EGU26-4956, updated on 18 Nov 2026
https://doi.org/10.5194/egusphere-egu26-4956
EGU General Assembly 2026
© Author(s) 2026. This work is distributed under
the Creative Commons Attribution 4.0 License.

Electromagnetic & Cone Penetration Test Data Fusion on Soil Characterization

Dimitrios Madelis[1,2], Marios Karaoulis[2], and Philippe De Smedt[1]
[1]Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium (Dimitrios.Madelis@ugent.be, Philippe.DeSmedt@UGent.be)
[2]Faculty of Science - School of Geology, Aristotle University of Thessaloniki, Thessaloniki, Greece (mkaraoulis@geo.auth.gr)

Defining subsurface soil conditions in complex coastal settings requires the use of both geophysical and geotechnical datasets, each with different resolution and sensitivity. This study combined helicopter-borne electromagnetic (HEM) data, where large areas are spatially covered with limitations to vertical resolution, with cone penetration test (CPT) data, where high resolution can be achieved while the spatial resolution often is very sparse due to drilling associated costs. To formulate a continuous three-dimensional model of subsurface soil properties for levee risk assessment, these datasets were integrated. HEM data provides extensive covering resistivity profiles, while CPT provides high resolution, spatially limited measurements of mechanical soil behaviour.
It is known that resistivity as a soil property depends on many parameters (mostly water quality and soil type), and there is no straightforward method to directly translate it to soil, hence the use of ML. To deal with these complexities, we employed machine learning methods – Random Forests and neural networks – to merge heterogeneous datasets and predict continuous soil behaviour indices and discrete lithological types. We propose the use of multiple features, such as spatial coordinates, depths, distance from coast, soil types and local geological conditions. After pre-processing, machine-learning models were trained to fuse the datasets to ensure spatial consistency in the coastal environment. Afterwards, the Soil Behaviour Type Index (SBT) (Robertson, 1990) was calculated using the CPT measurements and then was discretized into lithological units.
A classical machine learning algorithm (Random Forest) and a PyTorch-based neural network were trained for regression (predicting the continuous SBT index) and classification (predicting soil types) tasks, and their performance was evaluated using standard statistical and visual metrics. Final models were retrained on the full dataset to increase generalizability and robustness. The final product is to map □□□□ values and lithological classes at every HEM point and ultimately to make a 3D subsurface soil model. The outcome for each process was validated against an 80%-20% test to ensure reasonable results.
While regression models had similar RMSE scores, classification models generally produced models with greater accuracy of dominant soil types but captured fewer underrepresented mixed lithologies. This work focuses on the interpretability of soil models through integrating data (i.e., not just purely statistical but spatial output) and ultimately continuity in the spatial domain (where

engineers are most concerned). The goal of this study is to develop a framework where continuous geophysical data, collected either by helicopters or drones can be combined with additional geological boreholes and CPTs and other geotechnical information, to enable us to image the subsurface beyond resistivity. One of the products of this study serves to represent an approach to providing a better product to those grappling with levee design and safety.