Evaluation and Updating of Ishihara's (1985) Model for Liquefaction Surface Expression, with Insights from Machine and Deep Learning

Rateria, G. [1] **and Maurer, B.W.** [2]

Abstract: Liquefaction surface-manifestation is a popular proxy of damage potential for infrastructure. Models for predicting it are thus commonly used, and often codified, in earthquake engineering practice. One such model is that of Ishihara (1985) who proposed empirical "H_1-H_2" curves considering the influence of the non-liquefied crust on surface expression. Yet, while widely used and cited, these curves were trained on just ~300 data points from two earthquakes. Accordingly, this study evaluates and updates the Ishihara (1985) model using 14,400 data points from 24 earthquakes, while also comparing against three other manifestation models from the literature. In addition to retraining the H_1-H_2 model via traditional regression, new variants are developed via machine- and deep-learning. Each of the new H_1-H_2 models outperforms the original in unbiased testing and is suitable for application. Ultimately, however, this paper also explores the limits of H_1-H_2 models and the apparent inefficiency and/or insufficiency of their predictor variables. In this regard, the models developed herein may perform better than any other, yet new models are still needed to account for factors influential in producing surface manifestation in a more explicit and mechanistic manner.

1. Introduction

The surface manifestation of liquefaction in the free field is a practical, general proxy of damage potential for near-surface infrastructure (e.g., shallow foundations and lifelines). Using this proxy, manifestation models have been proposed to link the safety factor against liquefaction triggering (FS_{liq}) at depth within a profile to damage potential at the surface, such that asset damage is more likely when liquefaction manifestation (e.g., ejecta) is expected. Towards this end, Ishihara (1985) recognized the influence of the non-liquefied capping layer, or crust, on surface expression. Plotting observations from the 1983 Nihonkai-Chubu earthquake using the thicknesses of the crust, H_1, and liquefied strata, H_2, Ishihara (1985) proposed boundary curves for predicting surface manifestation as a function of H_1, H_2, and peak ground acceleration (PGA). Ishihara (1985) originally developed a single curve, shown in Fig. 1a, using data from sites that experienced a PGA of 200 gal (note: 980.7 gal = 1 g). Reinterpreting data compiled by Gao et al. (1983) from the 1976 Tangshan earthquake, a second curve corresponding

---

[1]Graduate Research Assistant, Dept. of Civil and Environmental Engineering, University of Washington, USA
[2]Assistant Professor, Dept. of Civil and Environmental Engineering, University of Washington, USA

to ~400-500 gal was added, as shown in Fig. 1b. Ishihara then proposed a third curve (300 gal) by interpolation. Collectively, these curves suggest that for a given *PGA*, there is a limiting H_1 beyond which surface manifestations are not expected, regardless of H_2.

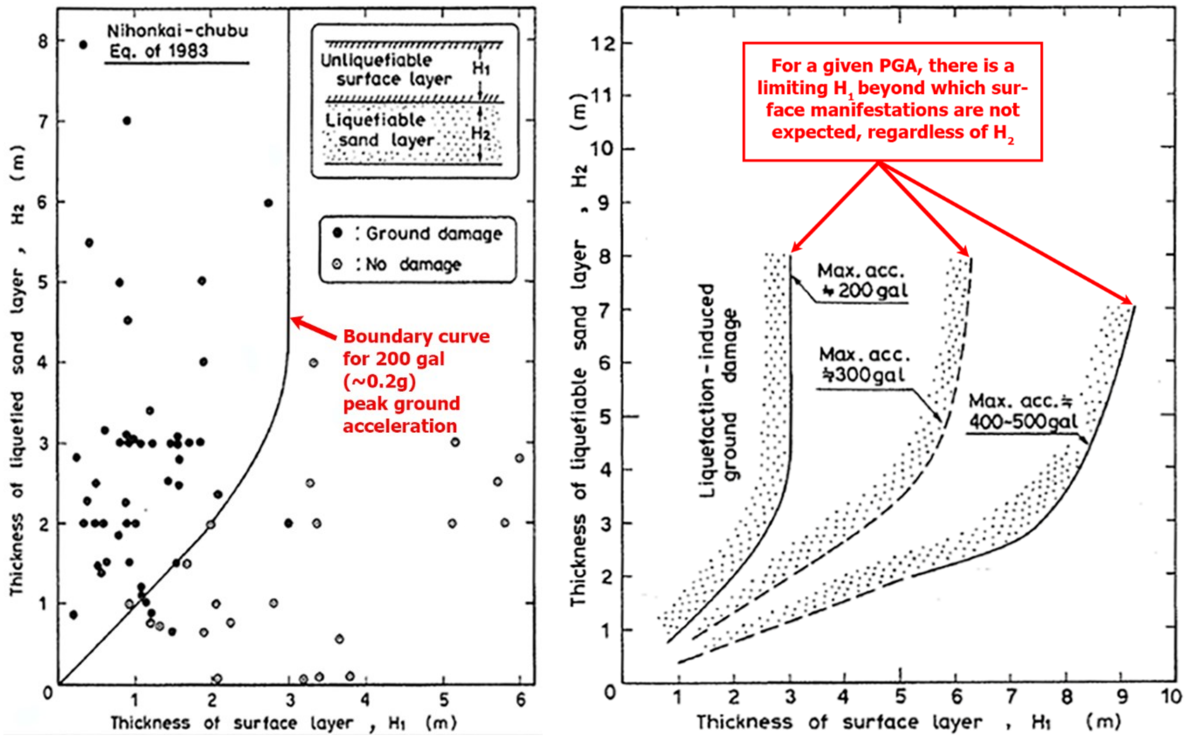


Fig. 1. (a) Conditions of subsurface soil stratification discriminating between the occurrence and non-occurrence of surficial liquefaction manifestation, given a *PGA* of 200 gal (~0.2g); and (b) boundary curves proposed for several different *PGA*s. After Ishihara (1985).

Ishihara's (1985) model, henceforth called the "H_1-H_2" model, has been widely cited since its inception and is programmed in popular software for modeling liquefaction hazards (e.g., *CLiq* by Geologismiki, 2020). However, it was derived from the limited data then-available (~300 data points from two earthquakes) and arguably has not been evaluated rigorously or calibrated since. In addition, alternative manifestation models have been proposed (e.g., Iwasaki et al., 1984; van Ballegooy et al., 2014a; Maurer et al., 2015a) but a test of the H_1-H_2 model against others is absent from the literature. While van Ballegooy et al. (2015) looked for correlations between the H_1-H_2 model and others using hypothetical soil profiles, the predictive efficiencies of these models have not been quantified and compared using case-history data from the field.

Meanwhile, recent events – in particular the 2010-2016 Canterbury, New Zealand, earthquakes – have significantly augmented the liquefaction case-history data available for model training and testing. As an example, the cone-penetration-test (CPT) based datasets of Geyin et al. (2021) and Geyin and Maurer (2021a) collectively contain ~15,000 case histories from 24 events. These data provide a unique opportunity to advance the science of predicting liquefaction and its effects. Accordingly, the objective of this study is to rigorously evaluate and update the seminal H_1-H_2 model for liquefaction surface expression. As part of this effort, new H_1-H_2 models are developed using both traditional analytical functions as well as machine- and deep-learning algorithms. Lastly, using unbiased test data, the original H_1-H_2 model and several updates developed herein are tested against three manifestation models from the literature. In effect, this paper explores the bounds of predicting liquefaction manifestations using H_1, H_2, and PGA. However, this should not be interpreted as an endorsement of the efficiency or sufficiency of these variables. In the following, the creation of training and test datasets is first explained, followed by a description of the models and methods that will be subsequently utilized.

2. Data

14,440 liquefaction case histories compiled from 24 earthquakes will be studied, as summarized in Table 1. However, since most of these cases were compiled from three earthquakes in Canterbury, New Zealand, data from these and the remaining 21 earthquakes will initially be separated. These respective datasets are henceforth called the "Canterbury" and "global" datasets. Each case history includes estimates of PGA and groundwater depth during an earthquake, CPT data, and observations of the presence or absence of liquefaction manifestations at the ground surface. The case history data studied herein are publicly available in digital format.

Table 1. Summary of Liquefaction Case-Histories Analyzed.

Date	Event	Country	Magnitude (M_w)	Number of Case Histories
16/6/1964	Niigata	Japan	7.60	3
9/2/1971	San Fernando	USA	6.60	2
4/2/1975	Haicheng	China	7.00	2
27/7/1976	Tangshan	China	7.60	10
15/10/1979	Imperial Valley	USA	6.53	7
9/6/1980	Victora (Mexicali)	Mexico	6.33	5
26/4/1981	Westmorland	USA	5.90	9
26/5/1983	Nihonkai-Chubu	Japan	7.70	2
28/10/1983	Borah Peak	USA	6.88	3
2/3/1987	Edgecumbe	New Zealand	6.60	23
24/11/1987	Elmore Ranch	USA	6.22	2
24/11/1987	Superstition Hills	USA	6.54	8

18/10/1989	Loma Prieta	USA	6.93	67
17/1/1994	Northridge	USA	6.69	3
16/1/1995	Hyogoken-Nambu	Japan	6.90	21
17/8/1999	Kocaeli	Turkey	7.51	16
20/9/1999	Chi-Chi	Taiwan	7.62	34
8/6/2008	Achaia-Ilia	Greece	6.40	2
4/4/2010	Baja	Mexico	7.20	3
11/3/2011	Tohoku	Japan	9.00	7
20/5/2012	Emilia	Italy	6.10	46
4/10/2010	Darfield	New Zealand	7.10	5371
22/2/2011	Christchurch	New Zealand	6.20	4806
14/2/2016	Christchurch	New Zealand	5.70	4771

The Canterbury data was sourced from Geyin et al. (2020b, 2021), who used the New Zealand Geotechnical Database (2016) to compile liquefaction case-histories from the: (i) M_w7.1, 4 Sept. 2010 Darfield; (ii) M_w6.2, 22 Feb. 2011 Christchurch; and (iii) M_w5.7, 14 Feb. 2016 Christchurch earthquakes. Of the 15,890 case histories compiled by Geyin et al. (2020b, 2021), 14,165 were ultimately analyzed in the present study. In reaching this number, cases were excluded if: (i) the depth at which the CPT began exceeded the depth of groundwater by at least 0.25 m, a situation that could arise when needing to bypass utilities; and (ii) the observed manifestation of liquefaction was lateral spreading, since the H_1-H_2 model is not intended to predict it (i.e., lateral spreading depends on factors not considered by the H_1-H_2 model). In addition, Geyin et al. (2020b, 2021) placed emphasis on compiling case histories from free-field level-ground sites, with the occurrence and severity of liquefaction defined primarily by liquefaction ejecta and ground cracking. Sites with other indications of liquefaction, such as foundation settlements or evidence from ground motions, were not considered. The definition of "surface manifestation" adopted herein is thus generally consistent with that used by Ishihara (1985) to develop the H_1-H_2 model. For the adopted case histories, the severity of manifestation was classified by Geyin et al. (2020b, 2021) as "none," "minor," "moderate," or "severe." To facilitate the evaluation of the H_1-H_2 model and others, these case histories are binomially reclassified as "No Manifestation" and "Manifestation," where the latter classification includes sites where the observed manifestation was at least of minor severity, indicating the presence of sporadic features covering up to 5% of the ground surface within a 10 m radial sample. Of the 14,440 cases analyzed from Canterbury, ~65% are "No Manifestation" and ~35% are "Manifestation." Additional details, and the digital data, may be found in Geyin et al. (2020b, 2021).

The global data was obtained from Geyin and Maurer (2021a), who digitized and merged 275 CPT case histories from numerous publications into a dataset having the same structure as that from

Canterbury. Because many historic case histories are documented in less detail than the recent events in Canterbury, the exact nature and severity of surface manifestation is not always known. Accordingly, while Geyin and Maurer (2021a) binomially classified case histories following the scheme mentioned above, there is undoubtedly some uncertainty. Of the 275 global case histories compiled by Geyin and Maurer (2021a) and analyzed herein, 42% are "No Manifestation" and 58% are "Manifestation." Additional details and the complete global dataset may be obtained from Geyin and Maurer (2021a). To properly recognize all sources of data used to compile the Canterbury and global datasets, a reference list appears in the Appendix for each of the 24 earthquakes.

