unknown parameters of the prediction engine and its domain.
$\overline{\boldsymbol{\beta}}, \underline{\boldsymbol{\beta}}$	Output weight vectors of the upper and lower bounds.
$G(\cdot), G_{up}(\cdot), G_{lo}(\cdot)$	Derivatives of overall lower-level function, $F_{up}(\cdot)$, and $F_{lo}(\cdot)$.

5-2 Introduction

Wind power prediction (WPP) models are amongst the most indispensable tools for addressing particular issues posed to power systems by uncertainties accompanying the power generation injected from wind farms [107]. The importance of a precise WPP model is further underlined by the ever-growing increase in wind power penetration. Based on the prediction horizon, WPP models can be divided into four categories, comprising long-, medium-, short-, and very short-term prediction [108].

Prediction models with horizons less than 1-hour ahead, which also belong to the short-term and very short-term WPP category [108], are of utmost importance to various power system operation problems, including regulation and control actions [108], real-time electricity market and power dispatch [95], optimal power flow [109], etc., and invaluable efforts have been made to develop accurate prediction models for these horizons (e.g., [12, 46, 110-113]). By virtue of the importance of WPP model accuracy, this paper zeros in on developing a WPP model for few minute-ahead to 1-hour ahead horizons.

Traditionally, WPP models provide the most likely wind power values in the next sample points as the deterministic estimation of wind power [12, 46, 110, 111]. These prediction models have been widely used in various power system operation and planning problems (e.g., [25, 26]). However, due to the chaotic and intermittent nature of wind power time series (TS), some degree of uncertainty in WPP is inescapable [112]. In this regard, many power system decision-making tasks (e.g., [95, 109]) call for WPP models that can estimate the uncertainties of wind power generation in a probabilistic manner [112].

Probabilistic WPP can be conducted using two different approaches: parametric [112] and non-parametric [113-116]. Amongst these, non-parametric probabilistic WPPs that do not limit the shape of the wind power probability distribution function (PDF) are preferred [113]. Non-parametric probabilistic WPPs can be categorized into four types [117]: quantile regression [114], kernel density estimation methods [115], ensemble approaches [116], and machine learning-based frameworks [113]. Some WPP approaches are an amalgamation of some or all of these types and do not exclusively belong to a specific category (e.g., [60]).

Owing to fruitful advancements in machine learning, significant attention has been devoted to machine learning-based probabilistic prediction [1, 45, 60, 113, 118-121]. WPP interval (WPPI) models are versatile probabilistic models that are used in a wide range of decision-making tasks [60]. The cost function and prediction engine are two major components of WPPI models [121]. In the optimization process, tailored cost functions are employed to train the prediction engine and find its unknown parameters; this is the most imperative step for acquiring a high-quality WPPI model and therefore is the focus of this work.

Reliability level (RL) and sharpness are two crucial evaluation factors in this context and have been used as primary blocks of various cost functions [1, 45, 60, 113, 118-121]. RL measures the probability that the constructed WPPIs encircle the future real wind power values while the sharpness of the WPPIs scales the deviation of their upper and lower bounds from real wind power values. A WPPI model with high RL and sharpness facilitates the economical and reliable integration of wind power to the power system [60].

In [118, 121], a single objective problem is put forward by means of a novel cost function, called the coverage width-based criterion (CWC), which reflects both RL and sharpness. In [1], the CWC is further extended to form a new CWC (NCWC) by taking the prediction interval root-mean-square-error ($PIMSE$) into consideration. In the CWC and its variants, an exponential term, containing a hyperparameter penalizes the violation from the RL requirement. The cost functions proposed in [119, 120] are constructed based on a weighted combination of RL and sharpness measurements to circumvent the challenge of tuning the hyperparameter in the CWC. In [45], the problem of maximizing sharpness and RL is formulated as a two-objective problem defined via fuzzifications. In [113], a constrained optimization problem is proposed, and the sharpness is maximized while RL meets the required RL.

As the previously proposed cost functions are non-differentiable [119], WPPI models in [1, 45, 113, 118-121] are inevitably forced to employ heuristic optimization techniques to find a large number of parameters in the prediction engine. However, heuristic optimization procedures may be trapped in local optima [60]; they are also highly dependent on initial values [60, 121], and the results of independent runs can differ from one another. Importantly, the optimization problem markedly scales up by increasing the size of the prediction engine [122]. In [60, 123], the WPPI

problem is formulated as convex programming using the quantile regression (QR) concept and extreme learning machine (ELM). However, the WPPI literature on differentiable and convex cost functions is still in its infancy. Besides, supervised control for examining and/or modifying the RL and sharpness according to the system operator (SO) preference is beneficial. However, it cannot be achieved by the thus-far single level convex programming due to the inherent non-convexity of RL as well as other prediction interval metrics.

To overcome the shortcomings of the cost functions developed in [1, 45, 113, 118-121] and their corresponding optimization difficulties in WPPI modeling, a bi-level formulation is proposed in this paper. The lower-level optimization problem is responsible for optimizing the prediction engine parameters by a differentiable formulation, which enables the use of derivative-based classical global optimization algorithms. Not being trapped in a local optimum and the availability of mature derivative-based global optimization packages are two of the most salient advantages of the proposed lower-level problem. Importantly, a few tunable hyperparameters, determined in the upper-level problem, influence the lower-level problem optimization.

The upper-level problem is devised so the operator's preference can be met by tuning hyperparameters that are upper-level variables. Maximizing RL and sharpness are set as the objectives of the upper-level problem. Non-dominated sorting genetic algorithm II (NSGA-II) [124]—a widely-used and adept meta-heuristic optimization in power system problems [125]—is adopted for this purpose. In the proposed WPPI model, only a few hyperparameters, which are defined in the upper-level optimization problem, are heuristically obtained while a large number of prediction engine parameters are derived from classical optimization. Therefore, although the proposed WPPI formulation utilizes heuristic optimization in the upper-level problem, the search space is limited due to the small number of upper-level variables, which makes finding the optimal

values of upper-level variables highly feasible [122]. Remarkably, the number of hyperparameters is independent of the scale of the prediction engine and the number of unknown parameters therein; in other words, scaling up the prediction engines will not increase the difficulty and complexity of the upper-level problem.

The bi-level optimization structure and the proposed cost function must be employed along with an appropriate prediction engine. In [1, 45, 113, 118, 121], the unknown parameters of neural networks (NNs) have been tuned heuristically such that the cost function is minimized. However, NNs are not only prone to becoming trapped in local minima [118], but they also suffer from high computational costs [119]. Unlike NNs, training of support vector machines (SVMs) obtains the global minimum [126, 127]. The main parameters of SVMs have been found via quadratic programming while hyperparameters are heuristically tuned by optimizing RL and sharpness. ELM is another learning algorithm, initially developed for training feed-forward NNs [128]. The hidden neuron parameters in ELM are randomly chosen while the output weights are found by means of a simple optimization with mild constraints. The authors of [129] show that ELM outperforms SVM in WPP and its training is thousands of times faster [128]; the simplicity of the optimization problem for finding ELM output weights are amongst the chief benefits of ELM. Additionally, employing ELM along with the proposed cost function leads to a differentiable and convex optimization problem; thus, its global solution can be attained. Hence, in this paper, ELM is employed as the prediction engine of the proposed WPPI.

Based on the above discussion, this paper addresses the challenges in tuning a large number of parameters in WPP that trace back to the cost function. The proposed WPPI model features a bi-level optimization structure and a novel cost function for tuning the prediction engine parameters. The upper-level variables control the RL and sharpness of the WPPI are fed into the lower-level

problem, which is differentiable and convex. Then, the lower-level variables, including the prediction engine parameters, are obtained via classical convex optimization. From the SO's viewpoint, knowing the value of wind power from the next multi-sample of wind power is of great importance. Hence, the WPPI model is developed as a multistep prediction. To evaluate the effectiveness and performance of the proposed WPPI model in comparison to well-developed benchmark models, two real-world wind power generation datasets, including aggregated wind power generation of a wide region and individual wind farm, with diverse characteristics are used. In short, the main contributions of this work are two-fold:

- A novel differentiable and convex cost function for tuning prediction engine parameters of the WPPI model. The prediction engine parameters are tuned globally and consequently, training the prediction models several times will lead to unique WPPIs for both training and test datasets.
- For the first time, a bi-level optimization structure for the WPPI is developed to supervise the quality of WPPIs based on the SO's criteria, while primary parameters of the prediction engines are tuned globally.

The remainder of this paper is structured as follows. In Section 5-3, the general formulation of the prevailing ELM-based WPPI models is reviewed, and the proposed cost functions are described. Section 5-4 describes the proposed WPPI formulation and its solution approach. Case studies and simulation results are described in Section 3-4. Finally, Section 3-5 contains salient remarks and conclusions from the work.

5-3 Conventional ELM-based WPPI and The Proposed Cost Function

5-3-1 Conventional ELM-based WPPI

Based on the outstanding advantages of ELM discussed earlier, an ELM is employed as the prediction engine without loss of generality. $\overline{y}_n$ and $\underline{y}_n$ can be obtained using (5-1) and (5-2) [128], respectively:

$$\overline{y}_n = \sum_{m=1}^{M} \left(\hbar_m (\boldsymbol{w}_m^T \boldsymbol{X}_n + b_m)\, \overline{\beta}_m \right) \tag{5-1}$$

$$\underline{y}_n = \sum_{m=1}^{M} \left(\hbar_m (\boldsymbol{w}_m^T \boldsymbol{X}_n + b_m) \underline{\beta}_m \right) \tag{5-2}$$

Hidden neuron weighting coefficients and bias are chosen randomly while the output weights must be selected optimally [128]. Therefore, for an ELM with M hidden neurons, $2 \times M$ prediction engine parameters (i.e., $\overline{\beta}_m$ and $\underline{\beta}_m, m = 1, \dots, M$) need to be tuned; therefore, the ELM formulations in (5-1)-(5-2) are linear with respect to unknown parameters. Note that this study utilizes historical wind power data as the input of the WPPI model. However, exogenous variables (e.g., wind speed, wind direction, etc.) can also be considered. Various feature selection (FS) approaches can be adapted to construct the feature vector ($\boldsymbol{X}_n$), e.g., [68, 92, 130, 131]. Due to the extreme nonlinearity of wind power TS [12], a state space representation of TS, based on the method of delay, is used in this paper. In this regard, the wind power TS is mapped to an embedding space by means of informative features, including a series of wind power lags [12]. The input vector, $\boldsymbol{X}_n$, is as follows:

$$\boldsymbol{X}_n = [y_{(n-1)-(D-1)\tau} \quad \cdots \quad y_{(n-1)-\tau} \quad y_{n-1}]_{1\times D} \tag{5-3}$$

where τ is found using the mutual information-based approach proposed in [130] and embedding dimension, D, is determined based on Cao's embedding dimension method [68].

The accuracy of the WPPI model is mainly assessed by RL and sharpness. The RL of a WPPI model is calculated as follows [45]:

$$RL = \frac{1}{L}\sum_{n=1}^{L} \delta_n, \tag{5-4}$$

where δ_n is defined as follows:

$$\delta_n = \begin{cases} 1, & \underline{y}_n \le y_n \le \overline{y}_n \\ 0, & otherwise \end{cases}.$$

The sharpness of a WPPI model is determined based on the prediction interval normalized average width ($PINAW$), which is defined as follows [45]:

$$PINAW = \frac{1}{\left(Max(\{y_n\}_{n=1}^{L}) - Min(\{y_n\}_{n=1}^{L})\right).L} \cdot \sum_{n=1}^{L}\left(\overline{y}_n - \underline{y}_n\right) \tag{5-5}$$

In the literature, prediction engines of WPPI models are trained such that RL is maximized or close to the preferred RL (RL^*), while $PINAW$ is close to zero ($PINAW \to 0$). RL is non-differentiable and non-convex due to the discontinuous piecewise function used for δ_n calculation. For this reason, heuristic optimization techniques are widely used for finding the prediction engines parameters in (5-1) and (5-2). Therefore, in total, $2 \times M$ parameters need to be found through a heuristic optimization process. The optimization problem also linearly scales up by increasing the number of hidden neurons due to a substantial increment of search space. In addition, heuristic optimization is highly dependent on initial values, and the outcomes for different runs can differ from each other.

In this regard, Section 5-3-2 aims to alleviate the aforementioned side effects by defining a convex and differentiable cost function for tuning the prediction engine's parameters. Hence, classical convex optimization can be employed for acquiring the prediction engine unknown parameters. This consequently brings about reproducibility in tuning prediction engine parameters and attaining high quality WPPIs.

5-3-2 Description of the Proposed Cost Function

To construct accurate WPPIs, upper and lower bounds should be determined such that their deviations from the actual wind power tend to zero, while the actual wind power is encircled by upper and lower bounds. Based on this, a cost function, which benefits from differentiability and convexity, is put forward to acquire prediction engine parameters.

The deviations of $\overline{y}_n$ and $\underline{y}_n$ from y_n are defined as follows:

$$\overline{\Delta y}_n = \overline{y}_n - y_n \tag{5-6}$$

$$\underline{\Delta y}_n = y_n - \underline{y}_n \tag{5-7}$$

According to (5-6) and (5-7), if the observed value is encircled by $\overline{y}_n$ and $\underline{y}_n$, then

$$\overline{y}_n \geq y_n \text{ or } \overline{\Delta y}_n \geq 0 \tag{5-8}$$

$$\underline{y}_n \leq y_n \text{ or } \underline{\Delta y}_n \geq 0 \tag{5-9}$$

Inequalities (5-8) and (5-9) describe the requirement for reliable upper and lower bounds, respectively. To efficiently train the prediction engine, substantial penalties need to be imposed if either of the inequalities (5-8) and (5-9) is violated. Meanwhile, depending on deviations of the lower and upper bounds from the actual value, the cost should be adaptively selected. In other words, $\overline{\Delta y}_n$ and $\underline{\Delta y}_n$ should tend to zero, while the actual points should lie within the upper and

lower bounds, for a high-quality WPPI. To this end, the cost function proposed for tuning the prediction engine unknown parameters is defined as follows:

$$C(\omega, \theta, \boldsymbol{a}) = \mathcal{E}(\omega, \theta) \cdot R(\theta, \boldsymbol{a}) \tag{5-10}$$

$$\mathcal{E}(\omega, \theta) = \begin{cases} \omega & \theta \geq 0 \\ 1 & \theta < 0 \end{cases} \tag{5-11}$$

$$R(\theta, \boldsymbol{a}) = \sum_{k=1}^{K} a_k \theta^{2k} \tag{5-12}$$

In (5-10), $\mathcal{E}(\cdot)$ accounts for the effects of any violation of (5-8) and (5-9) and is the primary term of the cost function. The second term, $R(\cdot)$, is set to make the whole cost function differentiable. θ is the variable of cost indicator and represents $\overline{\Delta y}_n$ or $\underline{\Delta y}_n$. ω is a hyperparameter that regulates both sharpness and RL. $\boldsymbol{a} = \{a_k, k = 1, \dots, K\}$ is a vector consisting of a set of polynomial coefficients. K is the number of non-zero coefficients.

In (5-11), ω is selected within the range of $[0\ \ 1]$. For a constant ω, $\mathcal{E}(\omega, \theta)$ is a piecewise function with output ω, if $\theta \geq 0$; otherwise, its output is 1. Therefore, $\mathcal{E}(\omega, \theta)$ can be used as a cost function for constructing reliable upper and lower bounds and tuning the prediction engine parameters. However, $\mathcal{E}(\omega, \theta)$ is still a non-convex function of θ, which makes it troublesome to find the set of prediction engine unknown parameters for a global optimal solution. Furthermore, $\mathcal{E}(\omega, \theta)$ is non-differentiable at $\theta = 0$, which is a barrier for utilization of classical convex optimization approaches. Multiplication of $\mathcal{E}(\omega, \theta)$ with certain convex functions is an efficient approach for convexification [132]. The multiplied function should be convex with its global optimum at $\theta = 0$ to reward sharp and reliable WPPIs and discourage construction of WPPIs with low sharpness. Amid the convex and differentiable functions, a combination of the even-degree terms, as shown in (5-12), not only meets the aforementioned requirements but also forms a generalized

structure that creates enough freedom to cover various shapes of the cost function. The convexity of $C(\cdot)$ is further discussed in Appendix C. The first derivative of (5-10) is as follows:

$$\frac{\partial C(\omega,\theta,\boldsymbol{a})}{\partial\theta}=\begin{cases}\omega\times\dfrac{\partial R(\theta,\boldsymbol{a})}{\partial\theta} & \theta\geq 0\\ \dfrac{\partial R(\theta,\boldsymbol{a})}{\partial\theta} & \theta<0\end{cases} \tag{5-13}$$

where

$$R(\theta=0,\boldsymbol{a})=0 \text{ and } \frac{\partial R(\theta=0,\boldsymbol{a})}{\partial\theta}=0$$

Therefore, Eq. (5-10) is differentiable for all θ values, and its first derivative with respect to θ is continuous. This feature will be used in Section 5-4-2 to tune the prediction engine parameters.

Figure 5-1 describes the impact of different ω on $C(\cdot)$ while a simple 2nd degree function, for instance, is considered for $R(\cdot)$. Once ω =1, the cost reaches a minimum at $\theta=0$. Employing such a cost function for training the prediction engine results in ignorance of the error sign; and therefore, the prediction engine parameters are tuned to minimize the absolute error of the prediction engine output from the real wind power value. This case is equivalent to a wind power point forecast with the mean squared error as the cost function. On the other hand, in the case of $\omega=0$,

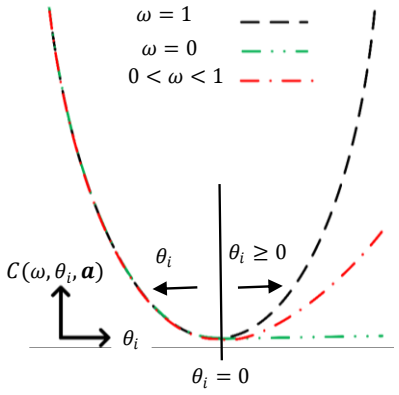

Figure 5-1: $C(\omega,\theta,\boldsymbol{a})$ for different ω.

the cost function evenly imposes the minimum cost if the inequality in (5-8) or (5-9) is met, and the sharpness of the prediction is disregarded; for $0 < \omega < 1$, some degrees of both sharpness and reliability are considered as described in Figure 5-1. As an example, let us assume the actual power at the nth point is $y_n = p$, and we have two possible upper bounds of $\overline{p}_1 \geq p$ and $\overline{p}_2 < p$. Based on Eq. (5-10), the penalties associated with them are $\omega \times \left(\overline{p}_1 - p\right)^2$ and $\left(\overline{p}_2 - p\right)^2$, respectively. As we set out to minimize Eq. (5-10), the value of ω is a determinant for selecting $\overline{p}_1$ or $\overline{p}_2$. In other words, ω decides whether RL or $PINAW$ matters. Therefore, the optimal ω needs to be found to construct optimal upper and lower bounds. Hereafter, ω related to the upper and lower bounds are respectively denoted by $\overline{\omega}$ and $\underline{\omega}$ and are hyperparameters in the proposed WPPI model.

Moreover, to optimally construct WPPIs, $\boldsymbol{a}$ related to constructing both upper bound ($\overline{\boldsymbol{a}}$) and lower bound ($\underline{\boldsymbol{a}}$) need to be aptly selected and are considered as hyperparameters. Sensitivity analysis should be carried out to find the proper degree of the polynomial; therefore, the number of hyperparameters can be substantially reduced.

According to the above discussion, Eq. (5-10) can be employed to construct proficient upper and lower bounds. The overall set of hyperparameters, u, is as follows:

$$u = \left\{\overline{\omega}, \underline{\omega}, \overline{\boldsymbol{a}}, \underline{\boldsymbol{a}}\right\} \tag{5-14}$$

Based on (5-6)-(5-12), (5-14), and substituting $\overline{\Delta y}_n$ and $\underline{\Delta y}_n$ for θ, the generalized cost function for tuning the prediction engine parameters is as follows:

$$F(u, \mathcal{k}) = \sum_{n=1}^{L} \left(C\left(\overline{\omega}, \overline{\Delta y}_n, \overline{\boldsymbol{a}}\right) + C\left(\underline{\omega}, \underline{\Delta y}_n, \underline{\boldsymbol{a}}\right)\right) \tag{5-15}$$

where $\mathscr{k}$ should be optimally found by minimizing (5-15) subject to (5-6) and (5-7). It should not escape our notice that the proposed cost function in (5-15) is differentiable, which facilitates the use of derivative-based optimization approaches; however, the convexity of (5-15) can be affected by the prediction engine formulation. Therefore, Eq. (5-15) is convex and differentiable if the prediction engine formulation is convex and differentiable with respect to its unknown variables. However, even if the prediction engine is non-convex, the proposed cost function can ameliorate the non-convexity of WPPI problem. As described earlier in Section 5.2.1, ELM possesses a linear formulation with respect to the unknown parameters. Hence, (5-15) becomes differentiable and convex, and the global solution can be determined by convex optimization techniques.

5-3-3 Specified Cost Function for ELM

In ELM, the prediction engine parameters related to the upper and lower bounds are independent of one another; hence, the cost function terms in (5-15) for the upper and lower bounds can be considered individually as shown in (5-16) and (5-17), respectively. In addition, minimizing the norm of output weights (i.e., $\overline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\beta}}$ in (5-1) and (5-2)) is also added to the cost function to minimize the output weights and avoid overfitting on training data in (5-16) and (5-17), similar to the prevalent ELM formulation in [128]:

$$F_{up}\left(\overline{u}, \overline{\boldsymbol{\beta}}\right) = c \,.\left(\sum_{n=1}^{L} C\left(\overline{\omega}, \overline{\Delta y}_n, \overline{\boldsymbol{a}}\right)\right) + \overline{\boldsymbol{\beta}}^T \overline{\boldsymbol{\beta}} \tag{5-16}$$

$$F_{lo}\left(\underline{u}, \underline{\boldsymbol{\beta}}\right) = c \,.\left(\sum_{n=1}^{L} C\left(\underline{\omega}, \underline{\Delta y}_n, \underline{\boldsymbol{a}}\right)\right) + \underline{\boldsymbol{\beta}}^T \underline{\boldsymbol{\beta}} \tag{5-17}$$

where c is the user-specified hyperparameter that should be optimally selected for trade-off between empirical and structural risks [128]. In ELM, output weights related to the upper bound

($\overline{\boldsymbol{\beta}} = \{\overline{\beta_1}, \dots, \overline{\beta_M}\}$) and lower bound ($\underline{\boldsymbol{\beta}} = \{\underline{\beta_1}, \dots, \underline{\beta_M}\}$) are unknown parameters of the prediction engine. $\overline{u} = \{\overline{\omega}, \overline{\boldsymbol{a}}, c\}$ and $\underline{u} = \{\underline{\omega}, \underline{\boldsymbol{a}}, c\}$ are sets of hyperparameters for (5-16) and (5-17), respectively.

5-4 The Proposed WPPI Model

In this section, the WPPI model formulation and its solution approach for different prediction steps are elucidated. It is worth mentioning that the direct prediction approach is applied to conduct multistep ahead prediction. Thus, for prediction of the jth-step ahead, J different prediction engines are trained and prediction of the jth-step ahead is independent of the $(j-1)$th-step ahead [12]. Direct WPPI has two chief merits. First, parallel computation can be employed, and therefore WPPIs related to different prediction steps can be simultaneously constructed. Second, for a nonlinear and chaotic wind power TS, direct prediction enhances the prediction accuracy and prevents accumulation of prediction errors that might occur in an iterative prediction [12].

5-4-1 WPPI Formulation

Considering the cost function in (5-16) and (5-17), the prediction engine parameters and the hyperparameters of the proposed WPPI for the jth-step ahead prediction can be formulated as a bi-level optimization, as described below:

$$\underset{u=\{\overline{u},\underline{u}\}\in U,\hbar=\left\{\overline{\boldsymbol{\beta}},\underline{\boldsymbol{\beta}}\right\}\in\mathcal{K}}{Min}\{-RL, PINAW\} \tag{5-18}$$

$$\hbar = \left\{\overline{\boldsymbol{\beta}}, \underline{\boldsymbol{\beta}}\right\} \in \arg\left\{\underset{\{\overline{z},\underline{z}\}\in\mathcal{K}}{Min}\left\{F_{up}(\overline{u},\overline{z}), F_{lo}(\underline{u},\underline{z})\right\}\right\} \tag{5-19}$$

Eq. (5-18) represents the upper-level optimization problem. To introduce the efficacy of the proposed WPPI model, in this paper, based on Section 5-3, a multi-objective optimization consisting of maximization of RL in (5-4) (i.e., minimization of $-RL$) and minimization of $PINAW$ in (5-5) is regarded as the upper-level optimization function. However, other objective functions, defined in the literature (e.g., skill score in [119]), can be employed based on SO preference. The lower-level optimization is presented in (5-19), and its cost function is described in Section 5.2.3 as (5-16) and (5-17). In (5-19), $\overline{z}$ and $\underline{z}$ correspond to prediction engine parameters and representative of $\overline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\beta}}$ in (5-16) and (5-17), respectively. The convexity of (5-19) is analyzed in Appendix C.

Figure 5-2 demonstrates the overall scheme of the proposed bi-level prediction model. The lower-level optimization function can be solved and $\overline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\beta}}$ can be found by classical convex optimization methods provided the variables in sets $\overline{u}$ and $\underline{u}$ are imported from the upper-level optimization problem.

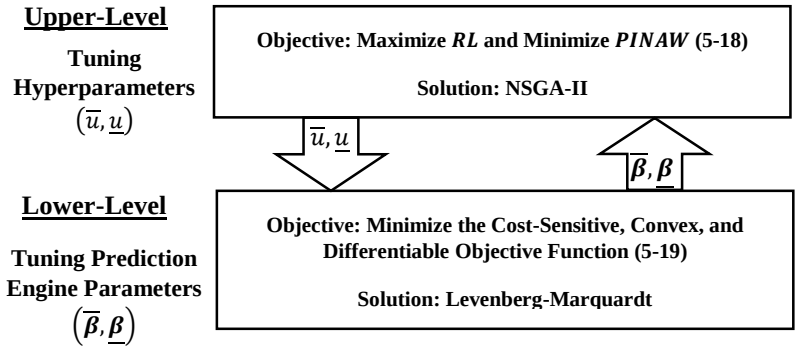

Figure 5-2: General scheme of the proposed WPPI.

5-4-2 Solution Approach

Classical convex optimization techniques are employed to solve the lower-level optimization, while the upper-level optimization is conducted heuristically by NSGA-II. Keen readers can find a description of this powerful meta-heuristic optimization tool in [124].

