Response to reviewers' comments on "Development of a Seismic Loss Prediction Model for Residential Buildings using Machine Learning – Christchurch, New Zealand"

Relevant changes made in the manuscript

- Line 18: Added hyphen to "government-backed"
- Line 26: Added an "s" at the end of building
- Line 40: Change the order of the list of the few notable ML studies
- Line 42: Added new studies published recently: (Ghimire et al., 2022; Harirchian et al., 2021; Kalakonas & Silva, 2022a, 2022b; Stojadinović et al., 2022)
- Line 53: Replaced "human-level performance" by "baseline performance"
- Line 53: Added "For some applications a ML can be benchmarked against the human-level performance (e.g., image recognition)."
- Line 56: Added a new reference (Ng, 2021)
- Line 67: Corrected the typo
- Line 77: Added an "s" to instance
- Line 81: Corrected the tense of the verbs
- Line 85: Added explanations regarding multiple claims
- Line 87: Changed the tense of the verb
- Line 88: Removed "As supervised ML requires a significant amount of data to be able to learn, only those events are selected for the development of the loss prediction model."
- Line 91: Changed "primary construction material" to "structural system"
- Line 93: Added "Key information such as the building height and primary construction material could not be taken from the EQC dataset."
- Line 97: Updated the author in the citation
- Line 99: Updated the author in the citation
- Line 100: Added an explanation regarding structural information
- Line 116: Change the tense of the verb
- Line 121 to 125: Added a new section on soil conditions
- Line 129: Added details regarding the EQC and RiskScape coordinates
- Line 138: Changed the tense of the verb
- Line 148: Changed the preposition from "in" to "with"
- Line 149: Added information about the percentage of points "(27.1% of the selected RiskScape data for Christchurch had three or more points within a LINZ property boundary)"
- Line 162: Added the preposition "of"
- Line 163: Changed "address" to "addresses"
- Line 166: Added "was"
- Line 173: Removed "the" before the date
- Line 190: Updated the citation as the link to the previous citation got turned down
- Line 209: Added "a" between "as" and "possible"
- Line 212: Updated the phrase "When possible similar categories were combined together (e.g., fibre cement plank and fibre cement sheet combined together in a category fibre cement) and categories with insufficient entries were discarded(e. g, corrugated iron, plastic, glass)." to "When possible similar categories were combined and categories with insufficient entries were discarded."
- Line 215: Changed "neighborhood" to "neighbourhood"
- Line 245: Rearranged the order of the attributes listed
- Line 251: Changed "20% for testing" to "10% for validation and 10% for testing."

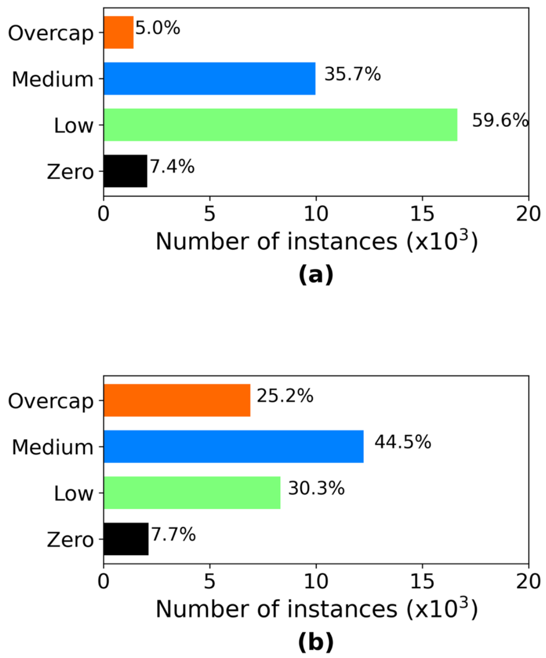
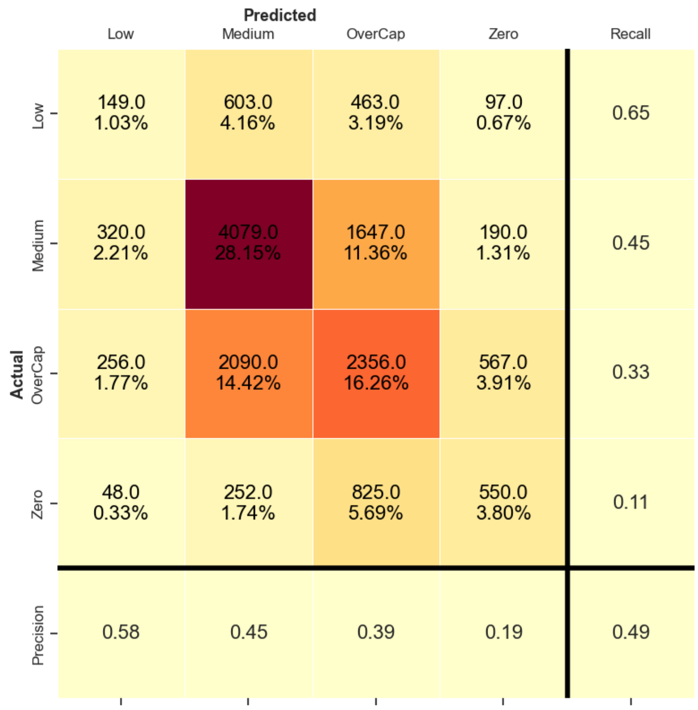
- Line 252: Updated the number of instances
- Line 256: Removed the phrase "Unlike the 'traditional approach' where the test set is held out from the same data as the training and validation set, the test set here employed came from another event in the CES (limited to the four main events). Testing the model using data from another earthquake (preprocessed in the same way as the training and validation set) enabled to evaluate the model capacity to generalize to other events. Thus, changing the earthquake from which the input and test data set comes from, it was possible to study multiple combinations and find the model which generalized best for the entire CES." to avoid confusion (clearer explanations are show in Figure 10)
- Line 264: Added "the"
- Line 266: Changed "are" to "can be"
- Line 275: Removed the hyphen from multiclass and added "l" to "trialled"
- Line 278: Removed the hyphen from overcap
- Line 284: Added "s" to task
- Line 286: Corrected "potential" to "potentially"
- Line 291: Added explanations on the hyperparameter tuning, especially regarding the RandomizedSearchCV function and the scoring using recall
- Line 299 to Line 306: Rearranged the generic explanation of the model evaluation
- Line 316 to 326: Added explanations of the confusion matrices
- Line 328: Changed "on" to "to"
- Line 331: Removed "This validated the probabilistic seismic loss estimation methodology which relies on PGA and the spectral acceleration at selected periods as intensity measures (IM) as the key input."
- Line 333: Removed "the" before the date
- Line 340: Added hyphens between "shaking-dominated" and 'liquefaction-dominated"
- Line 342: Added a new section on model application
- Line 367: Added a new section on model performance and error analysis
- Line 419: Removed "Despite the use of the Python imbalance toolbox to address the imbalance, having more instances in the over-cap category would be beneficial."
- Line 421: Added an hyphen in the "a more in-depth analysis"
- Line 448: Changed "from" to "of"
- Line 453: Added "The model application was demonstrated using a scenario whereby a ML model was trained on data from the 4 September 2010, 22 February 2011, and a representative sample from the 13 June 2011. The ML model was then used to make loss predictions on the rest of the building portfolio."
- Line 462: Changed "development of a seismic loss prediction models" to " rapid seismic loss assessment"

