nzsee
NEW ZEALAND SOCIETY FOR EARTHQUAKE ENGINEERING

Machine learning correction of overpredicted liquefaction manifestation using Liquefaction Severity Number

N.J. McDougall, B.R. Kennerley, J. Russell & V. Lacrosse
Tonkin & Taylor Limited, Auckland, New Zealand.

S. van Ballegooy
InfinityStudio.AI, Auckland, New Zealand.

M.E. Jacka
Tonkin & Taylor Limited, Christchurch, New Zealand.

ABSTRACT

Many state-of-practice methods for predicting liquefaction manifestation, such as the Liquefaction Severity Number (LSN), are known to suffer from significant overprediction in regions characterised by complex soil profiles comprising interbedded sands, silts, and clays. These methods typically analyse discrete soil layers, and sum results layer-by-layer, which does not properly account for system effects.

We demonstrate that machine learning techniques can be used to identify cases where simplified layer-by-layer liquefaction vulnerability indices give overestimated surficial liquefaction manifestation. Specifically, we have developed a convolutional neural network model. With the aim of better accounting for system effects, the model considers the full length of a CPT profile simultaneously to capture system effects of soil profiles such as interbedding, rather than processing and then aggregating information in discrete layers.

A database of over 47,000 case histories was used for training and model evaluation, spanning ten New Zealand earthquakes. Special techniques were used to address sampling bias and class imbalance. Finally, an adjustment procedure is proposed, which uses the machine learning model to improve the accuracy of LSN for specific site categories, resulting in significant accuracy improvements.

This research has been funded by Toka Tū Ake EQC to advance liquefaction science, and support a range of applications in New Zealand, including local government planning, public engagement and education, and loss modelling.

1 INTRODUCTION

A variety of state-of-practice methods exist to predict surficial liquefaction manifestation, typically based on subsurface test data such as that provided by a Cone Penetration Test (CPT) or a Standard Penetration Test (SPT). Manifestation is taken as a True-or-False value corresponding to whether the effects of liquefaction are visible at the ground surface (such as ejecta, ground surface distortion, etc.), or else manifestation is quantified on a severity spectrum using a liquefaction severity index correlated with manifestation, such as the Liquefaction Potential Index (LPI) of Iwasaki et al. (1981), and the Liquefaction Severity Number (LSN) of van Ballegooy et al. (2014).

It is known that in some areas, these liquefaction severity indexes can suffer from systematic overprediction of liquefaction (van Ballegooy et al., 2014), where the calculated index has a higher value than would be expected given the levels of manifestation observed from case histories. Specifically, overprediction has been demonstrated to occur on soil profiles having predominantly high fines-content, high-plasticity soil strata (Beyzaei et al., 2018; Upadhyaya et al., 2023) and for highly stratified deposits of interlayered sands, silts, and clays (van Ballegooy, 2018), especially those with low-permeability (Geyin and Maurer, 2021).

Manifestation prediction models tend to perform worse on profiles with multiple liquefied strata, and this issue is exacerbated for simplified methods which implicitly assume that liquefiable strata are independent entities, such as LSN (Rateria and Maurer, 2022). These simplified methods analyse discrete soil layers and sum the results layer-by-layer. This process does not properly account for system response effects. The system response of deposits is a fundamental consideration when assessing manifestation and can help explain why manifestation is overpredicted is by LSN (Cubrinovski et al., 2019).

The objective of this study is to use Machine Learning (ML) with CPT data to identify these cases where LSN is overpredicting manifestation and to develop a corrective procedure to improve LSN's accuracy in such cases. By allowing the entire CPT profile as input data to the ML model, rather than as a collection of individual layers analysed in isolation, the model can capture the system response of deposits.

Note that the purpose of the ML model is not to directly predict whether there would be manifestation. Instead, it is a "watchdog" model: it determines when LSN is likely to be significantly overpredicted, and then a separate correction procedure can be applied to the LSN values.

2 RELATED STUDIES

There are many existing studies which explore the use of ML models to predict liquefaction. Maurer and Sanger (2023) have undertaken a thorough review of these, and we agree with their conclusion that there are prevailing, severe methodological issues with much of the current literature. Despite this, they conclude there are good reasons to use ML techniques to empirically improve state-of-practice methodologies for surficial liquefaction manifestation prediction.

