Sieg, T., Vogel, K., Merz, B., Kreibich, H. (2017): Tree-based flood damage modeling of companies: Damage processes and model performance. - *Water Resources Research*, *53*, 7, pp. 6050—6068.

DOI: http://doi.org/10.1002/2017WR020784

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR020784

Tree-based flood damage modeling of companies: Damage processes and model performance

Tobias Sieg[1,2], Kristin Vogel[2], Bruno Merz[1,2], and Heidi Kreibich[1]

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

Key Points:

- Different damage-influencing variables are identified for the various company sectors and assets
- Prediction accuracies for random forests improve slightly with an increasing amount of training data
- A sector-specific consideration of flood damage is more effective than an increase in training data

Correspondence to:
T. Sieg,
tsieg@gfz-potsdam.de

Citation:
Sieg, T., K. Vogel, B. Merz, and H. Kreibich (2017), Tree-based flood damage modeling of companies: Damage processes and model performance, *Water Resour. Res.*, *53*, 6050–6068, doi:10.1002/2017WR020784.

Received 16 MAR 2017
Accepted 25 JUN 2017
Accepted article online 5 JUL 2017
Published online 25 JUL 2017

Abstract Reliable flood risk analyses, including the estimation of damage, are an important prerequisite for efficient risk management. However, not much is known about flood damage processes affecting companies. Thus, we conduct a flood damage assessment of companies in Germany with regard to two aspects. First, we identify relevant damage-influencing variables. Second, we assess the prediction performance of the developed damage models with respect to the gain by using an increasing amount of training data and a sector-specific evaluation of the data. Random forests are trained with data from two postevent surveys after flood events occurring in the years 2002 and 2013. For a sector-specific consideration, the data set is split into four subsets corresponding to the manufacturing, commercial, financial, and service sectors. Further, separate models are derived for three different company assets: buildings, equipment, and goods and stock. Calculated variable importance values reveal different variable sets relevant for the damage estimation, indicating significant differences in the damage process for various company sectors and assets. With an increasing number of data used to build the models, prediction errors decrease. Yet the effect is rather small and seems to saturate for a data set size of several hundred observations. In contrast, the prediction improvement achieved by a sector-specific consideration is more distinct, especially for damage to equipment and goods and stock. Consequently, sector-specific data acquisition and a consideration of sector-specific company characteristics in future flood damage assessments is expected to improve the model performance more than a mere increase in data.

1. Introduction

Extreme flood events like the riverine flood of 2013 in Europe have severe and manifold impacts on society, including huge financial damage to the economy [*Merz et al.*, 2014; *Schröter et al.*, 2015; *Thieken et al.*, 2016]. The total tangible damage caused in Germany in 2013 is estimated at 6.67 billion Euros, of which 1.48 billion Euros was suffered by private households and 1.32 billion Euros was suffered by the business sector [*Bundesministerium des Innern*, 2013]. Their share of about one fifth of the total damage reveals companies' large damage potential. Yet damage processes particularly in the business sector are not very well understood and are consequently difficult to model, resulting in an urgent need to gain more knowledge on flood damage accrued to companies [*Meyer et al.*, 2013; *Bubeck and Kreibich*, 2011]. Many different factors, such as the water level, the placement of equipment or goods, and the preparedness of the company, can affect the process leading to flood damage [*Kreibich et al.*, 2007].

Since flood risk analyses, including damage modeling, are an essential prerequisite for efficient flood risk management, the identification and quantification of the damage driving factors is highly important. Flood risk analyses are carried out at different spatial scales including the supranational (global), macro (national), meso (regional), and microscales (local) [*de Moel et al.*, 2015]. Many studies assessing flood risk on the micro to mesoscale model the damage in monetary terms on the basis of factors such as water depth, the contamination of the water or the land use of a certain area [*Apel et al.*, 2009; *Falter et al.*, 2015; *Gerl et al.*, 2014; *Huttenlau*, 2010; *Kim et al.*, 2012; *Koks et al.*, 2014a]. Some studies assessing flood risk, expressed as, e.g., the affected amount of the gross-domestic product and population, on the meso to macroscale include factors such as demography indices or other socioeconomic indicators to model the impacts of flood events [*Koks et al.*, 2014b; *Ward et al.*, 2013; *Winsemius et al.*, 2013]. More and more authors claim that societies' vulnerability must be taken into account in flood risk assessments in order to enable a more precise estimate of flood risk and identify effective adaptation measures [*Mechler and Bouwer*, 2014; *Jongman et al.*, 2015].

Commonly, the most detailed data are available at the microscale, enabling an in-depth assessment of the damage processes. Thus, improving the understanding of what influences damage and vulnerability on the microscale can support flood risk assessments on all spatial scales.

So far, several methods were used to determine flood damage-influencing factors to achieve a more precise description of the damage processes. *Zhai et al*. [2005], for instance, used a logistic and a multivariate regression model to estimate the flood damage to residential buildings and their contents, as well as to determine its influencing factors for the Tokai flood in Japan of 2000. Yet some factors that were considered important for the damage process, such as flood preparedness, were not taken into account for damage modeling due to their nonlinear effects [*Zhai et al*., 2005]. *Hudson et al*. [2014] aimed at identifying effective flood damage mitigation measures for private households by means of propensity score matching. One drawback of this method is the need for relatively large sample sizes to get reliable estimates for the effectiveness of the measures. *Merz et al*. [2013], for instance, applied bagging decision trees and regression trees to quantify the importance of various factors for the amount of damage and to model the flood damage to residential buildings.

