https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

Modelling Flood Losses to Microbusinesses in Ho Chi Minh City, Vietnam

Anna Buch[1,2,3], Dominik Paprotny[2], Kasra Rafiezadeh Shahi[3], Heidi Kreibich[3], Nivedita Sairam[3]

[1]Institute of Geography, University of Heidelberg, Heidelberg, 69117, Germany
[2]Potsdam Institute for Climate Impact Research (PIK), Transformation Pathways, Potsdam, 14473, Germany
[3]German Research Centre for Geosciences (GFZ), Section 4.4 Hydrology, Potsdam, 14473, Germany

Correspondence to: Anna Buch (a.buch@stud.uni-heidelberg.de), Nivedita Sairam (nivedita@gfz-potsdam.de)

Abstract. Microbusinesses are important sources of livelihood for low- and middle-income households. In Ho Chi Minh City (HCMC), Vietnam, many microbusinesses are set up in the ground floor of residential houses susceptible to urban floods. Increasing flood risk in HCMC threatens the financial resources of microbusinesses by damaging business contents and causing business interruption. Since flood loss estimations are rarely conducted at object-level resolution and are often focused on households or large companies, the losses suffered by microbusinesses are often overlooked. This study aims to derive the drivers of flood losses in microbusinesses by applying a Conditional Random Forest to survey data (content losses: n=317; business interruption losses: n=361) collected from microbusinesses in HCMC. The variability of content losses and business interruption were adequately explained by the revenues of the businesses from monthly sales, age of the building where the business is established and water depth in the building during the flood event. Based on the identified drivers, probabilistic loss models (non-parametric Bayesian Networks) were developed using a combination of data-driven and expert-based model formulation. The models estimated the flood losses for HCMC's microbusinesses with a mean absolute error of 3.8 % for content losses and 18.7 % for business interruption losses. The Bayesian Network model for business interruption performed with a similar predictive performance when it was regionally transferred and applied to comparable survey data from another Vietnamese city, Can Tho. The flood loss models introduced in this study make it possible to derive flood risk metrics specific to microbusinesses to support adaptation decision making and risk transfer mechanisms.

Plain Language Summary. Many households in Vietnam depend on revenues from microbusinesses (shop-houses). However, losses caused by regular flooding to the microbusinesses are not modelled. Business turnover, building age and water depth are found to be the main drivers of flood losses to microbusinesses. We built and validated probabilistic models (non-parametric Bayesian Networks) that estimate flood losses to microbusinesses. The results help in flood risk management and adaption decision making for microbusinesses.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

1 Introduction

Comprehensive risk management requires empirical evidence on drivers of risk and assessment of potential impacts. The lack of information on vulnerability of certain economic sectors or social groups, and their often-limited participation in local risk management, in turn foster a lack of awareness among decision-makers leading to biased risk management strategies. As impacts of climate change become more severe, comprehensive risk management that protects society as a whole is imperative - in particular the vulnerable and under-represented groups. However, it is often not plausible in low- and middle-income countries due to poor data availability. An example of a vulnerable economic sector in a society with a high flood risk, explored in this study, are microbusinesses in Ho Chi Minh City (HCMC), Vietnam.

HCMC is one of the world's most exposed cities to flood risk under current and future conditions (Hallegatte et al. 2013; Scussolini et al. 2017). Similar to other Vietnamese metropolises, HCMC lies in the great delta areas in the south of the country. These flat, riverine and coastal regions experience regular flooding in particular during the rainy season. In HCMC, these regular floods are often the result of compound events caused by the simultaneous occurrence of high tides, heavy rainfall and high flows of the Saigon and Dong Nai rivers and their tributaries (Tran 2014; Thuy et al. 2019). Other large cities in the delta areas of South Vietnam also experience regular urban flooding, for instance, Can Tho City in the Mekong Delta (abbreviated as Can Tho). Urban floods in Can Tho are predominantly fluvial in nature, such as a major flood event in 2011. Despite the ongoing efforts to improve protection and adaptation measures on private and municipal levels, climate change and the ongoing growth of these important economic centres increase their risk to urban flooding (Güneralp et al. 2015; Rentschler et al. 2022). The existing infrastructure and adaptation measures in these cities are unable to counterbalance the new risks caused by intensified flood events and ongoing urban pressure (e.g. Bouwer 2011; Jha et al. 2012; Formetta & Feyen 2019; Kreibich et al. 2022).

We define microbusinesses, including household-businesses, according to the definition of the World Bank, as very small businesses with less than ten employees. However, this general definition for microbusinesses needs to be adapted to the regional context of South-East (SE) and South (S)-Asia. Microbusinesses in these countries tend to employ usually less than three people. In most cases, they are located on the ground floor of a building with residences on the upper floors, commonly called shophouses in Vietnamese cities. Microbusinesses provide an important source of income for unemployed family members and people with limited opportunities on the labour market, likewise migrant workers and people who received less possibilities of schooling (Samantha 2018). Together with the operations of small and medium-sized companies (SMEs), microbusinesses drive the rapid economic developments of many SE-Asian states since the last decades (Trinh & Thanh 2017). According to Vietnam's economic census of 2017, it is estimated that around 75 % of all enterprises are microbusinesses in the country (General Statistics Office 2018). Vietnam's microbusinesses engage around 11 % of all employees (General Statistics Office 2018) and the density of microbusinesses is particularly high in economic centres such as HCMC and other delta-cities like Can Tho. The economic importance of HCMC becomes evident when the region's contribution to Vietnam's

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

total economic output is considered - the HCMC region accounts for approximately 40 % of the national GDP (General Statistics Office 2018). These values highlight the relevance of microbusinesses for Vietnam's local and national economy.

Microbusinesses are particularly vulnerable to the negative consequences of regular flooding due to their limited financial resources and inadequate support by local authorities and the government (JETRO 2017; Leitold et al. 2020). As a consequence, microbusinesses often rely on their neighbouring network to cope with flooding (Chinh et al. 2016; 2017; Leitold & Revilla Diez 2019; Leitold et al. 2021). Bank loans or microcredits are less common due to the usually rather low credit rating (Patankar 2019). In terms of flood losses this means that repair measures and other business investments are often directly financed by the savings of the microbusiness owner. Insufficient or missing flood insurance policies can further exacerbate the situation of flood-affected businesses (KPMG 2016; Patankar 2019). Besides temporal decline in revenues, repair costs or poor future prospects, worse case impacts may include business closures or unemployment among business owners and their employees (Jha et al. 2012).

In addition to the largely studied structural damages, the commercial sector also suffers directly from economic loss of business content (e.g. inventory, goods, equipment) and due to business interruption. The latter refers to the decline in business revenues due to interrupted business operations of flood-affected businesses during a reference period such as the flood month or period of flooding (Meyer et al. 2013; Chinh et al. 2016). However, our definition of interruption losses does not consider long-term losses or impacts on businesses outside the flood zone. The literature on commercial losses often focuses on companies of various sizes in Europe or the US and these studies indicate that indirect losses represent a significant share of flood consequences (e.g. Hallegatte 2008; Merz et al. 2010; Koks & Thissen 2016; Sieg et al. 2019 and Tsinda et al. 2019). Since the business structures and available resources for larger firms differ considerably from those of small- and micro-sized companies (JETRO 2017, Leitold & Diez 2019), the state-of-the-art approaches for commercial flood loss modelling are not generalizable to Vietnam's microbusinesses. However, the better the drivers of flood losses for a specific sector are understood, the more informed loss assessments can be made and investments towards flood adaptation improved (Sieg et al. 2017).

Modelling flood losses in low- and middle-income countries is often hampered by the lack of comprehensive and open-source data, which necessitates reliance on primary data collection campaign. The lack of information on flood losses among microbusinesses is explained by the fact that they mainly operate in the informal sector, which makes it difficult to record and thus to estimate their flood losses (Garschagen 2015; Rand & Tarp 2020). Despite these limitations, only some studies have analysed and modelled content losses to microbusinesses and SMEs in S- and SE-Asia (Chinh et al. 2016; Wijayanti et al. 2017; Samantha 2018). To the authors knowledge, there is no existing analysis elucidating the drivers of flood losses in microbusinesses in the context of low- and middle-income countries. However, the identification of the loss drivers is crucial to develop meaningful flood loss models that capture the role of the drivers in influencing losses (Rözer et al. 2019). The heterogeneity in flood loss processes at the object-level necessitates the development of multi-variable, probabilistic approaches capable of capturing non-linear effects (Schröter et al. 2014; Vogel et al. 2014; Rözer et al. 2019; Paprotny et al. 2020; 2021; Rafiezadeh Shahi et al. 2024). The absence of such probabilistic loss models in the contexts of microbusinesses impedes quantification and inclusion of uncertainties for adaptation decision making. Furthermore, multivariate flood loss

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

models are rarely evaluated under conditions other than those under which they were developed, consequently their applicability for spatial/temporal transfers remains unknown (Apel et al. 2009; Gerl et al. 2014; Ootegem et al. 2017; Vogel et al. 2018; Amadio et al. 2019). Our study aims to address these limitations in the state-of-the-art flood loss modelling approaches for microbusinesses in the context of low- and middle-income countries by deriving empirical evidence on the drivers of flood losses in microbusinesses in HCMC; calibrating and validating a process-based Bayesian Network (BN) models for HCMC that predict content and business interruption losses; and evaluating the transferability of the BN models by applying them on comparable data from a different city (Can Tho).

