12TH CANADIAN CONFERENCE ON EARTHQUAKE ENGINEERING
JUNE 17-20, 2019
CHÂTEAU FRONTENAC
QUÉBEC, QC.

12ÈME CONFÉRENCE CANADIENNE DU GÉNIE PARASISMIQUE
17-20 JUIN 2019

CCEE 2019

A Framework for Machine Learning-Assisted Design and Execution of Numerical Parametric Studies in Evaluating the Seismic Response of Soil-Structure Systems

Zach Bullock[1], Shideh Dashti[2], Abbie Liel[2], and Keith Porter[3]

[1]PhD Student, Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder – Boulder, CO, USA.
[2]Associate Professor, Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder – Boulder, CO, USA.
[3]Research Professor, Department of Civil, Environmental, and Architectural Engineering, University of Colorado Boulder – Boulder, CO, USA.

ABSTRACT

Numerical parametric studies are a critical tool for advancing the state of knowledge in the fields of structural and geotechnical earthquake engineering. Field observations can only be made post hoc and therefore (1) are limited by the occurrence of earthquakes and (2) typically do not include much, if any, data from instrumentation of the affected sites or structures. Physical experimentation addresses these limitations, but is still too time- and cost-intensive to use to generate large databases that explore the influence of all relevant parameters. Numerical studies must be validated using experimental or observational data, but are then limited only by the availability of computational resources, particularly when analyzing 3D, highly non-linear, soil-structure systems. Although extensive numerical studies are possible, efficiency and optimization of study design are critical as we increasingly need to model more complex systems and interactions. Typically, numerical studies are first designed (i.e., the parameters of all analyses are set), then executed, and then interpreted. This paper presents a framework for machine learning-assisted dynamic design of numerical studies based on iterative assessment of the information obtained from each analysis and maximization of the likelihood that further analyses will be informative. The steps of the framework are presented, and relevant statistical methods are reviewed. The steps include initial study design, assessment of the influence of individual analyses, and identification of the optimum subsequent analyses. A variety of machine learning techniques are applicable at each phase, and recommendations are made regarding model development and the quantification of the influence of each analysis in a dataset. A case study involving the consequences of soil liquefaction and its mitigation on building performance is considered to demonstrate the advantages of this method in terms of efficiency. The computational expense of a dynamically designed numerical study is compared to that of efficient statically designed studies, and the potential impacts of a priori uncertainty around final study quality are discussed. Lastly, the circumstances in which the dynamic approach offers significant advantages are identified.

Keywords: machine learning, parametric study design, numerical modeling, earthquake engineering, optimization.

INTRODUCTION

Numerical parametric studies are a commonly-used tool in structural and geotechnical earthquake engineering (e.g., [1-2]). This practice is commonplace because field observations of the performance of sites and structures is limited in quantity due to the practical reality of earthquake reconnaissance: they can occur only after earthquakes, only the affected structures are generally studied, and instrumentation is typically sparse. Although physical experimentation can address these shortcomings, it is also too costly to produce a database large enough to develop robust, descriptive models for the performance of built infrastructure. When properly validated and calibrated, numerical modeling addresses this limitation as well, and is limited only by the validity of the computational models used and the availability of computational power.

However, the design of numerical parametric studies is often somewhat arbitrary. Suites of models and loading conditions (usually represented by a set of ground motions) are selected in order to include a wide range of values of the parameters being investigated. While this study design paradigm can ensure that the combinations of parameters expected in the field are represented, it may result in studies that are inefficient and that fail to explore the most important range of response. This paper presents a framework for performing numerical parametric studies that addresses these limitations. Whereas traditional numerical studies are statically designed (i.e., the analyses to be performed are decided in advance), studies performed according to this framework are dynamically designed, using previous analyses to inform parameter combinations to test. Here, we focus on numerical parametric studies that will be used to develop a functional prediction of response.

Figure 1 shows a flow chart of the proposed framework. First, an initial study is proposed. The initial study is purposefully sparse in parameter space and consists of a number of numerically simulated data points that may be insufficient for model development. After the initial study is carried out, second, the relative information contained in each analysis is assessed. Third, further analyses are identified that are likely to inform modeling decisions. The second and third steps are repeated until a stopping criterion is reached, and, finally, models are produced. Steps 2 and 3 are absent from statically-designed studies, which proceed directly from running the prescribed set of analyses to interpreting the data and delivering the final model or conclusions.

