Flood risk and flood insurance in New Zealand

Jacob Pastor
Victoria University of Wellington
jacob.pastorpaz@vuw.ac.nz

Ilan Noy
Victoria University of Wellington
ilan.noy@vuw.ac.nz

Isabelle Sin
Motu
isabelle.sin@motu.org.nz

ABSTRACT

The standard framework for undertaking a risk assessment of a natural hazard involves analyzing the interaction of three components: Hazard data (in the form of maps), the elements exposed to the hazard (exposure), and measures of these elements' vulnerability (understood as the susceptibility to harm or damage). In New Zealand, national flood risk remains unquantified due to the absence of national flood inundation hazard map coverage. In this paper, we develop a methodology that aims to fill this gap by estimating instead the likelihood of a flood insurance claim for a stock of residential buildings. We estimate a non-linear limited-dependent variable model and using a set of fragility functions (also known as damage curves), we calculate the expected monetary losses under plausible flood depth scenarios. The outcome of this research could inform insurers of their potential liabilities and threats to their financial sustainability in the face of flooding events and storms.

Keywords: Flood risk, insurance, liabilities.

INTRODUCTION

Flood hazard maps are an essential input to undertake a risk assessment, which fundamentally consists of determining who or what can be affected by a flood (it could be people, buildings, land), estimating what is the degree or level of affection of the exposed elements could have (commonly measured in a scale between 0 and 1), and calculating how much money will this impacts –and its cascading effects will cost. Thus, the standard framework and conventional practice of a risk assessment involves analyzing the interaction of three components, i.e.: Hazard data (in the form of maps), elements at exposure, and measures of vulnerability (understood as a susceptibility to harm or damage) (NIWA, 2016). In New Zealand, national asset risk remains unquantified due to the absence of national flood inundation hazard map coverage based on realistic scenarios (Paulik, 2017). While it is true that local governments are legally required to produce hazard maps for land planning and risk management purposes (Resource Management Act, 2003), these maps are not necessarily available or accessible in appropriate formats. Moreover, there is not a national data catalog compiling and assembling 1D and 2D flood models in a single platform or spatial data infrastructure. Moreover, there is anecdotal information reporting that some of these hazard maps have been prepared by different consulting organizations with intellectual property data arrangements restricting access to the data.[1]

Traditional one dimensional (1D) and two dimensional (2D) flood modeling for multiple flood recurrence intervals require an enormous amount of inputs, and are developed sometimes with strong assumptions and simplifications of the real world due to their complexity. From a hydrological perspective flood modeling requires recorded gauge data, which is not necessarily available for all the important rivers or catchments, or if it

1 We understand that there is an initiative within the context of one of the twelve NZ National Challenges that aims to acquire from local governments appropriate one dimensional (1D) and two dimensional (2D) flood hazard maps for urban areas that identify flood prone land. Ultimately, the project aims to stitch the spatial data together to form a composite flood hazard map for identifying assets potentially at risk.

exists, the length of the time series is too short to estimate discharge and other hydrological parameters. Flood modeling also requires precipitation gauge measurements (collected by meteorological station networks), which sometimes do not cover the areas where the hazardous rivers actually pose a risk over urban settlements. From a hydraulic and civil engineering perspective, flood modelling requires information of the built-up infrastructure, such as: Culverts, storm water and pipe network, stop-banks, dams, roads, bridges. On top of this, data on the topography, land cover, land use, and soil data are fundamental to model how bodies of water behave. The process ends when the resulting flood maps are subjected to validation and calibration using historic records of the extent and intensity of previous floods. From a flood model's accuracy stand point, we have evidence of the mismatch between the flood model inundation area predictions -where insurance claims were expected to happen and the actual occurrence of the insurance claims. This mismatch between prediction and observed impact of weather-related events raises the question whether flood maps should be reliable and whether it is an investment worth doing given their ability to provide an estimate of the economic impact of a disaster.

We answer this question by providing a much cheaper and less resource-intensive alternative, as well as a solution to the absence of flood maps. In this paper, we develop an innovative methodology that aims to estimate the economic impact (direct losses) of floods by predicting the likelihood of an insurance claim. We do this by implementing a non-linear model using past weather-related claim data from New Zealand's public insurer (EQC), and a set of GIS (geospatial) data sets produced by Land Information New Zealand (LINZ), Landcare Research (LCR) and National Institute of Water and Atmospheric Research (NIWA). These datasets report on the physical (hydrological hydrographic, topographic, land cover, soil characteristics) and socio-economic (deprivation index, value of the house, building materials) characteristics of a set of building distributed in a region of New Zealand. Based on these likelihoods and a set of fragility functions, we estimate monetary losses under plausible flood depth scenarios. While it is true that our proposed model uses data that would also be used in a conventional flood modeling framework, we take a probabilistic approach rather than a physical one.

We showcase this methodology in a region that has been extensively and intensively impacted by extreme weather events, and displays the highest number of insurance claims in relation to the total asset stock of residential buildings. The results outputted by our model are statistically significant and consistent with what the theory predicts for most of our explanatory variables. Specifically, we see that slope and the hydrological characteristics of the soil (Flood Return Interval, Soil drainage, Permeability Profile and Profile Total Available Water) are significant. Moreover, post-estimation statistics and factual evidence seem to indicate the robustness of our model. Specifically, we assess the predictive power of our model using two post estimation tests. The first is a confusion matrix and a c-statistics value based on Receiver Operating Characteristics curve. The proposed methodology for flood risk assessment can be extended to other areas conditioned on the availability of geospatil data and a "balanced" relationship between the number of buildings with and without claims. to circumvent the problem of an unbalanced relationship, we suggest the creation of cluster of buildings with and without claims, where clusters can be manually created based on the geophysical features of the area of interest, or based on clustering algorithms such us the concave hull or the concave hull.

The remainder of this document is organized as follows. The first section of the paper describes the general traits of New Zealand's public risk transfer mechanism -the Earth Quake Commission (EQC) in the context of weather-related events. This section also analyses the spatial and temporal distribution of claims across New Zealand. The second section describes the covariates and inputs for the model, and the estimation method itself. Section three reports on the results, followed by a discussion. The last section concludes.

THE EARTHQUAKE COMMISSION (EQC)

The precursor of the current EarthQuake Commission (EQC) - the New Zealand's public insurer for natural disasters was established in 1945 under the name of the Earthquake and War Damage Commission, with the aim of providing compulsory war and earthquake insurance for all properties. In 1994, the Commission was reconstituted under the Earthquake Commission Act 1993 (Owen and Noy, 2017)**.** This Act contemplates insurance provision for any physical loss or damage occurring to residential buildings and contents[2] as a direct

2 Changes to the EQC Act have been announced, and the new Act will exclude contents coverage.

result of earthquakes, natural landslips, volcanic eruption, hydrothermal activity or tsunami. In the case of flood or storm events, there is provision for land damage.

