

**Ref. No.: NHESS-2016-281**

**Title: EVALUATING SIMPLIFIED METHODS FOR LIQUEFACTION ASSESSMENT FOR LOSS ESTIMATION**

5 **Author's Response to the Reviewers' Comments**

The authors would like to thank the reviewers for their efforts in reviewing this paper, the rapid review time and the constructive comments made. We have made a number of revisions to strengthen the paper in light of the reviewers' comments and considering the overall positive nature of the reviewers' comments to the first version  
10 of the manuscript, we are confident that this improved version will be adequate for publication.

In addition, whilst working through parts of the analysis again, we noticed an error in the process used to derive Eq. 9 from the original equations of Boore (2004). Eq. 9 has been corrected in the revised text and this affects the results corresponding to models LPI3 and LPI3b as follows:

- 15
- With the new form of Eq. 9, model LPI3 now does meet the diagnostic score criteria of  $TPR > 0.5$  and  $TNR > 0.5$  when the threshold is 5, albeit with some bias towards negative forecasts.
  - Model LPI3b still meets the criteria but with bias towards positive forecasts.
  - The optimum threshold for model LPI3 is now 4, while the optimum threshold for model LPI3b is now 10
  - Based on Youden's J-statistic, model LPI3 now outperforms model LPI3b
- 20
- In section 5, the liquefaction probability analysis of model LPI3b has been replaced by the equivalent analysis for model LPI3

Best Regards

Indranil Kongar, Tiziana Rossetto, Sonia Giovinazzi

25 **Reviewer #1:** The authors test various methods for the assessment of liquefaction using data collected following two recent earthquakes in New Zealand. The study is within the scope of Natural Hazards and Earth System Sciences, it is generally well written, testing these methods using a large database of observations is a valuable exercise and the analysis appears to be carefully performed. Therefore, I recommend that this paper is accepted for publication but only after the following editorial changes are made

30 *We thank the reviewer for his confirmation of the value of the research and recommendation for publication. Our responses to specific comments are included below.*

1. Abstract, first sentence: This sentence is grammatically incorrect. In addition, it is probably too long to be easy  
35 understandable.

*This sentence has been amended to correct the grammar and to make it easier to understand*

2. Abstract and throughout: “methods” or “procedures” are what is being talked about here. Therefore, these words should be used rather than “methodologies”, which are the principles that guide research practices.

5

*We have replaced “methodologies” with the word “method”*

3. Abstract and throughout: The word “data” is plural and hence the sentences should read “although these data may not” and “the input data are publicly”.

10

*We have amended the text to ensure that the word “data” is used as a plural*

4. P. 2, l. 8: This should probably read “both future risk assessments and post-event rapid response analyses”.

15 *We have amended the text as suggested by the reviewer*

5. P. 2, l. 9: Should it not be “liquefaction effects and physical damage” rather than “liquefaction risk and physical damage”?

20 *We have amended the text as suggested by the reviewer*

6. P. 5, l. 12: USGS are not the only group to publish assessments of the ground motion following earthquakes so this organisation should only be given as an example.

25 *We have added a note to make clear that the USGS is only one possible source of post-earthquake ground motion data*

7. P. 5, l. 13 (and elsewhere): Because of the high epistemic uncertainties in ground-motion prediction it is generally considered best practice to use a logic tree comprised of a set of ground-motion models rather than a single ground motion prediction equation. Hence I suggest slightly modifying this sentence.

30

*We have modified this sentence to reflect the reviewer's comment*

8. P. 5, l. 15: I do not understand the comment “Although the use of Vs negates the requirement for ground investigation” because to assess Vs requires measurements on site, although they could be non-invasive (e.g. based on ambient noise approaches) as well as invasive (from boreholes). Vs30 could be estimated from geology or topographical slope, for example, but these would be uncertain and ideally should not be considered for site-specific analyses (e.g. Lemoine et al., Bulletin of the Seismological Society of America, 102, 2585-2599 2012).

10 *We thank the reviewer for highlighting this mistake. The original comment “Although the use of Vs negates the requirement for ground investigation” applies only to the Christchurch case study for which Vs data already exist. However, more generally the reviewer is correct to point out that Vs assessment does require ground investigation. In this instance, the authors' intention was to talk about general approaches rather than a specific case study, and so we have amended the text accordingly.*

15 9. P. 5, l. 17: I would change “extrapolate” to “estimate” or “interpolate” as extrapolation should be avoided.

*We have amended the text as per the reviewer's suggestion*

20 10. P. 5, ll. 22-25: As noted in my comment 8 Vs30 from topographic slope (as provided by the USGS Global Vs30 Map Server) is uncertain because of the weak correlation between these variables. This should be commented on as a weakness of this approach.

*We have added a comment to highlight this weakness*

25 11. P. 5, ll. 30: Boore et al. (Bulletin of the Seismological Society of America, 101, 3046-3059, 2011) update the relationships of Boore (2004) and other authors have proposed relationship for other parts of the world since 2004. I recommend making a comment that such relationships ideally should be regionally calibrated. Some checking that the equations of Boore (2004) are appropriate for New Zealand would be useful.

30 *We acknowledge that ideally relationships should be regionally calibrated. However, the rationale of the paper is to test simplified methods to be used for loss estimation, principally by insurers. In this case, part of the definition*

of a 'simplified' method is that the relationships that underpin it already exist in the literature and no new model development is required. It should be noted that the relationships proposed by Boore et al. (2011) are not updates of the relationships of Boore (2004) – rather they are alternative relationships specific to Japan, while the original Boore (2004) relationships are specific to California. Nevertheless, we have added an investigation of the appropriateness of the Boore (2004) relationship to Christchurch in section 3.2.

12. P. 5, Equations 8 and 9: It is not statistically correct to invert equations based on standard regression analysis (it would be acceptable if orthogonal regression had been used). I recommend adding a note that this inversion could be a source of uncertainty. Ideally a set of equations predicting Vs0-10 from Vs30 and Vs10-20 from Vs30 should have been derived based on regression in the correct direction.

*The authors are not aware of any relationships that are derived in the correct direction, and as mentioned in the previous comment, the development of new relationships contradicts the objectives of this study. We have added a note to section 2.1 reflecting the reviewer's comment regarding uncertainty due to inversion.*

13. P. 7, ll. 5-6: There seems to be a problem with the phrase “and for the other zones are given”.

*We have amended the text to make this sentence clearer*

14. Section 2.3: Is it not circular to test this model on data from the Christchurch 2011 earthquake as data from this earthquake was used to develop it? I recommend adding a comment on this.

*We acknowledge that there is some circularity in this test. However, it is worth noting that the datasets used to develop the model and test the model have not come from the same source and so may not be identical. Furthermore the original model has been calibrated to optimise estimation of the areal extent of liquefaction whereas in the case study exercise, site specific predictions are tested. A comment to this effect has been added at the end of section 2.3.*

15. P. 8, l. 23: “comparing” should be “comparing” and “earthquake” should be “earthquake”. Please spell check before manuscript submission.

*We have run a spell check and corrected these errors*

16. P. 9, l. 3: What is the source of the moment magnitude of 6.2 for the 2011 earthquake? Both the USGS and Global CMT give Mw 6.1 for this event. Perhaps it is GeoNet. This should be stated.

5

*The source is indeed GeoNet and the reference (GNS Science, 2014) has been added at the end of the relevant sentence.*

17. Figure 2: What is the source of these contour maps?

10

*The source is the Canterbury Geotechnical Database and the reference has been added in the caption*

18. P. 9, ll. 24-25: Are the results of SPT after the ground has liquefied appropriate to assess whether the ground is liquefiable? I would have thought that SPT values would be changed by liquefaction.

15

*Firstly we should note that there is a mistake here since the LPI values have been obtained from CPT, not SPT, data and so this has been corrected. Historically it has been thought that after liquefaction occurs, soils densify and increase their resistance to future liquefaction. However, Lees et al. (Soil Dynamics and Earthquake Engineering, 79, 304-314, 2015) conducted an analysis comparing CPT-based strength profiles and subsequent liquefaction susceptibility at sites in Christchurch both before and after the February 2011 earthquake, and concluded that no significant strengthening occurred and that the liquefaction risk in Christchurch after the earthquake remained the same as it was beforehand. The study by Orense et al. (Geotech Eng J SEAGS & AGSSEA, 43(2), 8-17, 2012) came to similar conclusions and therefore our view is that the post-earthquake CPT data is appropriate for assessing liquefaction susceptibility. This explanation has been added to the text.*

20

19. P. 10, ll. 1-5: Are Vs profiles at only 13 points sufficient to estimate Vs profiles for the entire region? This may be appropriate if there are no changes in geology or topography across the city but it sounds too few for accurate results. A brief discussion on the uncertainties with this approach would be useful here. It would be useful to check the robustness of the interpolated profiles by removing one or more of the 13 profiles and comparing the results.

30

*We acknowledge that modelling Vs profiles for an entire city based on 13 samples has major limitations and the resulting estimations therefore carry significant uncertainty. We have added a comment to this effect. However we are not sure there is much value in carrying out the robustness check as suggested by the reviewer. Due to the aforementioned limitations, the check will almost certainly show that the interpolated profiles are not robust.*

5 *However, testing the limitations/weaknesses of input data is not the point of this study. On the contrary, the point of this study is to show what can be done with weak/limited datasets and more specifically, to investigate how a non-academic loss estimation analyst can use realistic, publicly available (but weak) data, together with existing models, to estimate liquefaction occurrence. The sample of 13 Vs profiles, acquired from an academic journal paper, is exactly the type of realistic dataset that a non-academic analyst can expect to have access to and the*  
10 *case study analysis has shown that despite its limitations, the resulting LPI model that corresponds to it performs better than most other simplified models. There is little that an analyst can do if an input dataset is weak and nothing better is available – they can either make use of what they have or not conduct the analysis at all. It is not clear what value there is to be gained simply by proving that the input dataset is weak.*

15 20. Section 4: It could help readability to split this section up into subsections for each of the tests.

*We agree that Section 4 would benefit from being split into sub-sections and this has been done*

21. Section 4: Why are the Zhu et al. (2015) models performing poorly when the Christchurch data was used in  
20 their development? There is a little discussion of this on pp. 14-15 but more discussion could be useful.

*There are two possible reasons for this. One is that the model development and test datasets have come from different sources and so may not be identical. The second, which has already been identified in the discussion, is that the Zhu et al. model has been optimised for quantification of the areal extent of liquefaction, whereas in the*  
25 *case study, site specific predictions are being tested. In light of this, the authors are keen to stress that the results of the case study do not contradict or invalidate the findings of Zhu et al. (2015) and a comment has been added to this effect.*

22. P. 5, l. 28: There is a problem with the grammar in the phrase “in both models though that the observed rates  
30 that are”.

*The sentence has been amended to make this clearer*

23. P. 17, l. 19, “tectonic uplift”: Could it not also be “tectonic subsidence”? What about just saying “tectonic movements”?

5

*We have amended the text accordingly*

24. P. 18, ll. 26-27: There is something missing from the sentence “To calculate duration, there are 19 strong-motion accelerograph stations in Christchurch that record ground motions at 0.02s intervals” as the stations are not there just to calculate duration.

10

*We have amended this sentence to make it clearer that duration can be measured from stations but this is not their primary purpose.*

15

25. P. 19, l. 18: It could be useful to say that even though the methods based on LPI are the best approaches tested that they still do not predict very well.

*We acknowledge this is a fair conclusion and have added a comment to this effect*

20

26. Figure 1: More information could be added to this map, e.g. the faults that ruptured in these earthquakes, the locations of the strong-motion stations used to estimate the durations, the locations of the 13 Vs profiles and the main areas of liquefaction (Figure 3). Currently, this map is not that useful and could be removed or combined with Figure 2 and/or Figure 3.

25

*We have added the suggested information to Figure 1*

27. Table 4: It would be useful to combine this with Table 5.

*Table 4 relates to section 3 (Model test application) and Table 5 relates to section 4 (Results). Furthermore,*

30

*Table 4 is referenced in the text much earlier than Table 5. We do not believe it would be appropriate to combine these tables*

28. Table 6: This could be added as an additional two lines to Table 5.

*We agree and these two tables have been combined*

5

29. Table 7: Is there not space to include these results in Table 5?

*It is important to stress the difference between Tables 5 and 7. The results in Table 5 are based on specified fixed thresholds, whereas the results in Table 7 are based on optimised thresholds. Therefore combining the two tables would necessitate the addition of two columns to identify the relevant threshold in each case. Therefore it may be possible to combine the tables if it can be published in landscape format - in portrait the combined table would appear very cluttered. At this stage we feel it is better and clearer to keep these two tables distinct.*

10

30. Table 9: Could these numbers be added to Table 3 after conversion to SI units (e.g. cm)?

15

*We have added this data to Table 3*

31. Tables 10 and 11: Give the units of the values reported here. Metres?

20 *Yes, this is metres and the tables have been updated accordingly*

32. In addition to my comments yesterday, it would also significantly improve this manuscript to include some maps showing the areas predicted to liquefy by each of the methods. That would help understanding of the differences between the results rather than relying on statistical tests.

25

*We have added the type of maps suggested by the reviewer, however due to the number of models tested in this paper, we believed it was appropriate to only include a selection of maps for the best performing models (LPI1, LPI3, ZHU1 and ZHU2)*

30

**Reviewer #2:** This is a review of "Evaluating Simplified Methods for Liquefaction Assessment for Loss Estimation". This paper is overall well written, interesting, and needing only minor revisions.

*We thank the reviewer for his/her confirmation of the quality of the paper and recommendation for publication. Our responses to specific comments are included below.*

5 My comments, in no order of importance are as follows:

(a) Abstract. Please put more quantitative description of data/results into the abstract, not just qualitative.

10 *As per the reviewer's suggestion, we have added more quantitative information to the abstract relating to data and comparison of model performance*

(b) Section 1. Although there is a good 'motivation' and letting the reader know how the paper is organized in the last paragraph, can you refer to what the paper is about early on? Perhaps in or at the end of the paragraph a sentence that says "Here we investigate ....".

15

*We have added a short description of the paper's topic at the end of the second paragraph of the introduction*

(c) Prediction. Please evaluate throughout the use of the word prediction (time, place, magnitude) if that is meant, or probabilistic forecasting. If prediction really is meant, then make this clear why, and to what degree.

20

25 *We do not use the word 'prediction' in the sense that the reviewer defines it and therefore acknowledge that the use of the word 'prediction' may be misleading and other terms should be used. Although the Hazus and Zhu models estimate liquefaction probabilities, the LPI models do not and furthermore, the final outputs from all the models are deterministic estimates of liquefaction occurrence, which are conditional on an earthquake of specified place and magnitude having occurred. Therefore the description of the forecasts as probabilistic is not necessarily appropriate here and we have therefore amended the text to replace instances of the word 'prediction' with either 'forecast' or 'estimate'.*

30 (d) Testing. I am not a fan of the use of the word 'testing' in the natural hazard community. See the following paper for why: [http://www.nssl.noaa.gov/users/brooks/public\\_html/feda/papers/Oreskes1.pdf](http://www.nssl.noaa.gov/users/brooks/public_html/feda/papers/Oreskes1.pdf)

*We thank the reviewer for pointing us towards this interesting and insightful article, which offers the view that words such as 'testing' are misused in the scientific community to give a false view of model perfection. With respect to our case study, we are clear throughout our paper that the objective is to compare the relative quality/performance of the range of models with respect to a set of observations. We are not claiming that any one model is 'right' and believe we are using the word 'testing' correctly. Nevertheless, it may be open to misinterpretation and for that reason, we have removed the words 'test' and 'testing' where appropriate (i.e. not where it refers to the name of a test) and have replaced with a suitable alternative. Most commonly we have used 'compare' or 'comparison' to emphasise this aspect of the study.*

10 (e) Where possible, avoid acronyms in Figure captions/Table headers (or spell them out the first time) to make the paper a tad less 'jargon' rich. Figures and tables should be as stand-alone as possible so if someone uses them (without the paper) one can tell from the figure caption/table header what it is about. Particularly important for this is your figure comparing all the methods—you need to then state what all the acronyms mean.