2.1 Training and Test Sets

This study aims to evaluate and update the H_1-H_2 model using case-history data with diverse geology, geomorphology, seismology, climate, etc., to the degree possible. Towards this goal, the Canterbury data presents a unique opportunity, given its unprecedented size, but also a modeling challenge, given that it dominates the overall dataset. To mitigate sampling bias from source and class imbalance, new training and test sets were created wherein the Canterbury and global data are given equal weighting, and where the quantities of "Manifestation" and "No Manifestation" cases are the same. Each of these biases (i.e., source and class bias) would otherwise be present in the results and in the derived models. To create these datasets, case histories from Canterbury and from other global events were respectively undersampled and oversampled, as illustrated in Fig. 2. The data to be analyzed were first separated into the Canterbury and global datasets, and then again based on whether a case was classified as "Manifestation" or "No Manifestation." These four data subsets were then each randomly split into training (70%) and test (30%) groups. Finally, from these eight subsets, random sampling with replacement was used to create a final dataset having the desired characteristics. This dataset consists of 70% training data (10,108 cases) and 30% test data (4,332 cases), where both are balanced with respect to source (i.e., 50% Canterbury, 50% global) and class (50% Manifestation; 50% No Manifestation).

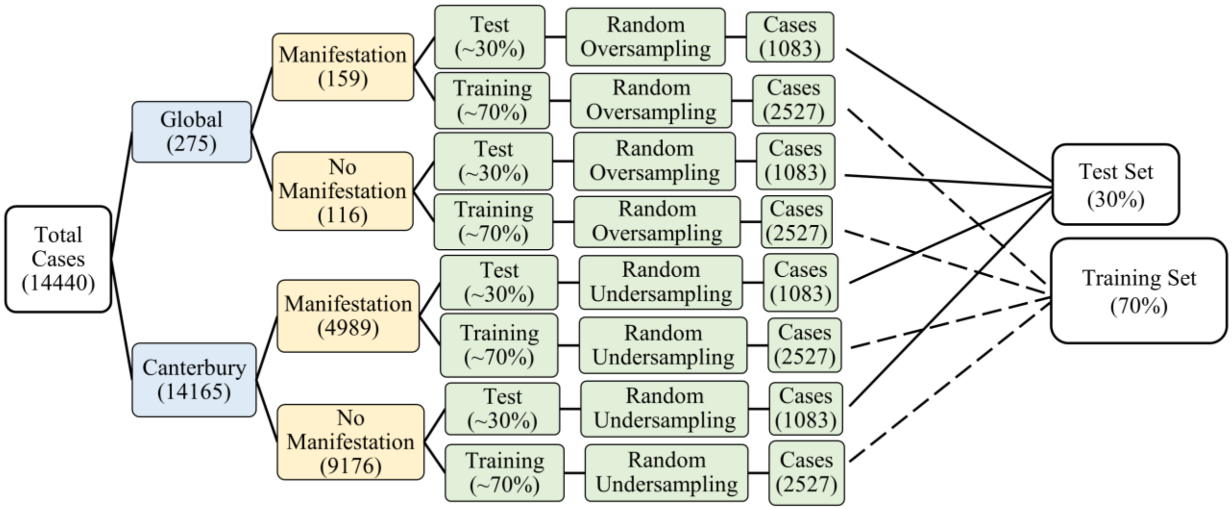


Fig. 2. Schematic illustrating the creation of the training and test sets analyzed herein.

3. Methodology

To evaluate and update the H_1-H_2 model, clear definitions of "H_1" and "H_2" are required. However, as previously discussed in the literature (e.g., van Ballegooy et al., 2015), the definitions originally used by Ishihara (1985) may be interpreted inconsistently, leading to different applications of the H_1-H_2 model. According to Ishihara (1985)**,** H_1 is the minimum depth at which liquefaction is expected (i.e., it is the thickness of the nonliquefied "crust" or "capping" layer) and H_2 is the thickness of soil expected to liquefy. Ishihara (1985) predicted, for example, that a soil subjected to a *PGA* of ~0.2g would likely liquefy if it had a Standard Penetration Test (SPT) blow count less than 10. It is less clear, however, whether H_2 was intended to be the thickness of the shallowest liquefied stratum or the cumulative thickness of all liquefied strata in a soil profile. While the selection of H_2 is straightforward for a profile with one liquefiable stratum, different interpretations may arise for interbedded profiles with multiple such strata. Because Ishihara (1985) predominantly studied profiles consistent with the former, the need to define H_2 in greater detail did not arise.

Accordingly, two definitions of H_2 are initially tested in this study and are demonstrated schematically in Fig. 3: (i) the thickness of the shallowest stratum predicted to liquefy, henceforth called "Case 1"; and (ii) the cumulative thickness of all strata predicted to liquefy, henceforth called "Case 2." Modern implementations of the H_1-H_2 model typically define H_2 in a manner consistent with "Case 2" (van Ballegooy et al., 2015; Geologismiki, 2020). A limiting depth of 10 m is adopted within these definitions for several reasons. First, prior studies of the H_1-H_2

model have used this limiting depth (van Ballegooy et al., 2015). Second, it is widely observed that liquefaction is more likely to manifest at the surface if it occurs at shallow depth (e.g., Iwasaki et al., 1984), yet the H_1-H_2 model does not explicitly account for this behavior when multiple liquefiable strata are present. A 1-m thick liquefied stratum, for example, may be viewed as having the same potential to manifest at the surface whether it is 2 m or 20 m below ground, all else being equal. The use of a limiting depth thus excludes from consideration soils that may liquefy, but which are unlikely to manifest at the surface. Lastly, different limiting depths were provisionally selected and H_1-H_2 models were trained and tested. The performance of these models exhibited a small but systematic decline as the limiting depth increased from 10 m to 20 m, which may be attributable to the shortcoming above. In addition to testing two definitions of H_2, model performance will also be parsed based on how many liquefied strata a site is interpreted to have. Whereas Ishihara (1985) analyzed profiles interpreted to have one liquefied stratum, many of the profiles studied herein are interpreted to have multiple such strata. The Ishihara (1985) H_1-H_2 model might thus perform worse on highly interbedded soil profiles, as has been observed of liquefaction models more generally (e.g., Geyin and Maurer, 2021b). The criteria for categorizing profiles in this manner, and the results of these analyses, will be introduced later in the paper.

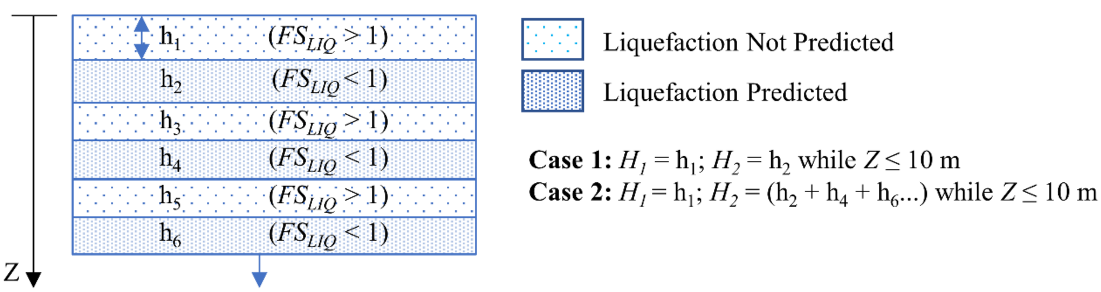


Fig. 3. Schematic demonstrating alternative definitions of "H_1" and "H_2" adopted herein.

To identify strata predicted to liquefy, the Boulanger and Idriss (2014) CPT-based triggering model was adopted, with liquefaction expected when the computed factor-of-safety against liquefaction (FS_{LIQ}) is less than one. Prior to its use, the soil-behavior-type index (I_c) (Robertson and Wride 1998) was used to infer liquefaction susceptibility, such that soils with $I_c > 2.50$ were assumed not susceptible, per Maurer et al. (2019). Ultimately, the most salient results of this study were insensitive to the exact I_c threshold chosen. Additionally, to estimate fines content (FC), a required input to the Boulanger and

Idriss (2014) model, two I_c-FC correlations – one specific to Canterbury (Maurer et al. 2019) and one intended for global application (Boulanger and Idriss 2014) – were used.

While the CPT has advantages over other in-situ tests used to predict liquefaction (e.g., the SPT) (National Research Council, 2016), its effectiveness is still potentially limited by the volume of soil mobilized around the cone. This mobilized zone can act as a low-pass filter, obscuring data from the high spatial frequencies (e.g., that defining a thin soil stratum or the boundary between two dissimilar materials). These filtering effects, often called "thin layer" effects, have been studied by many authors (e.g., Treadwell 1976; Ahmadi and Robertson 2005; van der Linden et al. 2018). Although chart-based correction procedures have been proposed, Boulanger and DeJong (2018) proposed what may be the first algorithmic solution. Termed an "inverse filtering and interface detection" procedure, it predicts "true" CPT data from measured values (i.e., it aims to remove thin layer effects via an inversion process). While current, limited studies do not suggest liquefaction models perform significantly better when using the Boulanger and DeJong (2018) procedure (Geyin and Maurer, 2021b; Yost et al., 2021), it: (i) may nonetheless become popular in practice; (ii) has never been tested in conjunction with the H_1-H_2 model; and (iii) was utilized by Geyin et al. (2021) and Geyin and Maurer (2021a) when compiling the Canterbury and global datasets, respectively, thereby allowing for its use to be easily evaluated. Accordingly, both measured and "true" CPT data will be analyzed and compared throughout this paper. While the reader is referred to Boulanger and DeJong (2018) for complete details, the procedure's default parameters were used to predict "true" CPT data from measured values. These defaults can conceivably be calibrated via site-specific investigation to adjust the "aggression" of the inversion, but such calibrations were not undertaken in the present study and have yet to be demonstrated elsewhere in the literature. As part of their processing methodology, Geyin et al. (2021) and Geyin and Maurer (2021a) used cross-correlation (Buck et al., 2002) to ensure that CPT tip and sleeve measurements were properly aligned.

When evaluating and updating the H_1-H_2 model, three existing, alternative manifestation models will also be implemented and analyzed for comparison. The first of these is the Liquefaction Potential Index (LPI) proposed by Iwasaki et al. (1984):

$$LPI = \int_0^{20\,m} F(FS_{liq}) \cdot w(z)\,\mathrm{d}z \tag{1}$$

where: $F(FS_{liq}) = 1 - FS_{liq}$ for $FS_{liq} \leq 1$ and $F(FS_{liq}) = 0$ otherwise; and $w(z) = 10 - 0.5z$, where z is depth. $F(FS_{liq})$ and $w(z)$ predict the respective influences of FS_{liq} and z on surface manifestation, which is assumed by LPI to depend on the thickness of all liquefied strata within the upper 20 m, the amount by which FS_{liq} is less than 1.0 in each stratum, and the proximity of those strata to the ground surface. Given this definition, the *LPI* domain ranges from zero to 100.

The second model is a modification of *LPI* proposed by Maurer et al. (2015a) and inspired by the H_1-H_2 model. Given its provenance, the result was named LPI_{ISH} and is defined by:

$$LPI_{ISH} = \int_{H_1}^{20\,m} F(FS_{liq}) \cdot w(z)\,\mathrm{d}z \tag{2}$$

where:

$$F(FS_{liq}) = \begin{cases} 1 - FS_{liq} & if\ FS_{liq} \leq 1\ \cap\ H_1 \cdot m(FS_{liq}) \leq 3 \\ 0\ otherwise & \end{cases} \tag{3}$$

$$m(FS_{liq}) = \exp\left(\frac{5}{25.56(1 - FS_{liq})}\right) - 1 \tag{4}$$

In Eqs. 3-4, $F(FS_{liq})$ and $w(z)$ serve the same objective as in *LPI*, but are defined differently, such that $F(FS_{liq})$ accounts for H_1 and $w(z)$ is defined by $w(z) = 25.56 \cdot z^{-1}$. Maurer et al. (2015a) recommended a minimum H_1 of 0.4 m, even if liquefiable soils exist at shallower depths. Using this constraint, the LPI_{ISH} domain also ranges from zero to 100.

The third is the Liquefaction Severity Number (*LSN*) proposed by (van Ballegooy et al., 2014a):

$$LSN = \int_{0}^{20\,m} \varepsilon_v \cdot w(z)\,\mathrm{d}z \tag{5}$$

where ε_v is volumetric strain (%) and $w(z) = 10 \cdot z^{-1}$. While various methods are available to estimate ε_v (e.g., Geyin and Maurer 2019), van Ballegooy et al. (2014a) adopted the Zhang et al. (2002) method, which is thus also adopted herein. The *LSN* domain ranges from zero to ∞ (if liquefiable soils are near the surface) but is typically less than 100.