In order to be entitled to the EQC cover, homeowners are required to have a private fire insurance contract. In New Zealand, the private insurers and the EQC operate under a dual scheme in the sense that they share obligations when it comes to operational matters and how they compensate homeowners (i.e. "ground - up" or “top-up") (NZ Treasury, 2015). The EQC premium is flat (as opposed to risk-based) and it costs fifteen cents for every 100 NZ$ insured, for a one-year period. Since the EQC building cover is capped to 100,000 NZ$, most residential buildings will have premiums of up to 150 NZ$. Contents are covered for an extra 30 NZ$ premium with an EQC cover of $20,000. Land is covered at no additional cost and has virtually no cap.[3]

Even though there is provision for land damage from flood or storm events for buildings with EQC cover, it is only certain land on which the residential building sits that has cover. Specifically, the EQC cover for land is limited to the land that is within the property boundary and includes: The land under the residential building and appurtenant structures; the land within 8 meters of the residential building and appurtenant structures; the land under or supporting the main access way from the boundary up to 60 meters from the house (but not the driveway surfacing). The EQC cover also provides some cover for: bridges and culverts located entirely within the areas referred to above; and some retaining walls that are necessary to support the residential building and appurtenant structures or insured land [4] (EQC Insurer’s Guide, 2016).

Regarding EQC’s monetary compensation, it pays the lesser of either: the cost to repair the damaged land (or in some cases the diminution in value of the land); or the value of the damaged land; or the value of 4000 square meters; or the value of the minimum-sized site allowed in the area where the damaged land is situated. Bridges, culverts, and retaining walls that support the residential building or insured land are covered by EQC for indemnity value, meaning that the valuation takes into account their age and condition. (EQC Insurer’s Guide, 2016). Since 1980, the EQC has received over half a million claims, where approximately 4% of them (21.241 claims) have been triggered by storms and floods (or landslips preceded by strong precipitation). The total expenditure from these events rises to approximately 300 million NZ$, which is equivalent to 3.22% of all the payouts ever made by the Commission since 1980 from all natural hazards insurance claims (see Table 1).

Table 1. EQC payouts from weather-related insurance claims.

Total number of claims received by the EQC 1980- 2017	21,241	
Total number of claims closed (to Oct - 2017)	18,194	
Total number of claims with complete information*	17,595	
Total amount paid for land damage**	$ 162,912,300	
Total amount paid for building damage**	$ 1,669,460	
Total amount paid for contents damage**	$ 74,674,460	

	Number of claims	Mean	S.D.
Paid amount –including zeros	17,595	$ 13,597.95	$ 41,965.37
Paid amount –excluding zeros	10,774	$ 22,206.80	$ 51,816.48

* This number includes only observations for which geospatial information was available, and observations from 2000 onwards are only considered due to data reliability considerations.
** Nominal values.

3 The Act provision for land cover with no cap implies growing liabilities for the Crown (tax payers) in the face of climate change and its effects on increased frequency and intensity of weather-related events.
4 It should be noted that EQC covers building and contents damage from landslips that are a result of weather-related events

In terms of the number of claims across time, there is no discernible trend in the evolution of the series (see figure 1a below), even though science predicts an increase in the frequency and severity of extreme events. Nevertheless, it should be noted that unsettled claims and recent events such as extra-tropical cyclones Fehi and Gita caused major flooding events in early 2018, and could possibly inflate the numbers substantially. Most of the claims match the seasonal changes in New Zealand, meaning that claims tend to rise at the beginning of the wet season (April) and decline towards the end of it (October) (see figure 1b).

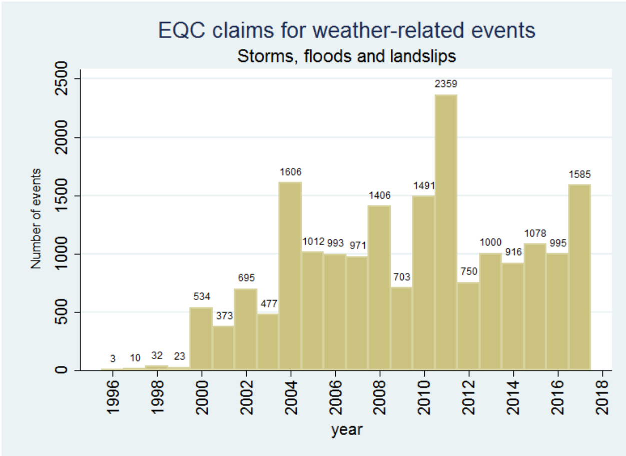


Figure 1a. Number of claims across years. This figure shows the annual number of weather-related claims (those classified as “Landslip/Storm/Flood”) made to New Zealand’s public natural hazard insurer (the New Zealand Earthquake Commission [EQC]) between Jan 1996 - Oct 2017.

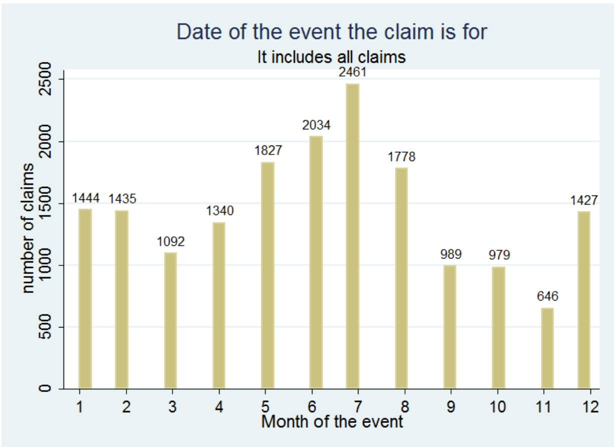


Figure 1b. Number of claims across months. Data excludes claims made in 2017.

In terms of the relationship between the total amount of payouts and the total number of claims, we see that large events dominate how many claims are made and how much is paid out. Figure 2 reports the number of claims on the right y-axis, and the total amount paid out on the left y-axis (in $NZ nominal), since the year 2000. The very first big spike of the series accounts for the "North Island Storm" in 2004, which meant for the EQC a total liability of 4.77 million (739 claims). The second highest spike is the "Bay of Plenty and Waikato Flooding" in 2005 with 12.3 million $NZ in losses triggered by 516 claims. The third highest peak in the series corresponds to the "North Island Weather Bomb" in 2008 with 581 claims and 13.29 $ NZ in losses. The remaining two highest peaks in the series reflect high damaging events that took place in 2011. The first was "extra-tropical cyclone Wilma", which triggered the highest amount of payouts ever made by the EQC as a consequence of a weather-related event with 13.96 million $ NZ in losses from 643 claims. The second event in 2011 was the "Tasman - Nelson Heavy rain and flooding", which in turn triggered the highest number of claims ever made to the EQC as a result of the impact of weather-related event, with 806 claims a losses for 13.02 million $NZ. Table 2 reports on the EQC top-10 most important events in relation to the size of the payouts made by the Commission resulting from floods, storms and landslips claims. It also reports on the amount that private insurers covered for the same events.