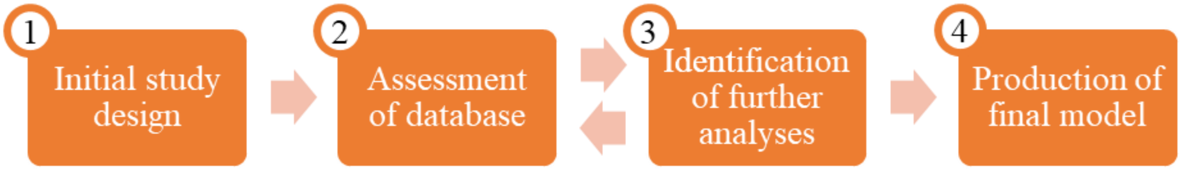

Figure 1. Flow chart of the steps in the proposed machine learning-assisted dynamic parametric study design paradigm.

The framework shown in Figure 1 requires the selection of a handful of methodologies and criteria to implement: (1) a sampling strategy for the initial study; (2) a modeling paradigm for the initial and iterative models; (3) a metric for evaluating the influence of each point in the database on the model; (4) a methodology for predicting the probability that hypothetical analyses will influence the model; and (5) a stopping criterion for iteration of the framework. Subsequent sections will review the options available for each of these selections and present a case study.

Terminology

The framework described above includes several pieces that could be described as "models," each of which include "parameters" or "features," and each of which are derived from or subjected to various types of "analyses." The situation is further complicated because each term is relevant to multiple steps outlined in Figure 1. The following discussion aims to assign a single meaning to each of these terms. We will draw examples from the numerical parametric study of Karimi et al. [2].

The primary goal of the framework is to reduce the computational expense of numerical parametric studies. We will refer to each case included in a numerical study as an *analysis*. For instance, a single nonlinear time-history analysis of a soil-foundation structure system from Karimi et al. [2] would be one analysis. An analysis consists of values of *input parameters* (such as foundation bearing pressure, q, or the peak ground acceleration of the applied motion, PGA) and *output parameters* (such as the average foundation settlement, S). The *database* consists of all of the analyses performed as of a given iteration.

The final product of a numerical parametric study is to use statistical tools to produce a *model* which characterizes trends in the output parameter(s) as a function of various input parameters. The model may be a regression model if the output parameter is a continuous quantity, or a classifier if the output parameter is a category. The terms in the functional form of the model are *features*. Features are functions of input parameters. For example, the foundation bearing pressure, q, is an input parameter, while features such as q, q^2, or $\ln(q)$ may appear in the model.

We must also use statistical tools to predict the probability that a hypothetical analysis will influence the model (Steps 2 and 3). We will refer to these tools as the *meta-model* in order to distinguish them from the model. The meta-model is a classifier, and may contain features similar to those in the model. Finally, various machine learning techniques for producing the model and meta-model have their own internal parameters. To differentiate these from the input and output parameters of the analyses, we will call them *hyperparameters*.

INITIAL STUDY DESIGN

Selection of study structure

Problems in earthquake engineering can typically be separated into two components: (1) the system, which is a representation of some kind of built infrastructure, soil, and/or rock; and (2) the loading, which is the demand applied to the system. There are two possible schemes for sampling sets of input parameters for analyses: (1) hierarchical, in which a sample of n systems is subjected to a sample of m loadings ($n \times m$ total analyses); or (2) dispersed, in which a sample consists of n system-loading combinations (n total analyses). The hierarchical study structure is needed if between-system comparisons for the same loading are desired. The dispersed structure may be impractical unless the loading can be stochastically generated (e.g., synthetic ground motions, earthquake scenarios, or stochastic displacement or force time histories).

Sampling of parameters

In order to develop the first iteration of both the model and the meta-model for predicting the influence of future analyses, an initial database of results must first be compiled. This database must reflect some degree of variability in all input parameters of interest, but need not be so extensive as to facilitate development of the final model. Several methodologies exist for generating minimally-sized samples which maximize the coverage of the input parameter space. Here, we review those which are most useful in the context of parametric studies in earthquake engineering.

Latin hypercube sampling (LHS; [3]) is a very popular methodology for optimizing samples of input parameters for civil engineering models, particularly in structural reliability (e.g., [4]) and hydrology (e.g., [5]). It has also been employed frequently in earthquake engineering, typically in combination with reliability methods or in the context of propagating uncertainty (e.g., [6-8]). LHS may be improved by algorithms to reduce the pairwise correlation coefficients between input parameters [9] or by using orthogonal arrays to assemble the study design [10].