In general, flood damage models use important damage-influencing variables as input to estimate the damage of elements at risk. Most models consider the type or use of the building or property and the water level as most important factors determining the damage [*Scawthorn et al*., 2006; *Smith*, 1994; *Emschergenossenschaft and Hydrotec*, 2004; *MURL*, 2000]. This concept goes back to the observation of *Grigg and Helweg* [1975] "that houses of one type had similar depth-damage curves regardless of actual value." Other models include additional factors to describe these processes, including precautionary measures, contamination, building quality, etc. [*Hasanzadeh Nafari et al*., 2016b; *Penning-Rowsell et al*., 2005; *Thieken et al*., 2008]. Recent studies used machine learning, multivariable and multivariate approaches to assess flood damage [*Kreibich et al*., 2016; *Merz et al*., 2013; *Hasanzadeh Nafari et al*., 2016c; *Poussin et al*., 2015; *Schröter et al*., 2014; *Vogel et al*., 2012]. *Schröter et al*. [2014] and *Merz et al*. [2013] claimed that tree-based models, such as random forests [*Breiman*, 2001], are suitable for flood damage modeling as they are able to capture nonlinear and even nonmonotonous dependencies between predictor and response variables and they take interactions between the predictors into account. Furthermore, they are able to identify the relevant predictor variables from the set of all considered variables and can be trained from data sets of various sizes, since intrinsic regularization criteria control the complexity of the derived model based on the available training data. However, while most of these studies cover damage to private households, flood damage to companies and its drivers are rarely assessed.

Regarding the flood damage estimation of companies, various models for different company sectors have already been suggested by previous studies. For instance, one of the first and very comprehensive approaches has been the blue manual of *Penning-Rowsell and Chatterton* [1977] which contains stage-damage curves for both residential and commercial property in the UK. In one of its successors, the multicolored manual, *Penning-Rowsell et al*. [2005] distinguished between the following four classes of nonresidential properties: retail, warehouse, office, and factory. Different stage-damage functions per business sector are provided in HAZUS-MH [*Scawthorn et al*., 2006]. The multivariable flood damage models FLEMOcs [*Seifert et al*., 2010] and FLFAcs [*Hasanzadeh Nafari et al*., 2016b] distinguish between different business sectors within the models.

However, to our knowledge, a sector-specific assessment of damage driving variables and estimation of flood damage to companies by means of machine learning has not yet been conducted. This may be due to a lack of suitable data sets, since this is particularly limiting data-driven flood damage assessments of companies [*Merz et al*., 2010; *Meyer et al*., 2013; *Molinari et al*., 2014]. The amount of available data is much smaller than for private households and the heterogeneity within the data is much greater due to the large variety of companies [*Kreibich et al*., 2005; *Merz et al*., 2010]. The questions of how much data are needed to build a reliable model and what can be gained from using more data have rarely been discussed so far. An exception based on private households is for instance the study by *Schröter et al*. [2016].

The objective of this study is a flood damage assessment of companies from different sectors on the microscale with respect to two aspects. The first aspect is the identification of damage-influencing variables to improve the understanding of the flood damage processes of companies. The second aspect is the analysis of the flood damage model performance with respect to increasing data set sizes. Both aspects should lead to a better idea of (1) what and (2) how much data are necessary to describe and quantify damage processes of companies. We propose the random forest approach as a powerful tool to identify relevant predictor variables with linear and nonlinear dependencies from limited data. A meaningful feature selection not only improves the understanding of the damage process, but also enables the development of suggestions for an improved data acquisition, which can then focus on the important damage determining variables.

2. Data and Methods

In the following, random forests are trained on postevent survey data to identify important predictor variables for the estimation of flood damage caused to different company assets: (a) buildings, (b) equipment, and (c) goods and stock. Within this context, a sector-specific consideration is realized by splitting the data into four subsets, following *Kreibich et al*. [2007], each representing one of the four considered sector types: (1) manufacturing, (2) commercial, (3) financial, and (4) service. Random forests are trained for each combination of sector type and company asset, as well as for the complete (sector-unspecific) data set. The prediction quality for the developed sector-specific and unspecific damage models is evaluated via cross-validation depending on the size of the training data set used. The derived models, and consequently, the identified predictors and the prediction quality depend strongly on the data used for training. To provide representative results, the construction of the random forests is repeated several times, with different data subsets sampled from the entire data set. Figure 1 illustrates the entire work flow of this study.

The following section 2.1 describes the survey data set used. The concept of random forests is explained in section 2.2, while section 2.3 describes the sampling scheme used for repeated model construction. The measures applied for the validation of the flood damage models are outlined in section 2.4.

2.1. Survey Data

The data sets used are taken from two surveys conducted after the floods in the Elbe and Danube catchments in the years 2002 and 2013 in Germany [*Kreibich et al*., 2005, 2007; *Thieken et al*., 2016]. The surveys were carried out by the SOKO Institute by means of computer-aided telephone interviews in October 2003, May 2004, and between May and July 2014. In total, 479 interviews were conducted for the flood in 2002 and 557 for the flood in 2013, whereby the interviewed companies were chosen from a site-specific random sample based on lists of affected streets in the corresponding areas [*Kreibich et al*., 2005]. The surveys of the 2002 and 2013 floods were conducted in a similar, comparable way. Questions about the following topics were asked in the surveys: flood impact parameters (e.g., contamination, water level), early warning, emergency measures, precautionary measures, company characteristics, flood damage, and flood experience. The person with the best knowledge about the flood damage was questioned for each company [*Kreibich et al*., 2005]. Given answers were cross-checked during the interview to improve the data quality and to clarify contradictory answers. See *Kreibich et al*. [2005, 2007] and *Thieken et al*. [2017] for further details about the survey and the data processing.

Table 1 shows the nine variables used in this study as potential flood damage predictors, which were derived from the data set. The variables were selected according to data availability and their potential to influence company flood damage according to previous studies [*Penning-Rowsell et al*., 2005; *Scawthorn et al*., 2006; *Kreibich et al*., 2007, 2010]. Variables describing the impact of the flood are the water level, the inundation duration and the contamination indicator, as used in other studies on damage modeling for companies by *Kreibich et al*. [2010] and *Seifert et al*. [2010]. The contamination indicator is the weighted sum of different contaminants such as oil, sewage water, or chemical substances, whereas contaminants which are expected to have a higher damaging potential are weighted accordingly [*Büchele et al*., 2006]. Variables characterizing companies' resistance to flooding are the adaptation ratio, the mitigation ratio, and the emergency indicator. The adaptation and mitigation ratios correspond to the fraction of implemented measures compared to all measures relevant for damage reduction. For example, the installation of flood-proof oil tanks is only relevant for companies that have oil tanks on their premises. Information about the relevance of the respective measures was requested in the survey, i.e., the companies were asked for each measure, if this measure is relevant for their company. A one was added to both numbers to avoid zeros in the fraction.

