https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

Deciphering the drivers of direct and indirect damages to companies from an unprecedented flood event: A data-driven, multivariate probabilistic approach

Ravi Kumar Guntu[1]*, Guilherme Samprogna Mohor[2], Annegret H. Thieken2, Meike Müller[3], and Heidi Kreibich[1]

[1]GFZ Helmholtz Centre for Geosciences, Section Hydrology, 14473 Potsdam, Germany

[2]Institute of Environmental Science and Geography, University of Potsdam, 14476 Potsdam, Germany

[3]Deutsche Rückversicherung AG, 40549 Düsseldorf, Germany

*Correspondence to: Ravi Kumar Guntu (ravikumar.guntu@gfz.de)

Abstract. Floods are among the most destructive natural hazards, causing extensive damage to companies through direct impacts on assets and prolonged business interruptions. The July 2021 flood in Germany caused unprecedented damages, particularly in North Rhine-Westphalia and Rhineland-Palatinate, affecting companies of all sizes. To date, no study has examined the factors influencing company damages during such an extreme event. This study addresses this gap using survey data from 431 companies affected by the July 2021 flood. Results show that 62% of companies incurred direct damages exceeding €100,000. Machine learning models and Bayesian network analyses identify water depth and flow velocity as the primary drivers of both direct damage and business interruption. However, company characteristics (e.g., premises size, number of employees) and preparedness also play critical roles. Companies that implemented precautionary measures experienced significantly shorter business interruption durations—up to 58% for water depths below 1 m and 44% for depths above 2 m. These findings offer important insights for policy development and risk-informed decision-making. Incorporation of behavioral indicators into flood risk management strategies and improving early warning systems could significantly enhance business preparedness.

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

1 Introduction

Understanding the damage processes of companies during unprecedented floods is essential to increase their resilience and avoid catastrophic economic disruption. Unprecedented floods are particularly destructive, as management measures often fail during events of a magnitude not experienced before by locals (Kreibich et al., 2022). In Europe, unprecedented flash floods caused €14.36 billion of damage in Spain in October 2024 (Munich Re, 2025) and about €33 billion of damage in Germany in July 2021 (Munich Re, 2022), additionally, such unprecedented floods are expected to become more frequent with increasing climate change (Blöschl et al., 2017; Hirabayashi et al., 2013; Merz et al., 2021).

Damages to companies constitutes a significant portion of the total flood loss (Schoppa et al., 2020). Direct damages arise from the immediate physical contact of the flood water with assets, such as damage to buildings, equipment, goods, and stock. The June 2013 flood in Germany revealed that 32.4% of the total damage in Bavaria and 13.9% of the total damage in Saxony were attributed to companies, respectively (Thieken et al., 2016). On the other hand, indirect damages stem from disruptions caused by the flooding, such as business interruptions and restrictions (Jongman et al., 2012). In surveys conducted after the floods, 88% of affected companies reported that they had been significantly affected by business interruptions (Thieken et al., 2016). The severity of indirect damages can be equally significant, often reaching a magnitude comparable to direct damages for low probability events (Koks et al., 2015; Pfurtscheller and Vetter, 2015; Sieg et al., 2019). For instance, Pfurtscheller and Vetter (2015) reported that indirect damages are often underestimated by companies, despite sometimes exceeding direct damages during rare flood events. Li et al., (2018) employed an Input-Output (IO) model to evaluate indirect economic losses among manufacturing firms in Shanghai, capturing how business interruptions and operational restrictions affected upstream and downstream sectors through interlinked value chains. Under extreme storm flood scenarios, the estimated indirect damage could reach up to \$60 billion. Sieg et al., (2019) used a supply-side IO model and identified the manufacturing, and financial sectors as particularly vulnerable to indirect damages. Koks et al., (2015) revealed that for rare, low-probability but high-impact events, indirect damages often surpass direct damages. Altogether, these studies underscore the relevance in deciphering the processes contributing to the direct and indirect damages to companies.

The process of understanding flood damage to companies is complex due to their heterogeneous nature and is influenced by several factors. Kreibich et al., (2010) examine factors such as water depth, sector, company size, precautionary measures, and contamination to assess direct flood damage. While the study provides valuable insights, it acknowledges that the impact of precautionary measures and contamination on flood damages is not fully understood. Seifert et al., (2010) estimated direct flood damage at the mesoscale and highlighted the need for a deeper understanding of damage processes in high water depth scenarios. Nafari et al., (2016), focusing on Australian commercial structures, demonstrated that considering building characteristics in addition to water depth led to improved model performance, with lower bias and mean absolute error. Schoppa et al., (2020) analyzed comprehensive survey datasets collected after major flood events between 2002 and 2013 in the Danube, Elbe, Oder, and Rhine catchments. Their study identified water depth and precautionary measures as primary

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

factors for building damage, while damage to equipment, goods, and stock was strongly influenced by company characteristics such as sector, size, and precautionary measures. Schoppa et al., (2022) developed a socio-hydrological model using water depth and precautionary measures for estimating building damage. The study revealed that companies in Dresden, Germany reduced vulnerability through the implementation of precautionary measures. Significant progress has been made in identifying the variables that have explanatory power in estimating direct damages to companies. While both the 2002 and 2021 floods in Germany were considered unprecedented in different ways, the 2021 event stands out due to its exceptional event magnitude, rapid onset, and high death toll (Rhein and Kreibich, 2025; Thieken et al., 2023a). Given its rare nature and distinct damage dynamics, this study focuses specifically on the 2021 flood event to better understand the factors contributing to direct damages.

Deciphering the factors of indirect damages, such as business interruption and business restriction, is also crucial for mitigating their contribution to the overall economic consequences. Yang et al., (2016) modeled business interruption losses using water depth data collected from business surveys conducted after the Tokai Heavy Rain in Japan. The model showed a better fit for reported losses at lower inundation levels, but it overestimated losses in areas with deeper inundation. Sultana et al., (2018) highlighted that company-specific attributes, such as the number of employees and emergency measures, often play a more critical role in estimating business interruption costs than water depth. Endendijk et al., (2024) investigated the relationship between flood characteristics and business interruptions using post-disaster survey data from the 2021 flood in the Netherlands. They identified water depth, delayed compensation, and regional connectivity as critical factors affecting business interruption duration, while building-level mitigation measures were found to have limited influence, highlighting an area for further exploration. The study by Kabirzad et al., (2024) found that proximity to the river and the profitability of business premises were significant factors contributing to indirect flood damages to company buildings in Peninsular Malaysia. Sakai and Yao, (2023) underscore the vulnerability of small companies, which suffer disproportionately higher damages relative to turnover compared to larger companies. Business interruption, largely driven by temporary closures and reduced sales, is identified as the most significant damage across sectors. Despite these advancements, a significant research gap persists in understanding the factors influencing indirect damages during unprecedented flood events.

Adaptation to flood risk encompasses a range of measures aimed at reducing vulnerability and exposure to flood impacts. These can be broadly categorized into short-term emergency responses, such as evacuation or temporary protection, and long-term precautionary strategies, including elevating buildings or relocating critical infrastructure (Neise and Revilla Diez, 2019). While emergency measures require a degree of preparedness, they are reactive and distinct from long-term adaptation strategies (Wutzler et al., 2022). Understanding the effectiveness of adaptation behaviors during unprecedented events is crucial in determining whether these measures can mitigate damages or fail. Kreibich et al., (2007) noted that the effectiveness of such measures depends on factors like prior flood experience, emergency plans, and early warning systems. Jehmlich et al., (2020) further investigated the drivers behind flood-adaptive behavior and reported that firsthand flood experience increases the likelihood of companies adopting precautionary measures. However, the lack of property ownership can hinder property-level adaptation, as companies are less inclined to invest in resilience measures for rented properties. In

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

fact, Hudson et al., (2022) found only little difference between the adaptation of small and medium companies on the one hand and private households on the other hand. Leitold et al., (2021) examined adaptation strategies and found that the manufacturing sector tends to adopt reactive or temporary measures rather than long-term, proactive strategies. Wutzler et al., (2022) identified perceived low self-efficacy as a barrier to proactive adaptation. The study also noted that property ownership plays a significant role, with property owners more likely to adopt adaptive measures than tenants. Companies with extensive flood experience and low response costs are more likely to engage in proactive adaptation. Furthermore, Hudson and Thieken, (2022) investigated the potential presence of moral hazard, suggesting that increased insurance coverage may discourage precautionary measures. Using German data between 2002 and 2013, it was found that there's an indication after 2005 that insurance coverage lowered businesses' intentions to employ more adaptation measures. Despite these findings, the interaction between adaptation strategies and flood damage remains unclear during unprecedented events.

This study aims to build on existing advancements to gain a deeper understanding of the processes underlying both direct and indirect flood damages, particularly in the context of unprecedented events. To achieve this, we analyze data collected in the aftermath of the 2021 flood in Germany. The objectives of this study are:

1. To assess the type and extent of flood damage across companies of varying sizes.
2. To identify the key factors influencing direct damages (to buildings, equipment, and goods & stock) and indirect damages (particularly business interruptions and restriction durations) using machine learning techniques (Random Forest, Elastic Net, and XGBoost).
3. To develop a multivariate probabilistic model using Bayesian networks to derive predictive density estimates of damages, including median values and uncertainty ranges, across a range of hazard and exposure scenarios.

2 Data and Methods

2.1 Survey data

The July 2021 flood event in Germany is widely seen as an unprecedented disaster (Mohr et al., 2023; Thieken et al., 2023b; Zander et al., 2023): it caused an estimated €33.1 billion in direct damages and €7.1 billion in indirect damages (Trenczek et al., 2022). In the affected regions of North Rhine-Westphalia (NRW) and Rhineland-Palatinate (RLP), thousands of companies were severely impacted. According to BMI & BMF (2022), approximately 7,000 companies in NRW and 3,000 in RLP were affected by the flood. The German Insurance Association (GDV, 2023) reported 27,000 insured claims from companies, with claims expenditures totaling €2.4 billion in NRW and €0.9 billion in RLP.

To assess the impacts of the July 2021 flood on companies in NRW and RLP, a telephone survey was conducted between November 8, 2022, and January 31, 2023. The goal of the survey was to collect data on damages, influencing factors, the reconstruction process, and the preparedness and precautionary measures undertaken by the companies. The survey questionnaire was adapted from former surveys (Kreibich et al., 2007; Thieken et al., 2017) to ensure consistency in data

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

collection. A total of 434 companies participated in the survey, with an average interview duration of 42 minutes. The response rate was approximately 14%, with 608 refusals, 76 cancelled or unarranged surveys, and 1,886 companies that could not be reached by telephone. Three responses, which were referred to multiple business locations, including educational institutions and administrative buildings, were excluded from the analysis, which ultimately included 431 valid responses. Of the companies surveyed, 258 (60%) were located in NRW and 173 (40%) were based in RLP.

