Utilizing the Monte-Carlo Capability in RS2 for Machine-Learning Applications

Amichai Mitelman(✉), Avshalom Ganz, and Alon Urlainis

Ariel University, Ariel, Israel
amichaim@ariel.ac.il

Abstract. Due to the uncertainty that stems from the heterogeneous nature of geological materials, probabilistic tools have been incorporated into geotechnical practice. A notable example is the Monte Carlo analysis method that is available as a built-in feature in the program RS2 (Rocscience in Phase2 version 6.020. Rocscience Inc. Toronto, 2007). In this paper, we demonstrate how to utilize the Monte Carlo method for enhancement of geotechnical analysis. The procedure consists of two primary stages: 1) data generation, where numerous finite-element (FE) models are generated based on an estimated range of input parameters, and: 2) data analysis, where machine-learning models are used to correlate input parameters and results of interest. The verified ML algorithm can be referred to as a surrogate model. An example of the implementation of a surrogate model is illustrated through an anchored sheet pile wall problem. For this problem, the surrogate model correlates between the forces in the first and second row of anchors, thus allowing for prediction and optimization of ground support during construction.

Keywords: Geotechnical engineering · machine-learning · surrogate models · Monte Carlo analysis

1 Introduction

Machine-learning (ML) algorithms have emerged as powerful tools for various applications that require the analysis of large data sets. ML algorithms can recognize patterns, which are in turn used to make predictions. ML methods have important advantages compared to traditional statistical methods, including an improved capability to generalize and adapt to previously unseen data [5]. An overview of different ML applications for geotechnical engineering is given by Morgenroth et al. [8]. The grander vision of ML with respect to rock and soil mechanics, is to establish greater predictive power for design purposes based on correlations between digital and quantifiable data collected in-situ, and actual ground behavior [12].

Nonetheless, despite their promise, for proper application of ML tools there are remain great challenges and obstacles where in-situ data is collected, amongst them:

- In many cases, project management officials are reluctant to share data due to matters of liability and confidentiality.

R. E. Hammah et al. (Eds.): RIC 2023, AHE 19, pp. 156–161, 2023.
https://doi.org/10.2991/978-94-6463-258-3_16

- Compared to mechanical solutions, ML models highly are sensitive to input parameter ranges. Generally, they perform poorly on extrapolated data. In other words, they fail to predict results for data that is outside the range of the learning process.
- Given the heterogeneity and uncertainty associated with rock and soil mechanical properties, it remains an open question whether learning monitoring data from one site would be useful for a different site, even for one that is situated in a similar geological environment.
- It is difficult to collect and organize high-quality data, as monitoring devices that deliver reliable and digitized data are costly, and require careful interpretation of recordings.
- The question of the amount of data required cannot be determined before-hand, and it is required to collect, train and test data until ML model performance is satisfactory. This problem applies to ML in general [1]. As a result, stakeholder may be reluctant to invest in a large-scale ML project as results cannot be guaranteed.

In this regard, surrogate models (sometimes referred to as approximation models, or black-box models) have recently been realized as a powerful analysis tool for different engineering fields. Surrogate models generally involve 3 primary steps: (1) artificial data generation, (2) data learning and testing, and (3) surrogate model execution. For geological engineering, surrogate models can be designed for different objectives, including instant computation of results, sensitivity analysis, and solving inverse problems [3]. Fortunately, the aforementioned drawbacks do not generally apply to surrogate models, as artificial data can be generated by automation of numerical modeling. Given so, it is argued that surrogate models are a great starting point for engineers interested in developing ML skills.

In this paper, the use of a surrogate model is demonstrated through the example of a practical problem: a design and optimization problem of an anchored sheet pile wall with tiebacks. The problem is first presented, and the process of data generation via the built-in Monte Carlo probabilistic analysis feature in the program RS2 [11] is then described. Finally, current limitations and suggestions for further development are highlighted.

2 Example Problem

Anchored tieback walls are highly ubiquitous structures, and represent a ground-support interaction problem best solved using numerical simulations [7]. For walls where more than one level of tiebacks is required, the observational approach, originally proposed by Peck [9], may be adopted. Under this approach, monitoring data collected in the initial excavation stage can be used to infer the wall and anchor behavior in the subsequent stage of construction.

