



Engaging soft computing in material and modeling uncertainty quantification of dam engineering problems

Mohammad Amin Hariri-Ardebili¹ · Fernando Salazar²

© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

Due to complex nature of nearly all infrastructures (and more specifically concrete dams), the uncertainty quantification is an inseparable part of risk assessment. Uncertainties might be propagated in different aspects depending on their relative importance such as epistemic and aleatory, or spatial and temporal. The objective of this paper is to focus on the material and modeling uncertainties, and to couple them with soft computing techniques aiming to reduce the computational burden of the conventional Monte Carlo-based finite element simulations. Several scenarios are considered in which the concrete and foundation material properties, the water level, and the dam geometry are assumed as random variables. Five soft computing techniques (i.e., random forest, boosted regression trees, multi-adaptive regression splines, artificial neural networks, and support vector machines) are employed to predict various quantities of interest based on different training sizes. It is argued that the artificial neural network is the most accurate algorithm in majority of cases, with enough accuracy as to be useful in reliability analysis as a complement to numerical models. The results with 200 samples in the training set are enough for reaching useful accuracy in most cases. For the simple prediction tasks, the results were predicted with less than 1% error. It is observed that increasing the number of input parameters increases the prediction error. The partial dependence plots provided most sensitive variables in dam design, which were consistent with the physics of the problem. Finally, several practical recommendations are provided for future applications.

Keywords Soft computing · Uncertainty quantification · Dam class · Machine learning · Response prediction · Big data

1 Introduction

Dams are complicated infrastructures whose responses are affected by many internal and external parameters (Moran et al. 2018). A proper safety assessment of dam structures requires a multi-disciplinary program combining different expertise such as structural engineering, geotechnical, hydraulic, and material. (Hariri-Ardebili 2018). On the other hand, dams are aging, and most of them were built at the time with limited technical information and computational tools. For example, American Society of Civil Engineers (ACSE) reports that by 2025, 70% of dams will be over 50 years old

in the USA (ASCE 2017). There is a similar situation in other countries.

In addition, with recent advances in computer science and information technology, there are large amount of data either directly gathered from the dam monitoring, or numerical simulations. The former one is usually used in the context of structural health monitoring (SHM) (Bukonya et al. 2014) to process the measured data during the life time of the structure. There is an increasing interest in the research community to integrate the overall response effectively (Salazar et al. 2015a).

Majority of the efforts in this area are focused on developing statistical tools to interpret the collected dam data. They are some times combined with numerical methods to improve the prediction (Fedele et al. 2005). Methods based on multiple linear regression (MLR), artificial neural networks (ANN), adaptive network-based fuzzy inference system (ANFIS), support vector regression (SVR), genetic algorithms (GA), multi-adaptive regression splines (MARS), and boosted regression trees (BRT), among others have been

Communicated by V. Loia.

✉ Mohammad Amin Hariri-Ardebili
mohammad.haririardabili@colorado.edu

¹ University of Colorado Boulder, Boulder, USA

² International Centre for Numerical Methods in Engineering (CIMNE), Barcelona, Spain

used for the characterization of dam behavior under environment loads (Mata 2011; Ranković et al. 2012, 2014; Stojanovic et al. 2013; Salazar et al. 2016).

1.1 Literature review: soft computing in dam engineering

On the other hand, the soft computing (SC) techniques can be directly fed with the synthetic results obtained from numerical simulations (more specifically finite element (FE) analysis). The main objective of this group of applications is to develop some structure-dependent surrogate models with limited number of initial simulations. This can be very helpful since majority of dam engineering problems have a multi-physics nature implying that they are computationally expensive to perform. More specifically, dealing with stochastic transient (e.g., seismic) simulations is computationally demanding, which limits both the application of probabilistic methods and also increases their bias due to lack of enough simulations. Thus, the SC methods seem to be appropriate alternative for the crude Monte Carlo-type simulations.

Unfortunately, this field is new in dam engineering and there is limited research devoted to probabilistic dam safety. In the following, the current applications are summarized:

- Chen et al. (2010) proposed an improved response surface meta-model (RSM) for a coupled foundation–dam system with linear behavior under sliding failure mode. Modulus of elasticity in rock and concrete is assumed to be random variables (RVs). The accuracy is evaluated as a function of number of iterations in the adopted weighted regression-based algorithm. It was reported that the proposed method not only saves the arithmetic operations, but also improves the efficiency of calculations and storage.
- Karimi et al. (2010) proposed an ANN-based procedure for system identification of gravity dams coupled with a hybrid FE-boundary element analysis to predict the dynamic characteristics of an empty dam. The gravity dam is modeled as a linear elastic system on a rigid foundation with unbounded reservoir. The conjugate gradient and the Levenberg–Marquardt algorithms were used for training the ANNs. The results of solving inverse problem were compared with numerical simulations to demonstrate the accuracy of the proposed method.
- Fan et al. (2010) combined the RSM and finite step method to fit the explicit performance function and reliability index calculation. The failure path and functionality failure mode were computed for a RCC dam. They simply reported the validity of the applied method compared to the traditional reliability index.
- Gaspar et al. (2014) developed a probabilistic thermo-chemo-mechanical model for some RCC's physical properties. A global sensitivity analysis was performed, and the heterogeneity on those parameters is also considered using a bi-dimensional random field. Using the random fields theory, a variance reduction in the output is observed. Among the considered RVs, the convection coefficient presents negligible sensitivity index.
- Cheng et al. (2015) adopted the kernel principle component analysis (KPCA) method for dam health monitoring under varying conditions, and to eliminate the effect of environmental variables. It was reported that using the data observed from ambient vibration testing of dams, the identification and warning capabilities of structural damage can be improved.
- Gu et al. (2015) developed a new method based on chaos genetic optimization algorithm to inverse the actual initial zoning deformation modulus and to determine the inversion objective function using the dam displacement measured data and FE method. No particular conclusion is presented in this paper.
- Gu et al. (2010) combined the concept of least squares support vector machine (LS-SVM) with back analysis technique to determine the mechanical properties of the RCC dams. A three-step approach is used: (1) uniform design of experiment is chosen for the initial samples, (2) train the samples using a transversely isotropic model, and (3) establish a complex nonlinear model between system output (i.e., relative values of hydraulic components of dam displacements) and the inputs (i.e., mechanical parameters).
- Su et al. (2016) performed a very similar research as well. A back analysis is performed for gravity dams including a sensitivity-based criterion for optimal parameter selection. A hybrid method of SVM, ANN, and uniform design methods was used to map the relationship between dam structural response and multi-material parameters at different regions.
- Rezaiee-Pajand and Tavakoli (2015) introduced an efficient method for crack detection in concrete gravity dams using a hybrid GA and FE methods. The GA identifies the location and magnitude of cracks in dams by minimizing the difference between the analytical responses and the measured ones.
- Xin and Chongshi (2016) performed the stability failure analysis of a gravity dam using the credibility theory. In this procedure, the stability is evaluated as a combined event of fuzziness and randomness of failure criterion, measured data, and design parameters.
- Cao et al. (2017) applied the fuzzy random event to study the stability of high arch dam abutments. The instability risk ratio models were proposed based on the credibility theory and were calculated using the Monte Carlo

simulation (MCS) and fuzzy random post-processing. A Gaussian random model was adapted to generate several realizations of spatial variability. The results showed that the proposed method is feasible with sound findings.

- Hariri-Ardebili and Pourkamali-Anaraki (2018a, b) showed the application of several SC techniques in multi-hazard analysis of gravity dams. Both the simplified models and the nonlinear damage analyses were performed. Seismic, hydrologic, and aging hazard were studied separately. They reported that SC techniques are useful when they are combined with FE analyses. These techniques can be used for prediction of future dam response, as well as the classification of safe/failed states. Several systematic algorithms are proposed in order to implement the real-world dam analysis data.
- Hariri-Ardebili and Barak (2020) compared several SC techniques in numerical seismic analysis of gravity dams. A large set of ground motion meta-features are extracted, and various feature selection techniques were applied to select the most efficient subset of ground motion unique signatures. They reported that overall ANN is the optimal method. Finally, Hariri-Ardebili and Sudret (2020) applied the polynomial chaos expansion (PCE) for material uncertainty quantification of four different case studies including an arch dam–reservoir–foundation coupled system, and gravity dams. A step-by-step framework is presented with practical recommendations. Impact of various number of initial samples as well as the engineering demand parameter is evaluated. Accuracy is compared with traditional ANN, and the advantage of PCE is shown.

1.2 Theoretical underpinning on probabilistic analyses

The objective of this paper is to focus on the material and modeling uncertainties in probabilistic FE analysis of concrete dams. This is, in deed, classified under the epistemic uncertainties according to Der-Kiureghian and Ditlevsen (2009). The classical structural reliability is based on quantitative evaluation of the limit state's (LS) exceedance probability, P_{LS} , given the uncertainties in the input parameters (Sørensen 2004). The LS function, G , divides the response domain into safe, $G > 0$, and failed, $G \leq 0$, regions:

$$G(X) = C(X) - D(X) \quad (1)$$

where $X \subset \mathbb{R}^M$ is a random vector of N basic RVs $X = X_1, X_2, \dots, X_N$. C and D refer to capacity and demand, respectively.

The time invariant probability of LS can be determined in terms of joint probability density function (PDF) of the all

contributing variables as (Melchers 1999):

$$P_{LS} = \int_{\{x : G(x) \leq 0\}} f_X(x) dx \quad (2)$$

Due to implicit nature of this integration, the direct estimation of the probability of LS is difficult. As an alternative, the simulation-based techniques are widely used (e.g., MCS) (Ditlevsen and Madsen 1996). The MCS is straightforward and based on the theory of large numbers:

$$\hat{P}_{LS} = \frac{1}{N_{sim}} \sum_{j=1}^{N_{sim}} \mathbf{I}_{LS}(x_j) = \frac{N_{exc}}{N_{sim}},$$

$$\mathbf{I}_{LS}(x) = \begin{cases} 1 & \text{if } G(x) \leq 0 \\ 0 & \text{if } G(x) > 0 \end{cases} \quad (3)$$

where N_{exc} and N_{sim} are, respectively, the number of simulations exceeding the LS, and the total number of simulations; the hat is the sign of estimation; and \mathbf{I}_{LS} is a binary classifier.

The major drawback in this technique is the large number of simulations to get the stable results which limits its application for transient analysis of dams. Variance reduction techniques such Latin hypercube sampling (LHS) (McKay et al. 1979) can be used, but they still require relatively large number of simulations.

1.3 Novelty and contribution

So far, the presented methods in Sect. 1.2 are all based on simulation-based techniques. As an alternative, a relatively small set of simulations can be used to feed into a SC model to develop its meta-model (Simpson et al. 2001). The developed meta-model can later be used to enlarge N_{sim} with negligible computational cost. Figure 1 illustrates this procedure schematically.

The key issue in this procedure is to predict the quantities of interest (QoIs) with high accuracy and limited initial simulations. There are multiple alternatives to build the meta-models (in terms of the SC technique, and the training options), and their performance (in terms of the prediction accuracy) depends on the problem at hand. A step-by-step procedure for the proposed method in Fig. 1 can be summarized as follows:

- Identify the most important physical features in the structure,
- Develop the finite element model which is as representative as possible (capturing the main response features),
- Identify all the uncertainty sources in material properties and assign most appropriate distributional model (either from literature, or from experimental tests),

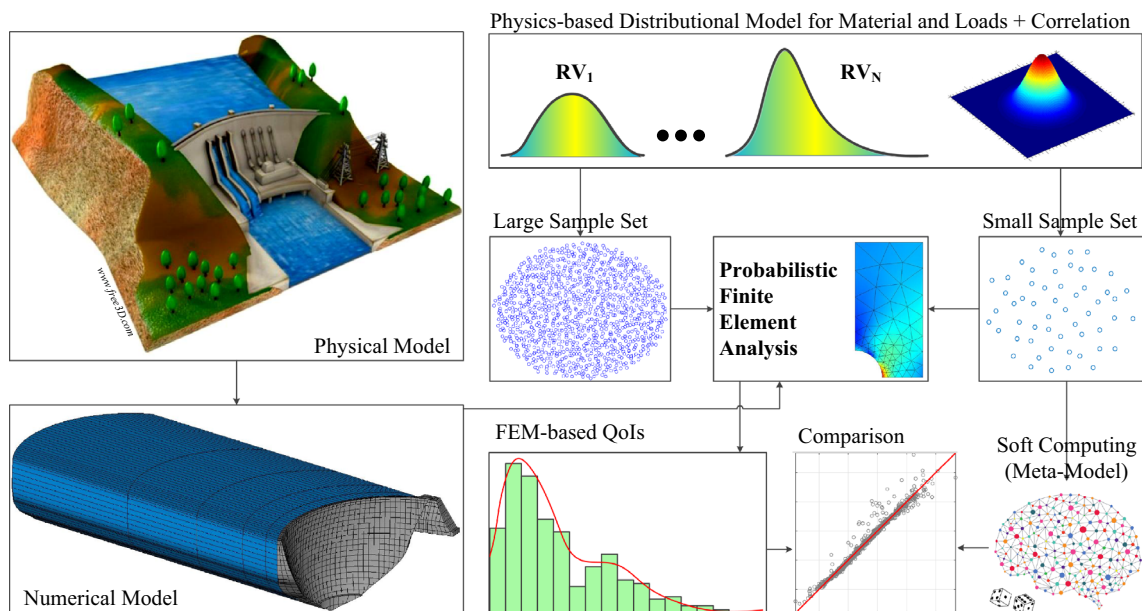


Fig. 1 Schematic explanation of the combined numerical simulations with soft computing techniques

- Draw sample a large number of realizations and perform a crude MCS or LHS. This database serves as a reference model for validation.
- Draw a (relatively) small number of samples. This database is used as a target meta-model.
- Perform probabilistic finite element simulations for both the small and large databases.
- Use the pair of input–output from small database in several soft computing algorithms and train several meta-models.
- Feed the same input of large database into the meta-models and “predict” the new outputs.
- Compare the “predicted” outputs with “exact” finite element based ones and determine the accuracy of the proposed SC method. If acceptable, recommend the SC method as an alternative method in response prediction of large-scale dam engineering problems.

