Delft University of Technology

On the use of Machine Learning in Geotechnical Engineering

Brinkgreve, Ronald; Zekri, Ashraf

Publication date
2024

Document Version
Final published version

Published in
Bundesanstalt für Wasserbau Kolloquium Numerik in der Geotechnik

Citation (APA)
Brinkgreve, R., & Zekri, A. (2024). On the use of Machine Learning in Geotechnical Engineering. In *Bundesanstalt für Wasserbau Kolloquium Numerik in der Geotechnik* (pp. 15-23). (BAWKolloquium). Bundesanstalt für Wasserbau.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

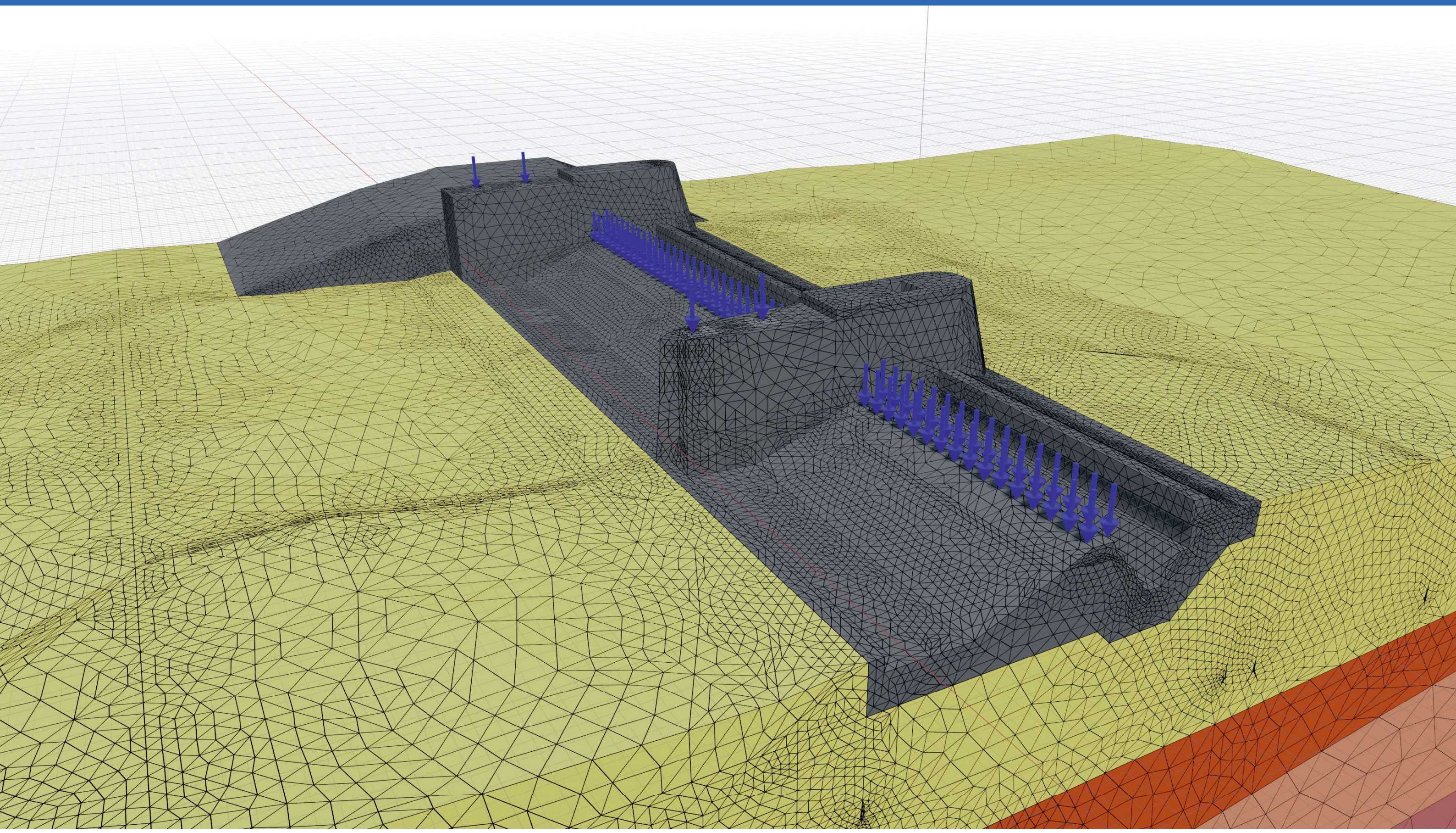
Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

BAW
Bundesanstalt für Wasserbau

Kompetenz für die Wasserstraßen

BAWKolloquium



Numerik in der Geotechnik

7. und 8. November 2024
Bundesanstalt für Wasserbau, Karlsruhe

Impressum

Herausgeber (im Eigenverlag):
Bundesanstalt für Wasserbau (BAW)
Kußmaulstraße 17, 76187 Karlsruhe
Postfach 21 02 53, 76152 Karlsruhe
Telefon: +49 (0) 721 9726-0
E-Mail: info@baw.de, www.baw.de

Creative Commons BY 4.0

https://creativecommons.org/licenses/by/4.0/

Soweit nicht anders angegeben, liegen alle Bildrechte bei der BAW.