The following example problem, shown in Fig. 1, is taken from the RS2 tutorial titled: “Anchored sheet piled walls”. The model consists of two excavation stages, where a row of tieback anchors is installed after excavation. It is proposed that the axial forces along the first row of tiebacks is correlated to the forces that develop in the subsequent row of tiebacks. Given that the predicted forces are small, a moderate factor of safety may be applied, and the extent of support can be reduced. In contrast, if the prediction

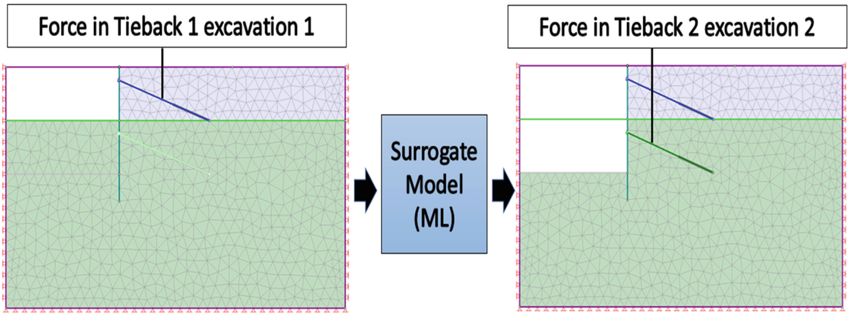

Fig. 1. Model excavation stages and surrogate model application

shows that large forces will develop (i.e. the ground is weaker than anticipated), a denser configuration of ground anchors must be considered.

It is emphasized that there is no trivial answer regarding what data to collect for this purpose. Since there are two soil layers, and many input parameters (Young's modulus, cohesion, friction angle, and more) that may vary greatly, the problem is complex and no straightforward solution can be applied. In addition, there is no a-priori guarantee that knowledge of the forces in the first-row anchors is sufficient for predicting the forces in the second row.

Hence, it is proposed that a surrogate model is built in order to analyze the data systematically. If the surrogate model succeeds in making accurate predictions, then it can be concluded that this approach is potentially valid. The process of data generation is described in the following section and illustrated in Fig. 1.

3 Data Generation via the Monte Carlo Method

Owing to the inherent uncertainty of geological engineers, statistical tools and probabilistic analyses have been widely applied in geotechnical practice for various purposes [6]. In contrast to traditional deterministic analyses where single values are used to compute a single model, in a probabilistic analysis the user selects a distribution of input parameters, and multiple models are randomly generated accordingly. One such method is the Monte-Carlo method, which involves a rigorous and iterative solving process. The great advantage of the Monte-Carlo method is its ability to work with any type of distribution (e.g. uniform, normal, exponential, etc.) and variables, whether independent or not. According to the RS2 user guide, the results of the Monte Carlo analysis in RS2 allow for the user to interpret contours of the mean, standard deviation, and coefficient of variation for all output variables, error plots for the range of output values, and yield zone probability.

For the current example, the Monte Carlo method is used as a tool for data generation. This data can then be analyzed using ML tools, and different correlations can be examined. Once the Monte Carlo option is selected, the number of models is assigned by the user. As discussed, there is no means of determining the number of FE models required to obtain accuracy, and this depends on the complexity of the problem [4].

Table 1. Input parameters varied for data generated via the Monte Carlo method

Material	Input parameter	Mean value	Range
Soil Layer #1	Young's modulus [MPa]	20	10
	Friction angle [degrees]	40	10
	Cohesion [MPa]	0.05	0.05
Soil Layer #2	Young's modulus [MPa]	50	25
	Friction angle [degrees]	30	10
	Cohesion [MPa]	0.2	0.1

The input parameters that are to be varied are selected, and their distribution and ranges must be assigned. Selection of reasonable values requires careful engineering judgment and experience. It is important to bear in mind that ML models are not capable of extrapolation, hence the ranges should be selected accordingly. For this example, the input parameter range is given in Table 1.

After the numerical model and statistical distributions are defined, the compute command in RS2 will carry out automated generation and solving of the FE models. Depending on the model, computer, and number of models to be solved, the solving process may require a long time. It is therefore advised to first generate a small number of models and verify that their results are reasonable, prior to solving several models.

Once the solving process is completed, a compressed output file is generated. Extracting this file allows axis to a number of files that contain results (stresses, strains, etc.) It is possible to write an external code and to create variables according to the desired inputs and outputs for the ML analysis. For this example, a Python code was written and the bolt forces were loaded from the corresponding result files. The code creates two variables:

1. B1- the axial forces along the first tieback after the first excavation.
2. B2- the axial forces along the second tieback after the second excavation.

