Geoscience Letters

RESEARCH LETTER

Open Access

Evaluation of residential building damage for the July 2021 flood in Westport, New Zealand

Ryan Paulik[1,2*], Alec Wild[3], Conrad Zorn[1], Liam Wotherspoon[1] and Shaun Williams[4]

Abstract

Reliable flood damage models are informed by detailed damage assessments. Damage models are critical in flood risk assessments, representing an elements vulnerability to damage. This study evaluated residential building damage for the July 2021 flood in Westport, New Zealand. We report on flood hazard, exposure and damage features observed for 247 residential buildings. Damage samples were applied to evaluate univariable and multivariable model performance using different variable sample sizes and regression-based supervised learning algorithms. Feature analysis for damage prediction showed high importance of water depth variables and low importance for commonly observed building variables such as structural frame and storeys. Overfitting occurred for most models evaluated when more than 150 samples were used. This resulted from limited damage heterogeneity observed, and variables of low importance affecting model learning. The Random Forest algorithm, which considered multiple important variables (water depth above floor level, area and floor height) improved predictive precision by 17% relative to other models when over 150 damage samples were considered. Our findings suggest the evaluated model performance could be improved by incorporating heterogeneous damage samples from similar flood contexts, in turn increasing capacity for reliable spatial transfer.

Keywords Flood, Damage, Residential building, Univariable model, Multivariable model

Introduction

Global economic losses from flood-related hazards demonstrate a year-on-year increase over the past few decades (Munich Re 2023). Fluvial flooding is the most frequent hazard and is expected to cause greater socio-economic exposure under future climate scenarios (Hallegatte et al. 2016). Physical damage and monetary loss estimation for high-value assets such as residential buildings are critical for evidence-based risk management solutions that limit social and economic harm to flood exposed communities (Aerts 2018).

Vulnerability is an important component in flood risk assessments. In this context, vulnerability often represents physical damage and/or tangible loss from element exposure to flood hazard characteristics and their intensities. Building damage models often represent a relative (i.e., ratio) or absolute (i.e., monetary value) damage response from increasing water depth (Merz et al. 2010). These 'depth-damage' functions or curves are a standard approach applied globally (Gerl et al. 2016), and are typically developed using empirical or judgement-based methods. Several recent studies have applied supervised learning models such as Bayesian

*Correspondence:
Ryan Paulik
ryan.paulik@auckland.ac.nz
[1] Department of Civil and Environmental Engineering, Faculty of Engineering, University of Auckland, 20 Symonds Street, Auckland 1010, New Zealand
[2] National Institute of Water and Atmospheric Research (NIWA), Greta Point, 301 Evans Bay, Wellington 6021, New Zealand
[3] Aon, Aon Centre Level 21/29 Customs Street West, Auckland CBD, Auckland 1010, New Zealand
[4] National Institute of Water and Atmospheric Research, 10 Kyle Street, Riccarton, Christchurch 8011, New Zealand

Springer Open

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

networks, artificial neural networks and Random Forests that consider multiple hazard and exposure variables explaining local damage processes (Wagenaar et al. 2018; Schröter et al. 2018; Cerri et al. 2021; Amadio et al. 2019; Carisi et al. 2018). Multivariable models evaluated in these studies demonstrate higher prediction precision for relative damage compared to univariable models based on water depth. This signals a need to consider a broad hazard and exposure explanatory variable range in damage model development for flood risk assessments.

Reliable building damage models closely replicate local flood damage processes. Flood damage observations are, however, limited to few areas or restricted to privately held databases. This hinders learning model approaches requiring large multivariable damage data sets, while univariable models can generalise flood damage processes leading to prediction underperformance (Cammerer et al. 2013; Wagenaar et al. 2017; Scorzini and Frank 2017). Flood damage models must balance empirical data collection or simulation requirements with prediction precision and reliability (Apel et al. 2009). Developing this knowledge requires consistent collection of hazard and exposure variables that explain local flood damage processes (Laudan et al. 2017; Paulik et al. 2022). Understanding the simple and complex damage model prediction performance also informs empirical data requirements for application in local flood risk assessments.

This study evaluates residential building flood damage in Westport, New Zealand. Westport's urban area sustained severe flooding from the Orowaiti and Buller Rivers between July 15 and 18 2021 (Fig. 1). Flooding covered ~ 3.8 km^2 and required over 2000 people to be evacuated (Buller Recovery 2023). More than 800 properties were inundated with 455 residential buildings damaged. Residential building and contents insurance claims exceeded 1300, with insured losses up to NZD $87.9 million (2022) (Insurance Council of New Zealand 2023). The Westport flood was highly significant for New Zealand's residential sector insurance losses in 2021, representing 50% of the NZD $174 million loss claims for all natural disasters that year.

Our flood damage evaluation has two objectives: (1) to describe the hazard, exposure and damage characteristics observed and recorded from on-site damage assessments and (2) to use the empirical damage data sample to develop and evaluate univariable and multivariable model damage prediction performance using variable

