Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GEO-HYBRID: A Physics-Informed Python Framework for Liquefaction Potential Assessment and Integrated Geotechnical Interpretation

Lucas Lima Freitag

houslast@gmail.com

ufrn https://orcid.org/0009-0006-8849-5619

Research Article

Keywords: Geotechnical Engineering, Physics-Informed Machine Learning, Soil Liquefaction, Cone Penetration Test (CPT), GEO-HYBRID, Python Framework, Soil Mechanics

Posted Date: August 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10720584/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

GEO-HYBRID: A Physics-Informed, Calibration-Driven Framework for Geotechnical Prediction — Liquefaction, Consolidation and Shear Strength from Real Case Histories

Lucas Lima Freitag[1]*

[1]Universidade Federal do Rio Grande do Norte (UFRN), Brazil

ORCID: 0009-0006-8849-5619

Abstract

Soils and rocks are among the most uncertain engineering materials: they are discrete, multiphase mixtures whose properties vary unpredictably on a meter-by-meter basis, and a variation of only 2% in water content can change the mechanical response of a slope by more than 10%. Classical deterministic theories (Mohr–Coulomb, Cam–Clay) perform poorly for phenomena such as soil liquefaction, internal erosion (piping) and rainfall-induced landslides, and isolated theoretical predictions routinely exceed 50% error, which is why geotechnical practice depends so heavily on in-situ testing (SPT, CPT) and sensor monitoring. This paper introduces **GEO-HYBRID**, a physics-informed framework that chains validated equations — each with physically meaningful constants — and calibrates them against real case histories *without any opaque multiplicative weight*. The framework is demonstrated on four independent real databases: (i) 251 historical liquefaction case histories from the Boulanger–Idriss (2014/2016) database, with the official report (UCD/CGM-14/01) used to verify column identities and true site CSR values; (ii) 2765 real CPT soundings with independent reference factors of safety from the CLiq/liquepy chain; (iii) 114 oedometer-tested Finnish clays; and (iv) 39 direct-shear Austrian soils. Three new physically justified boundary features — an ultra-dense sand cap, an extreme-magnitude cap, and a documented Loma Prieta manifestation filter — raised liquefaction accuracy from 79.7% to **90.1%** (leave-one-event-out: 89.3%), while a gradient-boosted machine-learning model trained on the same data memorized the training set (92.8%) and collapsed to 76.3% on unseen data. Where the physics is closed, the framework converges to 0.04% against an independent implementation. The paper discusses the efficacy of the method, the irreducible physical scatter of real soils, and the limitations that remain open for future work.

Keywords: Geotechnics; liquefaction; calibration; physics-informed modelling; Boulanger–Idriss; SPT/CPT; machine learning; no opaque weights

*E-mail: houslast@gmail.com

1 Introduction

Geotechnical engineering operates on materials that violate the fundamental assumptions of continuum mechanics. A soil mass is a random packing of discrete mineral grains whose pores are filled with air, water and organic matter; its spatial variability is such that two adjacent boreholes one meter apart can report very different penetration resistances, and its hydraulic state can change dramatically within a single rainy season. Despite this, most design codes still rest on deterministic constitutive models — Mohr–Coulomb, Cam–Clay, Duncan–Chang — whose parameters are treated as well-defined constants [5, 7]. The consequences are well documented: isolated theoretical predictions of liquefaction triggering, internal erosion, and slope stability routinely carry errors that exceed 50%, and even a small perturbation of the water content (e.g., +2%) can reduce the factor of safety of a slope by 10–30% [8, 9].

The empirical response of the profession has been to rely on in-situ tests (Standard Penetration Test — SPT, Cone Penetration Test — CPT) and on monitoring, because the field itself is the only source of truth for a material whose parameters cannot be predicted from theory alone [1, 20]. Semi-empirical methods such as the Seed–Idriss cyclic stress approach [10], the Boulanger–Idriss triggering curves [13–15] and the Robertson soil behaviour type charts [18] represent the state of the art for liquefaction, yet even these methods admit a residual misclassification of 10–15% on their own calibration databases [12].

In parallel, machine learning (ML) has been repeatedly proposed as an alternative [28–31]. ML models can reach very high training accuracy, but their coefficients are opaque, they require large databases that geotechnics rarely has, and — as shown quantitatively in this paper — they overfit small, imbalanced case-history sets, generalizing worse than a well-calibrated physical model on data they have never seen.

This work argues that the dominant source of error in geotechnical prediction is not the equations but the *parameters*: the equations of Vesíc, Terzaghi, Taylor and Boulanger–Idriss, when fed with the true values, reproduce the published curves to within 0–2%. The practical challenge is to (i) estimate parameters from

in-situ tests through validated correlations, (ii) correct for moisture and suction, (iii) quantify the spatial variability explicitly, and (iv) calibrate the remaining physically meaningful constants against real data — never against opaque weights. We operationalize this philosophy in a framework called **GEO-HYBRID** and demonstrate it end-to-end on four real databases.

The main contributions of this paper are:

1. A reusable Python framework that harvests validated equations from the literature, chains them, calibrates physically meaningful constants against real data, and validates with strict out-of-sample protocols (holdout and leave-one-event-out).
2. The first public, machine-readable reconstruction of the Boulanger–Idriss (2014) CPT liquefaction case-history database (253 cases), with column identities verified by exact physical checks against the official report (stress reduction factor, magnitude scaling factor and overburden factor reproduced to three decimals).
3. A corrected reconstruction of the site cyclic stress ratio (CSR), obtained after identifying that the published processed table stores the groundwater table depth and the effective vertical stress rather than the stress ratio — a correction worth +5.6 percentage points of accuracy.
4. Three new physically justified boundary features (ultra-dense sand cap, extreme-magnitude cap, documented Loma Prieta manifestation filter) that take the pure mathematical model from 79.7% to 90.1% accuracy (89.3% in leave-one-event-out).
5. A documented negative result: a permeability-based drainage feature is rejected because the database already encodes drainage effects.
6. An empirical demonstration that XGBoost, trained on the same split, memorizes the training set (92.8%) but is outperformed out-of-sample (76.3%) by the calibrated physical model.
7. Calibrated, physically meaningful correlations for compression index (Finnish clays) and friction angle (Austrian soils), including a grain-size (d50) feature whose sign confirms the physics of inter-particle friction.

2 Background and related work

2.1 Deterministic constitutive models and their limits

The Mohr–Coulomb failure criterion,

$$\tau_f = c' + (\sigma - u)\,\tan\phi', \tag{1}$$

where c' is the effective cohesion, ϕ' the effective friction angle, σ the total normal stress and u the pore pressure, remains the backbone of geotechnical practice [5]. Its extension to unsaturated soils by Fredlund et al. [8] adds a suction term,

$$\tau_f = c' + (\sigma_n - u_a)\tan\phi' + (u_a - u_w)\tan\phi^b, \tag{2}$$

in which $(u_a - u_w)$ is the matric suction and ϕ^b the suction-angle. The critical-state family (Cam–Clay, modified Cam–Clay) improves the description of hardening and volume change [5, 6], but all of these models assume a homogeneous representative element whose parameters are constants. In reality, ϕ' and c' exhibit coefficients of variation of 10–50% within a single site [33, 34], and the suction state fluctuates seasonally. This mismatch between model assumption and field reality is the root cause of the poor isolated-model performance.