As its name suggests, quasi-Monte Carlo sampling (QMC; e.g., [11]) is very similar to Monte Carlo sampling, but with a low-discrepancy series of numbers (e.g., the Sobol [12] sequence) being used to generate input parameter values rather than a pseudorandom series. Low discrepancy sequences of numbers are equally distributed in space for any sequence length. Use of a low discrepancy series ensures that the points in the sample are dispersed in input parameter space. Selecting cases in order from a low-discrepancy set of numbers allows two operations that are not convenient in LHS: (1) if we have already generated a QMC design with n cases, we can expand it to $n + n^*$ cases simply by analyzing the next n^* cases in the same low-discrepancy set; and (2) if we already generated a QMC design with n cases, we can easily subdivide this design into two QMC designs by splitting the cases but retaining their order (e.g., in order to split training and test data).

ASSESSMENT OF DATABASE

Steps 2 and 3 from Figure 1 seek to add analyses to the study iteratively. This step evaluates the current database in order to determine which analyses are most influential on model development. First, a prototypical version of the final model is produced. Many statistical tools can be employed to develop this model, depending on the quantity being modeled, each with several applicable machine learning techniques to assist with functional form selection. Subsequently, the analyses in the initial database are evaluated according to their relative influence on both form selection and final model coefficients. To accomplish these tasks, we must first select a modeling paradigm for our desired final model, and then select a metric for evaluating the influence of each analysis in the database. At each iteration, these two tasks will be repeated (i.e., a prototypical model will be produced, and the influence of each analysis will be evaluated).

Selection of a modeling paradigm

Choosing a statistical tool for model development depends primarily on the desired format of the output parameters in the study. Linear and nonlinear regression (e.g., [13]) are the preferred methods for generating estimates of continuous variables such as peak ground acceleration (e.g., [14]) or foundation settlement (e.g., [1], [16-17]). For estimating the probability of the occurrence (or nonoccurrence) of a specific event or condition, logistic regression [17] or deep learning (e.g., [18]) may be preferred. Lastly, for classification of points into multiple categories, multinomial logistic regression (e.g., [19]), naïve Bayseian classifiers (e.g., [20]), random forests [21], and deep learning are available. These suggestions are far from an exhaustive review of statistical methods for forward prediction.

For any of these methods, a variety of machine learning techniques are available to determine the optimal features to include in the model and to improve model performance. Methods such as ridge regression [22], lasso regression [23], and the elastic net [24] involve regularization of model coefficients and can be used to produce a parsimonious model from a large population of candidate terms. These methods are preferred for linear and logistic regression, and similar regularization can be implemented for nonlinear regression in many cases.

The performance of regression models and classifiers can be improved through bootstrap aggregation (bagging [25]) or boosting (e.g., [26]). These methods work by producing an ensemble of models that each produce estimates, which are then recombined. Although these methods can improve model performance, they may make the model impractical to describe in a publication and therefore difficult to communicate to other engineers (e.g., there may be hundreds or thousands of separate models, each with its own form, coefficients, and weight in the recombination). This problem applies to deep learning models even without use of ensemble methods, and random forests are naturally an ensemble method without the addition of bagging or boosting. Finally, many of these methods depend directly on using cross validation techniques (e.g., [27]), or can be improved substantially through their use.

Selection of an evaluation metric

In order to proceed to the next step in the framework, we must evaluate all analyses in the database according to their relative influence on model development. Ideally, the metric used to quantify the analyses' influence will capture both their influence of data on the coefficients of the selected model (within-model influence) and their influence on which model is chosen (between-model influence).

Any proxy metric for within-model influence should quantify the influence of each point in the database on the coefficients in the final model. Cook's distance [28] is a common measure of the within-model influence of points, and is defined by summing changes in the model predictions for all points in the database when a single point is removed. Cook's distance is related to the leverage of the points in the database. The leverage of a point is a function of the distance between the values of the predictors at that point to the values of the predictors for the population. Often, the leverage itself is used to identify influential points, since points with high leverage will tend to pull regression lines closer. However, raw leverage may not be applicable as a metric in this study, since all points on the periphery of the initial input parameter space should have similar leverage if LHS, QMC, or a similar method is used to generate the initial study sample.