The general flowchart of the proposed solution approach for tuning parameters of the proposed WPPI formulation, introduced in Section 5.3.1, is provided in Figure 5-3. As shown in the figure, wind power TS is preprocessed in the first step of the training procedure, and any gaps in historical data are detected. The data gaps are filled using interpolation and extrapolation techniques;

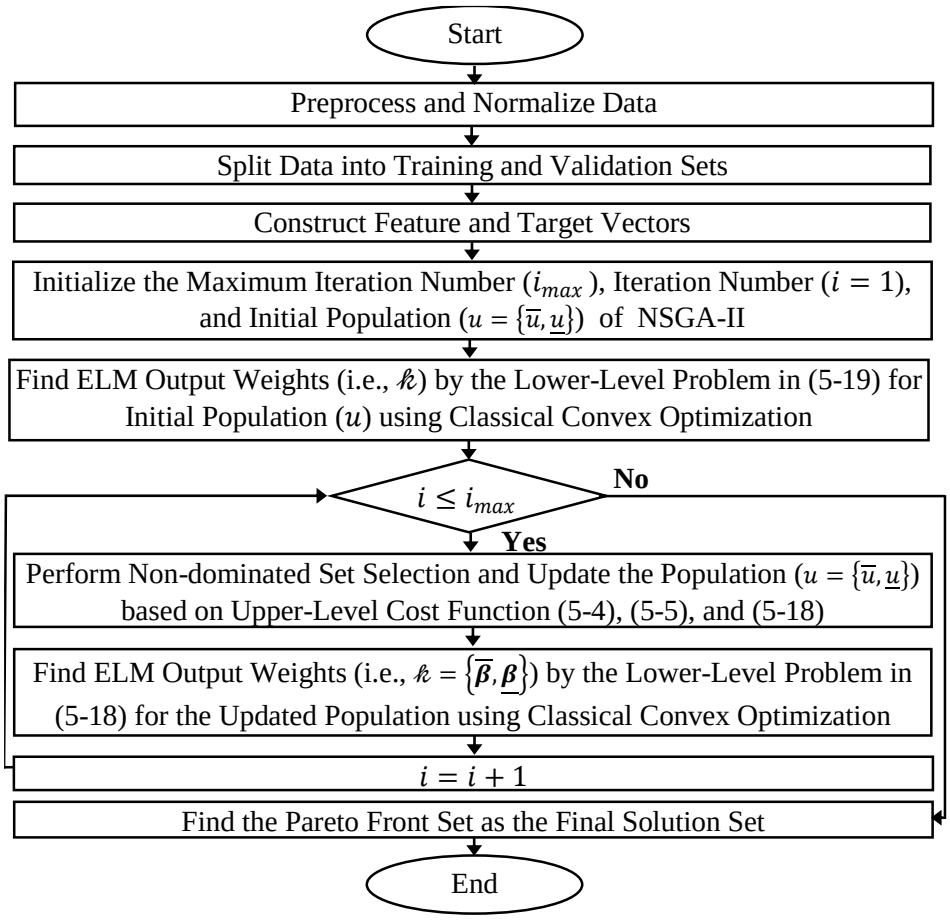

Figure 5-3: Flowchart of the proposed WPPI training procedure.

however, advanced outlier detection and removal techniques can be employed in this step without loss of generality. Detrending is conducted to reduce the nonstationarity and enhance the prediction accuracy. Data are normalized in the range of [0, 1] for a more efficient training procedure. In the second step, data are divided into training and validation sets. This helps to avoid overfitting while ensuring acceptable consistency in the performance of both training and validation sets. Next, the feature vector needs to be constructed using FS.

According to Section 5.2.1, Cao's embedding dimension, along with a mutual information-based approach for delay selection, is used. Afterward, NSGA-II is initialized, and a set of solutions for u are imported from the upper-level; then, the lower-level optimization described in (5-19) is solved for each set of u. Objective functions $F_{up}(\cdot)$ and $F_{lo}(\cdot)$ in the lower-level optimization problems are convex. Also, as shown in Appendix C, $F_{up}(\cdot)$ and $F_{lo}(\cdot)$ are differentiable. Therefore, global optimal solutions of $F_{up}(\cdot)$ and $F_{lo}(\cdot)$ (($\overline{\boldsymbol{\beta}}, \underline{\boldsymbol{\beta}}$)= ($\overline{k}, \underline{k}$)) are the root of $G(\overline{u}, \underline{u}, \overline{k}, \underline{k})$ defined as follows:

$$G(\overline{u}, \underline{u}, \overline{k}, \underline{k}) = \begin{bmatrix} G_{up}(\overline{u}, \overline{k}) \\ G_{lo}(\underline{u}, \underline{k}) \end{bmatrix} = \begin{bmatrix} \nabla_{k} F_{up}(\overline{u}, \overline{k}) \\ \nabla_{k} F_{lo}(\underline{u}, \underline{k}) \end{bmatrix} \tag{5-20}$$

where

$$\begin{aligned} G_{up}(\overline{u}, \overline{k}) &= c\left[\nabla_{\overline{k}} R(\theta, \overline{\boldsymbol{a}}) \circ \overline{\boldsymbol{\varepsilon}} + 2\overline{\boldsymbol{k}}\right]_{M\times 1} \\ G_{lo}(\underline{u}, \underline{k}) &= c\left[\nabla_{\underline{k}} R(\theta, \underline{\boldsymbol{a}}) \circ \underline{\boldsymbol{\varepsilon}} + 2\underline{\boldsymbol{k}}\right]_{M\times 1} \end{aligned} \tag{5-21}$$

$$\overline{\boldsymbol{\varepsilon}} = \left[\mathcal{E}(\overline{\omega}, \overline{\Delta y}_1) \quad \ldots \quad \mathcal{E}(\overline{\omega}, \overline{\Delta y}_L)\right] \tag{5-22}$$

$$\underline{\boldsymbol{\varepsilon}} = \left[\mathcal{E}\left(\underline{\omega}, \underline{\Delta y}_1\right) \quad \ldots \quad \mathcal{E}\left(\underline{\omega}, \underline{\Delta y}_L\right)\right] \tag{5-23}$$

where $\circ$ is the element-wise multiplication, known as the Hadamard product. The root of $G(\overline{u}, \underline{u}, \overline{k}, \underline{k})$ in (5-20) can be found using different nonlinear solvers. In this work, the MATLAB *Optimization Toolbox* [133] and *fsolve* function are put to use for finding the roots of $G(\overline{u}, \underline{u}, \overline{k}, \underline{k})$. The Levenberg-Marquardt method is employed to find the solution of (5-20) [39]. This algorithm is available in commercial packages and has several advantages, including robust performance, even in a near singular or singular Jacobian matrix, compared to other algorithms such as Newton's method. To improve the convergence, an appropriate initial set, $\{\overline{k}, \underline{k}\} = \{\underline{\boldsymbol{\beta}}^{(0)}, \overline{\boldsymbol{\beta}}^{(0)}\}$, is provided to *fsolve* using the ELM output weights in the deterministic prediction. The lower-level problem needs to be solved for different sets of hyperparameters provided from the upper-level. Parallel computation is performed to enhance the computation time. At every iteration, i, NSGA-II updates the non-dominated sets and population position $\{\overline{u}, \underline{u}\}$ based on its ELM output weights and its resultant upper-level objective function value. When $i > i_{max}$, this iterative process is stopped, and the Pareto front set is constructed from non-dominated sets. Therefore, an SO can choose the most appropriate solution set according to the problem. To reduce the computation time, the lower-level problem is solved in a parallel manner for different sets of $u = \{\overline{u}, \underline{u}\}$ in population.

5-5 Case Studies

5-5-1 Data Description

To evaluate the performance of the proposed WPPI, several realistic case studies are conducted using publicly available historical data, including accumulated wind power generation data from

the Alberta Electric System Operator (AESO) [62] and wind power generation for the Adelaide wind farm from the Independent Electric System Operator (IESO) [134].

AESO operates the power grid in Alberta, Canada and is authorized for the operating energy market. AESO wind power generation constitutes approximately 9% of installed generation capacity in Alberta, Canada. AESO's wind power generation data demonstrate different patterns in various months of year. To this end, evaluation of the proposed model for different months of a year and comparison to other WPPI models is conducted to assess the efficacy of the WPPI models. Eleven case studies related to different months (Feb. 2012 to Dec. 2012) are considered.

IESO operates the real-time power system and oversees the electricity market in Ontario, Canada. Individual wind farms possess more chaotic and intermittent behavior compared to aggregated wind power generation in a wide region. Therefore, prediction of wind power generation for the Adelaide wind power facility from 1 Jan. 2017 to 7 Apr. 2017 is considered to evaluate the performance of the proposed WPPI for individual wind farm power generation.

High-quality WPPI for intra-hour and hour-ahead resolution can address several severe impacts of the volatile and intermittent behavior of wind power on power system balance in the electricity market [45, 60, 108]. In this regard, this paper demonstrates the results of multistep ahead prediction for 1-hour ahead with 10-min resolution.

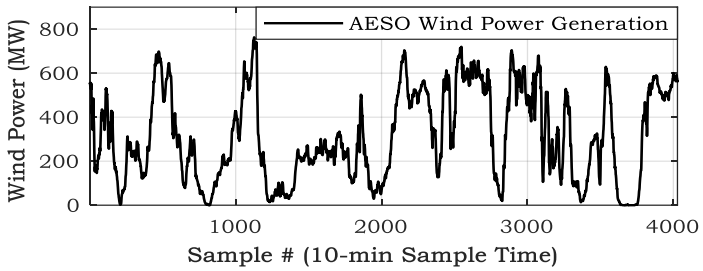

Figure 5-4: 10-min ahead wind power generation for AESO (Feb. 2012).

5-5-2 Training Dataset Length and Number of Hidden Neurons

Figure 5-4 shows the wind power generation in AESO in Feb. 2012, as an example. The wind power generation possesses a high level of volatility and intermittency, and therefore the proposed model is trained online in a daily manner to construct a model with historical data close to the prediction day. The WPPI model is not updated during a given day, as adding few points to the training dataset cannot significantly affect model accuracy yet increases the computational burden [92].

Training dataset length and number of hidden neurons are determined by analyzing data for the required prediction horizon. Here the process is described for 10-min ahead prediction in AESO wind power TS.

Figure 5-5 demonstrates the normalized root-mean-square error (NRMSE) [12] of both training and test datasets for 10-min ahead forecasting using ELM. Increasing the length of training dataset to more than 10 days results in minimal improvement but increases the computational burden. Also, the accuracy of prediction does not significantly improve significantly for approximately above 30 neurons. Notably, enlarging the training dataset and enhancing the number of neurons increase the computational burden of the training process. In this regard, 10 days of data for the

training dataset and 30 neurons are considered for 10-min ahead WPP prediction of the AESO. Empirical simulations for other prediction horizons verify that similar lengths of training dataset can lead to satisfactory training of the prediction model. However, the number of neurons varies based on the prediction horizon.

5-5-3 Benchmark Models and Evaluation Criteria

Persistence model (PM) is currently the most used classical benchmark model in renewable generation prediction [60]. Comparing the effectiveness of the proposed WPPI with this traditional benchmark can provide insightful information about the efficacy of the proposed approach in comparison to other models. The recently proposed QR-based ELM [60] is the second benchmark model considered. A prediction model called the hybrid intelligent algorithm (HIA) [119] is another ELM-based WPPI and is used as the third benchmark model. To compare the ability of the proposed ELM-based method against previous models, a feed-forward NN-based WPPI [121] called the lower upper bound estimation (LUBE) is used as the fourth benchmark model. Prediction models are trained with similar feature sets and training datasets; the exception is PM,

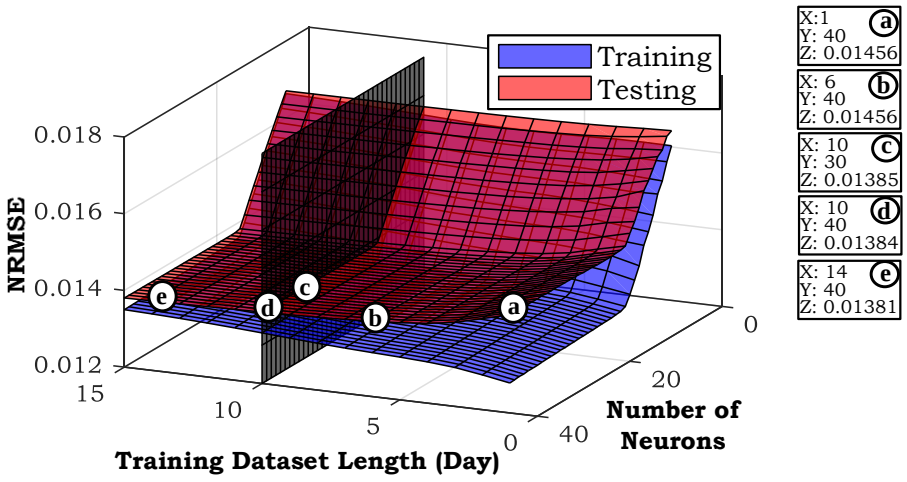

Figure 5-5: NRMSE for 10-min ahead forecasting.

for which the mean and variance of the latest observations are used to construct a Gaussian distribution representing the uncertainty of forthcoming sample points [119]|.

RL and $PINAW$, which are respectively described in (5-4) and (5-5) are widely used in literature [19], are the first two evaluation metrics. Skill score (Sc) is another evaluation metric that considers $PINAW$ along with RL; therefore, it provides an overall score for a WPPI model. Sc is used as the third evaluation metric and is defined as follows [119]:

$$Sc = \frac{1}{L}\sum_{n=1}^{L} \mathcal{K}\left(\underline{y_n}, \overline{y_n}, RL^*\right) \tag{5-24}$$

where

$$\mathcal{K}\left(\underline{y_n}, \overline{y_n}, RL^*\right) = \begin{cases} \mathcal{B} - 4\left(\underline{y_n} - y_n\right), & y_n < \underline{y_n} \\ \mathcal{B}, & otherwise \\ \mathcal{B} - 4(y_n - \overline{y_n}), & y_n > \overline{y}_n \end{cases}$$

$$\mathcal{B} = -2.(1 - RL^*) \cdot \left(\overline{y_n} - \underline{y_n}\right)$$

where RL^* is the preferred RL. Sc is a negative-sign evaluation metric, and a prediction model with Sc closer to zero possesses better performance from the Sc viewpoint. It is worth mentioning that all of these evaluation indices might need to be considered together to identify the most superior prediction model. For example, there might be a WPPI in which $|Sc| \to 0$ but $RL \ll RL^*$. In such cases, the operator would probably prefer a WPPI with less deviation from RL^* but featuring constructed WPPIs that are narrow ($PINAW \to 0$).

Table 5-1: Evaluation of the Proposed WPPI in 1-Step Ahead Prediction for Feb. 2012 and $RL^* = 90\%$.

Evaluation Metrics \ Max. Degree	2^{nd}	4^{th}	6^{th}	8^{th}
$PINAW$ (%)	3.8107	3.607	3.5901	3.5807

5-5-4 Test Results and Analysis for AESO

5-5-4-1 Sensitivity Analysis

As described in Section 5-3-2, the degrees of polynomials $R(\theta, \overline{\boldsymbol{a}})$ and $R(\theta, \underline{\boldsymbol{a}})$ in (5-21) impact the quality of the WPPI. In addition, increments in the degrees of polynomials can result in increases in the number of hyperparameters, which should be optimally selected from the upper-level problem. Hence, sensitivity analysis is conducted to determine the appropriate degrees of $R(\theta, \overline{\boldsymbol{a}})$ and $R(\theta, \underline{\boldsymbol{a}})$. WPPIs related to 1-step ahead prediction during Feb. 2012 for different degrees of polynomials are constructed, and the results for $RL^* = 90\%$ are summarized in Table 5-1. The resultant RLs, using different degrees of polynomials, are greater than 90%; hence, the RL criterion is met. Moreover, the $PINAW$ decreases by increasing the degree of the polynomial; however, the maximum increment occurs by the 4^{th}-degree polynomial. The number of hyperparameters increases by increasing the degree of the polynomial, and so the 4^{th}-degree polynomial is used in this paper as a tradeoff. Therefore, according to Section II-B, $\overline{\boldsymbol{a}}$ and $\underline{\boldsymbol{a}}$ only consist of two elements. In this paper, $\underline{\boldsymbol{a}} = \overline{\boldsymbol{a}}$, and therefore, considering $\overline{\omega}$, $\underline{\omega}$, and c, five parameters need to be tuned by the upper-level optimization.

5-5-4-2 Analysis of Diversity in Pareto Front

The first characteristic examined here is the Pareto frontier of the proposed WPPI model, so as to construct WPPIs that ensure that the Pareto set consists of solutions for a wide range of RL and

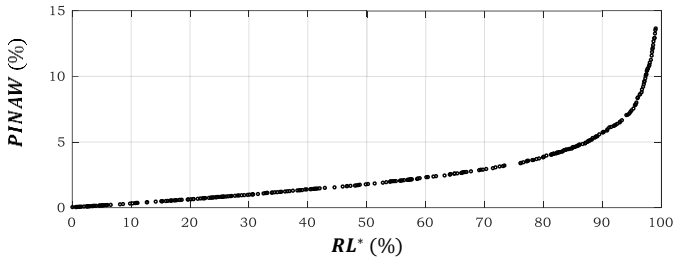

Figure 5-6: Pareto front derived by the proposed WPPI for 1-step ahead (10-min ahead) prediction.

$PINAW$. In this regard, a Pareto frontier for the AESO for 10-min ahead prediction is shown in Figure 5-6. This figure shows optimal WPPI models can be achieved for various values of RL^*, which is of great importance to SOs, and as expected when training WPPI models with lower RL^*, $PINAW$ is much lower than its value for realizing higher RL^*.

5-5-4-3 Steadiness in WPPI for Training and Testing

Consistency in performance of the proposed WPPI model in training and testing is of particular interest. Any overtraining or undertraining is detected by consistency analysis. To conduct this analysis, the WPPI was run for Feb. 2012 to Dec. 2012, and Sc values for $RL^* = 90\%$ calcauted for both the training and test datasets in each month are calculated. Figure 5-7 demonstrates the Sc values for different months in 1-step (10-min) ahead prediction and shows a strong consistancy between Sc values related to the training and test datasets. The average and maximum deviation

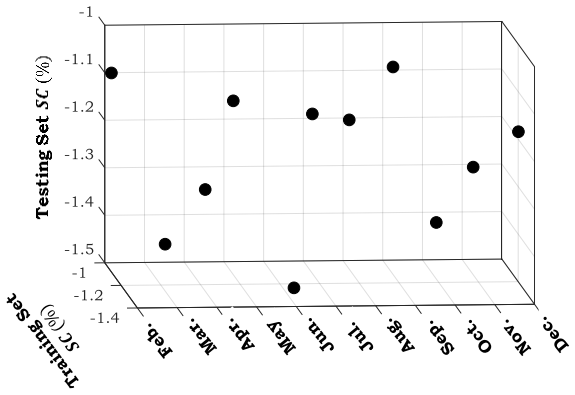

Figure 5-7: Sc values for various months in training and test datasets for $RL^* = 90\%$.

between the Sc values of the training and test datasets are about 0.07 and 0.2%, respectively. Thus, empirical results confirm that the model is not overtrained or undertrained. In this regard, the SO can estimate the Sc of the test dataset from the Sc of the training dataset.

5-5-4-4 Comparison with Benchmark Models for Various Months

After investigating the capability of the proposed WPPI for constructing WPPIs with any RL^* in Section 5.4.4.2 and observing the consistency of the WPPI model in Section 5.4.4.3, we now compare our model against other models.

Due to the high importance of $RL^* = 90\%$ in many power system operation problems, evaluation of WPPI models for $RL^* = 90\%$ is well-known in the literature [1, 45, 127]. Figure 5-8 demonstrates Sc and $PINAW$ values of WPPIs for $RL^* = 90\%$ in different prediction steps that are averaged from Feb. 2012 to Dec. 2012. Amongst WPPI models, the proposed WPPI presents superior performance in terms of Sc. In this figure, the result of PM is not reported due to very large $PINAW$ and Sc thereof which makes the comparison with other WPPIs difficult. QR demonstrates the best performance in terms of providing sharp WPPIs for different prediction steps

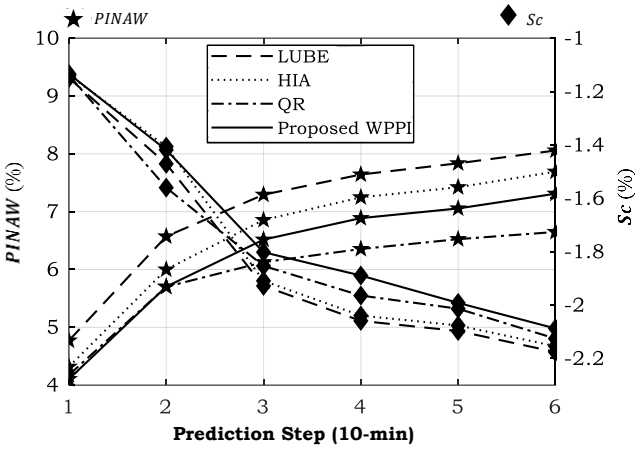

Figure 5-8: Average $PINAW$ and Sc evaluation of WPPIs with $RL^* = 90\%$ for the proposed WPPI and benchmark models for different prediction steps.

Table 5-2: Performance Evaluation of Different Prediction Models for Various Months in 6-step Ahead Prediction for $RL^* = 90\%$

Month	PM		LUBE		HIA		QR		Proposed WPPI		Avr. $PINAW$ Improvement (%)
	PINAW (%)	***RL***(%)	***PINAW*** (%)	***RL***(%)	***PINAW*** (%)	***RL***(%)	***PINAW*** (%)	***RL***(%)	***PINAW*** (%)	***RL***(%)	
Feb.	8.77	96.67	6.66	90.05	6.53	92.67	4.90	88.98	**5.55**	**90.76**	22.80
Mar.	11.54	93.06	7.09	88.07	6.45	85.84	7.05	86.13	**8.70**	**91.07**	24.63
Apr.	9.53	95.35	7.61	91.01	6.77	91.42	5.36	87.74	**6.14**	**90.66**	21.43
May	9.33	95.52	6.97	89.71	6.23	90.49	5.21	88.23	**5.70**	**90.33**	23.62
Jun.	11.57	95.30	8.05	89.01	7.06	88.20	7.56	90.05	**7.05**	**90.03**	22.90
Jul.	8.34	94.89	6.09	90.01	6.25	91.27	5.88	89.07	**5.20**	**90.20**	23.05
Aug.	9.05	95.27	6.89	93.05	5.30	89.81	5.45	88.25	**5.60**	**91.07**	28.14
Sep.	7.01	92.79	6.30	91.05	6.97	89.90	5.80	90.02	**5.47**	**92.06**	13.59
Oct.	10.64	94.11	8.80	89.40	6.67	87.98	6.06	86.79	**6.76**	**90.10**	36.42
Nov.	9.73	94.72	4.95	86.91	7.47	90.45	6.24	87.06	**6.28**	**90.01**	25.73
Dec.	8.02	92.72	7.85	89.77	6.73	90.43	5.58	88.51	**6.40**	**91.00**	12.56

but is unable to construct WPPIs with $RL \geq RL^*$, which results in higher $|Sc|$ compared to the proposed WPPI. The proposed WPPI model ranked second in terms of $PINAW$ while $RL^* \geq RL$ for all prediction steps. In this regard, the proposed model outperforms benchmark models in the case where meeting RL^* becomes crucial with minimum $PINAW$ for the SO.

Table 5-2 summarizes the performance of different WPPI models in 6-step (1-hour) ahead prediction of wind power in various months for $RL^* = 90\%$. This table shows that PM constructs the most conservative WPPIs, and $RL \gg RL^*$. Consequently, the sharpness has been substantially

sacrificed. Due to the nonstationarity of the data, the performance of the WPPI models can vary for different months.

The results related to the WPPI with the sharpest WPPIs for $RL \geq RL^*$ are presented in bold. The proposed WPPI model meets the RL requirement ($RL \geq RL^*$) while its $PINAW$ is substantially lower than benchmark models that satisfy the RL requirements.

The last column in Table 5-2 demonstrates the average relative improvement in reducing the $PINAW$ as a result of employing the proposed WPPI. The PINAWs related to WPPI models in which $RL < RL^*$ are excluded when calculating the average $PINAW$ improvement for every month. The WPPIs constructed by the proposed model improve the $PINAW$ by at least 12.56%. Additionally, it can be concluded from Table 5-2 that the proposed WPPI model demonstrates more consistency in prediction during different months in contrast to the benchmark models.

Figure 5-9 presents the constructed 1-hour ahead WPPIs for the 2nd week of Feb. 2012. To perceive the effectiveness of the proposed WPPI in comparison to the benchmark models, their resultant WPPIs are demonstrated in Figure 5-9 for two periods, in which QR and HIA have the best performance among the benchmark models. This figure clearly shows that the wind power changes significantly during this week; however, the proposed WPPI is proficient at constructing WPPIs for this period while HIA and QR may result in abrupt and irregular changes.

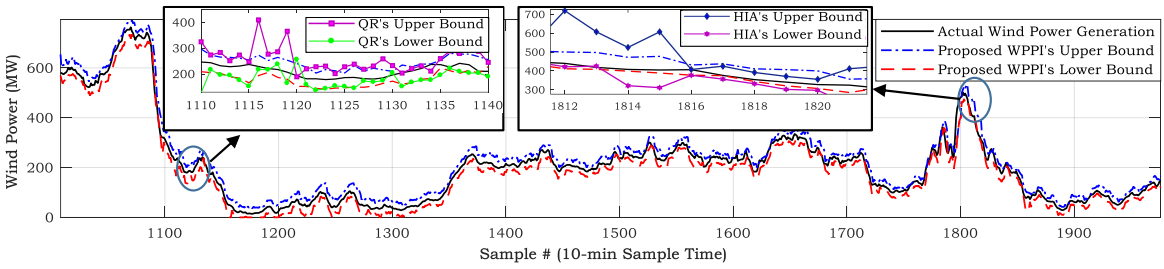

Figure 5-9: WPPIs with $RL^* = 90\%$ constructed by the proposed WPPI in the 2nd week of Feb. 2012.

Table 5-3: Summary of Performance Evaluation of WPPI Models in 1-Step Ahead for Multi Runs and $RL^* = 90\%$

Prediction Models	$PINAW(\%)$		$RL(\%)$	
	$Mean$	Std	$Mean$	Std
LUBE	6.014	1.853	91.051	1.010
HIA	5.871	1.221	92.963	0.350
Proposed WPPI	4.010	0.004	90.861	0.151

5-5-4-5 Reproducibility of WPPIs and Independency of the Proposed WPPI to Initial Values

As mentioned in this paper, the main motivation behind developing the proposed WPPI stems from issues caused by heuristically tuning a large number of parameters in prevalent WPPI. As stated in Section 5-5-4-1, the proposed WPPI still features five hyperparameters that must be tuned heuristically. Table 5-3 demonstrates the mean and standard deviation of $PINAW$ and RL associated with 10 runs of LUBE, HIA, and the proposed model in conducting 1-step (10-min) ahead prediction during Feb. 2012 with $RL^* = 90\%$. The proposed model outperforms other heuristic optimization-based models in terms of reproducibility of high-quality WPPIs in different runs. Therefore, the proposed model can lead to a reduced possibility of becoming trapped in local minima. Notably, similar results were observed for other case studies and prediction horizons.

5-5-4-6 Convergence and Computation Burden

Convergence and the resulting computational burden of WPPI for short-term wind power prediction are of great importance. Figure 5-10 describes the convergence of solving the convex optimization demonstrated in Eq. (5-20). As shown, the lower-level problem can be solved within

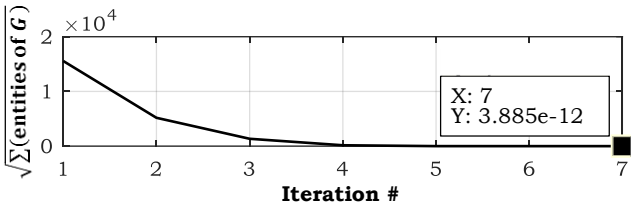

Figure 5-10: Convergence of convex optimization of Eq. (5-20).

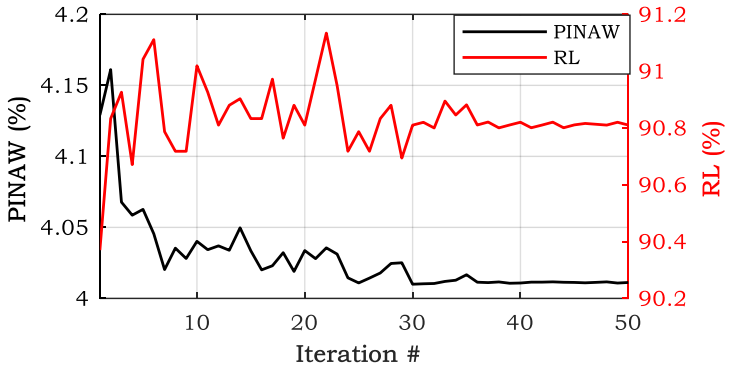

Figure 5-11: Convergence of proposed WPPI for $RL^{*} = 90\%$.

a few iterations. Figure 5-11 shows the overall convergence of the proposed model in 1-step ahead prediction of AESO during Feb. 2012, where the proposed WPPI model can be tuned and converged within limited iterations.