The manuscript is organized into the following sections: Section 2 explains the empirical survey dataset used in the study, Sect. 3 the methodology implemented including feature selection and the development of the probabilistic flood loss model, Sect. 4 presents and discusses the results of this study, followed by conclusions in Sect. 5.

2 Data – Post-flood survey of microbusinesses

The flood loss models for microbusinesses are built using empirical data from HCMC and the transferability of the models is evaluated using empirical data from Can Tho. Both datasets are based on in-person structured surveys undertaken with flood-affected microbusinesses. The owner or the manager of the microbusiness was asked to respond to the survey.

2.1 Ho Chi Minh City

The survey at HCMC was conducted during September-October 2020 and collected responses of 250 microbusinesses which experienced flooding between 2010 and the time of the survey (2020). The interviewees could respond to questions on two flood events – the most severe and the most recent event – which leads to 397 loss records in the HCMC dataset. The questionnaire covers aspects relating to the economic flood losses of microbusinesses and their potential drivers, i.e. flood characteristics, building conditions, business characteristics, undertaken emergency and precautionary measures, the respondent's risk perception and their socio-economic profile. The majority of microbusinesses surveyed in HCMC are shops or retailers (76 %) mostly selling groceries or other everyday objects. Around 17 % are services, such as restaurants or for reparations, and only 7 % produce consumer goods or processes raw materials. The presented shares of the business sectors in the HCMC survey are representative for entire Vietnam (General Statistics Office 2018).

2.2 Can Tho

Between August and December 2011 severe flooding affected several districts of Can Tho causing damages to various economic sectors. The survey was undertaken in January-February 2012 and received responses from 373 microbusinesses out of which 313 furnished information on losses. The questionnaire was comparable to the survey undertaken in HCMC. The value distributions of common variables queried by the HCMC and Can Tho survey is shown in the Supplementary Information, Figure S1. They consisted of about 88 questions covering topics including flood hazard characteristics of the

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

2011 flood event, flood preparedness, warning and emergency measures, flood losses to business contents and losses due to business interruption, risk perceptions, and the business and socio-economic characteristics of the respondents. Compared to the HCMC survey, the Can Tho dataset includes fewer microbusinesses operating in the trading sector (46 %). Consequently, more respondents provide services (45 %) or belong to Can Tho's manufacturing sector (9 %).

Details on the pre-processing of the survey data is provided in the Supplementary Information (Sect. 1). In order to derive the drivers of flood loss and develop the loss model, the 14 pre-processed candidate predictors from HCMC (Table 1) are used.

Table 1: Candidate predictors and response variables of flood losses from HCMC

Candidate predictor	Value range [Mean, Median]	Explanation
Water depth	1 – 150 [34; 30]	Water depth [cm] refers to the measured flood water level above the ground floor of the shophouse
Inundation duration	0.2 – 240 [11; 3]	Duration [h] of flood inundation of the shophouse
Contamination (indicator)	0: no visible 1: light 2: heavier [x; 1]	Type of visible contamination of the flood water
Flow velocity (m/s)	0.1 – 0.5 (calm – turbulent) [x; 0.3]	Flow velocity [m/s] of flood water on the street
Structural precautionary measures (indicator)	0.0 – 1.0 [0.2; 0.0]	Ratio between number of implemented measures and number of possible measures. These measures are often implemented during major renovations or building constructions. They comprise the usage of water-resistant building material and the elevation of the building or parts of it.
Non-structural precautionary measures (indicator)	0.0 – 1.0 [0.4; 0.3]	Ratio of number of implemented measures and number of possible measures. These measures need to be purchased before the flood event. They are quite affordable compared to structural measures and comprise wet-proofing of valuables, installation of the electricity control system at a higher level, acquisition of mobile water-barriers and pumping equipment.
Emergency measures (indicator)	0.0 – 1.0 [0.4; 0.5]	Ratio of number of implemented measures and number of possible measures. These measures can be applied shortly before or during the flood event. They comprise saving of documents, relocation of furniture, vehicles or products., usage of sandbags and sealing of doors and windows.
Building age (years)	0 – 100 [20; 18]	The age of the shophouses at the time of flooding [years]
Building area (sqm)	12 – 850 [87; 74]	Building footprint of the shophouse [sqm]
Flood experience	3 – 151 [82; 76]	Number of experienced floodings between 2010 and 2020 [n]
Flood resilience (indicator)	0 – 5 (weak – strong) [x; 3]	Interviewee's appraisal of support by authorities or the neighbourhood
Number of employees (number)	1 – 9 [x; 2]	Number of employees [n]
Average monthly income (Euro)	18 – 3314 [430; 295]	Available monthly income [Euro 2020] of the interviewee, in most cases the owner of the microbusiness

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

Average monthly sale (Euro)		92 – 2762 [370; 276]	Averaged revenue from monthly sales and production [Euro 2020]. The variable is representative for the value and quantity of goods and products hold by an individual microbusiness, i.e. it reflects the business size and type.
Flood loss variables			
Relative business interruption loss [%]		0 – 100 [18.2; 10.0]	Decline in revenues due interrupted business operations (e.g. reduced production and sales) during the flood event [%]. The decline is relative to the potential revenue that would be generated without the flood. 0 % represents no business interruption; 100 % a complete business downtime during the flood event.
Relative content loss [%]	Chance of content loss	0, 1 (zero-loss, loss) [x; 0]	Chance of flood losses to business content. 0 represents the absence of content loss (zero-loss case); 1 the occurrence of content loss (loss case).
	Degree of content loss [%]	0.2 – 93.5 [12.3; 4.0]	Flood losses relative to the value of business content, only loss cases [%]; Values close to 0 % represent minor flood losses to business content; 100 % the entire loss of business content.

3 Methodology

Our approach for modelling flood impacts specific to microbusinesses consists of two components. First, we identify the drivers of content and interruption losses to HCMC's microbusinesses based on the set of candidate predictors (Table 1). For this feature selection, a variant of Random Forest was chosen which provides a feature importance method not biased towards correlated predictors (see, Sect. 3.1.1). Second, we calibrate probabilistic loss models specific to microbusinesses based on the identified drivers (see, Sect. 3.2).

Since more than half of the businesses in both cities reported no or only marginal content losses (see, Supplementary Information Figure S3), we model the chance of loss to business content separately from the degree of loss. The former represents the absence or presence of content loss to microbusinesses and is binary (absence/presence), while the latter represents the severity of experienced loss and is a continuous value (0, 100]. Since only 40 % reported zero interruption losses, the aspects of business interruption loss (chance and degree of interruption loss) were not considered separately (see, Supplementary Information Figure S2).

The predictive performances of the Machine Learning (ML)-model used for feature selection and the flood loss models were assessed by the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Bias Error (MBE) and Symmetric Mean Absolute Percentage Error (SMAPE). The MAE metric is chosen due to its outlier robustness as a selection criterion for the cross-validation of the ML-based models (Chicco et al. 2021).

3.1 Feature Selection

3.1.1 Conditional Random Forest

The ML-model utilised for feature selection is Conditional Inference Trees, which were initially introduced by Hothorn et al. (2006) and extended by Strobl et al. (2007) to an ensemble of trees, so-called Conditional Inference Random Forest (CRF).

Each tree is grown only by a subset of features, which were identified before as significant based on their p-values (Hothorn et al. 2006).

The importance of each feature for model predictions is assessed by an unbiased version of the permutation-based feature importance method - the Conditional Permutation Importance (CPI, Debeer & Strobl 2020). The CPI accounts on linear and non-linear interactions of correlated predictors using a Chi-square test (Debeer & Strobl 2020). Though the CPI is a measure well suited for the feature selection from CRF models (Levshina 2020), the method is rather computationally expensive, but applicable for the presented approach due to the rather small sets of training samples.

The training and evaluation of the CRF model was done via nested cross-validation. Nested cross-validation is a state-of-the-art technique for an unbiased generalisation ability of a model (Krstajic et al. 2014). It is particularly recommended for relatively small datasets (Brill 2022; Liu et al. 2022). Repeated 10 inner folds were used for hyperparameter tuning and 10 outer folds for performance evaluation of the estimators. Of these 10 evaluated estimators, the estimator with the best performance (smallest MAE-score) was used for feature selection, i.e. for identifying the drivers for the degree of content loss and relative interruption loss to microbusinesses.

3.2 Probabilistic Flood Loss Models for Microbusinesses

3.2.1 Probabilistic Logistic Regression

The chance of content loss, as one component of relative content loss, is modeled using a probabilistic logistic regression model, applied on the candidate predictors from Table 1. To prevent model overfitting, probabilistic logistic regression incorporates L1 and L2 regularization, which effectively manage multicollinearity in the feature space. The model returns the probability of assigning a microbusiness to either zero-loss or loss category. However, the sample sizes between both categories are imbalanced (see, Supplementary Information Figure S3). To overcome this imbalance, the logistic regression model was trained on a weighted sample of zero-loss and loss cases. Similar as the CRF, the logistic regression model was also trained and evaluated by nested cross-validation consisting of 10 inner and 10 outer folds. However, we used all validated classifiers for modelling chance of content loss, rather than a single classifier due to their rather moderate predictive performance.