15 *We acknowledge that the paper does contain a large number of acronyms and have added more information to some figure captions and table headers. Where tables/figures contain model acronyms, writing out the full model names is no more informatve and writing out the full descriptions is impractical and so instead we have added a reference to Table 4 in the caption.*

20 (f) Would it be possible to provide an overview table of all the acronyms, and what data is being put into each one? This would be a nice 'tutorial' table that is more likely to be cited by people.

*We are not entirely clear what the reviewer means by this since a list of acronyms and list of variables are already provided in Tables 1 and 2 respectively. However we have added a column to Table 2 identifying the variable input data that goes into evaluating each of the listed variables.*

25 (g) Sensitivity of models to data input. It would be very nice to see more on how much the outputs (what you call prediction) are sensitive to slight changes in the inputs. Again, a comparison between different types of liquefaction models would be very useful.

30

*We have carried out two sensitivity tests involve variation of shear wave velocity and peak ground acceleration. However due to the number of models tested in this paper, we have only carried out the sensitivity testing for two of the best performing models (e.g. the two best models or similar) and have added a summary to the revised manuscript. We would also like to point out that the optimisation of threshold, already carried out for the*

5 *original paper, is also a type of sensitivity test.*

# Evaluating Simplified Methods for Liquefaction Assessment for Loss Estimation

Indranil Kongar<sup>1</sup>, Tiziana Rossetto<sup>1</sup>, Sonia Giovinazzi<sup>2</sup>

<sup>1</sup>Earthquake and People Interaction Centre (EPICentre), Department of Civil, Environmental and Geomatic Engineering,  
5 University College London, London, WC1E 6BT, United Kingdom

<sup>2</sup>Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, 8140, New Zealand

*Correspondence to:* Indranil Kongar (ucfbiko@ucl.ac.uk)

**Abstract.** Currently, some catastrophe models used by the insurance industry account for liquefaction simply by applying a simple factor to shaking-induced losses. The factor is based only on local liquefaction susceptibility and -this highlights so  
10 there is a need for a more sophisticated approach to incorporating the effects of liquefaction in loss models -is needed.  
This study compares eleven unique models, each based on one of three principal simplified liquefaction assessment  
methodologies methods: liquefaction potential index (LPI) calculated from shear-wave velocity; the HAZUS software  
methodology method; and a methodology method created specifically to make use of USGS remote sensing data. Data from  
the September 2010 Darfield and February 2011 Christchurch earthquakes in New Zealand are used to compare observed  
15 liquefaction occurrences to predictions forecasts from these models using binary classification performance measures. The  
analysis shows that the best performing model is the LPI calculated using known shear-wave velocity profiles, which  
correctly forecasts 78% of sites where liquefaction occurred and 80% of sites where liquefaction did not occur, when the  
threshold is set at 7. However, -although the seis data may not always be available to insurers. The next best model is also  
based on LPI, but uses shear-wave velocity profiles simulated from the combination of USGS  $V_{S30}$  data and empirical  
20 functions that relate  $V_{S30}$  to average shear-wave velocities at shallower depths. This model correctly forecasts 58% of sites  
where liquefaction occurred and 84% of sites where liquefaction did not occur, when the threshold is set at 4. These scores  
increase to 78% and 86% respectively when forecasts are based on liquefaction probabilities that are empirically related to  
the same values of LPI. This model is potentially more useful for insurance since the input data are is publicly available.  
HAZUS models, which are commonly used in studies where no local model is available, perform poorly and incorrectly  
25 forecast 87% of sites where liquefaction occurred, even at optimal thresholds. This paper also considers two models  
(HAZUS and EPOLLS) for estim prediction of the scale of liquefaction in terms of permanent ground deformation but finds  
that both models perform poorly, with correlations between observations and forecasts lower than 0.4 in all cases, and  
thus Therefore these models potentially provide negligible additional value to loss estimation analysis outside of the regions  
for which they have been developed.

## 1 Introduction

The recent earthquakes in Haiti (2010), Canterbury, New Zealand (2010-11) and Tohoku, Japan (2011) highlighted the significance of liquefaction as a secondary hazard of seismic events and the significant damage that it can cause to buildings and infrastructure. However, the insurance sector was caught out by these events, with catastrophe models underestimating the extent and severity of liquefaction that occurred (Drayton and Verdon, 2013). A contributing factor to this is that the method used by some catastrophe models to account for liquefaction is based only on liquefaction susceptibility, a qualitative parameter that considers only surficial geology characteristics. Furthermore, losses arising from liquefaction are predietestimated by adding an amplifier to losses predietestimated due to building damage caused by ground shaking (Drayton and Vernon 2013). There is a paucity of past event data on which to calibrate an amplifier and consequently, significant losses from liquefaction damage will only be predietestimated if significant losses are already predietestimated from ground shaking, whereas it is known that liquefaction can be triggered at relatively low ground shaking intensities (Quigley et al., 2013).

Therefore there is scope within the insurance and risk management sectors to adopt more sophisticated approaches for predietforecasting liquefaction for both future risk assessments and post-event rapid response analyses. It is also important to develop a better understanding of the correlation between liquefaction effects/risk and physical damage of the built environment, similar to the fragility functions that are used to prediet-estimate damage associated with ground shaking. This is particularly the case for critical infrastructure systems since, whilst liquefaction is less likely than ground shaking to be responsible for major building failures (Bird and Bommer, 2004), it can have a major impact on lifelines such as roads, pipelines and buried cables. Loss of power and reduction in transport connectivity are major factors affecting the resilience of business organizations in response to earthquakes as they can delay the recommencement of normal operations. Evaluating the seismic performance of infrastructure is therefore critical to understanding indirect economic losses caused by business interruption and to achieve this it is necessary to assess the liquefaction risk in addition to that posed by ground shaking. Therefore in this paper we investigate the performance of a range of models that can be applied to forecast the occurrence and scale of liquefaction based on simple and accessible input datasets. The performances are evaluated by comparing model forecasts to observations from the 2010-11 Canterbury earthquake sequence.

Bird and Bommer (2004) surmised that there are three options that loss estimators can select to deal with ground failure hazards. They can either ignore them; or use a simplified approach; or conduct a detailed geotechnical assessment. The first of these options will likely lead to underestimation of losses in earthquakes where liquefaction is a major hazard and lead to recurrence of the problems faced by insurers following the 2010-11 Canterbury earthquakes in particular. The last option, detailed assessment, is appropriate for single-site risk analysis but is impractical for insurance loss estimation purposes because: 1) insurers are unlikely to have access to much of the detailed geotechnical data required as inputs to these

methods; 2) they may not have the in-house expertise to correctly apply such methods and engaging consultants may not be a viable option; and 3) loss estimation studies are often conducted on a regional, national or supra-national scale for which detailed assessment would be too expensive and time-consuming.

5 | There are three stages to ~~predict-forecasting~~ the occurrence of liquefaction and its scale (Bird et al., 2006). First it is necessary to determine whether soils are susceptible to liquefaction. Liquefaction susceptibility is based solely on ground conditions with no earthquake-specific information. This is often done qualitatively and currently this is also the full extent to which liquefaction risk is considered in some catastrophe models (Drayton and Verdon, 2013). The next step is to determine liquefaction triggering, which determines the likelihood of liquefaction for a given earthquake based on the  
10 | susceptibility and other earthquake-specific parameters. Finally the scale of liquefaction can be ~~predicted-estimated~~ as a permanent ground deformation (PGDf). Since current catastrophe modelling practice is to consider only the first stage, liquefaction susceptibility, this paper focuses primarily on the extension of this practice to include liquefaction triggering.

The models assessed in this paper have been selected because their input requirements are limited to data that are in the  
15 | public domain or could be easily obtained without significant time or cost implications, arising for example from detailed site investigation. Furthermore, the models are appropriate for regional-scale analysis and although some engineering judgment is required in their application, they do not require specialist geotechnical expertise. In section 2, each of the models assessed in this paper are described and section 3 presents a summary of the liquefaction observations from the Canterbury earthquake sequence and the ~~methodologymethod~~ used to ~~test-compare~~ the ~~predictive-capabilities-of-the~~ model  
20 | ~~forecasts~~ against observations. The results and statistical analysis of the model assessment are presented in section 4, in relation to deterministic ~~predictionsforecasts~~, and in section 5, in relation to probabilistic ~~predictionsforecasts~~. Finally section 6 briefly considers the ~~predictive-capabilityperformance~~ of simplified models for quantifying PGDf.

## 2 Liquefaction assessment models

25 | Nine liquefaction ~~prediction-forecasting~~ models are ~~tested-compared~~ in this paper, including three alternative implementations of the liquefaction potential index method proposed by Iwasaki et al. (1984); three versions of the liquefaction models included in the HAZUS<sup>®MH</sup> MR4 software (NIBS, 2003); and three distinct models proposed by Zhu et al. (2015). This section summarises how each of the models are applied to make site-specific liquefaction  
| ~~predictionsforecasts~~. This paper presents a large number of acronyms and variables. For clear reference, Table 1 lists the acronyms used in this paper and Table 2 lists the variables used.

## 2.1 Liquefaction potential index

The most common approach used to ~~predict-forecast~~ liquefaction triggering is the factor of safety against liquefaction,  $FS$ , which is defined as the ratio of the cyclic resistance ratio,  $CRR$ , and the cyclic stress ratio,  $CSR$ , for a layer of soil at depth,  $z$  (Seed and Idriss, 1971).  $CSR$  can be expressed by:

$$CSR = 0.65 \left( \frac{a_{\max}}{g} \right) \left( \frac{\sigma_v}{\sigma'_v} \right) r_d \quad (1)$$

where  $a_{\max}$  is the peak horizontal ground acceleration;  $g$  is the acceleration of gravity;  $\sigma_v$  is the total overburden stress at depth  $z$ ;  $\sigma'_v$  is the effective overburden stress at depth  $z$ ; and  $r_d$  is a shear stress reduction coefficient given by:

$$\begin{aligned} r_d &= 1 - 0.00765z, \text{ for } z < 9.2\text{m} \\ r_d &= 1.174 - 0.0267z, \text{ for } z \geq 9.2\text{m} \end{aligned} \quad (2)$$

$CRR$  is normally calculated from geotechnical parameters based on cone penetration test (CPT) or standard penetration test (SPT) results. However, Andrus and Stokoe (2000) propose an alternative method for calculating  $CRR$  based on shear-wave velocity,  $V_s$ , where:

$$CRR = \left[ 0.022 \left( \frac{V_{s1}}{100} \right)^2 + 2.8 \left( \frac{1}{V_{s1}^* - V_{s1}} - \frac{1}{V_{s1}^*} \right) \right] \times MSF \quad (3)$$

where  $V_{s1}$  is the stress-corrected shear wave velocity;  $V_{s1}^*$  is the limiting upper value of  $V_{s1}$  for cyclic liquefaction occurrence, which varies between 200-215m/s depending on the fines content of the soil; and  $MSF$  is a magnitude scaling factor.  $V_{s1}$  is given by:

$$V_{s1} = V_s \left( \frac{P_a}{\sigma'_v} \right)^{0.25} \quad (4)$$

where  $P_a$  is a reference stress of 100kPa. The magnitude scaling factor is given by:

$$MSF = \left( \frac{M_w}{7.5} \right)^{-2.56} \quad (5)$$

Field Code Changed

where  $M_w$  is the moment magnitude of the earthquake. Liquefaction is ~~predicted-forecast~~ to occur when  $FS \leq 1$ , and ~~predicted-forecast~~ not to occur when  $FS > 1$ . However Juang et al. (2005) found that Eq. (3) is conservative for calculating  $CRR$ , resulting in lower factors of safety and over-~~prediction-estimation~~ of the extent of liquefaction occurrence. To correct for this, they propose a multiplication factor of 1.4 to obtain an unbiased estimate of the factor of safety,  $FS^*$ :

$$FS^* = 1.4 \times \frac{CRR}{CSR} \quad (6)$$

$FS^*$  is an indicator of potential liquefaction at a specific depth. However, Iwasaki et al. (1984) noted that damage to structures due to liquefaction was affected by the severity of liquefaction at ground level and so propose an extension to the factor of safety method, the liquefaction potential index,  $LPI$ , which ~~predicts-estimates~~ the likelihood of liquefaction at surface-level by integrating a function of the factors of safety for each soil layer within the top 20m of soil. They calculate  $LPI$  as:

$$LPI = \int_0^{20} F^* (10 - 0.5z) dz \quad (7)$$

where  $F^* = 1 - FS^*$  for a single soil layer. The soil profile can be sub-divided into any number of layers (e.g. twenty 1m layers or forty 0.5m layers), depending on the resolution of data available. Using site data from a collection of nine Japanese earthquakes between 1891 and 1978, Iwasaki et al. (1984) calibrated the  $LPI$  model and determined guideline criteria for determining liquefaction risk. These criteria propose that liquefaction risk is very low for  $LPI = 0$ ; low for  $0 < LPI \leq 5$ ; high for  $5 < LPI \leq 15$ ; and very high for  $LPI > 15$ .