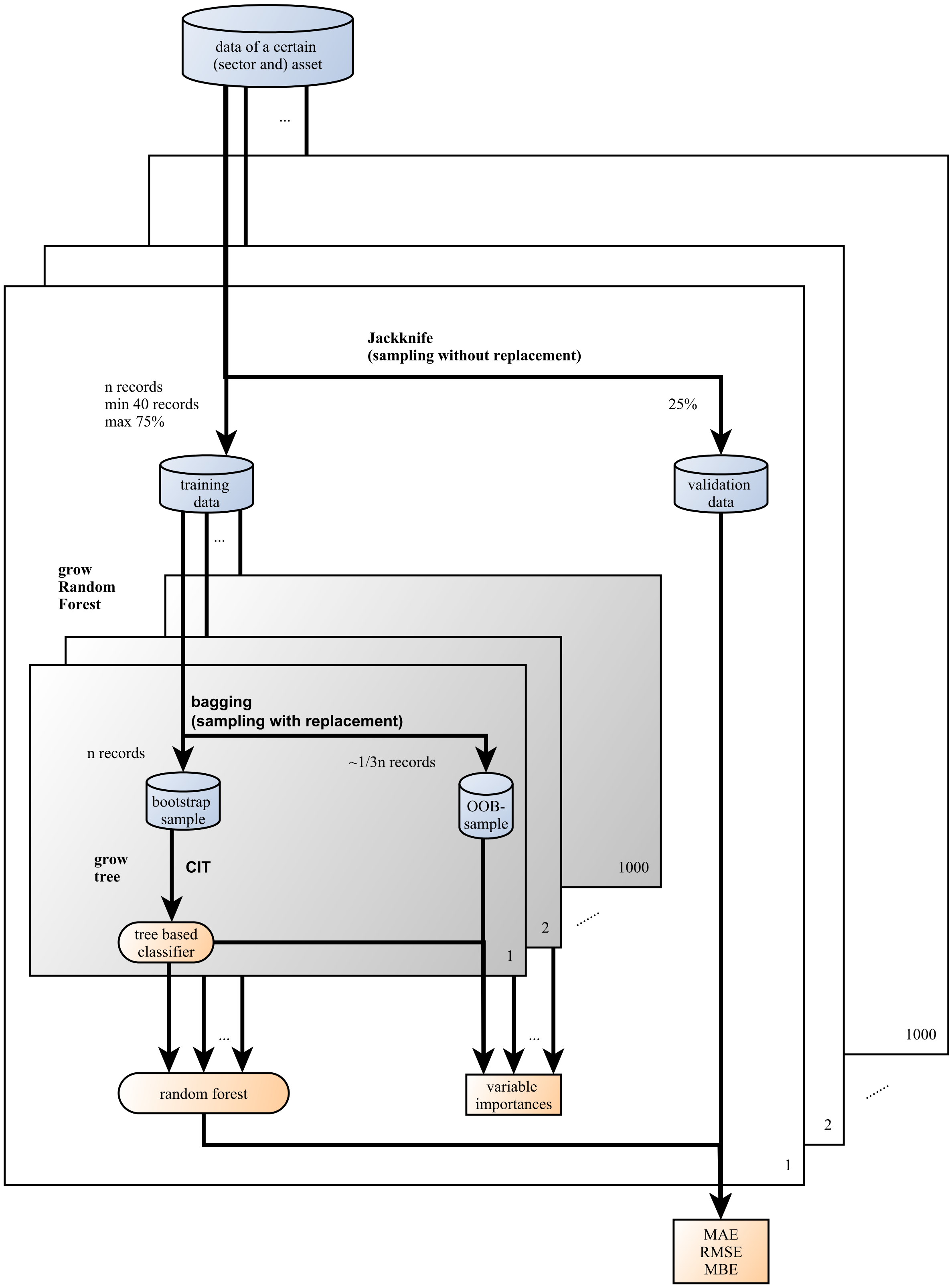

Figure 1. Flowchart of the sampling schemes used in this study.

Table 1. Variables Used in the Models[a]

Predictor Variable	Abbreviation	Values (Scale and Range)
Flood Impact		
Water level	wst	C: 0–960 cm above ground
Inundation duration	d	C: 0–1440 h
Contamination indicator	con	O: 0 = no contamination to 6 = heavy contamination (7 classes)
Damage Reduction		
Adaptation ratio	adapt	O: 0.25 = low adaptation to 1 = high adaptation (6 classes)
Mitigation ratio	mitig	O: 0.16 = low mitigation to 1 = high mitigation (11 classes)
Emergency indicator	emerg	O: 0–4 emergency measures undertaken (5 classes)
Company		
Size	size	C: 1–800 employees
Spatial situation	spatial	O: 1 = business premises with more than one building 2 = one entire building used by the company 3 = at least one floor in an externally used building 4 = less than one floor in an externally used building
Response Variable		
Damage		
Relative damage of buildings	rloss	C: 0 to 1 damage ratio
Relative damage of equipment	rloss	C: 0 to 1 damage ratio
Relative damage of goods and stock	rloss	C: 0 to 1 damage ratio

[a]C: continuous; O: ordinal.

$$Ratio = \frac{Measures_{undertaken} + 1}{Measures_{relevant} + 1} \tag{1}$$

Hence, a ratio of 1 indicates that all relevant measures were implemented.

