



Data Integration for the Development of a Seismic Loss Prediction Model for Residential Buildings in New Zealand

Samuel Roeslin^(✉), Quincy Ma⁽ⁱ⁾, Joerg Wicker⁽ⁱ⁾, and Liam Wotherspoon⁽ⁱ⁾

The University of Auckland, Auckland 1010, New Zealand
s.roeslin@auckland.ac.nz

Abstract. In 2010–2011, New Zealand experienced the most damaging earthquakes in its history. It led to extensive damage to Christchurch buildings, infrastructure and its surroundings; affecting commercial and residential buildings. The direct economic losses represented 20% of New Zealand's GDP in 2011. Owing to New Zealand's particular insurance structure, the insurance sector contributed to over 80% of losses for a total of more than NZ\$31 billion. Amongst this, over NZ\$11 billion of the losses arose from residential building claims and were covered either partially or entirely from the NZ government backed Earthquake Commission (EQC) cover insurance scheme. In the process of resolving the claims, EQC collected detailed financial loss data, post-event observations and building characteristics for each of the approximately 434,000 claims lodged following the Canterbury Earthquake sequence (CES). Added to this, the active NZ earthquake engineering community treated the event as a large scale outdoor experiment and collected extensive data on the ground shaking levels, soil conditions, and liquefaction occurrence throughout wider Christchurch. This paper discusses the necessary data preparation process preceding the development of a machine learning seismic loss model. The process draws heavily upon using Geographic Information System (GIS) techniques to aggregate relevant information from multiple databases interpolating data between categories and converting data between continuous and categorical forms. Subsequently, the database is processed, and a residential seismic loss prediction model is developed using machine learning. The aim is to develop a 'grey-box' model enabling human interpretability of the decision steps.

Keywords: Seismic loss · Christchurch earthquake sequence · Data aggregation using GIS

1 Background

1.1 The Christchurch Earthquake Sequence

In 2010–2011 New Zealand suffered the costliest natural disaster of its history with a series of earthquakes known as the Canterbury Earthquake sequence

© Springer Nature Switzerland AG 2020

P. Cellier and K. Driessens (Eds.): ECML PKDD 2019 Workshops, CCIS 1168, pp. 88–100, 2020.

https://doi.org/10.1007/978-3-030-43887-6_8

(CES). The CES led to 182 fatalities and extensive building damage across the region, with over NZ\$50 billion of economic losses accounting for 20% of New Zealand's GDP [1,24]. The CES began on 4 September 2010 with the Mw 7.1 Darfield earthquake. The Darfield earthquake was centered approximately 40 km west of Christchurch Central Business District (CBD) [12]. It affected mainly unreinforced masonry buildings, induced liquefaction in wider Christchurch and luckily, no lives were lost. In the next 15 months, the Canterbury region experienced numerous aftershocks with around 60 earthquakes above Mw 5 and hundreds over Mw 4, some of these such as the Mw 4.7 aftershock on 26 December 2010 resulted in further damage. Then on 22 February 2011 12.51 pm local time, a Mw 6.2 shallow aftershock occurred directly under Christchurch CBD at a depth of 5 km [13]. This was the most significant event in the CES. It happened near lunch time when office and street pedestrian occupancies were at their peaks. It caused collapses of unreinforced masonry buildings that were not already removed from earlier aftershocks, irrecoverably damaged to many mid-rise and high-rise buildings, and collapse of two notable concrete buildings that led to 135 of the total 182 human casualties in the event [18]. It also prompted liquefaction in Christchurch CBD and eastern residential areas which exacerbated building damage due to foundation displacement. Following this, there were a number of other aftershocks that led to further building damage. In total there were 11,200 aftershocks in the CES.

The CES highlighted a number of civil and earthquake engineering challenges, importance of liquefaction, short-term heightened seismicity, rock slope stability but also impacted the reconstruction and recovery [10]. An estimate of 70% of the Christchurch CBD was demolished or partly reconstructed. Significant parts of the CBD were cordoned off from public access for over 2 years from February 2011 until June 2013 [19]. The CES, being the fourth most costliest insurance event in history globally at the time, also extensively affected the local and global insurance sector regarding seismic building damage [20].