One of the critical considerations for insurers is availability of model input data. For post-event analysis, ground accelerations may be available from various online sources, with one example being the USGS ShakeMaps (USGS, 2014a). ~~However if they are not, then-but otherwise~~ it would be necessary to apply engineering judgment in the selection of ~~an~~ appropriate ground motion prediction equations (either a single equation or multiple equations applied in a logic tree). The  $LPI$  model also requires water table depth and soil unit weights. If these are not known exactly, engineering judgment needs to be applied to estimate these based on information in existing literature. For the specific case study presented in this paper, ~~some  $V_s$  data are available from published sources. Although the use of  $V_s$  negates the requirement for ground investigation, However, more generally  $V_s$  data itself-areis not ~~commonly~~ in the public domain and would require ground investigation to acquire. Even in cases where  $V_s$  data are available, may not necessarily be available across the entire study area, thus requiring geostatistical techniques to ~~extrapolate~~interpolate. Consequently this method may only be applicable in a small number of study areas.~~

Field Code Changed

Field Code Changed

Formatted: Font: Italic

Formatted: Subscript

Formatted: Font: Italic

Formatted: Subscript

To extend the applicability of the *LPI* model, two approaches are proposed to approximate  $V_S$  from more readily available data. The first approach uses  $V_{S30}$ , the average shear wave velocity across the top 30m of soil, as a constant proxy for  $V_S$  for all soil layers. Global estimates for  $V_{S30}$  at approximately 674m grid intervals are open-access from the web-based US Geological Survey Global  $V_{S30}$  Map Server (USGS, 2013), so this is an appealing option for desktop assessment. ~~TheOne~~ disadvantage of this approach is that the likelihood of liquefaction occurrence in the *LPI* method is controlled by the presence of soil layers near the surface with low  $V_S$ . Furthermore there is a maximum value of  $V_S$  at which liquefaction can occur. Hence the use of  $V_{S30}$  as a proxy for all layers will result in an overestimation of  $V_S$ , *CRR* and  $FS^*$  at layers closer to the surface, and therefore an underestimation of *LPI* and liquefaction risk. ~~This is compounded by the weakness of the USGS  $V_{S30}$  dataset, since the data are estimated from topographic slope and the correlation between these two variables is weak.~~

Formatted: Font: Italic  
Formatted: Subscript

The second approach proposes the ~~manipulation~~use of the same  $V_{S30}$  data ~~but manipulates it~~ to simulate a more realistic  $V_S$  profile in which velocities decrease towards the surface rather than being constant. Boore (2004) proposes ~~simple linear~~ empirical functions to extrapolate  $V_{S30}$  values in situations where shear wave velocity data are only known up to shallower depths, based on observations from the United States and Japan. ~~It is proposed to use-invert the Boore (2004)se~~ empirical functions in reverse ~~and use them-~~ to back-calculate shallower average shear wave velocities from  $V_{S30}$  data from the USGS Global Server (USGS, 2013). ~~However, it should be noted that since the original function was not developed using orthogonal regression, this inversion is an additional source of uncertainty.~~ For simplicity it is proposed to only use the empirical functions to calculate  $V_{S10}$  (average shear wave velocity across top 10m) and  $V_{S20}$  (average shear wave velocity across top 20m). The calculated value for  $V_{S10}$  can then be used as a proxy for  $V_S$  at all soil layers between 0-10m depth and both the  $V_{S10}$  and  $V_{S20}$  values can be used to determine an equivalent proxy for all soil layers between 10-20m. From manipulation of the Boore (2004) empirical functions and the formula for calculating averaged shear wave velocities, the following equations determine the proxies to be used in the two depth ranges:

$$V_{S(0-10)} = 10^{\left(\frac{\log V_{S30} - 0.042062}{1.0292}\right)} \quad (8)$$

Field Code Changed

$$V_{S(10-20)} = 2 \times 10^{\left(\frac{\log V_{S30} - 0.025439}{1.0095}\right)} - V_{S(0-10)} \quad (9)$$

Field Code Changed

In this study, both of these approximations are adopted in addition to the use of known  $V_S$  profiles, resulting in the assessment of three implementations of the *LPI* model.

## 2.2 HAZUS

HAZUS<sup>®MH</sup> MR4 (from here on referred to as HAZUS) is a loss estimation software package produced by the National Institute of Building Sciences (NIBS) and distributed by the Federal Emergency Management Agency (FEMA) in the United States. The software accounts for the impacts of liquefaction and the Technical Manual (NIBS, 2003) describes the methodology used to evaluate the probability of liquefaction.

HAZUS divides the assessment area into six zones of liquefaction susceptibility, from none to very high. This can be done by either, interpreting surficial geology from a map and cross-referencing with the table published in the Manual, or by using an existing liquefaction susceptibility map. Surface geology maps are generally not open-access or free to non-academic organizations and some basic geological knowledge is required to be able to cross-reference mapped information with the zones in the HAZUS table. Hence, the first approach may be problematic for insurers who do not have the requisite in-house expertise. Where liquefaction susceptibility maps are available, unless they use the same zonal definitions as HAZUS, it will be necessary to make assumptions on how zones translate between the third party map and the Manual.

For a given liquefaction susceptibility category, the probability of liquefaction occurrence is given by (NIBS, 2003):

$$P[Liq] = \frac{P[Liq | PGA = a]}{K_M K_W} P_{mi} \quad (10)$$

where  $P[Liq | PGA = a]$  is the conditional probability of liquefaction occurrence for a given susceptibility zone at a specified level of peak horizontal ground motion,  $a$ ;  $K_M$  is the moment magnitude correction factor;  $K_W$  is the ground water correction factor; and  $P_{mi}$  is the proportion of map unit susceptible to liquefaction, which accounts for the real variation in susceptibility across similar geologic units. The conditional probability is zero for the susceptibility zone 'None', and for the other susceptibility zones, the conditional probabilities are given by linear functions of acceleration (distinct for each zone), which are not repeated here. The moment magnitude and ground water correction factors are given by:

$$K_M = 0.0027M_w^3 - 0.00267M_w^2 - 0.2055M_w + 2.9188 \quad (11)$$

$$K_W = 0.022d_w + 0.93 \quad (12)$$

where  $d_w$  is the depth to ground water. The map unit factor is a constant for each susceptibility zone, with values of 0.25, 0.20, 0.10, 0.05, 0.02 and 0, going from 'Very high' to 'None'. In addition to the problems identified for determining

Field Code Changed

Field Code Changed

Field Code Changed

liquefaction susceptibility, the HAZUS method also requires water table depth to be known or estimated and judgment on selection of appropriate ground motion prediction equation if ShakeMap or equivalent data ~~are~~ not available.

### 2.3 Zhu et al. (2015)

Zhu et al. (2015) propose empirical functions to ~~predict-estimate~~ liquefaction probability specifically for use in rapid response and loss estimation. They deliberately use predictor variables that are readily accessible, such as  $V_{S30}$  and do not require any specialist knowledge to be applied. The functions have been developed using logistic regression on data from the earthquakes that occurred in Kobe, Japan on January 17<sup>th</sup> 1995 and in Christchurch, New Zealand on February 22<sup>nd</sup> 2011.

~~Forecasts from T~~he resulting functions have been ~~tested-oncompared to~~ observations from the January 12<sup>th</sup> 2010 Haiti earthquake. ~~Since these functions have been developed using data from the Christchurch earthquake, there is an element of circularity in assessing their performance against observations from the same event. However, it is worth noting that the datasets used to develop these functions have not come from the same source as the observations used in this case study. Furthermore, the functions have been calibrated to optimise estimation of the areal extent of liquefaction, whereas in this case study, it is the ability of the functions to make site specific forecasts that is being assessed.~~

For a given set of predictor variables, the probability of liquefaction is given by the function:

$$P[Liq] = \frac{1}{1 + e^{-X}} \quad (13)$$

where  $X$  is a linear function of the predictor variables. Zhu et al. (2015) propose three linear models that are applicable to the Canterbury region and are adopted in this study: a specific local model for Christchurch; a regional model for use in coastal sedimentary basins (including Christchurch) and a global model that is applicable more generally. For the global model, the linear predictor function,  $X_G$ , is given by:

$$X_G = 24.1 + \ln PGA_{M,SM} + 0.355CTI - 4.784 \ln V_{S30} \quad (14)$$

where,  $CTI$  is the compound topographic index, used as a proxy for saturation, and which can be obtained globally from the USGS Earth Explorer web service (USGS, 2014b);  $V_{S30}$  is obtained from the USGS Global Server (USGS 2013); and  $PGA_{M,SM}$  is the product of the peak horizontal ground acceleration from ShakeMap estimates (USGS 2014a) and a magnitude weighting factor,  $MWF$ , given by:

Field Code Changed

Field Code Changed

$$MWF = \frac{M_w^{2.56}}{10^{2.24}} \quad (15)$$

For the regional model, the linear predictor function,  $X_R$ , is given by:

$$X_R = 15.83 + 1.443 \ln PGA_{M,SM} + 0.136CTI - 9.759ND - 2.764 \ln V_{S30} \quad (16)$$

5 where additionally,  $ND$  is the distance to the coast, normalized by the size of the basin, i.e. the ratio between the distance to the coast and the distance between the coast and inland edge of the sedimentary basin (soil/rock boundary). The location of the inland edge can be estimated from a surface roughness calculation based on a digital elevation model (USGS, 2014b) or by using  $V_{S30}$  data such that the inland edge is assumed to be the boundary between NEHRP site classes C (soft rock) and D (stiff soil), (i.e. at  $V_s = 360\text{m/s}$ ). For the Christchurch-specific local model, the linear predictor function,  $X_L$ , is given by:

$$10 \quad X_L = 0.316 + 1.225 \ln PGA_{M,SM} + 0.145CTI - 9.708ND \quad (17)$$

For applicability within the insurance sector, this model presents an advantage over LPI and HAZUS since the only parameter that requires engineering judgment is the selection of ground motion prediction equation if ShakeMap or equivalent data ~~are~~ not available.

### 15 **3 Model ~~test-assessment~~ application**

This section summarises the procedure for comparing the model ~~predictions-forecasts~~ to observations from the Canterbury ~~earthquake~~ earthquake sequence. A brief description is provided of the liquefaction observation dataset and the additional datasets accessed in order to provide the required inputs to the nine models. This is followed by a discussion on the conversion of quantitative model outputs to categorical ~~liquefaction~~ liquefaction ~~predictions-forecasts~~ and an ~~explanantion~~ explanation of the ~~test~~ diagnostics used to assess model performance.

#### 20 **3.1 Liquefaction observations**

The methods described in the previous section are ~~tested-compared~~ for two case studies from the Canterbury earthquake sequence: the  $M_w$  7.1 Darfield earthquake on September 4<sup>th</sup> 2010 and the  $M_w$  6.2 Christchurch earthquake on February 22<sup>nd</sup> 2011 ([GNS Science, 2014](#)), as identified in [Figure 1](#) ~~Figure 1~~. The corresponding peak horizontal ground acceleration contours for each earthquake are shown in Figure 2.

Surface liquefaction observation data have been obtained from two sources: ground investigation data provided directly from Tonkin & Taylor, geotechnical consultants to the New Zealand Earthquake Commission (EQC) (van Ballegooy et al., 2014) and maps stored within the Canterbury Geotechnical Database (CGD, 2013a), an online repository of geotechnical data and reports for the region set up by EQC for knowledge sharing after the earthquakes. The data provided by Tonkin & Taylor includes records from over 7,000 geotechnical investigation sites across Christchurch. After each earthquake, a land damage category is attributed to each site, representing a qualitative assessment of the scale of liquefaction observed. There are six land damage categories, but since this study only investigates liquefaction triggering, the categories are converted to a binary classification of liquefaction occurrence. These data are supplemented by the maps from the CGD which show the areal extent of the same land damage categories. To ensure equivalence in the test study, all models are applied to the same test study area for each earthquake, which is the region for which the input data for all models are available. The test study area is divided into a grid of 100m x 100m squares, generating 25,100 test observation sites. It is noted however that at some locations within Christchurch, no liquefaction observations are available so these sites are excluded from the subsequent analysis. As a result, the test study area consists of 20,147 sites for the Darfield earthquake and 22,803 sites for the Christchurch earthquake. The observations from the two events are shown in Figure 3.

### 3.2 Prediction-Forecast model inputs

This study includes three implementations of the LPI model: 1) using known  $V_s$  profiles (referred to as LPI1 in this paper); 2) using  $V_{s30}$  as a proxy for  $V_s$  (LPI2); and 3) using 'realistic'  $V_s$  profiles simulated from  $V_{s30}$  and the Boore (2004) functions (LPI3). The geotechnical investigation data provided by Tonkin & Taylor also includes values of  $LPI$  calculated at each site from CSPT data rather than  $V_s$ . Although this approach is not feasible for insurers, for reference its predictive-forecasting power is also tested-compared here, and this implementation is referred to as LPIref. Historically it has been thought that after liquefaction occurs, soils densify and increase their resistance to future liquefaction. However, Lees et al. (2015) conducted an analysis comparing CPT-based strength profiles and subsequent liquefaction susceptibility at sites in Christchurch both before and after the February 2011 earthquake. They concluded that no significant strengthening occurred and that the liquefaction risk in Christchurch after the earthquake remained the same as it was beforehand. The study by Orense et al. (2012) came to similar conclusions and therefore for the purposes of this case study, post-earthquake CPT data is appropriate for assessing liquefaction susceptibility.

A water table depth of 2m has been assumed across Christchurch, reflecting the averages described by Giovinazzi et al. (2011) – 0-2m in the eastern suburbs and 2-3m in the western suburbs – and soil unit weights of 17kPa above the water table and 19.5kPa below the water table are assumed, as suggested by Wotherspoon et al. (2014).  $V_{s30}$  data for LPI2 and LPI3 are taken from the USGS web server, with point estimates on an approximately 674m grid.

Formatted: Font: 10 pt, Not Italic

Wood et al. (2011) have published  $V_S$  profiles for 13 sites across Christchurch obtained using surface wave testing methods. These sites are identified in Figure 1. In GIS, the profiles are converted to point data for each 1m depth increment from 0-20m, so that each point represents the  $V_S$  at that site for a single soil layer and there are a total of 13 points for each soil layer. Ordinary kriging (with log transformation to ensure non-negativity) is applied to the points in each soil layer to create interpolated  $V_S$  raster surfaces for each layer. Interpolation over a large area from such a small number of points is likely to result in estimations carrying significant uncertainty. However from the perspective of commercial loss estimation, this is typical of the type of data that an analyst may be required to work with and so there is value in investigating its efficacy. Whilst Andrus and Stokoe (2000) advise that the maximum  $V_{S1}$  can range from 200-215m/s depending on fines content, subsequent work by Zhou and Chen (2007) indicates that the maximum  $V_{S1}$  could range between 200-230m/s. In the absence of specific fines content data, a median value of 215m/s is assumed to be the maximum. In practice, a soil layer may have a value of  $V_{S1}$  below this threshold but not be liquefiable because the soil is not predominantly clean sand. Because of the regional scale of this analysis though, site-specific soil profiles (as distinct from  $V_S$  profile) are not taken into account in determining whether a soil layer is liquefiable. Goda et al. (2011) suggest the use of 'typical' soil profiles to determine the liquefaction susceptibility of a soil layer at a regional scale. Borehole data at sites close to the 13  $V_S$  profile sites are available from the Canterbury Geotechnical Database (CGD, 2013c). These indicate that in the eastern suburbs of Christchurch, soil typically consists predominantly of clean sand to 20m depth, with some layers of silty sand. On the western side of Christchurch however there is an increasing mix of sand, silt and gravel in soil profiles, particularly at depths down to 10m. Therefore it is possible, particularly in western suburbs, that the calculated  $V_{S1}$  values may indicate liquefiable soil layers when they are in fact not, which would lead to overestimation of  $LPI$  and the extent of liquefaction.

For the implementation of model LPI3, it could be argued that rather than using the Boore (2004) relationships to estimate  $V_S$  profiles at shallower depths from  $V_{S30}$ , the local  $V_S$  data published by Wood et al. (2011) could be used to develop a locally calibrated model. This would be preferable from a purely scientific perspective. However, the purpose of this study is to investigate the potential for a simple 'global' model for commercial application, and this is defined in part as a model that makes use of methods already in the literature and does not require additional model development. Nevertheless, when using existing models it is useful to assess their applicability to a study area, and the  $V_S$  profiles published by Wood et al. (2011), can be used to assess the suitability of the Boore (2004) relationships in Christchurch. Figure 4 shows plots of  $V_{S30}$  against  $V_{S10}$  and  $V_{S20}$  as calculated from the observed profiles and compares these to the Boore (2004) functions. The plots show that that the relationships exhibit a small bias towards the underestimation of  $V_{S30}$ . When inverted, the application of these relationships to Christchurch may therefore result in the overestimation of  $V_S$  at shallower depths and therefore underestimate liquefaction occurrence. However, the majority of observed values are within the 95% confidence intervals and so the relationships can be deemed to be applicable.