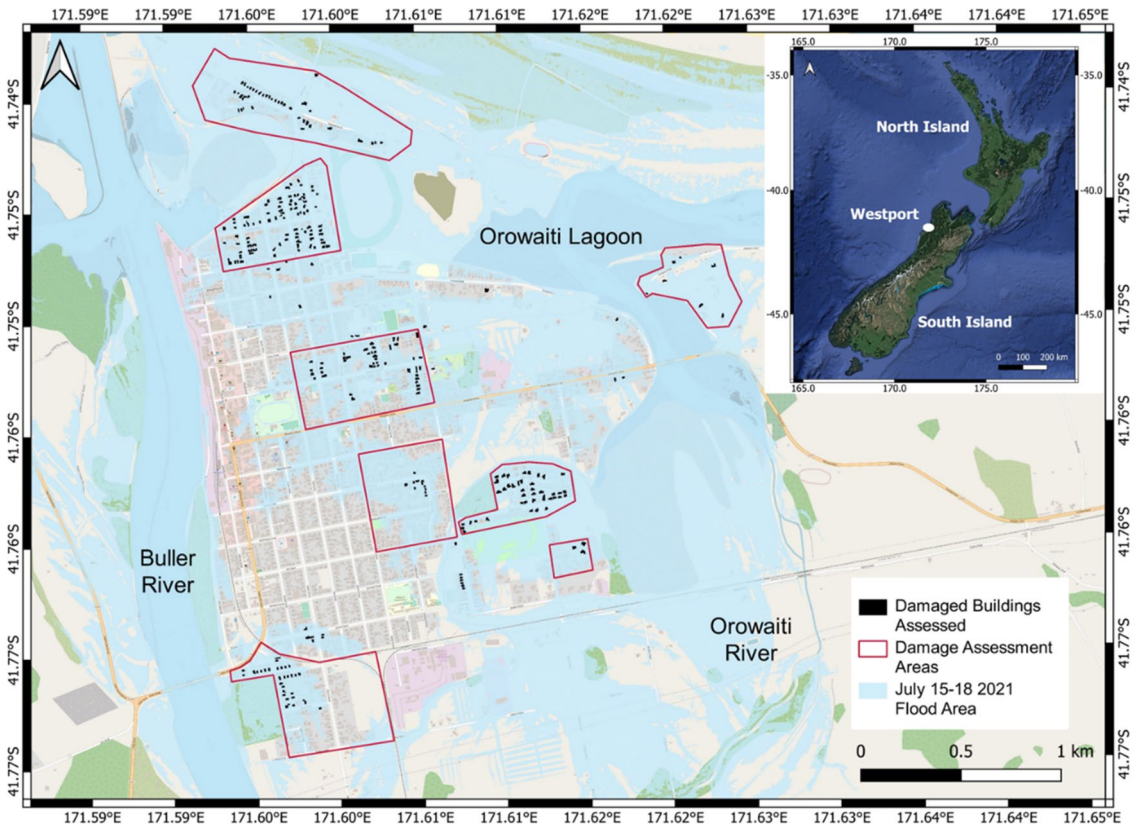

Fig. 1 Residential building damage assessment locations in Westport for the 15–18 July 2021 flood event

sample sizes and regression-based supervised learning algorithms with different complexity. This paper is structured by first describing empirical damage data collection methods in accordance with Paulik et al. (2022) followed by univariable and multivariable damage model development and evaluation methods. Hazard, exposure and damage observations are enumerated and reported to provide insight on explanatory variable influence on damage model performance. Three univariable and multivariable model approaches are each learned and evaluated for damage prediction using variable sample sizes. Conclusions are drawn from our evaluation on empirical data collection implications for local flood damage prediction.

Methods

Flood damage assessment

On-site residential building damage assessments were conducted in the Westport urban area from August 10th to 12th, 24 days after the 2021 flood event. Initial hazard and damage reports identified several residential building damage areas (Fig. 1). Data collected included: flood hazard characteristics, physical and non-physical building attributes, and physical building component damage using methods described by (Paulik et al. 2022). Building attributes observed on-site include dwelling type, structural frame, floor height, foundation, construction period, storeys and wall cladding (Table 1). These attributes form nominal and interval-scaled data structures, with a ratio-scale applied for floor height. Apart from floor height, building attributes were spatially mapped and validated on site. Measured water depths represent maximum inundation depth above ground level and above floor level (Table 1). Where access permitted (~80% of buildings) water depth above floor level was measured as indicated by highwater marks and debris lines visible on internal walls. Where access was not permitted depths were measured from an external doorstep. Other flood hazard variables including flow velocity, debris deposition (e.g., sediment) and presence of contamination could not be reliably quantified on-site 24 days after the event though building damage likely to be caused by these characteristics was recorded when observed.

Relative damage ratios (DR_b) were calculated from component and sub-component material damage summarised in Table 1. DR_b is a non-dimensional parameter between 0 and 1 representing the relative damage as the 'cost to repair/cost to replace'. On-site DR_b was estimated using: (1) an observed damage ratio (O_{DR}) for sub-components based on an ordinal scale between 0 and 1 increasing in 0.25 (i.e., 0%–25%) intervals and (2) a construction cost ratio (C_{CR}) for sub-components based on their 'replacement value/total building replacement value'. C_{CR} is estimated here based on area, dwelling type, foundations, storeys, structural frame and wall cladding. Sub-component damage ratios (C_{DR}) were then calculated based on maximum estimated O_{DR} and C_{CR} as

$$C_{DR} = O_{DR} \cdot C_{CR} \tag{1}$$

DR_b was then computed as the sum of $i = 1 \ldots n$ component damage ratios (C_{DRi}) enumerated from corresponding $j = 1 \ldots m$ sub-component damage ratios (C_{DRij}) as follows:

Table 1 Hazard, exposure and damage variables assessed in Westport (adapted from Paulik et al. 2022)

Variable		Types or Description	Data Type	Unit or Value
Hazard	Water depth above ground level	Maximum water depth above ground level	Decimal	m
	Water depth above floor level	Maximum water depth above first finished floor level	Decimal	m
	Flow Velocity	Presence of flow velocity damage on building	Boolean	0 = false; 1 = true
	Debris	Presence of debris damage on building	Boolean	0 = false; 1 = true
	Contamination	Presence of contamination damage on building	Boolean	0 = false; 1 = true
Exposure	Area	Building roof outline area	Integer	m^2
	Dwelling Type	Detached; Joined; Attached; Apartment	Text	4 classes
	Structural Frame	Brick masonry; Concrete masonry; Timber; Steel	Text	4 classes
	Floor Height	First finished floor level height above ground level	Decimal	m
	Foundation	Concrete slab; Pile; Solid wall; Mixed	Text	4 classes
	Construction Period	< 1900; 1900–1920; 1920–1940; 1940–1960; 1960–1980; 1980–2000; 2000–2020	Text	7 classes
	Storeys	Number of complete building floor levels	Integer	1 to ∞
	Wall Cladding	Brick masonry; concrete block; fibre–cement; Fibrolite; Mixed material; Roughcast; Sheet metal; Weatherboard	Text	8 classes
Damage	Damage Ratio	Relative damage of the residential building or its components	Decimal	0 to 1

$$DR_b = \sum_{i=1}^{n} C_{DRi} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} C_{DRij} \quad (2)$$

where n and m are the components (structure, external finishes, internal finishes and service) and sub-components, as described in Table 2, respectively. Building (DR_b), component (C_{DRi}) and sub-component (C_{DRij}) damage ratios provided ratio-scale values for evaluating relationships between hazard and building variables that influence residential building damage.

Vulnerability modelling

Univariable and multivariable regression models were learned for DR_b prediction performance in Westport using hazard and exposure variables presented in Table 1. A study objective was to evaluate model performance for variable damage sample sizes and different regression-based learning algorithms. This would provide insight on empirical data requirements for precise and reliable model DR_b prediction. Model performance was evaluated using ~ 50 sample increments (i.e., 50, 100, 150, 200, 247), with samples selected using a random shuffling procedure. Models were then learned for DR_b prediction and evaluated using samples selected for each shuffle. This procedure was performed using 100 iterations, ensuring complete data sample consideration for model learning in each increment. Model performance was evaluated for each shuffle using leave-one-out cross-validation to calculate regression metrics (Kohavi 1995). Cross-validation learns and evaluates model performance for each data sample. This creates a robust model performance analysis for smaller data samples as each sample represents the entire validation data set.

Univariable models

Univariable regression models correlated DR_b with maximum water depth above ground level. We selected several linear and non-linear functions based on their applications for DR_b prediction in international studies (Kreibich et al. 2008; Elmer et al. 2010; Scorzini and Frank 2017; Carisi et al. 2018; Arrighi et al. 2020). These included linear, second-order polynomial and square root functions, learned to determine water depth and DR_b relationships. Hazard variables with categorical values such as flow velocity were not considered for regression analysis, and could be the subject of future investigations on univariable model DR_b prediction performance.