4. Results and Discussion

4.1 Evaluating the H_1-H_2 Model

The H_1-H_2 model has several limiting traits that complicate its application and evaluation. In addition to the previously discussed interpretation of H_2, which can be ambiguous, it: (i) is a graphical, rather than analytical, solution; and (ii) was defined only for three discrete values of *PGA*, as shown in Fig. 1. Because very few of the 14440 compiled case histories experienced one of these exact *PGAs*, it is advantageous to convert the H_1-H_2 model into a continuous analytical function, such that it may be applied under all circumstances.

Accordingly, several simple functional forms were fit to the H_1-H_2 model (i.e., to the three curves proposed by Ishihara (1985)). Among these, a bilinear function defined by Eqs. 6-8 resulted in: (i) the closest fit to Ishihara's (1985) three curves, as measured by mean square error; (ii) heuristically sensible behavior when extrapolating beyond the curves (i.e., the resulting functions appear plausible at low and high *PGA* based on existing field observations); and (iii) the best prediction performance, as measured on case-history data and discussed subsequently. These equations thus represent a close approximation and plausible extrapolation of the original H_1-H_2 model, as shown in Fig. 4. While the equations are plotted for eight values of *PGA*, only the three curves proposed by Ishihara (1985) and digitized in Fig. 4 were used in the fitting process.

$$H_2 = \begin{cases} m * H_1 & if\ H_1 < H_{1,Lim} \\ \infty & if\ H_1 \geq H_{1,Lim} \end{cases} \tag{6}$$

where:

$$m = 2.13 * e^{-3.751*PGA} \tag{7}$$

$$H_{1,Lim} = 23.234 * PGA - 1.5 \tag{8}$$

In these equations, m represents the initial slope of the H_1-H_2 model and $H_{1,Lim}$ is the limiting value of H_1 beyond which surface manifestations are not expected, regardless of H_2. In developing Eqs. 6-8, various univariate functions (e.g., linear, bilinear, exponential, power) were evaluated. While the adopted model fits the three Ishihara (1985) H_1-H_2 curves well, it is nonetheless one of many models that could be justified based on the curves. It should thus be emphasized that we herein evaluate a fit of the original H_1-H_2 model, rather than the original model itself (which does not lend itself well to evaluation).

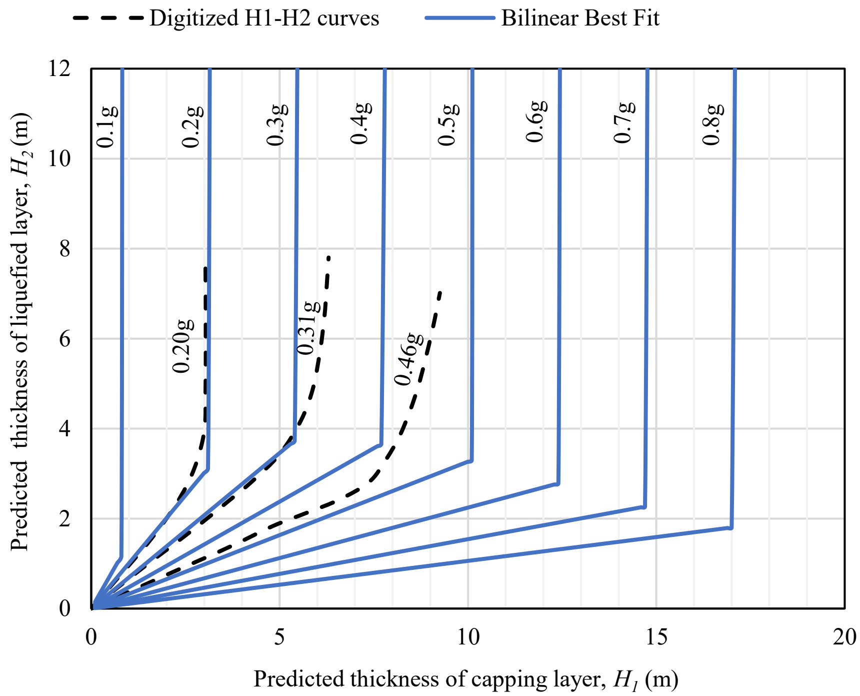


Fig. 4. Comparison of the Ishihara (1985) H_1-H_2 model and the bilinear model fit to the Ishihara (1985) curves and defined by Equations 6-8.

Using the bilinear model fit in Eqs. 6-8, the performance of the H_1-H_2 model was next evaluated using the test set of 4,332 case histories described previously and depicted in Fig. 2. The results, expressed in terms of overall accuracy (*OA*), are summarized in Table 2. *OA* describes the percentage of cases correctly classified as "Manifestation" or "No Manifestation." It is the sum of "true positive" predictions (i.e., manifestations are predicted and observed) and "true negative" predictions (i.e., manifestations are not predicted and are not observed) divided by the total number of cases analyzed. It may be seen that: (i) the "Case 2" definition of H_2 (i.e., the cumulative thickness of strata predicted to liquefy within the upper 10 m) results in a ~15% higher *OA* than the "Case 1" definition of H_2 (i.e., the thickness of the most shallow, discrete stratum predicted to liquefy), for which the H_1-H_2 model performs akin to random guessing; and (ii) the Boulanger and DeJong (2018) CPT correction procedure increases *OA* 2.9% if the "Case 2" definition of H_2 is used but decreases *OA* 3.7% if the "Case 1" definition of H_2 is used. Considering these results, the "Case 2" definition will henceforth be exclusively adopted as the H_1-H_2 model is further tested and improved. Given the inconclusive results of the Boulanger and DeJong (2018) procedure, its use will continue to be evaluated throughout the paper.

Table 2. Performance of the H_1-H_2 model-fit on the test set compiled herein.

Test Dataset	Number of True Negatives	Number of True Positives	Total Number of Cases	Overall Accuracy (*OA*)
Case 1 H_2 + Measured CPT Data	2087	493	4332	0.596
Case 1 H_2 + "True" CPT Data	2160	261	4332	0.559
Case 2 H_2 + Measured CPT Data	1612	1498	4332	0.718
Case 2 H_2 + "True" CPT Data	1637	1597	4332	0.747

4.2 Re-Regressing an Analytical H_1-H_2 Model

While the H_1-H_2 model provided relatively useful predictions in the prior test, its performance ($OA \leq 0.75$) was as near or nearer to random guessing ($OA = 0.5$) as to a perfect model ($OA = 1.0$). To evaluate whether the H_1-H_2 model could be improved if trained on a larger dataset than available to Ishihara (1985), the bilinear functional form in Eqs. 6-8 was next re-regressed using the newly compiled training set, both for measured and true CPT data, via optimization on *OA*. Because the dataset is balanced with respect to positive and negative observations, optimizing on *OA* produces a model without class bias (i.e., it is neither conservative nor unconservative, but seeks to minimize the total rate of mispredictions). The resulting model is defined by Eqs. 9-11:

$$H_2 = \begin{cases} m * H_1 & if\ H_1 < H_{1,Lim} \\ \infty & if\ H_1 \geq H_{1,Lim} \end{cases} \tag{9}$$

where:

$$m = a * PGA^b \tag{10}$$

$$H_{1,Lim} = c * PGA^d \tag{11}$$

In Eqs. 10-11, m and $H_{1,Lim}$ have the same meaning as in Eqs. 7-8 but use different functional forms. When trained on measured CPT data, the model coefficients are $a = 0.1436$; $b = -0.9321$; $c = 27.9483$; and $d = 1.0139$. When trained on "true" CPT data, they are $a = 0.1399$; $b = -0.9881$; $c = 31.1370$; and $d = 0.9908$. Both models are plotted against the Ishihara (1985) curves in Fig. 5. In either case, it can be seen that retraining resulted in an outward shift of the H_1-H_2 functions relative to those proposed by Ishihara (1985). That is, the newly trained model predicts that a profile is more hazardous (i.e., more likely to produce surface manifestation) for a given combination of H_1, H_2, and *PGA*. Given that the H_1-H_2 curves of Ishihara (1985) were drawn to minimize the total rate of mispredictions (i.e., the curves were intended to be neither conservative nor unconservative) – mirroring the approach used herein –

it is unlikely that the shift in perceived hazard is due to the method of optimization. Other, more likely causes are discussed as follows. *First*, Ishihara (1985) studied data from just two earthquakes, each with relatively uncertain ground motions, and constructed one of the three H_1-H_2 curves via interpolation. That the resulting model performs as well as it does despite such limited data is impressive. Nonetheless, the data from these two earthquakes might not be representative of liquefaction data more generally and might therefore not produce a model optimal for global application. *Second*, the CPT-based triggering model used herein to identify liquefied strata (i.e., Boulanger and Idriss, 2014) may deviate from the SPT-based approach used by Ishihara (1985). Specifically, the observed shift in perceived hazard could result if the Boulanger and Idriss (2014) model tends to predict less liquefaction than the method of Ishihara (1985), resulting in a larger H_1 and smaller H_2 for the same profile. *Third*, while the definition of "manifestation" adopted in this work appears consistent with that of Ishihara (1985), it is possible the perceived shift in hazard could result from differing criteria. Specifically, if our threshold for classifying "manifestation" is less severe than that of Ishihara (1985), the resulting H_1-H_2 models would tend to predict that a profile is more hazardous (i.e., more likely to produce surface manifestation). *Lastly*, our definition of H_2 might differ from that of Ishihara (1985), which is to say that Ishihara (1985) studied profiles interpreted to have one liquefied stratum, whereas most profiles in this study are interpreted to have multiple such strata.

It can also be seen from Fig. 5b that use of "true" data results in an additional, minor outward shift of the H_1-H_2 functions. This is likely a consequence of the Boulanger and DeJong (2018) procedure's average tendency to increase the computed FS_{liq}, as studied in detail by Geyin and Maurer (2021b). All else being equal, an increase in FS_{liq} increases H_1 and decreases H_2. Given that this is the most common outcome of the Boulanger and DeJong (2018) procedure, the optimal H_1-H_2 functions correspondingly shift in a similar fashion. Lastly, it may be observed that the re-regressed model is dependent upon *PGA*, even though the adopted functional form allows for a lack of dependence to occur, if statistically supported. This dependence occurs even though *PGA* is already considered within the liquefaction triggering model. The authors hypothesize that *PGA* provides confidence to an otherwise binomial prediction of liquefaction triggering, similar to how *LPI* and other models account for FS_{LIQ}. *LPI*, for example, assumes that an FS_{LIQ} of 0.1 is more likely to produce surface manifestation than an FS_{LIQ} of 0.5, all else being equal, even though from a mechanistic or laboratory perspective (e.g., Yoshimine et al., 2006), the expected outcomes may be the same. The performance of these and other models yet to be developed will be discussed subsequently.