Table 1: List of factors influencing direct and indirect flood damages to companies. The variable type "*c*" stands for continuous, "*o*" for ordinal and "*n*" for nominal. Further details for the variables with * symbol can be found at Thieken et al., (2005) and Kreibich et al., (2010). See Supplementary information for the calculation of precaution.

	Abbreviation	Variable	Type	Range
Hazard	wd	Water depth	c:	1 cm to 963 cm above ground
	d	Inundation duration	c:	1 to 1200 hours
	v	Velocity	o:	1 = low flow to 3 = torrential flow
	con	Contamination*	o:	0 = none to 4 = heavy contamination
emergency measures	wt	Warning lead time	c:	0 to 336 hours
	ws	Early warning source*	o:	0 = no warning to 4 = official warning through authorities
	ew	Early warning received	n:	0 = no, 1 = yes
	me	Emergency measures undertaken	n:	0 = no, 1 = yes
	ep	Emergency plan	n:	0 = no, 1 = yes
	kh	Knowledge about hazard	n:	0 = no, 1 = yes
	ms	Emergency measures success*	o:	0 = no measure undertaken, 1 = completely ineffective to 3 = very effective
precaution	fe	Flood experience*	o:	0 = no experience to 3 = recent flood experience
	pr	Precaution	o:	0 = no precaution, 1 = medium precaution, 2 = Good precaution.
	in	Insurance	n:	0 = no, 1 = yes
company characteristics	sp	Size premise	c:	100 to 4,400,000 m^2
	sec	Sector	n:	1 = Agriculture, 2 = Manufacturing, 3 = Commerce, 4 = Financial, 5 = Private and public services
	ss	Spatial situation	o:	1 = several buildings, 2 = entire building, 3 = one or more floors, 4 = less than one floor
	own	Ownership	n:	1 = building owned, 2 = rented, 3= partly owned/ partly rented
	emp	Number of employees	c:	1 to 920
damage type	bdam	Relative damage to building	c:	degree of damage between 0 and 1
	edam	Relative damage to equipment	c:	degree of damage between 0 and 1
	gsdam	Relative damage to goods & stock	c:	degree of damage between 0 and 1

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

bid	Business interruption duration	c:	0 to 540 days (cases with 540 days reflect the end of the survey. Durations beyond that point are not available as numeric value)
brd	Business restriction duration	c:	0 to 540 days (cases with 540 days reflect the end of the survey. Durations beyond that point are not available as numeric value)

The surveyed variables were grouped into five categories: hazard, emergency measures, precaution, company characteristics, and damage type. The variable types and ranges are outlined in Table 1. All variables were included in the data-driven analysis to identify the most influential factors for each damage type. For each damage type, the percentage of missing values per variable was less than 10%, as some companies did not provide responses (Fig. S1). To avoid reducing the sample size, we employed the k-nearest neighbor technique with $k = 5$ to impute the missing data across the dataset at once (Zhang and Tian, 2025). We conducted a sensitivity analysis using k values of 1, 3, 7, and 9, and the findings were insensitive to the choice of k.

2.2 Variable Selection

2.2.1 Elastic Net

Elastic Net (EN) balances variable selection and model fitting, making it suitable for handling multicollinearity (Tay et al., 2023). It combines the strengths of both Lasso and Ridge regression. Lasso promotes sparsity by driving less important coefficients to zero, effectively performing variable selection. Whereas, Ridge shrinks all coefficients to stabilize the model in the presence of highly correlated variables. The EN objective function is given by (Zou & Hastie, 2005):

$$Obj(\beta) = \frac{1}{2n}\sum_{i=1}^{n}(y_i - X_i\beta)^2 + \alpha\left(\lambda\sum_{j=1}^{p}\left|\beta_j\right| + \frac{1-\lambda}{2}\sum_{j=1}^{p}\beta_j^2\right) \tag{1}$$

Where n is the number of samples (excluding one fold for cross-validation), p is the number of variables (19 in this case), y_i represents the response for i^{th} sample, and X_i is the corresponding variable vector. The coefficient β_j represents the effect of the j^{th} variable. The parameter α controls the strength of the regularization, and λ determines the balance between Ridge ($\lambda = 0$) and Lasso ($\lambda = 1$) regression. Optimal values of α and λ were obtained by minimizing $J(\beta)$. We implemented EN using the *ElasticNet* package from *scikit-learn* python library (Pedregosa et al., 2011). The optimal hyperparameters were selected based on the lowest mean absolute error (MAE) obtained from the nested cross validation (see Text S1). Predictions for the test dataset (X_t) were computed as:

$$y_t = X_t\beta \tag{2}$$

Where y_t represents the predicted values. EN is a powerful linear model and handles multicollinearity, but it cannot model nonlinear relationships.

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

2.2.2 Random Forest

Random Forest (RF) is an ensemble learning method that improves predictive performance and prevents overfitting by aggregating multiple decision trees (Breiman, 2001a). Individual decision trees tend to have high variance due to their sensitivity to data variability. RF addresses this limitation by constructing multiple decision trees, each trained on a bootstrap sample of the data (Aria et al., 2021). Additionally, at each root node, RF selects a random subset of variables for splitting, reducing correlation among trees and enhancing generalization.

We implemented RF using the *RandomForestRegressor* package from *scikit-learn* python library (Pedregosa et al., 2011). A randomized search was conducted over predefined hyperparameter ranges, including the number of trees, maximum tree depth, and the number of variables per split (see TextS1). Using the optimal hyperparameters, predictions for the test dataset (X_t) were obtained by averaging predictions from all individual trees:

$$y_t = \frac{1}{R}\sum_{r=1}^{R} f_r(X_t) \tag{3}$$

Where $f_r(X_t)$ is the prediction from tree r, and R is the total number of trees. RF is well-suited for heterogeneous data and often outperforms linear model (Schoppa et al., 2020; Sieg et al., 2017). However, it can be computationally expensive for large datasets with numerous variables and deep trees.

2.2.3 XGBoost

XGBoost (Extreme Gradient Boosting, XGB) is an optimized gradient boosting algorithm designed for speed and efficiency (Chen and Guestrin, 2016a). XGB also handles missing values, whereas RF requires explicit imputation. The objective function for XGB is defined as (Chen and Guestrin, 2016a):

$$Obj^B = \sum_{k=1}^{u} L\left(y_k, y_k^B\right) + \sum_{b=1}^{B} \Omega(f_b) \tag{4}$$

Where $L(y_k, y_k^B)$ is the loss function measuring the difference between the actual value y_k, and the predicted value y_k^B at boosting iteration B. The updated prediction for the k^{th} sample after B iterations is:

$$y_k^B = y_k^{B-1} + f_B(x_k) \tag{5}$$

Where, y_k^{B-1} is the prediction for the k^{th} sample after $B-1$ iterations. $f_B(x_k)$ is the prediction made by the model at iteration B for the k^{th} sample. Unlike RF, which constructs trees independently and in parallel, XGB builds trees sequentially, where each new tree corrects the residual errors of the previous ones (Narin, 2025). Additionally, XGB incorporates both Lasso and Ridge regularization to control overfitting (Ma et al., 2021). The regularization term $\Omega(f_b)$ for the b^{th} model is defined as:

$$\Omega(f_b) = \gamma T + \frac{1}{2}\Lambda \sum_{m=1}^{T} w_m^2 \tag{6}$$

Where T is the number of terminal nodes in the tree f_b. γ is a regularization parameter that penalizes the number of leaves in the tree (encouraging simpler trees with fewer leaves). w_m represents the weight associated with the m^{th} leaf of the tree. Λ is a regularization parameter that penalizes the squared weights of the leaves. This term helps prevent overfitting by controlling the weights of the leaves. We implemented XGB using the *XGBRegressor* package from *scikit-learn* python library (Pedregosa et al., 2011). Predictions for test dataset (X_t) using optimal hyperparameter selection were computed as:

$$y_t = \sum_{b=1}^{B} f_b(X_t) \tag{7}$$

2.2.4 Variable importance

To assess variable importance across the three models, we used the Permutation Variable Importance (PVI) technique. PVI quantifies the contribution of each variable by measuring the change in model performance when its values are randomly permuted while keeping all others unchanged (Breiman, 2001a). Variables that cause a greater increase in error upon shuffling are considered more important (Fisher et al., 2019). This approach is model-agnostic and provides a consistent framework for comparing variable importance across different predictive models.

For each damage type, PVI scores for all variables in each model were rescaled to a range of 0 to 100 using min-max normalization. The median MAE from nested cross-validation was obtained for each model, denoted as MAE_{EN}, MAE_{RF}, and MAE_{XGB}. The model weights were computed as follows:

$$Weight_{EN} = \frac{\frac{1}{MAE_{EN}}}{Total}, Weight_{RF} = \frac{\frac{1}{MAE_{RF}}}{Total}, Weight_{XGB} = \frac{\frac{1}{MAE_{XGB}}}{Total} \tag{8}$$

Where, $Total = \frac{1}{MAE_{EN}} + \frac{1}{MAE_{RF}} + \frac{1}{MAE_{XGB}}$. The PVI scores from each model are weighted based on the respective model weights. The final variable importance was computed as the sum of the weighted scores, ranging from 0 to 100. If all three models identified the same variable as the most important, its importance would be 100. Based on the combined weighted importance scores, variables were ranked accordingly.

2.3 Bayesian Networks for multivariate probabilistic modeling

Bayesian networks (BNs) are probabilistic graphical models that represent dependencies among multiple variables and enable multivariate predictive density estimation (Sucar, 2021). A BN is a directed acyclic graph (DAG), $G = (V, E)$, where V denotes the set of variables and E represents the directed edges encoding conditional dependencies. The dataset comprises both continuous and categorical variables (see Table 1). Continuous variables are discretized using an equal-frequency

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

binning approach, where the number of bins is determined iteratively to optimize network learning while preserving data characteristics (Kitson et al., 2023). The network structure is learned through a data-driven approach based on the Tabu Search algorithm (Glover, 1986; Goudet et al., 2018), which iteratively explores possible network configurations by adding, removing, or reversing edges. The optimal structure is selected by maximizing the Bayesian Dirichlet Equivalent (BDe) score (Heckerman et al., 1995), which balances model complexity and goodness of fit while incorporating prior knowledge.

We developed five separate BNs corresponding to different damage types. During model development, we observed that for some damage types, the direct connections to the target variable (i.e., damage) involved up to four variables. To ensure consistency across BNs and to maintain model interpretability and parsimony, we selected the top four variables based on the combined weighted importance scores. The learned BN enables probabilistic inference, allowing computation of the posterior probability of any variable X_i given observed evidence E (Pearl, 1988):

$$P\left(\frac{X_i}{E}\right) = \frac{P\left(\frac{E}{X_i}\right)P(X_i)}{P(E)} \; with \; P(E) = \sum_{X_i} P\left(\frac{E}{X_i}\right)P(X_i) \quad (9)$$

Where $P\left(\frac{E}{X_i}\right)$ is the likelihood of evidence given X_i, and $P(X_i)$ is the prior probability of X_i. A detailed step-by-step procedure of the BN learning process, Conditional Probability Tables (CPTs), and Bayesian inference is provided in Text S2. The posterior probability of flood damage given the observed evidence E is discrete in nature. To derive a continuous distribution of direct and indirect damages, we fit a probability distribution based on weighted sampling of the empirical damage data, following the approach of Schoppa et al. (2020).