3.2.2 Bayesian Network

Bayesian Networks are probabilistic, graphical models with many applications to flood loss modelling (Vogel et al. 2014; Wagenaar et al. 2018; Rözer et al. 2019; Paprotny et al. 2020; 2021; Rafiezadeh Shahi et al. 2024). They have the benefit of explicitly representing the dependency structures, quantifying uncertainty and the possibility of including expert knowledge alongside data.

In this study, non-parametric Bayesian Networks (BNs), were chosen for modelling the degree of content losses and for modelling the relative business interruption losses. As the term "non-parametric" indicates, this type of Bayesian Network does not rely on prior assumptions about the distribution of the data (Du & Swamy 2019). Non-parametric BNs were first introduced by Kurowicka & Cooke (2006) and later extended by Hanea et al. (2006; 2015). They rather make use of the ranks of the empirical data which is favourable in terms of the varying distributions of the flood loss and its potential drivers. These drivers are used to construct the graphs of the BNs. Confirmed by the Cramer-von Mises measure for the single variable-pairs of the BN graphs, the joint distributions of the variables are represented by Gaussian copulas.

The constructed flood loss models are calibrated and internally validated on the flood losses reported in HCMC. The performance of the single Bayesian Network model for relative interruption loss was determined by 5-fold cross-validation, while the performance of the modelling approach used for relative content loss was assessed by calculating the prediction bias directly between the reported losses and their probabilistic estimates.

The transferability of these models is assessed based on their performance in predicting flood losses in Can Tho. The performance of the models for each prediction task is benchmarked against the performance of a reference Random Forest (RF) model (Chinh et al. 2017).

4. Results and discussion

The results are structured as follows. Firstly, the performance metrics of the CRF model are reported and the most important flood loss drivers for microbusinesses are derived (see, Sect. 4.1). Subsequently, the identified drivers are used to construct the Bayesian Network models (see, Sect. 4.2). The flood loss models are validated (see, Sect. 4.3). Finally, the transferability of the models to other delta-cities is tested using the survey data from Can Tho as a case study (see, Sect. 4.4).

4.1 Drivers of flood losses to Microbusinesses

The cross-validation of the CRF model showed that all its estimators, validated on the outer folds of the nested cross-validation, had similar moderate performances in predicting the degree of content losses and the relative interruption losses. Furthermore, the similar sets of hyperparameter values across the validated estimators shows that the applied ML-algorithm is suitable for both prediction tasks. The prediction of the degree of content losses resulted in an averaged MAE of 12.8 %, RMSE of 18.4 %, MBE of -0.2 % and SMAPE of 51.4 %, while the prediction of relative interruption losses led to an averaged MAE of 17.5 %, RMSE of 22.6 %, MBE of 0.3 % and SMAPE of 59.9 %. However, high SMAPE scores are caused by less severe cases of content loss being overestimated, while moderate and severe loss cases are often underestimated by the estimators. The same applies to the prediction of interruption losses.

Revenues returned from business operations (mthly. sales) are the most influencing factor for the severity (degree) of loss to business content, while the number of applied emergency measures has the greatest impact on interruption losses. Further main drivers for the degree of content loss and relative interruption loss are the age of the shophouse (building age), hydrological variables and the monthly income (Figure 1).

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

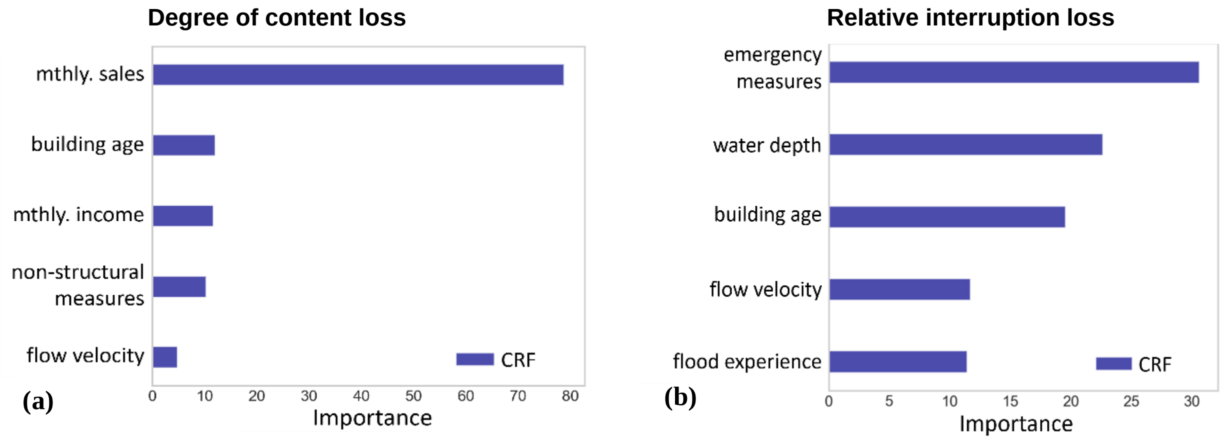

Figure 1: Feature Importance of the best-performed Conditional Random Forest estimator for predicting (a) the degree of losses to business content and (b) for relative interruption losses to microbusinesses in HCMC. Only the five most predictive features are shown.

The identified drivers of flood losses to microbusinesses in HCMC differ partly from those of less flood-experienced companies in high-income countries. The companies' flood experience, its number of employees and the building area are identified in this research as less important loss factors for microbusinesses but were identified as relevant for larger companies in Europe (e.g. Kreibich et al. 2007 (flood experience); Sieg et al. 2017 (employees- content loss); Sultana et al. 2018 (employees- interruption loss); Schoppa et al. 2020 (building area)). The missing impact of flood experience could be explained by HCMC's regular floodings which lead to a high level of adaptive behaviour across the residents.

4.2 Bayesian Network flood loss models

The graph of the non-parametric Bayesian Network for estimating the degree of loss to business content consists of six nodes, the graph for relative business interruption loss out of five nodes; the structures of the graphs are visualised in the Figures 2 and 3. The first parent node of each BN graph was set based on the strongest unconditional rank correlation between a predictor and the flood loss variable for degree of content loss and relative interruption loss, respectively. This highest unconditional correlation coefficient exists for both constructed BNs for the variable-pair of water depth in the building to the corresponding flood loss variable (Spearman's rank coefficient value for degree of content loss: 0.37; for relative interruption loss: 0.24). However, in the feature space for relative interruption losses exists an equally strong correlation to the indicator of emergency measures. This feature was identified by the CRF model as the most predictable for estimating relative interruption losses. However, the conditionalization of the BN with emergency measures showed that information about emergency measures becomes unimportant when adding further features to the BN graph. Thus, the BN graph for relative interruption loss was constructed without it. The variables for the remaining parent nodes (2. – 5. parent node) were selected based on the strongest conditional ranking correlation. During this process, the CRF ranking was used as a guideline so that the most important drivers were tested first as potential parent nodes.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

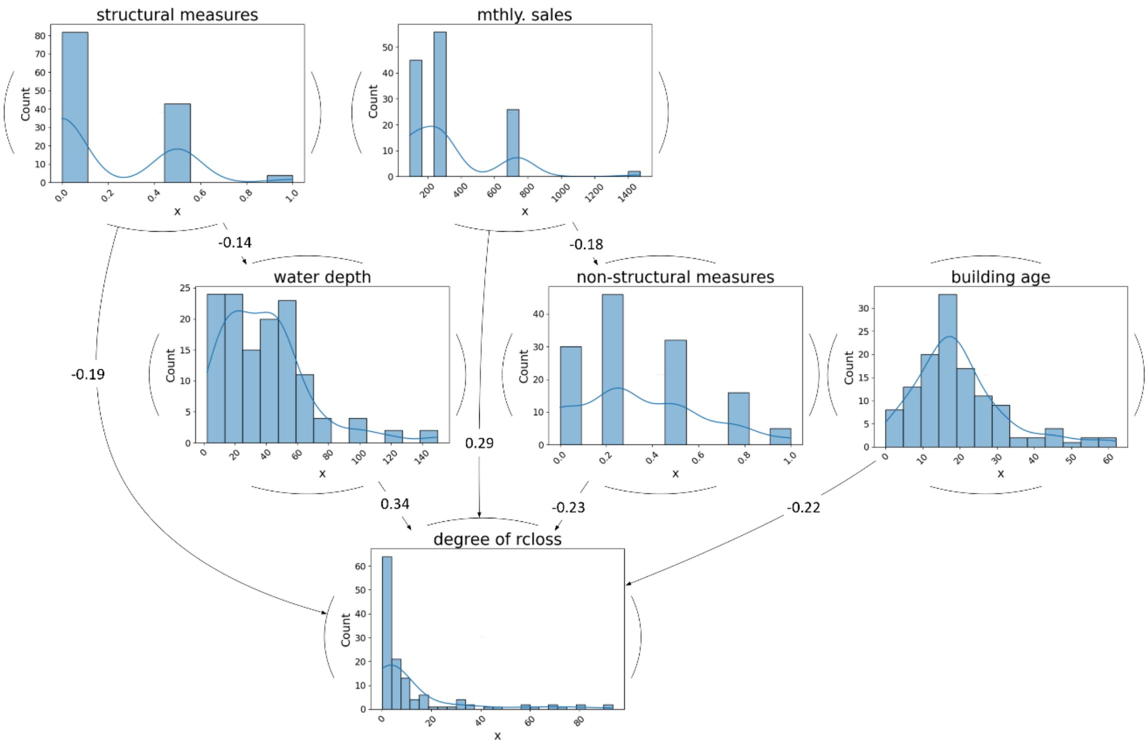

Figure 2: Structure of the Bayesian Network for predicting the degree of loss to business content (degree of rcloss).