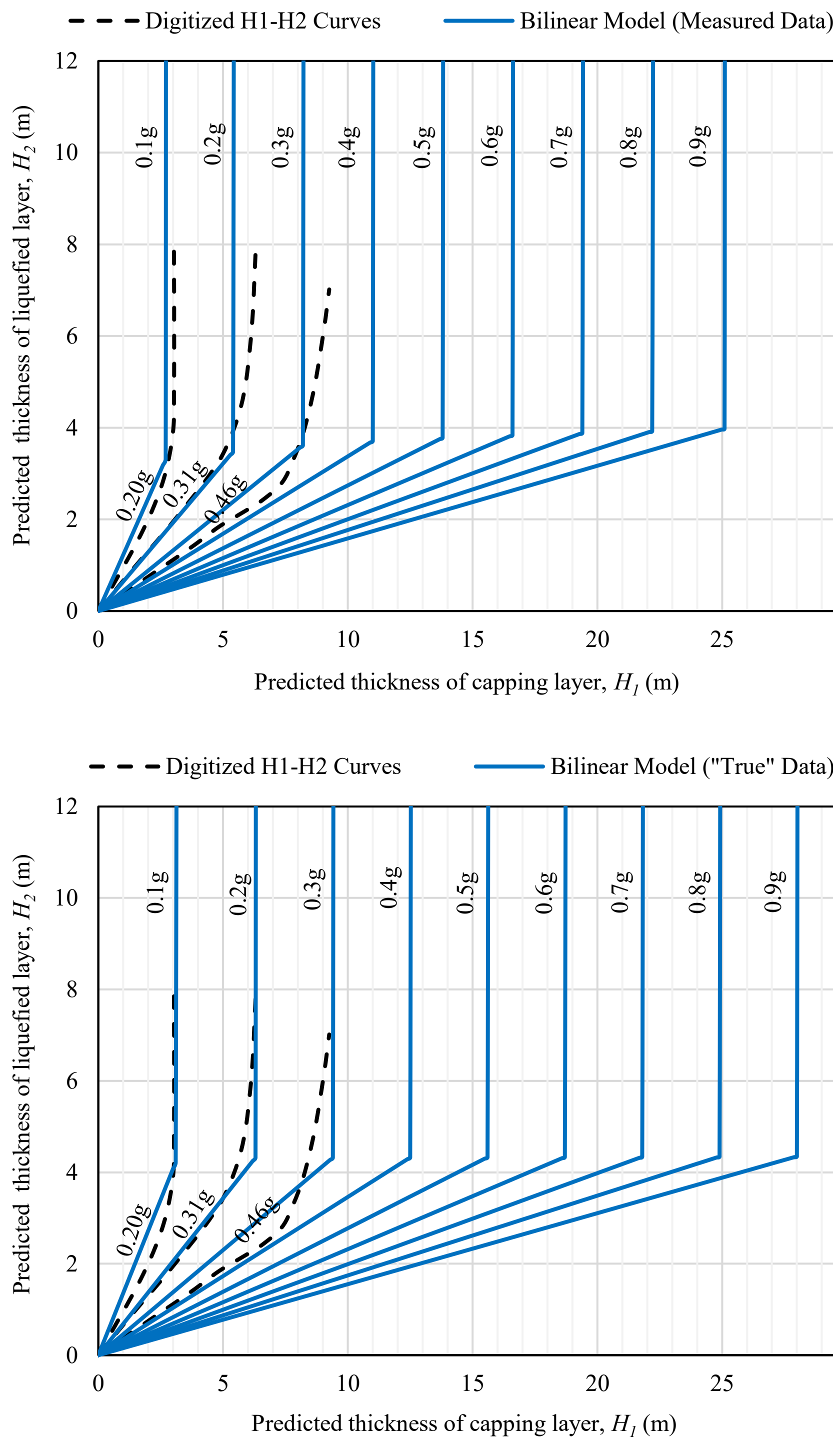


Fig. 5. Comparison of the Ishihara (1985) H_1-H_2 model and the optimized bilinear model defined by Equations 9-11, as trained on **(a)** measured CPT data; and **(b)** "true" CPT data.

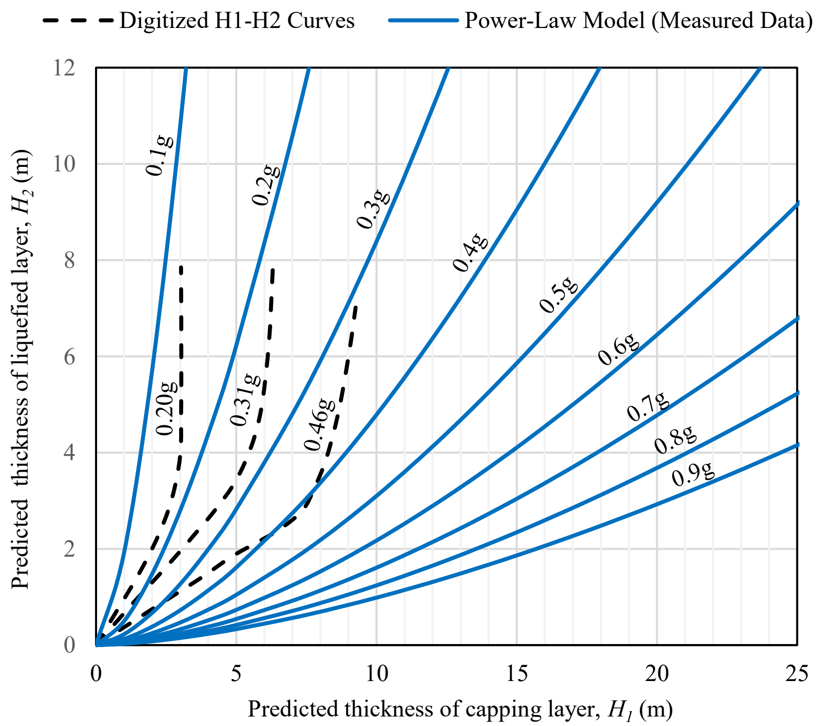
While a bilinear model was found to best fit the Ishihara (1985) curves, other functional forms could potentially result in better predictions of liquefaction manifestations. In re-regressing an analytical H_1-H_2 model, a power-law relation was also developed as an alternative and is defined by Eqs. 12-13:

$$H_2 = m * H1^c \tag{12}$$

where:

$$m = a * PGA^b \tag{13}$$

When trained on measured CPT data, the model coefficients are a = 0.0217; b = -1.9481; and c = 1.5688. When trained on "true" data, a = 0.1087; b = -1.0430; and c = 1.2162. Both models are plotted against the Ishihara (1985) curves in Fig. 6. Like the bilinear model in Fig. 5, the power-law model predicts that a profile is more hazardous (i.e., more likely to produce surface manifestation) for a given combination of H_1, H_2, and PGA, as compared to the Ishihara (1985) curves. It can also be seen that the Boulanger and DeJong (2018) CPT correction procedure most often results in an additional outward shift of the H_1-H_2 functions, as was also the case in the bilinear model. In comparing the new bilinear and power-law models, it is evident the models are in relative agreement when H_1 and H_2 are small (e.g., < 5 m), and particularly when PGA is also small, but tend to have large discrepancies otherwise. This can be attributed to the paucity of training data (i.e., case histories) with relatively large H_1, H_2, or PGA. Shown in Fig. 7 are the distributions of "no manifestation" and "manifestation" data binned on PGA. The lack of data within the aforementioned parameter space is readily apparent, particularly as PGA increases. As a result, models that are very different outside the limits of the empirical data can have the same prediction efficiency. While the proposed models are herein routinely shown beyond these limits to illustrate their extrapolation beyond the data, the models should not be relied on when H_1 or H_2 exceeds 10 m, or when PGA exceeds ~0.7 g. It follows that a more mechanistic-based approach (e.g., Cubrinovski et al., 2019; Bassal and Boulanger, 2021; Hutabarat and Bray, 2021) could be particularly beneficial for resolving predictions of liquefaction response in scenarios lacking prior empirical insights. It can also be surmised from Fig. 7 that the prediction of surface manifestation via predictor variables H_1, H_2 and PGA is unlikely to be highly efficient (say, OA > 0.90), regardless of a model's exact formulation. In other words, the "Manifestation" and "No Manifestation" data points have considerable overlap and are unlikely to be separated without a nontrivial rate of misprediction. Nonetheless, the possibility exists that algorithmic learning, which is not constrained to any functional form, could provide more efficient predictions than the two analytical models proposed in Eqs. 9-13. This possibility is explored in the following section.



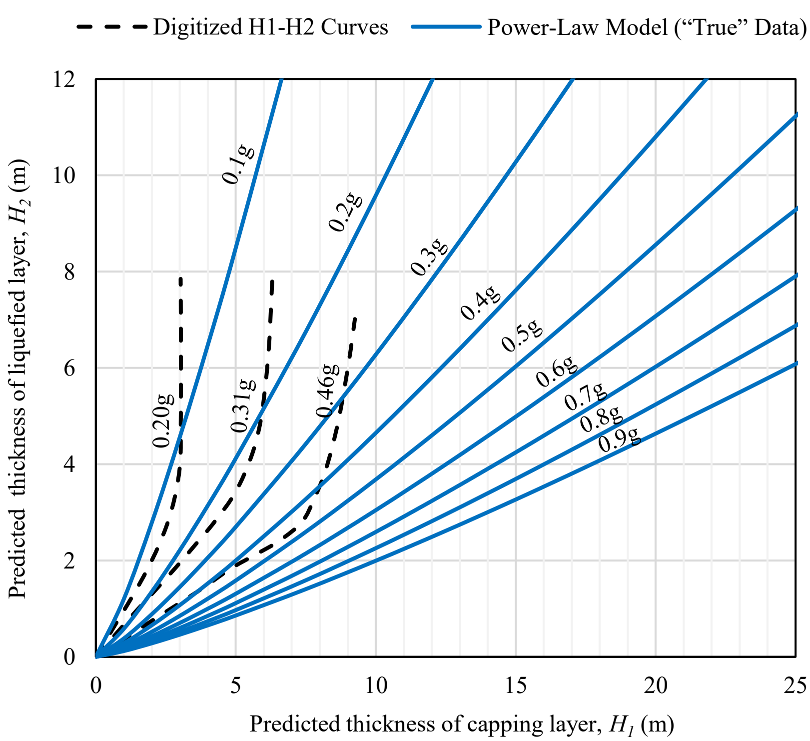


Fig. 6. Comparison of the Ishihara (1985) H_1-H_2 model and the re-regressed power-law model defined by Equations 12-13, as trained on **(a)** measured CPT data; and **(b)** "true" CPT data.

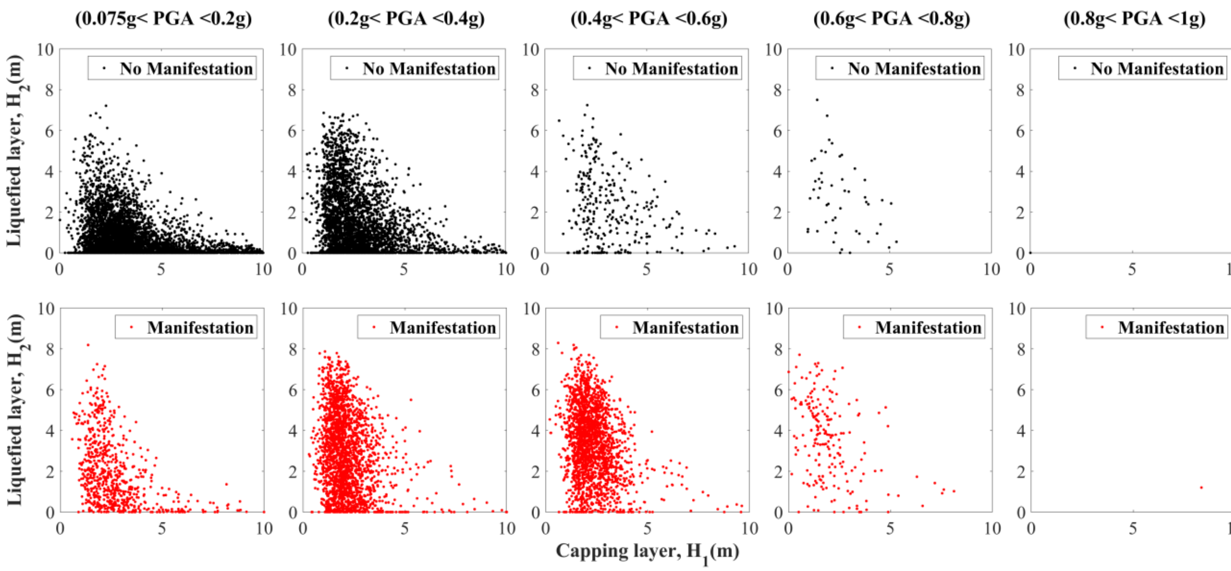


Fig. 7. Distributions of "No Manifestation" and "Manifestation" data, binned on *PGA*.

4.3 Improving the H_1-H_2 Model with Algorithmic Learning

To assess whether algorithmic learning could produce a more effective classifier model via variables H_1, H_2 and *PGA*, various machine and deep learning (ML/DL) algorithms were explored. These included support vector machines (SVM) (e.g., Vapnik, 1995), decision trees (e.g., Rokach and Maimon, 2008) and tree ensembles with random forests, bagging, or gradient boosting (e.g., Breiman, 1996; Piryonesi et al., 2021), Gaussian process models (e.g., Rasmussen and Williams, 2006), and neural networks (e.g., Glorot and Yoshua, 2010). In general, modeling techniques that are interpreted more easily (e.g., single decision trees) tend to have lesser prediction efficiency and are prone to overfitting, whereas those with higher efficiency and better transferability are often relatively complicated to interpret (e.g., ensembles of decision trees). Each technique has various internal options, or "hyperparameters," that can be optimized via an automated search scheme once promising models are identified. Like the previously developed models, prospective ML/DL models were trained on the compiled training set, both for measured and "true" CPT data. Because ML/DL models are especially susceptible to overfitting, 5-fold cross-validation was used, wherein the training data is partitioned into five random subsets of equal size. One subset is used to validate the model trained on the remaining four. This process is repeated five times, such that each subset is used once for validation. The result is a resampling-based ensemble of models trained on different subsets of the training data.