Table 2 gives an overview of all measures obtained by the survey and their classification as adaptation, mitigation, or emergency measures. Measures are classified as adaptation measures if the use or location of an asset/object is changed, that is, if an area is used in a different way or dangerous substances are relocated from areas which are prone to flooding. Measures are classified as mitigation measures if the use of an asset/object remains, but is protected in a certain way. An example of this would be the use of flood-proof oil tanks in flood-prone areas. The emergency indicator is the sum of the number emergency measures adopted, whereby eight different measures were named in the surveys and are therefore counted. However, the emergency indicator varies between zero and four, since the maximum number of emergency measures undertaken by a company was four measures. Variables describing the companies' characteristics are the number of employees and the spatial conditions of the company, indicating whether a company owns premises with more than one building or less than one floor in an externally used building. It can be assumed that the damage processes are different for businesses in a shopping street that own only a few rooms than for companies, that own entire premises. The relative loss (rloss) is calculated as the recorded asset damage divided by the recorded asset value. Damage ratios were calculated for three types of assets: (1) buildings, (2) equipment, and (3) goods and stock. The damage ratio could not be calculated for each

Table 2. Precautionary Measures and Their Classification

Classification	Measure
Adaptation	Adapted use of the flood-prone area Relocation of susceptible equipment Relocation of dangerous substances
Mitigation	Flood-proof oil tanks Flood-proof silos Flood-proof air conditioning Stable building foundation, waterproof-sealed cellar, etc.
Emergency	Emergency plan Number of emergency exercises Installation of water barriers Installation of water pumps Installation of emergency power Saving equipment and goods Preventing contamination Switching off machines, power etc.

record, since not every interviewee answered the question on the respective asset damage and/or asset value. Records with missing values for either asset damage or asset value were discarded for the respective asset. The resulting data set used for the analysis does not contain any missing values. If companies declared that certain assets were not damaged by the flood, the corresponding rloss values were assumed to be zero. Around 11% of the interviewed companies declared damage to all three asset types.

The analysis was undertaken separately for companies from different sectors. Companies were divided into four sectors following NACE Rev. 2 (Nomenclature statistique des Activites economiques dans la Communaute Europenne) according to the European statistical classification of economic activities in the European Community [*Eurostat*, 2008]: the manufacturing sector (Mining and Quarrying, Manufacturing, Electricity, Gas, Steam, and Air Conditioning Supply, Water Supply; Sewerage, Waste Management and Remediation Activities, Construction; NACE classes B-F), the commercial sector (Wholesale and Retail Trade; Repair of Motor Vehicles and Motorcycles, Transportation and Storage, Accommodation and Food Service Activities; NACE classes G-I), the financial sector (Information and Communication, Financial and Insurance Activities, Real Estate Activities, Professional, Scientific and Technical Activities, Administrative and Support Service Activities; NACE classes J-N), and the service sector (Public Administration and Defence; Compulsory Social Security, Education, Human Health and Social Work Activities, Arts, Entertainment and Recreation, other Service Activities; NACE classes O-S).

2.2. Random Forests

In this study, random forests are used to identify important damage-influencing variables by means of the variable importance and to model the flood damage. A random forest is an ensemble of *n* tree-based classifiers, whereby every tree is grown from a randomly sampled subset of the input data set. Tree-based models are suitable for flood damage modeling, as they allow for nonlinearities, predictor interactions and the use of categorical and continuous variables [*Merz et al*., 2013; *Schröter et al*., 2014; *Kreibich et al*., 2016]. In the following, we give a basic insight into tree-based classifiers and random forests. For a detailed introduction, we refer to *Breiman* [2001].

Figure 2 shows an exemplary tree to support the following introduction of random forests. The input training data sample corresponds to the root node of a single tree and is split recursively (branching) into

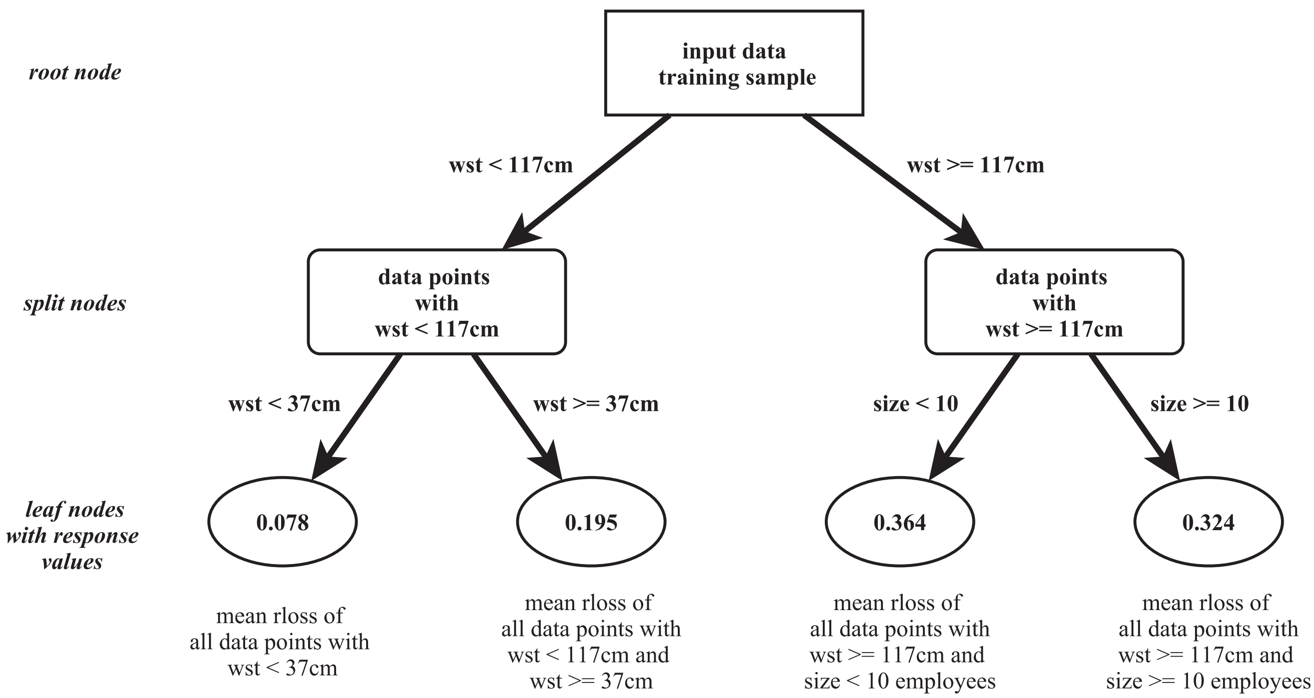

Figure 2. Exemplary representation of a single tree of a random forest for visualizing and explaining the approach. The tree consists of one root node, two split nodes and four leaf nodes.

subsamples that form the nodes of the tree. Each split is guided by a threshold value of a predictor, which is chosen such that the resulting subsamples minimize the heterogeneity of the response variable. The final subsamples form the leaf nodes, from which the response value is derived (Figure 2). For the prediction of the response variable of a certain data point the values of the predictor variables determine the leaf node that needs to be considered. For a categorical response variable (classification tree) the response value corresponds to the most frequent class of the leaf node's subsample. In case of a continuous response variable (regression tree), the mean value of the leaf node's subsample is returned (Figure 2). The predicted response value of a random forest is derived from the response values of the single classification or regression trees, by taking the mode or the mean value, respectively. As we use random forests with continuous response variables, from now on we will mainly focus on the aspects which are relevant for regression trees.