The predictors of flood losses and their assumed dependencies in the BN graphs are presented in the following:

- The degree of losses to business content and relative interruption losses, correlate with water depth in the shophouse (**water depth**). It is the predictor with the strongest rank correlation to both flood loss types and was also previously identified as a relevant predictor by the CRF model. Rising water levels in the building directly increase the potential damage to low-lying goods, equipment and machinery (Kreibich et al. 2010; Chinh et al. 2015; Sieg et al. 2017). Apart from (non-)structural damages, the flooding of business premises itself or indirectly through power outages potentially leads to business interruptions (Kreibich et al. 2009; Sultana et al. 2018).
- High flow velocities (**flow velocity**) on the streets are associated in the BN graphs with more severe business interruptions but are not important for modelling the degree of content losses. Business activities are potentially affected when high velocities hamper the transportation, such as by relocated objects blocking streets, or damage infrastructure, such as the energy systems (Jha et al. 2012). Additionally, flow velocities have a direct effect on the water level in buildings by pressing water through openings in windows or doors, as also expressed in the BN graph for relative interruption losses of Figure 3. However, the missing impact of flow velocity on the degree of content loss is explained by the high level of preparedness of HCMC's residents, such as the relocation of vehicles before potential flooding (Chinh et al. 2016), whereas business activities, especially those of shops and small retailers, cannot or can only partially be relocated to other premises.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

- Age of the shophouse (**building age**) and degree of content loss have a negative relationship in the BN graph. The majority of shophouses in the HCMC samples were built in the last 30 years before the flood event, i.e. mainly between the 1980s and late 2000s. These "newer" shophouses reported the most severe content losses and can be explained by the strong urban pressure in these decades. The findings are confirmed by Downes & Storch (2014), Chinh et al. (2015) and Nguyen et al. (2016), who highlight that "newer" buildings in HCMC are more flood-exposed than "older" ones.
- The revenues from business operations (**mthly. sales**) are positively correlated with the degree of content loss in the respective BN graph, but only a weak positive correlation exists to relative interruption losses. Monthly sales are seen as an indicator for the microbusiness size and its type of business content. Higher sales have effect on both exposure and vulnerability, as sale volume reflects a heterogeneity among companies in type of contents (Schoppa et al. 2020). The variable of monthly sales has a negative correlation with the indicator for non-structural precautionary measures in the graph for degree of content loss. This is theoretically explained by the connections within the data: businesses with limited revenues are more likely to acquire non-structural measures before the flood event, as loss of contents would have existential consequences for small retailers compared to more prosperous businesses.
- More implemented non-structural precautionary measures (**non-structural measures**) reduce the severity (degree) of content losses in microbusinesses, though they are not relevant for modelling relative interruption losses. The impact of precautionary measures on reducing commercial content losses is well studied (Kreibich et al. 2007; 2010; Chinh et al. 2016; Sieg et al. 2017; Schoppa et al. 2020). Non-structural measures usually prevent water from infiltrating into the building, but not in all cases. For instance, Chinh et al. (2016) found that in Can Tho flood water can also come from the sewage system and thus bypass implemented precautionary measures. Consequently, there is no link with water depth in our model due to weak correlation between water depth and non-structural measures.
- The implementation of structural precautionary measures (**structural measures**) has mitigating effects on the severity of content and interruption losses in microbusinesses. The moderate dependencies in the BN graphs are in line with the findings of various studies, which highlight the usage of structural measures as an efficient individual precautionary measure (Scussolini et al. 2017; Trinh & Thanh 2017; Du et al. 2020; Harish et al. 2023). The efficiency of these measures is represented in the BN graphs indirectly by lower water levels in the shophouses and directly in the flood loss variables, e.g. in elevated buildings, there is less chance that flood water will enter the building through overloaded drainage systems.
- A higher number of employees (**no. employees**) is linked with lower interruption losses in the respective BN graph. Despite its rather weak negative rank correlation it improves the predictive accuracy of the BN model. The number of employees refers to the availability to human resources on which the business owner can draw on, which in turn affects the possibility to keep the business running during the flood event, e.g. by relocating important business processes. The findings of Sultana et al. (2018) confirm the number of employees as an important predictor of interruption losses in German companies, though they identify a positive association which is contrary to the findings of this study.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

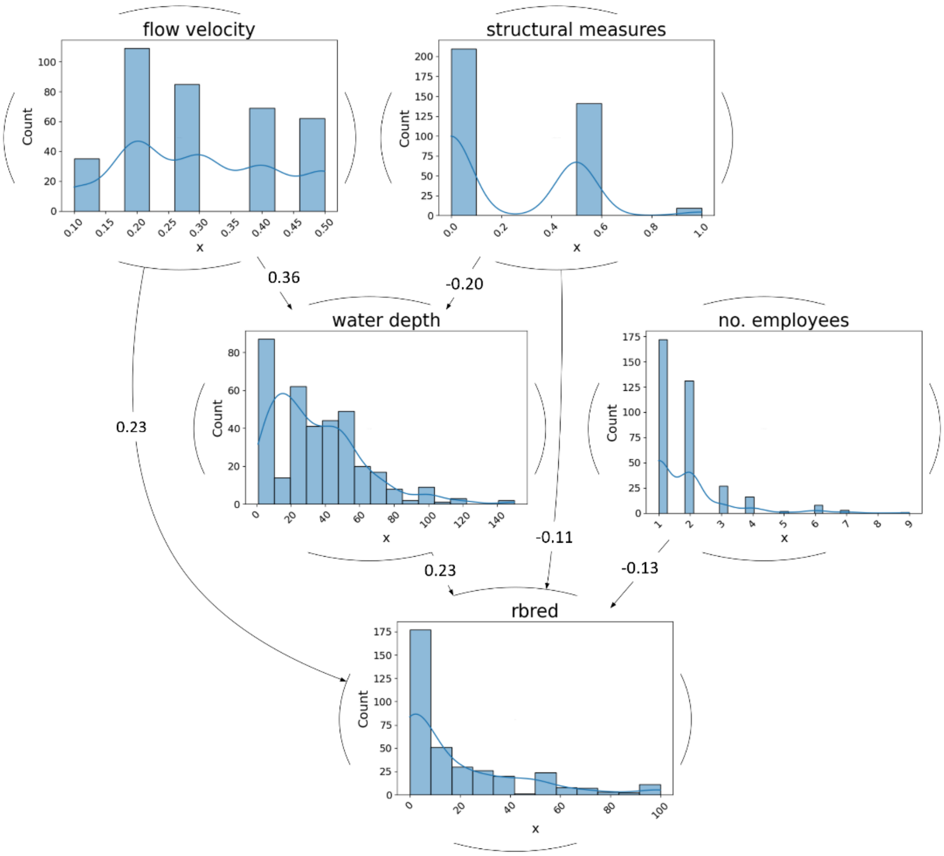

Figure 3: Structure of the Bayesian Network for predicting relative interruption losses (rbred).

4.3 Flood loss model validation

4.3.1 Relative content loss

At the first glance, the modelling approach consisting of logistic regression and Bayesian Network seems to perform quite well when predicting relative content losses (MAE of 3.8 %, RMSE of 12.3 %). It marginally underestimates losses (MBE: -2.4 %) and has a remarkable low SMAPE of 16.3 % indicating a good preciseness of the estimates. The mean value of the modelled relative content losses is of similar magnitude as the observed loss ratios (observed mean: 4.7 %, predicted mean: 4.6 %), as shown also by the clustering of the data points in the lower value range in Figure 4.a. However, the figure also shows that more severe losses to business content are consistently underestimated by the models.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

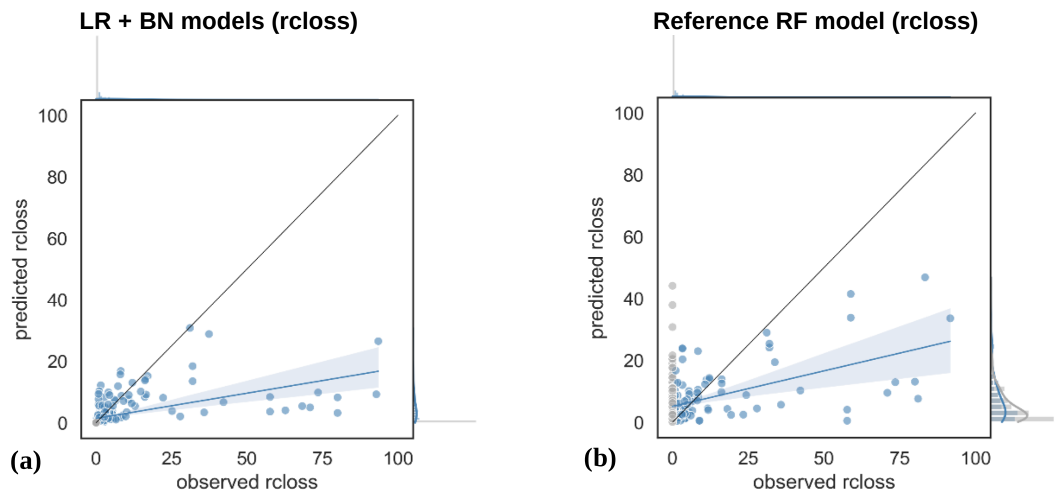

Figure 4: Scatterplots of observed and modelled relative content losses (rcloss) to HCMC's microbusinesses for (a) the combination of logistic regression (LR) and Bayesian Network (BN) and (b) the reference Random Forest (RF) model used for benchmarking. The grey points represent the observations of zero-loss. The ML-based classifiers assigned to most cases an absence of content loss (zero-loss), thus only one grey point seems to be visualised in Figure a.