- Fig 6: New figure 6
- Fig 10: Updated figure 10
- Fig 11: Updated figure 11
- Fig 12: Updated figure 12
- Fig 13: Added the proper figure 13
- Fig 14: New figure

- Table 2: New table
- Table 3: New table

Reply to Referee #1

	Referee #1 Comments	Responses
1	a) Dataset. For better prediction results, the authors should preserve (and demonstrate) the original data distribution from the initial dataset when merging and filtering instances. For the same reasons, this reviewer believes that it is necessary to include the class of undamaged buildings in the dataset (with 0$ compensations), which is unavoidable when mapping damage states.	Thanks for the comment. This model has been developed for insurance purposes with the aim of helping the Earthquake Commission (EQC) to get an understanding of the possible loss distribution across Christchurch for any future earthquake. EQC's interest is concentrated on damaged buildings for which a claim might be lodged. It must be reminded that the data used to train the machine learning (ML) model pertains to buildings for which one or multiple claims have been lodged as part of the Canterbury Earthquake Sequence (CES). Getting details information on undamaged buildings was not part of the scope. Thus, extracting reliable data related to the undamaged category is not straightforward. Simply assuming that a building was not damaged because no EQC claims were lodged is not satisfactory, as it is not possible to affirm that the remaining buildings had proper insurance coverage. Moreover, for some buildings which suffered slight damage, the building owners might have decided to cover the cost of the reparations by themselves to avoid paying the excess. Nevertheless, following the suggestion, we tried to include a fourth category for undamaged buildings and explored the influence on the ML model performance. The EQC data includes a few instances with zero compensation (BuildingPaid = NZ$0). Figure i shows the number of instances for 4 Sep 2010 and 22 Feb 2011. With 7%, the number of instances is very limited compared to the "low" and "medium" categories. Despite the low number in the category with no damage, a new machine learning model has been retrained (considering class imbalance). Figure ii shows the confusion matrix for the Random Forest algorithm trained using four categories. It can be seen that the overall accuracy dropped and that the model is having difficulties making predictions for the zero damage category. It was thus decided to keep only three categories (low, medium, and overcap).

		 Figure i: Number of instances in BuildingPaid categorical in the filtered data set including the category with zero compensation: (a) 4 Sep 2010, (b) 22 Feb 2011  Figure ii: Confusion matrix for Random Forest algorithm including the category with BuildingPaid = 0
2	b) Building representation. For a better presentation of the mapping problem, the authors should show the data distribution for more features	Thanks for the comment **Re: data distribution** Thanks for the suggestion. A new table has been added showing the nine attributes used in the model. The table gives information about the type and distribution of each attribute.

<table>
<tr>
<td></td>
<td>(like in Figure 6 for construction type).

Surprisingly, the height of buildings isn’t included in the building representation even though it captures the dynamics of the building. A feature like “number of floors” could possibly be informative.</td>
<td>Re: height of the buildings
Thanks for the remark. The authors of this paper are aware of the inclusion of the building height (or number of stories) as an attribute in the ML models in similar studies (Ghimire et al., 2022; Harirchian et al., 2021; Mangalathu et al., 2020; Stojadinović et al., 2022). The non-inclusion of the building height in this study was dictated by the availability of the information in the dataset and not based on a deliberate choice. While the original EQC data has an attribute for the number of storeys, many instances are missing. Figure iii shows the number of instances available for each storey category. It can be seen that for 4 Sep 2010 and for 22 Feb 2011, storey information is missing for 58% and 93% of the instances respectively. The information available related to the building height is thus very limited. When available, the data for 4 Sep 2010 shows that 10,37 buildings have one storey, 2,677 two storeys, and 140 three storeys. Similarly, for the 22 Feb 2011, 1,522 of the buildings have one storey, 727 two storeys, and 91 three storeys. Selecting only instances where the number of storeys is known would have been very limiting from the aspect of training a ML model. It should be reminded that the EQC cover is limited to residential dwellings. This study is thus limited to residential buildings only which for Christchurch are mostly one-story height houses.
Accurate information on the number of storeys could not be obtained from the RiskScape database either (RiskScape, 2015). The attribute ‘Storeys’ is reported as a float which seems to have been calculated from the building floor area 'BLDGFL_1' divided by the building footprint 'FOOTP_1'.
In the aim of retaining high data accuracy it has been decided to not include the number of storey from the EQC dataset in this study. This has been clarified in section 3.2 of the paper.

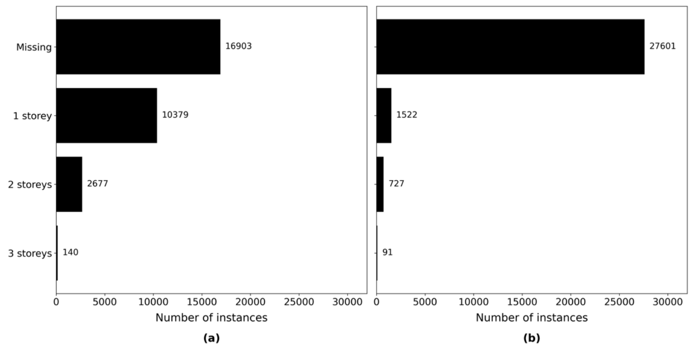


Figure iii: Number of instances for each storey category (EQC data): (a) 4 Sep 2010, (b) 22 Feb 2011</td>
</tr>
<tr>
<td>3</td>
<td>c) More discussion on the confusion matrix would be helpful. For example, how do the authors explain the accuracy difference between b) and c) in Figure 12? Does it have to do with PGA ranges of</td>
<td>Thanks for the comment.
Section 8 of the paper has been rewritten. It now includes a discussion of the model performance using the results presented in the confusion matrices.</td>
</tr>
</table>