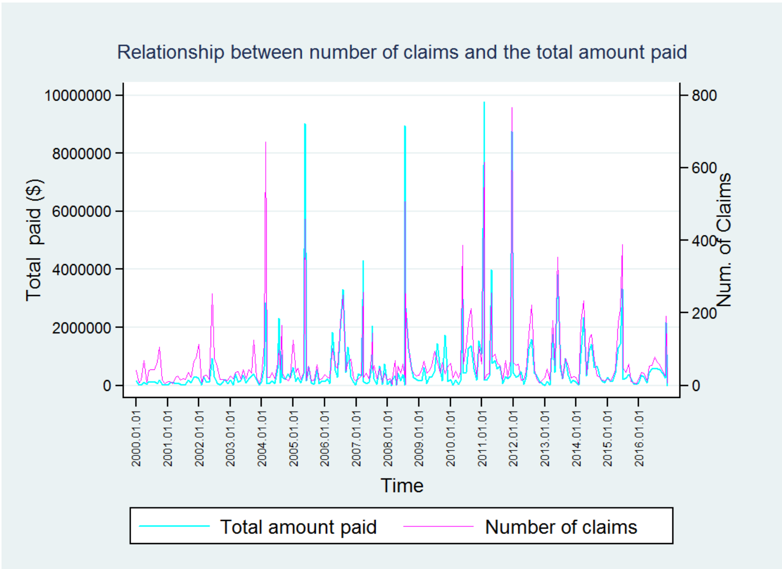


Figure 2. Total pay-outs and number of claims. This figure shows the number of claims and total pay-outs for weather-related claims made to New Zealand's public natural hazard insurer (the New Zealand Earthquake Commission [EQC]) from January 2000 to October 2017. Weather-related claims are those classified in the single category of "Landslip/Storm/Flood", which excludes earthquake-related landslips

An important attribute of the claims is the spatial dimension. Figure 3 shows the spatial distribution of the claims, where most of them are concentrated in urban centers of the most populated cities in New Zealand such as Auckland, Wellington, and Christchurch. It is noteworthy that a good deal of the claims affected the North Island. This could be explained by the higher exposure that this part of the country has to extra-tropical cyclones relative to the more southern areas. Nevertheless, there are areas in the South Island that, despite being less populated display a high concentration of claims (e.g. Bay of plenty, Abel Tasman).

DATA and METHODS

Data

The sources, datasets and the processes to construct a set of variables to feed our non-linear model are described in the following paragraphs. Most of the data are publicly available, but insurance data and residential asset inventory have been provided under data-sharing agreements with the Earthquake Commission (EQC), and CoreLogic and the National Institute of Water and Atmospheric Research (NIWA), respectively.

Insurance (Earthquake Commission -EQC)

We use variables that account for the temporal distribution of claims, their economic importance, the nature of the event and the exposure to previous events. The temporal dimension refers to the date of the event the claim is made for. The economic dimension is the actual payment or compensation in $ NZ. The dataset contains a series of variables reflecting monetary liabilities such "assessment of the repair cost", "total valuation of amount damaged", and "total remediated", among others. However, based on the recommendation of the data collector (EQC), we have restricted our attention to actual monetary payouts for the concept of land damage i.e. "land paid" variable.

Table 2. Public and Private insurance pay-outs for the most costly weather-related event in NZ

Date of event (1)	Name of event - characteristics (2)	Number of claims (3)	Total value of EQC claims (mill. $NZD) (4)	Paid by private insurance (mill. $NZD)
2005.05.18	Bay of Plenty and Waikato Flooding – heavy rain	795	21.4	28.5
2008.07.26	North Island Weather Bomb - high winds, seas and rainfall in several regions of the country	890	15.9	26.7
2011.04.25	Hawke's Bay Flooding – four days of heavy rain	429	15.7	6.4
2011.12.14	Tasman - Nelson Heavy Rain and Flooding	964	15.6	16.8
2011.01.29	Ex-tropical Cyclone Wilma - two days of heavy rain affecting the north of the country	815	15.0	19.8
2007.03.29	Northland Flooding – three days of heavy rain	630	9.4	12.5
2017.03.07	North Island Heavy Rain and Flooding - seven days of heavy rain	525	6.1	61.7
2016.11.10	Lower North Island flooding/wind	461	5.4	9.1
2007.07.09	Upper North Island Flooding and High Winds – three days of heavy rain	323	5.4	68.6
2004.02.16	North Island Storm - six days of heavy rain	1329	5.0	112
2015.06.20	New Zealand Storm – one week of intense rain in western areas of the South and North Islands	440	5.0	41.5

This table contains information on the weather events in New Zealand between 2000 and 2017 which led to the highest total pay-out from New Zealand's public insurer (the Earthquake Commission (EQC)). Column (1) contains date information in YYYY.MM.DD form for the first day of the weather event. Column (2) contains the name and characteristics reported in the NZ Historic Weather Events Catalog (NIWA 2018a). Column (3) contains the count of EQC "landslip/flood/storm" claims which are linked to an event matching the date in column (1). Column (4) contains the sum of EQC claim pay-outs expressed in inflation and GST adjusted 2017 NZ dollar values, and rounded to the nearest hundred thousand. Note the total number of EQC claims and EQC pay-outs are only lower-bound figures - these show the values linked to a single day. Column (5) contains information from the Insurance Council of New Zealand (2018) for the amount paid by NZ private insurance following the full weather event.

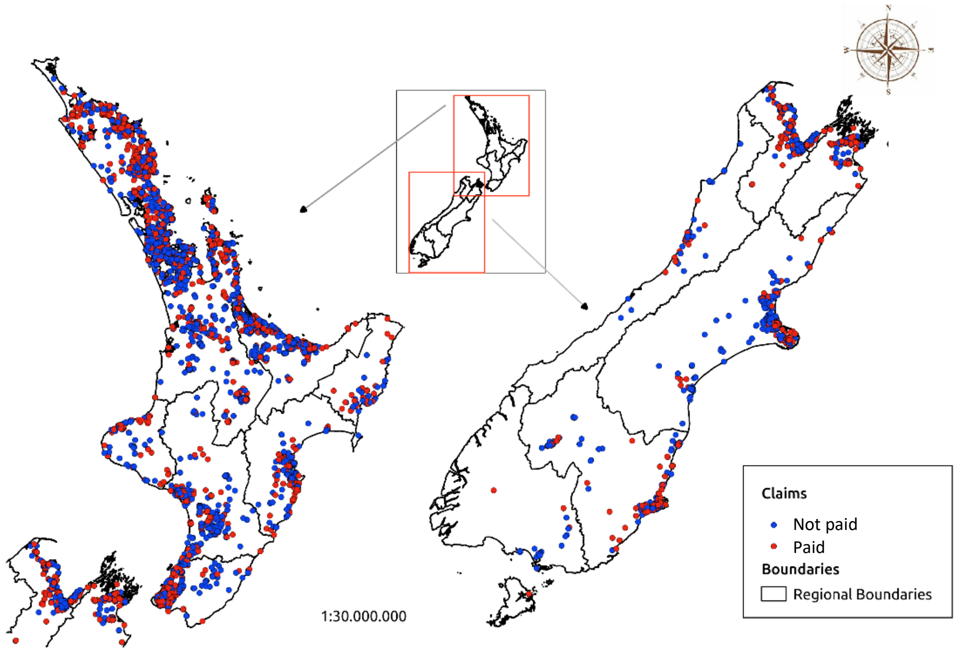


Figure 3. Spatial distribution of EQC claims across NZ (paid and not paid). The reasons for non paid claims include: absence of private insurance, claims referred to damages already assessed and covered, damage outside EQC's limit of 8 meters around the covered buildings, damage cost under excess ($500 for land damages), etc.

With regard to the nature of the event, we implement an algorithm using the “claims status" variable and the “EQC coverage definitions" -outlined in the EQCover Insurer's Guide (2016) to distinguish between claims that have been lodged as a result of a flood or storm, from claims that have been made as a result of a landslip. Specifically, if a claim has not received a payout for building or contents damage, but has an open or close claim for land damage, we classify that claim as a "Flood - Storm" claim. This is done on the basis that weather-related insurance claims are collected/registered/stored under a single category i.e. "Flood/Storm/Land/Slip". Note that we consider all flood claims regardless of the triggering mechanism i.e. coastal inundation, storm surge, riverine inundation, flash flood, or even extra-tropical cyclones. Finally, the exposure to previous events is calculated as the number of claims ever made to the EQC.
To account for the spatial dimension of residential building insurance claims, which are expressed in geographical coordinates (latitude and longitude), we join the EQC dataset with the CoreLogic dataset using the key variable "portfolioid.