The general well predictive performance of the modelling approach is caused by the usually low probability values for the chance of content loss. Having a critical look to the predicted probabilities of chance of content loss, it becomes clear that the observed small prediction bias is caused by the circumstance that the logistic regression estimated instances of chance of content loss usually as zero-loss cases, and thus assigns low probability of losses to most predictor combinations (see, centric, red histogram of Figure 5.a). The high share of observations of content loss wrongly predicted as zero-losses further illustrates this (see, False Positives in the lower left corner of Figure 5.b); only 25 % of the experienced content losses (loss cases) are correctly predicted by the ML-classifiers (see, True Negatives in the lower right corner of Figure 5.b).

As a consequence, the estimates for relative loss to business content are mostly reduced by more than half as soon as they are multiplied with the predicted probabilities for chance of loss. In particular, the estimates of severe cases of content loss are reduced in their magnitudes. Furthermore, the ML-based classifiers could hardly distinguish between cases with an absence of loss (zero-loss) and small loss fractions (near zero-loss), which further deteriorated the calibration and performance of the classifiers.

In comparison to the modelling approach, the reference Random Forest model (Chinh et al. 2017) less captures reported cases of zero-loss as such. This is shown when comparing the predicted values of zero-loss cases from the modelling approach (grey dots in Figure 4.a) with the ones from the reference RF model (grey dots in Figure 4.b). However, the general predictive performance is only marginally worse (Table 2). The cross-validated RF-estimators have on average similar magnitudes in the RMSE (12.4 %) and MBE (1.3 %) as the modelling approach, but higher MAE (7.2 %) and SMAPE (78.9 %).

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

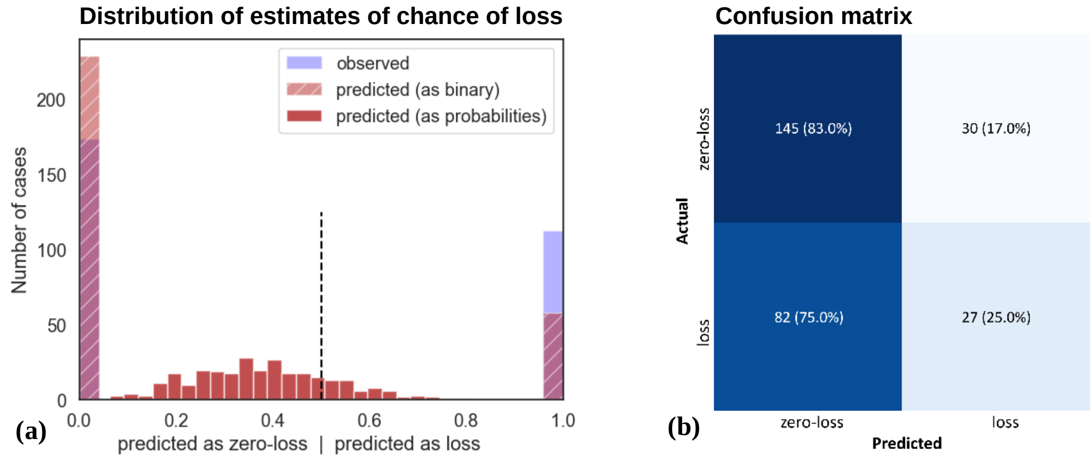

Figure 5: (a) Distribution of predicted probabilities of chance of content loss from the ML-based classifiers and (b) the corresponding confusion matrix for chance of content loss. The values in front of the brackets are the sample numbers; values in the brackets the sample numbers normalized over the observations.

4.3.2 Relative interruption loss

The cross-validation of the BN model for relative interruption losses results in an averaged MAE of 18.7 %, RMSE of 24.5 %, MBE of 0.17 % and SMAPE of 61.9 %. The modelled mean value in the interruption losses is almost equal to the observed mean of around 18.5 %, yet the variation in the observations is not well represented in the model estimates, as visualised in Figure 6.a. Nearly all reported cases of interruption loss are predicted by the BN with loss fractions between 10 % and 40 %. This is much narrower compared to the variation seen in the reported loss ratios ranging between 0 % to 100 % decrease in business revenues. Additionally, the figure shows that more severe cases of interruption loss are underestimated by the BN, despite their rather frequent occurrence.

The reference RF model results in similar high prediction errors as the BN (Table 2). They particularly overestimate cases of zero- and near zero-loss and underestimate severe loss cases (Figure 6.a & 6.b).

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

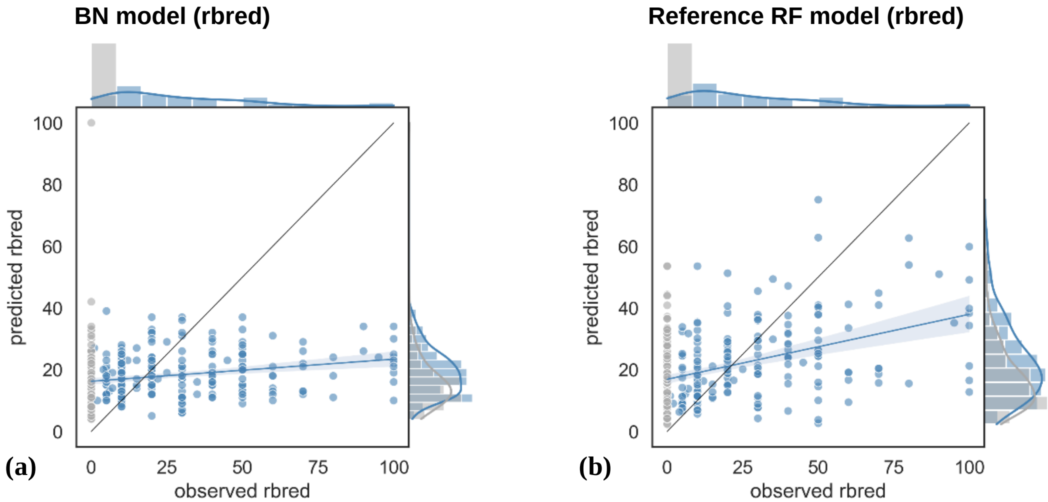

Figure 6: Satterplots for observed and modelled relative interruption losses (rbred) to HCMC's microbusinesses for (a) the Bayesian Network and (b) the reference Random Forest model used for benchmarking. The grey points represent the observations of zero-loss, i.e. the absence of interruption loss.

4.4 Transferability of the flood loss model

In this subsection the results of the transferability of the flood loss models are presented and discussed. The interruption loss model calibrated on microbusinesses in HCMC was applied to predict interruption losses in Can Tho using a comparable survey dataset. Cumulative distribution functions (CDFs) were chosen to visualise the main aspects of the model performances. The generalisation ability of the Bayesian Network model on the Can Tho samples results in similar prediction errors than during training on the HCMC samples, except for the SMAPE score. The transfer of the BN leads to a MAE of 17.9 %, RMSE of 23.5 %, MBE of 0.2 % and SMAPE of 23.2 %. The error scores show that the model's capacity to estimate interruption losses remains unchanged when transferred to Can Tho, in contrast to the transferability of the reference Random Forest model which resulted in a degraded performance (Table 2). The CDFs for the BN in Figure 7.a and for the reference RF model in Figure 7.b reflect these findings. The probability of the BN to predict a Can Tho sample precisely (prediction bias < ±10 %) remains unchanged (Figure 7.a) but drops for the reference RF model from around 45 % (HCMC samples) to 25 % (Can Tho samples) (Figure 7.b). In most cases the reference RF model underestimated the reported interruption losses, as shown when comparing the MBE scores between the transferred reference RF model and the transferred BN in Table 2.

The presented results of this subsection show that the reference RF model is less transferable than the BN, despite both models performing similarly well in their calibration site (i.e., HCMC).