BAWKolloquium: ISSN 2698-6841

Veranstaltungen der BAW verfügbar unter:
https://henry.baw.de/handle/20.500.11970/99403

Karlsruhe · November 2024

Titelbild: FE-Modell Wehr Geesthacht

On the use of Machine Learning in Geotechnical Engineering

Ronald Brinkgreve, Delft University of Technology & Seequent – The Bentley Subsurface Company
Ashraf Zekri, Seequent – The Bentley Subsurface Company

1. Introduction

The use of Artificial Intelligence (AI) and Machine Learning (ML) have significantly increased over the last couple of years. ML is driven by the availability of data. Although geotechnical engineering is generally not among the first in picking up new technologies, there is a lot of data in our profession, and therefore, lots of opportunities to apply ML.
In this article some examples are given of how ML can be used to facilitate and automate the geotechnical engineering workflow. Examples of soil identification, CPT interpretation, 3D soil stratification, parameter determination and surrogate modelling are given, and some other applications are mentioned.

2. Machine Learning

Machine Learning is a subset of methods within the realm of Artificial Intelligence. It can roughly be subdivided into *supervised learning* and *unsupervised learning* (other categories are beyond the scope of this article). In supervised ML, the computer is trained to find a complex relationship between a set of input parameters and the expected results (so-called *labelled data*). This can be further subdivided into *classification* (where results end up in predefined categories or classes) and *regression* (where the results are real numbers, such as the outcome of a mathematical function fitted to data). In unsupervised ML, the computer is expected to find similarities in given (*unlabelled*) data all by itself, without specifying any classes. An example of unsupervised learning is *clustering*, where the computer tries to group data points showing a certain correspondence, like clusters of points in a multi-dimensional space.
Several different ML methods (algorithms) exist. This article is not meant to give a complete introduction into ML and the various data processing methods; a nice ML overview can be found on Wikipedia (https://en.wikipedia.org/wiki/Machine_learning).
The recent availability of dedicated Python libraries for ML, such as Scikit-learn and PyTorch, has boosted the use of ML in all kinds of disciplines; not only in science and research, but also in (engineering) practice. What used to be a relatively new and complex field of expertise for mathematicians, data scientists and computer programmers, has now become routine for enthusiastic engineers, as part of their primary job activities, with endless application opportunities.

3. Data in Geotechnical Engineering

Although there is much uncertainty in the ground, there is still a lot of data going around in geotechnical engineering. Data obtained from site investigation and testing are interpreted and processed into parameters for geotechnical analysis and design purposes. Data obtained from monitoring and in-situ measurements (thanks to *The Observational Method*) are used to control the progress and safety of a construction project, and used to adapt the design, if necessary. Databases and case histories (soil properties, test results, (numerical) solutions, etc.) can be used as a reference for new situations. All these data can become even more useful in ML applications.

The storage, processing and management of (geotechnical) data related to a project or an asset can be conveniently and securely done in a cloud-based Digital Twin (DT) environment; not only for engineering, design and construction purposes, but also during its entire lifetime, for observations, maintenance, extension, retrofit, and finally decommissioning of the asset. The DT contains the most recent information about the state of an asset and its history, provided the data is kept up to date. Since a DT has facilities to securely manage a lot of data, and it provides interfaces to various data sources as well as applications that make use of the data, it can be easily seen how ML and DT can benefit from each other.
In the next Sections different types of applications using ML in geotechnical engineering will be discussed and demonstrated.

4. CPT Interpretation

Cone Penetration Testing (CPT) is a relatively cheap way to explore the upper part (~30m) of the sub-soil for geotechnical engineering and design purposes. CPT interpretation is generally based on Robertson's method (Robertson & Cabal, 2015) in which the core CPT parameters (cone resistance q_c, sleeve friction f_s (or friction ratio R_f) and pore pressure u_2, with depth) are translated into Soil Behaviour Type (SBT). This is subsequently used to extract soil profiles by clustering layers with similar SBT, and to infer soil properties and model parameters using correlations with the core CPT parameters.
Only a few lines of Python code can already demonstrate a simple use of ML by taking a number of data points (for example 1000) as combinations ($\log(q_c)$, $\log(R_f)$) using Latin Hypercube Sampling with the corresponding SBT obtained from Robertson's chart to train a ML model, and let the trained ML model predict the SBT for arbitrary sets of CPT parameters. Figure 1 shows the results using Gaussian Process classification: Left is the original Robertson chart and right is the contour plot from the ML model, including the sample points. Apart from some 'artefacts', there is good similarity with a coefficient of determination R^2 = 0.98. This application of ML may not be very useful, since equations exist to identify the various zones in Robertson's chart, hence SBT can be directly obtained from combinations of $\log(q_c)$ and $\log(R_f)$. However, it is a nice exercise for someone who would like to start using ML in geotechnical engineering.

