Probabilistic Flood Loss Models for Companies

Lukas Schoppa[1,2]*, Tobias Sieg[1,2], Kristin Vogel[2], Gert Zöller[3], and Heidi Kreibich[1]

[1] Section Hydrology, GFZ German Research Centre for Geosciences, Potsdam, Germany
[2] Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany
[3] Institute of Mathematics, University of Potsdam, Potsdam, Germany

6 May 2020

*lukas.schoppa@gfz-potsdam.de

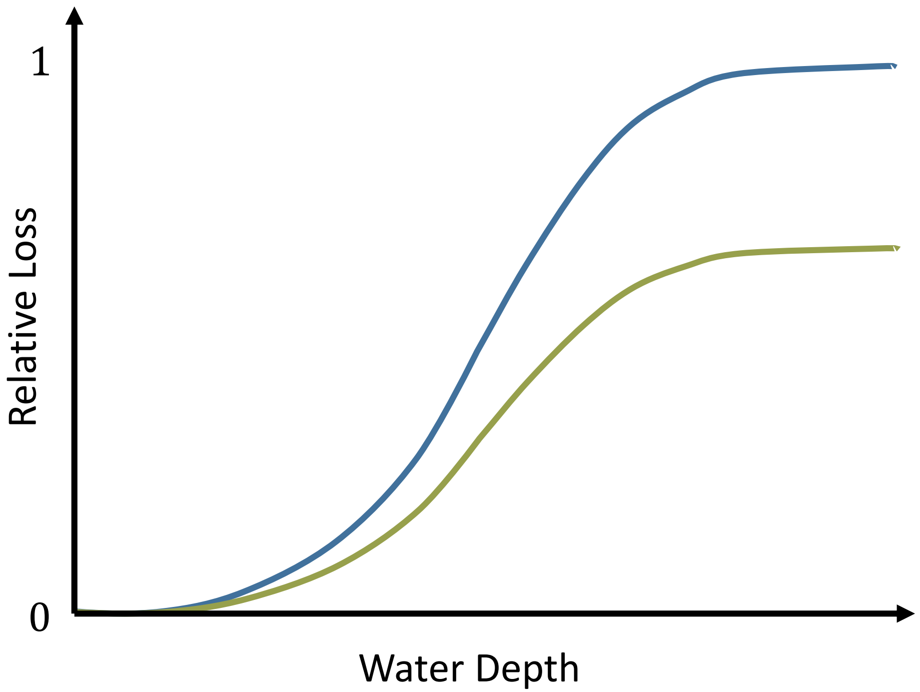
Universität Potsdam

GFZ Helmholtz-Zentrum Potsdam

Natural Hazards and Risks in a Changing World

Status Quo in Flood Loss Modeling

- Commonly involves stage-damage functions
- Most stage-damage functions are univariable models: loss depends only on water depth
- Few existing flood loss models account for uncertainty in loss estimates



New Approaches to Flood Loss Modeling

- Towards multivariable and probabilistic models; e.g.:
  - Multivariate generalized regression
  - Rule-based models
  - Decision trees
  - Bayesian networks
- Despite significant contributions of companies to total flood losses, development focused on private households

→ Multivariable, probabilistic flood loss models for companies are lacking

Study objectives

- Development and validation of three multivariable, probabilistic flood loss models for companies for the assets
  - Building (BUI)
  - Equipment (EQU)
  - Goods and stock (GNS)
- Comparing predictive performance of Bayesian networks, Bayesian regression and random forest
  1) against an established benchmark model: probabilistic square-root stage damage function (SDF)
  2) against each other

