RISK FROM EXTREME PRECIPITATION AND CLIMATE CHANGE ON NEW ZEALAND'S RESIDENTIAL PROPERTY

By

JACOB PASTOR-PAZ

A thesis

Submitted to the Victoria University

of Wellington in fulfilment of the requirements

for the degree of Doctor of Philosophy

Victoria University of Wellington

2021

Most respectfully dedicated to my beloved parents

For their endless love, support, and encouragement

Table of contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I dedicate this doctoral thesis to my parents, Augusta Paz-Noboa and Antonio Pástor-Novillo for the gift of life, for sustaining me with their love and for inspiring me with their tenacity, intelligence, and wisdom to fructify in this world. This thesis is also dedicated to my virtuous and gifted siblings, Ana, José, and Fernanda.

I would like to thank my supervisor, Prof. Ilan Noy, for his guidance, experience, and his continuous and reliable support during my PhD studies. I am also forever grateful to Dr. Isabelle Sin, my secondary supervisor, for sharing her expertise and providing invaluable insights and advice. I would like to express my immense gratitude to Juan Pablo Erraez, Dr. Jane Drummond, Dr. Jose Marrero-Linares, Rosario Maldonado and Luis Guasgua for their mentorship during my academic journey.

I also thank the members of my dissertation committee: Dr. Yigit Salam, Dr. Dáithí Stone, and Dr. Junko Mochizuki, for reviewing and providing useful comments on this thesis.

I thankfully acknowledge the financial support received for my PhD. I am sincerely grateful to the Chair in the Economics of Disasters, VUW for the fully funded scholarship which had made this doctorate study possible. I also thank to QuakeCore - New Zealand Centre for Earthquake Resilience and Victoria University of Wellington for providing a PhD thesis submission scholarship. I want to acknowledge the National Secretaria of Science and Technology of Ecuador (SENESCYT) for the fully funded master's scholarship to undertake studies in the University of Glasgow.

I would like to thank to all the educators that generously poured science and ethics into my heart. I want to acknowledge the invaluable treasures learnt from former supervisors and managers, and the rich institutional legacy I inherit from the organizations I worked for. Special thanks to Pablo Cabezas, Carlos Cordova, Alfredo Intriago, Dr. Alejandro Medina-Giopp, Ramiro Pazmino, Prof. Ilan Noy, Dr. Olga Fillipova and Dr. Isabelle Sin, for giving me the opportunity to develop and enhance my professional and academic career.

I would like to express gratitude towards my friends and colleagues for their company and support during my doctorate. Special thanks to Teisa Moce, Peter McDonald, Samuel Garratt, David Pomeroy, Pascal Moreno, Fady Elkersh, Amy Qiao, Ilkin Huseynov, Liu Pengfei, Thomas Dudek and Merzan Wadia. I am especially grateful to Farnaz Pourzand for her unconditional friendship and support she gave me throughout my PhD journey.

Abstract

Three manuscripts form the basis of this dissertation exploring the effect of extreme precipitation and climate change on residential property in New Zealand. The first manuscript investigates the public insurer's expected future liabilities, given future climate projections. Specifically, it examines the effect of extreme precipitation on direct property damage associated with rainfall-induced landslides, storms and floods. This study applies a fixed-effects panel regression model using claim data linked to extreme precipitation data over 2000-2017 and future climate change scenarios until 2100. The results show that liabilities will increase more if future greenhouse gasses emissions are higher. At the aggregate level, the percent change between past and future liabilities ranges between an increase of 7 to 8% higher in the next 20 years, and an increase between 9 to 25% increase by the end of the century, depending on the greenhouse gases emissions scenario.

The second manuscript examines the risk of property damage from landslides associated with extreme precipitation. The focus is on the Nelson region as it displays the highest number of claims and pay-outs relative to its population and residential stock asset, and two thirds of the pay-outs come from a single event. The focus is on this event. This research combines past insurance claim data with geographic and sociodemographic data to estimate probability of damage, which is then combined with property replacement values and damage-ratio information to calculate the expected loses and map the spatial distribution of risk. The study integrates into the risk estimates the impact of climate change on precipitation based on an 'attribution' study. The analysis shows that slope and social deprivation play a significant role in the probability of damage. Furthermore, higher expected losses are associated with higher property values.

The third manuscript studies the current and future risk of property damage from floods associated with extreme precipitation and climate change. The focus is on the most expensive event on record. This study applies a logistic cross-sectional regression model that exploits spatial variation of rainfall intensity-duration-frequency (with and without the effect of climate change), while controlling for other factors that might make a property more or less likely to experience damage. The expected monetary losses are calculated by factoring in the likelihood of flood damage derived from the regression model, property replacement values, and property vulnerability (based on flood-depth fragility functions). The results show that highest losses are associated with lowest annual exceedance probabilities (AEPs), still, sizeable losses are associated with higher AEPs. In this case, the effect of climate change for different emissions scenarios is too small to cause an economically meaningful increase in risk levels in the next 80 years (2100).

Introduction

The standard framework to assess the risk from weather-related hazards and climate change involves analysing the interaction of three components: hazard data (with and without the effect of climate change), the elements exposed to the hazard, and measures of these elements' vulnerability -understood as the susceptibility to harm or damage- to a hazard of a given intensity, spatial scope, and temporal frequency (IPCC, 2012; UNDRR, 2016; Bouwer, 2013). In Aotearoa New Zealand (NZ), risk assessment of such hazards remains largely unquantified due to the absence of comprehensive, accessible, homogeneous, or accurate hazard maps.

In NZ, the public insurer (the Earthquake Commission – henceforth the EQC) provides cover against residential property damage that arises from some weather-related hazards, including floods, storms, and rainfall-induced landslides. Yet, even though the public sector is therefore bearing much of this risk, it is still unquantified.

In this dissertation, I develop a methodology that aims to fill this knowledge gap by using detailed data of past EQC insurance claims to assess the current and future risk for residential properties in NZ from various hazards associated with extreme precipitation. In particular, we focus on flooding and landslides. Unlike the conventional 'catastrophe modelling' framework that relies on hazard maps to calculate risk, I use extreme precipitation data as a surrogate for hazard information and link it to georeferenced claim-level data and other geospatial datasets to produce spatially explicit estimates of the current and future risk associated with extreme precipitation and climate change. The proposed methodology provides a statistically-based assessment of current and future risk as an alternative and complement to the physical-based conventional assessment, and as a substitute for hazard maps when these are not available. Furthermore, this dissertation is the first collection of studies that estimate the risk from extreme precipitation associated with climate change in New Zealand.

In this research, I aim to inform where, when, and how much monetary liability can be expected to be generated from various weather-related hazards associated with extreme precipitation and climatic change. Overall, this dissertation aims to answer three questions: What is the risk to residential property from extreme precipitation? What will be the effect of climate change, through changes in extreme precipitation, on residential property risk? What is the spatial and temporal distribution of risk from extreme precipitation and climate change?

The public insurer of New Zealand can use the outcomes of these studies to assess its future potential financial liabilities. The detailed spatial and temporal representation of risk can inform

tailored disaster risk reduction efforts, future risk pricing, and potential areas where managed retreat may occur. Private insurers can use this collection of studies to identify clusters or hotspots of weather-related risk and to design and target risk-based insurance schemes at various spatial scales (property and grid level) and future time periods (2020-2100). Local Councils could use these studies to inform their land use planning and zoning of areas prone to damage. Other exposed locations, currently not subject to weather-related hazards, could also employ this methodology to estimate potential liabilities. Furthermore, these studies could inform regulators and policymakers who are assessing the future performance of both the public and private insurers that cover weather-related risks in the face of climatic change.

In the first chapter of this dissertation, I look at the whole country to investigate the effects of extreme precipitation events and their increased frequency because of climate change on direct property damage. The region-specific studies for the second and third paper are motivated by the large observed EQC pay-outs from property damage in two specific events/hazards/locations. In the second paper I assess the current risk from landslides in Nelson, and the third paper examines the current and future risk from floods in the Bay of Plenty.

In chapter one, I investigate the effect of extreme precipitation (defined as daily rainfall events above the 95th, 98th and 99th percentiles) on the amount (in NZ$) of property damage from the impact of storms, floods, and rainfall-induced landslides. I implement a fixed-effects panel regression framework using EQC claim data (for the years 2000 to 2017) linked to grid-level observed rainfall data. I exploit variation over space and time while controlling for spatial and temporal unobserved heterogeneity with fixed-effects panel regressions. Then, I use the estimated relationship, together with climate projections based on future greenhouse gases concentration scenarios from six different dynamically downscaled Regional Climate Models, to predict the impact of future extreme precipitation events on EQC liabilities for different time horizons up to the year 2100. The analysis shows predicted adverse impacts that vary over time and space. The change between projected and past damages—the climate change signal—ranges between an increase of 7% to 8% in liabilities for the period 2020 to 2040, to between 9% and 25% higher for the period 2080 to 2100, based on one day of accumulated precipitation for the 99^{th} percentile. I find that exposure plays an important role in the level of observed property damages. Locations (grids) with damages are closer to the shoreline and waterways (rivers); are within flood prone areas and lower elevations; and have higher number and value of residential assets than locations without damages.

The second chapter quantifies the risk of property damage from extreme rainfall-induced landslides. When comparing NZ's regions, Nelson has the highest number of EQC claims and pay-outs relative to the size of its population and stock of residential assets. Two thirds of the pay-outs in the Nelson region came from a single event, the 2011 Golden Bay Storm. I focus on this event. First, I find the functional relationship between the distribution of past geo-referenced insurance claim-level data and landslide conditioning factors. I estimate a logistic regression model at property level where I identify the drivers of residential property damage and their relative effect on landslide susceptibility, and I estimate the probability of residential property damage from landslides as result of extreme rainfall. Then, I combine the predicted probabilities from the logistic regression model with property replacement values and damage-ratio information to calculate the spatial distribution of risk. The results show that risk levels are driven by moderate and high slopes, lower social deprivation levels, and high property values. It is households living in high-value properties and with lower levels of social deprivation that can afford the geotechnical and engineering works that allow them to settle on coastal hills with steep slopes, areas that are prone to failure. Thus, visual amenities play a role in determining the risk of rainfall-induced landslides in Nelson. The contribution here is the quantification of risk with a less-resource intensive methodology, as it does not depend on landslide inventories that are typically derived from satellite or aerial imagery. Instead, it only uses historical georeferenced claim-level data. Furthermore, the contribution relates to the use of social and not only physical vulnerability factors as predictors of landslide susceptibility.

The third and last chapter quantifies the current and future risk of direct property damage from floods associated with extreme precipitation. The motivation for this work relates again to the data on EQC claim payments, which show numerous low-impact events and few high-impact events. I calculate that the costliest 20% of events account for 85% of the damages. This figure highlights the importance of understanding the effect of low-probability high-impact events. I study the most expensive event for the EQC, the Bay of Plenty event in 2005. I calculate the current and future risk from floods for a range of return periods and climate change and flood depth scenarios. First, I run a logistic cross-sectional regression model where I exploit spatial variation of rainfall intensity-duration-frequency, while controlling for other factors that might make a property more or less likely to experience damage. Second, I apply the historical relationship between extreme precipitation and flood-related damage calculated in the first step to extreme precipitation data that incorporates the effect of climate change. Here, precipitation data projections are based on future greenhouse gases concentration scenarios from six different dynamically downscaled Regional Climate Models. Third, I combine the predicted probabilities from the logistic regression model,

with and without the effect of climate change, with property replacement values and damage-ratio information to calculate the spatial distribution of risk for a range of flood depth scenarios, return periods and climate change scenarios. I find that the largest losses are associated with lowest return periods. The mean, minimum and maximum percentage difference between the predicted probabilities of the benchmark model 'without climate change' and the models that incorporate the response of precipitation to climate change, increase as the greenhouse gases emissions increase. However, the change in the probability of damage as a result of climate change is too small to cause economically meaningful changes in the expected losses from flood-related damage in the Bay of Plenty.

Chapter One

Projecting the effect of climate change on residential property damages caused by extreme weather events

Abstract

New Zealand's public insurer for natural hazards, the Earthquake Commission (EQC), provides residential insurance for some weather-related damage. Climate change and the expected increase in intensity and frequency of extreme weather-related events are likely to translate into higher damages and thus an additional financial liability for the EQC. We project future insured damages from extreme precipitation events associated with future projected climatic change. We first estimate the empirical relationship between extreme precipitation events and the EQC's weather-related insurance claims based on a complete dataset of all claims from 2000 to 2017. We then use this estimated relationship, together with climate projections based on future greenhouse gases concentration scenarios from six different dynamically downscaled Regional Climate Models, to predict the impact of future extreme precipitation events on EQC liabilities for different time horizons up to the year 2100. Our results show predicted adverse impacts that vary over time and space. The per cent change between projected and past damages—the climate change signal—ranges between an increase of 7% to 8% in liabilities for the period 2020 to 2040, and between 9% and 25% higher for the period 2080 to 2100. We also provide detailed caveats as to why these quantities might be misestimated. The projected increase in the public insurer's liabilities could be used to inform private insurers, regulators, and policymakers who are assessing the future performance of both the public and private insurers that cover weather-related risks in the face of climatic change.

1.1 Introduction

Anthropogenic warming, as a result of greenhouse gas (GHG) emissions, is expected to produce changes in the frequency and intensity of weather extremes (IPCC, 2012). Extreme weather will cause damage and create additional liabilities for public insurance systems that cover everyone, and cannot withdraw from areas that are becoming riskier. Great efforts have been made to produce climate projection data that improve our understanding of potential implications of climate change on the environment, economy and society (Mullan et al., 2018). However, no study had used such projection data together with past detailed insurance claims data to project future monetary losses from damages caused by weather-related extreme precipitation events.

New Zealand offers a convenient case study, as the national public insurer (the Earthquake Commission - henceforth EQC) provides residential insurance for weather risk. Specifically, it covers land damage resulting from floods and storms, and buildings, contents, and land damage that occur due to rainfall-induced landslides. These weather-related hazards have already cost the EQC NZ$450 million (using 2017 values) since the year 2000. The expectation that the frequency and intensity of extreme weather will be amplified by climate change ultimately implies additional liability for the EQC, and it is these future additional liabilities that we attempt to quantify.

A body of literature addresses projections of future losses from weather-related events. These studies differ in their approach, type of hazard, spatial scope, changes in hazard, and climate scenarios, as well as in how they consider future changes in exposure and vulnerability (Bouwer, 2013). In contrast with this previous literature, we use a risk modelling approach based on an econometric analysis of past insurance claims data to model the empirical relationship between weather-related damages to residential property and extreme precipitation events (the hazard), while controlling for exposure and vulnerability risk factors. Previous papers (e.g. Pinto et al., 2007; Leckebusch et al., 2007, Klawa and Ulbrich, 2003) have generally used simple damage functions obtained from first principles and laboratory and field testing, but their models incorporate limited information on exposure and vulnerability.

Unlike other papers, the individual damage records we use also allow us to exploit the time dimension in our data, as every claim can be linked with a time-specific weather event. The time/grid-cell structure of the data also permits us to isolate contemporaneous variation while controlling for exposure and vulnerability through the use of grid-cell fixed effects. We can thus isolate the impact of anthropogenic climate change.We use outputs from six Regional Climate Model (RCM)

simulations to assess the changes in hazard under the four main greenhouse gas (GHG) concentrations scenarios (Representative Concentration Pathways (RCP): 2.6, 4.5, 6.0, and 8.5). The main advantage of using downscaled RCM output is that it allows us to identify the climate change signal with spatial detail, since climate change impact on precipitation is heterogeneous across space in a country the size of New Zealand.

We ultimately aim to answer two questions: What are the EQC's expected future liabilities, given future climate projections? And, how much more will the EQC have to pay in the future as a consequence of anthropogenic induced climate change?

To address these questions, we start by identifying the empirical relationship between insurance claim payouts and the number of extreme precipitation events using a longitudinal geo-coded dataset of all insurance claims for the period 2000-2017. The historical extreme precipitation events are identified based on grid-cell threshold values of the 95th, 98th and 99th percentiles of the distribution of daily rainfall taken from observation-based gridded dataset. We calculate the number of extreme rainfall events based on the same percentile values for durations of up to five days of accumulated precipitation to consider the antecedent moisture conditions of the soil, and the persistent rainfall that might lead to an insurance claim.

The empirical historical relationship identified between insurance claims and extreme rainfall, identified in the damage regressions, is then applied to past and future climate projections data to identify the predicted change in EQC liabilities – i.e., the climate change signal. Our results reveal a moderate climate change signal, where the per cent change in the expected annual losses relative to the baseline past ranges from 7.1% to 25.5% between 2020 and 2100, for the mean model ensemble, with considerable variability between the individual regional climate models. The impact of climate change on the levels of losses is heterogeneous across time and space. Some locations are predicted to experience increases in extreme precipitation events and thus in damages, while others are predicted to experience decreases in extreme events and consequently damages from them.

The paper is organized as follows. Section 1.2 provides a short literature review to benchmark our methodology. Section 1.3 describes the unique data we use, while Section 1.4 describes the results obtained from the regression models we estimate. Section 1.5 applies the estimated regression coefficients to future climate projections and thus quantifies the climate change signal. The last section provides some caveats and concluding remarks.

1.2 Literature Review

Projecting damages from future weather extreme events implies considering the changes in the frequency and intensity of weather-related hazards, but also changes in exposure and vulnerability of populations and assets. Bouwer (2013) summarizes some of the basic features of the studies that estimate future projected losses as a consequence of human-driven climatic change. These features include estimation method, hazard type, hazard (probability) change and climate scenarios, region (or spatial coverage), exposure (or socioeconomic scenario) and vulnerability (damage function estimations).

Estimation methods commonly used in the research of projected damages include Integrated Assessment Models (e.g., Narita et al., 2009; Narita et al., 2010), Computable General Equilibrium Models (e.g., OECD, 2015), and risk models (e.g., Klawa and Ulbrich, 2003; Leckebusch et al., 2007; Pinto et al., 2007). Integrated Assessment Models (IAM) describe the interactions between the economy and the biophysical system under analysis. Similarly, Computable General Equilibrium (CGE) models describe the "relations between different economic actors and contain a full description of the economic system using multiple economic sectors" (OECD, 2015). They focus mostly on modelling the overall economy of a region or a country but are less detailed about the links to the bio-physical systems. In contrast, Integrated Assessment Models contain only a more simplified description of the economy but more detail about the links to the bio-physical systems. IAMs and CGEs are used to estimate the economy-wide effects whereas we implement a risk model that evaluates the direct damage (in NZ$) in a framework where damage is determined by hazard, exposure, and vulnerability.

The studies using these approaches mainly make projections of damages from tropical cyclones, extra-tropical cyclones, or river flooding. However, no study projects the damages from extreme precipitation. The changes in these hazards are measured using Global and Regional Climate Models for different scenarios. Regarding the spatial scale, some of the studies cover single countries or regions (e.g., Schwierz et al., 2010), while others are global (e.g., Pielke, 2007).

Projections of changes in exposure are rarely incorporated, but the studies that do include these consider mainly changes in the value of assets and/or changes in population (e.g., Strader et al., 2017; Bouwer et al., 2010). Finally, vulnerability estimations "…involve a simple relationship described by a damage curve... or a loss model that specifies different damage categories" (Bouwer, 2013).

In New Zealand, the Ministry of Environment, in collaboration with Crown Research Institutes, has produced guidance documents for local government to address climate change impacts and their assessments (Mullan, 2008; NIWA et al., 2012). However, these guidance documents are

prescriptive about what local governments can or should do and typically only reference local case studies as examples. Bell et al. (2015) provide an exposure analysis of low-lying areas at the national level and the impact that rising seas – as a result of a warming climate might have on coastal risk levels. Rosier et al. (2015) estimate the anthropogenic climate change influence on the probability of occurrence of a single flooding event. They do this using a large ensemble of simulations of a regional climate model to recreate different realizations of possible weather with and without human influence. Frame et al. (2020) use these estimates to attribute past damage from flood events to human influence on the climate.

Fleming et al. (2018), a precursor to this paper, describes the EQC's weather-related claims between 2000-2017 and the geophysical and socioeconomic context of individual residential buildings. More recently, Paulik et al. (2019a) undertake a comprehensive exposure analysis of population and assets regarding pluvial and fluvial inundation hazards. In another study, Paulik et al. (2019b) quantify exposure to extreme sea-level elevation for 1% annual exceedance probability for present-day and future higher sea levels.

However, none of these studies investigate the current and future risk associated with climate change. In this research, we project future damages by implementing a risk model for damages caused by extreme precipitation events (the hazard). The most important and unique component of our analysis is the reliance on spatially and temporally detailed records of all residential damage insurance claims in New Zealand for the years 2000-2017. We use several past and future climate models to quantify the change in hazard for four greenhouse gas concentration scenarios: a mitigation scenario (RCP 2.6), two stabilization scenarios (RCP 4.5 and RCP 6.0), and one scenario with a high greenhouse gas concentration (RCP8.5). Our analyses are performed for all the inhabited areas of New Zealand, and the projections assume no future changes in exposure or vulnerability in order to isolate the impact of anthropogenic climate change.

1.3 Data and summary statistics

We conduct our investigation using a longitudinal dataset of all individual weather-related insurance claims in New Zealand (2000-2017) and extreme rainfall events aggregated at grid-cell level. There are two possible physical processes underlying each insurance claim: a flood or storm, or a rainfall-induced landslip. Although we cannot differentiate between the two in the claim dataset, the set of covariates that we include in the chapter three and four are intended to capture the generating processes for both. In the second paper I assess the current risk from landslides in Nelson, and the third paper examines the current and future risk from floods in the Bay of Plenty. The region-specific

studies for the second and third paper are motivated by the large observed EQC pay-outs from property damage in two specific events/hazards/locations.

1.3.1 The EQC insurance scheme

In New Zealand, public natural hazard insurance is provided to residential property homeowners by the Earthquake Commission (EQC). In spite of its name, the EQC also insures some weather risk and currently provides insurance cover for buildings and for land (until July 2019, it also covered contents). Specifically, it covers residential land damage caused by a storm or a flood, and both residential building and land damage caused by rainfall-induced landslips. The land cover policy includes damages that occurred to the land underneath the building, the land underneath appurtenant structures, an 8 meters buffer around these buildings and structures, and the land underneath the main access point to the house. Other covers related to the land include damage to retaining walls, bridges, and culverts (EQCover Guide, 2016).

In order to have this insurance, homeowners need to purchase private fire insurance and pay a flat yearly premium to the EQC as a compulsory supplement to the private insurance premium. During the time we cover in this research (2000-2017), the EQC's cover for residential buildings provided the first NZ$ 100,000 of the replacement values for each insured dwelling. Damages above this amount were covered by the private insurers. In contrast, the EQC land cover cap is set at the land's assessed market value and is thus different across insured households. No premium is charged on land cover (Owen and Noy, 2019).

The insurance data contain a total of 15,196 weather-related settled (completed) claims between 2000-2017. We remove claims whose status is reported as: Open, Re-open, Declined, Not accepted, Withdrawn, Invalid, Field Work in Progress, Field Work Complete and Accepted. The remaining claims amount to NZ$ 449,730,984 (in 2017 NZ$) where about 67% of the payouts are because of land damage, 32% for building damage, and 1% for contents damage.

The evolution over time of the EQC payouts in absolute values, shown in Figure 1. 1, demonstrates no increasing trend; rather, the series is dominated by specific extreme events, such as the Bay of Plenty and Waikato heavy flooding in 2005, the North Island' weather bomb' in 2008; and the Tasman-Nelson heavy rain and flooding event in 2011. The shares of payouts per cover over time, we see that the land damage share has been trending upwards, which could be driven by increasing land prices, or changes in hazard, exposure, or vulnerability. The seasonality of damages i.e., distribution of total losses per month is shown in Figure 1. 2. Larger losses tend to occur during autumn and winter (April to August). However, significant damages also occur even in peak summer

(December-January). The distribution of total damages follows a negative exponential distribution: small compensation values are quite frequent, and high or extreme compensation values are very rare.

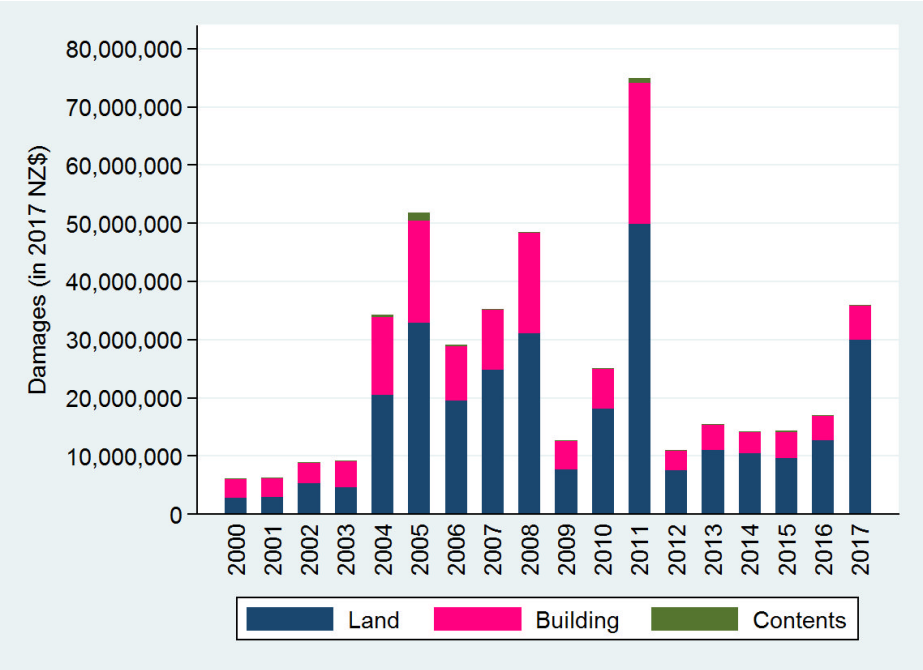


Figure 1. 1 Total value of claims paid out by the EQC for weather-related claims 2000-2017, per cover

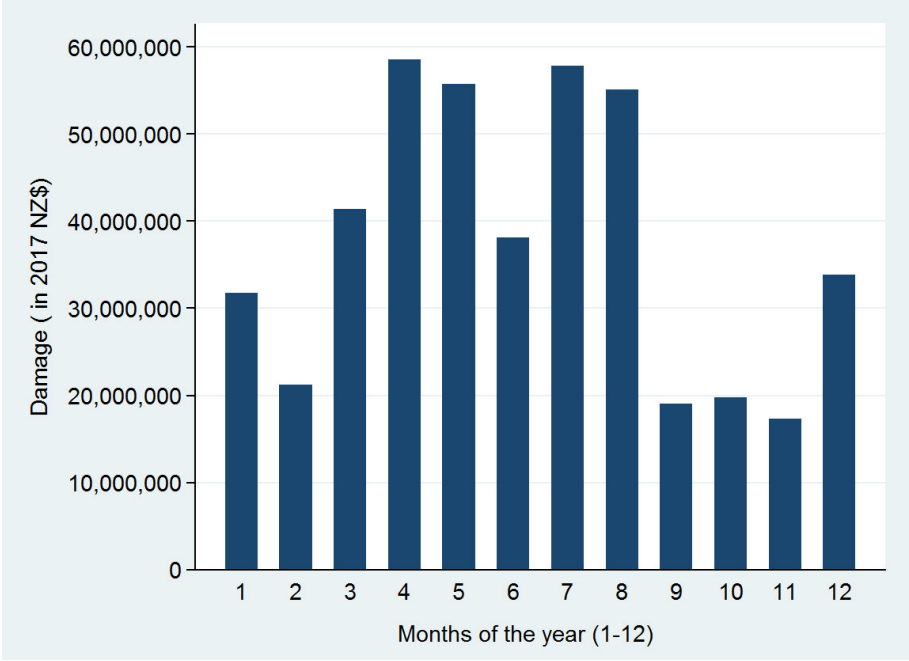


Figure 1. 2 Total value of claims paid out by the EQC for weather-related claims 2000-2017, by month

For our regressions, we drop from the sample any claim without a geospatial reference, which leaves us with 11,339 records. We also drop 2,945 claims for which the precipitation information is not available. For the regressions described in the next section, we are thus left with 8,394 claims

lodged to the EQC between 2000 and 2017, totaling NZ$180,404,945, representing about 40% of the total payouts ever made by the EQC for weather-related risk. We do not have information about any over-cap private insurance claims that were paid in these instances. However, these are not required to estimate the relationship between extreme precipitation events and EQC liabilities.

We aggregate claims data to the grid cell by year level to match the geographic level at which precipitation data are available. Grid-cell/year is thus our unit of observation. A gridded dataset is a regular data structure at which precipitation data is produced. We aggregate property-level claims to the grid cell-year-level in three ways such that we can capture the likelihood, frequency, and intensity of insurance claims that result from extreme precipitation (the hazard).

Likelihood of a claim is measured using a binary variable for whether a claim was lodged because of land, building and/or contents damage in the grid-cell/year. Frequency is measured by the total number of claims in the grid-cell/year. Finally, intensity is measured by the total value of paid claims in real NZ$ from land, building and/or contents damage in the grid-cell/year. These three spatially explicit measurements form our dependent variables.

1.3.2 Extreme precipitation (the hazard)

The precipitation data we use are an 18-year historic time-series (2000-2017) of observed daily precipitation, available for 5km-by-5km grid-cells (Tait et al., 2006). These data were produced by the National Institute of Water and Atmospheric Research (NIWA) as the Virtual Climate Station Network data.

As defined by the Intergovernmental Panel on Climate Change, extreme weather is defined as "the occurrence of a value of a weather or climate variable above (or below) a threshold value near the upper (or lower) ends of the range of observed values of the variable" (IPCC, 2012). We thus define extreme events based on the 95th, 98th, and 99th percentiles of the historical precipitation distribution for one day of accumulated precipitation. As in Griffiths (2007), the percentile thresholds are defined separately for each grid cell. To account also for the antecedent conditions that may lead to weather-related claims (for instance, saturated soil or waterways), we also calculate the same thresholds for percentiles for up to five days of accumulated precipitation. Only wet days are considered in the percentile calculations, as in Carey-Smith et al. (2010). We use these thresholds to construct, at the grid-cell/year level, the number of extreme events defined by three alternative

percentile values (95th, 98th, and 99th percentiles) and five alternative durations (from one day up to five days). [1]

We perform these calculations for inhabited grids only, which represent 56% of the entire gridded precipitation dataset. Extreme precipitation events occurring in uninhabited grids could affect adjacent grids with inhabited properties and vice versa (and similarly extreme precipitation events in inhabited grid cells could affect adjacent inhabited grid cells). However, accounting for such effects is beyond the scope of this paper. The consequences of this simplification are discussed in the last section.

In Figure 1.3 we examine the evolution of the per cent changes of the number and value of insurance claims and the number of extreme events. We can see that at a national level, the three time-series are correlated.

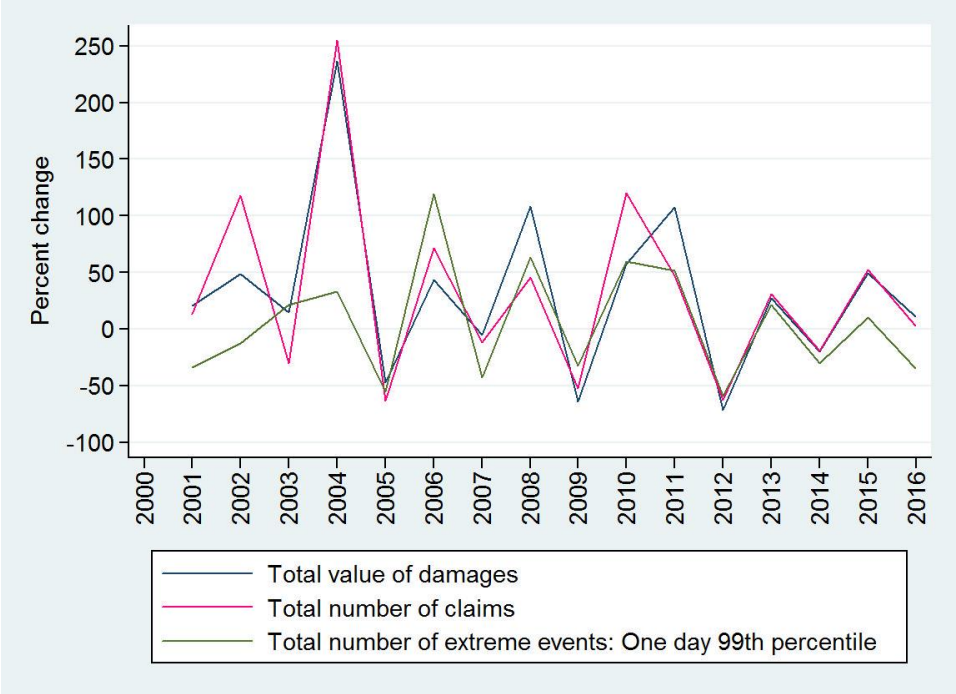


Figure 1. 3 Percent change in the number of extreme events and number and value of claims

1.3.3 Exposure and vulnerability variables

We use many variables to capture the extent of exposure and vulnerability to extreme precipitation (the hazard) for each grid-cell/year observation (these variables are reported in Appendix Table 1.8.1). The vulnerability and exposure measures come from cross-sectional data, and

[1] Because we define extreme events for each individual grid-cell, different grids are likely to display different percentile values. This is not problematic as properties and infrastructure in each location are constructed to cope/withstand with local climatic conditions.

although the underlying features are inherently dynamic in nature, they are not measured regularly for operational, financial, or practical reasons. Bouwer (2006, 2013) highlights the issue of ignoring changes in exposure and vulnerability, which may be driven by changes in adaptation and mitigation policies, as one of the main limitations of this research literature.

We aggregate all our continuous control variables (e.g., slope, elevation) from the property level to the grid cell level by taking averages and convert indicator variables to percentages (e.g. percentage of properties located in areas with poor soil drainage). Building exposure is captured by: the number of residential properties, the total value of assets in each grid-cell (building, land, contents, appurtenant, structures), the share of buildings in urban areas, and the total land area exposed. We approximate the land exposure (in km^2) by using building outlines that are "a representation of the roof outline of a building, classified from aerial imagery" (Land Information New Zealand, 2020). We spatially overlay the building outlines on the residential property dataset, and calculate an 8-metre buffer, because the EQC covers not only the land underneath the building but the surrounding land up to 8 meters. Because not all properties can be linked to an outline (presumably due to the 70-metre anonymization offset applied in the geolocation of the residential buildings), we calculate the average land exposed and multiply it by the number of properties within each grid. Due to the complexity of the cover offered by the EQC, it is not feasible to capture other related land exposures (retaining walls, access paths, etc.). We source the data from EQC (2017), CoreLogic (2017) and LINZ (2009).

Furthermore, to capture building exposure, we consider the characteristics of the soil where properties are located. Specifically, we consider the soil flood return period, drainage, readily available water, and permeability. We consider each characteristic and the categories that are likely to be associated with damages (or amplify them) under extreme precipitation events. Thus, we calculate the share of residential buildings per grid cell that are located: on soils with flood return periods ranging from 'slight' to 'very severe'; on soils with 'very poor', 'poor' and 'imperfect' drainage; on soils with 'very high', 'high' and 'moderately high' profiles of readily available water; and on soils with a 'slow' rate of water movement through saturated soil. The source for the soil datasets is Landcare Research (Newsome et al., 2008).

We calculate inundation-related exposure variables such as the share of residential buildings that are located in pluvial and fluvial flood-prone areas and the share of properties located in storm surge areas with a 1% annual exceedance probability. The flood zones were collated by NIWA with data from Local Councils (Paulik et al., 2019a), while the storm surge maps were constructed by NIWA directly (Paulik et al., 2019b). All the soil data were obtained from Newsome et al. (2008). We also measure the average distance of residential buildings from large rivers, small rivers, lakes,

and the coastline. The distances were calculated by the authors from data obtained directly from the topographic map series (LINZ, 2009)

We develop a landslip susceptibility-exposure measure based on the slope of the terrain and the type of soil on which the property is located. Specifically, we create an indicator variable for any property located on terrain with a slope greater than five degrees, and located on any of the following types of soil: "very poor", "poor" or "imperfect" soil drainage; soil with a "slow" rate of water movement through saturated soil; soil with "very high", "high" or "moderately high" profile of readily available water; fluvial soil; and, soil with flood return periods ranging from slight (less than 1 in 60-year event) to very severe (greater than 1 in 5-year event). We define the slope threshold as five degrees based on Dellow (2011), who reports probabilities of landslip hazard for slopes greater than five degrees and a rainfall index between 0 and 25 millimeters. We aggregate this property-level measure to the grid-cell level by taking the percentage of properties located on land with these characteristics and a slope greater than five degrees. Results for a second alternative approach to measure landslip susceptibility, which we also examined, are not reported here. Specifically, we used a landslip database (Geological and Nuclear Sciences, 2019) to approximate landslip exposure hazard maps. We created buffers of varying diameters around all the landslides - represented as points that were triggered by intense precipitation. However, after consultation with the institute's experts, we concluded that the buffers were not large enough to overcome the uncertainties associated with the geolocation of the rainfall-induced landslips.