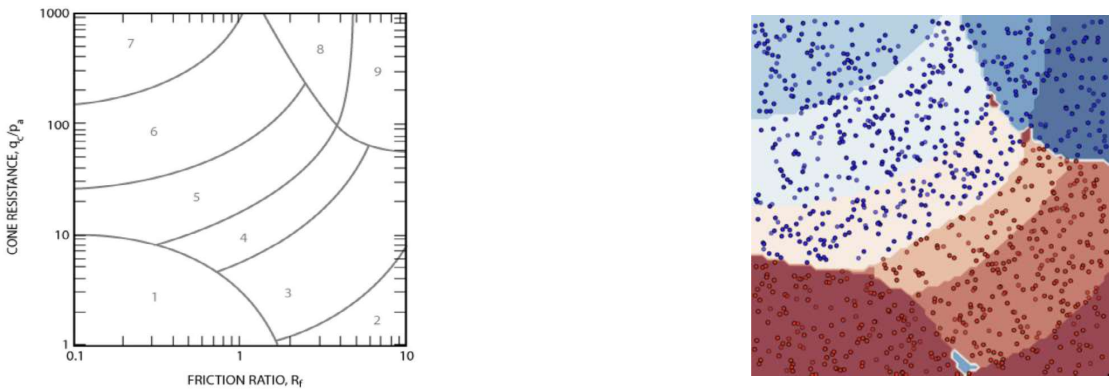


Figure 1. ML model of SBT according to Robertson (2010).
Left: Original Robertson chart with SBT zones Right: Contour plot from ML model

5. 3D Stratification

After determining the soil profile in individual CPTs, it is still a challenge to compose a 3D subsoil model, since the layering may change from one CPT to another and layers with the same SBT do not necessarily belong to the same geological formation. A ML clustering algorithm can be used to find similar layers across multiple CPTs (Brinkgreve et al., 2023). The clustering is based on a 'signature' of the layer sections as identified from individual CPTs. Besides the average core or derived CPT parameters, the signature may involve average depth, (effective) stress and/or stochastic characteristics (e.g. variance or scale of fluctuation) of the layer sections.

Figure 3 shows an example of four CPTs from a site in Rotterdam (NL), displaying the q_c -profile with depth together with the clustering information (five different colors in the 'boreholes'). The clustering is based on *K-Means*, using the stratified layer sections from individual CPTs with a minimum thickness of 0.35m and mean values of depth d (in reference coordinates), log(q_t/p_a) and log(R_f) as input parameters (Figure 2).

