Proceedings of the XVIII ECSMGE 2024
GEOTECHNICAL ENGINEERING CHALLENGES
TO MEET CURRENT AND EMERGING NEEDS OF SOCIETY

ISBN 978-1-032-54816-6
DOI 10.1201/9781003431749-472

ECSMGE 24
XVIII EUROPEAN CONFERENCE ON SOIL MECHANICS AND GEOTECHNICAL ENGINEERING

Application of Machine Learning algorithms to drilling data for deep foundations and soil treatments

Application d'algorithmes d'apprentissage automatique aux données de forage pour les fondations profondes et les traitements de sol

E. Martínez García*
Menard España/Departamento de Ingeniería Civil: Construcción, E.T.S de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Madrid, Spain

M. García Alberti
Departamento de Ingeniería Civil: Construcción, E.T.S de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Madrid, Spain

A.A. Arcos Álvarez
Departamento de Ingeniería y Morfología del Terreno, E.T.S de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Spain

**emartinez@menard.es*

ABSTRACT: Modern machinery for executing deep foundations and soil treatments is equipped with sensors that allow the measurement of execution parameters such as torque, thrust, and speed (Measurement While Drilling, MWD). These parameters are related to soil characteristics and, by using machine learning (ML) algorithms, it is possible to translate them into a penetrometric soil profile. Therefore, each perforation can provide analogous information of that obtained from a penetrometer. However, in order to apply these algorithms in the specific case of pile drilling and soil treatment, several factors must be considered in both ML techniques and specific to geotechnics. These factors include the presence of anomalous values in the data, the distance between the penetrometers and the drilling points or the selection of the most appropriate algorithms, among others. This paper proposes a specific methodology that outlines the various steps required in order to apply ML algorithms to drilling data for deep foundations and ground treatments.

RÉSUMÉ: Les machines modernes pour l'exécution de fondations profondes et de traitements du sol sont équipées de capteurs qui permettent de mesurer les paramètres d'exécution tels que le couple, la poussée et la vitesse (Measurement While Drilling, MWD). Ces paramètres sont liés aux caractéristiques du sol et, en utilisant des algorithmes d'apprentissage automatique (ML), il est possible de les traduire en un profil penétrométrique du sol. Ainsi, chaque perforation peut fournir une information équivalente à celle d'un pénétromètre. Cependant, afin d'appliquer ces algorithmes dans le cas particulier du forage de pieux et du traitement des sols, plusieurs facteurs doivent être considérés à la fois généraux à ML et spécifiques à la géotechnique. Ces facteurs incluent la présence de valeurs anormales dans les données, la distance entre les pénétromètres et les points de forage, les algorithmes les plus appropriés, entre autres. Cet article propose une méthodologie spécifique qui décrit les différentes étapes nécessaires pour appliquer les algorithmes ML aux données de forage pour les fondations profondes et les traitements de sol.

Keywords: Machine learning; measurement while drilling; penetrometer; rigid inclusions.

1 INTRODUCTION

Machine Learning (ML) can be defined as the extraction of information from large data sets by using mathematical algorithms applied by computers. One of the advantages of this approach is its ability to reveal hidden relationships between variables that are difficult to observe otherwise.

Geotechnical engineering is a field characterized by complex and often empirical relationships between variables, making it an ideal domain for the application of ML. Consequently, ML has gained increasing attention in recent years, leading to the establishment of a dedicated Technical Committee, TC309, by the International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE, 2022).

There are numerous applications of ML in geoscience, including soil and rock classification, soil stratification, settlement analysis, bearing capacity estimation, slope stability assessment, dam monitoring, and more (Liu and Lacasse, 2022).

The objective of this paper is to show how to apply ML to correlate drilling data from deep foundations and soil treatments with the values of penetrometers.

2 PENETROMETERS

The penetrometer is a soil investigation technique that measures the resistance to perforation by using a standardized cone. There are two main types of penetrometers: static and dynamic (CEN, 2008; ISO/TC 182 Geotechnics, 2012). An example of the output of a penetrometer can be seen in Figure 1.

One advantage of penetrometers is their ability to provide continuous information along their length. As a result, they are well-suited for correlation with drilling parameters.