All prediction models are developed using a 3.4 GHz Intel Core i7-6700 CPU with 16 GB of RAM. Average training required by all prediction models for 1-step ahead prediction are reported in Table 5-4.

The training process is conducted once a day when prior day data are available. In the AESO case, wind power samples are available every 10 minutes; the model can be trained at 23:50 every

Table 5-4: Summary of Computation Time for Training

Prediction Models	Training Time (s)
PM	0.01
LUBE	51.01
HIA	40.11
QR	27.63
Proposed WPPI	160.04

day to predict the wind power from 00:00 to 23:59 of the next day. During the test, it is observed that computations in all prediction models can be completed within fractions of a second. Thus, all prediction models are feasible for short-term WPPI.

5-5-4-7 Test Results and Analysis for IESO

To evaluate the performance of the proposed model in prediction of generated power related to a single wind farm in IESO, a one-hour ahead WPP with $RL^* = 90\%$ is conducted for the Adelaide wind power generation facility. To make the proposed WPP comparable with two state-of-the-art prediction models [1], specifically the NCWC-based Elman and NCWC-based nonlinear auto-regressive exogenous (NARX) WPP models, for this case study the models are evaluated for wind power generation from 1 Jan. to 7 Apr. 2017; this is similar to [1] from which results quoted.

To make the comparison more comprehensive and in line with [1], $PIMSE$ [1], which is defined in (5-25), is measured.

$$PIMSE = \frac{1}{L}\sum_{n=1}^{L}\left[(\overline{y_n} - y_n)^2 + (\underline{y_n} - y_n)^2\right] \tag{5-25}$$

This metric signifies large deviations from actual wind power and, from the $PIMSE$ prospective, a prediction model is more skillful if it leads to lower $PIMSE$. Table 5-5 summarizes the performance of the proposed model and two state-of-the-art models. The proposed model leads to significantly sharper WPPIs ($\frac{68.63-44.66}{68.63} \times 100 \cong 35\%$ perfection) with the required RL, and its

Table 5-5: Performance Evaluation of NCWC-based Elman and NARX WPP Models and the Proposed WPPI for Prediction for $RL^* = 90\%$

Evaluation Metrics	NCWC-based Elman [1]	NCWC-based NARX [1]	Proposed WPPI
$\boldsymbol{RL}$(%)	93.81	93.04	**90.38**
$\boldsymbol{PINAW}$ (%)	63.71	68.63	**44.66**
$\boldsymbol{PIMSE}$	1.42	1.64	**0.73**

$PIMSE$ is also substantially lower than other benchmark models. This case study is a further testimony to the efficiency of the proposed model.

5-5-4-8 Discussion

As experimental results elucidate, the proposed WPPI model possesses superior performance for short-term WPPI compared to several state-of-the-art well-established prediction models.

The proposed WPPI model benefits from a low number of heuristically tuned hyperparameters which results in a marked reduction in the required search space for the heuristic optimization process while the main parameters are tuned through a convex optimization process. The reproducibility and convergence of the proposed model are verified.

The proposed model presents robust performance for different datasets, while such robustness not observed in benchmark models. In line with [1, 45, 60, 118-121, 123], this study only presents WPPIs for horizons up to 1-hour, which make up a portion of short-term WPPI [117]. Wind power is closely related to meteorological variables, such as wind speed, wind direction, etc. Numerical weather prediction results are also beneficial for prediction for longer horizons. Performing accurate prediction for a look-ahead time of more than one hour requires advanced FS due to a large number of feature candidates. Moreover, as revealed in the case studies, the performance of WPPI models for a single wind farm is lower than observed for accumulated wind power generation. The main reason is because the TS of AESO is affected by spatial smoothing, whereas the TS of Adelaide wind farm is volatile with all 18 wind turbines located in a small area. Spatiotemporal modeling of this wind farm, accounting for meteorological features from the vicinity, might improve the prediction accuracy.

5-6 Conclusion

This paper presents a framework for short-term WPPI that benefits from a bi-level optimization structure along with a convex and differentiable cost function for optimally tuning unknown parameters of the prediction engine. In the proposed WPPI, prediction engine parameters can be globally tuned via classic optimization while only a few hyperparameters are tuned with heuristic approaches from the upper-level optimization problem. Hyperparameters are found based on a multi-objective optimization procedure, which results in a Pareto front with a wide range of $PINAW$ and RL. The comprehensive case studies reveal consistency in the results of WPPI evaluations for training and test datasets. Moreover, case studies validate the superior effectiveness of the proposed WPPI framework compared to traditional benchmark and state-of-the-art prediction models. The proposed WPPI model is competent with respect to forming sharp and reliable WPPIs for different datasets. SOs can benefit greatly from the proposed WPPI in various short-term power system operation tasks.

Further research could be conducted to adapt the proposed approach to solar prediction. Moreover, application of the proposed bi-level optimization framework in considering various SO-defined objective functions can be investigated.

6- TIDAL CURRENT AND LEVEL UNCERTAINTY PREDICTION VIA ADAPTIVE LINEAR PROGRAMMING[7]

Short-term uncertainty prediction modeling of tidal power generation supports power systems in reserve and regulation markets. In tidal power generation via various tidal energy harvesting technologies, tidal current and level are the most influential factors. This paper addresses a nonparametric prediction interval (NPI)-based uncertainty model thereof. The proposed model adapts a bi-level optimization formulation, based on extreme learning machine (ELM) prediction engine and quantile regression (QR). The quantile probabilities are asymmetrically and adaptively chosen in the upper-level optimization to make prediction intervals (PIs) sharper for a specific reliability level (RL). Besides, the training process of ELM is improved by adaptively selecting ELM's hidden neurons via upper-level optimization. The lower-level optimization finds ELM's output weighting coefficients through linear programming of QR. The heuristic optimization, consisting of gray wolf optimizer and simplex method, is designed to facilitate the NPI with high exploration and exploitation capabilities in upper-level optimization. The performance of the proposed NPI is examined using empirical data recorded in three different sites, located in North America. The results of case studies show that the proposed NPI can provide sharper PIs in comparison to the well-tailored rival models whilst a pre-specified RL criterion is met.

6-1 Nomenclature

Abbreviations

[7] © 2018 IEEE. Reprinted, with permission from [135] N. Safari, S. M. Mazhari, B. Khorramdel, and C. Y. Chung, "Tidal Current and Level Uncertainty Prediction via Adaptive Linear Programming," *IEEE Trans. Sustain. Energy, Early Access.*

TL	Tidal level.
TCS	Tidal current speed.
TCD	Tidal current direction.
NPI	Nonparametric prediction interval.
SVM	Support vector machine.
PI	Prediction interval.
RL	Reliability level.
QR	Quantile regression.
LP	Linear programming.
CDF	Cumulative distribution function.
TS	Time series.
AAQR	Asymmetrically adaptive QR.
LSSVM	Least squares SVM.
ELM	Extreme learning machine.
GWO	Grey wolf optimizer.
GA	Genetic algorithm.
PSO	Particle swarm optimization.
PINAW	Prediction interval nominal average width.

6-2 Introduction

Due to growing concerns about climate change and fossil fuel depletion, renewable energy sources are considered as a feasible substitute for conventional environment-polluting fossil-

fuel-based power plants. To this end, various alternative energy sources have been found and successfully deployed all over the world. Amongst them, tidal energy has recently grabbed remarkable attentions across governments and power industries [136], owing to high predictability feature of influencing factors in tidal energy harvesting [137].

With the increasing penetration of tidal power generation, output fluctuation and variation—known as intrinsic characteristics of this technology—impose severe and unprecedented challenges on power systems operation [78]. Hence, the further need for employing expensive dispatchable fast generation units for operational reserves and regulation purposes is raised. Accurate short-term uncertainty prediction tools provide reliable estimates of tidal energy and significantly decrease the system uncertainty. Consequently, their employment causes a noticeable reduction in the total cost of power systems operation with requiring less backup generation units [138]. By virtue of short-term uncertainty prediction importance, this work is devoted to this topic.

Tidal power generation depending on the tidal energy harvesting technology is mainly influenced by tidal level (TL), tidal current speed (TCS), and tidal current direction (TCD) [139]. These three factors are the chief culprits for uncertainty and fluctuation in tidal power generation, and their direct relations with tidal power are shown in [78], [140, 141]. Ergo, researchers zero in on developing prediction models for these factors in tidal energy harvesting applications (e.g. [137], [138], [140]). It is worth mentioning that there are similar trends in modeling any other variable renewable energy resources. As instances, in [42, 110], wind speed prediction models are developed for wind energy application, while in [142, 143], solar irradiance prediction models are developed for solar energy harvesting. Despite the importance of uncertainty modeling of both tidal current and TL, the most recent studies have mainly focused on TCS and

TCD (e.g., [79, 80, 127, 137, 138, 144]).

Several models have been developed in recent years for tidal prediction. Approaches based on harmonic analysis, which provide rough estimates of tidal current and level, have been traditionally used for this problem [145]. However, these methods suffer from two critical drawbacks: a) they require large historical datasets (several years) [145] and b) they have inefficient performances with respect to predicting sudden and aperiodic patterns resulting from weather-based factors [140, 146]. To overcome these barriers and consider the nature of tides and the weather-based irregularities, machine learning-based prediction models have been proposed [145]. Owing to striking advancements in this topic, several powerful tidal prediction models have been proposed and successfully applied in recent years. These studies can be categorized into two main groups: deterministic [79, 127, 137, 138, 145] and probabilistic [127, 144] prediction models.

While deterministic prediction models focus on minimizing the prediction error associated with a single output [137], probabilistic prediction models provide sufficient information about the prediction uncertainty, which is most welcome by power systems operators and planners [144]. In this regard, the probabilistic prediction is also referred as uncertainty prediction [1, 144]. Uncertainty prediction modeling of tidal is in its infancy, but uncertainty prediction of other renewable energy sources is well discussed with respect to different types of renewable energy generation, including wind [1] and solar [147]. Moreover, widely used parametric uncertainty prediction models are not applicable to tidal prediction as any assumptions with respect to the probability distribution of tidal data can result in erroneous uncertainty prediction [127]. Hence, few nonparametric prediction interval (NPI) models, which fit into the probabilistic prediction domain, have been proposed for TCS and TCD [127, 144]. In NPI,

uncertainty prediction of tidal data is conducted, relying upon historical data in lieu of making assumptions about the probability distribution shape of tidal data [127]. In [144], a feed-forward neural network is heuristically trained based on nonlinear and non-differentiable cost function; however, such process may become trapped in local solutions [127]. In [127], support vector machine (SVM) is trained based on the traditional quadratic programming problem formulation, while few hyperparameters are heuristically tuned to construct prediction intervals (PIs) with a required reliability level (RL). In comparison to the training process of prediction engines in which their parameters can be acquired by solving a linear system (e.g., [128, 148]), tuning SVM parameters might be a time-consuming task [128]. An NPI formulation can be formed by a linear representation using the quantile regression (QR) concept [60, 123, 149]; the problem can then be solved by the state-of-the-art linear programming (LP) packages, leading to global optima. Nevertheless, to the best of our knowledge, any QR-based NPI with/without LP has not yet been introduced for tidal prediction.

In QR-based NPI, models are trained for creating upper and lower quantiles, known as prediction intervals (PIs), to encircle the observed value with a specific RL [150]. For a specified RL, determined by the system operator, in QR-based NPI, the probabilities of the quantiles pair are symmetrically chosen with respect to the probability of the median from the cumulative distribution function (CDF) for wind and solar power prediction [60, 147]. In [123], the quantile probabilities are selected asymmetrically in an offline procedure for wind power prediction. Besides, in all the thus far QR-based NPI models, the difference between the tuning parameter of the upper and lower quantiles probabilities is equal to RL [147]. However, suchlike selection of quantile probabilities may result in low-quality uncertainty prediction for two main reasons. Firstly, the QR-based NPI suffers from inevitable regression error comparable to any other

regression tools [151]. Since QR-based NPI model does not directly evaluate empirical RL and sharpness of the constructed PIs, the resultant PIs might not satisfy the power system operator's requirements. Secondly, as elaborated within this paper, symmetrical selection of quantile probabilities might not bring about the sharpest PIs with the required RL. Notably, due to non-stationarity of tidal time series (TS) [144], offline selection of quantile probabilities might not result in an efficient NPI model, particularly for test dataset. To circumvent these impediments, in this paper, a bi-level optimization framework is put forward to select quantile probabilities asymmetrically and adaptively by evaluating the RL and sharpness of the PIs constructed by QR-based NPI. In this regard, the proposed NPI model is called asymmetrically adaptive QR (AAQR)-based NPI.

The prediction engine is one of the essential building blocks of an NPI and is adopted considering the cost function structure. Because the AAQR-based NPI, employed in this paper, is developed by an LP formulation; the prediction engine should also have linear formulation. Amid prediction engines with linear formulations, least squares SVM (LSSVM) [148] and extreme learning machine (ELM) [128] have been successfully applied in a wide range of applications (e.g. [147]). The specialized literature shows that ELM is not only more computationally efficient than LSSVM but also provides higher accuracy in prediction applications [128]. However, the ELM's performance depends on the random numbers selected as the hidden neuron parameters; and consequently, there might be inconsistency in ELM performance for different runs [152]. Hence, adaptive approaches are recommended to adjust these parameters [152]. However, this has not be taken into consideration in tidal uncertainty prediction.

This paper devotes specific consideration to gray wolf optimizer (GWO) and simplex method to adaptively tune the hidden neurons parameters of ELM. GWO is a powerful swarm intelligence optimization method with high exploration ability that benefits from fast convergence and few adjustable parameters compared to traditional meta-heuristic optimization techniques, such as differential evolution, genetic algorithm (GA), etc. [153]. GWO leads to satisfactory results in many power system problems compared to other traditional meta-heuristic optimization techniques, such as GA, particle swarm optimization (PSO), etc. [154, 155]. The simplex method is an optimization technique that is popular due to its high exploitation capability and its ability to fine-tune parameters [148]. Therefore, a self-adaptive ELM is addressed based on a hybrid optimization strategy, incorporating GWO and simplex method to result in a heuristic optimization algorithm with high exploration and exploitation capability.

This paper proposes an NPI model for tidal prediction application. Inspired by the QR and LP formulation, AAQR-based NPI is put forward. It employs hyperparameters to select the optimal quantile probabilities by an upper-level optimization automatically. A hybrid GWO and simplex method is devised to set the hidden neuron parameters, as well as the hyperparameters in the upper-level optimization. While LP conducts the lower-level optimization to find the ELM's output weighting coefficients using the parameters, injected from upper-level optimization. A wide range of tidal data is used to demonstrate the effectiveness of the developed approach, followed by a presentation of results with comparison to those of prior research. The main contributions of this work are three-fold:

- A QR-based NPI with LP formulation for tidal uncertainty prediction;
- A bi-level optimization framework for the AAQR-based NPI; and
- A hybrid GWO and simplex algorithm for parameter tuning.

The remainder of the paper is organized as follows: Section 6-3 sheds light on the problem in tidal uncertainty prediction. In Section 6-4, the QR-based PI model for tidal uncertainty prediction is described. Section 6-4 clarifies the need for asymmetrically adaptive QR formulation and developing the AAQR-based NPI in tidal prediction. Section 6-6 describes the proposed AAQR-based NPI. Comprehensive empirical evaluation and analysis are provided in Section 6-7. Finally, the main concepts and conclusions are provided in Section 6-8.

6-3 Problem Description of Tidal Uncertainty Prediction in Power System Operation

Tidal energy can be converted into electrical energy via three main technologies, including tidal range, tidal stream, and hybrid technologies [139]. Hybrid technology is the combination of the first two technologies and is still in the research and development stage. In this paper, we mainly focus on the first two developed technologies.

Tidal range is the most traditional harvesting technology. A general scheme of a tidal barrage, which is the common form of tidal range technology, is shown in Figure 6-1. As illustrated in this figure, a dam-like construction is required, and sluice gates are devised at both sides of the dam. There are three different operating modes for this type of tidal energy harvesting technology, including ebb generation, flood generation, and two-way generation. As an example, Figure 6-1 describes the ebb generation mode; the gates are open, and the water can flow into the reservoir in

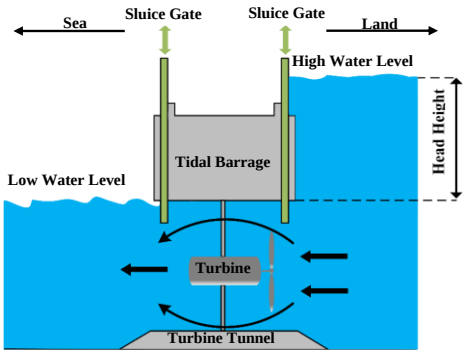

Figure 6-1: Conversion of potential energy of tides to electrical energy in tidal range.

the landside and fill it up during the flood tide. Once water reaches the maximum level, the gates close and the water is trapped into the landside of the dam [156]. As the tides ebb in the seaside, a gradual head height difference occurs between water levels in different sides of the dam. Then, the gates open and the turbine produces electric power by the potential energy resulted from such difference. The tidal electrical power can be estimated as follows [30]:

$$P = \rho g H \eta Q_t, \tag{6-1}$$

where ρ is the water density ($\approx$1025 kg/m^3), g is the acceleration related to earth's gravity, and $H = Landside\ Water\ Level - Seaside\ Water\ Level$. η and Q_t are turbine's efficiency coefficient and discharge across the turbine, respectively. η and Q_t are determined based on a hill chart providing that H is available [30]; but H is specified by TL. As can be found from (6-1) and above descriptions, predicting TL is considerably imperative for estimating available tidal power in advance.

Figure 6-2 shows a sketch of a tidal stream turbine. The power as a result of moving masses of water due to tidal current can be expressed by [78]:

$$P = \begin{cases} 0 & 0 \le W < W_{cutin} \\ 0.5 C_p \rho A W^3 & W_{cutin} \le W < W_{rated} \\ P_{rated} & W_{rated} \le W \end{cases} \tag{6-2}$$

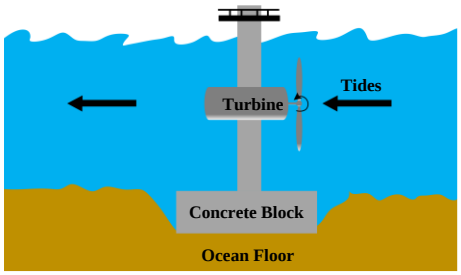

Figure 6-2: Conversion of the kinetic energy of tides to electrical energy in a tidal stream.

where A is the swept area of the turbine's blades, and W is the effective TCS, faced by blades. C_P ($0.4 \leq C_p \leq 0.5$ [78]) is turbine's power coefficient. Models have been developed to estimate C_P based on the effective TCS [29]. The effective TCS can be obtained based on the TCS and TCD measurements from measuring units. W_{cutin} and W_{rated} are cut-in and rated effective TCS, respectively. As can be perceived from Eq. (6-2), the output power of tidal stream is proportionate to velocity cubed. There-fore, relatively small errors in the effective TCS, and TCD prediction can result in an operationally significant error [146].

Tidal TS possesses non-stationarity [144], and correspondingly any assumption about tidal TS uncertainty distribution is erroneous. In timespans less than or equal to one hour, the non-harmonic components exist in tidal data as a result of supra-tidal fluctuations, which cannot be modeled by conventional harmonic analysis-based prediction models [146]. Precise short-term (e.g., 10-minute and 1-hour ahead) uncertainty prediction of tidal power helps to firm up regulation actions and reserve market. It also assists system operators in better managing the dispatchable generation and realizing the generation-load balance with the optimal cost [147, 157]. The facts, mentioned above, motivate us to focus on short-term uncertainty prediction of TL, TCS, and TCD data using a nonparametric PI model.

6-4 QR-based NPI Model for Tidal Uncertainty Prediction

Hereafter the tidal TS, which can be TCS, TCD, or TL, is presented by $\{y(nT_S)\}_{n=1}^{n=N}$. N is the total number of tidal data samples that are recorded every T_S. At time $t_0 = (n-1)T_S$, the uncertainty in $y(t)$, related to $t = nT_S$, can be predicted by its predictive probability distribution function (PDF), or CDF (F_t). As an instance, let assume that the predictive PDF of $y(t)$ is a uniform distribution. This assumption is made for illustration purposes, while there is no supposition underlying the proposed prediction model. Considering CDF of the uniformly distributed random variate, $y(t)$, in Figure 6-3, with the probability of $\bar{\alpha}$, the tidal TS value at time t is less than $\overline{y}(t)$. Also, one can note that with the probability of $\underline{\alpha}$, the tidal TS value at time t is less than $\underline{y}(t)$. To this end, the tidal TS value at time t may lay in the interval of $[\underline{y}(t), \overline{y}(t)]$ with the probability of $CL = \bar{\alpha} - \underline{\alpha}$. $\overline{y}(t)$ and $\underline{y}(t)$ are quantiles of tidal TS at time t, and they are respectively known as upper and lower bounds. In the literature [147], to acquire preferred reliability level (RL^*) for encircling TS values with upper and lower bounds, the relationships among RL^*, $\bar{\alpha}$, and $\underline{\alpha}$ are as follows:

$$\bar{\alpha} = 1 - \underline{\alpha} \tag{6-3}$$

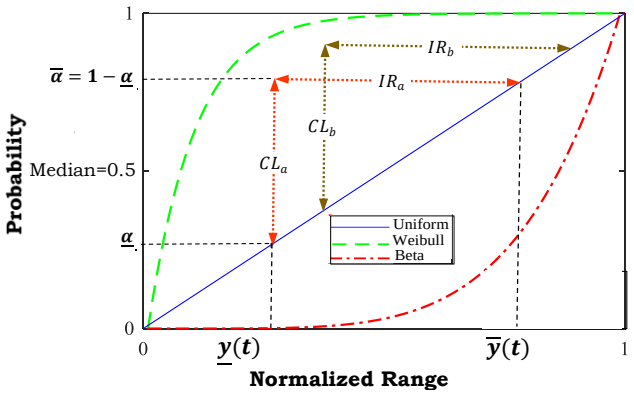

Figure 6-3: CDF of various probability distributions.

where

$$\underline{\alpha} = \frac{1 - RL^*}{2}. \tag{6-4}$$

From (6-3)-(6-4) it can be perceived that probabilities of upper and lower bounds are chosen symmetrically with respect to the probability of median in CDF (0.5) [147] as shown in Figure 6-3. Such selection of quantile probabilities leads to $CL = RL^*$ which will be further discussed in this paper.

In practice, finding the predictive distribution is troublesome and may result in some inaccurate assumptions about the PDF or CDF of the TS. But without any assumption about the predictive CDF and using QR-based NPI [150], the tidal TS uncertainty prediction can be formulated as follows:

$$\underset{\underline{\boldsymbol{B}},\overline{\boldsymbol{B}}}{Min} \quad \sum_{t=T_S}^{N.T_S} \left(\rho_{\underline{\alpha}}(y(t) - \underline{y}(t)) + \rho_{\overline{\alpha}}\left(y(t) - \overline{y}(t)\right) \right) \tag{6-5}$$

$s.t.$

$$\underline{y}(t) \leq \overline{y}(t) \tag{6-6}$$

$$\begin{bmatrix} \overline{y}(t) \\ \underline{y}(t) \end{bmatrix} = \begin{bmatrix} [q(\boldsymbol{a}_1.\boldsymbol{X}_t + b_1) & \dots & q(\boldsymbol{a}_M.\boldsymbol{X}_t + b_M)].\overline{\boldsymbol{B}} \\ [q(\boldsymbol{a}_1.\boldsymbol{X}_t + b_1) & \dots & q(\boldsymbol{a}_M.\boldsymbol{X}_t + b_M)].\underline{\boldsymbol{B}} \end{bmatrix} \tag{6-7}$$

where $\underline{y}(t)$ and $\overline{y}(t)$ are obtained based on a prediction engine, which is ELM [128] in this paper. The ELM's unknown parameters need to be tuned using the optimization problem, defined in (6-5)-(6-7). The relation between $\underline{y}(t)$ and $\overline{y}(t)$ and input features of the prediction engine $\boldsymbol{X}_t$ is demonstrated in (6-7). $q(\cdot)$ is the activation function of each hidden neuron. $\boldsymbol{a}_i$ and $b_i, i = 1, .., M$

are the hidden neuron parameters of ELM, and they are chosen randomly in the conventional ELM. While $\overline{\boldsymbol{B}}$ and $\underline{\boldsymbol{B}}$ are ELM's output weighting coefficients corresponding to upper and lower bounds, respectively. To construct $\boldsymbol{X}_t$, feature selection (FS) techniques need to be employed for finding the most appropriate features, which provide useful information about $y(t)$. In this paper, different lags of the tidal TS are considered as the input features, and then fed into the Cao's method [68], which is a simple and effective FS process. However, other feature candidates (such as various meteorological measurements), and other FS approaches can be utilized without loss of generality. In (6-5), $\rho_{\underline{\alpha}}(\cdot)$ and $\rho_{\overline{\alpha}}(\cdot)$ are asymmetrical weighting functions that consider different penalties for $\underline{y}(t)$ and $\overline{y}(t)$ based on the quantile probabilities:

$$\rho_{\underline{\alpha}}(y(t)-\underline{y}(t)) = \begin{cases} \underline{\alpha}\ .\ \left(y(t)-\underline{y}(t)\right) & \underline{y}(t) \le y(t) \\ (\underline{\alpha}-1)\ .\ (y(t)-\underline{y}(t)) & otherwise \end{cases} \tag{6-8}$$

$$\rho_{\overline{\alpha}}(y(t)-\overline{y}(t)) = \begin{cases} \overline{\alpha}\ .\ \left(y(t)-\overline{y}(t)\right) & \overline{y}(t) \le y(t) \\ (\overline{\alpha}-1)\ .\ \left(y(t)-\overline{y}(t)\right) & otherwise \end{cases}. \tag{6-9}$$

Based on (6-8) and (6-9), the optimization problem (6-5) imposes a penalty on ELM if the observed tidal TS value at t is beyond $[\underline{y}(t), \overline{y}(t)]$. The penalties are determined based on the probabilities of quantiles (i.e. $\underline{\alpha}$ and $\overline{\alpha}$). Accordingly, the ELM is heavily penalized if the observed tidal TS value does not lay within the upper and lower bounds by the probability of RL^*. In (6-8)-(6-9), both $\rho_{\underline{\alpha}}(\cdot)$ and $\rho_{\overline{\alpha}}(\cdot)$ are piecewise linear. The optimization problem described by (6-5)-(6-9) can be simply converted to an LP problem [149]. Then, the LP problem can be efficiently solved using off-the-shelf mathematical programming solvers. A detailed proof of this conversion is provided in [149]; in the following, we summarize the results of such a process. Introducing four

sets of non-negative auxiliary variables ($\underline{\theta_n}$,$\overline{\theta_n}$, $\underline{\gamma_n}$, and $\overline{\gamma_n}$) accompanied by (6-5)-(6-9), the linear representation of (6-5) can be rewritten as follows:

$$\underset{\substack{\underline{\boldsymbol{B}},\overline{\boldsymbol{B}} \\ \underline{\theta_n},\underline{\gamma_n},\overline{\theta}_n,\overline{\gamma}_n}}{Min} \quad \sum_{n=1}^{N}\left(\underline{\alpha}\,.\,\underline{\theta_n}+(1-\underline{\alpha})\,.\,\underline{\gamma_n}+\overline{\alpha}\,.\,\overline{\theta_n}+(1-\overline{\alpha})\,.\,\overline{\gamma_n}\right), \tag{6-10}$$

$s.t.$

(6-6),(6-7)

$$\overline{\theta_n},\overline{\gamma_n},\underline{\theta_n},\underline{\gamma_n}\geq 0 \tag{6-11}$$

$$y(nT_S)-\underline{y}(nT_S)-\underline{\theta_n}+\underline{\gamma_n}=0, \tag{6-12}$$

$$y(nT_S)-\overline{y}(nT_S)-\overline{\theta_n}+\overline{\gamma_n}=0 \tag{6-13}$$

The ELM's output weighting coefficients are found based on the available tidal historical data. The trained QR-based PI needs to be updated every day to alleviate the adverse impacts of tidal TS non-stationarity on the performance of prediction model.

