Automated CPT interpretation with a Convolutional Neural Network

Authors
Martina Pippi, Silvia Bersan, Ritchie Vink, Jacco Haasnoot

Affiliations
CEMS Crux Engineering MicroServices BV

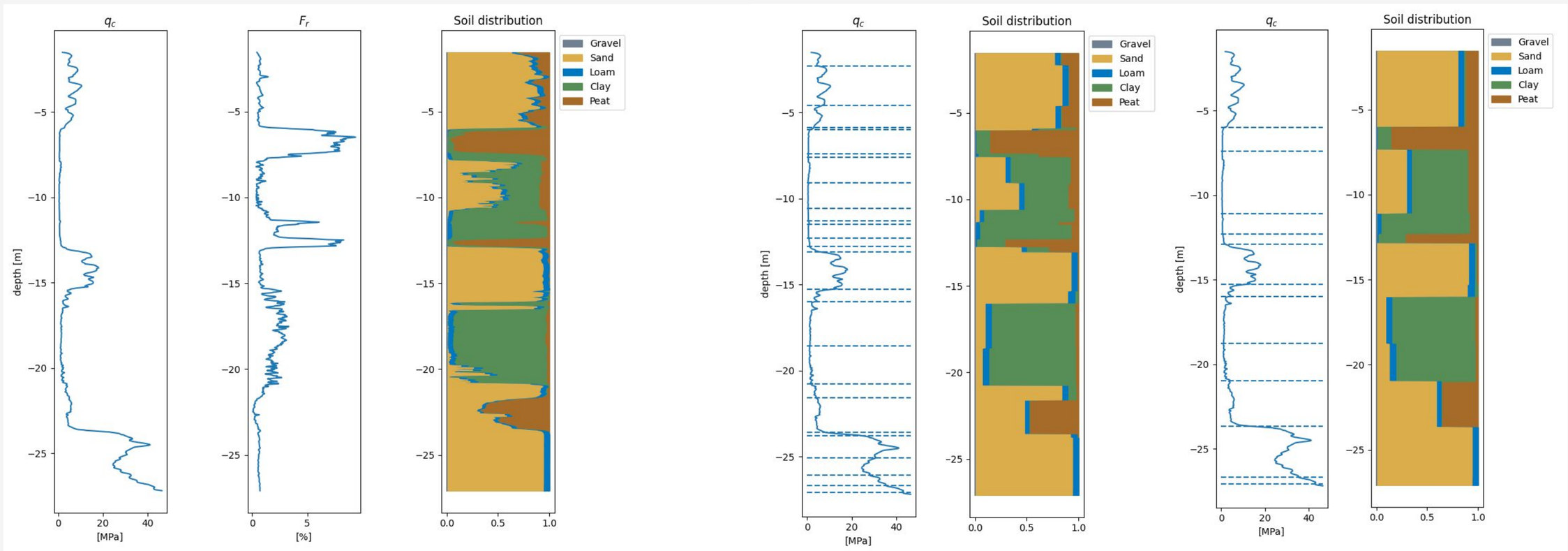

Figure 1: Cpt classification with different penalty parameter: 0, 1, 3

Introduction

Cone Penetration Tests are widely used in the Netherlands,as the amount of performed CPTs increases, an automated soil interpretation becomes more and more relevant. Attempts to automate soil classification have been done in the past, but the empirical formulas commonly used do not always provide a satisfactory interpretation for engineering purposes. Besides that the soil type is often not interpreted correctly, there is also the problem that the classification is provided for each measurement and no strategy is given to aggregate those tiny layers. This research shows how a data driven approach can yield better results than the traditional empirical methods.

The CEMS model

In order to train the model roughly 49000 CPTs and 40000 boreholes were checked, from which 1800 pairs met the condition of being less than 6 meters apart. These have been used as labeled data for the first model training. The model is retrained periodically whenever new data are available that meet the condition of being not more than 6 metres apart.

