Supplement of Nat. Hazards Earth Syst. Sci., 26, 103–118, 2026
https://doi.org/10.5194/nhess-26-103-2026-supplement

Natural Hazards and Earth System Sciences
Open Access
EGU

Supplement of

$\mathbf{FLEMO_{flash}}$ – Flood Loss Estimation MOdels for companies and households affected by flash floods

Apoorva Singh et al.

Correspondence to: Ravikumar Guntu (ravikumar.guntu@gfz.de)

S1. Bayesian Network

We employed the 'Tabu Search' algorithm to learn the directed acyclic graph structure of the Bayesian networks from the company and private household loss data. This optimization routine searches the space of candidate Bayesian network structures for a directed acyclic graph that maximizes a predefined goodness-of-fit score. Specifically, we opted for the Bayesian Dirichlet equivalent score. Once we obtained the Bayesian network structure, we learned the conditional probability tables from the survey dataset through Bayesian parameter estimation.

To make predictions with the fitted Bayesian network, we used exact Bayesian inference by querying the conditional probability of the target variable, relative loss, conditioned on the observed predictors. For further details on Bayesian network theory, we refer to the literature (Jensen & Nielsen, 2007; Nagarajan et al., 2013). The proposed Bayesian networks were implemented in R using the packages 'bnlearn' (Scutari & Denis, 2021) and 'gRain'(HøJsgaard, 2012).

S2. Model validation

The model validation process follows the methodology outlined in Schoppa et al., (2020). All models provide probabilistic predictions rather than deterministic loss estimates. However, they do not offer analytical predictive distributions but simulated approximations in the form of samples. For each model we sampled 1000 values from the conditional response distribution and evaluated this probabilistic response in terms of accuracy, sharpness, and calibration. Within each asset dataset, we estimated the model test errors through repeated cross validation to obtain robust estimates of true model performance. We initiated 100 independent runs of ten-fold cross validation with varying, random data partitioning. In each of the tenfold cross-validation runs, every company is held out of the training set for prediction exactly once. We validate model performance for each cross-validation fold using three performance metrics:

1. The mean absolute error (MAE) for the mean of the predictive distribution. The MAE evaluates the accuracy of a point forecast and averages the absolute differences between the observed response and predicted point estimate over the number of observations.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{S1}$$

Where n is the number of observations, x_i is the predicted point estimate and y_i is the observed response for observation i.

2. The mean bias error (MBE), which quantifies model overestimation and underestimation in the mean of predictive distributions.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \tag{S2}$$

Where n is the number of observations, x_i is the predicted point estimate and y_i is the observed response for observation i. Unlike the MAE, since the absolute value of error is not taken in the MBE, it consists of both positive and negative values, and it serves as a measure of average prediction bias.

3. The continuous ranked probability score (CRPS), which is a proper scoring rule that evaluates the entire continuous distribution of a probabilistic forecast. It jointly assesses the sharpness and calibration of the predictive distribution and generalizes the absolute error. Hence the error can be compared directly to the MAE. The CRPS for one observation y_i is defined as

$$CRPS_i(F_i, y_i) = \int_{-\infty}^{\infty} (F_i(x) - 1\{y_i \leq x\})^2 dx \quad \text{(S3)}$$

Where $F_i(x)$ is the CDF of the predictive distribution $f_i(x)$ and $1\{.\}$ Is the indicator function. We compute the CRPS with an empirical CDF estimated from samples of $f_i(x)$. For details on the numerical implementation of the CRPS for simulated forecasts, we refer to the corresponding literature (Gneiting & Katzfuss, 2014; Jordan et al., 2019; Krüger et al., 2021). For the proportional response variable, $rloss$, the CRPS is defined on the interval $[0,1]$ with the optimum at 0. Note that the CRPS is calculated individually for each observation. For the comparison with the MAE, we computed the mean CRPS value in each cross-validation fold.