Multivariable models

Supervised learning algorithms for regression were used for DR_b prediction. Selected learning algorithms included Decision Trees (Quinlan 1986), Random Forests (Breiman 2001) and Multi-Layer Perceptron (Bishop 1995). Learning algorithms were implemented using scikit-learn libraries in the Python programming language (Pedregosa et al. 2011).

Decision Trees (DT) and Random Forests (RF) were tree-based algorithms selected for their capacity to apply ensemble methods. Decision Trees have a hierarchical structure that includes root, branch, and leaf nodes to amalgamate parameters. The 'top root node' of this structure signifies the most crucial feature for categorising data labels. The DT partitioning process at the root node and progress through the branches until they arrive at the leaf nodes positioned at the bottom of the tree, yielding the regression outcome. The DT

Table 2 Residential building components and sub-components assessed in Westport (adapted from Paulik et al. 2022)

Component	Sub-component	Description	Unit or Value
Structure	Substructure	All foundation and floor-supporting structures below the underside of the lowest floor finish	0 to 1
	Structural Frame	All column and beam framework above lowest floor finish, major roof framing members	0 to 1
	Upper Floor	Suspended floors, mezzanine floors, balcony floors and roof slabs	0 to 1
External Finishes	Roof	Complete weatherproof covering for all types of roofs	0 to 1
	Walls	All work to exterior walls, including applied or in situ finishes	0 to 1
	Windows and doors	All windows and doors in exterior walls, including vertical or near vertical glazing	0 to 1
Internal Finishes	Stairs	Flights and intermediate landings including integral finishes	0 to 1
	Walls	All non-structural internal walls	0 to 1
	Doors	All interior doors including frames, architraves, finishes and hardware	0 to 1
	Floor finishes	Includes all preparatory work, screeds, surface finishes and raised floors	0 to 1
	Wall finishes	Includes all preparatory work and finishes to interior walls	0 to 1
	Ceiling finishes	Includes all preparatory work and finishes to ceilings	0 to 1
	Fixtures and fittings	Joinery fittings, built-in or fixed in position, includes glass, hardware and finishes	0 to 1
Services	Plumbing	Hot and cold-water supply, including hot water cylinder, sanitary fittings, soil, waste and vent pipes	0 to 1
	Electrical and heating	All electrical services providing lighting, power, heating and air-conditioning	0 to 1

Table 3 Regression model performance metrics for predicted (*pred*) compared to observed (*obs*) relative damage

Performance Metric	Formula
Mean Squared Error (MSE)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(pred - obs)^2$
Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n} pred - obs$
Mean Bias Error (MBE)	$MBE = \frac{1}{n}\sum_{i=1}^{n}(pred - obs)$
Quantile Range (QR)	$QR = \frac{1}{n}\sum_{i=1}^{n}(pred_{q95_i} - pred_{q5_i})/pred_{q50_i}$
Hit Rate (HR)	$HR = \frac{1}{n}\sum_{i=1}^{n} h_i; h = \begin{cases} 1 & \text{where } pred_{q5_i} \le obs_i \le pred_{q95_i} \\ 0 & otherwise \end{cases}$

algorithm is prone to overfitting for small data sample sizes or deeply grown trees. We attempt to reduce overfitting by setting a maximum tree depth of 4 leaf nodes and random selection of variable subsets for prediction at root nodes to the square root of variables in the learning data sample (Merz et al. 2013).

Random Forests are an ensemble-based method using a bagging algorithm to generate numerous damage predictions. This is accomplished by constructing multiple trees through random sampling of variable combinations. We formed RF ensembles from a learning data bootstrap sample. Predictor variable subsets were selected at each tree split (i.e., node), growing each tree to either a minimum number of nodes or maximum depth without pruning. The RF ensemble method tolerates outliers and noise in learning data samples (Hapfelmeier et al. 2014) while analysing non-linear interactions between variables (Brieman 2001). Despite these advantages the RF algorithm is sensitive to hyperparameters. Here we configured the algorithm to minimise out-of-bag (OOB) errors, calculated as the sum of squared residuals. RF tree and predictor variable combinations were tested with 100 reproductions to determine the lowest OOB error. We applied 1000 trees and 6 variables randomly sampled at each node for RF algorithm learning.

The Multi-layer Perceptron (MLP) is a neural network model employed for regression analysis to unveil non-linear relationships concealed within data samples (Gardner and Dorling 1998). The MLP comprises interconnected nodes, with connecting nodes governed by a function representing the sum of node input variables, further modified by a non-linear activation function (Bishop 1995). The architecture of MLP involves multiple layers of nodes, incorporating hidden layers positioned between the input and output layers, which help unravel complex variable interactions leading to a more accurate prediction (Amadio et al. 2019). In recognising MLP sensitivity to hyperparameters and feature scaling that creates risk of overfitting, efforts were made to mitigate this issue. A random search was conducted to identify optimal values for the number of hidden layers and nodes per layer. Feature scaling was implemented as a preprocessing step to normalise input variables and enhance the model's robustness against overfitting (Popescu et al. 2009). The Min–Max scaling method was employed, transforming explanatory variables to values between 0 and 1 by subtracting the minimum variable value and dividing by the variable value range. The normalisation approach ensured explanatory variables contributed proportionally to the model's learning process. The MLP algorithm showcased its highest performance when configured with 3 hidden layers, each consisting of 100 nodes, and utilizing the rectified linear (ReLU) activation function.

The selected tree-based learning algorithms determine feature importance for DR_b prediction within the data sample. Here, feature importance was measured by mean decrease accuracy (Breiman 2001) for DT, RF and MLP algorithms learned on the entire data sample. This required model accuracy calculations for all features, then each feature by removing other model features. Accuracy reduction was measured for each feature relative to model accuracy overall. Larger reductions result in higher mean decrease accuracy scores indicating features of greater importance for DR_b. Model accuracy was calculated for all features then calculated for each feature by removing other model features. Relative accuracy reduction for each individual feature was then measured with higher mean decrease accuracy indicating greater feature importance for model prediction.

Model performance evaluation

Model performance was evaluated using precision and reliability metrics in Table 3. Precision was evaluated using mean squared error (MSE), mean absolute error (MAE) and mean bias error (MBE). MSE measures the average squared deviation between observed (*obs*) and predicted (*pred*) DR_b, with smaller values indicating better model performance. MAE is an absolute metric for the mean error between *obs* and *pred*. The mean difference between *obs* and *pred* is measured by

MBE, with positive and negative values respectively indicating relative over or under prediction. Reliability was determined from model DR_b distributions using the quantile range (QR) for variation and hit rate (HR) for reliability (Wagenaar et al. 2018). QR represents DR_b prediction range between 5% (q5) and 95% (q95) quantiles (i.e., 90% quantile range), with larger QR demonstrating higher prediction uncertainty. HR is the DR_b prediction proportion within the observed QR, with values of 0.9 indicating high prediction reliability (Cerri et al. 2021; Gneiting and Raftery 2007).

