Tracing the value of data for flood loss modelling

Kai Schröter[1,a], Stefan Lüdtke[1], Kristin Vogel[2], Heidi Kreibich[1] and Bruno Merz[1]

[1]*Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences Section 5.4 Hydrology, 14473 Potsdam, Germany*
[2]*University of Potsdam Institute of Earth and Environmental Science, 14476 Potsdam, Germany*

Abstract. Flood loss modelling is associated with considerable uncertainty. If prediction uncertainty of flood loss models is large, the reliability of model outcomes is questionable, and thus challenges the practical usefulness. A key problem in flood loss estimation is the transfer of models to geographical regions and to flood events that may differ from the ones used for model development. Variations in local characteristics and continuous system changes require regional adjustments and continuous updating with current evidence. However, acquiring data on damage influencing factors is usually very costly. Therefore, it is of relevance to assess the value of additional data in terms of model performance improvement. We use empirical flood loss data on direct damage to residential buildings available from computer aided telephone interviews that were compiled after major floods in Germany. This unique data base allows us to trace the changes in predictive model performance by incrementally extending the data base used to derive flood loss models. Two models are considered: a uni-variable stage damage function and RF-FLEMO, a multi-variable probabilistic model approach using Random Forests. Additional data are useful to improve model predictive performance and increase model reliability, however the gains also seem to depend on the model approach.

1 Introduction

Flood loss modelling is associated with considerable uncertainty which is due to an incomplete knowledge about the damaging process and to the inherent variability of quantities involved [1]. In view of large uncertainty in flood loss model predictions, the reliability of model outcomes is questionable and challenges the practical usefulness of model results; particularly when this affects the quality of decision as for instance on the investment in flood defences [2]. Therefore, it is of high importance to complement model outcomes with quantitative information about prediction uncertainty [3]. In comparison to traditional stage-damage functions which simply relate flood loss to inundation depth, multi-variable flood loss models, which take additional factors as for instance building characteristics, precaution, contamination etc. into account are an improvement to explain the variability of observed flood loss data [4, 5]. In spite of this, uncertainty ranges of flood loss predictions are still large, and thus probabilistic modelling approaches which take uncertainty into account and provide quantitative information about model prediction uncertainty are required [3].

A key problem in flood loss estimation is the transfer of models to geographical regions and to flood events that may differ from the ones used for model development [6]. Variations in local characteristics and continuous system changes require regional adjustments by updating the model with local evidence [3].

In this light, the demand for more and systematically collected data is an obvious conclusion. However, the acquisition of information on flood loss and influencing factors is elaborate and costly. Therefore, it is of relevance to assess the value of additional data in terms of model reliability improvement.

We use empirical flood loss data on direct damage to residential buildings available from computer aided telephone interviews (CATI) that were compiled after major floods in Germany [7]. This unique data base allows us to trace the changes in predictive model performance and reliability within a split-sample validation test by incrementally extending the data base used to derive flood loss models. Further, it offers the possibility to gain insight into the benefit of incorporating local evidence to a flood loss model. To study the implications of additional data on model prediction uncertainty, the analysis is conducted for probabilistic flood loss modelling approaches.

2 Data and Models

2.1 Empirical flood loss data

Empirical data of direct flood damage to residential buildings and related damage influencing variables are available from CATI that were carried out after the floods in 2002, 2005, 2006, 2010, 2011 and 2013 in Germany. A compilation of loss cases broken down for events and river basins is provided in Table 1. In total 2,254

[a] Corresponding author: kai.schroeter@gfz-potsdam.de

empirical loss cases are available. Regional focuses of loss cases are in the Elbe and Danube river basins.

Set	River basin/region	Event year	Number of loss cases
E02	Elbe	2002	661
D02	Danube	2002	286
D05	Danube	2005	116
E06	Elbe	2006	43
De10	Germany	2010	242
De11	Germany	2011	103
E13	Elbe	2013	649
D13	Danube	2013	130
De13	Germany outside Elbe of Danube	2013	24
		Total:	2254

Table 1. Loss cases for different flood events and regions in Germany.

Within the CATI campaigns a broad range of information were gathered covering damage influencing aspects related to flood impacts, building characteristics, socio economic status, precaution and early warning. From this extensive data set, 28 candidate variables were preselected to be used in a modelling context for predicting the relative loss ratio of residential buildings (*rloss*). These candidate variables were selected according to experiences from previous analyses [7, 8]. This selection of variables for flood loss modelling is narrowed further in this study based on the analysis of the out-of-bag feature importance as an indicator for the relevance of individual variables [8]. Accordingly, the variables water depth (*wst*), building value (*bv*), floor space of building (*fsb*), contamination indicator (*con*), return period of flood peak discharge (*rp*), inundation duration (*d*), precautionary measures indicator (pre), emergency measures indictor (*em*), age of interviewed person (*age*) and indicator of flood warning information (*wi*) are used to predict *rloss*. Further details about the variables are documented in Merz et al. 2013 [8].