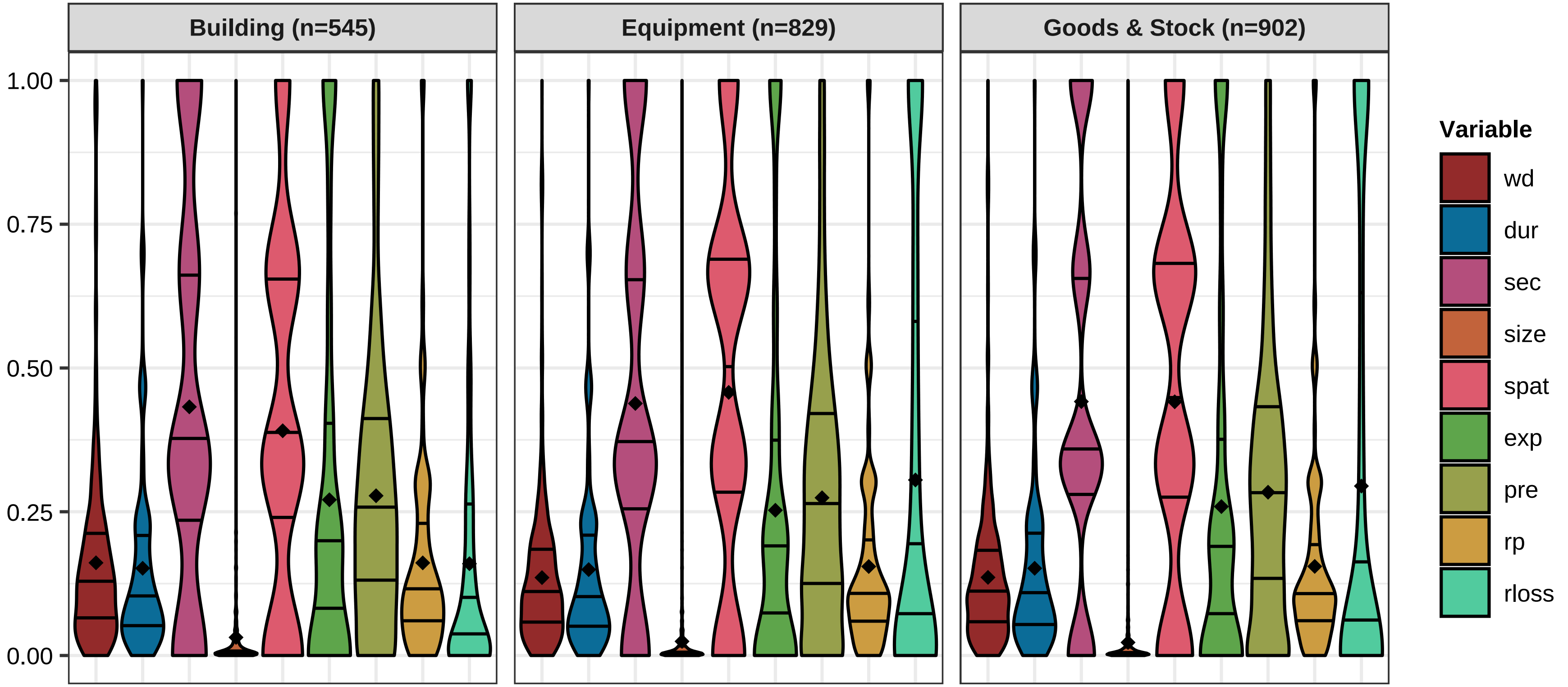
Survey Data

- Company loss data (n=1306) stems from four telephone surveys after major floods in Germany in the period 2002 – 2013
- Collected data cover flood intensity, company characteristics, warning and emergency measures, flood experience, and private precaution

Predictor (n=8) and response (n=1) variables

Variable	Abbreviation
Predictor Variables	
Water depth	wd
Inundation duration	dur
Return period	rp
Business sector	sec
Company size	size
Spatial situation	spat
Flood experience	exp
Precaution ratio	pre
Response Variables	
Relative loss building	rloss
Relative loss equipment	rloss
Relative loss goods/stock	rloss

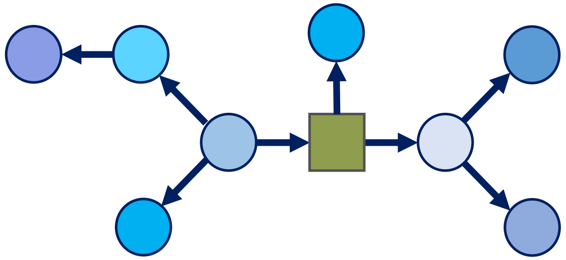
Model variables



Kernel density estimates of model data for the three company assets building, equipment, and goods and stock. For comparability, all variables are scaled from zero to one in this plot. The lines in the violin plots indicate the quartiles while the dot represents the mean. The predictor set is comprised by nominal, ordinal and continuous variables. The response variable, relative loss (rloss), is defined on the interval [0, 1] and contains significant shares of zeros and ones, which correspond to companies with no and total loss.