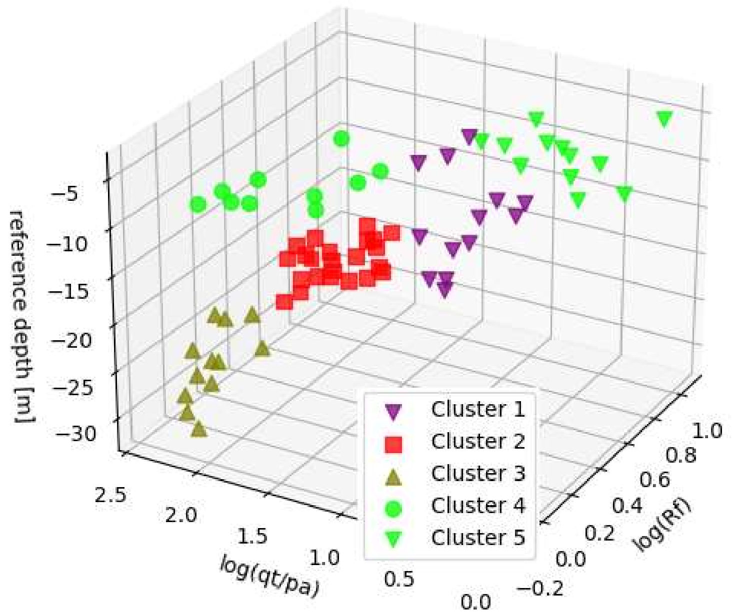


Figure 2. Clustering of layer sections into five clusters

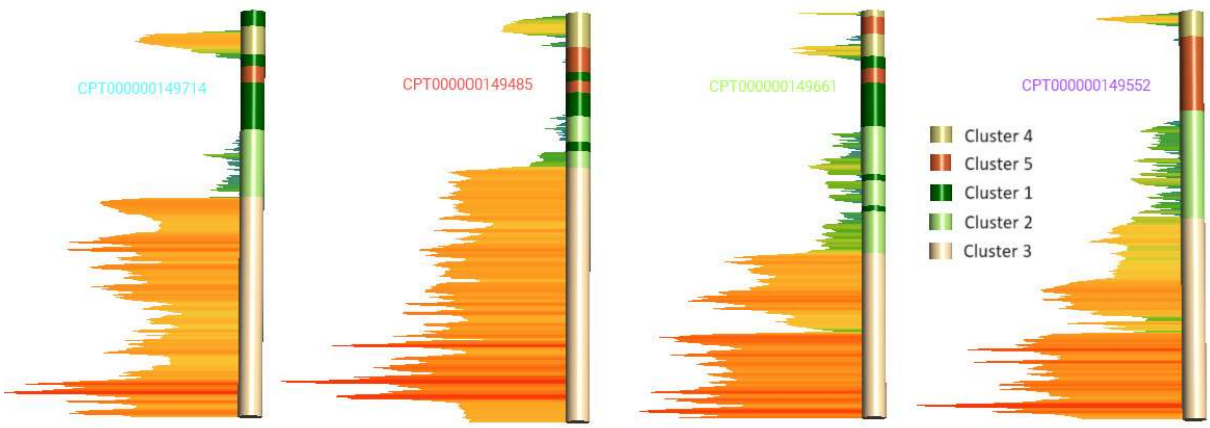


Figure 3. Four CPTs from a site in Rotterdam with cone resistance q_c and clustering of layer sections

In this way, clustering of layer sections can be done for multiple CPTs in a project. The CPT interpretations together with the cluster information (rather than Soil Behaviour Type, SBT) is sent as 'boreholes' to a geological modelling package (Leapfrog), which is subsequently used to create a 3D sub-soil model, from which 2D or 3D sections can be taken for numerical analysis purposes.

6. Parameter determination from field test data

CPT data can also be used to infer soil and model parameters. Although there is no theoretical relationship between the core CPT parameters and soil properties, many correlations exist to estimate soil properties or parameters from CPT data. It is important to recognize and obey the conditions and validity of a correlation before using it in arbitrary situations, as several of those are soil(behaviour) type and site-specific.

In a research collaboration on Automated Parameter Determination (APD) between TU Delft, TU Graz, Witteveen + Bos, and Seequent, Graph theory (a form of AI) was used to automatically create 'paths of correlations' (Van Berkom et al., 2022), all the way from core and derived CPT parameters via intermediate soil properties and parameters to final soil and model parameters, thereby obeying the correlation conditions. APD can be applied on individual CPT readings as well as on averaged CPT data per soil layer (1D, 2D or 3D).

After a period in which the outcomes of APD had been validated and fine-tuned (Marzouk et al., 2022; 2023; 2024), an attempt was made to see if the results from APD could be used to train a ML model based on a large data set of core and derived CPT parameters. The preparation of data and the training of the ML model itself takes a bit of time (depending on which ML algorithm is used), but once this is done and the ML model has been validated, it can be stored and reused in other applications. This costs only a few tens of kilobytes, and it is extremely fast.