Flood damage processes vary by region, flood type, and asset type (Mohor et al., 2020; Sairam et al., 2019; Wagenaar et al., 2018). To derive the drivers of flash flood losses, this study adopts a data-driven feature selection approach to the empirical data. Feature selection involves identifying variables that have the highest influence on the target variable (i.e. relative loss). We train multiple models – nonlinear models: Random Forest (RF), Extreme Gradient Boosting (XGBoost), and linear model: Elastic Net (EN).

RF is an ensemble machine learning method primarily used for classification and regression tasks, developed by Breiman, (2001). RF generates an ensemble of decision trees, each trained on a random subset of the data using bootstrap sampling. At each node within these trees, a random subset of features is considered for splitting. The final prediction for a given input is obtained by averaging the predictions from all individual trees. This approach helps RF reduce overfitting and enhances the model's generalization ability. XGBoost, similarly to RF, is an ensemble learning algorithm that benefits from a decision tree-based structure. However, the key difference compared to RF is that in XGBoost, each tree corrects the errors from the previous ones. The process starts with a simple model and iteratively adds trees that focus on the residuals or errors made by the existing ensemble. With its efficient implementation, XGBoost demonstrates superior performance and handles large-scale data more effectively than RF (Chen and Guestrin, 2016b). While RF and XGBoost are non-linear models, EN is a

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

regularization technique used in linear regression, combining both Lasso (L1) and Ridge (L2) regularization penalties. It effectively addresses multicollinearity in datasets by shrinking the less influential predictors toward zero (Lasso) while additionally providing some degree of regularization to prevent overfitting (Ridge). EN's ability to handle correlated features and select relevant predictors makes it a valuable tool in regression tasks (Zou and Hastie, 2005).

During training, we employed a nested cross-validation framework with 10 splits and 10 repeats, resulting in a total of 100 evaluations. We selected the best set of hyperparameters, which obtained the least mean absolute error, which was then applied to the final feature selection. From each resulting final model, we derived the feature importance. Next, we calculated each variable's weighted feature importance and overall rank. The final selection of the variables (Fig 1) is elaborated upon in the results section.

3 Results and Discussion

3.1 Overview of affected companies in the 2021 flood event

This section provides an in-depth analysis of the affected companies, focusing on their demographic profiles, the types of damage sustained, the extent of business interruptions, and the financial implications across various damage categories. Sales figures were requested but often not reported, due to this the number of employees is used as the measure of company size. The companies range from micro-companies with up to nine employees to large companies with 250 and more employees, according to the European classification (Destatis, 2003). The majority of the companies surveyed are therefore classified as micro-companies (1–9 employees) followed by small companies with 10 to 49 employees and medium-sized companies with 50 to 249 employees (Fig. 1). Large companies with 250 and more employees rarely participated.

Figure 1a illustrates the distribution of companies across sectors, showing a relatively balanced representation except for agricultue. Based on the WZ2008 economic classification (Destatis, 2008), all companies surveyed were assigned to one of five economic sectors: 1) agriculture (n = 14); 2) manufacturing (n = 81); 3) commercial (n = 126); 4) corporate and financial services (further: financial) (n = 81) and 5) public and private services (further: services) including educational, health and social services (n = 129). Micro and small companies dominate the sample, which aligns with the typical business landscape of many European countries (Eurostat, 2024). Figure 1b indicates a clear relationship between company size and the size premise of the companies. Micro-companies predominantly operated from size premise ≤ 5000 m^2, whereas medium and large companies were more likely to occupy a higher size premise > 5000 m^2. Large size premise inherently increased exposure to floodwaters, which partly explains the heightened damages among medium and large companies. A significant observation from Figure 1c is the generally low implementation of precautionary measures, particularly among micro and small companies.

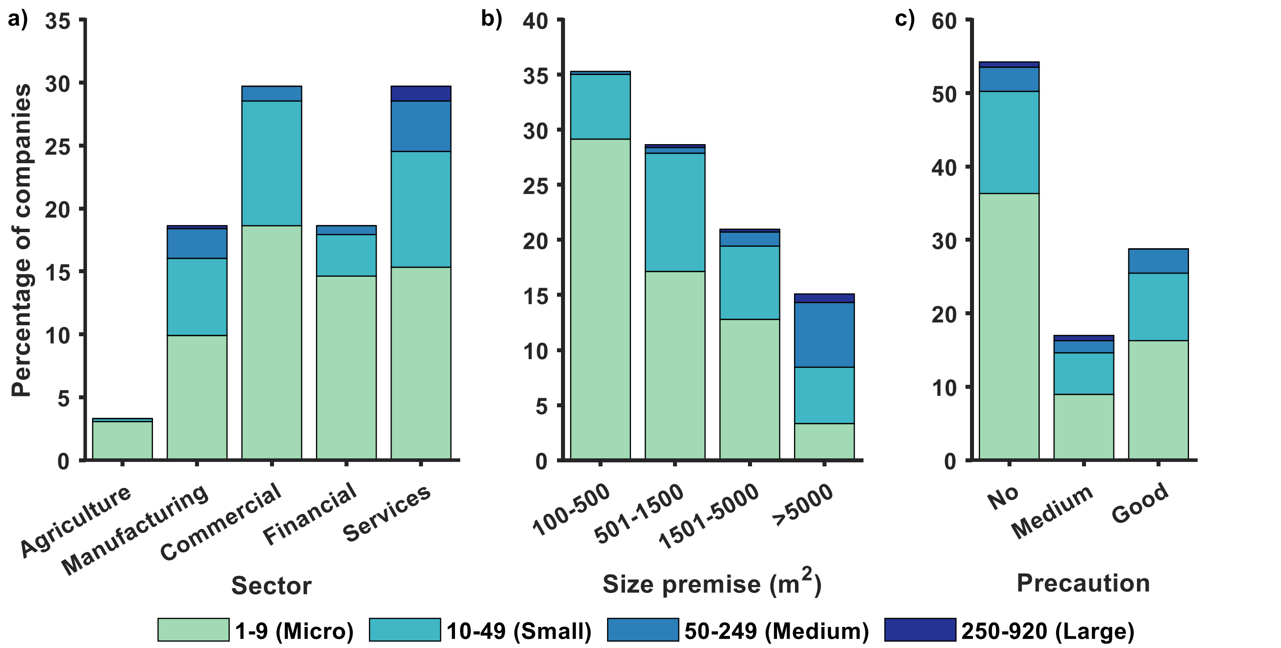


Figure 1: Bar chart showing the percentage of companies with varying numbers of employees by (a) sector, (b) size premise, and (c) precaution.

Floods not only cause damage to tangible assets through hydrodynamic forces and chemical contamination but also lead to significant disruptions in supply chains and transportation. These disruptions can result in partial or complete business interruptions, triggering consequences ranging from loss of sales to bankruptcy (Thieken et al., 2016). Figure 2 illustrates the percentage of companies affected by various types of impacts, categorized by company size. The results reveal clear differences in vulnerability and exposure levels across different company sizes. Damage to buildings emerges as the most frequently reported impact, with nearly 100% of companies across all size categories affected. Larger companies report the highest exposure to equipment damage (100%) and loss of goods and stock (over 80%), suggesting that companies with larger operational setups have more assets at risk. In contrast, the micro-companies report slightly lower, yet still significant, impacts in these categories, with equipment damage close to 90% and goods and stock losses around 70%.

Business interruption is another major consequence reported consistently across all company sizes, reaching 100% among large companies (Fig. 2). This suggests that larger operational scales correlate with increased disruption potential. Business restrictions due to regulatory or environmental constraints are reported less frequently but remain relevant, particularly for medium and large companies, with a frequency exceeding 60%. Interruptions in utility services are a widespread issue, affecting 90% to 100% of companies across all size categories. This finding highlights the universal dependency of businesses on essential services such as electricity, water, and telecommunications. Loss of customers and employee delays are also commonly reported impacts. Micro-companies experience customer losses of around 60%, underlining the challenges to business continuity and client retention following flood events. In contrast, employee delays affect

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

approximately 80% of companies, except micro-companies, reflecting disruptions in workforce mobility. Regarding supply chain disruptions, problems with suppliers affect between 40% and 80% of companies, with the highest impacts reported by larger businesses (80%; Fig. 2). This suggests greater vulnerability due to complex supply dependencies. Conversely, delivery problems are reported less frequently, with medium-sized companies experiencing the lowest impact (40%).

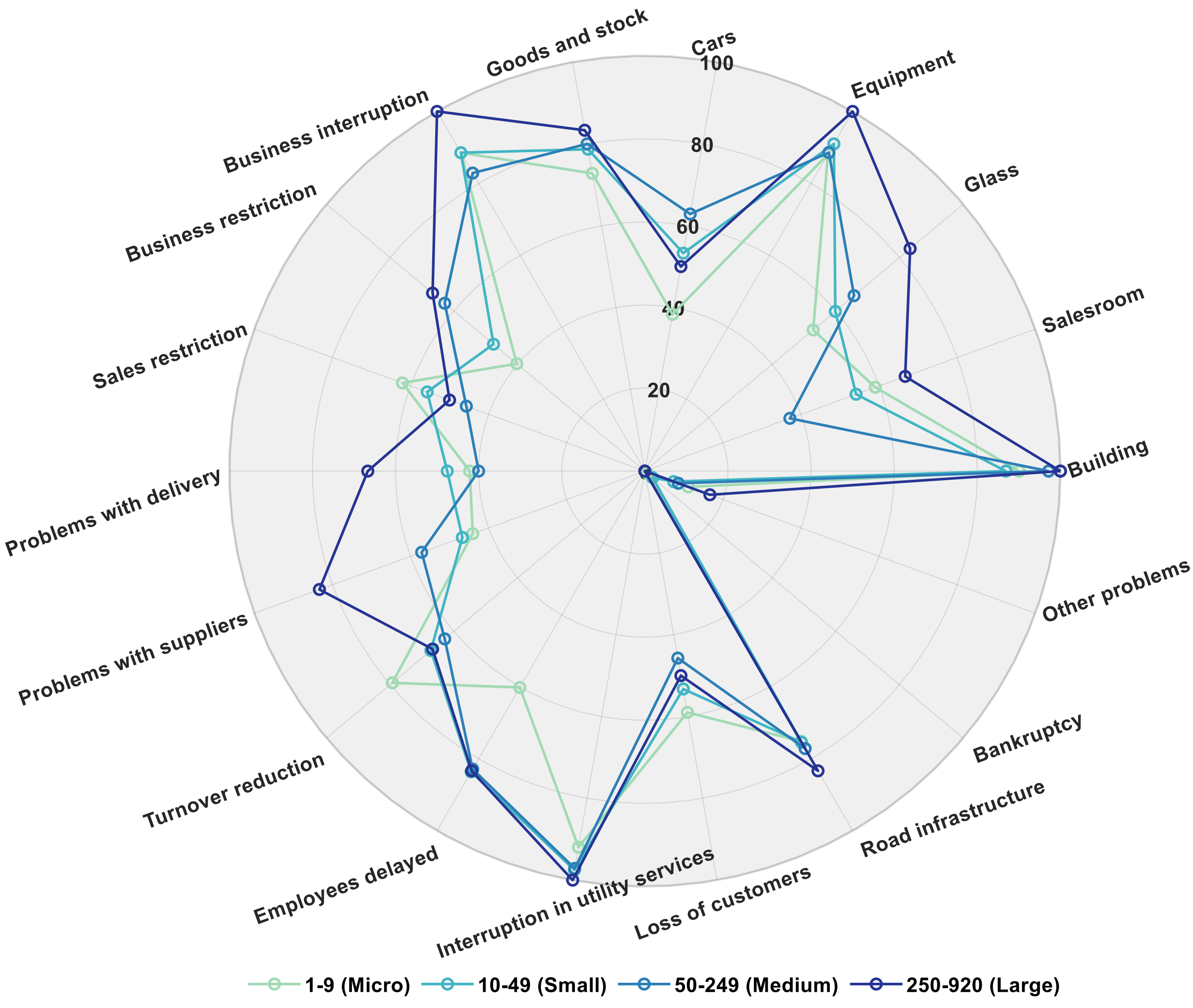


Figure 2: Spider chart illustrating the percentage of companies experiencing different types of flood impacts, categorized by the number of employees.