The main objectives of this work are:

- Performing a comparative study of the predictive capacity of five SC algorithms in the frame of probabilistic analysis of dams.
- Account for both the material and modeling uncertainty with 14 random variables.
- Identification of the effect of the RVs in the displacement and stress responses of the system.
- Propose a geometry-based meta-model for initial sketch out of gravity dams in an optimal way.

1.4 Organization of paper

The paper is organized in the following order: a general introduction and detailed literature review is presented in Sect. 1. A brief review on five SC techniques used in this paper is presented in Sect. 2. The case study and the computational model are presented in Sect. 3.1. Various prediction tasks with increasing complexities are introduced next, Sect. 3.2, and the results are discussed in Sect. 4. Finally, the paper is wrapped up with general conclusions, recommendations, and future works, Sect. 5.

2 Soft computing techniques

The great development of data science has resulted in the existence of a huge community of scientists and engineers who constantly propose new techniques or improved versions of the existing ones to generate predictive models. Most of them can be used almost indistinctly for the stated objectives: the goal is identifying a univocal relationship between some input variables (i.e., RVs) and certain system responses (i.e., QoIs). It is therefore a conventional problem of regression.

A complete analysis of all possible alternatives is infeasible to perform. Therefore, five of the most popular SC techniques, which have been previously used in dam engineering problems (Salazar et al. 2015b), are selected for the comparison. Since the theoretical fundamentals of these techniques are well documented in other articles, only succinct

introductions are included herein, together with the considered training options.

All techniques feature some random component in the training process. To account for that, as well as for the variability in the training sample, one hundred repetitions of each model were fitted, and their accuracy analyzed for comparison.

In spite of that, the result of this work cannot be considered as the final word in terms of the appropriateness of each SC technique for this problem. Although we tried to perform a fair comparison, for which we used similar degree of parameter tuning for every algorithm, this analysis is empirical, and different results may be obtained when other kind of numerical simulation, random variables, or training options are employed.

2.1 Random forests

Random forest (RF) is an ensemble model based on the idea of the superior performance of a large dataset of simple models over a single complex one. Thus, the prediction of a RF model is computed as the average prediction of (typically) hundreds of regression trees (Breiman 2001; Hastie et al. 2011). The training process includes features to control over-fitting and capturing the existing patterns in the training dataset. It can be summarized as follows:

- For $i = 1, \dots, N_t$, where N_t is the number of trees in the RF, a random subsample of the trainings set is taken.
- A regression tree is fitted to the subsample without pruning, taking a random subsample of the input variables for each split.
- The model output is computed as the average predictions of each individual tree.

Although these models have shown to be specially appropriate for prediction tasks where the number of input variables is large compared to the training samples Díaz-Uriarte and De Andres (2006) they have also offered useful results in more conventional settings Salazar et al. (2015b). Since they are insensible to input variables with different ranges of variation, they require a simple pre-processing. In addition, variable selection is automatically performed during the model training. Therefore, the predictions are not affected by the presence of non-informative variables.

In this work, the library `randomforest` (Liaw and Wiener 2002) in the R environment (R Core Team 2018) is used (the same environment is used for all the SC techniques), with default training parameters.

2.2 Boosted regression trees

Boosted regression trees (BRTs) are categorized in ensemble methods, because their outcome is combination of multiple models (Friedman 2001). Although boosting can be used in combination with base models of different nature, the regression trees are used in this work. Therefore, BRTs share some of the advantages of RFs in terms of preprocessing effort and capability for handling the interactions and nonlinearities.

Unlike RFs, the prediction is now computed as the sum of the individual trees, since each one is fitted to the residual of the previous model. Thus, the process starts with a simple model m_1 fitted to the training data; the residual r_1 is computed, a new model m_2 is fitted to estimate r_1 , and its result is added (with opposite signed) to the prediction of m_1 , obtaining $M_2 = m_1 + m_2$. The next model is generated to estimate the residual of M_2 , denoted by r_2 . The process continues until obtaining the complete model $M_N = \sum_{i=1}^N m_i$.

Another important difference between BRT and RF is the use of very simple trees in the ensemble, typically including a few branches. The complexity of these base learners is controlled by the `interaction.depth` parameter in the `gbm` library used (Greenwell et al. 2018). This parameter is set to 2 in this paper, to consider two-way interaction among the inputs.

The result of each model is down-weighted by a shrinkage parameter ν to control over-fitting. With the same purpose, each sub-model is trained on a random subsample of the original training dataset. A maximum of 2000 trees was used in this work, and the optimal value was selected after five-fold cross-validation. The shrinkage parameter was set to 0.01, and the remaining parameters were set to their default values.

2.3 Multi-adaptive regression splines

Multi-adaptive regression splines (MARS) is an adaptive technique proposed by Friedman (1991), Friedman and Roosen (1995) whose outcome is computed as the combination of piecewise linear functions. These are defined in *reflected pairs* which take the form $\max(0, X_i - t)$ and $\max(t - X_i, 0)$, where X_i is one of the input variables, and t is a constant often called *knot*. All functions of this kind are defined as candidate functions, being t all observed values for each X_i in the training set. The function which results in greater minimization of the residual error is chosen and progressively added until the model includes some pre-defined amount of terms. This typically results in over-fitting; therefore, the final model is obtained after some terms are pruned using generalized cross-validation (GCV) criterion: a version of the residual sum of squares which considers the number of terms in the ensemble (Friedman et al. 2001).

These models are also acknowledged to be appropriate to handle nonlinear effects of inputs while being interpretable. In addition, the algorithm implicitly includes variable selection, since the candidate functions involving non-informative inputs are not selected to be part of the final model, and vice versa. The library `earth` is adopted, and all the default parameters were considered except the number of terms in the final model (parameter `nprune`), which was selected via fivefold cross-validation repeated five times.

2.4 Artificial neural networks

Artificial neural networks (ANN) have been extensively applied in different fields, including dam safety (Mata 2011) and other problems in dam engineering such as hydraulic analyses (Salazar et al. 2013). They are appropriate for nonlinear effects and complex interactions. In contrast to other algorithms considered, input normalization is necessary before model training, which requires de-normalizing the model outcome to compute accuracy. The models resulting from the training process depend on the random initialization of the weights.

A simple version of the multilayer perceptron with one hidden layer was used. Preprocess included scaling and centering of input variables. Since there is not a standard procedure to determine the number of units in the hidden layer, we tried networks including 2, 4, and 6 neurons. Overfitting is controlled by means of the regularization parameter `decay`. We used grid search to obtain the best combination of hyper-parameters (`size` and `decay`). For the latter, we tried $10e^{-3}$, $10e^{-4}$ and $10e^{-5}$. Thus, we considered nine different settings, which were evaluated by means of fourfold cross-validation repeated three times. The combination of network architecture and regularization parameter with highest accuracy was later used to fit a new model with the entire training set.

The `caret` library (Kuhn 2008) was used for parameter selection, and the `nnet` library (Venables and Ripley 2002) for fitting the ANN models.

2.5 Support vector machines

Support vector machines (SVM) also stands among the most popular SC techniques. Although they were initially developed for classification purposes, they can be used for regression as well. Multiple options can be considered, not only with regard the training parameters, but also in terms of the nature of the kernel function used (Scholkopf and Smola 2001). The fundamentals of the method can be found in different sources (Smola and Schölkopf 2004; Moguerza and Muñoz 2006). The algorithm is based on performing

linear regression on modified input variables (often termed *features*), which are obtained from a nonlinear transformation of the original inputs. A differential characteristic of this method is the use of an ϵ -insensitive error function, i.e., residuals below ϵ are neglected during the fitting process.

We followed a similar procedure as that described for ANNs to select the value of the hyper-parameters. In this case, the library used (`e1071`; Dimitriadou et al. 2008) includes a function for parameter tuning. We chose a radial basis function as kernel, which depends on one parameter, γ . Again, we used grid search to compute the best combination of the training parameters, trying 0.01 and 0.001 for γ and 100, 500, and 1000 for the `cost` parameter. Here, we used fourfold cross-validation repeated five times. The remaining parameters were kept at their default values.

3 Numerical model

3.1 Case study description

Koyna Dam is used as pilot case study for numerical simulations (Bhattacharjee and Leger 1994). The dam includes 56 blocks in which 7 are overflow blocks. The height of the central non-overflow monoliths is 103 m, the thickness at the base and at the crest is 70.2 m and 14.8 m, respectively. Figure 2a shows the 3D slice view of the geometry model for the dam (including the base dimensions) prepared in ANSYS (2007). Analysis of concrete gravity dams can be performed in two or three dimensions. Usually due to large dimension of dams in cross-stream direction, two-dimensional (2D) model is adopted with plane strain assumption (Fenves and Chopra 1984).

In the current paper, 2D plane strain model is used to reduce the computational cost. The finite element mesh for the dam–reservoir–foundation system is shown in Fig. 2b. Four-nodes linear elastic elements are used in dam and foundation domain, while the water is modeled using pressure-based elements. Solid and fluid elements are in interaction at the dam–water interface, as well as the foundation–water interface.

The most optimal mesh size in structural engineering applications is those with aspect ratio close to one. This may imply a very fine mesh all over the system and thus unnecessarily increases the run time. On the other hand, one may take the advantage of a “smart mesh” in which the mesh density is only refined at the critical locations (e.g., high stress zones, and potential failure zones), while it remains relatively coarse for other zones. Such a technique is used in Fig. 2b where the mesh density is high for the neck area and dam–foundation interface [two potential failure modes for gravity dams (Ghanaat 2004)]. This mesh

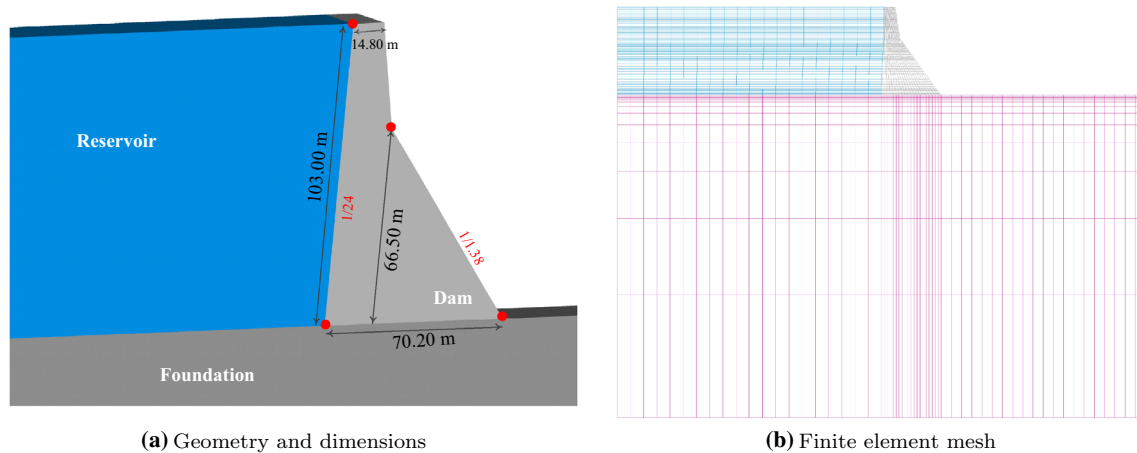


Fig. 2 Description of case study, Koyna Dam, including the dam body, foundation, and reservoir

ensures the trade-off between moderate computational cost (a large number of simulations were required) and refinement in the areas of interest. Such a non-uniform mesh for dam body is already reported by other researchers as well (Bhattacharjee and Leger 1993; Guanglun et al. 2000; Zhang et al. 2013). On the other hand, the direct responses (i.e., displacement and stress) are not required in the foundation and reservoir domains. We only need to consider their “impact” and “interaction” with the dam body. Therefore, the elements in foundation region (and far away from dam) have large ratio distortion grid. Similarly, for the reservoir, we just transfer the water pressure to dam upstream face, and thus, a non-uniform mesh is acceptable. Non-uniform meshing in reservoir and foundation domains is also reported by several researchers (Arabshahi and Lotfi 2008; Ghaemian et al. 2014; Segura et al. 2019; Ganji et al. 2019).