To capture additional vulnerability, we use: the share of residential buildings of "deficient condition"; the buildings' average floor height from the ground, and the average elevation (above sea level); average slope on which houses are situated; the share of residential buildings that are constructed with materials that are vulnerable to water damage. Specifically, we include brick masonry buildings since "houses in New Zealand normally have a timber frame and plasterboard wall linings in the inside, which makes them highly vulnerable to flooding" (Reese and Ramsay, 2010). We also calculate the percentage of timber and brick and masonry buildings with a deficient quality. Data on the building floor height, condition and construction materials come from the RiskScape asset inventory (NIWA and GNS, 2017). Topographic data are sourced from LINZ (2009). Finally, we calculate the share of residential buildings located in areas with no agricultural land use capability as a proxy for economic activity, sourced from Newsome et al. (2008).

1.3.4 Summary Statistics

Table 1.1 presents summary statistics at grid-cell/year level for the subsamples of observations with and without insurance claims during the study period (2000-2017). The mean number of extreme events in grids with claims is always statistically significantly higher than the mean number in grids without claims. However, since extreme events are constructed from percentile thresholds that are calculated separately for each grid, a relevant question is whether grids with claims have different rainfall thresholds to grids without claims. Thus, we examine the percentile threshold values for grid-year cells with and without claims. We find that grid cells with claims have significantly higher mean threshold values than do grid-cells without claims for all percentile values and durations. That is, despite the fact an event in a grid-cell with claims must have higher rainfall to be classified as extreme, such grid-cells still have higher numbers of extreme weather events.

The differences between the two samples are also observable in exposure and vulnerability measures at grid-cell level, as shown in Appendix Table 1.8.1. For instance, the average number of properties exposed is about 22 times higher for grids with claims than grids without claims. Similarly, the mean amount of land exposed and the mean value of assets (building, land, appurtenant structures, and contents) are approximately 10 and 27 times higher, respectively, in grids with claims than in grids without claims. Similarly, the share of buildings located in urban areas is eight times higher in grids with claims than in grids without claims. This is in line with the findings of prior studies where it is shown that damages from extreme weather events are strongly associated with exposure (Bouwer, 2018; Miller et al., 2009; Pielke et al., 2008).

Regarding the inundation-vulnerability and landslip-vulnerability measures, the differences in means between grids with claims and grids without claims are statistically significant for all variables except for the average distance of residential buildings from lakes. For instance, the average percentage of properties in fluvial and pluvial flood-prone areas in grids with claims is 3.3 percentage points higher than the average percentage of properties in grids without claims. All these differences are statistically significant. Finally, we find statistically significant differences in means for four additional vulnerability-related measures: average elevation, average floor height, the share of buildings in a deficient condition, and the share of buildings located in areas with agricultural land.

	Grid-cells with claims (N=2,370)				Grid-cells without claims (N=109,788)			
	Mean	**SD**	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**
Total paid, in 1,000 $NZ (adjusted for 2017)	76.1	264.6	0.00	9,147.9				
Total number of claims	3.54	9.99	1.00	238.00	-	-	-	-
Ratio of total paid to total value exposed	0.00	0.03	0.00	0.90	-	-	-	-
Ratio of total paid to total value exposed in 100k terms, in 2017 $NZ	305.79	3,252.18	0.00	90,028.88	-	-	-	-
Break down of total paid, per cover								
Total paid for land damage, in 1,000 $NZ (adjusted for 2017)	49.1	170.9	-7.3	5,847.6	-	-	-	-
Total paid for building damage, in 1,000 $NZ (adjusted for 2017)	26,518.0	140,966.00	-0.3	3,148.7	-	-	-	-
Total paid for contents damage, in 1,000 $NZ (adjusted for 2017)	474.18	4,582.56	-0.9	151.6	-	-	-	-
Percentile threshold values, for precipitation durations (1 to 5 days), in mm.								
95th percentile one day	36.39	9.46	17.10	65.70	33.15	12.31	16.30	191.40
98th percentile one day	52.67	14.46	23.00	92.90	47.10	17.70	22.20	268.60
99th percentile one day	66.46	18.93	29.50	122.20	58.91	22.22	27.30	318.90
95th percentile two days	49.44	13.14	23.25	92.10	45.40	18.24	19.30	282.00
98th percentile two days	71.35	20.05	30.80	133.70	63.91	25.64	28.30	372.20
99th percentile two days	89.22	26.00	39.70	170.30	79.59	31.77	36.40	460.50
95th percentile three days	59.48	15.96	27.80	114.80	54.77	22.95	21.70	346.20
98th percentile three days	84.08	23.73	38.65	161.80	76.02	31.31	31.70	450.60
99th percentile three days	103.35	29.76	46.30	201.60	93.41	38.32	41.00	578.20
95th percentile four days	68.13	18.30	32.10	133.50	62.87	26.98	23.60	402.60
98th percentile four days	94.68	26.57	44.10	179.20	86.17	36.33	35.70	533.90
99th percentile four days	115.07	32.94	50.80	218.60	104.69	43.62	42.70	674.90
95th percentile five days	76.15	20.65	36.30	151.40	70.36	30.82	25.40	453.70
98th percentile five days	103.70	28.87	48.60	197.90	95.04	40.59	36.70	606.20
99th percentile five days	125.22	35.40	55.30	235.70	114.88	48.46	45.00	721.10
Number of extreme precipitation events based on percentiles values and durations (1 to 5 days)								
95th percentile one day	39.24	133.99	0	3,570	6.17	3.05	0	23
98th percentile one day	11.84	40.62	0	1,190	2.47	1.82	0	12
99th percentile one day	6.86	27.10	0	952	1.24	1.23	0	8
95th percentile two days	39.24	133.99	0	3,570	8.85	4.36	0	30
98th percentile two days	18.13	66.36	0	1,904	3.50	2.65	0	18
99th percentile two days	10.58	39.71	0	1,044	1.75	1.80	0	12
95th percentile three days	48.69	162.81	0	4,522	10.81	5.58	0	37
98th percentile three days	23.18	85.06	0	2,380	4.26	3.37	0	23
99th percentile three days	13.33	51.23	0	1,392	2.12	2.31	0	16
95th percentile four days	56.79	187.74	0	4,998	12.32	6.60	0	40
98th percentile four days	26.84	99.29	0	2,610	4.86	4.05	0	26
99th percentile four days	15.99	62.48	0	1,666	2.43	2.78	0	19
95th percentile five days	63.24	214.18	0	5,950	13.48	7.50	0	45
98th percentile five days	30.48	119.18	0	3,306	5.33	4.62	0	31
99th percentile five days	18.47	76.36	0	2,088	2.67	3.21	0	22

Table 1.1 **Descriptive statistics for insurance claims and extreme precipitation – per grid-cell/year**. We distinguish between grids that have made weather-related claims and grids without any claim. We observe statistically significant differences for all hazard measures between the two sub-groups.

1.4 Regression models

To estimate the historical relationship between extreme weather events and claims, we use the equations:

$$L_{it} = \frac{e^{\beta_1 Extreme\ rainfall_{it} + \emptyset_i + \gamma_t + \epsilon_{it}}}{1 + e^{\beta_1 Extreme\ rainfall_{it} + \emptyset_i + \gamma_t + \epsilon_{it}}} \quad (1)$$

$$F_{it} = e^{\beta_1 Extreme\ rainfall_{it} + \emptyset_i + \gamma_t + \epsilon_{it}} \quad (2)$$

$$I_{it} = \beta_1 Extreme\ rainfall_{it} + \emptyset_i + \gamma_t + \varepsilon_{it} \quad (3)$$

where *i* denotes a grid-cell, *t* denotes year, and the dependent variable L_{it} is the likelihood, F_{it} is the frequency and I_{it} is the intensity of claims as described in the previous section (measured at the grid-cell/year). The term $extreme\ rainfall_{it}$, is a vector measuring the number of extreme precipitation events (at the grid-cell/year), Φ_i is the fixed-effects term (at the grid-cell) and γ_t is the time fixed-effects term (at the grid-cell). We use a logistic regression when looking at likelihood [1], a Poisson regression for frequency [2], and an OLS regression for intensity [3]. Depending on the model, the coefficients are expressed as incidence rate ratios (IRR), odds ratios (OR) or conventional coefficients, respectively. The regressions also include time fixed-effects to capture the climate variability; as we see in the precipitation data, some years have more extreme rainfall than others. Exposure and vulnerability variables (reported in Appendix Table 1.8.1) are time-invariant and therefore are not introduced into the fixed effect models.

We estimate fixed-effects models rather than random-effects models because we are interested in analysing the effect of variables that vary over time (given our interest in projecting climate change impacts). By using fixed-effects models, we remove the effect of observed and unobserved time-invariant characteristics. “Then, any change in one of the three dependent variables must be due to influences other than these fixed characteristics.” (Torres-Reyna, 2007). We use Huber and White robust standard errors to allow for heteroskedasticity. We ran a series of Hausman tests for all percentiles and days of accumulated precipitation to confirm whether fixed-effects are preferable to random-effects. The results show a fixed-effects model is more appropriate for the logistic regression models, whereas for the Poisson regression and OLS regression the Hausman tests are inconclusive. Specifically, the models fitted do not meet the asymptotic assumptions of the test. All the estimations were produced using Stata/MP 13 and are presented in Table 1. 2.

1.4.1 Likelihood model for the probability of a claim

Column (1) of Table 1.2 presents the results of a series of logistic regressions in which the dependent variable is an indicator for whether a claim occurred in a grid-cell/year (L_{it}) and the main control variable of interest is the number of extreme weather events for the various percentile thresholds and durations. Each presented coefficient comes from a separate regression, run separately for each percentile thresholds and duration, and is expressed as an odds ratio. Exposure and vulnerability are controlled for with grid-cell fixed effects.

The first coefficient presented in this column shows that a one-unit increase in the number of extreme events (as defined at the 95th percentile for one day of accumulated precipitation) is associated with a 21.3% increase in the odds ratio of an insurance claim. Across the different definitions of extreme events, the estimated increase in the odds of an insurance claim from an additional extreme precipitation event range from 9.3% to 46.3%. In each case, the coefficient is statistically significant at the 0.01 level.

1.4.2 Frequency model for the number of claims

Column (2) of Table 1.1 presents results from a series of Poisson regressions in which the dependent variable is the number of claims in the grid cell year (F_{it}). We opt for a Poisson model rather than a negative binomial one because the negative binomial fixed effect estimator is not a true fixed effects estimator (Wooldridge, 1999). The coefficients of the estimated model are expressed as incidence rate ratios (IRR). Our exposure variable, required for count models, is the number of properties per grid-cell. The first-row coefficient shows that if the number of extreme events increases by one unit (as defined at the 95th percentile for one day of accumulated precipitation), its incidence rate ratio is expected to increase by a factor of 1.24 (a 24% increase in the incidence rate). For the different definitions of extreme weather event, the IRR range from 1.09 to 1.41, and are all statistically significant at the 0.01 level.

1.4.3 Intensity model for the total value of claims

Column (3) of Table 1.2 presents results from a series of OLS regressions of the value of total payouts, adjusted for inflation to 2017 NZ$ values. Using our first definition of an extreme event, rainfall above the 95th percentile for one day of accumulated precipitation, we estimate that one additional extreme event in a grid cell and year is associated with a NZ$ 319 increase in payouts. As we vary our definition of an extreme event in the subsequent rows of the table, the estimated coefficients range from NZ$ 132.4 to NZ$ 887.9; all are statistically significant at the 0.01 level. An alternative model is to use as the dependent variable the ratio of the total payouts (for each grid-

cell/year) to the total value of residential assets exposed (building, land, contents, and appurtenant structures). As column (4) of Table 1.2 shows, with this ratio as the dependent variable, we find non-significant results for most models, except for the model with 95th percentile one day of accumulated precipitation (0.303), and the model 95th percentile two days accumulated precipitation (0.366). We argue that the normalized figures of damage (value of claims over total value exposed) are significantly associated with one and two days of accumulated precipitation for two reasons. In the first case, we argue that one day of accumulated precipitation is related with flooding events, which are generally triggered by sub-daily durations. In the second case, we see that about 12% of the pay-outs ever made by the EQC were triggered by 2-day rainfall durations. We address these two events in Chapter two and Chapter three. For the loss projection undertaken in the next section, we use the models where the dependent variable is the total payouts (column 3). [2]

[2] We note that intensity model can be considered as a censored variable, since in most of the grid cells, there may not be any claims, so the intensity is zero, while in some grid cells, the intensity would be positive due to claims. To correct for the bias, a Tobit model can be used to estimate the relationship when there is censoring in the dependent variable. Alternatively, it is possible to estimate a two-stage model where the probability of making a claim is first estimated, and then, the value of claim is estimated, conditional on having made a claim. However, we leave this issue for future research.

	(1)	(2)	(3)	(4)
Model type	Logit (Probability)	Poisson (Frequency)	OLS (Intensity)	OLS (Intensity)
Dependent variable	Indicator for at least one claim in grid/cell	Number of claims in grid/cell	Value of claims in grid/cell	Value of claims relative to exposed assets in grid/cell
Coefficient type	Odds Ratio (OR)	Incidence Rate Ratio (IRR)	OLS	OLS
95th percentile one day	1.213***	1.241***	319.0***	0.303***
	(0.0141)	(0.0237)	(72.46)	(0.0600)
98th percentile one day	1.404***	1.411***	538.1***	1.314
	(0.0238)	(0.0492)	(89.42)	(0.716)
99th percentile one day	1.597***	1.569***	887.9***	2.502
	(0.0364)	(0.0805)	(163.0)	(1.825)
95th percentile two days	1.157***	1.170***	250.5***	0.366**
	(0.00915)	(0.0138)	(45.32)	(0.132)
98th percentile two days	1.295***	1.275***	441.4***	1.253
	(0.0145)	(0.0368)	(70.27)	(0.800)
99th percentile two days	1.463***	1.376***	634.1***	1.803
	(0.0235)	(0.0500)	(90.74)	(1.150)
95th percentile three days	1.128***	1.127***	187.6***	0.484
	(0.00690)	(0.0126)	(32.72)	(0.290)
98th percentile three days	1.238***	1.221***	355.4***	1.000
	(0.0109)	(0.0248)	(52.28)	(0.630)
99th percentile three days	1.359***	1.260***	486.9***	1.454
	(0.0166)	(0.0322)	(69.25)	(0.890)
95th percentile four days	1.107***	1.105***	153.7***	0.359
	(0.00551)	(0.0104)	(24.15)	(0.197)
98th percentile four days	1.192***	1.172***	261.7***	0.836
	(0.00867)	(0.0198)	(39.57)	(0.545)
99th percentile four days	1.298***	1.255***	432.0***	1.266
	(0.0132)	(0.0231)	(63.26)	(0.738)
95th percentile five days	1.093***	1.090***	132.4***	0.284
	(0.00467)	(0.00988)	(21.54)	(0.147)
98th percentile five days	1.175***	1.152***	237.4***	0.651
	(0.00741)	(0.0144)	(37.76)	(0.379)
99th percentile five days	1.250***	1.239***	383.2***	1.134
	(0.0108)	(0.0167)	(57.04)	(0.679)
Year fixed-effects	Yes	Yes	Yes	Yes
Grid-cell fixed-effects	Yes	Yes	Yes	Yes
N	14,238	14,238	112,158	112,158

Table 1.2 **Functional relationship between EQC insurance claims and extreme precipitation**. This table presents the coefficients on extreme weather events from a series of regressions of claims on extreme events. Each coefficient comes from a separate regression. The dependent variable and regression type vary by column, and the definition of extreme event differs by row. For the intensity models, the dependent variable is the total amount of payouts from damage for all the insurance covers (column 3), and the total amount of payouts from damage to all the insurance covers divided by the total value of assets exposed, divided by a hundred thousand (column 4). The stars *** denote statistical significance at the 1% level.

1.5 Applying the damage function to future climate projections

Our next step is to use the damage functions we estimated in the previous section to project the value of insurance claims in the future, using the available predictions about the future impact of climate change on the occurrence of extreme precipitation events. We use output from simulations

of NIWA's dynamical downscaling setup driven by ocean temperatures from the Coupled Model Intercomparison Project (CMIP-5) climate models.

The six CMIP-5 models used in this study are: HadGEM2-ES from the UK; NorESM1-M from Norway; CESM1-CAM5, GFDL-CM3, and GISS-E2-R from the US; and BCC-CSM1.1 from China. The six different representations of the climate have been built-up to reflect the past climate (1971-2005) and project future climate under the different greenhouse-gas emissions scenarios known as Representative Concentration Pathways (RCPs 2.6, 4.5, 6.0, 8.5) and periods (2006-2100). RPCs are scenarios that include time series of emissions and concentrations of the full suite of greenhouse gases and aerosols and chemically active gases, as well as land use/land cover (Moss et al., 2008). Each model thus yields a different realization of possible future precipitation conditional on an emissions scenario. Some of these models predict the climate to 2120, but we restrict our predictions to 2100. Further details related to the climate models are provided in Mullan et al. (2018) and Sood (2014).

1.5.1 Projecting losses

We project losses for up to the year 2100 by applying the historical relationship between extreme precipitation and weather-related claims that we estimated in Section 1.4, to the modelled past and future weather data. The projection is done for all RCP scenarios in 20-year time slices for all percentiles and days of accumulated precipitation, and for all climate models; altogether, this implies 360 projections for each 20-year time slice. We avoid making predictions for short time-spans (e.g. 5 years) because these will be too volatile and may be affected by cyclical phenomena such as the timing of the tropical Pacific Ocean oscillations (El Niño and La Niña).

We count the future number of extreme precipitations as the number of times modelled future rainfall exceeds the percentile thresholds calculated from the modelled past data from the same simulation. This allows us to establish the appropriate benchmark against which we can calculate future climate change impact.

The model simulations of the past rainfall produce 95th, 98th, and 99th percentiles thresholds of past extreme events that are considerably lower than the corresponding percentiles of the past observed rainfall. We therefore cannot use the thresholds calculated from past observed rainfall, but calculate new thresholds from the modelled data of the past. It is those thresholds that are then used to identify and count the number of projected future extreme events (given the percentile and duration thresholds we obtained from the modelled data).

We project future losses assuming no changes in exposure (e.g. number and value of residential property) or vulnerability (e.g. construction materials). The main constraint preventing us from considering different scenarios for changes in exposure and vulnerability is the detailed spatial resolution at which we operate. The 5km x 5km grid-cells we use are much smaller than administrative units or regions at which socioeconomic pathway scenarios are generally developed for. However, the study provides a detailed baseline of potential future losses that the EQC could face given no further growth of residential areas. As the country continues to economically grow and develop, the projected losses here are of course likely to be higher.

Given the 360 projections we produced per each 20-year, we present only a subset of these. In Table 1.3 we present the results of the projections for one of the climate models (GFDL-CM3) and for only two durations (one and five days). In Appendix Table 1.8.2 we present results for all 6 climate models, but only for one duration (one day) and one percentile threshold (99%). All other results are available upon request.

Several observations about the results presented in Table 1.3 and Appendix Table 1.8.2 are noteworthy. In Table 1.3, we observe that predicted liabilities are largest when we use the 95% percentile 1-day model and decrease as we increase the duration or the percentile threshold we use. Essentially, this is because there are more events for these lower thresholds (e.g., 95 percentile) than there are, in the modelled data, for the higher thresholds (e.g., 99 percentile, in terms of either duration or percentile threshold). When we compare across the climate models, we see that the differences across models are not very large, though some models do have a flatter profile across time than others (e.g., the Norwegian model). We also observe, as can be expected, the differences between the RCP scenarios are more pronounced later in the century than they are in the near future (2020-2040). We next use these results to estimate by how much anthropogenic climate change will likely change future EQC liabilities from extreme weather events.

		One day of accumulated precipitation			Five days of accumulated precipitation		
		95th percentile	98th percentile	99th percentile	95th percentile	98th percentile	99th percentile
2020-2040	RCP 2.6	1,620	1,284	1,182	628	616	623
	RCP 4.5	1,600	1,289	1,199	621	606	611
	RCP 6.0	1,640	1,316	1,217	646	633	641
	RCP 8.5	1,523	1,199	1,099	563	556	558
2040-2060	RCP 2.6	1,583	1,271	1,169	598	583	583
	RCP 4.5	1,605	1,296	1,224	626	628	651
	RCP 6.0	1,605	1,296	1,212	621	623	646
	RCP 8.5	1,635	1,326	1,246	641	648	676
2060-2080	RCP 2.6	1,685	1,374	1,286	673	678	705
	RCP 4.5	1,583	1,299	1,222	601	593	601
	RCP 6.0	1,633	1,316	1,241	631	616	631
	RCP 8.5	1,720	1,426	1,359	698	701	725
2080-2100	RCP 2.6	1,660	1,326	1,224	641	623	626
	RCP 4.5	1,643	1,354	1,279	631	631	648
	RCP 6.0	1,625	1,336	1,269	641	651	678
	RCP 8.5	1,643	1,396	1,359	671	710	760

Table 1. 3 **Projected Future Liabilities with the GDFL-CM3 for the changing hazard (in NZ$ Millions).** Projected losses for 20-year aggregates for the percentiles 95th, 98th and 99th percentile values and one and five days of accumulated precipitation, and all Representative Concentration Pathways, using the GDFL-CM3 (NOAA-USA) climate model. These results assume no future changes in exposure or vulnerability. The projected liability figures were inflated by a correction factor of 2.50. The need for an adjustment rises as a result of the claims omitted from the regression analysis. The factor is calculated such that we add the value of the claims included and the value of the claims omitted and divide that over the value of the claims omitted.

1.5.2 Quantifying the climate change signal

To quantify the expected impact of climate change on damages, we compare the predicted damages using the past model of the climate for the years 1986 to 2005 with the losses based on future climate change projections, for each of the periods 2020-2040, 2040-2060, 2060-2080, and 2080-2100. We repeat this for all percentile thresholds (95% to 99%), days of accumulated precipitation (1 to 5 days), RCPs (4.5 - 8.5), and six climate models. The climate change signal is calculated with the following:

$$CC\_Signal_{pd} = 100 * \sum_{i=1}^{6,231}(CFuture_{ipd} - CPast_{ipd}) / \sum_{i=1}^{6,231}(CPast_{ipd}) \quad (4)$$

It is the percentage change of the sum, aggregated across 6,231 inhabited grid-cells of the future liabilities of the EQC, based on the modelled data, minus the past modelled liabilities, based on the same climate model. In Figures 1.4 we present the estimated increase in liabilities attributed to the impact of climate change in each 20-year future period, relative to the 20-year period 1986-2005, for the 99th percentile of one-day rainfall duration. These estimates are averages across the six climate models and presented for the four RPCs. We chose this 1-day duration because the time of concentration (TC) for most catchments in New Zealand is less than a day. This means that intensity rainfall duration (IRD) and the time for a drop of water to reach the coast occurs over sub-daily

periods. Thus, in the New Zealand hydro-geographical context, one day of accumulated precipitation is more appropriate to use over any longer durations.[3]

These projections reveal a modest climate change-driven increase in the value of EQC insurance claims that are projected in the future. Even towards the end of the century (2080-2100), we see that difference in losses that range spatially, depending on the climate model, from: -0.53% to 18.73% for RCP 2.6; -0.58% to 21.43% for RCP 4.5; 4.02% to 25.38% for RCP 6.0; and -4.43% to 27.17% for RCP 8.5. This range reflects the 'consensus' prediction that some parts of New Zealand will become dryer and thus experience fewer claims related to extreme precipitation.

These results can best be summarized by averaging across the different climate models for the same RPCs and time horizons; as shown in Figure 1. 4. The results from averaging climate change signal across the six different climate models appear consistent with our intuition. Overall, liabilities will increase more if future GHG emissions are higher (higher RCPs). The climate signal for the low emissions scenario (RCP 2.6) is lower and progressively increases with more emissions (higher RCP). For the low emission scenarios (RCP 2.6 and 4.5) the liabilities actually decrease toward the end of the century, when GHG concentrations in the atmosphere are assumed to decrease. In contrast, the time profile of the highest-emissions RCP 8.5 is much steeper, with the climate signal (the insurance liabilities) more than doubling between 2020-2040 and 2080-2100. Since model ensembles are used, we show variability in addition to the mean estimates in Appendix Table 1.8.3.

While Figure 1. 4 averages the climate change signal across New Zealand, Figure 1. 5 maps this information for each inhabited grid-cell for RCP 6.0 (similar maps for RCP 4.5 and 8.5 are available in the Appendix Figure 1.8.4 and 1.8.5). Similarly, to Figure 1. 4, the information in Figure 1. 5 is the average of the six climate models, for the 1-day 99% percentile thresholds. We find that a lot of the increase in liabilities for the EQC is concentrated in the Southern and Northern regions of the South Island, and along the South-West coast of the North Island from Wellington to Taranaki. The change we observe in the most densely populated region around Auckland is less pronounced.

[3] This reasoning if confirmed by the results shown in table 1.3, where the normalized figures of damage (value of claims over total value exposed) are statistically significant for one day of accumulated precipitation. However, two-day duration events are significantly related to levels of damage. In chapter two and three we observe that 12% of the pay-outs ever made by the EQC and they both were triggered by 2-day rainfall durations.

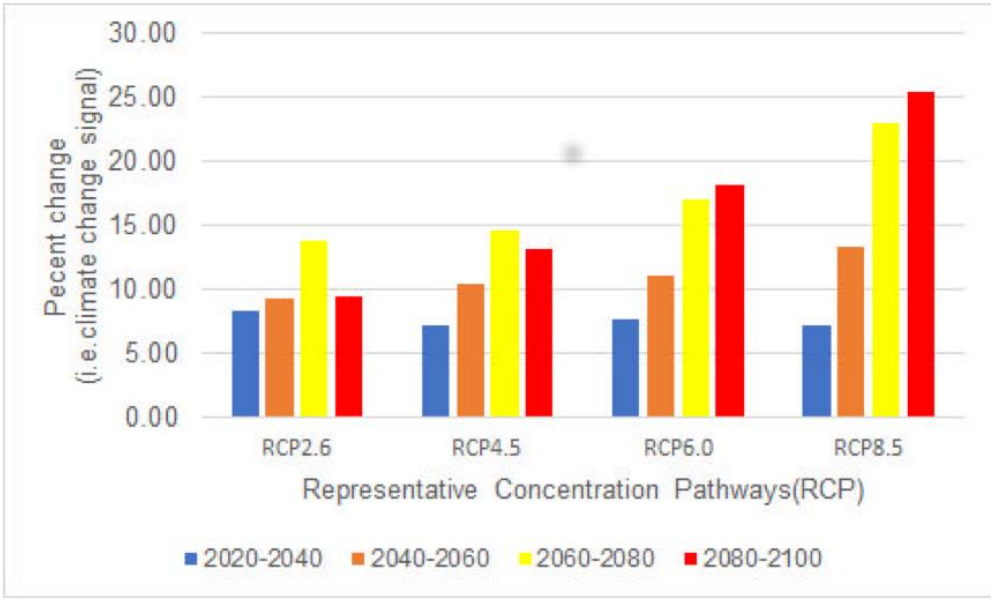


Figure 1. 4 **Increase in EQC liabilities due to climate change: average of all climate models (in %).** These results are calculated for the average one day of accumulated precipitation and 99th across six climate models, for each RCP and time horizon.

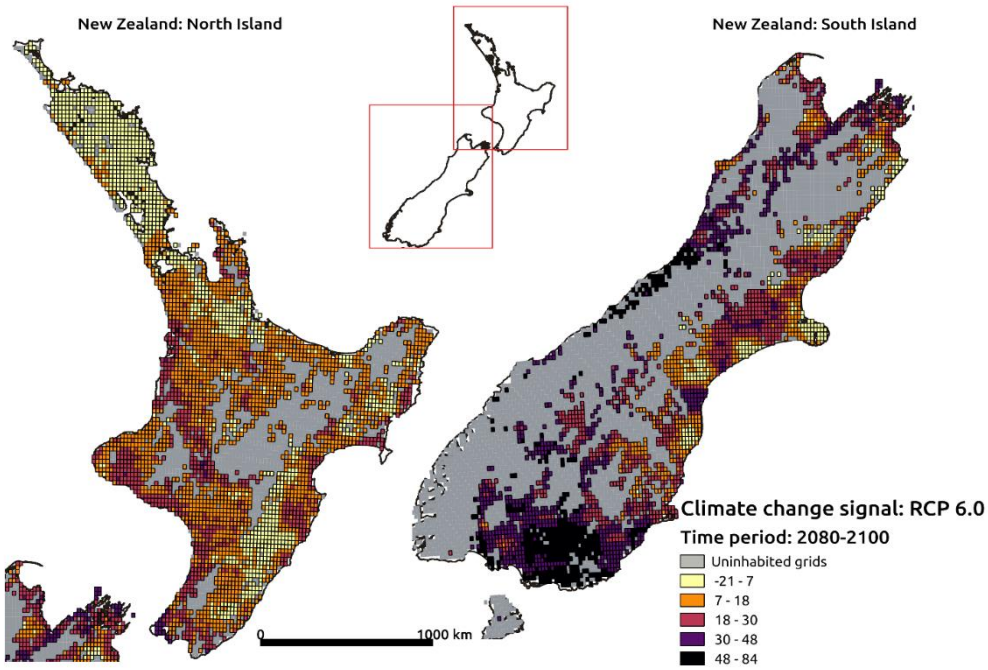


Figure 1. 5 **Average Predicted Change in EQC liabilities for RCP 6.0 in every grid-cell (in %)**

1.6 Conclusions and Caveats

Here, we combined residential insurance claim data, other geospatial records, and historical and projected precipitation data to project future liabilities of New Zealand's public insurer from extreme precipitation events. We calculate these future liabilities for four different Representative Concentration Pathways, for the output from six climate models, and using a range of definitions of

‘extreme precipitation events. We show that the climate signal (i.e., the per cent difference between future and past liabilities for the EQC) will range - depending on the GHG emissions scenario- between 7% and 8% higher in the period 2020-2040 and between 9% and 25% higher in the period 2080 to 2100 as a result of climate change-induced increases in extreme precipitation events.

The estimated climate change signal follows, approximately, the GHG concentration trajectories according to each RCP so that higher GHG concentrations are generally associated with larger increase in liabilities. New Zealand’s population and the value of its residential building stock have grown steadily over the past few decades (RBNZ, 2019) and both are projected to continue to increase. This suggests that the future liabilities may be higher than our estimates, since we assume constant exposure.

However, public policy can make a difference by either reducing exposure (e.g., through better land-use planning), or reducing vulnerability (e.g., through better construction standards). With the right policies and well-targeted investments, the public insurer’s liabilities can instead decrease. Another important policy consideration that should be explored is changes in what the public insurer covers, or whether our findings suggest a policy change, for example, in the amount of premiums the EQC collects annually.

These potential policy focus areas raise many difficult questions around responsibility, risk sharing, distributional concerns, procedural fairness, and political viability. They are all issues that the economic analysis presented here cannot resolve without resorting to ethical, philosophical, and political considerations that are best left for future research. For example, even a seemingly innocuous policy, like reducing exposure through managed retreats from increasing-risk areas is contentious politically, and raises some difficult questions about who is responsible for this increased risk, and who should pay for it. Should this increased risk continue to be covered by the public insurer? Should the homeowners bear some of it (if not all)? Should the taxpayers assist the homeowners affected? Is insurance the best way of dealing with this problem? Or, should this be kicked down the road for future generations to resolve? For examples of some of these discussions in the New Zealand context, see Ellis (2018), James et al. (2020), Noy (2020), Owen et al. (2019), and Storey and Noy (2017).

Given the obvious complicated policy space in which these results matter, some important caveats should be considered when using our results, or when implementing the methodology we proposed for estimating the future impacts of climate change.

First, because of missing records, 60% of the total damages between 2000-2017 had to be omitted from the regression analyses and thus from our projections. About 35% correspond to losses

that could not be georeferenced. The remaining 25% correspond to claims without precipitation information, as the precipitations records are not spatially complete. To correct for this, the projected liability figures were inflated by a correction factor of 2.5. If, however, the omitted data is somehow different, and the sample we have is not representative of the missing data, that correction factor might not be accurate enough.

Second, our data show that the EQC paid for some landslip/flood claims in grids that did not experience any measured precipitation. This is most likely because the intense precipitation happened upstream, but the damages (claims) occurred downstream, or because these were dry landslips (triggered by other factors). In order to be able to identify claims that have been caused by precipitation upstream, we require a complete hydrological modelling of all the watersheds in New Zealand. Such a modelling is not available and is unlikely to be available in the next few years. We did not remove these 'zero precipitation' claims from the regressions, but as long as the occurrence of these events is orthogonal to the wet landslide events, this should not bias our results.

Third, because of the difference between past modelled precipitation and the past observed precipitation percentile threshold values, future damage assessment of extreme precipitation events may be inaccurately projected. The biases in the precipitation extremes in the climate model simulations, which mostly are due to internal variability of the climate system, may lead to potential overestimation or underestimation of future losses. In order to deal with randomness associated with occurrence of climate extremes, one should use the mean model ensemble rather than results from a single simulation, which is what we presented. Even though six simulations may not be sufficient to adequately assess the extremes with a probabilistic approach, the use of carefully selected multiple models allows us to determine the range of the ensemble as an estimate of variability. As long as improved higher-resolution validated datasets and more simulation data products are not available, we see no other way of overcoming this problem (see also Sood, 2014).

Fourth, the predictions we make about future climate change costs assume constant exposure and vulnerability over time. Population projections are generally produced for large regions or administrative areas, but our estimations and calculations are produced at 5km-by-5km grid. We doubt any reliable modelling of the future distribution of population throughout New Zealand on a grid-cell basis is currently possible, so we do not attempt to account for that in our estimates. We also are not aware of any attempt to forecast future vulnerability for the housing stock. We therefore assume that vulnerability is constant over time. There are reasons to expect both increases and decreases in both exposure and vulnerability, so our *ceteris paribus* assumption appears plausible.

In a related matter, we expect the EQC cover policy change cover (whereby contents are not covered since 2019) to have a negligible impact on the projections as contents represent about 1% of the total value of claims.

Fifth, although the EQC dataset does not explicitly classify insurance claims as being caused by a flood or a rainfall-induced landslip, we deduce that more than 70% of the weather-related insurance claims are most likely related to landslips. We differentiate between claims triggered by storm/floods from the claims triggered by landslips by examining the "claim status" variable the dataset, and, based on the EQC policy coverage. However, this algorithm does not identify the cause of the damage accurately. As discussed in Section 1.3.3, we approximate landslip susceptibility by combining slope and soil type and based on an adaptation of the algorithm developed by Dellow et al. (2011). This measure approximates landslip hazard, though not as accurately as would be possible using actual landslip hazard maps. If rainfall-induced landslip hazard maps were to become available, future research could improve our estimates of the potential future liabilities of the EQC.

Sixth, our definition of extremes is based on a short time series (20 years) and only few simulations in a non-stationary system. The limited number of simulations imply that the climate signal of extremes is not statistically robust considering the levels of uncertainty. Conventionally, extreme events are defined as such when their return periods are low, and their threshold value is high. For instance, an event with a 100-year return period would qualify as an extreme event; in contrast, our definition of extremes for the lowest percentile and day of accumulated precipitation renders a total of 18 extreme precipitation events in a given year (assuming it rained every day). This issue can be addressed by using the modelled extreme precipitation with low return periods and examining the relationship with weather-related claim data. For such matter, we intend to use of the High Intensity Rainfall Design System (HIRDSv4) dataset, which provides a range of annual recurrence intervals for various durations (daily and sub-daily)[4]. An alternative set of extreme weather projections can also be obtained from more modest models which might be more suitable for identifying the distribution of the tails of the distribution, such as the *weather@home* project. We propose to take these challenges on in future research.

[4] Chapter three addresses the relationship between flood-related property damage and extreme precipitation from the HIRDSv4 dataset.

1.7 References

Bell, R.G., Paulik, R. and Wadwha, S., 2015. National and regional risk exposure in low-lying coastal areas: Areal extent, population, buildings and infrastructure. HAM2015-006, National Institute of Water & Atmospheric Research Ltd, Hamilton. https://www.pce.parliament.nz/media/1384/national-and-regional-risk-exposure-in-low-lying-coastal-areas-niwa-2015.pdf (accessed 18 June 2018)

Bender, M.A., Knutson, T.R., Tuleya, R.E., Sirutis, J.J., Vecchi, G.A., Garner, S.T. and Held, I.M., 2010. Modeled impact of anthropogenic warming on the frequency of intense Atlantic hurricanes. Science, 327, pp.454-458.

