Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L. M., Kreibich, H. (2018): Regional and Temporal Transferability of Multivariable Flood Damage Models. - *Water Resources Research*, *54*, 5, pp. 3688—3703.

DOI: http://doi.org/10.1029/2017WR022233

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE

10.1029/2017WR022233

Key Points:
- Multi-variable flood damage models can be transferred between locations, provided the training data are similar
- Flood damage collection efforts should focus on acquiring heterogeneous data, instead of collecting a large quantity of data only for a single event in one location
- There is a high potential to develop more broadly applicable flood damage models, by training multi-variable models with heterogeneous data from multiple flood events

Correspondence to:
D. Wagenaar,
dennis.wagenaar@deltares.nl

Citation:
Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L. M., & Kreibich, H. (2018). Regional and temporal transferability of multivariable flood damage models. *Water Resources Research*, *54*, 3688–3703. https://doi.org/10.1029/2017WR022233

Received 13 NOV 2017
Accepted 9 APR 2018
Accepted article online 17 APR 2018
Published online 23 MAY 2018

Regional and Temporal Transferability of Multivariable Flood Damage Models

Dennis Wagenaar[1], Stefan Lüdtke[2], Kai Schröter[2], Laurens M. Bouwer[1], and Heidi Kreibich[2]

[1]Deltares, Delft, The Netherlands, [2]GFZ German Research Centre for Geosciences, Potsdam, Germany

Abstract Reliable flood damage assessment is important for decision-making in flood risk management. Flood damage assessment is often done with damage curves based only on water depth. These depth-damage curves are usually developed based on data from a specific location and specific flood conditions. Such depth-damage curves tend to be applied outside the scope of their validity. Validation studies show that in such cases depth-damage curve are not very reliable, probably due to excluded influencing variables. The expectation is that the inclusion of more variables in a damage function will improve its transferability. We compare multi-variable models based on Bayesian Networks and Random Forests developed on the basis of flood damage data sets from Germany and The Netherlands. The performance of the models is tested on a validation sub-set of both countries' data. The models are also updated with data from the other country and then tested again. The results show that the German models (BN/RF-FLEMOps) perform better in the Netherlands than the Dutch models (BN/RF-Meuse) perform in Germany. This is probably because the FLEMOps models are based on more heterogeneous data than the Meuse models. The FLEMOps models, therefore, are better able to capture damages processes from other events and in other locations. Model performance improves via updating the models with data from the location to which the model is transferred to. The results show that there is high potential to develop improved damage models, by training multi-variable models with heterogeneous data, for example from multiple flood events and locations.

1. Introduction

Flood risk management is becoming increasingly risk-based, and flood damage estimation is therefore increasingly important in flood risk assessment studies (Merz et al., 2010). Reliable flood damage estimates support policy makers in making sound cost-benefit analyses when assessing and prioritizing risk reduction measures (Kreibich et al. 2014). An example in which flood damage estimates played an essential role is the Dutch Delta Programme, where they were used to determine the optimal protection level for the levee systems in the Netherlands (Kind, 2013; Van der Most et al., 2014).

Flood damage models typically estimate damage based on water depth and average building value for different types of buildings. When validated, such flood-damage models often perform poorly (e.g., Jongman et al., 2012) and there exists a lot of unexplained variation among damage functions in the literature (Gerl et al., 2016; Wagenaar et al., 2016). Merz et al. (2004) showed that water depth alone explains only a part of the variation among flood damage observations. Therefore, the most likely reason for the differences between damage functions is that these models contain implicit assumptions about variables not included (Wagenaar et al., 2016). Examples of such variables are: Flood duration, flow velocity, precautionary measures on household level, contamination of the flood water and household size. Including other additional variables in flood damage models can substantially improve the reliability of flood damage estimates (Kreibich et al., 2017; Schröter et al., 2014; Wagenaar et al., 2017). Thieken et al. (2008) and Kreibich et al. (2010) found that contamination and precautionary measures are important for predicting the flood damage in Germany, and introduced rule based loss estimation models (FLEMOps, FLEMOcs) with correction factors to adjust for these variables. Applications and validations of these models on the micro- and meso-scale are described by Seifert et al. (2010) and Falter et al. (2015, 2016). Merz et al. (2013) introduced tree based models to learn multi-variable damage models from data. Vogel et al. (2014) introduced the application of Bayesian Networks for the same purpose. These multi-variable flood damage models perform better than traditional flood damage models, particularly in a spatial and temporal transfer setting (Schröter et al., 2014).

Detailed data on flood damages are only rarely recorded, for instance after some flood events via surveys in Germany, France and Italy (Kienzler et al., 2015; Molinari et al., 2014; Poussin et al., 2014). Thus, for low-probability events, data often is not available or is outdated. In the application of flood damage models, a spatial and/or temporal transfer of models is therefore often necessary (Cammerer et al., 2013). For example, the Dutch standard model for flood damage estimation SSM2015 is based on transferred damage functions in both space (other countries) and time (1953 data) (De Bruijn et al., 2014). Other examples of such studies that transfer damage models are: Dahm et al. (2017), Tollenaar et al. (2016), Bouwer et al. (2017). This however can easily lead to errors, as the depth-damage functions are often not valid for the area to which they are transferred.(Meyer et al., 2013; Papathoma-Köhle et al., 2011) For transferability of flood damage models, it is therefore advantageous to describe the complexity of the damage process with multiple variables. A larger number of variables in the model could account for some of the implicit assumptions in the flood damage model and therefore make the model better transferable to other areas. Schröter et al. (2014) found that a complex multi-variable flood damage model outperforms simple models, when both are transferred to other areas. However, such comparisons are rare, and so far no study has compared the transfer of multi-variable damage models between two countries with different types of data sets.

The aim of this study is to test how well multi-variable models perform in a temporal and spatial transfer, using data originating from different countries with different collection methods. We used (1) A German data set based on telephone interviews conducted after several different flood events in Germany (Kienzler et al., 2015; Thieken et al. 2016) and (2) a data set from The Netherlands based on compensation data from the 1993 Meuse flood in the Netherlands (Wind et al., 1999). Both data sets are used to train damage estimation models based on Random Forests and Bayesian Networks which are then verified using the other data set. This study will focus entirely on residential damages relative to the building value.

2. Data and Methods

2.1. Data Sets

2.1.1. German Flood Damage Data

The German data set contains flood damage data collected through surveys after floods in 2002, 2005, 2006, 2010, 2011 and 2013. Each flood event is briefly described at the end of this section. Four surveys were carried out, using computer-aided telephone interviews, with private households which had been affected by the 2002, 2005/2006, 2010/2011 and 2013 floods in Germany. Computer-aided telephone interviews were undertaken by polling institutes with the VOXCO software package (www.voxco.com). The person with the best knowledge of the flood damage was interviewed. The surveys contained about 180 questions addressing a broad range of topics: flood impact (e.g., water depth in highest affected floor, perceived flow velocity, flood duration at building), flood warning, emergency measures, evacuation, cleaning up, building characteristics (e.g. building type and age, floor area for living, footprint area of building, availability of basement), damage to household contents and building, recovery, precautionary measures (e.g. which measures were implemented before the flood), flood experience (e.g. how many floods experienced before event and how long ago, experience with flood damage), and socio-economic variables (e.g. household size). To avoid errors as much as possible, only meaningful answers were accepted by the system. Wherever possible, answers were cross-checked. In case of conflicting answers, the interviewee was informed about a contradiction and prompted to clarify the situation. On the basis of the information provided in the survey, indicators for flow velocity (score from 0 = still to 3 = high velocity), precautionary measures (score 0 = no measures undertaken to 38 = many, efficient measures undertaken) and flood experience (score 0 = no experience to 9 = recent flood experience) were developed as described in Thieken et al. (2005). On the basis of information provided about water depth and building characteristics, the water depth relative to the ground level was calculated, with negative values in case of basement flooding only. The relative building and contents damage were calculated by dividing the actual damage as given in the survey by the building and contents values (replacement costs as at the event), which were estimated by using valuation methods of the insurance industry (Dietz, 1999).

