Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)

Application of machine learning to the identification of quick and highly sensitive clays from cone penetration tests

Cristian GODOY, Ivan DEPINA, Vikas THAKUR

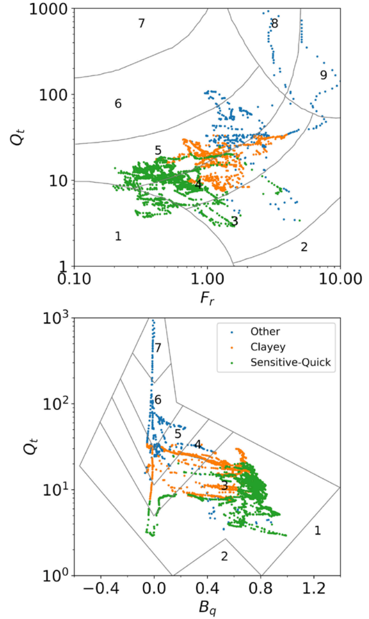
Cite this as: Cristian Godoy, Ivan Depina, Vikas Thakur, 2020. Application of machine learning to the identification of quick and highly sensitive clays from cone penetration tests. *Journal of Zhejiang University-SCIENCE A (Applied Physics & Engineering)*, 21(6):445-461. https://doi.org/10.1631/jzus.A1900556

Introduction

- The presence of highly sensitive and quick clays is one of the primary geotechnical concerns in Norway.

- The use of CPTu tests for soil classification is a common practice, mostly through chart classifications. However, the classification of non-textbook soils, as in the case of quick or highly sensitive clays, is a major challenge.

- This study investigates the potential of machine learning techniques to improve the identification of highly sensitive and quick clay soils using CPTu.



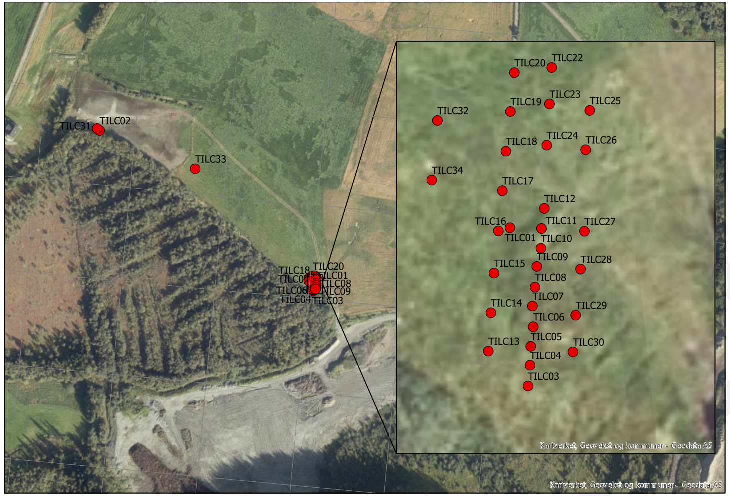
Robertson (1990) Classification Chart and the classification of highly sensitive and quick clays (green dots)

Methodology

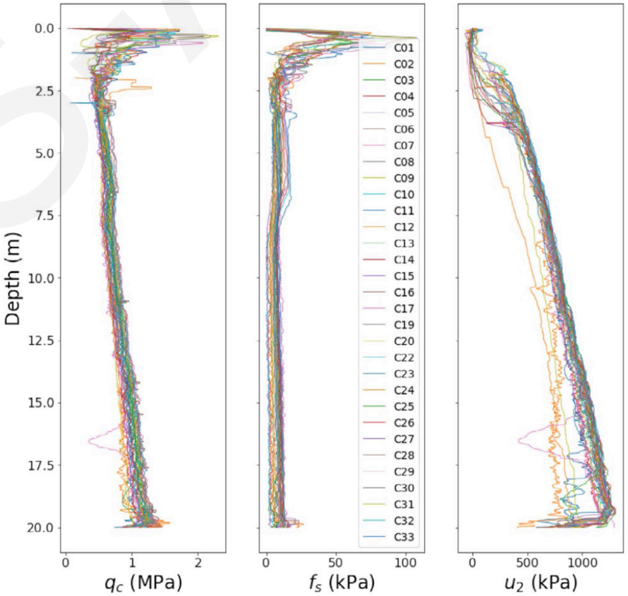
- The methodology followed considers the analysis of two CPTu datasets wherein highly sensitive and quick clays were encountered and wherein the layering at each test location is known.
- The CPTu data were then used to classify the samples using well-known classification charts and machine learning methods.
- Finally, the results were compared against the actual layering and performance measurements were computed to compare the different approaches.