Bankruptcy risks remain generally low across all company sizes, except for two isolated cases. This indicates that while the damages are widespread, most businesses in our sample manage to avoid insolvency. Turnover reduction is moderately reported (60%–80%) without a distinct size-based pattern, although micro-companies appear more affected, with rates around 80%. Damage or inaccessibility of road infrastructure is reported by approximately 80% of companies, underscoring systemic exposure that affects businesses regardless of size. Car damage is less frequently reported but shows slightly higher percentages (above 60%) among medium-sized companies. Sales restrictions exhibit variability, with micro-companies reporting higher percentages (over 60%), possibly due to their greater dependence on physical sales venues. Glass damage is moderately reported across companies but is notably higher among larger companies (around 80%), likely due to their larger commercial structures and exposure. Overall, the results illustrate the complex and diverse impacts of flooding on companies, varying by size and operational characteristics. They highlight the need for tailored risk management and resilience strategies, especially for micro and small companies that are more susceptible to supply chain disruptions and sales restrictions, while larger companies face higher asset-related risks.

July 2021 flood event had long-lasting impacts on businesses, severely disrupting operations for months or even years. Figure 3 presents the distribution of business interruption duration and business restriction duration (both measured in days) across companies of varying sizes. The boxplots reveal clear differences in the duration of these impacts based on company size. For business interruption duration, micro-companies (1–9 employees) experienced the longest disruptions overall, with a median duration of approximately 40 days. However, the range of reported durations for this group was highly variable, with several extreme cases extending beyond 365 days, as reflected by numerous outliers. This finding underscores the particular vulnerability of micro-enterprises to prolonged operational disruptions following flood events, likely due to their limited resources and reduced adaptive capacity. In contrast, small, medium, and large companies reported comparatively shorter business interruption durations. The median interruption durations for these groups ranged between 10 and 30 days, with fewer extreme cases observed. Notably, medium-sized companies demonstrated shorter interruption periods overall, suggesting better resilience or recovery capacity. This may be attributed to diversified operations, greater financial buffers, or the presence of formal contingency plans that facilitate faster recovery.

The pattern shifts when examining the duration of business restrictions. Both micro and small companies reported significantly prolonged periods of business restrictions, with median durations exceeding 100 days. In some cases, restrictions extended up to 365 days, again marked by several extreme values. The persistence of these restrictions may reflect regulatory, environmental, or logistical hurdles encountered during the recovery phase, particularly by smaller companies that often lack the influence or flexibility to expedite resolution. Interestingly, medium-sized companies reported relatively shorter business restriction durations, with a median significantly lower than that of micro and small companies. Most data points for this group clustered below 100 days, indicating a more efficient recovery from regulatory or operational constraints. For large companies, only a few values were reported, which likely explains the narrower distribution observed. Overall, these results highlight that company size is a critical factor influencing the duration of operational disruptions following extreme events. Micro and small companies are particularly vulnerable to prolonged indirect impacts, such as extended business restrictions and interruptions. In contrast, medium and large companies tend to recover more quickly, likely benefiting from greater resilience, diversified operations, and access to more substantial resources.

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

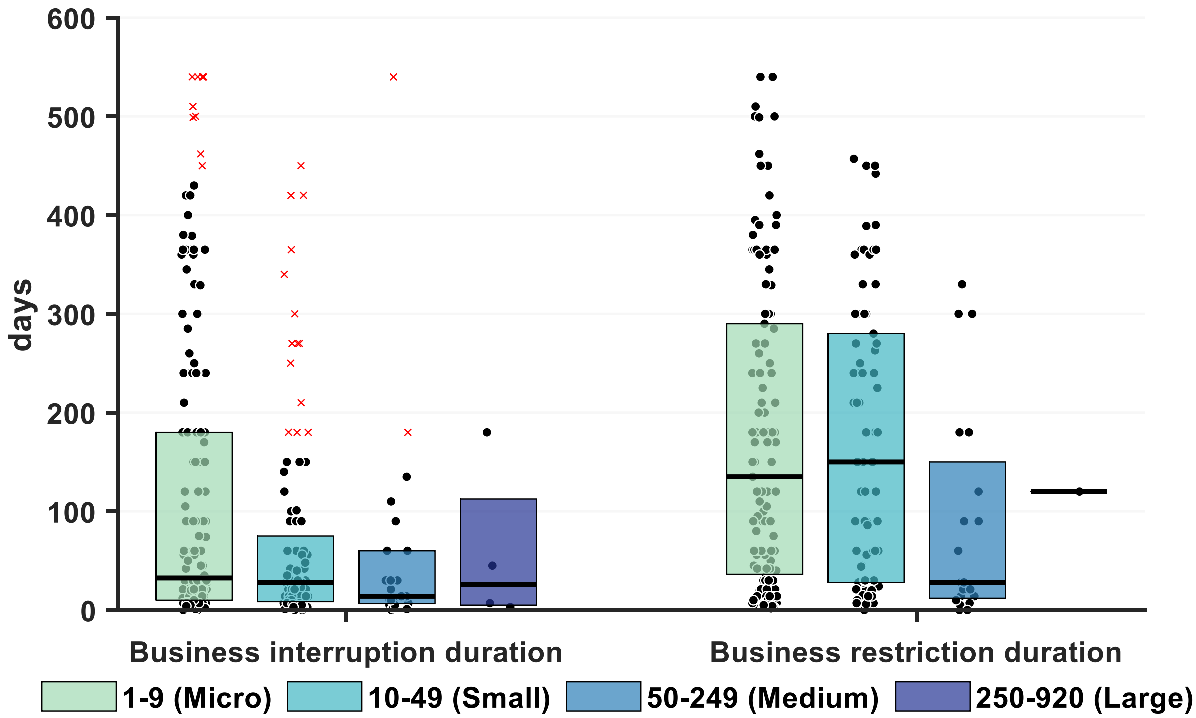


Figure 3: Boxplot of (a) Business interruption duration (days) and (b) Business restriction duration (days) for companies categorized by the number of employees. Black circular markers represent individual data points, and red crosses indicate outliers.

The survey recorded specific damage amounts across three categories of direct property damage: (1) building, (2) equipment, (3) goods & stock, as well as financial losses due to business interruptions. In most cases (approximately 62%), the direct damages amounted to more than €100,000 , while around 24% of companies reported damages reaching into the millions. The average costs (in euros) for each company size are presented in Table 2, alongside medians and the number of companies (n) contributing to each calculation. Building damages accounted for the highest average costs across all company sizes, particularly impacting medium and large companies. Micro companies reported average building damages of €711,459, with a median of €250,000. This wide gap between the mean and median suggests that while many small firms experienced moderate losses, a few outliers faced severe damages. For small companies, the average building damage increased to €908,482 (median €500,000). Medium companies faced substantial building-related losses, averaging €2,838,103 with a median of €1,350,000. Large companies, though represented by a very small sample (n = 4), reported the highest average building damages of €7,350,000, reflecting the scale of structures at risk within large industrial facilities.

In terms of equipment damages, micro companies incurred an average loss of €297,854, while small companies experienced significantly higher average costs of €541,898. Medium companies reported the highest average equipment losses at €3,630,652, likely driven by the presence of high-value machinery. Interestingly, large companies recorded a comparatively lower average equipment loss of €160,000, though this is based on a very small sample size (n = 3). Lower median values

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

across groups suggest the presence of extreme cases skewing the mean, particularly among medium-sized companies. Goods and stock damages were generally lower across all company sizes (Table 2). Micro companies faced average losses of €159,422, while small companies reported similar average damages of €134,470. Medium companies experienced higher average losses of €1,503,250, indicating greater inventory exposure. Large companies reported much smaller average losses of €55,000. Lower median values that most companies incurred relatively less damages in this category, with a few outliers. Business interruption losses also varied by company size. Micro companies faced average interruption costs of €139,931, while small companies reported higher average losses of €311,173. Medium companies were the most affected, with average losses of €703,250. Large companies, despite the small sample size (n = 3), recorded an average business interruption cost of €400,000, with the median even higher at €500,000, reflecting significant operational disruptions. Overall, the financial costs associated with building, equipment, goods & stock, and business interruption showed that larger companies typically incurred more significant costs. However, substantial variance within each category highlights the influence of extreme cases. Importantly, the limited number of large companies surveyed suggests that these results should be interpreted cautiously, as they may not fully representative.

Table 2: Average financial costs (in euros) incurred for building, equipment, goods and stock, and business interruption categorized by the number of employees (values in brackets represent medians, and n denotes the number of companies included in the calculation of the means and medians)

Number of employees (Company size)	Building	Equipment	Goods & stock	Business interruption
1-9 (Micro)	711,459 (250,000) n = 167	297,854 (50,000) n = 203	159,422 (30,000) n = 154	139,931 (30,000) n = 143
10-49 (Small)	908,482 (500,000) n = 83	541,898 (150,000) n = 96	134,470 (31,500) n = 82	311,173 (100,000) n = 74
50-249 (Medium)	2,838,103 (1,350,000) n = 29	3,630,652 (600,000) n = 23	1,503,250 (150,000) n = 20	703,250 (200,000) n =16
249-920 (Large)	7,350,000 (1,700,000) n = 4	160,000 (200,000) n = 3	55,000 (10,000) n = 3	400,000 (500,000) n = 3
Total	1,080,999 (350,000) n = 283	604,528 (100,000) n = 325	254,083 (30,000) n = 259	215,910 (50,000) n=236

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

3.2 Data-driven analysis of factors influencing direct and indirect flood damages

Understanding the complex processes driving flood damage is crucial for developing effective risk reduction measures for companies. To date, most insights into damage mechanisms stem from studies on private households affected by riverine floods (Gerl et al., 2016; Thieken et al., 2022). This analysis seeks to close the knowledge gap on the factors driving direct and indirect damages to companies during unprecedented flood events. Based on expert knowledge and data availability, 19 potentially relevant influencing factors were selected, covering hazard characteristics, emergency measures, precautionary actions, and company characteristics (Table 1). The dataset exhibited less than 7% missing data for 18 out of 19 variables (Fig. S1), which were imputed. Figure 4 presents the pairwise Spearman rank correlations between influencing factors and the five damage types—relative damage to buildings (bdam), equipment (edam), goods & stock (gsdam), business interruption duration (bid), and business restriction duration (brd).