2.2 Loss models

Two flood loss models are considered: (i) uni-variable stage-damage function and (ii) multi-variable RF-FLEMO which is based on the machine learning technique of random forests (9).

The model structure of the stage-damage function (sdf) for the estimation of *rloss* is defined as a two parameter (*a, b*) square root function of water depth (*wst*) as given in Equation 1:

$$rloss = a + b(wst)^{1/2} \qquad (1)$$

To evaluate predictive uncertainty, the sdf model is cast in a Bayesian modelling framework using a Monte Carlo Markov Chain approach for Gaussian linear regression using Gibbs sampling. Within this framework, the posterior distributions of model parameters *a* and *b* are used to sample the predictive distribution of *rloss* which describes model predictive uncertainty. Calculations are carried out using the R-package MCMCpack [10].

The multi-variable model RF-FELMO is built using the machine learning technique of random forests (RF). RF is an ensemble of Regression Trees (RT) derived by generating many bootstrap replicas of the data set and by growing a RT on each replica. RTs are tree-building algorithms for predicting continuous dependent variables [11]. They recursively sub-divide the predictor data space into smaller regions in order to approximate a nonlinear regression structure. At each split the data set is partitioned into two sub-spaces in such a way that the improvement in predictive accuracy is maximised. Bootstrapping captures the effects of data variability as one source of uncertainty in flood loss modelling [12]. The ensemble of candidate RT composing the RF represents a variety of model structures reflecting model structure uncertainty. The sample of *rloss* predictions provided by RF-FLEMO represents the prediction uncertainty of the multi-variable modelling framework. RF-FLEMO model derivation and *rloss* predictions are carried out using the R-package randomForest [13].

3 Analysis framework

The implications of additional data on model performance are investigated in a split-sample validation test framework. Basically this involves a splitting of available data into two sub-sets which are either used for model derivation or for an independent evaluation of model predictive performance also referred to as model validation [14]. Evaluation of model predictive performance and predictive uncertainty is implemented by means of a set of performance criteria for accuracy, reliability, and sharpness and prediction skill.

3.1 Split-sample validation experiments

Two validation experiments are designed. The first investigates the value of data within a gradual learning setting. In this case, the amount of data available for model derivation is incrementally increased and model performance and predictive uncertainty are evaluated using an independent split-sample of the data. This validation sample is randomly drawn from the complete sample and is not used to derive the models.

The second experiment examines the effect of using local evidence, i.e. regional specific observations to update a basic model. In this context basic flood loss sdf and RF-FLEMO models are derived using randomly selected loss cases from the complete data sample and then gradually including regional data available from specific local data sets of different CATI campaigns in the derivation of the models. Accordingly, the regional

updating of the basic models for instance in the Danube catchment includes a model derivation sequence using the data sub-set D02, next the data sub-sets D02 and D05 and finally the sub-sets D02, D05 and D13. Model predictive performance is evaluated using the observations of the sub-set D13. All experiments conducted are compiled in Table 2.

Experiment	Derivation set	Validation set
Incremental split-sample	100 increments of ca. 20 loss cases	200 loss cases randomly selected from complete sample
Regional updating Danube	• 200 random loss cases from complete sample • D02 • D02 + D05 • D02 + D05 + D13	D13
Regional updating Elbe	• 200 random loss cases from complete sample • E02 • E02 + E06 • E02 + E06 + E13	E13

Table 2. Split-sample validation experiments for incremental model derivation.

3.2 Performance criteria and scores

Model predictive performance and model predictive uncertainty are evaluated in terms of mean bias error (equation 2), mean absolute error (equation 3), quantile range of predictions (equation 4), hit rate (equation 5) and the interval score (equation 6); where Q_{50} refers to the median (50-quantile), Q_{95} to the 95-quantile and Q_{05} to the 5-quantile of the predictive distribution and O represents observed values of the predicted variable *rloss*.

