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

ISBN 978-1-032-54816-6
DOI 10.1201/9781003431749-43

Machine learning to improve the ground model for road settlement prediction

Apprentissage automatique pour améliorer le modèle de sol pour la prévision des tassements routiers

M. Brček*
Department of Geotechnics, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Slovakia

J. Panuška
PanGEO s.r.o., Zvolen, Slovakia

M. Kopecký
Department of Geotechnics, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Slovakia

**martin.brcek@stuba.sk*

ABSTRACT: The article discuss data processing and development of predictive ground model on the basis of data from extensive engineering geological survey. Basic model inputs are spatial coordinates (X, Y, elevation). Ground model with assigned soil groups, deformation properties and ground water level is created with input of spatial coordinates. Basic statistical and machine learning algorithms (regression, classification, clustering) were used in Python script to develop predictive ground model. Predictions are compared with the ground model developed by engineering geologist. ground model was used to calculate the settlement of the embankment subgrade in the highway route D1 for international and domestic transport.

RÉSUMÉ: L'article traite du traitement des données et du développement d'un modèle prédictif du sol sur la base de données provenant d'une étude géologique approfondie. Les entrées de base du modèle sont les coordonnées spatiales (X, Y, élévation). Le modèle de sol avec les groupes de sols assignés, les propriétés de déformation et le niveau de la nappe phréatique est créé à partir des coordonnées spatiales. Des algorithmes statistiques de base et d'apprentissage automatique (régression, classification, regroupement) ont été utilisés dans un script Python pour développer un modèle de sol prédictif. Les prédictions sont comparées au modèle de sol développé par l'ingénieur géologue. Les résultats montrent une grande précision de prédiction avec une erreur acceptable, mais avec une plus grande efficacité si l'apprentissage automatique est utilisé. Les profils géologiques d'ingénierie obtenus ont été utilisés pour calculer le tassement de la plate-forme de remblai sur le tracé de l'autoroute.

Keywords: Ground model; data processing; settlement; machine learning.

1 INTRODUCTION

Application of machine learning algorithms is on the rise in geotechnical engineering. This is due to the huge amount of data produced (laboratory measurements, geological survey, in-situ testing etc.) and due to empirical nature of the profession. Machine learning was used to predict small strain shear modulus (Cruz, Santos, & Cruz, 2013), predict slope movements (Yang, Yin, & Liu, 2019) and liquefaction potential (Goh & Goh, 2007), (Samui & Sitharam, 2011). Additionally, there are articles that discuss the classification of subsoil or the mapping of some geological features (Šapina, 2016), (Zhou & Wu, 1994), (Tsiaousi, Travasarou, Drosos, Ugalde, & Chacko, 2018). The development of ground models through machine learning is somewhat comparable with geostatistical algorithms such as Kriging (Cressie, 1990).

In this study, machine learning is utilized to develop ground model for calculating settlement in a large infrastructural project (highway of length 15 km). The automatically generated ground model for settlement analysis must include deformation properties (oedometric modulus E_{oed}) of various soil groups. Oedometric moduli were assigned to different soil groups based on soil type and its physical state (consistency or density). Settlement calculations were automatically performed in 111 ground profiles obtained solely by inputting spatial coordinates X, Y, Z.

2 METHODOLOGY

A dedicated application was developed to process and analyse the available data and create an engineering geological model with the settlement calculation. This application incorporates a suite of interconnected machine learning and neural network algorithms.