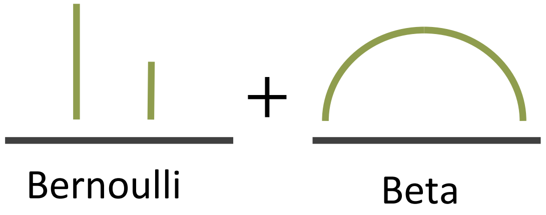
Multivariable Models

Bayesian Networks (BN)



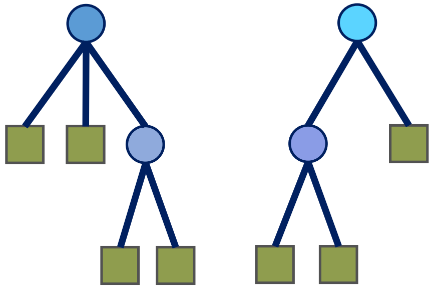
- Graphical probabilistic model
- Network structure learned from data

Bayesian Regression (BR)



- Zero-and-one inflated beta distribution
- Inflation accounts for zero and one loss cases

Random Forest (RF)

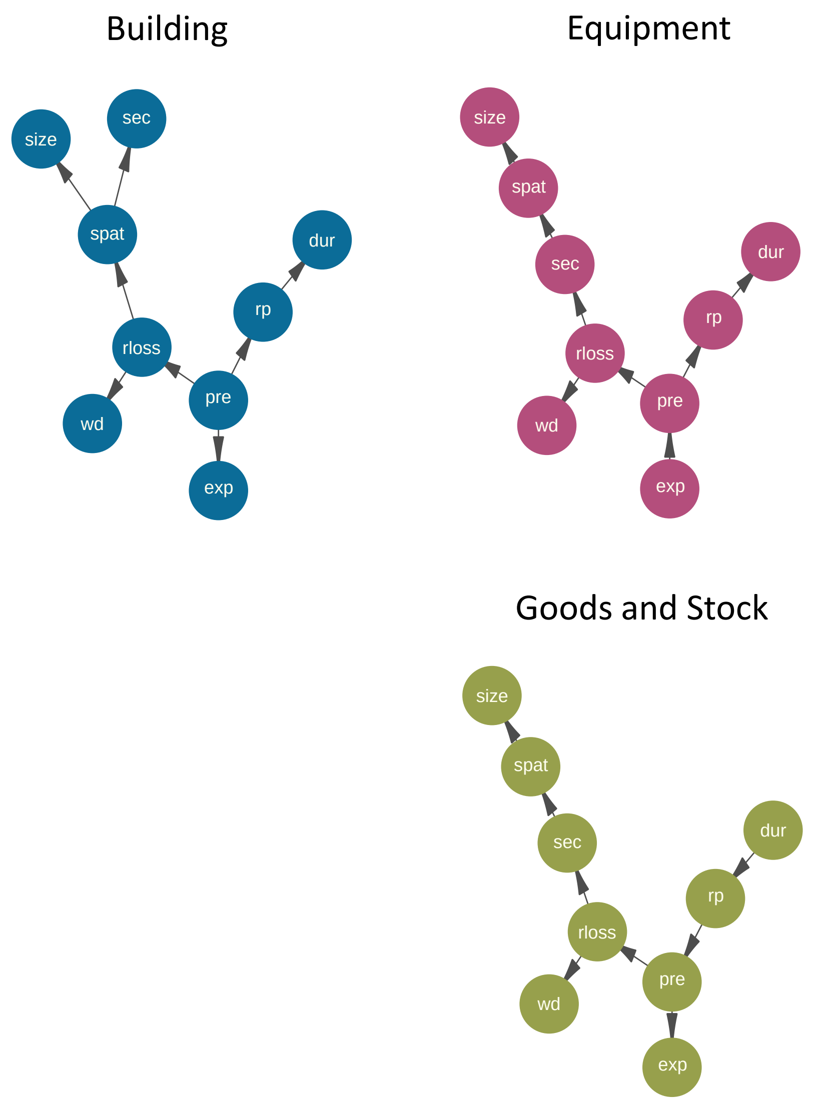


- Quantile regression forest
- Conditional inference tree algorithm

The multivariable models and the univariable stage-damage function all return probabilistic predictions