6-5 Evidence on the Need for Asymmetrically Adaptive Quantile Regression for Tidal TS

After constructing a prediction model for tidal uncertainty prediction, its performance should be evaluated using historical tidal TS. Empirical RL and PI normalized average width ($PINAW$) are two widely used evaluation metrics, used in tidal uncertainty prediction modeling [127]. Empirical RL and $PINAW$ are defined as follows:

$$RL(\%)=\frac{100}{N}\,.\,\sum_{t=T_S}^{N.T_S}\left(\mathbf{1}_{\left[\underline{y}(t),\overline{y}(t)\right]}(y(t))\right), \tag{6-14}$$

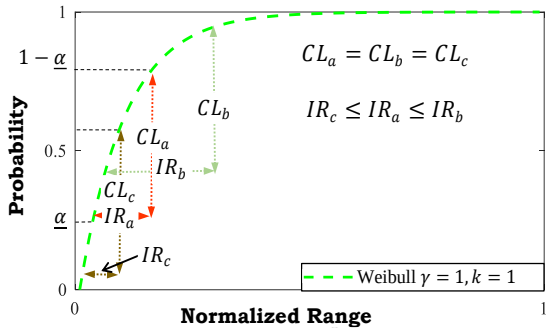

Figure 6-4: An illustrative example of non-uniform CDF.

$$PINAW(\%) = \frac{100}{N.(y_{Max}-y_{Min})} \cdot \sum_{n=T_S}^{N.T_S} (\overline{y}(t) - \underline{y}(t)), \tag{6-15}$$

$$\mathbf{1}_{[l_1,l_2]}(l) = \begin{cases} 1 & l \in [l_1, l_2] \\ 0 & otherwise \end{cases}, \tag{6-16}$$

where $\mathbf{1}_{\left[\underline{y}(t),\overline{y}(t)\right]}(y(t))$ is an indicator function as defined in (6-16). In (6-15), y_{Max} and y_{Min} are the maximum and minimum values observed in the tidal TS, respectively.

RL of the NPI model, developed as the result of solving (6-10)-(6-13), is not assessed during the optimization process. In this regard, the constructed PIs might not fulfill the system operator's preference, or the empirical RL becomes more than operator'spreference at the price of higher $PINAW$ and sacrificing the sharpness. Regardless of this fact, in the literature [19], for acquiring a specific RL^*, $\underline{\alpha}$ and $\bar{\alpha}$ are selected such $RL^* = CL = \bar{\alpha} - \underline{\alpha}$.

Moreover, as described in section 6-4 and expressed in (6-3) and (6-4), in the prevalent QR, $\underline{\alpha}$ and $\overline{\alpha}$ are symmetrically selected with respect to the median probability. Because of this selection, the difference between corresponding quantiles, known as interquantile range (IR), may not be a small value; ipso facto, $PINAW$ increases, and the PIs' sharpness decreases. Only in a uniform distribution, shown in Figure 6-3, there is no difference in selecting $\underline{\alpha}$ and $\overline{\alpha}$ with respect to median either symmetrically or asymmetrically ($IR_a = IR_b$). However, IR may significantly vary based

on the symmetrical or asymmetrical selection of quantile probabilities providing that the distribution is non-uniform, as shown in Figure 6-4.

A uniform distribution is not a precise assumption for tidal TS [127]. Besides, tidal TS possesses nonstationary characteristics so that the TS may take different shapes of CDFs as discussed in Section 6-3 and shown in [144]. Hence, for obtaining small *IR*, the quantile probabilities need to be asymmetrically chosen based on tidal TS. CDFs illustrated in Figure 6-3 are related to uniform, Weibull ($\gamma = 1, k = 1$), and Beta ($\alpha = 5, \beta = 1$) distributions, where $\gamma, k, \alpha,$ and β are parameters of the Weibull and Beta distributions [158]. It is worth mentioning that these two specific types of the Weibull and Beta distributions are only used for demonstration purpose.

To this end, in the proposed AAQR-based NPI, described in Section 6-6, an asymmetrically adaptive process for finding the appropriate quantile probabilities is devised to address the error in regression and also find a pair of quantiles probabilities, which can result in the lowest $PINAW$.

6-6 The Proposed AAQR-based NPI for Tidal TS

The general scheme of the proposed NPI is shown in Figure 6-5, and the solution approach is depicted in Figure 6-6. In this section, problem formulation of the proposed NPI for tidal TS is presented first, followed by the overall solution approach.

6-6-1 Formulation of the Proposed NPI

As shown in Figure 6-5, the proposed AAQR-based NPI has a bi-level optimization platform to address the issues mentioned in Section 6-5. In the proposed AAQR-based NPI, the upper-level optimization determines two hyperparameters ($\overline{\tau}$ and $\underline{\tau}$) which are employed to modify the QR-based NPI model (6-10)-(6-13) by slightly deviating $\underline{\alpha}$ and $\bar{\alpha}$ from their symmetrical positions. In addition to the hyperparameters, the upper-level optimization problem also adaptively tunes weighting coefficients input vector and bias ($\boldsymbol{a}_i,\ b_i, i = 1, \dots, M$). This adaptive selection of ELM parameters, which are conventionally selected randomly, enhances the effectiveness of the ELM training. The output weight coefficients of ELM ($\overline{\boldsymbol{B}}, \underline{\boldsymbol{B}}$) are obtained from the lower-level optimization by an LP of QR and considering the parameters imported from the upper-level optimization. In this regard, the proposed AAQR-based NPI is expressed by a bi-level optimization problem as follows:

$$\underset{u\in\mathcal{U}, w\in\mathcal{W}}{Min} \quad \mathcal{F}(u, w) \tag{6-17}$$

$$s.t.$$

$$-\underline{\alpha} \le \overline{\tau}, \underline{\tau} \le \underline{\alpha} \tag{6-18}$$

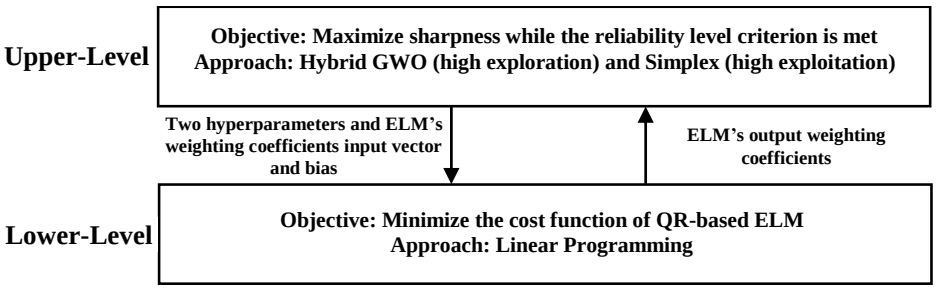

Figure 6-5: The general scheme of the proposed AAQR-based NPI model.

$$\mathcal{w} \in \arg\underset{z\in\mathcal{W}}{Min}\Big\{\mathcal{G}(\mathcal{u},z): 0 \le \underline{y_n} \le \overline{y_n} \le 1, \quad \text{(6-19)}$$

$$\overline{\theta_n}, \overline{\gamma_n}, \underline{\theta_n}, \underline{\gamma_n} \ge 0, y_n - \underline{y_n} - \underline{\theta_n} + \underline{\gamma_n} = 0, y_n - \overline{y_n} - \overline{\theta_n} + \overline{\gamma_n} = 0,$$

$$n \in \{1, \dots, N_T\}\Big\}$$

$$\mathcal{F}(\mathcal{u},\mathcal{w}) = K{\cdot}\mu_{RL}(RL) + \mathbf{1}_{[0,1]}(\mathcal{D}){\cdot}\big(1-\mu_{PINAW}(PINAW)\big) \quad \text{(6-20)}$$

$$\mathcal{G}(\mathcal{u},\mathcal{w}) = \sum_{n=1}^{N_T}\Big((\underline{\alpha}+\underline{\tau}).\underline{\theta_n} + (1-\underline{\alpha}-\underline{\tau}).\underline{\gamma_n} + (1-\underline{\alpha}+\overline{\tau}).\overline{\theta_n} + (\underline{\alpha}-\overline{\tau})\cdot\overline{\gamma_n}\Big) \quad \text{(6-21)}$$

$$\mathcal{D}(\%) = (RL - RL^*)\times 100 \quad \text{(6-22)}$$

where $\mathcal{u} = \{\boldsymbol{a}_i,\ b_i, \overline{\tau}, \underline{\tau}\}$ and $\mathcal{w} = \Big\{\overline{\boldsymbol{B}}, \underline{\boldsymbol{B}}, \overline{\theta_n}, \underline{\theta_n}, \overline{\gamma_n}, \underline{\gamma_n}\Big\}$ are sets of unknown parameters that are optimized by upper- and lower-level optimization problems, respectively. In (6-20), K is a hyperparameter, which emphasizes the importance of meeting RL criterion. Variables corresponding to the upper and lower bounds are denoted by overbars and underbars, respectively. Furthermore, $\mathcal{U}$ and $\mathcal{W}$, in (6-17), respectively refer to the domain of unknown sets $\mathcal{u}$ and $\mathcal{w}$. $\overline{\tau}$ and $\underline{\tau}$, in (6-18) and (6-21), are a pair of hyperparameters that control the quantiles probabilities asymmetricity for the upper and lower quantiles, respectively.

The upper-level optimization objective function ($\mathcal{F}(\cdot)$) in (6-17) is described in (6-20) and is defined based on Section 6-5 to evaluate the quality of constructed PIs from lower-level optimization problem (6-19). The first term of $\mathcal{F}(\cdot)$ is related to RL of PIs, while the second term corresponds to the sharpness ($PINAW$). In (6-20), RL and $PINAW$ are mapped in fuzzy domain via fuzzy membership functions (FMFs) $\{\mu_{RL}, \mu_{PINAW}\}$ as follows:

$$\mu_{\mathcal{o}}(\mathcal{o}) = \begin{cases} \dfrac{\mathcal{o}^* - \mathcal{o}}{\mathcal{o}^*} & \mathcal{o} \le \mathcal{o}^* \\ 0 & \mathcal{o} > \mathcal{o}^* \end{cases} \quad \text{(6-23)}$$

where $\mathcal{o}$ can be replaced by RL and $PINAW$, while their corresponding $\mathcal{o}^*$ are the minimum preferred reliability (RL^*) and the maximum acceptable $PINAW$ ($PINAW^*$), respectively. It is

worth mentioning that different kinds of FMFs can be defined based on power system operator's preference without loss of generality.

The lower-level optimization function is shown in (6-21). Comparing (6-21) with (6-10), one can note that the objective function of the lower-level optimization is similar to the formulation of linear QR with subtle changes to include deviation from the symmetrically chosen quantile probabilities ($\underline{\alpha}$ and $\overline{\alpha}$) using $\underline{\tau}$ and $\overline{\tau}$.

As shown in Figure 6-5, minimizing $\mathcal{F}(\cdot)$, which is the main cost function, is nested with minimization of $\mathcal{G}(\cdot)$, which is an LP optimization problem in the bi-level optimization formulation shown in (6-17)-(6-22). Some parameters used in $\mathcal{G}(\cdot)$, ($\overline{\boldsymbol{a}_i}, \underline{\boldsymbol{a}_i}, \overline{b_i}, \underline{b_i}, \overline{\tau}, \underline{\tau}$), are imported from the upper-level optimization; thus, solving this problem requires an iterative process. Moreover, the lower-level optimization can be solved by state-of-the-art LP packages, which not only guarantee achieving global solutions but also make the iterative optimization problem computationally feasible.

6-6-2 Solution Approach of the Proposed NPI

In this section, a solution approach for finding the parameters of the proposed NPI model is put forward. ELM [128], GWO [153], and simplex method [159] are three main building blocks, and keen readers may refer to related documents for comprehensive illustrations.

As put on show in Figure 6-6, in the first step of the proposed AAQR-based NPI, the data are normalized, and any gap in data is filled by linear interpolation. Further preprocessing is done by the first order differencing data to make it less non-stationary. Then, available data is split into training and validation sets. The training set is used in the training process and in solving the lower-

level optimization problem. In other words, ELM's output weighting coefficients (in set $\mathcal{w}$) are obtained using the training set, while both training and validation sets are employed in assessing the developed AAQR-based NPI in the upper-level problem to avoid overtraining and undertraining.

Due to nonlinearity and non-differentiability of the upper-level optimization and linearity of the lower-level optimization, a hybrid heuristic and LP optimization techniques are employed to solve the bi-level optimization for tuning AAQR-based NPI model parameters. In this regard, as shown in Figure 6-6, first, an initial population of the variables, $\mathcal{u}$, which need to be determined by the upper-level is first generated. Then, the lower-level LP problem is solved to find ELM's output

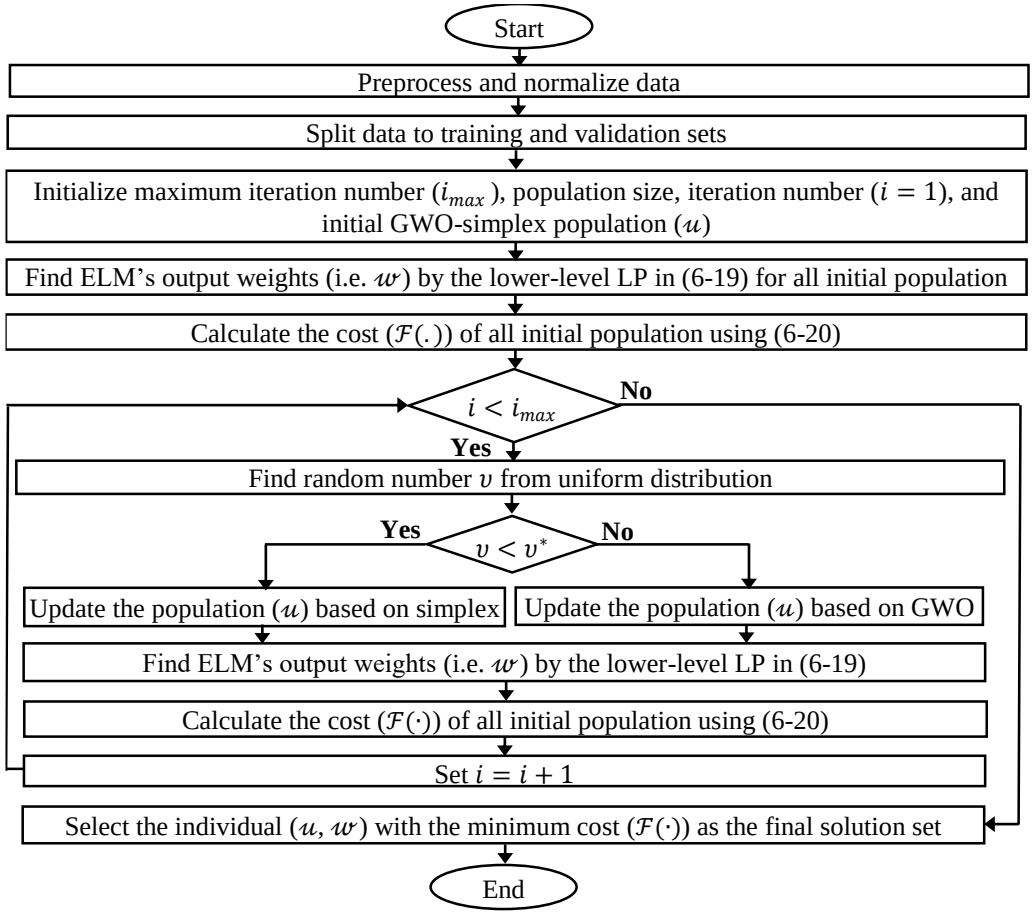

Figure 6-6: Flowchart of the proposed AAQR-based NPI model.

weighting coefficients which are in set $\mathcal{w}$ using training set. The quality of the PIs for both training and validating sets is then assessed by the upper-level optimization cost function ($\mathcal{F}(\cdot)$), and the heuristic optimization algorithm updates the population based on the PIs quality. This process is repeated until the stopping criteria are met. GWO and simplex approaches are integrated and used as the heuristic optimization algorithm of the above-mentioned process to simultaneously take advantage from exploration capability of GWO and exploitation ability of simplex. At each iteration, a random number υ is generated based on a uniform distribution. As shown in Figure 6-6, relaying on this random number, GWO or simplex is selected for population modification. If simplex is selected, all strategies except shrinkage are repeated for an inner iteration number equal to the population size. If GWO is elected, the population is modified based on a set of top three agents found in the past iteration (obtained from GWO or simplex). This process is continued until the stopping criteria are satisfied and the optimal solution is the best agent among the top three.

6-7 Experimental Results and Comparisons

Three different tidal datasets are used in this study. The first is hourly TL data from the Port Dover (located in Ontario, Canada) recorded from March 1st to April 1st [160]. This dataset is publicly available and possesses high variability, which makes it appropriate for assessing the efficacy of a prediction model. The second dataset is from Akutan Pass (located in Alaska, USA) from July 1st to August 1st, 2010 with a resolution of 6 minutes [161]. This location is amongst the high potential tidal generation sites in Alaska with a kinetic power density of 2870 W/m^2 [162]. The third is a 10-minute resolution data from the Bay of Fundy (located in Nova Scotia, Canada) recorded from September 1st to October 1st, 2008. The Bay of Fundy also has a high potential for tidal stream power generation [144], so this dataset is used to evaluate the

performance of the proposed framework for prediction of TCS and TCD. Please note that in this paper the entire dataset is normalized between 0 and 1. The NPI is trained every 12 hours with 15 days of historical data, which includes both training and validation datasets.

For the sake of performing a fair comparison with other existing PI models, a persistence model [147] is employed as the first benchmark model. Persistence model provides a rough estimation of TS and is a popular benchmark model in renewable energy prediction. To show the advantages of the AAQR-based NPI, the commonly used QR-based PI [60] is developed as the second benchmark model. Besides, to assess the efficacy of the proposed AAQR-based NPI, the obtained results are compared to those of the most recent prediction models in tidal application [127]. In this paper, the model, proposed in [127], which is based on support vector regression (SVR) and lower upper bound estimation (LUBE) is called SVR-LUBE for convenience in statement. RL in (6-14) and $PINAW$ in (6-15), which are the most important evaluation criteria in uncertainty prediction, are used for comparing the performance of the proposed AAQR-based NPI models against other prediction models. Based on these criteria, the best PI model is PI that can achieve the minimum required RL, asked by the power system operators, while the PIs are the sharpest. It is worth to remind that the lower $\mathcal{S}$ means the sharper PIs. In this paper, models are developed using MATLAB and CPLEX 12.7 [163] is employed for conducting LP by calling *cplexlp* function in MATLAB. All the prediction models are tested in an Intel 3.4-GHz CPU with 16 GB of RAM.

In this section, the main contributions of this paper are examined numerically in scenarios. The first scenario evaluates the overall performance of the AAQR-based NPI model for predicting TL, TCD, and TCS and illustrates its advantages over the existing PI models. The second scenario is devoted to exploring the impacts of the main components of the AAQR-based NPI, including the

asymmetrically adaptive selection of quantile probabilities, as well as the hybrid GWO-Simplex optimization approach.

6-7-1 First Scenario

In this scenario, the PI is analyzed for two different prediction horizons: 1 hour and 10 minutes due to the importance of these horizons as mentioned in Section 6-3. In all simulations, PIs are constructed with required reliability, RL^{*}=90%, which is of great importance for system operators [144].

The results obtained for the 1-hour horizon are shown in Figures 6-7 -6-11 and Table 6-1. Here, the prediction results are visually presented for three days (72 sample points with 1-hour sample time) for ease of illustration. These figures represent estimations of the upper and lower bounds predicted by the proposed approach and show that the proposed NPI can provide sharp and reliable PIs for all types of tidal data. Table 6-1 reports the performance of different PI models for this case; results of the best PI model, which can simultaneously provide sharp and reliable PIs, are bolded.

The persistence model could not construct reliable PIs for both the TCD and TCS of the Bay of Fundy, while the constructed PIs as the result of the persistence model are not sharp for other

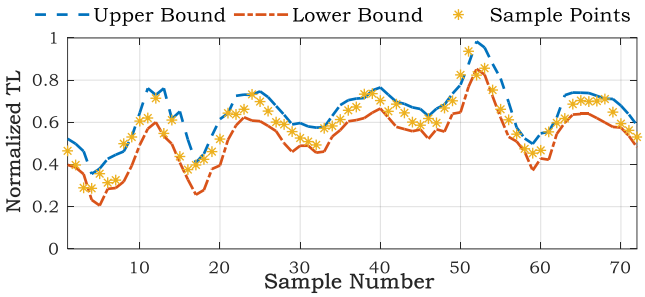

Figure 6-7: 1-hour ahead prediction of the TL using the proposed AAQR-based NPI for Port Dover.

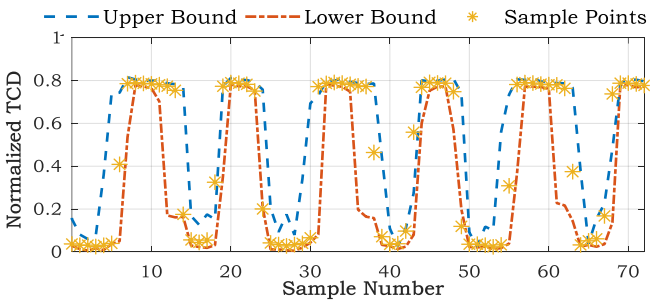

Figure 6-8: 1-hour ahead prediction of the TCD using the proposed AAQR-based NPI for Akutan Pass.

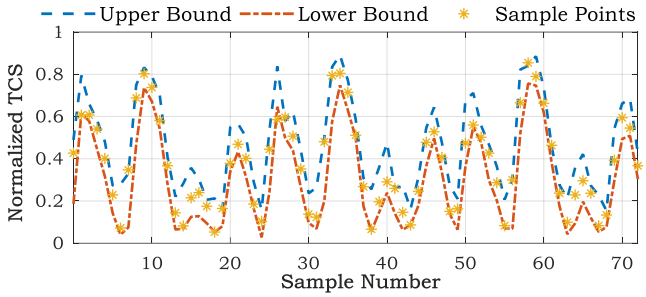

Figure 6-9: 1-hour ahead prediction of the TCS using the proposed AAQR-based NPI for Akutan Pass.

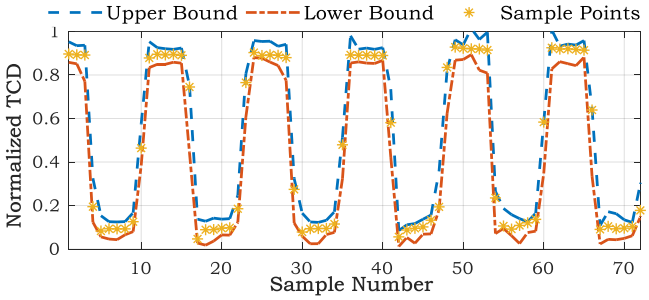

Figure 6-10: 1-hour ahead prediction of the TCD using the proposed AAQR-based NPI for the Bay of Fundy.

datasets. SVR-LUBE in TL prediction and Akutan Pass TCS almost met the reliability criterion; for the other case studies the reliability criterion was absolutely met in tidal prediction by SVR-LUBE. QR-based PI met the reliability criterion in all the cases except in prediction of TCS related to Akutan Pass. Amongst all prediction models, the proposed AAQR-based NPI satisfied the

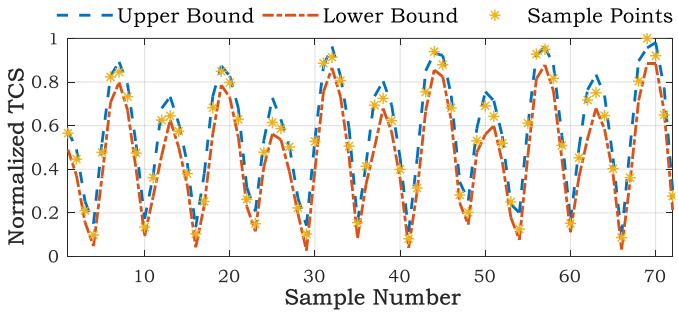

Figure 6-11: 1-hour ahead prediction of the TCS using the proposed AAQR-based NPI for the Bay of Fundy.

reliability requirement in all cases with the maximum sharpness, as reported in Table 6-1. The PIs obtained from the proposed AAQR-based NPI are $(25.0078 - 17.5841)/25.0078 =$

29.6% sharper than QR-based PI for TL. Also, AAQR-based NPI model results in 13.5% and 20.7% reduction in $PINAW$ and therefore better sharpness compared to the QR-based PI in TCD and TCS prediction of Akutan Pass, respectively. Similarly, 26.0%, and 25.2% increases in sharpness of TCD and TCS for the Bay of Fundy TS are respectively achieved using AAQR-based PI compared to QR-based PI. The proposed approach also outperforms SVR-LUBE by meeting the required RL with sharper PIs. Figures 6-7-6-11 show that TL for the 1-hour horizon possesses irregular and non-stationary patterns compared to TCD and TCS; thus, its prediction seems challenging [145]. Despite this, the proposed NPI can significantly enhance the sharpness of TL prediction while maintaining the required RL (Table 6-1). Figure 6-9 and Figure 6-11 also show that the TCS of Akutan Pass and the Bay of Fundy have different characteristics. Akutan Pass has mixed semidiurnal characteristics while the Bay of Fundy is more semidiurnal; therefore, the highly accurate PIs obtained from the proposed AAQR-PI confirm the efficacy thereof for different tidal TS types.

Table 6-1: Results of 1-hour ahead PI construction of tidal data (Minimum RL^*=90%)

Data		Prediction Model	RL	$PINAW$
TL in Port Dover		Persistence	95.8333	39.1848
		QR-based PI	94.7917	25.0078
		SVR-LUBE	89.0610	19.0701
		AAQR-based NPI	**91.6667**	**17.5841**
TCD	Akutan Pass	Persistence	93.7500	75.2905
		QR-based PI	95.6666	30.3951
		SVR-LUBE	90.7111	28.0810
		AAQR-based NPI	**90.4464**	**26.2801**
	Bay of Fundy	Persistence	79.2208	54.6104
		QR-based PI	92.8230	16.2408
		SVR-LUBE	90.0120	13.0025
		AAQR-based NPI	**97.9167**	**12.0148**
TCS	Akutan Pass	Persistence	90.8333	58.8451
		QR-based PI	88.0208	19.7717
		SVR-LUBE	89.0185	20.8466
		AAQR-based NPI	**91.8533**	**15.6747**
	Bay of Fundy	Persistence	89.6104	68.7521
		QR-based PI	93.0555	17.0712
		SVR-LUBE	91.7100	17.6627
		AAQR-based NPI	**90.0625**	**12.7576**

The results obtained for the 10-minute ahead PI are reported in Table 6-2. Considering the unavailability of tidal TS at Port Dover and Akutan Pass with the required resolution, PIs are only constructed for TCD and TCS of the Bay of Fundy. Table 6-2 shows that the QR-based PI cannot provide a reliable PI for TCS; in contrast, the SVR-LUBE and the proposed NPI constructed reliable PIs, but the proposed NPI offers 35.16% sharper PIs compared to SVR-LUBE.