Bouwer, L.M., Bubeck, P. and Aerts, J.C., 2010. Changes in future flood risk due to climate and development in a Dutch polder area. Global Environ. Chang. 20, pp.463-471.

Bouwer, L.M., 2013. Projections of future extreme weather losses under changes in climate and exposure. Risk. Anal., 33, pp.915-930.

Bouwer, L.M., 2018. Observed and projected impacts from extreme weather events: Implications for loss and damage, in: Mechler R., Bouwer L., Schinko, T., Surminski, S., Linnerooth-Bayer, J. (Eds.), Loss and damage from climate change: Concepts, methods and policy options. Springer, pp. 63–82.

Carey-Smith, T., Dean, S., Vial, J., Thompson, C. 2010. Changes in precipitation extremes for New Zealand: climate model predictions. Weather and Climate, 30, pp.23-48.

[dataset] Core Logic, 2017. Quotable Value. Accessed under a data sharing agreement.

Dellow, S. 2011. Tool 2.3.3.: Modelling present-day and future landslide potential. Impacts of Climate Change on Urban Infrastructure and the Built Environment - A Toolbox. Access link: https://www.niwa.co.nz/sites/niwa.co.nz/files/tool_2.3.3_modelling_landslide_potential.pdf.

Dorland, C., Tol, R.S. and Palutikof, J.P., 1999. Vulnerability of the Netherlands and Northwest Europe to storm damage under climate change. Climatic Change, 43, pp.513-535.

Earthquake Commission Act, 1993. New Zealand. http://www.legislation.govt.nz/act/public/1993/0084/latest/DLM305968.html (accessed 7 July 2017)

Earthquake Commission, 2016. EQCover insurers' Guide. Wellington, New Zealand. https://www.eqc.govt.nz/what-we-do/eqc-insurance/insurers-guide (accessed 7 July 2017)

Ellis, E., 2018. How should the risks of sea-level rise be shared? Deep South Research Report.

Fleming, D., Noy, I., Pástor-Paz, J. and Owen, S., 2018. Public insurance and climate change (part one): Past trends in weather-related insurance in New Zealand. Available at SSRN 3477038. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3477038 (accessed 19 January 2020)

Frame, D., Rosier, S., Noy, I., Harrington, L., Carey-Smith, T., Sparrow, S., Stone, D. and Dean, S., 2020. Climate change attribution and the economic costs of extreme weather events: a study on damages from extreme rainfall and drought. Climatic Change, forthcoming. https://doi.org/10.1007/s10584-020-02729-y.

Griffiths, G.M., 2007. Changes in New Zealand daily rainfall extremes 1930-2004. Weather and Climate, 27, pp.3-44.

[dataset] Geological and Nuclear Sciences(GNS), 2019. New Zealand Lanslides database. http://data.gns.cri.nz/landslides/index.html (accesed 2 February 2019)

Handmer, J., Honda, Y., Kundzewicz, Z.W., Arnell, N., Benito, G., Hatfield, J., Mohamed, I.F., Peduzzi, P., Wu, S., Sherstyukov, B. and Takahashi, K., 2012. Changes in impacts of climate extremes: human systems and ecosystems, in: Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.K. Plattner, S.K. Allen, M. Tignor, P.M. Midgley (Eds.), Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York, pp. 231-290.

IPCC., 2012: Glossary of terms, in: Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen,M. Tignor, and P.M. Midgley (Eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on ClimateChange (IPCC). Cambridge University Press Cambridge and New York, pp. 555-564.

James, V., Iorns, C., Gerard, P., 2020. Sea level rise and local government. Policy Q. 16(1).

Klawa, M. and Ulbrich, U., 2003. A model for the estimation of storm losses and the identification of severe winter storms in Germany. Nat. Hazard Earth Sys., 3, pp.725-732.

[dataset] Land Information New Zealand (LINZ), Building outlines, 2020. https://data.linz.govt.nz/layer/101290-nz-building-outlines/metadata/?type=dc (accessed 19 April 2019)

[dataset] Land Information New Zealand (LINZ), Topographic maps series 1:50.000., 2009. https://data.linz.govt.nz/data/category/topographic/nz-topo-50-data/ (accessed 18 July 2018)

Leckebusch, G.C., Ulbrich, U., Fröhlich, L. and Pinto, J.G., 2007. Property loss potentials for European midlatitude storms in a changing climate. Geophys. Res. Lett.**,** 34.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., ... & Wilbanks, T. J. (2010). The next generation of scenarios for climate change research and assessment. Nature, 463(7282), 747-756.

Miller, S., Muir-Wood, R., & Boissonnade, A. (2008). An exploration of trends in normalized weather-related catastrophe losses, in: H. Diaz & R. Murnane (Eds.), Climate Extremes and Society (pp. 225-247). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511535840.015

Mullan, A.B.; Wratt, D.; Dean, S, and Hollis, M., 2008. Climate change effects and impacts assessment A Guidance Manual for Local Government in New Zealand – 2nd Edition. NIWA Client Report WLG2007/62 for Ministry for the Environment https://www.mfe.govt.nz/publications/climate-change/climate-change-effects-and-impacts-assessment-guidance-manual-local-6 (accessed 7 August 2018)

Mullan, A.B.; Sood, A.; Stuart, S.; Carey-Smith, T. 2018. Climate Change Projections for New Zealand: Atmosphere Projections based on Simulations from the IPCC Fifth Assessment, 2nd Edition. NIWA Client Report for Ministry for the Environment, updating the June 2016 report with a section of projections of extreme rainfall changes. WLG2015-31. June 2018. https://www.mfe.govt.nz/sites/default/files/media/Climate%20Change/Climate-change-projections-2nd-edition-final.pdf (accessed 12 May 2019)

Narita, D., Tol, R.S. and Anthoff, D., 2009. Damage costs of climate change through intensification of tropical cyclone activities: an application of FUND. Clim. Res., 39, 87-97.

Narita, D., Tol, R.S. and Anthoff, D., 2010. Economic costs of extratropical storms under climate change: an application of FUND. J. Environ. Plann. Man. 53, 371-384.

[dataset] National Institute of Water and Atmospheric Research (NIWA), 2018. Storm surge 1% Annual Exceedance Probability (AEP). Accessed under a data sharing agreement.

[dataset] National Institute of Water and Atmospheric Research (NIWA), 2019. Virtual Climate Station Network (VCSN). Accessed under a data sharing agreement.

[dataset] National Institute of Water and Atmospheric Research (NIWA) and Geological and Nuclear (GNS) Sciences, 2017. RiskScape Asset Inventory. Accessed under a data sharing agreement.

[dataset] Newsome, P.F.J., Wilde, R.H., Willoughby, E.J., 2008. Land Resource Information System Spatial data layers. https://lris.scinfo.org.nz/data/category/environment/ (accessed 09 September 2018)

NIWA, MWH, GNS and BRANZ (2012) Impacts of Climate Change on Urban Infrastructure and the Built Environment: Toolbox Handbook. http://www.niwa.co.nz/climate/urban-impacts-toolbox (accessed 12 April 2019)

Noy, I., 2020. Paying a price of climate change: Who pays for managed retreats? Current Clim. Change Rep., 6, 17-23.

OECD, 2015. The Economic Consequences of Climate Change, OECD Publishing, Paris, https://doi.org/10.1787/9789264235410-en (accessed 10 October 2019)

Owen, S. and Noy, I., 2019. Regressivity in public natural hazard insurance: A quantitative analysis of the New Zealand case. w/ Sally Owen. Econ. of Disasters and Clim. Change, 3, 235-255.

Owen, S., Noy, I., Pástor-Paz, J. and Fleming, D., 2019. EQC and extreme weather events (part 2): Measuring the impact of insurance on New Zealand landslip, storm and flood recovery using nightlights. Motu Working Paper 19-19.

Paulik, R., Craing, D., Collins, D., 2019a. New Zealand Fluvial and Pluvial Flood Exposure. National Institute of Water & Atmospheric Research Ltd.https://www.deepsouthchallenge.co.nz/sites/default/files/201908/2019118WN_DEPSI18301_Flood%20Exposure_Final%20%281%29.pdf (accessed 25 March 2020)

Paulik, R., Stephens, S., Wadhwa, S., Bell, R., Popovich, B., Robinson, B., 2019b. Coastal Flooding Exposure Under Future Sea-level Rise for New Zealand. Institute of Water & Atmospheric Research Ltd. https://www.deepsouthchallenge.co.nz/sites/default/files/2019-08/2019119WN_DEPSI18301_Coast_Flood_Exp_under_Fut_Sealevel_rise_FINAL%20%281%29_0.pdf (accessed 25 March 2020)

Pielke Jr, R.A., Gratz, J., Landsea, C.W., Collins, D., Saunders, M.A., Musulin, R., 2008. Normalized hurricane damage in the United States: 1900–2005.Nat. Hazards Rev. 9, 29-42.

Pielke Jr, R.A., 2007. Future economic damage from tropical cyclones: sensitivities to societal and climate changes. Philos. T. R. Soc. A.365, 2717-2729.

Pinto, J.G., Fröhlich, E.L., Leckebusch, G.C., Ulbrich, U., 2007. Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GMC. Nat. Hazard Earth Sys. 7, 165-175.

Reese, S. and Ramsay, D., 2010. RiskScape: Flood fragility methodology. NIWA Technical Report: WLG2010-45. https://www.wgtn.ac.nz/sgees/research-centres/documents/riskscape-flood-fragility-methodology.pdf (accessed 5 January 2018)

Reserve Bank of New Zealand20, 2019. Macroeconomic and financial market trends in New Zealand. https://www.rbnz.govt.nz/statistics/key-graphs (accessed 22 November 2019)

Rosier, S., Dean, S., Stuart, S., Carey-Smith, T., Black, M.T., Massey, N., 2015. Extreme Rainfall in Early July 2014 in Northland, New Zealand—Was There an Anthropogenic Influence?. B. Am. Meteorol. Soc.96, S136-S140.

Schwierz, C., Köllner-Heck, P., Mutter, E.Z., Bresch, D.N., Vidale, P.L., Wild, M., Schär, C., 2010. Modelling European winter wind storm losses in current and future climate. Climatic Change 101, 485-514.

Sood, A., 2014. Improved Bias Corrected and Downscaled Regional Climate Model Data for Climate Impact Studies: Validation and Assessment for New Zealand. https://www.researchgate.net/publication/265510643_Improved_Bias_Corrected_and_Downscaled_Regional_Climate_Model_Data_for_Climate_Impact_Studies_Validation_and_Assessment_for_New_Zealand (accessed 4 February 2020)

Storey, B., and Noy, I., 2017. Insuring property under climate change. Policy Q. 13, 68-74.

Strader, S.M., Ashley, W.S., Pingel, T.J., Krmenec, A.J., 2017. Projected 21st century changes in tornado exposure, risk, and disaster potential. Climatic Change 141, 301-313.

Tait, A., Henderson, R., Turner, R., Zheng, X.G. 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. Int. J. Climatol. 26, 2097–2115.

Torres-Reyna, O., 2007. Panel data analysis fixed and random effects using Stata (v. 4.2). Data & Statistical Services, Princeton University. https://www.princeton.edu/~otorres/Panel101.pdf (accessed 6 January, 2020)

Wooldridge, J.M., 1999. Distribution-Free Estimation of Some Nonlinear Panel Data Models. J. Econometrics 90, 77–97.

1.8 Appendices

Appendix Table 1.8.1: Descriptive statistics for exposure – per grid-cell

Residential property exposure measures	Grid cells with claims (n=791) Mean	SD	Min	Max	Grid cells without claims (n=5,440) Mean	SD	Min	Max
Total number of residential buildings exposed	1,262	3,107	2	25,604	57	327	1	9,761
Total area of res. land exposed (km2)	17	34	0	203	2	5	0	77
Total value of assets *	783	2,360	0	29,500	29	175	1	5,390
Land value (modelled) *	278	1,200	0	18,600	5	54	0	2,180
Building value (modelled) *	393	975	0	8,460	19	96	0	2,680
Appurtenant structure value (modelled) *	15	33	0	318	1	4	0	108
Contents value (modelled) *	98	243	0	2,060	5	25	0	718
Share of res. bldgs. located in urban areas	30	37	0	99	4	15	0	99
Inundation and landslip exposure measures								
Share of res. bldgs. in flood-prone areas	9	17	0	100	6	17	0	100
Share of res. bldgs. exposed to storm surge	2	9	0	100	1	8	0	100
Distance of res. bldgs. from big rivers (m)	4,419	3,987	60	29,819	5,117	4,570	4	43,212
Distance of res. bldgs. from small rivers (m)	295	260	26	2,519	354	817	0	12,064
Distance of res. bldgs. from lakes (m)	1,652	1,809	120	17,652	1,699	1,731	0	18,736
Distance of res. bldgs. from shoreline (m)	13,495	19,336	34	105,482	32,149	25,754	8	114,088
Share of res. bldgs. with landslip susceptibility	21	25	0	100	16	26	0	100
Share of res. bldgs. on soils with flood return periods from slight to very severe	17	26	0	100	18	30	0	100
Share of res. bldgs. on very poor to imperfect soil drainage	32	35	0	100	29	37	0	100
Share of res. bldgs. on soils with a 'slow' rate of water movement in saturated soil	6	16	0	100	6	18	0	100
Share of res. bldgs. on soils with very high to moderately high available water	33	37	0	100	28	39	0	100
Vulnerability measures								
Share of res. bldgs. with vulnerable materials	95	7	0	100	96	8	0	100
Share of res. bldgs. in deficient condition	22	12	0	100	21	17	0	100
Average elevation (above mean sea level)	99	111	2	741	214	185	1	2,336
Average slope	5	4	0	27	5	5	0	52
Average floor height (above ground)	1	0	0	1	1	0	0	2
Share of res. bldgs. located in areas with no agriculture	81	33	0	100	98	11	0	100

* Note: In Million 2017 NZ$. All the modelled values were constructed by the EQC.

Appendix Table 1.8.2: Projected Future Liabilities with all climate models for the changing hazard (in NZ$ Millions)

	One day of accumulated precipitation, 99th percentile						
	Climate models	**GFFL CM3(10) NOAA-USA**	**GISS-E2 R(14) NASA-USA**	**NorESM-M(9) NCC-Norway**	**HadGEM 2ES(2) MOHC-UK**	**CESM1 CAM5(1) NSF-USA**	**BCCCSM1.1(17) BCC-CHINA**
2020-2040	RCP 2.6	1181.9	1330.5	1342.4	1244.7	1191.6	1191.6
	RCP 4.5	1198.3	1334.4	1219	1193.6	1230.2	1230.2
	RCP 6.0	1215.5	1196.8	1353.9		1213	1213
	RCP 8.5	1099.9	1257.2	1347.2	1222.3	1234.2	1234.2
2040-2060	RCP 2.6	1169.7	1182.6	1343.2	1235.2	1306.8	1306.8
	RCP 4.5	1223	1367.4	1397	1203.6	1213.5	1213.5
	RCP 6.0	1211.8	1304.5	1292.3		1292.1	1292.1
	RCP 8.5	1245.5	1379.8	1321.7	1276.6	1299.8	1299.8
2060-2080	RCP 2.6	1285.3	1230.5	1365.9	1255.4	1361.1	1361.1
	RCP 4.5	1221	1252.7	1340.2	1308	1393	1393
	RCP 6.0	1240.5	1355.4	1432.7		1354.9	1354.9
	RCP 8.5	1359.1	1420.2	1485.8	1292.3	1467.6	1467.6
2080-2100	RCP 2.6	1223.3	1219.5	1291.8	1144.7	1337.9	1337.9
	RCP 4.5	1279.1	1306	1397	1144.2	1345.7	1345.7
	RCP 6.0	1268.9	1342.2	1299.8		1443.1	1443.1
	RCP 8.5	1359.1	1454.1	1473.1	1451.4	1463.8	1463.8

Note: Projected losses for 20-year aggregates for the 99th percentile value (p=99) and one day of accumulated precipitation (d=1), all Representative Concentration Pathways and all climate models. These results do not consider future changes in exposure or vulnerability. Results for the UK climate model and RCP 6.0 were dubious and thus not included in the table. The projected liability figures were inflated by a correction factor of 2.50. The need for an adjustment rises as a result of the claims omitted from the regression analysis. The factor is calculated such that we add the value of the claims included and the value of the claims omitted and divide that over the value of the claims omitted.

Appendix Table 1.8.3: Climate change signal mean and variability: Regional Climate Models' Ensemble.

Average climate change signal: 99th percentile one day of accumulated precipitation

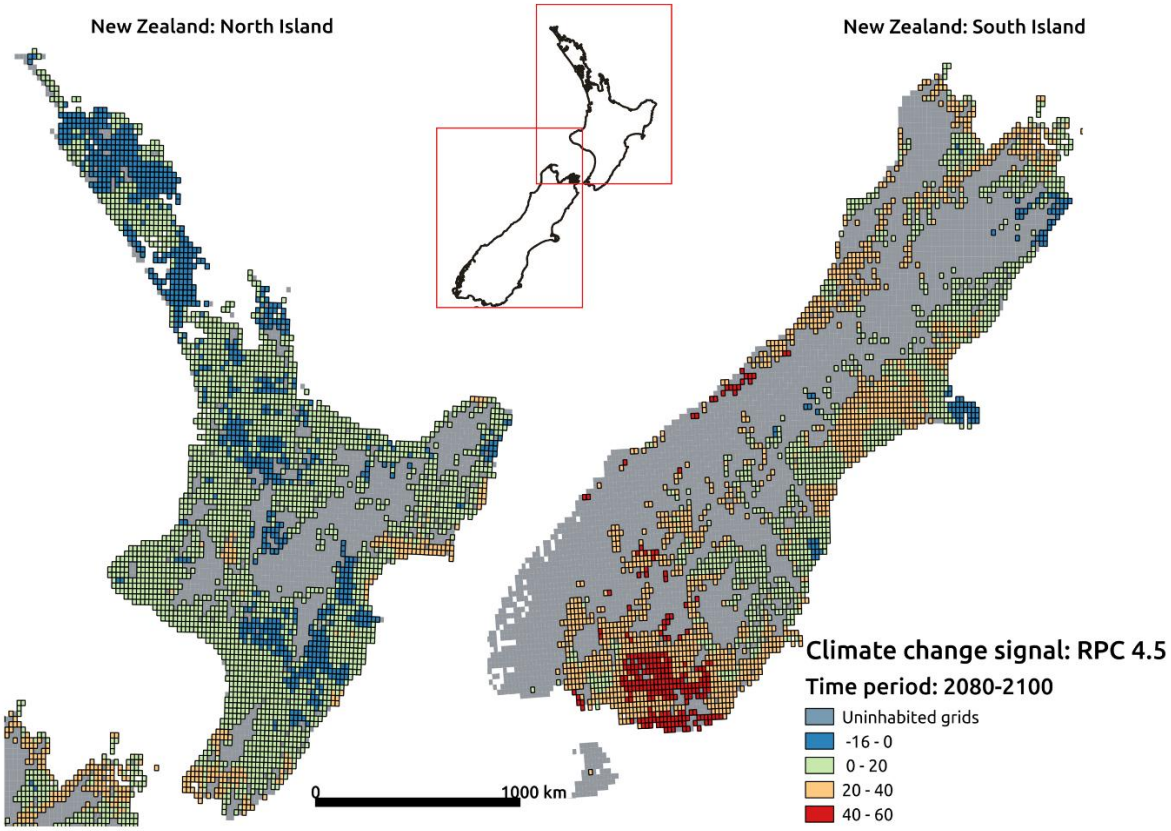
	2020-2040	2040-*2060	2060-2080	2080-2100
RCP2.6	8.37%	9.26%	13.83%	9.42%
RCP4.5	7.25%	10.33%	14.53%	13.22%
RCP6.0	7.62%	11.10%	17.10%	18.13%
RCP8.5	7.10%	13.30%	23.00%	25.49%

Climate change signal variability (standard deviation): 99th percentile one day of accumulated precipitation

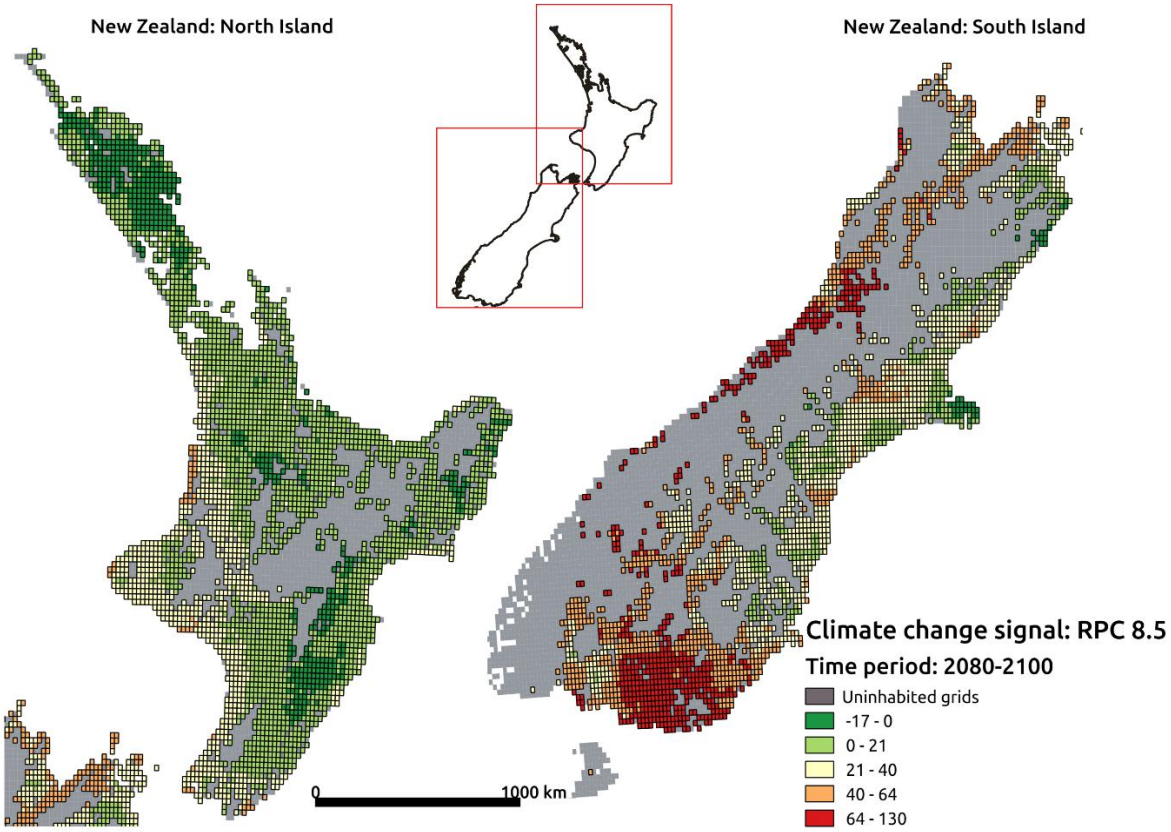
	2020-2040	2040-*2060	2060-2080	2080-2100
RCP2.6	6.34%	6.27%	5.25%	6.66%
RCP4.5	4.48%	7.66%	6.21%	7.61%
RCP6.0	5.66%	3.28%	5.99%	7.00%
RCP8.5	6.90%	3.95%	6.60%	3.69%

Climate change signal. The first table displays the average climate models' ensemble values while the second table shows the variability (standard deviation) around the mean.

Appendix Figure 1.8.4: Average Predicted Change in EQC liabilities for RCP 4.5 in every grid-cell (in %)



Appendix Figure 1.8.5: Average Predicted Change in EQC liabilities for RCP 8.5 in every grid-cell (in %)



Chapter Two

Mapping rainfall-induced landslide risk using insurance claim data

Abstract

Predicting the risk properties face from landslide is important for several purposes, including risk-based insurance pricing, land-use planning and managed retreats. Yet, rainfall-induced landslide hazard and risk maps are not available for most areas in New Zealand and typically are based on the availability of landslide inventories of past events, conventionally derived from satellite/aerial images or fieldwork. We focus on the region of Nelson (New Zealand) and address the lack of hazard information by using widely available property characteristics, combined with claim data from a particular event, to predict the expected monetary losses from landslides. We first implement a multivariate logistic regression model to quantify the relative effect of drivers that account for exposure and social and physical vulnerability on residential property damage. Then, we calculate the expected monetary losses (landslide risk) by factoring the likelihood of landslide damage derived from the regression model, property replacement values, and property vulnerability expressed as damage ratios. We find that risk levels are driven by moderate and high slopes, lower social deprivation levels, and high property values. The estimated average expected loses represents 90% of the actual cost of the event for the public insurer and range between an interval of 60 to 120 percent. We provide an additional estimate of the expected losses for a scenario without the effect of anthropogenic-driven climate change using a ‘Fraction of attributable risk- FAR’ study. The methodology we develop to produce a 'rainfall-induced landslide risk map at the property level' is useful for public and private insurers to identify potential losses or target-risk reduction efforts. The statistical significance and predictive power of the statistical model suggest that past insurance claim data can help circumvent the absence of rainfall-induced landslide hazard maps.

2.1 Introduction

The standard framework for undertaking a risk assessment of a natural hazard involves analyzing the interaction of three components: hazard data, the elements exposed to the hazard, and measures of these elements' vulnerability (understood as the susceptibility to harm or damage) to a hazard of a given intensity, spatial scope, and temporal frequency (Parise, 200l; Varnes, 1984, UNDRR, 2016). However, in many places, the risk assessment of rainfall-landslide hazard remains unquantified. In this paper, we develop a methodology that aims to fill this gap by using past insurance claims from the New Zealand public insurer to estimate rainfall-induced landslide susceptibility and risk from the impact of a 1-in-250-year precipitation event.

In New Zealand, the public insurer (the Earthquake Commission, EQC) provides cover against residential property damage that arises from some weather-related hazards, including rainfall-induced landslides. A landslide is the movement of a mass of rock, debris or soil down a slope under gravity's influence (Crude and Varnes, 1996; Guzzetti et al., 1999). Landslide hazard is a function of susceptibility (the spatial propensity for landslide activity) and the return period (the frequency) of landslide activity caused by wet (extreme rainfall) or dry (seismic or volcanic) events (Dai et al., 2002; Highland and Bobrowsky, 2008; UNDRR, 2016). Landslide susceptibility is a function of the topography, geology, hydrography, soil, land cover, geophysical processes, and human intervention in the landscapes (Soeters and van Westen, 1996). These variables are referred to as conditioning factors or predictors of landslide activity. Methods to assess the landslide susceptibility can be quantitative, qualitative, or a combination of both methods (Guzzetti et al., 2012)

We use a quantitative statistical approach to find the functional relationship between slope, conditioning factors, and the distribution of past insurance claims. We estimate a logistic regression model with two aims: 1) to identify the drivers of residential property damage; and 2) to estimate and map the expected monetary losses resulting from extreme rainfall-induced landslides.

Previous research that implements multivariate statistical methods (e.g. discriminant analysis, OLS or MLE regressions, support vector machine, neural networks) uses landslide inventories that are constructed using aerial or satellite imagery (e.g. Carrara et al., 1983; Pradhan, 2010; Santoso et al., 2011; Park et al., 2013). Instead, in this research, we use georeferenced insurance claim data. Unlike conventional landslide inventories, insurance records already exist and provide detailed information on the landslide location, date, and impact. Because of the high insurance penetration in New Zealand (~ 98%) (Nguyen and Noy, 2020), insurance claim data is a good proxy of landslides' costs. To the best of our knowledge, no existing research uses detailed insurance claim data to estimate the risk from a rainfall-induced landslide on residential property quantitatively.

Unlike prior research that implements multivariate logistic regression models to estimate landslide susceptibility (Ohlmacher and Davis, 2003; Lee, 2004; Ayalew and Yamagishi, 2005; Pradhan, 2010), we address the non-independence of observations -due to the spatial nature of the data- by clustering the errors. We cluster the errors using two different sets of clustering areas. We group homogenous households/buildings using enumeration units from the census for the first set of clusters. The second set of clusters groups buildings that are located in topographically homogenous areas.

Previous research papers that implement statistical methods to estimate landslide susceptibility incorporate combinations of continuous and discrete variables that include slope, geology, lithology, hydrology land cover, and soil, among others (e.g. Baeza and Corominas, 2001; Dong et al., 2009; Yao et al., 2008). In our research, in addition to the physical conditioning factors, we incorporate predictors that account for the sociodemographic characteristics of households by using a multidimensional indicator of well-being. Specifically, we use a social deprivation index that aggregates individual variables to reflect eight different social deprivation dimensions related to access to social services, employment status, qualifications, economic dependence and housing conditions (Salmond et al., 2007).

We showcase a risk assessment methodology of rainfall-induced landslide hazard in Nelson, the region of New Zealand with the highest number of weather-related claims and payouts relative to the region's population and residential asset stock. In this region, about two-thirds of the total value claimed between 2000-2017 was caused by the 2011 Golden Bay Storm. We focus on this event.

First, we identify a set of 29 predictors that capture the underlying and surrounding geophysical characteristics of residential properties and the sociodemographic traits of the households that inhabit them. Then, in a bivariate analysis, we test whether damaged and undamaged buildings are statistically different in each of their characteristics.

Second, we estimate a multivariate logistic regression model, where the dependent variable is an indicator/dummy variable that takes the value of one for damaged buildings and zero otherwise. The model includes the variables whose underlying distributions are statistically different for properties with and without damage. We implement different model specifications to identify the regression model with the best fit and performance measured by the R^2 and Area Under the Curve (AUC) statistics.

Third, we estimate the expected monetary losses, i.e., risk, by factoring the likelihood of damage (derived from the logistic regression model), the property replacement value, and the

property's vulnerability. Here, we define vulnerability as the degree of damage expressed as the ratio between the cost of repair and the replacement value of the property. We construct two damage-ratio measures as "upper and lower bounds" for the damage ratio based on the available data.

Finally, we map the expected monetary losses that the EQC could expect at the property level should an event of similar intensity impact Nelson's region. The expected losses are grouped using a classification method known as the Jenks Natural Breaks algorithm. The algorithm produces groups of data based on their similarity values such that the variance within groups is minimized, while the variance between groups is maximized. The map is a cartographic product that displays the risk levels in a sequential colour scheme so that lighter hues denote lower risk, and the darker hues indicate higher risk.

We find that slope, social deprivation, and the condition of being outside denser residential areas have the highest relative effects on landslide susceptibility. As measured by the R^2 (0.50) and AUC (0.98), the landslide susceptibility model's performance suggests that the model is capturing the physical and sociodemographic factors that make a property more likely to experience damage. However, when we integrate the likelihood of damage, along with property values and damage information, to calculate risk, we find that the highest risk is not only driven by higher likelihood values but is mainly driven by higher property values. We argue that households living in high-property values can afford geotechnical and engineering works that allow them to settle on coastal hills to enjoy desirable views.

This research's output is a detailed landslide risk map for the Nelson region at the property level. The map displays the spatial distribution of the expected monetary losses resulting from a 250-year extreme precipitation event. We demonstrate that insurance claim data can be used as an alternative to landslide inventories, based on aerial imagery or field observations, to evaluate landslide susceptibility using a statistical method. We note that statistical methods do not explain and describe the failure mechanisms, the physical set up of landslides, or the engineering mechanisms of property damage, so they are better suited to assess the damage's potential value and to identify priority locations for disaster risk reduction endeavours.

A spatial planner can use these results to introduce disaster risk reduction efforts in areas of high risk. Insurers can use them to inform risk-based insurance schemes and assess the potential financial liability brought about by extreme precipitation events. Local Councils could use this type of information to inform their land use planning and zoning of areas prone to damage. Other exposed locations currently not subject to landslips, or locations with similar geographies, could also employ

this methodology for predictions regarding their susceptibility to landslides triggered by low-probability precipitation events.

The paper remaining is organized as follows: Section 2.2 provides a literature review of the methods to estimate landslide susceptibility. Section 2.3 describes the New Zealand public insurance scheme concerning weather-related hazards and presents data on claims' spatial and temporal distribution. Section 2.4 describes the input datasets and variables derived and the results from the bivariate analysis. Section 2.5 presents the estimation method and the calculation of risk. Section 2.6 presents the regression models' results, evaluations of their statistical performance, and a discussion. Section 2.7 concludes with remarks about the findings.

2.2 Literature review

Soeters and Van Westen (1996) propose a classification of methods to assess the probability of land sliding: landslide inventory maps, deterministic or physical-based (geotechnical) models, heuristic-based methods, and statistical-based models. Guzzetti et al. (1999) group these methods into qualitative and quantitative approaches.

Landslide inventory maps are a collection of historical landslide occurrences. These maps display the spatial distribution of the number and size of historical events. Inventory maps are generally produced through photo interpretation using aerial and satellite imagery. Inventory maps are considered the most basic form of landslide mapping and form the basis for most landslide susceptibility maps (Wieczorek, 1984; Parise, 2001; Guzzetti et al., 2012).

Deterministic methods describe and model the physical laws that lead to slope instability. In this methods, "the physical properties of a particular slope are obtained from field investigations and laboratory tests" (Park et al., 2013; Montgomery and Dietrich, 1994; Terlien et al., 1995). Heuristic methods are the only qualitative method as they rely on expert judgement and knowledge of the physical processes acting upon the terrain. Here, the factors contributing to slope failure are ranked and weighted but do not get incorporated into a probability metric (Soeters and van Westen, 1996; Guzzeti et al., 1999). Statistical methods evaluate the relationship between landslide occurrence and the underlying and surrounding geophysical, topographic, hydrologic and land use characteristics (Soeters and Van Westen, 1996; Dai et al., 2002). Statistical methods aim to find the functional relationship between slope instability factors and the distribution of past and present landslides to predict areas susceptible to landslides (Guzzeti et al., 1999; Mahalingam et al., 2016).

Statistical methods for landslide hazard zoning can be implemented in a bivariate or multivariate modelling framework. Bivariate methods reported in the literature include weights of

evidence (WofE) and frequency ratio (FR) (Neuhauser, 2007; Yilmaz et al., 2012). Multivariate methods estimate models by linear and non-linear approaches via ordinary least squares (OLS) or maximum likelihood (ML) estimation. Logistic regression analysis, a non-linear ML method, has been used by Ohlmacher and Davis (2003), Lee (2004), Ayalew and Yamagishi (2005) and Pradhan (2010). Other multivariate methods include discriminant analysis (Carrara et al., 1983; Baeza and Corominas, 2001; Dong et al., 2009), Bayesian methods (Pham et al., 2016), random forests classification algorithms (Stumpf and Kerle, 2011), neural networks (Choi et al., 2010; Pradhan and Lee, 2010) and support vector machine methods (Yao et al., 2008; Marjanovic et al., 2011). Other papers develop a combination of deterministic and statistical approaches, where predictors for landslide susceptibility are obtained from laboratory tests and are then evaluated using probability models (Baum et al., 2005; Park et al., 2013; Santoso et al., 2011).

In New Zealand, Dellow (2011) proposes a probabilistic rainfall-induced landslide hazard model (PRILHM) framework based on previous works developed by Glade (1998) and Joyce et al. (2009). For a Wellington case study, Dellow (2011) calculates the probability of landslides based on the areal distribution of landslides in the landscape (i.e. frequency ratio method) using insurance claim data that has been temporally and spatially aggregated. The probability is derived by counting the number of landslide initiation sites per square kilometre for a given rainfall while controlling for slope bands, geology and vegetation (Dellow, 2011).

In contrast, we estimate property damage probabilities using a multivariate statistical non-linear model at the property level, where insurance claim data have not been spatially nor temporally aggregated. The detail of the insurance data allows us to identify the date when the damage occurred and thus enable us to link property damage to precipitation data. Furthermore, due to the great spatial detail of insurance claim data, we can exploit variation amongst properties across space using property-level information that account for the underlying and surrounding characteristics of residential property and their inhabitants.

The Geological and Nuclear Sciences Institute developed a landslide inventory of New Zealand. It is an open database that provides information on the triggering event, date, approximate location, size and damage of significant landslides (Rosser et al., 2017). Nevertheless, to the best of our knowledge, rainfall-induced landslide risk maps as such are nonexistent. Our work can be viewed as a pilot study to develop methods that can be used to create such risk maps.

2.3 Insurance and extreme events

2.3.1 Insurance

In New Zealand, residential property owners purchase public insurance to cover for physical damage occurring to buildings, the land beneath or around them, and home contents as a direct result of earthquakes, landslips, volcanic eruption, hydrothermal activity or tsunami. In the case of flood or storm events, public insurance only includes a provision for land damage. The Earthquake Commission (EQC) provides public insurance coverage to which homeowners pay a flat yearly premium. The EQC coverage is conditional on homeowners having private fire insurance; these private insurers are then liable for any damage to the building for flood/storm or any above the EQC cap cover for other types of insured hazard events (EQC Insurer's Guide, 2016; Earthquake Commission Act, 1993).5

Over the period 2000-2017, EQC's weather-related liabilities amount to a total of NZ$ 449.7 million (in 2017 NZ$). From the total value of EQC payouts, the highest share comes from land damage remediation (67%), followed by building damage (32%) and contents (1%).