Results and discussion

Damage observations

Damage samples were collected for 247 flood damaged residential buildings in Westport (Fig. 1). Water depths above ground level measured between 0.1 m and 1.4 m, and depths above floor level from − 0.1 m–0.8 m (Fig. 2a). Nearly 90% of buildings were exposed to water depths above floor level < 0.5 m. Higher water depths above floor levels were observed for concrete slab (μ 0.27 m, σ 0.13 m) foundations compared to piles (μ 0.25 m; σ 0.16 m) and solid walls (μ 0.22 m; σ 0.15 m). Compared to other foundation construction types, lower floor heights for concrete slab foundations (Fig. 2b) caused higher water depths above floor level. Damage presence from other flood hazards was not (i.e., flow velocity (n = 0)) or rarely (i.e., debris (n = 2), contamination (n = 2)) observed.

Buildings in the sample were constructed over a 120-year period with most comprising a timber structural frame (96%) and with one storey (93%). Over 60% were constructed on pile, solid wall or mixed foundations, with the remaining on concrete slab. Buildings constructed prior to 1960 were constructed with native hardwood timber floor finishes and often incorporated modern building materials and services. Suspended flooring systems for buildings constructed after 1960 were often constructed with floor finishes comprising composite timber materials, e.g., low or medium density fibreboard (BRANZ 2023). Buildings with concrete slab foundations were more frequently constructed after 1980 and the most common (75%) foundation after 2000. Building areas after 1980 were larger on average than earlier construction periods. Building area increased by 101 m^2 on average between the 1900–1920 and 2000–2020 construction periods. Multi-storey and non-timber structural frame buildings only accounted for 8% of observations.

Relative building (DR_b) and component (C_{DRi}) damage is presented in Fig. 3a and sub-components (C_{DRij}) in Fig. 3b–d. DR_b ranged from 0.01 to 0.5 (μ 0.29, σ 0.1), with internal finishes contributing the highest C_{DRi} (μ 0.22, σ 0.07). Internal finishes sub-components observe a C_{DRi} range between 0 and 0.35 (Fig. 3b). A high proportion of damage occurs for 0 m–0.5 m water depth above floor level whereby several internal finishes sub-components (i.e., internal doors, floor finishes, wall finishes, and fittings and fixtures) were damaged on water contact. Physical damage of composite timber materials used for floor finishes occurred, often resulting in indirect damage to other internal finishes (e.g., internal walls, ceiling finishes) and services (e.g., plumbing) during clean-up and repair activities. Native hardwood timber floor finishes, although more resistant to damage from direct water contact, usually required replacement upon drying due to warping. Internal finishes in buildings constructed with suspended floors sustained a slighly higher C_{DRi} (μ 0.23, σ 0.07) compared to those with concrete slab foundations (μ 0.2, σ 0.07).

External finishes observed relatively low C_{DRi} (μ 0.03; σ 0.02) in comparison to internal finishes (Fig. 3a). Sub-components damage susceptibility varies, with windows and doors more frequently damaged (90%) compared to

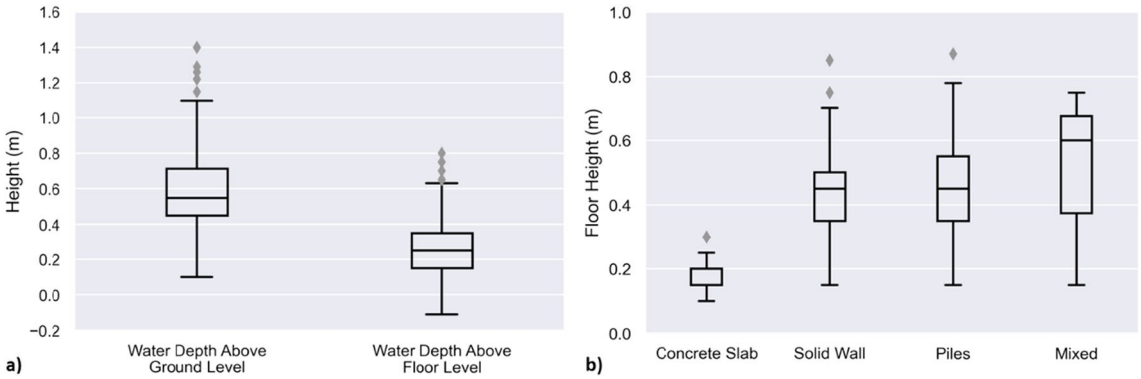

Fig. 2 Distributions of **a** water depth variables, and **b** foundation types and floor heights observed at Westport

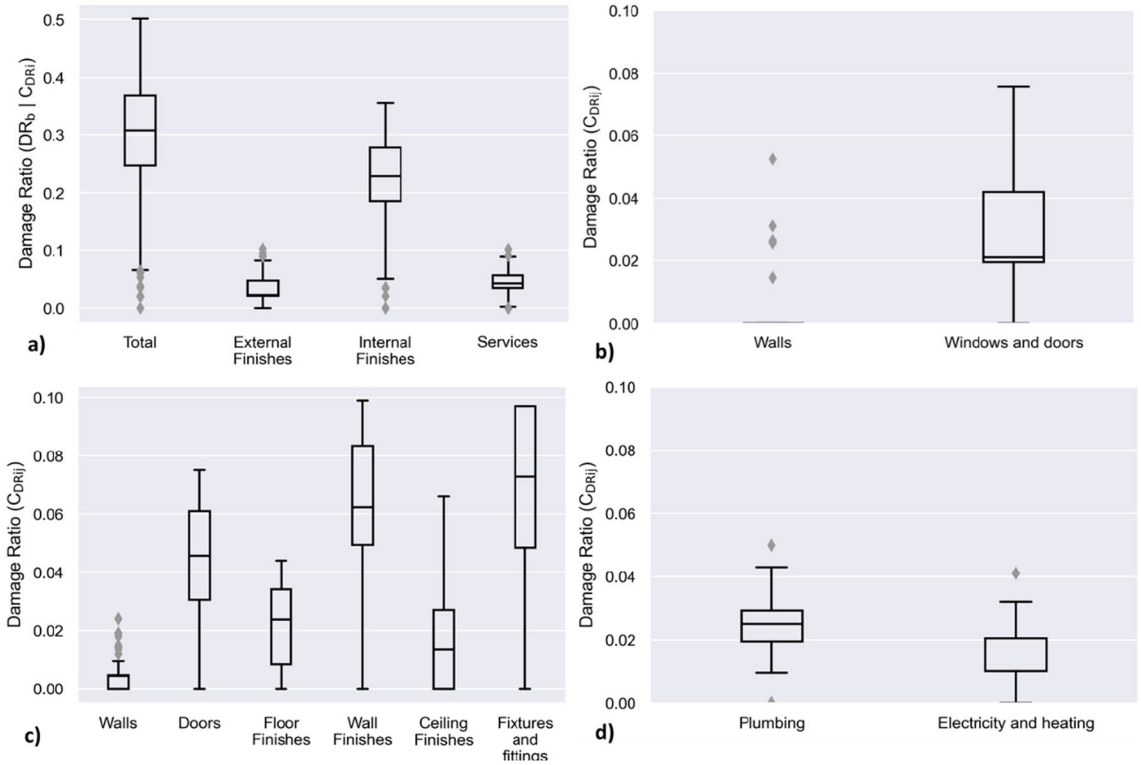

Fig. 3 Relative damage distributions observed for **a** buildings and components, and **b** external finishes, **c** internal finishes, **d** services sub-components. Structure components and stairs sub-component of internal finishes are omitted due to no observed damages

external walls (28%). Materials such as timber doors and glass panes were highly susceptible to physical damage on water contact. Despite external windows and doors being frequently damaged their relative contribution to DR_b is low on average (μ 0.03, σ 0.01) compared to sub-components comprising internal finishes (Fig. 3c).