Datasets

- NGTS Site (Trondheim, Norway)



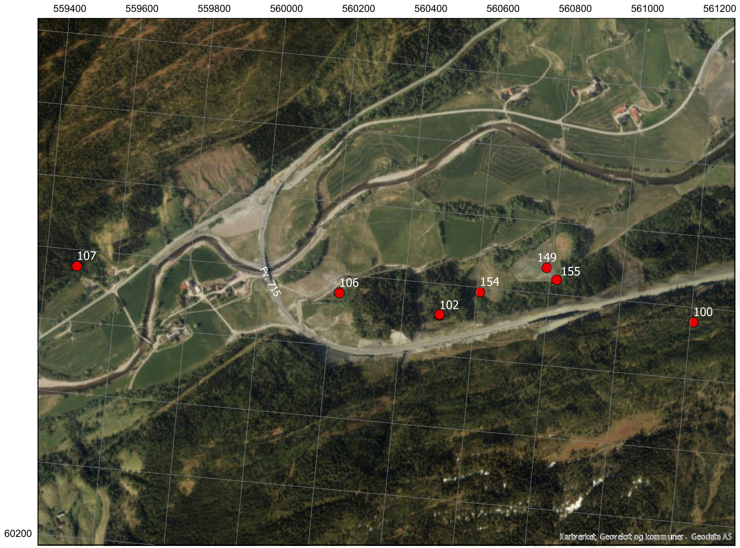
Layout of the CPTu tests at the NGTS site in Norway



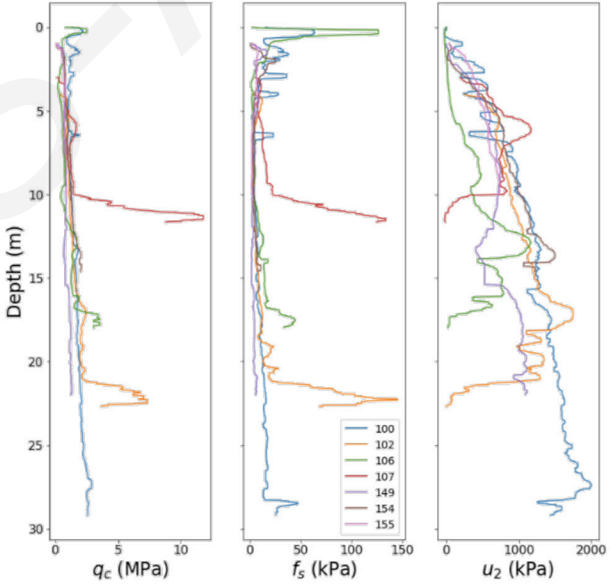
Summary of the CPTu tests results at NGTS

Datasets

- Vegvesen Site (County Road 715, Norway)



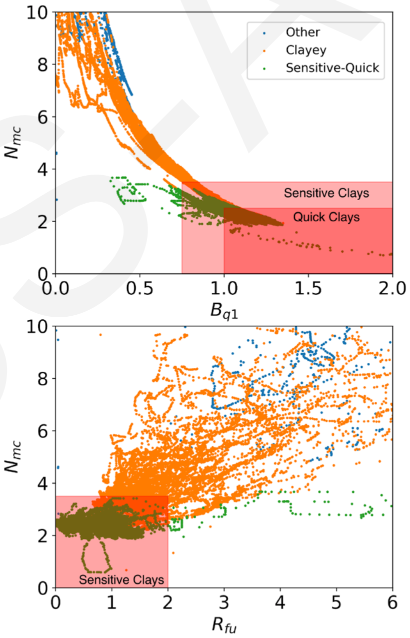
Layout of the CPTu tests at the Vegvesen site in Norway