A successful example of this was the work of Rateria and Maurer (2022) who undertook re-regressions of the H1–H2 method of Ishihara (1985) against up-to-date case history data. For these re-regressions, ML models were used, and they were found to outperform an approach where LSN is compared against a decision threshold (e.g., surface ground damage predicted to be present if LSN > 15). These methods are used to predict manifestation as a True-or-False value and do not account for severity.

It was observed by van Ballegooy et al. (2018) that there are systematic visual differences between the plotted CPT data at sites where LSN is giving accurate predictions and those where it is underpredicting or overpredicting. This provides evidence that it is possible, in principle, for a well-developed machine learning model to provide an improvement on LSN.

3 METHODOLOGY

3.1 Data Preparation

3.1.1 New Zealand Liquefaction Manifestation Case History Dataset

110,000 Liquefaction Case Histories were compiled from 10 historical New Zealand earthquakes, mostly from the Canterbury Earthquake Sequence (CES) but also three other events.

Each case history comprised a CPT, a modelled groundwater depth (GWD), a modelled Peak Ground Acceleration (PGA), and the event's moment magnitude (M_w). These were used as the input data for the machine learning model. Each case history also included a land damage observation: either of manifestation, or absence thereof (i.e. it was directly observed that no manifestation was present). These case histories are enumerated in Table 1, and the source of data for these case histories is enumerated in Table 2.

The CPTs were sourced from data in the New Zealand Geotechnical Database as of 6th June 2023 and supplemented by additional proprietary data provided by Tonkin & Taylor Limited. CPTs were subjected to minimum data quality requirements, and some basic data cleaning was applied. CPTs were not included when they were performed at sites with ground improvement.

The model was only trained on 47,700 of the case histories (with 21,900 distinct CPTs), namely the ones where the PGA was between 0.2g and 0.8g. Below 0.2g, only 788 of case histories (1%) show overprediction using LSN. This is too few case histories with which to properly train the ML model. Similarly, there are not enough case histories with PGAs above 0.8g.

Table 1: Case history data used in this study.

Event Name	GeoNet Public ID	M_w	PGA Data Source*	GWD Data Source*	Damage Observation Data Source*	Case History Count
Edgecumbe 1987	04228	6.5	[1]	[1]	[1]	395
CES Sept 2010	3366146	7.1	[2]	[3] (Sept)	[5]	19,619
CES Oct 2010	3391440	4.8	[4]	[3] (Sept)	[6]	28,473
CES Feb 2011	3468575	6.2	[2]	[3] (Feb)	[5]	18,930
CES Apr 2011	3497857	5.0	[2]	[3] (Feb)	[7]	2,863
CES Jun 2011 A	3528810	5.3	[2]	[3] (Jun)	[7]	2,849
CES Jun 2011 B	3528839	6.0	[2]	[3] (Jun)	[5]	5,204
CES Dec 2011 B	3631380	6.0	[2]	[3] (Dec)	[5]	12,321
Christchurch 2016	3528839	6.0	[8]	[9]	[10]	17,208
Kaikōura 2016	2016p 858000	7.8	[4]	[11]	[12]	2,158

* See Table 2

Table 2: Case History Data Sources used in this study.

ID	Case History Data Source
[1]	Mellsop et al. (2017).
[2]	Bradley and Hughes (2012) and Bradley (2012), in some areas choosing to defer to kriging on strong motions station data (O'Rourke and Milashuk, 2011).
[3]	Tonkin + Taylor (2013), Appendix C.
[4]	2022 National Seismic Hazard Model hindcasting models created by GNS Science.
[5]	Tonkin + Taylor (2015).
[6]	Comprehensive reconnaissance mapping performed by Sjoerd van Ballegooy. Damage was localised to Hoon Hay.
[7]	Reconnaissance mapping around the Avon River performed by Tonkin + Taylor.
[8]	Model developed using a similar methodology to [2] by Bradley (University of Canterbury, 2016).
[9]	Model developed using a similar methodology to [3] by Tonkin + Taylor.
[10]	Comprehensive reconnaissance mapping performed by Tonkin + Taylor.
[11]	In Wellington: Tonkin + Taylor (2021). In Marlborough: Ogden (2018)
[12]	In Wellington: reconnaissance mapping at CentrePort (Cubrinovski et al., 2017). In Marlborough: Bastin et al. (2021).