The distribution of the EQC payouts over time is characterized by high peaks associated with extreme weather-events. The highest spikes of damages (see Figure 2. 1) occur in 2005 (Bay of Plenty and Waikato heavy flooding), in 2008 (North Island 'weather bomb'), and in 2011 (Tasman-Nelson heavy rain and flooding; Hawke's Bay flooding; and extra-tropical cyclone Wilma). See Fleming et al. (2018) for a list of the costliest weather events for the public and private insurance industry in New Zealand.

[5] This public cover is provided as a mandatory add-on to private insurance. Almost all owners purchase public/private home insurance policies (estimated take-up is 98% - see Nguyen and Noy, 2020). Changes in the coverage for the public component were introduced in July 2019 to elevate the building cover cap (from NZ$100,000 to NZ$150,000) and to remove the contents cover (previously capped at NZ$ 20,000). Land is covered up to its assessed market value. Land cover is capped as the EQC 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 4,000 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 cover as specified above are covered for indemnity value (EQC Insurer's Guide, 2016; Earthquake Commission Act, 1993)

Figure 4: EQC weather-related claim pay-outs over time

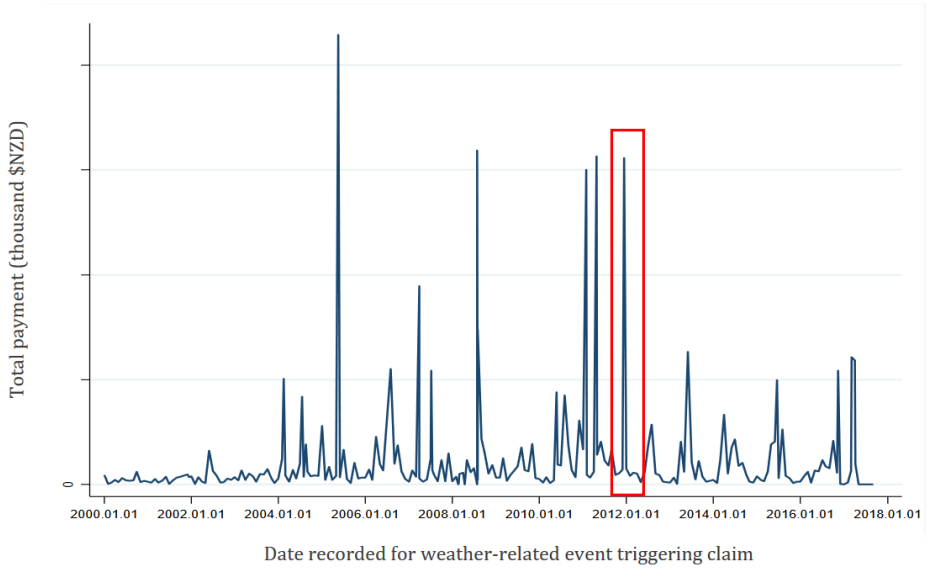


Figure 2. 1 **EQC's total weather-related claim payouts over time.** The highest levels of damages, i.e. the highest liabilities for New Zealand's public insurer, result from rainfall-induced landslides, floods, and storms in 2005, 2008 and 2011. The spike enclosed in the red rectangle corresponds to the Golden Bay Storm of 2011, which affected the Nelson and Tasman regions.

The distribution of EQC payouts across the regions of New Zealand is heterogeneous. For comparability purposes, we normalize the data by considering each region's exposure in terms of its population and residential property units. We calculate the number of claims per 1,000 properties and per 1,000 people; and the average weather-related claim payout per property ($) and per person ($). The results of normalizing the data allow us to rank each region's historical liability profile in relation to the impact of weather-related events. From the seventeen regions of New Zealand (see Table 2. 1), Nelson ranks number one as the region with the highest weather-related risk relative to its population and residential housing stock.

Region	No. of claims per 1,000 properties	No. of claims per 1,000 people	Avg. claim payout per property ($)	Avg. claim payout per person ($)
Nelson	48.97	18.13	661	245
Wellington	27.52	9.30	188	63
Tasman	22.62	8.55	230	87
Bay of Plenty	20.95	7.55	191	69
Northland	20.75	7.77	272	102
Hawke's Bay	13.24	4.62	177	62
Gisborne	12.18	4.00	222	73
Marlborough	11.22	4.96	146	65
Manawatu-W.	10.80	4.03	88	33
Otago	10.14	3.99	83	33
West Coast	9.10	4.00	66	29
Waikato	8.46	3.06	85	31
Taranaki	6.85	2.53	34	13
Auckland	6.64	1.89	92	26
Canterbury	5.07	1.83	33	12
Southland	0.57	0.22	2	1

Table 2. 1 **Ranking of normalized weather-related property damages in New Zealand's regions between 200-2017.** Nelson is the region with the highest number of claims and pay-outs relative to its population and residential asset stock. Taken from Fleming *et al.* (2019). See Appendix Figure 2.9.1 and Appendix Figure 2.9.2 for a graphical representation of this table.

Property owners in Nelson have received in pay-outs a total of 15.7 million NZ$. In Figure 2. 2, the time series of damages does not show an apparent trend. Instead, the time series is characterized by a large spike in 2011, where 97% of the damages in that year were caused by the Golden Bay storm in mid-December. We focus on this event. The descriptive statistics indicate that this event represents 62% of the total value of the claims paid by EQC to homeowners of the Nelson Region.[6] , where the average payouts for land and building damage were 18,253 and 8,644 NZ$ (in 2011 NZ$), respectively. In total, 352 claims were made to the EQC for this event.

[6] The private insurance industry paid 16.8 million NZ$ (in 2011 NZ$) as a result of the same event (Insurance Council of New Zealand, 2017).

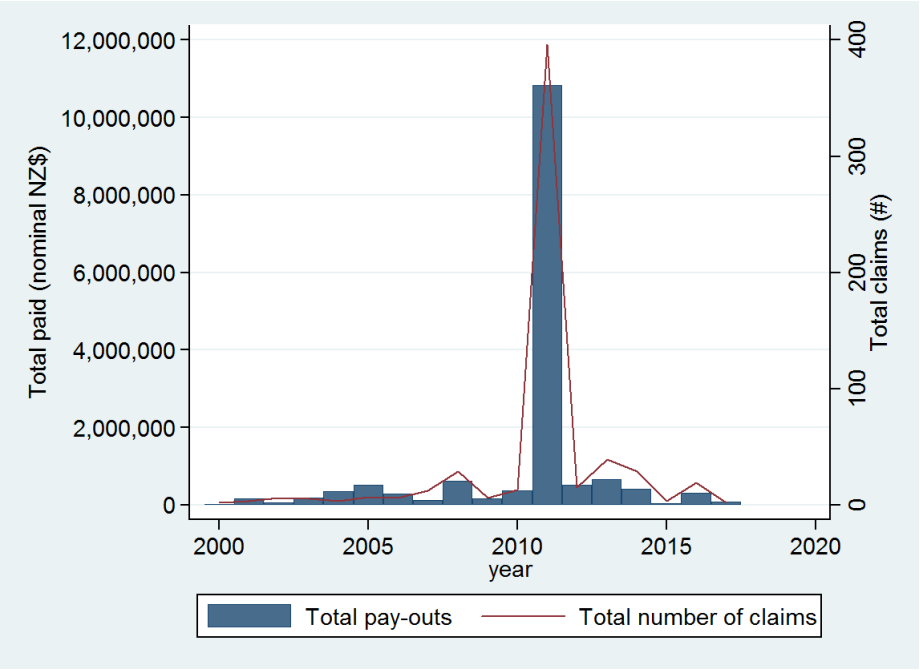


Figure 2. 2 **Nelson's time series of yearly insurance total claim payouts and the total number of claims from weather-related damage between 2000 and 2017.** The 2011 spike accounts for the damages paid by the EQC due to the 2011 Golden Bay storm event. The damages of this event represent 62% of the total value of the claims ever paid by the EQC to homeowners of the Nelson Region.

2.3.2 Extreme precipitation: The 2011 Golden Bay Storm

A low-pressure weather system triggered the Golden Bay storm event. Nelson region's precipitation levels were the highest 48-hour rainfall recorded in urban New Zealand until 2013 (Dean et al., 2013). The observed precipitation return period was estimated as a 250-year event for Nelson and a 500-year event for the Tasman-Takaka region (Ashraf and Jones, 2013). Figure 2. 3 depicts the spatial distribution of precipitation intensity (in mm) in the northern tip of the South Island over the 48-hour duration of the event. Unlike what happens in most of New Zealand's big storms, the air coming from the sub-tropics was unusually moist in the lower altitude, and as it rose over the hills, it cooled down near the coastline, rather than far inland, and the coastal foothills became saturated, and landslides occurred on waterlogged slopes (Dean et al., 2005; Terry, 2012)

Precipitation translated into landslides and floods that impacted road networks, water and sewage systems, and residential buildings. Saturated hillsides "gave away", triggering debris slides and creating dams that later burst (Terry, 2012). The damage concentrated on hills and very few low-lying areas, and most of the rain fell on coastal hills "such that none of the major rivers with headwaters in the mountains reached extreme flood levels" Dean et al. (2013).

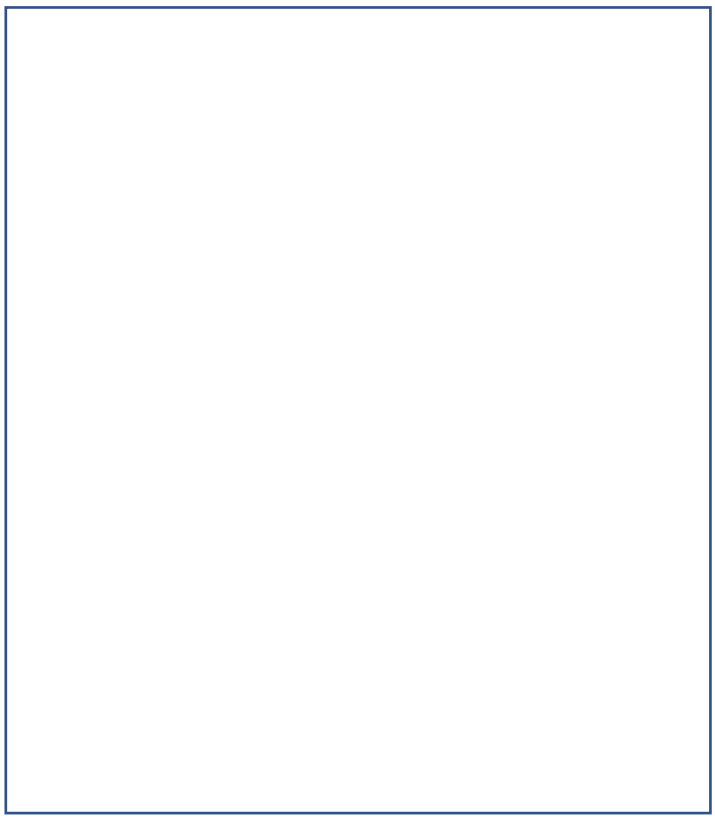


Figure 2. 3 **Isohyet 48-hour rainfall map of the 2011 Golden Bay Storm between 13 and 15 of December.** The precipitation return period was estimated to be a 250-year event for Nelson and a 500-year event for the Tasman-Takaka regions. Taken from Asbah and Jones (2013)

Although there were reports of swollen streams, "…the floodwaters were generally a result of streams becoming choked with debris from the slips, then spilling water as the debris dams released. On some occasions, very high surges occurred, with the water, mud, logs and silt travelling down at speed." (Terry, 2012). Asbah and Jones (2013) report 1,519 landslides with an average area of 170 m2 due to the event. Furthermore, news reports pointed at slips as the major problem as a direct manifestation of extreme precipitation (Stuff, 2011a; Stuff, 2011b). Two days after the event, 139 properties were issued yellow and red notifications, followed by building inspections geotechnical assessments carried by the EQC and the NT-CDEM (Terry, 2012; Stuff, 2011c). Houses on the slip-prone hills were rapidly issued with red stickers.

According to the Nelson Tasman Civil Defence Emergency Group (NT-CDEM), landslides are a significant threat in the area. Although hard rock types are dominant in the Golden Bay area, and high-intensity rainstorms do not lead to slope instability, "small scale superficial failures are common when the ground becomes water-saturated" (NT-CDEM, 2020). Figure 2. 4 shows the approximate spatial distribution of insurance claims lodged to the EQC and the residential housing stock distribution in the Nelson region and Nelson city.

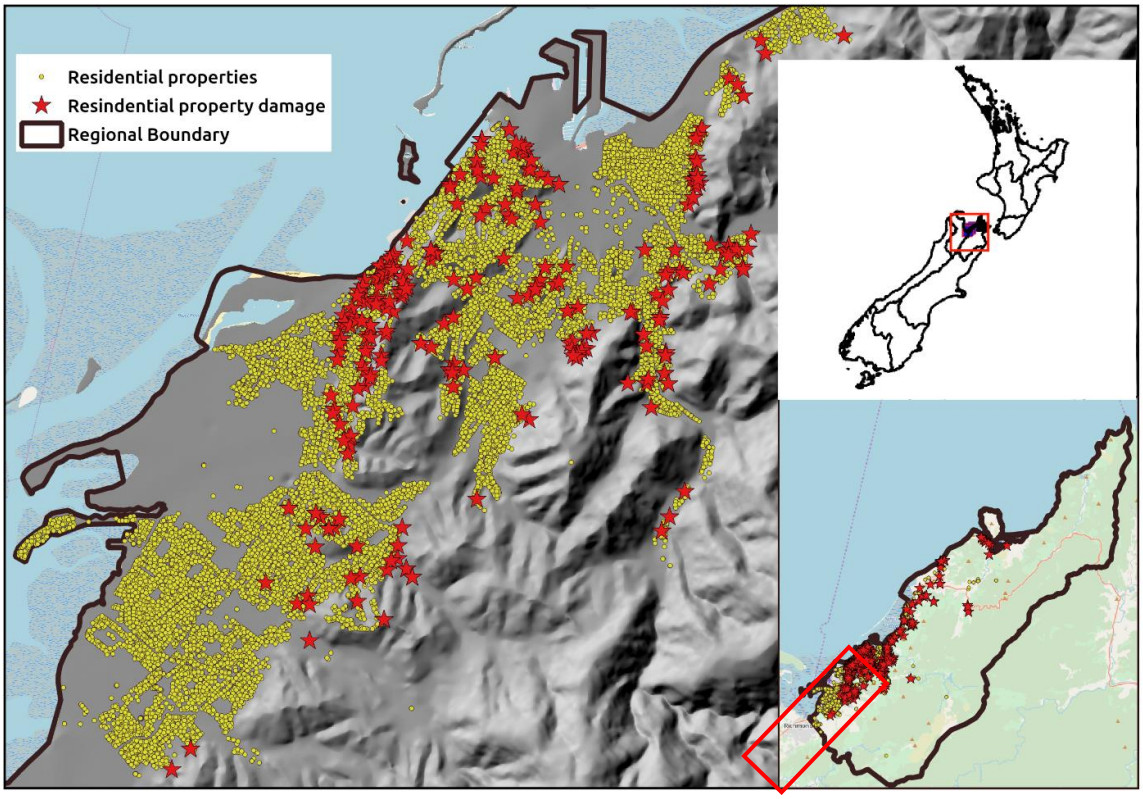


Figure 2. 4 **Spatial distribution of the residential housing stock and the approximate locations of weather-related property damage resulting from the 2011 Golden Bay storm in the Nelson region.** The impacts predominantly occur in non-flat areas and locations located along foothills. Figure elaborated by the authors.

2.4 Data and bivariate analysis

2.4.1 Data

We use a cross-sectional dataset of all individual weather-related insurance claims triggered by the 2011 Golden Bay Storm and a full set of residential building assets in the Nelson region. We create a profile for all buildings (with and without claims) in terms of their geophysical characteristics and their inhabitants' sociodemographic characteristics. The set of covariates we assess in the bivariate analysis are intended to capture conditioning factors for the observed damage. Furthermore, all the statistically significant variables at 0.05 level from the bi-variate analysis will be included in the regression model in section 2.5.1. Thus, the model will include the variables whose underlying distributions are statistically different for properties with and without damage. We group covariates into three categories for reporting purposes and based on the conceptualization of risk: hazard, exposure, and vulnerability (UNISDIR, 2016).

We can link information on each building to the surrounding and underlying environment, thanks to each residential property's geolocation information. We link geospatial data in two ways:

based on proximity and overlay.[7] Then, in a bivariate analysis, we test whether there are statistically significant differences in each risk factor between buildings with damage and without damage. We run different statistical tests depending on the measure of the variable (interval, nominal). Table 2. 2 reports the summary statistics of damaged and undamaged properties for each of the damage conditioning factors. In total, the number of properties with and without damage is 352 and 18,000, respectively.

The data we use are sourced from the Earthquake Commission (insurance claims), CoreLogic (property values), National Institute of Water and Atmospheric Research (precipitation, storm tide flood maps, pluvial flood maps, building asset inventory), Land Information New Zealand (topography, hydrography, land cover), Landcare Research (soil data), and from Salmond et al. (2007) (New Zealand's Social Deprivation Index).

2.4.2 Bivariate analysis

2.4.2.1 Precipitation

In New Zealand, precipitation data come from an extensive network of meteorological stations. Each of these stations takes measurements at the station-point level and are then converted into a continuous gridded surface (of 5x5 km^2) through interpolation (Tait et al., 2006). The nature of the precipitation data presents two limitations. First, interpolation smooths out extreme (high or low) observations. Second, because of the differences in spatial detail between the precipitation data (grid) and buildings data (point), about 82% of Nelson's residential property points fall within three grids. The remaining properties are scattered in a few other grids.

The bivariate analysis presented in Table 2.2 shows that differences in average precipitation between properties with damage (226mm) and without damage (236mm) are 10mm, and statistically significant.

2.4.2.2 Topography

We calculate the slope and elevation of the locations where a property is located using a high-resolution high-accuracy Lidar-based digital elevation model (DEM). The DEM's accuracy value is +/- 0.15 meters above mean sea level and a has grid resolution of 2x2m^2. Following Dellow (2011),

[7] Proximity can be measured as the distance between two features i.e. distance from a residential building point to a specific feature such as a shoreline or a stream. Here, the distance algorithm measures the shortest distance in meters assuming a two-dimensional space (i.e. Euclidean plane). Overlay is a measure that indicates the relationship between features that occupy or intersect the same geographic space. In the context of our study, the outcome of an overlay operation indicates whether a residential building point is inside or outside of an area (e.g. flood plain, soil type, slope).

we classify the slope values into 5-degree slope bands. We also evaluate the relationship between damage and slope, but without classifying the data into slope bands.

The results from the bivariate analysis in Table 2.2 show that, as the slope bands increase, the ratio between the proportion of properties with damage and the proportion of buildings without damage becomes larger, implying that households with steeper slope are more likely to experience damage. The differences in the proportion of properties with and without damage are statistically significant for all slope bands. These results support our intuition that the dummy variables represent a good classification of the slope values and the importance of slope itself in predicting landslide susceptibly. Figure 2.5 shows the proportion of properties with and without damage located in 5-degree slope bands

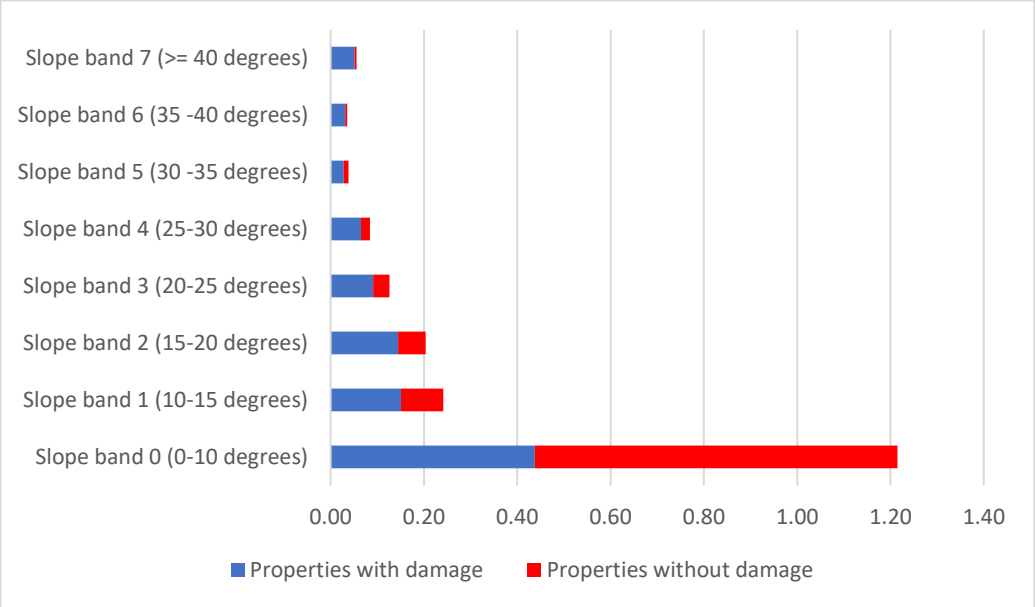


Figure 2. 5 **The proportion of properties with and without damage located in 5-degree slope bands**. As the slope bands increase, the ratio between the proportion of properties with damage and the proportion of properties without damage becomes larger, implying that households with steeper slope are more likely to experience damage. The differences in the proportion of properties with and without damage are statistically significant for all slope bands. Slope data was calculated using a Lidar-derived digital elevation model with an accuracy of +/-0.15 meters above mean sea level and a grid spatial resolution of $2x2m^2$.

2.4.2.3 Flood hazard

Flood maps delimit areas that are prone to experience pluvial, fluvial, and coastal inundation. Because the insurance dataset does not identify the identify the hazard that causes the damage, we evaluate the effect that flood hazard might have had on observed damage. To evaluate the relationship between damage and flood hazard, we use two flood hazard maps. The first map results from hydrological and hydraulic modelling and reflects a 100-year flooding event (Paulik et al., 2019a).

The second hazard map depicts storm tide inundation areas, reflecting a 100-year event (Paulik et al., 2019b). We also use a fluvial soil map of the region.

We associate residential properties with flood areas using two conceptualizations of the relationships in space, overlay and proximity. For instance, we identify whether a property is inside or outside the inundation area (overlay), but we also measure the distance (proximity) between a property and the flood map perimeter. The rationale for relating information using distance as a metric of proximity stems from field observations, which suggest that flood damage might have occurred in the vicinity of the flood hazard footprint areas. Furthermore, we use proximity because of the known inaccuracy of flood hazard maps reported in other places (e.g. Bernet et al., 2017).[8]

The bivariate analysis results in Table 2. 2 indicate that properties with damage are less likely to be located in 100-year riverine flood plains and are located further away from the flood perimeter. The proportion of properties without damage in 100-year flood plains is twice as large as the proportion of properties with damage located in 100-year flood plains. If we consider the proximity to flood plains, we see that properties with damage are, on average, twice as far from the flood plain perimeters than properties without damage.

As for coastal inundation, results show the proportion of properties without damages located within 100-year storm tide areas is about seven times higher than the proportion of properties with damage within storm tide areas. Finally, properties closer to fluvial soils are more likely to experience damage. See Figure 2.6, where we show the spatial extent of flood hazards and approximate property damage locations.

2.4.2.4 Exposure

We capture physical exposure by measuring the distance between properties to waterways (small and big rivers) and the shoreline; distance to an exotic and native forest; elevation above sea level; and the condition of being in urban areas and outside denser residential areas. By exposure, we also refer to the monetary values that make up the property's value, i.e., building, land, contents, and appurtenant structures.

Exposure metrics in relation to the elevation features show that properties with damage are located on higher ground (above mean sea level) than properties without damage. We find no difference in distance to the shoreline between properties with and without damage. However, properties with damage are closer to small and big rivers than properties without damage. In terms of

[8] In Chapter 3, I study a flooding event where I calculate that almost 92% of claims fall outside flood prone areas and only 0.08% of properties within flood-prone areas experienced damage as a result of flooding triggered by extreme precipitation.

land cover, properties closer to the native and exotic forest are more likely to experience damage. We see that damage is associated with being located outside denser residential areas and urban areas.

As for the exposure measured by monetary values, the bivariate analysis shows that the average building replacement value is 22% higher for properties with damage than for properties without damage. Similarly, the average value of appurtenant structures is 28% higher for properties with damage than for properties without damage. We find no statistical difference in land and contents exposure values.[9]

In Appendix Table 2.9.3, we provide further exposure summary statistics using soil classification data representing rural areas only. We associate property information with soil characteristics that include permeability profile, drainage capacity, water availability and soil flood return intervals. On average, properties without damage are located outside soils prone to inundation and further away from soils prone to inundation; in soils with 'moderate' drainage; in soils with 'moderate', 'moderately high' and 'high' water availability; and, in soils with 'moderate drainage.'

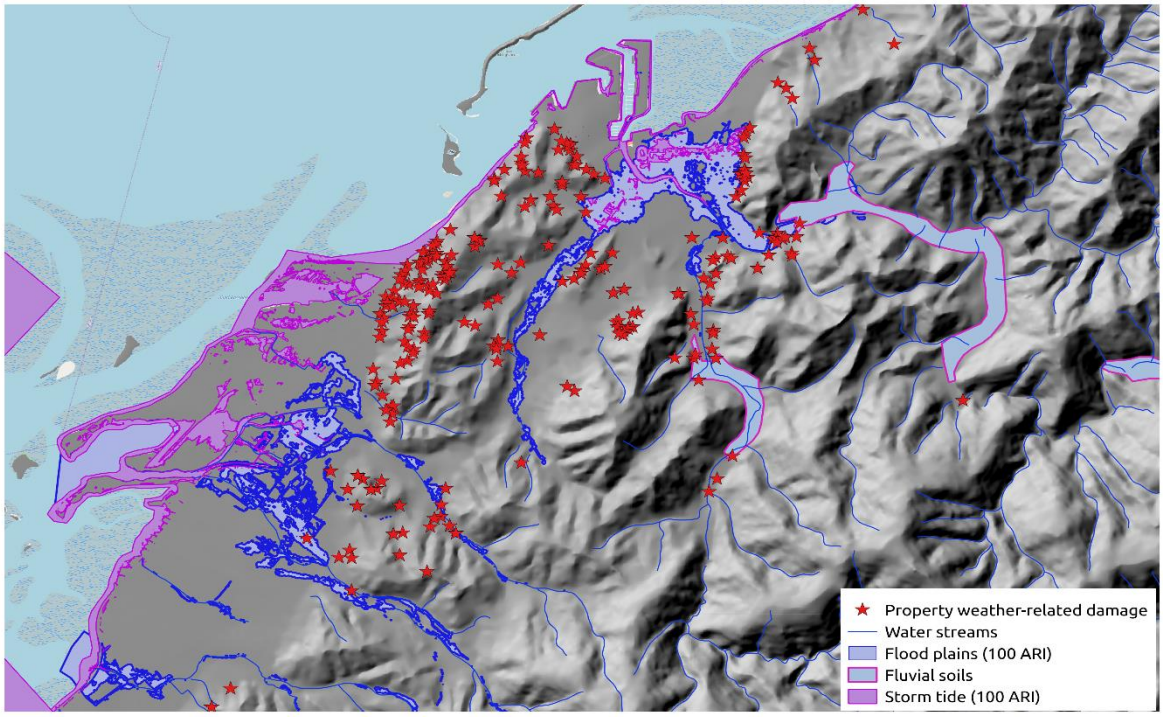


Figure 2. 6 **Flood hazard and property damage. It shows the spatial extent of storm tide and fluvial flood hazard and approximate property damage locations due to the 250-year Golden Bay storm.** From the bivariate analysis, property damage is likely to happen in properties that: fall outside and are further away from pluvial flood plains, are outside and further away from storm tide areas, and are closer to fluvial soils.

[9] The land and contents values are the result of modelling done by the EQC.

2.4.2.5 Vulnerability: physical and social

We develop a set of measures to proxy for the physical vulnerability of buildings. Physical vulnerability is captured by the buildings' characteristics, including construction materials, year of construction, the number of floors, floor height above ground, and condition (sound or deficient).

We also consider social vulnerability as measured by a multidimensional indicator of well-being. To account for social vulnerability, we use the New Zealand Deprivation Index (NZDI), which is constructed using data from the 2006 New Zealand Census. Salmond et al. (2007) describe the index as a multidimensional metric that aggregates individual variables to reflect eight different social deprivation dimensions. The index combines dimensions related to access to services, employment status, number of residents and space available, single-parent families' economic dependence, and qualifications. The index is provided as an ordinal scale with values that range from 1 to 10, where the former indicates "least deprivation" and the latter "most deprivation". The index is calculated for statistical areas of varying size called meshblocks, which contain multiple properties. We assign the meshblock-specific deprivation index to all the buildings in the same meshblock.

The bivariate analysis shows that properties with lower levels of social deprivation are more likely to experience damage. For instance, the proportion of damaged properties with a Deprivation Index of 1 is 2.73 times higher than the proportion of undamaged properties with a Deprivation Index of 1. As social deprivation levels increase, the ratio between the proportion of properties with damage and the proportion of buildings without damage becomes smaller, implying that households with higher social deprivation levels are less likely to experience damage. The differences between properties with and without damage are statistically significant except for the deprivation index 5. See figure 2.7.

The results for the physical vulnerability indicate no evidence of significant differences between properties with and without damage for the floor height (above ground), building condition (sound or deficient) or construction type (brick masonry, timber, concrete masonry, reinforced concrete shear wall, reinforced concrete moment resisting frame). Results show that, on average, taller and older buildings are more likely to experience damage.

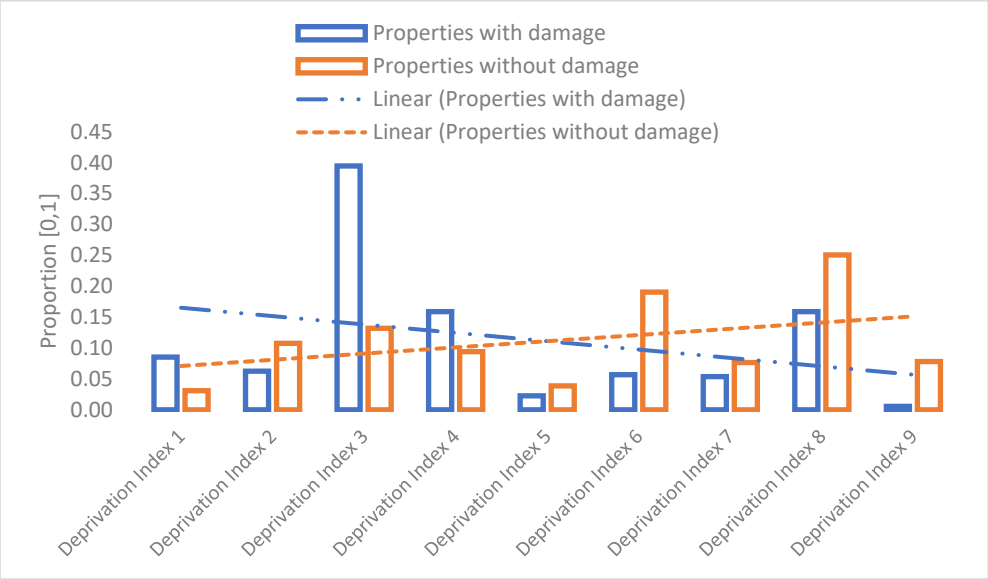


Figure 2. 7 **The proportion of properties with and without damage and social deprivation**. As social deprivation levels increase, the ratio between the proportion of properties with damage and the proportion of buildings without damage becomes smaller, implying that households with higher social deprivation levels are less likely to experience damage. The differences between properties with and without damage are statistically significant except for the deprivation index 5. The dashed lines indicate the linear trends of the proportion of properties with and without damages across Deprivation Index categories.

In summary, from the bivariate analysis, we find significant statistical evidence that property damage is likely to happen in properties that: are located in higher slope bands; fall outside and are further away from flood plains; are outside and further away from storm tide areas; are closer to fluvial soils and have lower rainfall. Damage is likely to happen in properties that are located in higher elevations (above mean sea level); are closer to small and big rivers/streams; are closer to the native and exotic forest. In terms of physical and social vulnerability, property damage is likely to happen in older properties (year of construction) and taller (number of stories) and properties whose occupants have among the lowest levels of social deprivation.

Sig. (*)		Residential properties **WITH** insurance claims					Residential properties **WITHOUT** insurance claims				
		Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std. Dev.	Min	Max
	Insurance										
	Land pay-out	352	18,253.36	32,462.24	0.00	298,000					
	Building pay-out (includes appurtenant structures)	352	8,644.08	22,180.31	0.00	114,000					
	Contents pay-out	352	499.65	2,277.94	0.00	22,800					
	Total pay-out	352	27,397.08	49,839.36	46.25	412,050					
	Number of claims (previous events)	352	0.06	0.29	0.00	2					
	Hazards										
	Precipitation										
*	Rain (mm)	352	225.97	14.96	203.10	256.30	18,000	235.82	18.65	178.20	256.30
	Fluvial and pluvial flooding										
	Properties in fluvial soils	352	0.05	0.23	0.00	1.00	18,000	0.04	0.20	0.00	1.00
*	Distance to fluvial soils	352	1,993.14	1,347.44	0.00	6,517.13	18,000	3,093.02	2,030.78	0.00	8,412.36
*	Properties in flood Plain (100-ARI)	352	0.07	0.26	0.00	1.00	18,000	0.13	0.33	0.00	1.00
*	Distance to flood plain (100 ARI)	352	452.14	416.38	0.00	2,408.51	18,000	262.96	307.47	0.00	4,544.09
	Coastal flooding										
*	Properties in storm surge (100-ARI)	352	0.00	0.05	0.00	1.00	18,000	0.02	0.14	0.00	1.00
	Distance to storm surge (100-ARI)	352	919.51	1,155.67	0.00	6,821.18	18,000	696.00	630.64	0.00	8,280.71
	Exposure										
	Property values (in 2011 NZ$)										
*	Building replacement value	352	338,246.40	223,849.60	66,739.20	1,578,226.00	18,000	275,335.90	176,487.80	16,684.80	2,413,757
*	Building value	343	320,441.70	127,279.50	55,908.00	939,793.00	450	347,037.80	151,221.60	78,317.00	1,650,658
*	Appurtenant structures value	343	12,727.71	19,110.22	0.00	104,272.00	450	9,875.08	20,534.18	0.00	298,094.00
	Land value 8-meter buffer	352	163,518.30	118,628.60	0.00	601,600.00	476	168,065.90	119,054.00	0.00	715,215.90
	Contents value	352	71,044.79	60,896.65	0.00	304,656.80	476	76,561.53	67,916.00	0.00	379,027.80
	Physical exposure										
	Topography										
*	Slope (in degrees)	352	14.37	11.56	0.236514	47.70194	18,000	6.38	7.84	0	87.20712
*	Elevation (m) - above mean sea level	352	49.55	37.82	8.04	259.71	18,000	32.15	26.38	0.00	625.43

Sig. (*)		Residential properties **WITH** insurance claims					Residential properties **WITHOUT** insurance claims				
		Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std. Dev.	Min	Max
	Hydrography										
	Distance to the shoreline (m)	352	1,158.57	1,014.74	32.45	5,876.10	18,000	1,128.46	763.71	0.29	8,467.81
*	Distance to small rivers (m)	352	297.36	245.76	3.22	1,133.55	18,000	318.43	234.43	0.05	1,204.56
*	Distance to big rivers (m)	352	2,336.56	1,613.33	13.76	8,090.45	18,000	3,630.40	2,415.76	2.01	10,043.08
	Land Cover										
*	Distance to native forest	352	528.40	324.71	0.00	1,313.75	18,000	575.54	317.99	0.00	1,720.54
*	Distance to exotic forest	352	513.41	366.19	0.00	1,597.10	18,000	762.83	447.15	0.00	2,052.98
*	Urban areas	352	0.50	0.50	0.00	1.00	18,000	0.60	0.49	0.00	1.00
*	Outside denser residential areas	352	0.85	0.36	0.00	1.00	18,000	0.98	0.14	0	1
	Vulnerability										
	Physical vulnerability										
*	Number of floors	352	1.30	0.49	1.00	3.00	18,000	1.21	0.42	1.00	10.00
*	Year of construction (<=1960)	352	0.42	0.49	0.00	1.00	18,000	0.32	0.46	0.00	1.00
*	Year of construction (>1960 & year_1<1980)	352	0.19	0.39	0.00	1.00	18,000	0.26	0.44	0.00	1.00
	Floor height (m) – above ground	352	0.60	0.08	0.20	0.74	18,000	0.60	0.24	0.20	8.00
	Building condition (sound)	352	0.74	0.44	0.00	1.00	18,000	0.72	0.45	0.00	1.00
	Building type (Reinforced concrete sheer wall)	352	0.04	0.20	0.00	1.00	18,000	0.04	0.21	0.00	1.00
	Building type (Reinforced concrete moment resisting frame)	352	0.00	0.00	0.00	0.00	18,000	0.00	0.01	0.00	1.00
	Building type (Timber)	352	0.91	0.29	0.00	1.00	18,000	0.91	0.28	0.00	1.00
	Building type (Brick masonry)	352	0.00	0.05	0.00	1.00	18,000	0.00	0.05	0.00	1.00
	Building type (Concrete masonry)	352	0.05	0.21	0.00	1.00	18,000	0.04	0.20	0.00	1.00

Table 2. 2 **Bivariate analysis. The table compares residential properties in terms of their physical and sociodemographic characteristics for two subgroups properties: properties without damage and properties with damage resulting from a 250-year extreme precipitation event: the 2011 Golden Bay Storm.** The information is grouped such it reflects the components of risk: hazard, exposure, and vulnerability. Significant variables are marked with an asterisk [*]. The statistical significance test (at 0.05) is calculated with different methods depending on the variable's measure (interval, nominal).