Fluid–structure interaction (FSI) is an important issue in dynamic analysis of concrete dams. In the adopted technique, the unknown variables are the displacements (in the solid domain) and the pressures (in the fluid domain). The water is linearly compressible, and its viscosity is neglected. The coupled equation of motion is solved using second-order differential equations in both domains, Fig. 3. A direct time integration scheme is used to find the displacement and hydrodynamic pressure at the end of the time increment $i + 1$ given those quantities at time i . The α -method is implemented for discretization (implicit–implicit technique). Finally, the staggered solution is adopted to solve the coupled system (Hariri-Ardebili and Mirzabozorg 2013).

Table 1 represents the material properties for mass concrete and the foundation rock. It includes the mean, coefficient of variation (COV), and the truncation bound. Normal distribution is assumed in all cases (any other model might be used based on the detailed field investigations).

Applied loads on the system are: self-weight, hydrostatic pressure (normal water level (NWL) is at 91.75 m for the pilot model), bottom sediment (wave reflection coefficient for the reservoir bottom materials is assumed to be 0.75), and earthquake-induced loads (both inertia and hydrodynamic).

The seismic excitation is applied at the foundation base, whereas the recorded earthquake signal is on the free-field. Therefore, a de-convolution process is required to determine the motion at the rigid base boundary (Saouma et al. 2010). Loma Prieta earthquake of 1989 at San Francisco 1295 Shafter station with 6.9 magnitude (PEER 2014) is used for all the transient simulations. The ground motion record-to-record variability is not considered in this paper, as it is a topic of separate study under fragility analysis (Hariri-Ardebili and Saouma 2016).

3.2 Prediction tasks

First, the conventional MCS is performed using the advanced LHS technique where $N_{sim} = 10,000$. Six material properties in Table 1 are assumed to be RVs. Although it is possible to adopt any distributional model for the RVs, the truncated normal distribution is used in all cases for simplicity.

Maximum crest displacement and maximum vertical stress at heel are chosen as two QoIs. Note that σ_{yy} is the governing stress component in 2D analysis of gravity dams which leads to crack opening at the dam heel. On the other hand, the Δ_{max} represents the overall stability. Figure 4a, b illustrates dependency of the mean and 95% confidence intervals (CI) to the sample size in displacement-based and stress-based results, respectively. It seems that stability of the stress-based results is faster.

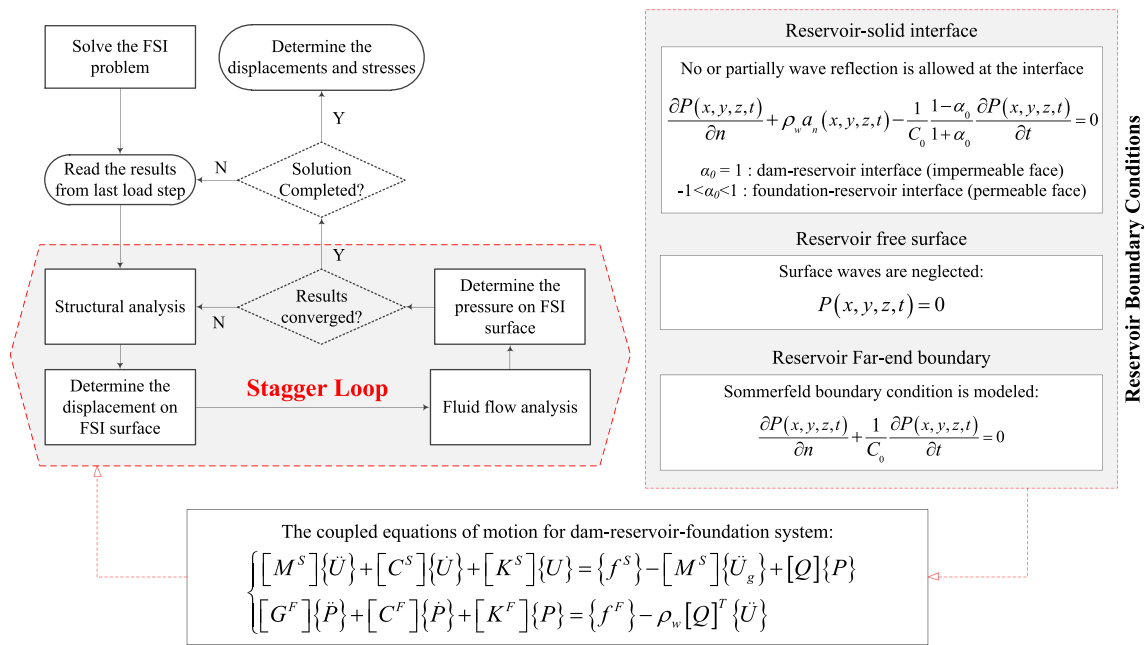
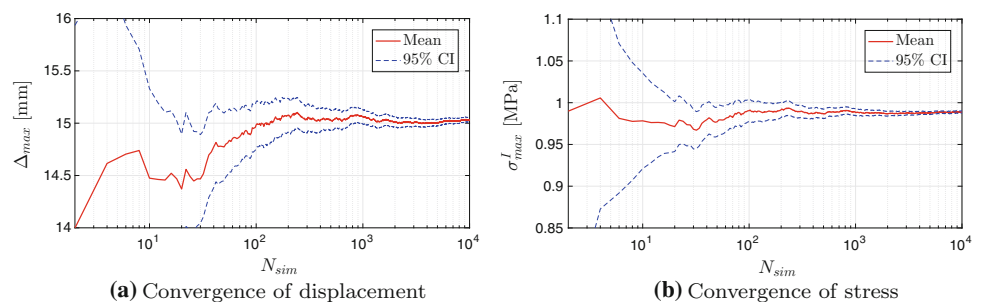


Fig. 3 Stagger solution for FSI problem (Hariri-Ardebili and Mirzabozorg 2013)

Table 1 Material properties for the concrete and foundation

ID	Quantity	Symbol	Unit	Mean	COV	Truncation
M_1	Concrete modulus of elasticity	E_c	GPa	24	0.10	[19.1 28.8]
-	Concrete Poisson's ratio	ν_c	-	0.2	-	-
M_2	Concrete mass density	ρ_c	kg/m ³	2470	0.1	[2160 2790]
M_3	Concrete hysteretic damping	η_c	-	0.06	0.33	[0.02 0.10]
M_4	Foundation modulus of elasticity	E_f	GPa	21.5	0.10	[17.7 25.4]
-	Foundation Poisson's ratio	ν_f	-	0.33	-	-
M_5	Foundation mass density	ρ_f	kg/m ³	2680	0.1	[2320 3040]
M_6	Foundation hysteretic damping	η_f	-	0.05	0.40	[0.02 0.08]

Fig. 4 Uncertainty quantification based on LHS with 10,000 simulations



The objective is to compare different SC techniques in prediction of structural response, as well as their relative performance. The prediction accuracy is assessed by means of

the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the average relative variance (ARV) (Weigend et al. 1992):

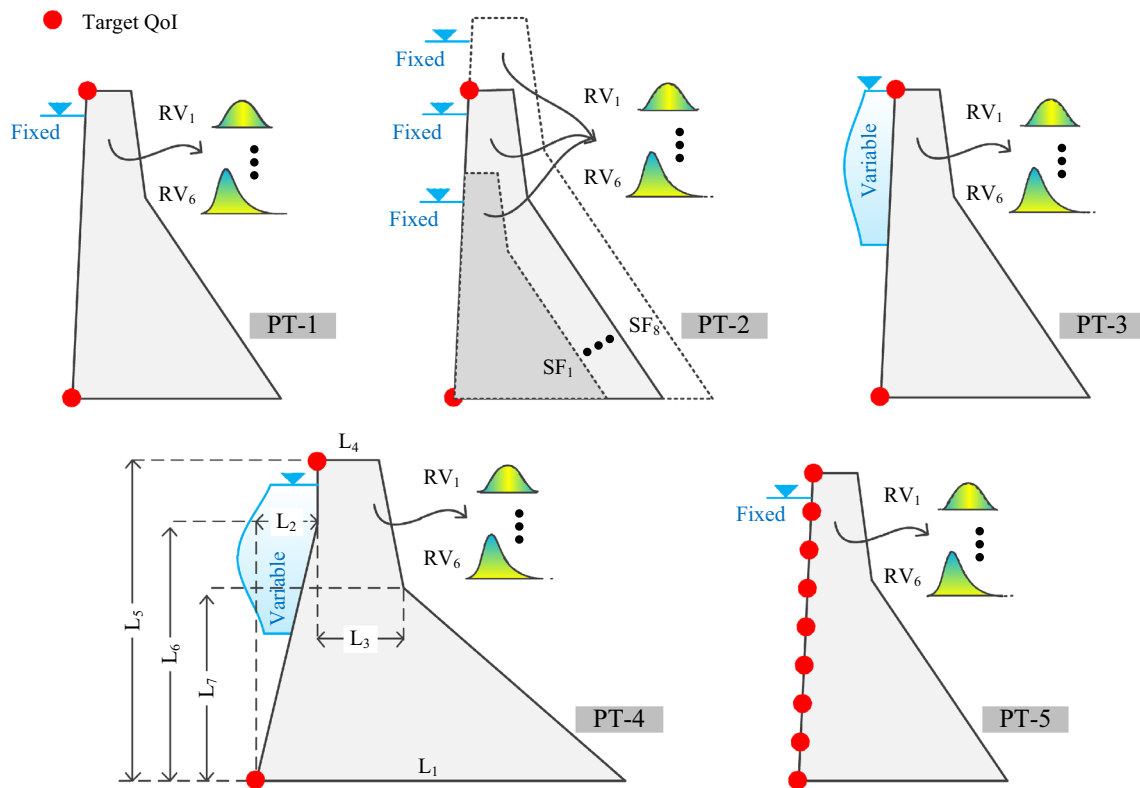


Fig. 5 Schematic presentation of different prediction tasks

$$\begin{aligned}
 MAE &= \frac{\sum_{i=1}^N |QoI_i^{FE} - QoI_i^{SC}|}{N} \\
 MAPE &= \frac{100}{N} \sum_{i=1}^N \left| \frac{QoI_i^{FE} - QoI_i^{SC}}{QoI_i^{FE}} \right| \\
 ARV &= \frac{\sum_{i=1}^N (QoI_i^{FE} - QoI_i^{SC})^2}{\sum_{i=1}^N (QoI_i^{FE} - \bar{QoI}^{FE})^2}
 \end{aligned} \tag{4}$$

where N is the size of the test set, QoI_i^{FE} are the observed outputs from finite element, \bar{QoI}^{FE} is the mean value, and QoI_i^{SC} are the predicted values by SC.

Prediction tasks (PT) can be summarized as follows with increasing complexities, Fig. 5:

- PT-1: The pilot dam with fixed reservoir at NWL, and six material properties as RVs. 10,000 realizations were used for algorithm comparison.
- PT-2: Scaled pilot dam with fixed reservoir at NWL, and six material RVs. Eight different scale factors (SF) are used for scaling the dam size up and down.
- PT-3: The pilot dam with variable water level (simulating the flooding conditions), and also six material properties as RVs.

PT-4: A class of different dam shapes (with seven RVs controlling the geometry), variable water level, and six random material properties.

PT-5: The pilot dam is analyzed to assess the sensitivity of the predictive models on multiple points along the height of the dam (and not a single-targeted task).

4 Results and discussion

4.1 PT-1: pilot model

The first prediction task (i.e., PT-1) is the full analysis of the Koyna Dam in Fig. 5. Assumptions and procedure are summarized as follows:

- The geometry of the dam and the water level is fixed.
- Initially, 10,000 simulations are performed in which only the six material properties are RVs.
- Five different SC methods are used for prediction.
- Two QoIs are studied, i.e., crest displacement and heel stress.
- Three metrics are used to evaluate the accuracy of prediction, Eq. 4.
- Different sample sizes are used for training, namely 50, 100, 200, 300, 500, 1000, and 2000.