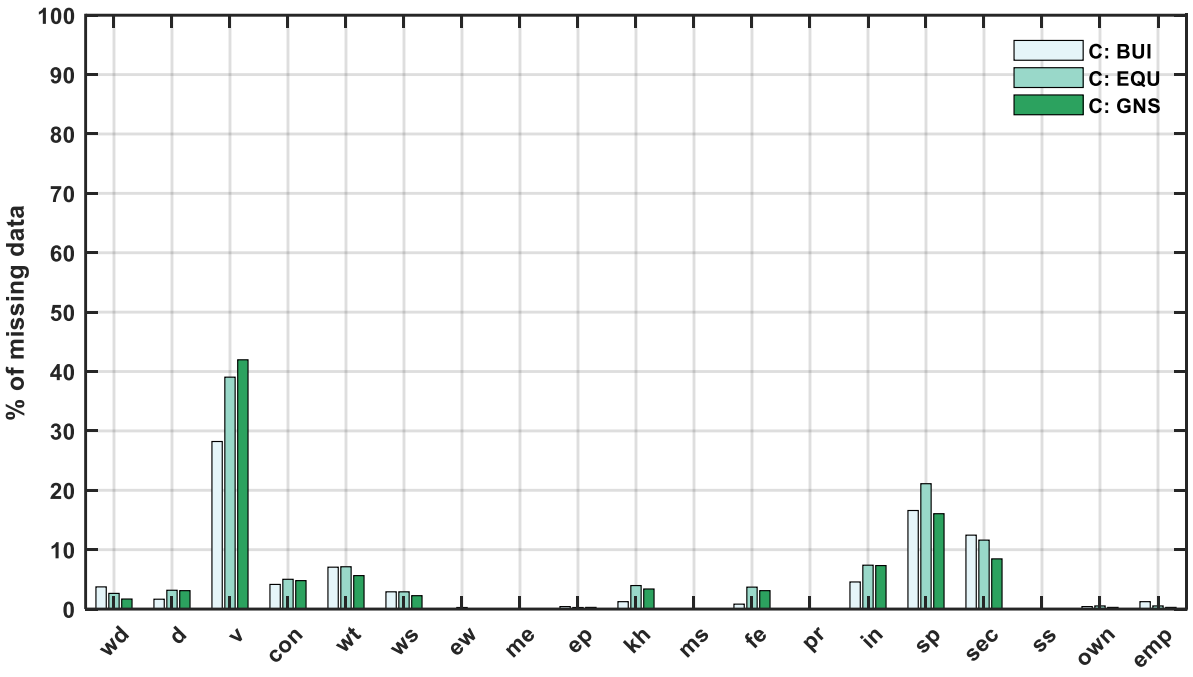

Fig. S1: Illustration of missing data in each variable for developing company damage models. For instance, in the $C{:}\,BUI$ loss category, a value of 0.10 indicates that 24 observations out of 241 are missing for the respective variable.