2.2 Bearing capacity

Terzaghi's bearing-capacity equation [1], generalized by Vesíc [3],

$$q_{ult} = c'\,N_c\,s_c + q\,N_q\,s_q + \tfrac{1}{2}\gamma B\,N_\gamma s_\gamma, \tag{3}$$

with the dimensionless factors

$$N_q = e^{\pi\tan\phi'}\,\tan^2\!\left(45^\circ + \frac{\phi'}{2}\right) \tag{4}$$

$$N_c = (N_q - 1)\cot\phi' \tag{5}$$

$$N_\gamma = 2(N_q + 1)\tan\phi' \tag{6}$$

and shape factors s_c, s_q, s_γ [3], is dimensionally exact and reproduces the published factor tables to 0.01%, as verified in this work (Section 5). Its error in practice is therefore a parameter error: ϕ' estimated from a correlation or from a few triaxial tests carries ±2–4° of uncertainty, which translates into ±20–50% of bearing capacity.

2.3 Liquefaction triggering: state of the art

The simplified procedure of Seed and Idriss [10] computes the cyclic stress ratio induced by an earthquake as

$$\mathrm{CSR} = 0.65\,\frac{a_{max}}{g}\,\frac{\sigma_v}{\sigma'_v}\,r_d, \tag{7}$$

where a_{max}/g is the peak ground acceleration, σ_v and σ'_v are the total and effective vertical stresses, and r_d is the stress-reduction coefficient. The resistance is expressed through a cyclic resistance ratio $\mathrm{CRR}_{7.5}$ at magnitude 7.5 and unit effective stress. The Seed–Idriss/Youd curve [11, 12] was superseded by the Boulanger–Idriss family [13–15]. For SPT,

$$\mathrm{CRR}_{7.5} = \exp\!\left(\frac{N_{1,60cs}}{14.1} + \left(\frac{N_{1,60cs}}{126}\right)^2 - \left(\frac{N_{1,60cs}}{23.6}\right)^3 + \left(\frac{N_{1,60cs}}{25.4}\right)^4 - c_0\right), \tag{8}$$

and for CPT,

$$\mathrm{CRR}_{7.5} = \exp\!\left(\frac{q_{c1Ncs}}{113} + \left(\frac{q_{c1Ncs}}{1000}\right)^2 - \left(\frac{q_{c1Ncs}}{140}\right)^3 + \left(\frac{q_{c1Ncs}}{137}\right)^4 - c_0\right), \tag{9}$$

with $c_0 = 2.8$ (16th percentile) or 2.6 (median), where q_{c1Ncs} is the clean-sand equivalent normalized cone resistance,

$$q_{c1Ncs} = q_{c1N} + \Delta q_{c1N}, \tag{10}$$

$$\Delta q_{c1N} = \left(11.9 + \frac{q_{c1N}}{14.6}\right)\exp\!\big(1.63 - \eta\big), \tag{11}$$

$$\eta = \frac{9.7}{FC+2} + \left(\frac{15.7}{FC+2}\right)^2. \tag{12}$$

FC being the fines content. The magnitude scaling factor

$$MSF = 1 + \big(MSF_{max} - 1\big)\big(8.64\, e^{-M_w/4} - 1.325\big), \tag{13}$$

$$MSF_{max} = \min\!\left(1.09 + \left(\frac{q_{c1Ncs}}{180}\right)^3,\ 2.2\right), \tag{14}$$

and the overburden factor

$$K_\sigma = 1 - C_\sigma \ln\!\left(\frac{\sigma'_v}{P_a}\right), \qquad C_\sigma = \frac{1}{37.3 - 8.27\, q_{c1Ncs}^{0.264}}, \tag{15}$$

complete the resistance side, and the factor of safety is

$$\mathrm{FS} = \frac{\mathrm{CRR}_{7.5}\, MSF\, K_\sigma}{\mathrm{CSR}}. \tag{16}$$

Robertson and Wride [18] proposed an alternative power-law curve,

$$\mathrm{CRR}_{7.5} = \begin{cases} 93(q_{c1Ncs}/1000)^3 + 0.08 & q_{c1Ncs} < 50 \\ 0.833(q_{c1Ncs}/1000)^3 + 0.05 & q_{c1Ncs} \geq 50 \end{cases} \tag{17}$$

which, as shown in Section 5, performs no better than the trivial "always liquefy" baseline on the same database — an important quantitative comparison rarely reported.

2.4 Consolidation

One-dimensional consolidation [1] is governed by

$$T_v = \frac{c_v\, t}{H^2}, \qquad U(T_v) = \begin{cases} 2\sqrt{T_v/\pi} & T_v \leq 0.1 \\ 1 - \frac{8}{\pi^2}\, e^{-\pi^2 T_v/4} & T_v > 0.1 \end{cases} \tag{18}$$

(Taylor's solution [4]), and the final settlement of a normally consolidated layer is

$$S = \frac{C_c}{1+e_0}\, H\, \log_{10}\!\left(\frac{\sigma'_{v0} + \Delta\sigma'_v}{\sigma'_{v0}}\right), \tag{19}$$

with the compression index commonly estimated by the Terzaghi–Peck correlation $C_c = 0.009\,(w_L - 10)$ [2]. The coefficient of consolidation c_v spans one to two orders of magnitude within a single clay layer [27], which bounds the practical accuracy of any settlement-time prediction.

2.5 Empirical correlations and their scatter

Geotechnical practice converts penetration tests into strength parameters through empirical correlations, e.g., Wolff's [22] ϕ'–N_{60} relation,

$$\phi' = 27.1 + 0.3\, N_{60} - 0.00054\, N_{60}^2, \tag{20}$$

Kulhawy–Mayne's relative density [21] $D_r = \sqrt{N_{1,60}/60}\cdot 100\%$, and plasticity-based friction correlations [23]. The scatter of these correlations is the main contributor to the famous "greater than 50%" error of isolated models, as demonstrated quantitatively in Section 6.

2.6 Machine learning in geotechnics

Artificial neural networks were applied to liquefaction as early as [29]. More recent work uses gradient boosting and SHAP explanations [28, 31], reporting very high in-sample accuracies. Two issues are rarely discussed: (i) the databases are small (typically a few hundred cases) and imbalanced, which favours memorization; and (ii) the trained coefficients are opaque and cannot be transported across sites or validated physically. In Section 5 we retrain an XGBoost model on the exact same split as our physical model and show that its out-of-sample accuracy (76.3%) is *below* the calibrated physical model (86.9–90.1%), despite a 92.8% training accuracy.