of relative flood loss in the form of samples. We developed individual models for losses to the assets building, equipment, and goods and stock.

Comparing Model Fits

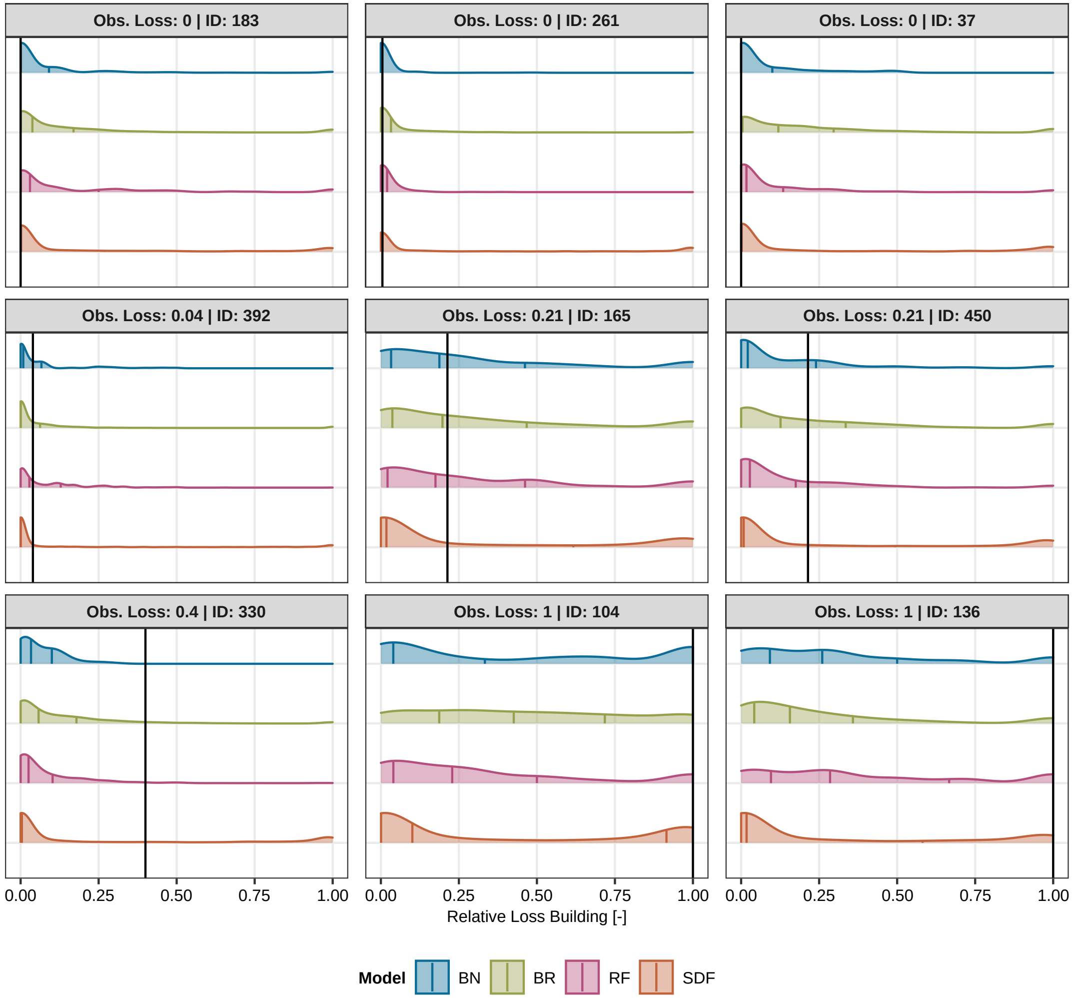
- Water depth and precaution are dominant predictor variables for all assets and models
- The predictor importance measures of the BN, BR, and RF are plausible and agree with previous findings
- The candidate models consistently identify the same predictors as most relevant
- Damage processes differ across assets: separate models are justified