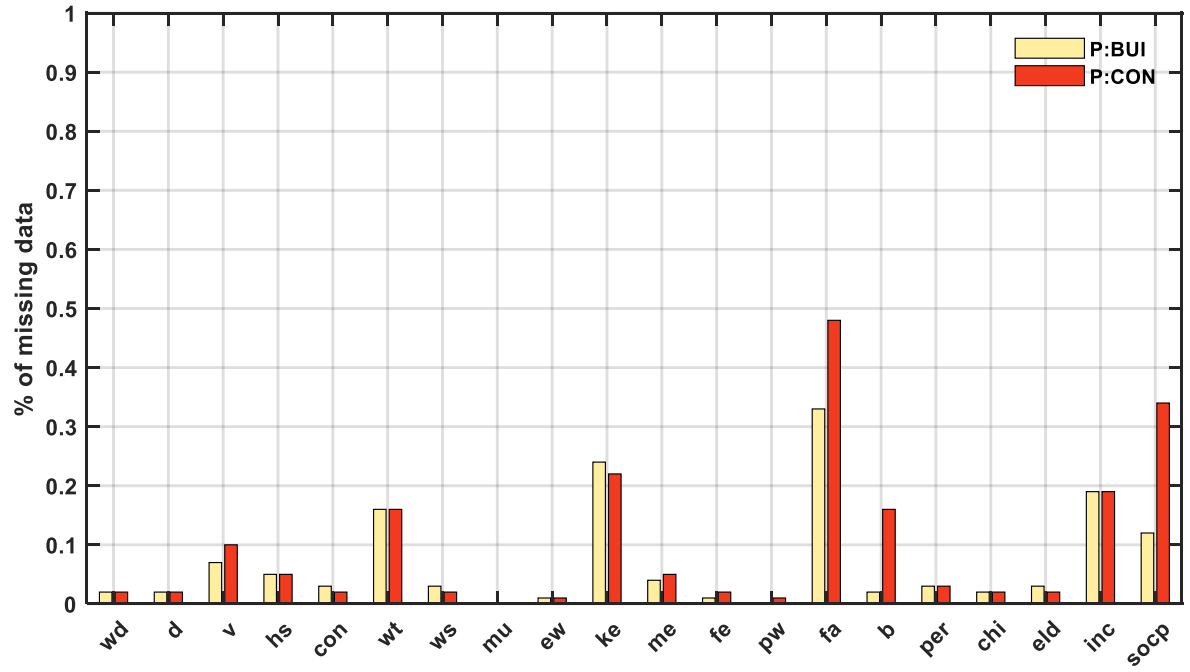

Fig. S2: Same as Fig. S1 but for private households.

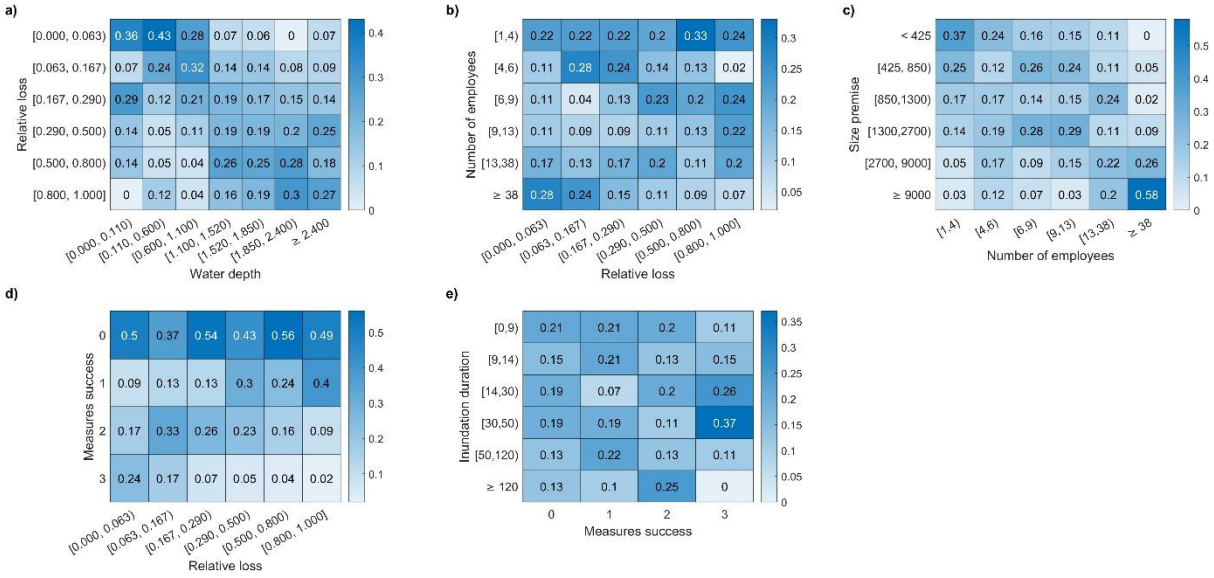

Fig. S3: Conditional probability table of C:BUI Bayesian network. X-axis denotes the parent node, while Y-axis denotes the children node within the network. Each heatmap illustrates the conditional probabilities between a child and its parent node, with increasing intensity of blue indicating higher probability values. Numerical values are displayed in each cell, and a colorbar in each subplot shows the corresponding probability scale.