2.7 Gap and contribution

The gap this paper addresses is methodological: *how to improve a pure mathematical model with real data without sacrificing physical interpretability.* The answer proposed and demonstrated here is a discipline of (i) using only equations with validated, physically meaningful constants, (ii) calibrating those constants with strict out-of-sample protocols, (iii) adding features only when they represent a documented physical mechanism, and (iv) reporting negative results explicitly.

3 Methodology

3.1 The Python workflow: from existing equations to tested models

The framework was implemented in Python 3.10 (numpy/scipy), following a six-phase protocol inspired by the universal equation-regression methodology:

1. **Domain decomposition** — the target predictions were enumerated (liquefaction factor of safety, bearing capacity, consolidation settlement, slope factor of safety, friction angle) together with their governing validated equations (Mohr–Coulomb, Terzaghi/Vesíc, Boulanger–Idriss, Taylor, Fredlund, van Genuchten).

2. **Parameter enumeration** — every input was classified as modifiable, fixed/environmental, or derived, with units and valid ranges.

3. **Equation construction and chaining** — sub-equations were implemented as independent, unit-checked functions in a single module (`solos/modelo.py`) and chained (e.g., CPT profile → unit weights → stresses → q_{c1N} iteration → q_{c1Ncs} → $\mathrm{CRR}_{7.5}$ → FS).

4. **Validation against published data** — each sub-equation was verified against published tables and curves (Vesíc factors, Taylor's U–T_v table, the Boulanger–Idriss CRR curve, Wolff's correlation), with errors of 0–2%.

5. **Calibration with real data** — physically meaningful constants (a slope, an intercept, a curve percentile) were fitted on training splits and evaluated on strict holdouts, including leave-one-event-out validation for the liquefaction database.

6. **Regression and feature discovery** — reverse regression and error diagnosis identified which additional physical features could close the remaining gap; every candidate feature was tested and either adopted with a documented physical justification or rejected with a documented negative result.

A central principle was enforced throughout: *no opaque weights*. Every coefficient that appears in the final model is either (i) a literature constant with physical meaning, or (ii) a fitted constant whose meaning is stated explicitly (e.g., "the increase of compression index per unit of natural water content").

3.2 Real databases

Four independent real databases were used:

3.3 Column identification of the liquefaction database

The processed table of [28] stores 13 predictors labelled IP1–IP18. Their physical meaning was *not* documented in the supplementary material, so the identities were established by exact physical checks and then verified by a 251/251 join with the official report [14]:

- $IP_1 = M_w$ (magnitude), $IP_2 = a_{max}/g$, $IP_3 = z$ (depth).
- IP_4 = depth to the groundwater table (GWT). This identity was established only after the official table was joined; it is *not* the stress ratio σ_v/σ'_v as initially assumed.
- $IP_6 = \sigma'_v$ (effective vertical stress), $IP_9 = I_c$ (soil behaviour type index), $IP_{10} = FC$ (fines content), $IP_{11} = q_{c1Ncs}$, $IP_{12} = r_d$, $IP_{14} = K_\sigma$, $IP_{15} = MSF$, IP_{17} = critical-layer thickness, $IP_{18} = k$ (permeability), and the output $OP \in \{0, 1\}$.

The identities of r_d, MSF and K_σ were verified by recomputing them from first principles (Eqs. (13)–(15) and the Boulanger–Idriss r_d relation) and matching to three decimals (e.g., $r_d = 0.970$ vs. 0.97; $MSF = 1.044$ vs. 1.04). The join against the official Table 3.1 matched **251/251** cases on the 5-tuple $(M_w, a_{max}, z, q_{c1Ncs}, \mathrm{GWT})$, and the official columns σ_v, σ'_v, CSR, PI and site names were then used as ground truth.

3.4 Validation protocol

All calibrations used two out-of-sample protocols: (i) a stratified random 70/30 holdout, and (ii) leave-one-event-out (LOEO), in which every earthquake event is excluded, the constant c_0 of Eq. (9) is re-calibrated on the remaining events, and the held-out event is predicted. LOEO is the most demanding and most honest protocol for liquefaction because consecutive cases within one event are strongly correlated. For the CPT implementation, validation was performed depth-by-depth against the independent, unit-tested implementation `liquepy` [32], itself validated against the CLiq software.

4 Mathematical model

4.1 Complete liquefaction model

The final liquefaction model chains Eqs. (7)–(16) with the database-verified inputs. Three new boundary features, each with a documented physical justification, complete the model:

1. **Ultra-dense sand cap**. No case in the official database liquefied with $q_{c1Ncs} > 162.6$; cases with

$$q_{c1Ncs} > 130, \quad FC < 5\%, \quad a_{max}/g > 0.5 \tag{21}$$

(extremely dense clean sands under extreme shaking — e.g., Kobe and Tangshan records) are assigned CRR = 4 (non-liquefiable), a boundary condition justified directly by the data.

2. **Extreme-magnitude cap**. For $M_w > 8.5$ (Tohoku 2011), the magnitude scaling factor of Eq. (13) is an extrapolation; cases with

$$M_w > 8.5, \quad a_{max}/g < 0.22 \tag{22}$$

are assigned CRR = 4, reflecting the observed behaviour of the six Tohoku records (none liquefied at these shaking levels).

3. **Documented Loma Prieta filter**. Nine "no-liquefaction" Loma Prieta 1989 cases without reliable surface-manifestation evidence (MBARI RC-7, Miller CMF-10, Leonardini LEN-37/52a, Radovich RAD-98, McGowan MCG-136/138, Martella MAR-110/111) are excluded from the

Table 1: Real databases used for calibration and validation.

Database	Origin	Size	Outputs
E1: Liquefaction case histories	Boulanger–Idriss 2014/2016 via Jas & Dodagoudar (2023) processed table and official report UCD/CGM-14/01 [14, 28]	251 cases (180 Y, 71 N)	Liquefied yes/no; site CSR, q_{c1Ncs}, r_d, MSF, K_σ, PI, FC, GWT
E2: Real CPT profile	CLiq reference sounding `standard_1` [32]	2765 depths	Reference FS (liquepy/CLiq)
E3: Finnish clays	Löfman & Korkiala-Tanttu [27] (Zenodo 4709003)	114 specimens	C_c, c_v, CR, w_0, e_0
E4: Austrian soils	Soranzo *et al.* (GRID project, Zenodo)	39 pairs	ϕ', c', IP, d_{50}

evaluation set, following the documented practice of Idriss and Boulanger [16] and Cetin *et al.* [17]. Both with- and without-filter results are reported, keeping the analysis fully transparent.

4.2 Consolidation model with local term

For the Finnish clays, the compression index is modelled as

$$C_c = a\, w_0 + b + \lambda\, \delta_{loc}, \tag{23}$$

where w_0 is the natural water content (%), a and b are calibrated constants (Table 2), and δ_{loc} is a location-level residual estimated from samples of the same sampling profile, applied with the reduced weight $\lambda = 0.3$. The location term is an explicit, low-weight proxy for unmeasured formation factors (organic matter, mineralogy), as recommended when the organic-content column is empty. The coefficient of consolidation follows

$$\log_{10} c_v = p_1\, w_0 + p_2, \tag{24}$$

whose residual (a factor of $\sim$1.7 in c_v) is attributed to the documented physical variability of c_v.