Service finishes (i.e., electrical and heating) were frequently damaged below floor level due to water contact of near-to ground condenser units for heating and air-conditioning systems. Electrical and heating services damaged below floor level have a minor contribution to C_{DRi} ($\mu < 0.01$) for services (Fig. 3d). Electrical fittings were replaced on water contact, while electrical wiring in buildings constructed after 1960 was reusable. Services C_{DRi} (μ 0.05, σ 0.01) between 0 m and 0.5 m water depth above floor level often resulted from indirect damage to electrical and heating and plumbing services caused during internal finishes clean-up and repair.

Vulnerability models

Hazard and exposure variable importance for DR_b was evaluated for DT and RF models. Water depth above ground and floor level show high importance (Fig. 4), which agrees with numerous international studies on its primary influence for DR_b (Gerl et al. 2016; Mohor et al. 2020). Water depth above floor level had highest importance. This variable has limited application in international damage models (Gerl et al. 2016), as floor height is rarely collected during damage assessments. Other hazard variables had low DR_b importance which was expected as no component damage from high flow velocities was observed. Area showed the highest building attribute importance, which may occur from its larger homogeneity relative to other exposure variables. Dwelling type and structural frame were more homogenous within the study area and showed low DR_b importance. These variables are often used to classify residential buildings for univariable flood damage models (Merz et al. 2010). In Westport the low importance of these variables affirmed water depth as the key driver for flood damage.

Understanding damage data sample size effect on model prediction performance was a key study objective. Model precision and reliability changes in response to sample size variability is presented in Fig. 5. With increasing sample size, univariable and multivariable models showed improved reliability overall for DR_b predictions within observed 90% quantile ranges (QR), and hit rates (HR) between 0.8 and 0.9. Precision metrics

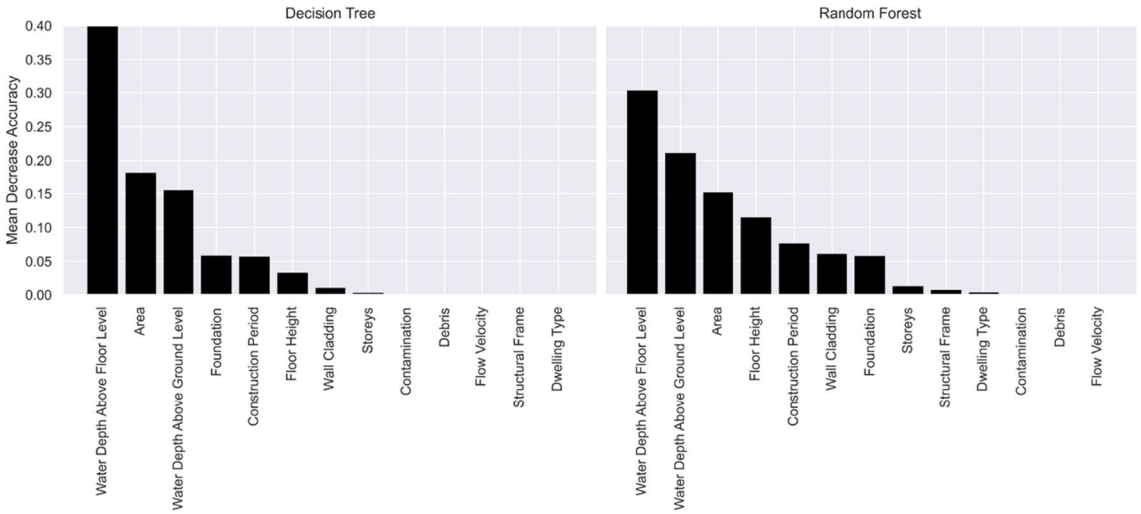

Fig. 4 Feature importance for relative damage (DR_b) to residential buildings in Westport estimated from Decision Tree and Random Forest models

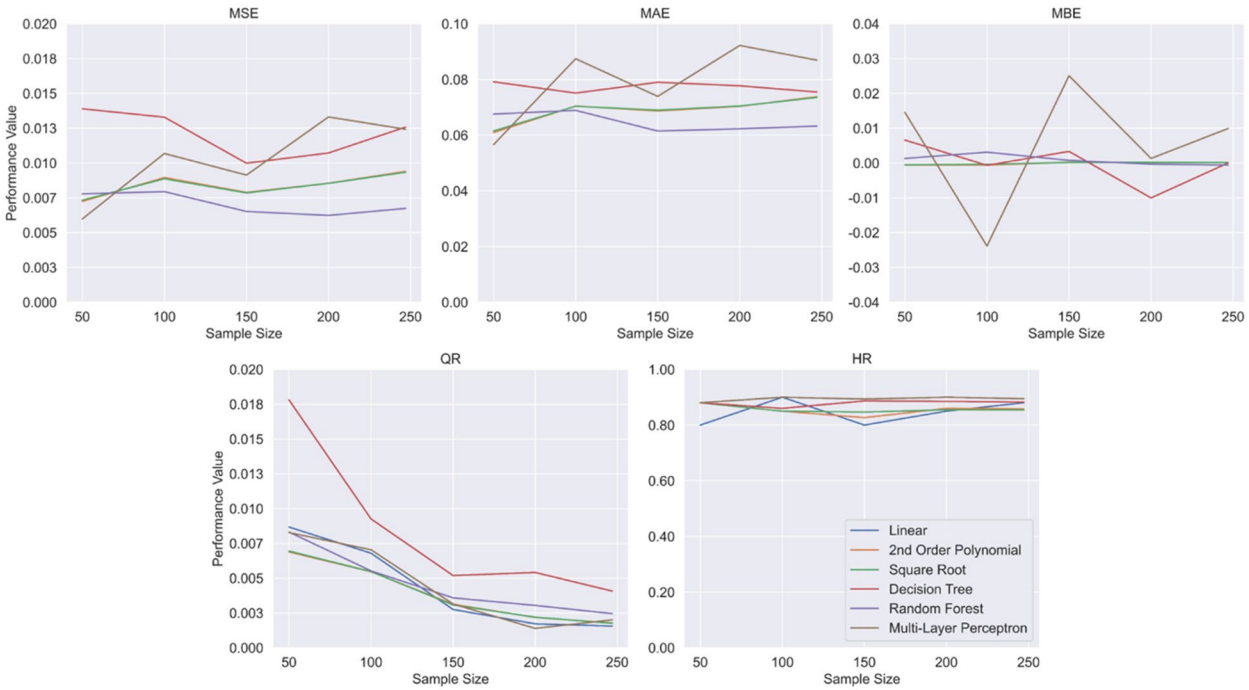

Fig. 5 Predictive performance of univariable and multivariable models based on variable damage sample size