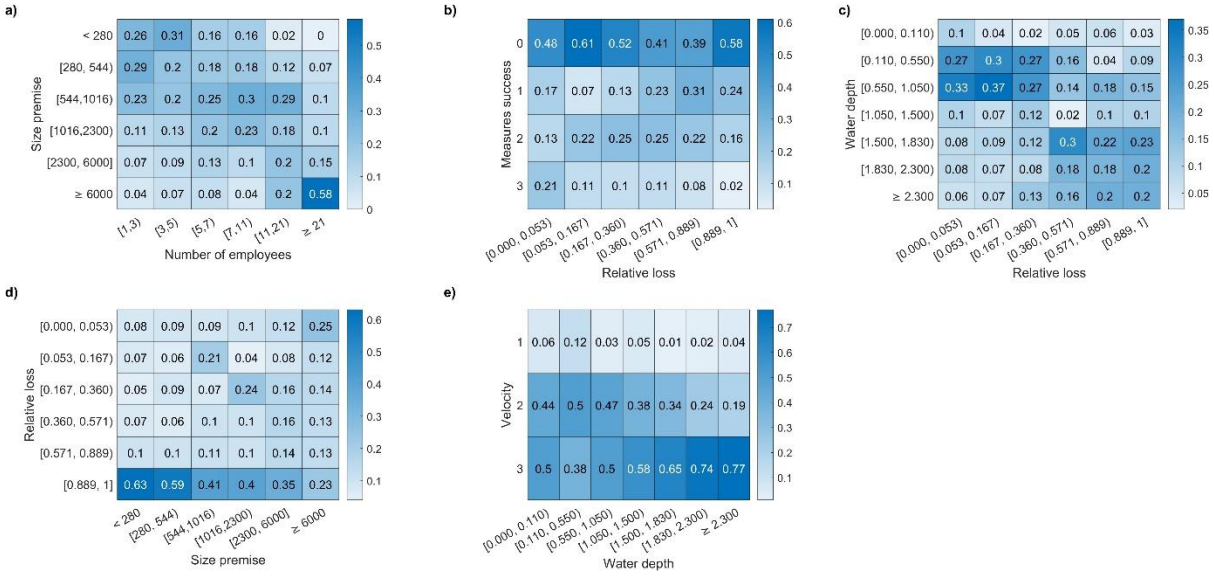

Fig. S4: same as Fig.S1, but for C:EQU

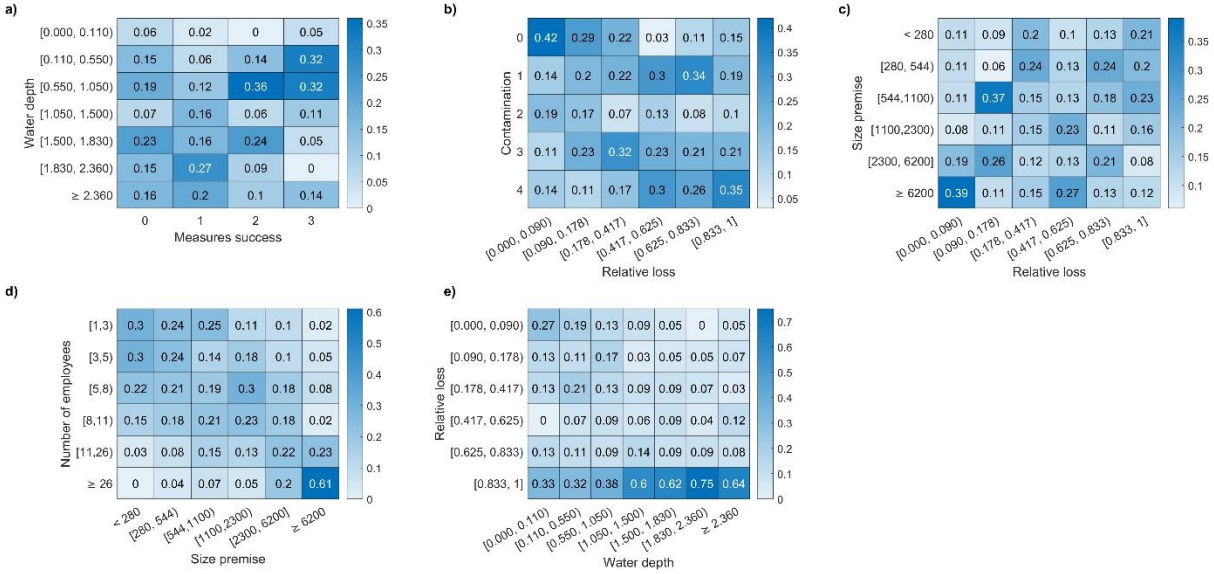

Fig. S5: Same as Fig.S1 but for C: GNS

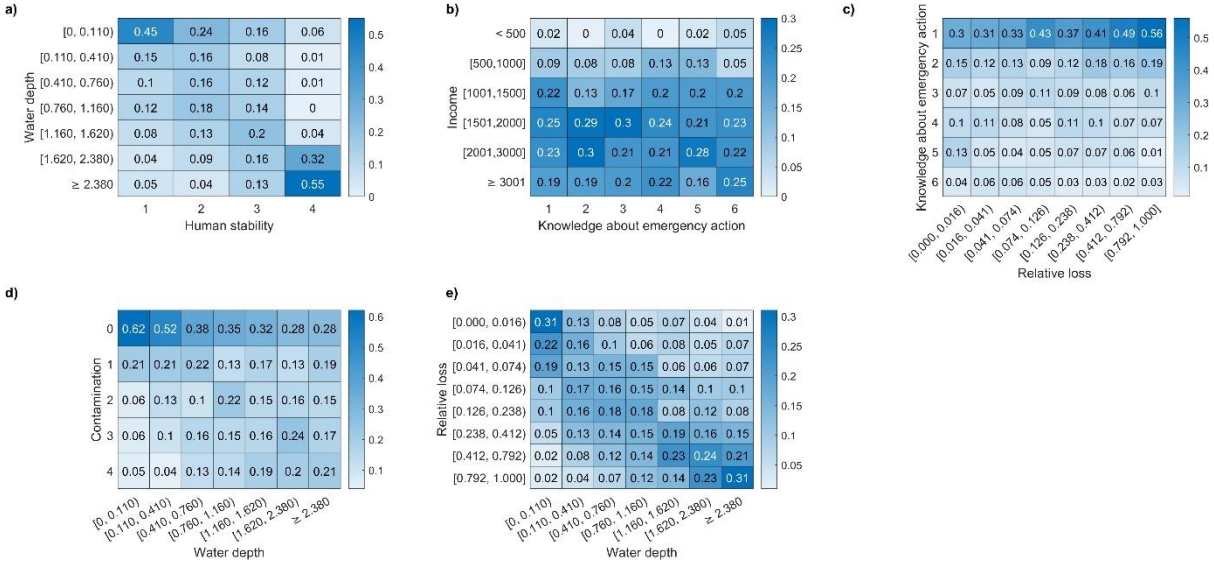

Fig. S6: same as Fig. S1, but for P:CON

Table S1: Overview of the company variables, including abbreviations, full variable names, survey questions, response options, coding, and index construction.

Predictors		**Survey question**	**Response**
wd	Water depth	At maximum water level, how high was the water above the Earth's surface on your company premises in cm?	Continuous variable
d	Inundation duration	For how many hours did water remain on the company premises?	Continuous variable
v	Velocity indicator	How strong was the water current in the immediate vicinity of your company?	• 1 – Calm/slowly flowing • 2 • 3 • 4 • 5 • 6 – Wild/violent current <u>Recoded categories (used in the analysis):</u> 1. Low flow (original categories 1–2) 2. Moderate flow (original categories 3–4) 3. Torrential flow (original categories 5–6)
con	Contamination	Did contamination from the following substances entered your company during the flood event?	*<u>Response (with multiple options possible):</u>* • Oil/Gasoline • Chemicals • Sewage • No contamination <u>Recoded categories (used in the analysis):</u> 0. No contamination 1. Sewage or Chemicals only 2. Oil/Gasoline only

			3. Oil/Gasoline + Sewage, or Oil/Gasoline + Chemicals 4. Oil/Gasoline + Chemicals + Sewage
wt	Warning lead time	How many hours before the arrival of the flash flood or heavy rainfall did the warning reach your company?	• Number of hours • No warning received
ws	Early warning source	From which source did your company receive the flood warning?	*Response (with multiple options possible):* • Loudspeaker announcements • App or SMS • Telephone call • Radio report • TV report • Newspaper report • Social media • Own research • Own observation • No warning Recoded categories (used in the analysis): 0. No warning 1. Own research 2. Contacts (employees, acquaintances, other companies, phone calls) 3. Media (radio, TV, newspaper, online, social media) 4. Official authorities (direct official warning, apps/SMS, civil protection, loudspeaker announcements, regional services)
ew	Early warning received	Did your company receive an early warning of the flood event?	0. No 1. Yes