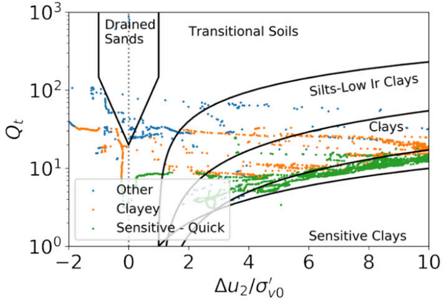
Summary of the CPTu tests results at the Vegvesen site

CPTu Chart Classification

- The chart classifications used were the ones proposed by Robertson (1990 and 2016), Eslami and Fellenius (1997), Schneider et al. (2008) and Gylland et al. (2017).
- The accuracy score of the classification ranged between 27% and 87%.
- Gylland classification gave the highest accuracy for NGTS dataset (87%), while Schneider gave the best results on Vegvesen dataset (75%).



Gylland classification for NGTS dataset, accuracy score = 87%

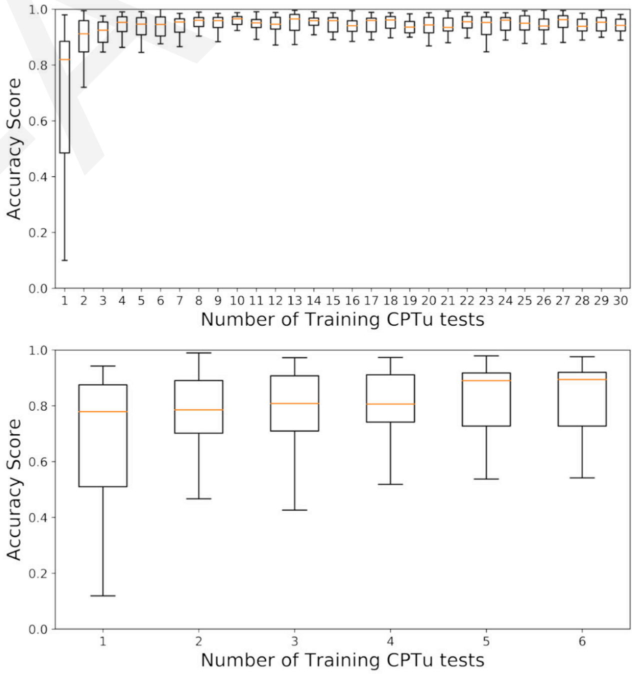


Schneider classification for Vevesen dataset, accuracy score = 75%

NTNU

Machine Learning Classification

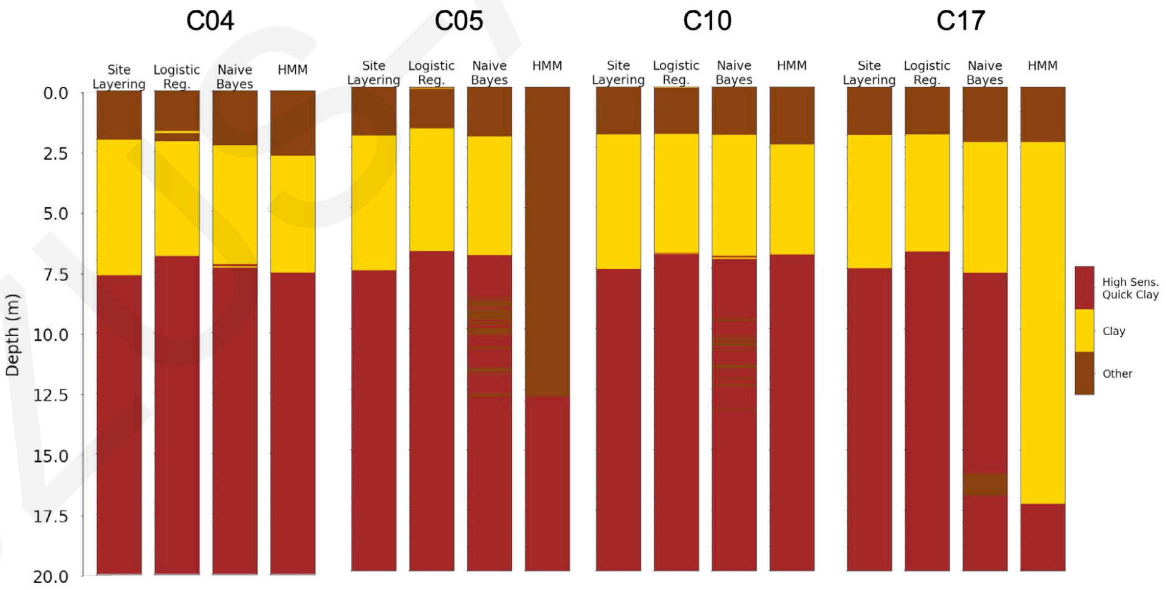
- The machine learning algorithms for the classification used in this work were logistic regression, naive Bayes, and a hidden Markov model (HMM)
- The model's performance was measured through the accuracy score.
- The training of the algorithms was done in a sequential manner, ie, the training dataset size is increased incrementally, in order to check how many CPTu tests are needed to be able to predict properly. This is done randomly and repeated several times to capture the variability of the process.