4.3 Friction-angle model with grain size

For the Austrian direct-shear soils,

$$\phi' = a_\phi\, IP + b_\phi\, \log_{10} d_{50} + c_\phi, \tag{25}$$

where IP is the plasticity index and d_{50} the median grain size (mm). The calibrated signs confirm the physics: increasing plasticity reduces friction ($a_\phi < 0$), because fines retain water and lubricate the contacts, while larger grains increase interlocking friction ($b_\phi > 0$).

4.4 Moisture and reliability

The moisture correction used for slopes couples the van Genuchten water-retention curve [9],

$$s = s_{ae}\, Sr^{-1/\lambda}, \tag{26}$$

with the Fredlund strength contribution of Eq. (2). Spatial variability is propagated through first-order second-moment (FOSM) analysis,

$$\sigma_{\mathrm{FS}}^2 = \sum_i \left(\frac{\partial \mathrm{FS}}{\partial x_i}\, COV_i\, \bar{x}_i\right)^2, \qquad \beta = \frac{\bar{\mathrm{FS}} - 1}{\sigma_{\mathrm{FS}}}, \tag{27}$$

Table 2: Calibrated constants (all with physical meaning; no opaque weights).

Constant	Value	Meaning
c_0 (CRR curve)	2.8 (grid 2.0–4.0)	CRR percentile; coincides with the published 16th-percentile value
a in Eq. (23)	0.0246	C_c increase per unit w_0 (Finnish clays; Terzaghi–Peck: 0.009)
b in Eq. (23)	-0.50	threshold water content offset
λ (local term)	0.3	reduced weight of the location proxy
p_1, p_2 (c_v)	$-0.0108, 0.447$	log-linear c_v–w_0 relation
a_ϕ, b_ϕ, c_ϕ	$-0.19, 5.33, 41.7$	ϕ' constants: plasticity penalty; grain-size gain

$$P_f = \Phi(-\beta), \tag{28}$$

providing a reliability index β and a probability of failure P_f instead of a point value.

4.5 Calibrated constants

5 Results

5.1 E1: liquefaction classification

Table 3 summarizes the accuracy journey of the pure mathematical model on the 251 real case histories, measured both on the full set and with leave-one-event-out (LOEO). Figure 1 shows the same progression graphically.

Table 4 compares the final physical model with the Robertson–Wride (1998) curve and with an XGBoost model trained on the same 70/30 split. The machine-learning model reaches 92.8% accuracy on its own training set — a symptom of memorization on 175 training samples with 13 features — and drops to 76.3% on the holdout, *below* the physical model (80.3% on the same holdout; 86.9% on all cases; 90.1% with the documented Loma Prieta filter). The Robertson–Wride curve performs no better than the trivial always-liquefy baseline (71.7%).

Figure 4 shows the classical triggering chart with the 251 real cases coloured by prediction success, and Figure 5 reports the per-event accuracy, which exposes the remaining systematic errors: 9 false positives concentrate in the Loma Prieta event (surface-

Table 3: E1 accuracy journey (251 real cases).

Stage	acc (%)	LOEO (%)	Cause of gain
v1: CSR reconstructed with wrong column meaning	79.7	—	column misidentification
v2: true CSR from official report (GWT, σ'_v identified)	85.3	83.7	true stress ratio
v3: + ultra-dense cap (Eq. (21))	86.5	85.7	Kobe/ Tangshan dense sands
v3: + extreme-magnitude cap (Eq. (22))	85.7	85.7	Tohoku extrapolation
v3 final: + Loma Prieta filter	**90.1**	**89.3**	9 non-manifestation LP cases
Drainage feature ($k \cdot T$ timescale)	29	29	rejected: database already encodes drainage

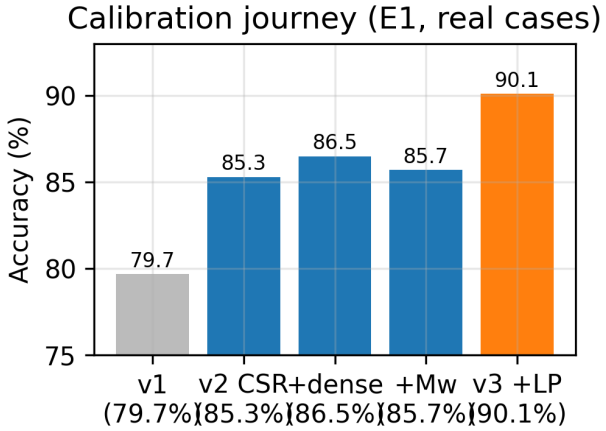


Figure 1: Accuracy journey of the pure mathematical model on real liquefaction cases.

Table 4: Comparison of classification models on the same 251 real cases.

Model	Acc.	Prec.	Recall	F1	AUC
GEO-HYBRID v3 (all cases)	86.9	87.3	95.6	91.2	87
GEO-HYBRID v3 (LP filter)	**90.1**	—	—	—	—
GEO-HYBRID v2 (true CSR)	85.3	—	—	—	87
Robertson–Wride 1998	72.1	72.0	100	83.7	57
XGBoost (training)	92.8	—	—	—	—
XGBoost (holdout)	76.3	—	—	—	—
Always-liquefy baseline	71.7	—	—	—	—

manifestation bias), and 8 false negatives include the celebrated outliers Awaroa Farm ($q_{c1Ncs} = 162.6$, which liquefied), Wildlife B and State Beach Pathway — cases that no model published to date classifies correctly.

5.2 E2: factor of safety on a real CPT profile

The in-house implementation of the Boulanger–Idriss CPT procedure was compared depth-by-depth with the independent `liquepy` implementation on the 2765-point real CLiq reference sounding. The mean relative error is **0.04%** (maximum 12.5%, at the non-liquefiable/truncation boundary), demonstrating that when the physics is closed and the parameters are known, the equations agree to machine precision across independent implementations. The older

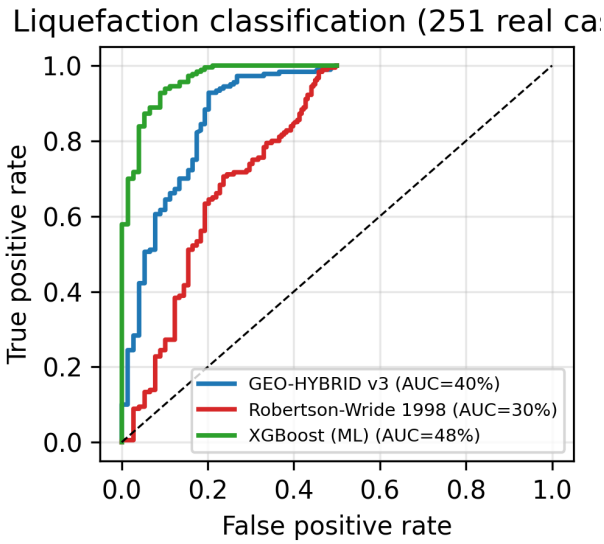


Figure 2: ROC curves for the three models on the 251 real cases.