me	Emergency measures undertaken	Were measures to reduce damage undertaken in your company before or during the flood event?	0. No 1. Yes
ep	Emergency plan	At the time of the flood event, did your company have an emergency or flood protection plan?	0. No 1. Yes
kh	Knowledge about hazard	Had this site already been flooded before? Were you aware that your company is located in a flood-prone area?	0. No 1. Yes
ms	Emergency measures success	Were measures to reduce damage undertaken in your company before or during the flood event? How effective were these mitigation measures?	• No measure undertaken • Not effective at all • Only partly effective • Mostly effective • Completely effective <u>Recoded categories (used in the analysis):</u> 0. No measure undertaken 1. Completely ineffective, 2. Partly effective, *3.* Mostly/ completely effective
fe	Flood experience	Q1: Had this company site already been flooded before the event? If yes, how many times?	<u>Number of previous floods</u>: 0. Never 1. Once 2. Twice 3. Three times 4. Four times 5. More than four times
		Q2: When was the company site last affected by a flood prior to the event? (Year)	<u>Time elapsed since the last flood</u>: 1. 25 years ago 2. 10–25 years ago 3. 5–10 years ago 4. 2–5 years ago 5. 0–2 years ago
		Flood experience was calculated from the number of previous floods (Q1) and the time elapsed since the last flood (Q2).	• If only one value (Q1 or Q2) was available, that value was used. • If both values were available, the flood experience score was

			calculated as the mean of the two.
pr	Precaution indicator	*Measures included* V1. Company insured against flood damages. V2. Heating system adjusted (converted or flood-protected). V3. Emergency plan in place. V4. Frequency of emergency drills conducted before the flood. V5. Tanks, silos, or storage facilities securely anchored. V6. Stationary or mobile water barriers installed. V7. Sensitive equipment relocated to higher floors. V8. Water-hazardous substances relocated to higher floors. V9. Use of flood-prone areas adapted to risk. V10. Air conditioning/ventilation system flood-proofed. V11. Building flood safety improved (e.g., sealing basements, strengthening stability).	Conversion: • Each measure was coded as 1 if implemented prior to the flood, 0 otherwise. • For drills, any positive frequency (≥1 per year) was coded as 1, absence as 0. Weighting scheme: • Low impact / basic preparedness (weight = 1): V1 to V4 • Medium impact / protective but limited scope (weight = 5): V5 to V8 • High impact / comprehensive protection (weight = 10): V9 to V11 Calculation of weighted score (p): $p = v1 + v2 + v3 + v4 + \big(5 \times (v5 + v6 + v7 + v8)\big) + \big(10 \times (v9 + v10 + v11)\big)$ Precaution Indicator (pr): 0. No precautionary measures 1. Medium precaution $(p: 1 - 5)$ 2. Very good precaution $(p \geq 6)$
in	Insurance	Is the company insured against flood damages before the flood event?	0. No 1. Yes
sec	Sector	Which sector does your company belong to?	1. Agriculture 2. Manufacturing 3. Trade

			4. Finance 5. Services
ss	Spatial situation	Which description best fits the spatial situation of this flood-affected company site?	1. Business premises with several buildings belonging to the company 2. Entire building fully used by the company 3. One or more floors in a building otherwise used for non-business purposes 4. Less than one floor in a building otherwise used for non-business purposes
own	Ownership	Are the buildings or rooms owned by the company or rented?	1. Owned 2. Rented 3. Partly owned / partly rented
emp	Number of employees	How many people were employed in the previous month?	Continuous variable
sp	Size premise	How large is the property on which your company is located?	Continuous variable

Table S2: Overview of the private household variables, including abbreviations, full variable names, survey questions, response options, coding, and index construction.

Predictors		**Survey question**	**Response**
wd	Water depth	At the maximum water level: How high did the water stand approximately outside the building?	Continuous variable
d	Inundation duration	For how many hours did the water remain inside the building in total?	Continuous variable
v	Velocity scaled	How strong was the water current in the immediate vicinity of your house?	0. No flow 1. Calm flowing 2. . 3. . 4. .