DFBETA [29] is a measure of the change in model coefficients when a single point is removed. The definition of DFBETA does not sum over points or predictors, so there is a DFBETA statistic value for each coefficient-data point pair. DFBETA-based metrics can be tailor-made for specific situations, particularly when certain predictors are considered critical. In the context of earthquake engineering, an example might be the sum of DFBETA over all points over all predictors related to ground motion intensity.

Cook's distance is most directly applicable to linear and nonlinear regression models. Analogous influence diagnostics exist for logistic regression [30] which can also be applied to multinomial logistic regression [31]. DFBETA can be easily applied in these cases as well. However, quantifying the influence of points in the database on the formulation of more complex models (e.g., deep learning) is substantially more difficult (e.g., [32]). Some of the machine learning techniques for model selection described above (e.g., ridge or lasso regression) have existing metrics that reflect the between-model influence of points in the database (e.g., [33-36]). However, calculation of between-model metrics may be computationally expensive, particularly if they require selecting functional forms for the database with each point excluded one at a time.

Novel or problem-specific influence metrics may be preferred or even required, depending on the situation. Metrics based on Cook's distance may be the most general and the easiest to interpret and develop. However, metrics related to DFBETA may be preferred if a specific input parameter is of particular interest (e.g., if there is one critical input parameter which may be augmented by the remaining input parameters).

IDENTIFICATION OF SUBSEQUENT ANALYSES

Step 2 quantified the relative influence of all analyses in the database at a given iteration. The next step in the framework is to identify which analyses should be added to the database for the next iteration. First, we must develop the meta-model for predicting the influence of hypothetical additional analyses. The restrictions on more complex model types (i.e., the need to clearly interpret and communicate the trends identified in regression), such as deep learning or random forests, are less consequential during this phase since there is little need to communicate the features, coefficients, and outputs of the meta-model at each iteration compared to those of the final model. Then, the meta-model must be applied to determine the analyses to be added. After these analyses are performed and their output parameters are recorded, we return to Step 2.

We propose reducing this phase to a probabilistic classification problem: what is the likelihood that a hypothetical point will be material to the final model? Because this is a binary problem (i.e., a point is material or it is not), regular logistic regression is a potential candidate for developing the meta-model, although boosted logistic regression or deep learning may improve performance. In order to represent the problem in this way, some threshold value of the selected within- and between-model influence metrics must be defined. The threshold could be absolute (e.g., Cook's distance exceeding 1.0) or relative (e.g., the top 10% of points in terms of influence). Further, this threshold should incorporate measures of both within- and between-model influence, and it may therefore be desirable to sum two or more metrics, where each may be normalized by their average or maximum value or otherwise weighted. Potential analyses are then identified by finding sets of input parameter values that maximize the predicted likelihood of influence. The inputs to the meta-model should be basic variables (i.e., the input parameters identified by the LHS or QMC sample in the initial study) rather than intermediate variables calculated subsequently. Further constraints should be applied to penalize the selection of points that are too close to existing points in the database.

Next, we must decide how many analyses to add in each iteration (i.e., before reevaluating each point's influence and rebuilding the meta-model). This decision may be influenced by the relative computational expense of one analysis and the development of the meta-model. For very large databases of relatively cheap analyses, iterating in increments of multiple analyses may be optimum for reducing overall computational expense, but the cost of a single very complex analysis (e.g., a time history analysis of a finely meshed nonlinear soil-foundation-structure model such as those in Karimi et al. [2]) may well significantly exceed the cost of developing the meta-model. In those situations, iterating after single analyses may be preferable.

Finally, the framework requires a stopping criterion. Practically, the final database size will likely be constrained by the availability of computing time, so a hard limit may be imposed on the number of analyses to be performed. Alternatively, adding more points will eventually fail to add an appreciable amount of information to the database (e.g., [37]). Convergence to this limit could be identified if the variance of the influence metrics falls below a certain threshold (i.e., all points in the database are equally relevant to the outcomes) or if the performance of the meta-model falls below a certain threshold (i.e., areas in the input parameter space containing informative analyses are too difficult to identify). If a set of test data (i.e., data that are not used in development of the model) can be afforded, the marginal improvement in the model's test error may fall below a tolerance, after which the iterations can be stopped.