Other data were also used or collected in addition to the survey. Information from affected communities, flood reports, press releases, as well as with the help of flood masks derived from satellite data (DLR, Centre for Satellite Based Crisis information, www.zki.caf.dlr.de) were used to compile lists of affected streets. These provided the basis for generating property-specific random samples. Return periods of the flood at the

Table 1
Overview of the Variables Used in This Study

<table>
<tr><th rowspan="2">Abbreviation</th><th colspan="2">Variable</th><th rowspan="2">Unit</th></tr>
<tr><th>Dutch data set</th><th>German data set</th></tr>
<tr><td>rsd</td><td colspan="2">Relative building damage[a,b]</td><td>Local currency</td></tr>
<tr><td>rcd</td><td colspan="2">Relative content damage[a,b]</td><td>Local currency</td></tr>
<tr><td>wdf</td><td colspan="2">Water depth relative to floor[a,b]</td><td>meter</td></tr>
<tr><td>bt</td><td colspan="2">Building type[a,b]</td><td></td></tr>
<tr><td>fa</td><td colspan="2">Footprint area of the building[b,c]</td><td>Square meter</td></tr>
<tr><td>wdt</td><td>Water depth relative to DEM[d]</td><td></td><td>cm</td></tr>
<tr><td>bs</td><td>Basement[a,b]</td><td></td><td>1=Yes, 0=No</td></tr>
<tr><td>hs</td><td>Household size[a,b]</td><td></td><td>number</td></tr>
<tr><td>fv</td><td>Flow velocity[b,d]</td><td></td><td>Estimated from score to m/s</td></tr>
<tr><td>ba</td><td>Building age[b,c]</td><td></td><td>Year</td></tr>
<tr><td>fal</td><td>Floor area for living[b,c]</td><td></td><td>Square meter</td></tr>
<tr><td>fd</td><td colspan="2">Flood duration[b,d]</td><td>Hour</td></tr>
<tr><td>rp</td><td colspan="2">Return period[b,d]</td><td>Year</td></tr>
<tr><td>fe</td><td></td><td>Flood experience[b]</td><td>Score</td></tr>
<tr><td>pre</td><td></td><td>Precautionary measures[b]</td><td>Score</td></tr>
</table>

[a]WL Delft, 1994 (Meuse). [b]Flood event surveys (Germany). [c]Basisregistraties Adressen en Gebouwen (BAG), version 2011 (Kadaster website) (Meuse). [d]2-D flood simulation data using WAQUA (Meuse).

affected residential buildings were estimated on basis of the annual maximum series of discharge of all gauges in the study areas using extreme value statistics as described by Elmer et al. (2010).

The variables used for this study are provided in Table 1. Further details about the surveys and the data processing are published by Thieken et al. (2005, 2017) and Merz et al. (2013). In total, data of 4,368 households are contained in the data set, with the following distribution across events: August 2002: 1,697 households; August 2005: 305 households; April 2006: 156 households; August 2010: 349 households; January 2011: 209 households; June 2013: 1,652 households. Cases with missing values were excluded for the analyses, resulting in 1.456 complete records for building damage and 1,324 for contents damage.

In August 2002, a cyclonic depression from the southern direction led to an extreme flood event in Germany, Austria, the Czech Republic and Slovakia. Record-breaking precipitation occurred; for instance 312 mm within 24 hours at the gauge Zinnwald-Georgenfeld in the Ore Mountains, Germany (Engel, 2004; Ulbrich et al., 2003). Highly dynamic floods occurred in the Ore Mountains, e.g. at the rivers Mulde, Weieritz, Schwarze Elster. Discharge return periods were estimated to be about 150–200 years at the Dresden gauge on the river Elbe, 200–300 years at the Mulde River, and 100–300 years at the Regen River, a left-bank tributary of the Danube (IKSE (International Commission for the Protection of the Elbe), 2004; Ulbrich et al., 2003). The flood caused 21 fatalities and overall financial damages of 11.6 billion Euros (price level 2002) in Germany (Thieken et al., 2006).

In August 2005, a considerable flood affected the German part of the Danube catchment. Cyclone 'Norbert' induced prolonged rainfall with notably high amounts within 12 to 24 hours in Switzerland, northern Italy, Austria and southern Germany, such as 216 mm in 24 hours in Balderschwang, Germany (LfU (Bavarian Environment Agency), 2007). The alpine foothills were affected by flash floods. Inundations occurred both along the river Danube and its southern tributaries. Return periods were classified to less than 100 years at the Iller, Schmutter, Amper, Inn and Isar rivers and to 20 to 50 years at the rivers Lech, Loisach and Mangfall. At the Danube River, the highest return periods occurred at the cities of Ingolstadt and Kelheim in the range of 20 to 50 years (LfU (Bavarian Environment Agency), 2007). The total economic damage was estimated to be about 175 million Euros (price level 2005) in Germany (Kron, 2009).

The flood in the spring of 2006 occurred mainly in the Elbe catchment. High amounts of water were stored as snow at the beginning of 2006 in the upper Elbe catchment (Korndörfer et al., 2006). At the end of March, heavy rainfall occurred and temperatures rose rapidly from 5 to 15°C leading to a complete snowmelt also in the upper parts of the middle hills (BfG (Federal Institute of Hydrology), 2006). At the Dresden gauge the return period was estimated to 15 years (Kreibich & Thieken, 2009). But the flood situation downstream of the Havel confluence was comparable to or even worse than in August 2002. For instance, at the Neu Darchau gauge, the flood discharge of 3,600 $m^3\ s^{-1}$ in 2006 was the second highest in 100 years and exceeded the 2002 flood discharge of 3,400 $m^3\ s^{-1}$ (BfG (Federal Institute of Hydrology), 2006). The total resulting damage in Germany was estimated at 120 million Euros (price level 2006) (Kron & Ellenrieder, 2008).

Three heavy rainfall events in August and September 2010 resulted in extreme floods in the Odra and Elbe catchments (Walther et al., 2013). The flood situation was aggravated significantly due to the breach of the dam Niedow at the Witka River, which is a tributary of the river Lausitzer Neiße, on 7 August (Jelonek et al., 2010). In the upper parts of the Schwarze Elster and Spree catchments, the highest peak flows occurred at the beginning of August with return periods of up to 500 years at the Spree and up to 200 years at the Schwarze Elster. At their lower reaches, the highest flows occurred at the end of September with return periods of 50 to 100 years (Walther et al., 2013). Particularly high damage occurred in the upper reaches of the Lausitzer Neiße and Spree as well as at the Mandau River. The total resulting damage in Germany was reported to be 839 million Euros (price level 2010) (EC, 2014).

Processes leading to flooding in January 2011 were comparable with the flood in 2006. Due to massive snowfall in winter a lot of water was stored as snow. A temperature increase and heavy rainfall led to snow melt and a first increase in river discharges between 5 and 6 January 2011. In the following days, between 12 and 14 January 2011, large-scale, intense rainfall fell on already saturated soils which led to a second flood wave with water levels above the flood warning levels at many gauges (Axer et al., 2012). Nearly all large catchments in Germany were affected, e.g. the catchments of the Rhine, the Danube, the Weser and the Elbe (Axer et al., 2012). Particularly high discharges occurred at the rivers Main and Saale and in the upstream part of the Weser catchment. The total damage was estimated to be more than 100 million Euros in Germany (price level 2011) (Axer et al., 2012).

The flood in June 2013 was mainly driven by the combination of high catchment wetness due to a strong rainfall anomaly during the month of May and spatially extended high but not extraordinary precipitation (Merz et al., 2014; Schröter et al., 2015). Nearly all main river basins in Germany were affected, but particularly severe flooding occurred along the Danube river in the federal state of Bavaria and along the Elbe River and its tributaries Saale and Mulde in the federal states of Saxony and Saxony-Anhalt (Schröter et al. 2015). In Passau, the highest water level since 1501 was observed, due to the superposition of the flood waves from the Inn and Danube rivers (Blöschl et al., 2013; Deutsches Komitee Katastrophenvorsorge e.V. (DKKV), 2015). Due to the large spatial extent of flood peaks with high magnitudes, the June 2013 flood was in hydrological terms the most severe flood in Germany at least for the last six decades (Schröter et al., 2015). The flood caused 14 fatalities and overall financial damages of about 8 billion Euros (price level 2013) (Thieken et al., 2016).

2.1.2. Dutch Flood Damage Data

The Dutch data set is based on the Meuse River flood in December 1993, in the province of Limburg near the German and Belgian border. This was a river flood in an area that was protected by its natural elevation rather than dikes, which is uncommon in the Netherlands. Forecasting of water levels in the Meuse River was at that time relatively difficult, due to the limited ability to predict precipitation in the Meuse basin and the quick response of the river (Wind et al., 1999). Therefore the local authorities were caught by surprise, and there was insufficient time to warn the public (Wind et al., 1999).

The flood caused 254 million Guilder (price level 1993) in damages within the Netherlands, which is about 180 million Euros (price level 2016). The flood event inundated about 180 km^2 of land, which is 8% of the Province of Limburg. About 32% of the damage was to residential buildings and content, which are the damage categories considered in this study.

After the Meuse flood of 1993 the Dutch national government compensated the damage. The damage was collected by insurance experts working for a governmental organization called "Stichting Watersnood 1993". Shortly after the compensation the data were shared with WL Delft (now Deltares). The data were used in 1994 to create a flood damage model for the Meuse basin, in order to inform decision makers with respect to required flood protection works. WL Delft also applied several manual adjustments to the data set (WL Delft, 1994) because the structural damage to rental houses was incomplete. What these adjustments were exactly is unknown. The only data set now available is the data set reported in the WL Delft study. More information about the background of these data and adjustments are reported in WL Delft (1994) and Wagenaar et al. (2017).

The data set contains 4,398 complete damage records. Apart from the total flood damage, the data set also contains the individual damages for structural building and content damage. Besides the damage variables the data set also contains variables on water depth relative to floor level, household size, whether the house is detached and whether the house has a basement.

Wagenaar et al. (2017) added the following additional variables to this data set, based on 2D flood simulations for the 1993 event: Water depth, flow velocity, flood duration and flood return period at the building location. Based on the Cadaster data the following variables were added: building age, building footprint area and building floor area. These variables were added based on the location of the building and there was significant uncertainty in this process because the building location was only available as a 6 digit postal code. More information about this is provided in Wagenaar et al. (2017). Table 1 shows the variables in the Dutch data set.