1.2 Seismic Insurance Following the Canterbury Earthquake Sequence

Many countries located near tectonic plate boundaries are exposed to frequent earthquakes. However, insurance uptake for geophysical events remains low (2% in Italy, 5% in Turkey, 9% to 11% in Japan, 10% in Mexico, 26% in Chile, 38% in US, and 80% in New Zealand [1]). New Zealand is an exception with an insurance penetration of 80% [1,20]. Over the two years of the CES, major earthquake events and multiple aftershocks led to 77 events for which more than 650,000 insurance claims have been lodged [17]. Apportionment of the losses by sector is as follow: 59% account for the residential sector and 41% for the commercial sector [2]. Most of the claims for residential buildings were lodged for the main events of the 4 September 2010 and 22 February 2011. However, it was difficult to assess the exact impact of each earthquake and aftershocks on buildings. As the time between the event was too short to permit detailed building assessments following each event, especially for such a large number of affected buildings. This

also led to significant legal challenges between claimants, insurers and reinsurers about the damage apportionment between events. Reports shows that 61% of the residential insurance claims were settled by the Earthquake Commission (EQC) and 39% by private insurers [2]. This distribution points the significant participation of EQC.

1.3 The Earthquake Commission

The Earthquake Commission (EQC) is a Crown entity which has for its mission to provide natural disaster insurance for residential property. EQC also manages the Natural Disaster Fund (NDF) and promotes research and education on solutions for reducing the impact of natural disasters. EQC involvement is particularly visible with the EQC insurance EQCover [5]. EQCover provides home and land insurance for natural disaster for every home that is covered by private fire insurance. At the time of the CES, EQC provided coverage for the first NZ\$100,000 + 15% Goods and Service Tax (GST) of the building damage, NZ\$20,000 + GST for contents and land damage up to the value of the damaged land (since 1 July 2019 the cap for residential building cover was increased to NZ\$150,000 but do not include the cover for contents anymore). EQC accessed the NDF and its reinsurance cover to settle the claims. Before the CES, the NDF had a value of NZ\$6.1 billion (more than US\$4 billion) though this has now been significantly depleted to less than NZ\$180 million following the CES and a smaller Kaikoura earthquake in 2016 [8, 11].

The CES brought major changes for New Zealand, especially for the insurance industry [16]. EQC increased the annual levy in order to replenish the NDF [4]. Owing to the largely unexpected losses for the private insurers since the CES, there had been a trend of increased scrutiny of the risk profile of any insurance cover. Private insurers are now currently applying risk-based premium pricing for earthquake covers. This had led to increased premiums and at times unavailability of earthquake insurance for some regions in New Zealand.

1.4 EQC's Catastrophe Loss Models

Loss models are important for the insurance and reinsurance sector for quantifying probable losses to ensure adequate provisions in case of a catastrophe. EQC similarly relies on hazard and loss models for adjusting base cover, investment and reinsurance strategies and general planning for response to natural catastrophe [23].

In early attempts to quantify the risk for New Zealand, EQC actuaries estimated possible annual claims from historical data, and probable earthquake intensities. With the evolution of individual computers in the 1980s, new modelling opportunities arose. EQC first employed a computer-based modeling software for loss simulation in 1993. In the past, EQC relied on two models that work in tandem: a system dynamics model (SDM) called 'Logjam' for the management of the claims and a hazard and financial risk management system called 'Minerva' [23]. EQC employed Minerva for estimating claims numbers and losses

following a major disaster, as well as for the predicting earthquake loss risk over 10 years in the future to design EQC levy structures and deductibles and to maintain the reserves in the NDF. Minerva relied on an internal database as well as external sources such as the EQC Building Costs or Aon Soils database (Fig. 1a). An earthquake loss subsystem which entails an attenuation and a vulnerability model combined to simulate the losses for any one earthquake event (Fig. 1b). Additionally, it has source models for New Zealand as well as 10-year portfolio models that enable to predict the loss frequency data. Outputs from these possible scenarios are stored in the Minerva database which can then be accessed by the financial management sub-system [27]. Nowadays, EQC works closely with reinsurance companies to ensure that New Zealand retains the necessary international support in case of a disaster [7]. EQC still uses Minerva as an impact estimation tool to predict likely losses for single events and one-year probabilistic analyses.

