THE PRIORS OF LIQUEFACTION ENGINEERING

Achyut Tiwari[1], Robb Moss[1,2], and Ashok Kumar Gupta[1,3]

[1]GeoLiquefy LLC, San Francisco, California, USA
[2]Department of Civil and Environmental Engineering, California Polytechnic State University, San Luis Obispo, CA, USA
[3]Department of Civil Engineering, Jaypee University of Information Technology, Waknaghat, Solan, 173234, Himachal Pradesh, India

May 27, 2026

ABSTRACT

Liquefaction assessment is a Bayesian problem. Before any calculation, the engineer already holds a prior: depositional history, regional seismicity, the documented performance of analogous sites. The simplified procedure produces a factor of safety, and that prior is discarded. This is the central error. The factor of safety—defined as the cyclic resistance ratio divided by the cyclic stress ratio— compresses a multidimensional problem into a scalar whose false precision obscures uncertainty rather than quantifies it. The CRR curves embed a prior, frozen inside regression coefficients the practicing engineer inherits but cannot update. Similarity-based retrieval recovers what scalar compression destroys. By matching design sites to analogous case histories in the NGL database on CPT resistance, fines content, groundwater depth, and PGA, the GeoLiquefy Liquefaction Hazard Engine returns documented field outcomes, explicit consequence data, and visible uncertainty structure. A matched case history tells you what happened at the most similar site the profession has ever recorded. That is qualitatively different from a factor of safety of 0.92. For sites near the triggering threshold, this is the information that should drive the decision.

1 THE PRIOR

Before you open a CPT file, you already know something. The soil was hydraulically placed or naturally deposited. The deposit is young Holocene alluvium or stiff Pleistocene material. The region has or has not produced liquefaction in prior earthquakes. The fines are plastic or granular. This is your prior, the structured, professional knowledge about how soils like this behave under shaking like this, accumulated across decades of post-earthquake field observation.

Bayesian reasoning is precise about what to do with it: the posterior, your belief after the site investigation, is proportional to the prior multiplied by the likelihood of the new data. A strong prior should not be casually overridden by a model output whose uncertainty band spans ± 0.15 to 0.25 FS units [1]. The precision of an index is not evidence of its accuracy.

> "A better screening analysis is simply to ask: has there ever been liquefaction with engineering consequences in a similar soil profile in an area with similar seismicity? If the answer is no, then there is no problem." Pyke (2026)

Two stories make this concrete. At Treasure Island, hydraulically placed fill and native shoal sands showed nearly identical CPT profiles, yet responded completely differently under vibro-compaction [6]. The simplified analysis treated them as equivalent. The prior, rooted in depositional history and soil fabric, would have distinguished them before the first calculation. In the Marina District, the drainage system that survived the 1989 Loma Prieta earthquake was designed from a clear prior about what loose saturated sand near the water table does in a Bay Area earthquake, not from a computed factor of safety. No formal analysis was run. The prior was sufficient [6].

The prior is not noise. It is signal, and the standard workflow discards it.

2 THE INDEX PROBLEM

$FS = CRR/CSR$ is a compression. It maps a multidimensional physical problem, depositional history, soil fabric, fines plasticity, shaking duration and frequency content, onto a single number. Everything that does not survive that compression does not become irrelevant. It becomes invisible.

What is lost: the frequency content and duration of the ground motion; the depositional and aging state of the soil; the plasticity of the fines (which can dramatically improve cyclic resistance regardless of tip resistance, [6]; [7]); the documented performance of every analogous site in the historical record. What remains is a number that is reproducible but may be deeply uninformative about this specific site.

$FS = 0.92$ and $FS = 0.88$ feel meaningfully different. They are not. Both sit inside the model's own uncertainty band. The precision is false. False precision is the most dangerous form of information loss, it creates the appearance of knowledge where there is uncertainty.

The CRR curves are derived from global case history regressions [10, 4, 5]. Every regression embeds a prior: the functional form, the weighting of observations, the correction factors applied. That prior belongs to the procedure's developers. The practicing engineer inherits it silently and cannot inspect, challenge, or update it. The prior is frozen inside the coefficients.

The practical cost is systematic conservatism. When the prior favors non-liquefaction but $FS < 1.0$, the engineer must either report the number and recommend expensive mitigation, or argue against the calculation in a regulatory environment that demands a threshold exceedance. Pyke [6] describes predictions of ten inches of settlement for sites where analogous depositional environments had never shown significant disturbance. At the 2nd Ishihara Colloquium, Youd described a single project where more rigorous site screening avoided approximately $100 million in unnecessary ground improvement. The index was the obstacle.