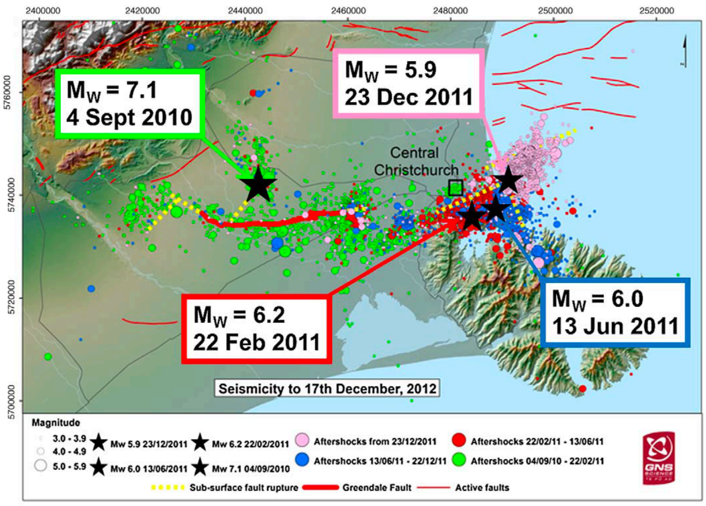
	earthquakes? The worse prediction seems to be "Predicted Medium / Actual OverCap". How to explain this?	
4	d) It would be interesting to evaluate the prediction accuracy for the sum of compensations for all buildings. It is reasonable to expect good prediction accuracy for total cost since errors would cancel out each other. But it is difficult to perform without precise "OverCap" values.	Thanks for the suggestion. Building claims larger than NZ$100,000 (+GST) were handled by private insurers. Unfortunately, private insurers were not inclined to make their data available for this research work. As mentioned in paragraph 5.2, the current data is thus soft-capped at NZ$100,000 (+GST) making the analysis of the total costs not possible with the currently available data.
5	e) Finally, how do the authors evaluate the usefulness of the research and model implementation for new earthquakes? Namely, what about the changing value of money over time and frequent changes in market prices? How to implement the model without the class of undamaged buildings (this version could work just if damaged buildings were pre-selected)?	**Re: model implementation for new earthquakes** Thanks for the comment. The ML model presented in this paper was designed to be easily retrainable. This enables the addition of new instances in the training set after the occurrence of a new earthquake. However, one of the challenges, is that claim amounts are not readily available after an earthquake as on-site assessment of building damage can be spread over a long period of time. To circumvent this issue, building damage should be assessed on a representative set of buildings. This subset where the damage extent will be known should be added to the training set of the ML model that can be retrained. The ML model can then be used to make predictions on the entire building portfolio. This approach has been schematically described in a new version of Fig 10. The selection of a representative training set can be made following an event based on the effects of the earthquake or prior to the event using a predetermined representative subset of the residential building in Christchurch. Special care should be applied to ensure that the selected buildings can be used to produce a satisfactory seismic loss assessment at the city level. Recent discussions with experts highlighted the uniqueness of the Canterbury region. They mentioned that the analysis of damage observations across the CES showed that the main earthquake events affect different areas in Christchurch. They thus suggested that the selection of a representative set of buildings should take into account the geographical characteristics, the liquefaction setting, and building characteristics. When expert opinion is not available, similar studies showed that ML could even be employed in the selection of a representative building set (Mangalathu & Jeon, 2020). The actual selection of a representative set of buildings for Christchurch is beyond the scope of this study. **Re: Change in value of money** Thanks for the remark. The authors agree with the necessity to consider the evolution of the market over time. Here again, the ease of retraining the ML model when new or

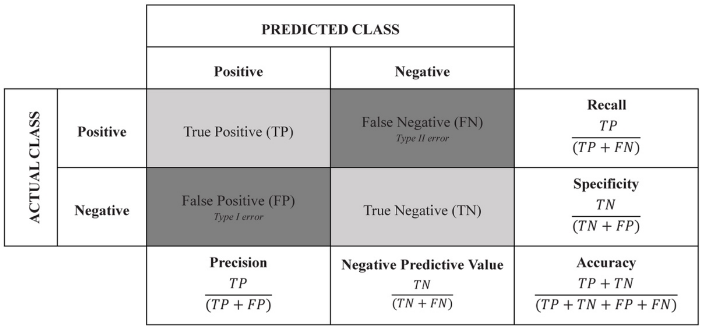
		updated training data gets available should be highlighted. The authors are aware of the step change related to the value of the EQC cap over time (e.g., since 1 Oct 2022, the new cap is at NZ$300,000 + GST (Earthquake Commission (EQC), 2022). Nevertheless, Fig 8 showed that for 4 Sep 2010 earthquake most of the claims fell in the 'low' and 'medium' categories. Even for the 22 Feb 2011 earthquake, which was unprecedented in the damage extent caused in the Canterbury region, many claims still relate to the 'low' and 'medium' categories. It is thus believed that the value of the model lies in its ability to make predictions for those categories ('low' reflecting the limit of initial cash settlement consideration, 'medium' for building having more damage but where claims are still fully addressed by EQC only, and 'overcap' where private insurer come into consideration for higher level of damages). **Re: no undamaged buildings class** Thanks again for the comment. Please see our reply to comment #1. This ML model has been developed with the purpose of being used in the insurance setting. The focus is thus on being able to predict the possible damage and loss extent for buildings having EQC claims in future earthquakes.
6	Technical corrections There is a significant number of needed technical corrections. Some examples are highlighted in the attached file. The authors should carefully check the paper.	Thanks for having highlighted those typos. The errors have been corrected.

Reply to Referee #2

We would like to thank referee #2 for the detailed and constructive feedback. We are grateful for the thoughtful comments and suggestions regarding the application of machine learning to real-world data. We are currently in the process of exploring and trying to implement the requested changes.

	Referee #2 Comments	Responses
1	The selection of the test set is unclear in the paper. It also appears that the test set has been erroneously used as a validation set. If that is the case, then it is difficult to assess the generalizability of the authors' conclusions. It would be helpful to clarify how the test set was selected and used in this study. Additional comments regarding test set are also included below with specific line references.	Thanks for the comment. The selection of the training, validation, and test set has been revised. The updated version is schematically shown in Fig 10. The training and the test set are now sourced from the same earthquake. The validation set is implemented using k-fold cross-validation as part of the hyperparameter tuning. Models have been trained and tested using data from the four main earthquakes in the CES (4 Sep 2010, 22 Feb 2011, 13 Jun 2011, 23 Dec 2011).