PERFORMANCE OF THE FRAMEWORK FOR AN EARTHQUAKE ENGINEERING APPLICATION

Problem description

In this section, we apply the framework to an existing problem and compare its performance to that of statically-designed parametric studies. The case study consists of an example discussed in Bullock et al. [38]: development of a model for replacing certain inputs in the Bullock et al. [15] procedure for estimating foundation settlement with a function of the liquefaction potential index (LPI). The data analyzed in the case study consist of random field realizations of borehole logs of soil density ($N_{1,60}$) and fines content (FC) and randomly generated earthquake scenarios. Each realization is described by 63 random variables that dictate the number, geometry, soil type, and density of layers in the soil profile, all relevant characteristics of the hypothetical earthquake scenarios, and which methods are used to estimate ground acceleration and the factor of safety against liquefaction. Table 1 provides the ranges considered for all input parameters.

Table 1. Input parameters for the framework case study.

Input parameter	Values considered
Number of susceptible layers	1 to 5
Thickness of an individual layer	2 to 30
Depth to water table, z_{wt} (m)	0 to 4
Unit weight of layers, γ (kg/m^3)	14 to 22
SPT blow count of susceptible layers (mean), $\mu(N_{1,60})$ (blows)	5 to 40
SPT blow count of susceptible layers (coefficient of variation), $CV(N_{1,60})$ (%)	0 to 40
SPT blow count of susceptible layers (correlation length), $\theta(N_{1,60})$ (m)	1 to 7
Fines content of susceptible layers (mean), $\mu(FC)$ (%)	0 to 30
Fines content of susceptible layers (coefficient of variation), $CV(FC)$ (%)	0 to 40
Fines content of susceptible layers (correlation length), $\theta(FC)$ (m)	1 to 7
Earthquake moment magnitude, M_W	4.0 to 8.0
Earthquake focal depth, H (km)	0 to 30
Earthquake dip angle, δ (degrees)	0 to 90
Earthquake distance to rupture, R_{rup} (km)	0.1 to 300
Earthquake rupture mechanism	Normal, reverse, strike-slip or unknown

In a given analysis, the parameters in Table 1 are used to calculate the liquefaction potential index (LPI) and the soil term for the Bullock et al. [15] procedure for estimating foundation settlement (F_{soil}). Calculating F_{soil} requires calculating the median cumulative absolute velocity (CAV) predicted by the Bullock et al. [39] ground motion prediction equations. The goal is to produce a database of analyses to use to develop a model that predicts F_{soil} as a function of LPI and other input parameters. The lasso is used to select features for regression models and to determine their coefficients [23]. Possible features include the natural logarithm and first four orders of LPI, and the natural logarithm and first two orders of CAV, the earthquake moment magnitude (M_W), and the earthquake distance to rupture (R_{rup}).

Framework parameter selection

The case study uses a LHS design with 100 analyses for the initial database and a test database consisting of 100 separate analyses. We use the model residuals as a simple proxy for the influence of each point, and we consider a point "influential" if its absolute residual ($|r|$) is larger than the 75th percentile absolute residual ($\widetilde{|r|}_{75}$). We predict the probability of a point being influential by training a random forest with adaptive boosting [21,40]. At each iteration, ten analyses (defined by selecting the input parameters in Table 1) are added which optimize the expression in Equation 1, where d is a vector of the Euclidean distances between the hypothetical points and all of the existing points in parameter space. We track changes in the test mean squared error (MSE) as data are added, and stop adding data when the change in test MSE changes by less than 1% for three consecutive iterations.

$$\text{maximize: } P\left(|r| > \widetilde{|r|}_{75}\right) \times \min(d) \qquad 1$$

Framework performance

Figure 1 shows the test MSE as a function of the current database size as random data are added, as data are from a QMC sample are added sequentially, and as data are added using the proposed framework. Adding data randomly or from a QMC sample generally fails to improve the model's test error, even when the total database size is increased six-fold from the initial database. However, adding data by optimizing the likelihood that each point will be influential on the final model results in improved test error and the algorithm stops after the database size has tripled. At the very beginning of the analysis, the test error is increased before falling sharply. This initial increase occurs because the initial database is sparse enough that the first few added points are selected primarily because they have very large values of $\min(d)$ rather than $P\left(|r| > \widetilde{|r|}_{75}\right)$, per Equation 1 (i.e., they fill relatively sparse areas in the parameter space).