Precipitation (National Institute of Water and Atmospheric Research -NIWA)

We use precipitation as one of our main predictors in the likelihood of a claim since rain fall is the driving force behind flooding events (DHI, 2018). Precipitation data is estimated through the Virtual Climate Station Network (VCSN), which outputs data estimates of climatic variables such as daily rainfall, potential evapotranspiration, relative humidity, wind speed, soil moisture among others. The data are provided at a ~5km spatial resolution and gridded structure covering the whole of New Zealand. "The estimates are produced every day, based on the spatial interpolation of actual data observations made at climate stations located around the country." (NIWA, 2017). We have computed the accumulated precipitation for a time frame that includes five days before the date of the claim. This measure aims to capture the intensity of precipitation that can presumably translate into a flooding event. Moreover, by having a wide window before the claim we account for rain and the possibility it saturates the soil it falls on (Abba Sood, 2017). We have attached the accumulated precipitation values to the buildings using a geospatial algorithm that extracts grid attributes and attaches them to the building point features within the grid.

Soil characteristics (Land Care Research -LCR)

We use a series of geospatial dataset (polygons) that report on the soil's water-related characteristics that we expect will have an impact on in the likelihood of an insurance claim. These layers are: Annual Water Deficit, Flood Return Interval, Soil drainage, Permeability Profile and Profile Total Available Water. These sets of geospatial data have been modelled based on the soil's physical and chemical properties, and a various sets of environmental variables for the whole of New Zealand. These datasets cover mostly country side (rural) areas and have minor coverage or urban areas. Each of the characteristics of the soil -either expressed as categories or continuous value have been transferred to the residential buildings sitting on them by using a spatial algorithm that "adds polygon attributes to point features".

Annual water deficit**.** This layer is used as an indicator of soil dryness. "The data layer was derived from surfaces fitted to monthly data describing daily average temperature, daily solar radiation and monthly rainfall. The rainfall surface was fitted using elevation, and a model describing relationships between topography and westerly winds"(LENZ, 2002). The units for this layer are in mm, higher values are areas that have a larger deficit.

Flood Return Interval**.** This layers contains flood return interval (FRI) categories that go from Nil to Very severe. The description of each category is expressed as an Annual Exceedance Probability (AEP), for instance a 1 in 60 year for a moderate FRI, or 1 in 20–1 in 60 for a moderate FRI, etc. A *1 in a "x" year event* is a measure commonly used among to describe the rarity of an event. For instance, a 1% AEP flood is the flood that has a 1% chance of occurring or being exceeded every year, and is sometimes known as the 1 in 100 year flood. Therefore, low probability events are associated with extreme and more damaging events. The FRI categories are as follows:

Table 3. Flood Return Interval categories and values

Description	Flood return interval (years)
Nil	Nil
Slight	<1 in 60
Moderate	1 in 20–1 in 60
Moderately severe	1 in 10–1 in 20
Severe	1 in 5–1 in 10
Very severe	>1 in 5

Soil drainage**.** "Drainage classes are assessed using criteria of soil depth and duration of water tables inferred from soil colours and mottles(spots or streaks)... or from reference to diagnostic horizons... Drainage classes used here are the same as those used in the NZ Soil Classification (Hewitt 1993), and outlined by Milne *et al.* (1995)" (LNZRI, 2008). The drainage classes with their descriptions are as follows:

Table 4. Drain class values and categories

Drain class	Description
1	Very poor
2	Poor
3	Imperfect
4	Moderately well
5	Well

Permeability profile. "Permeability is the rate that water moves through saturated soil. The permeability of a soil profile is related to potential rooting depth, depth to a slowly permeable horizon and internal soil drainage. Permeability classes are from Clayden and Webb (1994)" (LNZRI, 2008). Permeability values and their description are as follows:

Table 5. Permeability values and categories

Permeability	Description
S	Slow
M	Moderate
R	Rapid
NA	Not applicable

Profile total available water**.** **"**It is a classification of profile total available water for the soil profile to a depth of 0.9 m, or to the potential rooting depth (whichever is the lesser). Values are weighted averages over the specified profile section (0–0.9 m) and are expressed in units of mm of water. The classes originate from the work of Gradwell and Birrell (1979), Wilson and Giltrap (1982) and Griffiths (1985), and are described more fully in Webb and Wilson (1995)" (LNZRI, 2008). Profile total available water classes and their corresponding values are as follows:

Table 6. Profile total available water values and categories

Class	Description
1	Very high
2	High
3	Moderately high
4	Moderate
5	Low
6	Very low

Topography, hydrography, and land cover (Land Information New Zealand -LINZ)
We use a set of geospatial datasets produced under the 1:50k Topographic Map Series, which provide information on topography, hydrography and land cover of whole of New Zealand.

Elevation. We use a digital elevation model (DEM) to obtain the elevation of each residential building with respect to a vertical datum i.e. mean sea level. We obtain these values at building level by applying a spatial algorithm that extracts the DEM values and attaches them to the buildings.
Slope. We derive the slope of the terrain (in degrees) where the buildings sit on by applying a 10 parameter- 3rd order polynomial algorithm based on Haralick (1983).

Up until this point, all the previous variables have been constructed by using a spatial algorithm where residential buildings adopt the attribute of the area that contains them. In contrast, the following variables have been constructed using a distance-algorithm relationship, which measures the shortest distance from the residential buildings to different features e.g.: Coastline, water bodies (rivers, lakes) and forest cover..

<u>Hydrography</u>. We use three different sets of hydrographic features i.e.: big and small rivers, water bodies (lakes, ponds) and the coastline. Their inclusion in the model is motivated by the role these earth's surface features play in the hydrological cycle and because, they are an essential input in any flood modelling and risk assessment endeavor.
Distance to big and small rivers. Rivers are defined as "a natural, flowing body of water emptying into an ocean, lake or other body of water and usually fed along its course by converging tributaries." (LINZ topo50, 2013).
Distance to coastline. The coastline is defined as "The line forming the boundary between the land and sea, defined by mean high water."
Distance to lakes and ponds.They are defined as "any standing body of fresh inland water" (LINZ topo50, 2013)

<u>Land Cover</u>. We use three layers of vegetated land cover, i.e.: Exotic forest, native forest and scrub. Their inclusion in the model is driven by the known protective role that forests can have (Ferreira and Ghimire, 2012). We considered the inclusion of swamps and mangroves as they can also play a protective role during flooding events, but they are absent in our region of interest.
Distance to exotic forest and native forest. It is defined as "a tract of land covered by trees not native to New Zealand, or trees native to New Zealand" (LINZ topo50, 2013).
Distance to scrub. It is defined as "a tract of land covered by vegetation less that 3m high" (LINZ topo50, 2013).

Socio-Demographics (building inventory).
We leverage of the abundant number of building attributes available in the building inventory published by RiskScape[5]. The attributes include construction type (e.g. Concrete, Timber), year of construction, number of stories, floor height, replacement cost and the deprivation index. We link this inventory to the already merged georeferenced insurance claims dataset, using a "nearest-neighbor" spatial algorithm.