(i.e., MSE, MAE) demonstrated model overfitting when sample size increased. Overfitting occurred for most models when more than 150 samples were used, and more than 50 samples for univariable models and MLP. Several reasons may cause overfitting. Learned univariable models could be affected by 'noise' caused by limited water depth and DR_b variability in the data sample, while multivariable models may learn on variables with lower feature importance. These observations have broader implications for empirical models derived from event-specific data samples. Univariable models representing homogenous residential building classes and multivariable models should be limited to exposure variables of higher importance for observed damage. This supports the findings of Cerri et al. (2021) in Germany, observing that geometric building variables with low

DR_b importance caused RF model precision loss. Limited DR_b sample heterogeneity at Westport may further affect model capacity to predict conditions where relatively lower and higher DR_b was observed. DR_b values between 0.2 and 0.4 formed over 70% of the damage data sample, raising the potential for model learning on a DR_b value range. Additional DR_b data samples outside this range may have caused precision loss suggesting future flood damage assessments should focus on increasing DR_b heterogeneity in the local damage data sample.

Multivariable models analyse non-linear interactions between hazard and exposure variables to evaluate DR_b from multiple interdependent relationships (Merz et al 2013). In Westport, evaluated model capacity to predict observed DR_b is illustrated in Figs. 6, 7. The RF model showed up to 17% higher precision (MAE) compared to other models when more than 150 samples were

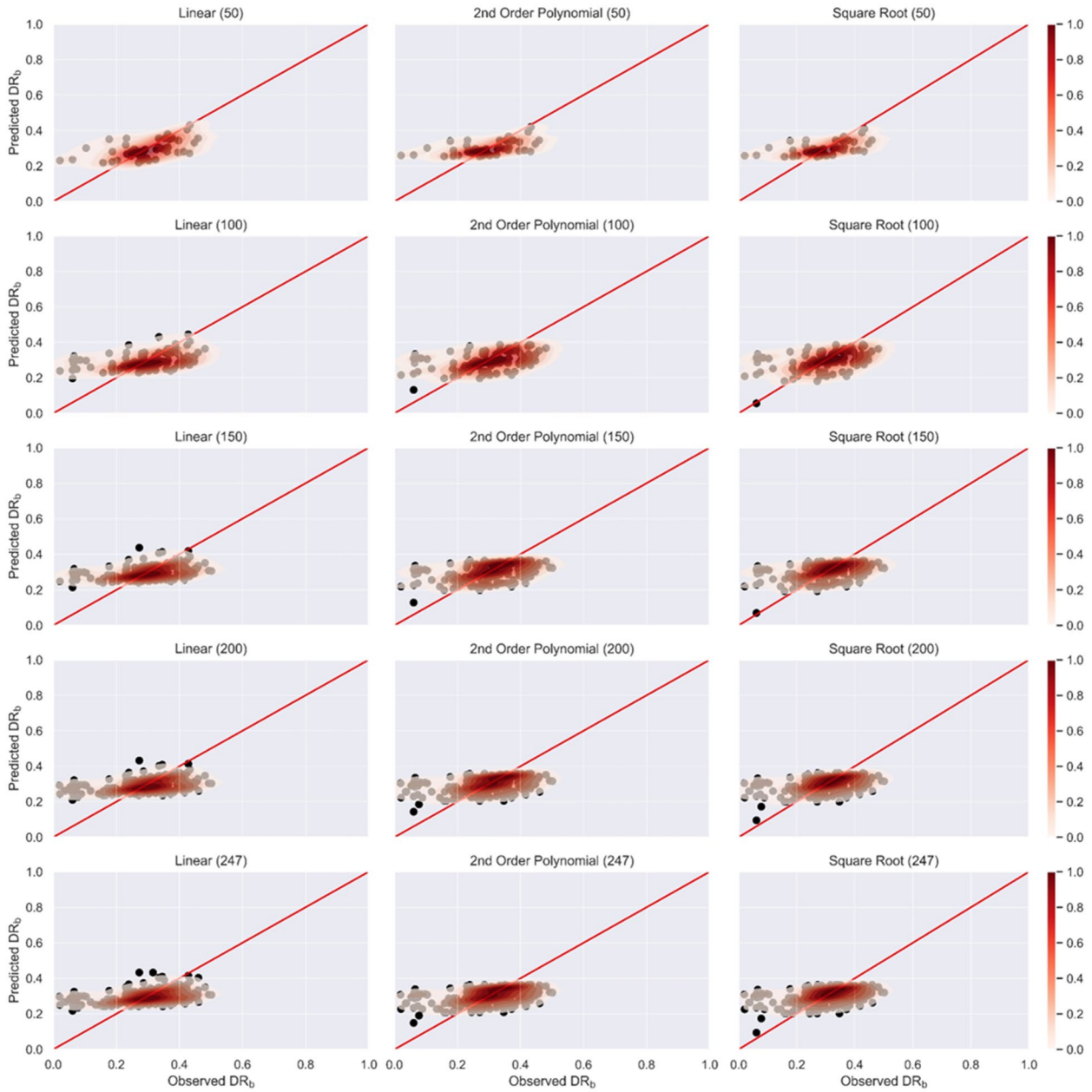

Fig. 6 Relative damage (DR_b) to residential buildings in Westport estimated from univariable models. Damage sample size used for model learning is indicated in parentheses