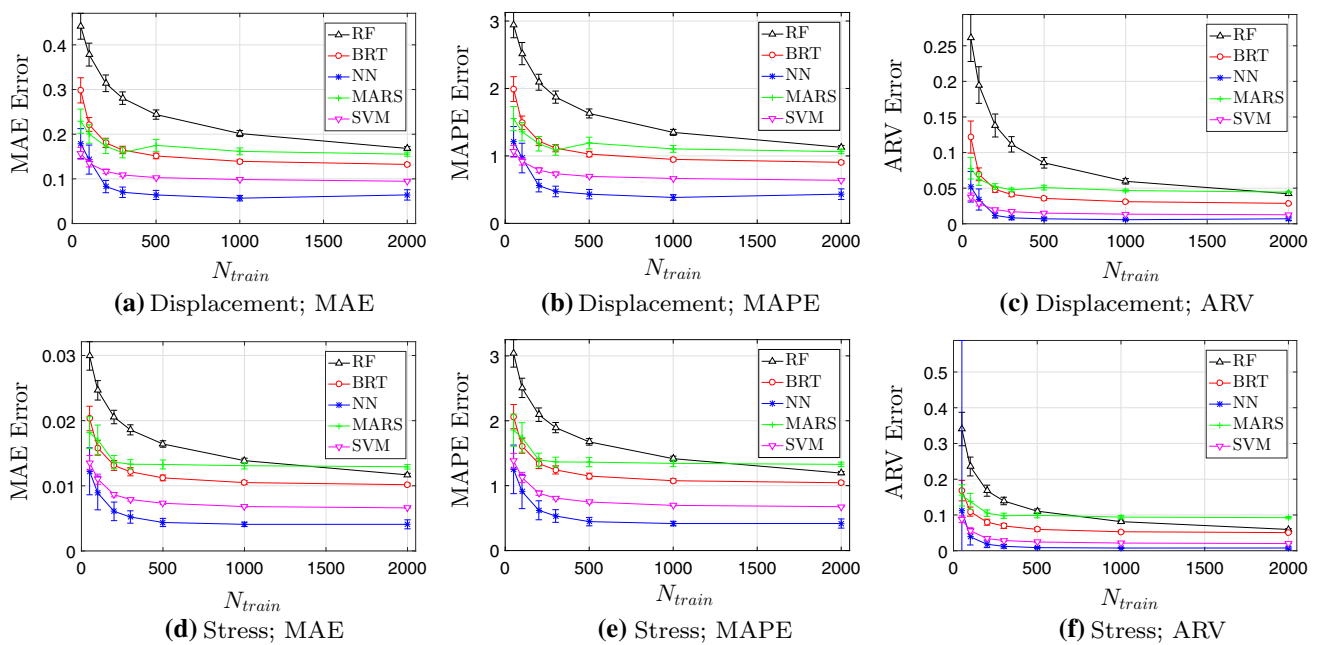


Fig. 6 Prediction accuracy of the algorithms in PT-1 considered for the validation set (8000 cases) and 100 iterations

- The rest of the simulations (8000) are taken as external validation to control the accuracy.
- Evaluation is based on the validation set.
- For each training set, 100 iterations are repeated to account for the randomness in realizations of the RVs, as well as in the training process of the algorithms. Results are presented as mean and mean \pm standard deviation (STD).
- To prevent poor performance of the SC models when extrapolating, the maximum and minimum values of each RV are always included in the training set.

Figure 6 presents the variation of the prediction accuracy for different training sets, SC algorithms, QoIs and metrics. Major observations are:

- In general (as expected), increasing the number of training samples reduces the mean and STD of the error for all metrics and SC techniques.
- ANN provides the highest accuracy for both outputs, followed by SVM. The RF is the worst case, specially for the smaller N_{train} .
- The ANN, SVM, and BRT reach the stability after about $N_{train} = 500$. However, similar results are obtained with $N_{train} = 300$, which can be useful for practical/engineering purposes.
- Results are similar for both displacement and stress in relative terms (MAPE and ARV). The differences of displacement and stress in MAE are due to the range of variation of both variables.

The results for ANN were further analyzed by computing the partial dependence plots for the input RVs and the both QoIs. They present the average effect of each predictor on the response. The plots show the average influence in solid blue and the standard deviation intervals in dashed-red, Fig. 7.

From the observation of the results for displacements, it can be concluded that concrete properties are more relevant than those of the foundation: The range of the vertical axis for M_{1-3} is higher than those for M_{4-6} . In addition, there is higher variance in the results for the foundation parameters in relative terms. As expected, higher E_c increases the stiffness and therefore reduces the displacement. This input has the greater relevance, with the average response ranging from 13.5 mm for the lower bond of E_c to 17.5 mm for the upper one. Concrete density has opposite effect, though less important. Finally, the model captured the expected effect of damping, with high values resulting in lower displacements. The influence of all RVs in displacement is barely linear except for E_f , which nonetheless has very low relevance.

For the heel stress, the major influence is observed for M_2 , with similar shape as for displacements. This implies that the dynamic effect of higher concrete density is not fully compensated by the corresponding increase in compression stress. The effect of M_1 has a local minimum at about $E_c = 25$ GPa, though the average influence is around 25% of that observed for M_2 . Damping also showed greater effect in this case than E_c . Both M_4 and M_5 showed low influence.

Last but not least, the sensitivity of the six RVs is assessed with respect to each other. Only the ANN model is used in conjunction with the Olden index (Olden and Jackson

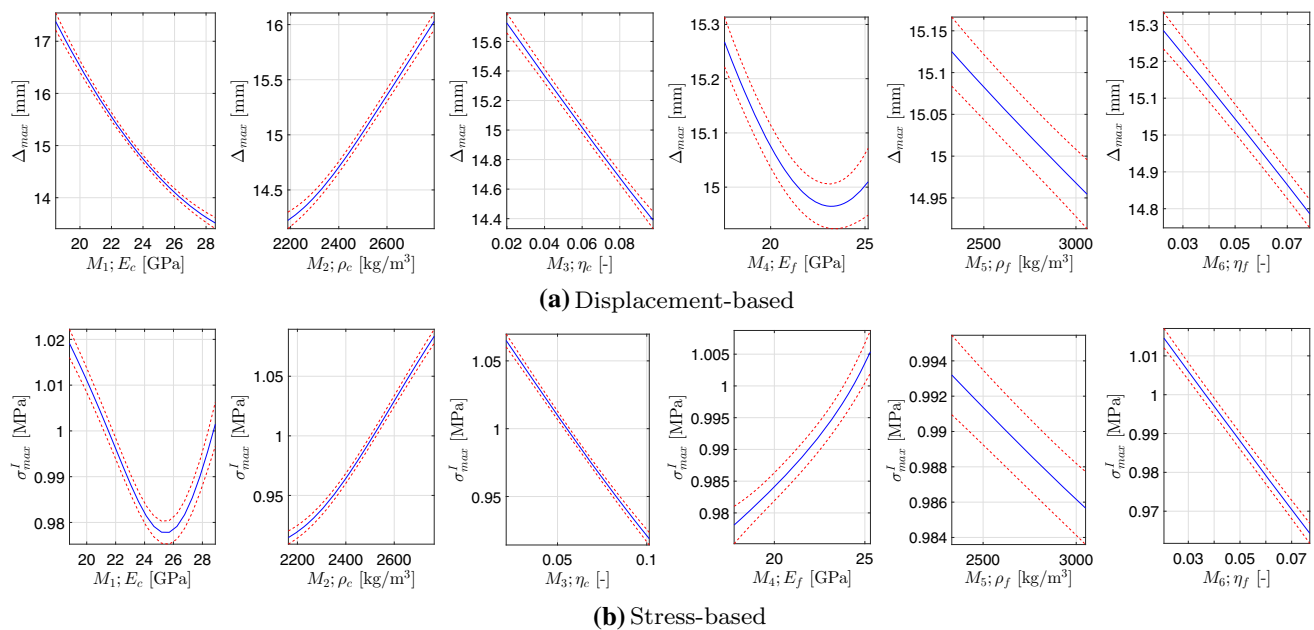


Fig. 7 Partial dependence plots for six RVs in PT-1

2002). This is a measure of the effect of each input in the system response based on the value of the connection weights between neurons that accounts for their signs. Therefore, it allows distinguishing between positive and negative correlation between inputs and response. Since previous applications of this index showed some dependency on the random component of the training algorithm (Salazar and Crookston 2019), a ANN model with 300 random samples is trained for 100 times. This reduces the uncertainty and dependency of the Olden index. Results are reported in Fig. 8 for both QoIs.

The most important parameters to control the displacement are concrete modulus of elasticity and mass density. On the other hand, the stress-based importance analysis reveals interesting findings: (1) three concrete properties are important than three foundation ones, (2) there is a very similar ratio among three concrete and three foundation properties, and (3) rock mass density is least effect of the results. Finally, the concrete modulus of elasticity has completely different behavior on displacement and stress. This conflicting criteria might be important in displacement-based and stress-based design procedure.

4.2 PT-2: dam class effect

The second task is to investigate the impact of dam class on the accuracy of the prediction task. For this purpose, the same shape (geometry) of the pilot dam is preserved with following scale factors (SF): 0.25, 0.50, 0.75, 1.00 (pilot dam), 1.25, 1.50, 1.75, and 2.00. All the assumptions are similar to PT-1; however, only 1000 finite element simulations are

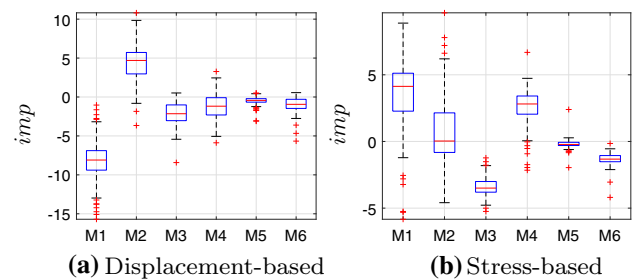


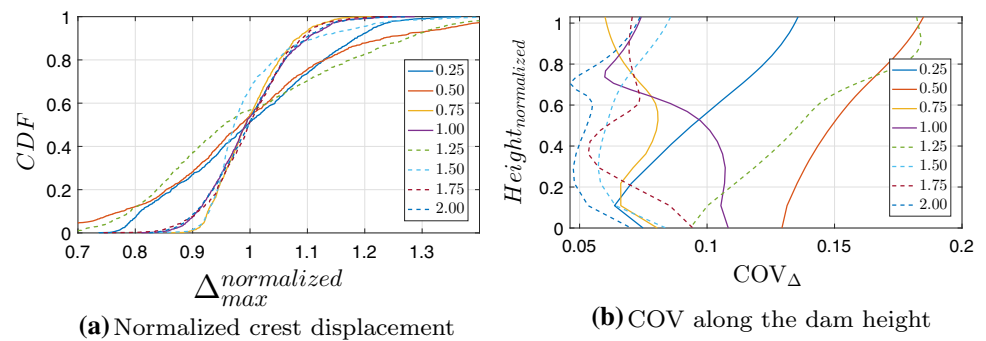
Fig. 8 Relative influence of the predictors for 100 training repetitions in PT-1

conducted for each of those 8 dam classes (totally 8000 simulations). Since the number of initial simulations is reduced, the number of training size is also limited to no more than 500. Moreover, the uncertainty in the sample selection for training is neglected, as the results of the previous task showed that it had no influence in the algorithm comparison.

The concept of dam class is very important, because not all the QoIs are linearly scaled up/down by scaling the dam geometry (Hariri-Ardebili and Saouma 2018). Detailed response variation of different dam classes is shown in Fig. 19 along the upstream face. As clear, the dispersion and pattern of both displacement and stresses change with dam size. More interestingly, it is the location of dominant stress which moves from the heel to the upper location by increasing the dam size. For the dam with SF = 0.25, the critical zone is heel, while for the one with SF = 2.00 it is located at the height of about 120 m from the heel.

Figure 9 explores the dam class dependency of the displacement responses more in detail. Figure 9a shows the

Fig. 9 Summary uncertainty quantification of different dam classes



cumulative distribution function (CDF) of the normalized (with respect to mean) displacement at the crest. The mean crest displacement increased from 0.7, 6.1, 11.4, 15.0, 26.3, 33.1, 43.0, to 41.6 mm for SF of 0.25 to 2.00, respectively. There is a considerable variation among the CDF curves (both in shape and dispersion). Results are further propagated along the upstream face through the coefficient of variation (COV), Fig. 9b. As seen, dispersion is quite large not only among the dams but also along the dam face.

Figure 10 shows the comparison of the algorithms in terms of prediction accuracy (based on MAPE) and dam class. The following observations can be drawn:

- Again, increasing the size of the training set reduces the MAPE error. However, there is a little decrease in accuracy for training sets above 100–200 samples. This can be useful for practical purposes.
- For all the dam classes, the ANN model offers best prediction.
- For majority of the dam classes, the SVM is also competing with ANN model.
- The worst performance is recorded for RF, nearly for all the dam classes.
- In general, the stress-based prediction models lead to lower MAPE than the displacement-based models.
- There is no particular correlation among the accuracy of prediction and dam class. In fact, results in terms of MAPE are similar for different SF and training set size. The greater differences observed correspond to models trained with 25–50 samples, which are clearly insufficient for practical applications.

Similar observations can be reported for the other metrics. In order to present a big picture of “normalized” performance of all SC models, the training size of 300 is selected (which is minimum stable and practical size for FE simulations). For each of two QoIs, and three metrics, the normalized accuracy with respect to largest errors in that comparison group is presented. Figure 11 illustrates the matrix presentation of the normalized performance. The six matrices are based on three performance metric and two QoIs. The 8×5 grid inside

each matrix refers to eight SFs for dam classes and five SC algorithms. For each subplot, the maximum error is also provided under the caption. One may observe that:

- Interestingly, the accuracy of the ANN model is barely insensitive to the dam class.
- The SVM also has minimum reaction to the dam class.
- Three other algorithms (and specially RF and MARS) are highly affected by dam class, with poorer performance on average for high SF values.
- The maximum displacement-based errors are larger than stress-based ones.