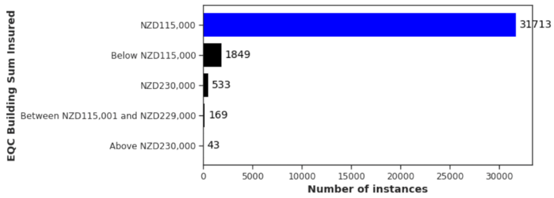
2	While it is a suitable approach to only select the 4 events with the highest number of claims during model training, the other events with fewer claims could be used for testing purposes. This would not only ensure that no data leakage occurred between the training and test sets, but also enable the authors to validate the generalizability of their models more effectively.	Thank you for the suggestion. As mentioned in the reply to comment #1, the process of selecting the training, validation, and test set has been reviewed. The training and test set are now clearly separated thus ensuring that no data leakage can occur. Additionally, the way the model should be applied to future earthquakes has been clarified in Fig 10 as well as in the paper. We are here deliberately focusing on the claim data related to the four main earthquakes rather than the claim data pertaining to the aftershocks. Figure iv shows the situation in the area around Christchurch as of December 2012 (one year after the end of the CES). It can clearly be seen that the earthquakes with the larger magnitudes were: 4 Sep 2010, 22 Feb 2011, 13 Jun 2011, 23 Dec 2011. Severe buildings damage, and hence higher claims, are mainly related to those four earthquakes as the seismic intensity and liquefaction occurrence were larger for those events, hence the selection of those earthquakes in this study.  Figure iv: Map of the region around Christchurch showing the epicentre location of the four main earthquake and multiple aftershocks in the Canterbury earthquake sequence (CES) (O'Rourke et al., 2014, originally from GNS)
3	It would improve the paper if the authors added their thoughts on some of the potential use cases of this research. While the authors' conclusions indicate a promise for using ML within this domain, it was unclear how this model and approach could be used in the future. For example, if training data is needed each time an earthquake occurs, then is one of the use cases to manually collect a subset of ground truth data for building losses, train a model, and then apply it widely to the rest of the buildings?	Thanks for the suggestions Please see our reply to comment #5 of reviewer #1 concerning the implementation of the model. A new section on the model application has been added to the paper.

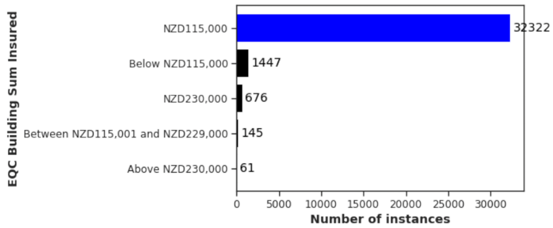
4	Further discussion of the model metrics such as recall and precision would be helpful. For example, a recall of only 20% for overcap, and 49% for low loss category indicates that 80% and 51% of these losses, respectively would be missed when implementing this model. Depending on the model's use cases, this could have a significant impact on the model's utility. Further discussion of the most appropriate metric (or their combinations), given the model's use cases would also improve the paper. For example, why was accuracy selected as the primary evaluation metric for choosing the best performing model?	Thanks for the comment. As studies applying machine learning in an earthquake engineering context (see section 2.1 of the paper) are now more common, many publications include background explanation on theoretical part related to ML (e.g., Harirchian et al. (2021)). The authors are aware of the different metrics related to a classification (Figure v) but decided to not include generic explanations related to ML to keep the paper to reasonable length. One of the main reason for conveying the model performance using the accuracy was to enable the benchmarking of this ML model against the performance of other ML models for damage prediction. While a direct comparison would be improper as the earthquake selected, model attributes, and algorithm are not the same, most of the current studies report the model performance using the accuracy (Ghimire et al., 2022; Harirchian et al., 2021; Mangalathu et al., 2020; Stojadinović et al., 2022). The model has been retrained using recall as the scoring. The RandomizedSearchCV from scikit-learn was applied with k-fold validation. More information on the hyperparameter tuning has been added to section 7.  Figure v: Details of a confusion matrix for a binary class problem
5	I appreciated that the authors listed the distribution imbalance of different features, such as construction type. However, the paper could be further improved by adding the model performance in those different feature categories. This would enable the reader to understand in which categories the model performs better than others.	Thanks for the suggestion. This is currently work in progress. Once the model retrained, we will have a closer look at the model performance for each of the categories. We will add interesting findings to the paper.
6	Given the relatively low performance of the ML model (as highlighted above for recall), adding a section on error analysis would substantially improve the paper. In error analysis for ML, the objective	Thanks for the suggestion. A new section on model performance and error analysis has been added to the paper.

	is to identify the cases in which the model does not perform well. This error analysis is often used in ML modeling to improve model performance and generalizability.	
7	Figure 13 is missing, and appears to be a repeat of Figure 12. Hence, Section 9 - Insights - could not be reviewed.	Thanks for note. Fig 13 has been updated.
8	It would further improve the paper if the authors added some information about their hyperparameter tuning methodology, and which search strategy they used.	Thanks for the suggestion. For the hyperparameter tuning we used the randomized search cross-validation approach. We selected RandomizedSearchCV over GridSearchCV as for larger datasets RandomizedSearchCV often outperforms the results from a GridSearchCV. As suggested in comment #4, recall is currently being considered. This is done during the hyperparameter tuning process adjusting the scoring for recall. Clarifications have been added to section 7 of the paper.
9	Line 50 - While the authors are completely correct in the paragraph at line 50, this paper deals with ML for structured data, for which the goal is often to surpass human performance since humans are generally unable to identify all patterns in millions of data points with hundreds of features, often found in these problems. Hence the paragraph does not apply to the ML scope of this paper. It may be suitable to remove the paragraph within the scope of this paper, or change “human-level performance” to “baseline model”, which would be a more suitable term in this case	Thanks for the remark. We have reformulated the paragraph to make clearer that in this case the ML model can surpass the human-level performance. As explained in section 11, the performance of this ML model was assessed against the outputs of the software RiskScape v1.0.3.
10	Line 65 - latter -> later	Thanks for the note. The typo has been corrected.
11	Line 73 - Suggest adding reference/url for the source of the data.	Thanks for the suggestion. Unfortunately the EQC data is not public. We signed a confidentiality agreement with EQC which made the data available to us for research purposes only.
12	Line 83 - It would be helpful to further describe Figure 2. Why is there a difference in the number of claims and buildings?	Thanks for the comment. In some cases, multiple claims might have been lodged for the same earthquake event. Clarification has been added in line 88 of the paper.
13	Line 95 - I was curious about the accuracy of the Riskscape dataset. For example, are the building characteristics determined statistically from Census data similar to HAZUS in the US, or was it based on collecting data from	The building characteristics data was obtained from the RiskScape - Asset Module Metadata (RiskScape, 2015). A copy of the document has been attached in Appendix A. According to the documentation the “data in this inventory is partly derived from information purchased from Quotable Value (QV) Ltd, together with ‘industry knowledge’ and data gathered from surveying the area. All QV data is applied at