2.1.3. Comparison and Harmonization of Variables

The compilations of the German and the Dutch data sets differed in respect to flood events and data acquisition methods. The German data were collected via surveys after six flood events, including the two most severe floods in Germany at least since 1950 (Schröter et al. 2015), while the Dutch data were collected by insurance experts during the compensation process of one flood event at the Meuse River. Thus, the German data cover a substantially larger variety of damage processes, including extreme situations. This is also reflected in the range and densities of the variable values, as shown for three examples in Figure 1. The variables in the German data set show a considerably larger spread and often cover a larger range of values. For example, the German data contain a relatively high amount of cases with negative water depth (i.e. basement flooding only) but also with water depth above 200 cm (Figure 1). In contrast, most cases in the Dutch data set have a water depth between zero and 200 cm. The German data set also contains many more extreme values. Thus, also the mean relative building and content damages of the German data set are higher in comparison with the Dutch data set (Figure 1).

Due to the different data acquisition methods, the definitions and units of the variables differ between the data sets. Thus, a harmonization of the variables was necessary to enable cross validation and updating procedures in which data of different data sets is mixed. Adjustments were made to the flow velocity and return period variables. The flow velocity in the German data set is translated from its intensity class to a meter per second value to match the Dutch data set. The return period definition of the Dutch data set has been replaced with the German definition. In the German data set the return period represents the return

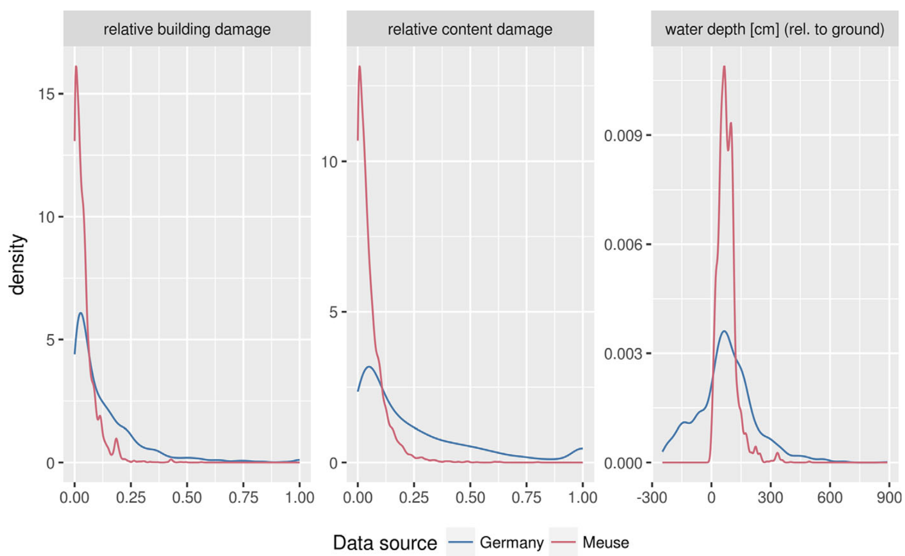

Figure 1. Comparison of the value ranges and densities of the relative building and content damage as well as water depths for the Dutch and German data sets.

period of the flood that caused the damage, so it says something about the magnitude of the flood. In the Dutch data set it is the flooding probability of any flood at that location. By switching to the German definition all records in the Dutch data set get the same value, namely the return period of the 1993 event, which is 40 years. Table 1 shows an overview of the variables used in this study.

2.2. Damage Models

2.2.1. Bayesian Networks Based Damage Models

In Bayesian Networks, random variables and their conditional dependencies are represented in a directed acyclic graph (DAG) structure. Each variable can be either observed or represented as a prior probability distribution. Dependencies between variables are represented with edges representing joint probability distributions. The edges in a Bayesian Network are directed which means there is a direction in which the influence of one variable flows to the other. From this network, the probability distributions of any variable can be predicted based on knowledge about the other variables.

The Bayesian Networks in this paper all use discretized variables. This means that each variable can take only a limited number of values. The advantage of this is that the joint probability distributions can be represented as tables (conditional probability tables) rather than as functions. The probability of a set of variable values can then be calculated with the following formula:

$$P(X_i, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{parents}(X_i))$$

Where X_i are the variables and $\text{parents}(X_i)$ is the set of variables directed to X_i. The probability of a single variable value can be obtained by taking the sum of all the probabilities that contain the variable value of interest. $P(X_i|\text{parents}(X_i))$ can be looked up from the conditional probability tables.

Conditional probability tables show for each combination of parent variable values the probability of each possible output value. The tables can be trained with data using maximum likelihood estimation. In this method the conditional probabilities are computed based on the number of observations that have the particular output value. The drawback of this method is that the number of possible combinations of parent variable values and output values can get very large. In such cases the number of observations is likely to become insufficient for the maximum likelihood estimation (Koller & Friedman, 2009). Discretization can therefore not become too fine and one variable shouldn't have too many edges pointing toward it.

2.2.1.1. German Bayesian Network Based Damage Model (BN-FLEMOps)

The German Bayesian Network-based damage model is referred to as BN-FLEMOps, which stands for Bayesian Network Flood Damage Estimation MOdel for the private sector. The structure of BN-FLEMOps network is based on all complete records with respect to the variables water depth, relative damage, return period, flood duration, building area, building type, precautionary measures, and flood experience from the German database described in section 2.1.1. A set of three different algorithms implemented in the R-package "bnlearn" (Scutari, 2010) were used in a bootstrap approach to learn the network structure shown in Figure 2. With arandom sample of 950 observations (out of 1,456) the algorithms Fast-IAMB (Fast Incremental Association (Yaramakala & Margaritis, 2005), Inter-IAMB (Interleaved Incremental Association (Tsamardinos et al., 2003) and a hill-climbing approach using the Bayesian Dirichlet Equivalent (BDE) (Heckerman et al., 1995) were initialized 500 times. The result set of 1,500 network structures in total provided the basis to define the connections between the variables .All arcs, and their associated directions, that occurred in at least 80 percent of the cases within the result set have been used to support the development of the expert network BN-FLEMOps. The bootstrap approach that uses only the subset of 950 observations was used to avoid over fitting and to ensure a robust network

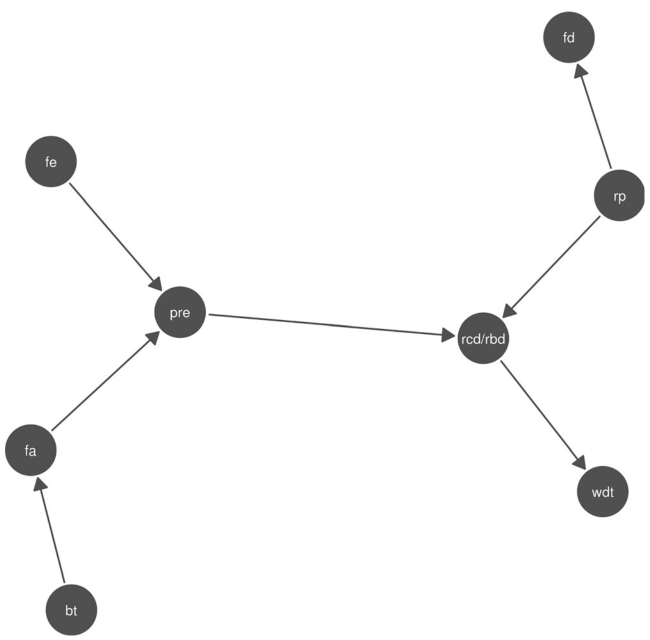

Figure 2. Structure of the German Bayesian Network based Flood Damage Estimation MOdel for the private sector BN-FLEMOps

structure.The variables were discretized on an equal frequency basis with water depth and relative damage in 10 classes, return period and duration in 5 classes and building area in 3 classes. The other variables are discrete by definition.

2.2.1.2. Dutch Bayesian Network (Meuse-BN)

In the Dutch Bayesian Network, Meuse-BN, both the content and building damage are included in the same network, and the network is used to determine both values simultaneously (conditional on the other variables in the network). The Meuse-BN model works with the libpgm Python library (Cabot, 2012), which is based on the methodology described in Koller and Friedman (2009). The discretization is based on the principle of having an equal number of observations per bin, because this is an efficient way of using the limited number of observations. The number of bins per variable value is determined by testing several configurations and then picking the one with the smallest mean absolute error (MAE). The result is 6 bins for the output variables, 4 bins for the water depth variables, 3 bins for the living area and other hydraulic variables and 2 bins for the rest of the variables. More details about the Meuse-BN model can be found in Wagenaar et al. (2017).

The network structure of the Dutch Bayesian Network was learned on the basis of all available 4,398 records from the Dutch data set and the predefined discretization. A hill climbing approach together with the BDE (Heckerman et al., 1995) was used to find the structure with the highest BDE score. The approach uses a sequence of arc additions, arc removals and arc inversions to maximize the BDE score, initially starting with no arc between the variables at all. This approach was repeated 1,000 times to avoid local minima in the search space. The structure of the resulting Bayesian network is shown in Figure 3.

2.2.2. Random Forest-Based Damage Models

Random Forests are ensembles of regression trees. A regression tree is a series of binary question nodes (e.g. water depth > 0.5m) about the input variables leading to an output value of the target variable (i.e. relative building damage or relative content damage). These trees are typically created with regression tree learning based on a data set with observations.