Bayesian network structures, which were learned from the survey data, for the three assets. The BN structures show that the damage processes differ between building loss and losses to equipment, and goods and stock.

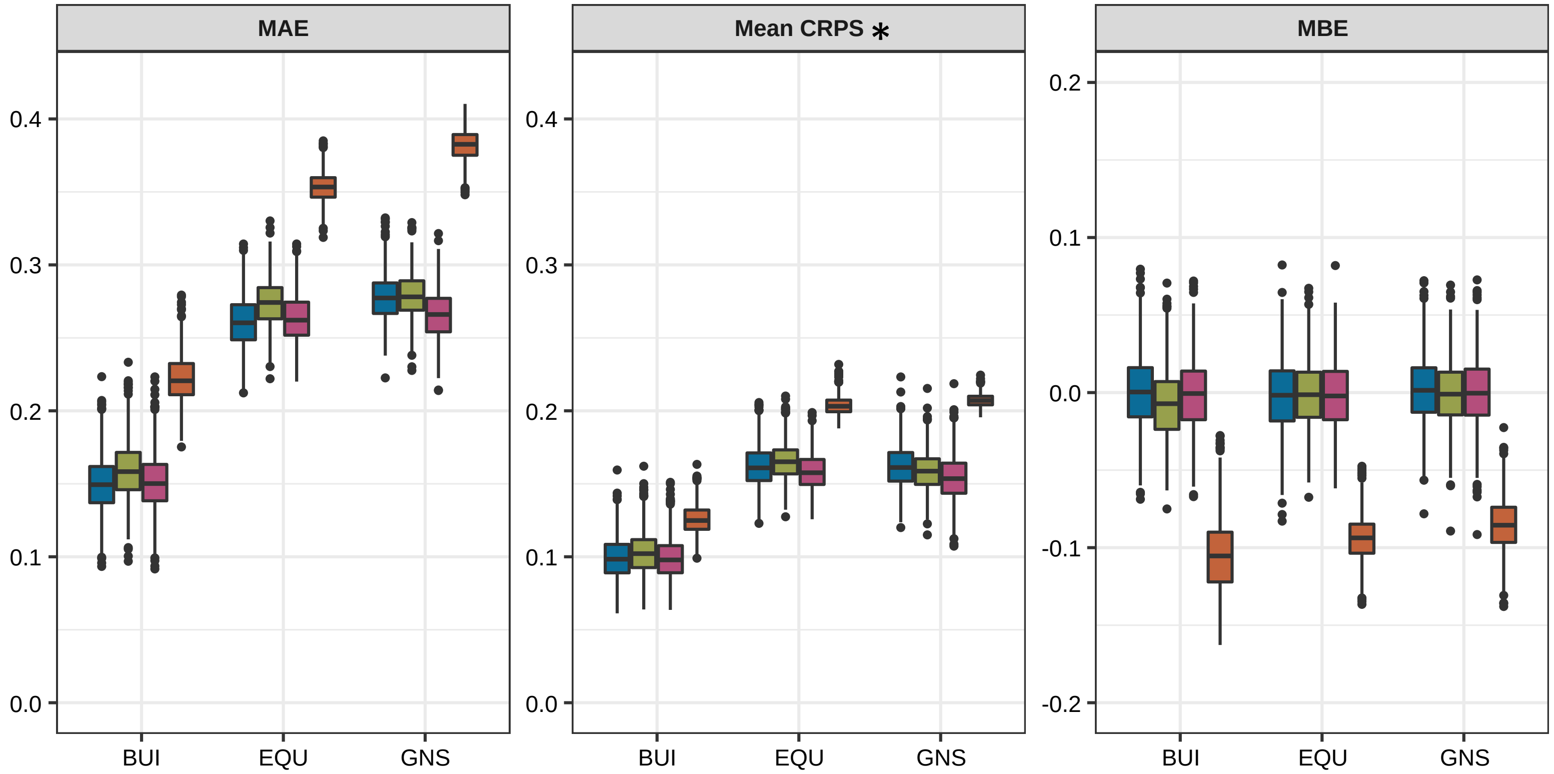
Probabilistic Predictions

- Prediction accuracy and sharpness appear to be higher for minor losses than for severe losses
- Stage-damage function (SDF) predictions are often bimodal (high density at 0/1)
- Predictive densities of the multivariable models (BN, BR, RF) are more flexible; e.g. de-/inflation of predictive densities at 0/1