Formatted: Font: Italic

Formatted: Subscript

For application of the HAZUS method, liquefaction susceptibility zones have to be identified to determine the values of model input parameters. In this paper liquefaction susceptibility zones are adopted from the liquefaction susceptibility map available from the Canterbury Maps web-resource operated by Environment Canterbury Regional Council (ECan, 2014). From the map it is possible to identify four susceptibility zones: 'None', 'Low', 'Moderate' and 'High'. However, six susceptibility zones are defined by HAZUS (NIBS, 2003). Since the Canterbury zones cannot be sub-divided, it is necessary to map the Canterbury zones onto four of the HAZUS zones. In HAZ1 the zones are mapped simply by matching names; in HAZ2, the 'Low' and 'High' zones in Canterbury are mapped to the more extreme 'Very low' and 'Very high' zones in HAZUS; and in HAZ3, the relevant input parameters for each zone are taken to be the average of those identified in HAZ1 and HAZ2. The mapping between susceptibility zones in each of the implementations described in Table 3. As with the LPI model, depth to water table is assumed to be 2m across Christchurch.

Three models proposed by Zhu et al. (2015) are ~~tested-compared~~ in this paper: 1) the global model (referred to as ZHU1); 2) the regional model (ZHU2); and 3) the local model (ZHU3). The *PGA* 'shakefields' from the Canterbury Geotechnical Database (CGD, 2013b) are used as equivalents to the USGS ShakeMap. *CTI* (USGS, 2014b), at approximately 1km resolution and  $V_{s30}$  (USGS, 2013) are downloaded from the relevant USGS web resources. In total nine model implementations are being ~~tested-compared~~, based on three general approaches (see Table 4).

### 3.3 Site-specific ~~prediction-forecasts~~

When using probabilistic ~~prediction-forecasting~~ frameworks, one can interpret the calculated probability as a regional parameter that describes the spatial extent of liquefaction rather than discrete site specific ~~predictions-forecasts~~, and indeed Zhu et al. (2015) specifically suggest that this is how their model should be interpreted. So for example, one would expect 30% of all sites with a liquefaction probability of 0.3 to exhibit liquefaction and 50% of all sites with a liquefaction probability of 0.5, etc. However, when using liquefaction ~~predictions-forecasts~~ as a means to estimate structural damage over a wide area, it is useful to know not just the number of liquefied site but also where these sites are. This is particularly important for infrastructure systems since the complexity of these networks means that damage to two identical components can have significantly different impacts on overall systemic performance depending on the service area of each component and the level of redundancy built in.

There are two ways to generate site specific ~~predictions-forecasts~~ from probabilistic assessments. One approach is to group sites together based on their liquefaction probability, and then randomly assign liquefaction occurrence to sites within the group based on that probability, e.g. by sampling a uniformly distributed random variable. This method is good for ensuring that the spatial extent of the site specific ~~predictions-forecasts~~ reflect the probabilities, but since the locations are selected

randomly it has limited value for comparison of ~~predictions-forecasts~~ to real observations from past earthquakes. It can be more useful for generating site specific ~~predictions-forecasts~~ for simulated earthquake scenarios.

Another method is to set a threshold value for liquefaction occurrence, so all sites with a probability above the threshold are ~~predicted-forecast~~ to exhibit liquefaction and all sites with a probability below the threshold are ~~predicted-forecast~~ to not exhibit liquefaction. The disadvantage of this approach is that the resulting ~~predictions-forecasts~~ may not reflect the original probabilities. For example if the designated threshold probability is 0.5 and all sites have a calculated probability greater than this (even if only marginally), then every site will be ~~predicted-forecast~~ to liquefy. Conversely if all sites have a probability below 0.5, then none of the sites will be ~~predicted-forecast~~ to liquefy. However since there is no random element to the determination of liquefaction occurrence, the ~~predictions-forecasts~~ are more definitive in spatial terms and hence more useful for the ~~model-testing-in~~ this ~~comparative site-specific~~ study. Although not strictly a probabilistic framework, thresholds can also be used to assign liquefaction occurrence based on *LPI*, by determining a value above which liquefaction is assumed to occur.

For all of the methods however, the issue arises of what value the thresholds should take. No guidance is given for HAZUS, whilst Zhu et al. (2015) propose a threshold of 0.3 to preserve spatial extent, although they also consider thresholds of 0.1 and 0.2. In their original study, Iwasaki et al. (1984) suggest critical values of *LPI* of 5 and 12 for liquefaction and lateral spreading respectively. However other localized studies where the *LPI* method has been applied have found alternative criteria that provide a better fit for observed data as summarized by Maurer et al. (2014). Since there is uncertainty in the selection of threshold values, this study ~~tests-investigates~~ a range of values for each model. Both the observation and ~~prediction-forecast~~ datasets are binary classifications, so standard binary classification measures based on 2 x 2 contingency tables are used to ~~test-compare~~ performance.

### 3.4 ~~Test-Performance~~ diagnostics

Comparison of binary classification ~~predictions-forecasts~~ with observations is made by summarizing data into 2 x 2 contingency tables for each model. The contingency table identifies the true positives (*TP*), true negatives (*TN*), false positives (*FP*, Type I error) and false negatives (*FN*, Type II error). A good ~~predictive-forecasting~~ model would ~~predict~~ ~~forecast~~ both positive (occurrence of liquefaction) and negative (non-occurrence of liquefaction) results well. Diagnostic scores for each model can be calculated based on different combinations and functions of the data in the contingency tables. The true positive rate (*TPR* or sensitivity) is the ratio of true positive ~~predictions-forecasts~~ to observed positives. The true negative rate (*TNR* or specificity) is the ratio of true negative ~~predictions-forecasts~~ to observed negatives. The false positive rate (*FPR* or fall-out) is the ratio of false positive ~~predictions-forecasts~~ to true negatives. ~~The best~~ ~~A useful~~ model would have a high *TPR* and *TNR* (> 0.5) and low *FPR* (< 0.5).

The results presented in a contingency table and associated diagnostic scores assume a single initial threshold value. However, further statistical analysis is undertaken to optimize the thresholds in accordance with the observed data. For a single model, at a specified threshold, the Receiver Operating Characteristic (ROC) is a graphical plot of  $TPR$  against  $FPR$ .  
 5 The line representing  $TPR = FPR$  is equivalent to random guessing (known as the chance or no-discrimination line). A good model has a ROC above and to the left of the chance line, with perfect classification occurring at (0,1). The diagnostic scores for each model are re-calculated with different thresholds and the resulting ROC values are plotted as a curve for the model. Since better models have points towards the top left of the plot, the area under the ROC curve,  $AUC$ , is a generalized measure of model quality that assumes no specific threshold. Since the diagonal of the plot is equivalent to random guessing,  
 10  $AUC = 0.5$  suggests a model has no value, while  $AUC = 1$  is a perfect model. For a single point on the ROC curve, Youden's  $J$ -statistic is the height between the point and the chance line. The point along the curve which maximizes the  $J$ -statistic represents the  $TPR$  and  $FPR$  values obtained from the optimum threshold for that model.

As well as comparing the performance of simplified models to each other, it is also useful to measure the absolute quality of  
 15 each model. Simply counting the proportion of correct ~~predictions-forecasts~~ does not adequately measure model performance since it does not take into the account the proportion of positive and negative observations, e.g. a negatively biased model will result in a high proportion of correct ~~predictions-forecasts~~ if the majority of observations are negative. The Matthews correlation coefficient,  $MCC$ , is more useful for cases where there is a large difference in the number of positive and negative observations (Matthews, 1975). It is proportional to the chi-squared statistic for a 2 x 2 contingency table and its  
 20 interpretation is similar to Pearson's correlation coefficient, so it can be treated as a measure of the goodness-of-fit of a binary classification model (Powers, 2011). From contingency table data,  $MCC$  is given by:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (18)$$

#### 4 Results

25 This section summarises the results of the model applications using contingency table analysis. The results are first presented for analysis using a set of initial assumed thresholds for positive forecasts and subsequently for analysis in which thresholds are optimised for performance. The sensitivity of the forecasts to variation of  $V_{S30}$  and PGA inputs is also assessed.

Field Code Changed

Formatted: Normal, Line spacing: single

Formatted: Font: Italic

Formatted: Subscript

#### 4.1 Contingency table analysis – initial thresholds

An initial set of results using 5 as a threshold value for the LPI models, 0.3 as a threshold for the ZHU models and 0.5 as a threshold value for the HAZUS models is shown in Table 5, alongside the corresponding diagnostic scores.

- 5 The LPI1 and LPIref models are the only models that meet the criteria of having  $TPR$  and  $TNR > 0.5$  and  $FPR < 0.5$ , with the LPI1 model performing better despite being based on  $V_S$  rather than ground investigation data. Table 3 shows that all HAZUS models are very good at ~~predicting-forecasting~~ non-occurrence of liquefaction. However, this is only due to the fact that they are ~~predicting-forecasting~~ no liquefaction all the time, and so their ability to ~~predict-forecast~~ the occurrence of liquefaction is extremely poor. The high  $TNR$  but relatively low  $TPR$  of the three ZHU models indicate that they all show a bias towards ~~the prediction/forecasts~~ of non-occurrence of liquefaction. The difference between  $TPR$  and  $TNR$  is indicative of the level of bias in the model and this regard ZHU2, the regional model shows less bias than in ZHU1, the global model, as would be expected. The bias in the ZHU2 and ZHU3 models is approximately similar although ~~the predictive power of~~ ZHU2 ~~performs~~ slightly better.
- 15 The LPI2 model, using  $V_{S30}$  as a proxy, also shows a very strong bias towards ~~predicting-forecasting~~ non-occurrence, which is expected since  $V_{S30}$  generally provides an overestimate of  $V_S$  for soil layers at shallow depth. At sites where the soil profile of the top 30m is characterized by some liquefiable layers at shallow depth with underlying rock or very stiff soil (e.g. in western and central areas close to the inland edge of the sedimentary basin),  $V_{S30}$  will be high. Hence, this leads to false classification of shallow layers as non-liquefiable. The LPI3 model with simulated  $V_S$  profiles exhibits good performance in ~~the prediction-offorecasting~~ non-occurrence of liquefaction ~~but-onlyand~~ correctly ~~predicts-forecasts just over~~ half of the positive liquefaction observations, indicating bias towards negative ~~predictions/forecasts~~. Although the  $V_S$  profiles generated through this approach are more realistic than using a constant  $V_{S30}$  value, the  $V_S$  at each layer is related to  $V_{S30}$ . Therefore, at sites characterized by a high  $V_{S30}$  value with low  $V_S$  values at shallow depths, even using Eq. (8) and Eq. (9) may not ~~predict~~ ~~estimate~~ sufficiently low values of  $V_{S1}$  to classify the shallow layers as liquefiable. Another factor in the LPI models is the use of the bias-correction factor proposed by Juang et al. (2005). Whilst this correction factor is appropriate when actual  $V_S$  profiles are used, as in LPI1, it may not be appropriate for LPI2 and LPI3 where non-conservative proxies for  $V_S$  are used and the resulting misclassification of liquefiable soil layers balances the conservativeness of the Andrus and Stokoe (2000)  $CRR$  model. The sensitivity of the models to the correction factor is ~~tested-investigated~~ by reproducing the contingency tables for LPI2 and LPI3 with the same threshold values but ignoring the correction factor for  $FS$ . These models are referred to as LPI2b and LPI3b and the new contingency table analysis is presented in ~~Table-6~~Table 5.
- 25
- 30

These results show that not using the bias correction makes little difference to the performance of LPI2, as LPI2b still exhibits an extremely strong bias towards ~~predicting-forecasting~~ non-occurrence of liquefaction. For LPI3 however, the

5 difference is more significant. Without the correction factor, ~~the *TPR* and *TNR* values for LPI3b reverse, with only just over half negative liquefaction occurrences being correctly forecast~~ meets the criteria of having both *TPR* and *TNR* values greater than 0.5. ~~LPI3b therefore exhibits a The-bias towards positive liquefaction forecasts and so it confers no advantage over LPI3, towards the prediction of non-occurrence of liquefaction is still present although greatly reduced.~~ However, the improved capability of predicting the occurrence of liquefaction is offset by a reduction in the capability of predicting non-occurrence of liquefaction.

#### 4.2 Contingency table analysis – optimised thresholds

10 The results in Table 3 and Table 4 demonstrate the performance of each model with a single initial threshold value. ROC analysis is used to optimize the thresholds and curves for the eleven simplified models and reference model are generated using the ROC package in R (Sing et al., 2005), as shown in ~~Figure 5~~Figure-4. For this study, the threshold for the LPI models is assumed to be a whole number, while for the HAZ and ZHU models, the threshold is assumed to be a multiple of 0.05 subject to a minimum value of 0.1, which is the minimum ~~tested-applied~~ by the Zhu et al. (2014). The *AUC* values, maximum *J*-statistics, optimum thresholds and corresponding *TPR* and *TNR* values for all models are shown in ~~Table 6~~Table 7.

15 With optimized thresholds all the LPI models, except LPI2 and all the ZHU models meet the *TPR* and *TNR* criteria (>0.5). All HAZ models and both versions of LPI2 have *AUC* values closer to the ‘no value’ criterion, suggesting that the problems with these models lie not just with threshold selection, but more fundamentally with their composition and/or relevance to the case study ~~being-tested~~ (noting that the HAZ models have been developed for analysis in the United States). The reason 20 these are to the left of the chance line is because they are ~~predicting-forecasting~~ non-occurrence of liquefaction at nearly every site and hence they are guaranteed a low *FPR* value. LPI1 is the best performing model according to both of the ROC diagnostics and although the optimum threshold value of 7 is higher than proposed by Iwasaki et al. (1984), it is within the range for marginal liquefaction – 4 to 8 – proposed by Maurer et al. (2014) and so may be considered plausible. The two 25 versions of the LPI3 model perform similarly and have reasonable diagnostic scores but LPI3b, with ~~out~~ the correction factor, produces a more plausible optimum threshold value of 4. It is noted however that although the optimum threshold for LPI3b is 10, the *TPR* and *TNR* criteria are met with a threshold of 4 but with a lower model performance and greater ~~positiveneegative prediction-forecast~~ bias (*J*-statistic = 0.403344, *TPR* = 0.523806, *TNR* = 0.880538).

30 The ZHU1 and ZHU2 models perform reasonably with *AUC* values and *J*-statistics slightly lower than the LPI3 models, but the optimum thresholds are at the minimum of the ~~tested-range~~ that has been investigated, confirming the degree to which these models under-predict-estimate liquefaction occurrence. The ZHU2 model also meets the *TPR* and *TNR* criteria with a threshold value of 0.2, albeit with a greater ~~prediction-forecast~~ bias (*J*-statistic = 0.370, *TPR* = 0.555, *TNR* = 0.815). The ZHU3 model, despite being specific to Christchurch, does not perform as well as ZHU1 or ZHU2. ~~There are potential~~

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Heading 2

reasons for this anomaly, such as that may be because the ZHU models were calibrated to preserve the extent of liquefaction rather than to make site-specific predictions/forecasts or because the data used to develop the models has not come from the same source as the observation data used for comparison. Therefore, these results do not contradict or invalidate the original findings of Zhu et al. (2015).

The absolute quality of models is tested/evaluated by calculating  $MCC$ . In the preceding analysis, the best performing model is LPI1 and this has a value of  $MCC = 0.48$ . The correlation is only moderate, but nevertheless indicates that the model is better than random guessing. As part of a rapid assessment or desktop study for insurance purposes, this may be sufficient. LPI3 and LPI3b have  $MCC = 0.38039$  and  $0.35748$  respectively, whilst LP1ref has  $MCC = 0.29$ .