3 MEASUREMENT WHILE DRILLING

Measurement While Drilling (MWD) refers to the continuous collection of data during the drilling process. Modern geotechnical machinery and equipment are equipped with ample instrumentation that enables the measurement of various drilling parameters, including drilling speed, rotary torque, rotary speed, and applied crowd.

These parameters should be correlated with the characteristics of the soil being traversed, and several authors have developed compound parameters to establish these relationships (Laudanski et al, 2012). There is a standard specific to MWD (CEN, 2016) that primarily focuses on boreholes but can also be applied to geotechnical equipment.

Instead of relying on compound parameters, an alternative approach is to utilize ML to correlate the drilling feedback at a certain depth with the previous penetrometer value at the same level.

4 DATA FORMAT

The information obtained from penetrometers and drilling can be provided in various formats or, in some cases, without any specific format. Therefore, to effectively utilize ML, it is recommended to standardize all the information into a consistent format.

For geotechnical data, a suitable option is to employ the AGS format (Association of Geotechnical & Geoenvironmental Specialists, 2023a). This format can be easily converted and utilized with a dedicated library (Association of Geotechnical & Geoenvironmental Specialists, 2023b).

5 DATA PREPROCESSING

Data preparation and cleaning are crucial steps when using ML algorithms, which include handling missing values. Additionally, it is essential to detect outliers, as this is a common task in all ML problems. However, the criteria for identifying outliers can vary depending on the specific case.

5.1 Penetrometers outlier detection

An approach to identify outliers could be the distance from the mean value. But this is not possible due to the difference in nature of soil layers. A value of 15 blows could be an outlier at depth 4 m, but no at a depth of 10 m (Figure 1).

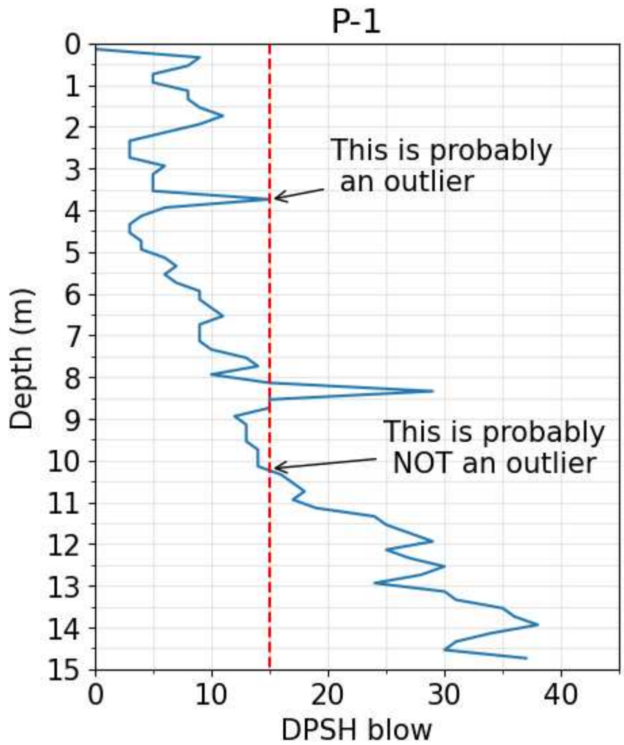

Figure 1. Example of a DPSH penetrometer output.

To assess this issue, a moving average of the penetrometer can be used, with a window that depends on the desired level of smoothness. It is very difficult to know if a strange value in the penetrometers is an outlier (i.e. the presence of some gravels) or a true soil condition. Using a moving average allow to keep the value while softening the profile.

5.2 Drilling data outlier detection

In the case of the drilling data, the detection of outlier becomes more complicated as is a multivariate problem. A high value of one parameter can be normal or not depending on the values of the other parameters.

For this problem, a specific outlier detection, such as isolation forest (Liu et al, 2008), can be used. Isolation forest classifies outliers by how difficult the sample is to isolate, with easier to isolate meaning more probability to be an outlier.

5.3 Data labelling

To correlate the drilling data to the penetrometric profile, a value of it should be assigned to each sample of the data. This could seem easy enough: assign the value from the closer penetrometer as the proper level.

But the closer penetrometer could be quite far. In fact, these correlations are interesting because the geotechnical data are usually scarce. In Figure 2, inclusions 1 and 2 are closer to the penetrometer P1. At inclusion 1 there is a small error, but at inclusion 2 the error is much greater due to distance and changes in the soil.