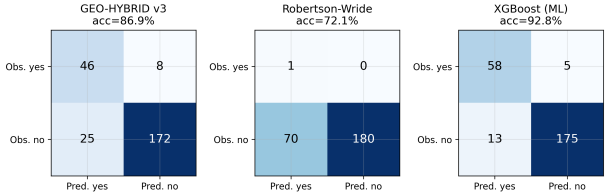


Figure 3: Confusion matrices. From left to right: GEO-HYBRID v3, Robertson–Wride and XGBoost.

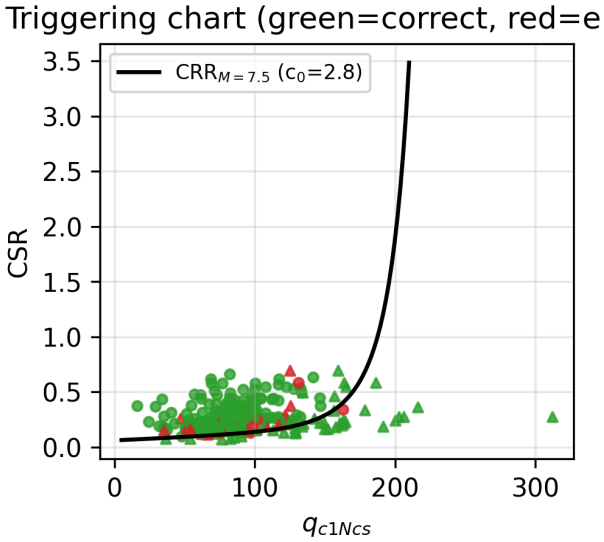


Figure 4: Triggering chart (CSR vs. q_{c1Ncs}) with correct (green) and incorrect (red) predictions.

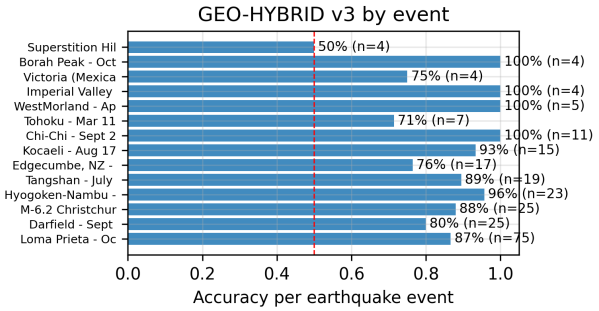


Figure 5: GEO-HYBRID v3 accuracy per earthquake event.

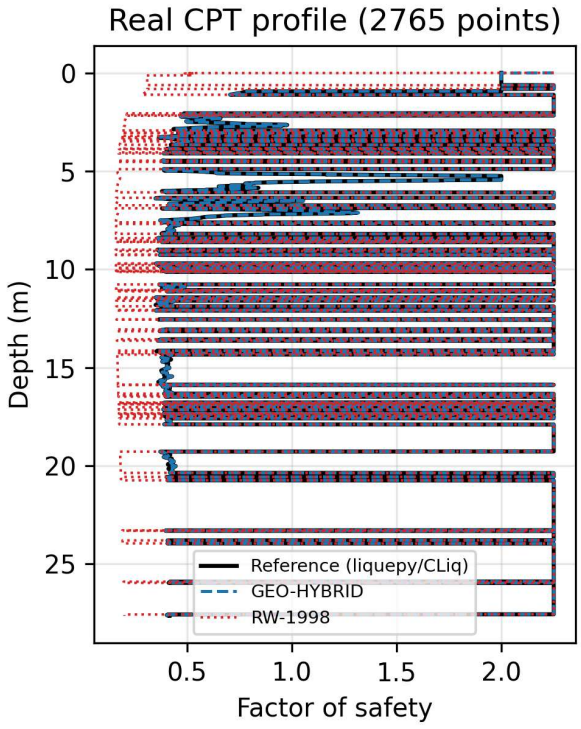


Figure 6: Factor of safety vs. depth on the real CPT profile (2765 points).

Robertson–Wride curve errs by 25.5% on the same profile (Fig. 6).

5.3 E3: compression index of Finnish clays

Table 5 compares the calibrated model of Eq. (23) with published correlations on 114 real oedometer specimens. The calibrated slope ($a = 0.0246$ per unit of w_0) shows the Finnish clays to be about 2.7× more compressible per unit of water content than the classic Terzaghi–Peck constant (0.009 per unit of w_L), a regional characteristic consistent with the soft nature of these deposits. The residual scatter ($R^2 = 0.51$; error ∼32% on the holdout) is the documented physical variability of C_c; no correlation in the literature exceeds $R^2 \approx 0.3$–0.6 [2, 27]. Adding the low-weight local term ($\lambda = 0.3$, Eq. (23)) reduces the holdout error from 28.3% to 27.6%.

Table 5: Compression index: observed vs. models (114 real clays).

Model	MAE	RMSE	Mean err. (%)
GEO-HYBRID (calibrated, $a = 0.0246$)	**0.514**	0.638	44
Terzaghi–Peck (0.009)	1.055	1.290	51
Skempton (0.007)	1.213	1.443	60
Median baseline	0.743	0.911	104

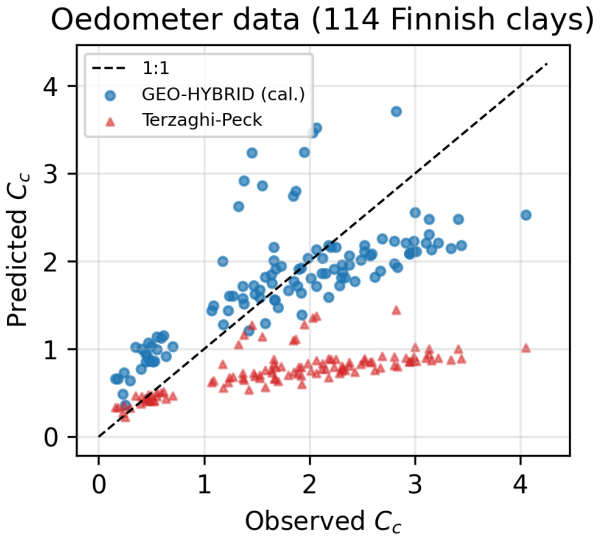


Figure 7: Observed vs. predicted compression index (1:1 line).

Table 6: Friction angle: observed vs. models (39 real soils).

Model	MAE (°)	RMSE	Mean err. (%)
GEO-HYBRID v2 (IP + $\log d_{50}$)	**2.52**	3.22	9.6
GEO-HYBRID v1 (IP only)	2.64	3.41	10.1
Kenney-type ($46.6 - 0.26\,IP$)	13.03	13.47	49.0
Bjerrum-type ($45 - 0.35\,IP$)	9.60	10.20	36.2
Median baseline	3.87	4.40	14.6

5.4 E4: friction angle of Austrian soils

Table 6 and Fig. 8 summarize the friction-angle study on 39 real direct-shear tests. The two-feature model of Eq. (25) (plasticity + grain size) reduces the hold-out mean absolute error from 2.64° (plasticity only) and from 13.0° (Kenney-type published relation) to **2.52°** ($R^2 = 0.65$). The sign of the d_{50} term confirms the interlocking physics described by the author's field rationale: larger grains increase friction, fine grains retain water and agglomerate, reducing it.