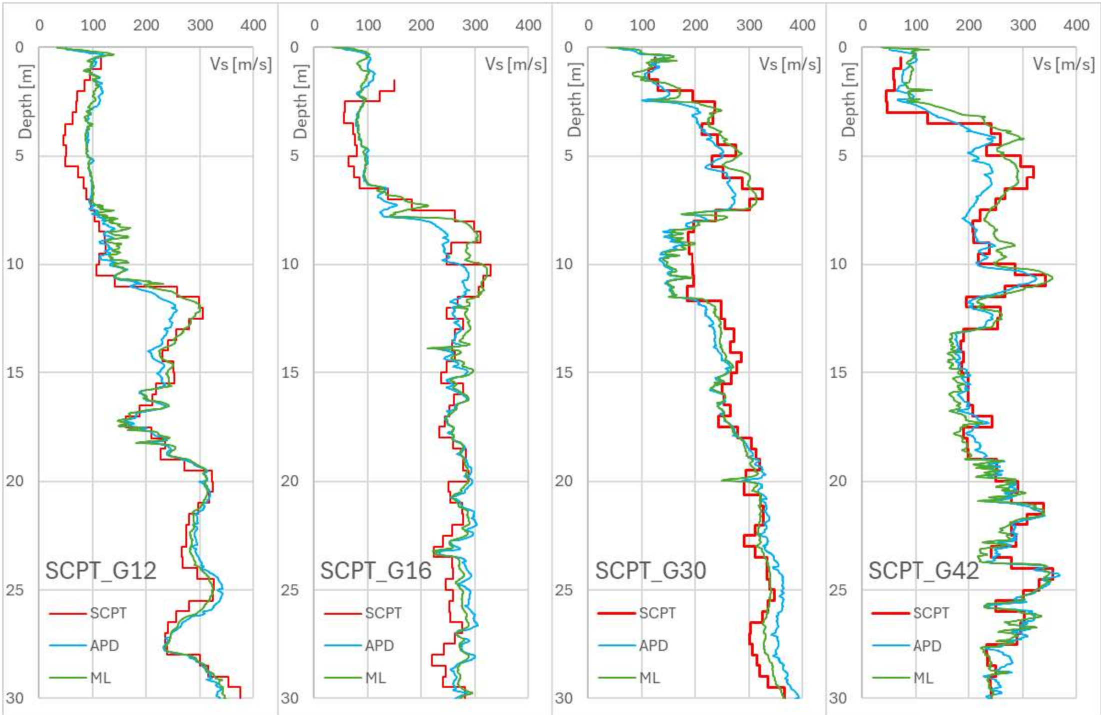


Figure 4. *Comparison of Vs profile obtained from APD, ML and SCPT measurements for four different locations in Groningen, The Netherlands*

As an example, a Neural Network-based ML model was created to determine the shear wave velocity V_s based on the input of depth d, $\log(q_t/p_a)$, $\log(R_f)$ and vertical effective stress σ'_y obtained from CPT data. V_s is used to calculate the small-strain shear modulus G_0 as a model parameter in advanced constitutive models. The Neural Network was trained on V_s results from APD based on 34 CPTs from the Groningen region (NL) (Kruiver et al., 2021) together with V_s measurements from 31 Seismic CPTs (SCPT), with a total of 42875+1657 = 44532 data points, resulting in R^2 = 0.991. Figure 4 shows that the results obtained from the APD correlations already give a reasonable prediction of the measured V_s profile from the SCPTs, while the ML results are even better. Note that the SCPT gives only one value per meter depth, while CPT readings (and hence the V_s obtained via APD and ML) is determined for every 2 centimeters.
Further research is being undertaken to investigate the use and accuracy of ML models for CPT-based parameter determination as an alternative to correlations.

7. Parameter determination from lab test data

Where field testing is performed in a preliminary design stage of a project, lab testing is generally done in a somewhat later stage to provide more detailed soil properties and parameters, facilitating a more refined design. Soil and model parameters inferred from lab testing data are considered more accurate, although this can also be misleading for different reasons (sample disturbance, unrealistically high strength properties of organic soils, etc.), which are beyond the scope of this article.
Geotechnical labs may possess large databases of lab testing data, which can be very interesting for ML applications. For example, a student from Delft University of Technology (Stals, 2023; Stals et al., 2024) executed his MSc graduation project at a big geotechnical lab in The Netherlands where he had access to a large database of lab testing data and corresponding soil index properties. He numerically simulated a series of undrained triaxial tests using the Hardening Soil small-strain constitutive model and optimized the model parameters to match the real triaxial test results. He added the optimized model parameters to the database, such that it linked the model parameters to the corresponding soil index properties. He then trained a Neural Network model to predict the model parameters based on the input of the soil index properties, without using the lab testing data. In this way, he created a lab testing-based ML model that could be used together with a field testing-based ML model to facilitate the determination of soil and model parameters in an early project stage, when limited soil testing data is available.

8. Surrogate modelling in probabilistic analysis

Another use of ML in geotechnical engineering is surrogate modelling. In surrogate modelling, a ML regression model is trained on the results from more time-consuming numerical analysis to provide similar answers, but much faster. This is particularly useful in the case of repetitive calculations in which only some (material or geometric) parameters vary. Examples of such applications are (geometric) optimization or probabilistic analysis. Meanwhile, several applications of surrogate modelling in geotechnical engineering have been reported in literature.
As an example, a surrogate model was built to perform probabilistic slope stability analysis. A probabilistic stability analysis involves the generation of trials for every slip surface, where each trial has a set of properties that are generated by sampling from a cumulative density function

(CDF). The parameters in the probabilistic stability analysis can be soil strength parameters, piezometric surface properties, reinforcement parameters or loading properties. Several kinds of CDF functions are available. The factor of safety (FoS) should be calculated for each trial of every slip surface. The computation time of this phase can be prohibitive, especially in 3D geometries.

For a given analysis project, a ML model was trained using a subset of trials of each surface, solved by GeoStudio's slope stability package SLOPE3D based on the limit equilibrium method. The trained model was utilized to estimate the FoS of the rest of the trials of that surface. This geometry-agnostic surrogate model accelerates the analytical process, facilitating efficient and flexible what-if scenarios.

Figure 5 demonstrates the location of 50 slip surfaces on an open pit structure, while Figure 6 shows the histograms of the probabilistic parameters. A ML model was developed for each slip surface, utilizing 25% of the data from 5000 trials solved by the SLOPE3D software. A comparison of the FoS of these slip surfaces has proven that the results from the surrogate model are very close to the values calculated by the slope stability program. This observation is substantiated by the ML models' robust performance, with an R^2 value greater than 0.990 for all evaluated surfaces.