			5. . 6. Torrential flow
hs	Human stability	Do you think an average man could have stood upright in the flood near your house?	1. Person can stand effortlessly in calm water, 2. Should make effort to stand, 3. Person would have been swept away, 4. Too deep to stand
con	Contamination	Was your affected property contaminated by the following substances?	*Response (with multiple options possible):* • Oil/Gasoline • Chemicals • Sewage • No contamination Recoded categories (used in the analysis): 0. No contamination 1. Sewage or Chemicals only 2. Oil/Gasoline only 3. Oil/Gasoline + Sewage, or Oil/Gasoline + Chemicals 4. Oil/Gasoline + Chemicals + Sewage
ew	Early warning received	How did you become aware that the flood danger was becoming acute for you?	0. No warning received 1. Warning received
wt	Warning lead time	How many hours before the onset of flooding did the warning reach you, or did you yourself become aware of the danger?	Continuous variable
ws	Warning source	How did you become aware that the flood danger would become acute for you?	0. No warning received 1. Own observation 2. Contacts 3. Media 4. Official warning through authorities
ke	Knowledge about emergency action	Before the flood danger became acute: Did you know how you and your household could protect	1. It was completely unclear to me 2. . 3. . 4. .

		yourselves against flooding from heavy rainfall?	5. . 6. It was completely clear to me
me	Emergency measures undertaken	Did you – or someone else – take measures to reduce damages in your house?	0. No 1. Yes
mu	Number of emergency measures undertaken	Did you – or someone else – take measures to reduce damages in your house?	<u>(Nominal: 0 = No, 1 = Yes)</u> • Secured documents and valuables • Moved/secured furniture and movable items • Secured oil tanks or other containers • Pumped out or scooped water • Brought animals to safety • Moved vehicles to flood-safe place • Protected building against water intrusion • Redirected water flow on property • Received help from outside • Unplugged electronic devices • Dismantled fixed electrical installations • Shut off gas/electricity manually • Gas/electricity shut off centrally by authorities • No measure taken $\text{Score} = \text{documents} + \text{furniture} + \text{oil} + \text{pump} + \text{pets} + \text{car} + \text{building} + \text{redirect} + \text{help} + \text{unplugged} + \text{dismantled} + \text{gas}_{\text{self}} + \text{gas}_{\text{authority}}$ • Minimum = 0 (No measure undertaken) • Maximum = 13 (All measures undertaken)
fe	Flood experience	Q1: How often were you personally affected by heavy rainfall or floods before the event?	<u>Number of previous floods</u>: 0. Never 1. Once 2. Twice 3. Three times 4. Four times 5. More than four times
		Q2: When was the last time you were affected	<u>Time elapsed since the last flood</u>: 1. 25 years ago

		by a flood or heavy rainfall-related inundation? (Year)	2. 10–25 years ago 3. 5–10 years ago 4. 2–5 years ago 5. 0–2 years ago
		Flood experience was calculated from the number of previous floods (Q1) and the time elapsed since the last flood (Q2).	• If only one value (Q1 or Q2) was available, that value was used. • If both values were available, the flood experience score was calculated as the mean of the two.
pw	Precaution indicator	*Measures included* V1. I find out how to protect my house/flat against flooding. V2. I take out insurance against flood damage V3. I participate in neighborhood flood assistance. V4. I use flood-prone floors in a low-value way (adapted use). V5. I avoid valuable permanent fittings in flood-prone storeys and use water-resistant/renewable materials (adapted furniture). V6. I relocate the heating system and/or electrical supply to higher floors. V7. I change the heating system or flood-protect the oil tank. V8. I improve the safety of the building (e.g. seal basements) V9. I install stationary or mobile water barriers.	<u>Conversion:</u> • Each measure was coded as 1 if implemented prior to the flood, 0 otherwise. <u>Weighting scheme:</u> • Low impact (weight = 1): V1 to V4 • Medium impact (weight = 5): V6 to V10 • High impact (weight = 10): V4, V5 <u>Calculation of weighted score (p):</u> $p = v1 + v2 + v3 + v4 + \big(5 \times (v6 + v7 + v8 + v9 + v10)\big) + \big(10 \times (v4 + v5)\big)$ <u>Precaution Indicator (pw):</u> 0. No/Low precaution ($p < 7$) 1. Medium precaution ($7 \leq p < 25$) 2. Very good precaution ($p \geq 25$)