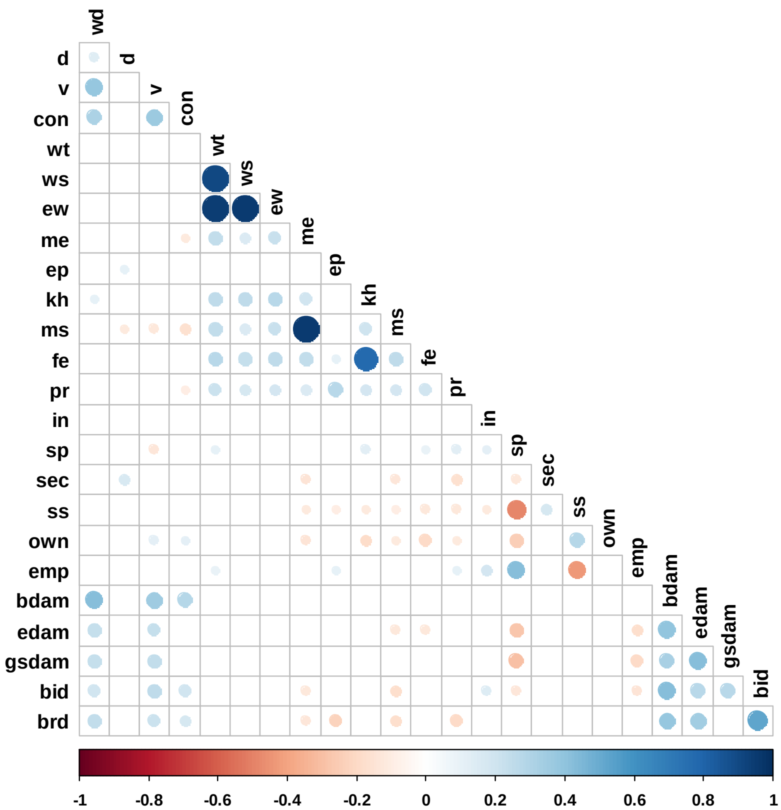
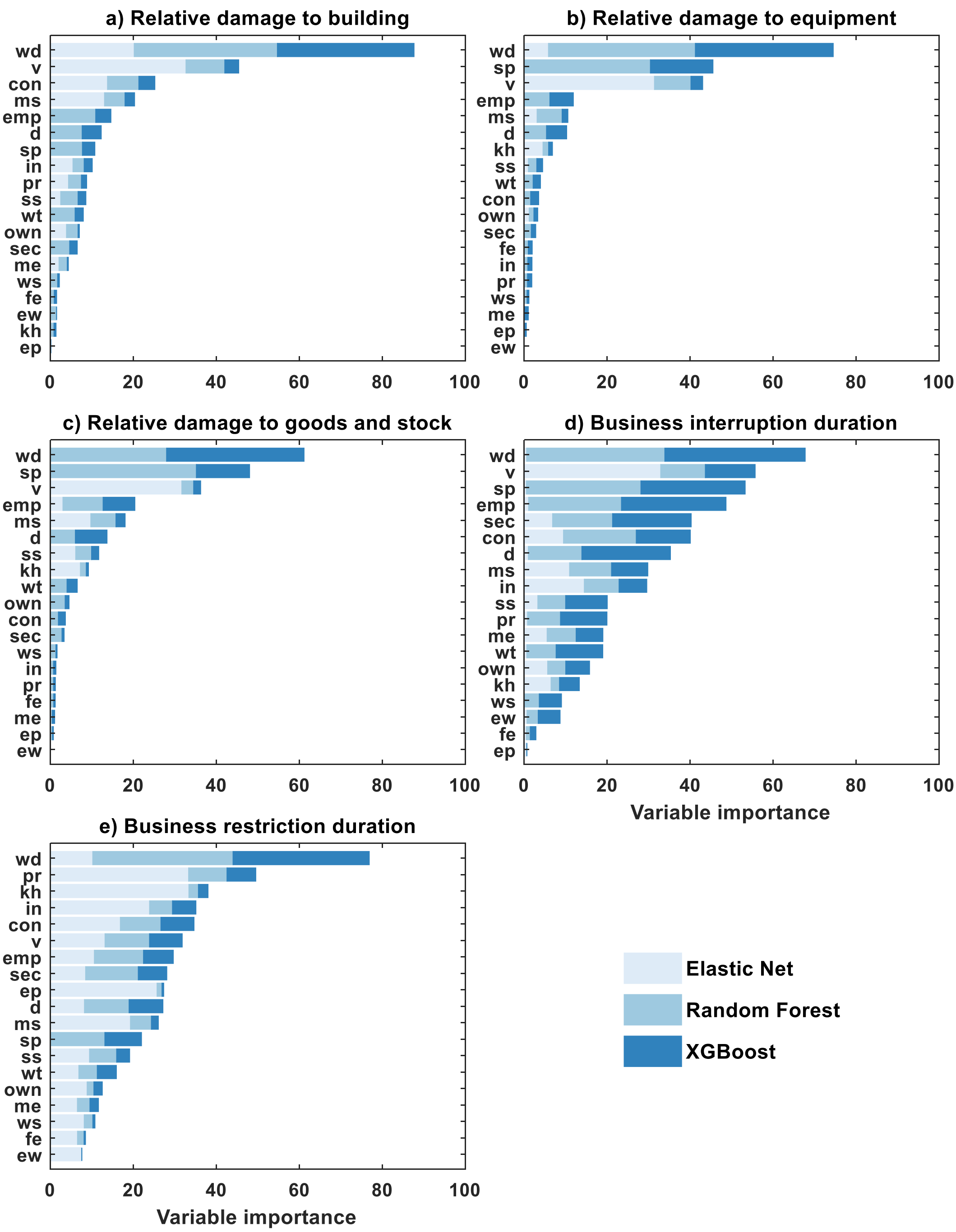


Figure 4: Spearman rank correlation coefficients between 19 influencing factors and five damage types. Only significant correlations (p-value < 0.05) are displayed, providing insights into key factor-damage relationships. See Table 1 for abbreviations.

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025



385 **Figure 5: Importance of influencing variables (see Table 1 for abbreviations) for damage types: (a) buildings, (b) equipment, (c) goods & stock, d) business interruption duration, and e) business restriction duration. The x-axis shows the weighted importance of each variable, as determined by the three models (Random Forest, Elastic Net, and XGBoost).**

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

High positive correlations exist between water depth (wd), flow velocity (v), and various damage types, highlighting the critical role of flood intensity in driving both direct and indirect damages. Interestingly, factors characterizing emergency response and preparedness exhibit negative correlations with business interruption and restriction durations. Specifically, the successful implementation of emergency measures (ms), and precaution (pr) are associated with reduced indirect impacts (Fig. 4). Additionally, company characteristics such as the size of the premises (sp) and the number of employees (emp) show negative correlations with equipment and goods & stock damages, as well as business interruption duration, reflecting the role of operational scale and exposure in shaping flood impacts. Furthermore, considerable inter-correlations are observed between various influencing factors, reinforcing the need for a multivariate approach.

To account for these interactions and robustly identify the most influential factors, a data-driven modelling framework was implemented. Three machine learning models (Random Forest, Elastic Net, and Extreme Gradient Boosting (XGBoost)) were trained on the empirical survey data. Ten repetitions of a ten-fold cross-validation based on random partitioning were carried out. In each model, the hyperparameter combination model yielding the lowest MAE was used to derive the variable importance. The combined variable importance scores from all three models are illustrated in Figure 5. As expected, water depth (wd) consistently emerged as the most important driver across all damage types, aligning with previous findings (Schoppa et al., 2020; Sieg et al., 2017). Notably, flow velocity (v) ranked as the second or third most influential factor, particularly for the 2021 flood event. This prominence of velocity reflects the dynamic nature of the flood, contrasting with large-scale, slowly rising river floods where factors like contamination typically dominate damage (Kreibich et al., 2007; Sieg et al., 2017). In addition, company characteristics such as the size of the premises (sp) and the number of employees (emp) also played significant roles (Fig. 5). The success of emergency measures (ms) further influenced damage, ranking fourth or fifth in importance for direct damages (Fig. 5a-c). Interestingly, business restriction duration (brd) was primarily influenced by preparedness-related variables, i.e., amount of precaution taken (pr), prior knowledge about the hazard (kh), and the company's insurance status (Fig. 5e). This finding underscores the critical role of proactive risk management in minimizing operational disruptions, even during unprecedented events like the 2021 flood.

3.3 Quantifying direct and indirect flood damages using multivariate probabilistic modelling

Understanding interdependencies among influencing factors and damage types is crucial for reliable flood damage estimation. Figure 6 presents the kernel density estimations of the top four influencing factors considered for multivariate probabilistic damage modelling across five different types of damage. The violin plots illustrate the probability density of scaled variables (ranging from 0 to 1), with quartile lines indicating central tendencies and variability. The presence of skewed distributions and multimodal characteristics highlights the complexity of flood damage relationships across different damage types.

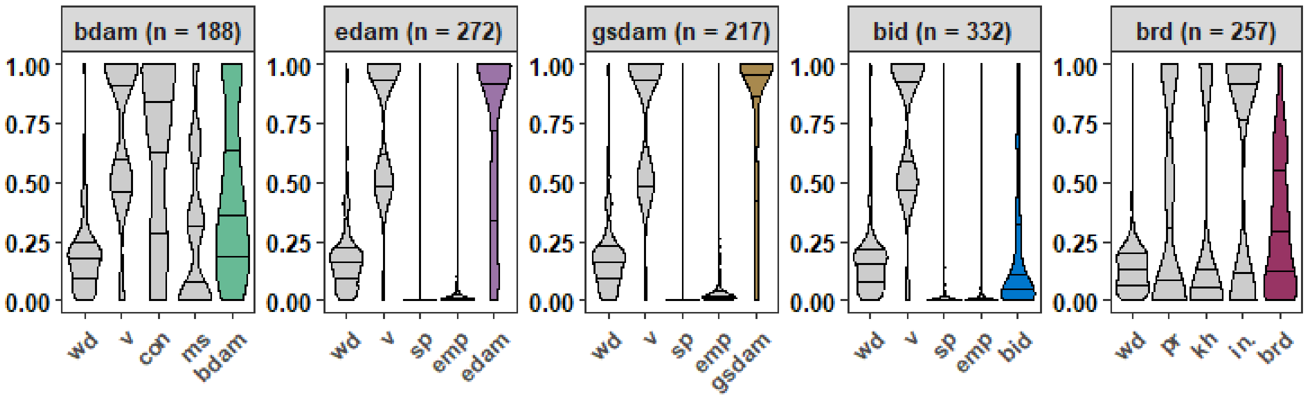


Figure 6: Kernel density estimations of influencing factors and damage types, with all variables scaled between 0 and 1. The lines in the violin plots indicate the quartiles. For abbreviations see Table 1.