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

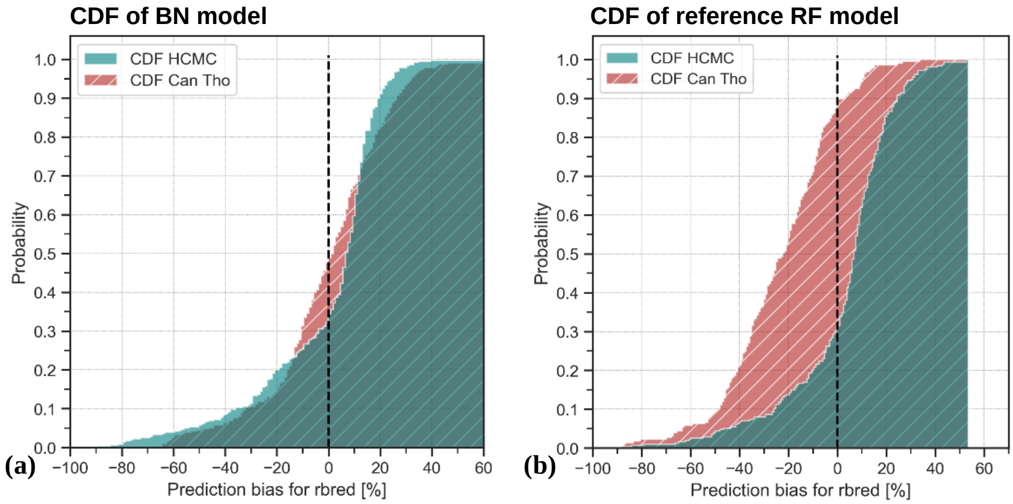

Figure 7: Cumulative distribution function (CDF, normalized) of prediction errors for modelling business interruption losses (rbred) in HCMC and in the transfer region, Can Tho. (a) CDF of the Bayesian Network performances; (b) CDF of the reference Random Forest model performance. The CDF for the reference RF model is cut by 55 % as no larger prediction errors exist.

Transfer experiments on (Bayesian Network) flood loss models have shown that model transfer usually leads to a stagnation or drop in the model's performance, in particular, when the new conditions differ remarkably from the calibration region (Schröter et al. 2014; Wagenaar et al. 2018). However, a drop in model performance when regional transferred could not be observed in this study due to very similar local conditions between the calibration (HCMC) and application site (Can Tho). These local conditions are reflected in the similar predictor ranges and distributions of both survey datasets (see, Supplementary Information Figure S1). Additionally, the high heterogeneity in the HCMC samples, in particular in the hydrological, building- and business-related predictors, has the potential to increase the model robustness for new study sites (Wagenaar et al. 2018).

Table 2: Model validation of flood loss models in HCMC and in the transfer region (Can Tho). The different sample sizes are due to the differences in the number of cases reported and the way in which incomplete samples are treated in the models. rcloss: relative loss to business content, rbred: relative loss due to business interruption, LR: probabilistic logistic regression, BN: Bayesian Network, RF: reference Random Forest, x: not applicable

	Model validation [sample size]	MAE [%]	RMSE [%]	MBE [%]	SMAPE [%]
HCMC					
rcloss	LR + BN [284]	**3.8**	12.3	-2.4	16.3
	RF [284]	7.2	12.4	1.3	78.9
rbred	BN [360]	18.7	24.5	0.17	61.9
	RF [314]	**16.4**	21.8	1.7	58.6
Can Tho (transfer region)					
rcloss	LR + BN [266]	x	x	x	x
	RF [266]	**13.5**	19.6	0.8	75.0
rbred	BN [313]	**17.9**	23.5	0.2	23.2
	RF [267]	25.7	32.6	-23.5	41.1

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

4.5 Limitations and uncertainties

The results have indicated high uncertainty in reconstructing flood losses from survey data. One possible further analysis would be comparing the model with other studies. However, comparability is limited by the fact that the flood losses were determined at object-level, while flood loss modelling in low- and middle-income countries is mainly carried out on meso- or macroscale (Booij 2004; Aerts et al. 2020; Tierolf et al. 2021), with commercial losses reported only in absolute values (Wijayanti et al. 2017; Patankar 2019; Tsinda et al. 2019) and often without validation (Ke et al. 2012; Patankar & Patwardhan 2015; Yang et al. 2016).

The regional transfer of the BN did not affect the width in model uncertainties. The mean values of the empirical flood losses are for both response variables within the uncertainty ranges (within 95 % confidence interval), independent from the region of application. However, as seen before the majority of interruption-related losses are underestimated by the flood loss models remarkably. The example of regional transfer illustrates the potential of non-parametric, continuous Bayesian Network models compared to a Random Forest model. However, since the transfer capability was validated for only one case study, the models presented here need to be calibrated under further local and temporal conditions to truly estimate specific flood impacts on microbusinesses in new regions.

5. Conclusions

We proposed a first approach to estimate flood losses to microbusinesses by combining expert knowledge with survey data of flood-affected microbusinesses from HCMC and Can Tho in Vietnam. A Conditional Random Forest model was applied to obtain the main drivers of content and interruption losses from a set heterogeneous samples and potential predictors which are partly correlated to each other. The identified drivers were used to calibrate knowledge-based probabilistic loss models consisting of non-parametric, continuous Bayesian Networks and logistic regression. The findings of this study indicate that information on business revenues from monthly sales and production, building age, and hydrological characteristics of the flood is crucial in estimating content and interruption losses for microbusinesses. The probabilistic flood loss models performed in the calibration region (HCMC) with a MAE of 3.8 % for relative content losses and 18.7 % for relative interruption losses. The interruption model was transferred to another city – Can Tho. The developed loss models are openly provided and integrating them to flood risk assessments will advance risk management decision making with a focus on microbusinesses.

Data and Code Availability. The survey data will be made openly available in the in the HOWAS21 database (https://howas21.gfz-potsdam.de/) after an embargo of three years after the end of the project (in 2027). The data can be accessed in the meantime from the authors. Source code (python) is openly available at https://github.com/A-Buch/flood-loss-

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

models-4-HCMC/tree/microbusiness-paper. The BN flood loss models are created with the PyBanshee toolbox (Koot et al. 2023, https://github.com/mike-mendoza/py_banshee.git); the Conditional Random Forest model is based on the R package partykit (Hothorn & Zeileis 2015, https://www.jmlr.org/papers/v16/hothorn15a.html).

Competing interests. Some authors are members of the editorial board of NHESS.

Funding. The data collection was undertaken by project DECIDER (grant nos. 01LZ1703G and 01LZ1703A) and WISDOM II ("Water related Information System for a Sustainable development of the Mekong Delta") funded by the BMBF (German Ministry for Education and Research). Nivedita Sairam is funded by project HI-CliF (grant no. 01LN2209A) by the BMBF.

References

Aerts, J. P. M., Uhlemann-Elmer, S., Eilander, D., and Ward, P. J.: Comparison of estimates of global flood models for flood hazard and exposed gross domestic product: a China case study, Natural Hazards and Earth System Sciences, 20, 3245–3260, https://doi.org/10.5194/nhess-20-3245-2020, 2020.

Amadio, M., Scorzini, A. R., Carisi, F., Essenfelder, A. H., Domeneghetti, A., Mysiak, J., and Castellarin, A.: Testing empirical and synthetic flood damage models: the case of Italy, Natural Hazards and Earth System Sciences, 19, 661–678, https://doi.org/10.5194/nhess-19-661-2019, 2019.

Apel, H., Aronica, G. T., Kreibich, H., and Thieken, A. H.: Flood risk analyses—how detailed do we need to be?, Nat Hazards, 49, 79–98, https://doi.org/10.1007/s11069-008-9277-8, 2009.

Booij, M. J.: Flood Damage Assessment and Modelling in the Red River basin in Vietnam, International Workshop on Flood Controls Decision Support Systems (FLOCODS), Hanoi, Vietnam, available at: https://www.utwente.nl/en/et/cem/research/wem/people-attachments/booij/publications/44.pdf, 2004.

Bouwer, L. M.: Have Disaster Losses Increased Due to Anthropogenic Climate Change?, Bull. Amer. Meteor. Soc., 92, 39–46, https://doi.org/10.1175/2010BAMS3092.1, 2011.

Brill, F. A.: Applications of machine learning and open geospatial data in flood risk modelling, PhD Thesis, University Potsdam, https://doi.org/10.25932/PUBLISHUP-55594, 2022.

Chicco, D., Warrens, M. J., and Jurman, G.: The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Computer Science, 7, https://doi.org/10.7717/peerj-cs.623, 2021.

Chinh, D. T., Gain, A., Dung, N. V., Haase, D., and Kreibich, H.: Multi-Variate Analyses of Flood Loss in Can Tho City, Mekong Delta, Water, 8, 6, https://doi.org/10.3390/w8010006, 2015.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

Chinh, D. T., Bubeck, P., Dung, N. V., and Kreibich, H.: The 2011 flood event in the Mekong Delta: preparedness, response, damage and recovery of private households and small businesses, Disasters, 40, 753–778, https://doi.org/10.1111/disa.12171, 2016.

Chinh, D. T., Dung, N. V., Gain, A., and Kreibich, H.: Flood Loss Models and Risk Analysis for Private Households in Can Tho City, Vietnam, Water, 9, 313, https://doi.org/10.3390/w9050313, 2017.

Debeer, D. and Strobl, C.: Conditional permutation importance revisited, BMC Bioinformatics, 21, 307, https://doi.org/10.1186/s12859-020-03622-2, 2020.

Downes, N. K., Storch, H., Schmidt, M., Nguyen, T. C. V., Dinh, L. C., Tran, T. N., and Hoa, L. T.: Understanding Ho Chi Minh City's Urban Structures for Urban Land-Use Monitoring and Risk-Adapted Land-Use Planning, in: Sustainable Ho Chi Minh City: Climate Policies for Emerging Mega Cities, edited by: Katzschner, A., Waibel, M., Schwede, D., Katzschner, L., Schmidt, M., and Storch, H., Springer International Publishing, Cham, 89–116, https://doi.org/10.1007/978-3-319-04615-0_6, 2016.