<table>
<tr><td></td><td></td><td>V10. I prepare for emergencies (e.g. water pumps, generator).</td><td></td></tr>
<tr><td>fa</td><td>Building footprint area</td><td>What is your estimate of the building's floor area?</td><td>Continuous variable</td></tr>
<tr><td>b</td><td>Basement</td><td>Does the building have a full or partial basement?</td><td>0. No basement
1. Partial basement
2. Full basement</td></tr>
<tr><td>per</td><td>Household size</td><td>How many people live permanently in your household, including yourself and all children?</td><td>Continuous variable</td></tr>
<tr><td>chi</td><td>Number of children</td><td>How many children under 14 years of age live in your household?</td><td>Continuous variable</td></tr>
<tr><td>eld</td><td>Number of elders</td><td>How many people in your household are older than 65?</td><td>Continuous variable</td></tr>
<tr><td>inc</td><td>Monthly net income in classes</td><td>What is the approximate total monthly net income of your household in euros?</td><td>1. < 500 €
2. 500-1000
3. 1001-1500
4. 1501-2000
5. 2001-3000
6. > 3000 €</td></tr>
<tr><td rowspan="2">$socp$</td><td rowspan="2">Socioeconomic status according to Plapp, (2003)</td><td>What is your highest educational qualification?</td><td>1. No school degree
2. Lower secondary
3. Secondary school
4. Vocational or technical qualification
5. Higher education</td></tr>
<tr><td>Living condition: Derived from ownership structure and building type</td><td>Ownership structure:
1. Tenant
2. Apartment owner
3. House owner
Building type:
1. Single-family house
2. Multi-family house
3. Semi-detached house
<table><tr><th>Ownership</th><th>Building type</th><th>Living condition</th></tr></table></td></tr>
</table>

			1 (Tenant)	2 (multiple)	1
				1 (single)	2
				3 (semi-detached)	2
			2 (Apartment owner)		3
			3 (House owner)		4
		What is the total usable living area of the house (all floors together, but without the basement)?	$living\ space = \frac{usable\ area}{household\ size}$ 1. Less than 25% 2. 25% to < 50% 3. 50% to < 75% 4. 75% or more		
		$Socp = Education + Living\ condition + Living\ space$	• Minimum value: 3 (if all indicators are at their lowest) • Maximum value: 13 (if all indicators are at their highest)		

Table S3: Prior distributions for probabilistic stage-damage function classified according to model parameters and predictors. Each prior contains a note motivating the prior choice.

Model parameters	**Predictors**	**Prior**	**Notes**
α – Intercept	-	$Student - t(3,0,10)$	Weakly informative standard prior for intercepts in brms-package.
β – regression parameter	wd	$Normal(1,1.5)$	Higher water depth causes larger flood loss.
λ – zero-and-one-inflation	-	$Beta(1,1)$	The parameter represents a probability and, hence, is constrained to the interval [0,1].
γ – conditional one-inflation	-	$Beta(1,1)$	The parameter represents a probability and, hence, is constrained to the interval [0,1].
ϕ – beta precision	-	$Gamma(0.01,0.01)$	The precision of the beta distribution has to be positive.

References

Gneiting, T., & Katzfuss, M. (2014). Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, *1*(1), 125–151. https://doi.org/10.1146/annurev-statistics-062713-085831

HøJsgaard, S. (2012). Graphical Independence Networks with the **gRain** Package for *R*. *Journal of Statistical Software*, *46*(10). https://doi.org/10.18637/jss.v046.i10

Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*. New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-68282-2

Jordan, A., Krüger, F., & Lerch, S. (2019). Evaluating Probabilistic Forecasts with **scoringRules**. *Journal of Statistical Software*, *90*(12). https://doi.org/10.18637/jss.v090.i12

Krüger, F., Lerch, S., Thorarinsdottir, T., & Gneiting, T. (2021). Predictive Inference Based on Markov Chain Monte Carlo Output. *International Statistical Review*, *89*(2), 274–301. https://doi.org/10.1111/insr.12405

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). *Bayesian Networks in R: with Applications in Systems Biology*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-6446-4

Plapp, S. (2003). Risk perception of natural catastrophes–an empirical investigation in six endangers areas in south and west Germany. *Karlsruher Reihe II–Risikoforschung Und Versicherungsmanagement*, *2*.

Schoppa, L., Sieg, T., Vogel, K., Zöller, G., & Kreibich, H. (2020). Probabilistic Flood Loss Models for Companies. *Water Resources Research*, *56*(9), e2020WR027649. https://doi.org/10.1029/2020WR027649

Scutari, M., & Denis, J.-B. (2021). *Bayesian Networks: With Examples in R* (2nd ed.). Boca Raton: Chapman and Hall/CRC. https://doi.org/10.1201/9780429347436