3.1.2 LSN Calculation Methodology

All LSN calculations in this study were performed using the Boulanger and Idriss (2014) methodology, using assumed values for soil unit weight (γ) of 18 kN/m^3 and fines content estimation calibration parameter (CFC) of 0. Some CPTs did not have recorded groundwater measurements from the time of investigation, in which case the modelled level from the time of the earthquake event was imputed. The 50th percentile liquefaction cyclic resistance curve ($CRR_{PL=50\%}$) was used for this study since it is more appropriate for back-analysis purposes. An inverse filtering procedure to correct for thin layer effects was not applied.

Analysis was performed to a depth of 20m, and soil within any surface predrill and below the depth of CPT termination was assumed to be non-liquefiable in the LSN calculation. In such cases the LSN is artificially low, meaning that the case history is less likely to be deemed as a case of over-prediction.

3.1.3 LSN Decision Threshold

When LSN is above a decision threshold, but the site does not experience any manifestation, this corresponds to overprediction (i.e. the True-or-False estimate for manifestation is True when it should be False). Usually, LSN is calculated using the 15th percentile cyclic resistance curve ($CRR_{PL=15\%}$) rather than the 50th percentile used in this study.

For the 15th percentile, the optimal decision threshold is around 15 under the assumption that false positives and false negatives are equally undesirable. The same assumption gives a decision threshold of around 10 using the 50th percentile cyclic resistance curve. However, for this study, a higher decision threshold of 15, rather than 10, has been adopted. This means the model focuses on the more extreme cases of overprediction. This reduces the scope of the model but allows for potentially higher accuracy. Further work to calibrate this decision threshold could potentially increase the usefulness of the model without significantly decreasing accuracy.

3.1.4 Train-Validation-Test Dataset Split

As is typical when training a machine learning model, the case histories were divided into three split sets: one for training the model (70% of the data), another for iteratively validating it while it is being developed (20%), and another for testing the final model at the very end of the model development process (10%).

To avoid train-test contamination of individual CPTs shared across multiple case histories, a stratified sampling scheme was adopted for the division into split sets. All case histories associated with a particular CPT were forced to be confined to the same split set; this means that it is impossible for the model to be tested against a CPT which it has been trained against, even across different case histories.

The testing set was not used until after the model was finalised. During model development and fine-tuning, performance was evaluated on the validation set, which was not trained upon directly.

3.1.5 Oversampling Algorithm

In the dataset, there are various forms of sampling bias, and insufficiency of data. As can be seen in Table 1, some events (and hence M_w values) have significantly more data than others. Likewise, some PGA and groundwater depth parameter combinations have more data than others. The rate at which LSN overpredicts manifestation varies across these input parameters.

To address these concerns, a synthetic oversampling algorithm was developed to train the model. It is based on a physical insight: for any case history with liquefaction manifestation, a case history with higher PGA and/or M_w could be assumed to also give rise to liquefaction manifestation, all else being equal, i.e., for the same GWD and the same CPT. Conversely, the inverse holds true for cases without manifestation; a lower PGA and/or M_w could be assumed to also not give rise to liquefaction.

The oversampling procedure ensures an equal balance between cases of manifestation and those without it; as well as between correct prediction with misprediction. It is also a form of data augmentation to help the model avoid overfitting to the data.

3.2 Developing the Machine Learning Model

This study has been focused on Convolutional Neural Network (CNN) architectures. Partly, this is because the length of CPTs vary significantly, whereas most ML architectures require a fixed input size.

Also, there are good theoretical reasons to justify the use of CNNs for this problem. Simplified methods to assess liquefaction analyse each layer in isolation (Rateria and Maurer, 2022). Similarly, the initial convolutional filter is trained to analyse small numbers of adjacent layers (e.g., 8cm), and this filter is re-used down the entire soil profile. This reduces the number of trainable parameters, which in turn decreases the chance of overfitting. However, deeper into the network, as the convolutions condense, system effects can be better captured.

A variety of CNN architectures were considered, and extensive hyperparameter experiments were undertaken on many of them. The best performing model architecture is shown in Figure 1.