Ultimately, two ML/DL models were selected for testing, with each separately trained on measured and "true" CPT data. The first is a bagged ensemble of decision trees, wherein multiple relatively weak learners are aggregated to form a stronger model. For brevity, we henceforth refer to it as the "ML model." The theory underlying this approach – which is commonly included in machine learning toolkits (e.g., Scipy, TensorFlow) – is explained in detail by Breiman (1996). The growth of a decision tree involves the establishment of recursive binary splits, such that specific combinations of model inputs (i.e., H_1, H_2, or PGA) map to a predicted output – in this case, a binomial classification of 1 (Manifestation) or 0 (No Manifestation). However, because a single tree tends not to be very accurate and is prone to overfitting, ensembles of decision trees are advantageous. In "bagging," multiple versions of a training set are formed by bootstrap sampling, thereby generating multiple models. The predictions from those models are then ensembled, which for a classification problem is the majority vote (i.e., 0 or 1). Using this approach, the ML model was trained via optimization on *OA*, like the previously developed models. While a single decision tree could be elucidated in a schematic, simple interpretations of a tree ensemble are infeasible, given that its strength is derived from the aggregate of numerous models.

The second model is a multilayer feed-forward artificial neural network (NN). For brevity, we henceforth refer to it as the "DL" model (although the defining characteristics of deep vs. shallow learning are debatable). Dating to the 1980's (e.g., Hopfield, 1982), this now ubiquitous approach mimics the perceived structure of the human brain, with layers of interconnected nodes. At the most basic level, NNs have four components: inputs, weights, a threshold, and an output. Connections between nodes are modeled as weights, such that positive and negative weights indicate excitatory and inhibitory connections, respectively. During training, the weights are iteratively adjusted to optimize model performance. If the output from an individual node is above a specified threshold, the node is activated, sending data to the next layer of the network. An activation function then controls the amplitude of the output at each node. The above process is repeated multiple times, with each layer potentially passing information from the previous layer to the next. For this work, the NN was trained using the Levenberg-Marquardt algorithm (Hagan and Menhaj, 1999) which combines the classical gradient descent and Gauss-Newton minimization methods. In addition, a sigmoid function (e.g., Han and Morag, 1995) was adopted for the activation function of the output neuron. This results in output values that are estimates of the probability of the input belonging to a specified class. Thus, unlike the binomial ML model, the output from the DL model is the probability of a positive class (i.e., the probability of surface manifestation). However, like the ML (bagged decision tree) model, NNs are quite convoluted, rendering comprehensive interpretations of the detailed inner workings difficult, since single node weights have little physical meaning, and since many thousands or millions of connections may be present.

4.3.1 Model Availability and Implementation

An obvious limitation of most models trained via algorithmic learning is the lack of a defined equation easily ported and executed via hard copy. Simple depictions of model structure and form are also generally absent. While these detractions can be significant, it is clear the use of algorithmic learning will only grow, given its demonstrated capabilities when provided with large datasets. It is critical, however, that trained ML/DL models be provided as code. Despite this necessity, enumerable ML/DL models have been published without code, meaning that while a model may be available for use by the respective developers, it is not easily accessed by the broader community, and is therefore not applied, tested, or improved upon by others. To facilitate user adoption and evaluation, the trained models are provided via an electronic supplement as Matlab code (i.e., in .m format). The only required inputs are values of H_1, H_2, and PGA in an l x 3 matrix, X, where l is the number of sites at which a prediction is requested. An example of matrix X is also provided in the electronic supplement as an excel file (i.e., in .xls format). The program "H1H2_Execute" requests this input file and makes predictions using any of the models developed in this paper (i.e., both the ML and DL models, as well as the analytical models developed previously). As part of this process, the user is prompted to specify whether they are studying CPT data processed via the Boulanger and DeJong (2018) correction procedure, in which case the models corresponding to its use are selected.

4.4 Analysis and Discussion of Model Performance

Following from the prior performance evaluation of the original H_1-H_2 model, all subsequently trained variants of the H_1-H_2 model were similarly evaluated in terms of OA using the compiled test set. It is worth restating that: (i) all H_1-H_2 models adopt the "Case 2" definition of H_2 in training and testing; and (ii) the tests use data entirely separate from model training and are thus unbiased, to the degree possible (e.g., test and training data do come from the same earthquakes, so some "bias" may be present in that the geomorphic settings and seismic loadings where liquefaction occurs are similar). The results of these performance evaluations are summarized in Fig. 8.

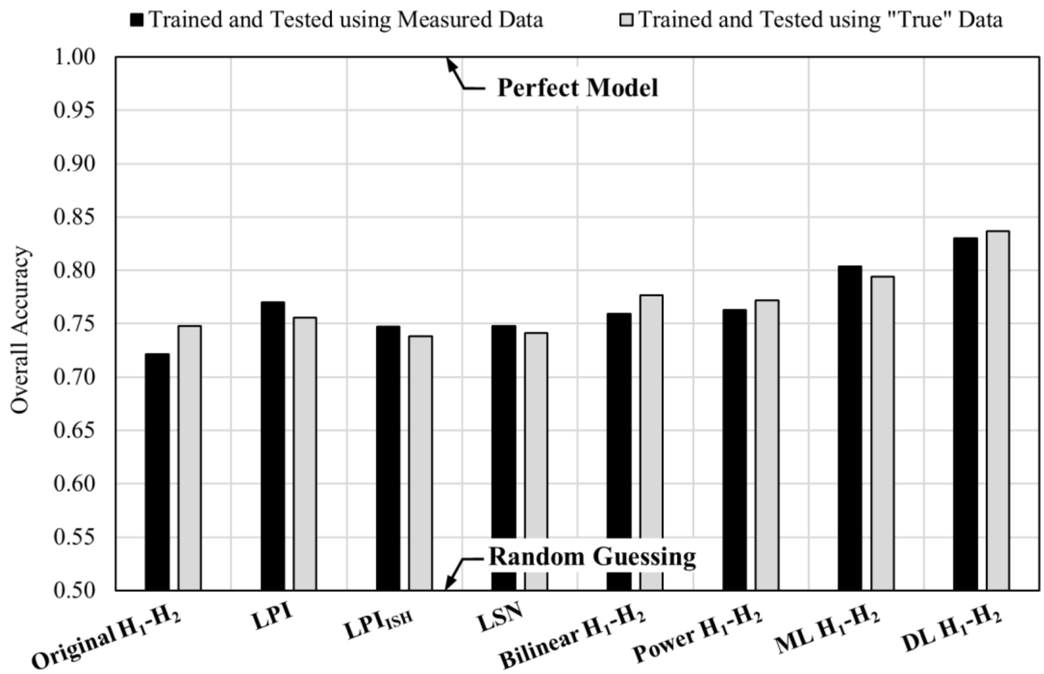


Fig. 8. Model performance (*OA*) measured on the compiled test set.

Notable findings from Fig. 8 are as follows. *First*, the bilinear and power-law H_1-H_2 models outperformed the original H_1-H_2 model (Ishihara, 1985), as fitted herein. Whereas the original model had an average *OA* of 0.732 (i.e., when averaging over the measured and "true" model variants) the bilinear and power models each had average *OA*s of 0.767. In the context of the compiled test set of 4,332 case histories, this amounts to an additional 151 cases for which the predicted and observed responses agree. This improvement is likely attributable to having vastly more data for model training than was originally available when the H_1-H_2 model was devised. *Second*, none of the alternative manifestation models from the literature (i.e., *LPI*, LPI_{ISH}, or *LSN*) performed better on the test set than the newly trained analytical H_1-H_2 models. To compute an *OA* using *LPI*, LPI_{ISH}, or *LSN*, the classification threshold at which *OA* was maximized on the training set (say, $LPI = 5$) was applied to the test set to make forward predictions. With this approach, manifestations are expected for computed values above this threshold and are not expected otherwise. While the relative performance of these models might vary if a different test set were used, it is nonetheless notable that the H_1-H_2 concept performs as well or better than alternative models when quantified via *OA*. It should be emphasized, however, that as a binomial predictor, the H_1-H_2 concept (at least, as originally devised) is not well suited for predicting the *severity* of manifestation. In addition, the degree of misprediction cannot be judged, since all mispredictions are equally erroneous. For this reason, continuous index models like

LPI, LPI_{ISH}, or LSN may have advantages that cannot be quantified via OA. Moreover, the relative performance of LPI, LPI_{ISH}, and LSN might change if a different, more comprehensive metric of performance were used (Geyin et al., 2020a).

Third, the ML and DL versions of the H_1-H_2 model showed further improvement, with average OAs of 0.798 and 0.832, respectively. Relative to the original H_1-H_2 model, this is an increase in OA upwards of 10%, meaning that liquefaction response was correctly predicted at 434 additional sites in the test set. *Lastly*, it can be seen in Fig. 8 that use of the Boulanger and DeJong (2018) CPT correction procedure was inconclusive with respect to improving prediction efficiency. In four of the eight models evaluated, the use of this procedure decreased OA, and in four other models, it increased OA. In general, these changes were relatively minor (±1.2% on average). It may thus be observed that this correction procedure is unlikely to significantly improve the performance of current liquefaction models, a conclusion similarly reached by Geyin and Maurer (2020; 2021b) using different methods. Although not shown in Fig. 8, it was found that inputting measured CPT data to a model trained on "true" data, or inputting "true" CPT data to a model trained on measured data, typically resulted in a more significant decline in performance. Accordingly, the authors recommend neither the use nor disuse of the Boulanger and DeJong (2018) procedure but do recommend that the H_1-H_2 models developed herein be used in a manner consistent with their training. Measured and "true" CPT data and models should not be mixed.

To investigate the specific conditions under which some models perform better or worse, the OA of each model was next computed for various bins of H_1 and H_2 within the test set, as shown in Fig. 9 for models trained and tested on measured CPT data. An analogous figure for models trained on "true" data is not shown because the results are very similar. Shown in Fig. 10 are the quantities of cases occupying each bin, where it can be seen that relatively few cases have H_1 or H_2 exceeding 5 m; the most populous bin is that with 2.5 m < H_1 < 5 m and $H_2 < 2.5$ m. The OA results in Fig. 9 should thus be viewed in the context of Fig. 10 since each bin contributes differently to overall model performance. Notable findings from Figs. 9 and 10 are as follows.

First, all models perform relatively better when either H_1 or H_2 is relatively large. In these situations, the predicted liquefied stratum is either thin and at large depth, or thick and at shallow depth. The presence or absence of surface manifestation is more easily predicted in either case. In contrast, all models perform relatively worse when H_1 and H_2 are relatively small (but nonzero). In this situation, a relatively thin liquefied stratum is predicted at relatively shallow depth, in which case it is more difficult to predict whether surface manifestations will, or will not, be observed. In fact, all models are nearer to random guessing than to a perfect model when H_1 and H_2 are between 0 m and 2.5 m. This is a significant finding, given that many such cases are encountered in practice (and represent ~20% of the test set in this study). In such cases, a more advanced, asset-specific assessment of liquefaction hazard may be particularly useful. Ultimately, this points to the apparent inefficiency and/or insufficiency of predictor variables H_1, H_2, and PGA for predicting surface manifestation. It is known, for example, that low permeability soils interbedded within a profile may complicate that profile's expected response by altering the triggering of liquefaction and/or the morphology of surface expression (e.g., Maurer et al., 2015b; Cubrinovski et al., 2019; Bassal and Boulanger, 2021; Hutabarat and Bray, 2021). Yet, while the presence and sequencing of low permeability soils are known to affect pore pressure generation and transmission, these variables are not explicitly included in the H_1-H_2 concept nor in LPI, LPI_{ISH}, or LSN. Moreover, these methods implicitly assume that liquefiable strata are independent entities, and that liquefaction is concurrent throughout a profile, when in reality a profile's system response may give rise to time-varying values of H_1 and H_2 (e.g., Cubrinovski et al., 2019; Hutabarat and Bray, 2021).