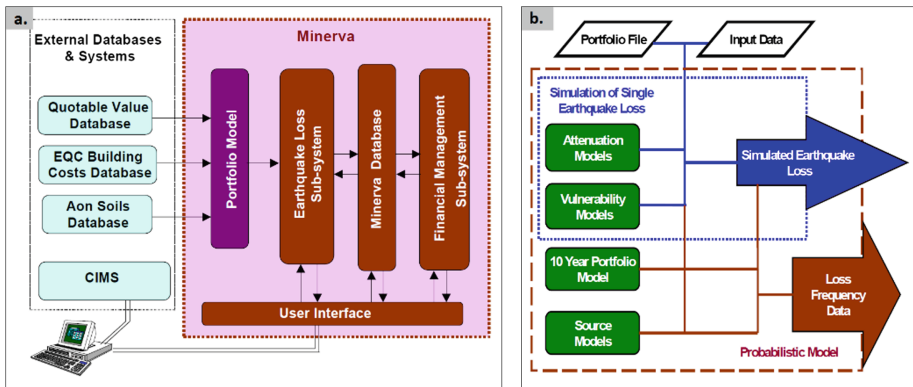


Fig. 1. (a) Overall Minerva system architecture, (b) Schematic diagram of the Earthquake Loss sub system used in Minerva [27]

Without minimizing the great improvement that these tools offered to the New Zealand insurance sector, limitations are still present. Since EQC offers natural disaster insurance for residential building on top of existing private insurance, EQC does not retain a database of its policyholders. It thus uses New Zealand records of real estate property as a base of its calculation [23]. This led to limitations regarding the accuracy of the exact loss prediction per asset. Moreover, the CES highlighted that the existing loss models did not accurately capture liquefaction. Additionally, the models usually took the building stock as undamaged at the time of the earthquake. But in the CES, the time between the events was too short such that the structures could not have been repaired or rebuilt. Cumulative damage occurred in reality but was not taken into account by the loss models [3].

1.5 Earthquake Commission Amendment Bill

On the 18 February 2019, the Earthquake Commission Amendment Bill 2018 (37-2) obtained royal assent [26]. The EQC Amendment Bill introduced changes including an increase in the time limit to lodge a claim following an earthquake event from three months to two years, the removal of the insurance cover for content, but an increase in the cap for the building cover from NZ\$100,000 to NZ\$150,000. At the same time, the bill brought revisions to the information sharing provision. EQC is now allowed to share information about the residential property claims, which have been lodged with EQC. Homeowners and prospective buyers can now ask EQC to provide them with information on residential property damage due to a natural disaster [6]. The bill also enables EQC to share information for public good purposes [26] which is favorable to the here presented project. While access to EQC’s property and claim database was granted since November 2017, difficulties arose due to anonymized building coordinates. Before March 2019, the latitude and longitude of each building in EQC’s property database were rounded to approximately 70 m to protect privacy. This led to the difficulty to relate each claim with a specific street address thus making impossible to merge EQC’s claim information with additional databases. The Earthquake Commission Amendment Bill 2018 (37-2) loosened the rules. EQC is now able to share the exact building location for each claim. This change in legislation enabled new opportunities for this research. The accurate building location enabled spatial joining and merging with new information on liquefaction, soil conditions, and building characteristics.

2 Developing a Loss Prediction Model Using EQC’s Residential Claim Database

2.1 Exploration of the Database

Following the changes brought by the 2019 Earthquake Commission Amendment bill, EQC provided access to the claim database for research purposes only. The exploration made in this paper uses the March 2019 version of the EQC claim database. Over 95% of the insurance claims for the CES have been settled by that time. However, revision of the event apportionment is still subjected to review meaning that the division of the cost between EQC and the private insurers can still change in future.

The EQC claim database is a wide dataset with 62 variables. It contains the relevant information related to the claims such as the date of the event, the opening and closing date of a claim, a unique property number, and the amount of the claim for the building, content and land. At the time of the CES in 2010–2011, EQC’s liability was capped to the first NZ\$100,000 (+GST) of building damage. Costs above this cap are borne by private insurers if building owner previously subscribed to adequate insurance coverage. Private insurance could not disclose information on private claim settlement, leaving the claim database

for this study soft-capped at NZ\$100,000 for properties with over NZ\$100,000 damage.

CES insurance claims are organized according to the event date when the damage is purported to have stemmed from. For the CES, the EQC database entails 77 different earthquake events. Figure 2 shows the number of claims against the 13 most significant events with more than 1,000 claims lodged. The two most significant events are the 4 September 2010 earthquake (145,000 claims) and 22 February 2011 aftershock (144,300 claims). Among the 62 variables, the database also includes building features. However, not all meta-data were collected in every instance and this led to incomplete data as highlighted in Fig. 3. The original EQC database has 85% of the values missing for critical features regarding the building characteristics (e.g. construction year, primary construction material, number of stories). Furthermore, the building characteristics may be subjective to individual assessor’s visual observation.