Examples of predictive densities from the four models (color-coded) for the building loss of nine randomly selected companies (identified by ID). The observed loss is indicated by the black lines.

Predictive Performance



- Multivariable models outperform stage-damage functions for all assets/performance metrics
- Performance differences among multivariable models are small

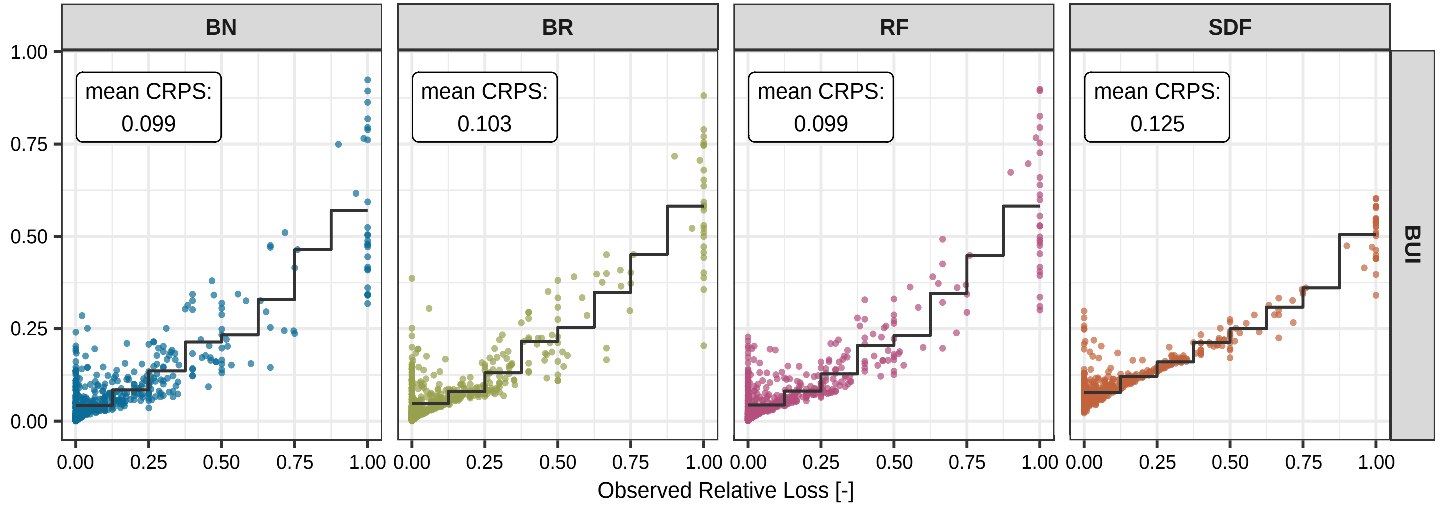
Performance metrics mean average error (MAE), mean continuous ranked probability score (CRPS), and mean bias error (MBE) for the four models (color-coded) and assets (x-axis). Each boxplot summarizes 100 repetitions of a 10-fold cross-validation with varying data partitioning.