	building records so that it is expected to be fairly accurate? If possible, it would be helpful in the paper to include some information describing Riskscape's data collection methodology and comment on its expected accuracy.	the meshblock level, and the RiskScape attributes are derived from this information so as to provide a suitable model of the actual building stock." Further information can also be found in (King & Bell, 2006; Reese et al., 2007).
14	Line 115 - Although a reference is provided to the authors' previous work, it would be helpful to summarize the major reasons for incorrect merging using direct spatial joins within this paper to help understand the issue without having to read the previous work.	Thanks for the comment. The paragraph has been rewritten to include explanations related to the location of the EQC and RiskScape coordinates. It is now clearly stated in the paper that the coordinates provided as part of the EQC dataset relate to the location of the street address while the RiskScape coordinates are located in the actual centre of the footprint of a building. In some cases, buildings from neighbouring properties are located closer to the street address than the actual building. For those cases, the use of spatial join functions and spatial nearest neighbour joins led to unsatisfactory outputs.
15	Line 122 - It would be helpful if the authors added the percentage of addresses in each of the 3 categories - 1-1 match with titles, 0-1 match, and many-1 match.	Thanks for the comment. As suggested in comment #18, the percentages have been added to Table 1.
16	Line 131 - It would be helpful if the authors added the percentage of RiskScape data that was discarded.	Thanks for the comment. A comment mentioning that 27.1% of the RiskScape instances were having 3 or more RiskScape datapoints within a LINZ property title as been added at the end of the phrase. Additionally, as suggested in comment #18, the percentages have also been added to Table 1.
17	Line 132 - I was unable to understand the intent described in this paragraph, especially the first and the last sentences.	Thanks for your comment. Fig 3 showed that some LINZ property titles included more than one point (i.e. building) per property title (polygon). This poses challenges for an automated data merging process as it needs to be ensured that the RiskScape and EQC information get assigned to the correct building. As mentioned on line 121: "The merging process was thus started with instances having a unique street address per property". Those instances having a unique building per polygon were used to merge RiskScape attributes and EQC info to the buildings constraining the merging within a property title. The actions for merging scenarios related to property titles having 1 point per LINZ property are listed in Table 1 rows 1 to 3. It is known that the training of an ML model is often benefiting from more training data. Thus, options to merge more data were explored. It was found that among all the LINZ data, 7% of the LINZ property titles have two street address points (e.g. two buildings within one polygon). We explored if there would be an automated way to merge RiskScape attributes for such cases and thus get more data. For instances having 2 buildings (LINZ points) per property and 2 RiskScape points, the automated merging was done joining the RiskScape attribute to the closest LINZ point and

		filtered using the building floor area. Among the instances where a LINZ property title had 2 buildings it was also found that some the RiskScape data only had one point per property title. While a human could make an educated guess to find which of the two buildings the RiskScape attribute were pertaining to, no satisfying automated approach could be developed to merge these instances. This case was thus discarded. More details can be found in section 5.6.6 Properties with two street addresses and one or multiple RiskScape Instances of the thesis (Roeslin, 2021). Section 5.6.6 also includes multiple figures showing examples of properties having two LINZ NZ street addresses per polygon. To keep this paper to a reasonable length such figures were not included.
18	Table 1 - The table is very helpful. However, the action taken for 2 points LINZ and 1 point Riskscape was unclear. The above mentioned percentages of data could also be added to Table 1 instead.	Thanks for the comment. Re: action for 2 LINZ points and 1 RiskScape The wordiness has been removed. In short, those points were discarded Re: Percentage to Table 1 Thanks for the suggestion. The percentages have been added in brackets for each LINZ and RiskScape scenario of the Table 1. Out of the selected instances in Christchurch, 89% have 1 street address point per LINZ property title, 7% two addresses per property, and 4% have three instances or more. For the RiskScape data (RiskScape, 2015), after merging in ArcMap it was found that 29.3% of the properties have one RiskScape instance, 43.6% have two instances, and 27.1% have three RiskScape points or more.
19	Line 150 - It would be helpful if the authors added the methodology used to merge soil conditions, and liquefaction occurrence with street address. Did they use the same inverse distance weighted interpolation as seismic demand?	Thanks for the comment. The seismic demand captured through the peak ground acceleration (PGA) was interpolated using the inverse distance weighted (IDW) technique, applying the IDW spatial analyst function in ArcMap (Esri, 2021). In the paper, the explanation on how the seismic demand was interpolated from the ground motion recordings obtained from GeoNet is explained in section 3.3.
20	Line 172 - The reason for discarding claims with maximum value lower than or higher than \$115,000 is unclear. Is it because this wasn't possible and hence the data is erroneous?	Thanks for pointing that out. The EQC data entails a feature called 'EQC Building Sum Insured'. As EQC provided a maximum cover of NZ\$100,000 (+ GST) at the time of the Canterbury Earthquake Sequence, it was expected that all the individual dwellings for which one or multiple claims have been lodged during the CES would have an EQC Building Sum Insured = NZ\$115,000. However, the initial exploratory data analysis (EDA) revealed that for some of the instances the building sum insured was not at NZ\$115,000 (see Figure vi). As only a few instances were not equal to NZ\$115,000 (2,594 instances for 4 Sep 2010 and 2,329 instances for 22 Feb 2011), it was decided not to include those instances (n the objective of retaining accurate data for the training of the ML model).