Accuracy scores for the naive Bayes classification of the NGTS (up) and Vegvesen (down) datasets, using a sequential training

Results for the NGTS dataset

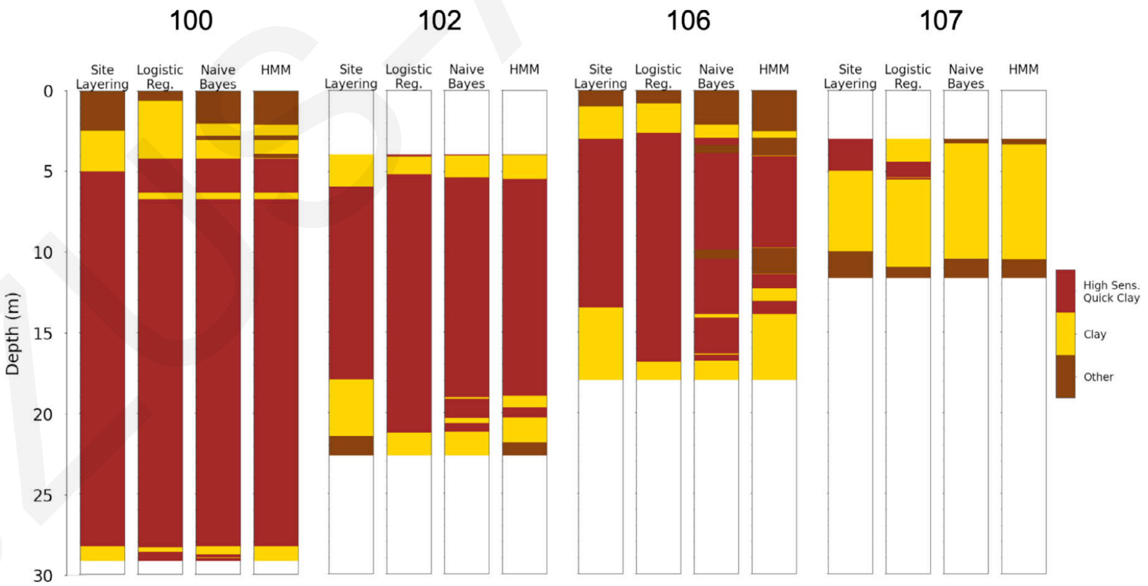
- Accuracy Score above 96% for all three algorithms. Being naive Bayes the one with the best performance (97%).
- The site profiles estimated using the machine learning methods show a good agreement with the site layering.



Estimation of site layering with different machine learning methods at the NGTS site

Results for the Vegvesen dataset

- Accuracy Score above 79% for all three algorithms. Again being naive Bayes the one with the best performance (91%).
- The site profiles estimated using the machine learning methods show a good agreement with the site layering, considering the variability of the soil layering.



Estimation of site layering with different machine learning methods at the Vegvesen site

Conclusions

- While the accuracy score of chart classifications ranged between 27% and 87%, the machine learning algorithms showed a significative increase in the performance of the classification with accuracy scores above 90% for NGTS site (homogeneous soil layering) and above 80% for Vegvesen (variable soil layering). The best results were obtained with the naive Bayes classificator.

- The results demonstrated the abilities of these methods to learn from the data, with high classification accuracies reached after only 3-4 training CPTus.

- The major challenge with this approach is obtaining enough data to train the models as well as enough laboratory results to verify them.