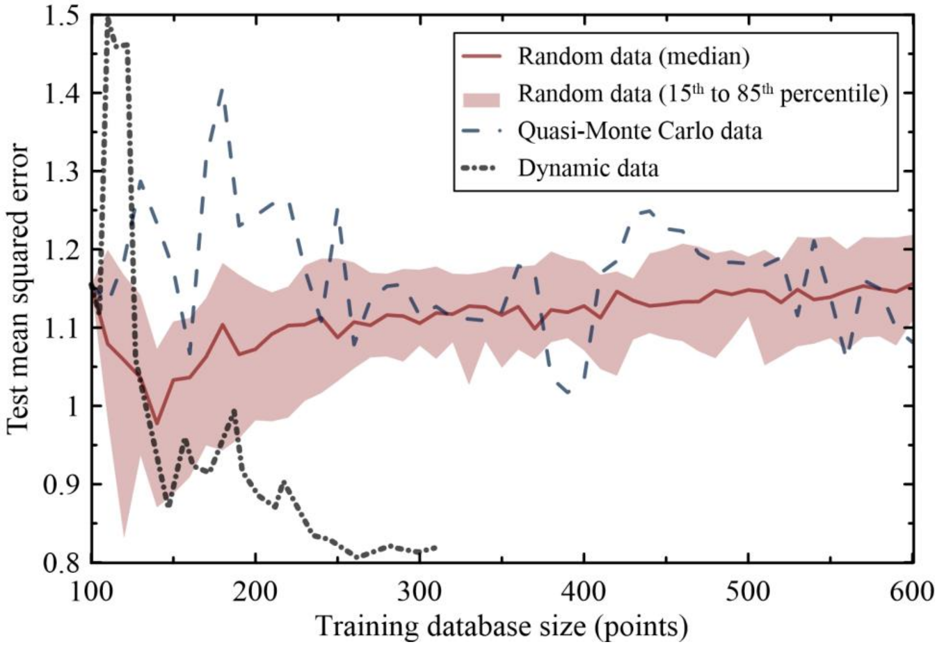

Figure 1. Performance of the framework as compared to randomly selected data and quasi-Monte Carlo data.

CONCLUSIONS

This paper outlines a framework for minimizing the total expense of numerical parametric studies in earthquake engineering. The framework is shown to improve model performance while reducing the amount of data needed for model development. However, additional work is needed to determine the best practices for selection of several hyper-parameters, including: (1) the optimum proxy for within- and between-model influence; (2) the optimum threshold for considering a point "influential;" (3) the optimum initial database size, and how to generalize the initial database size for a parametric study of arbitrary complexity; and (4) the optimum stopping criterion.

Although the framework proposed here can be applied to any numerical parametric study, it is particularly useful for geotechnical earthquake engineering. Numerical modeling of soil-foundation-structure systems subjected to seismic loading can be computationally expensive. Therefore, we are often constrained in terms of the number of analyses that can practically be performed. However, many input parameters of such models can be varied, and it is often uncertain which input parameters (and in what ranges) will be the most influential on the final results. Application of machine learning in the described framework can help reduce this uncertainty and produce higher quality models at reduced computational expense.

ACKNOWLEDGMENTS

This research is supported partly the United States Department of Education under award number P200A150042 and partly by the Civil, Environmental, and Architectural Engineering department at the University of Colorado, Boulder.

REFERENCES

[1] Bray, J. D., & Macedo, J. (2017). 6th Ishihara lecture: Simplified procedure for estimating liquefaction-induced building settlement. *Soil Dynamics and Earthquake Engineering*, *102*, 215-231.

[2] Karimi, Z., Dashti, S., Bullock, Z., Porter, K., & Liel, A. (2018). Key predictors of structure settlement on liquefiable ground: a numerical parametric study. *Soil Dynamics and Earthquake Engineering*, *113*, 286-308.

[3] McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, *21*(2), 239-245.

[4] Novák, D., Teplý, B., & Keršner, Z. (1998). The role of Latin Hypercube Sampling method in reliability engineering. In *Proc. of ICOSSAR* (Vol. 97, pp. 403-409).

[5] Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of hydrology*, *324*(1-4), 10-23.

[6] Liel, A. B., Haselton, C. B., Deierlein, G. G., & Baker, J. W. (2009). Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. *Structural Safety*, *31*(2), 197-211.