<table>
<tr><td></td><td></td><td>


(a) 4 September 2010


(b) 22 February 2011

Figure vi: Number of instances in the attribute EQC Building Sum Insured (categorised to simplify the visualisation)
</td></tr>
<tr><td>21</td><td>Line 240 - It is unclear which event was selected as the test set. From my understanding of line 243, one of the 4 events was selected as test set, and the other 3 events as training+validation sets. However, it also appears from the sentence that in different instances of the model, a different event was selected as a test set so as to determine the most generalizable model. If that is the case, the test set was erroneously used as a validation set, since the model cannot be changed at any point after evaluating its performance on the test set. It would be helpful to clarify the selection of the test set, and ensure that it was only used once at the end to evaluate the performance of the final developed model.</td><td>Thanks for the remark.
As mentioned in the reply to comment #1, the selection of the training, validation, and test set has been revised. All of those sets are now coming from the same earthquake event. For new earthquakes, a sample of buildings will be selected and added to the training set. Figure 10 has been updated to reflect those changes.</td></tr>
<tr><td>22</td><td>Line 254 - It would be helpful if the authors added how the min-max scaling was implemented with respect to training, validation, and test sets.</td><td>Thanks for the comment.
The min-max scaling of the numerical features was performed using the sklearn.preprocessing.MinMaxScaler available in scikit-learn (Pedregosa et al., 2011). A pipeline containing the min-max scaler was created. All sets were passed through the same pipeline.</td></tr>
<tr><td>23</td><td>Line 286 - It is unclear which limitations related to random forest model the authors are referring to.</td><td>Thanks for pointing that out.
The limitations mentioned in this section refer to the overall limited model performance not to the limitations of the random forest algorithm itself. The first section of the phrase has been rewritten to clarify that it is related to the overall model accuracy.</td></tr>
</table>

24	Figure 11 - The SVM model does not appear to have been modeled correctly as its output prediction is always the medium category, hence it has been reduced to a trivial model.	Thanks for pointing that out. We are currently having a deeper look at the retraining of the SVM model.
25	Line 295 - It appears that the model was selected based on the best performing model on the test set. This indicates that the test set was not used correctly, as the model selection can only be done using validation sets. The test set must only be used to show the performance of an already selected model on it.	Thanks for the comment. Sorry for the confusion. The paragraph has been rewritten and Fig 10 has been improved to clarify that the training, validation, and testing are initially done on one earthquake only.
26	Line 326 - While the authors raise an accurate point about the lack of claims information exceeding \$115,000, it is not clear how that data could have benefitted this study since the claims have been bucketed and all those claims greater than \$115,000 are already expected to be included in the over-cap category.	Thanks for the remark. The ML model currently presented in this study is categorical. Yet, the target attribute BuildingPaid is initially numerical but soft-capped at NZ\$100,000 (+GST) or NZ\$115,000. Having the actual numerical distribution of the extent of the claims might enable a deeper analysis of the target attribute and can possibly enable better performance for a regression model. This might alleviate the need to transform BuildingPaid from a numerical attribute to a categorical feature.

References

Earthquake Commission (EQC). (2022). *EQC Insurance Overview*. https://www.eqc.govt.nz/what-we-do/insurance-overview/

Esri. (2021). *ArcMap - IDW (Spatial Analyst) - Documentation*. https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/idw.htm

Ghimire, S., Guéguen, P., Giffard-Roisin, S., & Schorlemmer, D. (2022). Testing machine learning models for seismic damage prediction at a regional scale using building-damage dataset compiled after the 2015 Gorkha Nepal earthquake. *Earthquake Spectra*. https://doi.org/10.1177/87552930221106495

Harirchian, E., Kumari, V., Jadhav, K., Rasulzade, S., Lahmer, T., & Raj Das, R. (2021). A Synthesized Study Based on Machine Learning Approaches for Rapid Classifying Earthquake Damage Grades to RC Buildings. *Applied Sciences*, *11*(16), 7540. https://doi.org/10.3390/app11167540

Kalakonas, P., & Silva, V. (2022a). Seismic vulnerability modelling of building portfolios using artificial neural networks. *Earthquake Engineering and Structural Dynamics*, *51*(2), 310–327. https://doi.org/10.1002/eqe.3567

Kalakonas, P., & Silva, V. (2022b). Earthquake scenarios for building portfolios using artificial neural networks: part II—damage and loss assessment. *Bulletin of Earthquake Engineering*. https://doi.org/10.1007/s10518-022-01599-2

King, A., & Bell, R. (2006). Riskscape New Zealand - A Multihazard Loss Modelling Tool. *Proceedings of 2006 New Zealand Society for Earthquake Engineering Conference, Napier, New Zealand*, *Paper 34*, 9. https://nzsee.org.nz/db/2006/Paper30.pdf

Mangalathu, S., & Jeon, J.-S. (2020). Regional Seismic Risk Assessment of Infrastructure Systems through Machine Learning: Active Learning Approach. *Journal of Structural Engineering*, *146*(12). https://doi.org/10.1061/(ASCE)ST.1943-541X.0002831

Mangalathu, S., Sun, H., Nweke, C. C., Yi, Z., & Burton, H. v. (2020). Classifying earthquake damage to buildings using machine learning. *Earthquake Spectra*, *36*(1), 183–208. https://doi.org/10.1177/8755293019878137

Ng, A. (2021). CS230 Deep Learning - C3M1: ML Strategy (1). In *PowerPoint Slides - CS230 Deep Learning - Stanford*. https://cs230.stanford.edu/files/C3M1.pdf

O'Rourke, T. D., Jeon, S. S., Toprak, S., Cubrinovski, M., Hughes, M., van Ballegooy, S., & Bouziou, D. (2014). Earthquake response of underground pipeline networks in Christchurch, NZ. *Earthquake Spectra*, *30*(1), 183–204. https://doi.org/10.1193/030413EQS062M

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research 12*, *12*, 2825–2830. https://doi.org/10.1007/s13398-014-0173-7.2

Reese, S., Bell, R., & King, A. (2007). RiskScape: a new tool for comparing risk from natural hazards. *Water & Atmosphere*, *15*(3), 24–25. https://niwa.co.nz/sites/niwa.co.nz/files/import/attachments/tool.pdf

RiskScape. (2015). *Asset Module Metadata*. https://github.com/samroeslin/samroeslin.github.io/blob/gh-pages/assets/pdf/RiskScape-Asset_Module_Metadata-2015-12-04.pdf

Roeslin, S. (2021). *Predicting Seismic Damage and Loss for Residential Buildings using Data Science* [University of Auckland]. https://hdl.handle.net/2292/57074

Stojadinović, Z., Kovačević, M., Marinković, D., & Stojadinović, B. (2022). Rapid earthquake loss assessment based on machine learning and representative sampling. *Earthquake Spectra*, *38*(1), 152–177. https://doi.org/10.1177/87552930211042393