Number of claims per event in the Canterbury Earthquake Sequence

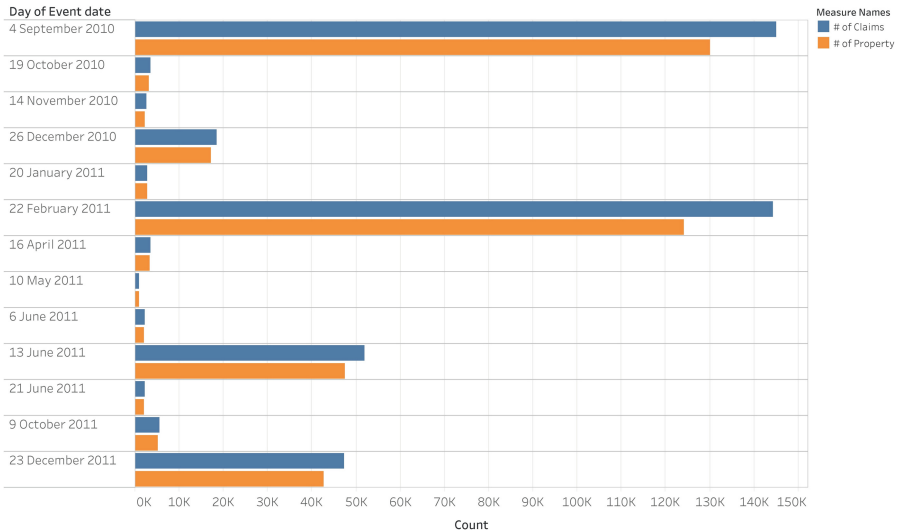


Fig. 2. Number of claims per event in the Canterbury Earthquake Sequence (Source: EQC database for claims on residential buildings)

2.2 Merging of Multiple Databases

To develop a loss prediction model using machine learning, it is necessary to overcome the limitations of missing data for key variables. This is addressed by combining information available in other sources. Figure 4 shows a schematic overview of the databases that are combined with the EQC database.

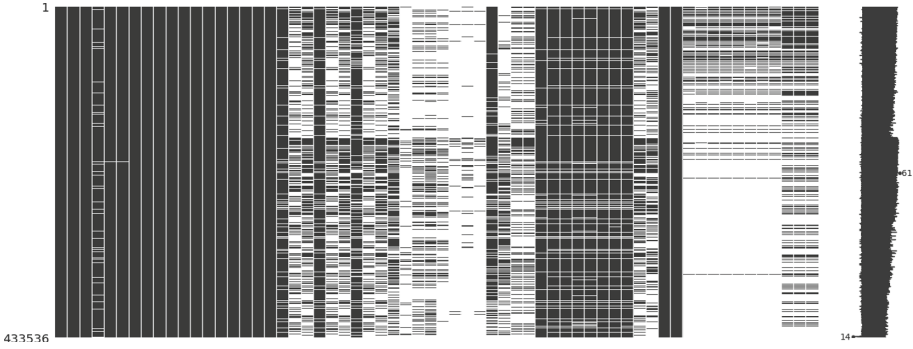


Fig. 3. Graphical overview of the data in the EQC claim database for the Canterbury Earthquake sequence. Each column represent a variable and each claim is a row. White areas represent missing values.

The RiskScape database [15] delivered critical information on buildings characteristics. It contains detailed information on the construction type, use category, building year, floor area, and deprivation index for every building in New Zealand. The Canterbury maps [25] and the New Zealand Geotechnical Database (NZGD) [9] provided records of the location and severity of liquefaction occurrence during CES based on interpretation of observations and LIDAR surveys. Land Information New Zealand (LINZ) [21] and Land Resource Information Systems (LRIS) [22] databases provided further topographical and soil conditions for the buildings of interest. Finally, the GeoNet [14] database provided strong motion seismograph recordings of all events in the CES as recorded at 14 recording stations located throughout Christchurch. This study focused on summary data such as peak ground acceleration (PGA), peak ground velocity (PGV) and peak ground displacement (PGD). This data enabled interpolation layers for all Christchurch to be created through the use of GIS software. Figure 5 presents an example of such an interpolated PGA map.