2.1 Material

Engineering geological data from bore logs are crucial for learning phase of model development. Following data were used for analysis:

- Text documentation of realized and archive boreholes
- Spatial coordinates of boreholes (*X, Y*) and their elevation *Z* (m a.s.l.)
- Ground water level
- Atterberg limits, grain size distribution curves, oedometric moduli

Visual Basic for Applications was used to extract:

- Borehole name
- Soil class USCS (American Society for Testing and Materials, 2017)
- Depth boundaries of single layer
- Soil consistency (set as lower consistency) → description (keyword) “soft to stiff “ or “stiff to soft“ → resultant consistency soft with symbol T
- Soil was set to organic (symbol O) in case of organic matter (keywords: detrit, rotten wood, wood, plants)
- Homogenized soil group f [CH-CV-CE, MH-MV-ME, CE, ML-MI], fs [CS, MS, SC, SM], fg [CG, MG, GC, GM], g [GF, GP, GW], s [SF, SP, SW], r [soft rocks], OR, Y
- origin/genesis of soil: quaternary sediment D (deluvial), F (fluvial) and Neogene sediments N (Figure 1)
- spatial coordinates of borehole X, Y, Z
- ground water level

A total of 444 boreholes were processed with 3558 soil layers. Laboratory tests, Atterberg limits (liquid and plasticity limit), percentage of fines, and grain size distribution curves were used to develop quasi homogenous soil groups. In total, 1071 samples were processed, with 109 oedometric moduli.

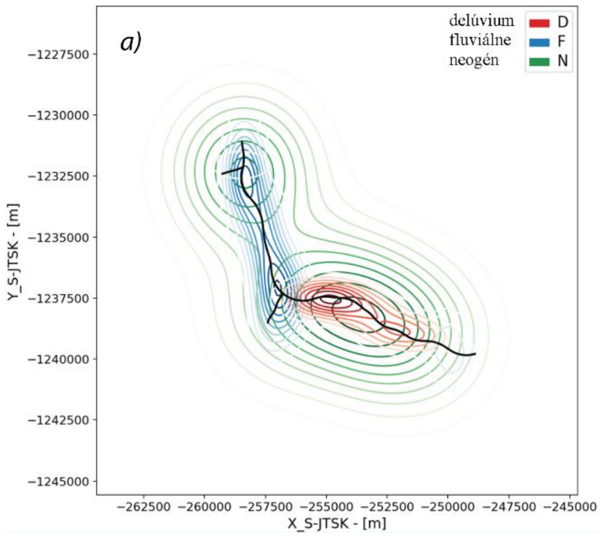

Figure 1. Distribution of soils in the highway route D1.

2.2 Methods of machine learning

Methods of supervised learning and unsupervised learning were employed to predict ground model including ground water level. The K-means clustering (KMC) algorithm was utilized to identify critical quasi-homogeneous sections along the highway route. The algorithm's inputs consist of the spatial coordinates of terrain surface & elevation and oedometric modulus. An elbow diagram was used to determine the optimal number of clusters, which was found to be 9 in this study.

The accuracy of the model is calculated as follows:

$$ACCURACY = TP/(TP+FP) \quad (1)$$

where *TP* is number of correctly assigned data points, *FP* is number of misclassified points.

2.3 Prediction of ground water level

Ground water level was predicted by deep feed forward artificial neural network (ANN). Algorithm inputs are spatial coordinates of terrain surface & elevation. For training, 75% of the total 405 groundwater level data points were used, while 25% were used for model validation.

2.4 Prediction of soil genesis

The Random Forest (RF) algorithm was trained for the classification task. Inputs for classification included the elevation, and the depth of the center of the homogeneous layer from the surface, in addition to spatial coordinates. Each borehole was subdivided into elements with thickness of 0.1 m (oversampling). The model consists of 10 decision trees, with equal weights for single classes and entropy as the optimization criterion. The number of data points for each genesis are shown in Table 1. Overall, 65% of the data were used for learning, and 35% for testing. The accuracy of the model (Table 1) shown very good prediction of soil genesis.

Table 1. Number of inputs for model with soil genesis.

Soil genesis	km 0.0-6.0	Accuracy of model (%)
F - fluvial	8700	98.92
N-neogene	3860	96.29
D-deluvium	590	97.61
Overall	*13 150*	*98.09*

A different procedure was utilized to predict soil consistency. The Support Vector Machine (SVM) algorithm with bootstrap aggregation and One vs. One classification strategy was employed. Input parameters consisted of the same parameters as in previous models, supplemented by inputs for genesis and soil consistency. Furthermore, no subdivision was applied to soil layers (no oversampling of elements with thickness of 0.1 m). A regularization parameter C=50, with equal weights for classes, and 200 sub-models were used

Oedometric modulus and unit weights were assigned to soils based on their consistency, soil group, genesis, and highway part A – F.