In order to test the correlation between variables B1 and B2, different ML algorithms can be trained and then tested on this data. Note that no variable that consists of the soil input parameters is created, and the impact of soil strength is accounted for implicitly, through B1 and B2.

A detailed overview of the procedure for the application and verification of ML analysis will not be given here, as this is beyond the scope of the current paper. In brief, it is found that ML models perform successfully on a data set of 100 FE models.

As in any engineering application, implementation of the proposed analysis for practical applications requires an awareness to methodological limitations. Surrogate models are constrained to the simplifying assumptions of the data they have been trained on (i.e. numerical models). Therefore, drawing conclusions for practical problems should be made with caution. Monitoring data collected in-situ is regularly found to be considerably noisy. Arguably, a significant amount of work is still needed in order to increase the overall predictive power of numerical models in geotechnical engineering. In order to enhance the proposed method, it is possible to synthesize the proposed analysis with real-world data (e.g. monitoring devices in actual projects, laboratory readings, etc.) [2].

4 Summary and Ideas for Future Development

On the basis of various examples and applications, Furtney et al. [3] anticipate that surrogate models will become an important tool in engineering practice. In this paper, an example of the utilization of a surrogate model for optimization of support in a staged excavation problem is presented.

The Monte Carlo feature in RS2 allows users to readily generate a large number of FE models based on statistical distributions. It is shown that this data can then be analyzed with external ML tools. However, there are currently a number of technical limitations that constrain the wider application of surrogate models.

The process of properly building the data sets for ML analysis from the RS2 input and output files may require significant efforts, depending on the desired task. Note that other RocScience programs such as Slide2, for slope stability, and RocSupport for tunnel support design, include built-in features that allow users to readily export input and output data into tabular Excel format [10]. These methods are based on closed-form solutions and require considerably less computational resources than numerical codes. Therefore, they are very suitable for probabilistic analyses.

In contrast, the 3D numerical program RS3, currently lacks the Monte Carlo feature, but requires significant computational costs. Due to the latter issue, many engineers choose to avoid 3D modeling. It is hypothesized that surrogate models can be used to develop powerful techniques for accounting for 3D effects in 2D simulations.

To summarize, it is generally encouraged that 2D and 3D numerical packages allow for interface capabilities that help researchers and engineers accelerate the integration of ML tools into geotechnical engineering.

References

1. Balki, I., Amirabadi, A., Levman, J., Martel, A. L., Emersic, Z., Meden, B., Garcia-Pedrero, A., Ramirez, S. C., Kong, D., & Moody, A. R. (2019). Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. Canadian Association of Radiologists Journal, 70(4), 344–353.
2. Elmo, D., Mitelman, A., & Yang, B. (2022). Examining Rock Engineering Knowledge through a Philosophical Lens. Geosciences, 12(4), 174.
3. Furtney, J. K., Thielsen, C., Fu, W., & le Goc, R. (2022). Surrogate Models in Rock and Soil Mechanics: Integrating Numerical Modeling and Machine Learning. Rock Mechanics and Rock Engineering, 55(5), 2845–2859.
4. Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. "O'Reilly Media, Inc."
5. Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.
6. Mitelman, A., & Elmo, D. (2018). A proposed probabilistic analysis methodology for tunnel support cost estimation depending on the construction method. 52nd US Rock Mechanics/Geomechanics Symposium.

7. Mitelman, A., & Elmo, D. (2019). Analysis of tunnel-support interaction using an equivalent boundary beam. Tunnelling and Underground Space Technology, 84. https://doi.org/10.1016/j.tust.2018.11.021
8. Morgenroth, J., Khan, U. T., & Perras, M. A. (2019). An overview of opportunities for machine learning methods in underground rock engineering design. Geosciences, 9(12), 504.
9. Peck, R. B. (1969). Advantages and limitations of the observational method in applied soil mechanics. Geotechnique, 19(2), 171–187.
10. Rocscience. (2004). Rocscience software products: Dips, Slide, Phase2. Rocscience Toronto.
11. Rocscience. (2007). Phase2 version 6.020. Rocscience Inc. Toronto.
12. Yang, B., Mitelman, A., Elmo, D., & Stead, D. (2022). Why the future of rock mass classification systems requires revisiting their empirical past. Quarterly Journal of Engineering Geology and Hydrogeology, 55(1).