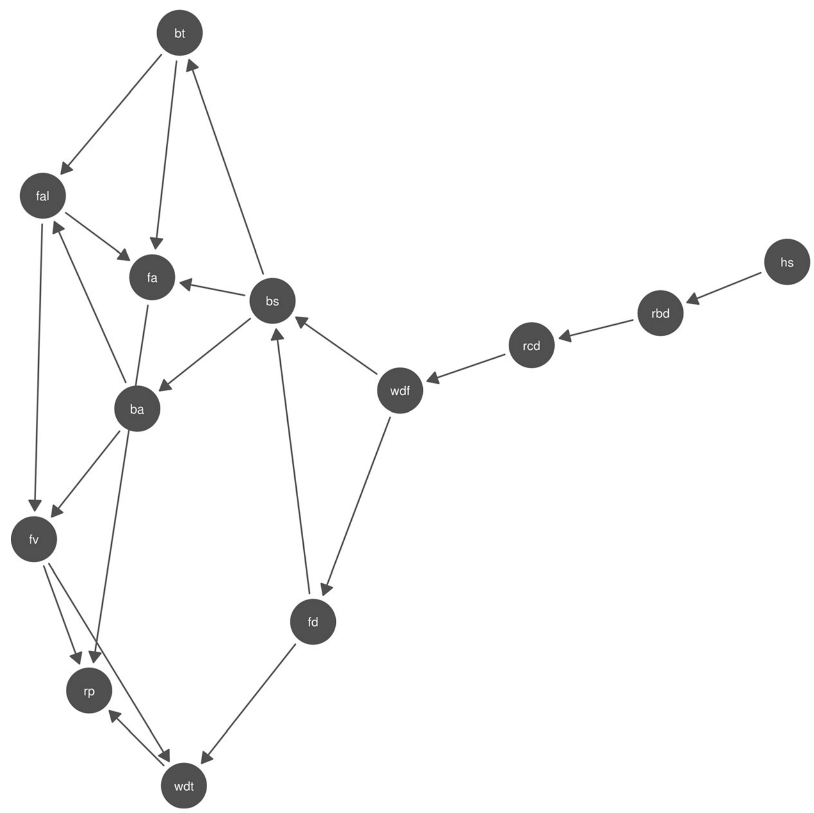

Figure 3. Structure of the Dutch Bayesian Network based model Meuse-BN

A regression tree is a serie of questions about the data (e.g. water depth > 0.5 m?) on which the data are split into different branches forming a tree of options. This tree can be seen as a model to make new predictions. During regression tree learning, an algorithm is searching for variables and values within the data set on which to split the data in order to maximize improvement in homogeneity of the output variable value observations. After each splitting question the tree is further build up and the resulting branches with observations get more homogenous. A central question of regression tree learning is when to stop splitting up the observations. If the tree gets too complex, it will perform poorly on predicting new observations. Several methods are available to stop the tree from growing too complex.

One of the advantages of a random forest is that it helps to prevent overfitting and hence improve the predictive capacity of the model. In a random forest a large number of regression trees are built based on resampling the observations. Furthermore, at each split of each tree in a random forest a number of variables are left out as potential splitting candidates. The average of the predictions of all trees is the prediction of the entire random forest. The variation among the trees represents the probability distribution of the prediction.

2.2.2.1. German Random Forest Based Damage Model (RF-FLEMOps)

The models for the German Random Forest based Flood Damage Estimation MOdel for the private sector (RF-FLEMOps) were learned using the R-package "randomForest" (Liaw & Wiener, 2002). A sensitivity analysis was conducted to find the most robust configuration for the number of trees and the number of variables that were considered at each node. The sensitivity analysis was done visually, to evaluate the convergence of the "out of bag error.". The "out of bag" error is the prediction error of an independent part of the training data that is not used to build an individual tree but instead used to validate the performance. Convergence was reached with 5,000 trees and the default value using 1/3 of the variables at each node for splitting.

2.2.2.2. Dutch Random Forest (Meuse-RF)

The Random Forest trained on the Meuse data (Meuse-RF) works with the Python library "Sci-Kit learn". Overfitting is prevented by limiting the number of splits in each tree to 25, the value which resulted in the smallest MAE on validation data. During the random forest learning 2/3 of the variables are randomly left out as potential splitting candidates at each split. This is a standard Sci-Kit learn configuration for Random Forests. One hundred different regression trees are built in the random Forest.

2.3. Cross-Country Model Transfer Setup

2.3.1. Sampling of Training and Validation Data Sets

The German models BN-FLEMOps (for building damage) and RF-FLEMOps (for content damage) were initially trained on 350 randomly selected German data points. Similarly, the Dutch models Meuse-BN and Meuse-RF were initially trained on 350 Dutch data points. The German and Dutch models were then validated twice using 200 data points sampled from both the German and Dutch data sets. This set-up is shown in Figure 4. The training and validation process was repeated 1,000 times, resulting in a distribution of model performance results.

Some input variables of the damage models BN-FLEMOps and RF-FLEMOps have no (near) equivalent in the Dutch data set, such as the precautionary measures and flood experience indicators. Missing variables in the validation data set increase the uncertainty and decrease the probability that a damage observation is similar to the modeled damage value. To address the problem of missing values in Bayesian Networks, which can calculate with any number of unknown variables, the missing data are treated as an extra unknown variable (together with the relative damage). The BN calculates a posterior distribution for the missing variable (conditional on the remaining variables), from which a value is randomly sampled. For the Random Forests, no assumptions about missing data are made and the missing values are filled by drawing from a uniform distribution that is defined by the minimum and maximum of the original variables with no missing values.

2.3.2. Updating Models With Different Data Sets

In another analysis all the Dutch (Meuse-BN, Meuse-RF) and the German (BN-FLEMOps, RF-FLEMOps) models are in a first step updated with 350 randomly selected extra data points from the data set of the other country and in another second step again updated with additionally 350 data points of the data set of the other country. The models are then again validated after the first and the second updating step based on 200 independently randomly selected validation data points from the data set of the other country. This process is also repeated 1,000 times to get rid of any randomness in the sampling. The hypothesis is that the models will perform better in this cross-country transferability test when they have been updated with (independent) data from the country for which they are validated on.

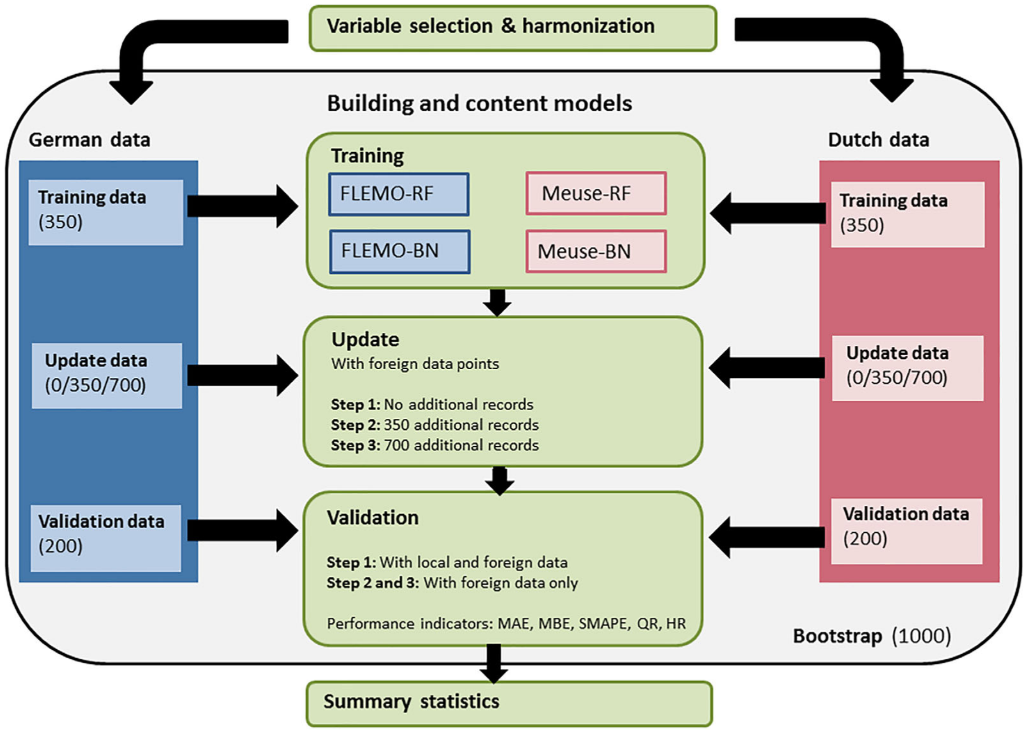

Figure 4. Overview of the cross-country model transfer set-up

Since some input variables of the damage models BN-FLEMOps and RF-FLEMOps have no equivalent in the Dutch data set, dealing with missing values during the updating procedure is necessary. In case of the model BN-FLEMOps, missing data were filled following the joint probability distribution of the model without the new data. This approach ensures that existing dependencies between variables defined by the Bayesian network structure are considered and applied. Thus, existing knowledge about damage processes is transferred during the updating process. In case of the model RF-FLEMOps, a linkage between variables does not exist in such an explicit structure, but is rather hidden under the total number of trees in the Random Forest. Because this structure is hard to address, no assumptions about missing data are taken and the missing values are filled by drawing from a uniform distribution that is defined by the minimum and maximum of the original variables with no missing values.