Table 6-2: Results of 10-minute ahead PI construction of tidal data (Minimum RL^*=90%)

Data	Prediction Model	RL	$PINAW$
TCD	Persistence	77.0833	26.1472
	QR-based PI	96.5301	15.0492
	SVR-LUBE	93.1167	13.5658
	AAQR-based NPI	**93.0550**	**11.5875**
TCS	Persistence	95.8333	37.4140
	QR-based PI	85.7638	14.5235
	SVR-LUBE	92.1416	24.2846
	AAQR-based NPI	**90.0321**	**15.7441**

The above descriptions and discrepancies clearly show that the proposed model outperforms the existing benchmark models in terms of providing sharp and reliable PIs for 1-hour and 10-minute horizons. Empirical observations from other simulations indicate similar performance is realized for other prediction horizons. Furthermore, although TL prediction is rarely considered in the literature, its variability in the 1-hour ahead horizon seems aperiodic and irregular, which makes its prediction complicated. The simulation results demonstrate that the proposed model has the capability to provide a reliable estimate for TL but has an intermittent pattern.

6-7-2 Second Scenario

This part investigates the effectiveness of different components of the proposed NPI, including an asymmetrically adaptive selection of quantile probabilities, adaptive selection of ELM parameters, and the GWO-Simplex optimization method.

The effects of the proposed hyperparameters ($\overline{\tau}$, $\underline{\tau}$) on asymmetrical selection of quantile probabilities are reported in Figure 6-12; the results obtained are associated with the 1-hour ahead TCS data shown in Figure 6-11. The values of $\overline{\tau}$ and $\underline{\tau}$ are updated continuously every 12 hours, to find the optimal TCS prediction reported in Table 6-1. Moreover, not only are $\overline{\tau}$ and $\underline{\tau}$ asymmetric with respect to the median probability but their deviations from the traditionally employed quantile probabilities are also asymmetric. For instance, the values of $\overline{\tau}$ and $\underline{\tau}$ for sample # 40 in Figure 6-12 are respectively optimized to 0.0117 and -0.0317, which result in $CL = 0.8566$. This shows that the traditionally chosen $RL^* = CL = \bar{\alpha} - \underline{\alpha}$ may not be optimal in several cases and consequently demonstrates the need for the proposed AAQR.

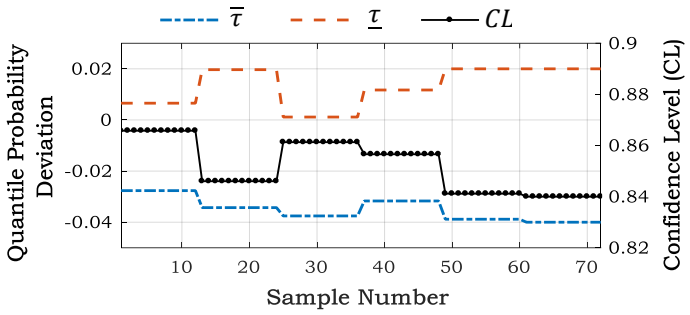

Figure 6-12: $\overline{\tau}$ and $\underline{\tau}$ for 1-hour ahead TCS prediction.

Another factor that impacts on the performance of the proposed NPI is tuning kernel function parameters. The results obtained with and without adaptive tuning of the ELM kernel function parameters for 1-hour ahead TCS data are provided in Table 6-3. Adaptive tuning improves the sharpness of the ELMs, but the reliability is mostly unchanged. Similar behavior is empirically observed for other tidal data for different prediction resolutions.

For further examination, the performance of the hybrid GWO-Simplex algorithm is assessed on its own as well as using two well-known optimization techniques: GA and PSO. In the case of the GA, a roulette wheel selection strategy is utilized to select between adaptive crossover and mutation operators. The classical PSO is used, in which inertia weight is fixed to 1 and accelerator coefficients are both set to 2. Maximum iteration number and convergence are chosen as the stopping criteria for all optimization approaches. For the same population size, all programs are

Table 6-4: Results of Different Heuristic Optimization Techniques for the Proposed NPI of TCS Data (Minimum RL^*=90%)

Optimization Algorithm	Best Solution		Ave. Solution		Ave. Time (s)
	RL	*PINAW*	*RL*	*PINAW*	
GA	92.3150	14.5196	96.0076	16.3719	291.7060
PSO	92.0315	14.0102	95.3861	15.9107	211.0576
GWO	91.7082	14.4501	93.8640	15.2761	95.50736
Hybrid GWO & Simplex	90.0625	12.7576	91.1573	13.2907	137.05513

Table 6-3: Effects of ELM Adaptive Tuning on the Proposed NPI of TCS Data

Training Strategy	*RL*	*PINAW*
Without Adaptive Tuning	91.5157	14.0901
With Adaptive Tuning	**90.0625**	**12.7576**

implemented for 10 independent runs, with the results reported in Table 6-4. The rates of success to reach RL^* in all runs were 80% and 90% for PSO and GWO, respectively, and 100% for others. Thus, the average values reported in Table 6-4 are related to those methods for the successful cases. From this table, it can be concluded that hybrid GWO and simplex outperforms other benchmark heuristic optimization approaches in terms of meeting the RL^* with sharp PIs. Furthermore, its computation time is comparable with GWO and is lower that GA and PSO.

6-8 Conclusion

This paper proposes an AAQR-based NPI model for short-term uncertainty prediction of tidal TS. The shortcomings of traditional QR-based NPI, including error in the prediction engine and symmetrical selection of a pair of quantiles probabilities, are discussed and addressed by proposing two hyperparameters. Hyperparameters and prediction engine parameters are tuned through a bi-level optimization framework. The upper level optimization problem is solved by a hybrid heuristic optimization, while LP is conducted for the lower-level problem. The model developed has been successfully tested on a wide range of tidal data; the results obtained, and comparisons reported in the paper show that the proposed approach can be used as an effective tool for maximizing tidal energy deployment by providing more accurate uncertainty modeling of the most inflectional and uncertain factors.

Further research can be conducted to investigate the effects of meteorological variables on tidal power generation prediction. In addition, the impacts of the asymmetrical and adaptive selection of quantile probabilities in power system operation problems can be scrutinized.

7- A SECURE DEEP PROBABILISTIC DYNAMIC THERMAL LINE RATING PREDICTION[8]

Accurate short-term prediction of overhead line (OHL) transmission ampacity can directly affect the efficiency of power system operation and planning. Any overestimation of dynamic thermal line rating (DTLR) can lead to lifetime degradation and failure of OHL, safety hazards, etc. This paper presents a secure yet sharp probabilistic prediction model for an hour ahead forecasting of DTLR. It is based on an augmented deep learning architecture which makes use of a wide range of predictors, including the his-torical climatology data and latent variables, obtained during DTLR calculation. Besides, by introducing three hyperparameters, a customized cost function is introduced to train a deep neural network to consider the DTLR security confidence level criterion while minimizing deviations of the predicted DTLRs from the actual values. The hyperparameters are tuned according to the power system operator's requirement adopting a grid search algorithm. The proposed probabilistic DTLR is developed and verified using recorded experimental data. The simulation results validate the superiority of the proposed DTLR compared with the state-of-the-art prediction models using well-known evaluation metrics.

7-1 Introduction

Thermal line rating (TLR) is known as the primary culprit behind limitation of current carrying capability of overhead line (OHL) [164]. Both IEEE [165] and CIGRE [166] put forward TLR calculation methods. Although these approaches provide almost similar results, the one proposed

[8] © To be submitted, N. Safari, S. M. Mazhari, C. Y. Chung, and S. B. Ko, "A secure deep probabilistic dynamic thermal line rating prediction," *IEEE Trans. Power Syst.*

by IEEE is more simplified and easier to use [167]. TLR of OHL is a weather dependent variable and prevalently calculated by the heat balance equation under the worst-case weather scenarios [168]. Despite conservativeness of such static TLR (STLR), there might be some cases in which the STLR exceeds the real OHL thermal constraint. Consequently, the OHL might be exposed to damages due to lack of monitoring thermal line rating in the STLR-based approaches [169]. However, routinely, the STLR is less than the actual OHL ampacity [168, 169].

To overcome the shortcomings of STLR, dynamic TLR (DTLR) is proposed in which the thermal condition of OHL can be monitored. Ergo, DTLR results in unlocking the additional capacity headroom of the current OHLs, thereby addressing the network congestion and postponing/eliminating the need for transmission expansion [168]. As a significant byproduct, DTLR facilitates delivering the highly variable and uncertain power of renewable energy systems (RESs) to the end-user due to perceptible correlation between RESs generation with the additional capacity provided by DTLR. For such reasons, DTLR has recently grasped attention of governments and transmission companies; and it is considered as an enabling tool for enhancing the penetration of RESs [170-172].

DTLR is a function of several climatology variables, such as wind speed, wind direction, etc. [165]; therefore, its value for up-coming hours need to be forecasted. The DTLR forecast can be employed in various power system problems, such as unit commitment, economic dispatch, optimal power flow, etc., [167, 173]. In this respect, specialized research communities have devoted momentous endeavors to develop DTLR monitoring and prediction models [171-181]. Since DTLR monitoring and prediction may not be feasible through the entire OHL, critical spans are identified for this mission. In [182, 183], heuristic approaches are brought forward to identify

the number and locations of the required monitoring stations to make the OHL fully observable from the thermal rating perspective.

Relying on the literature, DTLR prediction has been interpreted with two different viewpoints. In a group of studies, researchers predict the maximum allowable current, in which the OHL temperature is equal to the OHL thermal limit [171-180]. However, on the other side, DTLR prediction refers to estimating the future value of an OHL temperature providing that it carries a certain amount of current [181, 182]. While both perspectives bring about intriguing advancements to the field and provide insightful information about the OHL thermal constraints, this paper zeroes in on the DTLR prediction based on the first definition.

DTLR methods can be broadly divided into direct [171, 173-180], indirect [180], and hybrid [179, 182] approaches. In direct DTLR calculation, the required values are calculated based on straightforward computation of the conductor's maximum allowable current while taking the maximum permissible temperature into account. In this respect, the process is accomplished by employing weather data and heat balance equation. This is while some measured values, such as sag position, mechanical tension, etc., are used in indirect DTLR calculation. Nowadays, direct approaches are of remarkable interest as they are less dependent to external equipment and lead to low-cost outcome compared to their counterparts [170].

From the prediction output point of view, DTLR prediction models can be divided into deterministic [178, 180] and probabilistic [174-178] categories. In deterministic prediction, a single quantity associated with the most likely value of DTLR in the next sample point is considered as the output; while in the probabilistic approach, information about uncertainty in DTLR prediction can be acquired. Since any overestimation of DTLR may lead to unprecedented

issues, probabilistic prediction is most welcome by power system operators [168] and accordingly is the focus of this work.

In [175, 176], parametric distributions model their associated uncertainties with climatology variables. In [177], Taylor series expansion is employed to find the mean and variance of DTLR in the coming hours on the basis of the forecasted values of mean and variance corresponding to the meteorological variables. To consider the interdependency among meteorological variables in DTLR forecast in [184], multivariate Gaussian mixture distributions, resulted from different meteorological variables, are used in a Monto Carlo simulation process to extract DTLR distribution.

In [175-177, 184], the meteorological variables are assumed to possess a parametric representation for their uncertainties. However, this assumption may be erroneous due to high non-stationarity of meteorological time series (TS) [135]. To this end, authors in [174] proposes a non-parametric autoregression framework for probabilistic DTLR prediction based on quantile regression (QR). Authors in [168, 178] put forward a non-parametric DTLR prediction based on quantile regression forest (QRF), of which meteorological measurements and numerical weather predictions (NWPs) compose the input features of the prediction model. As the literature reveals, studies on non-parametric DTLR prediction is still in its early stage; thus, this paper is devoted to further contribute to this literature.

The weather-based DTLR prediction models mainly limit their inputs to those features, obtained directly from meteorological measurements. However, from DTLR formulations pro-posed in [165, 166], one can note that the relations of meteorological variables with DTLR value are too complex and non-linear. Meanwhile, there are many latent variables (e.g., convection cooling,

radiated heat loss rate, etc.) which are acquired in the process of DTLR calculation and may provide information about the complex relationship between DTLR value and meteorological variables. However, to the best of authors knowledge, these important predictors have not been considered in DTLR forecast thus far. Thence, this paper scrutinizes the impacts of latent variables in DTLR forecast. Moreover, according to CIGRE recommendations, DTLR values should be estimated such that actual DTLR is equal to or greater than the estimated value in 99% of instances [167]. However, this CIGRE recommendation has not been considered in assessing the thus far proposed prediction models.

The remarkable advancement in deep learning and its successful implementation in prediction result in forecasting accuracy enhancement in a range of meteorologically dependent variables of power systems [110, 185, 186]. In deep neural networks (DNNs) architectures, highly efficient unsupervised dimension reduction blocks (i.e., autoencoder variants) can be employed to tackle with the high dimension feature space issue, posed due to numerous meteorological and latent variables. Using DNNs, more complex patterns, which are not viable to identify in shallow networks, can be perceived. Compared to various building blocks of DNNs, long short-term memory (LSTM) [187]—a recurrent NN (RNN)—has demonstrated superior performance in wind speed as well as meteorological variables prediction problems [185]. The benefit of LSTM compared to the conventional RNNs is its capability to capture long- and short-term dependencies in a sequence with addressing the vanishing and exploding gradient problems of prevalent RNNs [187]. Despite the successful applications of RNNs, specifically LSTM, and benefits of DNNs, they have not been adapted to DTLR prediction. In this paper, LSTM is employed in developing the DTLR prediction, while the stacked denoising autoencoder (SDAE) [188] is developed for feature extraction and reducing the feature space dimension. To the extent of authors' knowledge,

the so far developed DNN models are trained deterministically in power systems applications [110, 185, 186]). A major impediment which made the past methods incapable of offering a probabilistic model in DNN may originate from lack of proficient probabilistic cost function; such function should consider reliability level and sharpness together so that the DNN parameters are tuned according to the system operator's preference. To address this need, the current paper presents an innovative cost function for DNN training.

This paper proposes a probabilistic DTLR prediction which takes benefits from the accessible latent variables in addition to meteorological measurements. The high dimension input feature space in DTLR is reduced using an SDAE. A DTLR model is developed using LSTM units in a DNN architecture. The prediction engine is trained by considering a novel cost function to meet the CIGRE requirement while the sharpness is maximized. The performance of the proposed models is compared with the state-of-the-art DTLR prediction models using publicly available data. Briefly, the main contributions of this work are three-fold:

- For the first time, latent variables are introduced in DTLR forecast as valuable predictors;
- For the first time, a deep learning architecture is constructed for DTLR forecasting;
- For the first time, deep learning model is trained for probabilistic forecasting in power systems application.

7-2 Proposed Dynamic Thermal Line Rating Prediction

Hereafter, $\{WS_t\}_{t=1}^{N}$ and $\{WD_t\}_{t=1}^{N}$ are TS associated to wind speed, wind direction, respectively. N is the length of TS, and index t refers to the tth sample of TS. Also, the TS of the wind speed components, decomposed by Cartesian coordinate system, are demonstrated by

$\{WSX_t\}_{t=1}^{N}$ and $\{WSY_t\}_{t=1}^{N}$. Besides, TS of ambient temperature and solar irradiance are represented by $\{AT_t\}_{t=1}^{N}$ and $\{SI_t\}_{t=1}^{N}$, respectively.

Figure 7-1 demonstrates the overall framework of the proposed DTLR prediction model. The suitable number of lags (l) to form the input vector is identified using the Cao's embedding dimension approach [68]. From the measurements, l lags of meteorological TS are imported to DTLR calculation model, described in Section 7-2-1. Using the DTLR calculation model, $[DTR_{t-l+1} \quad \dots \quad DTR_t]$, $[qc_{t-l+1} \quad \cdots \quad qc_t]$, and $[qr_{t-l+1} \quad \cdots \quad qr_t]$, which respectively represent DTLR, convection cooling, and radiated heat loss rate for lags of meteorological inputs, are calculated. The obtained features from this step along with the meteorological input vector are utilized in feature reduction and extraction stage based on SDAE with LSTM building blocks, as

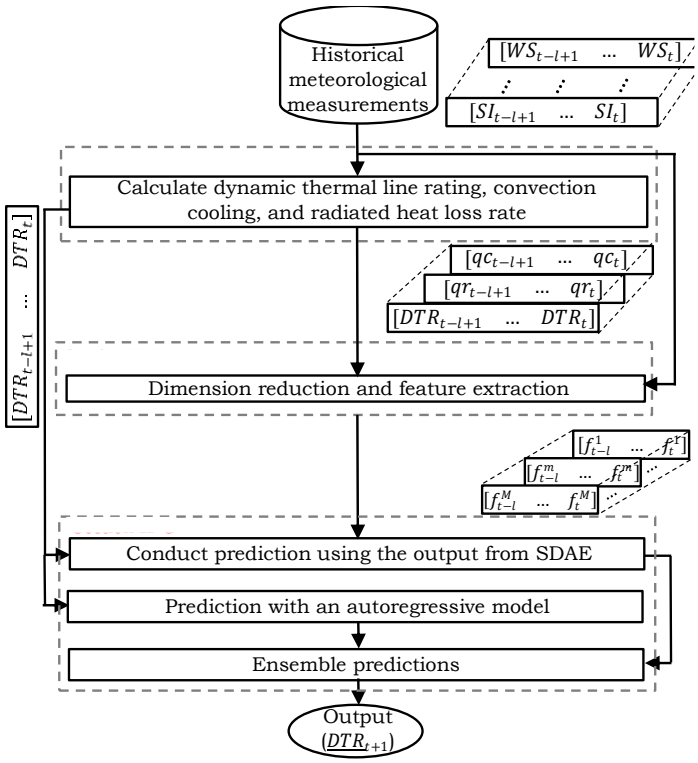

Figure 7-1: General scheme of the proposed DTLR prediction.

discussed in Section 7-2-2. Afterwards, various trained models are employed to acquire the final DTLR prediction ($\underline{DTR_{t+1}}$ as elaborated in Section 7-3).

7-2-1 Dynamic Thermal Line Rating Calculation

In this study, it is assumed that the power system is under normal operation; and therefore, the current fluctuations in the OHLs can be considered negligible, providing that the system does not require any abrupt and temporary switching [181]. Besides, studies reveal that the maximum time required to reach the steady state because of a step change in current is approximately 30 minutes [189]. Coming from the mentioned points, the transients in DTLR can be ignored, and DTLR can be estimated for every hour.

Figure 7-2 schematically represents the influential factors in DTLR of OHL. As per IEEE Std 738-2012, in the steady-state the heat balance equation for an OHL at tth sample can be written as follows [165]:

$$qc_t + qr_t = qs_t + I_t^2 \cdot R(T_{con}) \tag{7-1}$$

where qc_t and qr_t are convection and radiated heat loss rates per unit length, respectively. qs_t in (7-1) is heat gain rate from sun, and $R(T_{con})$ is alternating current (AC) resistance associated to

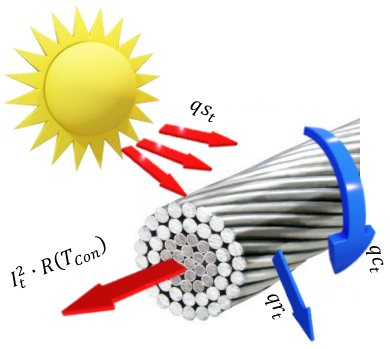

Figure 7-2: A schematic diagram of DTLR.

conductor temperature, T_{con}. From (7-1), the allowable conductor current, I_t, at T_{con} can be simply obtained as follows:

$$I_t = \sqrt{\frac{qc_t + qr_t - qs_t}{R(T_{con})}} \tag{7-2}$$

From (7-1)-(7-2) one can observe that qc_t and qr_t are the cooling elements in the heat-balance equation, and their increase can help to attain more OHL ampacity, while qs_t is a heating component and is a culprit of ampacity reduction. qc_t is a function of WS_t, WD_t, and AT_t. qc_t is calculated as follows [165]:

$$qc_t = \max\left(qc_t^{f1}, qc_t^{f2}, qc_t^{n}\right) \tag{7-3}$$

$$qc_t^{f1} = K_{angle} \cdot \left[1.01 + 1.35 \cdot {N_{Re}^t}^{0.52}\right] \cdot k_f \cdot (T_{con} - AR_t) \tag{7-4}$$

$$qc_t^{f2} = K_{angle} \cdot 0.754 \cdot {N_{Re}^t}^{0.6} \cdot k_f \cdot (T_{con} - TA_t) \tag{7-5}$$

$$qc_t^{n} = 3.645 \cdot \rho_f^{0.5} \cdot D_o^{0.75} \cdot (T_{con} - TA_t) \tag{7-6}$$

$$N_{Re}^t = \frac{D_o \cdot \rho_f \cdot WS_t}{\mu_f} \tag{7-7}$$

where K_a is wind direction factor. ρ_f and μ_f are air density and absolute air viscosity, respectively. D_o is the outside diameter of the conductor. As it can be observed from (7-3)-(7-7), qc_t is a nonlinear and complicated function of meteorological variables while the relationship between qc_t and I_t is simple as shown in (7-2). Therefore, considering $[qs_{t-l+1} \quad \cdots \quad qs_{t-l}]$ as elements of predictor set for forecasting $DTLR_{t+1}$ can be beneficial. Moreover, qr_t in (7-1)-(7-2) can be calculated as follows [165]:

$$qr_t = 17.8 \cdot D_o \cdot \epsilon \cdot \left[\left(\frac{T_{con} + 273}{100}\right)^4 - \left(\frac{TA_t + 273}{100}\right)^4\right] \quad (7\text{-}8)$$

where ϵ is emissivity and has a value of [0.23, 0.91] which increases by conductor age. As can be seen from (7-8), like qc_t, qr_t has relation with the fourth power of TA_t. Consequently, the relation between TA_t and I_t is intricate; and therefore, considering the simple vector of $[TA_{t-l+1} \quad ... \quad TA_t]$ as the elements of the feature set may not be adequate to reflect the importance of ambient temperature and radiative heat loss in DTLR prediction, and it is worthful to analyze the influence of considering historical data of radiated heat loss rate in DTLR prediction. It is worth noting that for a solar irradiance at time t, SI_t, the rate of solar heat gain, qs_t, can be estimated by a linear function of SI_t [165]; and therefore, its consideration as a feature cannot be informative.

After calculating $[DTR_{t-l+1} \quad ... \quad DTR_t]$, $[qc_{t-l+1} \quad ... \quad qc_t]$, and $[qr_{t-l+1} \quad ... \quad qr_t]$, all the thus far obtained features are used in a feature extraction and feature space reduction stage as elucidated below.

7-2-2 Feature Reduction and Extraction in Dynamic Thermal Line Rating Prediction

As elaborated in Section 7-2-1, several climatic variables, including wind speed, wind direction, wind speed Cartesian components, ambient temperature, and solar irradiance, highly influence DTLR value. A tensor, formed by series of lags associated with these variables, contains the potential informative predictors for DTLR. Besides, it is clarified that the latent variables (i.e., convective cooling and radiated heat loss rate) may also contain valuable information about the complex relationship with climatic variables and DTLR. Also, the historical DTLR values could

also therefore, nine types of feature candidates and their corresponding lags are considered to form the feature pool.

7-2-2-1 Feature Reduction

Properly optimizing the input features of the prediction engine by eliminating the non-informative and redundant features and constructing more features, which can demonstrate the DTLR pattern more efficiently, is a principal stage in DTLR prediction. To this end, we first employ a feature reduction stage based on minimal-redundancy-maximal-relevance (mRMR) [190]. mRMR is a mutual information (MI) based approach which is employed in various power system problems [191] for identifying the subset of features, providing the most amount of information about the observation (target variable). MI is widely used in feature selection literature to evaluate the degree of uncertainty which a predictor can alleviate from observation by measuring the mutual relevancy of predictor and target variable. For two random variables with domains A and B the MI is defined as follows [190]:

$$MI(A;B) = \sum_{a\in A}\sum_{b\in B} P(a,b)\cdot \log\left(\frac{P(a,b)}{P(a)\ \cdot P(b)}\right) \tag{7-9}$$

where $P(a,b)$ represents the joint probability density function, and $P(a)$ and $P(b)$ are individual probability density functions of a and b random variables, which are the discretized format of continuous predictor and target variables. Based on (7-9), the iterative mRMR algorithm is carried out by the following optimization problem [190]:

$$\max_{\substack{a_j^{t-p}\in\mathcal{A}-\boldsymbol{\Omega_{n-1}},\\ j=\{1,\dots,9\},p=\{0,\dots,l-1\}}} \left[MI\left(a_j^{t-p}; \boldsymbol{DTR_{t+1}}\right) - \frac{1}{n-1} \sum_{\substack{a_i^{t-q}\in\boldsymbol{\Omega_{n-1}}\\ i=\{1,\dots,9\},q=\{0,\dots,l-1\}}} MI\left(a_j^{t-p}; a_i^{t-q}\right) \right] \tag{7-10}$$

$$\begin{aligned}\mathcal{A} = \{ & a_1^{t-l+1} = \boldsymbol{WS_{t-l+1}}, \dots, a_1^t = \boldsymbol{WS_t}, \\ & \dots, \\ & a_7^{t-l+1} = \boldsymbol{DTR_{t-l+1}}, \dots, a_7^t = \boldsymbol{DTR_t}, \\ & \dots, \\ & a_9^{t-l+1} = \boldsymbol{qr_{t-l+1}}, \dots, a_9^t = \boldsymbol{qr_t}\}\end{aligned} \tag{7-11}$$

where $\boldsymbol{\Omega_{n-1}}$ is the subset containing selected features at the $n-1th$ iteration. Bold notations represent the random variables of various predictors. a_j^{t-p} is the random variable describing the pth lag of jth feature. $\mathcal{A}$ is a set consisting of the random variables associated to feature candidates. In the first iteration of solving (7-10), $n = 1$, and $\boldsymbol{\Omega_0} = \emptyset$. In this work, the optimization problem in (7-10) is iteratively solved till MI of the last component selected from solving (7-10) with $\boldsymbol{DTR_{t+1}}$ is larger than or equal to the one tenth of the MI of the first component selected from solving (7-10) with $\boldsymbol{DTR_{t+1}}$.

After conducting mRMR, the feature types which are not among the selected features in $\boldsymbol{\Omega_n}$ or constitute negligible portion of $\boldsymbol{\Omega_n}$ are removed from the feature pool. Therefore, the number of feature types will be reduced to less than nine, which is the initial number of feature types. Then, the selected features along with their corresponding l lags are used to build the tensor input of the deep learning-based feature extraction, described in Section 7-2-2-3.

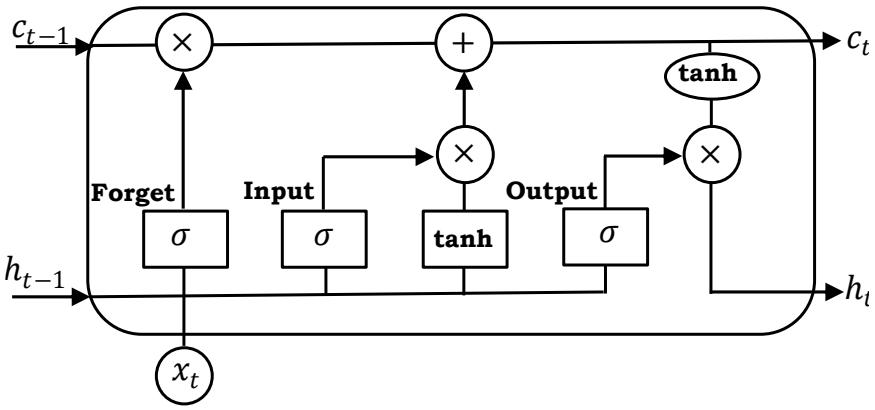

Figure 7-3: An LSTM unit

7-2-2-2 Long Short-Term Memory

LSTM is a unit of RNNs in which the temporal dependency among elements of the TS can be captured. An LSTM block consists of a memory cell, an input gate, an output gate, and a forgetting gate. The memory cell stores values for arbitrary time intervals. In LSTM, the three gates are neurons with activation functions. Figure 7-3 represents an LSTM unit. In this figure, c_{t-1} and h_{t-1} denote the cell memory state and hidden state at the previous time, respectively. The input vector (x_t), c_{t-1}, and h_{t-1} are used to update the memory state, c_t, and attain the output, h_t, corresponding to x_t.