### 4.3 Mapping of model forecasts

The maps in Figure 6 and Figure 7 show how forecasts of liquefaction occurrence, relating to the Darfield and Christchurch earthquakes respectively, are distributed across the city for four of the best performing models identified in Table 6: LPI1, LPI3, ZHU1 and ZHU2. Figure 3 shows that a greater extent of liquefaction was observed in the Christchurch earthquake than in the Darfield earthquake and this is reflected by all four models represented in Figure 6 and Figure 7. However for both earthquakes, each of the models forecasts a greater extent of liquefaction than was observed. In the Darfield earthquake, most of the liquefaction was observed in the north and east of the city. Whilst to some degree, this spatial distribution is matched by model LPI1, the remaining models do not represent the observed distribution well, with in particular the models ZHU1 and ZHU2 estimate a greater proportion of liquefaction in the south of the city. In the Christchurch earthquake, liquefaction was mostly observed in the eastern suburbs of the city. All the models forecast the majority of liquefaction to occur in these areas, although model ZHU2 forecasts more liquefaction occurring in western suburbs than actually occurred, while model ZHU1 forecasts no liquefaction occurring to the west of the city at all. The spatial distributions of the forecasts from the LPI models exhibit only limited accuracy, yet they are better than the forecasts from the two ZHU models. This can be explained partly by the fact the LPI method is designed for site-specific estimation, whereas the ZHU models have been calibrated to optimise the extent rather than the location of liquefaction.

### 4.4 Sensitivity test – $V_{S30}$

The sensitivity of the forecasts to variation in  $V_{S30}$  is assessed for models LPI3 and ZHU1. LPI3 is the best performing model that requires  $V_{S30}$  and ZHU2 is the best performing ZHU model. The forecasting procedure and contingency table analysis for the two models are repeated for two scenarios, one where  $V_{S30}$  is decreased by 10% at all sites and one where  $V_{S30}$  is increased by 10% at all sites.

In the scenario where  $V_{S30}$  is decreased, the  $TPR$  for model LPI3 increases to 0.819 with a threshold of 4 (the optimised threshold from Table 6), while the  $TNR$  decreases to 0.536, effectively reversing the bias demonstrated by the original

Formatted: Heading 2

Formatted: Font: Italic

Formatted: Heading 2

Formatted: Subscript

Formatted: Font: Italic

Formatted: Font: Italic

model. The  $J$ -statistic reduces significantly to 0.356 indicating lower performance than the original model. With the new  $V_{S30}$  values, the optimised threshold increases to 9, with  $J$ -statistic = 0.426, which is higher than the original model,  $TPR = 0.654$  and  $TNR = 0.773$ . When  $V_{S30}$  is increased,  $TPR = 0.308$  with a threshold of 4, which is lower than criterion for good performance ( $TPR > 0.5$ ),  $TNR = 0.974$ . This demonstrates a strengthening of the negative bias in the original model and poor performance since the  $J$ -statistic reduces to 0.282. The optimum threshold changes to 1, yet even with this threshold, while the  $J$ -statistic improves to 0.388,  $TPR = 0.489$ , which is still below the performance criterion. These results show that LPI3 forecasts are sensitive to variation in  $V_{S30}$ . Therefore, although currently the optimum LPI3 threshold for Christchurch has been identified as 4, if in the future more accurate  $V_{S30}$  becomes available, then the analysis presented in this paper should be repeated to recalibrate model LPI3 with a new optimum threshold.

Model ZHU2 experiences much smaller changes as a result of changes to  $V_{S30}$ . When  $V_{S30}$  is decreased, and with a threshold of 0.1 (the optimised threshold from Table 6),  $TPR = 0.820$ ,  $TNR = 0.532$  and  $J$ -statistic = 0.352. When  $V_{S30}$  is increased,  $TPR = 0.700$ ,  $TNR = 0.662$  and  $J$ -statistic = 0.362. For both scenarios all performance criteria are met and there only small reductions in  $J$ -statistic. When the models are optimised, the thresholds change to 0.25 for the decrease ( $J$ -statistic = 0.370) and to 0.15 for the increase ( $J$ -statistic = 0.368). These results suggest that ZHU2 forecasts are relatively stable in response to variations in  $V_{S30}$ , but if more accurate  $V_{S30}$  becomes available in the future, then some performance improvement can be achieved through recalibration of the optimum threshold.

#### 4.5 Sensitivity test – PGA

The sensitivity of the forecasts to uncertainty in PGA measurements is also assessed for models LPI3 and ZHU1. In the two sensitivity test scenarios, The forecasting procedure and contingency table analysis are repeated for two scenarios, where PGA is decreased by 10% at all sites and where PGA is increased by 10% at all sites.

In the scenario where PGA is decreased, the  $TPR$  for model LPI3 decreases to 0.503 with a threshold of 4, while the  $TNR$  increases to 0.905 and there is only a small reduction in  $J$ -statistic to 0.408. The optimised threshold decreases to 2, with  $J$ -statistic = 0.424, which is higher than the original model,  $TPR = 0.594$  and  $TNR = 0.830$ . When PGA is increased,  $TPR = 0.652$  with a threshold of 4 and  $TNR = 0.765$ , with corresponding  $J$ -statistic = 0.417. The optimum threshold changes to 6, with  $J$ -statistic = 0.419,  $TPR = 0.576$  and  $TNR = 0.843$ . In general, changes in PGA do affect the scores but in all cases, the changes are relatively small, particularly with respect to the  $J$ -statistic, and the performance criteria are still met.

Model ZHU2 also experiences small changes as a result of changes to PGA. When PGA is decreased, and with a threshold of 0.1,  $TPR = 0.725$ ,  $TNR = 0.637$  and  $J$ -statistic = 0.362. When PGA is increased,  $TPR = 0.798$ ,  $TNR = 0.574$  and  $J$ -statistic = 0.372, which is a small increase over the original model. For both scenarios all performance criteria are met and there only small changes to  $J$ -statistic. When the models are optimised, the threshold changes to 0.2 for the decrease scenario ( $J$ -

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Subscript

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Subscript

Formatted: Heading 2

Formatted: Font: Italic

statistic = 0.369), but for the increase scenario, the optimum threshold is still 0.1. These results suggest that both LPI3 and ZHU2 forecasts are relatively stable in response to variations in PGA and so while small uncertainties in PGA measurements will change the rates of true positive and true negative forecasts, overall performance in terms of  $J$ -statistic remains similar.

## 5 Probability of liquefaction

5 | ~~When the~~ threshold-based approach to liquefaction occurrence ~~identification using is applied to~~ the LPI models, ~~it~~ provides a deterministic ~~prediction forecast~~. This may be considered sufficient for the simplified regional-scale analyses conducted for catastrophe modelling and loss estimation. However, a modeller may also want to establish a probabilistic view of liquefaction risk by relating values of  $LPI$  to probability of liquefaction occurrence. Since the occurrence of liquefaction at a site is a binary classification variable, it can be modelled by a Bernoulli distribution with probability of liquefaction,  $p$ ,  
10 | which depends on the value of  $LPI$ . With data from past earthquakes, functions relating  $p$  to  $LPI$  can be derived using a generalized linear model with probit link function. The probability of liquefaction occurring given a particular value of  $LPI$ ,  $\lambda$ , is given by:

$$p_{LPI=\lambda} = P[Liq | LPI = \lambda] = \Phi(Y^*) \quad (19)$$

15 | where  $\Phi$  is cumulative normal probability distribution function and  $Y^*$  is the probit link function given by:

$$Y^* = \beta_0 + \beta_1 \lambda \quad (20)$$

The link function is a linear model with  $LPI$  as a predictor variable and is derived from the individual site observations. ~~Figure 8~~ ~~Figure-5~~ displays the relationships between liquefaction probability and  $LPI$  fit by this method for the two best  
20 | performing LPI models, LPI1 and LPI3b, including 95% confidence intervals. The relationships are accompanied by plots of the observed liquefaction rates, aggregated at each value of  $LPI$ . The plot for model LPI3b shows greater scatter of observed rates around the fit line than the plot for model LPI1, although in both cases the confidence interval is very narrow, which is a reflection of the large sample size. The confidence interval for LPI1 ( $\pm 0.0014$ ) is slightly narrower than the confidence interval for LPI3b ( $\pm 0.0021$ ), indicating that LPI1 is the better ~~predictor model of for estimating~~ liquefaction probability, just  
25 | as it is better at ~~predicting forecasting~~ liquefaction occurrence by  $LPI$  threshold. ~~For both models, it is noted in both models though that~~ the observed rates that are furthest away from the ~~best-fit line~~ are ~~mostly-predominantly~~ those ~~that are~~ based on smaller sample sizes (arbitrarily defined here as 100), ~~and which therefore~~ ~~These~~ have less influence on the regression – since the use of individual site observations implicitly gives more weight to observations in the region of  $LPI$  values for which sample sizes are larger. Furthermore, the observed rates are themselves more unreliable for smaller sample sizes. For

example, for model LPI1, observations based on more than 100 samples have an average margin of error of 0.05 whereas the average margin of error for smaller samples is 0.19. For model LPI3b, observations based on more than 100 samples have an average margin of error of 0.05 and this increases to 0.2234 when considering the observations based on smaller samples.

5 The Hosmer-Lemeshow test (Hosmer and Lemeshow, 1980) is a commonly used procedure for assessing the goodness of fit of a generalized linear model when the outcome is a binary classification. However, Paul et al. (2013) show that the test is biased with respect to large sample sizes, with even small departures from the proposed model being classified as significant and consequently recommend that the test is not used for sample sizes above 25,000. Pseudo-R<sup>2</sup> metrics are also commonly used to test model performance (Smith and McKenna, 2013), but these compare the proposed model to a null intercept-only  
10 model rather than comparing the model ~~predictions-forecasts~~ to observations. Although the purpose of the analysis in this section is to relate *LPI* to liquefaction probabilistically, contingency table analysis with a threshold probability to determine liquefaction occurrence remains an appropriate technique to test the fit of the model (Steyerberg et al., 2010). Assuming a threshold probability of 0.5, ~~Table 7~~Table-8 presents summary statistics from the contingency table analysis of each model and also the coefficients of the corresponding probit link function.

15 Both models have values of *TPR* and *TNR* above 0.5 and the values are of a similar order to those obtained in Table 5 and ~~Table 6~~ for the same models. ~~The exception is the *TPR* for model LPI3, which is significantly higher when the threshold probability is used and eliminates much of the bias towards negative forecasts. but using a value of *LPI* as a threshold. In particular however, values of *TNR* are now higher, which indicates that the probabilistic *LPI* model is better at predicting non-occurrence of liquefaction. This is important as over 80% of observations in this study are negative.~~ The difference in values of *AUC* between ~~Table 6~~Table-7 and ~~Table 7~~Table-8 are negligible but the *J*-statistic for LPI3b with a threshold probability of 0.5 is considerably higher than the *J*-statistic ~~for that~~ the optimal threshold found for LPI3b in ~~Table 6~~Table-7. This suggests that LPI3b is best implemented as a probabilistic model for liquefaction occurrence. Overall these statistics indicate that both of the probabilistic *LPI* models proposed are good fits to the observed data.

## 25 **6 Permanent ground deformation**

The preceding sections have analysed methods for ~~predicting-forecasting~~ liquefaction triggering, but for assessing the fragility of structures and infrastructure, it is more informative to be able to ~~predict-estimate~~ the scale of liquefaction, in terms of the permanent ground deformation (PGDf). In fact, fragility functions for liquefaction-induced damage are commonly expressed in these terms (Pitilakis et al., 2014). A summary of the available approaches for quantifying PGDf is provided by Bird et al. (2006), which also compares approaches for lateral movement, settlement and combined movement  
30 (volumetric strain). The majority of these approaches require detailed geotechnical data as inputs, (e.g. median particle size, fines content). The likelihood that insurers possess or are able to acquire such data is low, which means that these approaches

are not suitable for regional-scale rapid assessment. The lack of simplified models is not surprising given the small number of models that exist for liquefaction triggering assessment and that by definition measuring the scale of liquefaction is more complex. From the available models in the literature, there are three that can be applied without the need for detailed geotechnical data: the EPOLLS regional model for lateral movement (Rauch and Martin, 2000) and the HAZUS models for lateral movement and vertical settlement (NIBS, 2003). To demonstrate the challenge faced by insurers looking to improve their liquefaction ~~predictionmodelling-~~ capability, these models are compared to PGDF observations from the Darfield and Christchurch earthquakes. It should be noted that the HAZUS model has been developed specifically for the United States and the empirical data used to develop its constituent parts comes mainly from California and Japan. The EPOLLS model is based on empirical data from the United States, Japan, Costa Rica and the Philippines.

### 6.1 Vertical settlement

A time series of LiDAR surface data for Christchurch has been produced from aerial surveys over the city, initially prior to the earthquake sequence in 2003, and subsequently repeated after the Darfield and Christchurch earthquakes. The surveys are obtained from the Canterbury Geotechnical Database (2012a). The LiDAR surveys recorded the surface elevation as a raster at 5m-cell resolution. The difference between the post-Darfield earthquake survey and the 2003 survey represents the vertical movement due to the Darfield earthquake. Similarly the difference between the post-Christchurch earthquake and the post-Darfield earthquake surveys represents the movement due to the Christchurch earthquake. In addition to liquefaction, elevation changes recorded by LiDAR can also be caused by tectonic ~~movementsuplift~~. Therefore, to evaluate the vertical movement due to liquefaction effects only,  $PGDf_v$ , the differences between LiDAR surveys have been corrected to remove the effect of the tectonic movement. Tectonic ~~uplift-movement~~ maps have been acquired from the Canterbury Geotechnical Database (2013d). The only simplified method for calculating vertical settlement is from HAZUS (NIBS, 2003) in which the settlement is the product of the probability of liquefaction, as in Eq. (10), and the expected settlement amplitude, which varies according to liquefaction susceptibility zone, as described in ~~Table 3Table 9~~.

The HAZUS model is applied with each of the three implementations used for ~~predicting-forecasting~~ liquefaction probability in the liquefaction triggering analysis. Summary statistics of the  $PGDf_v$  ~~predictions-estimates~~ from each implementation are presented in ~~Table 8Table 10~~. This shows that the HAZUS model significantly underestimates the scale of liquefaction, regardless of how liquefaction susceptibility zones are mapped between the Canterbury and HAZUS classifications. The residuals have a negative mean in each implementation indicating an underestimations bias. Furthermore, the maximum value ~~predicted-estimated~~ by HAZ1 and HAZ3 is smaller than the observed lower quartile. The coefficient of determination is also extremely low in each case, implying that there is little or no value in the ~~predictionsestimates~~. It is important to note that there is a measurement error in the LiDAR data itself of up to 150mm, as well as a uniform probability prediction interval around the HAZUS ~~predictionestimates~~. However, even when using the upper bound of the HAZUS ~~prediction~~

Field Code Changed

estimates (two times the mean), only around 50% of ~~predictions-estimates~~ fall within the observation error range. These results suggest that the HAZUS model for ~~predicting-estimating~~ vertical settlement is not suitable for application in Christchurch.

## 6.2 Lateral spread

5 The LiDAR surveys for Christchurch also record the locations of reference points within a horizontal plane and the differences between these data have been used to generate maps identifying the lateral displacements caused by each earthquake on a grid of points at 56m intervals. Similarly to the elevation data, the lateral displacements have to be corrected for tectonic movements, although in this case the corrected maps have been obtained directly from the Canterbury Geotechnical Database (2012b).