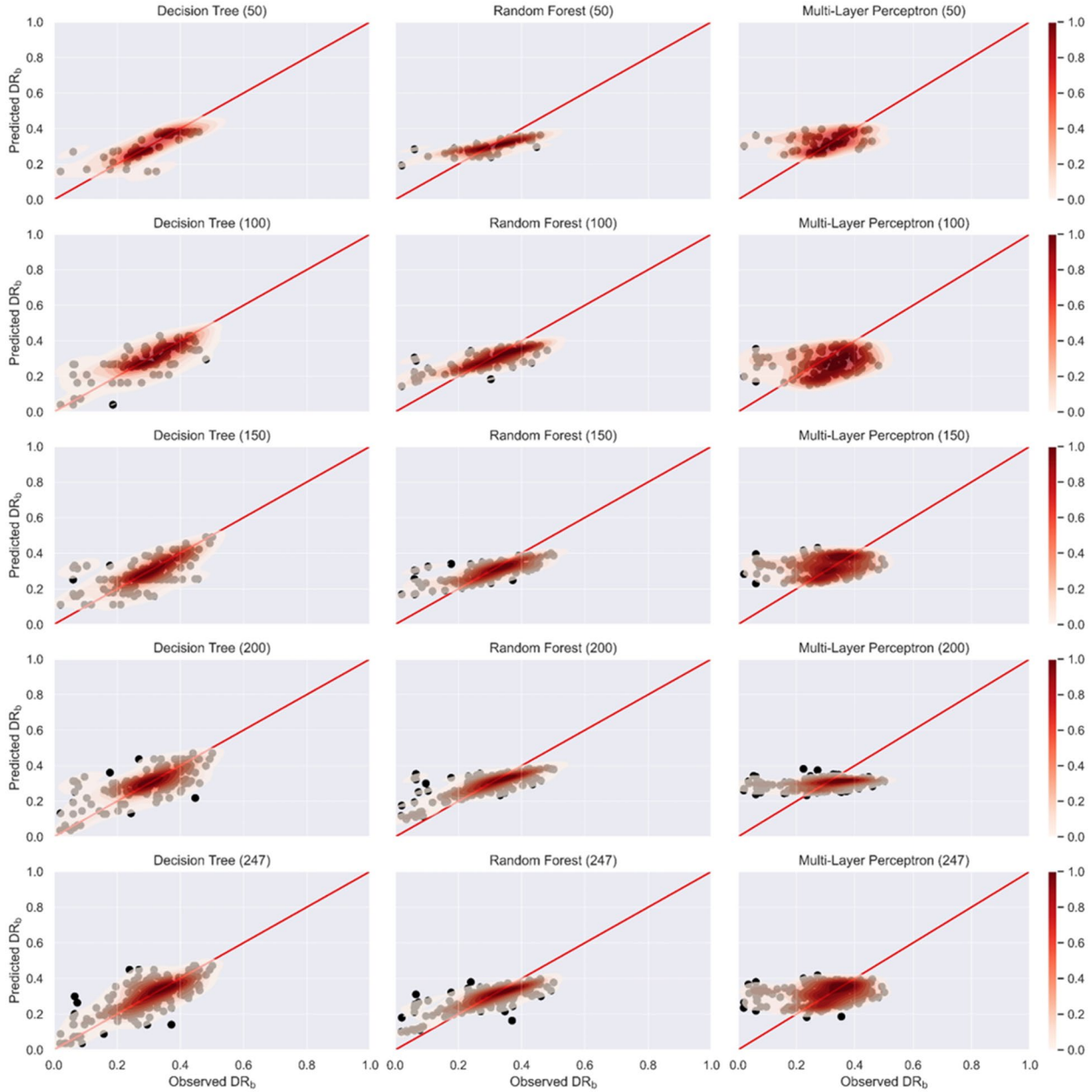

Fig. 7 Relative damage (DR_b) to residential buildings in Westport estimated from multivariable models. Damage sample size used for model learning is indicated in parentheses

evaluated. The ensemble method has lower susceptibility to overfitting as shown by increasing DR_b prediction density along the identity line in Fig. 7 as sample size increases. Hit rates (HR) close to 0.9 and lower QR with larger sample sizes (Fig. 7) further indicates high reliability for RF model predictions. The RF algorithm has greater tolerance to 'noise' within data samples (Brieman 2001) and the learned algorithm in Westport demonstrated greater capacity to predict DR_b despite relatively few samples represent $DR_b < 0.2$ and > 0.4. This supports findings from several international studies whereby RF showed superior damage prediction performance over other univariable or multivariable models when learned on 10 s or 100 s of data samples (Carisi et al. 2018; Amadio et al. 2019; Malgwi et al. 2021).

Univariable models predicted DR_b with higher overall precision than multivariable models DT and MLP. Square

root and second-order polynomial models demonstrated similar precision and reliability (Fig. 5), with MSE and MAE up to 25% and 10% higher respectively than DT and MLP when data sample size exceeded 150. Non-ensemble model performance improvement fluctuated as sample sized increased indicating potential overfitting. Square root and DT models were able to predict DR_b values < 0.2 (Figs. 6 and 7), while DT also predicted DR_b > 0.4. While DT predicted a broader DR_b range compared to other evaluated models, higher uncertainty (QR) was observed (Fig. 5). This observation reinforces a need for further model learning based on broader DR_b heterogeneity from future flood events at Westport or other locations with comparatively similar flood hazard and residential building exposure characteristics.

Residential building damage for the July 2021 flood in Westport shows the importance of representing local damage processes in empirical damage models. Building variables showed homogenous attributes with 92% comprising one storey and timber structural frame. Storey and structural frame variables had low DR_b importance at Westport (Fig. 4), leading to DT and MLP model overfitting when learned on larger damage samples (Fig. 5). Water depth variables showed highest importance and may have influenced higher prediction precision for RF models, even when learned on fewer than 150 damage samples. These findings suggest improvements to univariable and multivariable model performance could be gained by limiting model learning to highly importance predictor variables (Cerri et al. 2021) and extending damage sample heterogeneity (Wagenaar et al. 2018) for water depth variables corresponding to DR_b values < 0.2 and > 0.4. Achieving the latter requires model relearning on either damage samples collected from future Westport flood events or incorporating samples from other locations demonstrating homogeneous damage processes (Di Bacco et al. 2023). Empirical models learned on heterogeneous damage samples could improve both prediction precision and reliability and capacity for spatial transfer to other flood hazard contexts.

Conclusions

This study contributes to the growing global understanding of univariable and multivariable model applications for flood damage assessment. Residential building damage data collected for 247 residential buildings after the July 2021 Westport flood event in New Zealand has demonstrated utility for univariable (linear, second-order polynomial, square root) and multivariable (Decision Trees, Random Forest, Multi-layer Perceptron) flood model development. We used variable sample sizes from the empirical data set to evaluate relative damage prediction performance change for selected models.

Flood damaged buildings at Westport exhibited homogenous attributes and component damage caused primarily from water depth. Most buildings comprised a single storey and timber structural frame, and were exposed to < 0.5 m water depth above floor level. Internal finishes were frequently damaged, while no structure damage indicated water alone was likely the primary hazard characteristic causing damage. This was confirmed by feature importance analysis that demonstrated high importance of water depth variables and low importance for building variables such as structural frame and storeys.

Limited damage sample heterogeneity affected model learning outcomes for residential building damage prediction. Models improved reliability as damage sample size increased for model learning. Univariable and non-ensemble multivariable models (Decision Trees, Multi-layer Perceptron), however, were susceptible to overfitting as demonstrated by precision loss in response to increasing sample size. Overfitting resulted from limited heterogeneity of relative damage observed, and variables of low importance affecting model learning. The ensemble-based Random Forest algorithm considered multiple important variables (water depth above floor level, area and floor height) and return predictive precision by 17% relative to other models when over 150 damage samples were considered. Our findings suggest empirical model prediction performance for the empirical models evaluated could be improved by incorporating heterogeneous damage samples from similar flood contexts, in turn increasing capacity for reliable spatial transfer.