7-2-2-3 Recurrent Neural Network-based Stacked Denoising Autoencoder

AE is a type of neural networks employed for dimensionality reduction in a wide range of applications [188]. An autoencoder (AE) is composed of two fragments, the encoder and decoder. The input of the encoder is the original feature tensor, which is mapped to different space in the output of encoder, while decoder use the mapped features as the input and reconstruct the original feature space. Using the encoder output of the trained AE can provide more valuable information

in DTLR prediction problem, which suffers from high dimension feature space even after the feature reduction stage, discussed in Section 7-2-2. To make AE more robust to outlier and noisy data, DAE is proposed in [188]. In DAE, the original input features are reconstructed from a corrupted one, while stacking several DAEs, results in a deep feature extraction.

At time t, all features which are remained after the feature reduction stage, described in Section 7-2-2, are used to construct the input vector of DAE, x_t, a three-dimensional tensor $(1 \times l \times m)$ where m is the number of feature variants. To capture the sequential correlation of the TS, LSTM, described in Section 0, are used as building blocks of DAEs in this paper. After layer-wise training, SDAE is fine-tuned by using adaptive learning rate method in a mini-batch manner [192]. An RNN-based DAE can be formulated as follows:

$$\min_{\theta,\theta'}(L_{AE}(\boldsymbol{x},\boldsymbol{z})) \tag{7-12}$$

$$\tilde{x} \sim q_{\mathcal{D}}(x_t) \tag{7-13}$$

$$y_t = f_\theta(\tilde{x}_t, m_{t-1}) \tag{7-14}$$

$$z_t = g_{\theta'}(y_t, m'_{t-1}) \tag{7-15}$$

where $L_{AE}(\cdot)$ is the loss function, which needs to be minimized. In this paper, mean-squared-error is employed as the loss function. The tensor $\boldsymbol{x}$ $(n \times l \times m)$ contains all tensors $x_t, t = 1, \dots, n$, where n is the number of available points in the validation set; while tensor $\boldsymbol{z}$ consists of all tensors z_t, $t = 1, \dots, n$, which are the output of a DAE corresponding to input $\boldsymbol{x}$. Eq. (7-13) represents the stochastic process of destroying some elements of input vector, x_t, to form the corrupted input, $\tilde{x}$. In (7-14) and (7-15), y_t and z_t are the output of encoder and decoder blocks, respectively. $\boldsymbol{m}_{t-1}$ and $\boldsymbol{m}'_{t-1}$ are information passed from the calculation of y_{t-1} and z_{t-1}, respectively, as the result of recurrent units. As can be observed from Figure 7-3, if LSTM is used in DAE, $\boldsymbol{m}_{t-1} =$

$[c_{t-1}, h_{t-1}]$. In (7-14) and (7-15), $f_\theta(\cdot)$ and $g_\theta(\cdot)$ are the functions associated with encoder and decoder

blocks, respectively.

7-3 Proposed Training Framework for Probabilistic Dynamic Thermal Line Rating Prediction

Training of the proposed DTLR is conducted in several steps according to the scheme presented in Figure 7-4. First, the data is divided into three parts of training, validation, and test datasets. The training dataset is used to tune the model parameters, while the validation dataset is used to assess the performance of the model during training. The test dataset is used to evaluate the trained model. The redundant or non-informative features are removed as described in Section 7-2-2-1. Next, layer-wise training of the DAE, as explained in Section 7-2-2-3, is performed; then, SDAE is formed, and its parameters are fine-tuned by mean-sqaured-error (MSE) loss function. Afterward, the autoregressive prediction model, named as Model 1, is trained using a series of DTLR TS lags, simultaneously, a many to one prediction model, labeled as Model 2 is tainted using the feature tensor obtained from the SDAE output and DTLR TS lags. It is worth mentioning both Models 1 and 2 are based on RNNs and trained using MSE as the cost function, while the weights of SDAE are frozen. Then, an ensemble model which employs the predictions of Models 1 and 2 as the input to provide a final prediction of DTLR is trained deterministically using MSE as the cost function. The ensemble model is named as Model 3.

The thus far developed model is trained prognostically. The cost function, presented in (7-16)–(7-19), is put forward to fine-tune the overall model such that the DTLR model provides the lower

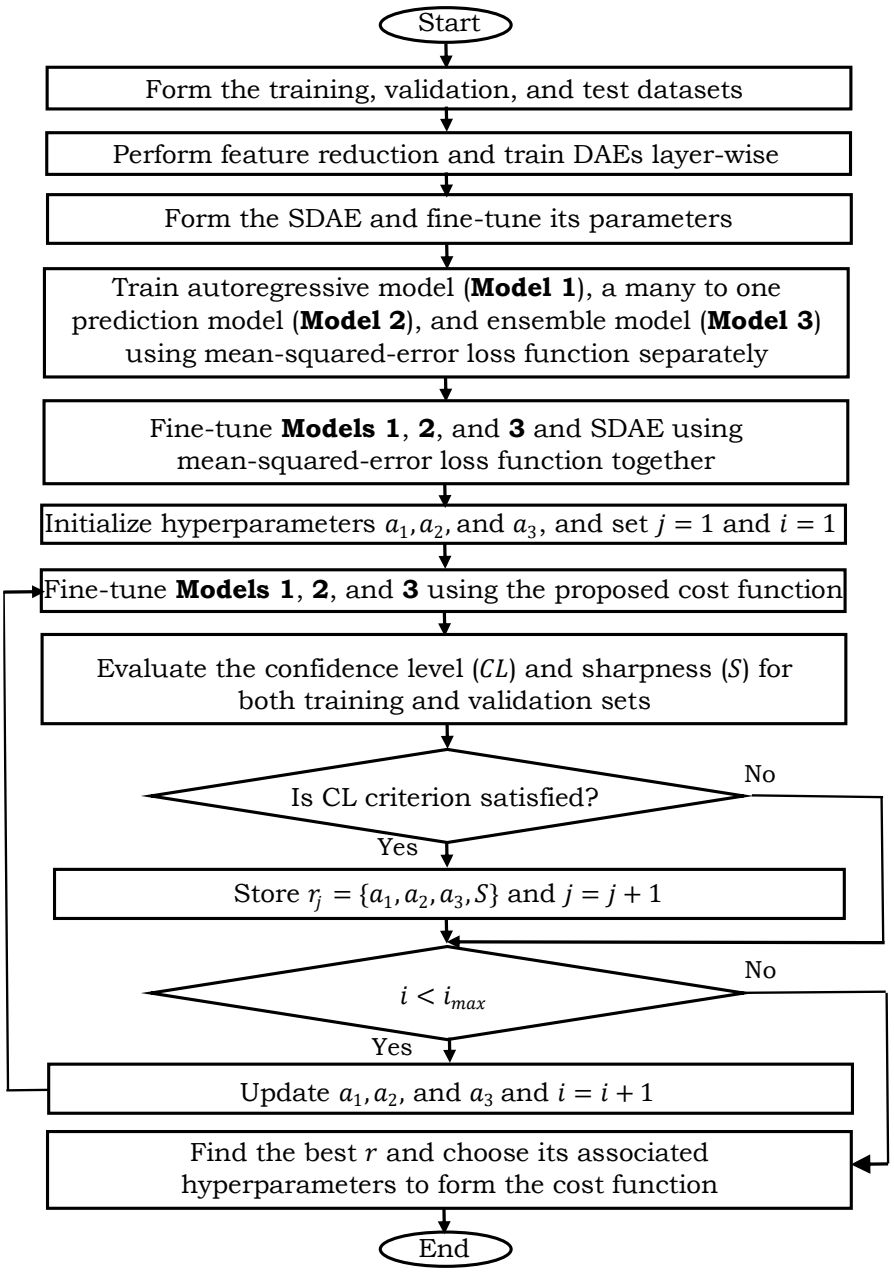

Figure 7-4: Flowchart of the training the proposed DTLR prediction.

bound of DTLR values with specified confidence level (CL) while the sharpness (S), defined as the average deviation of predicted DTLR from actual DTLR, is minimized.

$$C\left(DTR_{t+1}, \underline{DTR_{t+1}}\right) = \sum_{i=1}^{i=3} a_i(sign(\delta_i) + 1) \times \delta_i^2 \tag{7-16}$$

$$\delta_1 = \underline{DTR_{t+1}} - DTR_{t+1} \tag{7-17}$$

$$\delta_2 = DTR_{t+1} - \underline{DTR_{t+1}} \tag{7-18}$$

$$\delta_3 = \underline{DTR_t} - DTR^{\alpha} \tag{7-19}$$

where $\underline{DTR_{t+1}}$ is the predicted DTLR value, which is expected to provide a lower bound for the actual DTR_{t+1}. In (7-16), the cost function is composed of three terms, associated to $\delta_i, i = 1,2,3$. $a_i, i = 1,2,3$ are the hyperparameters, which determine the importance of each term. The term related to δ_1 penalizes the prediction model when prediction model results in $\underline{DTR_{t+1}}$ values above DTR_{t+1}. When $\underline{DTR_{t+1}} < DTR_{t+1}$, the term related to δ_1 in the cost function will become zero. The term corresponds to δ_2 penalizes the prediction model if the prediction model can provide a reliable $\underline{DTR_{t+1}}$; in another word, $\underline{DTR_{t+1}} < DTR_{t+1}$. While the last term penalizes prediction model if $\underline{DTR_{t+1}}$ is lower than the αth percentile of DTLR (DTR^{α}) in the training dataset, where $\alpha = 1 - CL$. This term is introduced to enhance the sharpness (S) of the prediction model defined as follows:

$$S = \frac{\frac{1}{N}\sum_{t=1}^{N}\left|DTR_{t+1} - \underline{DTR_{t+1}}\right|}{R} \tag{7-20}$$

where N is the number of points in training, validation, or test datasets, and R is the range of DTLR values.

The values of hyperparameters $a_i, i = 1,2,3$ are determinant of the CL and S as described above. Therefore, a grid search is conducted with the maximum iteration i_{max}. The hyperparameter sets which meet the CL requirement are stored as the set r_j, where j is the number of sets stored. After running the grid search for i_{max} iterations, the set r which results in the lowest S is used to form the cost function.

7-4 Case Studies and Comparisons

7-4-1 Data Description

We performed the analysis based on a 5-year data (Jan. 1, 2010 to Jan 1, 2015), recorded from M2 met tower in National Wind Energy Center (NWEC) [31], located in Denver, US. It is assumed that the measurements correspond to an OHL constructed from 400 mm^2 Drake 26/7 ACSR conductor. The elevation of the conductor from sea level is set to 1861 m.

The specification of this conductor is summarized in Table 7-1 [165], where, STLR is calculated based on low-speed perpendicular wind, 0.6 m/s, high ambient temperature, $40^{\circ}C$, and full solar heating, 1000 W/m^2 [193].

7-4-2 Analyzing Feature Candidates

To evaluate the dependency of DTR to different feature candidates, a 5-year meteorological data are used to generate historical DTR, qc, and qr TS, using the procedure described in Section II-A.

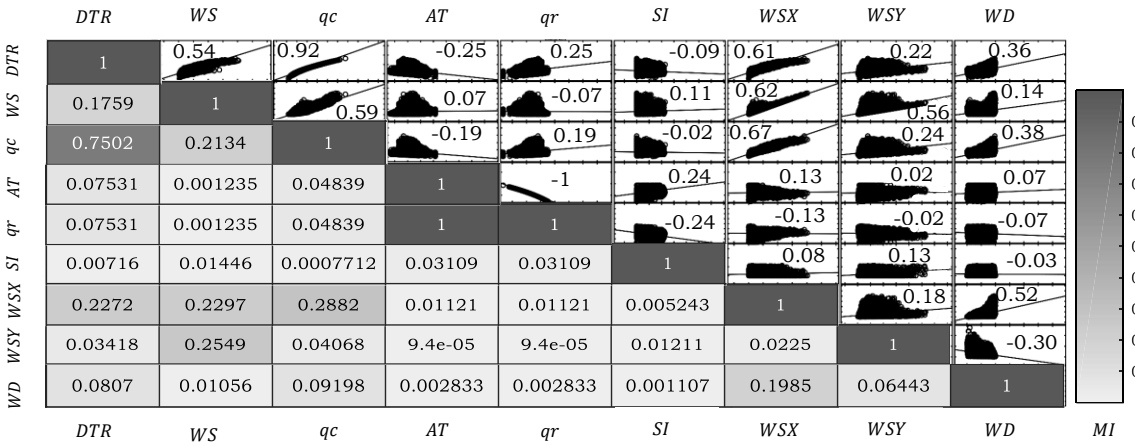

Figure 7-5: Dispersion of DTLR with respect to different feature types and Kendall τ rank coefficient and mutual information of different features and DTLR.

Table 7-1: Specifications of The Conductor

STLR	685 A
Outside diameter of conductor	28.12 mm
Minimum conductor temperature	25 C
Maximum conductor temperature	75 C
Conductor resistance at minimum temperature	0.07284 Ω/km
Conductor resistance at maximum temperature	0.08689 Ω/km
Solar absorptivity and emissivity	0.5

Figure 7-5 represents the dispersion of DTR with respect to various feature types in the upper triangle. In this figure, WSX and WSY represent the wind speed decomposed in X and Y axes, respectively. Kendall τ rank coefficient is widely employed as a non-parametric statistical test in hypothesis test to identify the statistical dependency between two random variables [194]. The detailed explanations on Kendall τ rank coefficient can be found in [194]. The presented quantities in the subplots, forming the upper triangle, represent the Kendall τ rank coefficient of different features vis-à-vis each other and DTR. Based on the calculated Kendall τ rank coefficients, which represent the relations between DTR and other features, and also their associated p-values, it is

observed that the null hypothesis at the significance level of (0.05) is rejected; and therefore, DTR possesses relations with other feature species.

Besides, from this figure one can perceive that qc has the most influential factor in DTR compared to all features. This strong relation can be also found from the MI values which are presented in diagonal subplots and the lower triangle of Figure 7-5. The MI values testify the high dependency of DTR to qc. Also, from this figure, one can note that among meteorological variables, WS demonstrates the strongest correlation with DTR, while $MI(WS; DTR)$ is significantly lower than $MI(qc; DTR)$. On the other hand, it is found that there is strong correlation between qr and AT; and therefore, for the sake of preventing from curse of dimensionality and unnecessary increment of computation burden [195], considering only AT could be sufficient. Moreover, from $MI(DTR; SI)$, it can be recognized that SI can provide the minimum information in DTLR forecasting. To this end, this feature species is eliminated in the feature reduction stage described in Section II-B. It is worth mentioning that the data are discretized with respect to median values of different features in order to calculate the MI values. This type of discretization is widely used in calculating MI among continuous variables [195].

7-4-3 Description of Benchmark Models

Three benchmark models are utilized in this paper, including persistence [135], QR [174], and QRF [178] prediction models. The persistence model (PM) is a conventional prediction model which is widely used for short-term prediction of meteorological variables [12, 135]. As DTLR is also a meteorologically dependent variable, we make use of PM as one of our benchmark models. The simplicity of PM can facilitate the comparison of the proposed model with other prediction models [135]. QR is recently proposed for non-parametric probabilistic prediction of DTLR, which

is the focus of this study. Therefore, comparing the efficacy of the proposed model with QR provides some insights about the superiority of the proposed model with respect to other non-parametric probabilistic DTLR prediction models. As the last benchmark model, the state-of-the-art QRF-based probabilistic DTLR prediction model [168, 178] is employed. To carry out a fair comparison, the proposed model and benchmark models utilize similar input variables, as elaborated in Section 7-2. A notable exception is PM, for which the mean and variance of the latest wind power observations form a Gaussian distribution representing the uncertainty of upcoming samples [135].

7-4-4 Evaluation Metrics

Three evaluation metrics are used to appraise the performance of different DTLR models. CL of the prediction model is the most imperative evaluation criterion for probabilistic DTLR prediction. Any deviation of CL from the preferred CL, CL^*, can lead to unprecedented issues. S, defined in (7-20), is another metric which is utilized as a supplementary evaluation metric. In a perfect DTLR prediction, $CL = CL^*$ while $S \to 0$. Root-mean-squared-error (RMSE) [41], which is a valuable measure and signifies large deviations of DTLR prediction from its actual value, is also employed as another evaluation metric.

7-4-5 Numerical Studies

A six-month data (Jan. 1, 2010 to Jul. 1, 2010) are used for numerical comparisons. 85% of data are used for training and validation, while 15% is employed for testing. In this section, first, the

Table 7-2: Comparing the Performance of the Proposed Model with and without Latent Variables as Input

	$\boldsymbol{CL}(\%)$	$\boldsymbol{S}(\%)$	$\boldsymbol{RMSE}$
Without latent variable	99.19	24.93	230.53
With latent variable	99.01	23.67	221.52

effectiveness of considering latent predictors (i.e., convective cooling and radiated heat loss rate) is investigated. Afterwards, the proposed DTLR is compared with the benchmark models for different CL preferences, using various evaluation metrics.

7-4-5-1 The Impact of Latent Predictors

In Section 7-2, it was shown that in comparison to directly observed meteorological variables, radiated heat loss rate has more pretentious relation with DTLR.

To investigate empirically the efficacy of considering a series of lags associated with the mentioned latent variables as predictors, in this section a case study is conducted using the proposed DTLR prediction with and without the latent predictor. Table 7-2 summarizes the conducted case study for $CL^* = 99\%$. As can be observed from this table, the experiential results are in line with the theoretically expected outcome. From this table, one can notice that considering latent variables can reduce S and $RMSE$, while the CL criterion is satisfied. So, it can be surmised that radiated heat loss rate information can provide further information about the DTLR pattern, and therefore, the prediction can be performed more precisely.

7-4-5-2 Numerical Comparisons of Different Prediction Models

The performance evaluations of different prediction models for $CL^* = 95\%$ and 99% are tabulated in Table 7-3. As can be observed from this table, neither PM and nor QR provide a reliable prediction for $CL^* = 95\%$. On the other hand, QRF and the proposed model can securely

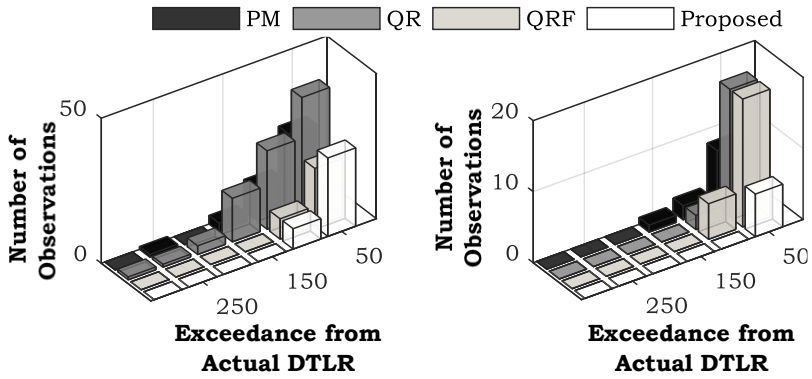

Figure 7-6: Exceedance of DTLR values as the result of various prediction models; ; (a) $CL^* = 95\%$ and (b) $CL^* = 99\%$.

Table 7-3: Performance Evaluation of Different Prediction Models

	Prediction Models	***CL* (%)**	***S* (%)**	***RMSE***
$CL^* = 95\%$	PM	93.87	21.93	246.52
	QR	91.87	15.99	160.02
	QRF	95.86	20.28	195.73
	Proposed Model	95.10	19.40	186.18
$CL^* = 99\%$	PM	99.08	33.44	329.69
	QR	96.93	20.97	201.01
	QRF	96.63	22.74	214.98
	Proposed Model	99.01	23.67	221.52

predict the DTLR for $CL^* = 95\%$, but the proposed model results in improvement in sharpness (S) and $RMSE$. For $CL^* = 99\%$, QR and QRF models are unable to meet the reliability level. PM provides a reliable prediction by sacrificing the sharpness, while the proposed model can yield a sharp and reliable predictions for $CL^* = 99\%$, which is one of the requirements stated in the literature for implementing a secure DTLR [167]. Thus, from this case study, it can be observed that the proposed model can facilitate secure DTLR employment.

To further investigate the performance of the proposed model compared to benchmark models, the bar graph, describing the number of exceedances from the actual DTLR value for different levels of deviation as the result of different prediction models in $CL^* = 95\%$ and $CL^* = 99\%$, are presented in Figure 7-7. For $CL^* = 95\%$ QRF and the proposed model provides comparable

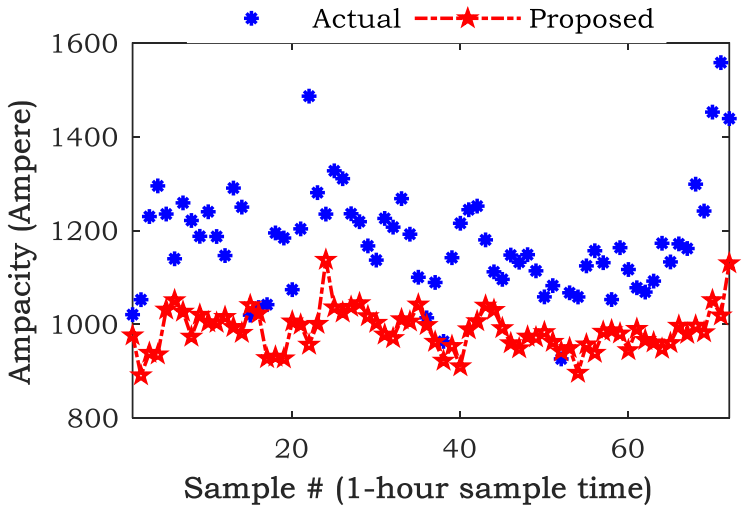

Figure 7-7: Prediction of DTLR using the proposed model for $CL^* = 95\%$.

performance, while the proposed model result in the minimum exceedances in $CL^* = 99\%$. Therefore, employing the proposed DTLR leads to minimal overheating the OHL conductors.

Figure 7-6 depicts a sample episode for prediction with the proposed DTLR and $CL^* = 95\%$. The figure covers the prediction for three successive days to provide the view of the high volatility in DTLR values and efficacy of the proposed model to track this volatile pattern.

7-5 Conclusion

This paper proposed a deep learning-based probabilistic DTLR prediction model for hour-ahead power system operation problems. The latent variants, attained in the process of calculating DTLR value, have been considered as new predictors of the proposed model. The large number of feature candidates result in high feature space dimension. Therefore, SDAE is employed to reduce the feature dimension and extract the meaningful ones. A training strategy is devised to train the DNN, and a cost function is put forward to train the prediction model in a probabilistic manner. The proposed prediction framework considers no hypothesis about the uncertainty of DTLR.

Simulation results confirm the efficacy of the proposed model and its superiority compared to benchmark models.

Further research can be conducted to include the NWP results in the proposed DTLR prediction model and form a long-term prediction framework. Also, studies can be carried out to adopt the proposed DNN-based probabilistic prediction framework for prediction of other meteorologically dependent variables in power systems.

8- CONCLUSIONS AND FUTURE WORK

8-1 Conclusion

In this thesis, I have developed various forecasting tools, which benefit from advanced signal processing techniques, machine learning tools, and novel problem formulations. The forecasting tools are tailored for short-term prediction of two important large-scale VRESs (i.e., wind and tidal) and DTLR. Accurate prediction of large-scale VRESs facilitates enhancing the renewables penetration in power systems by lowering the uncertainties in power system operation problems, while the DTLR forecast alleviates the need for transmission systems expansion, required for delivering the generated power from renewables to load center, by employing the untouched additional headroom of existing OHLs.

The significance of this thesis is its explicit focus on the short-term prediction of meteorologically dependent variables of power systems. SOs can employ the developed models for daily problems, such as optimal power flow, generation-load balancing, short-term multi-step unit commitment, etc. The detailed conclusion of Chapters 0-7- are summarized next.

Chapter 0 presents a short-term multi-step WPP, taking the chaotic nature of wind power TS component into account in both decomposition and prediction stages. In the decomposition stage, using chaotic TS analysis the chaotic components obtained from EEMD are identified; and then, the chaotic components become more predictable by eliminating high-frequency variations that are small in amplitude through SSA theory. In the prediction stage, the proposed WPP makes use of the localized direct and iterative prediction modules for chaotic and non-chaotic components, respectively. The numerical studies confirm that the proposed MSSSA is more capable than

EEMD for prediction of chaotic components. This is mainly due to introducing SSA in the decomposition wind power TS. Using the SSA enables maintaining the general trend of the chaotic components, while omitting the high-frequency, random-looking, and low amplitude sub-components of chaotic components. Numerical comparisons testify the efficacy of the MSSSA-LSSVM for predicting aggregated wind power generation of a region, as well as wind power generation in different locations for a single wind farm. The proposed model demonstrates superior performance compared to widely-used benchmark models, as well as the state-of-the-art models.

In Chapter 3-, the proposed model of Chapter 0 is adapted for multi-step prediction of TCS and TCD which are the determining factor in tidal energy harvesting using tidal stream technology. Similar to TS associated with meteorological variables (e.g., wind speed, wind direction, solar irradiance, etc.), tidal data also possesses nonlinearity and nonstationarity. To develop the proposed model a variant of EMD, named ICEEMDAN, is established and employed as a decomposition approach. The EMD suffers from mode mixing problems, which deteriorate the capability of the EMD in reducing the non-stationarity of TS. Besides, a variant of EMD, known as EEMD, which is developed to address the mode mixing problem, poses new issues, as elaborated this chapter. As case studies reveal, the proposed ICEEMDAN-based prediction model results in lower prediction error compared to the EEMD-based model. Also, to further enhance the prediction accuracy, the aggregation of the predicted components, employed in the proposed prediction model in Chapter 0, is further improved by devising an error correction stage, composed of a combination of ELMs. The empirical results indicate that considering the combination of ELMs for aggregation of predicted components leads to improvement in final prediction accuracy.

There are circumstances, in which several wind farms are located in the vicinity of each other. In the case that either the wind farm owners might be interested in sharing their wind power generation data with each other, or all the wind farms belong to the same owner. Therefore, spatiotemporal correlations among wind farms can be considered to enhance the prediction accuracy of each individual wind farm. Therefore, the proposed model in Chapter 0 can be adapted to benefit from informative features, provided by other wind farms to improve the prediction accuracy. To this end, Chapter 4- puts forward time-Frequency-Based spatiotemporal modeling of WPP. The proposed WPP benefits from information with different frequency range, provided from various wind farms; thus the WPP accuracy is significantly enhanced. Also, it has been observed that considering the spatiotemporal correlation might deteriorate the WPP accuracy unless the decomposition approaches are employed.

The proposed models developed in Chapters 2-4- are developed with the aim of enhancing the deterministic prediction accuracy, which is of great importance for power system operation problems, which are solved by deterministic optimization. However, there are many power system operation problems, which make use of interval optimization or robust optimization. PI is of great importance for such optimization problems. For this reason, Chapter 5- presents a novel WPPI benefiting from a bi-level optimization structure along with a convex and differentiable cost function for tuning the unknown parameters of prediction engine. Global tuning of prediction engine parameters via classic optimization and alleviating the need for heuristic optimization are the main features of the proposed WPPI formulation. In the proposed WPP, hypermeters are found based on a multi-objective optimization procedure, which results in a Pareto front with a wide range of $PINAW$ and RL. The comprehensive case studies reveal the consistency in the results of

WPPI evaluations for training and testing datasets. Moreover, the case studies verify the superiority of the proposed WPPI framework compared to traditional benchmark and state-of-the-art prediction models. The proposed WPPI model provides sharp and reliable WPPIs in an extensive case study. SOs can benefit greatly from the proposed WPPI in a wide range of short-term power system operation tasks.