Random forests apply a bootstrap sampling called bagging internally to define the training subsamples of the single trees. Only about two thirds of the data sample is used to build a single tree, while one third of the sampled data subset is left out. The data points which are not taken into account for the training of the classifier are called Out-of-Bag observations (OOB). The OOB observations are used internally to calculate quality measures of the resulting model and to estimate the variable importance.

The literature provides different algorithms, such as the Classification And Regression Tree (CART) algorithm, THAID, C4.5 [*Quinlan*, 1986] and the Conditional Inference Tree (CIT) algorithm [*Hothorn et al*., 2006] to build the individual trees [*Wei et al*., 2015]. One of the most popular and widely used algorithms is CART. However, many studies have observed a bias in the CART algorithm with respect to variable selection toward variables with different scales and many possible splits [*Kass*, 1980; *Segal*, 1988; *White and Liu*, 1994; *Jensen and Cohen*, 2000; *Shih*, 2004; *Strobl et al*., 2007], which affects the interpretability of the models [*Hothorn et al*., 2006]. The CIT algorithm was developed to reduce this bias.

The main differences between CART and CIT are the methods used to select and split variables (splitting criterion) and to identify leaf nodes (stop criterion). CART uses an exhaustive search method on a randomly chosen set of m variables to identify the variable with the best split based on a measure of node impurity. The node impurity is usually measured as the mean square error MSE of the response values in the respective parts. The splitting is stopped either if a certain threshold of node impurity is reached or if no further splitting is possible. The OOB observations are used for an internal cross-validation, which intends to avoid overfitting. CIT makes use of hypothesis tests to identify the splitting variable at each node, whereby the dependence between the variables and the response is assessed by multiple procedure tests. At each node a randomly chosen set of variables can be used as candidate variables for splitting. The variable with the strongest association, measured by the p-value of the hypothesis test, to the response variable is selected as the splitting variable. If no association between the response variable and the covariates in the current node can be stated this node is defined as a leaf node. *Hothorn et al*. [2006] showed structural differences between the models resulting from the two algorithms, while reaching similar prediction accuracies. In addition, trees grown with the CIT algorithm are less prone to the problem of overfitting, since the variable selection and the stopping of the tree growth is done by appropriate statistical testing [*Hothorn et al*., 2006]. The algorithm CIT allows for an unbiased variable selection for variables with different scales and many possible splits, which improves the interpretability of the trees.

To our knowledge, previous studies that used regression trees for the estimation of flood damage made use of the CART algorithm [*Merz et al*., 2013; *Schröter et al*., 2014; *Hasanzadeh Nafari et al*., 2016c]. However, since the data sets used contain variables with different scales as well as many possible splits, and since an unbiased variable selection is key for the identification of damage-influencing variables, the algorithm used in this study is CIT. The analysis was done with R (version 3.3.2)—A language and environment for statistical computing [*R Core Team*, 2016]. The package "party" (version 1.2) was used to compute the random forests [*Hothorn et al.*, 2006, 2015; *Strobl et al.*, 2007]. Each random forest consists of 1000 trees (ntree = 1000) and three variables were randomly chosen as candidate variables at each node for splitting (mtry = 3). Each terminal node consists of at least seven observations.

2.3. Variable Importance

Apart from modeling applications, random forests can also be used to identify relevant predictor variables from a set of input predictor variables. The relevance can be assessed by the so-called variable importance. In the case of regression, this importance can be estimated by a random permutation of the values of the corresponding predictor variable, simulating the absence of this particular variable. The difference of the prediction error calculated by means of the OOB observations with and without the permutation indicates whether or not the predictor variable is important for the prediction. The rationale behind this is that the prediction accuracy will decrease if a relevant predictor variable is permuted randomly. Therefore, the increase of the prediction error with the permutation of the corresponding predictor variable can be interpreted as a measure for the variable importance.

2.4. Sampling of the Data Sets

Due to the large heterogeneity of the data, the learned random forest and its predictions depend on the respective data sample used for training. In addition to the sampling which takes place within the random forests, a further data sampling is applied before the training of the models to provide stable and comparable results. Hence, many random forests are trained with different data samples and the averaged results are provided in section 3. The results shown in section 3 are therefore an outcome of many differently trained random forests.

The sampling method used is the Jackknife, which was developed to assess the stability of estimates [*Rodgers*, 1999]. To assess the effect of the data set size on the model performance, the size of the samples is increased stepwise by one data point until a maximum of 75% of the respective data set is reached. The data points are sampled without replacement from the original data set. The 25% of the data set which is not used for the sample is used for the validation of the random forest, trained with this particular sample. For the calculation of the variable importance measures 75% of the data points of the respective data sets are used. The data sets are sampled 1000 times per asset, sector, and data size step. Hence, 1000 random forests are built per asset, sector and data size step.

2.5. Flood Damage Model Performance

Random forests trained with data from only one sector (sector-specific) and those trained with data from all sectors (sector-unspecific) are built and compared with each other. A leave-p-out cross-validation is performed to evaluate the results of the random forests. The validation of the predicted relative damage is done with 25% of the respective data set, which was not used for the training of the model. Three measures are used to evaluate the performance of the models:

The mean absolute error (MAE)

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|est-obs|, \tag{2}$$

The root mean square error (RMSE)

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(est-obs)^2}, \tag{3}$$

The mean bias error (MBE)

$$MBE=\frac{1}{n}\sum_{i=1}^{n}est-obs. \tag{4}$$

The MAE describes the average deviation from the predicted to the observed values, while the RMSE considers the square of the errors. Compared to the MAE, the RMSE is more strongly affected by large deviations. In the ongoing discussion on the choice of MAE or RMSE [*Willmott and Matsuura*, 2005; *Willmott et al.*, 2009], *Chai and Draxler* [2014] suggest considering both metrics in the model validation. In addition, the MBE describes a systematic overestimation or underestimation of the model.