To address this problem, Bunieski (2022) proposed two different methods. The variability method establishes a maximum distance where the error is considered acceptable. The continuous method uses the mean value of close penetrometers but establishing and information level depending on distance. Points with low information are discarded. Both methods consider how the soil profile changes in the available penetrometers.

6 CHOOSE ML ALGORITHM

The selection of the algorithm to be used is very case dependent. Some of them need some preprocessing of the data for being used (for example a normalization of the values). Others can avoid preprocessing but would benefit from previously applied algorithms. On top on that, the hyperparameters that control the process should be tuned for each problem, greatly affecting the performance.

As can be seen, finding the most adequate approach is no easy task.

One possible solution is to use an auto-ML library, such as TPOT (Olson et al., 2016), which will suggest the best possible pipelines after a series of iterations.

A disadvantage for this approach is that it can be very computationally demanding and time consuming, depending on the data. The number of iterations can be limited, but that will decrease the possibility of reaching the best pipeline.

7 PROPOSED PIPELINE

The proposed pipeline for the use of ML with MWD data is as follows:

- Transform the available information to a consistent data format.
- Data preprocessing.
    - Outlier detection
    - Data labelling (special care to distance)
- Auto-ML and selection of algorithms.
- Evaluation of the obtained results (correlation coefficient, root of mean squared error, etc.)
- Auto-ML and selection of algorithms.
- Evaluation of obtained results (correlation coefficient, root of mean squared error, etc.)

Figure 3 shows the scheme of this process.

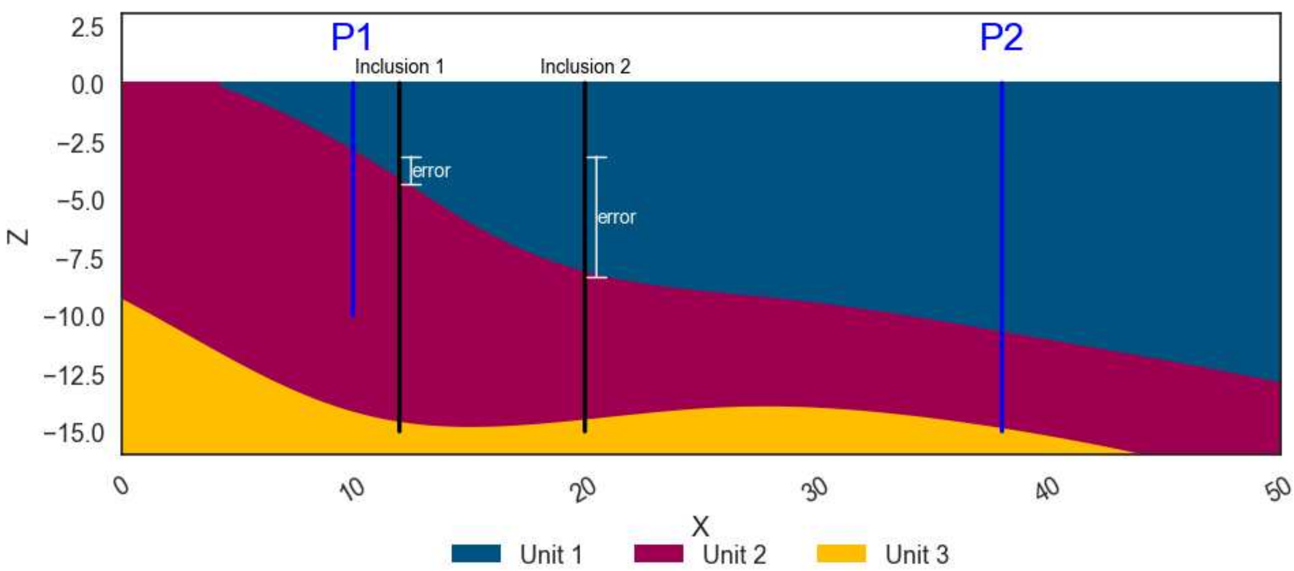

Figure 2. Soil cross section. Error introduced by distance between inclusion and penetrometer.