2.5 Methods

In Nelson, rainfall-induced landslip hazard mapping is currently nonexistent. To fill this gap, we first develop a rainfall-induced landslip susceptibility logistic regression model using insurance claim data. We aim to find the functional relationship between slope instability factors and past landslides' distribution to predict properties susceptible to damage. In a second step, we calculate the expected monetary losses (i.e. risk) by factoring the likelihoods produced from the regression model, property replacement values, and vulnerability of the buildings – expressed as damage ratios. Finally, we map at the property level the spatial distribution of risk.

2.5.1 Regression analysis

We estimate a logistic model of residential property damage with two aims. First, to identify the risk factors associated with residential property damage; and second, to predict the probability, i.e. likelihood that a weather-related claim will be made to the New Zealand public insurer in the event of a 250-year precipitation event. We use the following model:

$$LC_i = \frac{e^{\beta_1 Haz_i + \beta_2 Exp_i + \beta_3 Vul_i + \epsilon_i}}{1 + e^{\beta_1 Haz_i + \beta_2 Exp_i + \beta_3 Vul_i + \epsilon_i}} \qquad (1)$$

where LC_i is a binary variable that indicates whether an insurance claim has been made by property *i*. The terms Haz_i, Exp_i *and* Vul_i are vectors that account for hazard, exposure and vulnerability of property *i*. The model includes all the statistically significant variables at 0.05 level from the bi-variate analysis reported in Section 2.4.2. The components (variables) that make up each term in equation one are reported in Table 2.3. We refer to this regression as the 'full model'. We also implement a stepwise approach that gradually eliminates variables from the 'full model' based on the regression coefficients' statistical significance values (i.e. 0.10 and 0.05). In total, we implement three logistic regressions to identify the regression model that best explains the data in terms of the magnitude of the R^2 and the Area Under the Curve (AUC) statistic. The regression coefficients are expressed as odds ratios.

To address multicollinearity among the independent variables, the models exclude any variable with inflation factors (VIF) exceeding 4 (Hair et al., 1998). We cluster the errors by two different geographic areas to address the non-independence of observations due to the spatial nature of the data and autocorrelation in the errors. By clustering the errors, we allow for intragroup correlation, thus relaxing the usual requirement that the observations be independent. Thus the observations are independent across groups (clusters) but not necessarily within groups. We cluster the errors by meshblock, which are small geographical areas, and in general, households within a

meshblock will be relatively similar in terms of their sociodemographic characteristics (StatsNZ, 2020). We cluster the error by areas topographically homogenous. We create the latter using building-level slope information to create a continuous surface by interpolating the properties' location slope values over a regular gridded surface of 100x100m^2 (QGIS 2.18.16, 2020a). The outcome is geographic units that group relatively similar buildings in terms of the steepness of the terrain.

2.5.2 Expected damages (risk)

This section describes the calculation of expected monetary losses at the property level to ultimately map the spatial distribution of rainfall-induced landslide risk. To this end, we factor the likelihoods produced by the regression model, property replacement values, and damage ratios, such that:

$$ExpectedDamage_i = \widehat{LC_i} * RV_i * DR_i \qquad (2)$$

where *i* is a residential property, $\widehat{LC_i}$ is the estimated likelihood of damage, RV_i is the building replacement value, and DR_i is the damage ratio.

The damage that a property experiences in the event of a slide depends on the landslide type and intensity (e.g. volume and velocity of sliding, Area affected, etc.) and the property characteristics (floor materials, wall materials, retaining walls, etc.). For a given landslide type and intensity, damage can be expressed as the cost to repair or the ratio between the cost of repair and the property's replacement value, i.e., damage ratio. A damage ratio is measured on a scale of 0 to 1, where 0 means no damage and one represents that the damages are at least equal to the building's replacement value (Meyer et al., 2013).

We construct two versions of the damage-ratio term, and thus we produce two estimates of risk given the data availability for the second term of the equation (2). Specifically, for the replacement value (RV_i), we can observe property values for only a small subset of records in the dataset. Here, the property values reflect the land, appurtenant structures and buildings. In contrast, we have data on the building replacement value for all the records, but these data do not reflect all the other components that make up the residential asset's value.

In the first version of the damage ratio, the denominator includes the entire residential property asset's value. In the second version, the denominator includes only the building replacement value. The numerator in both versions is the total cost of repair (i.e. total payouts). The first and

second versions of the damage-ratio have mean values of 0.057 and 0.114, respectively.[10] We calculate the expected losses for all building assets within the Nelson region and not for only for the properties that lodged a claim. The rationale is that all properties experienced extreme precipitation, and will thus have a non-zero probability of damage.

2.6 Results and Discussion

2.6.1 Results

The relationship between property damage and the set of predictors for all model specifications is not due to chance (as shown by the Log-likelihood significance Chi-2 figures). The R^2 values, around 0.50, tell us how much of the variation in the outcome variable can be explained. The results are shown in Table 2.3.

The main conditioning factors associated with landslide susceptibility resulting from a 250-year precipitation event are slope, social vulnerability and the condition of being located outside denser residential areas. The odds of experiencing damage for properties located in the slope band 10-15 degrees is twice as large as the odds of a building sitting on flatter terrain with a slope band between 0-10 degrees, the reference category.[11] The magnitude of the odds ratio of damage increases from 2 to 5 as the slope band increases. For instance, the odds of damage for a property sitting in a slope band of 35 to 40 degrees is 5.29 times as large as the odds of a property sitting in the reference category. Overall, five out of the seven slope bands display statistically significant effects on the likelihood of experiencing property damage from landslide when compared to the reference category.

The influence of social deprivation on the likelihood of damage is statistically significant for four out of eight social deprivation levels. For the lower levels of social deprivation, that is, properties with deprivation indexes of 3 and 4 (DI 3 and DI 4), the odds of property damage are 7.11 and 11.11 times as large as the odds of property damage in a household in the properties with deprivation index 9, respectively. Here, the reference category is DI 9, which reflects the most socially deprived households. The impact of social vulnerability on the likelihood of damage is also statistically significant such that the odds of property damage for households with DI 7 is 9.17 times as large as

[10] Buxton et al. (2013) simulate damage ratios from rainfall induced landslides for different building types for all of New Zealand using EQC claim data. For timber buildings, which make up 91% of Nelson's building stock, they estimate an average damage ratio of 0.047. For concrete buildings, which make the remaining 9% of Nelson's building stock, they estimate a damage ratio of 0.46. On average, the damage ratio estimates from Buxton et al. (2013) vary from the first version of our damage ratio by 0.009.

[11] The reference category (0-10 degrees) is chosen based on Dellow (2011) who finds that the average probability of landslip for the slopes bands 0-5 and 5-10 is zero regardless of the rain intensity. The estimates are computed for a case study in Wellington – New Zealand.

the odds of property damage of households with DI 9, and the odds of DI 8 are 4.92 times larger than the odds of DI9. We claim that this relationship is due to visual amenities. We argue that households living in high-property values can afford geotechnical and engineering works that allow them to settle on coastal hills with steep slopes prone to failure.

In terms of exposure, we see that damage is more likely to occur to more expensive buildings. As for the relationship between damage and flood hazard, we find that the condition of being located within storm tide or pluvial flood-prone areas is unrelated to the odds of property damage. However, the distance to the perimeters of fluvial flood plains increases the odds of property damage by 1.07 for every additional meter. The odds of property damage per meter of distance to small rivers increase by 0.89. In section 2.6.2, we discuss the role of flood hazard in detail and explain these results.

Independent variable (full model) Binary: Claim = 1, No claim = 0	(1)	(2)
Dependent variables		
Slope band (0-10 degrees) omitted category		
Slope band 1 (10-15 degrees)	2.064***	2.064***
	(0.467)	(0.512)
Slope band 2 (15-20 degrees)	2.174**	2.174***
	(0.709)	(0.647)
Slope band 3 (20-25 degrees)	1.431	1.431
	(0.619)	(0.597)
Slope band 4 (25-30 degrees)	3.068**	3.068***
	(1.372)	(1.318)
Slope band 5 (30 -35 degrees)	1.552	1.552
	(1.033)	(1.009)
Slope band 6 (35 -40 degrees)	5.291***	5.291***
	(2.755)	(2.679)
Slope band 7 (>= 40 degrees)	4.703***	4.703***
	(2.630)	(2.434)
Rain (mm)	0.980**	0.980**
	(0.009)	(0.008)
Properties in flood Plain (100-ARI)	0.643	0.643
	(0.344)	(0.316)
Distance to flood plain (100 ARI) (in hundreds of meters)	1.076***	1.076***
	(0.030)	(0.028)
Properties in storm tide (100-ARI)	0.620	0.620
	(0.454)	(0.563)
Building value (in tens of thousands NZ$)	1.132***	1.132***
	(0.009)	(0.008)
Elevation (m) - above mean sea level	0.990**	0.990**
	(0.005)	(0.004)
Distance to small rivers (in hundreds of meters)	0.890**	0.890**

	(0.047)	(0.046)
Distance to exotic forest (in hundreds of meters)	0.952	0.952
	(0.031)	(0.029)
Distance to native forest (in hundreds of meters)	0.951	0.951
	(0.053)	(0.046)
Properties in outside denser residential areas	2.693***	2.693**
	(0.975)	(1.157)
Properties in urban areas	0.728	0.728
	(0.175)	(0.176)
Deprivation Index 1	1.600	1.600
	(1.521)	(1.638)
Deprivation Index 2	0.774	0.774
	(0.760)	(0.743)
Deprivation Index 3	7.112**	7.112**
	(5.476)	(5.467)
Deprivation Index 4	11.117***	11.117***
	(8.914)	(8.770)
Deprivation Index 5	1.585	1.585
	(1.332)	(1.368)
Deprivation Index 6	3.298	3.298
	(2.637)	(2.706)
Deprivation Index 7	9.717**	9.717***
	(8.786)	(8.523)
Deprivation Index 8	4.920**	4.920**
	(3.678)	(3.660)
Deprivation Index 9		
(omitted category)		
Number of floors	0.578**	0.578**
	(0.133)	(0.137)
Year of construction (<=1960)	1.006	1.006
	(0.215)	(0.231)
Year of construction (>1960 & year_1<1980)	0.756	0.756
	(0.205)	(0.193)
Constant	0.704	0.704
	(1.426)	(1.299)
Observations	18,277	18,275
Chi2-value	840.1	744
Prob_chi2	0	0
Number of clusters	363	2144
Pseudo R-squared	0.501	0.501

Table 2. 3 **Logistic regression results: 'full model'.** Only statistically significant variables (at 0.05) from the bivariate analysis are included in the model. The results in columns (1) and (2) differ in the make-up of the areas that the standard errors are clustered for. The first geographic unit (known as meshblock), which are small geographical areas, and in general, households within a meshblock will be relatively similar in terms of their sociodemographic characteristics. The second geographic unit delimits areas that are topographically homogenous in terms of the steepness of the terrain. The results are very similar for both clustering methods (i.e., the difference in the magnitude of the standard errors in column (1) and (2) is minimal). The coefficients are expressed as odds ratios, and the clustered standard errors are shown in parentheses for the significance levels *** p<0.01, ** p<0.05, * p<0.1

As for extreme precipitation, the odds of property damage per unit-increase in rain intensity increase by 0.98. Despite the significant negative effect of rain on the likelihood of property damage, we argue that the effect is rather small and mainly explained by the data's resolution. In section 2.4.1,

we argued that the precipitation data's resolution was too coarse to exploit variation across records since 82% of the residential stock falls within three grids.

The condition of being located outside denser residential areas also plays a significant role in the likelihood of property damage. For properties located outside residential areas, the odds of damage are 2.69 as large as the odds of buildings located inside denser residential areas. The odds of property damage per additional floor unit increase by 0.57. In terms of exposure, the odds of property damage are 1.12 per unit increase in the property's value. The results from the stepwise regression models are reported in Appendix Table 2.9.4 and Appendix Table 2.9.5.

2.6.1.1 Evaluation of the models' performance

We assess the models' statistical significance by evaluating the Log-likelihood value and its associated Chi-2 value. To assess the models' predictive power, we look at the size of the R^2. Also, to compare the models' performance, we produce a Receiver Operating Characteristics (ROC) graph and its associated area Under the Curve (AUC) value, based on a series of confusion matrices.

The results produced by the probabilistic modelling are statistically significant based on the Chi2 values at significance levels >99.9%. The R^2 values outputted by the models (Table 2. 4) show figures between 0.49 and 0.51. The R^2 is the proportion between the intercept-only model's log-likelihood and the log-likelihood of a model with all the predictor variables. The R^2 values range from 0 to 1, where higher values indicate a better fit of the model.

A confusion matrix accounts for a classification model's performance and evaluates the observed and the model's predicted state, i.e., property damage. A confusion matrix classifies a property in the "damaged group" or the "undamaged group" depending on the logistic regression model's likelihoods. It is based on a conventionally set threshold at 0.5. Observations with likelihoods above the threshold are classified as 'damaged', and observations below the threshold are classified as undamaged. Two model performance metrics can be calculated from the confusion matrix: Sensitivity or true positive rate (TPR) and specificity or true negative rate (TNR).

Sensitivity and specificity show the proportion of correctly classified records by the model based on the threshold value of 0.5. Sensitivity shows the percentage of properties that had damage and were correctly classified by the models. Similarly, sensitivity shows the percentage of properties that had no damage and were correctly classified by the models. We use a threshold value of 0.0192 (~2%), which is the share of properties in the Nelson region that experienced damage (i.e., 352 properties out 18,352). The results in Table 2. 4 shows that sensitivity and specificity values are around 95% for all models.

A Receiver Operating Characteristics (ROC) graph summarizes confusion matrices that result from setting different thresholds. Thus, a ROC graph avoids the arbitrariness of setting a classification cutoff to evaluate the models' performance. A ROC graph uses pair-wise combinations of sensitivity and the false positive rate (FPR = 1 – specificity) for a range of thresholds. We calculate the Area Under the Curve (AUC) to select the model with the best performance from the resulting graph. A model with no predictive power has an AUC value of 0.5, whereas a perfect model AUC of 1. The results from the models show AUC values that range from 0.9780 to 0.9808. Therefore, the three models perform uniformly well, but the highest variation explained by a model is achieved with the 'full model'. Table 2.4 provides a summary of the models' performance.

	Performance metrics of the logistic regression models							
	Log-likelihood	Degrees of freedom	Chi-2	Significance	R-squared	Area under curve (AUC)	True positive rate (TPR)	False positive rate (FPR)
Full model	-8045.65	29	920.69	>0.999	0.5013	0.9780	95.02	94.91
Full model stepwise regression 0.10	-812.60	13	670.8	>0.999	0.4970	0.9793	95.95	95.31
Full model stepwise regression 0.05	-825.08	12	656.56	>0.999	0.498	0.9808	95.95	95.61

Table 2. 4 **Performance metrics of three logistic regression models.** All the models are statically significant at the 0.001 level. The R2 values range from 0.491 to 0.508 the AUC values are nearly 1, suggesting a good model performance to predict the likelihood of rainfall-induced property damage.

To compare our results with the actual monetary losses, we compare the actual payouts made by the EQC and our estimates of risk. Depending on the version of the damage ratio employed, the event's estimated cost represents 70% and 120% of the actual cost depending on the version of the damage ratio used. The average estimate of the cost using both damage ratios represent 97% of the event's actual cost. Appendix Table 2.9.6 reports these calculations.

In Appendix 2.9.7 we provide a table with aggregated losses with and without the effect of climate change. According to Dean et al. (2013), 5% of the 2011 Golden Bay storm event's intensity can be attributed to the influence of anthropogenic climate change. Dean et al. (2013) estimate the effect of climate change by simulating the conditions of the world's weather and climate with and without anthropogenic greenhouse gasses so that inferences about the effect of human activity on climate can be discerned. Specifically, Dean et al. (2013) attribute the effect of climate change by

> *'using a methodology that is capable of identifying all occurrences of the synoptic situation matching the Golden Bay/Nelson event within large ensembles of climate-model simulations that alternatively exclude and include the impact of GHGs. If the observed precipitation distribution for these events is well simulated by the model for the past climate, then any change in the rainfall distribution due to GHGs can be considered.*'

2.6.2 Discussion

This section discusses the results in light of the statistical analysis and their cartographic representation to provide context and explain the spatial distribution of risk from rainfall-induced landslides in the Nelson region and city. The discussion is further informed by knowledge of the areas impacted and accounts of the storm event.

Figure 2.8 displays the spatial distribution of the average expected damages at the property level across the Nelson region. The highest damages (coloured in black and purple points), as defined by the class values [4,975 - 12,411] and [12,411 - 27,299] (NZ$), occur in areas with high slopes, lower levels of social deprivation, high building replacement values, and outside denser residential locations.

The first group of properties with the highest risk levels is an urban settlement located on steep hills overlooking Tasman Bay and the Southern Alps (enclosed in a blue ellipsoid in Figure 2.8). The group overlaps predominantly, although it is not restricted to the suburbs of Tasman and Britannia Heights. These properties display high levels of risk determined by the high building values, which we argue are mainly due to desirable views. We argue that households living in high-property values can afford geotechnical and engineering works that allow them to settle on coastal hills with steep slopes prone to failure. We thus claim that visual amenities play a role in the risk of rainfall-induced landslides in Nelson. It is worth noting that the households that self-select into landslide-prone areas represent a higher liability for the EQC, although their contribution to the insurance pool of collected yearly premiums is the same as households' contribution in lower property values. This finding implies that the EQC's exposure to weather-related risk is higher for better-off households (as measured by the property's value). Evidence of the EQC insurance scheme's regressive nature has been reported by Owen and Noy (2019). They show that mean household income and claim payouts from the Canterbury Earthquakes impact are positively related.

In addition to high property values, we note that high-risk levels are also explained by higher probability predictions driven by large coefficients for lower social deprivation levels and higher slopes. Moreover, higher probability estimates are also explained by the condition of being located outside denser residential areas. Figure 2.8 shows groups of buildings with the highest risk levels (enclosed in black rectangles), which spread outside denser residential areas, lower social deprivation levels, and sit on moderate slopes on coastal hills or relatively flat grounds upon the foothills.

The map's scale in figure 2.8 might give the impression that the highest risk levels might be related to coastal flooding due to their apparent proximity to the shoreline. We argue that coastal

flooding, as measured by proximity to storm tide areas (or by being inside a storm tide area), played no role in the impact of extreme precipitation on property damage. We discuss two possible mechanisms that explain the absence of the impact of coastal flooding on property damages. First, Nelson has the largest tidal range in New Zealand (LINZ, 2020). The tidal range is defined as the area between the highest and lowest tide marks. It follows that the smaller the tidal range, the more exposed buildings are to storm tide. This is because there is a bigger chance that the storm tide will be bigger than the tidal range. Second, the downpour during the Golden Bay storm was described to have occurred within 2 to 5 km of the shoreline (Dean et al., 2013), and thus we argue that damages from a swollen sea are unlikely to have caused damage. In addition to these two factors, no historical records of the event (Dean et al., 2013; Ashbah and Jones, 2013; Terry, 2012; Stuff, 2011a; Stuff, 2011b, Stuff, 2011c; Stuff, 2013) report on the role of coastal flooding in the levels of damages experienced. Thus, we dismiss the hypothesis that the model incorporates records triggered by coastal flooding.

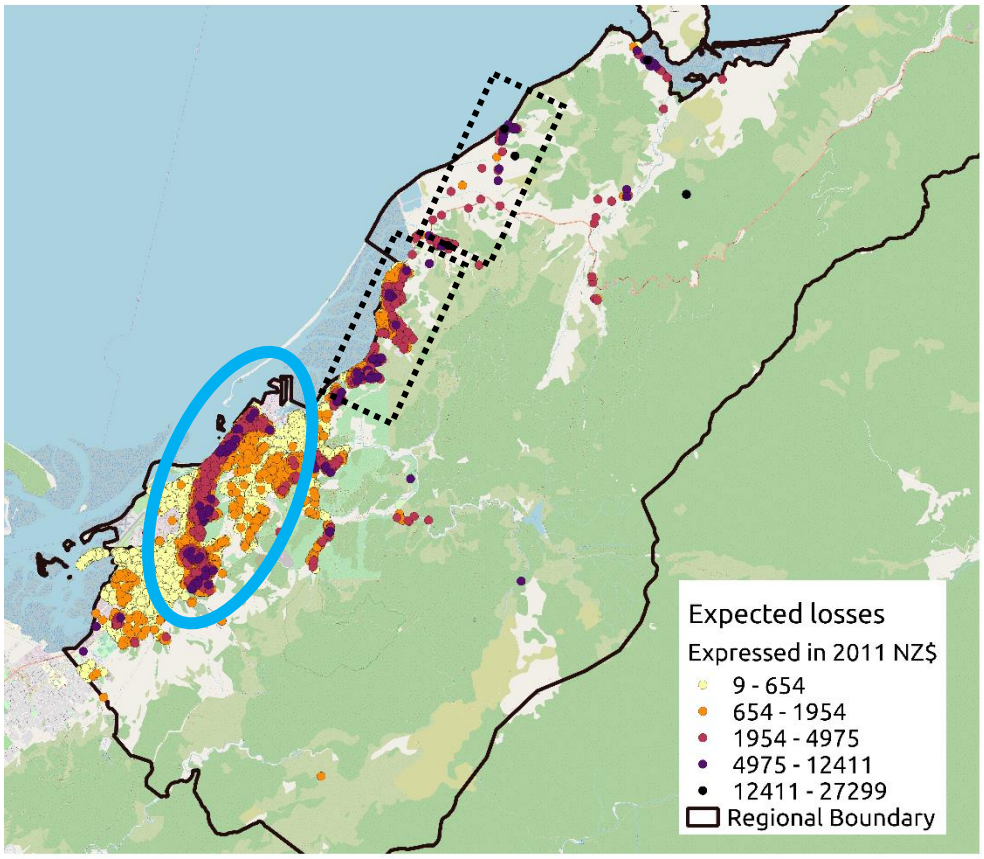


Figure 2. 8 **Spatial distribution of rainfall-induced landslide risk in the Nelson region.** Risk is calculated as the product of the likelihood of damage derived from the logistic regression model, the property's replacement value, and the damage-ratio. The map shows the average value of the expected monetary losses from two damage-ratios. The map shows the clusters (enclosed in blue and black) with the highest levels of risk. The first cluster (enclosed in a blue ellipsoid) overlaps the areas with lower social deprivation levels and high property values (enclosed in a blue ellipsoid). The second set of clusters (enclosed in black rectangles) is located outside denser residential, with lower social deprivation levels, high property values, moderate slopes, or foothills along the coast. The classes we used to group expected losses are

obtained by applying the Jenks Natural Breaks classification algorithm. The algorithm produces groups of data based on their similarity values such that the variance within groups is minimized, while the variance between groups is maximized. We assign to each group a sequential colour scheme so that lighter hues denote lower risk, and the darker hues indicate higher risk.

In terms of the role of fluvial flooding in the levels of risk, we now discuss the results by looking at the spatial extent of flood hazard in Nelson city in Figure 2.9. This area is of special focus as it contains the highest share of residential assets and damage in the entire Nelson region. Figure 2.9 shows that buildings with the lowest risk (coloured in yellow and orange) are located inside areas demarked as flood plains and nearby low-lying areas. This aligns with the regression model's findings and bivariate analysis. We show that properties with damage are less likely to happen in the 100-year flood plain and are located further away from the flood hazard footprint perimeter.

Nevertheless, during the field trip, we identified some claims that could have resulted from the Brook stream or the Matai river overtopping their banks. Since the insurance data do not provide information on a claim's triggering hazard, it wasn't entirely clear whether the damage to properties was because of their location in the waterways' vicinity or whether the damage was from landslides. We do not fully dismiss the possibility of both hazards impacting the properties along the Brook stream. However, Terry (2013) notes that any flooding that might have occurred resulted from dams created by landslips that eventually ruptured and brought with them debris and mud. We argue that the cause of damage during the 2011 Golden Bay storm were landslides and not riverine nor coastal flooding. We thus argue that our estimates capture landslide risk.

In terms of our model's compliance with logistic regression models' assumptions, we claim that the true conditional probabilities are a logistic function of the independent variables. We base this claim on the model's statistical significance of the Chi-2-value and the performance metrics (R2 =0.50, AUC=0.98). We claim that no important variables are omitted. The regression model incorporates predictors of landslide susceptibility (29 in total), which the literature report as fundamental for modelling slope instability processes (Soeters and Van Westen, 1996; Dai et al., 2002).

We claim that the independent variables are not linear combinations of each other as evaluated in the multicollinearity analysis. The non-independence of observations has been addressed by clustering the models' errors. We cannot fully claim, however, that the independent variables are measured without error. This is because properties' location, as defined by the x and y coordinates, has a 70-meter offset from their actual location. The data producer (the EQC) introduces this offset in the location to anonymize the information. We argue that a 70- meter offset is a short span of space

and does not bias our estimations. Still, we note that risk profiles can differ despite the proximity of properties, particularly in heterogeneous landscapes and urban set-ups.

The methodology we develop is useful to forecast the spatial distribution of future risk in case of another triggering event with a similar magnitude. For the EQC, this is particularly important given that the highest exposure to weather-related events come from low-probability high-impact events. In the case of rainfall-induced landslides, the EQC's liability will be amplified by the complexities that entail land and building remediation and the effects of climate change on extreme precipitation increases.

We note that the number of claims relative to the number of properties within the sample have implications in the probability estimates. We note that computing probabilities of rare events using logistic regression analysis can lead to underestimates of the probability. A strategy to deal with the imbalance in the number of properties with and without claims would entail a bootstrapping exercise where: (1) we randomly select a reasonable (e.g., 1,000 properties) subset from the 18,000 properties without a claim, (2) run the logistic regression with this smaller subset plus 352 claims, and (3) repeat the first two steps many times. Then, we can check if the coefficient estimates are stable across different sets of results. Firth (1993) and King and Zeng (1999) propose alternative methods to reduce the bias using what is known as ‘penalized maximum likelihood estimation’ (PMLE). We leave for future research the implementation of alternative methodologies to address the imbalance of properties with and without a claim.

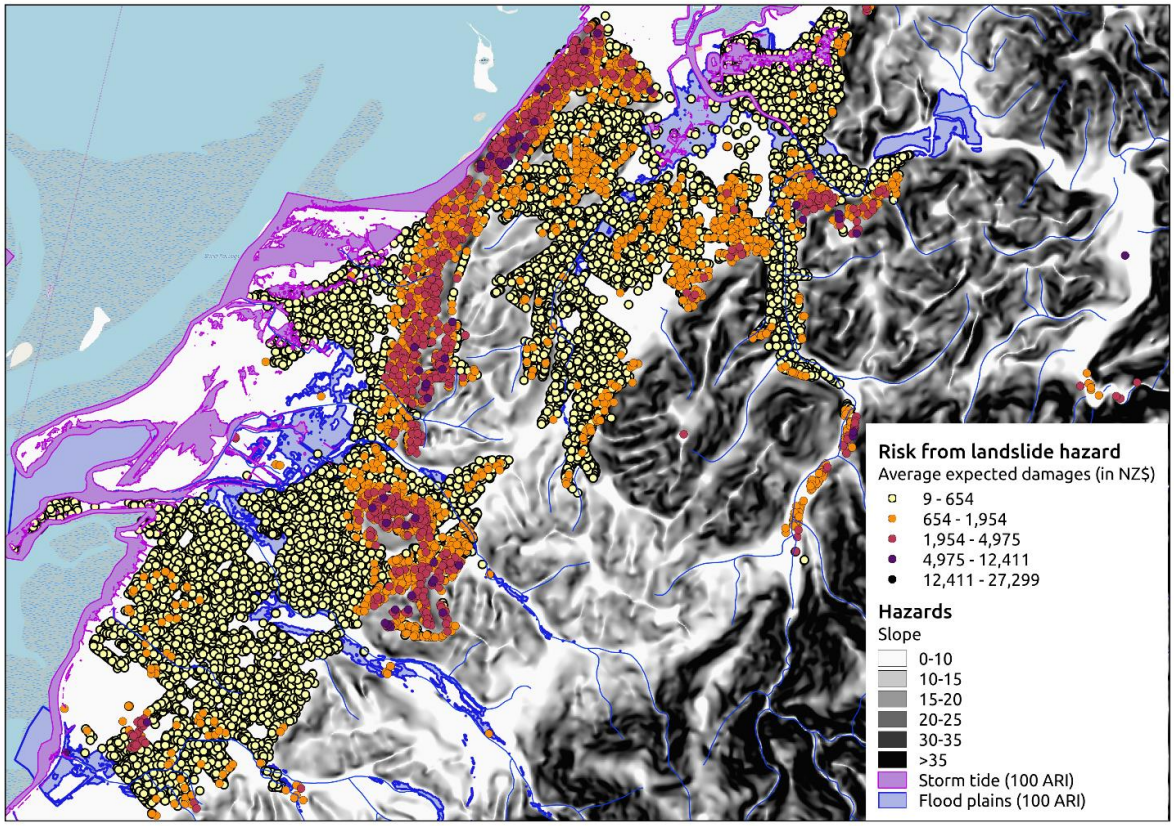


Figure 2. 9 **Spatial distribution of rainfall-induced landslide risk, and the spatial scope of flood inundation from flood hazards (coastal and riverine) in the Nelson city.** Risk is calculated as the product of the likelihood of damage derived from the logistic regression model, the property's replacement value, and the damage-ratio. The map shows the average value of the expected monetary losses from two damage-ratios. The highest risk occurs in moderate to high slope bands, with lower social deprivation and high property values. Nevertheless, high risk also occurs in the relatively flat ground but surrounded by hills. The classes we used to group expected losses are obtained by applying the Jenks Natural Breaks classification algorithm. The algorithm produces groups of data based on their similarity values such that the variance within groups is minimized, while the variance between groups is maximized. We assign to each group a sequential colour scheme so that lighter hues denote lower risk, and the darker hues indicate higher risk.

7 Conclusion

We estimate and map the risk of residential property damage due to extreme rainfall-induced landslides using insurance claim data and identify the drivers of property damage. We find that landslides susceptibility and risk levels are driven by moderate and high slopes, lower social deprivation levels, and high property values. The estimated average expected loses represents 90% of the actual cost of the event for the public insurer and range between an interval of 60 to 120 percent.

The methodology proposed is based on a less resource-intensive and cheaper alternative than conventional landslide susceptibility models that rely on historical landslides' inventories from aerial and satellite imagery. Other highly exposed locations could exploit disaster loss databases (e.g., Desinventar) and employ this methodology to predict future damages from an extreme rainfall event.

We showcase a landslide risk mapping methodology that circumvents the absence of rainfall-induced landslide hazard maps. It leverages off georeferenced historical insurance claims and sets of geospatial data to estimate expected monetary losses. This study's outcome is a cartographic product that reflects the spatial distribution of risk from rainfall-induced landslide hazard associated with a 250-year extreme storm event on property damage. The results produced by the probabilistic modelling are statistically significant and consistent with factual evidence of the nature of the event and on-site assessments.

The proposed model fills a gap of knowledge and provides a statistical-based assessment of rainfall-induced landslide risk as an alternative to the physical-based conventional assessment and as a substitute for landslide risk maps when these are not available. The estimation of expected monetary losses for insurers is fundamental for insuring decisions, determining premiums and excess values, and informing insurance firms' potential financial liability. The public insurer can use these results to introduce disaster risk reductions efforts in areas of high risk. Local Councils could use this type of information to inform their land use planning and zoning of risk-prone areas.

2.8 References

Ashraf, S.; Jones, K.E. 2013. Processing and classifying WV-2 satellite imagery to assess the December 2011 landslide storm damage in the Nelson area. GNS Science Consultancy Report2013/13. 19p+ CD.

Ayalew, L. and Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, *65*(1-2), pp.15-31.

Baeza, C. and Corominas, J., 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, *26*(12), pp.1251-1263.

Baum, R., McKenna, J., Godt, J., Harp, E. and McMulle, S., 2005. Hydrologic monitoring of landslide-prone coastal bluffs near Edmonds and Everett, Washington. *US Geological Survey open-file report*.

Bernet, D.B., Prasuhn, V. and Weingartner, R., 2017. Surface water floods in Switzerland: what insurance claim records tell us about the damage in space and time. *Natural Hazards and Earth System Sciences*, *17*(9), pp.1659-1682.

Buxton, R., Dellow, G. D., Matcham, I. R., Smith, W. D., Rhoades, D. A. 2013. A New Zealand framework for predicting risk due to rainfall-induced landslides, GNS Science Report 2012/22. 12 p.

https://www.gns.cri.nz/static/pubs/2012/SR%202012-022.pdf (accessed 2 February 2019)

Carrara, A., 1983. Multivariate models for landslide hazard evaluation. *Journal of the International Association for Mathematical Geology*, *15*(3), pp.403-426

Choi, J., Oh, H.J., Won, J.S. and Lee, S., 2010. Validation of an artificial neural network model for landslide susceptibility mapping. *Environmental Earth Sciences*, *60*(3), pp.473-483.

[dataset] Core Logic, 2017. Quotable Value. Accessed under a data sharing agreement.

Cruden, D.M., 1993. Cruden, DM, Varnes, DJ, 1996, Landslide Types and Processes, Transportation Research Board, US National Academy of Sciences, Special Report, 247: 36-75.

Dean, S.M., Rosier, S., Carey-Smith, T. and Stott, P.A., 2013. The role of climate change in the two-day extreme rainfall in Golden Bay, New Zealand, December 2011. *Bulletin of the American Meteorological Society*, *94*(9), pp.S61-S63.

Dellow, G., 2011. Tool 2.3. 3: Modelling present-day and future landslide potential. In *Impacts of Climate Change on Urban Infrastructure and the Built Environment-A Toolbox*.

Dai, F.C., Lee, C.F. and Ngai, Y.Y., 2002. Landslide risk assessment and management: an overview. *Engineering geology*, *64*(1), pp.65-87

Dong, J.J., Tung, Y.H., Chen, C.C., Liao, J.J. and Pan, Y.W., 2009. Discriminant analysis of the geomorphic characteristics and stability of landslide dams. *Geomorphology*, *110*(3-4), pp.162-171.

Earthquake Commission Act, 1993. New Zealand. http://www.legislation.govt.nz/act/public/1993/0084/latest/DLM305968.html (accessed 7 July 2017)

Earthquake Commission, 2016. EQCover insurers' Guide. Wellington, New Zealand. https://www.eqc.govt.nz/what-we-do/eqc-insurance/insurers-guide (accessed 7 July 2017)

[dataset] Earthquake Commission, 2018. Insurance claim dataset. Accessed under a data sharing agreement.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27-38.

Fleming, D.A., Noy, I., Pástor-Paz, J. and Owen, S., 2018. *Public insurance and climate change (part one): Past trends in weather-related insurance in New Zealand* (No. 1124-2019-2362). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3477038 (accessed 19 January 2020)

[dataset] Rosser, B., Dellow, S., Haubrock, S. and Glassey, P., 2017. New Zealand's national landslide database. *Landslides*, *14*(6), pp.1949-1959. http://data.gns.cri.nz/landslides/index.html (Accessed 4 Abril 2019)

Frame, D.J., Wehner, M.F., Noy, I. and Rosier, S.M., 2020. The economic costs of Hurricane Harvey attributable to climate change. *Climatic Change*, *160*(2), pp.271-281.

Glade, T., 1998. Establishing the frequency and magnitude of landslide-triggering rainstorm events in New Zealand. *Environmental Geology*, *35*(2-3), pp.160-174.