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\left(Q_{50i} - O_i\right) \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Q_{50i} - O_i\right| \tag{3}$$

$$QR_{90} = \frac{1}{n}\sum_{i=1}^{n}\left(Q_{95i} - Q_{05i}\right) \tag{4}$$

$$HR = \frac{1}{n}\sum_{i=1}^{n}h_i \quad ; h_i = \begin{cases}1, & if\ O_i \in \left[Q_{95i}, Q_{05i}\right]\\ 0, & otherwise\end{cases} \tag{5}$$

$$IS = QR_{90} + \frac{1}{n}\sum_{i=1}^{n}\frac{2}{\beta}\left(Q_{05i} - O_i\right)|\{O_i < Q_{05i}\} + \frac{2}{\beta}\left(O_i - Q_{95i}\right)|\{O_i > Q_{95i}\} \tag{6}$$

Mean bias (*MBE*) and mean absolute error (*MAE*) give information about the accuracy and the exactness of model predictions. *MBE* and *MAE* are calculated using the 50-quantile (Q_{50}) of the predictive distributions for each loss case. The quantile range (QR_{90}), Hit rate (*HR*) and interval score (*IS*) enable to evaluate the sharpness, reliability and

overall performance of the interval predictions generated by the probabilistic models and are outlined in detail in [15] and [16]. QR_{90} is defined as the mean quantile range of model predictions between the 95-quantile and 5-quantile. It is an accuracy measure of the prediction interval with smaller values representing smaller ranges of the model prediction interval. *HR* is an indicator for model reliability by quantifying the ratio of observations within the prediction interval. *IS* combines sharpness and reliability scores in a single overall performance measure. The second term in Equation 6 penalises observations outside the prediction interval by the distance to either the upper or lower boundary of the interval. This penalty depends on the quantile range considered in terms of the coefficient β which is defined as $\beta = 1 - 0.9$ for the 95 to 5-quantile range as upper and lower boundaries of the prediction interval. Smaller *IS* values indicate a narrow quantile range and higher coverage of observations within the prediction interval.

4 Results and Discussion

The results of the incremental split-sample validation test are shown in Figure 1 for the sdf-model and in Figure 2 for RF-FLEMO in terms of performance traces for the different criteria and scores depending on the sample size used for model derivation.
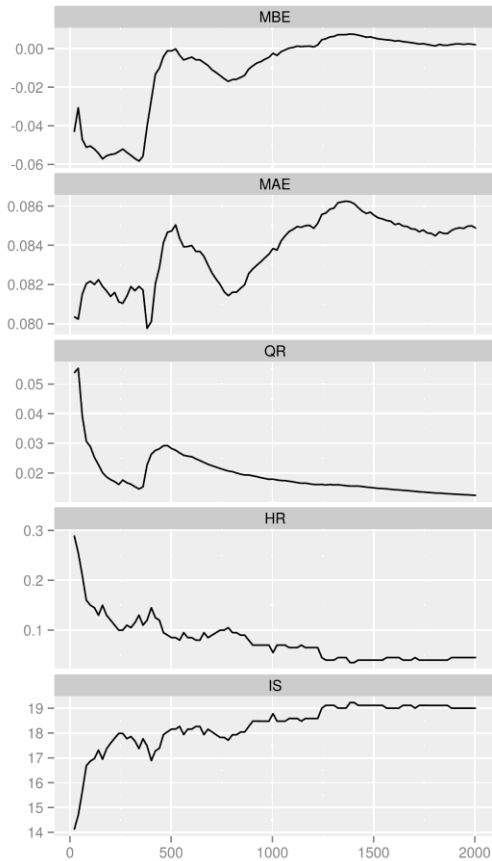
Figure 1. Performance traces for incremental split-sample evaluation of sdf model, x-axis: sample size, y-axes performance values.