Second, it can be seen for the newly developed models that improvement is relatively uniform across the parameter domain. That is, when OA increases, it tends to increase similarly across a range of H_1 and H_2 values. As such, we do not recommend that a model be selected on the specific basis of H_1 and H_2. With very few exceptions, the four newly trained H_1-H_2 models are more efficient than the original model across all H_1 and H_2. These new models could be ensembled (e.g., averaged), however, as is common in the prediction of other hazards (e.g., ground motions and storm tracks). The ensembling of models with different forms can have the effect of stabilizing predictions (i.e., minimize large swings on account of which model is chosen), and potentially, provide benefits unrealized during testing. Accordingly, the aforementioned "H1H2_Execute" program provides the option to average predictions from all newly proposed models.

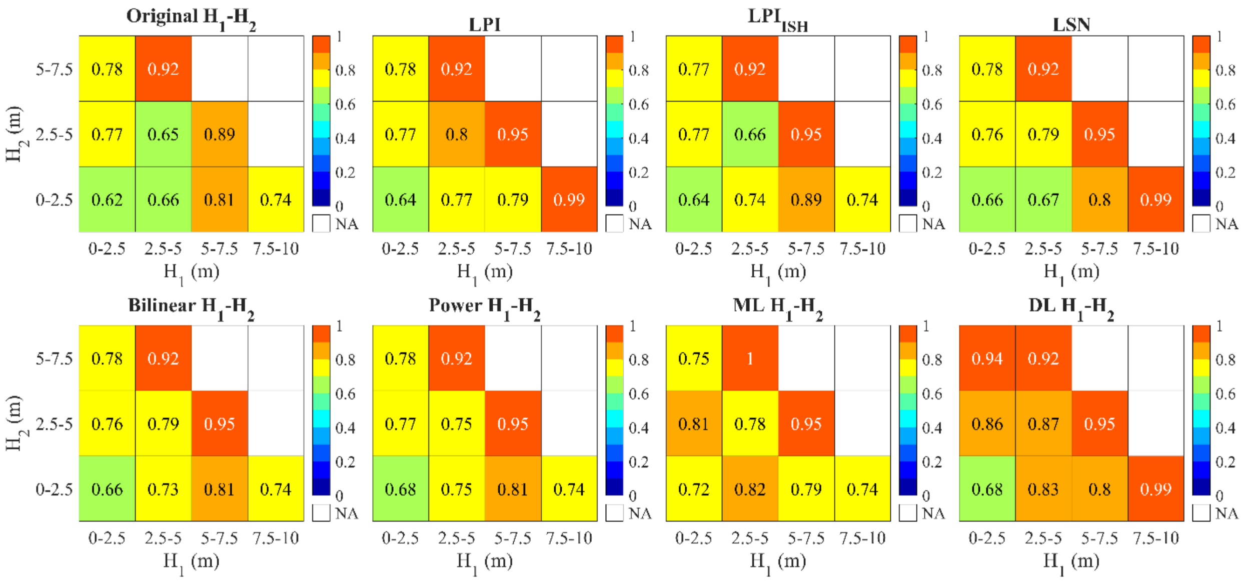


Fig. 9. Model performance (*OA*) on the compiled test set, binned on H_1 and H_2 values. All models are trained and tested on measured CPT data. The value within each bin is the *OA*.

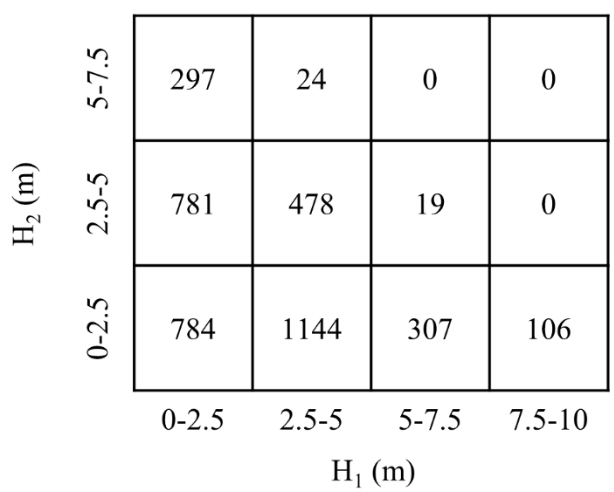


Fig. 10. Distribution of data in the compiled test set, binned on H_1 and H_2 values.

Finally, to investigate whether the H_1-H_2 models perform relatively better on some types of profiles than on others, model performance was parsed based on the number of strata with computed $FS_{liq} < 1.0$. The results are shown in Fig. 11 for models trained and tested on measured CPT data. An analogous figure for "true" data is not shown because the results demonstrate the same trends. All H_1-H_2 models perform relatively worse when multiple liquefied strata are present, as has been observed of models more generally (e.g., Geyin and Maurer, 2021b), and which is also true of *LPI*, LPI_{ISH}, and *LSN*. Across all models evaluated in Fig. 11, the *OA* declines an average of 8.6% when the interpreted quantity of liquefied strata exceeds one. Additionally, and as previously discussed, Ishihara (1985) analyzed profiles interpreted to have one such stratum, whereas in practice many profiles are interpreted to have multiple such strata. It can be seen in Fig. 11 when comparing the original H_1-H_2 model to the newly trained analytical models (i.e., the bilinear and power-law models), that relatively more improvement is achieved on profiles with multiple liquefied strata. In other words, the original H_1-H_2 model performs nearly as well as others if applied only to the types of profiles studied by Ishihara (1985). The overall improvement of the bilinear and power-law models may thus be partly attributable to differences between the case histories available to Ishihara (1985) and those studied herein. However, it can also be seen that the ML and DL models improve *OA* not only on interbedded profiles, but across all profiles investigated. Ultimately, regardless of which H_1-H_2 model is adopted, it can be expected that performance will degrade in highly interbedded profiles.

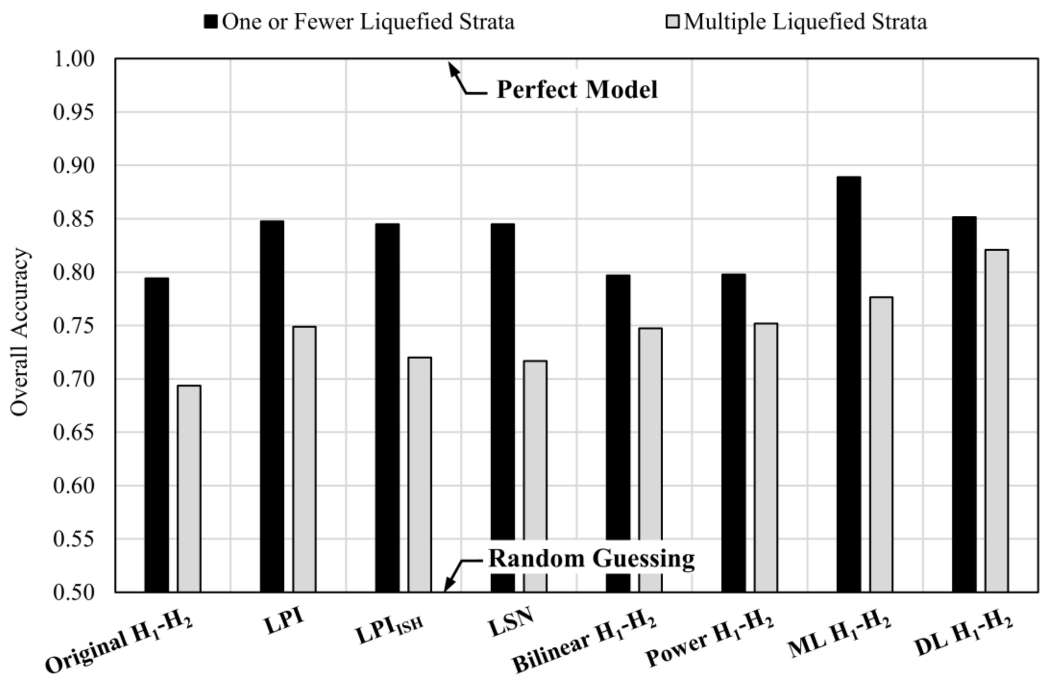


Fig. 11. Model performance (*OA*) measured on the compiled test set (measured CPT data), parsed by the number of liquefied strata a profile is interpreted to have.

4.5 Limitations and Uncertainties

It should be emphasized that “surface manifestation” – the prediction target of the analyses presented herein – refers to free field surface ejecta, settlement, and cracking on ground that is generally level, consistent with the definition adopted by Ishihara (1985). While these effects are a popular and pragmatic proxy for the damage potential of infrastructure, such damage may not necessarily occur. Similarly, liquefaction could trigger at depth and damage infrastructure without otherwise expressing in the free field. The limitations of the presented H_1-H_2 models for predicting damage to specific assets, or for predicting other manifestations of liquefaction (e.g., lateral spreading) should be understood by users. It should also be understood that our evaluation of the original H_1-H_2 model is that of a model fit to the three curves proposed by Ishihara (1985). While this fit is plausible, it is nonetheless one of many models that could be justified from the three curves. In addition, the models and findings presented herein are dependent on the data available for analysis. Due to the relative paucity of case histories available outside of Canterbury, New Zealand, the global and Canterbury datasets were respectively upsampled and downsampled to mitigate source and class bias. Inherent to these datasets, it should also be noted that the relative timing of in-situ testing varies (i.e., in some cases it was performed before an earthquake while in others it was performed after). Also present are cases in which a site subjected to multiple earthquakes was used to develop multiple case histories. As discussed in detail by Geyin et al. (2021), this practice is consistent with past precedent but also worthy of further investigation, given that variations in the relative timing of in-situ testing could introduce uncertainty to the models. Invariably, any changes to the training or test sets could result in changes to the models and model performance. The applicability of the findings presented herein to other datasets or other earthquakes elsewhere, especially in regions unrepresented in model training, is thus unknown. Similarly, all analyses were performed using the Boulanger and Idriss (2014) CPT-based triggering model, from which H_1 and H_2 were computed. While this model is popular and has been shown to be at least as effective as all others (Geyin et al., 2020a), adoption of a different model would invariably change the results, albeit any change would likely be inconsequential to the overall conclusions of the work. The use of SPT-based models to compute H_1 and H_2 for input to the proposed models would likewise introduce uncertainty. While it has been suggested that CPT-based characterizations may result in higher hazards than those based on the SPT (Lenz and Baise, 2007), such investigations are rare, and the issue remains unresolved. Thus, while the proposed H_1-H_2 manifestation models could be used with triggering models other than Boulanger and Idriss (2014), the appropriateness of doing so is unknown and subject to further evaluation. Ideally, and as expanded upon below, the H_1-H_2 models would be applied in a manner consistent with their training to mitigate potential bias.

5. Conclusions

While the H_1-H_2 model (Ishihara, 1985) for predicting surficial manifestations of liquefaction has been widely used since its inception, it has arguably not been rigorously evaluated or calibrated since.

Accordingly, this study evaluated and updated the seminal H_1-H_2 model using a large database of CPT-based liquefaction case histories compiled from global earthquakes. Some of these data were omitted from model training to perform unbiased testing. Using this test set, a fit of the original H_1-H_2 model had relatively similar (albeit lesser) *OA* when compared to other, often newer manifestation models from the literature. Benefiting from the large increase in liquefaction case histories available for model training, four new H_1-H_2 models were developed using the same predictor variables as Ishihara (1985). Two were trained via traditional regression and are provided in Eqs. 9-13. Two others were trained via ML/DL algorithms and are provided as simple-to-use codes. Relative to the original H_1-H_2 model, these new variants all increased *OA* on the unbiased test set, especially on profiles with multiple liquefied strata, with the ML and DL models performing best. Two versions of each model were developed to allow for either measured or "true" CPT data to be used, where the latter is data processed using the Boulanger and DeJong (2018) thin layer and interface correction procedure. While the results do not indicate a consistent performance improvement using this procedure, it is important that the H_1-H_2 models be employed in a manner consistent with their development. That is, measured CPT data should not be input to a model trained on "true" data, or vice versa. Similarly, the models should be used in conjunction with the Boulanger and Idriss (2014) triggering model and the definitions of H_1 and H_2 adopted herein. Given these definitions, and considering the limits of the empirical data, the models should not be relied on when H_1 or H_2 exceeds 10 m, or when *PGA* exceeds ~0.7 g. Ultimately, while this paper proposed manifestation models that appear, based on *OA*, to outperform existing models, it also explored the limits of predictor variables H_1, H_2, and *PGA*. As seen in Fig. 7, sites with and without manifestations cannot be separated in a highly efficient manner using these predictors, no matter how complex a modeling technique is used. In the short term, this paper provides updated H_1-H_2 models suitable for application. In the long term, new manifestation models are needed to account for influential factors in a more explicit and mechanistic manner (e.g., the effects of strata permeability, sequencing, depth, and thickness on pore pressure gradients and transmission).