Du, K.-L. and Swamy, M. N. S.: Neural Networks and Statistical Learning, Springer London, London, https://doi.org/10.1007/978-1-4471-7452-3, 2019.

Du, S., Scussolini, P., Ward, P. J., Zhang, M., Wen, J., Wang, L., Koks, E., Diaz-Loaiza, A., Gao, J., Ke, Q., and Aerts, J. C. J. H.: Hard or soft flood adaptation? Advantages of a hybrid strategy for Shanghai, Global Environmental Change, 61, https://doi.org/10.1016/j.gloenvcha.2020.102037, 2020.

Formetta, G. and Feyen, L.: Empirical evidence of declining global vulnerability to climate-related hazards, Global Environmental Change, 57, 2019, https://doi.org/10.1016/j.gloenvcha.2019.05.004, 2019.

Garschagen, M.: Risky Change? Vietnam's Urban Flood Risk Governance between Climate Dynamics and Transformation, pac aff, 88, 599–621, https://doi.org/10.5509/2015883599, 2015.

General Statistics Office: Results of the 2017 economic census, Statistical Publishing House, available at: https://www.gso.gov.vn/en/data-and-statistics/2019/03/results-of-the-2017-economic-census/, 2018.

Gerl, T., Bochow, M., and Kreibich, H.: Flood Damage Modeling on the Basis of Urban Structure Mapping Using High-Resolution Remote Sensing Data, Water, 6, 2367–2393, https://doi.org/10.3390/w6082367, 2014.

Güneralp, B., Güneralp, I., and Liu, Y.: Changing global patterns of urban exposure to flood and drought hazards, Global Environmental Change, 31, 217–225, https://doi.org/10.1016/j.gloenvcha.2015.01.002, 2015.

Hallegatte, S.: An Adaptive Regional Input-Output Model and its Application to the Assessment of the Economic Cost of Katrina, Risk Analysis, 28, 779–799, https://doi.org/10.1111/j.1539-6924.2008.01046.x, 2008.

Hallegatte, S., Green, C., Nicholls, R. J., and Corfee-Morlot, J.: Future flood losses in major coastal cities, Nature Clim Change, 3, 802–806, https://doi.org/10.1038/nclimate1979, 2013.

Hanea, A. M., Kurowicka, D., and Cooke, D.: Hybrid Method for Quantifying and Analyzing Bayesian Belief Nets, Quality & Reliability Eng, 22, 709–729, https://doi.org/10.1002/qre.808, 2006.

Hanea, A. M., Morales Napoles, O., and Ababei, D.: Non-parametric Bayesian networks: Improving theory and reviewing applications, Reliability Engineering & System Safety, 144, 265–284, https://doi.org/10.1016/j.ress.2015.07.027, 2015.

Harish, T. V., Sairam, N., Yang, L. E., Garschagen, M., and Kreibich, H.: Identifying the drivers of private flood precautionary measures in Ho Chi Minh City, Vietnam, Natural Hazards and Earth System Sciences, 23, 1125–1138, https://doi.org/10.5194/nhess-23-1125-2023, 2023.

Hothorn, T. and Zeileis, A.: partykit: A Modular Toolkit for Recursive Partytioning in R, Journal of Machine Learning Research, 16, 3905–3909, 2015.

Hothorn, T., Hornik, K., and Zeileis, A.: Unbiased Recursive Partitioning: A Conditional Inference Framework, Journal of Computational and Graphical Statistics, 15, 651–674, https://doi.org/10.1198/106186006X133933, 2006.

JETRO: Policies supporting SMEs – experience from Japan, Workshop on Policies Supporting SMEs – Experience from Japan, Hanoi, Vietnam, 2017.

Jha, A. K., Bloch, R., and Lamond, J.: Cities and Flooding: A Guide to Integrated Urban Flood Risk Management for the 21st Century, The World Bank, https://doi.org/10.1596/978-0-8213-8866-2, 2012.

Ke, Q., Jonkman, S. N., Van Gelder, P. H. A. J. M., and Rijcken, T.: Flood damage estimation for downtown Shanghai – sensitivity analysis, in: Conference of the international society for integrated disaster risk management, Conference of the international society for integrated disaster risk management, Beijing, China, 260–273, available at: http://resolver.tudelft.nl/uuid:bf75cdab-8a0d-4dbf-ae8e-61f59d3e5d86, 2012.

Koks, E. E. and Thissen, M.: A Multiregional Impact Assessment Model for disaster analysis, Economic Systems Research, 28, 429–449, https://doi.org/10.1080/09535314.2016.1232701, 2016.

Koot, P., Mendoza-Lugo, M. A., Paprotny, D., Morales-Nápoles, O., Ragno, E., and Worm, D. T. H.: PyBanshee version (1.0): A Python implementation of the MATLAB toolbox BANSHEE for Non-Parametric Bayesian Networks with updated features, SoftwareX, 21, https://doi.org/10.1016/j.softx.2022.101279, 2023.

KPMG: Preparing MSMEs for effective disaster management, KPMG, available at: https://assets.kpmg.com/content/dam/kpmg/pdf/2016/05/Disaster-Management-Preparedness-SME.pdf, 2016.

Kreibich, H., Müller, M., Thieken, A. H., and Merz, B.: Flood precaution of companies and their ability to cope with the flood in August 2002 in Saxony, Germany, Water Resour. Res., 43(3), https://doi.org/10.1029/2005WR004691, 2007.

Kreibich, H., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., Merz, B., and Thieken, A. H.: Is flow velocity a significant parameter in flood damage modelling?, Nat. Hazards Earth Syst. Sci., 9, 1679–1692, https://doi.org/10.5194/nhess-9-1679-2009, 2009.

Kreibich, H., Seifert, I., Merz, B., and Thieken, A. H.: Development of FLEMOcs – a new model for the estimation of flood losses in the commercial sector, Hydrological Sciences Journal, 55, 1302–1314, https://doi.org/10.1080/02626667.2010.529815, 2010.

Kreibich, H., Van Loon, A. F., Schröter, K., Ward, P. J., Mazzoleni, M., Sairam, N., Abeshu, G. W., Agafonova, S., AghaKouchak, A., Aksoy, H., Alvarez-Garreton, C., Aznar, B., Balkhi, L., Barendrecht, M. H., Biancamaria, S., Bos-Burgering, L., Bradley, C., Budiyono, Y., Buytaert, W., Capewell, L., Carlson, H., Cavus, Y., Couasnon, A., Coxon, G., Daliakopoulos, I., De Ruiter, M. C., Delus, C., Erfurt, M., Esposito, G., François, D., Frappart, F., Freer, J., Frolova, N., Gain, A. K., Grillakis, M., Grima, J. O., Guzmán, D. A., Huning, L. S., Ionita, M., Kharlamov, M., Khoi, D. N., Kieboom, N., Kireeva, M., Koutroulis, A., Lavado-Casimiro, W., Li, H.-Y., LLasat, M. C., Macdonald, D., Mård, J., Mathew-Richards, H., McKenzie, A., Mejia, A., Mendiondo, E. M., Mens, M., Mobini, S., Mohor, G. S., Nagavciuc, V., Ngo-Duc, T., Thao Nguyen Huynh, T., Nhi, P. T. T., Petrucci, O., Nguyen, H. Q., Quintana-Seguí, P., Razavi, S., Ridolfi, E., Riegel, J., Sadik, M. S., Savelli, E., Sazonov, A., Sharma, S., Sörensen, J., Arguello Souza, F. A., Stahl, K., Steinhausen, M., Stoelzle, M., Szalińska, W., Tang, Q., Tian, F., Tokarczyk, T., Tovar, C., Tran, T. V. T., Van Huijgevoort, M. H. J., Van Vliet, M. T. H., Vorogushyn, S., Wagener, T., Wang, Y., Wendt, D. E., Wickham, E., Yang, L., Zambrano-Bigiarini, M., Blöschl, G., and Di Baldassarre, G.: The challenge of unprecedented floods and droughts in risk management, Nature, 608, 80–86, https://doi.org/10.1038/s41586-022-04917-5, 2022.

Krstajic, D., Buturovic, L. J., Leahy, D. E., and Thomas, S.: Cross-validation pitfalls when selecting and assessing regression and classification models, J Cheminform, 6, 10, https://doi.org/10.1186/1758-2946-6-10, 2014.

Kurowicka, D. and Cooke, R.: Uncertainty Analysis with High Dimensional Dependence Modelling, 1st ed., Wiley, https://doi.org/10.1002/0470863072, 2006.

Leitold, R. and Revilla Diez, J.: Exposure of manufacturing firms to future sea level rise in Ho Chi Minh City, Vietnam, Journal of Maps, 15, 13–20, https://doi.org/10.1080/17445647.2018.1548385, 2019.

Leitold, R., Revilla Diez, J., and Tran, V.: Are we expecting too much from the private sector in flood adaptation? Scenario-based field experiments with small- and medium-sized firms in Ho Chi Minh City, Vietnam, Climatic Change, 163, 359–378, https://doi.org/10.1007/s10584-020-02888-y, 2020.

Leitold, R., Garschagen, M., Tran, V., and Revilla Diez, J.: Flood risk reduction and climate change adaptation of manufacturing firms: Global knowledge gaps and lessons from Ho Chi Minh City, International Journal of Disaster Risk Reduction, 61, https://doi.org/10.1016/j.ijdrr.2021.102351, 2021.