2.3.3. Evaluation Criteria

To evaluate the performance of the models 5 different evaluation criteria are applied (see Table 2). These are: Mean Absolute Error (MAE), Mean Bias Error (MBE), Symmetric Mean Absolute Percentage Error (SMAPE), Hit Rate (HR) and Quantile Range (QR). Both, Bayesian Networks and Random Forests provide multiple estimations for every individual validation point or prediction. For Bayesian Networks, this is the result of conditional probabilities, for the Random Forests, it is given by the single prediction of each tree. MAE, MBE and SMAPE evaluate the precision of the models using the median of the simulations for every individual validation point. The MAE shows the mean error among all observations used for the validation. The errors are made absolute before taking the mean and therefore an overestimation cannot be used to balance out an underestimation. This error metric is especially useful when one is interested in the model performance on individual objects, which is of interest for insurance purposes, for instance. This is also an absolute error metric in the sense that the error is expressed in relative damage. Therefore a large error on a large damage is equal to the same large error on a small damage. The SMAPE error metric corrects for this by calculating a relative error. The sum of the absolute errors is divided by the sum of all observed and simulated damages.

Table 2
Formulas of the Error Metrics[a]

Error metric	Formula
Mean Absolute Error (MAE)	$MAE=\frac{1}{N}\sum \lvert RL_{sim,n}-RL_{obs,n}\rvert$
Mean Bias Error (MBE)	$MBE=\frac{1}{N}\sum RL_{sim,n}-RL_{obs,n}$
Symmetric Mean Absolute Percentage Error (SMAPE)	$SMAPE=\frac{\sum \lvert RL_{sim,n}-RL_{obs,n}\rvert}{\sum \lvert RL_{sim,n}\rvert+\lvert RL_{obs,n}\rvert}$
Quantile Range (QR)	$QR=\frac{1}{N}\sum RL_{sim,n,q95}-RL_{sim,n,q5}$
Hit Rate (HR)	$HR=\frac{Observations\ \ within\ \ QR}{N}$

Note. $RL_{sim,n}$ is the nth simulated relative damage, $RL_{obs,n}$ is the nth observed simulated damage. N is the total number of observations.

This means that a large error on a small damage is more important than the same large error on a large damage. This is relevant here because the average Dutch damages are much lower than the German damages (see 2.1.4).

The MBE is an error metric in which overestimates are allowed to be compensated by underestimates. This is relevant when the purpose of the model is to estimate total damage such as often is the case in studies related to public policymaking on flood risk, and for societal benefit-cost analysis.

The quality of the simulations is further evaluated by the QR (Quantile Range) and the HR (Hit Rate). The QR evaluates the variation within the model predictions and the HR the reliability of the models. The QR is the distance between the 5% and 95% quantiles of the estimated flood damage. A large QR means the uncertainty of the prediction is high according to the model. The HR is the fraction of the observations that lie within this 90% QR. A HR of 0.9 means a perfect reliability (Thordarson et al., 2010). A HR larger than 0.9 indicates that the QR is probably too large (10% of the observations should be outside the 90% QR). A smaller HR shows that the QR is too small. All models are trained and validated with 1,000 different sample combinations, which allows the construction of a distribution of the observed quantile ranges and hit rates.

3. Results and Discussion

In this section first the results are shown from the cross-country model transfers. These results are compared with the model performance on local data and implications of these results are discussed. Then the results are shown from the models updated with data from the country it is applied in and the improvement in results is shown and are again discussed.

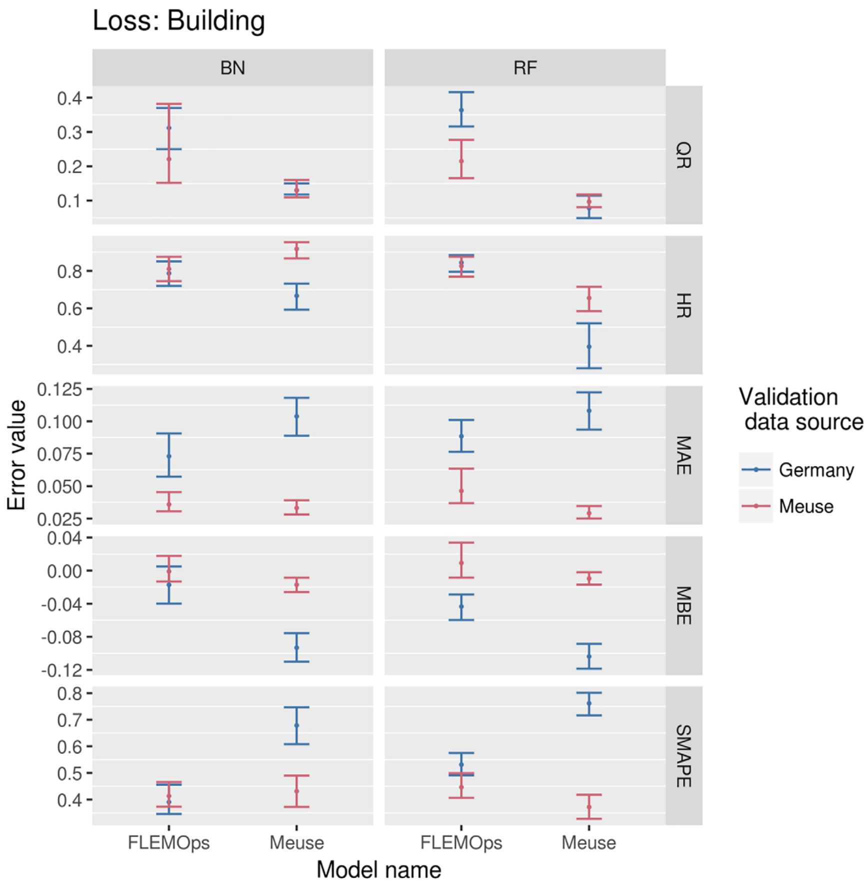

Figure 5. Validation results for building damage. Left the results for the Bayesian Networks and right the results for the Random Forests.

3.1. Initial Model Performances

The models BN-FLEMOps and RF-FLEMOps, which were trained on the German data predict the Dutch damages much better than the other way around (i.e. the Meuse models predicting the German damage). This is consistent over all precision error metrics (MAE, MBE and SMAPE). In a few instances the BN-FLEMOps even outperforms (by a small but significant margin) the Meuse-BN in estimating the Dutch damage (SMAPE and MBE error metrics in Figures 5 and 6). The probable reason for this is that the German data set is much more heterogeneous than the Dutch data set. The cause of the larger heterogeneity is the mix of various flood events and large, diverse regions from which data are included. The Elbe catchment (former East Germany) and Danube catchment (Bavaria) have very different building stock and socio-economic characteristics. Some of the observations are therefore possibly representative of the Dutch situation. This allows the FLEMOps models to do well in the Netherlands because the spectrum of flood damage data that it was trained on is much larger. The Dutch data come from only one flood event in a single relatively small region, with relatively homogenous building stock. Thus, the damages and other variables are more homogenous and the size of the spectrum of its observations are smaller. Therefore, there will be many observations in the German data outside the small spectrum of the Dutch data and hence the Meuse models need to extrapolate when trying to predict these observations. Most Dutch damage observations will be within the large spectrum of German observations and hence, the FLEMOps models mainly only need to interpolate when estimating Dutch damages.

The Dutch data set has a larger number of data points than the German data set, 4,398 (building and content) in comparison with 1,456 (building) and 1,324 (content), respectively. Because of the sampling setup in this study this potential advantage is not used in the current study. However, several tests with taking more data points for training have shown no significant improvements when sampling a larger number of data points for training. It is therefore expected that extra data points are only valuable when they add to

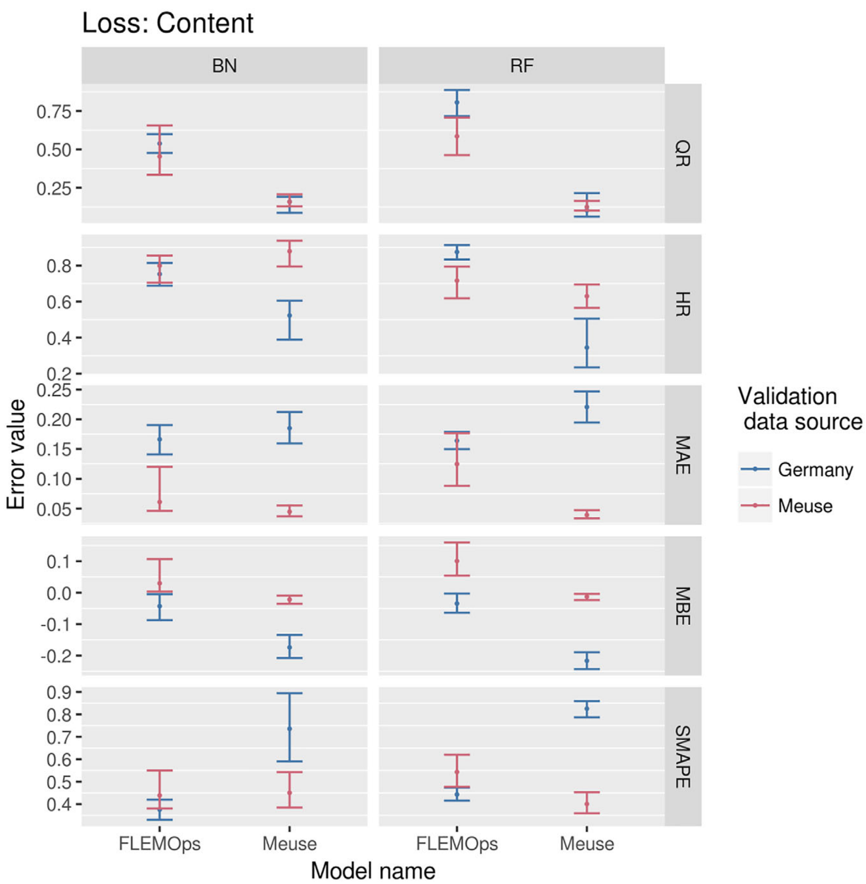

Figure 6. Validation results for content damage. Left the results for the Bayesian Networks and right the results for the Random Forests.