3 RESULTS

3.1 Identification of homogenous units on highway route and prediction of groundwater level

The highway route, with 9 classes identified by K-means clustering (KMC) and resultant 6 quasi homogenous parts are depicted in Figure 2. These six quasi-homogeneous parts, labeled A – F, were identified across two sections of the highway route. The first part, A – C, predominantly consists of fluvial sediments, while the second part, D – F, is mostly characterized by deluvial sediments.

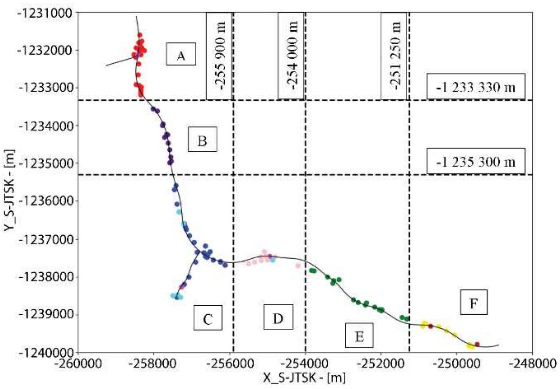

Figure 2. Result of KMC clustering algorithm.

Only results from part one are presented in this paper.

The maximum error between predicted and measured water level is 7.24 m, with an average error of 1.99 m.

3.2 Comparison of ground models of homogenous units on highway route

Differences in depths of observed parameters between machine created ground models (ML) and ground models developed by engineering geologist (EG) are shown in Figure 3. Differences were found in determining the depth of the surface cover (f/fs), thickness of gravelly (gZ, gK) sediments, depositional depth of the Neogene sediments (N), and groundwater level (HPV). The depth difference in estimating soil layers or groundwater levels is generally up to 2.0 meters. The largest error, 3.5 m can be found in the estimation of Neogene depth in profile km 13.550. The errors in profile km 8.475 are interesting due to the fact that no learning points were provided in this section (there are only cuts in this part and no embankments, thus no need for settlement calculation). The model was not trained within a radius of 1.0 km from this point; however, the error is only slightly above 2.0 meters. The ground models are developed for settlement analysis, thus, the comparison of settlement in chosen profiles is of interest.

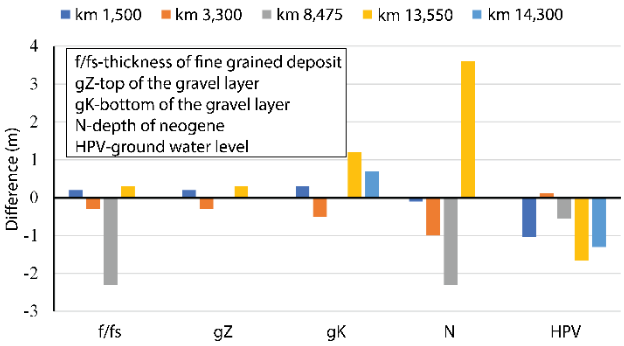

Figure 3. Comparison of profiles created by ML algorithm and human made (EG).

3.3 Comparison of measured and predicted settlements

Calculated settlements for ground models developed by machine and by hand are summarized in Table 2. A difference of up to 10% is considered very good and appears promising for the analysis. The comparison of predicted and measured settlements (by horizontal inclinometers) is presented in Table 2 and Figure 4. The maximum difference is slightly above 20%. It must be noted that such difference could be caused not only by the misprediction of ground model but also by the use of a simple analytical calculation model for ground settlement analysis. Finally, a difference of 20% is considered acceptable, particularly if the model is intended for preliminary quantitative analysis.