Estimation Method
We implement a non-linear limited-dependent variable model (logit) using past weather-related claims made to the New Zealand public insurer (EQC), and a the set of physical and socio-economic variables described in the data section. A non-linear logit model allows us to establish a relationship between a binary outcome variable (a claim) and a group of predictor variables (physical and socio economic variables), where the parameter values are estimated via maximum likelihood estimation (UCLA, 2017). The model has the following fucntional form:

$$\text{Logit}(p) = \log(p/(1-p)) = \beta_0 + \beta_1 * x_1 + \ldots + \beta_k * X_k$$

In terms of probabilities, the equation above is translated into

5 RiskScape is a software that was developed to meet the demand for natural hazard impact and loss modelling in New Zealand (RiskScape, 2017)

$$p= \exp(\beta_0 + \beta_1 * X_1 + \ldots + \beta_k * X_k)/(1+\exp(\beta_0 + \beta_1 * X_1 + \ldots + \beta_k * X_k)).$$

Our main interest is the predicted probabilities outputted by the model rather than the relative importance of a variable in the likelihood of a claim. This is because our ultimate goal is to estimate the expected monetary losses under plausible flood depth scenarios. It is noteworthy that most of our covariates are included in conventional hydraulic engineering modeling of floods.

All the independent variables i.e. $X_1, X_2, \ldots X_k$, have been spatially modeled as the *buildings' distance to rivers / native forests / shoreline, etc.* or as *buildings within drainage soil categories / buildings within permeability soil categories*, etc. All the categorical variables have been transformed into dummies and a comparison category has been selected by dropping it from the analysis. For the analysis, we use all observed flood claims regardless of the nature of the flood (coastal inundation, riverine flooding, flash flood, storm surge).

RESULTS AND DISCUSSION

Results

The overall model is statistically significant - given by the Prob > chi2 being less than the critical value 0. 05. The signs of the coefficients are aligned with what theory predicts and what we expected for the majority of our independent variables. In the following paragraphs we provide an interpretation for all the significant coefficients.

Soil characteristics. We find that buildings are more likely to make a claim if they are located in areas with "slight" and "moderate" *flood return intervals*, compared to buildings in areas with "Nil" flood return interval. When it comes to *drainage*, buildings are less likely to make a claim if they are located in "well" and "moderately-well" drainage areas, compared to buildings located in areas with "poor" drainage. Regarding the *permeability* of the soil, buildings sitting in areas with a "rapid" rate at which the water moves through saturated soil, are less likely to make a claim compared to buildings sitting on "slow" rate areas. When it comes to the total *water availability* of the soil, buildings within areas with "very low to moderate" and "moderately high" water availability are more likely to make a claim compared to the buildings in areas with "high to very high" total available water.

Hydrography. We find that the further the properties are from the coast line, the more likely they are to make a claim. This result is somewhat counter intuitive but we argue that this outcome could be the result of two seemingly contradictory clauses in the EQC Act (1993). Specifically, the Act states that erosion is not covered, whereas damage from storms and floods is covered . We argue that the EQC Act is not very sophisticated to make a conceptual distinction between the damage caused by floods and storms, and erosion. Based on this, we manage two hypothesis. First, people close to the coast are indeed experiencing damage, but they believe they are not entitled to the EQC cover (possibly because of the advice of their lawyer or private insurer). Second, the tidal range in Nelson is the largest in NZ (LINZ, 2017), which means that houses are less exposed to coastal inundation arising from storm surge (PCE, 2015). Regarding rivers, we find that the further the buildings are from big rivers, the less likely they are to make a claim, which is consistent with what we would expect. We also find that the larger the distance of a building to a lake, the more likely it is to make a claim. This is somewhat counterintuitive, but in the case of Nelson there is one large dammed-reservoir classified as lake. The conventional wisdom lead us to argue that the closer you are to a dam, the more managed the water flow is. Therefore, the further a building is from the dammed-reservoir, the more it is exposed to natural variation. Nevertheless, we believe that our model does not necessarily capture this relationship.

Topography. We find that slope is significant in the likelihood of making a claim, that is, the steeper the terrain the more likely it is for a building to lodge a claim.

Socio-demographics. We find that buildings within residential areas are less likely to make a claim compared to buildings within rural areas. The rationale behind this relationship is the nature of the EQC cover for land damage, which covers the land within an 8-meter buffer. It is therefore expected that residential areas have less land area exposed compared to rural areas. Regarding the deprivation-index, we find that people residing in buildings categorized as poor (or less well-off) are less likely to make a claim compared to the people living in buildings classified as the richest ones. We argue that poor households are less likely to own land. Moreover, we argue that the richest households are more likely to make claims as they are more likely to be informed on how to exercise their rights, or be able to access advice or information. Moreover, the wealthier could be more likely to make claims due their ability to construct houses in hazardous areas because of the amenities value (view), given that they can afford the costs associated to geotechnical and civil engineering studies. This the case in Nelson, where a good deal of claims occur in wealthy areas with steep slopes. Regarding the replacement cost of a building, we find that the higher the replacement value, the more likely the building is to have a claim. The logic behind this finding is that buildings with higher replacement values are more likely to have a larger building footprint, which means that more land area is exposed -even when the 8-meter buffer is fixed.

Exposure to previous events. We find that buildings that have made previous claims are more likely to make a claim, compared to buildings without previous insurance claims. This result reveals higher risk -either explained by exposure or vulnerability. It could also be explained by the learning experience of the having put a claim through.

Logit post-estimation

To assess the predictive power of our model we ran two post estimation tests. The first is a confusion matrix reporting on the sensitivity, specificity, errors of the classification (Error type I and Error type II), and the overall accuracy. The second post estimation test also accounts for the performance of the classification model by outputting a c-statistic value based on Receiver Operating Characteristics Curve.
A confusion matrix accounts for the performance of a classification model. The final model reports an overall accuracy of 98.08%; a sensitivity value (or true positive rate) of 11.20%; a specificity value (or true negative rate) of 99.86%; an Error Type II (positives that have been classified as negatives) of 88.80%; and an Error Type I (negatives that have been classified as positives) of 0.14%.[6] However, the overall accuracy value can be misleading as it is sensitive to the distribution of the "true" values in the data, which is our case for the "true negatives". Moreover, if we were to improve our sensitivity values by trying different cutoff values, we will be trading off specificity to get more sensitivity (e.g. 0.1, 0.05). Ideally, a specific cutoff "should yield a combination of sensitivity and specificity that is on balance.

In order to avoid the arbitrariness of setting a classification cutoff to obtain better performance values in confusion matrix, we use a Line Receiver Operating Characteristics curve and the underlying c-value, which "is a single statistic that summarizes the ability of the test to distinguish cases from non-cases and it inherently takes into account all possible sensitivity-specificity trade-offs." (StataListForum, 2016). A model with no predictive power has c-value of 0.5 and a perfect model has c-value of 1. "A model with no predictive power would be a 45 degree line. The greater the predictive power, the more bowed the curve, and hence the area beneath the curve is used as a measure of the predictive power (Stata, 13).

We ran four different specifications in order to identify the one outputting the highest c-statistic value. The first specification includes all variables (es described in the Data section), the second specification includes only soil, land cover and hydrographic variables (geophysical variables), the third specification includes only socio economic variables. The last is a parsimonious specification that includes only the variables that turned out to be significant in previous three specifications. The results of the post estimation test show that the best specification is the first one -the full model with a c-statistic value of c=0.8310. See Figure 6 below.