10

The HAZUS model (NIBS 2003) for ~~predicting-estimating~~ ground deformation due to lateral spread is given by:

$$PGDf_H = K_{\Delta} \times E[PGD | (PGA / PL_{SC}) = a] \quad (21)$$

15 where,  $K_{\Delta}$  is a displacement correction factor, which is a cubic function of earthquake magnitude, and the term on the right hand side is the expected ground deformation for a given liquefaction susceptibility zone, which is a function of the normalized peak ground acceleration (observed PGA divided by liquefaction triggering threshold PGA for that zone). The formulae for calculating these terms are not repeated here but can be found in the HAZUS manual (NIBS, 2003).

20 The EPOLLS suite of models for lateral spread (Rauch and Martin, 2000) includes proposed relationships for ~~predicting-estimating~~ ground deformation at a regional scale (least complex), at site specific scale without detailed geotechnical data and at site specific scale with detailed geotechnical data (most complex). In the regional EPOLLS model,  $PGDf_H$  is given by:

$$PGDf_H = (0.613M_w - 0.0139R_f - 2.42PGA - 0.01147T_d - 2.21)^2 + 0.149 \quad (22)$$

25 where,  $R_f$  is the shortest horizontal distance to the surface projection of the fault rupture, and  $T_d$  is the duration of ground motion between the first and last occurrence of accelerations  $\geq 0.05g$  at each site. ~~To calculate duration Durations have been calculated from ground motion records (at 0.02s intervals) obtained, there are from~~ 19 strong-motion accelerograph stations in Christchurch, ~~identified in Figure 1 that record ground motions at 0.02s intervals~~. The records from each station for both earthquakes are available from the GeoNet website (GNS Science, 2014).  $T_d$  is calculated at each station and then the value at intermediate sites is interpolated by ordinary kriging. Summary statistics of the ~~predictions-estimates~~ from the regional

EPOLLS and HAZUS models are presented in [Table 9Table 11](#). The statistics show that none of the models ~~predict-estimate~~  $PGDf_H$  well. The EPOLLS model overestimates the scale of liquefaction, while the HAZUS models each show an underestimation bias. The mean residuals and root mean square error (RMSE) are higher for the EPOLLS model, suggesting that the HAZUS models perform slightly better, but this is of little significance since the coefficients of determination of the HAZUS models are all extremely low. A mitigating factor is that the LiDAR data ~~has~~ a very large error – up to 0.5m – in the horizontal plane. Taking this into account, over 90% of HAZUS ~~predictions-estimates~~ are within the observation error range, although this needs to be interpreted in the context of the mean observed  $PGDf_H$  being 0.269m. Since the HAZUS model underestimates  $PGDf_H$ , and  $PGDf_H$  cannot be negative, the fact that so many ~~predictions-estimates~~ are within this error range is more a reflection of the size of the error relative to the values being observed. Consequently the statistics in [Table 9Table 11](#) are more informative and these show that the simplified models all perform poorly.

## 7 Conclusions

This study compares a range of simplified desktop liquefaction assessment methods that may be suitable for insurance sector where data availability and resources are key constraints. It finds that the liquefaction potential index, when calculated using shear-wave velocity profiles (LPI1) is the best performing model in terms of its ability to correctly ~~predict-forecast~~ liquefaction occurrence both positively and negatively, ~~although it must be noted that its predictive power is not high~~. Shear-wave velocity profiles are not always available to practitioners and it is notable therefore that the analysis shows that the next best performing model is the liquefaction potential index calculated with shear-wave velocity profiles simulated from USGS  $V_{330}$  data (LPI3b). Since it is based on USGS data, which ~~are~~ publicly accessible online, this method is particularly attractive to those undertaking rapid and/or regional scale desktop assessments.

The HAZUS ~~methodology-method~~ for estimating liquefaction probabilities performs poorly irrespective of triggering threshold. This is significant since HAZUS ~~methodologies-methods~~ (not only in respect to liquefaction) are often used as a default model outside of the US when no specific local (or regional) model is available. Models proposed by Zhu et al. (2015) perform ~~well-reasonably~~ and since they are also based on publicly accessible data, represent another viable option for desktop assessment. The only issue with these models is that they perform optimally with a low threshold probability of 0.1, which may lead to over-~~prediction-estimation~~ of liquefaction when applied to other locations.

As an extension of the liquefaction triggering analysis, this study also uses the observations to relate  $LPI$  to liquefaction probability for the two best-performing models. In the case of LPI3b, the model performance (as measured by Youden's  $J$ -statistic) actually improves significantly when employed with a threshold based on corresponding probability rather than based directly on  $LPI$ . The final stage of liquefaction assessment is to measure the scale of liquefaction as  $PGDf$ . This study only briefly considers this aspect but shows that existing simplified models perform extremely poorly. Existing models show

very low correlation with observations and strong ~~prediction~~ bias – underestimation in the case of HAZUS and overestimation in the case of regional EPOLLS. Based on this analysis the ~~predictions-estimations~~ from these simplified models are highly uncertain and it is questionable whether they genuinely add any value to loss estimation analysis outside of the regions for which they have been developed.

## 5 Competing interests

The authors declare that they have no conflict of interest.

## Acknowledgments

The authors are grateful to Sjoerd van Ballegooy of Tonkin and Taylor for providing data on observed land damage categories and liquefaction metrics. Funding for this research project has been provided by the UK Engineering and Physical Sciences Research Council and the Willis Research Network, through the Urban Sustainability and Resilience Doctoral Training School at University College London.

## References

- Adachi, T. and Ellingwood, B. R., (2008), Serviceability of earthquake-damaged water systems: effects of electrical power availability and power backup systems on system vulnerability, *Reliab. Eng. Syst. Safe.*, 93, 78-88.
- 15 Andrus, R. D. and Stokoe, K. H., (2000), Liquefaction resistance of soils from shear-wave velocity, *J. Geotech. Geoenviron.*, 126(11), 1015-1025.
- [Beaven, J., Motagh, M., Fielding, E. J., Donnelly, N. and Collett, D., \(2012\), Fault slip models of the 2010-2011 Canterbury, New Zealand, earthquakes from geodetic data and observations of postseismic ground deformation, \*New Zealand Journal of Geology and Geophysics\*, 55\(3\).](#)
- 20 Bird, J. F. and Bommer, J. J., (2004), Earthquake losses due to ground failure, *Eng. Geol.*, 75, 147-179.
- Bird, J. F., Bommer, J. J., Crowley, H. and Pinho, R., (2006), Modelling liquefaction-induced building damage in earthquake loss estimation, *Soil Dyn. Earthq. Eng.*, 26, 15-30.
- Boore, D. M., (2004), Estimating  $V_{s30}$  (or NEHRP site classes) from shallow velocity models (depths < 30m), *B. Seismol. Soc. Am.*, 94(2), 591-597.
- 25 Canterbury Geotechnical Database. (2012a) Vertical Ground Surface Movements, Map Layer CGD0600 – 23 July 2012. <https://canterburygeotechnicaldatabase.projectorbit.com/>. Accessed 9 October 2014.

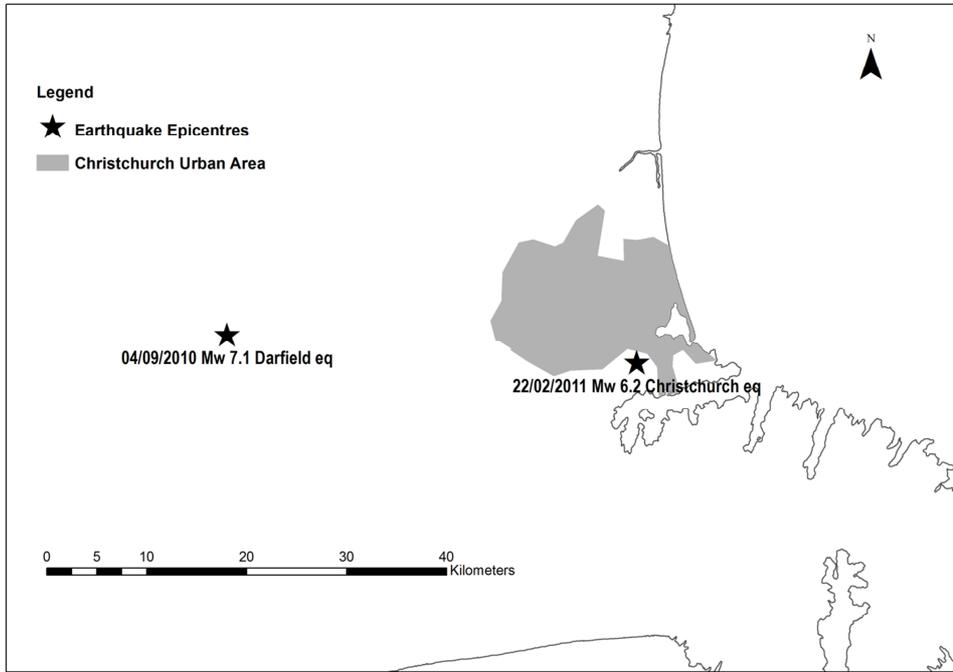
- Canterbury Geotechnical Database. (2012b) Horizontal Ground Surface Movements, Map Layer CGD0700 – 23 July 2012. <https://canterburygeotechnicaldatabase.projectorbit.com/>. Accessed 9 October 2014.
- Canterbury Geotechnical Database. (2013a) Liquefaction and Lateral Spreading Observations, Map Layer CGD0300. <https://canterburygeotechnicaldatabase.projectorbit.com/>. Accessed 20 September 2014.
- 5 Canterbury Geotechnical Database. (2013b) Conditional PGA for Liquefaction Assessment, Map Layer CGD5110. <https://canterburygeotechnicaldatabase.projectorbit.com/>. Accessed 21 February 2014.
- Canterbury Geotechnical Database. (2013c) Geotechnical Investigation Data, Map Layer CGD0010. <https://canterburygeotechnicaldatabase.projectorbit.com/>. Accessed 20 September 2014.
- Canterbury Geotechnical Database. (2013d) LiDAR and Digital Elevation Models, Map Layer CGD0500 – 03 August 2013.
- 10 <https://canterburygeotechnicaldatabase.projectorbit.com/>. Accessed 9 October 2014.
- Drayton MJ, Verdon CL (2013) Consequences of the Canterbury earthquake sequence for insurance loss modelling. Proc. 2013 NZSEE Conf., NZSEE, Wellington, NZ.
- ECan (2014) OpenData Portal. Canterbury Maps. <http://opendata.canterburymaps.govt.nz/>. Accessed 20 September 2014.
- Giovinazzi, S., Wilson, T., Davis, C., Bristow, D., Gallagher, M., Schofield, A., Villemure, M., Eiding, J., Tang, A.,  
15 (2011), Lifelines performance and management following the 22<sup>nd</sup> February 2011 Christchurch earthquake, New Zealand: highlights of resilience, Bull. NZ. Soc. Earthq. Eng., 44(4), 402-417.
- GNS Science (2014) Strong-Motion Data. [http://info.geonet.org.nz/display/appdata/Strong-Motion+Data\\_](http://info.geonet.org.nz/display/appdata/Strong-Motion+Data_) Accessed 15 November 2014.
- Goda, K., Atkinson, G. M., Hunter, A. J., Crow, H., Motazedian, D., (2011), Probabilistic liquefaction hazard analysis for  
20 four Canadian cities, Bull. Seismol. Soc. Am., 101(1), 190-201.
- Hosmer, D. W. and Lemeshow, S., (1980), A goodness-of-fit test for multiple logistic regression model, Commun. Stat., A10, 1043-1069.
- Hwang, H. H. M., Lin, H. and Shinozuka, M., (1998), Seismic performance assessment of water delivery systems, J. Infrastruct. Syst., 4(3), 118-125.
- 25 Iwasaki, T., Arakawa, T. and Tokida, K.-I., (1984), Simplified procedures for assessing soil liquefaction during earthquakes, Soil Dyn. Earthq. Eng., 3(1),49-58.
- Juang, C. H., Yang, S. H. and Yuan, H., (2005), Model uncertainty of shear wave velocity-based method for liquefaction potential evaluation, J. Geotech. Geoenviron., 131(10), 1274-1282.
- [Lees, J. J., Ballagh, R. H., Orense, R. P. and van Ballegooy, S., \(2015\), CPT-based analysis of liquefaction and re-liquefaction following the Canterbury earthquake sequence, Soil Dyn. Earthq. Eng., 79, 304-314.](#)
- 30 Matthews, B. W., (1975), Comparison of the predicted and observed secondary structure of T4 phage lysozyme, Biochimica et Biophysica Acta – Protein Structure, 405(2), 442-451.
- Maurer, B. W., Green, R. A., Cubrinovski, M. and Bradley, B. A., (2014), Evaluation of the liquefaction potential index for assessing liquefaction potential in Christchurch, New Zealand. J. Geotech. Geoenviron., 140(7), 04014032.

- National Institute of Building Sciences (NIBS) (2003) HAZUS<sup>®MH</sup> Technical Manual. NIBS, Washington, D.C.
- Orense, R. P., Pender, M. J. and Wotherspoon, L. M., (2012), Analysis of soil liquefaction during the recent Canterbury (New Zealand) earthquakes. Geotechnical Engineering Journal of the SEAGS and AGSSEA, 43(2), 8-17.
- Paul, P., Pennell, M. L. and Lemeshow, S., (2013), Standardizing the power of the Hosmer-Lemeshow goodness of fit tests in larger datasets, Stat. Med., 32, 67-80.
- Pitilakis, K., Crowley, H. and Kaynia, A. (eds), (2014), SYNER-G: Typology Definition and Fragility Functions for Physical Elements at Seismic Risk , Vol. 27 Geotechnical, Geological and Earthquake Engineering. Springer, Netherlands.
- Powers, D. M. W., (2011), Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, J. of Mach. Lear. Tech, 2(1), 37-63.
- Quigley, M., Bastin, S. and Bradley, B. A., (2013), Recurrent liquefaction in Christchurch, New Zealand, during the Canterbury earthquake sequence, Geology, 41(4), 419-422.
- Rauch, A. F. and Martin, J. R., (2000), EPOLLS model for predicting average displacements on lateral spreads, J. Geotech. Geoenviron. Eng., 126(4), 360-371.
- Seed, H. B. and Idriss, I. M., (1971), Simplified procedure for evaluating soil liquefaction potential, J. Geotech. Engrg. Div., 97(9), 1249-1273.
- Sing, T., Sander, O., Beerenwinkel, N. and Lengauer, T., (2005), ROCr: visualizing classifier performance in R, Bioinformatics, 21(20), 3940-3941.
- Smith, T. J. and McKenna, C. M., (2013), A comparison of logistic regression pseudo-R<sup>2</sup> indices, Multiple Linear Regression Viewpoints, 39(2), 17-26.
- Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J. and Kattan, M. W., (2010), Assessing the performance of prediction models: a framework for some traditional and novel measures, Epidemiology, 21(1), 128-138.
- USGS (2013) Global V<sub>s</sub><sup>30</sup> map server, Earthquake Hazards Program. <http://earthquake.usgs.gov/hazards/apps/vs30/>. Accessed 10 November 2013.
- USGS (2014a) ShakeMaps, Earthquake Hazards Program. <http://earthquake.usgs.gov/earthquakes/shakemap/>. Accessed 13 March 2014.
- USGS. (2014b) Earth Explorer. <http://earthexplorer.usgs.gov/>. Accessed 13 October 2014.
- 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., (2014), Assessment of liquefaction-induced land damage for residential Christchurch, Earthq. Spectra, 30(1), 31-55.
- Wood, C., Cox, B. R., Wotherspoon, L. and Green, R. A., (2011), Dynamic site characterization of Christchurch strong motion stations, B. NZ Soc. Earthq. Eng, 44(4), 195-204.