The proposed PI model in Chapter 5- can address the issues, posed by non-differentiability and non-convexity of the prevalent PI models, which requires heuristic optimization for tuning a large number of prediction engine parameters. However, the proposed cost function is nonlinear. In Chapter 6-, an AAQR-based PI model, which benefits from a linear formulation, is developed. The proposed model is adapted for uncertainty modeling of a variety of tidal TS, including TCS, TCD, and TL. The shortcomings of traditional QR-based NPI, including the error in the prediction engine and symmetrical selection of a pair of quantiles probabilities, are discussed and addressed by introducing two hyperparameters. Hyperparameters and prediction engine parameters are tuned through a bi-level optimization framework. The upper-level optimization problem is solved by a hybrid heuristic optimization, while LP is carried out for the lower-level problem. The developed model has been successfully tested on a wide range of tidal data; the results obtained, and comparisons reported in Chapter 6- show that the proposed approach can be used as a useful tool for maximizing tidal energy deployment by providing more accurate uncertainty modeling of the most inflectional and uncertain factors.

In Chapter 7-, a prediction model for DTLR prediction is put forward. In this chapter, the importance of considering features, associated with latent variables (e.g., convective cooling rate) are discussed. The empirical case studies validate the informativeness of the convective cooling rate in DTLR prediction. To address the issue, regarding the high dimension of feature space,

SDAE based on LSTM is adapted for feature extraction, and a DNN-based prediction model is developed. A training strategy is devised such that the proposed model first deterministically trained; and subsequently, a novel differentiable cost function is employed to train the model for probabilistic prediction based on the required secure CL utilizing gradient descent approaches.

The proposed models developed in Chapters 0-7- have mainly proposed with the aim of addressing a variety of barriers in enhancing the prediction accuracy of wind power generation, tidal energy, and DTLR. Using several real-world case studies, the proposed models, introduced in this thesis, have been tested; and their performance compared to both traditional and state-of-the-art prediction models. The results reveal the high superiority of the proposed approach in forecasting large-scale renewables and DTLR. The prediction models could facilitate the power system operation problems under high penetration of large-scale renewables, in which considerable of the level of uncertainty is observed.

8-2 Future Work

Further possible extensions to this work are as follows:

1. Chapter 2 could be extended for long-term prediction. This could be realized by developing a hybrid model and taking the NWP results and meteorological variables into attention. Considering these factors can significantly increase the number of potential predictors, from which the suitable predictors need to be selected or extracted. Therefore, it is worth to develop FS and feature extraction approaches and adapt the proposed model in Chapter 2 for long-term prediction.
2. The work in Chapter 2 could be used as the input for conditional probabilistic forecasting. In the literature, there are several probabilistic wind power forecasting models, in which the

deterministic wind power prediction values of upcoming samples are considered as inputs. The more accurate the wind power is deterministically predicted; the sharper uncertainty model can be realized with the required *RL*. As shown in Chapter 2, the proposed model can provide more precise prediction compared to different prediction models. Therefore, the study presented in Chapter 2 could be extended to conditional probabilistic prediction.

3. In Chapters 2, 4, and 5, direct WPP tools have been developed. In such prediction models, the wind power is directly forecasted without any mapping stage in which prediction of meteorological variables are mapped to the wind power values using power curve models. As stated earlier, direct WPP benefits from several advantages, namely preventing from the erroneous mapping of meteorological variables to wind power due to lack of a precise power curve model. But it is also worth mentioning that wind power TS is more prone to outlier due to the possibility of wind power curtailment occurrence. Besides, due to cut-off and cut-in speed thresholds, the wind power TS can be more volatile and irregular in comparison to the meteorological variables affecting wind power. In other words, prediction of meteorological variables might be more precise. To this end, using the models proposed in Chapters 2, 4, and 5 as the foundations, further study could be conducted to predict the meteorological variables, while a detailed power curve model needs to be developed to prevent from imprecise mapping of meteorological variables to wind power.
4. Chapters 5-7 put forward efficient uncertainty modeling approaches for short-term horizons. The proposed models in Chapters 5-7 could be used as platforms for long-term prediction, in which more predictors such as NWP results are taken into consideration.
5. Chapters 5-7 propose different formulations for probabilistic modeling of uncertainty in wind power, tidal energy, and DTLR, respectively. In the developed models, the upper-level

optimization problem seems to be capable of taking any custom cost function, which can be defined based on the power system operator's preference. However, in this work, only RL and $PINAW$ are considered as the cost functions in the upper-level optimization problem. Therefore, further research could be carried out to understand the efficacy of the proposed model in forming optimized PIs based on power system operation problems (e.g., optimal power flow, economic dispatch, etc.).

6. The impacts of enhancement in accuracy of renewable and DTLR prediction as the result of the proposed models in power system problems could be conducted in future studies. Mostly, in the literature, the power system operation problems are performed using naïve prediction models; however, accurate prediction can play a crucial role in those problems. Therefore, assessing the proposed prediction tools in power system operation problems could provide more depth insight into the benefits of the proposed models.
7. The efficiency of the models proposed in Chapters 2-7 can be further improved by adapting state-of-the-art prediction engines. Prediction engines are the core components of prediction models. As the literature reveals adapting advanced prediction engines improves the accuracy of prediction models. There are evermore advances in machine learning, which can be beneficial for improving prediction accuracy.
8. The proposed models can be adapted for other prediction tasks in power systems, e.g., solar forecasting. Solar energy is another source of renewables, which can be harvested in both large- and small- scales. Solar irradiance is the most influential factor in solar energy harvesting. Similar to wind power and tidal data, solar irradiance TS posses non-stationarity and non-linearity; and therefore the proposed models in the thesis can be used for initial

modeling of solar prediction; however, adaptions is necessary based of understanding the unique characteristics of the solar irradiance data.

9. The models introduced in Chapters 2-7 can be used together as an ensemble prediction tool. Every prediction tool may have the best performance compared to other prediction models for a specific dataset. Therefore, it is beneficial to integrate the developed models to form a generalized model, in which depending on dataset some prediction models become more determinant compared to others.

Appendix A

Lyapunov Exponents

Lyapunov exponents provide qualitative and quantitative descriptions of the dynamic behavior in a TS. They evaluate how the value of the next element in a TS is related to the previous values. Lyapunov exponents were originally determined for dynamic systems, represented by differential equations, but were later extended to the analysis of historical data [67].

Lyapunov exponents represent the average exponential rates of divergence or convergence of a pair of nearby orbits in the reconstructed multi-dimensional space. In a chaotic TS, a pair of nearby points that has very similar states in the reconstructed multi-dimensional space can result in different future states. Such behavior deteriorates the predictability of the TS. The existence of at least one positive Lyapunov exponent can guarantee the chaosity of the time series. Hence, the value and sign of the MLE are of great importance.

For a nonlinear TS $\{x_i\}_{i=1}^{N}$ which is mapped into multi-dimensional space, by employing MOD, the process of finding the Lyapunov exponent is briefly illustrated in the following. For a point in the multi-dimensional space, the nearest neighbor is detected, and the Euclidean distance between these two points is denoted by $D(t_0)$. With time, these two points evolve in different orbits. Hence, at the next sample time, the Euclidean distance will change to $D'(t_1)$. The Lyapunov exponent can be calculated as follows:

$$\lambda = \frac{1}{M\tau}\sum_{i=1}^{M}\log_2\frac{D'(t_k)}{D(t_{k-1})} \tag{A-1}$$

where τ is the sample time and M is the number of evolutions.

Appendix B

DETERMINING THE EMBEDDING DIMENSION

The appropriate embedding dimension, or number of lags in the trajectory matrix, can help us to find the pattern of a nonlinear TS. Cao's method, proposed in [68], is a widely used embedding dimension selection approach. Let us assume *d* is an adequate number of lags and construct the trajectory matrix based on MOD with *d* elements in each row. Then, represent each row as a point in *d*-dimensional space. If a suitable number for *d* is chosen, the point that is the nearest neighbor for another will remain the nearest point in (*d*+1)-dimensional space as well.

For a TS $\{x_j\}_{i=1}^{N}$ with a trial dimension of *d*, the trajectory matrix of $\boldsymbol{X}$, which is $(N-d+1)$ by *d*, can be constructed. Two parameters ($E(d)$,$E^*(d)$) corresponding to *d* are defined as follows:

$$E(d) = \frac{1}{N-d}\sum_{i=1}^{N-d}\frac{\left\|X_i^{d+1} - X_{n(i,d)}^{d+1}\right\|}{\left\|X_i^{d} - X_{n(i,d)}^{d}\right\|}$$
$$E^* = \frac{1}{N-d}\sum_{i=1}^{N-d}\left|X_{i+1}^{d}(d) - X_{i+1(i,d)}^{d}(d)\right| \qquad \text{(B- 1)}$$

where $X_i^{d+1} = \left[X_i^d, X_i^d(d)\right]$, $X_{n(i,d)}^{d+1} = \left[X_{n(i,d)}^d, X_{n(i,d)+1}^d\right]$. In Eq. (B- 1), X_i^d is the *i*th row vector of $\boldsymbol{X}$ with d elements and $X_{n(i,d)}^d$ is its nearest neighbor, and $X_{i+1}^d(d)$ is *d*th element of *i*+1th row vector. If the rate of changes in $E(d)$ and $E^*(d)$ for two successive embedding dimensions of *d* and $(d+1)$ are negligible, *d* can be considered as the appropriate embedding dimension.

Appendix C

CONVEXITY AND DIFFERENTIABILITY OF COST FUNCTION

The function $C(\cdot)$ in (5-10) is convex with respect to θ if for all $\theta_1, \theta_2 \in \mathbb{R}_\theta$ and $0 < \gamma < 1$, the following equality is satisfied [196]:

$$C(\omega, \gamma \cdot \theta_1 + (1-\gamma) \cdot \theta_2, \boldsymbol{a}) \leq \gamma \cdot C(\omega, \theta_1, \boldsymbol{a}) + (1-\gamma) \cdot C(\omega, \theta_2, \boldsymbol{a}) \qquad \text{(C- 1)}$$

As elucidated in [196], Eq. (C- 1) graphically means that for any θ_1, $\theta_2 \in \mathbb{R}_\theta$, all points $\big(\theta_3, C(\omega, \theta_3, \boldsymbol{a})\big)$, where $\{\theta_3 = \gamma \cdot \theta_1 + (1-\gamma) \cdot \theta_2 : 0 \leq \gamma \leq 1\}$, are below or on the line connecting $(\theta_1, C(\omega, \gamma \cdot \theta_1 + (1-\gamma) \cdot \theta_1, \boldsymbol{a}))$ and $\big(\theta_2, C(\omega, \gamma \cdot \theta_2 + (1-\gamma) \cdot \theta_1, \boldsymbol{a})\big)$. Figure C-1 depicts $C(\omega, \theta, \boldsymbol{a})$ for arbitrary ω. As can be seen, $C(\cdot)$ is convex with respect to θ. Moreover, as one of the important properties of a convex function, $C(\cdot)$ is also convex with respect to an affine map of θ [196]. From (5-1), (5-2), (5-6), and (5-7), θ has a linear relationship to $\overline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\beta}}$; therefore, cost functions in (5-16), (5-17), and subsequently (5-19) are convex.

Furthermore, as $C(\cdot)$ is continuous and has derivatives for all $\theta \in \mathbb{R}_\theta$, as shown in (2-13), $C(\cdot)$ is also differentiable [197]. Then, due to the linear relationships of θ with $\overline{\boldsymbol{\beta}}$ and $\underline{\boldsymbol{\beta}}$, the lower-level cost function in (5-19) is also differentiable.

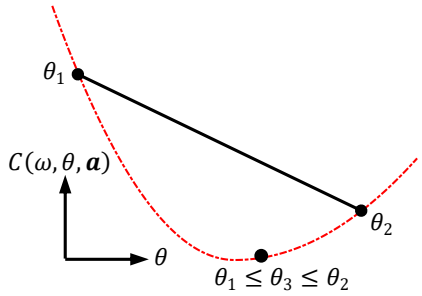

Figure C-1: Graph of convex function $C(\omega, \theta, \boldsymbol{a})$.

APPENDIX D

COPYRIGHT PERMISSION LETTERS

To Whom It May Concern:

I, C. Y. Chung, hereby grant permission to Mr. Nima Safari to reuse the below four articles in his thesis titles “Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems”.

1- N. Safari, C. Chung, and G. Price, “A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis,” IEEE Trans. Power Syst., vol. 33, no. 1, pp. 590-601, Jan. 2018.

2- N. Safari, B. Khorramdel, A. Zare, and C. Y. Chung, “An advanced multistage multi-step tidal current speed and direction prediction model,” in Proc. IEEE Electrical Power and Energy Conference (EPEC), 2017, pp. 1-6.

3- N. Safari, Y. Chen, B. Khorramdel, L. Mao, and C. Chung, “A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction,” in Proc. IEEE Electrical Power and Energy Conference (EPEC), 2017, pp. 1-6.

4- N. Safari, S. M. Mazhari, B. Khorramdel, and C. Y. Chung, “Tidal current and level uncertainty prediction via adaptive linear programming,” IEEE Trans. Sustain. Energy, *Early Access*.

5- N. Safari, S. M. Mazhari, and C. Y. Chung, “Very Short-Term Wind Power Prediction Interval Framework via Bi-level Optimization and Novel Convex Cost Function,” *IEEE Trans. Power Syst., Early Access*.

Date:

Signiture:

To Whom It May Concern:

I, G. Price, hereby grant permission to Mr. Nima Safari to reuse the below article in his thesis titles "Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems".

1- N. Safari, C. Chung, and G. Price, "A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis," IEEE Trans. Power Syst., vol. 33, no. 1, pp. 590-601, Jan. 2018.

Date:

Signiture:

To Whom It May Concern:

I, B. Khorramdel, hereby grant permission to Mr. Nima Safari to reuse the below three articles in his thesis titles “Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems”.

1- N. Safari, B. Khorramdel, A. Zare, and C. Y. Chung, “An advanced multistage multi-step tidal current speed and direction prediction model,” in *Proc. IEEE Electrical Power and Energy Conference (EPEC)*, 2017, pp. 1-6.
2- N. Safari, Y. Chen, B. Khorramdel, L. Mao, and C. Chung, “A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction,” in *Proc. IEEE Electrical Power and Energy Conference (EPEC)*, 2017, pp. 1-6.
3- N. Safari, S. M. Mazhari, B. Khorramdel, and C. Y. Chung, “Tidal current and level uncertainty prediction via adaptive linear programming,” *IEEE Trans. Sustain. Energy, Early Access*.

Date:

Signiture:

To Whom It May Concern:

I, A. Zare. Chung, hereby grant permission to Mr. Nima Safari to reuse the below article in his thesis titles “Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems”.

1- N. Safari, B. Khorramdel, A. Zare, and C. Y. Chung, “An advanced multistage multi-step tidal current speed and direction prediction model,” in *Proc. IEEE Electrical Power and Energy Conference (EPEC)*, 2017, pp. 1-6.

Date:

Signiture:

To Whom It May Concern:

I, L. Mao, hereby grant permission to Mr. Nima Safari to reuse the below article in his thesis titles "Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems".

1- N. Safari, Y. Chen, B. Khorramdel, L. Mao, and C. Chung, "A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction," in *Proc. IEEE Electrical Power and Energy Conference (EPEC)*, 2017, pp. 1-6.

Date:

Signiture:

To Whom It May Concern:

I, Y. Chen, hereby grant permission to Mr. Nima Safari to reuse the below article in his thesis titles "Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems".

1- N. Safari, Y. Chen, B. Khorramdel, L. Mao, and C. Chung, "A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction," in *Proc. IEEE Electrical Power and Energy Conference (EPEC)*, 2017, pp. 1-6.

Date:

Signiture:

To Whom It May Concern:

I, S. M. Mahari, hereby grant permission to Mr. Nima Safari to reuse the below article in his thesis titles “Forecasting Models for Integration of Large-Scale Renewable Energy Generation to Electric Power Systems”.

1- N. Safari, S. M. Mazhari, B. Khorramdel, and C. Y. Chung, “Tidal current and level uncertainty prediction via adaptive linear programming,” *IEEE Trans. Sustain. Energy, Early Access*.
2- N. Safari, S. M. Mazhari, and C. Y. Chung, “Very Short-Term Wind Power Prediction Interval Framework via Bi-level Optimization and Novel Convex Cost Function,” *IEEE Trans. Power Syst., Early Access*.

Date:

Signiture:

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

REFERENCES

[1] Z. Shi, H. Liang, and V. Dinavahi, "Direct interval forecast of uncertain wind power based on recurrent neural networks," *IEEE Trans. Sustain. Energy,* vol. 9, no. 3, pp. 1177-1187, Jul. 2018.

[2] Statistics Canada. [Online]. Available: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo02a-eng.htm, Accessed on Oct 28, 2018.

[3] Environment and Climate Change Canada Data. [Online]. Available: http://donnees.ec.gc.ca/data/substances/monitor/national-and-provincial-territorial-greenhouse-gas-emission-tables/C-Tables-IPCC-Sector-Provinces-Territories/?lang=en, Accessed on Oct 28, 2018.

[4] Union of Concerned Scientist. [Online]. Available: https://www.ucsusa.org/global-warming/science-and-impacts/science/each-countrys-share-of-co2.html, Accessed on Oct 28, 2018.

[5] Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Group_of_Seven, Accessed on Oct 28, 2018.

[6] The Gaurdian. [Online]. Available: https://www.theguardian.com/world/2015/jun/08/g7-leaders-agree-phase-out-fossil-fuel-use-end-of-century, Accessed on Oct 28, 2018.

[7] Independent. [Online]. Available: http://www.independent.co.uk/news/world/americas/canada-renewable-energy-catherine-mckenna-coal-fired-power-2030-a7430471.html, Accessed on Oct 28, 2018.

[8] H. Chitsaz, H. Shaker, H. Zareipour, D. Wood, and N. Amjady, "Short-term electricity load forecasting of buildings in microgrids," *Energ. Buildings,* vol. 99, pp. 50-60, Jul. 2015.

[9] N. Amjady, "Short-term hourly load forecasting using time-series modeling with peak load estimation capability," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 798-805, Aug, 2001.

[10] M. Rafiei, T. Niknam, J. Aghaei, M. Shafie-khah, and J. P. Catalão, "Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine," *IEEE Trans. Smart Grid,* vol. 9, no. 6, pp. 6961-6971, Nov. 2018.

[11] H. Sangrody and N. Zhou, "An initial study on load forecasting considering economic factors," in *Proc. IEEE Power and Energy Society General Meeting (PESGM),* 2016, pp. 1-5.

[12] N. Safari, C. Chung, and G. Price, "A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 590-601, Jan. 2018.

[13] N. Safari, O. A. Ansari, A. Zare, and C. Chung, "A novel decomposition-based localized short-term tidal current speed and direction prediction model," in *Proc. IEEE Power and Energy Society General Meeting 2017*, 2017, pp. 1-5.

[14] J. Zhan, C. Chung, and E. Demeter, "Time series modelling for dynamic thermal rating of overhead lines," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2172-2182, May 2017.

[15] Global Wind Energy Council. [Online]. Available: http://gwec.net/global-figures/graphs/

[16] K. D. Orwig, M. L. Ahlstrom, V. Banunarayanan, J. Sharp, J. M. Wilczak, J. Freedman, *et al.*, "Recent trends in variable generation forecasting and its value to the power system," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 924-933, 2015.

[17] P. Zamanidehkordi, "Data analytics in competitive electricity markets to uncover the impact of emerging technologies," *University of Calgary*, 2017.

[18] Canadian Wind Eenergy Association. [Online]. Available: http://canwea.ca/, Accessed on Oct 28, 2018.

[19] M. Lumley, "Annual Report 2016," *Fundy Ocean Research Center for Energy*, NS, CA, 2016.

[21] Natural Resources Canada. [Online]. Available: http://www.nrcan.gc.ca/energy/funding/current-funding-programs/cef/4955, Accessed on Oct 28, 2018.

[22] H. Chitsaz, "Developing energy forecasting tools in power systems: application to microgrids," *University of Calgary*, 2017.

[23] *Xcel Energy*. [Online] Available: https://www.xcelenergy.com/energy_portfolio/renewable_energy/wind, Accessed on Oct 28, 2018.

[24] K. Parks, Y.-H. Wan, G. Wiener, and Y. Liu, "Wind energy forecasting: A collaboration of the National Center for Atmospheric Research (NCAR) and Xcel Energy," *National Renewable Energy Laboratory (NREL)*, Golden, CO.2011.

[25] J. Hetzer, C. Y. David, and K. Bhattarai, "An economic dispatch model incorporating wind power," *IEEE Trans. Energy Convers.*, vol. 23, no. 2, pp. 603-611, Jun. 2008.

[26] H. Bitaraf, S. Rahman, and M. Pipattanasomporn, “Sizing energy storage to mitigate wind power forecast error impacts by signal processing techniques,” *IEEE Trans. Sustain. Energy,* vol. 6, no. 4, pp. 1457-1465, Oct. 2015.

[27] W. Wu, J. Chen, B. Zhang, and H. Sun, “A robust wind power optimization method for look-ahead power dispatch,” *IEEE Trans. Sustain. Energy,* vol. 5, no. 2, pp. 507-515, Apr. 2014.

[28] L. Bai, F. Li, H. Cui, T. Jiang, H. Sun, and J. Zhu, “Interval optimization based operating strategy for gas-electricity integrated energy systems considering demand response and wind uncertainty,” *Applied Energy,* vol. 167, pp. 270-279, Apr. 2016.

[29] M. Kuschke, S. Pertzsch, and K. Strunz, “Modeling of tidal energy conversion systems for primary response testing,” in *Proc. IEEE Power and Energy Society General Meeting IEEE 2012,* 2012, pp. 1-6.

[30] J. Xia, R. A. Falconer, and B. Lin, “Impact of different operating modes for a severn barrage on the tidal power and flood inundation in the Severn Estuary, UK,” *Applied Energy,* vol. 87, no. 7, pp. 2374-2391, Jul. 2010.

[31] National Wind Energy Center. [Online]. Available: https://www.nrel.gov/nwtc/index. html, Accessed on Oct 28, 2018.

[32] Y. Zhang, H. H.-C. Iu, T. Fernando, F. Yao, and K. Emami, “Cooperative dispatch of BESS and wind power generation considering carbon emission limitation in Australia,” *IEEE Trans. Ind. Inform.,* vol. 11, no.6, pp. 1313-1323, Dec. 2015.

[33] H. Bitaraf, “Mitigating impacts of high wind energy penetration through energy Storage and demand response,” Virginia Tech, 2016.

[34] H. Aly, “Forecasting, modeling, and control of tidal currents electrical energy systems,” Dalhousie University, 2012.

[35] R. Karki, S. Thapa, and R. Billinton, “A simplified risk-based method for short-term wind power commitment,” *IEEE Trans. Sustain. Energy,* vol. 3, no. 3, pp. 498-505, Jul. 2012.

[36] L. Landberg, “A mathematical look at a physical power prediction model,” *Wind Energy,* vol. 1, pp. 23-28, Sep. 1998.

[37] J. Catalao, H. Pousinho, and V. Mendes, “Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal,” *IEEE Trans. Sustain. Energy,* vol. 2, no. 1, pp. 50-59, Jan. 2011.

[38] K. Bhaskar and S. Singh, "AWNN-assisted wind power forecasting using feed-forward neural network," *IEEE Trans. Sustain. Energy,* vol. 3, no. 2, pp. 306-315, Apr. 2012.

[39] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, "Wind power forecasts using Gaussian processes and numerical weather prediction," *IEEE Trans. Power Syst.,* vol. 29, no. 2, pp. 656-665, Mar. 2014.

[40] G. Sideratos and N. D. Hatziargyriou, "An advanced statistical method for wind power forecasting," *IEEE Trans. Power Syst.,* vol. 22, no. 1, pp. 258-265, Feb. 2007.

[41] M. Khalid and A. V. Savkin, "A method for short-term wind power prediction with multiple observation points," *IEEE Trans. Power Syst.,* vol. 27, no. 2, pp. 579-586, May 2012.

[42] Q. Hu, S. Zhang, M. Yu, and Z. Xie, "Short-term wind speed or power forecasting with heteroscedastic support vector regression," *IEEE Trans. Sustain. Energy,* vol. 7, no. 1, pp. 241-249, Jan. 2016.

[43] D. Villanueva and A. Feijóo, "Normal-based model for true power curves of wind turbines," *IEEE Trans. Sustain. Energy,* vol. 7, no. 3, pp. 1005-1011, Jul. 2016.

[44] J. Shi, X. Qu, and S. Zeng, "Short-term wind power generation forecasting: direct versus indirect ARIMA-based approaches," *Int. J. Green Energy,* vol. 8, pp. 100-112, 2011.

[45] A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "A new fuzzy-based combined prediction interval for wind power forecasting," *IEEE Trans. Power Syst.,* vol. 31, no. 1, pp. 18-26, Jan. 2016.

[46] J. Wu, T. Ji, M. Li, P. Wu, and Q. Wu, "Multistep wind power forecast using mean trend detector and mathematical morphology-based local predictor," *IEEE Trans. Sustain. Energy,* vol. 6, no. 4, pp. 1216-1223, Oct. 2015.

[47] A. Kusiak, H. Zheng, and Z. Song, "Wind farm power prediction: a data-mining approach," *Wind Energy,* vol. 12, no. 3, pp. 275-293, Apr. 2009.

[48] Q. Wu and C. Peng, "Wind power generation forecasting using least squares support vector machine combined with ensemble empirical mode decomposition, principal component analysis and a bat algorithm," *Energies,* vol. 9, no. 4, p. 261, Apr. 2016.

[49] Y. Liu, J. Shi, Y. Yang, and W.-J. Lee, "Short-term wind-power prediction based on wavelet transform–support vector machine and statistic-characteristics analysis," *IEEE Trans. Ind. Appl.,* vol. 48, no. 4, pp. 1136-1141, Jul. 2012.

[50] H. Liu, H.-q. Tian, and Y.-f. Li, “Comparison of new hybrid FEEMD-MLP, FEEMD-ANFIS, wavelet packet-MLP and wavelet packet-ANFIS for wind speed predictions,” *Energ. Convers. Manage.,* vol. 89, pp. 1-11, Jan. 2015.

[51] Y. Ren, P. Suganthan, and N. Srikanth, “A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods,” *IEEE Trans. Sustain. Energy,* vol. 6, no. 1, pp. 236-244, Jan. 2015.

[52] D. Labate, F. La Foresta, G. Occhiuto, F. C. Morabito, A. Lay-Ekuakille, and P. Vergallo, “Empirical mode decomposition vs. wavelet decomposition for the extraction of respiratory signal from single-channel ECG: a comparison,” *IEEE Sens. J.,* vol. 13, no. 7, pp. 2666-2674, Jul. 2013.

[53] Y.-H. Wang, C.-H. Yeh, H.-W. V. Young, K. Hu, and M.-T. Lo, “On the computational complexity of the empirical mode decomposition algorithm,” *Physica A: Statistical Mechanics and its Applications,* vol. 400, pp. 159-167, Apr. 2014.

[54] X. An, D. Jiang, M. Zhao, and C. Liu, “Short-term prediction of wind power using EMD and chaotic theory,” *Commun. Nonlinear Sci Numer. Simul.,* vol. 17, no. 2, pp. 1036-1042, Feb. 2012.

[55] H.-B. Xie, T. Guo, B. Sivakumar, A. W.-C. Liew, and S. Dokos, “Symplectic geometry spectrum analysis of nonlinear time series,” *Proc. R. Soc. A,* vol. 470, p. 20140409, Oct. 2014.