8 CONCLUSIONS

This paper has highlighted the specific challenges associated with using ML for MWD data, particularly emphasizing the significance of considering the distance between the perforation and penetrometer, as well as the detection of outliers. It is evident that expert knowledge plays a critical role in addressing ML problems, especially in the geotechnical field where it is difficult to account for every variable that may affect the process. But, with the proper pipeline, ML is a very interesting tool to extract information from perforations.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Menard Spain for supplying the data used in this study.

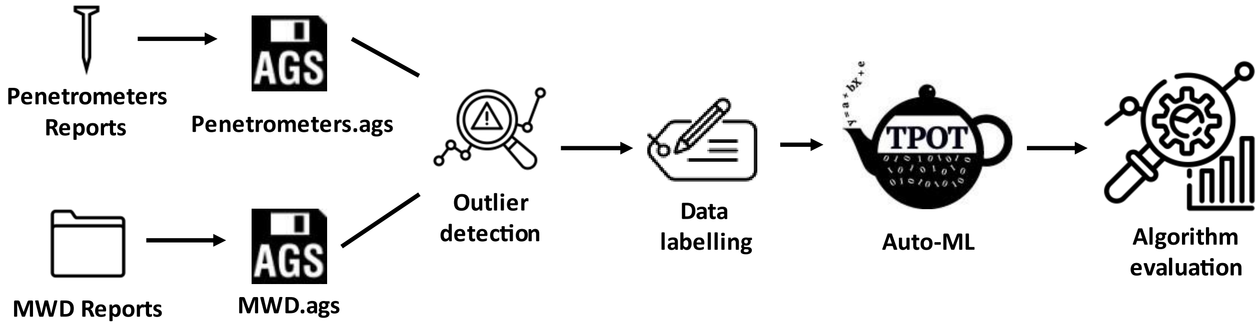

Figure 3. Proposed pipeline.

REFERENCES

Association of Geotechnical & Geoenvironmental Specialists (2023a) *AGS Data Format*. Available at: https://www.ags.org.uk/data-format/.

Association of Geotechnical & Geoenvironmental Specialists (2023b) *AGS Python Library*. Available at: https://gitlab.com/ags-data-format-wg/ags-python-library.

Bunieski, S. (2022) 'Practical methodologies to apply machine learning algorithms to ground improvement data', in.

CEN (2008) 'UNE-EN ISO 22476-2. Investigación y ensayos geotécnicos. Ensayos de campo. Parte 2: Ensayo de penetración dinámica'. AENOR.

CEN (2016) 'UNE-EN ISO 22476-15. Investigación y ensayos geotécnicos. Ensayos de campo. Parte 15: Registro de parámetros de perforación'. AENOR.

ISO/TC 182 Geotechnics (2012) 'Geotechnical investigation and testing. Field testing. Part 1: Electrical cone and piezocone penetration test (ISO 22476-1:2012)'.

ISSMGE (2022) *Reference List for Machine Learning and its Applications in Geotechnical Engineering*. Available at: https://www.issmge.org/news/reference-list-for-machine-learning-and-its-applications-in-geotechnical-engineering-part-i-need-to-know-knowledge0 (Accessed: 25 June 2022).

Laudanski, G., Reiffsteck, J.-L. and Benoît, J. (2012) 'Experimental study of drilling parameters using a test embankment', in. *Geotechnical and Geophysical Site Characterization 4*.

Liu, F.T., Ting, K.M. and Zhou, Z.-H. (2008) 'Isolation Forest', in *2008 Eighth IEEE International Conference on Data Mining*. *2008 Eighth IEEE International Conference on Data Mining (ICDM)*, Pisa, Italy: IEEE, pp. 413–422. Available at: https://doi.org/10.1109/ICDM.2008.17.

Liu, Z. and Lacasse, S. (2022) 'Machine learning in geotechnical engineering: Opportunities and applications', in M. Rahman and Jaksa Mark (eds) *Proceedings of the 20th ICSMGE-State of the Art and Invited Lectures*. Sydney, Australia: Australian Geomechanics Society, pp. 543–558.

Olson, R.S. *et al.* (2016) 'Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science', in *Proceedings of the Genetic and Evolutionary Computation Conference 2016. GECCO '16: Genetic and Evolutionary Computation Conference*, Denver Colorado USA: ACM, pp. 485–492. Available at: https://doi.org/10.1145/2908812.2908918.