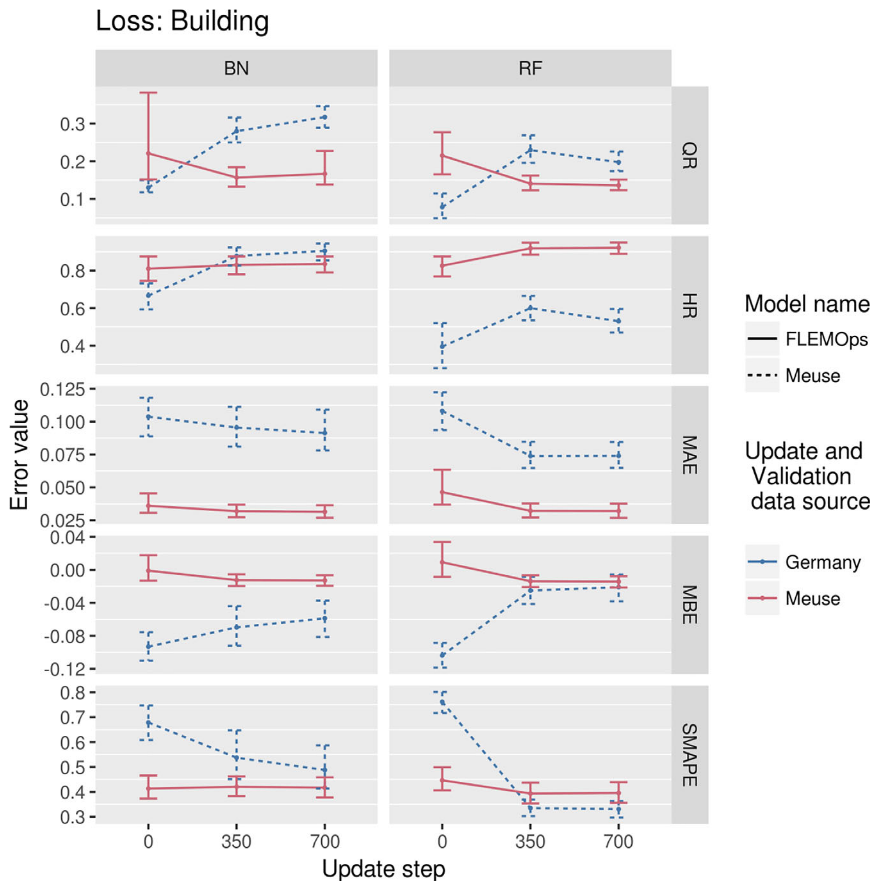

Figure 7. Validation results for building damage after update with foreign data. Left the results for the Bayesian Networks and right the results for the Random Forests.

the spectrum of the model. This is more evidence, that when aiming on development of an empirical flood damage model, for flood damage data collection the heterogeneity of the collected data, and thus the reflection of different damage processes, is more important than the size of the data set.

The Quantile Range results show a similar story. The Meuse models are trained on relatively homogenous data and are therefore much more confident about their predictions, which is reflected in a small QR. The consequence however is a poor HR of the Meuse models validated on the German data. While Meuse-BN almost scores a perfect HR of 0.9 when validated on the Dutch data, reflecting a high reliability, the FLEMOps models have good HRs around 0.8, on both the Dutch and German data.

A notable result is that FLEMOps performs for the MAE significantly better in predicting the Dutch damage than in predicting the German damage. This is unexpected because a model is believed to perform better in the area its training data originates from. This is probably because the damages in the Netherlands are on average much lower than in Germany (see 2.1.4) and therefore the absolute errors are also smaller. The relative error metric corrects for this and hence on the SMAPE error metric the FLEMOps models work better when predicting the German data 3 out of 4 times. Similar observations are given for the comparison between the estimates for the building and contents damage. MAEs are significantly larger for estimating relative contents damage in comparison with relative building damage for all models. Such a significant difference cannot be observed for the SMAPE. This shows that when comparing performances from different validation sets a relative error metric adds additional information and thus should complement the absolute error metric.

In the cross country validation test, i.e. FLEMOps validated with Dutch data and Meuse model validated with German data, the Bayesian Network based models perform better in comparison with the random forest based models.

3.2. Model Performance After Update

The updating (see Figures 7 and 8) test confirms our hypothesis: All models improve their performance when validated on the data set of the other country after they are updated with independent data from that country. This improvement is visible on almost all error metrics. The MBE for building damage is an exception here, the performance of the FLEMOps declines slightly (but significant) after an update with Dutch data (validated on Dutch data).

The improvement of model performance after updating is high for the Meuse models when updated and validated with German data. The performance improvement of the FLEMOps models is smaller in comparison with the Meuse models.

For instance, updating BN-FLEMOps hardly leads to improvements for estimating building damage, however, it does for estimating content damage. This can also be explained based on the homogeneity of the Dutch data. Because the Dutch data are homogenous it profits a lot by extra data from other events, while the German data already has data from many events and therefore profits less by the data of another event, even if this data comes from the same source as the validation data set.

Generally, the performances of the Bayesian Network models does not differ much from the ones of the Random forest models. However, during the updating procedure, the Random forest models seem to profit more from additional data than the Bayesian Network models. This is as expected because updating the Bayesian Network just updates the conditional probability tables, not the relation between the variables (i.e. the Bayesian Network structure). Thus newly introduced damage process are not fully reflected by the updated models. For Random Forests on the other hand, the update learns an entire set of new trees and thus damaging processes that are given by the new data at hand can be described by the updated models.

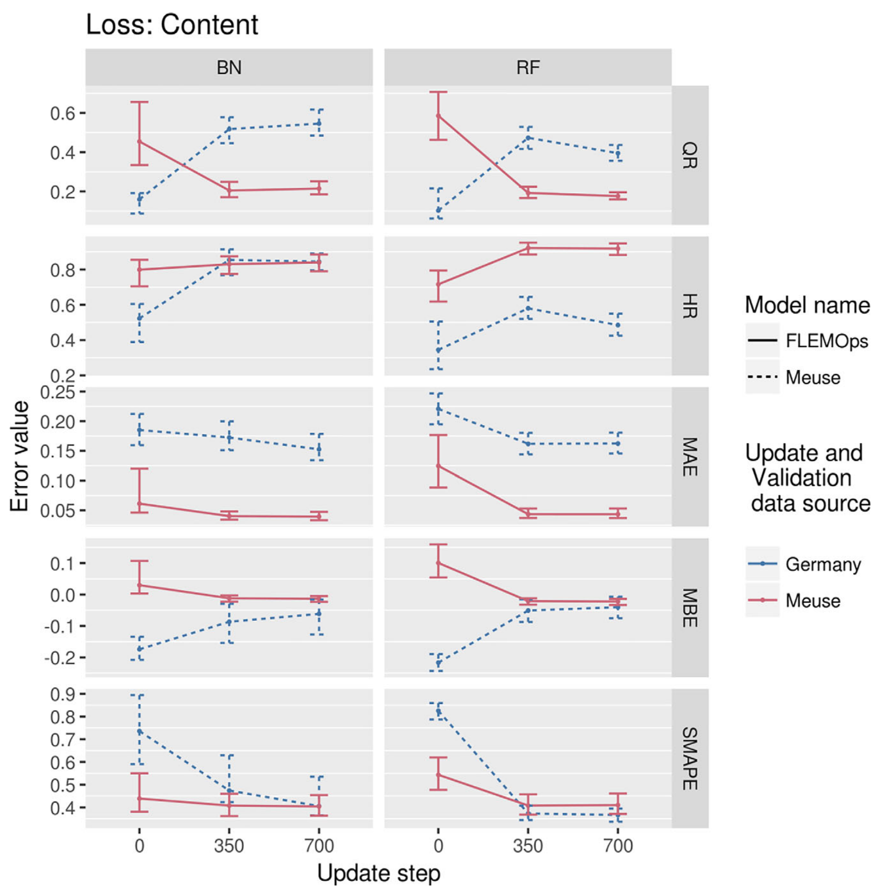

Figure 8. Validation results for content damage after update with foreign data. Left the results for the Bayesian Networks and right the results for the Random Forests.