Bayesian networks (BN) provide a probabilistic framework for understanding the interdependencies between the top four influencing factors and damage outcomes. We developed data-driven BN models using a score-based structure learning algorithm. The BN structure (Figure 7) provides a probabilistic representation of these relationships, allowing for improved estimation of both direct and indirect damages under uncertainty. The direction of the arrows represents associations between variables but does not necessarily imply causality (Sairam et al., 2020). The results align with previous studies while also offering new insights into key influencing factors.

Consistent with prior research (Kreibich et al., 2010; Seifert et al., 2010; Nafari et al., 2016; Sieg et al., 2017; Schoppa et al., 2020, 2022), our results confirm that water depth (wd) and velocity (v) are the most critical hazard predictors for direct damages, particularly for building damage (bdam). The direct link between these variables and bdam (Fig. 7a) underscores the predominant role of flood intensity. Furthermore, our study highlights the role of contamination (con) in influencing building damage, which is in agreement with Sieg et al. (2017). Intuitively, the perceived success of emergency measures (ms) is linked to the water depth (Fig. 7a). For equipment damage (edam) and goods & stock damage (gsdam), our results show that flow velocity and company characteristics such as size premises (sp) and number of employees (emp) play significant roles (Figures 7b and 7c). This supports the findings of Schoppa et al. (2020), who emphasized the importance of company-specific characteristics in explaining non-structural damages. Unlike previous studies that primarily focused on hazard intensity variables (Nafari et al., 2016; Sieg et al., 2017), our results provide a more nuanced perspective by demonstrating how company exposure modulate damage susceptibility.

BN structure of business interruption duration (bid) (Fig. 7d) indicates that wd, v, and emp are key predictors of bid, which is in agreement with Sultana et al. (2018), who found that company-specific factors (e.g., number of employees) often outweigh hazard characteristics in estimating business interruption costs. Additionally, our results support Endendijk et al. (2024), who identified water depth as a primary factor affecting business interruption. Moreover, our findings complement

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

those of Sakai and Yao (2023), who highlighted that small companies suffer disproportionately higher business interruption relative to turnover. Interestingly, the BN structure of business restriction duration (brd) (Fig. 7e) reveals the impact of precautionary measures (pr), which is linked to the knowledge about flood hazard (kh). This underscores the role of risk communication in shaping proactive behavior. While previous studies have acknowledged the importance of preparedness (Kreibich et al., 2010; Schoppa et al., 2022), our BN results explicitly quantify its role in reducing business restriction duration. The direct link between pr and brd suggests that proactive measures have a tangible effect on post-flood recovery.

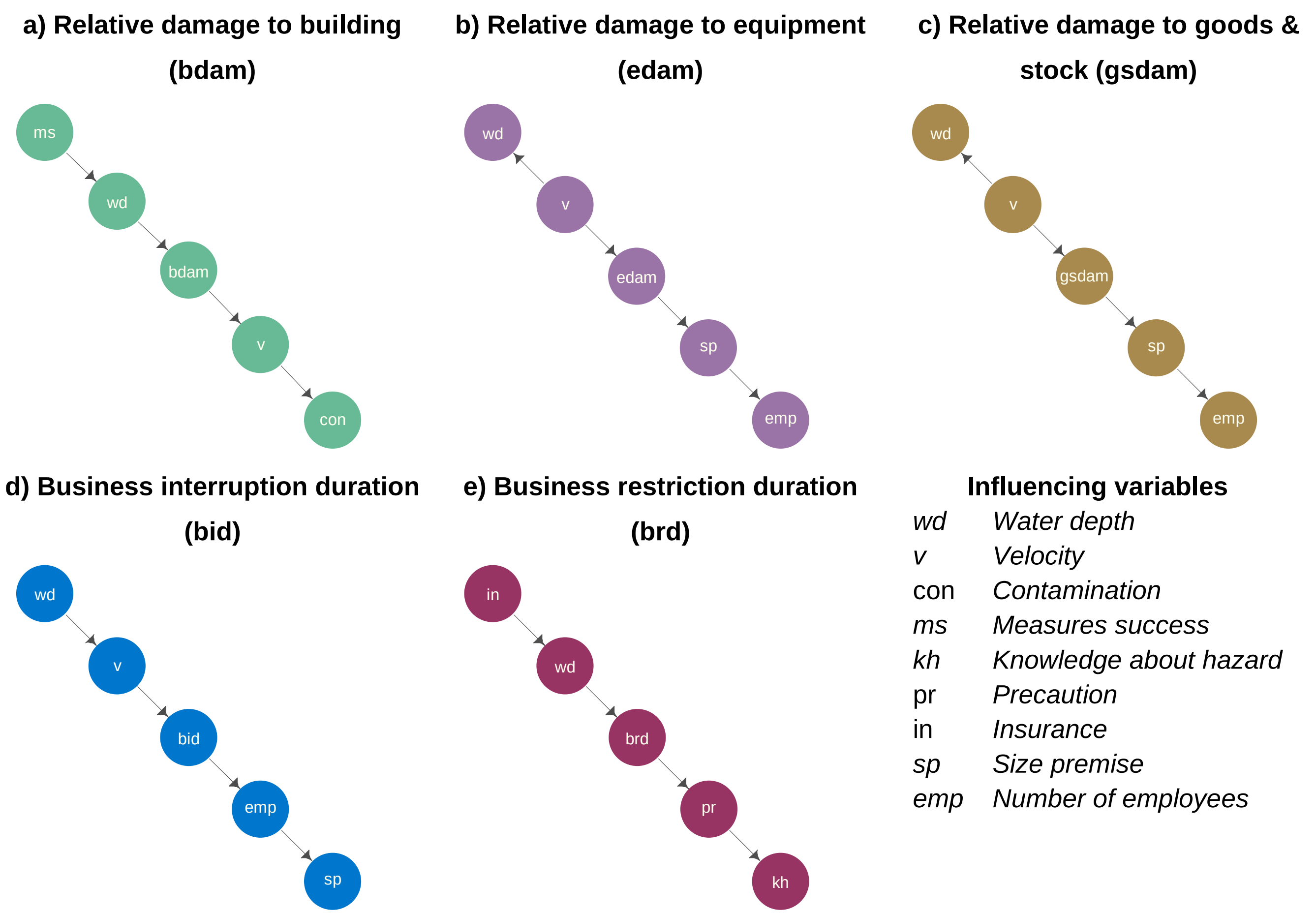


Figure 7: Bayesian network structure showing interdependencies among factors and damage types, derived from a score-based structure learning algorithm.

We used the BNs to estimate the damage under different hazard, exposure, and vulnerability scenarios. Figure 8 shows the distribution for five types of flood damage, where each damage type is probabilistically modelled using its respective Markov blanket. These distributions are derived from Conditional Probability Tables (Fig. S4). In Figure 8, horizontal solid lines represent the observed range of damage and business interruption/restriction durations, scaled from 0 to 1. Red dots

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

indicate the median (50th percentile) relative damage, while the dotted vertical lines denote the interquartile range (25th–75th percentiles), providing a measure of uncertainty. The relative damage to buildings is modeled as a function of water depth and flow velocity (Fig. 8a). As water depth increases, the median damage values generally rise, especially under moderate and torrential flow conditions. At low flow velocities, median damage remains relatively low across all depth levels, although the uncertainty (as indicated by the interquartile range) increases with depth, suggesting greater variability. Under moderate flow conditions, damage estimates increase slightly compared to low flow, with overlapping uncertainty bounds. In contrast, torrential flow conditions consistently lead to the highest damage estimates, particularly for water depths exceeding 2 meters, where the 75th percentile approaches near-total destruction. Notably, the uncertainty in damage estimates increases with both rising water depth and flow velocity, indicating heightened variability in damage outcomes under extreme flood conditions.

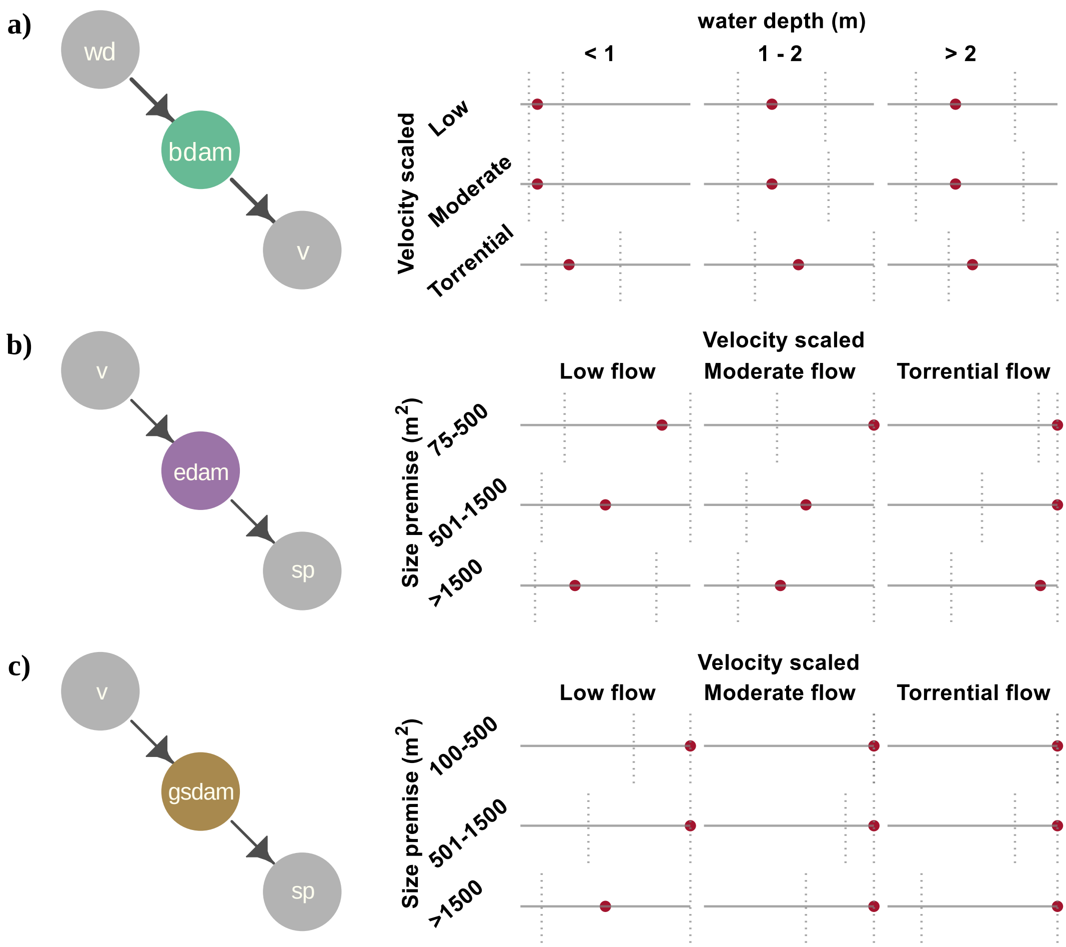
The relative damage to equipment (edam) is assessed as a function of flow velocity and company size premises (Fig. 8b). Flow velocity categories (Low, Moderate, Torrential) are arranged as columns, while classes of size premises (75–500 m², 501–1500 m², >1500 m²) are organized in rows. At low flow velocities, median damage is relatively high but decreases as the size premises increases, particularly for the largest category (>1500 m²). Under moderate and torrential flow conditions, equipment damage escalates significantly, with most cases reaching the maximum possible damage (100%) for companies with size premises smaller than 500 m². In addition, these companies exhibit a narrower range of damage values, indicating more consistent outcomes. In contrast, companies with premises larger than 1500 m² show greater uncertainty, especially under moderate and torrential flow conditions, where the 75th percentile reaches 100% damage in most cases. While low flow velocities present a moderate risk, particularly for companies with larger premises, torrential flows lead to severe damage, regardless of the size premises.