3. Results and Discussion

The following section contains the analysis of the data and the results of the random forests. Section 3.1 provides a descriptive analysis and a short discussion of the data sets used. The results of the random forests regarding the identification of damage-influencing variables are given in section 3.2, while the performance of the models is analyzed in section 3.3. Both sections are subdivided into the three different assets (buildings, equipment, goods, and stock) and followed by a discussion on the general findings.

3.1. Descriptive Analysis of the Data Sets

Figure 3 shows the distribution of the variables represented by violin plots and the number of points available per asset and sector. The distribution of the variables were estimated by means of Kernel Density Estimator. For this overview, all variables were scaled from 0 to 1. Only a few companies from the financial and service sectors have goods and stock. Therefore, the case numbers are very small. Nonetheless, these few case numbers are analyzed and shown, but the results should be considered with caution.

The values of the relative damage vary not only between the sectors, but also between the assets. Manufacturing companies show the highest mean value for relative building damage, while commercial companies show the highest mean value for relative equipment damage. Compared to the distributions of the relative building damage, distributions of the equipment as well as the goods and stock damage show a higher number of cases in which the entirety of the equipment or goods and stock of a company was damaged. The distributions of the water level and the inundation duration are relatively similar across all sectors and assets. For all assets, the mean values for the contamination index of manufacturing and commercial companies are slightly higher than the mean values of financial and service companies. Most companies were only marginally affected by contamination, as the contamination indices are low in general. The distributions of the mitigation ratio indicate that most companies did not undertake all the mitigation measures that they considered to be relevant. There are only slight differences between the sectors and assets. However, the distributions of the adaptation ratio reveal that many companies undertook all adaptation measures that they considered to be relevant. This can on the one hand be explained by the fact that the implementation of adaptation measures, such as changing the use of flood-prone areas within the business premises, demands less effort than most mitigation measures, such as retrofit building to make them flood-proof. On the other hand, some adaptation measures are rather specific, e.g., the relocation of hazardous substances, and are therefore not relevant for all of the companies. The number of the emergency

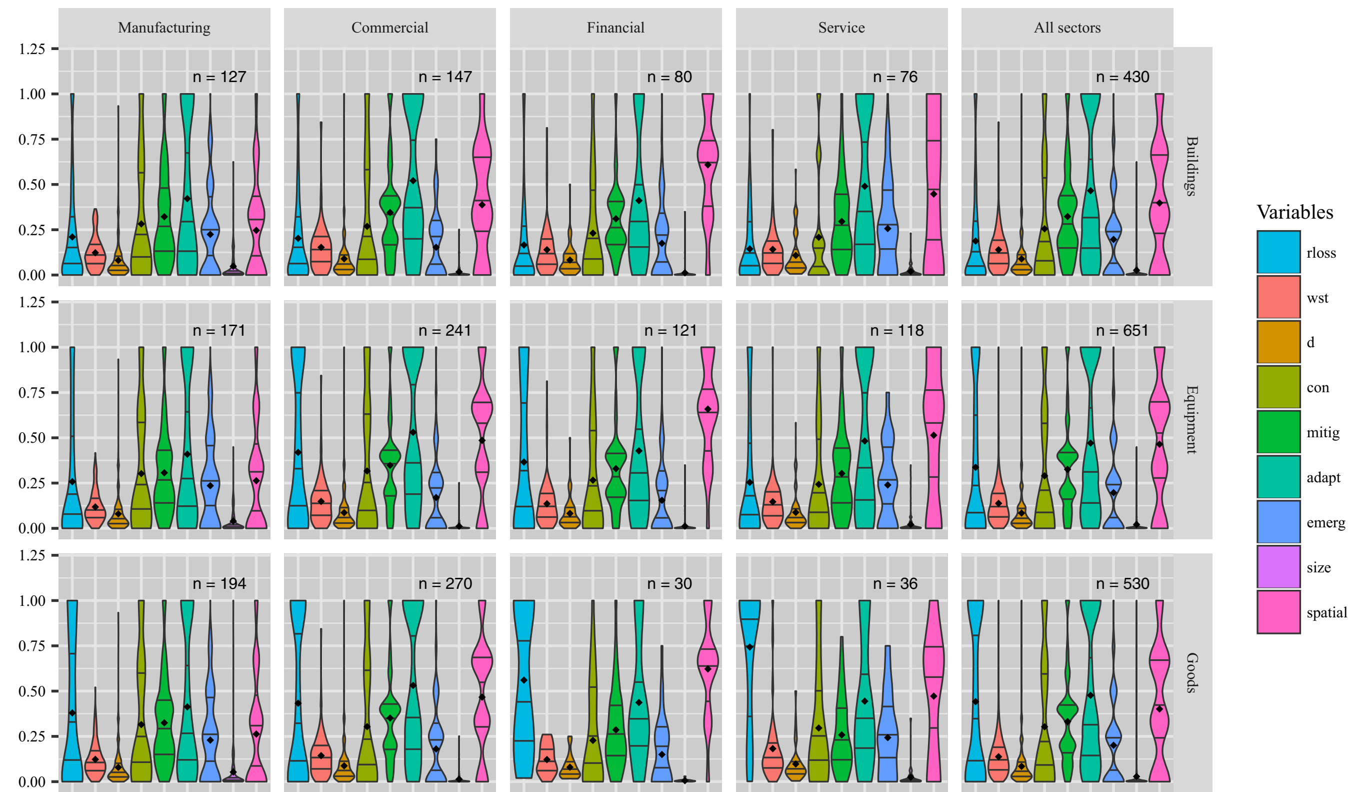

Figure 3. Kernel density estimations of the nine variables for all assets and sectors based on the available data sets. Variables are scaled from zero to one. The lines represent the first, second, and third quartiles. The dot represents the mean.

measures taken is slightly higher for manufacturing and service companies. The same can be observed for the number of employees. The distributions of the spatial situation show clear differences between the sectors. Manufacturing companies mainly own one or more buildings, while financial companies have mostly one floor or less. This is plausible, as most manufacturing companies have more employees and need space for storage and production sites.