2.3 Challenges and Lessons Learned

During the process of merging the databases together, several challenges were encountered. These challenges occurred primarily due to the non-exact matching of the coordinates between the databases. Figure 6 shows the location of the EQC claims compared to the actual location of the buildings taken from the RiskScape database. From the map it is to see that the points from the two databases are not close to each other. Additionally, for some property, it can be observed that the EQC database entails two points meaning that multiple claims have been lodged throughout the CES.

As shown on Fig. 4, it was first attempted to join the EQC claim data with RiskScape information using a spatial join function implemented in GIS software. However, due to the distance between the points from EQC and RiskScape the

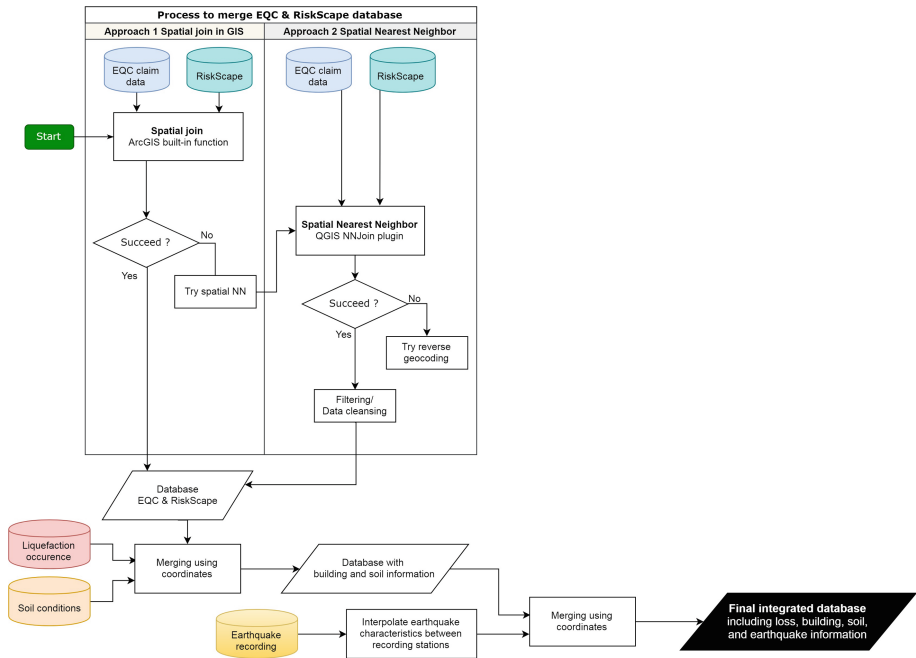


Fig. 4. Overview of the available databases and steps to the final integrated database

software was not able to successfully merge both databases together. It was thus decided to use a spatial nearest neighbor join (NNJoin) [28]. Nevertheless, the RiskScape database entails information for houses as well as secondary buildings such as garages and garden sheds. As shown on Fig. 7, multiple points might be present within the limits of one property tile. Thus, in certain cases the NNJoin led to the join of multiple buildings on one EQC claim. To reduce the number of buildings to the principal property it was not sufficient to filter the merged data by distance. Fortunately the RiskScape database includes information on the building footprint and floor area. It was then possible to select the principal house by filtering the data for each property title on the footprint area. However, it still left the possibility of neighbouring property being incorrectly joined up. To overcome this shortcoming, another approach applying reverse geocoding will be explored in future studies.

In its raw version, EQC's claim database is claim centric. This means one row of data corresponds to one claim, and the total damage to a property can consist of multiple claims or multiple rows of data filed at different dates, particularly due to the nature of multiple events in the CES. The combination of information with additional databases did not change the structure of the original EQC claim database. The final aggregated database retained a claim centric structure. The aim however, is to develop a machine learning model for the loss prediction on a building by building basis. It is thus necessary to have training data that contains