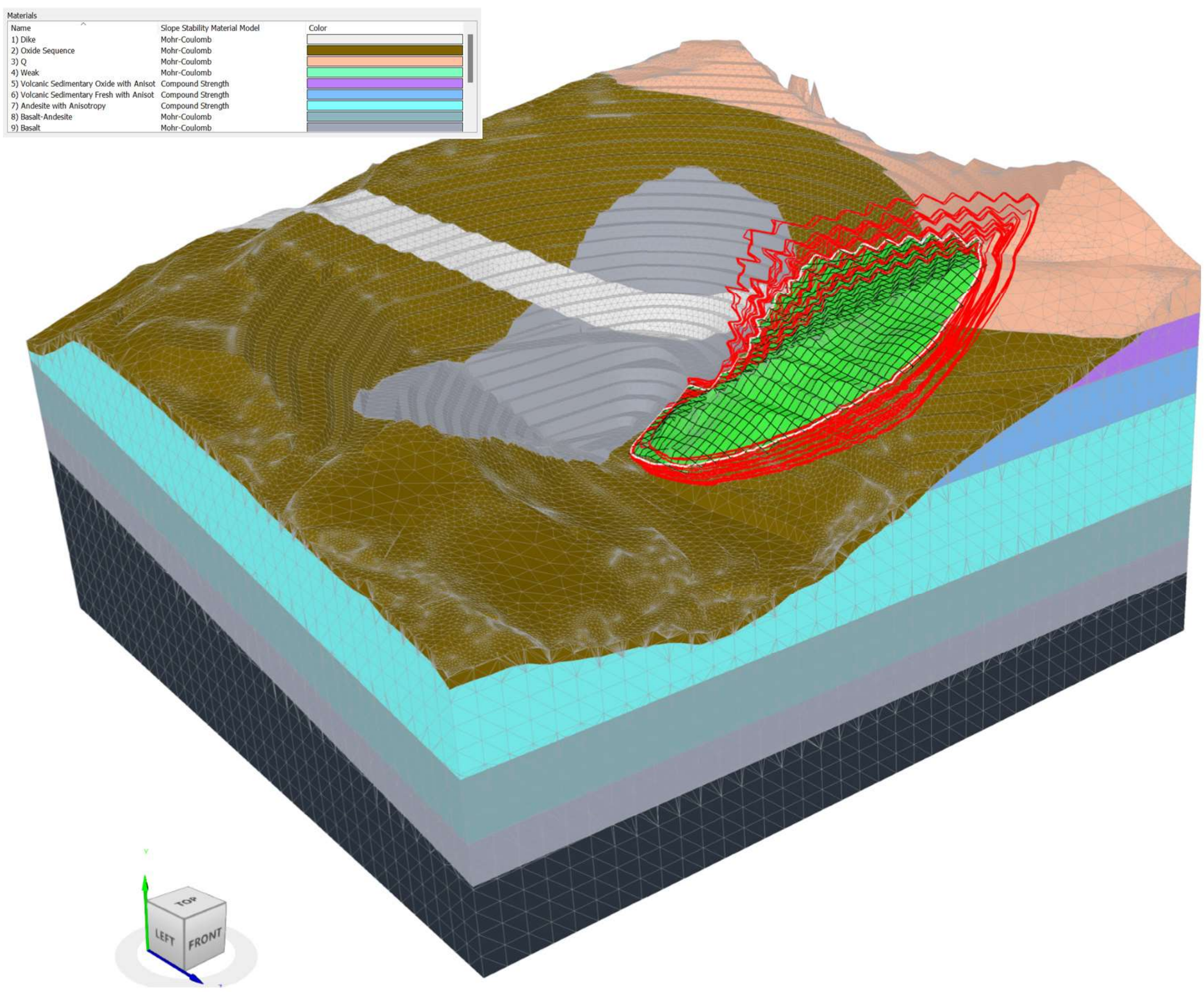


Figure 5. A specific slip surface in an open pit example

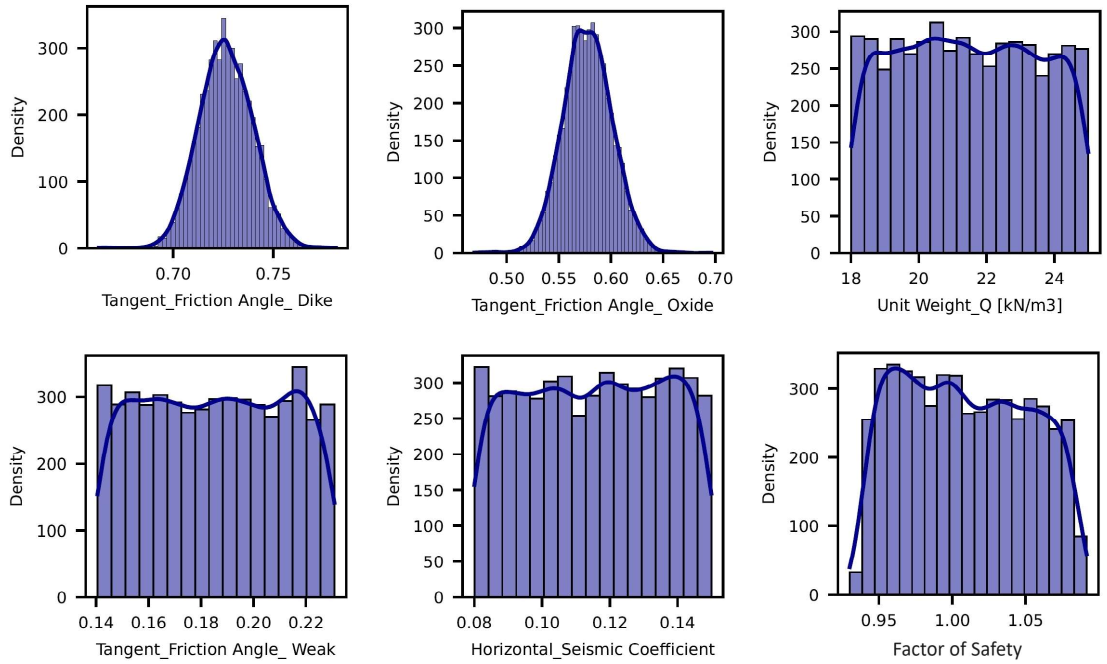


Figure 6. *Histogram of the probabilistic parameters used in the probabilistic slope stability analysis*

9. Other applications of Machine Learning

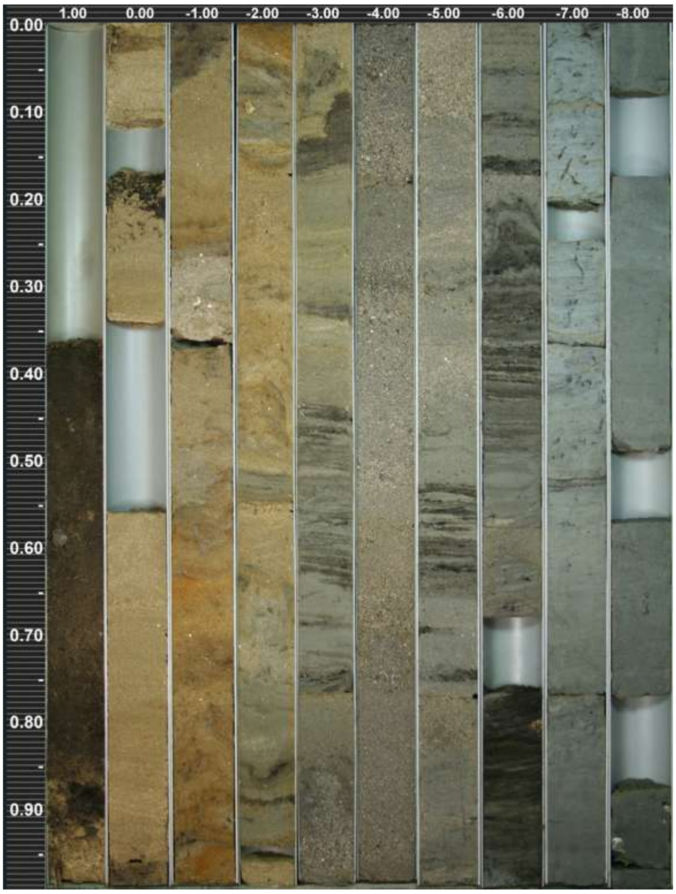
Besides the examples elaborated in the previous sections, there are many more opportunities for using ML in geotechnical engineering and design.

A typical application concerns the processing and interpretation of observation and monitoring data from project sites or from satellites (Synthetic Aperture Radar, SAR or InSAR) to find trends and to establish correlations (e.g. between excavation level and wall deflection, or between (ground)water level and land subsidence). ML models trained on such data can be used to make predictions and be integrated in early warning systems to enable timely maintenance or mitigation measures to be taken, preventing undesired scenarios from happening. In this way, as well as when trained on historical data and case studies, ML models can play an important role in geotechnical risk assessment.