4.3 PT-3: water-level effect

So far, only the material properties are assumed to be RVs, and the dam class is studied with parametric model. Reservoir water is another important factor that affects the seismic response of the dams. Both hydrostatic and hydrodynamic pressures change the magnitude and pattern of forces induced during the seismic excitation. Therefore, the prediction task in PT-1 is repeated again accounting for the water variations. A uniform distribution is used to the pool elevation with lower bound of $0.5H_d$ (minimum water level in winter) and upper bound of H_d (where H_d is dam height). Initial FE simulations were limited to 1000 runs.

The MAPE-based prediction accuracy for five SC algorithms is shown in Fig. 12 (two other metrics are skipped). One may draw the following observations:

- There is a clear decrease in prediction accuracy compared to PT-1, where water level was taken as constant.
- At the same time, the ANN model outperformed the remaining algorithms.
- Performance of the ANN model becomes better once at least 200 training samples are used.
- Again, the RF is the less accurate SC technique.

In view of these results, the ANN model was again selected to investigate the sensitivity of the QoIs to the RVs, also evaluated using Olden index, Fig. 13a, b. As before, these

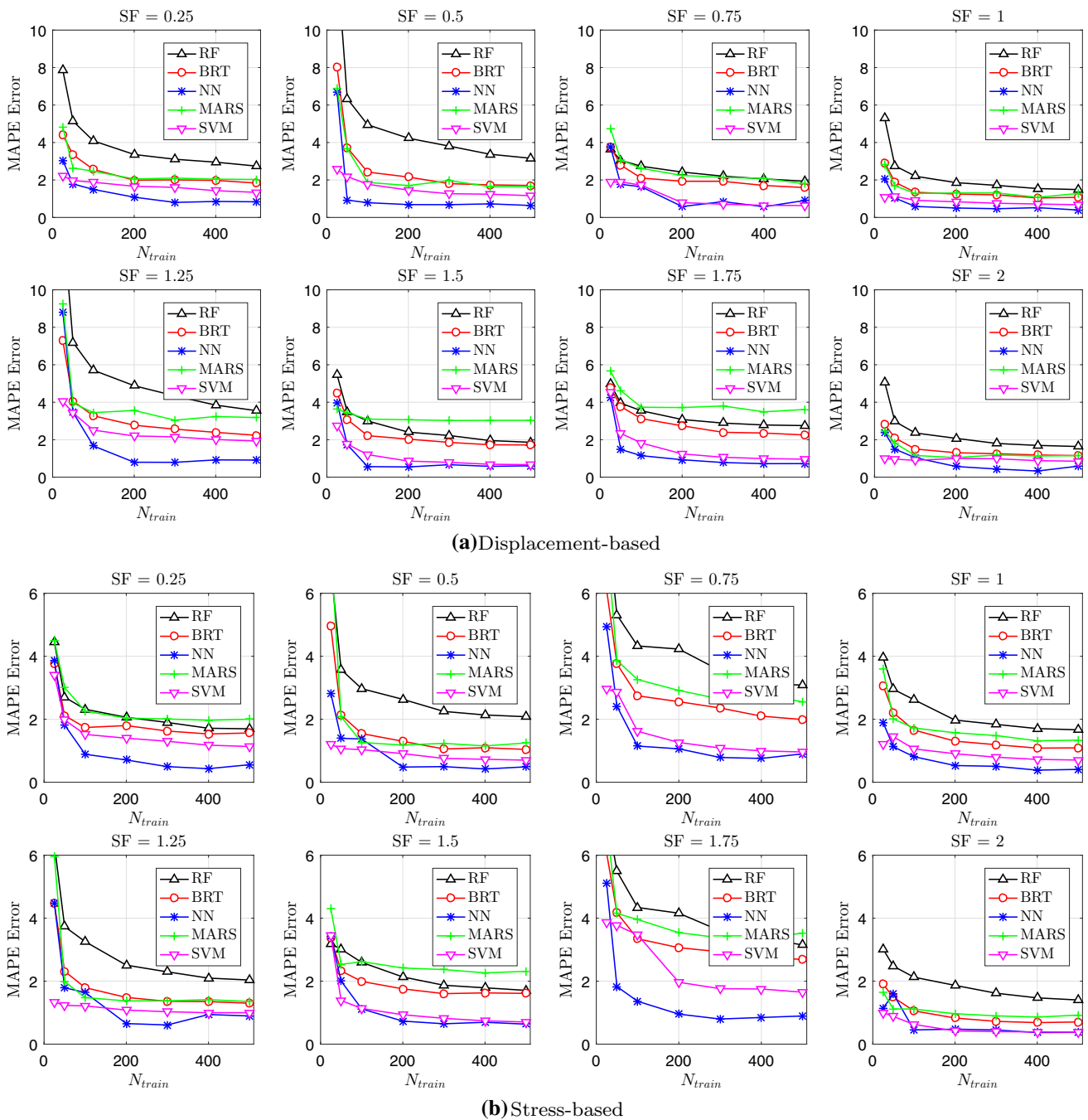


Fig. 10 Prediction accuracy of different dam classes in PT-2 only based on MAPE metric

results correspond to 100 repetitions of model fitting with 300 samples in the training set. Compared to PT-1 and PT-2, it reveals that the water-level variations have the greatest influence on the dam response. Since the index is positive, higher water level is associated with larger crest displacement and higher heel stress (which is consistent with the physics of the problem). In addition, concrete modulus of elasticity and mass density (i.e., RV M_1 and M_2) are also effective

(similar to the previous task). The mean influence of modulus of elasticity on stress response is nearly zero, similar to results in PT-1 (Fig. 7b).

Figure 13c, d shows the partial dependence of both QoIs on water level, which is the most influential variable. Displacement response seems to be more nonlinear than the stress. The partial dependency is skipped for other RVs, since they are similar to those reported in PT-1 and PT-2.

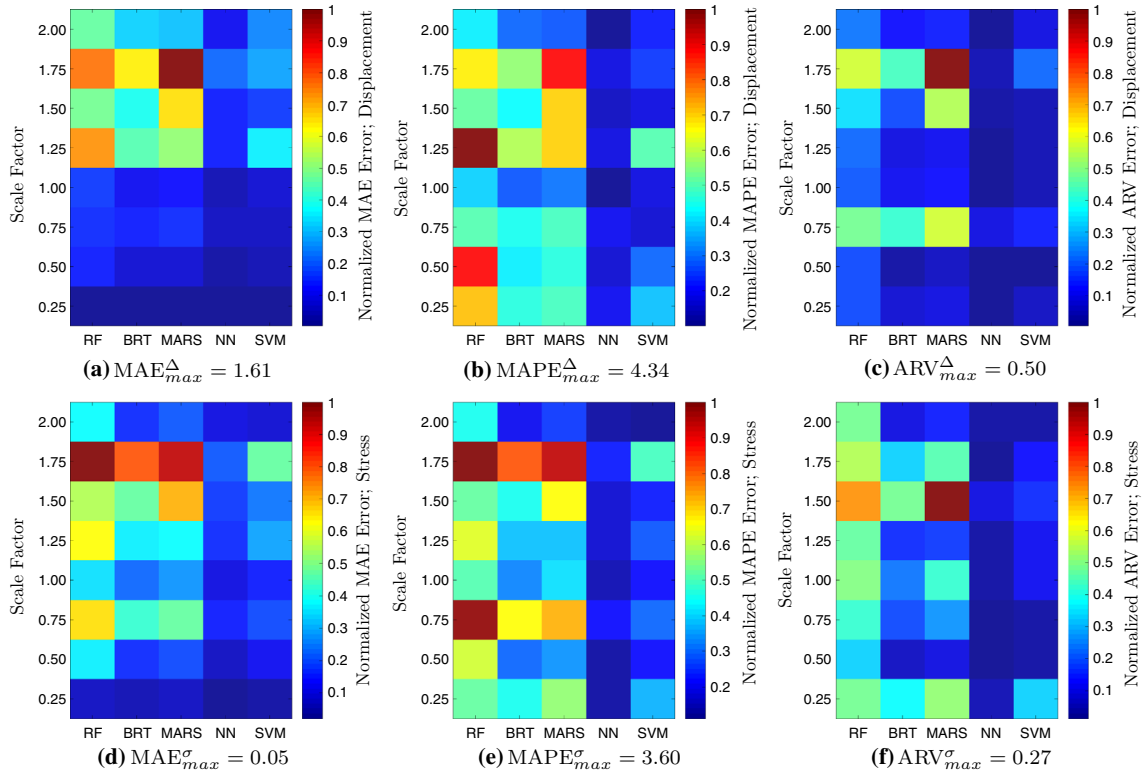


Fig. 11 Normalized performance of different SC techniques with two QoIs, and three metrics; training size is 300 in all cases; the maximum error is shown for each group

Fig. 12 Prediction accuracy for the algorithms in PT-3 only based on MAPE metric

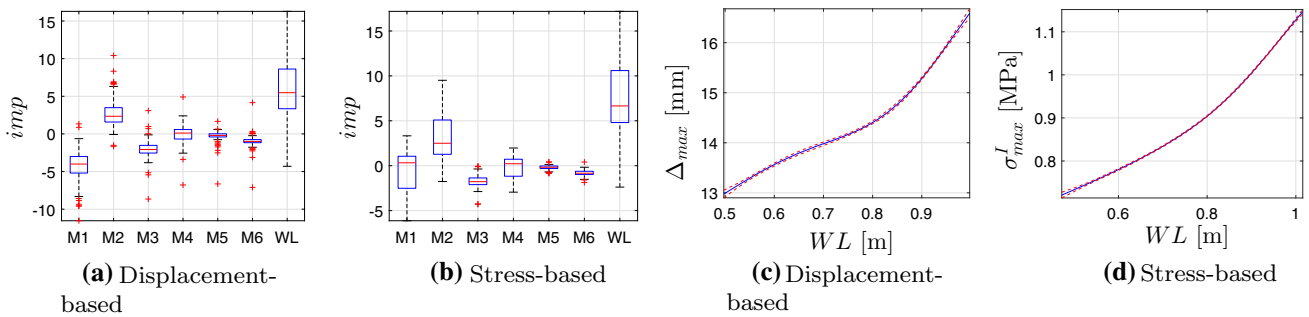
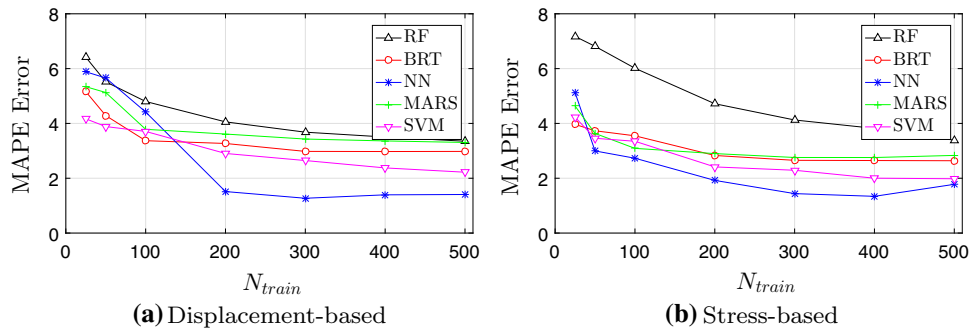


Fig. 13 Relative influence of the predictors for 100 training repetitions; and partial dependence on WL