22 Feb 2011 Christchurch earthquake event

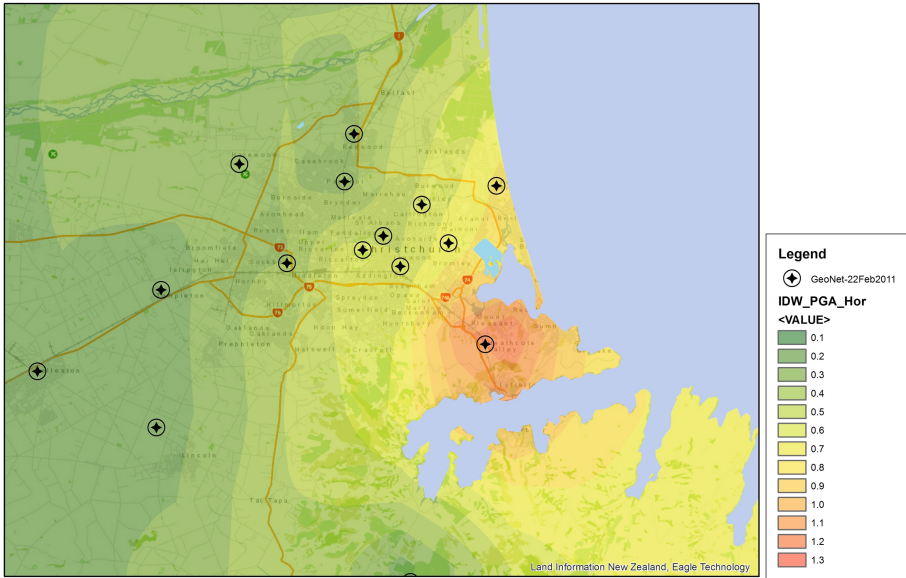


Fig. 5. Location of the GeoNet recording stations in Christchurch and interpolation of the PGA for the 22 February 2011 earthquake

22 Feb 2011 Christchurch earthquake event

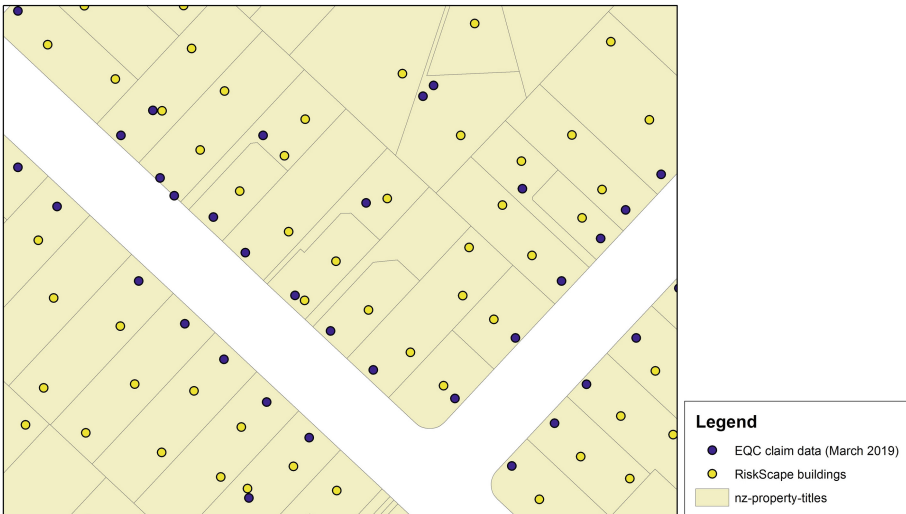


Fig. 6. Comparison of the spatial location of the EQC claim data (blue dots) and the building location from RiskScape database (yellow dots) (Color figure online)

22 Feb 2011 Christchurch earthquake event

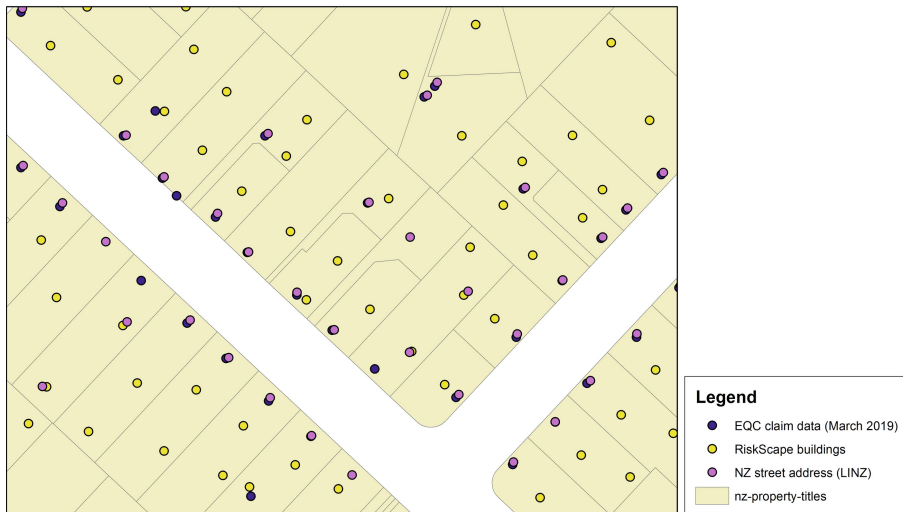


Fig. 7. Comparison of the spatial location of the EQC claim data (blue dots) and the location of NZ street address (pink dots) (Color figure online)

only one unique ID per property. This was achieved by pivoting the database to make it property centric.