Appendix A - RiskScape - Asset Module Metadata

Summary Information

Attribute	Details
Author(s)/Organisation	RiskScape (National Institute of Water and Atmospheric Research Ltd, GNS Science)
Contact Address	National Institute of Water and Atmospheric Research Ltd 301 Evans Bay Parade Greta Point Wellington 6021 GNS Science 1 Fairway Drive Avalon Lower Hutt 5011
Contact Email	info@riskscape.org.nz
Inventory Name	New Zealand Buildings
Inventory Description	The 'New Zealand Building' inventory developed for asset impact and loss modelling applications using RiskScape software. Building asset information relate to individual buildings defined as, a permanent enclosed structure including a roof, walls and one or more levels, which are used for a variety of activities. Asset attribute information is stored and presented as vector point features.
Inventory References	Cousins, W. J. 2009. RiskScape – development of a default assets model for Hawke's Bay, GNS Science Report 2009/50. 33 p. King, A.B., Bell, R.G. (Programme Managers). (2009). RiskScape Project: 2004 - 2008. NIWA Science Report 2009/75. 172p.
Inventory Use Case Description	The 'New Zealand Building' inventory has been applied in a number of impact and risk modelling case studies. A selection of relevant studies are provided in the text box below. Most applications have involved the inventory use in earthquake, flood, tsunami, storm-tide and wind impacts or loss modelling to building damage states and reinstatement costs as well as indirect impacts such as, displacement of people.
Inventory Use Case References	Bell, R. G., Wadwha, S., Paulik, R. (2015). National and regional exposure to coastal hazards and sea-level rise. Areal extent, population and assets. Prepared for Parliamentary Commissioner for the Environment, June 2015. NIWA Client Report No: HAM2015-006. Grace, E. S. (compiler). 2014. Gisborne District Risk Assessment, GNS Science Consultancy Report 2014/279. 149 p. Lane, E. M., Mountjoy, J., Power, W. L., Mueller, C., Paulik, R., Crowley, K. (2015). The hazard and risk of tsunami inundation due to submarine-landslide-generated tsunamis in Cook Strait Canyon. Coasts and Ports 2015 Conference, 15-18 September 2015, Auckland, New Zealand. Paulik, R., Lane, E. 2014: Future tsunami risk at Omaha Beach, Auckland. 7th Australasian Natural Hazards Management Conference 2014, 23-24 September 2014, Wellington, New Zealand. Paulik, R., Lane, E., Sturman, J. (2012). Tsunami impact modelling using RiskScape: Omaha Beach, Auckland. New Zealand Coastal

	Society Conference: Making waves, 20 years and beyond 14-16 November, 2012, Auckland, New Zealand. Paulik, R., Smart, G., Bind, J. (2014). 'Flockton Basin' building impact and loss estimates for the March 5th 2014 Christchurch flood event. 7th Australasian Natural Hazards Management Conference 2014, 23-24th September 2014, Wellington, New Zealand. Paulik, R., Turner, R., Sturman, J., Gray, S., Flay, R. (2013). Using RiskScape to estimate building impacts and loss from the 2012 Hobsonville Tornado, Auckland. "Water & Weather: Solutions for Health, Wealth and the Environment" The NZ Hydrological Society & The Meteorological Society of NZ Joint Conference. 19-22 November 2012, Palmerston North, New Zealand. Smart, G., Paulik, R. (2012). Prioritising Perils: A Case Study. 6th Australasian Natural Hazards Management Conference 2012, 21-22 August 2012, Christchurch, New Zealand. Wright, K. C., Johnston, D. M., Cousins, W. J., McBride, S. K. (2012). Estimating post-earthquake welfare and sheltering needs following a Wellington earthquake. New Zealand Society for Earthquake Engineering (NZSEE) Annual Conference "Implementing Lessons Learnt". April 13-15, Christchurch, New Zealand.
Inventory is a Component of a RiskScape Asset Module	Yes
RiskScape Asset Module Name	The 'New Zealand Buildings' inventory is currently disaggregated and supplied as regional modules for New Zealand.

Asset Information Summary

Attribute	**Primary Attribute**	**Secondary Attribute**	**Field Name**	**File Geodatabase Alias**	**Data Type**
RiskScape Asset Type	Buildings	NA	NA	NA	NA
RiskScape Asset Attributes	Building Earning Potential	2011 NZD$	NA	NA	Short Integer
	Condition	1: Sound 2: Deficient	NA	NA	Text
	Construction Type	1 : Reinforced Concrete Shear Wall 2 : Reinforced Concrete Moment Resisting Frame 3 : Steel Braced Frame 4 : Steel Moment Resisting Frame 5 : Light Timber 6 : Tilt Up Panel 7 : Light Industrial 8 : Advanced Design	NA	NA	Text

		9 : Brick Masonry 10 : Concrete Masonry 11 : Unknown Residential 12 : Unknown			
	Contents Value	2011 NZD$ 0-	NA	NA	Short Integer
	Deprivation Index	1 : DI 1 2 : DI 2 3 : DI 3 4 : DI 4 5 : DI 5 6 : DI 6 7 : DI 7 8 : DI 8 9 : DI 9 10 : DI 10	NA	NA	Short Integer
	Floor Area	0-	NA	NA	Short Integer
	Floor Height	0-	NA	NA	Double
	Floor Type	1 : Timber 2 : Concrete Slab	NA	NA	Text
	Footprint Area	0-	NA	NA	Short Integer
	Occupancy	0-	NA	NA	Double
	Parapet	1 : No Parapet 2 : Has Parapet	NA	NA	Text
	Replacement Cost	2011 NZD$ 0-	NA	NA	Short Integer
	Roof Cladding Class	1 : Clay/Concrete Tile 2 : Concrete Slab 3 : Membrane 4 : Metal Tile 5 : Other - Heavy 6 : Other - Light 7 : Sheet Metal	NA	NA	Text
	Roof Pitch	0-89°	NA	NA	Short Integer
	Storeys	1-	NA	NA	Double
	Use Category	1 : Residential Dwellings 2 : Commercial – Business 3 : Commercial - Accommodation 4 : Industrial - Manufacturing, Storage 5 : Industrial - Chemical, Energy, Hazardous 6 : Fast Moving Consumer Goods 7 : Government	NA	NA	Text