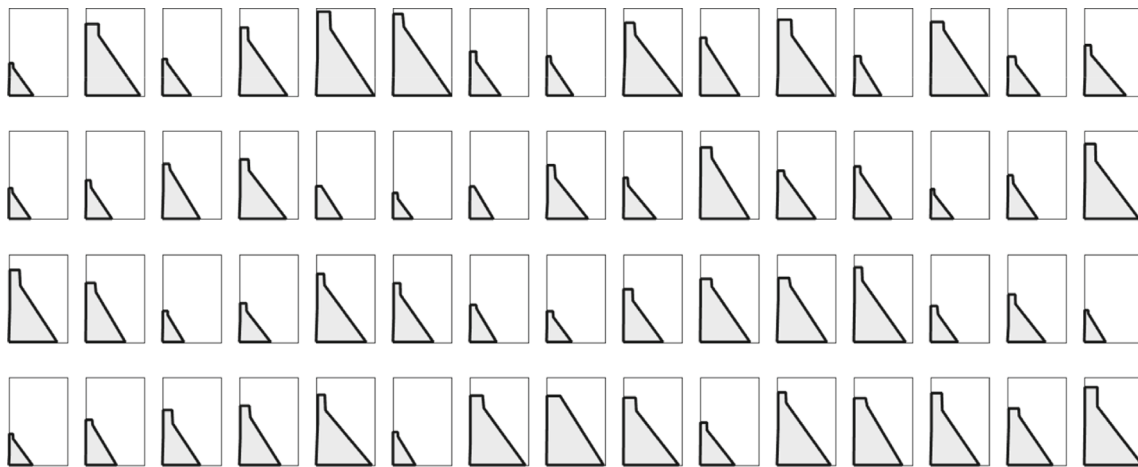
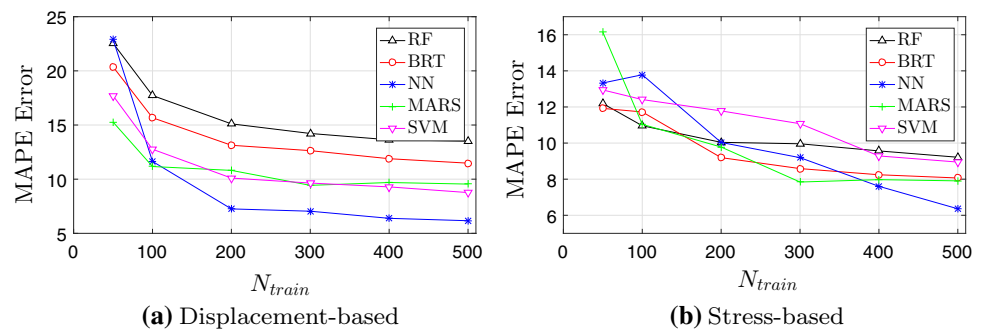


Fig. 14 The first 60 random gravity dam shapes generated based on a MATLAB code; the box size is 150×225 m in all cases

Fig. 15 Prediction accuracy for the algorithms in PT-4 (only based on MAPE)



4.4 PT-4: material and modeling effect

Finally, the shape of the gravity dam is also assumed to be unknown (i.e., variable) in addition to the material randomness and water level. This is, in fact, a very important topic for users dealing with a diverse portfolio of dams. Developing a general approach which can be used for different dam shapes is of interest for many big dam owners (e.g., USBR). There are several evidence that dam design companies try to correlate the response of dams in different size and shape.

To account for different dam shapes (note that this is different from SF used in PT-2), a generic form for the gravity dam is proposed with 7 variable lengths as shown in Fig. 5 (L_1 to L_7). The following relations are used in order to generate a random dam shape, Fig. 14:

- $L_1 = \text{rand}(50, 150)$ m.
- $L_2 = L_1 \times \alpha_1$; $\alpha_1 = \text{rand}(0.00, 0.03)$
- $L_3 = L_4 \times \alpha_3$; $\alpha_3 = \text{rand}(1.00, 1.15)$
- $L_4 = L_1 \times \alpha_2$; $\alpha_2 = \text{rand}(0.12, 0.24)$
- $L_5 = L_1 \times \alpha_4$; $\alpha_4 = \text{rand}(1.20, 1.50)$
- $L_6 = L_5 \times \alpha_5$; $\alpha_5 = \text{rand}(0.30, 0.70)$
- $L_7 = L_5 \times \alpha_6$; $\alpha_6 = \text{rand}(0.75, 0.90)$
- $WL = L_5 \times \alpha_7$; $\alpha_7 = \text{rand}(0.50, 1.00)$

Figure 15 shows the comparison of the algorithms in terms of prediction accuracy. A similar pattern is observed, with the ANN model featuring highest accuracy. With so much uncertainty, prediction accuracy decreases compared to previous tasks. The results suggest that the training set size should be enlarged for stable results (in particular for heel stress). However, the ANN model was capable of predicting both outputs with around 6% error. The displacement-based error variation for different SC algorithms is well-separated compared to stress-based errors.

Next, the results of the sensitivity analysis of the $6 + 1 + 7 = 14$ RVs and the ANN model are shown in Fig. 16 for both displacement and stress QoIs. The results are based on the model trained with 500 samples and 100 iterations. The major observations are:

- The impact of material randomness is overshadowed with modeling/geometry uncertainty.
- The most influencing RV is L_5 (dam height) for both QoIs, followed by L_7 (neck height) for displacements and by L_1 (base length) for heel stress.
- The water level, WL, has positive effect on both QoIs, with higher impact on stress. However, it has lower relevance than the mentioned geometrical parameters.

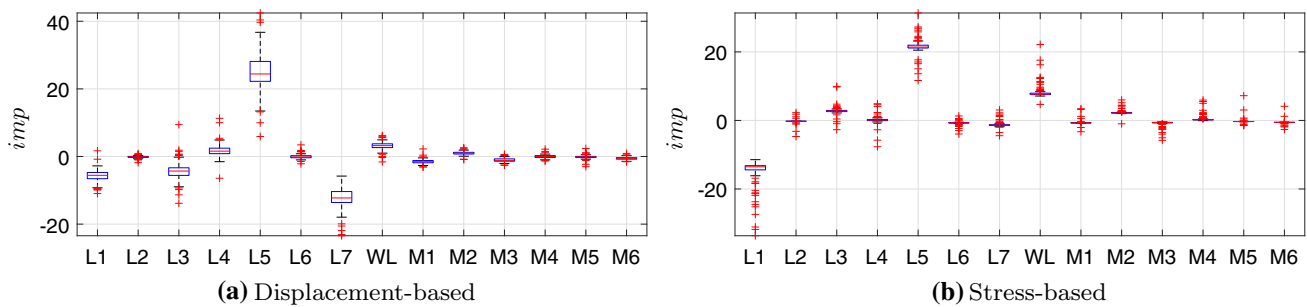


Fig. 16 Relative influence of the predictors for 100 training repetitions; and partial dependence on WL and shape

- The impact of each geometry is next discussed which is consistent with the physics of the problem in all cases:
 - L_1 (base length) has negative impact on both QoIs (mainly dominates stress).
 - L_2 and L_6 (location of upstream slope) have practically no impact on the response quantities.
 - L_3 (width of the lower neck) has negative impact on displacement as increasing it increases the stiffness and thus, reduces the displacement. On the other hand, it has a positive impact of stress as it increases the weight.
 - L_4 (width of the upper neck) has practically no impact on the base stress, and a slight positive impact on the crest displacement.
 - L_5 (dam height) is the most effective parameter and has a positive correlation with both QoIs.
 - L_7 (neck height) has a negative impact on displacement, as increasing it makes the neck shorter (and stiffer) and reduces the displacement. It does not have practically impact on base stress.
- The relative importance of 6 material RVs is qualitatively similar to those obtained in the pilot model, though lower as compared to the geometrical parameters.

The findings in Fig. 16 are further supported by detailed dependence plots in Fig. 17. General observations are similar to those explained for the sensitivity plots. However, the dependence plots show the linear or nonlinear trend of the QoIs with respect to input parameters. Of more importance is the sigmoid shape of the L_5 (so far the most important parameter) with respect to QoIs. The narrow upper/lower bound of water level is also confirms the direct and strong correlation of this external load on the dam response.

4.5 PT-5: multi-targeted problem

So far, all the prediction tasks focused on a scalar value for displacement or stress. However, in most of the cases, it is required to predict the response of multiple points in the finite

element model. One way (and probably the easiest way) is to process each point of interest (PoI) individually and independent from other ones. An alternative (and probably the most elegant) way to predict multiple points, is to account for potential correlation among the responses (this will be postponed to another paper with more details). This problem is referred to “multi-targeted prediction” (Spyromitros-Xiouf et al. 2016).

In this section, the pilot dam in PT-1 is studied again; however, the QoIs are predicted in multiple PoIs along the upstream face. Again, the ANN algorithm is selected as the best technique. For each PoI, 50 iterations of the training process are performed. The prediction accuracy is reported in 25 nodal points, and 24 elements along the upstream face, Fig. 18. Six training sizes from 50 to 500 are used for each of three metrics. As before, increasing the training sample results in improved accuracy, with low effect for more than 200 cases. The results are similar for all PoIs in relative terms (MAPE and ARV). The difference for MAE is due to the value of the corresponding QoI for each PoI. In this regard, it should be noted that all results are ordered from the top (blue) to the bottom (red) of the upstream face. Thus, MAE follows the same trend as the corresponding QoI along the upstream face: from top to bottom, it decreases for displacements and increases for stress.

4.6 Overall considerations

The obvious advantage of using SC techniques over numerical models is the reduction in computational cost. However, this is, indeed, an advantage in practical applications only in case the predictions of the QoIs are accurate enough. Establishing an error threshold to determine whether a model is useful in practice is not straightforward, and generally depends on the problem at hand. Some relevant aspects can nonetheless be mentioned:

- There will always be a discrepancy between the numerical and the SC model that will need to be properly quan-

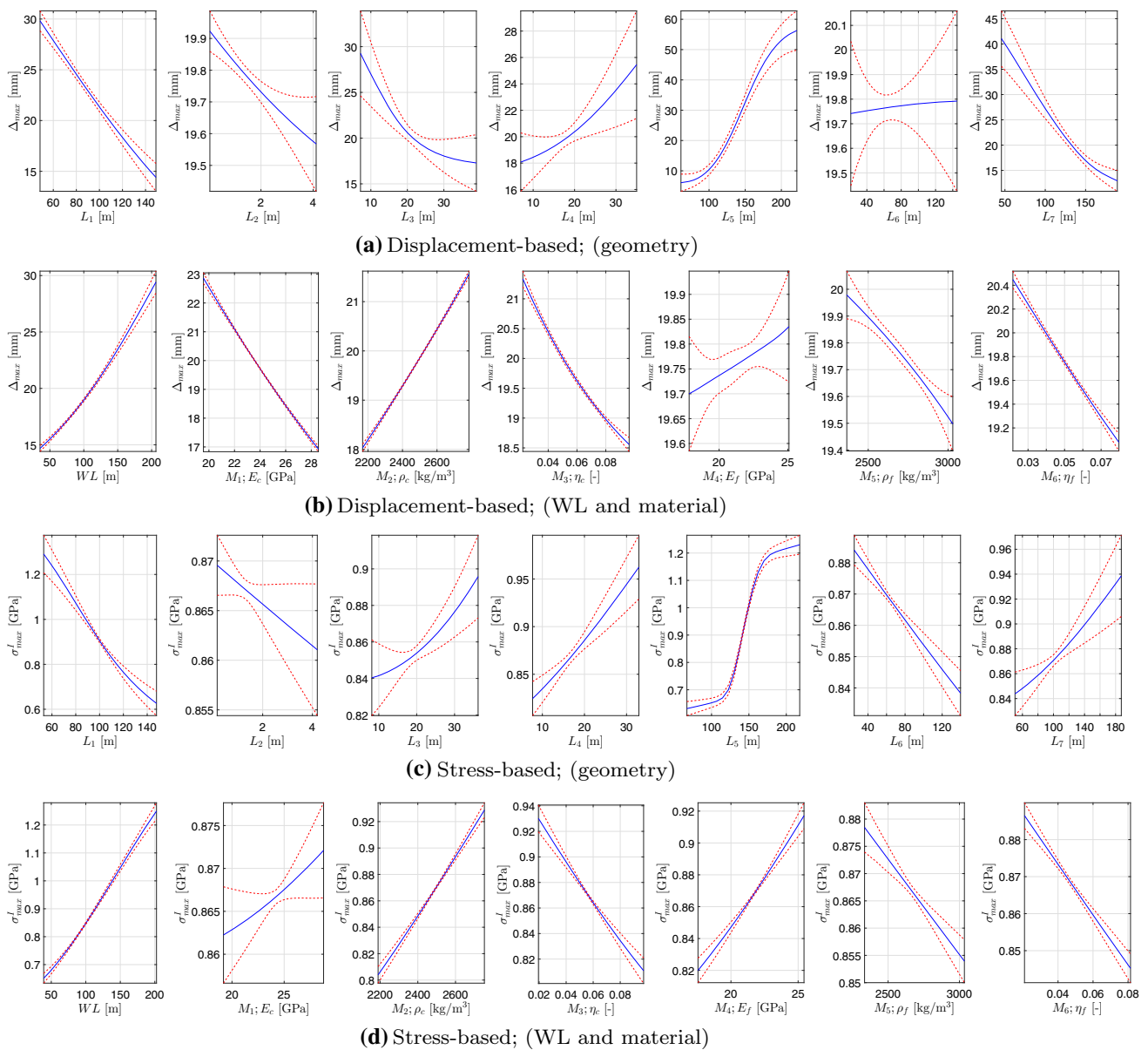


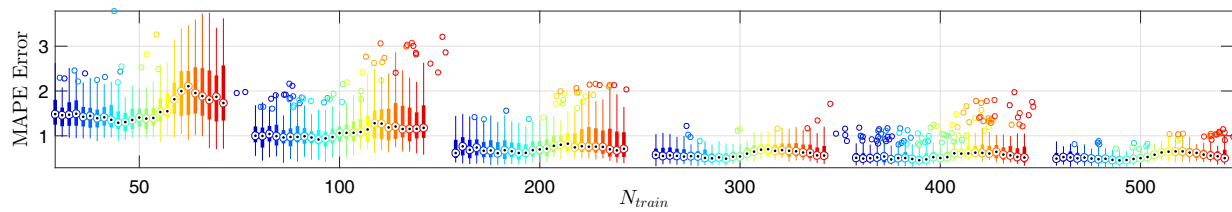
Fig. 17 Partial dependence plots for 14 RVs in PT-4

tified. Cross-validation is the most convenient approach to reduce the overall computational cost.