3 Future Model Development Using Machine Learning

The combined database will be used as an input for the development of a seismic loss prediction model for residential building in New Zealand. The additional variables obtained through data integration enrich EQC's claim database. Machine learning is applied to process many variables and 'learn' from a large number of instances. Both the 4 September 2010 and 22 February 2011 events led to more than 140,000 claims each. This combined database constitutes the input of a machine learning model for seismic loss prediction.

In the development of the machine learning model, several algorithms such as linear regression, decision tree, support vector machine (SVM), and random forest will be applied. Their prediction accuracy will be compared and the algorithm leading to the most accurate prediction will be retained. The machine learning will be able to extract patterns from the integrated database and evaluate the relative importance of each variables. Nevertheless, particular attention will also be paid to human interpretability of the model. Whenever possible, intrinsically interpretable algorithms are preferred. More complex algorithms are always applied in combination with post hoc methods to allow for human interpretation. The aim is to develop a 'grey-box' model that would produce intermediate output, which allow modelers to look through and validate the

predictions at various key intermediate steps. A ‘grey-box model’ would allow different stakeholders to extract information that matters to them. For instance, a Civil Emergency Manager could be interested in the number of inhabitable dwellings, whilst an insurer might be interested in monetary repair cost only.

A loss model built on machine learning offers the advantage to be retrained easily. Whenever new data becomes available, it will be possible to iterate and improve the model accuracy. The possibility to retrain a model also offers the opportunity to test different parameters and their influences on the final losses.

4 Conclusion

This paper demonstrated the complex process of combining data from multiple sources using GIS. The data integration process focused on having extensive information for each property damaged during the CES. It merged information about the building characteristics, soil type, liquefaction occurrence and seismic demand on top of EQC’s claim database. It resulted in a aggregated database that can later be used to develop a seismic loss prediction model for New Zealand using machine learning. It allows for a future analysis of the relationship between variables that are usually not directly considered in a building loss analysis.

Acknowledgments. We acknowledge EQC for generously providing the claim data to realize this study. Many thanks to Geoffrey Spurr for his interpretation, assistance and review of the paper. We gratefully acknowledge the New Zealand Society for Earthquake Engineering (NZSEE) for the financial support. Thanks also goes to Dr. Sjoerd Van Ballegooy for his insightful advices.

References

1. Bevere, L., Balz, G.: Lessons from recent major earthquakes. Technical report, Swiss Reinsurance Company Ltd., Zurich, Switzerland (2012). <https://www.swissre.com/institute/library/Expertise-Publication-lessons-from-recent-major-earthquakes.html>
2. Deloitte Access Economics: Four years on: Insurance and the Canterbury Earthquakes. Technical report, February, Deloitte Access Economics (2015). <https://www.vero.co.nz/documents/newsroom/deloitte-vero-four-years-on-insurance-canterbury-earthquakes-report-february-2015.pdf>
3. Drayton, M.J., Verdon, C.L.: Consequences of the Canterbury earthquake sequence for insurance loss modelling. In: 2013 New Zealand Society for Earthquake Engineering Conference, pp. 1–7 (2013). http://db.nzsee.org.nz/2013/Paper_44.pdf
4. Earthquake Commission (EQC): Budget Announcement: EQC levy to increase (2017). <https://www.eqc.govt.nz/news/budget-announcement-eqc-levy-to-increase>
5. Earthquake Commission (EQC): EQC Insurance (2019). <https://www.eqc.govt.nz/what-we-do/eqc-insurance>
6. Earthquake Commission (EQC): EQC welcomes Act changes and gets ready to respond (2019). <https://www.eqc.govt.nz/news/eqc-welcomes-act-changes-and-gets-ready-to-respond>