For the sdf-model larger datasets tend to reduce the bias of predictions. In contrast, a slight improvement but no substantial change is seen concerning *MAE*. The QR_{90} decreases clearly with larger datasets reflecting a higher sharpness of predictions. However, this smaller prediction interval hardly covers any of the observations used for model validation which can be recognized from the decreasing *HR*. Both developments are also reflected in a continuous deterioration of the interval score *IS* for the predictive distribution. QR_{90}, *HR* and *IS* seem to be stabilizing above a sample size of ca. 1,500 loss cases. Beyond that additional data do not contribute substantially to a further improvement.
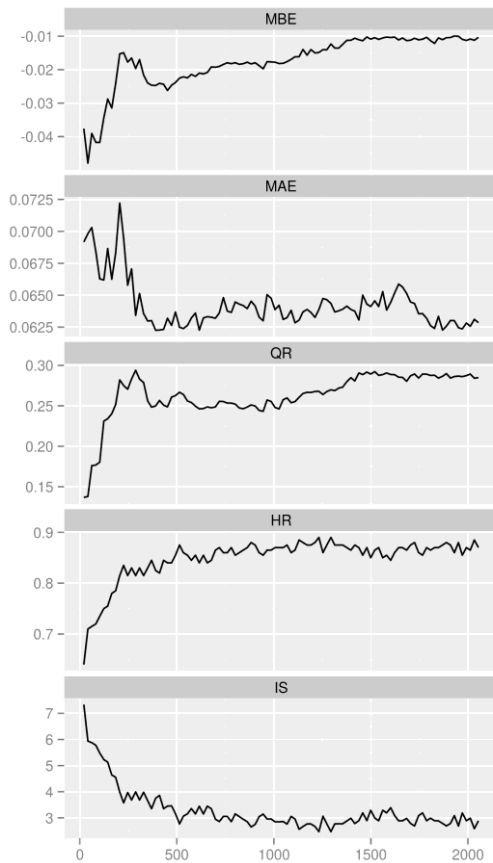

Figure 2. Performance traces for incremental split-sample evaluation of RF-FLEMO, x-axis: sample size, y-axes performance values.

For the RF-FLEMO model larger datasets reduce the bias and minimize the *MAE*. The QR_{90} of the RF-FLEMO prediction interval increases as additional data are taken into account for model derivation and indicates a larger uncertainty in the model predictions. However, this development is paralleled by a clear improvement of the reliability of model predictions which can be recognized from the *HR* score. *HR* approximates a value of 0.85 which is very close to the nominal coverage rate of 0.9 for the 95 to 5-quantile range and expresses an almost complete representation of uncertainty in the framework. The development of QR_{90} and *HR* results in a continuous improvement of the interval score *IS* for the prediction interval. As for the sdf-model the performance scores appear to be stabilizing with increasing sample size above

1,500 loss cases, yet a clear improvement of model performance is achieved already with ca. 500 loss cases indicating a more rapid learning curve for the multi-variable RF-FLEMO model in comparison to the uni-variable sdf model.

The results of the regional updating experiments in the Danube and Elbe catchments are summarized in Figure 3. Performance measures are shown for each updating step, i.e. starting from the basic model and gradually extending the data basis for deriving the models, cf. Table 2.
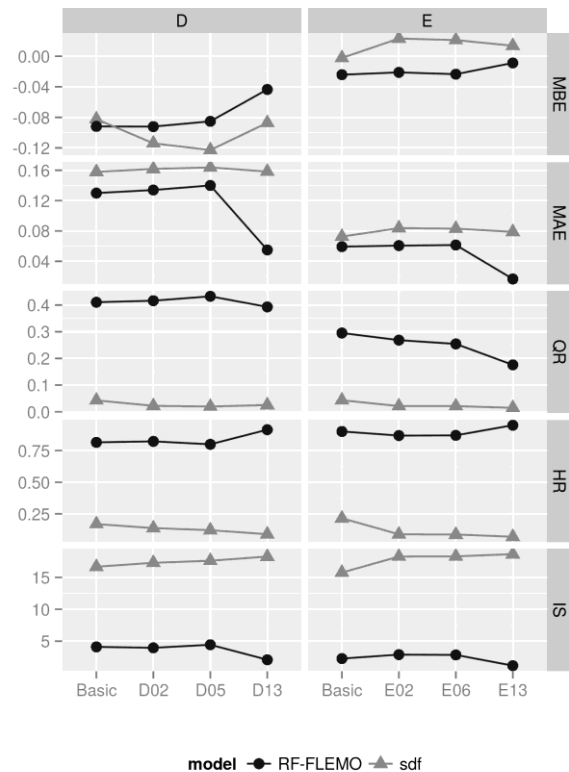
Figure 3. Model performance of sdf and RF-FLEMO models in regional updating in the Danube (D) and Elbe (E) catchments.

The improvement of model performance by including gradually local specific observations in the model derivation can be recognized for both models in terms of reduced bias (*MBE*) and smaller *MAE*. For RF-FLEMO the *HR* is continuously improved while QR_{90} is reduced which is also reflected in smaller *IS* values, indicating an improved reliability and sharpness of the predictive distribution. In contrast, similar to the incremental split-sample learning experiment for the sdf model additional regional data are not effective to increase model prediction reliability. The benefit of additional data is more obvious in the Elbe updating sequence. In the Danube catchment additional data appear to be of less value to achieve better predictions for the D13 validation set. A clear improvement is only achieved if the D13 sub-set is included in model derivation. A possible explanation for this is that in this specific sub-set many loss cases are related to inundation as a result of major dike breaches. This resulted in large inundation depths and high relative losses (not shown), which are not

represented well in the sub sets of the earlier campaigns D05 and D05.