The relative damage to goods & stock is also modeled as a function of flow velocity and size premises (Fig. 8c). In low flow scenarios, the median damage estimate ranges from 0.50 for companies with premises >1500 m² to 1.0 for companies with smaller premises (<1500 m²). Under moderate and torrential flow conditions, the majority of damage values concentrate around 1.0, indicating near-total damage to goods and stock under extreme flood conditions, irrespective of company size premises. In addition, for smaller premises (75–500 m²) the uncertainty is very less. In contrast, companies with premises exceeding 1500 m² exhibit greater variability, with the 75th percentile reaching 100% damage in most cases. These results underscore the importance of considering both company size premises and flow velocity when evaluating potential impacts on equipment and goods & stock during floods.

The duration of business interruption varies with flow conditions and company size. Micro-companies (1–9 employees) show a consistent pattern under low and moderate flow conditions, with median values of 0.04. However, under torrential flow conditions, interruption durations increase significantly, with median values reaching 0.11. Small companies (10–49 employees) exhibit a similar trend, though their interruption duration under torrential flow is slightly lower. Medium and large companies (>49 employees) demonstrate greater resilience, with median interruption durations ranging from 0.02 to 0.06 across all flow conditions. Their lower upper quartile values suggest more effective adaptation strategies. The results

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

indicate that small companies, especially micro-companies, are disproportionately affected by extreme flood events. The analysis of business restriction duration, based on the Markov blanket, demonstrates that companies without precautionary measures experience the longest restrictions. For instance, the median restriction duration for companies without precautions increases from 0.19 for water depths less than 1 meter to 0.39 for depths greater than 2 meters. Implementing medium precautionary measures (see supplementary information) results in a slight reduction in restriction duration, particularly for deeper water, where the 75th percentile decreases from 0.68 to 0.61. However, the most significant reduction is observed in companies with strong precautionary measures, where the median restriction durations remain below 0.28. The greatest improvements are seen in shallow floods (<1m), where effective precautionary measures reduce the 75th percentile to 0.33, compared to 0.44 in companies without precautions. These results highlight the effectiveness of precautionary measures in reducing business restriction durations.



https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

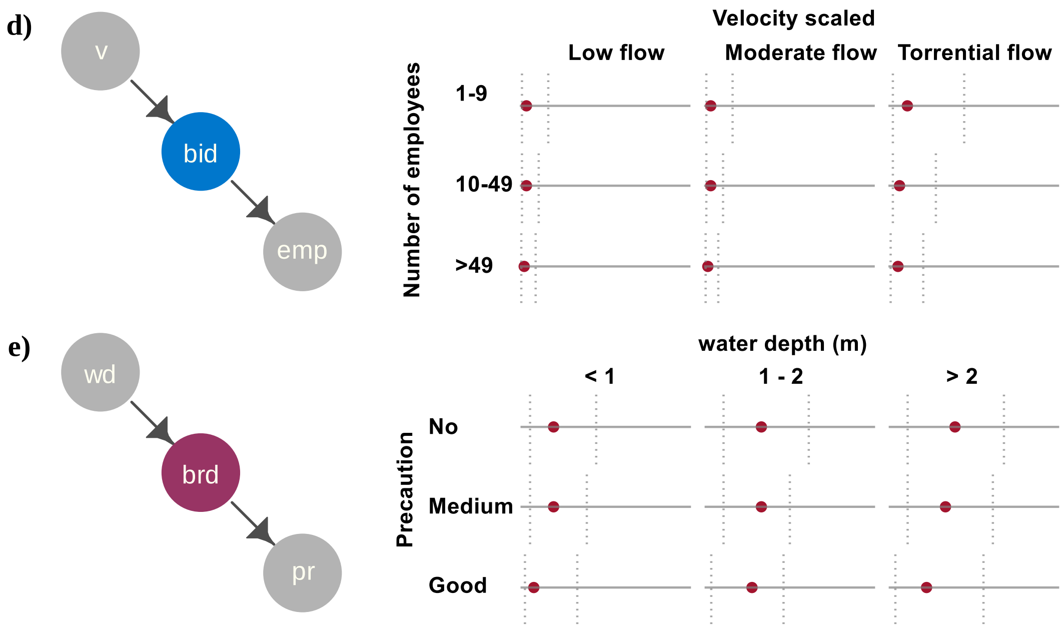


Figure 8: Predictive density plots for damage types considering Markov blankets: Relative damage to (a) buildings, (b) equipment, and (c) goods and stock, as well as the duration of (d) business interruption and (e) business restriction. The horizontal solid line represents the range of damage/duration on a scale from zero to one. Circular markers indicate the expected damage/duration (median), and vertical dotted lines represent uncertainty (25th and 75th percentiles).

4 Conclusions

The July 2021 flood in Germany highlighted the significant vulnerability of companies to extreme floods, with both direct and indirect damages resulting in substantial financial costs. A central question of this study is whether the influencing factors behind flood damage during the extreme July 2021 event differ from those in earlier floods from 2002 to 2016. Our findings indicate that core hazard related variables, including water depth, flow velocity, and contamination, remain consistent predictors of damage across different events. Similarly, company characteristics such as premises size and number of employees continue to play an important role. What sets the 2021 flood apart is the elevated importance of emergency preparedness and behavioural responses, particularly in shaping indirect damages such as business restriction duration. A novel insight from this study is the demonstrated link between knowledge about flood hazard and amount of precaution taken, highlighting its relevance in reducing business disruptions. To generalize these insights, further comparative studies using data from previous flood events and diverse regions are needed. The results underscore the critical role of emergency preparedness and risk communication during extreme events, serving as essential complements to structural protection measures that may be less effective under unprecedented conditions.

Code availability

Data analysis code is available upon request from the first author (Ravi Kumar Guntu).

Data availability

The survey data will be added into the German flood damage database, HOWAS21 (http://dx.doi.org/10.1594/GFZ.SDDB.HOWAS21), and will be partly accessible.

Author contributions

All authors contributed to the design of the study. RKG conducted the analysis and wrote the first draft. All authors reviewed and edited the final paper.

Competing interests

The author Heidi Kreibich is a member of the editorial board of Natural Hazards and Earth System Sciences.

Financial support

This research has been supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the AVOSS project (grant no. FKZ 02WEE1629C) and the KAHR project (grant no. 01LR2102I). Collection of the 2021 company data was undertaken by Section Hydrology, GFZ and Deutsche Rückversicherung AG, funded by the GFZ-HART initiative and Deutsche Rückversicherung AG.

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

References

Aria, M., Cuccurullo, C., and Gnasso, A.: A comparison among interpretative proposals for Random Forests, Machine Learning with Applications, 6, 100094, https://doi.org/10.1016/j.mlwa.2021.100094, 2021.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., Harrigan, S., Kireeva, M., Kiss, A., Kjeldsen, T. R., Kohnová, S., Koskela, J. J., Ledvinka, O., Macdonald, N., Mavrova-Guirguinova, M., Mediero, L., Merz, R., Molnar, P., Montanari, A., Murphy, C., Osuch, M., Ovcharuk, V., Radevski, I., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Živković, N.: Changing climate shifts timing of European floods, Science, 357, 588–590, https://doi.org/10.1126/science.aan2506, 2017.

BMI & BMF: Report on the 2021 flood disaster: disaster relief, reconstruction and evaluation processes, 2022.

Breiman, L.: [No title found], Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001a.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001b.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA, 785–794, https://doi.org/10.1145/2939672.2939785, 2016a.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 785–794, https://doi.org/10.1145/2939672.2939785, 2016b.

Destatis: Small and medium-sized enterprises (SME), 2003.

Destatis: Klassifikation der Wirtschaftszweige, Ausgabe 2008 (WZ 2008), 2008.

Endendijk, T., Botzen, W. J. W., De Moel, H., Slager, K., Kok, M., and Aerts, J. C. J. H.: Enhancing resilience: Understanding the impact of flood hazard and vulnerability on business interruption and losses, Water Resources and Economics, 46, 100244, https://doi.org/10.1016/j.wre.2024.100244, 2024.

Eurostat: Micro & small businesses make up 99% of enterprises in the EU, 2024.

Fisher, A., Rudin, C., and Dominici, F.: All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously, Journal of Machine Learning Research, 20, 1–81, 2019.

GDV: Facts and figures: Flood disaster "Bernd" July 2021. General Association of Insurers, Berlin, 2023.

Gerl, T., Kreibich, H., Franco, G., Marechal, D., and Schröter, K.: A Review of Flood Loss Models as Basis for Harmonization and Benchmarking, PLoS ONE, 11, e0159791, https://doi.org/10.1371/journal.pone.0159791, 2016.

Glover, F.: Future paths for integer programming and links to artificial intelligence, Computers & Operations Research, 13, 533–549, https://doi.org/10.1016/0305-0548(86)90048-1, 1986.

Goudet, O., Kalainathan, D., Caillou, P., Guyon, I., Lopez-Paz, D., and Sebag, M.: Learning Functional Causal Models with Generative Neural Networks, in: Explainable and Interpretable Models in Computer Vision and Machine Learning, edited by: Escalante, H. J., Escalera, S., Guyon, I., Baró, X., Güçlütürk, Y., Güçlü, U., and Van Gerven, M., Springer International Publishing, Cham, 39–80, https://doi.org/10.1007/978-3-319-98131-4_3, 2018.