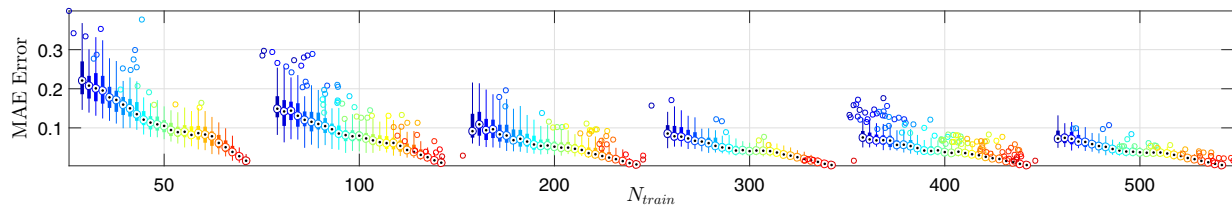
- The numerical model is also an approximation to reality that implies some degree of simplification/uncertainty. The accuracy of the model is more difficult to evaluate, unless past behavior of a dam is analyzed and monitoring data is available. Even in these cases, the measuring error can play a crucial role. Therefore, there is always uncertainty that must be taken into account when evaluating the results of this type of analyses.
- Working with SC-based meta-models, one should be cautious about extrapolating data. The predictions of these models in situations outside the range of variation of the training set are in general less reliable. In this work, we

avoided this effect by including the bounded (i.e., maximum and minimum) RVs in the training set. We thus could ensure that the prediction error computed on the validation set did not include cases outside that range. This procedure can be used when the training data are generated by numerical models, but that is not the case if monitoring data are used instead Salazar et al. (2017).

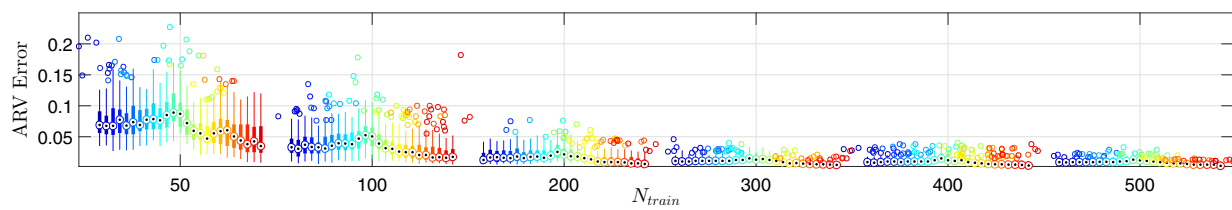
On another note, the examination of ANN models generated useful information on dam response and the effect of each of the input variables. The results of this work are in agreement with existing knowledge on dam behavior, which allowed verifying the validity of this analysis. This procedure can be more appropriate than conventional sensitivity analy-



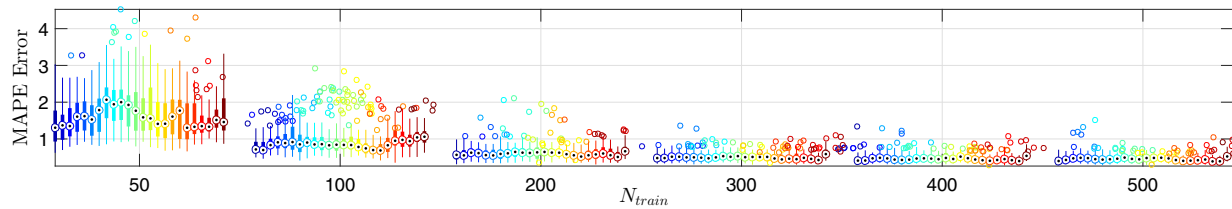
(a) Displacement-based; MAPE



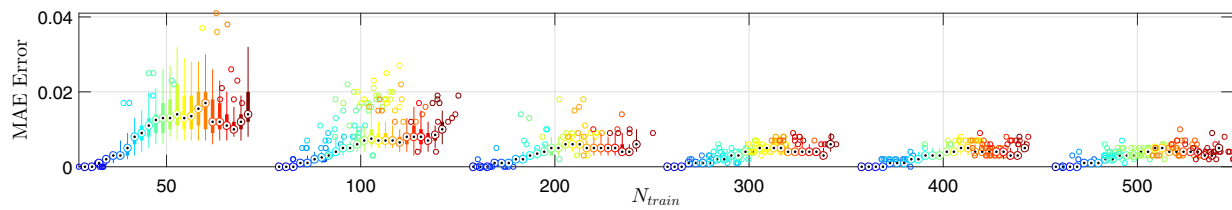
(b) Displacement-based; MAE



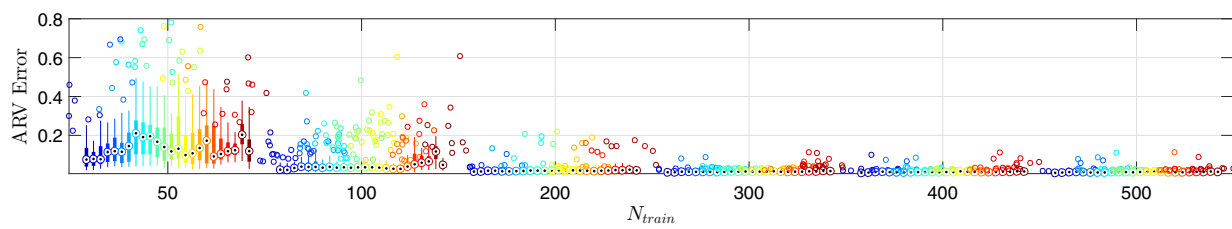
(c) Displacement-based; ARV



(d) Stress-based; MAPE



(e) Stress-based; MAE



(f) Stress-based; ARV

Fig. 18 Prediction accuracy of the ANN model for all PoIs. Results for 50 repetitions and variable training set size, computed for the validation set (8000 cases). Each set of boxplots for a given metric and training set size represent the results from the top (blue) to the bottom (red) of

the upstream face. Height of points (from top to bottom) is: 103 (blue), 101, 98, 96, 93, 91, 88, 85, 82, 80, 77, 74, 71, 67, 64, 61, 57, 52, 48, 42, 36, 29, 22, 11, 0 (red) m (color figure online)

sis based on numerical models, since SC techniques account for interaction among inputs.

As mentioned in Sect. 2, the comparison among algorithms is empirical and thus cannot be extrapolated to all kind of structural analysis of dams. In addition to possible effects of the particularities of the model, the training options may have an influence. Moreover, other existing algorithms were not considered, which offered useful results in other settings, e.g., PCE as reported by Hariri-Ardebili and Sudret (2020).

5 Conclusions

Results of the capability of five SC techniques for estimating QoIs in dynamic analysis of dams were presented. Although the options for fine tuning of each model cannot be fully considered, the results consistently point to ANN as the most accurate algorithm for this problem and all levels of complexity analyzed. The applied methodology allowed for reliable results in terms of prediction accuracy due to the relatively high number of cases used for validation. Although all algorithms can be further tuned for particular problems, the robustness of ANNs for this task with low level of adaptation can be an advantage to be applied by dam engineers, without the need for deep knowledge of the fitting options.

In this matter, special attention should be paid to control over-fitting. In this research, we run 1000–10,000 FE simulations to generate a large validation set. This allowed to obtain reliable conclusions in terms of prediction accuracy. In practice, however, the benefits of using SC models would be lost if such set of runs were made. Therefore, alternative procedures should be used such as cross-validation.

For the simplest task, with only 6 RVs involved, 100 realizations were enough to predict both crest displacement and heel stress with less than 1% error. This value was halved when the training set was enlarged up to 500 cases.

Prediction error was higher when additional parameters were considered as variables, as was the case with dam size and water level. In these cases, however, the ANN was again the most accurate technique, and it results the most robust in relation with the dam size.

SC models can also be useful for better understanding the effect of the input parameters (here, the RVs) on the system response. Partial dependence plots and measures of variable importance based on the Olden index were applied in this case for the ANN models. Results showed some degree of variation due to the random component of the process. In addition to the well-known dependency of ANN fitting to the random initialization of the weights, there is also a random component in this procedure when running RV realizations. The results of the tests performed suggest that conclusions made using one single ANN model can be misleading, but also that average results of 50–100 iterations provide stable results.

The interpretation of the ANN models can be used for selecting which input parameters to consider as RVs. This is also critical in reliability analysis, since the number of RVs strongly affects the amount of cases needed in Monte Carlo simulations.

The partial dependence plots provided interesting information for all settings. The average effect of each input on both QoIs was correctly captured by the ANN model. The results of the task in which geometrical parameters were also taken as RVs were of particular interest: Although the geometrical parameters featured strong correlation, the model was capable of correctly identifying dam height as the most influential variable for both QoIs. The results for other inputs were also in agreement with engineering knowledge. This feature can be useful in more complex settings, with high number of RVs involved and more uncertainty on their true effects.

Overall, the results suggest that ANN models can be useful in probabilistic dynamic analysis of dams for generating large datasets to be used for Monte Carlo simulation: The results with 200 samples in the training set were enough for reaching useful accuracy in most cases (below 1% error in PT-1, and below 2% for PT-2 and PT-3). This implies a relevant reduction in computational burden in a typical Monte Carlo analysis. Similar conclusions can be drawn for other QoIs, as shown in PT-5.

From a practical viewpoint, it can be concluded that SC techniques are ready to be applied in combination with numerical models for the probabilistic design of concrete dams. An efficient procedure would include further tuning of the SC model, a reduced number of numerical simulations and cross-validation. Such methodology would open possibilities to overcome the limitations of the conventional procedure based on numerical modeling due to the computational burden. For instance, the effect of the distribution of the material parameters could be assessed (shape and truncation bounds), which can be helpful in defining the measures for quality control during construction. Other options may include the consideration of the spatial variation of the material properties, as well as the assessment of the uncertainty of the loads.

The authors are working on more specific features for practical purposes: analysis of alternatives for more accurate multi-target prediction, a methodology for input selection based on model interpretation (partial dependence and variable importance) considering larger sets of RVs, and the integration of these tools in an efficient overall procedure for reliability analysis of dams.

Acknowledgements The second author acknowledges financial support from the Spanish Ministry of Economy and Competitiveness, through the “Severo Ochoa Programme for Centres of Excellence in R&D” (CEX2018-000797-S), and from the Spanish Ministry of Science, Innovation and Universities (Ministerio de Ciencia, Innovación

y Universidades) through the project TRISTAN (RTI2018-094785-B-100). The partial financial support by X-Elastica LLC is also acknowledged. The first author would like to express his sincere appreciation to his former advisor (and the current mentor), Professor Victor E. Saouma at the University of Colorado Boulder for his enthusiastic guidance and advice throughout this research.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

A Detailed response of dams with scale factor

See Fig. 19.