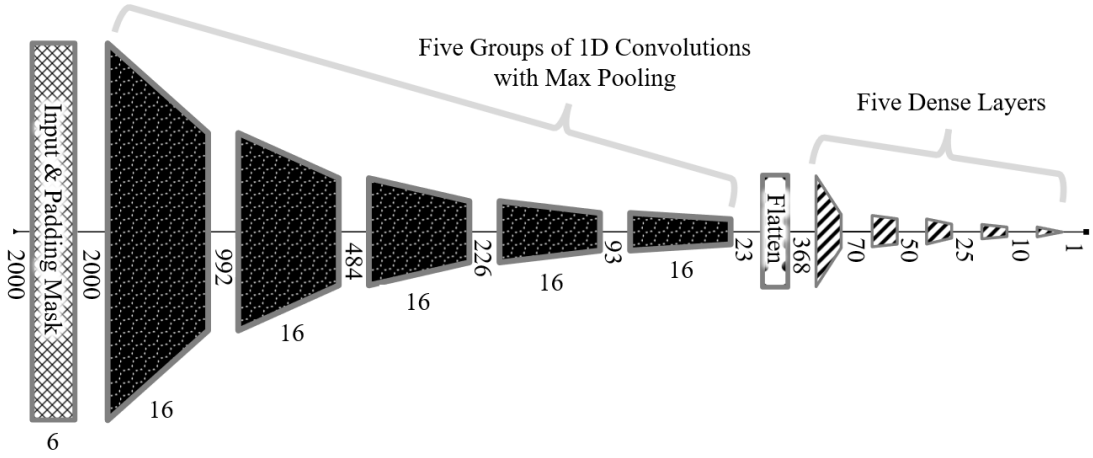

Figure 1: The CNN model architecture developed in this study. Numbers annotated between layer groups correspond to the length of the input/output array to/from each group. Numbers annotated below layer groups correspond to the number of channels in the group.

The input data was provided in an array of six channels: two for the 20m of raw CPT data in 1cm increments (tip resistance and sleeve friction; q_c and f_s), two for the PGA and M_w values, one for the GWD value, and one for a simplified saturation degree (S_r) model, which was equal to 100% for values below the groundwater table and 0% above the groundwater table. In future work, better results may be possible to be obtained by including the soil behaviour type index (I_c) and normalised cone tip resistance (q_{c1N}) as inputs.

4 MODEL PERFORMANCE AND CONCLUSION

When the ML model identifies a case of LSN overprediction, the LSN should be corrected by reducing it to a value below the decision threshold of 15. The precise value it is reduced to depends on the application; for design purposes, a value of 15 would potentially be appropriate as conservative choice. For loss modelling purposes, a value closer to zero would be more appropriate, to minimise bias (recall that overprediction cases correspond to cases where there was *no* manifestation observed – and hence no losses). More research is needed to provide a robust correction methodology.

We found that in practice, when the ML model *incorrectly* identifies a case of LSN overprediction, most of the nearby CPTs were correctly identified as *not* having overprediction. On this basis, we developed a geospatial consensus-based application of the model: the ML adjustment is only applied if at least 20% of CPTs within 300m had LSN identified as overpredicting, among those with LSN > 15. The choice of 20% was chosen to maximise accuracy based on data from the CES September 2010 and February 2011 events.

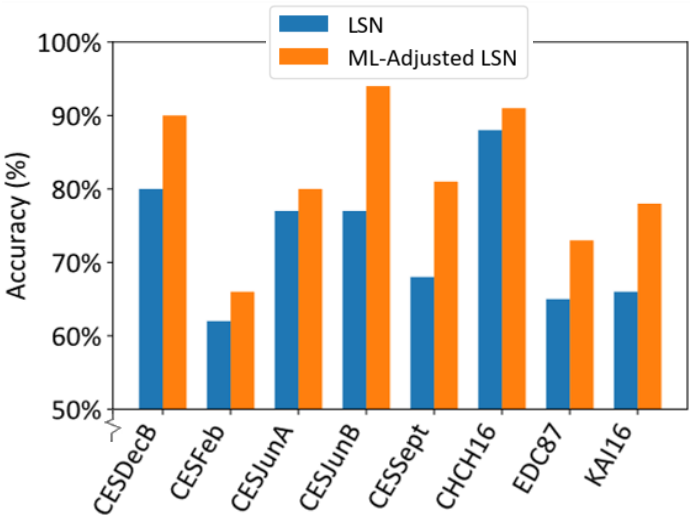

Figure 2: Weighted accuracy of the model on the entire dataset across the different events, when applied using the geospatial consensus methodology.

The accuracy of LSN with these ML-adjustments on the testing dataset is shown in **Error! Reference source not found.**, broken down for each event. This is given on a 300m grid-cell-by-grid-cell basis, rather than CPT-by-CPT, since there are some areas with much spatial density of CPTs which would bias the accuracy statistic.