The bootstrap approach assesses the sensitivity of the data selection for model training and validation. As the German data have a higher variability, the FLEMOps models trained on that data are sensitive to the selection because data might be chosen that is similar to the Dutch validation data and thus show low errors, or data might be chosen that is very different to the Dutch validation data and thus show higher errors. This would have been a smaller issue if more than 350 data points would be used for the zero step. As a consequence, the spread for the error measures at step 0 is bigger than at the following steps since the updated data used for the next steps is from the same population as the validation data, and as we add data from the same source as the validation set, we are more likely drawing data that are similar to the validation data. Thus the spread of the errors of the FLEMOps models is generally decreasing during the updating procedure. The Meuse models don't have this issue because the data are homogenous and therefore the performance depends less on what data are sampled (it performs consistently bad on the German validation data).

4. Conclusions

The multi-variable flood damage models BN- and RF-FLEMOps perform better when validated on data from The Netherlands than the Meuse models when validated on German data. This is probably because the German training data are more heterogeneous, since they are based on a large number of events from different regions. The FLEMOps models, therefore, seem to be better able to capture damage processes from other events and in other locations. The Dutch data are only based on a single flood event in a relatively small area. This also explains why the Meuse models benefit more from updating with data from Germany, than the other way around.

This implies that the heterogeneity of collected flood damage data for modeling purposes can be more important than the quantity of data points. Future flood damage collection efforts should therefore focus on acquiring heterogeneous data, for instance by targeting different locations in the same event, or several events for the same location, instead of collecting a large quantity of data only for a single event in one location. Such data collection efforts over several years and events may provide much more relevant information for improving risk analysis than an in-depth study for a single event.

The good performances of the FLEMOps models in the Netherlands show that the transfer of multi-variable flood damage models in space and time can at least in some cases be successful. The important condition for this seems to be that the training data contain observations collected under similar conditions as present at the location to which the model is transferred. Performance of models in transfer settings is additionally improved via updating the models using data from the location in which they are applied.

Acknowledgments

This research was supported by the European Union's Horizon 2020 research and innovation programme, through the IMPREX project (grant agreement 641811). Parts of this work have been supported by funding for the OASIS+ Future Danube: Multi Hazard and Risk Model demonstrator project through the Climate Knowledge and Innovation Community (Climate-KIC) by the European Institute of Innovation and Technology (EIT) and from the European Union's Horizon 2020 research and innovation program (grant agreement 730381). The surveys to collect the empirical damage data in Germany were supported by the German Research Network Natural Disasters (German Ministry of Education and Research (BMBF), no. 01SFR9969/5), the MEDIS project (BMBF; 0330688) the project "Hochwasser 2013" (BMBF; 13N13017) and by a joint venture between the German Research Centre for Geosciences GFZ, the University of Potsdam and the Deutsche Rückversicherung AG, Düsseldorf. We would like to thank the editor Jim Hall and two anonymous reviewers for their constructive and useful suggestions. We would also like to thank Kathryn Roscoe for proofreading the manuscript. Most of the German flood damage data are available via the German flood damage database HOWAS21 (http://howas21.gfz-potsdam.de/howas21/). The Dutch data are available in the supporting information of Wagenaar et al. (2017).

References

Axer, T., Bistry, T., Klawa, M., Müller, M., & Süßer, M. (2012). Sturmdokumentation 2011 Deutschland (Storm documentation 2011 Germany), Deutsche Rückversicherung, Düsseldorf [in German]. Retrieved from https://www.deutscherueck.de/fileadmin/user_upload/Sturmdoku_2011_WEB.pdf (last accessed 2.5.2018)

BfG (Federal Institute of Hydrology). (2006). Das Hochwasser der Elbe im Frühjahr 2006, BfG-1514, Koblenz, Germany, 2006 [in German].

Blöschl, G., Nester, T., Komma, J., Parajka, J., & Perdigão, R. A. P. (2013). The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. *Hydrology and Earth System Sciences*, *17*, 5197–5212. https://doi.org/10.5194/hess-17-5197-2013

Bouwer, L. M., Haasnoot, M., Wagenaar, D., & Roscoe, K. (2017). Assessment of alternative flood mitigation strategies for the C-7 Basin in Miami, Florida. *Deltares*, 1230718.

Cabot, C. (2012). Libpgm Python library. Retrieved from http://pythonhosted.org/libpgm/, accessed on 10-08-2016

Cammerer, H., Thieken, A. H., & Lammel, J. (2013). Adaptability and transferability of flood loss functions in residential areas. *Natural Hazards and Earth System Sciences*, *13*, 3063–3081.

Dahm, R., Verhoeven, G., Gehrels, H., Diermanse, F., Wagenaar, D., Dettrick, D., et al. (2017). *Strengthening disaster risk management and urban flood resilience in Lao PDR – Report 2 assessing flood risk and evaluating design options. Deltares* (Rep. 1230978-002). Delft, the Netherlands: Deltares.

De Bruijn, K. M., Wagenaar, D. J., Slager, K., De Bel, M., & Burzel, A. (2014). *Proposal for the new flood damage assessment method: SSM2015*. Rijkswaterstaat: Deltares.

Delft, W. L. (1994). *Onderzoek watersnood Maas, Deelrapport 1: Wateroverlast December 1993* [in Dutch]. Delft, the Netherlands: WL Delft.

Deutsches Komitee Katastrophenvorsorge e.V. (DKKV) (Ed.) (2015). *Das Hochwasser 2013 - Bewährungsprobe für das Hochwasserrisikomanagement in Deutschland* (Schriftenreihe des DKKV; 53). Bonn, Germany: DKKV. Retrieved from http://www.dkkv.org/fileadmin/user_upload/Veroeffentlichungen/Publikationen/DKKV_53_Hochwasser_Juni_2013.pdf (last accessed 2.5.2018)

Dietz, H. (1999). *Wohngebäudeversicherung Kommentar* (2nd ed.). Karlsruhe, Germany: VVW Verlag Versicherungswirtschaft GmbH.

EC (European Commission) (2014). Study on economic and social benefits of environmental protection and resource efficiency related to the european semester (701 pp.). Brussels, Belgium: Country Fiches, European Environment Agency.

Elmer, F., Thieken, A. H., Pech, I., & Kreibich, H. (2010). Influence of flood frequency on residential building losses. *Natural Hazards and Earth System Sciences*, *10*, 2145–2159. https://doi.org/10.5194/nhess-10-2145-2010

Engel, H. (2004). The flood event 2002 in the Elbe River basin: Causes of the flood, its course, statistical assessment and flood damages. *Houille Blanche*, (6), 33–36.

Falter, D., Nguyen, D., Vorogushyn, S., Schröter, K., Hundecha, Y., Kreibich, H., et al. (2016). Continuous, large-scale simulation model for flood risk assessments: Proof-of-concept. *Journal of Flood Risk Management*, *9*(1), 3–21.

Falter, D., Schröter, K., Nguyen, D., Vorogushyn, S., Kreibich, H., Hundecha, Y., et al. (2015). Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain. *Journal of Hydrology*, *524*, 182–193.

Gerl, T., Kreibich, H., Franco, G., Marechal, D., & Schröter, K. (2016). A review of flood loss models as basis for harmonization and benchmarking. *PLoS ONE*, *11*(7), e0159791. https://doi.org/10.1371/journal.pone.0159791

Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Machine Learning*, (1995) *20*, 197. https://doi.org/10.1023/A:1022623210503

IKSE (International Commission for the Protection of the Elbe). (2004). *Dokumentation des Hochwassers vom August 2002 im Einzugsgebiet der Elbe* (IKSE Report). Magdeburg, Germany: IKSE Report. Retrieved from http://www.ikse-mkol.org/fileadmin/media/user_upload/D/06_Publikationen/02_Hochwasserschutz/2004_IKSE-Dokumentation_Hochwasser_2002.pdf (last accessed 2.5.2018)

Jelonek, L., Wrzeszcz, L., Zawi'slak, J., Walther, P., Winkler, U., Wortha, S., et al. (2010). Gemeinsamer polnisch-deutsch-tschechischer Bericht zum Hochwasser vom 07. bis 10. August 2010 an der Lausitzer Neisse als Bestandteil der vorläufigen Risikobewertung gemäß Artikel 4 der Hochwasserrisikomanagementrichtlinie (2007/60/EG), Report Wrocław, Dresden, Prag. Retrieved from http://www.umwelt.sachsen.de/umwelt/wasser/download/06-12-10_LN_dt.pdf (last access: 22 January 2018)

Jongman, B., Kreibich, H., Apel, H. B., Bates, P., Feyen, L., Gericke, A., et al. (2012). Comparative flood damage model assessment: Towards a European approach. *Natural Hazards and Earth Sciences*, *12*, 3733–3752.

Kienzler, S., Pech, I., Kreibich, H., Müller, M., & Thieken, A. H. (2015). After the extreme flood in 2002: Changes in preparedness, response and recovery of flood-affected residents in Germany between 2005 and 2011. *Natural Hazards and Earth Sciences*, *15*, 505–526.

Kind, J. M. (2013). Economically efficient flood protection standards for the Netherlands. *Journal of Flood Risk Management*, *7*(2), 103–117.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: Massachusetts Institute of Technology Press.