An essential part of the classification is based on the location where the CPT is taken. The model is enhanced with location-based embedding, in this way the model could learn its own location embedding and could learn the probabilities of soil type conditional on a certain location. Furthermore, most of the bore-hole data show that layers consist of multiple soil components in variable percentages. Therefore we should predict the total soil distribution per layer.

In Figure 2 the locations of the clusters are indicated on the map of the Netherlands. The colors represent a similarity measure between the clusters based on the cosine similarity.

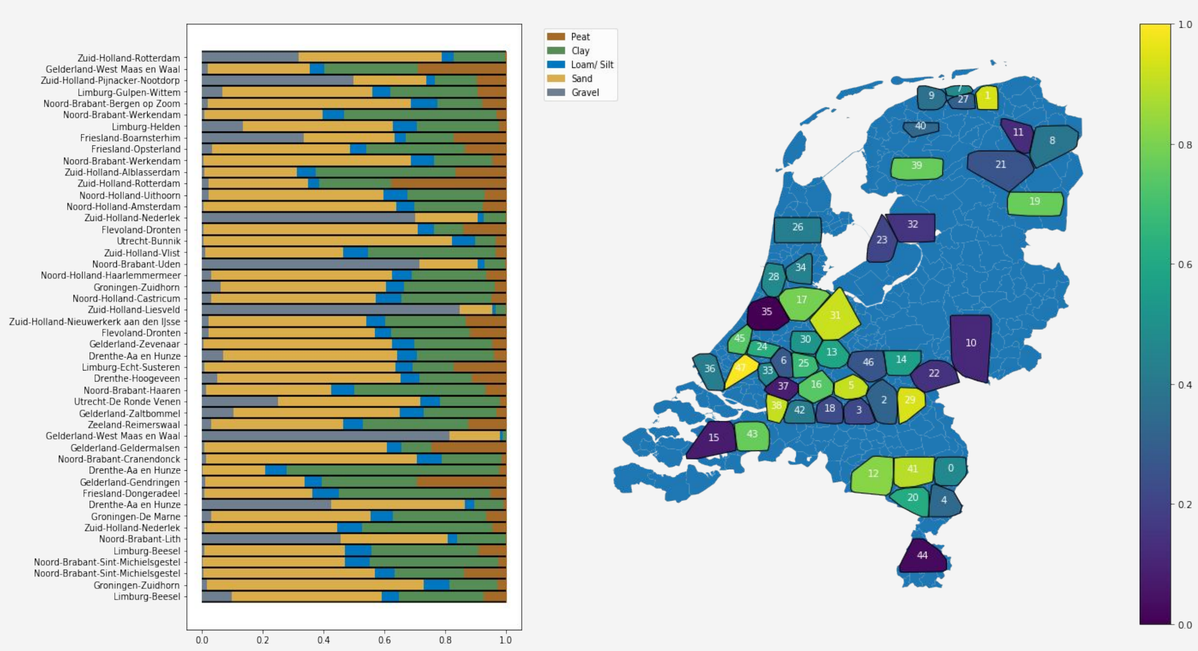

Figure 2:Biases per location and location clusters

The grouping algorithm

The output of the model is a prediction given for each measurement of the CPT. Although this is probably the most accurate representation of the soil stratigraphy, it is impossible to use 1000+ layers in the traditional software, hence we need to group these tiny layers into macro layers. This is done using an algorithm that is based on considering the CPT measurement as a signal.

Conclusions

A CPT interpretation based on data proved to be better than empirical formulations. The gain in terms of time when using it in an automated process for geotechnical design is considerable. The model has been trained only on Dutch soil data. Even though it would probably still give a reasonable prediction outside of the Netherlands, it cannot apply any biases for the location. Future development could include the expansion of the geographical limits of the model, by including CPTs and boreholes from other countries and the definition of location biases and clusters based on expert knowledge.

Do you want to try it?

The model can be accessed with a HTTPS call via a web REST API, if you are using the language Python go to https://github.com/cemsbv/cpt-model-teaser, launch the binder and follow the tutorial in the jupyter notebook to learn how to use it.