6 False positive (Error Type I): Positive prediction, but in reality a negative result. False negative (Error Type II): Negative prediction, but in reality a positive result.

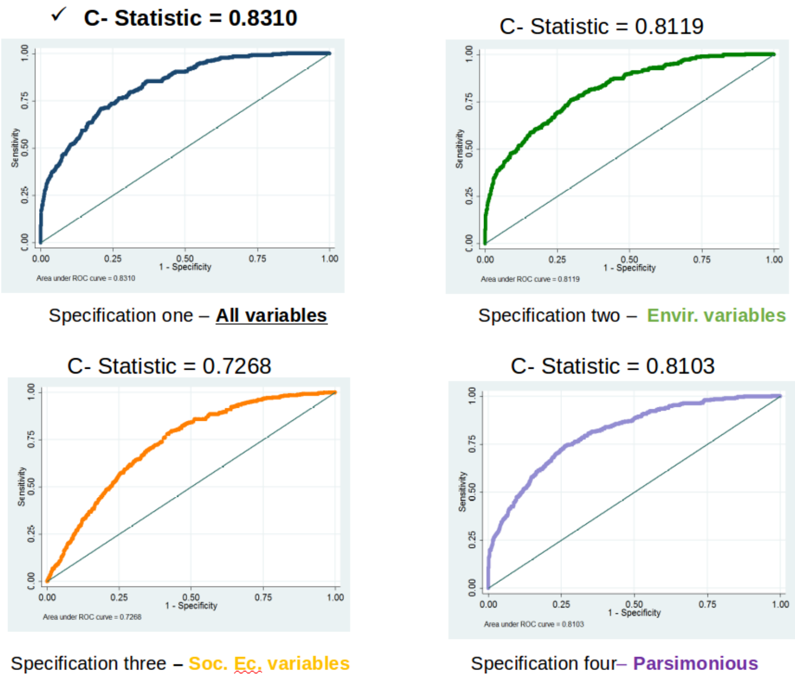


Figure 6. Predictive power of four different specifications using c - statistic value. The specifications are: All variables, Environmental variables, Socio-economic variables, and a parsimonious specification which only contains the variables that turned out significant in either of the three previous specifications.

Table 7. Logistic regression results

	VARIABLES	(1) Log odds - Specification 1	(2) Odds Ratio Specification 1
Socio-Economic	Floor Height above 0.61cm	0.31366***	1.36842***
		(0.11669)	(0.15968)
	Deprivation Index 5	-1.26992*	0.28086*
		(0.71248)	(0.20010)
	Deprivation Index 6	-1.63552**	0.19485**
		(0.64364)	(0.12541)
	Deprivation Index 7	-1.37574*	0.25265*
		(0.71271)	(0.18007)
	Deprivation Index 9	-2.07332**	0.12577**
		(0.81164)	(0.10208)
	Previous claims (exposure to previous events)	4.11741***	61.39970***
		(0.28487)	(17.49109)
Topography	Slope	0.06301***	1.06504***
		(0.01176)	(0.01252)
Soil Characteristics	Flood Return Interval - Slight (<1 in 60)	1.07746**	2.93722**
		(0.54698)	(1.60659)
	Flood Return Interval - Moderate (1 in 20-1 in 60)	1.40624**	4.08058**
		(0.58843)	(2.40115)
	Soil Drainage - Moderately well	-2.30879**	0.09938**
		(1.04831)	(0.10418)
	Soil Drainage - Well	-1.52828*	0.21691*
		(0.78705)	(0.17072)
	Permeability profile - Moderate to (Slow over Rapid)	-0.85907***	0.42356***
		(0.30578)	(0.12951)
	Profile available water - Moderately High	2.57105***	13.07959***
		(0.96948)	(12.68037)
	Profile available water - Moderate / Very Low	1.75769*	5.79902*
		(0.91514)	(5.30694)
	Within residential areas	-0.62785***	0.53374***
		(0.16855)	(0.08996)
	Constant	-3.56767***	0.02822***
		(1.29793)	(0.03663)
	Observations	18,352	18,352

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Only statistically significant coefficients at 0.1, 0.05 and 0.001 p-values are shown for simplicity of presentation. The first column indicates the log-odds value and column two indicates the odds ration value for the first specification, which includes all the set of independent variables described in the data section.

Expected losses

Given that we have confirmed the statistical significance and predictive power of our model, we now proceed to calculate the expected losses using the following inputs: the predicted probabilities (fitted values) of a flood insurance claim, the building's replacement value, and their susceptibility to harm or damage under different flood depth scenarios. The damage that a building can experience in the event of a flood depends on the flood intensity (given by the water depth, water velocity, the duration of the inundation) and the building characteristics (floor materials, wall materials, number of storeys, etc.). Flood building damage -and damage in general can be expressed as the cost to repair, or as the ratio between cost of repair and the replacement value of the building. The relationship between damage-ratio and flood intensity is described by what is known as fragility functions, fragility curves or damage curves (Reese and Ramsay, 2010). These functions are a method to calculate the potential direct damage, and are derived from field observations (damage repair cost estimates from surveys after flooding events) or derived from experts' opinion (synthetic curves), where different flood depths are associated with different damage levels in scale zero to one, where zero means no damage and 1 represents that the damages exceed the replacement value of the building (NZIER, 2004).

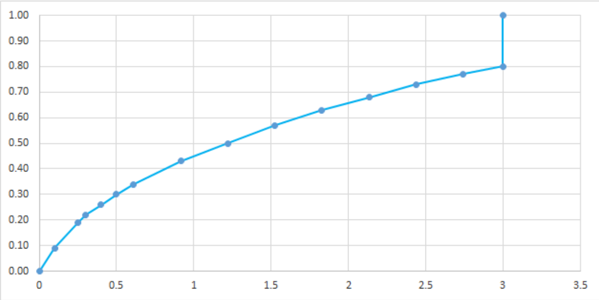
It should be noted that under the EQC cover scheme there is only provision for land damage resulting from floods or storms. However, we argue that regardless of the extent of the coverage of the EQC policy, the residential buildings that experienced land damage were also highly likely to have had some level of impact in their buildings. Moreover, we believe that the vulnerability of a building is much lower than the one of the land for the same flood depth. On these grounds, we estimate monetary expected losses resulting from building damage since either private insurers or homeowners will have to bear them.

In this paper, we implement a series of fragility functions that combine empirical and synthetic fragility curves that were estimated "through a review of international studies and consideration of how these studies apply to the housing attributes and flooding characteristics of New Zealand" (Reese and Ramsay, 2010). We restrict our attention to timber buildings only as they constitute approximately 91% of all the housing stock of Nelson. The fragility functions for timber buildings have been constructed considering the number of storeys (one or two) and the age of the building, where buildings constructed before 1960s and after 1980s are categorized as age-class one, and buildings constructed between the period 1960-1980 are categorized as age-class two. The distinction made in the years of construction is based on the materials that were most commonly used in those periods. For instance, age-class one buildings' floor type (pre-1960s and post-1980's) is mainly slab concrete, whereas age-class two buildings' floor type (1960-1980) is chipboard.