Guzzetti, F., Carrara, A., Cardinali, M. and Reichenbach, P., 1999. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, *31*(1-4), pp.181-216.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K.T., 2012. Landslide inventory maps: New tools for an old problem. Earth-Science Reviews, 112(1-2), pp.42-66.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E. and Tatham, R.L., 1998. *Multivariate data analysis* (Vol. 5, No. 3, pp. 207-219). Upper Saddle River, NJ: Prentice hall.

Highland, L. and Bobrowsky, P.T., 2008. The landslide handbook: a guide to understanding landslides (p. 129). Reston: US Geological Survey. https://pubs.usgs.gov/circ/1325/pdf/C1325_508.pdf (accessed 10 April 2020)

[dataset] Insurance Council of New Zealand (ICNZ). Cost of Natural Disasters. https://www.icnz.org.nz/natural-disasters/cost-of-natural-disasters/ (accessed 1 March, 2018)

King, G., & Zeng, L. (1999). Logistic regression in rare events data. Department of Government, Harvard University. *Online: http://GKing. Harvard. Edu.* (Accessed 1 June, 2021)

Joyce, K.E., Dellow, G.D. and Glassey, P.J., 2009, July. Using remote sensing and spatial analysis to understand landslide distribution and dynamics in New Zealand. In *2009 IEEE International Geoscience and Remote Sensing Symposium* (Vol. 3, pp. III-224). IEEE.

[dataset] King, A. B., Bell, R., Heron, D., Matcham, I., Schmidt, J., Cousins, W. J., Reese, S., Wilson, T., Johnston, D., Henderson, R., Smart, G., Goff, J., Reid, S., Turner, R., Wright, K., & Smith, W. D. (2009). RiskScape Project: 2004-2008.

[dataset] Land Information New Zealand (LINZ), Topographic maps series 1:50.000, 2009a. https://data.linz.govt.nz/data/category/topographic/nz-topo-50-data/ (Accessed 18 July 2018)

[dataset] Land Information New Zealand (LINZ). Tidal levels table, 2020.https://www.linz.govt.nz/sea/tides/tide-predictions/standard-port-tidal-levels (accessed 15 April 2020)

Lee, S., 2004. Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. *Environmental Management*, *34*(2), pp.223-232

[dataset] National Institute of Water and Atmospheric Research (NIWA), 2005. New Zealand's Historic Events Weather Catalog . https://hwe.niwa.co.nz/event/December_2011_Tasman-Nelson_Heavy_Rain_and_Flooding (Accessed 13 April 2017)

Marjanović, M., Kovačević, M., Bajat, B. and Voženílek, V., 2011. Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, *123*(3), pp.225-234.

Mahalingam, R., Olsen, M.J. and O'Banion, M.S., 2016. Evaluation of landslide susceptibility mapping techniques using lidar-derived conditioning factors (Oregon case study). *Geomatics, Natural Hazards and Risk*, *7*(6), pp.1884-1907.

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J.C., Bouwer, L.M., Bubeck, P., Ciavola, P., Genovese, E., Green, C. and Hallegatte, S., 2013. Assessing the costs of natural hazards–state of the art and knowledge gaps. *Natural Hazards and Earth System Sciences*, *13*(5), pp.1351-1373.

Montgomery, D.R. and Dietrich, W.E., 1994. A physically based model for the topographic control on shallow landsliding. *Water resources research*, *30*(4), pp.1153-1171.

Neuhäuser, B. and Terhorst, B., 2007. Landslide susceptibility assessment using "weights-of-evidence" applied to a study area at the Jurassic escarpment (SW-Germany). *Geomorphology*, *86*(1-2), pp.12-24.

[dataset] Newsome, P.F.J., Wilde, R.H., Willoughby, E.J., 2008. Land Resource Information System Spatial data layers. https://lris.scinfo.org.nz/data/category/environment/ (accessed 09 September 2018) (LANDCARE RESEARCH)

Nelson-Tasman Civil Defense Emergency Management (TN-CDEM) http://www.nelsontasmancivildefence.co.nz/regions/golden-bay/significant-threats/ (accessed 13 April 2020)

Nguyen, C.N. and Noy, I., 2020. Measuring the impact of insurance on urban earthquake recovery using nightlights. *Journal of Economic Geography*, *20*(3), pp.857-877.

Ohlmacher, G.C. and Davis, J.C., 2003. Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering geology*, *69*(3-4), pp.331-343

Owen, S. and Noy, I., 2019. Regressivity in public natural hazard insurance: a quantitative analysis of the New Zealand case. *Economics of Disasters and Climate Change*, *3*(3), pp.235-255.

Pham, B.T., Pradhan, B., Bui, D.T., Prakash, I. and Dholakia, M.B., 2016. A comparative study of different machine learning methods for landslide susceptibility assessment: A case study of Uttarakhand area (India). *Environmental Modelling & Software*, *84*, pp.240-250.

Park, H.J., Lee, J.H. and Woo, I., 2013. Assessment of rainfall-induced shallow landslide susceptibility using a GIS-based probabilistic approach. *Engineering Geology*, *161*, pp.1-15.

[dataset] Paulik, R., Craig, H. and Collins, D., 2019a. New Zealand Fluvial and Pluvial Flood Exposure. *Deep South National Science Challenge Report prepared by NIWA*. https://www.deepsouthchallenge.co.nz/sites/default/files/2019-08/2019118WN_DEPSI18301_Flood%20Exposure_Final%20%281%29.pdf (accessed 25 March 2020)

[Dataset] Paulik, R., Stephens, S.A., Wadhwa, S., Bell, R., Popovich, B. and Robinson, B., 2019b. Coastal flooding exposure under future sea-level rise for New Zealand. *NIWA Client Report 2019119WN, prepared for The Deep South Science Challenge, 2019*. https://www.deepsouthchallenge.co.nz/sites/default/files/2019-08/2019119WN_DEPSI18301_Coast_Flood_Exp_under_Fut_Sealevel_rise_FINAL%20%281%29_0.pdf (accessed 25 March 2020)

Parise, M., 2001. Landslide mapping techniques and their use in the assessment of the landslide hazard. *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, *26*(9), pp.697-703.

Pradhan, B., 2010. Landslide susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches. *Journal of the Indian Society of Remote Sensing*, *38*(2), pp.301-320.

Pradhan, B. and Lee, S., 2010. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. *Environmental Modelling & Software*, *25*(6), pp.747-759.

QGIS 2.18.16, 2020a, Software documentation of geospatial algorithms to interpolate point data into gridded data structure.

[dataset] Rosser, B., Dellow, S., Haubrock, S. and Glassey, P., 2017. New Zealand's national landslide database. *Landslides*, *14*(6), pp.1949-1959.

Salmond, C.E., Crampton, P. and Atkinson, J., 2007. NZDep2006 index of deprivation (Vol. 5541, pp. 1-61). Wellington: Department of Public Health, University of Otago. https://koordinates.com/layer/1066-nz-deprivation-index-2006/ (Accessed 26/05/2020)

Santoso, A.M., Phoon, K.K. and Quek, S.T., 2011. Effects of soil spatial variability on rainfall-induced landslides. *Computers & Structures*, *89*(11-12), pp.893-900.

Soeters, R. and Van Westen, C.J., 1996. Slope instability recognition, analysis and zonation. Landslides: investigation and mitigation, 247, pp.129-177.

StatsNZ. 2020. Terminology: meshblock http://archive.stats.govt.nz/methods/classifications-and-standards/classification-related-stats-standards/meshblock/definition.aspx#gsc.tab=0 (Accessed 20/07/2020)

Stuff, 2011a. Deluge: The toll mounts. Report by Karen Goodger http://www.stuff.co.nz/nelson-mail/news/6151631/Deluge-The-toll-mounts (Accessed 13 April 2020)

Stuff, 2011b . Slips causing major problems in Nelson http://www.stuff.co.nz/ipad-big-picture/6141806/Slips-causing-major-problems-in-Nelson (Accessed 13 April 2020)

Stuff, 2011c. Red stickers for 100 homes http://www.stuff.co.nz/nelson-mail/editors-picks/6161191/Red-stickers-for-100-homes (Accessed 13 April 2020)

Stuff, 2013. NZ's worst rainfall ever. Report by Matt Strewart.http://www.stuff.co.nz/nelson-mail/news/9145759/NZs-worst-ever-rainfall?rm=m (Accessed 13 April 2020)

Stumpf, A. and Kerle, N., 2011. Object-oriented mapping of landslides using Random Forests. *Remote sensing of environment*, *115*(10), pp.2564-2577.

[data set] Tait, A., Henderson, R., Turner, R. and Zheng, X., 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *26*(14), pp.2097-2115.

Terry, F. and Cuff, T. (2012) After the Flood. *New Zealand Geographic*, March/ April p.30-41. https://www.nzgeo.com/stories/after-the-flood/ (Accessed 4 February ,2020)

Terlien, M.T., Van Westen, C.J. and van Asch, T.W., 1995. Deterministic modelling in GIS-based landslide hazard assessment. In *Geographical information systems in assessing natural hazards* (pp. 57-77). Springer, Dordrecht.

United Nations Office for Disaster Risk Reduction (UNISDR) 2016. National Disaster Risk Assessment (NDRA) Guidelines. https://www.preventionweb.net/files/52828_nationaldisasterriskassessmentwiagu.pdf (Accessed 25/07/2020)

Varnes, D.J., 1984. *Landslide hazard zonation: a review of principles and practice* (No. 3).

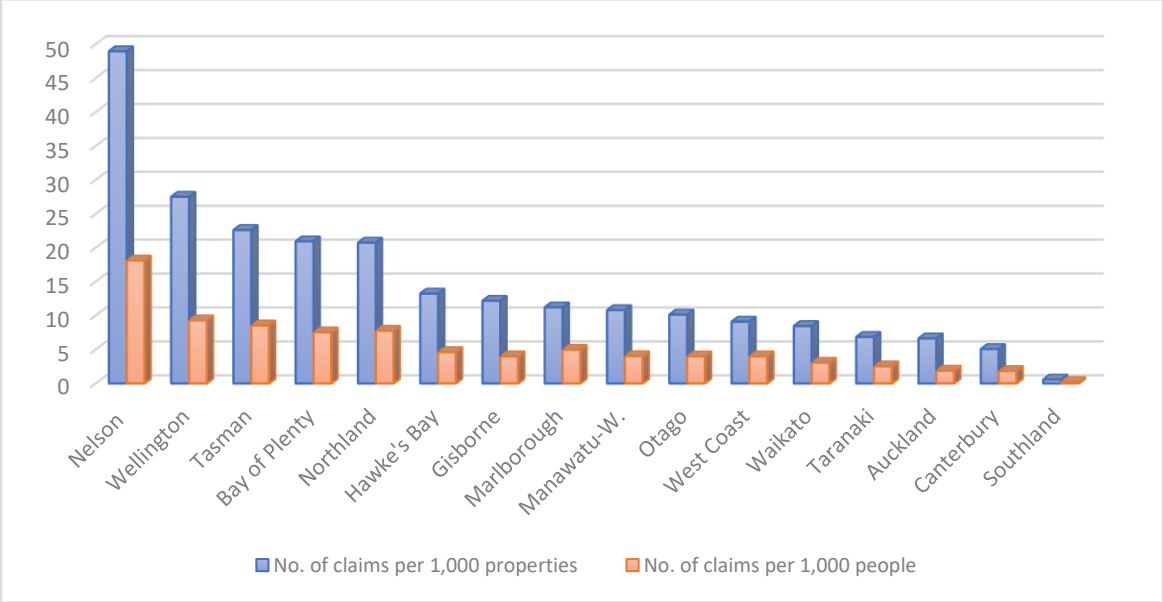
Wieczorek, G.F., 1984. Preparing a detailed landslide-inventory map for hazard evaluation and reduction. Bulletin of the Association of Engineering Geologists, 21(3), pp.337-342.

Yilmaz, C., Topal, T. and Süzen, M.L., 2012. GIS-based landslide susceptibility mapping using bivariate statistical analysis in Devrek (Zonguldak-Turkey). *Environmental earth sciences*, *65*(7), pp.2161-2178.

Yao, X., Tham, L.G. and Dai, F.C., 2008. Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology*, *101*(4), pp.572-582.

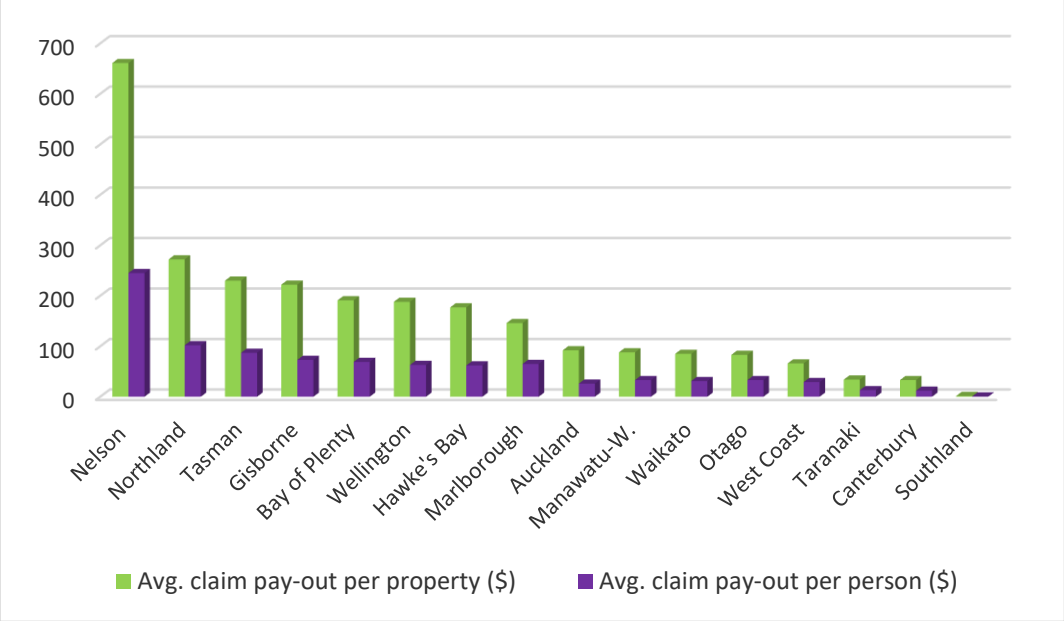
2.9 Appendices

Appendix Figure 2.9.1: Distribution of normalized weather-related EQC' per region between 2000-2017



Distribution of the number of weather-related claims per region. It is measured by the number of claims made to the Public Insurer of New Zealand between 2000-2017 relative to the Housing stock (per 1,000 buildings) (blue bars) and measured by the number of claims relative to the population (per 1,000 people) (orange bars).

Appendix Figure 2.9.2: Distribution of normalized weather-related EQC' per region between 2000-2017



Distribution of weather-related payouts per region. It is measured by the average payouts (in NZ$) made by the Public Insurer of New Zealand between 2000-2017 relative to the Housing stock (per 1,000 buildings) (green bars), and measured by the average claim payouts (in NZ$) relative the population (per 1,000 people) (purple bars)

Appendix Table 2.9.3: Bivariate analysis: Soil data summary statistics

Sig.		Residential properties **WITH** insurance claims					Residential properties **WITHOUT** insurance claims				
		Obs.	Mean	Std. Dev.	Min	Max	Obs.	Mean	Std.Dev.	Min	Max
	Soil data										
	Soil flood class										
*	Properties in flood class (nill)	351	0.42	0.49	0.00	1.00	10,717	0.29	0.46	0.00	1.00
	Properties in flood class slight *(<1 in 60)*	351	0.03	0.18	0.00	1.00	10,717	0.03	0.18	0.00	1.00
*	Properties in flood class – slight *(1-20 to 1-60)*	351	0.04	0.20	0.00	1.00	10,717	0.07	0.26	0.00	1.00
*	Distance to flood class areas *(<1 in 60; 1-20 to 1-60)*	352	1,093.33	817.32	0.00	2,973.13	10,717	930.36	737.10	0.00	7,236.91
	Permeability profile						10,717				
*	Moderate (M)	351	0.13	0.34	0.00	1.00	10,717	0.01	0.12	0.00	1.00
*	Moderate over Rapid (M/R)	351	0.02	0.14	0.00	1.00	10,717	0.11	0.31	0.00	1.00
*	Moderate over Slow (M/S)	351	0.34	0.48	0.00	1.00	10,717	0.26	0.44	0.00	1.00
*	Rapid (R)	351	0.00	0.00	0.00	0.00	10,717	0.01	0.11	0.00	1.00
	Drainage						10,717				
*	D – Poor	351	0.02	0.14	0.00	1.00	10,717	0.07	0.25	0.00	1.00
	D – Imperfect	351	0.03	0.18	0.00	1.00	10,717	0.03	0.18	0.00	1.00
*	D – Moderately	351	0.05	0.21	0.00	1.00	10,717	0.01	0.09	0.00	1.00
	Profile Water availability						10,717				
	PAW – Very high	351	0.00	0.00	0.00	0.00	10,717	0.00	0.01	0.00	1.00
*	PAW – High	351	0.07	0.26	0.00	1.00	10,717	0.11	0.31	0.00	1.00
-	PAW – Moderately high	351	0.26	0.44	0.00	1.00	10,717	0.23	0.42	0.00	1.00
*	PAW – Moderate	351	0.17	0.37	0.00	1.00	10,717	0.05	0.22	0.00	1.00
*	PAW – Low	351	0.00	0.00	0.00	0.00	10,717	0.01	0.11	0.00	1.00
	Water deficit						10,717				
*	Water deficit	347	77.13	23.18	22.00	162.00	10,717	82.93	21.92	0.00	167.00

Bivariate analysis. The table compares residential properties in terms of their soil characteristics for two subgroups properties: properties without damage and properties with damage resulting from a 250-year extreme precipitation event: the 2011 Golden Bay Storm. The soil data statistics are only representative of the rural areas. The remaining areas are impervious surfaces such as roads, parking lots, driveways, etc. Significant variables are marked with an asterisk [*]. The statistical significance test (at 0.05) is calculated with different methods depending on the variable's measure (interval, nominal).

Appendix Table 2.9.4: Full model: Stepwise regression model (sig* level = 0.05)

Independent variable Binary: Claim = 1, No claim = 0	(1)	(2)
Dependent variables		
Slope band (0-10 degrees) omitted category		
Slope band 1 (10-15 degrees)	1.63**	1.63**
	(0.33)	(0.37)
Slope band 4 (25-30 degrees)	2.42**	2.42**
	(1.00)	(1.04)
Slope band 6 (35 -40 degrees)	3.35**	3.35**
	(1.59)	(1.58)
Slope band 7 (>= 40 degrees)	3.83***	3.83***
	(1.92)	(1.88)
Rain (mm)	0.98***	0.98***
	(0.01)	(0.01)
Building value (in tens of thousands NZ$)	1.13***	1.13***
	(0.01)	(0.01)
Properties in outside denser residential areas	4.73***	4.72***
	(1.81)	(1.87)
Deprivation Index 9 omitted category		
Deprivation Index 3	5.33***	5.33***
	(2.01)	(1.95)
Deprivation Index 4	7.49***	7.49***
	(3.33)	(3.04)
Deprivation Index 6	2.53**	2.53**
	(0.97)	(1.11)
Deprivation Index 7	4.26**	4.26**
	(2.87)	(2.62)
Deprivation Index 8	3.43***	3.43***
	(1.29)	(1.24)
Distance to exotic forest (in hundreds of meters)	0.93***	0.93***
	(0.03)	(0.03)
Number of floors	1.76**	1.76**
	(0.43)	(0.44)
Constant	0.33	0.33
	(0.61)	(0.56)
Observations	18,277	18,275
Chi2-value	555.8	596.3
Prob_chi2	0	0
Number of clusters	363	2144
Pseudo R-squared	0.489	0.489

Stepwise regression results with probability 0.05. Only statistically significant variables (at 0.05) from the bivariate analysis are included in the model. The results in columns (1) and (2) differ in the make-up of the areas that the standard errors are clustered for. The first geographic unit (known as meshblock), which are small geographical areas, and in general, households within a meshblock will be relatively similar in terms of their sociodemographic characteristics. The second geographic unit delimits areas that are topographically homogenous in terms of the steepness of the terrain. The results are very similar for both clustering methods (i.e., the difference in the magnitude of the standard errors in column (1) and (2) is minimal). The coefficients are expressed as odds ratios, and the clustered standard errors are shown in parentheses for the significance levels *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 2.9.5: Full model: Stepwise regression model (sig* level = 0.10)

Independent Binary: Claim = 1, No claim = 0	(1)	(2)
Dependent variables		
Slope band (0-10 degrees)	.	.
omitted category	(.)	(.)
Slope band 1 (10-15 degrees)	1.96***	1.96***
	(0.41)	(0.47)
Slope band 2 (15-20 degrees)	2.06**	2.06**
	(0.63)	(0.59)
Slope band 4 (25-30 degrees)	2.91**	2.91**
	(1.22)	(1.23)
Slope band 6 (35 -40 degrees)	4.90***	4.90***
	(2.43)	(2.47)
Slope band 7 (>= 40 degrees)	4.41***	4.41***
	(2.26)	(2.25)
Rain (mm)	0.98***	0.98***
	(0.01)	(0.01)
Distance to flood plain (100 ARI) (in hundreds of meters)	1.08***	1.08***
	(0.03)	(0.03)
Distance to small rivers (in hundreds of meters)	0.91*	0.91*
	(0.05)	(0.05)
Distance to exotic forest (in hundreds of meters)	0.94*	0.94**
	(0.03)	(0.03)
Building value (in tens of thousands NZ$)	1.13***	1.13***
	(0.01)	(0.01)
Elevation (m) - above mean sea level	0.99*	0.99**
	(0.00)	(0.00)
Properties in outside denser residential areas	4.18***	4.18***
	(1.65)	(1.70)
Deprivation Index 9		
omitted category		
Deprivation Index 3	5.01***	5.01***
	(2.05)	(1.99)
Deprivation Index 4	9.66***	9.66***
	(4.36)	(4.07)
Deprivation Index 6	3.00***	3.00**
	(1.16)	(1.32)
Deprivation Index 7	6.33***	6.32***
	(4.29)	(3.85)
Deprivation Index 8	4.18***	4.18***
	(1.71)	(1.67)
Number of floors	1.76**	1.76**
	(0.43)	(0.45)
Constant	0.26	0.26
	(0.50)	(0.45)
Observations	18,277	18,275
Chi2-value	602.7	650.1
Prob_chi2	0	0
Number of clusters	363	2144
Pseudo R-squared	0.497	0.497

Stepwise regression results with probability 0.10 Only statistically significant variables (at 0.05) from the bivariate analysis are included in the model. The results in columns (1) and (2) differ in the make-up of the areas that the standard errors are clustered for. The first geographic unit (known as meshblock), which are small geographical areas, and in

general, households within a meshblock will be relatively similar in terms of their sociodemographic characteristics. The second geographic unit delimits areas that are topographically homogenous in terms of the steepness of the terrain. The results are very similar for both clustering methods (i.e. the difference in the magnitude of the standard errors in column (1) and (2) is minimal). The coefficients are expressed as odds ratios, and the clustered standard errors are shown in parentheses for the significance levels *** p<0.01, ** p<0.05, * p<0.1

Appendix Table 2.9.6: Expected losses versus actual damages

	(1) Expected losses	(2) Percentage (%) expected vs actual losses
Expected losses When damage ratio 1 = 0.05676	6,115,409	60.02%
Expected losses When damage ratio 2 = 0.11387	12,267,451	120.66%
Average expected losses (row 1 and row 2)	9,191,431	90.41%
Actual pay-outs Made by the public insurer -the EQC	9,658,327	-

Expected losses versus actual damages from rainfall-induced landslides caused by a 250-year storm event (Golden Bay 2011). The expected Losses (i.e. risk) are estimated by factoring the probability of damage from the logistic model, replacement values and damage ratios. Depending on the damage ratio, the expected losses range between an interval of 60 to 121 per cent of the actual cost. The average expected losses using two damage ratios represent 90% of the public insurer's actual cost event. The likelihood term used in the calculation is derived from the 'full model' .

Appendix Table 2.9.7: Expected losses with and without climate change

	(1) Losses **with** the effect of climate change (in 2011 nominal NZ$)	(2) Losses **without** the effect of climate change (in 2011 nominal NZ$)	(3) Percentage of the actual losses*
Expected Losses When damage ratio 1 = 0.056	6,115,409	5,809,639	60.02%
Expected Losses When damage ratio 2 = 0.114	12,267,451	11,654,079	120.66%
Average expected Losses (row 1 and row 2)	9,191,431	8,731,859	90.41%
Actual pay-outs Made by the public insurer -the EQC	9,658,327	9,175,411	-

Expected losses versus actual damages from rainfall-induced landslides caused by a 250-year storm event, with and without the effect of climate change. Dean et al. (2013) report that 5% of the 2011 Golden Bay storm event's intensity can be attributed to anthropogenic-driven climatic change. Thus, 5% of the total value of the estimated expected losses could be attributable to climate change. The likelihood term of the expected losses is derived from the 'full model'. However, we note that the 5% indicates the change in rainfall intensity and not the change in risk.

Chapter Three

Property risk from extreme precipitation, floods, and climate change

Abstract

Climate change is increasing the risk of floods, but few available flood maps incorporate this change in flood hazard probability, constraining the assessment of future risk. In this study, I use insurance claim data to quantify the risk of residential property flooding caused by extreme precipitation and climate change. Risk is defined as the likelihood of flood damage and the potential consequences (damage). Here the likelihood of property damage explicitly accounts for the intensity, duration, frequency, and spatial extent of extreme precipitation, with and without the effect of climate change, while controlling for other predictors of property damage. The potential consequences of flood hazard are measured by the properties' physical vulnerability (per cent damage) to varying flood depths. I first implement a multivariate logistic regression model to estimate the likelihood of damage with and without the effect of climate change. Then, I calculate the expected monetary losses by factoring in the likelihoods derived from the regression models, property replacement values, and a property's physical vulnerability to varying flood depth scenarios. I focus on a 2005 flood event in the region around New Zealand's Bay of Plenty, which received the highest number of insurance payouts from New Zealand's public insurer due to a weather-related event. I find that the highest monetary losses are associated with low-return periods, as expected. Nevertheless, high return periods (i.e., 2-year events) bring about sizeable losses. The likelihood of property damage resulting from climate-related changes in precipitation increases significantly, but the effect (climate change signal) is too small to cause an economically meaningful increase in risk levels. Public and private insurers can use this methodology to assess its current and future financial risk exposure to low-probability, high-impact weather-related events.

3.1 Introduction

Greenhouse gas emissions emitted by human activity are warming the atmosphere, and this is expected to produce changes in the frequency and intensity of extreme weather events (IPPC, 2012). In New Zealand, climate change is projected to increase extreme precipitation events across the country, defined as events with a recurrence interval of two years or greater (Mullan et al., 2018). Flood risk will increase as a result (Woods et al., 2010).

However, flood hazard maps are subject to multiple uncertainties, whether they include the impact of climate change or not. The uncertainties stem from the complexity of the hazard, the shortcomings of the methods, and inadequate data to model flood hazard (Meyer et al., 2013). Other sources of uncertainty in hazard data result from natural phenomena being highly variable and rare, and the changing probability of hazard extremes over time due to climate change (Bouwer, 2013; Kundzewicz et al., 2017).

Such uncertainties can make it difficult to estimate direct impacts accurately. For instance, several research studies report on the mismatch between predicted flooding areas and the actual extent of flooding events. Bernet et al. (2017) report that 60% of flood claims were outside a flood inundation zone in a canton in Switzerland. Bihan et al. (2017) report that only 6% and 9% of claims occurred inside the floodplains of two flooding events in France. Hunn et al. (2018) reported on Houston's flooding from Hurricane Harvey in 2017, where almost 75% of flooded properties were outside the 100-year flood zone. In New Zealand's 2005 Bay of Plenty event, almost 92% of claims fell outside flood-prone areas. Only 0.08% of properties within flood-prone areas experienced flood damage caused by extreme precipitation. The mismatch between the predicted and observed impacts of flood events suggests that traditionally constructed flood hazard maps alone are insufficient to understand flood risk and project the future direct economic impacts of weather-related disasters.

This paper proposes a new method based on an alternative to traditional flood hazard mapping that requires fewer intensive resources and information. I use insurance claim data and other geospatial datasets to quantify the current risk from extreme precipitation on residential property and future risk from changes in extreme precipitation caused by climate change. I focus on a region around New Zealand's Bay of Plenty. This region received the highest number of insurance payouts for a weather-related event from New Zealand's public insurer, the Earthquake Commission (EQC).

I quantify the risk of flood-related property damage, where risk is defined as the product of the probability (likelihood) of flood damage and the potential consequences (damage). The

probability of property damage explicitly accounts for the intensity, duration, frequency, and spatial extent of extreme precipitation, with and without the effect of climate change, while controlling for other predictors of property damage. The potential consequences of flood hazard are measured by the properties' physical vulnerability (per cent damage) to varying flood depths.

Unlike previous studies, I simultaneously address the selection bias and the unexplained variation that result from insurance data aggregation. I use microlevel georeferenced claim data to solve the unexplained variation resulting from aggregated insurance data (Gradeci et al., 2019). And I address the selection bias problem by including both affected and non-affected buildings in the estimation sample.

The proposed method addresses three questions: What is the risk to residential property from extreme precipitation? What will be the effect of climate change on residential property flood risk through changes in extreme precipitation? What is the spatial and temporal distribution of future flood risk due to climate change? To answer these questions, I do the following.

First, I estimate a logistic regression model using past insurance claim data from EQC and several geospatial datasets. This model estimates the empirical relationship between insurance claims and property characteristics. My two aims are to identify the relative effect of factors that might make a property more or less likely to experience flood-related damage, and to predict the probability of property damage. The datasets I use include information on physical characteristics (precipitation, topography, soil, inundation hazard) and socioeconomic characteristics (property value, and a multidimensional indicator of social deprivation) of the full set of residential properties within the footprint of the Bay of Plenty event. I identify the probability of damage by exploiting variation in precipitation across space while considering the influence of physical and socioeconomic factors.

In the logistic regression model, daily observed precipitation data come from the Virtual Climate Station Network (VCSN) (Tait et al., 2006). To approximate the return periods (frequency), I compare observed precipitation (VCSN) data with estimates of extreme precipitation from the High Intensity Rainfall Design (HIRDSv4) dataset (Carey-Smith et al., 2018). These estimates are referred as to depth-duration-frequency (DDF) values.[12] The premise is that the HIRDSv4 mapping from duration and return period to rainfall can be used to approximate the return period of observed (VCSN) rainfall at a given location. Thus, I compare VCSN rainfall values with HIRDSv4 DDF values and approximate the return period based on the smallest difference between the two datasets

[12] 'Depth' refers to rainfall amount (in mm) and 'frequency' refers to the return period. The return period gives the estimated time interval between events of a similar size or intensity. For example, the HIRDSv4 rainfall figure of given location for daily rainfall with a return period of 20 years is 284 mm of rain expected to fall within a 48-hour duration with a probability of 1/20 in any given year. This figure is referred as to depth-duration-frequency (DDF) value.

in each grid. The average absolute value between VCSN and the nearest HIRDSv4 rainfall data is 9.5%. Based on the proximity between the two datasets, the return period of the observed precipitation can be defined. It follows that the predicted probabilities of flood-related property damage show the relative effect of observed precipitation intensity-duration-frequency.

Second, I predict the impact of future extreme precipitation events on the distribution of the probability of property damage. To predict this impact, I use the empirical relationship estimated in the logistic regression model and climate projections based on future scenarios of greenhouse gas concentrations. Pastor-Paz et al. (2020) (Chapter one) and Cheng et al. (2013) also follow this approach to calculating the effect of climate change for extreme precipitation; and Pinto et al. (2007), Leckebusch et al. (2007) and Klawa and Ulbrich (2003) use this approach for severe windstorms. In this approach, the regression coefficient estimates obtained from the historical relationship between extreme precipitation and flood-related damage are applied to extreme precipitation data that incorporates response to climate change.

To calculate the response of precipitation to climate change, I multiply precipitation data by percentage change factors (augmentation factors) to project rainfall intensity derived from the current climate to a future climate that is 1 degree warmer. Then, I convert to fixed percentage changes by multiplying the augmentation factors by the relevant temperature change for each Representative Concentration Pathway (RCP) (Carey-Smith et al., 2018). In total, I calculate the effect of climate change on precipitation considering four future emission scenarios (RCPs 2.6, 4.5, 6.0, 8.5) and time horizons (2031-2050, 2056-2075, 2081-2100). In total, the RCP-time period scenarios render 12 distributions of precipitation that incorporate climate change response.

Third, I compare the predicted probabilities of damage in the first-stage baseline model with the predicted probabilities of damage from each RCP-time scenario from the second stage. The comparison shows that any differences between the probability distributions of damage can be attributed to climate change.

Fourth, I estimate the risk: expected losses expressed in monetary terms. Specifically, I calculate the risk by multiplying the probability of damage (derived from the algorithm described in the first two steps), property replacement values, and the physical vulnerability of a property. I use fragility functions (also known as damage curves) to calculate physical property vulnerability (the potential consequences of flood hazard). These functions relate the percentage damage (damage-ratio) to variations in flood hazard intensity, measured by flood depth. I use the fragility functions to identify the damage-ratios for a series of plausible flood depth scenarios for 11 residential construction types. Here, ‘plausible’ refers to the flood depths observed during the 2005 Bay of Plenty

event. The damage functions were derived from empirical field observations in New Zealand, using surveys after flooding events, and expert opinion (Ramsey and Reese, 2010). While most damages were caused by pluvial flooding, the insurance claim dataset does not separate out the cause of damage. Therefore, I also quantify the expected losses for a small subset of records for an area affected by rainfall-induced landslides. In the latter case, I estimate the expected losses using damage-ratios for four building construction types. These damage-ratios were derived by Buxton et al. (2013), using historical information on damages caused by rainfall-induced landslides in New Zealand.

Finally, I present the results in the form of loss exceedance probability curves (EPLC). These curves relate the expected losses as a function of annual exceedance probabilities (AEP), where the AEPs are calculated as the inverse of the return period. The EP curve is the basis upon which insurers estimate their likelihood of experiencing various levels of loss.

I find that the highest risk from floods is associated with rainfall with low-probability precipitation in any given year, as expected. Changes in extreme precipitation resulting from climate change will increase the probability of flood damage across all future emissions scenarios and periods. Despite finding statistically significant changes in the probability distribution of property damage due to climate change-induced extreme precipitation, the signal size is small, resulting in a negligible increase in expected monetary loss from flood-related property damage.

I argue that the increases in projected rainfall aren't large enough to reflect a substantial change in the probability of damage and increased risk levels. The small changes in probability are mainly due to the lack of information on climate change augmentation factors for the 250-year return period and the observed precipitation data's temporal detail.

The risk estimates are spatially explicit, which implies that the spatial distribution of property risk from flood hazard can be generated as a map that identifies 'hotspots' of risk from 'coldspots' by displaying expected monetary losses at property-point level. The detailed representation of risk can help to tailor disaster risk reduction efforts and risk-based pricing strategies.

3.2 Literature review

The existing body of literature that uses insurance data to explore the relationship between flood-related property damage and extreme precipitation is fairly new (Sampson et al., 2017; Spekkers et al., 2014; Grahn and Nyberg, 2017; Spekkers et al., 2015; Torgersen et al., 2015; Torgersen et al., 2017). These papers do not quantify the risk or the change in risk due to climate change, and only identify the factors associated with property damage.

Pastor-Paz et al. (2020) and Cheng et al. (2012) evaluate the role of climate change on rainfall and their effect on property damages using insurance data. Cheng et al. (2012) estimate a simple relationship between a rainfall index and insurance data. They then use this relationship with future rainfall simulations to project changes in the number of claims and losses. However, Cheng et al. (2012) use only information on properties that were damaged, where the claim data have been temporally and spatially aggregated in a cross-sectional way. In contrast, Pastor-Paz et al. (2020) (Chapter one) exploit the time and space dimension using a longitudinal georeferenced micro-level dataset of insurance claim data. Their data's longitudinal structure permits them to isolate contemporaneous variation of extreme precipitation, while controlling for exposure and vulnerability through the use of location (grid-cell) fixed-effects.[13]

Unlike previous studies, I simultaneously address the selection bias and the unexplained variation that result from insurance data aggregation. I use microlevel georeferenced claim data to solve the unexplained variation resulting from aggregated insurance data (Gradeci et al., 2019). Data aggregation is a typical practice that anonymises the exact location of damaged properties, which the insurance industry rarely discloses (Grahn and Nyberg, 2017). The problem with aggregation is that the areas are typically defined for administrative purposes or to deliver services (e.g., postal service). Therefore, the areas are not necessarily statistically homogenous units in terms of socioeconomic status, topography, and building construction types (Spekkers et al., 2014). Data aggregation eliminates the heterogeneity between properties within areas and limits the models' ability to capture the influence and the magnitude of the factors that make a property more or less prone to flood damage.