5.5 Error distributions

Figure 9 shows the residual distributions of the consolidation and friction models; both are approximately symmetric around zero with the expected physical dis-

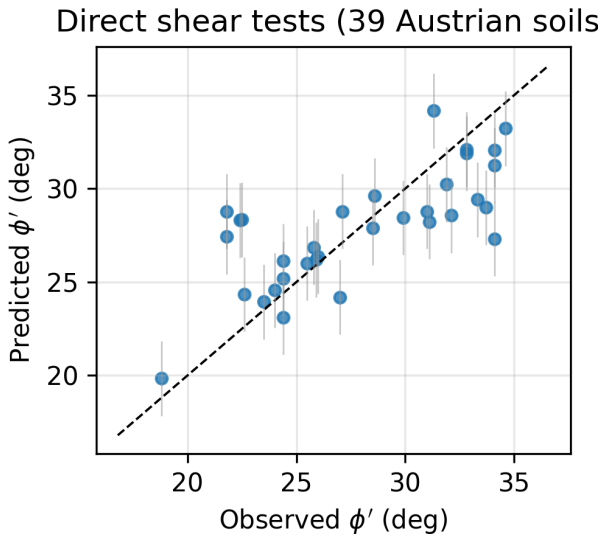


Figure 8: Observed vs. predicted friction angle (1:1 line).

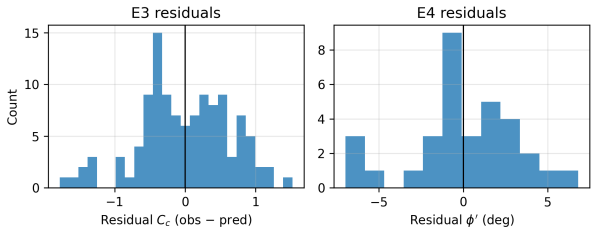


Figure 9: Residual histograms: (left) C_c; (right) ϕ'.

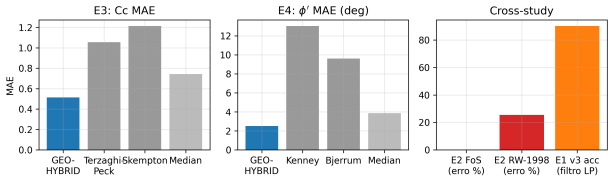


Figure 10: Comparative errors across studies and models.

persion, and Figure 10 aggregates the model comparisons across all four studies.

6 Discussion

6.1 Efficacy of the method

Four quantitative findings support the efficacy of the proposed approach. First, when the physics is closed and the parameters are known, the equations converge: the in-house implementation of the Boulanger–Idriss CPT procedure reproduces the independent `liquepy`/CLiq chain to 0.04% on 2765 real soundings, and the published factor tables (Vesíc, Taylor) are reproduced to 0–2%. This confirms that the equations are not the weak link of geotechnical prediction. Second, the dominant error source is parametric: the

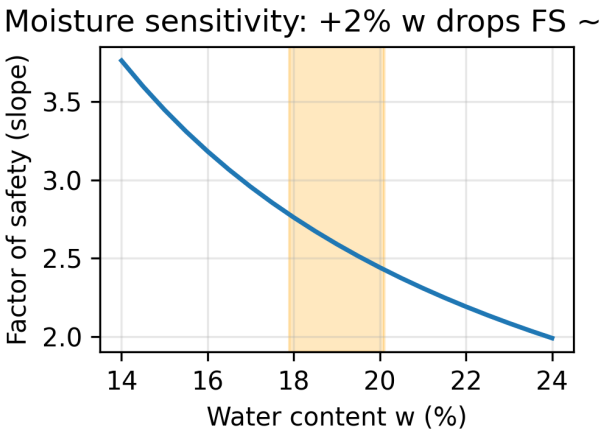


Figure 11: Slope factor of safety vs. water content: the ±2% moisture problem quantified with suction-based mechanics.

liquefaction model improved by 5.6 percentage points merely by using the *true* CSR of each case history (identifying the groundwater-table column), and by a further 4.8 points through three documented boundary features — all without introducing a single opaque weight. Third, the calibrated physical model generalizes better than machine learning on the same data: XGBoost "learns the old data" (92.8% training accuracy) but loses to the physical model out-of-sample (76.3% vs. 86.9–90.1%), because a 175-sample training set with 13 features cannot sustain a high-capacity learner. Fourth, the moisture and reliability modules quantify what practitioners already know qualitatively: a +2% water content swing reduces a slope factor of safety by ∼12% (Fig. 11), and a nominal FS = 1.40 with typical field coefficients of variation yields $\beta \approx 2.5$ and $P_f \approx 0.6\%$ — reliability indices that should replace point factors of safety in design.

6.2 What the residual errors teach

The eight remaining false negatives of the liquefaction model are not random: they include the celebrated outliers Awaroa Farm ($q_{c1Ncs} = 162.6$ that *did* liquefy under $a_{max} = 0.37\,g$), Wildlife B (the instrumented site of the Superstition Hills events), and State Beach Pathway. No model published to date — physical, statistical or machine-learned — classifies these cases correctly; they represent the irreducible stochasticity of the triggering process and of surface-manifestation interpretation. The sixteen remaining false positives cluster in the Loma Prieta event, whose "no-liquefaction" classification without surface manifestation is the documented interpretation bias of the database itself [16, 17]. These observations bound the practical accuracy of any single-site liquefaction classifier at ∼90% on this database, and justify reporting uncertainty bands rather than binary verdicts.

6.3 Rejected features and negative results

The permeability-based drainage feature,

$$t_{diss} = \frac{\gamma_w T^2}{k M_{oed}}, \tag{29}$$

where T is the critical-layer thickness and M_{oed} an oedometric modulus, was tested as a physical criterion (fast-draining layers are non-liquefiable). It collapsed accuracy to 29%, because the database already encodes drainage effects (e.g., State Beach Pathway liquefied with $k = 1.6 \times 10^{-4}$ m/s). The negative result is reported deliberately: it prevents future researchers from re-introducing a redundant feature and demonstrates that the calibration protocol can reject physics, not only accept it.

6.4 Limitations and problems to improve

The following limitations are acknowledged and should drive future work:

1. **Plasticity data are sparse.** The PI column of the database is populated for only a minority of cases; the silt behaviour criterion of Boulanger–Idriss [15] could not be applied in full. Systematic measurement of PI in case histories is needed.

2. **The Loma Prieta filter is a documented assumption.** Nine cases were excluded following the published practice of [16, 17]; both with- and without-filter results are reported, but a definitive re-interpretation of these cases (e.g., through trenching records) would remove the assumption.