[7] Dolsek, M. (2009). Incremental dynamic analysis with consideration of modeling uncertainties. *Earthquake Engineering & Structural Dynamics*, *38*(6), 805-825.

[8] Vamvatsikos, D., & Fragiadakis, M. (2010). Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty. *Earthquake engineering & structural dynamics*, *39*(2), 141-163.

[9] Iman, R. L., & Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics-Simulation and Computation*, *11*(3), 311-334.

[10] Tang, B. (1993). Orthogonal array-based Latin hypercubes. *Journal of the American statistical association*, *88*(424), 1392-1397.

[11] Niederreiter, H. (1992). Random number generation and quasi-Monte Carlo methods (Vol. 63). Siam.

[12] Sobol, I. Y. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, *7*(4), 784-802.

[13] Myers, R. H. (1990). *Classical and modern regression with applications* (Vol. 2). Belmont, CA: Duxbury press.

[14] Bozorgnia, Y., Abrahamson, N. A., Atik, L. A., Ancheta, T. D., Atkinson, G. M., Baker, J. W., ... & Darragh, R. (2014). NGA-West2 research project. *Earthquake Spectra*, *30*(3), 973-987.

[15] Bullock, Z., Karimi, Z., Dashti, S., Porter, K., Liel, A. B., & Franke, K. W. (2018a). A Physics-Informed Semi-Empirical Probabilistic Model for the Settlement of Shallow-Founded Structures on Liquefiable Ground. *Géotechnique*, 1-34.

[16] Bullock, Z., Dashti, S., Karimi, Z., Liel, A. B., Porter, K. A., and Franke, K. W. (2018b). Probabilistic models for residual and peak transient tilt of mat-founded structures on liquefiable soils. *Journal of Geotechnical and Geoenvironmental Engineering*.

[17] Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 215-242.

[18] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436.

[19] Böhning, D. (1992). Multinomial logistic regression algorithm. *Annals of the institute of Statistical Mathematics*, *44*(1), 197-200.

[20] Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46). New York: IBM.

[21] Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

[22] Tikhonov, A. N., Goncharsky, A., Stepanov, V. V., & Yagola, A. G. (2013). *Numerical methods for the solution of ill-posed problems* (Vol. 328). Springer Science & Business Media.

[23] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

[24] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(2), 301-320.

[25] Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123-140.

[26] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119-139.

[27] Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, *4*, 40-79.

[28] Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics, 19*(1), 15-18.

[29] Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity* (Vol. 571). John Wiley & Sons.

[30] Pregibon, D. (1981). Logistic regression diagnostics. *The Annals of Statistics*, *9*(4), 705-724.

[31] Lesaffre, E., & Albert, A. (1989). Multiple-group logistic regression diagnostics. *Applied Statistics*, 425-440.

[32] Koh, P. W., & Liang, P. (2017). Understanding black-box predictions via influence functions. *arXiv preprint arXiv:1703.04730*.

[33] Walker, E., & Birch, J. B. (1988). Influence measures in ridge regression. *Technometrics*, *30*(2), 221-227.

[34] Shi, L., & Wang, X. (1999). Local influence in ridge regression. *Computational Statistics & Data Analysis*, *31*(3), 341-353.

[35] Jang, D. H., & Anderson-Cook, C. M. (2017). Influence plots for LASSO. *Quality and Reliability Engineering International*, *33*(7), 1317-1326.

[36] Kim, C., Lee, J., Yang, H., & Bae, W. (2015). Case influence diagnostics in the lasso regression. *Journal of the Korean Statistical Society*, *44*(2), 271-279.

[37] Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1, No. 10). New York, NY, USA:: Springer series in statistics.

[38] Bullock, Z., Dashti, S., Liel, A. B., & Porter, K. A. (2019). Generating Synthetic Borehole Data for Applications in Site-Specific and Regional Evaluation of Liquefaction Consequences. In *Proceedings of Geo-Congress 2019*. Philadelphia, PA. March 27-29, 2019.

[39] Bullock, Z., Dashti, S., Liel, A., Porter, K., Karimi, Z., & Bradley, B. (2017). Ground-Motion Prediction Equations for Arias Intensity, Cumulative Absolute Velocity, and Peak Incremental Ground Velocity for Rock Sites in Different Tectonic Environments. *Bulletin of the Seismological Society of America*, *107*(5), 2293-2309.

[40] Schapire, R. E., & Freund, Y. (1999). A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, *14*(5), 771-780.