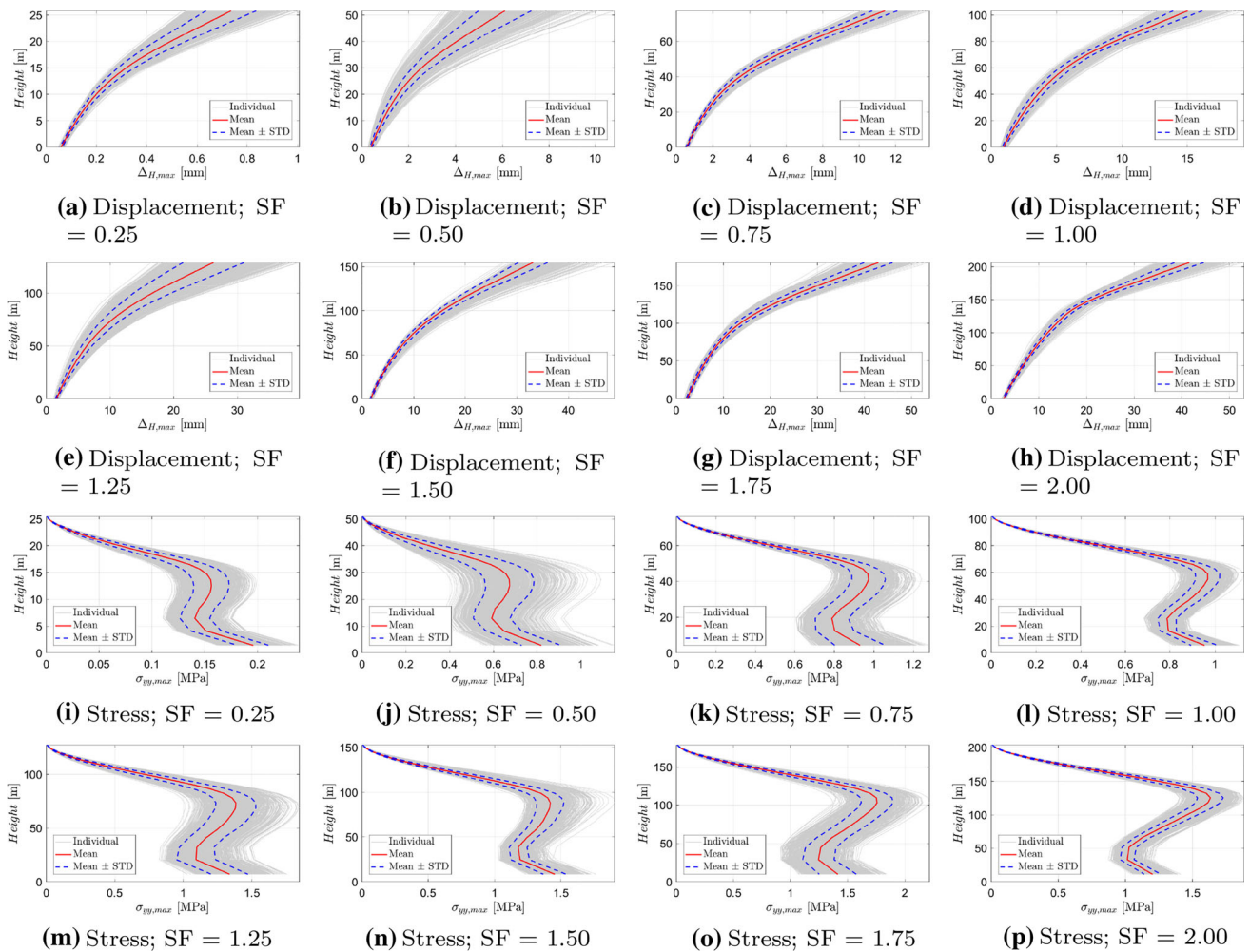


Fig. 19 Detailed uncertainty quantification of different dam classes

References

- ANSYS (2007) Ansys software reference manuals, release notes, mechanical apdl, elements reference, commands reference and theory reference, version release 11
- Arabshahi H, Lotfi V (2008) Earthquake response of concrete gravity dams including dam–foundation interface nonlinearities. *Eng Struct* 30(11):3065–3073
- ASCE (2017) 2017 report card for America's infrastructure; dams. American Society of Civil Engineers. <https://www.infrastructurereportcard.org/cat-item/dams/>. Last viewed January 2018
- Bhattacharjee S, Leger P (1993) Seismic cracking and energy dissipation in concrete gravity dams. *Earthq Eng Struct Dyn* 22:991–1007
- Bhattacharjee S, Leger P (1994) Application of nlfm models to predict cracking in concrete gravity dams. *ASCE Struct Eng* 120:1255–1271
- Breiman L (2001) Random forests. *Mach Learn* 45(1):05–32
- Bukenya P, Moyo P, Beushausen H, Oosthuizen C (2014) Health monitoring of concrete dams: a literature review. *J Civ Struct Health Monit* 4(4):235–244
- Cao X, Gu C, Zhao E (2017) Uncertainty instability risk analysis of high concrete arch dam abutments. *Math Probl Eng* 2017
- Chen J-Y, Xu Q, Li J, Fan S-L (2010) Improved response surface method for anti-slide reliability analysis of gravity dam based on weighted regression. *J Zhejiang Univ Sci A* 11(6):432–439
- Cheng L, Yang J, Zheng D, Li B, Ren J (2015) The health monitoring method of concrete dams based on ambient vibration testing and kernel principle analysis. *Shock Vib* 2015
- Der-Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? Does it matter? *Struct Saf* 31:105–112
- Díaz-Uriarte R, De Andres SA (2006) Gene selection and classification of microarray data using random forest. *BMC Bioinform* 7(1):3
- Dimitriadou E, Hornik K, Leisch F, Meyer D, Weingessel A (2008) Misc functions of the department of statistics (e1071), TU Wien. R Package 1:5–24
- Ditlevsen O, Madsen HO (1996) Structural reliability methods, vol 178. Wiley, New York
- Fan S-l, Chen J-y, Li J, Wu-qiang F (2010) Roller compacted concrete gravity dams reliability analysis based on response surface approach. In: *Earth and space 2010: engineering, science, construction, and operations in challenging environments*, pp 3355–3367
- Fedele R, Maier G, Miller B (2005) Identification of elastic stiffness and local stresses in concrete dams by in situ tests and neural networks. *Struct Infrastruct Eng* 1(3):165–180
- Fenves G, Chopra A (1984) EAGD-84: a computer program for earthquake analysis of concrete gravity dams. University of California, Earthquake Engineering Research Center
- Friedman J (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 1189–1232
- Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning. Springer series in statistics, vol 1. Springer, New York
- Friedman JH (1991) Multivariate adaptive regression splines. *Ann Stat* 1–67
- Friedman JH, Roosen CB (1995) An introduction to multivariate adaptive regression splines. *Stat Methods Med Res* 4(3):197–217
- Ganji HT, Alembagheri M, Khaneghahi MH (2019) Evaluation of seismic reliability of gravity dam-reservoirinhomogeneous foundation coupled system. *Front Struct Civ Eng* 13(3):701–715
- Gaspar A, Lopez-Caballero F, Modaressi-Farahmand-Razavi A, Gomes-Correia A (2014) Methodology for a probabilistic analysis of an rcc gravity dam construction. Modelling of temperature, hydration degree and ageing degree fields. *Eng Struct* 65:99–110
- Ghaemian M, Vafai A, Karimi Z (2014) Nonlinear seismic response of concrete gravity dams due to foundation fault movement. *Sci Iran Trans A Civ Eng* 21(5):1539
- Ghanaat Y (2004) Failure modes approach to safety evaluation of dams. In: *Proceedings of the 13th world conference on earthquake engineering*, Vancouver, BC, Canada
- Greenwell B, Boehmke B, Cunningham J, Developers G (2018) gbm: generalized boosted regression models. R Package Vers 2(1):4
- Gu C, Li B, Xu G, Yu H (2010) Back analysis of mechanical parameters of roller compacted concrete dam. *Sci China Technol Sci* 53(3):848–853
- Gu H, Wu Z, Huang X, Song J (2015) Zoning modulus inversion method for concrete dams based on chaos genetic optimization algorithm. *Math Probl Eng* 2015
- Guanglun W, Pekau O, Chuhan Z, Shaomin W (2000) Seismic fracture analysis of concrete gravity dams based on nonlinear fracture mechanics. *Eng Fract Mech* 65(1):67–87
- Hariri-Ardebili M, Saouma V (2016) Collapse fragility curves for concrete dams: comprehensive study. *ASCE J Struct Eng* 142(10):04016075
- Hariri-Ardebili MA (2018) Risk, reliability, resilience (r3) and beyond in dam engineering: a state-of-the-art review. *Int J Disaster Risk Reduct* 31:806–831
- Hariri-Ardebili MA, Barak S (2020) A series of forecasting models for seismic evaluation of dams based on ground motion meta-features. *Eng Struct* 203:109657
- Hariri-Ardebili MA, Mirzabozorg H (2013) A comparative study of the seismic stability of coupled arch dam-foundation-reservoir systems using infinite elements and viscous boundary models. *Int J Struct Stab Dyn* 13(6)
- Hariri-Ardebili MA, Pourkamali-Anaraki F (2018a) Simplified reliability analysis of multi hazard risk in gravity dams via machine learning techniques. *Arch Civ Mech Eng* 18(2):592–610
- Hariri-Ardebili MA, Pourkamali-Anaraki F (2018b) Support vector machine based reliability analysis of concrete dams. *Soil Dyn Earthq Eng* 104:276–295
- Hariri-Ardebili MA, Saouma VE (2018) Random response spectrum analysis of gravity dam classes: simplified, practical and fast approach. *Earthq Spectra* 34(2):941–975
- Hariri-Ardebili MA, Sudret B (2020) Polynomial chaos expansion for uncertainty quantification of dam engineering problems. *Eng Struct* 203:109631
- Hastie T, Tibshirani RJ, Friedman JH (2011) The elements of statistical learning: data mining, inference, and prediction. Springer, Berlin
- Karimi I, Khaji N, Ahmadi M, Mirzayee M (2010) System identification of concrete gravity dams using artificial neural networks based on a hybrid finite element-boundary element approach. *Eng Struct* 32(11):3583–3591
- Kuhn M (2008) Building predictive models in R using the caret package. *J Stat Softw* 28(5):1–26
- Liaw A, Wiener M (2002) Classification and regression by randomforest. *R News* 2(3):18–22
- Mata J (2011) Interpretation of concrete dam behaviour with artificial neural network and multiple linear regression models. *Eng Struct* 33(3):903–910
- McKay M, Beckman R, Conover W (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239–245
- Melchers RE (1999) Structural reliability analysis and prediction. Wiley, Hoboken
- Moguerza JM, Muñoz A (2006) Support vector machines with applications. *Stat Sci* 322–336
- Moran EF, Lopez MC, Moore N, Müller N, Hyndman DW (2018) Sustainable hydropower in the 21st century. *Proc Natl Acad Sci* 115(47):11891–11898

- Olden JD, Jackson DA (2002) Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecol Model* 154(1–2):135–150
- PEER (2014) Ground motion database. <http://ngawest2.berkeley.edu/>. Last viewed November 2014
- Core Team R (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Ranković V, Grujović N, Divac D, Milivojević N (2014) Development of support vector regression identification model for prediction of dam structural behaviour. *Struct Saf* 48:33–39
- Ranković V, Grujović N, Divac D, Milivojević N, Novaković A (2012) Modelling of dam behaviour based on neuro-fuzzy identification. *Eng Struct* 35:107–113
- Rezaiee-Pajand M, Tavakoli FH (2015) Crack detection in concrete gravity dams using a genetic algorithm. *Proc Inst Civ Eng Struct Build* 168(3):192–209
- Salazar F, Crookston BM (2019) A performance comparison of machine learning algorithms for arced labyrinth spillways. *Water* 11(3):544
- Salazar F, Morán R, Toledo MÁ, Oñate E (2015a) Data-based models for the prediction of dam behaviour: a review and some methodological considerations. *Arch Comput Methods Eng* 24:1–21
- Salazar F, Morán R, Rossi R, Oñate E (2013) Analysis of the discharge capacity of radial-gated spillways using CFD and ANN—Oliana Dam case study. *J Hydraul Res* 51(3):244–252
- Salazar F, Toledo M, Oñate E, Morán R (2015b) An empirical comparison of machine learning techniques for dam behaviour modelling. *Struct Saf* 56:9–17
- Salazar F, Toledo MÁ, Oñate E, Suárez B (2016) Interpretation of dam deformation and leakage with boosted regression trees. *Eng Struct* 119:230–251
- Salazar F, Toledo MÁ, González JM, Oñate E (2017) Early detection of anomalies in dam performance: a methodology based on boosted regression trees. *Struct Control Health Monit* 24(11):e2012
- Saouma V, Červenka J, Reich R (2010) Merlin finite element user’s manual. <http://civil.colorado.edu/~saouma/pdf/Software/users.pdf>. Accessed Feb 2015
- Scholkopf B, Smola AJ (2001) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge
- Segura RL, Bernier C, Durand C, Paultre P (2019) Modelling and characterizing a concrete gravity dam for fragility analysis. *Infrastuctures* 4(4):62
- Simpson TW, Poplinski J, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: survey and recommendations. *Eng Comput* 17(2):129–150
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222
- Sørensen JD (2004) Notes in structural reliability theory and risk analysis. Technical report
- Spyromitros-Xioufis E, Tsoumakas G, Groves W, Vlahavas I (2016) Multi-target regression via input space expansion: treating targets as inputs. *Mach Learn* 104(1):55–98
- Stojanovic B, Milivojevic M, Ivanovic M, Milivojevic N, Divac D (2013) Adaptive system for dam behavior modeling based on linear regression and genetic algorithms. *Adv Eng Softw* 65:182–190
- Su H, Wen Z, Zhang S, Tian S (2016) Method for choosing the optimal resource in back-analysis for multiple material parameters of a dam and its foundation. *J Comput Civ Eng* 30(4)
- Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York ISBN 0-387-95457-0
- Weigend AS, Huberman BA, Rumelhart DE (1992) Predicting sunspots and exchange rates with connectionist networks. In: Eubank S, Casdagli M (eds) Proceedings of the 1990 NATO workshop on nonlinear modeling and forecasting (Santa Fe, NM), vol 12. Addison-Wesley, Redwood, pp 395–432
- Xin C, Chongshi G (2016) Risk analysis of gravity dam instability using credibility theory monte carlo simulation model. *Springer-Plus* 5(1):778
- Zhang S, Wang G, Yu X (2013) Seismic cracking analysis of concrete gravity dams with initial cracks using the extended finite element method. *Eng Struct* 56:528–543

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.