		8 : Territorial Authority/Civil Defence 9 : Lifeline Utilities 10 : Police 11 : Hospital, Clinic 12 : Fire Station 13 : Community 14 : Education 15 : Resthome 16 : Religious 17 : Forestry, Mining 18 : Farm 19 : Lifestyle 20 : Parking 21 : Clear Site 22 : Other			
	Vehicle Value	2011 NZD$ 0-	NA	NA	Short Integer
	Vehicles	0-	NA	NA	Short Integer
	Wall Cladding Class	1 : Weatherboard 2 : Stucco, Roughcast 3 : Corrugated Iron 4 : Plastic 5 : Fibre Cement Sheet 6 : Fibre Cement Plank 7 : Reinforced Concrete 8 : Concrete Masonry 9 : Brick 10 : Glass 11 : Curtain Wall Glazing 12 : Sheet Metal 13 : Other Sheet - Combustible 14 : Other Sheet - Non-Combustible 15 : Other	NA	NA	Text
	Year of Construction	1800 -	NA	NA	Short Integer
Other Attributes	NZTME	-	NA	NA	Double
	NZTMN	-	NA	NA	Double

Caveats and Constraints

Attribute	Details
Use Caveats	The following data sources were used in the development of the New Zealand Building Inventory: • Quotable Value NZ or QV for meshblock-scale aggregated property data for all of New Zealand, and for individual property data for all

	properties larger than 2000 m2 in New Zealand. Property to land parcel links were available for most of the individual properties. • Statistics New Zealand (population and employment data). • Ministry of Education (school population data). • Local government (building footprint and earthquake prone building data). • User and stakeholder knowledge and investigation of local asset attributes. • RiskScape national building statistics generated from site investigations in New Zealand towns and cities (e.g. Westport, Christchurch, Hastings, Napier). This data in this inventory is partly derived from information purchased from Quotable Value (QV) Ltd, together with 'industry knowledge' and data gathered from surveying the area. All QV data is applied at the meshblock level, and the RiskScape attributes are derived from this information so as to provide a suitable model of the actual building stock. Due to the time required to individually survey buildings, very few buildings in this dataset are verified against their actual condition however, over time and with the cooperation of RiskScape users this dataset will be improved to include progressively more and more verified attribute information. Initial building data are derived from QV datasets. These have a number of issues, not least in the inexact and imprecise nature of the data they contain. There are many values with duplicate addresses while some of these have wrong locations and the building data where they exist is of unknown integrity. Issues associated with deriving national building statistics arose from field surveys. What can be seen from the street is limited to fields such as building storeys, wall and roof cladding, outbuildings etc. This is also true of buildings on rear sections. Often these are not of the same age class as those visible from the street, as infill housing typically happens in older suburbs. This could bias the observations by age class. Estimation of building width and depth also varies with observer's experience. Details of internal construction are educated guesses based on external features; e.g. floor construction whether slab or timber, internal structure, cladding etc. These observational problems are more pronounced in commercial/industrial areas. Here they are arguably more important as the value of individual buildings increases and has a large impact on the total risk profile. Floor level calculations are sensitive to ground slope, GPS location and inclinometer accuracy. In Westport 10% of floor levels calculated were below the ground at the house (King and Bell, 2009). In the Napier/Hastings survey, the equivalent figure was 50%. This may relate to the coarser DEM being used (as supplied by HBRC) in this region. Observational errors exist for the floor height measurements, but some could be detected by consideration of photographs, and by comparison with nearby values. Errors are probably present in other measurements too, but can be less easy to find. Alternatively, the necessary error bounds are simply larger (for variables such as building width and depth for example).

	Assignation of relevant variables to unmeasured buildings also is a difficult problem in regions lacking usable building attribute data. Very few related variables are available to construct models from without local data. For other variables, we have adopted either a constant value or a stochastic approach where values are assigned over the observed range and distribution, but randomly across the region. This approach gives overall regional risk at about the right level but may lead to bias in particular aggregation units (e.g. meshblocks) where the buildings are more similar and have a more restricted distribution of any particular variable. Due to the QV database being based around property rather than building, many 'buildings' are colocated in the source data. This is due to flats or other residences which share the same building. However the process of attribute derivation has a random element to it and thus may not assign the same attributes to two or more colocated assets. For this reason, any assets which are colocated in the dataset are re-located evenly on a circle of radius five metres around the original location. Despite the various problems and uncertainties overall the data seems to be reasonable accurate especially on a meshblock or regional level, where eventual errors level out. This is currently also the most comprehensive database known by the RiskScape project team to be available in New Zealand and contains some building attributes that cannot be found elsewhere (e.g. floor height). Asset attribute quality levels provide RiskScape users with certainty on asset data validity and can be used to quantify the quality of the final result of a model run. Asset data quality levels in RiskScape vary from guessed values estimated from a general understanding of the assets' attributes, to measured values resulting from an engineer's examination of asset attributes. Each asset attribute must be described by a quality level from the list below. This may be used by RiskScape to quantify the quality of the final result of an impact or risk analysis. 1. Global Knowledge: Guessed from general understanding of assets. 2. Derived - Low Reliability: Derived by random selection from distributions - low reliability. 3. Derived - High Reliability: Derived by random selection from distributions - high reliability. 4. Supplied: Supplied by a reliable agent on a building by building basis (e.g. council, QV, owner). 5. Observed: Observed by a walk-by survey where only part of the asset is visible. 6 Surveyed: Surveyed by detailed inspection of the specific asset. 7. Measured: Measured by reference to plans and engineering calculation.
Security Constraints	Data is currently provided only at an aggregated level.
Legal Constraints	RiskScape data and models are being supplied to you for your evaluation and for no other purpose. The data [or any information derived from the data] may not be used for commercial purposes. You may not under any circumstances copy, sell or supply the data [or any information derived from the data]. The data is proprietary information and you will hold the data in confidence. Installation of this product indicates your acceptance the terms under which you may evaluate the data.

	While all reasonable effort has been made to ensure that the data are as accurate as practicable, neither the RiskScape partners nor the other data source organisations can be held responsible for errors in the data, or for any actions taken based on the data. No warranties are given in relation to the data or its suitability. RiskScape partners and the other data source organisations therefore, to the full extent permitted by law, exclude liability, including for negligence, for any loss or damage, direct or indirect and howsoever caused resulting from any person or organisations use or reliance on the data.

History Summary

Revision No.	Date Published	Feature Count	Features Added	Features Modified	Features Deleted	Details
1	01/02/2013	2,275,809	0	0	0	First version of the 'New Zealand Building Inventory' for RiskScape completed.
2	04/12/2015	NA	NA	NA	NA	'New Zealand Building Inventory' metadata template developed for RiskScape Wiki.

Other Metadata

Attribute	Details
Date Added	01/02/2013
Last Update	04/12/2015
Progress	On Going
Data Type	Vector Point
Format	ESRI Point Shapefile
Projection	New Zealand Transverse Mercator (NZTM), NZGD 2000
Geographic Extent	New Zealand land area excluding Stewart Island and other offshore islands.

Originally obtained from (RiskScape, 2015)