Overall, even though the prediction uncertainty remains considerable, the performance measures achieved by sdf and RF-FLEMO model updated with local specific information are superior to the performance values achieved within the incremental split-sample validation. This supports the hypothesis that local information is useful to improve the reliability of flood loss estimation.

5 Conclusions

The value of data for the performance and reliability of flood loss predictions has been analysed within incremental split-sample and regional updating validation tests conducted for two probabilistic flood loss models. Both experiments demonstrate that additional data are useful to improve model predictive performance and reliability. The importance of local specific information to adjust loss models to regional characteristics is stressed. Comparing both model approaches the performance gain is higher for the multi-variable model. Even though predictive uncertainty ranges for RF-FLEMO are larger, which is due to the fact that with additional variables additional uncertainty sources are revealed, the modelling framework seems to more realistically represent the uncertainty. In contrast, the predictions provided by the sdf model are clearly less reliable. Apparently the uni-variable model approach is not capable to adequately describe the variability in loss observations. Therefore, one recommendation is to develop methods and to establish procedures which enable decision makers to take uncertain model predictions explicitly into account.

The effort to collect additional loss data seems to be justified as it is useful to improve model predictive performance and reliability. However, the gain also depends on the model approach applied. In this context, it is rewarding to not only focus on inundation depth but to comprehensively gather information for other loss influencing factors which enable the application of multi-variable flood loss models. As an extension of the analysis framework the connection of costs for data acquisition and monetary gains of more informed decisions should be investigated.

6 References

1. Merz B, Kreibich H, Thieken A and Schmidtke R. (2004). Estimation uncertainty of direct monetary flood damage to buildings. Nat Hazards Earth Syst Sci. **4(1)**,153–63.
2. Wagenaar D. J., Bruijn K. M. de, Bouwer L. M. and de Moel H. (2016). Uncertainty in flood damage estimates and its potential effect on investment decisions. Natural Hazards and Earth System Sciences. **16(1)**,1–14.
3. Schröter K, Kreibich H, Vogel K, Riggelsen C, Scherbaum F and Merz B. (2014). How useful are complex flood damage models? Water Resour Res. **50(4)**,3378–95.
4. Thieken A. H., Olschewski A, Kreibich H, Kobsch S and Merz B. (2008). Development and evaluation of FLEMOps - a new Flood Loss Estimation MOdel for the private sector. International Conference on Flood Recovery, Innovation and Response, London
5. Elmer F, Thieken A. H., Pech I and Kreibich H. (2010). Influence of flood frequency on residential building losses. Nat Hazards Earth Syst Sci. **10(10)**,2145–59.
6. Cammerer H, Thieken A. H. and Lammel J. (2013). Adaptability and transferability of flood loss functions in residential areas. Nat Hazards Earth Syst Sci. **13(11)**,3063–81.
7. Thieken A. H., Müller M, Kreibich H and Merz B. (2005). Flood damage and influencing factors: New insights from the August 2002 flood in Germany. Water Resources Research. **41(12)**,1–16.
8. Merz B, Kreibich H and Lall U. (2013). Multi-variate flood damage assessment: a tree-based data-mining approach. Nat Hazards Earth Syst Sci.**13(1)**,53–64.
9. Breiman L. (2001) Random Forests. Machine Learning.**45(1)**, 5–32.
10. Martin A. D., Quinn K. M. and Park J. H. (2011). MCMCpack: Markov Chain Monte Carlo in R. Journal of Statistical Software. **42(9)**, 22.
11. Breiman L, Friedman J, Stone CJ and Olshen R. A. (1994). Classification and Regression Trees. New Ed. Taylor & Francis Ltd; 368 p.
12. Breiman L. (1996). Bagging predictors. Machine Learning. **24(2)**, 123–40.
13. Liaw A and Wiener M. (2002). Classification and Regression by randomForest. R News. **2(3)**, 18–22.
14. Klemes V. (1986). Operational testing of hydrological simulation models. Hydrological Sciences Journal -Journal des Sciences Hydroliques. **31(1)**, 13–24.
15. Gneiting T and Raftery A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. Journal of the American Statistical Association. 102(477):359–78.
16. Thordarson F. Ö., Breinholt A, Møller J. K., Mikkelsen P. S., Grum M and Madsen H. (2012). Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion. Stoch Environ Res Risk Assess. **26(8)**, 1151–62.