Figure 4 shows the correlation matrices of the nine variables per asset and sector. The used correlation coefficient is Spearman's rank correlation coefficient. The first column of each matrix contains the correlation coefficients of the relative loss and the predictor variables. In general, correlation coefficients range from 0.48 to −0.50, whereas most of the correlations are around 0. Water level, inundation duration, and contamination have the highest positive correlation with the relative damage for all sectors and assets. The highest negative and significant correlations with relative damage has the variable adaptation ratio. Other variables significantly negatively correlated with relative damage are the mitigation ratio, emergency measures, and spatial situation.

The predictor variables are also correlated with each other. Since the data considered for the different assets are subsets of the same data set—sampled according to the available damage information of the respective asset—the correlations between the predictors show similar patterns over all assets of the same sector. Yet due to different sample sizes, significant correlations are partly missing especially for smaller subsamples (e.g., subsets considering the building damage and the financial or service sectors). For instance, in the manufacturing sector the significant negative correlation between adaptation ratio and contamination is not detected in the buildings subset, as well as the pairwise correlations between size, spatial situation and adaptation ratio. The correlation matrix is used to support the interpretation of the variable importances in section 3.2.

However, the correlation coefficients consider only pairwise and monotonic relationships. Random forests are able to capture nonmonotonic and multivariable relationships, since they consider dependencies

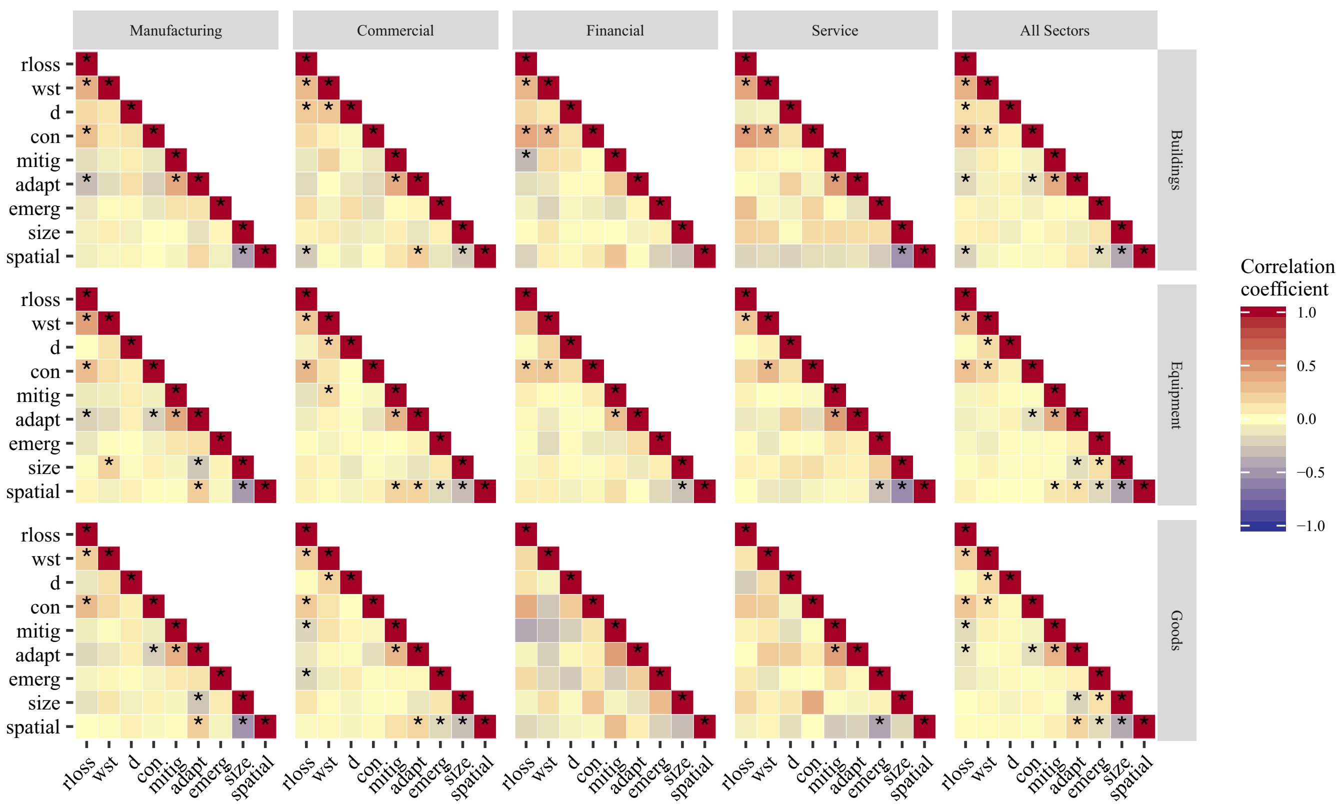

Figure 4. Pairwise Spearman correlation coefficient for the nine variables. Stars show significant correlation at the 1% significance level.

between the predictors as well. Therefore, influences of variables which cannot be detected by correlation coefficients might be detected by the variable importance measures of random forests.

3.2. Variable Importances

Figure 5 shows boxplots of variable importances of the eight predictor variables for different company sectors or all sectors together and different assets derived from random forests.

3.2.1. Buildings

The most important predictor variable for random forests predicting rloss of buildings is the water level, when considering companies from all sectors. This is consistent with many studies and existing models [*Gerl et al*., 2014; *Penning-Rowsell et al*., 2005; *Hasanzadeh Nafari et al*., 2016b]. However, the variable importance of the water level decreases in random forests trained with sector-specific data only. Furthermore, it can be observed that other variables apart from the water level influence the flood damage within the different sectors. Hence, certain influences can only be captured when distinguishing between the company sectors.

When predicting rloss of buildings from the manufacturing sector the variable adaptation ratio is slightly more important than the water level. This observation suggests that rloss is not only influenced by the water level, but also by the adaptation measures a company might have undertaken. Especially for manufacturing companies, adaptation measures such as the relocation of hazardous substances, could play an important role. This assumption is additionally supported by the negative correlation between rloss and the adaptation ratio (see Figure 4), indicating a damage-reducing effect of the abovementioned measures.