I address the selection bias problem by including both affected and non-affected buildings in the estimation sample. The selection bias in other studies manifests in two forms. The first form relates to the absence of undamaged properties. In this case, the relationship between damage and explanatory variables is assessed using only data from affected properties. The second form of selection bias relates to insurance penetration. In New Zealand, public insurance cover is almost universal, with penetration of residential properties estimated at 98% (Nguyen and Noy, 2020). In contrast, other studies use data representing a fraction of the market share (e.g. 35% in Grahn and Nyberg, 2017) or a fraction of the building stock (e.g. 6% in Spekkers et al., 2015). The exception is Norway, with 90% insurance penetration (Torgersen et al., 2015; Torgensen et al., 2017).

[13] Chapter 2 also evaluates the role of climate change on precipitation. The attribution of the impact of climate change on losses is quantified based on the 'Fraction of attributable risk -FAR. According to Dean et al. (2013), 5% of the intensity of the 2011 Golden Bay storm event can be attributed to the influence of anthropogenic climate change.

3.3 Data

I use past weather-related georeferenced claim data from New Zealand's public insurer (EQC) and a set of geospatial datasets produced by Land Information New Zealand (LINZ), Landcare Research (LCR) and the National Institute of Water and Atmospheric Research (NIWA). These datasets include information on the physical characteristics such as precipitation, topography, soil, and flood hazard maps. These datasets also include socioeconomic characteristics (property value and a multidimensional indicator of well-being) of a full set of buildings distributed across New Zealand, including the Bay of Plenty. Most of the data are publicly available. However, EQC insurance claim data, the residential building inventory dataset, and flood hazard maps have been provided under data-sharing agreements with EQC, CoreLogic, and NIWA, respectively.

3.3.1 Insurance sample construction

The EQC dataset contains full details on all individual weather-related claims submitted. I use the variables on the claim's geolocation, the compensation values per cover (building, land, and contents, in $NZ), the date of the claimed event, and the claim's status.[14]

I normalise the damage by inflation-adjusting the compensation figures paid out over 2000-2017 (in 2017 $NZ). I aggregate the payouts at the event level because I need to match them to the duration dimension in the HIRDSv4 data. Therefore, I need to identify the duration of each extreme event separately. Figure 3.1 shows spatial distribution of weather-related insurance claims for different events.

The normalised time series of damages per event shows numerous low-impact events and few high-impact events (Figure 3.2), implying a skewed distribution of damages. Because of this, I create a ranking of events (percentiles) and plot them against cumulative losses (Figure 3.3). The often-cited 80/20% relationship (Jagger et al., 2008), whereby the top 20% strongest weather events account for 80% of the damages, is also true for New Zealand's public insurer. Specifically, the costliest 20% of weather-related events represent around 85% of the damages. In this study, I focus on the most expensive event for EQC, the 2005 Bay of Plenty (BOP 2005) event, which represents approximately 8% of all the payouts made by EQC for weather-related claims.

The BOP 2005 event triggered losses of NZ$35.9 million for public insurer EQC, and NZ$36.5 million for private insurers (EQC claim dataset, 2017; Insurance Council of New Zealand dataset, 2005). The event we study was catalogued as 'phenomenal and unprecedented rainfall' that affected

[14] I keep the claims that have been settled, i.e., those claims with 'claim status' defined as 'completed payments'. Thus, I drop from the sample claims that have were declined or not accepted A total 182 claims were declined for reasons that could include: not insured, no coverage, damage unrelated to the event.

the Bay of Plenty, especially Tauranga and Matata where states of civil emergency were declared (Tauranga City Libraries, 2005; NIWA Weather Catalog, 2005)

> *...a moist, subtropical northeasterly flow prevailed over the northeast of the North Island and embedded in this were a number of 'convergence lines' which moved slowly east across the western Bay of Plenty during the day. Bands of cumulonimbus cloud formed along these convergence lines producing torrential downpours, notably in Tauranga and Matata (MetService, 2005).*

Most damage in Tauranga was caused by pluvial flooding (Ministry of Environment, 2005). The stormwater system could not cope with the waters as they were built to cope with a precipitation intensity corresponding to a 5-year return period. In contrast, damage in Matata was predominantly caused by debris brought down from landslides triggered in the hills (McSaveney et al., 2005).

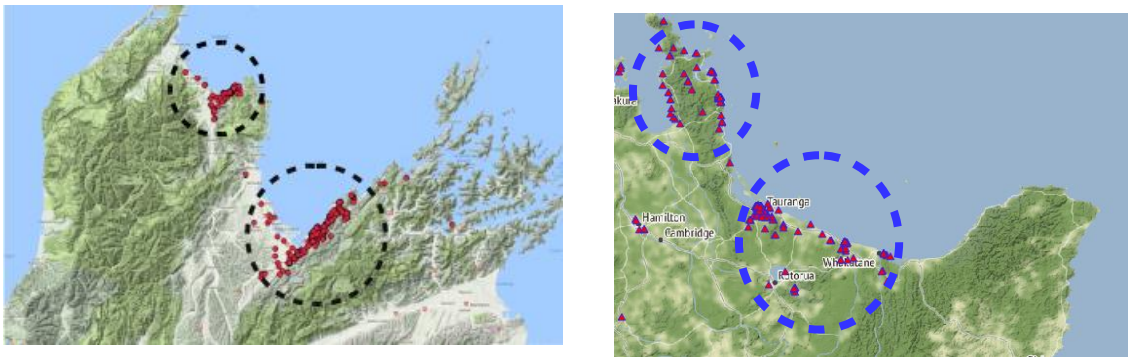


Figure 3. 1 **Spatial distribution of weather-related insurance claims for different events.** Each event has an associated spatial footprint, duration and meteorological characteristics. The figure on the left shows the Golden Bay Storm that affected the Tasman-Nelson regions in 2011. The figure on the right shows the Bay of Plenty event that affected Tauranga and Matata predominantly, and marginally the Coromandel Peninsula in 2005. Figure elaborated by the author using approximate locations of EQC claim-level data.

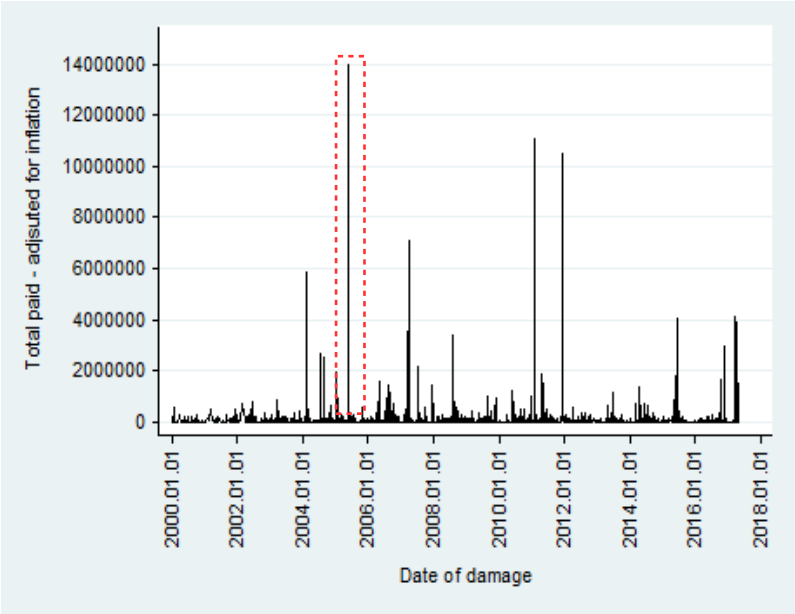


Figure 3. 2 **Time series of normalised total property damages (i.e., adjusted by inflation values in 2017 NZ$) per date of the weather-related event (storm, flood, landslide).** The 'boxed' event is the 2005 Bay of Plenty event.

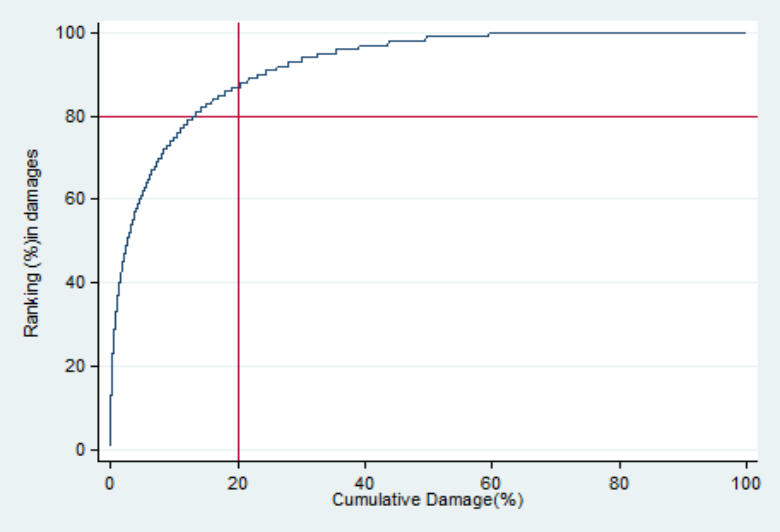


Figure 3. 3 **Cumulative percentage of total damage as a function of percentage ranking.** The reference lines indicate the often-cited 80/20% relationship, where the top 20% strongest weather events account for 80% of the damage. In New Zealand, the costliest 20% of weather-related events represent around 85% of the damage.

3.3.2 Precipitation

Precipitation data come from two datasets: the High Intensity Rainfall Design System (HIRDSv4) (Carey-Smith et al., 2018) and the Virtual Climate Station Network (VCSN) (Tait et al., 2006). VCSN data are provided as a regular gridded surface of 5x5 km^2, where each grid reports on the amount of observed (actual) daily rainfall (in mm) at a given date. In contrast, HIRDSv4 estimates high-intensity rainfall across New Zealand for a range of return periods and event durations. That is, for each duration and return period, the HIRDSv4 data give the millimetres of rainfall each grid location might experience. HIRDSv4 data are provided as a regular gridded surface of 2x2 km^2, where each grid reports multiple combinations of DDF rainfall values. Depth refers to the amount of rainfall in millimetres (mm); duration refers to daily (1-5 days), sub-daily (1-12 hours) or sub-hourly durations (10-30 minutes); and frequency refers to the return period (2-250 years). For example, the HIRDSv4 rainfall estimate for two-day rainfall with a return period of 20 years in a given grid is 284 mm of rain expected to fall within a 48-hour duration, with a probability of 1/20 in any given year.

I link insurance claim data corresponding to the 2005 BOP event to both HIRDS and VCSN gridded surfaces. Claims with an underlying HIRDS grid are assigned the precipitation value of the VCSN they fall in, or the closest precipitation value from a neighbouring VCSN grid for the date of the event.[15] In other words, I pair/match observed precipitation from the VCSN data with estimated

[15] About 30 claims do not have an underlying HIRDSv4 grid as the properties are located around the coastline and fall outside the spatial extent of the gridded surface. The damaged and undamaged properties without an underlying HIRDSv4 grid are not considered in the sample.

depth-duration-frequency rainfall values from the HIRDSv4 dataset. I found the 48-hour HIRDS surface with the rainfall depth closest to the observed VCSN amount and from this determined the return period of the rainfall amount.

I define the 2005 BOP event as a two-day duration event based on reports from New Zealand's Historic Weather Events Catalog (2005) and the MetService Newsletter (2005). I define the precipitation return period for each grid based on the smallest difference between observed and estimated precipitation. Figure 3.4 shows the comparison between observed and estimated precipitation across all the grids with at least one property with damage. Figure 3.4 shows a close match for central precipitation values (in the 150-300 mm range), slight discrepancy at the lowest values, and more pronounced differences at the highest precipitation values. Figure 3.5 shows the same comparison as in Figure 3.4, but the differences in precipitation are shown as a function of each grid's return period. The discrepancies for low and high return periods are due to observed rainfall amounts either being less than the HIRDS 2-year surface or higher than the 250-year surface. The average absolute value between VCSN and the nearest HIRDSv4 rainfall data is 9.5%.

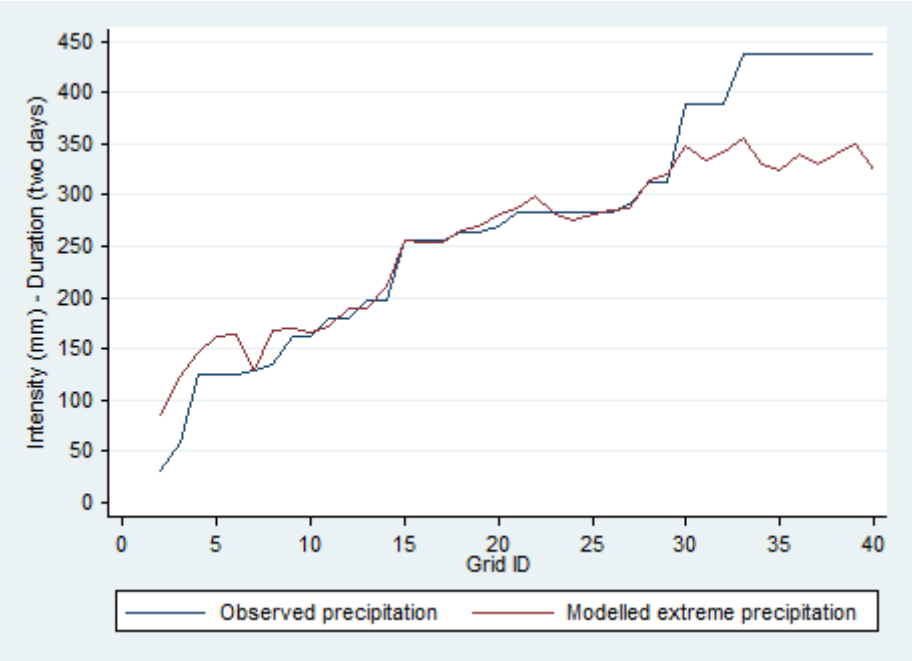


Figure 3. 4 **Comparison between observed and DDF precipitation values across all grids for a two-day duration event.**

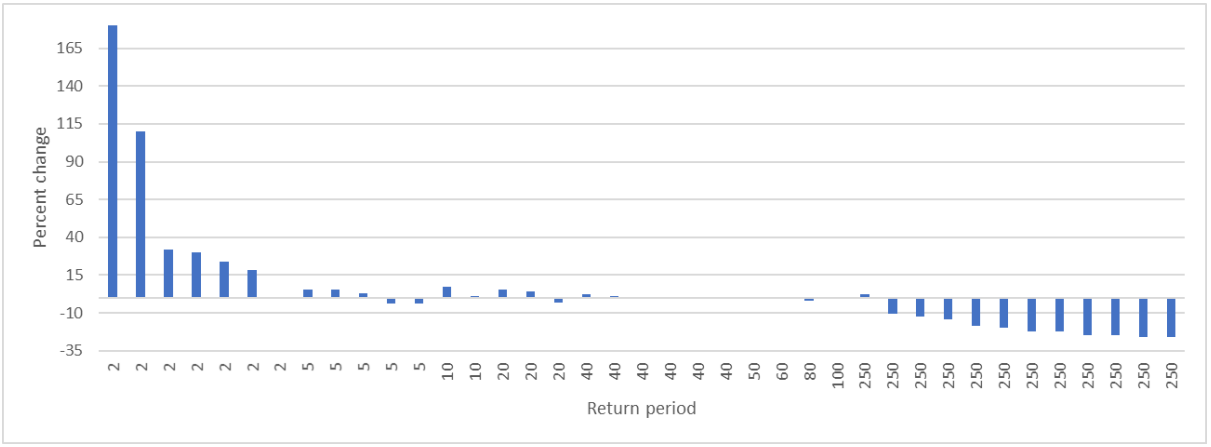


Figure 3. 5 **The percentage difference between observed and DDF precipitation values as a function of the return period (in years) for a two-day duration event**. The discrepancies for low and high return periods are due to observed rainfall amounts either being less than the HIRDSv4 2-year surface or higher than the 250-year surface.

The spatial footprint of the 2005 BOP event spreads over 40 HIRDSv4 grids (160 km^2) and 16 VCSN grids (400 km^2). I drop from the analysis any HIRDSv4 grid whose closest VCSN grid is further than 10 km, and thus the properties within. I drop from the sample any HIRDSv4 grid with fewer than two claims. This ensures variation across records in the logistic regression model's dependent variable: a binary variable that indicates whether a property has made an insurance claim. Thus, the event's final sample spreads over 16 HIRDSv4 grids (64 km^2) and 6 VCSN grids (150 km^2). The sample contains 22,027 properties, including 245 claims. Figure 3.6 shows the total number of grids per return period for the final sample.

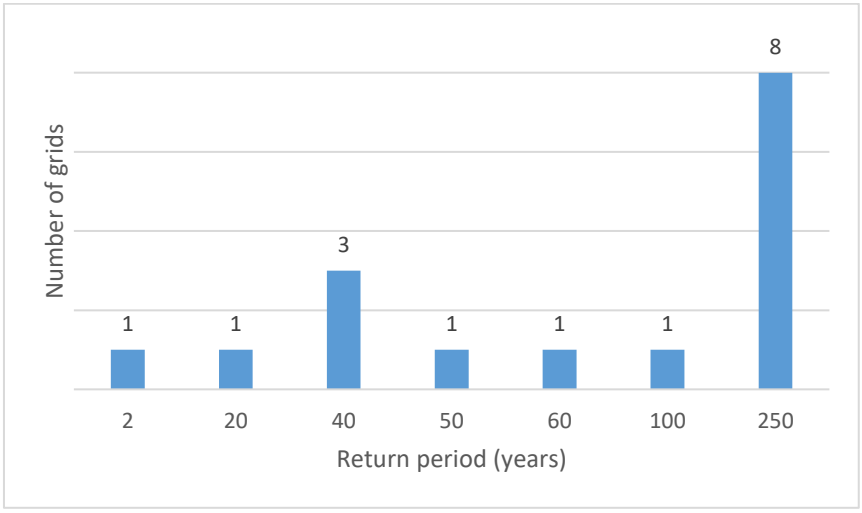


Figure 3. 6 **Number of grids per return period (in years) for the final sample.** The sample contains 22,203 properties, 245 claims that spread over 16 HIRDSv4 grids (64 km^2) and 6 VCSN grids (150 km^2).

3.3.3 Flooding hazard: Fluvial and coastal flooding

I identify the number of properties with and without damage located in areas subject to fluvial and storm tide flooding. As well as examining whether properties fall within the flood zones, I measure the distance between properties and the flood maps' perimeters (when they are located

outside the zone) as predictors of the probability of damage. I relate information using proximity because observations suggest that flood damage can occur in the vicinity of the flood hazard footprint and because of the known inaccuracies of flood maps reported in the literature.

The fluvial flood hazard map reflects the inundation extents of a 100-year return period (as of 2013). This map results from two-dimensional modelling (2D) of the catchment and incorporates the effect of land cover and storm water systems on flood hazard. As well as modelling for depth, the flood map also models floodwater speed and direction. The flood model was built using Lidar-derived topographic data. Details of the characteristics of the flood map can be found in Paulik et al. (2019a).

The coastal storm tide inundation map represents the spatial extent of a 100-year return period (as of 2018). Neither of these maps incorporates the effect of climate change. NIWA provided the flood maps under data-sharing agreements. Details of storm surge maps' characteristics can be found in Paulik et al. (2019b).

3.3.4 Topography

No rainfall-induced landslide hazard maps in the area of study are available. I calculate each property's slope and elevation (relative to mean sea level) using a high-resolution, high-accuracy Lidar-based digital elevation model (DEM). The topographic data are sourced from LINZ.

3.3.5 Soil characteristics

I identify the soil characteristics of properties as predictors of the probability of damage, specifically: soil flood return period, drainage, readily available water, permeability, and land use capability. I assess how these characteristics might be associated with damage or amplify it under extreme precipitation. Thus, I identify whether a property is located: on soils with flood return periods ranging from 'slight' to 'very severe'; on soils with 'very poor', 'poor' and 'imperfect' drainage; on soils with 'very high', 'high' and 'moderately high' profiles of readily available water; on soils with a 'slow' rate of water movement through saturated soil; on soils with poor land-use capability. Furthermore, I identify whether a property is located on fluvial soils and impervious surfaces (i.e., in urban areas). The soil data are sourced from Landcare Research (LCR) (Newsome et al., 2008).

3.3.6 Social vulnerability

To account for social vulnerability, I use the New Zealand Deprivation Index (NZDI), a multidimensional indicator of well-being constructed using data from the 2001 New Zealand Census. Salmond et al. (2007) describe the index as a multidimensional metric that aggregates individual variables to reflect eight different social deprivation dimensions. The index combines dimensions

related to access to services, employment status, number of residents and space available, single-parent families' economic dependence, and qualifications. The index is provided as an ordinal scale with values that range from 1 ('least deprivation') to 10 ('most deprivation') (Salmond et al., 2007). The index is calculated for statistical areas of varying size called meshblocks, which contain multiple properties. I assign the meshblock-specific deprivation index to all the buildings in the same meshblock. There are 493 meshblocks in the area of study.

To proxy for the socioeconomic status, I also compute the total value (in NZ$) of property: land, building, contents and appurtenant structure values. These monetary values are inflation-adjusted to 2017 NZ$ values.[16]

3.3.7 Summary statistics: bivariate analysis

Properties with damage and without damage are compared in a bivariate fashion to determine statistically significant differences (at 0.05 level) for each physical and socioeconomic conditioning predictor of damage. Depending on whether a variable is continuous or categorical, a 'ranksum' or a 'Chi-2' test is applied.

The bivariate results in Table 3.1 show, on average, a statistically significant difference between the underlying distributions of extreme precipitation for properties with and without damage.

Damage is unrelated to the location of property inside flood plains and storm surge inundation areas. However, damage is related to the proximity to the shoreline, and the flood plain and the storm tide's inundation perimeters. Damage is positively related to slope, with more damage where the average slope is higher than properties without damage. Damage is negatively related to altitude, i.e. damage occurs at lower altitudes.

Surprisingly, damage is unrelated to location on soils prone to flood, soils with poor drainage, soils with a slow water movement through saturated soil, soils with poor use capability, and impervious surfaces. No damage at all occurred in fluvial soils. Damage is positively associated with locations on soils with high water availability.

The bivariate analysis of social vulnerability and damage shows properties in the middle levels of social deprivation are more likely to experience damage. Finally, the bivariate analysis shows that damage is positively associated with higher property values. On average, the values of property with damage are 12% higher than properties without damage.

[16] Land and contents values have been modelled by EQC.

Sig.	Variables	Properties with damage					Properties **without** damage				
		Obs.	Mean	S. Deviation	Min.	Max	Obs.	Mean	S. Deviation	Min.	Max
	Rain										
*	Observed 48-hour rainfall (VCSN)	245	390.4	74.8	124.1	438.4	21,782	373.2	85.4	124.1	438.4
*	Estimated 48-hour rainfall (HIRDSv4)	245	321.4	32.7	163.9	356.3	21,782	310.2	38.9	163.9	356.3
	Flood hazard										
-	Properties located in storm surge areas (1%AEP)	245	0.0	0.1	0.0	1.0	21,82	0.0	0.1	0.0	1.0
-	Properties in flood-prone areas	245	0.07	0.3	0.0	1.0	21,782	0.1	0.3	0.0	1.0
*	Distance to the perimeter of flood-prone area	245	5,270.5	15172.4	3.2	59660.7	21,782	1,247.0	7,115.3	0.1	59,876.7
*	Distance to the perimeter of storm tide	245	226.9	168.4	0.5	704.6	21,782	284.9	249.2	0.2	2,022.6
	Hydrography										
*	Distance to shoreline	245	449.6	348.4	27.2	1670.9	21,782	629.7	494.1	2.0	2,462.2
	Topography										
*	Slope	245	3.9	3.3	0.1	17.3	21,782	2.7	3.5	0.0	36.5
*	Elevation	245	21.4	10.3	1.5	58.9	21,782	24.3	14.3	0.0	156.5
	Soil										
-	Properties located on soils with Slight (<1 in 60) to Very Severe Return Periods (>1 in 5)"	62	0.1	0.3	0.0	1.0	5,461	0.2	0.4	0.0	1.0
-	Properties located on 'very poor', 'poor' and 'imperfect' soil drainage	245	0.0	0.1	0.0	1.0	21,644	0.0	0.2	0.0	1.0
-	Properties located on soils with a 'slow' rate of water movement through saturated soil	245	0.0	0.1	0.0	1.0	21,644	0.0	0.1	0.0	1.0
*	Properties located on soils with 'moderate-high' to 'very-high' profile total available water for the soil profile to a depth of 0.9 m	62	0.7	0.5	0.0	1.0	5,461	1.0	0.2	0.0	1.0
-	Properties located on land with poor land use capability	245	0.3	0.4	0.0	1.0	21,644	0.3	0.4	0.0	1.0
-	Properties located on impervious surfaces (in urban areas)	245	0.9	0.3	0.0	1.0	21,782	0.9	0.3	0.0	1.0
	Socioeconomic										
*	Total property value	245	659,000	313,000	234,000	1,790,000	21,782	586,000	298,000	5,000	8,150,000

Sig.	Variables	Properties with damage					Properties without damage				
		Obs.	Mean	S. Deviation	Min.	Max	Obs.	Mean	S. Deviation	Min.	Max
-	Deprivation Index 1	245	0.024	0.155	0	1	21,728	0.037	0.189	0	1
-	Deprivation Index 2	245	0.061	0.24	0	1	21,728	0.05	0.218	0	1
*	Deprivation Index 3	245	0.061	0.24	0	1	21,728	0.103	0.304	0	1
-	Deprivation Index 4	245	0.159	0.367	0	1	21,728	0.12	0.325	0	1
*	Deprivation Index 5	245	0.184	0.388	0	1	21,728	0.138	0.345	0	1
-	Deprivation Index 6	245	0.127	0.333	0	1	21,728	0.114	0.317	0	1
*	Deprivation Index 7	245	0.216	0.413	0	1	21,728	0.147	0.354	0	1
*	Deprivation Index 8	245	0.037	0.188	0	1	21,728	0.113	0.316	0	1
-	Deprivation Index 9	245	0.127	0.333	0	1	21,728	0.123	0.328	0	1
*	Deprivation Index 10	245	0.004	0.064	0	1	21,728	0.055	0.227	0	1

Table 3. 1 **Bivariate analysis.** The table compares residential properties in terms of their physical and socioeconomic characteristics for two subgroups properties: properties without damage and properties with damage as result a of the Bay of Plenty event. ***** Marks statistically significant variables. The statistical significance is calculated at 0.05. Depending on the type of variable i.e. continuous or categorical, we use a ranksum and Chi-2 test.

3.4 Methods

The risk of a natural hazard is calculated as the product of the probability (likelihood) of the hazard and the potential outcomes or consequence (damage). The probability is associated with the frequency (recurrence interval), intensity and spatial extent of the hazard. The consequence refers to the degree of damage. In this study, I quantify the risk from floods on property damage, and the effect of climate change on flood risk, through changes in extreme precipitation. Here the probability explicitly accounts for the frequency, intensity, duration and spatial extent of extreme precipitation, and the potential property damage is simulated for various flood depths.

I first use a statistical approach to find the functional relationship between physical and socioeconomic predictors of property damage and the distribution of past flood-related damage to predict properties likely to experience damage from extreme precipitation. I implement a logistic regression model where identification comes from the spatial variation of precipitation values across grids while controlling for exposure and vulnerability measures.

Second, I apply the historical relationship between extreme precipitation and flood-related damage of the first step (the regression coefficient estimates) to extreme precipitation data that incorporates the change in extreme precipitation because of climate change.

Third, I calculate the climate change signal, i.e., changes in the likelihood of damage resulting from climate change under different emission scenarios and periods.

Fourth, I calculate physical property vulnerability using fragility functions.

Finally, I integrate the damage probability derived from the regression model, property replacement values, and property vulnerability to quantify the risk from flood hazard under a range of flood-depth scenarios. I present the results in the form of annual loss exceedance probability curves.

While most residential property damage during the 2005 BOP event was caused by pluvial flooding, I cannot identify the triggering cause of damage from the insurance claim dataset. Therefore, I also quantify the expected losses for a subset of records of an area known to have been predominantly affected by rainfall-induced landslide debris (Matata).

3.4.1 Likelihood of experiencing damage: the benchmark model

I estimate a logistic regression model of residential property damage to predict the likelihood of experiencing damage. I use the following model:

$$LC_i = \frac{e^{\beta_1 Haz_i + \beta_2 Exp_i + Vul_i + \epsilon_i}}{1 + e^{\beta_1 Haz_i + \beta_2 Exp_i + Vul_i + \epsilon_i}} \qquad (1)$$

where LC_i is a binary variable that indicates whether an insurance claim has been made by property *i*. The terms Haz_i Exp_i *and* Vul_i are vectors that account for the physical and socioeconomic characteristics of property *i*. Table 3.3 reports the variables included in the terms of the equation. All the model data are standardised by subtracting the mean value and dividing by the standard deviation for each value of each record.

In the logistic regression model, I use the observed rainfall values from the VCSN dataset. These precipitation values are the benchmark against which we will evaluate climate change's effect on the changes in the probability of damage. We cluster the standard errors by location based on the HIRDSv4 grid.

3.4.2 Likelihood of experiencing damage: climate change

We use three inputs to incorporate the response of extreme precipitation from climate change: rainfall values (VCSN), percentage change factors per degree of warming, and temperature increases for four emission scenarios (RCPs) and different periods. The second input is a percentage change factor (augmentation factor) that maps the relationship of an increase in temperature to an increase in rainfall for various return periods per degree of warming. The third input is the increase in temperature for each RCP. The percentage change factors per degree of warming are converted to fixed percentage changes for different future periods and climate change scenarios by multiplying the augmentation factors by the relevant temperature change. Appendix Figure 3.8.1 provides information on the second and third inputs, taken from Carey-Smith et al. (2018). In total, I calculate the effect of climate change on precipitation considering four future emission scenarios (RCPs 2.6, 4.5, 6.0, 8.5) and time horizons (2031-2050, 2056-2075, 2081-2100). In total, the RCP-time period scenarios render 12 distributions of precipitation that incorporate climate change response.

3.4.3 Climate change signal

I calculate the climate change signal, i.e., changes in the likelihood of damage resulting from climate change. To do so, I calculate the predicted likelihoods of all the logistic regression models with and without climate change. Then, I compare the predicted probabilities of the benchmark model (without climate change) with the predicted probabilities of each RCP-time combination. Any

statistically significant difference in the probability distributions of damage can be thus attributed to climate change. To test whether differences are statistically significant, I use a non-parametric version of a paired sample t-test known as 'Wilcoxon signed rank sum test'. This test assumes that the differences between the fitted probabilities of the model with and without climate change are not normally distributed but assumes that the difference is ordinal.

3.4.4 Property vulnerability (Fragility functions)

3.4.4.1 Floods

The damage a flood causes to a building depends on the flood intensity (water depth, water velocity, duration of inundation, debris, etc.) and the building characteristics (floor materials, wall materials, number of storeys, etc.). Flood building damage can be expressed as the cost to repair, or as the ratio between the cost of repair and the building's replacement value (percentage of damage). The relationship between damage-ratio and flood intensity is described by 'fragility functions': fragility curves or damage curves (Meyer et al., 2013). These functions are a method to calculate the potential direct damage, typically derived from empirical field observations (from surveys after flooding events), expert opinions, or experiments in laboratory settings. The functions associate different flood depths with different degrees of damage on a scale of 0 to 1, where 1 means the damage is equal to the building's replacement value.

EQC's cover scheme provides for damage to the land around and under residential buildings and appurtenant structures that result from floods or storms. Therefore EQC's dataset of claims does not include information on building damage. However, regardless of EQC's policy coverage, residential buildings that experienced land damage were also likely to be damaged as well. On these grounds, I estimate the expected monetary losses resulting from building damage, since either the private insurers or homeowners will have to bear them.

For these calculations, I use 11 flood-depth fragility functions with data provided by NIWA, reported in Reese and Ramsay (2010) (see Figure 3.7). Each fragility function considers a combination of the construction type (timber, concrete, masonry), the number of stories (1 or 2) and construction age (pre-1960, between 1960-1980, and post-1980). The distinction made in the construction year is based on the most commonly used materials in those periods. For instance, buildings constructed pre-1960s and post-1980s have mainly a slab concrete floor, whereas buildings

constructed between 1960-1980 have a chipboard floor type. This distinction matters because chipboard is much more susceptible to water damage (Reese and Ramsay, 2010).[17]

The Historic Weather Events Catalog, published by NIWA, is an important source that reports observed depth of the flooding events in the Bay of Plenty, and across New Zealand. The catalogue collects and compiles data from government institutions, newspapers, databases, councils, and various sources. Based on the reports, we use three flood depths as lower, middle and upper bound values: 0.25 metres, 0.5 metres and 1 metre. I compute the damage-ratios by fitting into the fragility functions the flood-depth scenarios for each combination of construction type, age and number of floors. I assume that all properties experience the same flood depth level. Differentiating flood depths for each property is beyond the scope of this paper.

Figure 3. 7 **Flood fragility functions**. Each fragility function considers a combination of the construction type (timber, concrete, masonry), the number of floors (1 or 2) and construction age (pre-1960, between 1960-1980, and post-1980). The inundation depth (above the floor level in metres) is shown along the horizontal axis, and the damage ratio is shown in the vertical axis. Based on these fragility functions, I simulate three flood depth scenarios as lower, middle and upper bounds of the observed flood depths during the 2005 Bay of Plenty event: 0.25 m, 0.5 m and 1 m. The figure is taken from Reese and Ramsay (pp. 7, 2010).

[17] Applying flood-depth fragility functions is an internationally accepted standard approach for evaluating physical damage. However, damage is likely explained not only by flood depth, but by water velocity and flood duration (Thieken et al, 2008). In New Zealand, given the data availability, building vulnerability can be assessed only in terms of flood depth.

3.4.4.2 Landslides

To quantify landslide damage, I source four different damage-ratios for four building types using the findings from Buxton et al. (2013). The building types are concrete, masonry, and two types for residential timber buildings with one and two storeys. Buxton et al. (2013) estimate the damage-ratios by simulating historical damage-ratio information from rainfall-induced landslides for various building types for all of New Zealand.

3.4.5 Quantification of risk

I calculate the expected losses (risk) by factoring the likelihoods produced from the regression model, property replacement values, and vulnerability of the buildings, such that:

$$ExpectedDamage_i = \widehat{LC_i} * RV_i * DR_i \qquad (2)$$

where *i* is a residential property, $\widehat{LC_i}$ is the estimated likelihood of damage, RV_i is the building replacement value, and DR_i is the calculated damage ratios.

I present the results in the form of annual exceedance probability loss curves. These curves relate the expected monetary losses and annual exceedance probabilities (AEP). AEPs are calculated as the inverse of the return period (AEP = 1/ return period). To calculate and graph the annual exceedance probability loss curves, I aggregate the expected monetary losses per grid and per AEP value.

3.5 Results and Discussion

3.5.1 Regression model

The relationship between flood-related property damage, extreme precipitation and the set of conditioning factors of damage is not due to chance (as shown by the significance of the Chi-2 test). Table 3.2 shows the results from the logistic regression model.

The factors significantly associated with property damage are property value, slope, distance to flood plain perimeters, distance to the shoreline, and observed precipitation at the time of the event. A standard deviation unit increase in precipitation intensity (in mm) is associated with an increase in the odds of property damage of 1.863. This result confirms the expected relationship between extreme precipitation and property damage, where higher precipitation values are related to the likelihood of flood damage.

We find that a standard deviation unit increase in slope (in degrees) is associated with an increase of 1.458 in the odds of property damage. The role of slope in the likelihood of property

damage has also been observed at the national level (Fleming et al., 2018; Pastor Paz et al., 2020) and at the regional level (as reported in Chapter 2). In the 2005 BOP event, rainfall-induced landslides were reported in Otumoetai and Matata (Roseer et al., 2017). Debris flows triggered structural damage to buildings in Matata (McSaveney et al. 2005). In Otumoetai, it is impossible to differentiate flood from landslide damage. One could argue that all damage in Matata was caused by landslides, based on accounts of the event. Because this area is quite localised and disconnected from the majority of properties impacted by the 2005 BOP event, I excluded all Matata properties from the sample (with and without damage). I excluded the Matata records so that the estimation sample would include and thus predict flood-related damage. However, without the Matata records the regression model cannot be estimated because one grid is lost, and the number of clusters is equal to the number of variables. Therefore, I estimate the regression model (Table 3.3) with all the records.