3 WHAT SIMILARITY RESTORES

The solution is not a better index. It is less compression. The GeoLiquefy Liquefaction Hazard Engine searches the Next Generation Liquefaction (NGL) database [2, 9] for the case histories most geotechnically similar to a design site, matched on q_{c1Ncs}, fines content, groundwater depth, and PGA, and returns their documented field outcomes. Not a model output. Evidence.

A case history matched on your CPT profile, fines content, groundwater depth, and shaking intensity tells you what happened at the most analogous site the profession has ever documented. A factor of safety cannot do that. It does not even try.

Three things are restored that the index suppresses. First, the prior: explicit, traceable to named cases and named earthquakes, not averaged into a coefficient. Second, consequences: triggering models predict triggering, nothing more. Matched cases come with documented settlement magnitudes, sand boil behavior, and structural damage from the NGL database [2]. Third, honest uncertainty: not a confidence interval on a regression parameter, but the spread of outcomes across the top matched cases. When all five matched cases liquefied, confidence is high. When outcomes diverge, the parameter driving the divergence is visible.

Carlton et al. [3] demonstrated this on the Canterbury Earthquake Sequence: case-based reasoning achieved accuracy comparable to or better than established simplified procedures, while remaining fully transparent, every prediction traceable to specific field observations. Rudin [8] makes the general argument: in high-stakes decisions, models that reason from explicit evidence are epistemically superior to accurate models whose outputs cannot be interrogated.

Index (FS / P_L)	**Similarity (Liquefaction Hazard Enginel)**
A number from a model	Observations from the field
Prior frozen in coefficients, invisible	Prior explicit, traceable to real cases
Uncertainty: $\pm$interval on a regression	Uncertainty: divergence among matched cases
No consequence prediction	Settlement, damage from matched cases
Silent when extrapolating	Flags when no close analog exists

The recommendation: for any site where FS sits within ~0.2 units of the triggering threshold, or where conditions fall outside the central calibration range, unusual fines, old deposits, hydraulic fill, deep layers, treat the similarity-based assessment as primary, not supplementary. The index tells you where the site falls in a statistical model. The matched cases tell you what the profession knows about what actually happens. In decisions affecting public safety and project economics, the latter is the information that matters.

References

[1] Boulanger, R.W. and Idriss, I.M. (2014). CPT and SPT Based Liquefaction Triggering Procedures. Report UCD/CGM-14/01, UC Davis.

[2] Brandenberg, S.J. et al. (2020). "Next-generation liquefaction database." *Earthquake Spectra*, 36(2), 939–959. doi:10.1177/8755293020902477

[3] Carlton, B., Geyin, M. and Engin, H.K. (2024). "Evaluation of case-based reasoning to estimate liquefaction manifestation." *Earthquake Spectra*. doi:10.1177/87552930231203573

[4] Cetin, K.O., Seed, R.B., Der Kiureghian, A., Tokimatsu, K., Harder, L.F., Kayen, R.E. and Moss, R.E.S. (2004). "SPT-Based Probabilistic and Deterministic Assessment of Seismic Soil Liquefaction Potential." *J. Geotech. Geoenviron. Eng.*, 130(12), 1314–1340.

[5] Moss, R.E.S., Seed, R.B., Kayen, R.E., Stewart, J.P., Der Kiureghian, A. and Cetin, K.O. (2006). "CPT-Based Probabilistic and Deterministic Assessment of In Situ Seismic Soil Liquefaction Potential." *J. Geotech. Geoenviron. Eng.*, 132(8), 1032–1051.

[6] Pyke, R. (2026). "Another Attempt to Promote the Use of Both Commonsense and Experience in Evaluating Earthquake-Induced Liquefaction and its Consequences." Unpublished manuscript, March 2026.

[7] Robertson, P.K. and Wride, C.E. (1998). "Evaluating cyclic liquefaction potential using the cone penetration test." *Canadian Geotechnical Journal*, 35(3), 442–459.

[8] Rudin, C. (2019). "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence*, 1(5), 206–215.

[9] Stewart, J.P., Kramer, S.L., Kwak, D.Y. et al. (2016). "PEER-NGL project: Open source global database and model development." *Soil Dynamics and Earthquake Engineering*, 91, 317–328.

[10] Youd, T.L., Idriss, I.M., Andrus, R.D. et al. (2001). "Liquefaction Resistance of Soils: Summary Report from the 1996 NCEER and 1998 NCEER/NSF Workshops." *J. Geotech. Geoenviron. Eng.*, 127(10), 817–833.