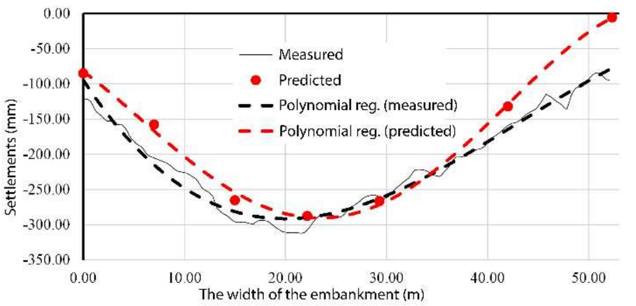

Figure 4. Comparison of calculated and measured settlements, HINK-17, km 6.416.

Bigger differences are only observed at the endpoints, where the measurements are affected by the design of the end of the guide casing embedded and the works in its vicinity.

Table 2. Comparison of calculated settlements for ground models developed by machine (ML) with Horizontal Inclinometer Measurements (HINK).

	HINK	**ML**
Section	Settlements (mm)	
km 1.500	106.14	126.72
km 2.081	123.44	113.20
km 6.416	292.64	287.48
km 6.653	270.96	257.58

4 CONCLUSIONS

This paper was focused on a real case of application of artificial intelligence in construction practice.

To determine ground settlement, it was necessary to create a spatial engineering geological ground model of the motorway section D1 Budimír-Bidovce, which included information about groundwater, soil horizons, and their geotechnical characteristics.

The presented predictive model can deliver robust and fast ground model development used for quantitative analysis of soil conditions at given site. The effectiveness of used approach was demonstrated on example of settlement analysis. Calculation of 111 settlement profiles takes only tens of minutes. The increase in robustness of the presented approach is enhanced by combining it with Building Information Modeling (BIM) and Geographic Information System (GIS) tools.

ACKNOWLEDGEMENTS

The authors are grateful for the financial support provided by Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic. The project presented in this article is supported by grant project VEGA-1/0745/21.

REFERENCES

American Society for Testing and Materials. (2017). ASTM D2487-17. Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System). West Conshohocken, PA, 2017: ASTM International.

Cressie, N. (1990). The Origins of Kriging. Mathematical Geology, 239-252. https://doi.org/10.1007/BF00889887

Cruz, M., Santos, J. M., & Cruz, N. (2013). Estimating the Maximum Shear Modulus with Neural Networks. Recent Trends in Applied Artificial Intelligence, Amsterdam, Netherlands, 684-693.

Goh, A., & Goh, S. (2007). Support vector machines: Thier use in geotechnical engineering as illsutrated using seismic liquefaction data. Computers and Geotechnics, 34, 410-421. https://doi.org/10.1016/j.compgeo.2007.06.001

Samui, P., & Sitharam, T. (2011). Machine learning modelling for predicting soil liquefaction susceptibility. Natural Hazards and Earth System Sciences, 11, 1-9. https://doi.org/10.5194/nhess-11-1-2011

Šapina, M. (2016). A Comparison of Artificial Neural Networks and Ordinary Kriging depth maps of the Lower and Upper Pannonian stage border in the Bjelovar Subdepression, Northern Croatia. The Mining - Geology - Petroleum Engineering Bulletin, 75-86.

Tsiaousi, D., Travasarou, T., Drosos, V., Ugalde, J., & Chacko, J. (2018). Machine Learning Applications for Site Characterization Based on CPT Data. Geotechnical Earthquake Engineering and Soil Dynamics, COnference, Austin, Texas, V, 461-472.

Yang, B., Yin, K., Lacasse, S. & Liu, Z. (2019). Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides, 16 (4), pp. 677-694. DOI 10.1007/s10346-018-01127-x

Zhou, Y., & Wu, X. (1994). Use of Neural Networks in the Analysis and Interpretation of Site Investigation Data. Computers and Geotechnics, 16 (2), 105-122. https://doi.org/10.1016/0266-352X(94)90017-5