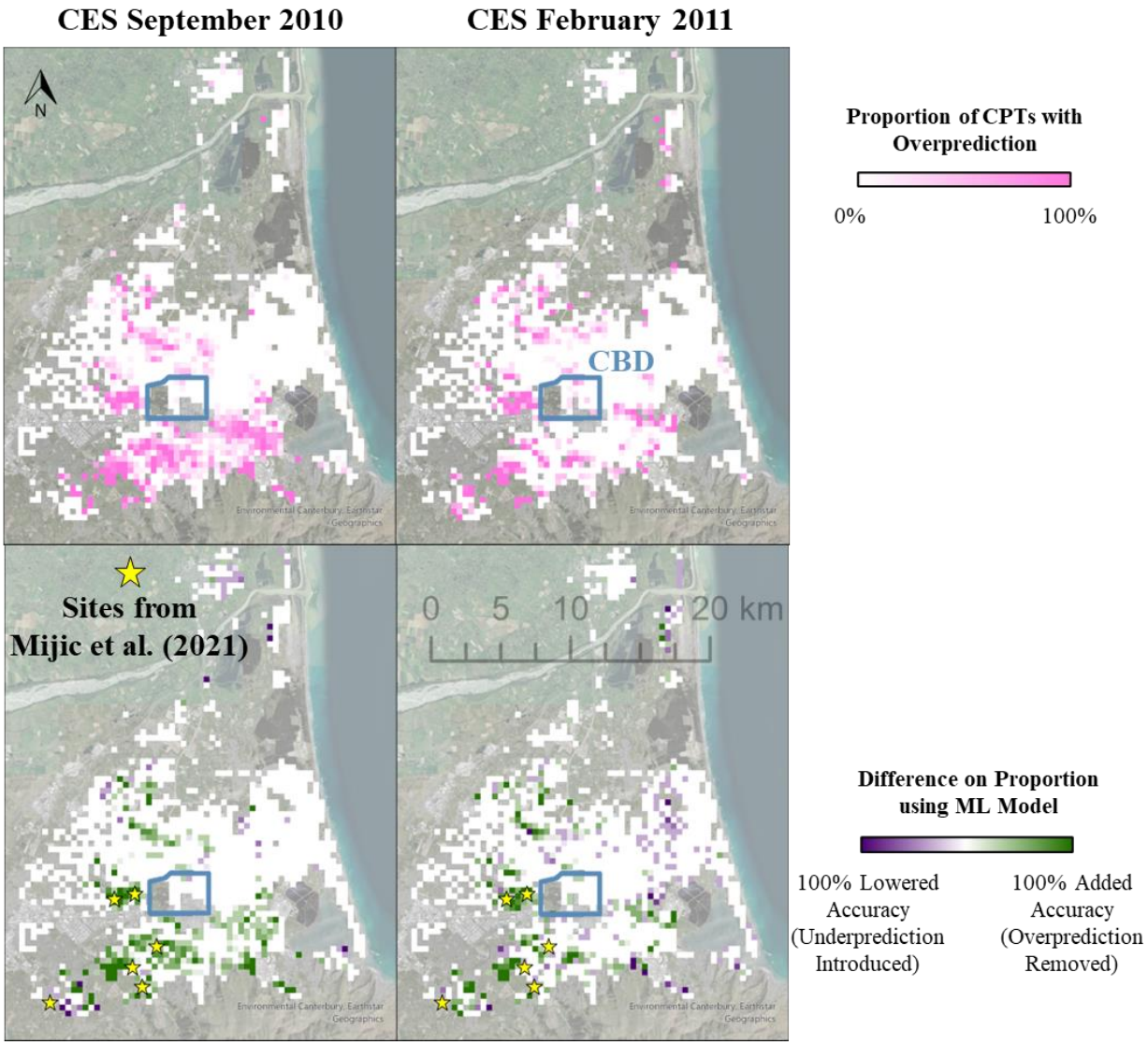

Figure 3: Effect of ML-Adjusted LSN for the CES September 2010 and February 2011 events, on a grid-cell-by-grid cell basis. The CBD is shown for context. Aerial Imagery is attributed to Environmental Canterbury and Earthstar Geographics.

Figure 3 shows the effect on accuracy of applying the ML correction for the CES September 2010 and February 2011 events. It shows that the ML-based correction can give significant improvements in the key areas where LSN is known to overpredict in the south-west of Christchurch. These areas align well with the areas with a high value for the interbedded index developed by Geyin and Maurer (2021). There are also areas where the ML-based correction makes things worse (by lowering LSN when it shouldn't be lowered, introducing underprediction), but the loss of accuracy in such cases is usually slight.