* CRPS: generalization of the MAE; optimum at zero

Predictive Performance

- The scatter plots confirm that model errors are larger for severe losses
- The variation in the predictions and, hence, the errors is larger for multivariable models than for the stage-damage function
- Still, on average, the multivariable models perform better

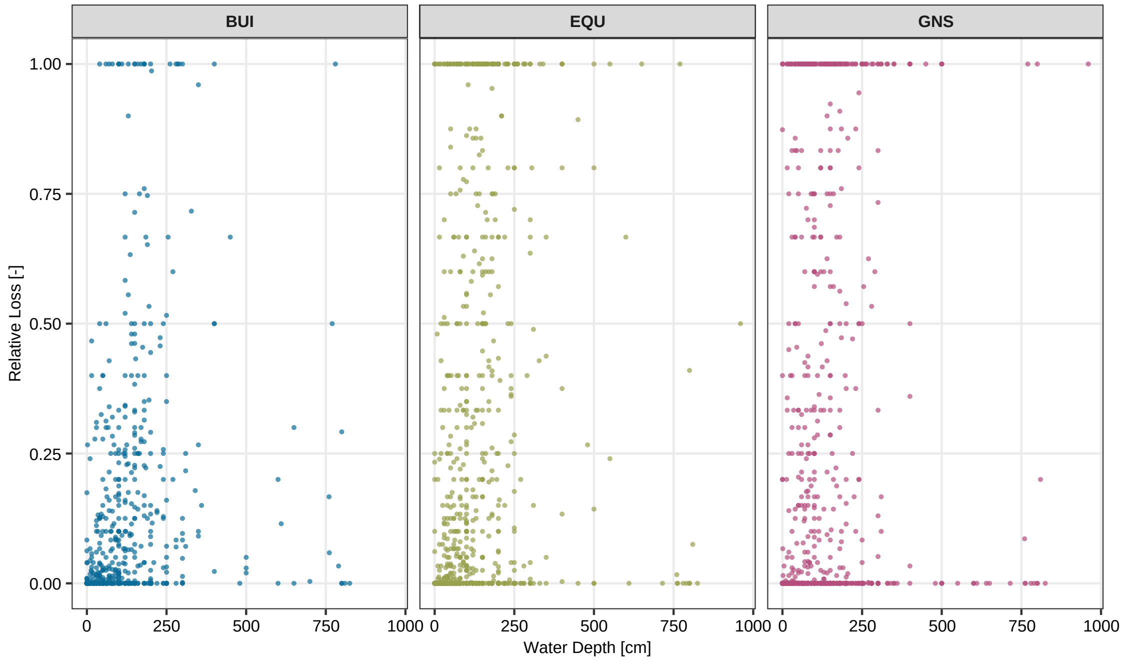
→ Bias-variance tradeoff



Scatter plots of observed relative loss versus cross-validated continuous ranked probability scores (CRPS) for all models (columns, color-coded) and building loss (BUI). Each symbol represents the prediction error incurred by the respective model for one company. The black, step-wise lines show the average CRPS in different intervals of observed relative loss. The labels in the top-left corner of each panel contain the mean CRPS over all predictions of the respective model. We observed similar predictive errors for the assets equipment, and goods and stock.

Advantages of Multivariable Models

- Water depth insufficiently discriminates between minor and major losses (see figure); additional predictors contain valuable information
- The structural complexity of the multivariable models allows for large flexibility in the predictive distributions, which is required to model the large shares of 0/1-cases in the data



Scatter plots of water depth versus relative loss for the assets building (BUI), equipment (EQU), and goods and stock (GNS). The relationship between water depth and relative loss is characterized by pronounced noise.

Conclusions

- Bayesian networks, Bayesian regression, and random forest outperform the stage-damage function
- Performance differences among multivariable models are small; model choice depends on data availability and study task
- Large predictive errors for severe losses could be caused by imbalances in the data: extreme losses are less frequent than minor losses
- Wide predictive densities suggest that the uncertainty in the predictive distributions is generally high and, hence, requires quantification through probabilistic loss models (especially for large losses)