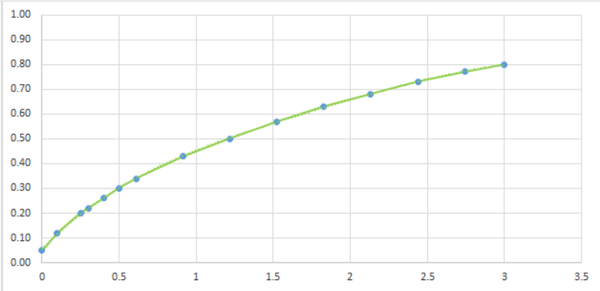
We implement four sets of fragility curves using data provided by NIWA and by fitting polynomial functions - as suggested by the data collectors (NIWA). It should be noted that these curves report the potential damage for different levels of flood-depths and not water velocity or duration of the inundation.

After fitting the fragility curves to the data, we investigated historical reports and records describing the extent of flooding events in Nelson. In particular we looked at any information that could account or proxy for the flood depths experienced in the Region. An important repository is the Catalog of Historic Weather Events published by NIWA, which collects and compiles data from government institutions, newspapers, databases, councils and a variety of sources. Based on the findings we decided to estimate the expected damage using precipitation values (in millimeters) and assuming these values as flood depths. This mainly because most of the reports -amalgamated in the catalog describe events in terms of total precipitation rather than observed flood depths. This said, we use two flood depths i.e. 0.25 meters and 0.67 meters, as lower and upper bound of all the events described as extreme or highly damaging[7]. Given the specified flood depth scenarios, we compute the expected damage for our set of fragility curves.

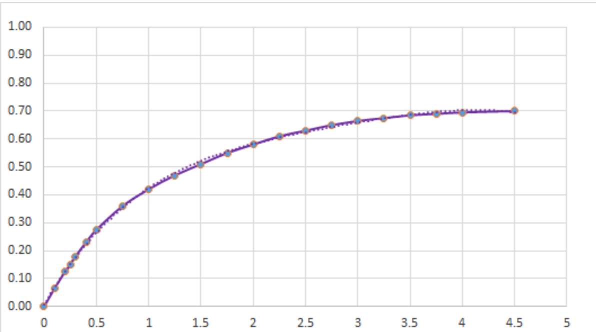
7 There is evidence of higher flood depth values - reaching up to 1.2meters for timber buildings in NZ, however, we did not find evidence for the area of Nelson and therefore kept our scenarios down to .67 meters as the maximum flood depth.



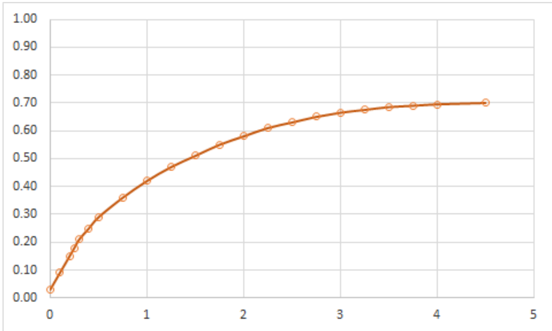
Timber - one storey - (pre-1960s and post-1980's)



Timber - one storey - (1960-1980)



Timber - two storey - (pre-1960s and post-1980's)



Timber - two storey - (1960-1980)

Figure 7. Fragility curves. The fragility curves were fitted through four-order polynomials using observed flood depth and damage stage for timber buildings with one and two story buildings.

At this stage we have all the inputs necessary to estimate the expected losses i.e.: Predicted probability of a claims, expected damage under a flood depth scenario and replacement value of the building. We therefore calculate expected losses by implementing the following formula

$$\Sigma^{n}{}_{i=1} = lc_i * dr_i * rv_i$$

where lc_i is the likelihood of a claim, dr_i is the damage ratio for a flood depth and rv_i is the replacement value. The expected monetary losses results are summarized in the tables below.

Table 8. Expected monetary losses for a flood depth scenario of 0.25 meters

Building type	Expected losses (in NZ $)
Timber, one storey - pre 1960 & post 1980	10, 138, 672.78
138Timber, one storey - between 1960 - 1980	1, 418, 220.05
Timber, two storey - pre 1960 & post 1980	7, 308, 319.68
Timber, two storey- between 1960 - 1980	2, 060, 810.48
Total losses	20, 926, 024

Table 9. Expected monetary losses for a flood depth scenario of 0.67 meters

Building type	Expected losses (in NZ $)
Timber, one storey - pre 1960 & post 1980	19, 125, 679.47
Timber, one storey - between 1960 - 1980	2, 630, 875.15
Timber, two storey - pre 1960 & post 1980	15, 877, 549.07
Timber, two storey- between 1960 - 1980	3, 998, 239.76
Total losses	41, 632, 344

Discussion

The proposed model fills a gap of knowledge and provides a probability-based assessment of flood risk as an alternative to the physical-based conventional assessment of flood risk, and also as a substitute of flood maps when these are not available. The estimation of expected monetary losses for insurers is fundamental for the actual decision to insure or not, determining premiums, excess, and informs the financial sustainability of the firms. Our model has included most of the variables -whenever available that hydraulic engineers input into their flood models.

Back Casting. In an effort to contrast our results with actual monetary losses experienced in the Nelson region, we refer to a dataset published by the Insurance Council of New Zealand (ICNZ). Specifically, they publish a list with all the "natural disasters that have occurred in New Zealand since 1968, and the cost to the insurance industry in paying claims for damage resulting from those events" (ICNZ, 1997). We also use the total EQC expenditure in the Nelson region from all weather-related claims.
The private insurers dataset reports two flooding events in Nelson alone, with a total expenditure of 16.89 million NZ$, and the EQC's figure rises up to 10.33 million NZ$ making up a total of 27.22 million in losses. The Insurance Council's dataset reports three other events that affected Nelson and simultaneously the areas of Plymouth, Bay of Plenty and Tasman with 50.15 million in losses. Similarly, our estimates for the two flood depth scenarios show expected losses between 20.92 and 41.63 million. Although the results from the model and factual data may not be fully comparable, the discrepancy between the two is not exorbitant.

Table 10. Weather-related events losses for the private insurance industry (private and public) for Nelson, and for Nelson and other areas

Year	Day and Month	Area	Cost ($ NZ) in million
2014	25 Jun	Nelson – Tasman floods	2.7
2013	19-22 Apr	Nelson/Bay of Plenty storm/floods	46.2
2011	15-16 Dec	Nelson floods	16.8
2007	23-May	Nelson/New Plymouth flooding	1.25
1998		Nelson floods	0.92

Applicability in other areas. The methodology can be extended to other areas mainly under two conditions: Availability of geospatial data, and a "balanced" relationship between the number of buildings with and without claims. The first condition is not always achievable as the production of these type of datasets is highly technical and expensive, particularly the ones referring the soil's characteristics (profile availability, return interval, drainage, etc). Moreover this methodology is likely to be implemented mostly by private insurers as New Zealand is quite an exceptional country in the sense that public insurance is ubiquitous with a penetration of about 95%. The second condition refers to the need of having a reasonable amount of buildings with claims with respect to the number of buildings without claims. In this paper we selected the Nelson region as it displayed this desired relationship -the highest ratio, which is not the case for the whole of New Zealand. To circumvent an unbalanced relationship, we suggest the creation of cluster of buildings with and without claims. This clusters can be manually created based on the geophysical features of the area of interest, or based on clustering algorithms such us the concave hull or the concave hull, which compute areas for a set of point in a plane.