A previous study by Mijic et al. (2021) compiled sites from the CES where state-of-practice methods gave overprediction compared with observations. From the compiled sites, those with highly stratified deposits of interbedded, low-permeability soils were selected in this study for further validation of the model. These are also shown in Figure 3. Five out of six sites fall within areas where the model's accuracy has indeed improved in the south-west of Christchurch. The sixth site lies beyond the scope of available data but is nearby. This demonstrates that the model seems to be able to address overprediction in areas dominated by highly stratified sands, silts, and clays.

This study has shown that an ML-based correction to LSN is a promising avenue of research. The neural network and associated correction procedure developed in this paper offer a significant improvement to standard LSN. However, there are many aspects which merit further research, especially new choices of model inputs (e.g. I_c and q_{c1N}), decision thresholds, vulnerability indices (e.g. LPI), and different correction methodologies.

Underprediction has not been considered as a part of this research. When CPT data is not available for the full 20m vertical soil profile, a low LSN may be caused by a failure to consider the soil layers where CPT data is unavailable, rather than underprediction the LSN calculation procedure itself. This fact significantly limits the number of suitable CPTs available for investigating underprediction.

5 ACKNOWLEDGEMENTS

The authors would like to acknowledge Toka Tū Ake EQC for providing funding and support for this research. We are especially grateful for their support of the New Zealand Geotechnical Database, without which this work would not be possible. We would also like to acknowledge GNS Science for the provision of earthquake shaking models for this research.

REFERENCES

Bastin SH, van Ballegooy S and Ogden M (2021). "The past is key to the future; Collating historical cases of liquefaction to supplement liquefaction hazard assessments". *21st New Zealand Geotechnical Society Symposium*, 24-26 March, Dunedin, New Zealand, 10pp. https://www.nzgs.org/libraries/proceedings-of-the-21st-nzgs-symposium/

Beyzaei CZ, Bray JD, Cubrinovski M, Riemer M and Stringer M (2018). "Laboratory-based Characterization of Shallow Silty Soils in Southwest Christchurch". *Soil Dynamics and Earthquake Engineering*, **110**: 93-109. https://doi.org/10.1016/j.soildyn.2018.01.046

Boulanger R and Idriss IM (2014). "*CPT and SPT Based Liquefaction Triggering Procedures*". UCD/CGM-14/01, Center for Geotechnical Modeling, Department of Civil and Environmental Engineering, University of California, Davis, California, 134pp. https://faculty.engineering.ucdavis.edu/boulanger/reports

Bradley BA (2012). "Strong Ground Motion Characteristics Observed in the 4 September 2010 Darfield, New Zealand Earthquake". *Soil Dynamics and Earthquake Engineering*, **42**: 32-46. https://doi.org/10.1016/j.soildyn.2012.06.004

Bradley B, Hughes M (2012). "*Conditional Peak Ground Accelerations in the Canterbury Earthquakes for Conventional Liquefaction Assessment*". Two part report for the Department of Building and Housing and Ministry of Business, Innovation and Employment, University of Canterbury, Christchurch, New Zealand, 41pp.https://www.researchgate.net/publication/272831749
https://www.researchgate.net/publication/272831755

Cubrinovski M, Bray JD, De La Torre C, Olsen MJ, Bradley BA, Chiaro G, Stocks E and Wotherspoon L (2017). "Liquefaction effects and associated damages observed at the Wellington CentrePort from the 2016 Kaikoura earthquake". *Bulletin of the New Zealand Society for Earthquake Engineering*, **50**(2), 152-173. https://doi.org/10.5459/bnzsee.50.2.152-173

Cubrinovski M, Rhodes A, Ntritsos N and van Ballegooy S (2019). "System Response of Liquefiable Deposits". *Soil Dynamics and Earthquake Engineering*, **124**: 212-229. https://doi.org/10.1016/j.soildyn.2018.05.013

Geyin M and Maurer BW (2021). "Evaluation of a Cone Penetration Test Thin-layer Correction Procedure in the Context of Global Liquefaction Model Performance". *Engineering Geology*, **291**: Article 106221, 1-15. https://doi.org/10.1016/j.enggeo.2021.106221