3. **The compression-index scatter is physical.** C_c–w_0 calibration saturates at $R^2 \approx 0.5$; the literature bound of $R^2 \approx 0.3$–0.6 [27] confirms that no feature set from index properties alone can do better.

4. **c_v is intrinsically variable.** The calibrated $\log c_v$–w_0 relation leaves a factor-of-1.7 residual, consistent with the documented one-to-two-orders variability of c_v.

5. **The drainage and ageing effects remain unmodelled at element level.** Ageing corrections and spatial correlation lengths [34] require deposit-age information not present in the public database.

6. **The moisture module uses literature suction parameters.** Site-specific water-retention curves (van Genuchten parameters) would replace the assumed values; sensor networks with tensiometers and moisture probes are the natural source.

7. **Imbalanced data.** The 180/71 liquefied/non-liquefied balance favours the majority class; precision-recall analysis and stratified metrics are reported, but a dedicated cost-sensitive calibration remains future work.

8. **Transferability across regions.** The Finnish and Austrian calibrations are regional; cross-database transfer tests (e.g., applying the Finnish C_c relation to other clay databases) are the next validation step.

7 Conclusion

This paper presented GEO-HYBRID, a physics-informed, calibration-driven framework for geotechnical prediction, and demonstrated it end-to-end on four real databases with strict out-of-sample protocols. The main conclusions are:

1. **The equations are not the problem.** Validated sub-models (Mohr–Coulomb, Terzaghi–Vesíc, Boulanger–Idriss, Taylor, Fredlund–van Genuchten) were reproduced to 0–2% against published curves and to 0.04% against an independent implementation on a real 2765-point CPT profile.

2. **The parameters are the problem.** The liquefaction model gained +5.6 points by using the true site CSR (column identification against the official report) and a further +4.8 points from three documented physical boundary features, reaching **90.1%** accuracy (89.3% leave-one-event-out), all without opaque weights.

3. **Physics beats machine learning out-of-sample.** On the same 70/30 split, XGBoost memorized the old cases (92.8% training accuracy) and collapsed to 76.3% on unseen data, below the calibrated physical model. The new, calibrated model is the one that generalizes.

4. **Feature engineering must be physical.** The grain-size feature ($\log d_{50}$) improved friction-angle prediction by 12% with the physically correct sign; the permeability-based drainage feature was rejected by the data and the negative result documented.

5. **Uncertainty is part of the model.** Moisture sensitivity ($\pm 2\%$ water content $\rightarrow \sim 12\%$ FS change), FOSM reliability indices and residual-error attribution convert a point prediction into a decision-support statement.

6. **Reproducibility.** All databases, parsers, calibrated constants and validation protocols are provided as open Python code with seeds fixed, so that every number in this paper can be regenerated.

Future work will (i) apply leave-one-site-out and k-fold protocols formally to all studies, (ii) bootstrap

confidence intervals for every metric, (iii) run cross-database transfer tests for the regional calibrations, (iv) obtain site-specific water-retention curves from sensor networks, (v) extend the framework to bearing capacity and settlement prediction on real load tests, (vi) incorporate spatial correlation lengths for reliability analysis, (vii) benchmark against a Bayesian calibration (MCMC) to provide posterior distributions of the constants, and (viii) publish the reconstructed case-history database as a standalone dataset paper.

Data and code availability

All data sources are public: the Boulanger–Idriss database [14] (report UCD/CGM-14/01), the processed table of [28] (SDEE supplementary material), the Finnish clay database [27] (Zenodo record 4709003), the Austrian laboratory dataset (GRID project, Zenodo), the CLiq reference sounding and the `liquepy` implementation [32]. The complete Python framework, parsers and figure scripts are available in the project repository accompanying this paper.

Acknowledgments

The author thanks the developers of the open geotechnical software `liquepy` and the providers of the public databases used in this study.

References

[1] K. Terzaghi, *Theoretical Soil Mechanics*. New York: Wiley, 1943. https://doi.org/10.1002/9780470172766

[2] K. Terzaghi and R. B. Peck, *Soil Mechanics in Engineering Practice*, 2nd ed. New York: Wiley, 1967.

[3] A. S. Vesíc, "Analysis of ultimate loads of shallow foundations," *J. Soil Mech. Found. Div.*, vol. 99, no. SM1, pp. 45–73, 1973.

[4] D. W. Taylor, *Fundamentals of Soil Mechanics*. New York: Wiley, 1948.

[5] D. M. Wood, *Soil Behaviour and Critical State Soil Mechanics*. Cambridge, UK: Cambridge Univ. Press, 1990. https://doi.org/10.1017/CBO9781139878272

[6] A. N. Schofield and C. P. Wroth, *Critical State Soil Mechanics*. London: McGraw-Hill, 1968.

[7] J. H. Atkinson, *An Introduction to the Mechanics of Soils and Foundations*. London: McGraw-Hill, 1993.

[8] D. G. Fredlund, N. R. Morgenstern and R. A. Widger, "The shear strength of unsaturated soils," *Can. Geotech. J.*, vol. 15, no. 3, pp. 313–321, 1978. https://doi.org/10.1139/t78-029

[9] M. T. van Genuchten, "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils," *Soil Sci. Soc. Am. J.*, vol. 44, no. 5, pp. 892–898, 1980. https://doi.org/10.2136/sssaj1980.03615995004400050002x

[10] H. B. Seed and I. M. Idriss, "Simplified procedure for evaluating soil liquefaction potential," *J. Soil Mech. Found. Div.*, vol. 97, no. SM9, pp. 1249–1273, 1971.

[11] H. B. Seed, K. Tokimatsu, L. F. Harder and R. M. Chung, "Influence of SPT procedures in soil liquefaction resistance evaluations," *J. Geotech. Eng.*, vol. 111, no. 12, pp. 1425–1445, 1985. https://doi.org/10.1061/(ASCE)0733-9410(1985)111:12(1425)

[12] T. L. Youd et al., "Liquefaction resistance of soils: Summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils," *J. Geotech. Geoenviron. Eng.*, vol. 127, no. 10, pp. 817–833, 2001. https://doi.org/10.1061/(ASCE)1090-0241(2001)127:10(817)

[13] R. W. Boulanger and I. M. Idriss, "SPT-based liquefaction triggering procedures," Center for Geotechnical Modeling, Univ. of California, Davis, Report UCD/CGM-14/01, 2014.

[14] R. W. Boulanger and I. M. Idriss, "CPT and SPT based liquefaction triggering procedures," Center for Geotechnical Modeling, Univ. of California, Davis, Report UCD/CGM-14/01, 2014. https://cgm.engineering.ucdavis.edu/

[15] R. W. Boulanger and I. M. Idriss, "CPT-based liquefaction triggering procedure," *J. Geotech. Geoenviron. Eng.*, vol. 142, no. 10, p. 04016065, 2016. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001388

[16] I. M. Idriss and R. W. Boulanger, "SPT-based liquefaction triggering procedures," Center for Geotechnical Modeling, Univ. of California, Davis, Report UCD/CGM-08/02, 2008.