Heckerman, D., Geiger, D., and Chickering, D. M.: Learning Bayesian networks: The combination of knowledge and statistical data, Mach Learn, 20, 197–243, https://doi.org/10.1007/BF00994016, 1995.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., and Kanae, S.: Global flood risk under climate change, Nature Clim Change, 3, 816–821, https://doi.org/10.1038/nclimate1911, 2013.

Hudson, P. and Thieken, A. H.: The presence of moral hazard regarding flood insurance and German private businesses, Nat Hazards, 112, 1295–1319, https://doi.org/10.1007/s11069-022-05227-9, 2022.

Hudson, P., Bubeck, P., and Thieken, A. H.: A comparison of flood-protective decision-making between German households and businesses, Mitig Adapt Strateg Glob Change, 27, 5, https://doi.org/10.1007/s11027-021-09982-1, 2022.

Jehmlich, C., Hudson, P., and Thieken, A. H.: Short contribution on adaptive behaviour of flood-prone companies: A pilot study of Dresden-Laubegast, Germany, Journal of Flood Risk Management, 13, https://doi.org/10.1111/jfr3.12653, 2020.

Jongman, B., Kreibich, H., Apel, H., Barredo, J. I., Bates, P. D., Feyen, L., Gericke, A., Neal, J., Aerts, J. C. J. H., and Ward, P. J.: Comparative flood damage model assessment: towards a European approach, Nat. Hazards Earth Syst. Sci., 12, 3733–3752, https://doi.org/10.5194/nhess-12-3733-2012, 2012.

Kabirzad, S. A., Rehan, B. M., Zulkafli, Z., Yusuf, B., Hasan-Basri, B., and Toriman, M. E.: Examining direct and indirect flood damages in residential and business sectors through an empirical lens, Water Science and Technology, 90, 142–155, https://doi.org/10.2166/wst.2024.202, 2024.

Kitson, N. K., Constantinou, A. C., Guo, Z., Liu, Y., and Chobtham, K.: A survey of Bayesian Network structure learning, Artif Intell Rev, 56, 8721–8814, https://doi.org/10.1007/s10462-022-10351-w, 2023.

Koks, E. E., Bočkarjova, M., Moel, H. de, and Aerts, J. C. J. H.: Integrated Direct and Indirect Flood Risk Modeling: Development and Sensitivity Analysis, Risk Analysis, 35, 882–900, https://doi.org/10.1111/risa.12300, 2015.

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 Resources Research, 43, 2005WR004691, https://doi.org/10.1029/2005WR004691, 2007.

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.

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, 102351, https://doi.org/10.1016/j.ijdrr.2021.102351, 2021.

Li, W., Wen, J., Xu, B., Li, X., and Du, S.: Integrated Assessment of Economic Losses in Manufacturing Industry in Shanghai Metropolitan Area Under an Extreme Storm Flood Scenario, Sustainability, 11, 126, https://doi.org/10.3390/su11010126, 2018.

Ma, M., Zhao, G., He, B., Li, Q., Dong, H., Wang, S., and Wang, Z.: XGBoost-based method for flash flood risk assessment, Journal of Hydrology, 598, 126382, https://doi.org/10.1016/j.jhydrol.2021.126382, 2021.

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., Bertola, M., Kemter, M., Kreibich, H., Lall, U., and Macdonald, E.: Causes, impacts and patterns of disastrous river floods, Nat Rev Earth Environ, 2, 592–609, https://doi.org/10.1038/s43017-021-00195-3, 2021.

Mohor, G. S., Hudson, P., and Thieken, A. H.: A Comparison of Factors Driving Flood Losses in Households Affected by Different Flood Types, Water Resources Research, 56, e2019WR025943, https://doi.org/10.1029/2019WR025943, 2020.

Mohr, S., Ehret, U., Kunz, M., Ludwig, P., Caldas-Alvarez, A., Daniell, J. E., Ehmele, F., Feldmann, H., Franca, M. J., Gattke, C., Hundhausen, M., Knippertz, P., Küpfer, K., Mühr, B., Pinto, J. G., Quinting, J., Schäfer, A. M., Scheibel, M., Seidel, F., and Wisotzky, C.: A multi-disciplinary analysis of the exceptional flood event of July 2021 in central Europe – Part 1: Event description and analysis, Nat. Hazards Earth Syst. Sci., 23, 525–551, https://doi.org/10.5194/nhess-23-525-2023, 2023.

Munich Re: Hurricanes, cold waves, tornadoes: Weather disasters in USA dominate natural disaster losses in 2021, 2022.

Munich Re: Climate change is showing its claws: The world is getting hotter, resulting in severe hurricanes, thunderstorms and floods, 2025.

Nafari, R. H., Ngo, T., and Lehman, W.: Development and evaluation of FLFAcs - A new Flood Loss Function for Australian commercial structures, International Journal of Disaster Risk Reduction, 17, 13–23, https://doi.org/10.1016/j.ijdrr.2016.03.007, 2016.

Narin, O. G.: Gap filling of water level time series with water area using remote sensing data: a comparative performance analysis of polynomial functions, XGBoost, Random Forest and Support Vector Machine, Hydrological Sciences Journal, 02626667.2025.2458555, https://doi.org/10.1080/02626667.2025.2458555, 2025.

Neise, T. and Revilla Diez, J.: Adapt, move or surrender? Manufacturing firms' routines and dynamic capabilities on flood risk reduction in coastal cities of Indonesia, International Journal of Disaster Risk Reduction, 33, 332–342, https://doi.org/10.1016/j.ijdrr.2018.10.018, 2019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

Pfurtscheller, C. and Vetter, M.: Assessing entrepreneurial and regional-economic flood impacts on a globalized production facility, J Flood Risk Management, 8, 329–342, https://doi.org/10.1111/jfr3.12102, 2015.

Rhein, B. and Kreibich, H.: Causes of the exceptionally high number of fatalities in the Ahr valley, Germany, during the 2021 flood, Nat. Hazards Earth Syst. Sci., 25, 581–589, https://doi.org/10.5194/nhess-25-581-2025, 2025.

Sairam, N., Schröter, K., Lüdtke, S., Merz, B., and Kreibich, H.: Quantifying Flood Vulnerability Reduction via Private Precaution, Earth's Future, 7, 235–249, https://doi.org/10.1029/2018EF000994, 2019.

Sakai, P. and Yao, Z.: Financial losses and flood damages experienced by SMEs: Who are the biggest losers across sectors and sizes?, International Journal of Disaster Risk Reduction, 91, https://doi.org/10.1016/j.ijdrr.2023.103677, 2023.

Schoppa, L., Sieg, T., Vogel, K., Zöller, G., and Kreibich, H.: Probabilistic Flood Loss Models for Companies, Water Resources Research, 56, e2020WR027649, https://doi.org/10.1029/2020WR027649, 2020.

Schoppa, L., Barendrecht, M., Sieg, T., Sairam, N., and Kreibich, H.: Augmenting a socio-hydrological flood risk model for companies with process-oriented loss estimation, Hydrological Sciences Journal, 67, 1623–1639, https://doi.org/10.1080/02626667.2022.2095207, 2022.

Seifert, I., Kreibich, H., Merz, B., and Thieken, A. H.: Application and validation of FLEMOcs – a flood-loss estimation model for the commercial sector, Hydrological Sciences Journal, 55, 1315–1324, https://doi.org/10.1080/02626667.2010.536440, 2010.

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, e0212932, https://doi.org/10.1371/journal.pone.0212932, 2019.

Sucar, L. E.: Probabilistic Graphical Models: Principles and Applications, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-61943-5, 2021.

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.

Tay, J. K., Narasimhan, B., and Hastie, T.: Elastic Net Regularization Paths for All Generalized Linear Models, J. Stat. Soft., 106, https://doi.org/10.18637/jss.v106.i01, 2023.

Thieken, A., Kreibich, H., Müller, M., and Lamond, J.: Data Collection for a Better Understanding of What Causes Flood Damage–Experiences with Telephone Surveys, in: Geophysical Monograph Series, edited by: Molinari, D., Menoni, S., and Ballio, F., Wiley, 95–106, https://doi.org/10.1002/9781119217930.ch7, 2017.

Thieken, A., Bubeck, P., and Zenker, M.-L.: Fatal incidents during the flood of July 2021 in North Rhine-Westphalia, Germany: what can be learnt for future flood risk management?, JCRFR, 2, https://doi.org/10.59490/jcrfr.2023.0005, 2023a.

Thieken, A. H., Müller, M., Kreibich, H., and Merz, B.: Flood damage and influencing factors: New insights from the August 2002 flood in Germany, Water Resources Research, 41, 2005WR004177, https://doi.org/10.1029/2005WR004177, 2005.

https://doi.org/10.5194/egusphere-2025-1715
Preprint. Discussion started: 25 April 2025

Thieken, A. H., Bessel, T., Kienzler, S., Kreibich, H., Müller, M., Pisi, S., and Schröter, K.: The flood of June 2013 in Germany: how much do we know about its impacts?, Nat. Hazards Earth Syst. Sci., 16, 1519–1540, https://doi.org/10.5194/nhess-16-1519-2016, 2016.

Thieken, A. H., Samprogna Mohor, G., Kreibich, H., and Müller, M.: Compound inland flood events: different pathways, different impacts and different coping options, Nat. Hazards Earth Syst. Sci., 22, 165–185, https://doi.org/10.5194/nhess-22-165-2022, 2022.

Thieken, A. H., Bubeck, P., Heidenreich, A., Von Keyserlingk, J., Dillenardt, L., and Otto, A.: Performance of the flood warning system in Germany in July 2021 – insights from affected residents, Nat. Hazards Earth Syst. Sci., 23, 973–990, https://doi.org/10.5194/nhess-23-973-2023, 2023b.

Trenczek, J., Lühr, O., Lukas, E., and Viktoria, L.: Damage from flash floods and floods in July 2021 in Germany. An ex-post analysis. Project report, Prognos AG, Düsseldorf, 2022.

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.

Wutzler, B., Hudson, P., and Thieken, A. H.: Adaptation strategies of flood-damaged businesses in Germany, Frontiers in Water, 4, https://doi.org/10.3389/frwa.2022.932061, 2022.

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.

Zander, K. K., Nguyen, D., Mirbabaie, M., and Garnett, S. T.: Aware but not prepared: understanding situational awareness during the century flood in Germany in 2021, International Journal of Disaster Risk Reduction, 96, 103936, https://doi.org/10.1016/j.ijdrr.2023.103936, 2023.

Zhang, Z. and Tian, H.: Hybrid imputation-based optimal evidential classification for missing data, Appl Intell, 55, 69, https://doi.org/10.1007/s10489-024-05950-9, 2025.

Zou, H. and Hastie, T.: Regularization and Variable Selection via the Elastic Net, Journal of the Royal Statistical Society. Series B (Statistical Methodology), 67, 301–320, 2005.