Extreme events. The area of Nelson-Tasman is particularly more exposed to storms and floods compared to other southern areas due to its shape U shape and the paths of meteorological currents (Frame, 2018). In 2011, Nelson was impacted by unprecedented weather -a 1 in 500 year event that triggered about 60% of all the EQC claims ever made in the Nelson Region. A *1 in a 500* year event (or alternatively an event with a 0.2% probability) is a measure commonly used to describe the rarity of an event, where low probability events are associated with extreme and more damaging events.[8] Having said this, our predicted likelihoods may be most likely driven by this extreme event.

The resolution of the treatment variable. Precipitation is estimated at 5k grid structure and almost the entire stock of buildings falls within three grids, and consequently the variable does not display variation from building to building. We understand that the spatial resolution of the precipitation is too coarse and has to be downscaled, which is one of our next goals in refining the inputs of the model.

Omitted variables. Pipelines, water courses, drainage, flood gates, stop-banks, dams, storm water systems and any other variable related to built-up structures dedicated to water management were not included in the analysis since they were not available for the area of Nelson in the Topographic 1:50k Map series. These variables are essential to assess flood risk, however they are not always available and even conventional hydraulic flood models exclude them from the analyses. We believe that some of this variables have not been collected in the 1:50k maps due to the spatial resolution (scale) of the maps.

Private insurance claims. Our model has only used insurance claims made to the public insurer - EQC and has therefore not considered private insurance claims. We argue that accessing past claims made to private insurers could improve the estimated model and lead us to better classification outcomes and understanding of the risk of floods.

Temporal and spatial variation of the covariates. The covariates accounting for the physical characteristics of the landscape (soil characteristics, hydrography, topography and land cover) have been produced in different epochs -most of them in 2013, 2011 and 2002, whereas the claims refer to a period encompassing 1998-2017. We understand that the physical landscape is dynamic but in order to leverage of the information available, we assume that changes that have occurred in the physical covariates are negligible. In a further version of this paper we aim to support this assumption with credible figures about the steadiness of the physical changes in Nelson. Regarding the spatial scope of the soil characteristics and land cover variables, we have observed that most of these layers do not fully cover residential areas. Moreover, the areas with data are quite extensive and therefore can include vast amounts of buildings, which can affect the variability of our covariates. Despite this seemingly non-variation of the covariates we find significant evidence of their impact in the likelihood of an insurance claim.

Building asset inventories. Even though the building asset inventory has been essential for the accomplishment of this paper, we recognize that is not necessarily up to date. This was reflected during the matching process

8 Most flood maps are produced assuming a 1 in 100 year event, but in the Netherlands they have produced 1 in 1250 year flood maps, presumably as as a precautionary approach, and based on this they have built up their infrastructure (DHI, 2018).

between RiskScape and the insurance EQC-QV data set. Specifically, there were insurance claims data without any surrounding building in a buffer of 10km. We "flagged" these claims to obtain a quality measure of the spatial linking process by setting a distance threshold of 70 meters. The threshold value has based on the off-set QV has placed in the coordinates of the buildings as a mean to maintain the appropriate anonymity levels of the data. Regardless of this issue and the potential confounding effects this could have in our estimation, we argue that the number of observations displaying this problem is relatively low (46 observations). In a further version of this paper we will report the difference in estimation values when including and excluding these observations.

CONCLUSION

We showcase a methodology in a region in New Zealand that has been extensively and intensively impacted by extreme weather events. The proposed methodology circumvents the absence of flood hazard maps and leverages of georeferenced historic insurance claims and sets of geospatial data, to provide estimates of expected monetary losses under plausible flood depth scenarios. The results of the paper inform on what could be potential insurers' liabilities and possibly the threats to their financial sustainability. Moreover, it could inform the potential losses that uninsured homeowners would bear in the face of a flood event.

ACKNOWLEDGEMENTS

We express our gratitude to Ryan Paulik (NIWA) for facilitating access to the residential building inventory data, and the flood depth and damage observations for the fragility curves derivation. We also express our gratitude to Abba Sood for facilitating access to observed precipitation data and advising on how to process it in the context of our research purposes.

We thank all authors, program and local committee members, and volunteers for their hard work and contributions to the ISCRAM conference. The layout of this format was originally adapted by Bartel Van de Walle for the ISCRAM 2005 conference in Brussels from a workshop document template created by Ping Zhang. Thank you Ping for your kind permission to use your template. Slight changes have been made by Murray Turoff for ISCRAM 2006 in Newark NJ. Further changes were made by Mark Pfaff for ISCRAM2014, primarily to conform to Google Scholar indexing criteria. The references cited in this paper are included for illustrative purposes only.

REFERENCES

Danish Hydraulic Institute. (2018) "Flood modelling and forecasting masterclass", *Floods, slips and coastal erosion management conference*. Wellington – New Zealand

Frame, D. (2018). "Climate change: Counting the Cost". *Climate change: Counting the Cost Conference.* Wellington – New Zealand.

Earthquake Commission Act, 1993. Wellington – New Zealand

Earthquake Commission (2016). EQCover insurers' Guide, 2016.

Ferreira, S., & Ghimire, R. (2012). Forest cover, socioeconomics, and reported flood frequency in developing countries. *Water resources research*,*48*(8).

Insurance Council of New Zealand (2017). The Cost of Natural Disasters in New Zealand. https://www.icnz.org.nz/media-resources/natural-disasters/. Last accessed: 08/03/18; Last updated: 03/12/2018

Land Care Research New Zealand (2002). Land Environments of New Zealand: A Technical Guide.

Land Care Research New Zealand (2008). Land Resource Information System Spatial data layers, Data dictionary.

Land Information New Zealand (LINZ), Topographic maps series 1:50.000 - Data dictionary, 2013

National Institute of Water and Atmospheric Research (2017). Virtual Climate Stations Network (VCSN) https://www.niwa.co.nz/climate/our-services/virtual-climate-stations. Last accessed: 08/03/18; Last updated: Undefined.

National Institute of Water and Atmospheric Research (2017). Historic Weather Catalog. https://hwe.niwa.co.nz/. Last accessed: 08/03/18; Last updated: Undefined.

Noy, I., 2016. Tropical storms: the socio-economics of cyclones. *Nature Climate Change*, *6*(4), p.343.

NZIER. (2004) Economic im0pacts on New Zealand of Climate Change - related extreme events. Report to the New Zealand Climate Chage Office, 2004)

Owen, S. and Noy, I. The unfortunate regressivity of the public natural disaster insurance: Quantifying distributional implications of EQC building cover for New Zealand. 2017.

Parliamentary Commissioner for the Environment. (2015) Preparing New Zealand for rising seas: Certainty and Uncertainty

Reese, S. and Ramsay, D. RiskScape: Flood fragility methodology. 2010

Ryan Paulik - Deep South Challenge (2017), Emergent exposure of flood inundation hazards under future climate change in New Zealand. - 2017

Ryan Paulik - Deep South Challenge (2018), Interview/meeting on the 2 March, 2018.

Sood, A. Climate Scientist at the National Institute of Water and Atmospheric Research (NIWA). Interview on processing observed precipitation on the November, 2017.

Stata List Forum(2016). Interpreting classification of Logistic model. https://www.statalist.org/forums/forum/general-stata-discussion/general/1327821-interpretating-classification-of-logistic-model. Last accessed: 08/03/18; Last updated: 12/03/2018.

Stata 13ME Resources.