Another application of ML is in various geotechnical design aspects, such as the design of soil retaining walls, soil reinforcements, shallow foundations, piles, tunnels, dams, embankments or excavated slopes (e.g. open pit mines). Given in-situ soil data (such as from field tests) and the required excavation depth and load conditions, a ML model can predict the required sheet-pile profile or wall/plate dimensions, the required number of piles, anchors or reinforcements and their lengths, or the maximum angle of the slope to be excavated. Moreover, when trained on the results of multiple loading tests with known soil conditions, a ML model can predict bearing capacity and load-displacement behaviour of axially and laterally loaded piles and footings.

Finally, a challenging application of ML would be in recognizing and classifying soils from images of soil core drillings (Figure 7). Besides field tests, core drillings are generally taken to identify soil types and geological formations, and to determine soil profiles, while some soil samples are used for lab testing purposes. In the current practice, high resolution photos are taken from the core drillings, but the interpretation is still a manual task. It would be good if soil recognition and classification could be further automated based on such images. For rock, automatic classification of core drillings and estimation of properties like Rock Quality Designation (RQD) is in progress, but for soil, automatic classification and identification of properties seems more difficult.

Depth [m]



Courtesy from Mos Grondmechanica

Figure 7. Example of a soil core drilling with ten 1m sections; some soil samples were taken for lab testing purposes, leaving 'holes' in the casing. ML could potentially be used for soil classification from such images.

10. Conclusions

Machine Learning (ML) offers endless opportunities to facilitate geotechnical engineering and design. The use of ML methods (algorithms) and the creation of ML models is relatively simple nowadays in the programming language Python, with the availability of dedicated software libraries. This article demonstrates the use of ML models for the interpretation of CPT data, finding corresponding layers across multiple CPTs, determination of soil and model parameters based on field and lab testing, and surrogate modelling for probabilistic analysis. It also mentions other applications of ML to facilitate the geotechnical engineering profession.

Although ML models allow for full automation of the geotechnical engineering workflow, it is important to emphasize the role and responsibility of the geotechnical engineer in validating the outcomes from ML models.

Literature

Brinkgreve, R.B.J.; Tschuchnigg, F.; Laera, A.; Brasile, S. (2023). Automated CPT interpretation and modelling in a BIM/Digital Twin environment. In: Zdravkovic, Kontoe, Taborda & Tsiampousi (eds), 10th Conf. on Numerical Methods in Geotechnical Engineering. London: Imperial College. https://doi.org/10.53243/NUMGE2023-111.

Kruiver, P.P.; de Lange, G.; Kloosterman, F.; Korff, M.; van Elk, J.; Doornhof, D. (2021). Rigorous test of the performance of shear-wave velocity correlations derived from CPT soundings: A case study for Groningen, the Netherlands. Soil Dynamics and Earthquake Engineering 140. https://doi.org/10.1016/j.soildyn.2020.106471.

Marzouk, I.; Tschuchnigg; F., Paduli, F.; Lengkeek, H.J.; Brinkgreve R.B.J. (2022): Determination of fine-grained soil parameters using an automated system. In: Gottardi & Tonni (eds), Cone Penetration Testing 2022. Taylor & Francis, 540–545. https://doi.org/10.1201/9781003308829-77.

Marzouk, I.; Tschuchnigg; Brinkgreve R.B.J. (2023): Expansion of an automated system for determining soil parameters using in-situ tests. In: Zdravkovic, Kontoe, Taborda & Tsiampousi (eds), 10th Conf. on Numerical Methods in Geotechnical Engineering. London: Imperial College. https://doi.org/10.53243/NUMGE2023-70.

Marzouk, I.; Oberhollenzer, S.; Tschuchnigg, F. (2024): An automated system for determining soil parameters: Case study. In: 8th Int. Symp. on Deformation Characteristics of Geomaterials, Porto. https://doi.org/10.1051/e3sconf/202454404002.

Robertson, P.K. (2010). Soil Behaviour Type from the CPT: an update. Proceedings: 2nd Int. Syposium on Cone Penetration Testing, Huntington Beach, USA, Vol. 2. 575–583.

Robertson, P.K.; Caval K.L. (2015). Guide to Cone Penetration Testing for Geotechnical Engineering – 6th Edition. Signal Hill, CA: Gregg Drilling & Testing, Inc.

Stals, M.T. (2023). Advanced constitutive model parameter determination, optimisation and selection using a database of triaxial tests and machine learning tools. MSc thesis. Delft: Delft University of Technology.

Stals, M.T.; Siderius, K.; Brinkgreve R.B.J. (2024). Advanced constitutive model parameter determination, optimisation and selection using a database of triaxial tests and machine learning tools. In: Eur. Conf. on Soil Mech. and Geotech. Eng., Lisbon (to be published).

Van Berkom, I.E.; Brinkgreve R.B.J.; Lengkeek H.J.; De Jong A.K. (2022). An automated system to determine constitutive model parameters from in-situ tests. In: Rahman & Jaksa (eds), 20th Int. Conf. on Soil Mechanics and Geotechnical Engineering. Sydney: Australian Geomechanics Society, 329–334.