Claim	Coef.	St.Err.	t-value	p-value	[95% Conf	Interval]	Sig
Precipitation	1.863	0.349	3.32	0.001	1.290	2.690	***
Slope	1.458	0.178	3.08	0.002	1.147	1.853	***
Distance to flood plain perimeter (fluvial)	1.560	0.339	2.05	0.041	1.019	2.389	**
Distance to storm surge perimeter	0.848	0.144	-0.97	0.331	0.607	1.183	
Distance to shoreline	0.646	0.099	-2.86	0.004	0.479	0.872	***
Property value	1.084	0.037	2.33	0.020	1.013	1.159	**
Deprivation Index 1 & 2 (base category)							
Deprivation Index 3	0.624	0.133	-2.21	0.027	0.411	0.948	**
Deprivation Index 4	1.708	0.756	1.21	0.226	0.718	4.065	
Deprivation Index 5	2.129	0.457	3.52	0.000	1.398	3.241	***
Deprivation Index 6	0.795	0.613	-0.30	0.767	0.176	3.605	
Deprivation Index 7	2.382	0.581	3.56	0.000	1.477	3.843	***
Deprivation Index 8	0.727	0.377	-0.61	0.540	0.263	2.011	
Deprivation Index 9	2.264	1.026	1.80	0.071	0.932	5.504	*
Deprivation Index 10	0.283	0.294	-1.22	0.224	0.037	2.163	
Constant	0.006	0.001	-19.40	0.000	0.003	0.009	***
Pseudo r-squared	0.078		Number of obs		21973		
Chi-square	4188.590		Prob > chi2		0.000		
*** p<0.01, ** p<0.05, * p<0.1							

Table 3. 2 **Regression results benchmark model.** This table contains results from a logistic regression of the benchmark model, where observed precipitation does not incorporate the effect of climate change. The left-hand side variables is an indicator variable that shows whether a property received insurance payouts from weather-related damage. Only statistically significant variables (at 0.05) from the bivariate analysis are included in the model. I combined the two lowest categories of social deprivation to be able to estimate the model. The model cannot be estimated if the number of variables and number of clusters in the model are the same. The standard errors are clustered using the HIRDSv4 grids. All variables have been mean-standardised. The coefficients are expressed as odds ratios.

The regression coefficient estimate shows that a standard deviation unit increase in the distance to the shoreline decreases the odds of property damage by 0.646. The role of proximity to the shoreline and property damage has also been observed at the national level (Fleming et al., 2018; Pastor Paz et al., 2020). In these two studies, however, distance to the shoreline is positively related

to property damage: the closer the more likely properties will experience damage. For damage and flood mapping, the model's estimates show that the odds of property damage increase by 1.560 per standard deviation unit increase in the distance between properties and pluvial flood plain perimeter. This result seems to reinforce the lack of predictive power observed in the flood maps.

The impact of social deprivation on the odds of property damage shows that households with deprivation index (DI) of five, seven and nine have 2.129, 2.382 and 2.264 higher odds of property damage than the least socially deprived households, respectively.[18] In contrast, the odds of property damage decrease by 0.624 for households with DI of three.

Overall, the model performs well. The R^2 value is 0.078, the true positive rate (TPR) is 70.61%, the true negative (FPR) is 61.87% and precision 2.04%. TPR shows the percentage of properties that had damage and were correctly classified by the model as damaged. TNR shows the percentage of properties that had *no* damage and were correctly classified by the model as undamaged. Precision is the proportion of positive results that were correctly classified. Precision is a more useful metric than the TNR when the phenomenon under study is rare. Precision does not include the number of true negatives in its calculation and is not affected by the imbalance of events. In the context of this study, the imbalance is reflected in that many more properties had no damage (21,782) than those with damage (245). In Chapter 2, I observe a similar figure for the Nelson region where 1.92% (352 claims out 18,000) of properties experienced damage due to extreme precipitation during the 2011 Golden Bay Storm. These findings highlight the fact that the likelihood of property damage is small even for high-impact, low-probability weather-related events.

To assess the model's performance, we also produce a Receiver Operating Characteristics (ROC) graph and its associated Area Under the Curve (AUC) value, based on a series of confusion matrices. A model with no predictive power has an AUC value of 0.5, whereas a perfect model AUC value is 1. The AUC value of the model is 0.71.

We note that the number of claims relative to the number of properties within the sample have implications in the probability estimates. We note that computing probabilities of rare events using logistic regression analysis can lead to underestimates of the probability. A strategy to deal with the imbalance in the number of properties with and without claims would entail a bootstrapping exercise where: (1) we randomly select a reasonable (e.g., 1,000 properties) subset from the 21,973 properties

[18] I combined the two lowest categories of social deprivation to be able to estimate the model. The model cannot be estimated if the number of variables and number of clusters in the model are equal. I also eliminated elevation as predictor and opted to preserve the social deprivation variables as they represent multiple sociodemographic dimensions of well-being.

without a claim, (2) run the logistic regression with this smaller subset plus 245 claims, and (3) repeat the first two steps many times. Then, we can check if the coefficient estimates are stable across different sets of results. Firth (1993) and King and Zeng (1999) propose alternative methods to reduce the bias using what is known as 'penalized maximum likelihood estimation' (PMLE). We leave for future research the implementation of alternative methodologies to address the imbalance of properties with and without a claim.

3.5.2 Regression models: climate change signal

I find statistically significant differences between the benchmark model's predicted probabilities and the predicted probabilities from each RCP-time model combination. Consequently, climate change increases the probability of property damage from floods caused by extreme precipitation. The mean, minimum and maximum percentage change in probability of damage increase as the RCP increases from 2.6 up to 8.5 for the periods 2031-2050, 2056-2075 and 2081-2100 (see Table 3.3).

Variable	Observations	Mean	Std.Dev.	Min	Max
Period 2031 - 2050					
RCP 2.6	21973	0.028	0.044	0.005	0.292
RCP 4.5	21973	0.035	0.055	0.007	0.362
RCP 6.0	21973	0.032	0.051	0.006	0.334
RCP 8.5	21973	0.040	0.063	0.007	0.413
Period 2056 - 2075					
RCP 2.6	21973	0.032	0.050	0.006	0.330
RCP 4.5	21973	0.048	0.076	0.009	0.502
RCP 6.0	21973	0.053	0.083	0.010	0.550
RCP 8.5	21973	0.073	0.115	0.014	0.755
Period 2081 - 2100					
RCP 2.6	21973	0.028	0.044	0.005	0.292
RCP 4.5	21973	0.055	0.087	0.010	0.572
RCP 6.0	21973	0.072	0.113	0.013	0.747
RCP 8.5	21973	0.107	0.168	0.020	1.107

Table 3. 3 **Summary statistics of the climate change signal:** shows the percentage difference between the predicted probabilities between the benchmark model 'without climate change' and the models that incorporate the response of precipitation to climate change. The mean, minimum and maximum percentage change in probability of damage increase as the RCP increases for all RCPs and time horizons. The differences in probability distributions are statistically significant based on the 'Wilcoxon signed rank sum test'.

The size of the climate change signal (the change in the probability of damage) is too small to cause economically meaningful changes in the expected losses from flood-related damage, although I can quantify statistically significant changes in the probability of property damage as a result of changes in precipitation caused by climatic change. I address two reasons for this.

First, information is lacking on climate change augmentation factors for the 250-year return period. Carey-Smith et al. (2018) provide augmentation factors up to 100-year return periods, so I assigned the 100-year augmentation factor to the 250-year return period. But in the estimation sample, 50% of the grids had a 250-year return period.[19]

Second, flood-related damage in New Zealand is mostly explained by sub-daily and sub-hourly precipitation durations. However, as observed data (VCSN) is available on daily time steps, I had to assume that damage results from daily precipitation. Carey-Smith et al. (2018) state that long-duration events are expected to increase much less than short-duration events. For instance, for 48-hour duration events the augmentation factors range between 6.7 to 7.5% change per degree of warming across all return periods. In contrast, for short-duration events (e.g. 1 hour), the increases range from 12.2 to 13.6% (Figure 3.8). Although observed sub-daily weather station data can be matched/paired with DDF rainfall value estimates, only two weather stations are available within the event's footprint. By only using two weather stations, the rainfall variability across the data records would be limited, making it difficult/impossible/not feasible to estimate the regression models. Figure 3.9 shows the observed precipitation in absolute values, as well as precipitation augmented by climate change per degree of warming, per RCP and per time horizon across grids.

Figure 3. 8 **Augmentation factors for extreme precipitation based on 1 degree of warming plotted as a function of the return period.** Long-duration events are expected to increase much less than short-duration events as a result of climate change. The solid lines show the median over New Zealand for different event durations, and the 5th and 95th percentiles are displayed as shaded bars for the 1- and 120-hour event durations. Taken from Carey-Smith et al. (2018). Appendix Figure 3.8.1 provides a table representation of the graph.

[19] I fitted a linear and exponential function to the augmentation factors to come up with a factor for the 250-year return period. The best fit is from an exponential function. The function is concave and flattens for higher return periods. Thus, the 250-year factor we infer is not substantially different from the 100-year augmentation factor.

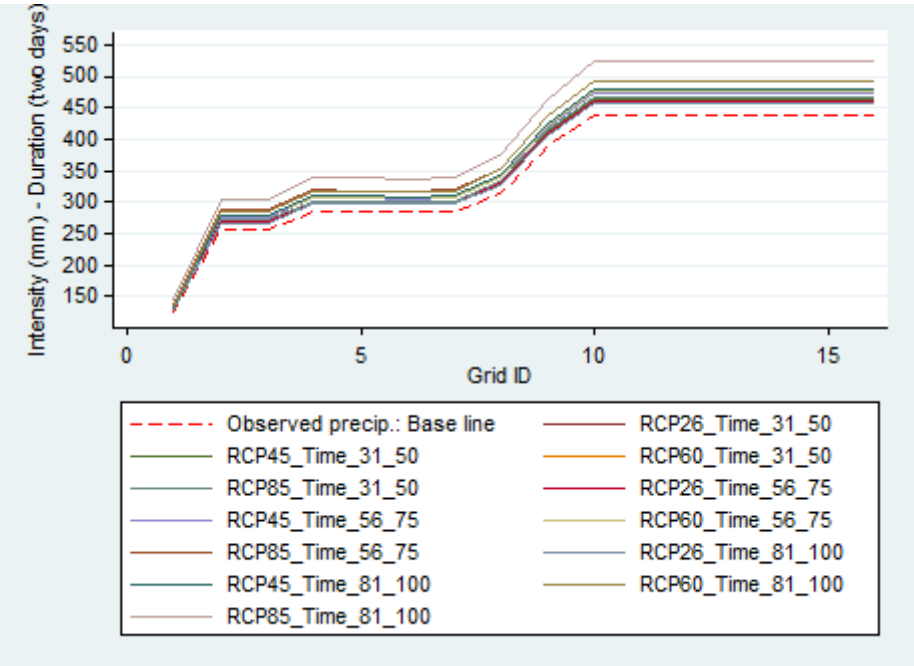
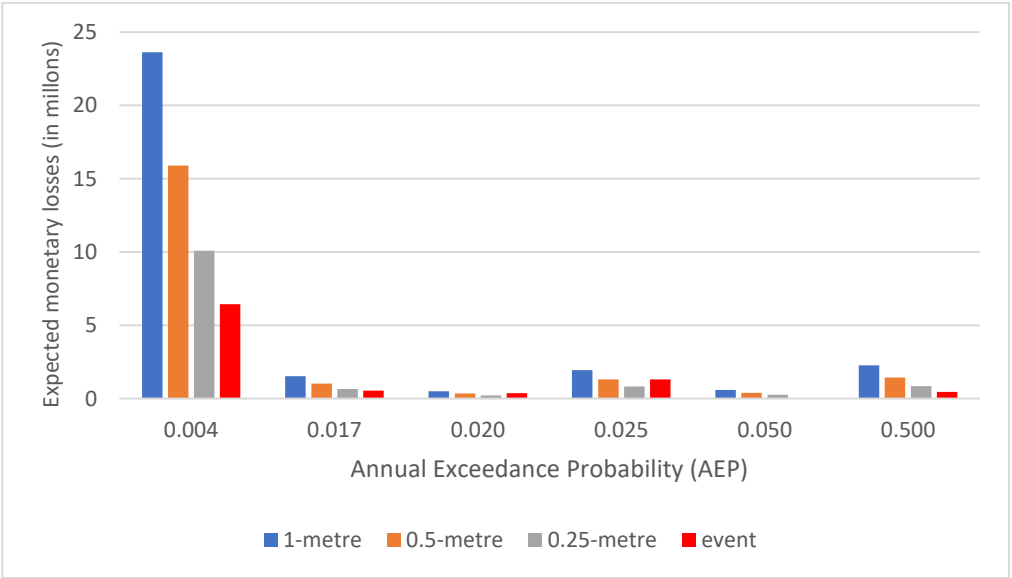


Figure 3. 9 **Observed precipitation and precipitation augmented by climate change per degree of warming, per RCP, and per time horizon across grids**.

3.5.3 Risk and its spatial distribution

Figure 3.10 shows the exceedance probability loss curves (EPLC) for the flood scenarios simulated. High monetary losses are associated with low AEP values, as expected EPLC are the result of aggregating losses per grid and AEP value. Appendix Table 3.8.2 provides the monetary figures of the estimated losses for 11 construction types. Even though I provide various scenarios instead of a single estimate of risk, a wider range of flood depth scenarios can be simulated to derive more loss probability curves.



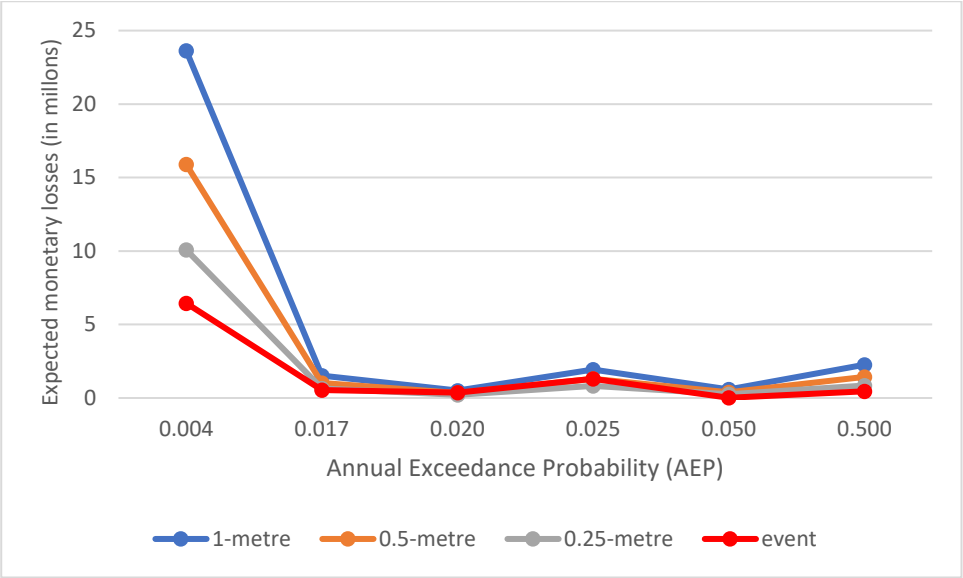


Figure 3. 10 **Annual exceedance probability loss curves for a range of flood-depth scenarios.** The relationship between losses and annual exceedance probability (AEP) values is negative, as expected. To calculate and graph annual exceedance probability loss curves, I aggregate the expected monetary losses per grid. Each grid has an associated AEP value and is calculated as the inverse of the return period. Then I aggregate the losses of the grids with the same AEP. Both graphs show the same information in the form of bars(up) and lines(down).

The risk estimates produced are calculated at the property level. The estimates are spatially explicit, which implies that the spatial distribution of property risk from flood hazard can be generated as a map that displays the expected monetary losses. The level of detail in the data allows providing estimates of the spatial distribution of risk such that 'hotspots' of losses can be differentiated from 'cold spots'. Figure 3.11 shows the spatial distribution of flood risk for a 1-metre flood depth scenario.

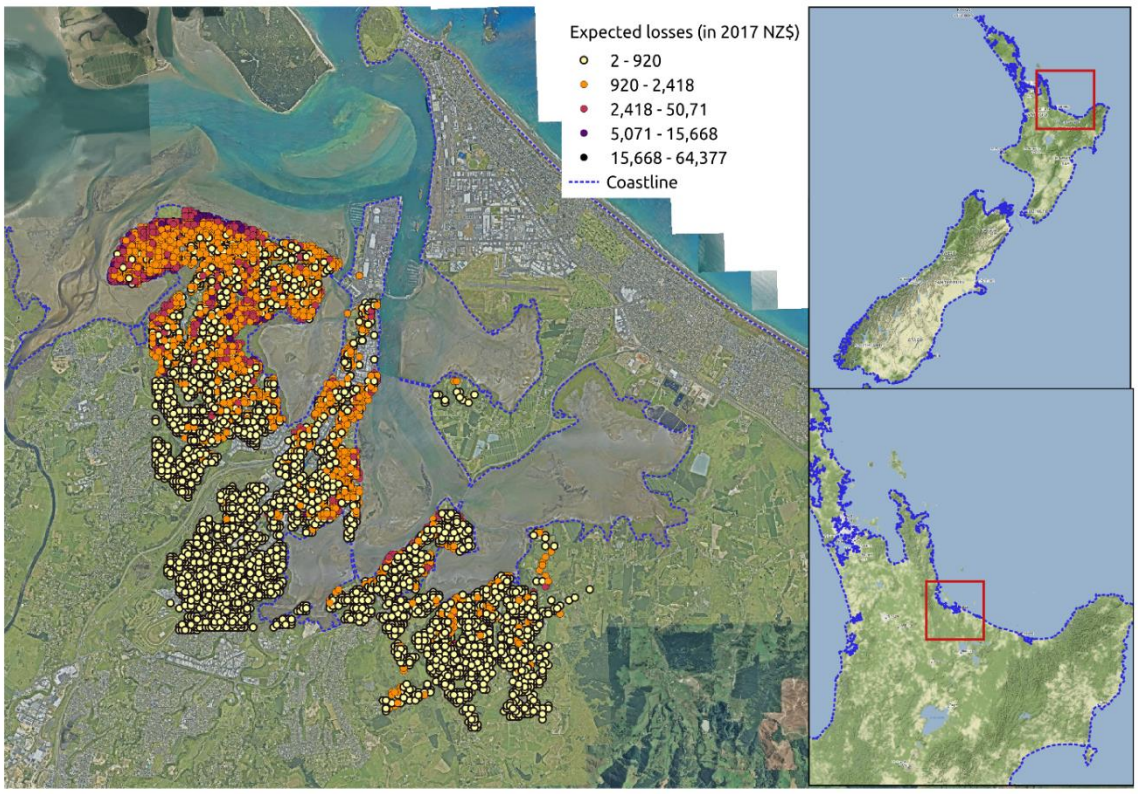


Figure 3. 11 **Spatial distribution of flood risk.** Risk is calculated as the product of the probability of damage as reported by the logistic model, the replacement value of the property and a damage ratio for a flood depth scenario of 1 metre. The classes used to group expected losses are obtained by applying the Jenks Natural Breaks classification algorithm. The algorithm produces groups of data based on their similarity values so that the variance within groups is minimised, while the variance between groups is maximised. I assign to each group a sequential colour scheme so that lighter hues denote lower risk and darker hues denote higher risk.

3.6 Conclusions

In this study, I use insurance claim data to estimate the empirical relationship between extreme precipitation and flood-related damage. Together, with projections of changes in extreme precipitation resulting from climate change, I predict the expected future costs of property damage from flooding. I use historical insurance claim data to quantify the risk of flood hazard unlike the conventional 'catastrophe modelling' framework that relies on hazard maps to calculate risk.

I show a flood risk quantification methodology that circumvents the inaccuracy or absence of flood hazard maps. It uses georeferenced historical insurance claims and sets of geospatial data that account for residential property's underlying and surrounding characteristics. The methodology uses variation in rainfall across space to identify the probability of damage and the effect of climate change on the damage probability distribution, while accounting for the influence of physical and socioeconomic factors.

I find that the highest monetary losses are associated with low-return periods, as expected. For instance, 250-year return period events are associated with 24, 16, and 10 million NZ$ losses for scenarios with a flood depth of 1, 0.5 and 0.25-meters, respectively. Nevertheless, high return periods (i.e., 2-year events) bring about sizeable losses ranging between 2 and 1 million NZ$ for the same flood depth scenarios. The likelihood of property damage resulting from climate-related changes in precipitation increases significantly, but the effect (climate change signal) is too small to cause an economically meaningful increase in risk levels.

The methodology proposed fills a gap in the literature by quantifying the risk of low-probability, high-impact flooding events caused by extreme precipitation. The methodology proposed enables the impact of climate change on flood risk to be evaluated under various emissions pathways, helping the public insurer assess its financial exposure from climate change. The level of detail in the data allows estimates of the spatial distribution of risk so that 'hot spots' of losses can be differentiated from 'cold spots'. Ultimately, a detailed representation of risk (in a map) can help tailor efforts in reducing disaster risk, future risk pricing, and potential areas where managed retreat may occur. Other highly exposed locations could exploit disaster loss databases (e.g., Desinventar) and employ this methodology to predict future damages from floods associated to extreme rainfall events.

3.7 References

Bernet, D.B., Prasuhn, V. and Weingartner, R., 2017. Surface water floods in Switzerland: what insurance claim records tell us about the damage in space and time. *Natural Hazards and Earth System Sciences*, *17*(9), pp.1659-1682.

Bihan, G.L., Payrastre, O., Gaume, E., Moncoulon, D. and Pons, F., 2017. The challenge of forecasting impacts of flash floods: test of a simplified hydraulic approach and validation based on insurance claim data. *Hydrology and Earth System Sciences*, *21*(11), pp.5911-5928.

Bouwer, L.M., 2013. Projections of future extreme weather losses under changes in climate and exposure. *Risk Analysis*, *33*(5), pp.915-930.

Buxton, R., Dellow, G. D., Matcham, I. R., Smith, W. D., Rhoades, D. A. 2013. A New Zealand framework for predicting risk due to rainfall-induced landslides, GNS Science Report 2012/22. 12 p.

Carey-Smith, T., Henderson, R. and Singh, S., 2018. High Intensity Rainfall Design System Version 4. *National Institute of Water and Atmospheric Research Ltd, Report 2018022CH, Christchurch, 73p*.

Cheng, C.S., Li, Q., Li, G. and Auld, H., 2012. Climate change and heavy rainfall-related water damage insurance claims and losses in Ontario, Canada. *Journal of Water Resource and Protection*, *4*(2), pp.49-62.

[dataset] Core Logic, 2017. Quotable Value. Accessed under a data sharing agreement.

[dataset] Earthquake Commission, 2018. Insurance claim dataset. Accessed under a data sharing agreement.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, 80(1), 27-38.

Fleming, D.A., Noy, I., Pástor-Paz, J. and Owen, S., 2018. *Public insurance and climate change (part one): Past trends in weather-related insurance in New Zealand* (No. 1124-2019-2362). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3477038 (accessed 19 January 2020)

Gradeci, K., Labonnote, N., Sivertsen, E. and Time, B., 2019. The use of insurance data in the analysis of Surface Water Flood events–A systematic review. *Journal of Hydrology*, *568*, pp.194-206.

Grahn, T. and Nyberg, L., 2017. Assessment of pluvial flood exposure and vulnerability of residential areas. *International Journal of Disaster Risk Reduction*, *21*, pp.367-375.

Handmer, J., Honda, Y., Kundzewicz, Z.W., Arnell, N., Benito, G., Hatfield, J., Mohamed, I.F., Peduzzi, P., Wu, S., Sherstyukov, B. and Takahashi, K., 2012. Changes in impacts of climate extremes: human systems and ecosystems, in: Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.K. Plattner, S.K. Allen, M. Tignor, P.M. Midgley (Eds.), Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge and New York, pp. 231-290.

Hunn, D., Dempsey, M. and Zaveri, M., 2018. Harvey's floods: Most homes damaged by Harvey were outside flood plain, data show. *Houston Chronicle. Retrieved from https://www. houstonchronicle. com/news/article/In-Harvey-s-deluge-most-damaged-homes-were-12794820. php*.

[dataset] Insurance Council of New Zealand (ICNZ). Cost of Natural Disasters. https://www.icnz.org.nz/natural-disasters/cost-of-natural-disasters/ (accessed 1 March, 2018)

[dataset] Insurance Council of New Zealand (ICNZ). Cost of Natural Disasters. https://www.icnz.org.nz/natural-disasters/cost-of-natural-disasters/ (accessed 1 March, 2018)

Jagger, T.H., Elsner, J.B. and Saunders, M.A., 2008. Forecasting US insured hurricane losses. *Climate extremes and society*, *189*, p.209.

King, G., & Zeng, L. (1999). Logistic regression in rare events data. Department of Government, Harvard University. *Online: http://GKing. Harvard. Edu.* (Accessed 1 June, 2021)

Klawa, M. and Ulbrich, U., 2003. A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Natural Hazards and Earth System Sciences*, *3*(6), pp.725-732.

[dataset] King, A. B., Bell, R., Heron, D., Matcham, I., Schmidt, J., Cousins, W. J., Reese, S., Wilson, T., Johnston, D., Henderson, R., Smart, G., Goff, J., Reid, S., Turner, R., Wright, K., & Smith, W. D. (2009). RiskScape Project: 2004-2008.

Kundzewicz, Z.W., Krysanova, V., Dankers, R., Hirabayashi, Y., Kanae, S., Hattermann, F.F., Huang, S., Milly, P.C., Stoffel, M., Driessen, P.P.J. and Matczak, P., 2017. Differences in flood hazard projections in Europe–their causes and consequences for decision making. *Hydrological Sciences Journal*, *62*(1), pp.1-14.

[dataset] Land Information New Zealand (LINZ), LIDAR

Leckebusch, G.C., Ulbrich, U., Fröhlich, L. and Pinto, J.G., 2007. Property loss potentials for European midlatitude storms in a changing climate. *Geophysical Research Letters*, *34*(5).

Mullan, A.B.; Sood, A.; Stuart, S.; Carey-Smith, T. 2018. Climate Change Projections for New Zealand: Atmosphere Projections based on Simulations from the IPCC Fifth Assessment, 2nd Edition. NIWA Client Report for Ministry for the Environment, updating the June 2016 report with a section of projections of extreme rainfall changes. WLG2015-31. June 2018. https://www.mfe.govt.nz/sites/default/files/media/Climate%20Change/Climate-change-projections-2nd-edition-final.pdf (accessed 12 May 2019)

McSaveney, M.J., Beetham, R.D. and Leonard, G.S., 2005. The 18 May 2005 debris flow disaster at Matata: Causes and mitigation suggestions. *Client Report*, *71*.

MetService News Letter https://www.metsoc.org.nz/app/uploads/2019/07/101_200506.pdf

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J.C., Bouwer, L.M., Bubeck, P., Ciavola, P., Genovese, E., Green, C. and Hallegatte, S., 2013. Assessing the costs of natural hazards–state of the art and knowledge gaps. *Natural Hazards and Earth System Sciences*, *13*(5), pp.1351-1373.

Ministry of Environment of New Zealand, 2005.https://www.mfe.govt.nz/publications/climate-change/tauranga-city-council-prepares-more-intense-rainfall/tauranga-city Accessed 20/04/2021)

[dataset] Newsome, P.F.J., Wilde, R.H., Willoughby, E.J., 2008. Land Resource Information System Spatial data layers. https://lris.scinfo.org.nz/data/category/environment/ (accessed 09 September 2018)

[dataset] National Institute of Water and Atmospheric Research (NIWA), 2005. New Zealand's Historic Events Weather Catalog https://hwe.niwa.co.nz/event/May_2005_Bay_of_Plenty_and_Waikato_Flooding (Accessed 20/04/2020)

Nguyen, C.N. and Noy, I., 2020. Measuring the impact of insurance on urban earthquake recovery using nightlights. *Journal of Economic Geography*, *20*(3), pp.857-877.

Pastor-Paz, J., Noy, I., Sin, I., Sood, A., Fleming-Munoz, D. and Owen, S., 2020. Projecting the effect of climate change on residential property damages caused by extreme weather events. *Journal of Environmental Management*, *276*, p.111012.

[dataset] Paulik, R., Craig, H. and Collins, D., 2019a. New Zealand Fluvial and Pluvial Flood Exposure. *Deep South National Science Challenge Report prepared by NIWA*. https://www.deepsouthchallenge.co.nz/sites/default/files/2019-08/2019118WN_DEPSI18301_Flood%20Exposure_Final%20%281%29.pdf (accessed 25 March 2020)

[dataset] Paulik, R., Stephens, S.A., Wadhwa, S., Bell, R., Popovich, B. and Robinson, B., 2019b. Coastal flooding exposure under future sea-level rise for New Zealand. *NIWA Client Report 2019119WN, prepared for The Deep South Science Challenge, 2019*. https://www.deepsouthchallenge.co.nz/sites/default/files/2019-08/2019119WN_DEPSI18301_Coast_Flood_Exp_under_Fut_Sealevel_rise_FINAL%20%281%29_0.pdf (accessed 25 March 2020)

Pinto, J.G., Fröhlich, E.L., Leckebusch, G.C. and Ulbrich, U., 2007. Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GCM. *Natural Hazards and Earth System Sciences*, *7*(1), pp.165-175.

[dataset] Reese, S. and Ramsay, D., 2010. RiskScape: flood fragility methodology. *Wellington, New Zealand. National Institute of Water and Atmospheric Research*, *42*.

[dataset] Rosser, B., Dellow, S., Haubrock, S. and Glassey, P., 2017. New Zealand's national landslide database. *Landslides*, *14*(6), pp.1949-1959.

Sampson, C.C., Fewtrell, T.J., O'Loughlin, F., Pappenberger, F., Bates, P.B., Freer, J.E. and Cloke, H.L., 2014. The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model. *Hydrology and Earth System Sciences*, *18*(6), pp.2305-2324.

Salmond, C.E., Crampton, P. and Atkinson, J., 2007. NZDep2006 index of deprivation (Vol. 5541, pp. 1-61). Wellington: Department of Public Health, University of Otago. https://koordinates.com/layer/1066-nz-deprivation-index-2006/ (Accessed 26/05/2020)

Spekkers, M.H., Kok, M., Clemens, F.H.L.R. and Ten Veldhuis, J.A.E., 2014. Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Natural Hazards and Earth System Sciences*, *14*(9), pp.2531-2547.

Spekkers, M.H., Clemens, F.H.L.R. and Ten Veldhuis, J.A.E., 2015. On the occurrence of rainstorm damage based on home insurance and weather data. *Natural Hazards and Earth System Sciences*, *15*(2), pp.261-272.

[data set] Tait, A., Henderson, R., Turner, R. and Zheng, X., 2006. Thin plate smoothing spline interpolation of daily rainfall for New Zealand using a climatological rainfall surface. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *26*(14), pp.2097-2115.

Tauranga City Libraries, 2005 http://tauranga.kete.net.nz/en/tauranga_local_history/topics/show/2874-bay-of-plenty-flooding-18-may-2005 (Accessed 20/04/2020)

Torgersen, G., Bjerkholt, J.T., Kvaal, K. and Lindholm, O.G., 2015. Correlation between extreme rainfall and insurance claims due to urban flooding–case study Fredrikstad, Norway. *Journal of Urban and Environmental Engineering*, *9*(2), pp.127-138.

Torgersen, G., Rød, J.K., Kvaal, K., Bjerkholt, J.T. and Lindholm, O.G., 2017. Evaluating flood exposure for properties in urban areas using a multivariate modelling technique. *Water*, *9*(5), p.318.

Woods, R., Mullan, A.B., Smart, G., Rouse, H., Hollis, M., McKerchar, A., Ibbitt, R., Dean, S. and Collins, D., 2010. Tools for estimating the effects of climate change on flood flow: A guidance manual for local government in New Zealand. *Wellington, New Zealand: Ministry for the Environment*.

3.8 Appendices

Appendix Figure 3.8.1: Climate change augmentation factors

Percentage change factors. Factors (%) to project rainfall depths derived from the current climate to a future climate that is 1 degree warmer. Taken from Carey-Smith et al. (2018)

Temperature increase. New Zealand land-average temperature increase relative to 1986—2005 for four future emissions scenarios. The three 21st century projections result from the average of six RCM model simulations (driven by different global climate models). The early 22nd century projections are based only on the subset of models that were available and so should be used with caution. Taken from Carey-Smith et al. (2018)

Example of the calculation of a climate change augmentation factor by combining percentage change factors and temperature increase

If the estimated rainfall for a 1-hour duration event and 10-year return period is 35mm, then the projected precipitation value for the RCP 2.6 (0.59) for the period 2031-2050 would be calculated as follows: 35mm × [(0.131 x 0.59) + 1] = 37.7mm.

Appendix Table 3.8.2: Expected losses

Expected losses are calculated as the probability of damage reported by the logistic model, the property's replacement value, and a damage ratio for a flood depth scenario of one meter for eleven construction types. The fragility functions to calculate the damage ratio have been sourced from Reese and Ramsay (2010).

	Flood risk: direct expected damages (in NZ$)		
	Flood depth 1m	**Flood depth 0.5**	**Flood depth 0.25**
Construction type			
Timber			
Timber, one storey - pre 1960 & post 1980	16,700,000	11,300,000	7,100,421
Timber, one storey between 1960 - 1980	10,100,000	6,721,215	4,335,146
Timber, two storey - pre 1960 & post 1980	3,317,625	2,067,451	1,179,422
Timber, two storeys between 1960 - 1980	1,215,919	792,082	494,296
Concrete shear wall			
Concrete -reinforced- shear wall, one storey pre 1960 & post 1980	19,390	7,301	3,070
Concrete -reinforced- shear wall, one storey between 1960 - 1980	58,168	39,856	34,970
Concrete -reinforced- shear wall, two storey - pre 1960 & post 1980	4,082	1,465	646
Concrete -reinforced- shear wall, two storeys between 1960 - 1980	41,473	23,367	17,879
Masonry (concrete & brick)			
Masonry concrete or brick, one storey - pre 1960 & post 1980	147,743	99,487	60,387
Masonry concrete or brick, one storey between 1960 - 1980	280,530	192,486	130,822
Masonry concrete or brick, two storeys between 1960 - 1980	621,602	525,866	447,799
	32,506,532	21,770,575	13,804,857

Expected losses are calculated as the probability of damage reported by the logistic model, the property's replacement value, and a damage ratio for four construction types. The damage ratio data have been sourced from Buxton et al. (2013). The expected losses represent a subset of records of an area (Matata) known to have been predominantly affected by rainfall-induced landslide debris.

	Damage ratio (Buxton et al., 2013)	**Rainfall-induced landslide risk: direct expected damages (in NZ$)**
Timber, one storey	0.054	34,990
Timber, two storey	0.049	37,702
Concrete	0.046	3,114
Masonry	0.056	2,331
Total		78,137

Conclusions

In this dissertation, I evaluated the effect of extreme precipitation events and climate change on residential property risk from floods, storms, and landslides in New Zealand. I applied econometric techniques that model insurance data in terms of extreme precipitation, geographic, and sociodemographic factors, along with climate data projections, to produce spatially explicit projections that reflect how much losses will increase in the future due to anthropogenic-induced climatic change. I demonstrate that insurance claim data is a good proxy for direct damage that can be exploited to evaluate and project the risk from weather-related hazards and climate change. I showcase a statistical methodology that circumvents the absence of hazard maps or their lack of accuracy to quantify risk. Overall, this dissertation answered three questions: What is the risk to residential property from extreme precipitation? What is the effect of climate change, through changes in extreme precipitation, on residential property risk? What is the spatial and temporal distribution of risk from extreme precipitation and climate change?

The effect of climate change will increase the risk of residential property damage in New Zealand. The future impacts are heterogeneous in time and space, depending on the climate change scenarios. The projections do not consider future changes in exposure and vulnerability. Thus, the calculated changes in future liabilities are driven exclusively by predicted changes in the hazard, due to climate change. Current risk tends to cluster in coastal areas and hills, steep and moderately steep terrain, areas with lower levels of social vulnerability, and areas with high concentration of properties and property values.

I believe that future risk assessments should rely on both statistical and deterministic methods, where deterministic approaches use hazard maps that account for the intensity, duration, frequency and spatial extent of extreme-precipitation-induced floods, landslides, and storms, with and without the effect of climate change. Risk assessments should incorporate future pathways of exposure or vulnerability as well. NZ's population and the value of its residential building stock have grown steadily over the past few decades, and both are projected to continue to increase. This suggests that the future risk may be higher than our estimates.

Public policy can make a difference by either reducing exposure (e.g., through better land-use planning), or reducing vulnerability (e.g., through better construction standards). With the right policies and well-targeted investments, the public insurer's liabilities can instead decrease. Another important policy consideration that should be explored is changes in what the public insurer covers, or whether our findings suggest a need for a policy change. For example, is there a need to change the premiums the public insurer collects annually? These potential policy focus areas raise many

difficult questions around responsibility, risk sharing, distributional concerns, procedural fairness, and political viability. They are all issues that the economic analysis presented here cannot resolve without resorting to social and political considerations that are best left for future research.