Acknowledgements

The presented study is based upon work supported by the National Science Foundation (NSF) under Grant No. CMMI-1751216. The authors also wish to acknowledge the numerous researchers who contributed to the liquefaction case-history data studied herein. Much of this data was collected under the sponsorship of either the New Zealand Earthquake Commission (EQC), the U.S. Geological Survey (USGS), the Pacific Earthquake Engineering Research Center (PEER), or NSF. However, any

opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF, EQC, USGS, OR PEER.

Appendix: Data Attribution

The curated Canterbury and global datasets are available in digital format, as cited within the text. These datasets were compiled from the following sources, parsed by event:

1964 M_w7.6 Niigata, JPN - Ishihara & Koga (1981), Farrar (1990), Moss et al. (2003); **1971 M_w6.6 San Fernando, USA** - Bennett et al., 1998, Toprak and Holzer (2003); **1975 M_w7.0 Haicheng, CHN** - Arulanandan et al. (1986), Shengcong & Tatsuoka (1984); **1976 M_w7.6 Tangshan, CHN** - Shibata & Teparaska (1988), Moss et al. (2009; 2011); **1979 M_w6.53 Imperial Valley, USA** - Diaz-Rodriguez (1984), Diaz-Rodriguez and Armijo-Palaio (1991), Moss et al. (2003); **1981 M_w5.9 Westmoreland, USA** - Bennett et al. (1984), Seed et al. (1984), Cetin et al. (2000), Moss et al. (2005); **1983 M_w7.7 Nihonkai-Chubu, JPN** - Farrar (1990); **1983 M_w6.88 Borah Peak, USA** - Andrus (1986), Andrus & Youd (1987), Moss et al. (2003); **1987 M_w6.6 Edgecumbe, NZ** - Christensen (1995), Moss et al. (2003); **1987 M_w6.54 Superstition Hills, USA** - Bennett et al. (1984), Cetin et al. (2000), Toprak & Holzer (2003), Moss et al. (2005), Holzer & Youd (2007); **1989 M_w6.93 Loma Prieta, USA** - Mitchell et al. (1994), Pass (1994), Bennett & Tinsely (1995), Boulanger et al. (1995; 1997), Kayen et al. (1998), Toprak & Holzer (2003), Youd & Carter (2005); **1994 M_w6.69 Northridge, USA** - Abdel-Haq & Hryciw (1998), Bennett et al., 1998, Holzer et al. (1999), Moss et al. (2003); **1995 M_w6.9 Hyogoken-Nambu, JPN** - Suzuki et al. (2003); **1999 M_w7.51 Kocaeli, TUR** - PEER (2000a), Youd et al. (2009); **1999 M_w7.62 Chi-Chi, TWN** - Lee et al. (2000), PEER (2000b); **2008 M_w6.4 Achaia-Ilia, GRC** - Batilas et al. (2014); **2008 M_w7.2 El Mayor-Cucapah, MEX** - Moss et al. (2005); CESMD (2016), Turner et al. (2016); **2011 M_w9 Tohoku, JPN** - Cox et al. (2013), Boulanger & Idriss (2014); **2012 M_w6.1 Emilia, ITA** - Papathanassiou et al. (2015), Facciorusso et al. (2015), Servizio Geologico (2016); **2010 M_w7.1, 2011 M_w6.2, and 2016 M_w5.7 Canterbury, NZ** - Bradley (2013); Green et al. (2014); Maurer et al. (2014); van Ballegooy et al. (2014b); Maurer et al. (2015c); New Zealand Geotechnical Database (2016); Quigley et al. (2016).

References

Abdel-Haq, A., and Hryciw, R. D., 1998. Ground settlement in Simi Valley following the Northridge earthquake. Journal of Geotechnical and Geoenvironmental Engineering 124(1), 80-89.

Ahmadi, M.M., and Robertson, P.K., 2005. Thin-layer effects on the CPT q_c measurement. Canadian Geotechnical Journal 42(5), 1302-1317.

Andrus, R.D., 1986. Subsurface investigations on a liquefaction induced lateral spread, Thousand Springs Valley, Idaho: liquefaction recurrence and a case history in gravel. M.S. Thesis, Dept. of Civil Engineering, Brigham Young University, Provo, Utah.

Andrus, R.D. and Youd, T. L. 1987. Subsurface investigation of a liquefaction-induced lateral spread Thousand Springs Valley, Idaho. Misc. paper GL-87-8, U.S. Army Corps of Engineers.

Arulanandan, K., Yogachandran, C., Meegoda, N.J., Ying, L., and Zhauji, S., 1986. Comparison of the SPT, CPT, SV and electrical methods of evaluating earthquake induced liquefaction susceptibility in Ying Kou City during the Haicheng Earthquake. Geotechnical Special Publication 6, 389–415, ASCE.

Bassal, P.C., and Boulanger, R.W., 2021. System response of an interlayered deposit with spatially preferential liquefaction manifestations. Journal of Geotechnical and Geoenvironmental Engineering, 147(12), 05021013.

Batilas, A., Pelekis, P., Vlachakis, V., and Athanasopoulos, G., 2014. Soil Liquefaction/ Nonliquefaction in the Achaia-Ilia (Greece) 2008 Earthquake: field evidence, site characterization and ground motion assessment. International Journal of Geoengineering Case histories 2(4), 270-287.

Bennett, M.J., McLaughlin, P.V., Sarmiento, J.S., and Youd, T. L., 1984. Geotechnical investigation of liquefaction sites, Imperial Valley, CA. U.S. Geological Survey, Open-File Report 84-252.

Bennett, M.J., and Tinsley, J.C. III, 1995. Geotechnical data from surface and subsurface samples outside of and within liquefaction-related ground failures caused by the October 17, 1989, Loma Prieta earthquake, Santa Cruz and Monterey Counties, CA. U.S. Geological Survey Open-File Report 95-663.

Bennett, M.J., Ponti, D.J., Tinsley, J.C., III, Holzer, T.L., and Conaway, C.H., 1998. Subsurface geotechnical investigations near sites of ground deformations caused by the January 17, 1994, Northridge, California, earthquake. U.S. Geological Survey, Open-File Report 98-373.

Boulanger, R.W., Mejia, L.H., and Idriss, I.M., 1997. Liquefaction at Moss Landing during Loma Prieta earthquake. Journal of Geotechnical and Geoenvironmental Engineering 123(5), 453–67.

Boulanger, R.W., Idriss, I.M., and Mejia, L.H., 1995. Investigation and evaluation of liquefaction related ground displacements at Moss Landing during the 1989 Loma Prieta earthquake. Report No. UCD/CGM-95/02, Center for Geotechnical Modeling, University of California, Davis, 231 pp.

Boulanger, R.W. and Idriss, I.M., 2014. CPT and SPT based liquefaction triggering procedures. Center for Geotech. Modeling, Report No. UCD/CGM-14/01, University of California, Davis.

Boulanger, R.W. and DeJong, J.T., 2018. Inverse filtering procedure to correct cone penetration data for thin-layer and transition effects. Cone Penetration Testing 2018, Hicks, Pisano, and Peuchen, eds., Delft University of Technology, The Netherlands, 25-44.

Bradley, B.A., 2013. Site-specific and spatially-distributed ground motion intensity estimation in the 2010-2011 Christchurch earthquakes. Soil Dynamics and Earthquake Engineering 48, 35-47.

Breiman, L.,1996. Bagging predictors. Machine Learning 24(2), 123–140.

Buck, J.R., Daniel, M.M, and Singer, A.C, 2002. Computer Explorations in Signals and Systems Using MATLAB®, 2nd Edition. Upper Saddle River, NJ: Prentice Hall.

CESMD (Center for Engineering Strong Motion Data), 2016. U.S. structural and ground response data; www.strongmotioncenter.org/

Cetin, K.O., Seed, R.B., Moss, R.E.S., Der Kiureghian, A.K., Tokimatsu, K., Harder, L.F., and Kayen, R.E., 2000. Field Performance Case Histories for SPT-Based Evaluation of Soil Liquefaction Triggering Hazard, Geotechnical Engineering Research Report No. UCB/GT-2000/09, UC Berkeley.

Christensen, S.A., 1995. Liquefaction of cohesionless soils in the March 2, 1987 Edgecumbe earthquake, Bay of Plenty, New Zealand, and other earthquakes, Master of Engineering Thesis, Department of Civil Engineering, University of Canterbury, Christchurch, New Zealand.

Cox, B.R., Boulanger, R.W., Tokimatsu, K., Wood, C., Abe, A., Ashford, S., Donahue, J., Ishihara, K., Kayen, R., Katsumata, K., Kishida, T., Kokusho, T., Mason, B., Moss, R., Stewart, J., Tohyama, K., and Zekkos, D., 2013. Liquefaction at strong motion stations in Urayasu City during the 2011 Tohoku-Oki earthquake. Earthquake Spectra 29(S1), S55-S80.

Cubrinovski, M., Rhodes, A., Ntritsos, N., and Van Ballegooy, S., 2019. System response of liquefiable deposits. Soil Dynamics and Earthquake Engineering, 124, 212-229.

Diaz-Rodriquez, J.A., 1984. Liquefaction in the Mexicali Valley during the earthquake of June 9, 1980. Eighth World Conference on Earthquake Engineering, San Francisco, 223-230.

Diaz-Rodriquez, J.A., and Armijo-Palacio, G., 1991. Liquefaction potential of fine cohesionless soils using the CPT. Soils and Foundations 31(3), 111-119.

Facciorusso, J., Madiai, C., and Vannucchi, G., 2015. CPT-Based Liquefaction Case History from the 2012 Emilia Earthquake in Italy. Journal of Geotechnical and Geoenvironmental Engineering 141(12), 05015002.

Farrar, J.A., 1990. Study of in situ testing for evaluation of liquefaction resistance, R-90-06, U.S. Department of the Interior, Bureau of Reclamation, Geotechnical Services Branch, Denver Office.

Gao, H., Hu, B., and Chang, D. 1983. Some geological considerations for the damage during the Tangshan earthquake." North China Earthquake Sciences 1, 64-72.

Geologismiki, 2020. CLiq v3, CPT soil liquefaction software, https://geologismiki.gr/products/cliq/. Geologismiki Geotechnical Software, Serres, Greece.

Geyin, M. and Maurer, B.W., 2019. An analysis of liquefaction-induced free-field ground settlement using 1,000+ case-histories: observations vs. state-of-practice predictions. ASCE Geotechnical Special Publication 308, 489-498.

Geyin, M. and Maurer, B.W., 2020. Fragility Functions for Liquefaction-Induced Ground Failure. Journal of Geotechnical and Geoenvironmental Engineering 146(12), 04020142.

Geyin, M., and Maurer, B.W. 2021a. CPT-Based Liquefaction Case Histories from Global Earthquakes: A Digital Dataset (Version 1). DesignSafe-CI. https://doi.org/10.17603/.

Geyin, M., and Maurer, B.W., 2021b. Evaluation of a cone penetration test thin-layer correction procedure in the context of global liquefaction model performance. Engineering Geology, 106221.

Geyin, M., Baird, A.J. and Maurer, B.W., 2020a. Field assessment of liquefaction prediction models based on geotechnical vs. geospatial data, with lessons for each. Earthquake Spectra 36(3), 1386–1411.

Geyin, M., Maurer, B., Bradley, B.A, Green, R., van Ballegooy, S., 2020b. CPT-Based Liquefaction Case Histories Resulting from the 2010-2016 Canterbury, New Zealand, Earthquakes: A Curated Digital Dataset (Version 2). DesignSafe CI. https://doi.org/10.17603/ds2-tygh-ht91.

Geyin, M., Maurer, B.W., Bradley, B.A., Green, R.A., & van Ballegooy, S., 2021. CPT-based liquefaction case histories compiled from three earthquakes in Canterbury, New Zealand. Earthquake Spectra 37(4), 2920-2945.