Korndörfer, C., Rümpel, A., Seifert, J. O., Kroll, H., Grebedünkel, D., Jakob, T., et al. (2006). *Bericht zum Frühjahrshochwasser 2006, Report of the Umweltamt Dresden* [in German]. Dresden, Germany: Environmental Agency of Dresden.

Kreibich, H., Botto, A., Merz, B., & Schröter, K. (2017). Probabilistic, multivariable flood loss modeling on the mesoscale with BT-FLEMO. *Risk Analysis*, *37*(4), 774–787.

Kreibich, H., Seifert, I., Merz, B., & Thieken, A. H. (2010). Development of FLEMOcs - A new model for the estimation of flood losses in companies. *Hydrological Sciences Journal*, *55*(8), 1302–1314.

Kreibich, H., & Thieken, A. H. (2009). Coping with floods in the city of Dresden, Germany. *Natural Hazards*, *51*(3), 423–436.

Kreibich, H., van den Bergh, J. C. J. M., Bouwer, L. M., Bubeck, P., Ciavola, P., Green, C., et al. (2014). Costing natural hazards. *Nature Climate Change*, *4*, 303–306.

Kron, W. (2009). Überschwemmungsüberraschung: Sturzfluten und Überschwemmungen fernab von Gewässern [in German]. Wasserwirtschaft, 6, 15–20.

Kron, W., & Ellenrieder, T. (2008). Zunehmende Wetterschäden: Was kostet das die Versicherungswirtschaft? Teil 1: Wetterkatastrophen – Ursachen und SchûÊden (in German). *Korrespondenz Wasserwirtsch*, *1*, 691–696.

LfU (Bavarian Environment Agency) (2007)., *Gewässerkundlicher Bericht Hochwasser August 2005* [in German]. Augsburg, Germany: Bayerisches Landesamt für Umwelt.

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22.

Merz, B., Elmer, F., Kunz, M., Mühr, B., Schröter, K., & Uhlemann-Elmer, S. (2014). The extreme flood in June 2013 in Germany. *Houille Blanche - Revue Internationale De L'eau*, *1*, 5–10.

Merz, B., Kreibich, H., & Lall, U. (2013). Multi-variate flood damage assessment: A tree-based data-mining approach. *Natural Hazards and Earth System Sciences*, *13*, 53–64. https://doi.org/10.5194/nhess-13-53-2013

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article "Assessment of economic flood damage." *Natural Hazards and Earth System Sciences*, *10*, 1697–1724. https://doi.org/10.5194/nhess-10-1697-2010

Merz, B., Kreibich, H., Thieken, A. H., & Schmidtke, R. (2004). Estimation uncertainty of direct monetary flood damage to buildings. *Natural Hazards and Earth System Sciences*, *4*(1), 153–163.

Meyer, V., Becker, N., Markantonis, V., Schwarze, R., van den Bergh, J. C. J. M., Bouwer, L. M., et al. (2013). Review article: Assessing the costs of natural hazards - state of the art and knowledge gaps. *Natural Hazards and Earth System Sciences*, *13*, 1351–1373. https://doi.org/10.5194/nhess-13-1351-2013

Molinari, D., Menoni, S., Aronica, G. T., Ballio, F., Berni, N., Pandolfo, C., et al. (2014). Ex post damage assessment: An Italian experience. *Natural Hazards and Earth System Sciences*, *14*, 901–916. https://doi.org/10.5194/nhess-14-901-2014

Papathoma-Köhle, M., Kappes, M., Keiler, M., & Glade, T. (2011). Physical vulnerability assessment for alpine hazards: State of the art and future needs. *Natural Hazards*, *58*, 645–680. https://doi.org/10.1007/s11069-010-9632-4

Poussin, J. K., Botzen, W. J. W., & Aerts, J. C. J. H. (2014). Factors of influence on flood damage mitigation behaviour by households. *Environmental Science & Policy*, *40*, 69–77. https://doi.org/10.1016/j.envsci.2014.01.013

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., & Merz, B. (2014). How useful are complex flood damage models?. *Water Resources Research*, *50*, 3378–3395. https://doi.org/10.1002/2013WR014396

Schröter, K., Kunz, M., Elmer, F., Mühr, B., & Merz, B. (2015). What made the June 2013 flood in Germany an exceptional event? A hydrometeorological evaluation. *Hydrology and Earth System Sciences*, *19*, 309–327.

Scutari, (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, *35*(3), 1–22.

Seifert, I., Kreibich, H., Merz, B., & Thieken, A. H. (2010). Application and validation of FLEMOcs - a flood loss estimation model for the commercial sector. *Hydrological Sciences Journal*, *55*(8), 1315–1324.

Thieken, A., Olschewski, A., Kreibich, H., Kobsch, S., & Merz, B. (2008). Development and evaluation of FLEMOps - a new Flood Loss Estimation MOdel for the private sector. In D. Proverbs, C. A. Brebbia, & E. Penning-Rowsell (Eds.), *Flood recovery, innovation and response I* (pp. 315–324). Southampton, UK: WIT Press.

Thieken, A. H., Bessel, T., Kienzler, S., Kreibich, H., Müller, M., Pisi, S., & Schröter, K. (2016). The flood of June 2013 in Germany: How much do we know about its impacts? *Natural Hazards and Earth System Sciences*, *16*, 1519–1540.

Thieken, A. H., Kreibich, H., Müller, M., & Lamond, J. (2017). Data collection for a better understanding of what causes flood damage – experiences with telephone surveys. In D. Molinari, S. Menoni, & F. Ballio (Eds.). *Flood damage survey and assessment: New insights from research and practice* (Chap. 7, pp. 95–106). Washington, DC: AGU. ISBN 978-1-119–21792-3. Retrieved from http://onlinelibrary.wiley.com/book/10.1002/978111921793. https://doi.org/10.1002/9781119217930.ch7 0, 2017.

Thieken, A. H., Müller, M., Kreibich, H., & Merz, B. (2005). Flood damage and influencing factors: New insights from the August 2002 flood in Germany. *Water Resources Research*, *41*, W12430. https://doi.org/10.1029/2005WR004177

Thieken, A. H., Petrow, T., Kreibich, H., & Merz, B. (2006). Insurability and mitigation of flood losses in private households in Germany. *Risk Analysis*, *26*(2), 383–395.

Thordarson, F. Ö., Breinholt, A., Moller, J. K., Mikkelsen, P. S., Grum, M., & Madsen, H. (2012). Evaluation of probabilistic flow predictions in sewer systems using grey box models and a skill score criterion. *Stochastic Environmental Research and Risk Assessment*, *26*(8), 1151–1162. https://doi.org/10.1007/s00477-012-0563Thordarson-3

Tollenaar, D., Dahm, R., & Wagenaar, D. (2016). *Flood risk assessment as input to informed decision making for future cities: The case of Bandung*. Deltares, The Netheerlands: Water Resources Institute.

Tsamardinos, I., Aliferis, C., & Statnikov, A. (2003). *Algorithms for large scale Markov Blanket discovery*. Palo Alto, CA: American Association for Artificial Intelligence (AAAI).

Ulbrich, U., Brücher, T., Fink, A. H., Leckebusch, G. C., Krüger, A., & Pinto, J. G. (2003). The central European floods of August 2002: part 1—Rainfall periods and flood development. *Weather*, *58*, 371–377.

Van der Most, H., Tanczos, I., De Bruijn, K. M., & Wagenaar, D. J. (2014). New, Risk-Based standards for flood protection in the Netherlands. Paper presented at 6th International Conference on Flood Management (ICFM6), Sao Paulo, Brazil. September 2014.

Vogel, K., Riggelsen, C., Kreibich, H., Merz, B., & Scherbaum, F. (2014). Flood Damage and Influencing Factors: A Bayesian Network Perspective, in: Proceedings of the 6th European Workshop on Probabilistic Graphical Models (PGM 2012). In *6th European Workshop on Probabilistic Graphical Models* (pp. 347–354). Granada, Spain.

Wagenaar, D., de Jong, J., & Bouwer, L. M. (2017). Multi-variable flood damage modelling with limited data using supervised learning approaches. *Natural Hazards and Earth System Sciences*, *17*, 1683–1696. https://doi.org/10.5194/nhess-17-1683-2017

Wagenaar, D. J., De Bruijn, K. M., Bouwer, L. M., & 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.

Walther, P., Görner, C., Gerber, S., & Wundrak, P. (2013). Ereignisanalyse der Hochwasser von 2010 und 2011 – meteorologische und hydrologische Ursachen, Prozesse, Schäden und Lehren. In *Wasserbaukolloquium 2013: Technischer und organisatorischer Hochwasserschutz – Bauwerke, Anforderungen, Modelle, Dresdner Wasserbauliche Mitteilungen* (Vol. 48, pp. 141–148) [in German]. Dresden, Germany.

Wind, H. G., Nierop, T. M., de Blois, C. J., & de Kok, J. L. (1999). Analysis of flood damages from the 1993 and 1995 Meuse floods. *Water Resources Research*, *35*, 3459–3466.

Yaramakala, S., & Margaritis, D. (2005). Speculative Markov blanket discovery for optimal feature. Paper presented at the Fifth IEEE International Conference on Data Mining. https://doi.org/10.1109/ICDM.2005.134