Wotherspoon, L., Orense, R. P., Green, R., Bradley, B. A., Cox, B. and Wood, C., (2014), Analysis of liquefaction characteristics at Christchurch strong motion stations. In: Orense, R. P., Towhata, I. and Chouw, N. (eds), Soil Liquefaction During Recent Large-Scale Earthquakes. Taylor and Francis, London, pp 33-43.

5 Zhou, Y.-G. and Chen, Y.-M., (2007), Laboratory investigation on assessing liquefaction resistance of sandy soils by shear wave velocity, *J. Geotech. Geoenviron.*, 133, 959-972.

Zhu, J., Daley, D., Baise, L. G., Thompson, E. M., Wald, D. J. and Knudsen, K. L., (2015), A geospatial liquefaction model for rapid response and loss estimation, *Earthq. Spectra*, 31(3), 1813-1837.



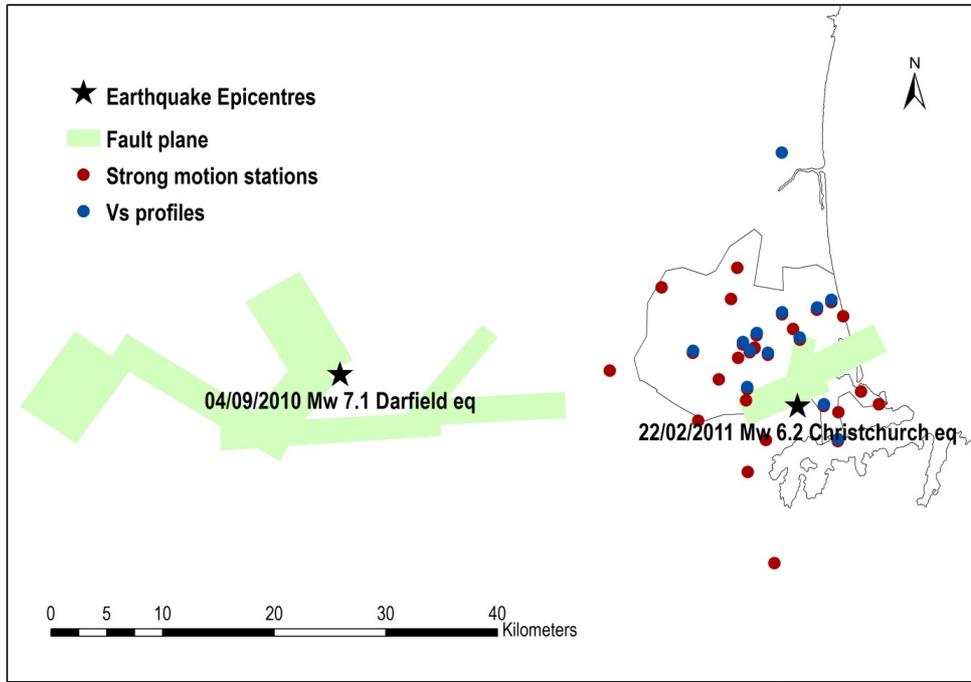


Figure 1 – Locations of epicentres and fault planes (Beaven et al., 2012) of the Darfield and Christchurch earthquakes, strong motion stations from which recordings are used to estimate shaking durations and locations at which shear wave velocity ( $V_s$ ) profiles are known (Wood et al., 2011). Note that locations of  $V_s$  profiles coincide with strong motion stations, in relation to the Christchurch urban area and central business district.

**Formatted:** Font: Italic

**Formatted:** Subscript

**Formatted:** Font: Italic

**Formatted:** Subscript

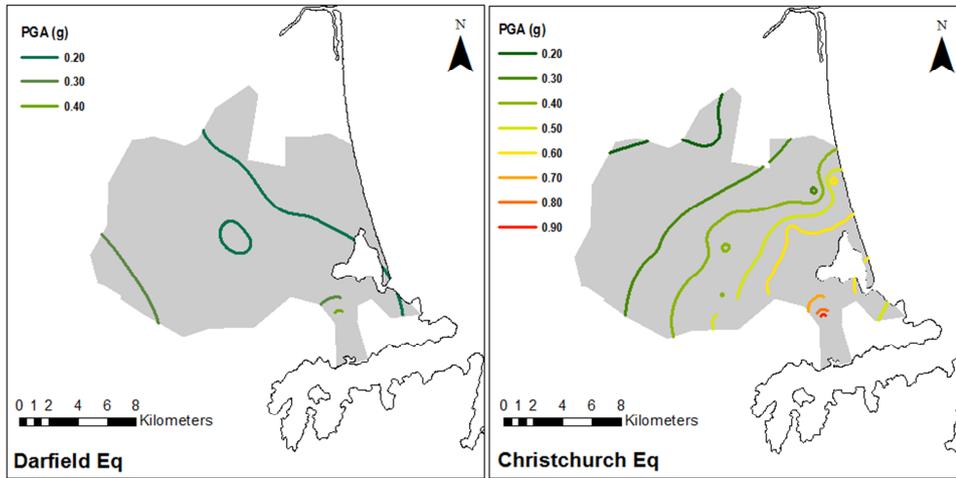


Figure 2 – Contours of peak horizontal ground acceleration (PGA) for the Darfield and Christchurch earthquakes (source: Canterbury Geotechnical Database, 2013b).

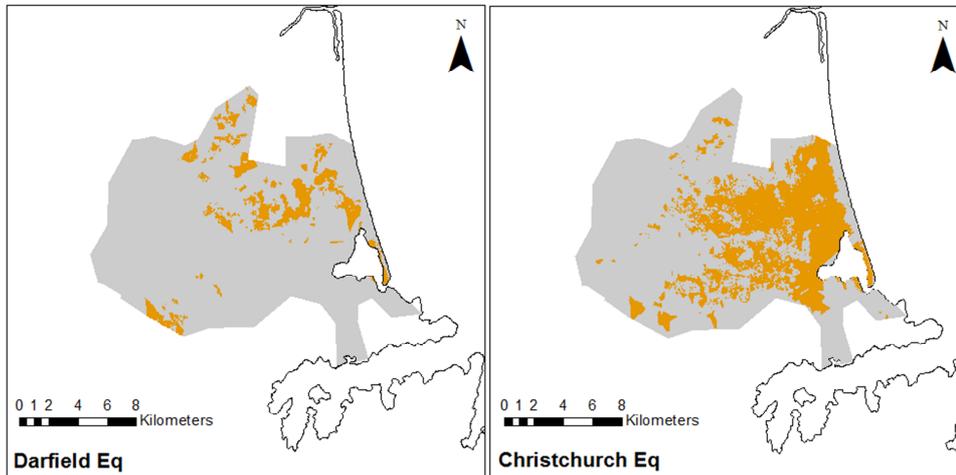


Figure 3 – Location of surface liquefaction observations (brown) in Christchurch and surrounding areas due to the Darfield and Christchurch earthquakes, based on data provided by Tonkin and Taylor and published within the Canterbury Geotechnical Database (CGD 2013a).

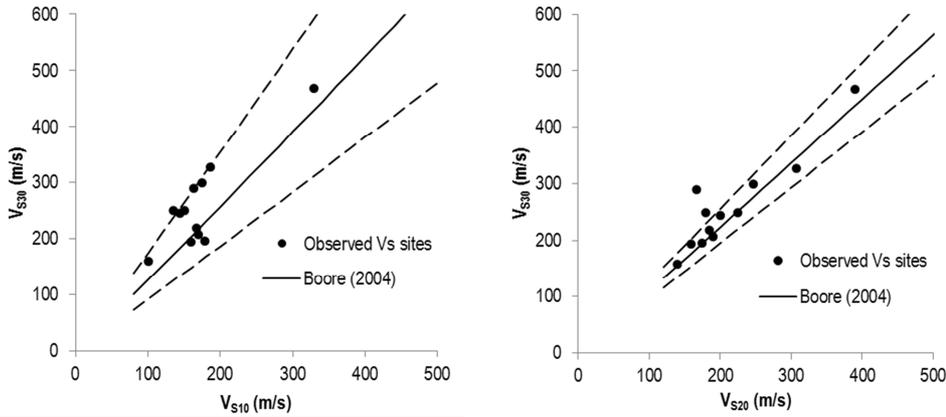


Figure 4 – Plots comparing observed  $V_{S30}$  with  $V_{S30}$  estimated from Boore (2004) equations, with respect to observed  $V_{S10}$  (left) and observed  $V_{S20}$  (right). The dashed lines represent the 95% confidence interval around the Boore (2004) relationships.  $V_{S30}$  is the average shear wave velocity in the top 30m of ground and  $V_{S10}$  and  $V_{S20}$  are the equivalents at 10m and 20m depth respectively.

Formatted: Font: Italic

Formatted: Subscript

Formatted: Normal, Line spacing: single

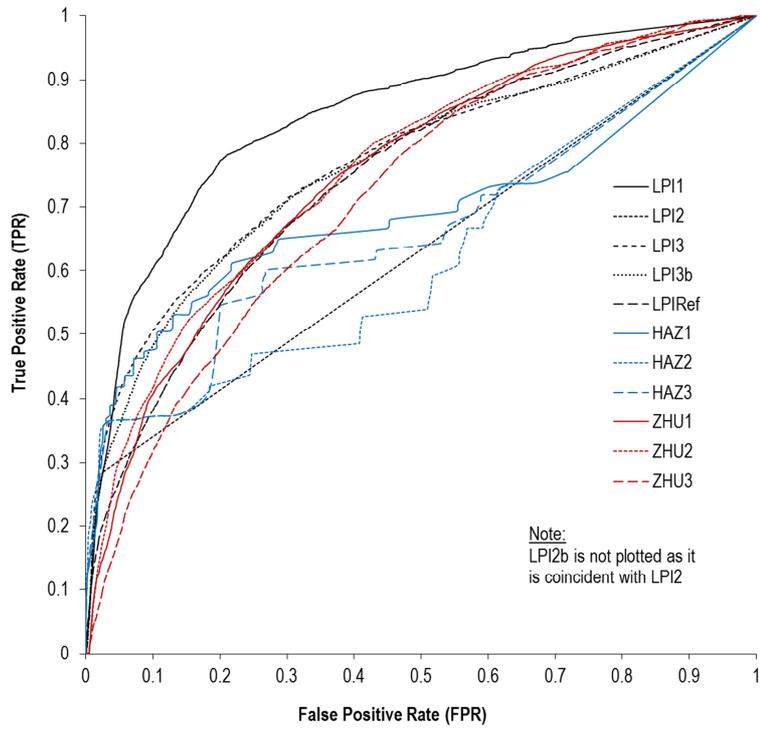
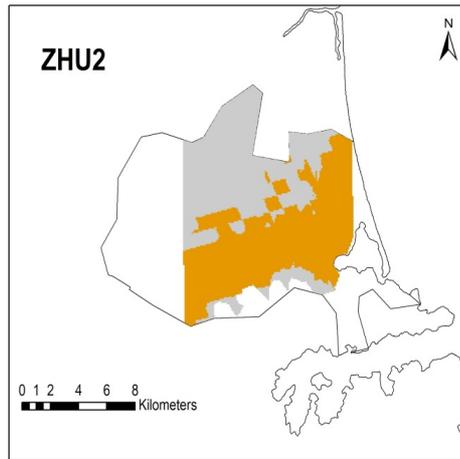
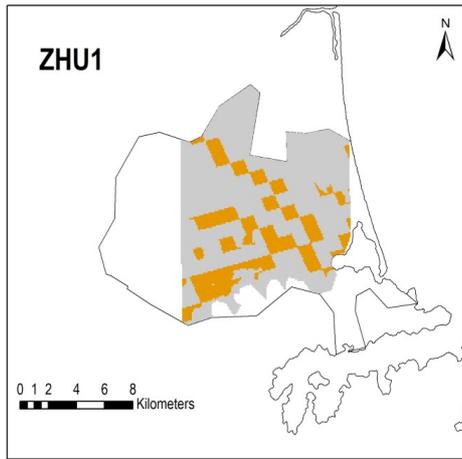
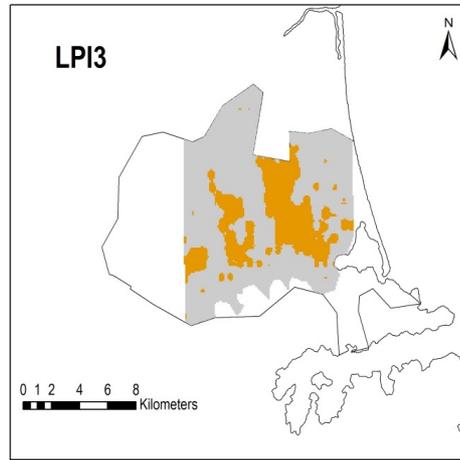
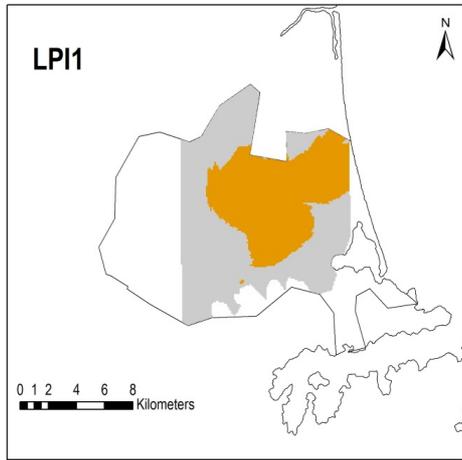


Figure 5 – Receiver Operating Characteristic (ROC) curves for prediction-forecasting models. Refer to Table 4 for descriptions corresponding to model acronyms.



