

PREDICTION OF SHEAR MODULUS FOR CLAYEY SOILS USING ENSEMBLE MACHINE LEARNING METHOD

Borderon Julien¹ Julie Régnier² and Nathalie Dufour³

¹ GeoCoD, Cerema, 30 Rue Albert Einstein, 13290, Aix-en-Provence, France
E-mail: j.borderon@cerema.fr

² Geoazur, University of Nice/Cerema, 500 route des Lucioles 06309, France
E-mail: julie.regnier@cerema.fr

³ GeoCoD, Cerema, 30 Rue Albert Einstein, 13290, Aix-en-Provence, France
E-mail: nathalie.dufour@cerema.fr

Key Words: *Seismic hazard, nonlinear soil behavior, Shapley values, ensemble boosting learning*

ABSTRACT

At a specific location the amplification of seismic waves, known as site effects, can occurs within the top few hundred meters of the ground and is influenced by soil geometry, properties, and seismic wave intensity due to nonlinear soil behavior. Predicting these site effects requires defining dynamic soil properties, typically determined through costly laboratory tests which entail uncertainty in their soil property evaluation. Consequently, to predict these properties engineers frequently use empirical models derived from geotechnical databases of dynamic parameters.

Traditionally, parametric regression has been used to predict these soil properties, such as shear modulus reduction curves, but non-parametric approaches offer more accuracy. However, existing models are not available in open access, not for clayey soils and moreover they typically handle data points from each test individually rather than as an ensemble. Here, we propose training a model for clayey soils using several databases. A key challenge arises in database selection, as combining datasets can sometimes reduce model performance. Using Shapley values in our XGBoostRegressor model^[1], we assessed each database's impact on overall performance. Our analysis revealed that the Cincimino^[2] and Facciorusso^[3] databases contributed significantly (Shapley value of 0.32) to model accuracy, compared to Gaudiosi^[4] (0.30), achieving an R^2 of 0.96 when using only the first two databases.

Moreover, we identified liquid limit (LL) and confining pressure (σ) as the most critical features, contrasting with empirical models that prioritize the Plasticity Index (PI). This approach provides a first framework for predicting full shear modulus reduction curves instead of isolated points.

REFERENCES

- [1] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
- [2] Cincimino, A. et al. (2023) The PoliTO–UniRoma1 database of cyclic and dynamic laboratory tests: assessment of empirical predictive models. Bulletin of Earthquake Engineering, pp. 1–33.
- [3] Facciorusso, J. (2020) An archive of data from resonant column and cyclic torsional shear tests performed on Italian clays. Earthquake Spectra, p. 8755293020936692.
- [4] Gaudiosi, I. et al. (2023) Shear modulus reduction and damping ratios curves joined with engineering geological units in Italy. Scientific Data, 10(1), p. 62.