Ishihara K (1985). "Stability of natural deposits during earthquakes". *11th International Conference on Soil Mechanics and Foundation Engineering*, 12-16 August, San Francisco, California, Paper Number 7, 56pp. https://www.issmge.org/uploads/publications/1/34/1985_01_0007.pdf

Iwasaki T, Tokida K and Tatsuoka F (1981). "Soil liquefaction potential evaluation with use of the simplified procedure". *1st International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics Proceedings*, 26 April – 3 May, St Louis, Missouri, 6pp. https://scholarsmine.mst.edu/icrageesd/01icrageesd/session02/12

Maurer BW and Sanger MD (2023). "Why "AI" Models for Predicting Soil Liquefaction Have Been Ignored, Plus Some that Shouldn't Be". *Earthquake Spectra*, **39**(3): 1883-1910. https://doi.org/10.1177/87552930231173711

Mijic Z, Bray JD and van Ballegooy S. (2021). "*Detailed Evaluation of Insightful Liquefaction Ejecta Case Histories for the Canterbury Earthquake Sequence, New Zealand*". USGS Award No. G20AP00079, University of California, Berkeley, California, 45pp. https://earthquake.usgs.gov/cfusion/external_grants/reports/G20AP00079.pdf

Mellsop N, Bastin S, Wotherspoon L and van Ballegooy S (2017). "Development of detailed liquefaction case histories from the 1987 Edgecumbe Earthquake". *20th NZGS Geotechnical Symposium*, 24-26 November, Napier, New Zealand. https://www.nzgs.org/libraries/proceedings-of-the-20th-nzgs-symposium

Ogden M (2018). "*Assessment of Liquefaction from New Zealand Case Histories: With Insights from Liquefaction-induced Land Damage, over the Lower Wairau Plains, Caused by the 2016 Kaikoura earthquake*". Masters Thesis, Department of Civil Engineering University of Auckland, Auckland, New Zealand, 327pp. http://hdl.handle.net/2292/37259

O'Rourke T and Milashuk S (2011). "Spatial Distribution of Ground Motion During Earthquakes affecting Christchurch, New Zealand".

Rateria G and Maurer BW (2022). "Evaluation and Updating of Ishihara's (1985) Model for Liquefaction Surface Expression, with Insights from Machine and Deep Learning". *Soils and Foundations*, **62**(3): Article 101131, 1-17. https://doi.org/10.1016/j.sandf.2022.101131

Tonkin + Taylor (2013). "*Liquefaction Vulnerability Study*". #52020.0200/v1.0, Earthquake Commission, Wellington, New Zealand, 59pp. https://canterburygeotechnicaldatabase.projectorbit.com

Tonkin + Taylor (2015). "*Canterbury Earthquake Sequence: Increased Liquefaction Vulnerability Assessment Methodology*". #52010.140.v1.0, Chapman Tripp acting on behalf of the Earthquake Commission, Wellington, New Zealand, 204pp. https://www.eqc.govt.nz/our-publications/ces-increased-liquefaction-vulnerability-assessment-methodology-t-t-report

Tonkin + Taylor (2021). "*Development of Earthquake Damage Fragility Functions for New Zealand Residential Buildings*". #1013637.0000.300.v1, Unpublished document issued to Toka Tū Ake EQC, 119pp.

University of Canterbury. *UC professor uses supercomputer to simulate quake*. UC News, 23 February 2016. https://www.canterbury.ac.nz/news/2016/uc-professor-uses-supercomputer-to-simulate-quake.html

Upadhyaya S, Maurer BW, Green RA, Rodriguez-Marek A, and van Ballegooy S (2023). "Surficial Liquefaction Manifestation Severity Thresholds for Profiles having High Fines-content, High-plasticity Soils". *Canadian Geotechnical Journal*, **60**(5): 642-653. https://doi.org/10.1139/cgj-2022-0092

van Ballegooy S (2018). "Use of geotechnical and liquefaction observation GIS databases for liquefaction hazard mapping". *5th International Conference on GIS and Geoinformation Zoning for Disaster Mitigation*, 15-17 November, Auckland, New Zealand. 16pp.

van Ballegooy S, Malan P, Lacrosse V, Jacka ME, Cubrinovski M, Bray JD, O'Rourke TD, Crawford SA and Cowan H (2014). "Assessment of Liquefaction-Induced Land Damage for Residential Christchurch". *Earthquake Spectra*, **30**(1): 31-55. https://doi.org/10.1193/031813EQS070M