[56] M. Ghanbarzadeh and M. Aminghafari, “A new hybrid-multiscale SSA prediction of non-stationary time series,” *Fluctuation and Noise Letters,* vol. 15, no.1, p. 1650005, Mar. 2016.

[57] L. Chen and X. Lai, “Comparison between ARIMA and ANN models used in short-term wind speed forecasting,” in *Proc. IEEE Asia-Pacific Power and Energy Engineering Conference (APPEEC),* 2011, pp. 1-4.

[58] M. G. De Giorgi, S. Campilongo, A. Ficarella, and P. M. Congedo, “Comparison between wind power prediction models based on wavelet decomposition with least-squares support vector machine (LS-SVM) and artificial neural network (ANN),” *Energies,* vol. 7, no. 8, pp. 5251-5272, Aug. 2014.

[59] L. Zunxiong and L. Jianhui, “Chaotic time series multi-step direct prediction with partial least squares regression,” *J. Syst. Eng. Electron.,* vol. 18, no. 3, pp. 611-615, Sep. 2007.

[60] C. Wan, J. Lin, J. Wang, Y. Song, and Z. Y. Dong, "Direct quantile regression for nonparametric probabilistic forecasting of wind power generation," *IEEE Trans. Power Syst.,* vol. 32, no. 4, pp. 2767-2778, Jul. 2017.

[61] J. Yan, Y. Liu, S. Han, Y. Wang, and S. Feng, "Reviews on uncertainty analysis of wind power forecasting," *Renew. Sust. Energ. Rev.,* vol. 52, pp. 1322-1330, Dec. 2015.

[62] Alberta Electric System Operator (AESO). [Online]. Available: http://www.aeso.ca/, Accessed on Oct 28, 2018.

[63] Y. Wen, W. Li, G. Huang, and X. Liu, "Frequency dynamics constrained unit commitment with battery energy storage," *IEEE Trans. Power Syst.,* vol. 31, no. 6, pp. 5115-5125, Nov. 2016.

[64] L. Xu and D. Tretheway, "Flexible ramping products," *California Independent System Operator (CAISO),* 2012.

[65] L. Zhang, W.-D. Zhou, P.-C. Chang, J.-W. Yang, and F.-Z. Li, "Iterated time series prediction with multiple support vector regression models," *Neurocomputing,* vol. 99, pp. 411-422, Jan. 2013.

[66] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in Adaptive Data Analysis,* vol. 1, no. 1, pp. 1-41, Jan. 2009.

[67] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica D: Nonlinear Phenomena,* vol. 16, no. 3, pp. 285-317, Jul. 1985.

[68] L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena,* vol. 110, no. 1-2, pp. 43-50, Dec. 1997.

[69] J. A. Suykens, T. Van Gestel, and J. De Brabanter, "Least squares support vector machines," World Scientific, 2002.

[70] J. Zhou, J. Shi, and G. Li, "Fine tuning support vector machines for short-term wind speed forecasting," *Energy Convers. Manag.,* vol. 52, no. 4, pp. 1990-1998, Apr. 2011.

[71] S. Xavier-de-Souza, J. A. Suykens, J. Vandewalle, and D. Bollé, "Coupled simulated annealing," *IEEE Trans. Syst., Man, Cybern., B,* vol. 40, pp. 320-335, Apr. 2010.

[72] A. Garulli, S. Paoletti, and A. Vicino, "Models and techniques for electric load forecasting in the presence of demand response," *IEEE Trans. Control Syst. Tech.,* vol. 23, no. 3, pp. 1087-1097, May 2015.

[73] Sotavento, [Online]. Available: http://www.sotaventogalicia.com/en, Accessed on Oct 28, 2018.

[74] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Trans. Power Syst.,* vol. 27, no. 1, pp. 134-141, Feb. 2012.

[75] N. Safari, B. Khorramdel, A. Zare, and C. Y. Chung, "An advanced multistage multi-step tidal current speed and direction prediction model," in *Proc. IEEE Electrical Power and Energy Conference (EPEC),* 2017, pp. 1-6.

[76] D. Magagna and A. Uihlein, "Ocean energy development in Europe: current status and future perspectives," *Inter. J. Marine Energy,* vol. 11, pp. 84-104, Sep. 2015.

[77] F. Birol, "World energy outlook 2010," *International Energy Agency,* vol. 1, 2010.

[78] M. Liu, W. Li, C. Wang, R. Billinton, and J. Yu, "Reliability evaluation of a tidal power generation system considering tidal current speeds," *IEEE Trans. Power Syst.,* vol. 31, no. 4, pp. 3179-3188, Jul. 2016.

[79] A. Kavousi-Fard and W. Su, "A combined prognostic model based on machine learning for tidal current prediction," *Trans. Geosci. Remote Sens.,* vol. 55, no. 6, pp. 3108-3114, Jan. 2017.

[80] A. Kavousi-Fard, "A hybrid accurate model for tidal current prediction," *Trans. Geosci. Remote Sens.,* vol. 55, no. 1, pp. 112-118, Jan. 2017.

[81] E. Pouranbarani, S. H. Sardoubie, and M. B. Shamsollahi, "Ictal EEG signal denoising by combination of a semi-blind source separation method and multiscale PCA," in *Proc. IEEE 1st International Iranian Conference on Biomedical Engineering (ICBME) 2016,* 2016, pp. 226-231.

[82] B. Khorramdel, H. Marzooghi, H. Samet, M. Pourahmadi-Nakhli, and M. Raoofat, "Fault locating in large distribution systems by empirical mode decomposition and core vector regression," *Inter. J. Electric. Power Energy Syst.,* vol. 58, pp. 215-225, Jun. 2014.

[83] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, "Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method," *Adv. Adapt. Data Anal.,* vol. 2, no. 2, pp. 135-156, Apr. 2010.

[84] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing," *Biomed. Signal Process. Control,* vol. 14, pp. 19-29, Nov. 2014.

[85] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE International Conference Acoustics, speech and signal processing (ICASSP) 2011,* 2011, pp. 4144-4147.

[86] National Oceanic and Atmospheric Administration. [Online]. Availabe: http://www.noaa.gov/, Accessed on Oct 28, 2018.

[87] H. Tang, K. Qu, G. Chen, S. Kraatz, N. Aboobaker, and C. Jiang, "Potential sites for tidal power generation: A thorough search at coast of New Jersey, USA," *Renew. Sust. Energ. Rev.,* vol. 39, pp. 412-425, Nov. 2014.

[88] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing,* vol. 70, no. 1-3, pp. 489-501, Dec. 2006.

[89] N. Safari, Y. Chen, B. Khorramdel, L. Mao, and C. Chung, "A spatiotemporal wind power prediction based on wavelet decomposition, feature selection, and localized prediction," in *Proc. IEEE Electrical Power and Energy Conference (EPEC),* 2017, pp. 1-6.

[90] A. S. Bubshait, A. Mortezaei, M. G. Simoes, and T. D. C. Busarello, "Power quality enhancement for a grid connected wind turbine energy system," *IEEE Trans. Ind. Appl.,* vol. 53, no. 3, pp. 2495-2505, May 2017.

[91] B. Khorramdel, H. Khorramdel, and H. Marzooghi, "Multi-objective optimal operation of microgrid with an efficient stochastic algorithm considering uncertainty of wind power," *Int. J. Rev. Mod. Sim,* vol. 4, pp. 3079-89, Dec. 2011.

[92] N. Amjady, F. Keynia, and H. Zareipour, "Wind power prediction by a new forecast engine composed of modified hybrid neural network and enhanced particle swarm optimization," *IEEE Trans. Sustain. Energy,* vol. 2, no. 3, pp. 265-276, Jul. 2011.

[93] Z. Wang, W. Wang, C. Liu, Z. Wang, and Y. Hou, "Probabilistic forecast for multiple wind farms based on regular vine copulas," *IEEE Trans. Power Syst.,* vol. 33, no. 1, pp. 578-589, Jan. 2018.

[94] M. A. Chitsazan, M. S. Fadali, A. K. Nelson, and A. M. Trzynadlowski, "Wind speed forecasting using an echo state network with nonlinear output functions," in *Proc. IEEE American Control Conference (ACC), 2017,* 2017, pp. 5306-5311.

[95] A. Botterud, Z. Zhou, J. Wang, J. Sumaili, H. Keko, J. Mendes, R. J., Bessa, and V. Miranda, "Demand dispatch and probabilistic wind power forecasting in unit commitment

and economic dispatch: a case study of Illinois," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 250-261, Jan. 2013.

[96] N. Zhang, C. Kang, Q. Xia, and J. Liang, "Modeling conditional forecast error for wind power in generation scheduling," *IEEE Trans. Power Syst.,* vol. 29, no. 3, pp. 1316-1324, May 2014.

[97] A. Lenzi, I. Steinsland, and P. Pinson, "Benefits of spatio-temporal modelling for short term wind power forecasting at both individual and aggregated levels," *Environmetrics,* vol. 29, no. 3, May 2017.

[98] L. Ye, Y. Zhao, C. Zeng, and C. Zhang, "Short-term wind power prediction based on spatial model," *Renewable Energy,* vol. 101, pp. 1067-1074, Feb. 2017.

[99] J. Dowell and P. Pinson, "Very-short-term probabilistic wind power forecasts by sparse vector autoregression," *IEEE Trans. Smart Grid,* vol. 7, no. 2, pp. 763-770, Mar. 2016.

[100] O. Abedinia, N. Amjady, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Trans. Power Syst.,* vol. 32, no. 1, pp. 62-74, Jan. 2017.

[101] S. Li, P. Wang, and L. Goel, "Wind power forecasting using neural network ensembles with feature selection," *IEEE Trans. Sustain. Energy,* vol. 6, no. 4, pp. 1447-1456, Oct. 2015.

[102] P. E. Meyer and G. Bontempi, "On the use of variable complementarity for feature selection in cancer classification," in *Workshops on Applications of Evolutionary Computation,* 2006, pp. 91-102.

[103] P. E. Meyer, C. Schretter, and G. Bontempi, "Information-theoretic feature selection in microarray data using variable complementarity," *IEEE J. Sel. Topics Signal Process.,* vol. 2, no. 3, pp. 261-274, Jun. 2008.

[104] L. Santos-Mayo, L. M. San-José-Revuelta, and J. I. Arribas, "A computer-aided diagnosis system with EEG based on the P3b wave during an auditory odd-ball task in schizophrenia," *IEEE Trans. Biomed. Eng.,* vol. 64, no. 2, pp. 395-407, Feb. 2017.

[105] S. Mallat, *A Wavelet Tour of Signal Processing*, San Diego, CA: Academic, 1998.

[106] N. Safari, S. M. Mazhari, and C. Y. Chung, "Very short-term wind power prediction interval framework via bi-level optimization and novel convex cost function," *IEEE Trans. Power Syst., Early Access.*

[107] Q. Wang, C.B. Martinez-Anido, H. Wu, A.R. Florita, and B.M. Hodge, "Quantifying the economic and grid reliability impacts of improved wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 7, no. 4, pp. 1525–1537, 2016.

[108] S. Soman, H. Zareipour, and O. Malik, "A review of wind power and wind speed forecasting methods with different time horizons," in *Proc. North American Power Symposium 2010*, Arlington, TX, 2010, pp. 1-8.

[109] R. A. Jabr, "Adjustable robust OPF with renewable energy sources," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4742-4751, 2013.

[110] M. Khodayar, O. Kaynak, and M. E. Khodayar, "Rough deep neural architecture for short-term wind speed forecasting," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2770-2779, Dec. 2017.

[111] Y. Zhao, L. Ye, P. Pinson, Y. Tang, and P. Lu, "Correlation-constrained and sparsity-controlled vector autoregressive model for spatio-temporal wind power forecasting," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5029-5040, Sep. 2018.

[112] P. Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions," *J. Royal Statist. Soc., Series C*, vol. 61, no. 4, 555-576, Aug. 2012.

[113] H. Quan, D. Srinivasan, and A. Khosravi, "Short-term load and wind power forecasting using neural network-based prediction intervals," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 303-315, Feb. 2014.

[114] G. Anastasiades and P. McSharry, "Quantile forecasting of wind power using variability indices," *Energies*, vol. 6, no. 2, pp. 662-695, Feb. 2013.

[115] R.J. Bessa, V. Miranda, A. Botterud, Z. Zhou, and J. Wang "Time-adaptive quantile-copula for wind power probabilistic forecasting," *Renewable Energy*, vol. 40, no. 1, pp. 29-39, Apr. 2012.

[116] P. Pinson and H. Madsen, "Ensemble-based probabilistic forecasting at Horns Rev," *Wind Energy*, vol. 12, no. 2, pp. 137-155, Mar. 2009.

[117] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255-270, Apr. 2014.

[118] A. Khosravi and S. Nahavandi, "Combined nonparametric prediction intervals for wind power generation," *IEEE Trans. Sustain. Energy*, vol. 4, no. 4, pp. 849–856, Oct. 2013.

[119] C. Wan, Z. Xu, P. Pinson, Z.Y. Dong, and K.P. Wong, "Optimal prediction intervals of wind power generation," *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1166-1174, May 2014.

[120] G. Zhang, Y. Wu, K.P. Wong, Z. Xu, Z.Y. Dong, and H-C. Iu, "An advanced approach for construction of optimal wind power prediction intervals," *IEEE Trans. Power Syst.*, vol. 30, no. 5, pp. 2706–2715, Sep. 2015.

[121] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction intervals for short-term wind farm power generation forecasts," *IEEE Trans. Sustain. Energy*, vol. 4, no. 3, pp. 602-610, Jul. 2013.

[122] R.V. Rao, V.J. Savsani, and D.P. Vakharia, "Teaching–learning-based optimization: an optimization method for continuous non-linear large-scale problems," *Information Sciences*, vol. 183, no. 1, pp.1-15, Jan. 2012.

[123] C. Wan, J. Wang, J. Lin, Y. Song and Z. Y. Dong, "Nonparametric prediction intervals of wind power via linear programming," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 1074-1076, Jan. 2018.

[124] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182-197, Apr. 2002.

[125] M. Alamaniotis, A. Ikonomopoulos, and L. H. Tsoukalas, "Evolutionary multiobjective optimization of kernel-based very-short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1477-1484, Aug. 2012.

[126] N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Prediction interval estimation of electricity prices using PSO-tuned support vector machines," *IEEE Trans. Ind. Informat.*, vol. 11, no. 2, pp. 322–331, Apr. 2015.

[127] A. Kavousi-Fard, "Modeling uncertainty in tidal current forecast using prediction interval-based SVR," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 708-715, Apr. 2017.

[128] G.B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[129] K. Mohammadi, S. Shamshirband, P.L. Yee, D. Petković, M. Zamani, and S. Ch, "Predicting the wind power density based upon extreme learning machine," *Energy*, vol. 86, pp. 232-239, Jun. 2015.

[130] A.M. Fraser and H.L. Swinney, “Independent coordinates for strange attractors from mutual information,” *Phys. Rev. A*, vol. 33, no. 2, pp. 1134–1140, Feb. 1986.

[131] A. Doostmohammadi, N. Amjady, and H. Zareipour, “Day-ahead financial loss/gain modeling and prediction for a generation company,” *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3360-3372, Sep. 2017.

[132] B. Efron, “Regression percentiles using asymmetric squared error loss,” *Statistica Sinica*, vol. 1, pp. 93–125, 1991.

[133] J.J. Moré, “The Levenberg-Marquardt algorithm: implementation and theory,” Numerical Analysis, Berlin, Germany: Springer-Verlag, 1978, pp. 105–116.

[134] Independent Electricity System Operator (IESO), [Online]. Available: http://www.ieso.ca/en/power-data/data-directory, Accessed on May 20, 2018.

[135] N. Safari, S. M. Mazhari, B. Khorramdel, and C. Y. Chung, “Tidal current and level uncertainty prediction via adaptive linear programming,” *IEEE Trans. Sustain. Energy, Early Access.*

[136] Y. Dai, Z. Ren, K. Wang, W. Li, Z. Li, and W. Yan, “Optimal sizing and arrangement of tidal current farm,” *IEEE Trans. Sustain. Energy,* vol. 9, no. 1, pp. 168-177, Jan. 2018.

[137] B. Papari, C. S. Edrington, and F. Kavousi-Fard, “An effective fuzzy feature selection and prediction method for modeling tidal current: a case of Persian Gulf,” *Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4956-4961, Sep. 2017.

[138] H. H. Aly and M. E. El-Hawary, “A proposed ANN and FLSM hybrid model for tidal current magnitude and direction forecasting,” *IEEE J. Ocean. Eng.*, vol. 39, no. 1, pp. 26-31, Jan. 2014.

[139] R. K. F. Neumann, “Tidal energy: technology brief,” International Renewable Energy Agency, 2014.

[140] Z. Ren, K. Wang, W. Li, L. Jin, and Y. Dai, “Probabilistic power flow analysis of power systems incorporating tidal current generation,” *IEEE Trans. Sustain. Energy,* vol. 8, no. 3, pp. 1195-1203, Jul. 2017.

[141] H. Lamb, *Hydrodynamics*, Cambridge, U. K., Cambridge Univ. Press, 1932.

[142] A. Shakya, S. Michael, C. Saunders, D. Armstrong, P. Pandey, S. Chalise, *et al.*, “Solar irradiance forecasting in remote microgrids using markov switching model,” *IEEE Trans. Sustain. Energy,* vol. 8, no. 3, pp. 895-905, Jul. 2017.

[143] A. S. B. M. Shah, H. Yokoyama, and N. Kakimoto, "High-precision forecasting model of solar irradiance based on grid point value data analysis for an efficient photovoltaic system," *IEEE Trans. Sustain. Energy,* vol. 6, no. 2, pp. 474-481, Apr. 2015.

[144] A. Kavousi-Fard, "A novel probabilistic method to model the uncertainty of tidal prediction," *Trans. Geosci. Remote Sens.,* vol. 55, no. 2, pp. 828-833, Feb. 2017.

[145] S. Rajasekaran, K. Thiruvenkatasamy, and T.-L. Lee, "Tidal level forecasting using functional and sequential learning neural networks," *Appl. Math. Model.,* vol. 30, no. 1, pp. 85-103, 2006.

[146] B. Polagye, J. Epler, and J. Thomson, "Limits to the predictability of tidal current energy," in *Proc. MTS/IEEE SEATTLE OCEANS, Seattle, WA, USA, Sep. 2010*, pp. 1-9.

[147] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation—with application to solar energy," *IEEE Trans. Power Syst.,* vol. 31, no. 5, pp. 3850-3863, Sep. 2016.

[148] K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, *et al.*, "LS-SVMlab Toolbox User's Guide: version 1.8," Katholieke Universiteit Leuven, 2010.

[149] A. Antonv,"Quantile regression through linear programming,"Mathematica for Prediction Algorithm. [Online]. Available: https://mathematica forprediction.wordpress.com, Accessed on Oct. 28, 2018.

[150] R. Koenker, *Quantile Regression*, Cambridge, U. K.: Cambridge Univ. Press, 2005.

[151] H. Bludszuweit, J. A. Domínguez-Navarro, and A. Llombart, "Statistical analysis of wind power forecast error," *IEEE Trans. Power Syst.,* vol. 23, no. 3, pp. 983-991, Aug. 2008.

[152] J. Cao, Z. Lin, and G.-B. Huang, "Self-adaptive evolutionary extreme learning machine," *Neural Process Lett.,* vol. 36, pp. 285-305, Dec. 2012.

[153] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.,* vol. 69, pp. 46-61, Mar. 2014.

[154] F. S. Gazijahani and J. Salehi, "Robust design of microgrids with reconfigurable topology under severe uncertainty," *IEEE Trans. Sustain. Energy,* vol. 9, no. 2, pp. 559-569, Apr. 2018.

[155] S. Mohanty, B. Subudhi, and P. K. Ray, "A new MPPT design using grey wolf optimization technique for photovoltaic system under partial shading conditions," *IEEE Trans. Sustain. Energy,* vol. 7, no. 1, pp. 181-188, Jan. 2016.

[156] *Tidal Power*. Available: http://www.energybc.ca/tidal.html

[157] B. Papari, C. S. Edrington, I. Bhattacharya, and G. Radman, "Effective energy management of hybrid AC-DC microgrids with storage devices," *IEEE Trans. Smart Grid, Early Access*.

[158] A. L. Bowley, *Elements of statistics*, New York, US, 1920.

[159] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The Computer J.*, vol. 7, pp. 308-313, 1965.

[160] *Fisheries and Oceans Canada*. [Online]. Available: http://www.dfo-mpo.gc.ca.

[161] *National Oceanic and Atmospheric Administration*. [Online]. Available: http://www.noaa.gov, Accessed on Oct. 28, 2018.

[162] K. A. Haas, H. M. Fritz, S. P. French, B. T. Smith, and V. Neary, "Assessment of energy production potential from tidal streams in the United States," *Georgia Tech Research Corporation*, 2011.

[163] *CPLEX Optimization Studio IBM ILOG*. [Online]. Available: https://www-01.ibm.com/software/websphere/products/optimization, Accessed on Oct. 28, 2018.

[164] D. M. Greenwood, *et al.*, "A comparison of real-time thermal rating systems in the US and the UK," *IEEE Trans. Power Del.*, vol. 29, pp. 1849-1858, Aug. 2014.

[165] *Standard for Calculating the Current-Temperature Relationship of Bare Overhead Conductors*, IEEE 738-2012, Jun. 2012, Power Energy Soc.

[166] R. Stephen, G. Pirovano, M. Tunstall, Y. Ojala, A. McCulloch, F. Jakl, *et al.*, "The thermal behaviour of overhead conductors. section 3: mathematical model for evaluation of conductor temperature in the unsteady state," *Electra,* vol. 174, pp. 58-69, 1997.

[167] T. Seppa and A. Salehian, "Guide for selection of weather parameters for bare overhead conductor ratings," *CIGRE Technical Brochure,* vol. 299, 2006.

[168] F. Teng, R. Dupin, A. Michiorri, G. Kariniotakis, Y. Chen, and G. Strbac, "Understanding the benefits of dynamic line rating under multiple sources of uncertainty," *IEEE Trans. Power Syst.,* vol. 33, no. 3, pp. 3306-3314, May 2018.

[169] D. M. Greenwood and P. C. Taylor, "Investigating the impact of real-time thermal ratings on power network reliability," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2460-2468, Sep. 2014.

[170] G. Murphy, "Implementation of real-time thermal ratings," *Scottish Power Energy Networks*, 2013.

[171] B. P. Bhattarai, J. P. Gentle, T. McJunkin, P. J. Hill, K. S. Myers, A. W. Abboud, *et al.*, "Improvement of transmission line ampacity utilization by weather-based dynamic line rating," *IEEE Trans. Power Del.,* vol. 33, no. 4, pp. 1853-1863, Aug. 2018.

[172] H. Hoekstra, C. Jansen, J. Hagen, J. Van Schuylenburg, J. Wisse, and W. Zitterrsteijn, "Weather forecasted thermal line rating model for the Netherlands," *Proc. Int. Conf. CIGRE Session,* Aug. 2012.

[173] C. Yanfei, T. Fei, R. Moreno, and G. Strbac, "Impact of dynamic line rating with forecast error on the scheduling of reserve service," in *Proc. 2016 IEEE Power and Energy Society General Meeting (PESGM),* 2016, pp. 1-5.

[174] Z. Wei, M. Wang, X. Han, H. Zhang, and Q. Zhang, "Probabilistic forecasting for the ampacity of overhead transmission lines using quantile regression method," in *Proc. IEEE Power and Energy Engineering Conference (APPEEC),* 2016, pp. 1632-1635.

[175] F. Fan, K. Bell, and D. Infield, "Probabilistic real-time thermal rating forecasting for overhead lines by conditionally heteroscedastic auto-regressive models," *IEEE Trans. Power Del.,* vol. 32, pp. 1881-1890, Aug. 2017.

[176] T. Ringelband, P. Schäfer, and A. Moser, "Probabilistic ampacity forecasting for overhead lines using weather forecast ensembles," *Electrical Engineering,* vol. 95, no. 2, pp. 99-107, Jun. 2013.

[177] X. Sun, P. B. Luh, K. W. Cheung, and W. Guan, "Probabilistic forecasting of dynamic line rating for over-head transmission lines," in *Proc. IEEE Power & Energy Society General Meeting,* 2015, pp. 1-5.

[178] J. L. Aznarte and N. Siebert, "Dynamic line rating using numerical weather predictions and machine learning: a case study," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 335-343, Feb. 2017.

[179] H.-M. Nguyen, J.-L. Lilien, and P. Schell, "Dynamic line rating and ampacity forecasting as the keys to optimise power line assets with the integration of res. the European project

twenties demonstration inside central western Europe," in *Proc. 22nd Int. Conf. Exhibit. Elect. Distrib.*, Jun. 2013, pp. 1–4.

[180] A. Salehian, "ARIMA time series modeling for forecasting thermal rating of transmission lines," in *Proc. IEEE Transmission and Distribution Conference and Exposition,* 2003, pp. 875-879.

[181] J. A. Jiang, Y. T. Liang, C. P. Chen, X. Y. Zheng, C. L. Chuang, and C. H. Wang, "On dispatching line ampacities of power grids using weather-based conductor temperature forecasts," *IEEE Trans. Smart Grid,* vol. 9, pp. 406-415, Apr. 2018.

[182] Y. Yang, R. Harley, D. Divan, and T. Habetler, "Adaptive echo state network to maximize overhead power line dynamic thermal rating," in *Proc. IEEE Energy Conversion Congress and Exposition,* 2009, pp. 2247-2254.

[183] J.-A. Jiang, J.-J. Wan, X.-Y. Zheng, C.-P. Chen, C.-H. Lee, L.-K. Su, *et al.*, "A novel weather information-based optimization algorithm for thermal sensor placement in smart grid," *IEEE Trans. Smart Grid,* vol. 9, no. 2, pp. 911-922, Mar. 2016.

[184] S. Karimi, A. M. Knight, P. Musilek, and J. Heckenbergerova, "A probabilistic estimation for dynamic thermal rating of transmission lines," in *Proc. IEEE Environment and Electrical Engineering International Conference (EEEIC),* 2016, pp. 1-6.

[185] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Trans. Power Syst.,* vol. 33, pp. 1087-1088, Jan. 2018.

[186] M. Khodayar and J. Wang, "Spatio-temporal graph deep neural network for short-term wind speed forecasting," *IEEE Trans. Sustain. Energy, Early Access.*

[187] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, pp. 1735-1780, 1997.

[188] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th international conference on Machine learning*, 2008, pp. 1096-1103.

[189] A. Pavlinić, V. Komen, and M. Uzelac, "Application of direct collocation method in short-term line ampacity calculation," *Electr. Pow. Syst. Res.,* vol. 155, pp. 216-224, Feb. 2018.

[190] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226-1238, Aug. 2005.

[191] S. M. Mazhari, N. Safari, C. Chung, and I. Kamwa, "A hybrid fault cluster and thévenin equivalent based framework for rotor angle stability prediction," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5594-5603, Sep. 2018.

[192] M. D. Zeiler, "ADADELTA: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.

[193] D. Douglass, W. Chisholm, G. Davidson, I. Grant, K. Lindsey, M. Lancaster, *et al.*, "Real-time overhead transmission-line monitoring for dynamic rating," *IEEE Trans. Power Del.*, vol. 31, no. 3, pp. 921-927, Jun. 2016.

[194] M. G. Kendall, "Rank correlation methods," Griffin, Oxford, 1955.

[195] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 306-318, Feb. 2009.

[196] S. Boyd and L. Vanderberghe, *Convex Optimization*, New York, NY, USA: Cambridge Univ. Press, 2004.

[197] F. Ayres Jr., *Theory and Problems of Differential and Integral Calculus*, New York, NY, USA, Schaum, 1964.