Acknowledgements

The work presented was funded from the University of Auckland Doctoral Scholarship, National Institute of Water and Atmospheric Research (NIWA) Strategic Scientific Interest Fund work programme on "Hazard Exposure and Vulnerability" (CARH2405), New Zealand Ministry of Business, Innovation, and Employment (MBIE) Endeavour Fund (CONT-69394-ENDRP-NIW) and Resilience to Nature's Challenges National Science Challenge Multihazard Risk Programme Grant GNS-RNC043. The authors would like to thank Buller District Council, West Coast Regional Council and West Coast Civil Defence Emergency Management Group for providing intelligence on flood damaged building locations and local communications with the Westport residents on the purpose of our damage assessment activities. Jake Langdon, Jo Patterson and Erica Andrews are especially thanked for their assistance and helpful advice in the field.

Author contributions

RP developed the concept for this study. RP, AW and SW performed the data collection analysis. RP, CZ and LW verified the analysis. RP wrote the original manuscript. All authors discussed the results and edited the manuscript.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Received: 3 September 2023 Accepted: 12 February 2024
Published online: 22 March 2024

References

Aerts JC (2018) A review of cost estimates for flood adaptation. Water. https://doi.org/10.3390/w10111646

Amadio M, Scorzini AR, Carisi F, Essenfelder AH, Domeneghetti A, Mysiak J, Castellarin A (2019) Testing empirical and synthetic flood damage models: the case of Italy. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-19-661-2019

Apel H, Aronica GT, Kreibich H, Thieken AH (2009) Flood risk analyses—how detailed do we need to be? Nat Hazards. https://doi.org/10.1007/s11069-008-9277-8

Arrighi C, Mazzanti B, Pistone F, Castelli F (2020) Empirical flash flood vulnerability functions for residential buildings. SN Applied Sci. https://doi.org/10.1007/s42452-020-2696-1

Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press

BRANZ (2023) Renovate – the technical resource for industry on the renovation of houses from different eras. https://www.renovate.org.nz/. Accessed 13 Mar 2023

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Buller Recovery (2023) Weather Events. https://bullerrecovery.org.nz/weather-events/. Accessed 28 Apr 2023

Cammerer H, Thieken AH, Lammel J (2013) Adaptability and transferability of flood loss functions in residential areas. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-13-3063-2013

Carisi F, Schröter K, Domeneghetti A, Kreibich H, Castellarin A (2018) Development and assessment of uni- and multivariable flood loss models for Emilia-Romagna (Italy). Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-18-2057-2018

Cerri M, Steinhausen M, Kreibich H, Schröter K (2021) Are openstreetmap building data useful for flood vulnerability modelling? Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-21-643-2021

Di Bacco M, Rotello P, Suppasri A, Scorzini AR (2023) Leveraging data driven approaches for enhanced tsunami damage modelling: Insights from the 2011 Great East Japan event. Environ Model Soft. https://doi.org/10.1016/j.envsoft.2022.105604

Elmer F, Thieken AH, Pech I, Kreibich H (2010) Influence of flood frequency on residential building losses. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-10-2145-2010

Gardner MW, Dorling SR (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos Environ. https://doi.org/10.1016/S1352-2310(97)00447-0

Gerl T, Kreibich H, Franco G, Marechal D, Schröter K (2016) A review of flood loss models as basis for harmonization and benchmarking. PLoS ONE. https://doi.org/10.1371/journal.pone.0159791

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. J Am Stat Assoc. https://doi.org/10.1198/016214506000001437

Hallegatte S, Vogt-Schilb A, Bangalore M, Rozenberg J (2016) Unbreakable: building the resilience of the poor in the face of natural disasters. World Bank Publications, Washington, DC

Hapfelmeier A, Ulm K, Strobl C (2014) A new variable importance measure for random forests with missing data. Stat Comput 24(4):513–528

Insurance Council of New Zealand (2023) Cost of natural disasters. https://www.icnz.org.nz/natural-disasters/cost-of-natural-disasters. Accessed 28 Apr 2023.

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai 14:1137–1145

Kreibich H, Thieken AH (2008) Assessment of damage caused by high groundwater inundation. Water Res Resear. https://doi.org/10.1029/2007WR006621

Laudan J, Rözer V, Sieg T, Vogel K, Thieken AH (2017) Damage assessment in Braunsbach 2016: data collection and analysis for an improved understanding of damaging processes during flash floods. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-17-2163-2017

Liaw A, Wiener M (2002) Classification and regression by Random Forest. R News 2(3):18–22

Malgwi MB, Schlögl M, Keiler M (2021) Expert-based versus data-driven flood damage models: a comparative evaluation for data-scarce regions. Int J Disaster Risk Reduct. https://doi.org/10.1016/j.ijdrr.2021.102148

Merz B, Kreibich H, Schwarze R, Thieken A (2010) Review article: assessment of economic flood damage. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-10-1697-2010

Merz B, Kreibich H, Lall U (2013) Multi-variate flood damage assessment: a tree-based data-mining approach. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-13-53-2013

Mohor GS, Hudson P, Thieken AH (2020) A comparison of factors driving flood losses in households affected by different flood types. Water Res Resear. https://doi.org/10.1029/2019WR025943

Munich Re (2023) Risks posed by natural disasters. Losses from natural disasters. https://www.munichre.com/en/risks/natural-disasters.html. Accessed 24 Apr 2023.

Paulik R, Wild A, Zorn C, Wotherspoon L (2022) Residential building flood damage: insights on processes and implications for risk assessments. J Flood Risk Manag. https://doi.org/10.1111/jfr3.12832

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N (2009) Multilayer perceptron and neural networks. WSEAS Trans Circ Syst 8:579–588

Quinlan JR(1986) Induction of decision trees Machine Learning 1(1):81–106 https://doi.org/10.1007/BF00116251

Schröter K, Lüdtke S, Redweik R, Meier J, Bochow M, Ross L, Nagel C, Kreibich H (2018) Flood loss estimation using 3D city models and remote sensing data. Env Model Soft. https://doi.org/10.1016/j.envsoft.2018.03.032

Scorzini AR, Frank E (2017) Flood damage curves: new insights from the 2010 flood in Veneto, Italy. J Flood Risk Manag. https://doi.org/10.1111/jfr3.12163

Wagenaar D, De Jong J, Bouwer LM (2017) Multi-variable flood damage modelling with limited data using supervised learning approaches. Nat Hazards Earth Syst Sci. https://doi.org/10.5194/nhess-17-1683-2017

Wagenaar D, Lüdtke S, Schröter K, Bouwer LM, Kreibich H (2018) Regional and temporal transferability of multivariable flood damage models. Water Res Resear. https://doi.org/10.1029/2017WR022233

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.