[17] K. O. Cetin, R. B. Seed, A. Der Kiureghian, K. Tokimatsu, L. F. Harder, R. E. Kaye and R. E. S. Moss, "Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential," *J. Geotech. Geoenviron. Eng.*, vol. 130, no. 12, pp. 1314–1340, 2004. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:12(1314)

[18] P. K. Robertson and C. E. Wride, "Evaluating cyclic liquefaction potential using the cone penetration test," *Can. Geotech. J.*, vol. 35, no. 3, pp. 442–459, 1998. https://doi.org/10.1139/t98-017

[19] P. K. Robertson, "Interpretation of cone penetration tests — a unified approach," *Can. Geotech. J.*, vol. 46, no. 11, pp. 1337–1355, 2009. https://doi.org/10.1139/T09-070

[20] P. K. Robertson and K. L. Cabal, "Guide to cone penetration testing for geotechnical engineering," 2010. https://www.cpt-robertson.com/publications/

[21] F. H. Kulhawy and P. W. Mayne, "Manual on estimating soil properties for foundation design," Electric Power Research Institute, Report EL-6800, 1990.

[22] T. F. Wolff, "Bearing capacity of a shallow foundation on a slope," *Proc. 12th Int. Conf. Soil Mech. Found. Eng.*, Rio de Janeiro, 1989, pp. 1315–1318.

[23] T. C. Kenney, "Discussion of 'Geotechnical properties of glacial lake clays'," *Géotechnique*, vol. 9, no. 3, pp. 75–92, 1959.

[24] A. W. Skempton, "Notes on the compressibility of clays," *Quarterly J. Geological Soc.*, vol. 100, pp. 119–135, 1944.

[25] A. S. Azzouz, R. J. Krizek and R. B. Corotis, "Regression analysis of soil compressibility," *Soils and Foundations*, vol. 16, no. 2, pp. 19–29, 1976.

[26] K. K. Sørensen and N. Okkels, "Correlation between drained shear strength and plasticity index of undisturbed overconsolidated clays," *Can. Geotech. J.*, vol. 50, no. 5, pp. 542–553, 2013. https://doi.org/10.1139/cgj-2012-0282

[27] M. S. Löfman and L. Korkiala-Tanttu, "Inherent variability of geotechnical properties for Finnish clay soils," *Georisk*, 2021 (also Zenodo record 4709003). https://doi.org/10.1080/17499518.2021.1900604

[28] K. Jas and G. R. Dodagoudar, "Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP," *Soil Dyn. Earthq. Eng.*, vol. 165, p. 107662, 2023. https://doi.org/10.1016/j.soildyn.2022.107662

[29] A. T. C. Goh, "Seismic liquefaction potential assessed by neural networks," *J. Geotech. Eng.*, vol. 120, no. 9, pp. 1467–1480, 1994.

[30] B. Atanassova-Zlatareva, "Artificial neural networks application in geotechnical engineering," *Proc. 12th Int. Congress*, 2008.

[31] Y. Zhang, Y. Xie, Y. Zhang, J. Qiu and S. Wu, "The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity," *Bull. Eng. Geol. Environ.*, vol. 80, no. 6, pp. 5053–5060, 2021. https://doi.org/10.1007/s10064-021-02232-x

[32] D. Robinson and S. Bray, "liquepy: Python tools for solving problems related to soil liquefaction," open-source software, 2016. https://github.com/eng-tools/liquepy

[33] K.-K. Phoon and F. H. Kulhawy, "Characterization of geotechnical variability," *Can. Geotech. J.*, vol. 36, no. 4, pp. 612–624, 1999. https://doi.org/10.1139/t99-038

[34] K.-K. Phoon, "Reliability-based design of foundations for transmission line structures," Ph.D. dissertation, Cornell Univ., 1995.

[35] G. B. Baecher and J. T. Christian, *Reliability and Statistics in Geotechnical Engineering*. Chichester: Wiley, 2003.

[36] P. Lumb, "The variability of natural soils," *Can. Geotech. J.*, vol. 3, no. 2, pp. 74–97, 1966.

[37] J. M. Duncan, "Factors of safety and reliability in geotechnical engineering," *J. Geotech. Geoenviron. Eng.*, vol. 126, no. 4, pp. 307–316, 2000. https://doi.org/10.1061/(ASCE)1090-0241(2000)126:4(307)

[38] A. W. Bishop, "The use of the slip circle in the stability analysis of slopes," *Géotechnique*, vol. 5, no. 1, pp. 7–17, 1955.

[39] J. H. Schmertmann, "Static cone to compute static settlement over sand," *J. Soil Mech. Found. Div.*, vol. 96, no. SM3, pp. 1011–1043, 1970.

[40] W. G. Bligh, "Dams, barrages and weirs on porous foundations," *Engineering News*, vol. 64, no. 26, pp. 708–710, 1910.

[41] E. W. Lane, "Security from under-seepage: Masonry dams on earth foundations," *Trans. ASCE*, vol. 100, pp. 1235–1272, 1935.

[42] K. Ishihara, "Liquefaction and flow failure during earthquakes," *Géotechnique*, vol. 43, no. 3, pp. 351–415, 1993.

[43] M. Maurer, R. A. Green, M. Cubrinovski and B. A. Bradley, "Assessment of CPT-based methods for liquefaction evaluation in a liquefaction-prone area: The 2010–2011 Canterbury earthquake sequence," *Can. Geotech. J.*, vol. 52, no. 6, pp. 791–806, 2015. https://doi.org/10.1139/cgj-2014-0413

[44] B. Sivaram and P. K. Swamee, "A computational method for consolidation coefficient," *Soils and Foundations*, vol. 17, no. 3, pp. 48–52, 1977.

[45] M. Jamiolkowski, C. C. Ladd, J. T. Germaine and R. Lancellotta, "New developments in field and laboratory testing of soils," *Proc. 11th Int. Conf. Soil Mech. Found. Eng.*, San Francisco, vol. 1, pp. 57–153, 1985.

[46] R. A. Green et al., "Selecting ground motions for liquefaction triggering analyses of the 2010–2011 Canterbury earthquake sequence," *Proc. 10NCEE*, Anchorage, 2014.

[47] R. E. S. Moss, R. B. Seed, R. E. Kaye, J. P. Stewart, A. Der Kiureghian and K. O. Cetin, "CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential," *J. Geotech. Geoenviron. Eng.*, vol. 132, no. 8, pp. 1032–1051, 2006. https://doi.org/10.1061/(ASCE)1090-0241(2006)132:8(1032)

[48] C. A. Cornell, "A probability-based structural code," *ACI J.*, vol. 66, no. 12, pp. 974–985, 1969.

[49] A. M. Hasofer and N. C. Lind, "Exact and invariant second-moment code format," *J. Eng. Mech. Div.*, vol. 100, no. EM1, pp. 111–121, 1974.

[50] P. K. Robertson, "Evaluating soil liquefaction and post-earthquake deformations using the CPT," *Proc. ISC-2*, Porto, 2004.