Levshina, N.: Conditional Inference Trees and Random Forests, in: A Practical Handbook of Corpus Linguistics, edited by: Paquot, M. and Gries, S. Th., Springer International Publishing, Cham, 611–643, https://doi.org/10.1007/978-3-030-46216-1_25, 2020.

Liu, X., Kounadi, O., and Zurita-Milla, R.: Incorporating Spatial Autocorrelation in Machine Learning Models Using Spatial Lag and Eigenvector Spatial Filtering Features, IJGI, 11(4), https://doi.org/10.3390/ijgi11040242, 2022.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A.: Review article "Assessment of economic flood damage," Nat. Hazards Earth Syst. Sci., 10, 1697–1724, https://doi.org/10.5194/nhess-10-1697-2010, 2010.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., Van Den Bergh, J. C. J. M., Bouwer, L. M., Bubeck, P., Ciavola, P., Genovese, E., Green, C., Hallegatte, S., Kreibich, H., Lequeux, Q., Logar, I., Papyrakis, E., Pfurtscheller, C., Poussin, J., Przyluski, V., Thieken, A. H., and Viavattene, C.: Review article: Assessing the costs of natural hazards – state of the art and knowledge gaps, Nat. Hazards Earth Syst. Sci., 13, 1351–1373, https://doi.org/10.5194/nhess-13-1351-2013, 2013.

Nguyen, T. B., Samsura, D. A. A., Van Der Krabben, E., and Le, A.-D.: Saigon-Ho Chi Minh City, Cities, 50, 16–27, https://doi.org/10.1016/j.cities.2015.08.007, 2016.

Ootegem, L. V., Herck, K. V., Creten, T., Verhofstadt, E., Foresti, L., Goudenhoofdt, E., Reyniers, M., Delobbe, L., Tuyls, D. M., and Willems, P.: Exploring the potential of multivariate depth-damage and rainfall-damage models, Journal of Flood Risk Management, 11, S916–S929, https://doi.org/10.1111/jfr3.12284, 2017.

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Terefenko, P., and Schröter, K.: Estimating exposure of residential assets to natural hazards in Europe using open data, Nat. Hazards Earth Syst. Sci., 20, 323–343, https://doi.org/10.5194/nhess-20-323-2020, 2020.

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Wagenaar, D., Castellarin, A., Carisi, F., Bertin, X., Merz, B., and Schröter, K.: A probabilistic approach to estimating residential losses from different flood types, Nat Hazards, 105, 2569–2601, https://doi.org/10.1007/s11069-020-04413-x, 2021.

Patankar, A.: Impacts of Natural Disasters on Households and Small Businesses in India, SSRN Journal, https://doi.org/10.2139/ssrn.3590902, 2019.

Patankar, A. and Patwardhan, A.: Estimating the uninsured losses due to extreme weather events and implications for informal sector vulnerability: a case study of Mumbai, India, Nat Hazards, 80, 285–310, https://doi.org/10.1007/s11069-015-1968-3, 2015.

Rafiezadeh Shahi, K., Sairam, N., Schoppa, L., Sang, L. T., Tan, D. L. H., and Kreibich, H.: BN-FLEMOΔ: A Bayesian Network-based Flood Loss Estimation Model for Adaptation Planning in Ho Chi Minh City, Vietnam, ESS Open Archive [preprint], https://doi.org/10.22541/essoar.172081523.38063336/v1, 12 July 2024.

Rand, J. and Tarp, F. (Eds.): Micro, Small, and Medium Enterprises in Vietnam, 1st ed., Oxford University Press, Oxford, https://doi.org/10.1093/oso/9780198851189.003.0001, 2020.

Rentschler, J., Salhab, M., and Jafino, B. A.: Flood exposure and poverty in 188 countries, Nat Commun, 13, 3527, https://doi.org/10.1038/s41467-022-30727-4, 2022.

Rözer, V., Kreibich, H., Schröter, K., Müller, M., Sairam, N., Doss-Gollin, J., Lall, U., and Merz, B.: Probabilistic Models Significantly Reduce Uncertainty in Hurricane Harvey Pluvial Flood Loss Estimates, Earth's Future, 7, 384–394, https://doi.org/10.1029/2018EF001074, 2019.

Samantha, G.: The Impact of Natural Disasters on Micro, Small and Medium Enterprises (MSMEs): A Case Study on 2016 Flood Event in Western Sri Lanka, Procedia Engineering, 212, 744–751, https://doi.org/10.1016/j.proeng.2018.01.096, 2018.

https://doi.org/10.5194/egusphere-2024-2340
Preprint. Discussion started: 19 August 2024

Schoppa, L., Sieg, T., Vogel, K., Zöller, G., and Kreibich, H.: Probabilistic Flood Loss Models for Companies, Water Resources Research, 56, https://doi.org/10.1029/2020wr027649, 2020.

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., and Merz, B.: How useful are complex flood damage models?, Water Resources Research, 50, 3378–3395, https://doi.org/10.1002/2013wr014396, 2014.

Scussolini, P., Tran, T. V. T., Koks, E., Diaz-Loaiza, A., Ho, P. L., and Lasage, R.: Adaptation to Sea Level Rise: A Multidisciplinary Analysis for Ho Chi Minh City, Vietnam, Water Resources Research, 53, 10841–10857, https://doi.org/10.1002/2017wr021344, 2017.

Sieg, T., Vogel, K., Merz, B., and Kreibich, H.: Tree-based flood damage modeling of companies: Damage processes and model performance, Water Resources Research, 53, 6050–6068, https://doi.org/10.1002/2017wr020784, 2017.

Sieg, T., Schinko, T., Vogel, K., Mechler, R., Merz, B., and Kreibich, H.: Integrated assessment of short-term direct and indirect economic flood impacts including uncertainty quantification, PLOS ONE, 14, https://doi.org/10.1371/journal.pone.0212932, 2019.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T.: Bias in random forest variable importance measures: Illustrations, sources and a solution, BMC Bioinformatics, 8, 25, https://doi.org/10.1186/1471-2105-8-25, 2007.

Sultana, Z., Sieg, T., Kellermann, P., Müller, M., and Kreibich, H.: Assessment of Business Interruption of Flood-Affected Companies Using Random Forests, Water, 10, 1049, https://doi.org/10.3390/w10081049, 2018.

Thuy, N. B., Tien, T. Q., Wettre, C., and Hole, L. R.: Monsoon-Induced Surge during High Tides at the Southeast Coast of Vietnam: A Numerical Modeling Study, Geosciences, 9, 72, https://doi.org/10.3390/geosciences9020072, 2019.

Tierolf, L., De Moel, H., and Van Vliet, J.: Modeling urban development and its exposure to river flood risk in Southeast Asia, Computers, Environment and Urban Systems, 87, https://doi.org/10.1016/j.compenvurbsys.2021.101620, 2021.

Tran, T. N.: Improvement of flood risk assessment under climate change in Ho Chi Minh City with GIS applications, PhD Thesis, University of Technology in Cottbus - Senftenberg, 186 pp., available at: https://opus4.kobv.de/opus4-btu/files/3078/Tran_Thong_Nhat_Thesis_2014.pdf, 2014.

Trinh, P. T. T. and Thanh, N. D.: Development Characteristics of SME Sector in Vietnam: Evidence from the Vietnam Enterprise Census 2006-2015 (Working Paper WP-18), VEPR [Viet Nam Institute for Economic and Policy Research, supported by the Friedrich Naumann Foundation for Freedom], Hanoi, Vietnam, available at: http://vepr.org.vn/upload/533/20171222/EN_VEPR%20WP%2018.pdf, 2017.

Tsinda, A., Kind, C., Hess, J. S., Mugsiha, R., and Bizoza, A. R.: Estimating damage costs of flooding on small- and medium-sized enterprises in Kigali, Rwanda, Jàmbá Journal of Disaster Risk Studies, 11, https://doi.org/10.4102/jamba.v11i1.755, 2019.

Vogel, K., Riggelsen, C., Scherbaum, F., Schröter, K., Kreibich, H., and Merz, B.: Challenges for Bayesian network learning in a flood damage assessment application, in: Safety, Reliability, Risk and Life-Cycle Performance of Structures and Infrastructures, CRC Press, 3123–3130, https://doi.org/10.1201/b16387-452, 2014.

Vogel, K., Weise, L., Schröter, K., and Thieken, A. H.: Identifying Driving Factors in Flood-Damaging Processes Using Graphical Models, Water Resources Research, 54, 8864–8889, https://doi.org/10.1029/2018wr022858, 2018.

Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L. M., and Kreibich, H.: Regional and Temporal Transferability of Multivariable Flood Damage Models, Water Resources Research, 54, 3688–3703, https://doi.org/10.1029/2017wr022233, 2018.

Wijayanti, P., Zhu, X., Hellegers, P., Budiyono, Y., and Van Ierland, E. C.: Estimation of river flood damages in Jakarta, Indonesia, Nat Hazards, 86, 1059–1079, https://doi.org/10.1007/s11069-016-2730-1, 2017.

Yang, L., Kajitani, Y., Tatano, H., and Jiang, X.: A methodology for estimating business interruption loss caused by flood disasters: insights from business surveys after Tokai Heavy Rain in Japan, Nat Hazards, 84, 411–430, https://doi.org/10.1007/s11069-016-2534-3, 2016.