Glorot, X. and Yoshua B., 2010. Understanding the difficulty of training deep feedforward neural networks. Proceedings of the thirteenth international conference on artificial intelligence and statistics, 249–256.

Green, R.A., Cubrinovski, M., Cox, B., Wood, C., Wotherspoon, L., Bradley, B, and Maurer, B., 2014. Select Liquefaction Case Histories from the 2010-2011 Canterbury Earthquake Sequence. Earthquake Spectra 30(1), 131-153.

Hagan, M.T., and M. Menhaj, 1999. Training feed-forward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks, 5(6), 989–993.

Han, J., and Morag, C., 1995. The influence of the sigmoid function parameters on the speed of backpropagation learning. In Mira, José; Sandoval, Francisco (eds.). From Natural to Artificial Neural Computation. Lecture Notes in Computer Science, 930, 195–201. doi:10.1007/3-540-59497-3_175.

Holzer, T.L., Bennett, M.J., Ponti, D.J., and Tinsley, J.C., 1999. Liquefaction and soil failure during 1994 Northridge earthquake. Journal of Geotechnical and Geoenvironmental Engineering 125(6), 438-452.

Holzer, T.L., and Youd, T.L., 2007. Liquefaction, ground oscillation, and soil deformation at the wildlife array, California. Bulletin of the Seismological Society of America 97(3), 961-976.

Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. U.S.A. 79(8), 2554–2558.

Hutabarat, D., and Bray, J. D., 2021. Effective stress analysis of liquefiable sites to estimate the severity of sediment ejecta. Journal of Geotechnical and Geoenvironmental Engineering, 147(5), 04021024.

Ishihara, K., 1985. Stability of natural deposits during earthquakes. In: Proceedings of the 11th International Conference on Soil Mechanics and Foundation Engineering. San Francisco, CA, USA, 1, 321–376.

Ishihara, K., and Koga, Y., 1981. Case studies of liquefaction in the 1964 Niigata earthquake. Soils and Foundations 21(3), 35-52.

Iwasaki, T., Arakawa, T., Tokida, K., 1984. Simplified procedures for assessing soil liquefaction during earthquakes. Soil Dynamics and Earthquake Engineering 3(1), 49-58.

Juang, C.H., Yang, S.H., Yuan, H., and Fang, S.Y., 2005. Liquefaction in the Chi-Chi earthquake-effect of fines and capping non-liquefiable layers. Soils and Foundations 45(6), 89-101.

Kayen, R.E., Mitchell, J.K., Seed, R.B., and Nishio, S., 1998. Soil liquefaction in the east bay during the earthquake, The Loma Prieta, California, Earthquake of October 17, 1989 – Liquefaction. Thomas L. Holzer, editor, USGS Professional Paper 1551-B, B61-B86.

Lee, D.H., Ku, C.S., and Juang, C.H., 2000. Preliminary investigation of soil liquefaction in the 1999 Chi-Chi, Taiwan, earthquake, International Workshop on Annual Commemoration of Chi-Chi Earthquake, Vol. III – Geotechnical Aspect, C.H. Loh and W.I. Liao, eds., National Center for Research on Earthquake Engineering, Taipei, Taiwan, pp. 140-151.

Lenz, J.A., and Baise, L.G., 2007. Spatial variability of liquefaction potential in regional mapping using CPT and SPT data. Soil Dynamics and Earthquake Engineering 27(7), 690-702.

Maurer, B.W., Green, R.A., Cubrinovski, M., and Bradley, B.A., 2014. Evaluation of the liquefaction potential index for assessing liquefaction hazard in Christchurch, New Zealand. Journal of Geotechnical and Geoenvironmental Engineering 140(7), 04014032.

Maurer, B.W., Green, R.A., and Taylor, O.D.S., 2015a. Moving towards an improved index for assessing liquefaction hazard: lessons from historical data. Soils and Foundations 55(4), 778-787.

Maurer, B.W., Green, R.A., Cubrinovski, M., and Bradley, B.A., 2015b. Fines-content effects on liquefaction hazard evaluation for infrastructure in Christchurch, New Zealand. Soil Dynamics and Earthquake Engineering 76, 58-68.

Maurer, B.W., Green, R.A., Cubrinovski, M., and Bradley, B., 2015c. Assessment of CPT-based methods for liquefaction evaluation in a liquefaction potential index framework. Géotechnique 65(5), 328-336.

Maurer, B.W., Green, R.A., van Ballegooy, S., and Wotherspoon, L., 2019. Development of region-specific soil behavior type index correlations for evaluating liquefaction hazard in Christchurch, New Zealand. Soil Dynamics and Earthquake Engineering, 117, 96-105.

Mitchell, J.K., Lodge, A.L., Coutinho, R.Q., Kayen, R.E., Seed, R.B., Nishio, S., and Stokoe, K. H., II 1994. Insitu test results from four Loma Prieta earthquake liquefaction sites: SPT, CPT, DMT and shear wave velocity. Report No. UCB/EERC-94/04. Earthquake Engineering Research Center, UC Berkeley.

Moss, R.E.S., Seed, R.B., Kayen, R.E., Stewart, J. P., Youd, T.L., and Tokimatsu, K. 2003. Field case histories for CPT-based in situ liquefaction potential evaluation. Geoengineering Research Rep. UCB/GE-2003/04.

Moss, R.E.S., Collins, B.D., and Whang, D.H., 2005. Retesting of liquefaction/nonliquefaction case histories in the imperial valley. Earthquake Spectra 21(1), 179–196.

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. Journal of Geotechnical and Geoenvironmental Engineering 132(8), 1032-1051.

Moss, R.E.S., Kayen, R.E., Tong, L.-Y., Liu, S.-Y., Cai, G.-J., and Wu, J., 2009. Re-investigation of liquefaction and nonliquefaction case histories from the 1976 Tangshan earthquake. Rep. No. 209/102, Pacific Earthquake Engineering Research (PEER) Center, Berkeley, CA.

Moss, R.E.S., Kayen, R.E., Tong, L.-Y., Liu, S.-Y., Cai, G.-J., and Wu, J. 2011. Retesting of liquefaction and nonliquefaction case histories from the 1976 Tangshan earthquake. Journal of Geotechnical and Geoenvironmental Engineering 137(4), 334-343.

National Research Council, 2016. State of the art and practice in the assessment of earthquake-induced soil liquefaction and its consequences. Committee on Earthquake Induced Soil Liquefaction Assessment. National Research Council, The National Academies Press, Washington, DC.

New Zealand Geotechnical Database (2016). <https://www.nzgd.org.nz/Default.aspx> Accessed 8/24/16. New Zealand Earthquake Commission (EQC).

Papathanassiou, G., Mantovani, A., Tarabusi, G., Rapti, D., and Caputo, R., 2015. Assessment of liquefaction potential for two liquefaction prone areas considering the May 20, 2012 Emilia (Italy) earthquake. Engineering Geology, 189, 1-16.

Pass, D.G., 1994. Soil characterization of the deep accelerometer site at Treasure Island, San Francisco, California. MS thesis, University of New Hampshire.

PEER (Pacific Earthquake Engineering Research Center), 2000a. Documenting incidents of ground failure resulting from the 1999 Kocaeli, Turkey Earthquake. < https://tinyurl.com/ujtpwrk>

PEER (Pacific Earthquake Engineering Research Center), 2000b. Documentation of soil conditions at liquefaction sites from 1999 Chi-Chi, Taiwan Earthquake. < https://tinyurl.com/sbh9cpg>

Piryonesi, S.M. and El-Diraby, T.E., 2021. Using machine learning to examine impact of type of performance indicator on flexible pavement deterioration modeling. Journal of Infrastructure Systems 27(2), 04021005.

Quigley, M., Hughes, M.W., Bradley, B.A., van Ballegooy S., Reid, C., Morgenroth, J., Horton, T., Duffy, B., Pettinga, J, 2016. The 2010-2011 Canterbury earthquake sequence: Environmental effects, seismic triggering thresholds, geologic legacy. Tectonophysics 672-673, 228-274.

Rasmussen, C.E. and Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press. Cambridge, Massachusetts.

Robertson, P.K., and Wride, C.E., 1998. Evaluating cyclic liquefaction potential using cone penetration test. Canadian Geotechnical Journal 35(3), 442-459.

Rokach, L. and Maimon, O., 2008. Data mining with decision trees: theory and applications. World Scientific Pub Co Inc. ISBN 978-9812771711.

Seed, H. B., Tokimatsu, K., Harder, L. F. Jr., and Chung, R., 1984. The influence of SPT procedures in soil liquefaction resistance evaluations. Report No. UCB/EERC-84/15. Earthquake Engineering Research Center, University of California, Berkeley.

Servizio Geologico, 2016. Database of the Emilia-Romagna region: geological, seismic, and soil Survey." http://ambiente.regione.emiliaromagna.it/geologia/cartografia/webgis-banchedati/.

Shengcong, F. and Tatsuoka, F., 1984. Soil liquefaction during Haicheng and Tangshan earthquake in China; a review. Soils and Foundations 24(4), 11-29.

Shibata, T., and Teparaska, W., 1988. Evaluation of Liquefaction Potential of Soils Using Cone Penetration Testing. Soils and Foundations 28(2), 49-60.

Suzuki, Y., Tokimatsu, K., Moss, R.E.S., Seed, R.B., and Kayen, R.E., 2003. CPT-based liquefaction case histories from the 1995 Hyogoken-Nambu (Kobe) earthquake, Japan. Geotechnical Engineering Research Report No. UCB/GE-2003/03.

Toprak, S., and Holzer, T.L., 2003. Liquefaction potential index: field assessment. Journal of Geotechnical and Geoenvironmental Engineering 129(4), 315-322.

Treadwell, D.D., 1976. The influence of gravity, prestress, compressibility, and layering on soil resistance to static penetration. Doctor of Philosophy dissertation, Univ. of California, Berkeley, CA.

Turner, B., Brandenberg, S., and Stewart, J., 2016. Case Study of Parallel Bridges Affected by Liquefaction and Lateral Spreading. Journal of Geotechnical and Geoenvironmental Engineering 142(7), 05016001.

van Ballegooy, S., Malan, P., Lacrosse, V., Jacka, M.E., Cubrinovski, M., Bray, J.D., O'Rourke, T.D., Crawford, S.A., and Cowan, H., 2014a. Assessment of liquefaction-induced land damage for residential Christchurch. Earthquake Spectra 30(1), 31-55.

van Ballegooy, S., Cox S.C., Thurlow C., Rutter H.K., Reynolds, T., Harrington, G., Fraser, J., and Smith, T., 2014b. Median water table elevation in Christchurch and surrounding area after the 4 September 2010 Darfield earthquake: Version 2. GNS Science Report 2014/18, 2014.

van Ballegooy, S., Green, R.A., Lees, J., Wentz, F., and Maurer, B.W., 2015. Assessment of various CPT based liquefaction severity index frameworks relative to the Ishihara (1985) H_1-H_2 boundary curves. Soil Dynamics and Earthquake Engineering 79, 347-364.

van der Linden, T.I., De Lange, D.A. and Korff, M., 2018. Cone penetration testing in thinly inter-layered soils. Proceedings of the Institution of Civil Engineers-Geotechnical Engineering, 171(3), 215-231.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Yoshimine, M., Nishizaki, H., Amano, K., & Hosono, Y., 2006. Flow deformation of liquefied sand under constant shear load and its application to analysis of flow slide of infinite slope. Soil Dynamics and Earthquake Engineering 26(2-4), 253-264.

Yost, K. M., Green, R. A., Upadhyaya, S., Maurer, B. W., Yerro-Colom, A., Martin, E. R., & Cooper, J., 2021. Assessment of the efficacies of correction procedures for multiple thin layer effects on Cone Penetration Tests. Soil Dynamics and Earthquake Engineering 144, 106677.

Youd, T.L., and Carter, B.L., 2005. Influence of soil softening and liquefaction on spectral acceleration, Journal of Geotechnical and Geoenvironmental Engineering 131(7), 811-825.

Youd, T.L., DeDen, D.W., Bray, J.D., Sancio, R., Cetin, K.O., and Gerber, T.M., 2009. Zero displacement lateral spreads, 1999 Kocaeli, Turkey, earthquake. Journal of Geotechnical and Geoenvironmental Engineering 135(1), 46-61.

Zhang, G., Robertson, P.K., and Brachman, R.W.I., 2002. Estimating liquefaction-induced ground settlements from CPT for level ground. Canadian Geotechnical Journal 39(5), 1168-1180.