**Figure 6 – Maps of liquefaction forecasts from selected models for the Darfield earthquake, where the brown areas represent positive liquefaction forecasts, the grey areas represent negative liquefaction forecasts and the white areas are where no forecast was made due to lack of input data. Refer to Table 4 for descriptions corresponding to model acronyms.**

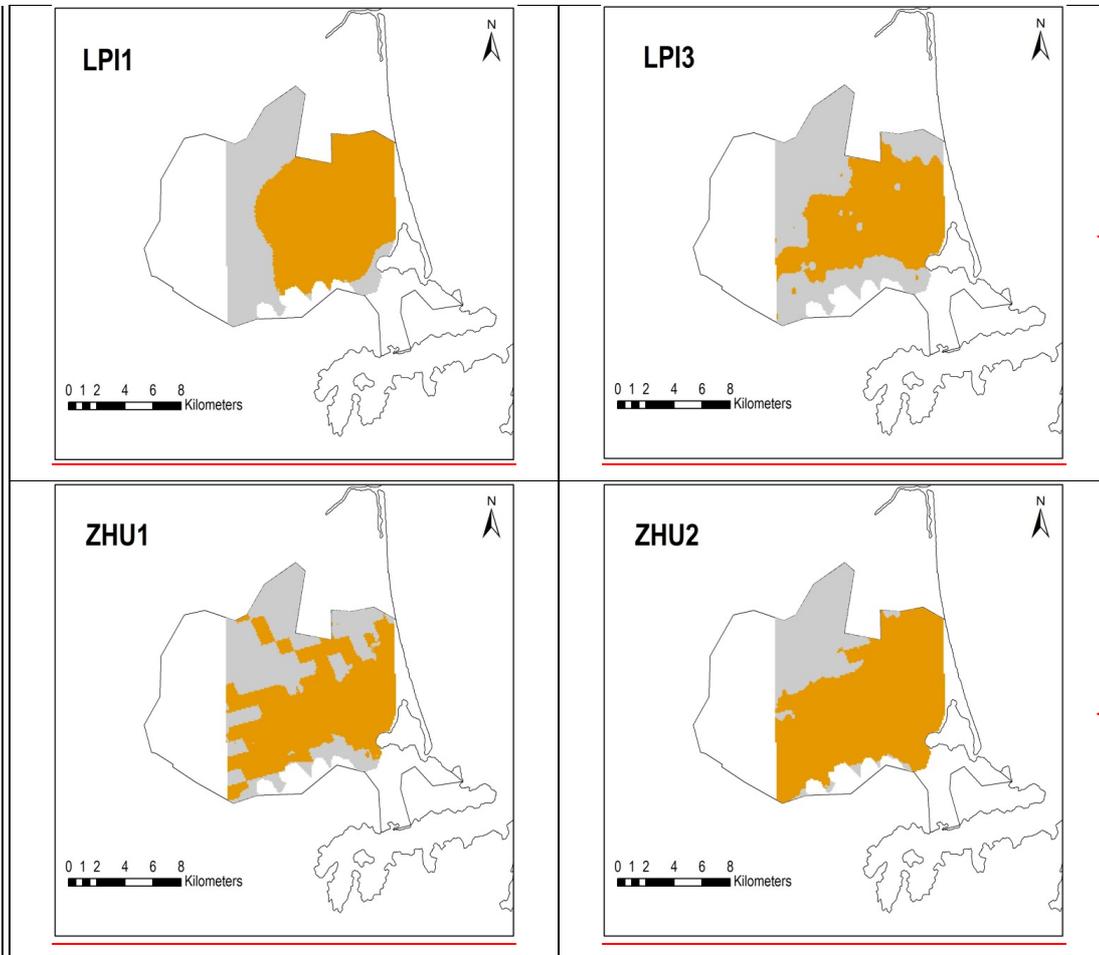
**Formatted:** Centered

**Formatted Table**

**Formatted:** Centered

**Formatted:** Centered, Keep with next

**Formatted:** Caption, Line spacing:  
1.5 lines



**Figure 7 – Maps of liquefaction forecasts from selected models for the Christchurch earthquake, where the brown areas represent positive liquefaction forecasts, the grey areas represent negative liquefaction forecasts and the white areas are where no forecast was made due to lack of input data. Refer to Table 4 for descriptions corresponding to model acronyms.**

Formatted: Centered

Formatted: Centered

Formatted: Centered, Keep with next

Formatted: Caption

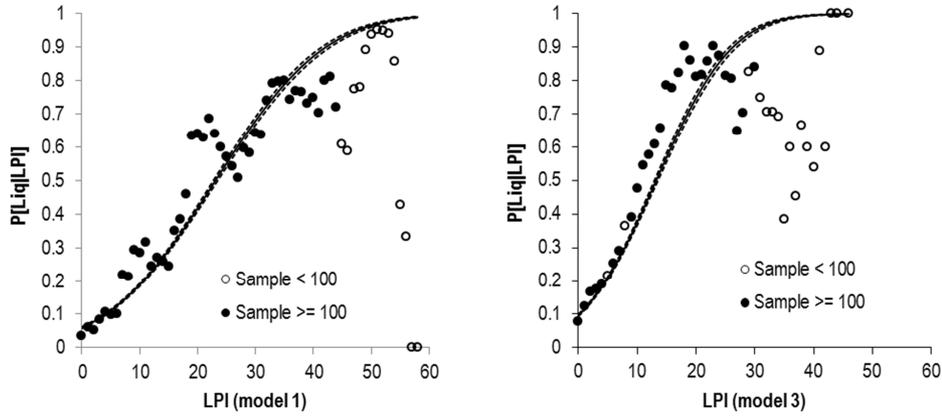


Figure 8 – Plots of liquefaction probability against **Liquefaction Potential Index (LPI)** derived from site specific observations by generalized linear model with probit link function for two best performing LPI models. Plots also display the observed liquefaction rates at each LPI value and classified by sample size

Formatted: Normal, Line spacing: single

**Table 1 – Reference list of acronyms used in this paper**

Acronym	Description
AOC	Area under ROC curve
CGD	Canterbury Geotechnical Database
CPT	Cone penetration test
CRR	Cyclic resistance ratio
CSR	Cyclic stress ratio
CTI	<del>Compound</del> Compound topographic index
EQC	Earthquake Commission
FN	False negative model <del>predictions-forecasts</del> (no.)
FP	False positive model <del>predictions-forecasts</del> (no.)
FPR	False positive rate ( = FP / Observed negatives)
FS	Factor of safety against liquefaction
LPI	Liquefaction potential index
MCC	Matthew’s correlation coefficient
MSF	Magnitude scaling factor
MWF	Magnitude weighting factor
NEHRP	National Earthquake Hazards Reduction Program
PGA	Peak ground acceleration
PGDf	Permanent ground deformation
PGDf <sub>H</sub>	Horizontal permanent ground deformation
PGDf <sub>V</sub>	Vertical permanent ground deformation
RMSE	Root mean square error
ROC	Receiver operating characteristic
SPT	Standard penetration test
TN	True negative model <del>predictions-forecasts</del> (no.)
TNR	True negative rate ( = TN / Observed negatives)
TP	True positive model <del>predictions-forecasts</del> (no.)
TPR	True positive rate ( = TP / Observed positives)
USGS	United States Geological Survey

**Table 2 – Reference list of variables used in this paper**

Variable	Description	Units	Input variables	
$a_{max}$	Peak horizontal ground acceleration	$m/s^2$	=	Formatted Table
$CRR$	Cyclic resistance ratio	-	$MSF, V_{S1}, V_{S1}^*$	Formatted: Left
$CSR$	Cyclic stress ratio	-	$a_{max}, \sigma_v, \sigma_v', L_d$	Formatted: Left
$CTI$	<del>Compound</del> topographic index	-	=	Formatted: Left
$d_w$	<u>Depth to groundwater</u>	<u>m</u>	=	Formatted: Font: Not Italic, Subscript
$FS$	Factor of safety against liquefaction	-	$CRR, CSR$	Formatted: Left
$K_M$	HAZUS moment magnitude correction factor	-	$M_w$	Formatted: Left
$K_W$	HAZUS ground water correction factor	-	$d_w$	Formatted: Font: Italic
$K_d$	<del>Displaedement</del> <u>Displacement</u> correction factor	-	$M_w$	Formatted: Left
$LPI$	Liquefaction potential index	-	$FS, z$	Formatted: Subscript
$MSF$	Magnitude scaling factor	-	$M_w$	Formatted: Left
$M_w$	Moment magnitude	-	=	Formatted: Left
$MWF$	Magnitude weighting factor	-	$M_w$	Formatted: Font: Italic
$ND$	Normalised distance to coast (Zhu et al., 2015)	-	=	Formatted: Left
$PGA$	Peak horizontal ground acceleration (non USGS)	g	=	Formatted: Left
$PGA_{M,SM}$	Peak horizontal ground acceleration (from USGS ShakeMap)	g	=	Formatted: Left
$PGD   (PGA / PLSC)$	HAZUS expected $PGD_f$ for a given <u>liquefeation</u> <u>liquefaction</u> susceptibility zone	m	$PGA, liquefaction susceptibility$	Formatted: Left
$PGD_f$	Permanent ground deformation	m	=	Formatted: Left
$PGD_{fH}$	Horizontal permanent ground deformation	m	=	Formatted: Left
$PGD_{fV}$	Vertical permanent ground deformation	m	=	Formatted: Left
$P_{ml}$	HAZUS proportion of map unit susceptible <del>to</del> <u>liquefaction</u> <u>for a given liquefaction</u> <u>susceptibility zone</u>	-	$Liquefaction susceptibility$	Formatted: Left
$r_d$	Shear stress reduction coefficient	-	$z$	Formatted: Left
$R_f$	Horizontal distance to surface projection of fault	km	=	Formatted: Font: Italic
				Formatted: Left

	rupture				
$T_d$	Duration between first and last occurrence of $PGA \geq 0.05g$	s	=		Formatted: Left
$V_S$	Shear wave <del>velocity</del> <u>velocity</u>	m/s	=		Formatted: Left
$V_{S1}$	Stress-corrected shear wave velocity	m/s	=	$V_{S1} \sigma_v'$	Formatted: Left
$V_{S1}^*$	Limiting upper value of $V_{S1}$ for cyclic liquefaction occurrence	m/s	=	Fines content	Formatted: Left
$V_{S(0-10)}$	Average shear wave velocity in top 10m	m/s	=		Formatted: Left
$V_{S(10-20)}$	Average shear wave velocity between 10m and 20m	m/s	=		Formatted: Left
$V_{S30}$	Average shear wave velocity in top 30m	m/s	=		Formatted: Left
$z$	Depth	m	=		Formatted: Left
$\sigma_v$	Total overburden stress	kPa	=	$z \cdot \text{soil density}$	Formatted: Left
$\sigma_v'$	Effective overburden stress	kPa	=	$\sigma_v \cdot z$	Formatted: Left

**Table 3 – Conversion between Canterbury and HAZUS liquefaction susceptibility zones for three implementations of HAZUS methodology. Refer to Table 4 for descriptions corresponding to model acronyms.**

Canterbury liquefaction susceptibility zones	Equivalent HAZUS liquefaction susceptibility zone and expected settlement amplitude					
	Model HAZ1		Model HAZ2		Model HAZ3	
	Zone	Expected settlement (cm)	Zone	Expected settlement (cm)	Zone	Expected settlement (cm)
None	None	<u>0</u>	None	<u>0</u>	None	<u>0</u>
Low	Low	<u>2.5</u>	Very low	<u>0</u>	Average low and very low	<u>1.25</u>
Moderate	Moderate	<u>5</u>	Moderate	<u>5</u>	Moderate	<u>5</u>
High	High	<u>15</u>	Very high	<u>30</u>	Average high and very high	<u>22.5</u>

5

**Table 4 – Liquefaction ~~prediction-forecasting~~ models tested-compared in this paper**

Model	Description
LPI1	<u>Liquefaction potential index (LPI)</u> with known <u>shear wave velocity</u> , $V_{s,z}$ profiles
LPI2	LPI with <u>average shear wave velocity in the top 30m</u> , $V_{s30z}$ as <u>a proxy for <del><math>V_s</math></del>-proxy</u>
LPI3	LPI with simulated $V_s$ profiles
LPIref	LPI calculated from <u>standard penetration test (SPT)</u> results
HAZ1	HAZUS with ‘direct’ conversion of susceptibility zones
HAZ2	HAZUS with ‘extreme’ susceptibility zones
HAZ3	HAZUS with ‘average’ conversion of susceptibility
ZHU1	Global model by Zhu et al. (2015)
ZHU2	Regional model by Zhu et al. (2015)
ZHU3	Local model by Zhu et al. (2015)

Formatted Table



**Table 6 – Model quality diagnostics and optimum threshold values for each model from ROC curves. Refer to Table 4 for descriptions corresponding to model acronyms.**

Model	<u>Area Under Curve (AUC)</u>	<u>J-statistic</u>	<u>Threshold</u>	<u>True Positive Rate (TPR)</u>	<u>True Negative Rate (TNR)</u>
LPI1	0.845	0.573	7	0.774	0.799
LPI2	0.630	0.122	1	0.131	0.991
LPI2b	0.630	0.206	1	0.224	0.982
LPI3	0.76272	0.41420	4	0.671581	0.742839
LPI3b	0.76466	0.4154	410	0.64617	0.76997
LPIref	0.748	0.366	6	0.689	0.678
HAZ1	0.679	0.238	0.1	0.073	0.999
HAZ2	0.608	0.316	0.1	0.134	0.997
HAZ3	0.661	0.315	0.1	0.133	0.998
ZHU1	0.753	0.355	0.1	0.556	0.799
ZHU2	0.760	0.371	0.1	0.767	0.604
ZHU3	0.718	0.306	0.1	0.712	0.594

- Formatted: Font: Not Italic
- Formatted: Font: Not Italic
- Formatted: Centered
- Formatted Table
- Formatted: Font: Not Italic
- Formatted: Font: Not Italic
- Formatted: Centered

5 **Table 7 – Coefficients of link function and summary of contingency table analysis for the two best performing LPI models. Refer to Table 4 for descriptions corresponding to model acronyms.**

Model	$\beta_1$	$B_0$	<u>True Positive Rate (TPR)</u>	<u>True Negative Rate (TNR)</u>	<u>J-statistic</u>	<u>Area Under Curve (AUC)</u>
LPI1	0.067	-1.555	0.683	0.869	0.551	0.843
LPI3b	0.08298	-	0.70484	0.86056	0.564641	0.7606
		1.376299				

- Formatted: Font: Not Italic
- Formatted: Font: Not Italic
- Formatted: Centered
- Formatted Table
- Formatted: Font: Not Italic
- Formatted: Font: Not Italic
- Formatted: Font: Not Italic
- Formatted: Centered
- Formatted: Centered

**Table 9 – Expected settlement amplitudes for liquefaction susceptibility zones from HAZUS methodology (NIBS, 2003)**

Liquefaction susceptibility zone	Expected settlement amplitude (inches)
Very high	12
High	6

Moderate	2
Low	1
Very low	0
None	0

**Table 8** – Summary statistics of vertical ground deformation (PGD<sub>v</sub>) predictions-estimates for Darfield and Christchurch earthquakes from HAZUS models. Refer to Table 4 for descriptions corresponding to model acronyms.

Score	Observed	<u>Vertical permanent ground deformation, PGD<sub>v</sub> (m)</u>		
		HAZ1	HAZ2	HAZ3
Pearson R <sup>2</sup>	n/a	0.064	0.051	0.058
Mean	0.118	0.003	0.008	0.005
Minimum	0.000	0.000	0.000	0.000
Lower quartile	0.051	0.000	0.000	0.000
Median	0.100	0.001	0.000	0.000
Upper quartile	0.162	0.004	0.004	0.004
Maximum	1.464	0.022	0.066	0.043
Residual mean	n/a	-0.114	-0.110	-0.112
<u>Root-mean-square errorRMSE</u>	n/a	0.146	0.142	0.144

Formatted: Centered

Formatted: Subscript

Formatted: Centered

5

**Table 9** – Summary statistics of horizontal permanent ground deformation (PGD<sub>h</sub>) predictions-estimates for Darfield and Christchurch earthquakes from EPOLLS and HAZUS models. Refer to Table 4 for descriptions corresponding to model acronyms.

Score	Observed	<u>Horizontal permanent ground deformation, PGD<sub>h</sub> (m)</u>			
		EPOLLS	HAZ1	HAZ2	HAZ3
Pearson R <sup>2</sup>	n/a	0.000	0.022	0.032	0.027
Mean	0.269	0.682	0.141	0.172	0.150
Minimum	0.001	0.149	0.000	0.000	0.000
Lower quartile	0.124	0.418	0.000	0.000	0.000

Formatted: Centered

Formatted: Subscript

Formatted: Centered

Formatted Table

Formatted: Centered

Formatted: Centered

Formatted: Centered

Median	0.206	0.748	0.084	0.050	0.067
Upper quartile	0.312	0.964	0.184	0.191	0.182
Maximum	3.856	1.989	1.872	3.205	2.443
Residual mean	n/a	0.413	-0.128	-0.096	-0.118
<u>Root-mean-square</u> errorRMSE	n/a	0.582	0.345	0.438	0.376

- ← Formatted: Centered
- ← Formatted: Centered
- ← Formatted: Centered
- ← Formatted: Centered
- ← Formatted: Centered