7. Earthquake Commission (EQC): International reinsurers continue to provide cover to EQC (2019). <https://www.eqc.govt.nz/news/international-reinsurers-continue-to-provide-cover-to-eqc>
8. Earthquake Commission (EQC): The Natural Disaster Fund (2019). <https://www.eqc.govt.nz/about-eqc/our-role/ndf>
9. Earthquake Commission (EQC), Ministry of Business Innovation and Employment (MBIE), New Zealand Government: New Zealand Geotechnical Database (NZGD) (2012). <https://www.nzgd.org.nz/Default.aspx>
10. Elwood, K.J., et al.: Preface earthquake spectra. *Earthq. Spectra* **30**(1), fmi-fmix (2014). <https://doi.org/10.1193/8755-2930-30.1.fmi>
11. Feltham, C.: Insurance and reinsurance issues after the Canterbury earthquakes (2011). <https://www.parliament.nz/en/pb/research-papers/document/00PlibCIP161/insurance-and-reinsurance-after-canterbury-earthquakes>
12. GeoNet: M 7.2 Darfield (Canterbury), 4 September 2010 (2010). <https://www.geonet.org.nz/earthquake/story/3366146>
13. GeoNet: M 6.2 Christchurch, 22 February 2011 (2011). <https://www.geonet.org.nz/earthquake/3468575>
14. GeoNet: GeoNet strong-motion FTP site (2012). <ftp://ftp.geonet.org.nz/strong/processed/>
15. GNS (Geological and Nuclear Sciences Ltd.): RiskScape User Technical Documentation Wiki (2017). <https://wiki.riskscape.org.nz/index.php/Overview>
16. Greater Christchurch Group - Department of the Prime Minister and Cabinet: Whole of Government Report: Lessons from the Canterbury Earthquake Sequence. Tech. rep., Greater Christchurch Group, Department of the Prime Minister and Cabinet, Christchurch, New Zealand (2017). <https://www.dPMC.govt.nz/sites/default/files/2017-07/whole-of-government-report-lessons-from-the-canterbury-earthquake-sequence.pdf>
17. Insurance Council of New Zealand (ICNZ): Canterbury Earthquakes (2019). <https://www.icnz.org.nz/natural-disasters/canterbury-earthquakes/>
18. Kam, W.Y., Pampanin, S., Elwood, K.: Seismic performance of reinforced concrete buildings in the 22 February Christchurch (Lyttelton) earthquake. *Bull. N. Z. Soc. Earthq. Eng.* **44**(4), 239–278 (2011). <http://hdl.handle.net/10092/9006>
19. Kim, J.J., Elwood, K.J., Marquis, F., Chang, S.E.: Factors influencing post-earthquake decisions on buildings in Christchurch, New Zealand. *Earthq. Spectra* **33**(2), 623–640 (2017). <https://doi.org/10.1193/072516EQS120M>
20. King, A., Middleton, D., Brown, C., Johnston, D., Johal, S.: Insurance: its role in recovery from the 2010–2011 Canterbury earthquake sequence. *Earthq. Spectra* **30**(1), 475–491 (2014). <https://doi.org/10.1193/022813EQS058M>
21. Land Information New Zealand (LINZ): LINZ Data Services (2019). <https://data.linz.govt.nz/>
22. Land Resource Information Systems (LRIS): Christchurch City Soil Map (2010). <https://lris.scinfo.org.nz/layer/48148-christchurch-city-soil-map/data/>
23. Middleton, D.A.: EQC’s use of computer modelling in a catastrophe response. In: Proceedings of 2002 Annual Conference of the New Zealand Society for Earthquake Engineering (2002). <https://www.nzsee.org.nz/db/2002/Paper31.PDF>
24. Munich RE: NatCatSERVICE (2019). <http://natcatservice.munichre.com/>
25. New Zealand Government: Canterbury maps - Liquefaction Susceptibility - Final (2017). <http://opendata.canterburymaps.govt.nz/datasets/mapped-liquefaction-feb-2011>

26. New Zealand Parliament: Earthquake Commission Amendment Bill (2019). https://www.parliament.nz/en/pb/bills-and-laws/bills-proposed-laws/document/BILL_77657/earthquake-commission-amendment-bill
27. Shephard, R.B., Spurr, D.D., Walker, G.R.: The Earthquake Commission's earthquake insurance loss model. In: Proceedings of 2002 Annual Conference of the New Zealand Society for Earthquake Engineering (2002). <https://www.nzsee.org.nz/db/2002/Paper32.PDF>
28. Tveite, H.: QGIS NNJoin plugin (2019). <https://github.com/havatv/qgisnnjoinplugin>