Important variables derived from random forests trained with company data from the commercial sector are the water level, spatial situation, emergency measures, and contamination. The importance of the variable spatial situation in combination with the negative correlation with rloss leads to the assumption that the damage suffered by commercial companies depends to a certain extent on their business premises.

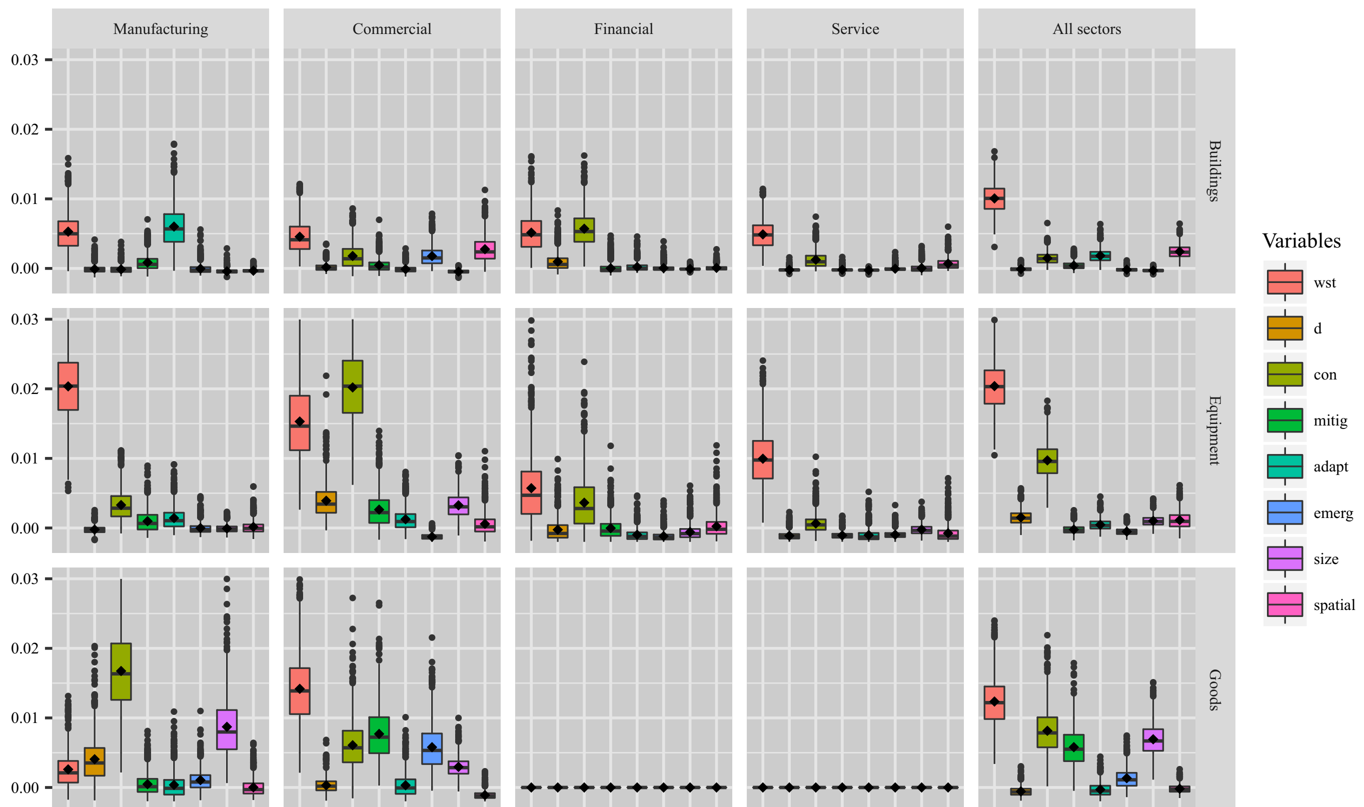

Figure 5. Variable importances derived by 1000 fandom forests trained with individually sampled data sets. The variable importance is measured by the increase of the mean squared error (MSE) with the random permutation of the respective variable. The lines in the individual boxes indicate the quartiles, while the diamond indicates the mean value of the respective variable. The number of available data points was not sufficient to estimate the variable importances for goods and stock from the financial and service sectors.

This cannot be observed for companies from other sectors. One reason could be that the spatial situations in this sector are more heterogeneous, and consequently a good separation can be reached by this variable.

The variable importances from financial and service-oriented companies are dominated by water level and contamination. A stable separation based on other variables could not be found. This could on the one hand be an effect of the relatively low number of data points available for these company sectors. On the other hand, the predictors considered might not be sufficient for damage predictions, and hidden (not yet considered) variables are needed for an adequate damage description.

3.2.2. Equipment

Water level and contamination are identified as the most important variables when predicting rloss of equipment for companies from all sectors. In general, the range of the variable importances identified for equipment damage is larger compared to those for building damage. This indicates that the heterogeneity within the data and the processes describing the damage to equipment is larger than the heterogeneity within the building data.

Considering manufacturing companies only, the variable importance of the water level is high, while the importance of contamination is lower. Relatively high variable importances can be observed for the mitigation and adaptation ratios. Measures like the adjusted use of flood-prone areas at the company site could potentially lead to a decrease in the damage suffered by the companies' equipment. In addition, the adaptation ratio and contamination are negatively correlated with each other, indicating another damage-reducing effect of the adaptation measures, like the relocation of hazardous substances.

Companies from the commercial sector show high variable importances for water level and contamination, followed by inundation duration, mitigation ratio and size. While the Spearman correlation also identifies a significant correlation of water level and contamination with the damage caused (Figure 4), the significance threshold is not reached for the remaining predictors. The importance of inundation duration, mitigation ratio, and size detected by the random forest approach might be due to a nonmonotonic relationship or—which we consider to be more likely—the impact on the relative loss becomes more obvious if multivariable dependencies are taken into account. Thus, a pairwise correlation could be blurred by interfering factors (e.g., water level), but becomes more distinct for sorted data subsets that are formed in the tree growing process.

Variable importances for the financial and service sectors are lower in general and less diverse. Water level and contamination are identified as important variables for the financial sector, while the water level is the only important variable for the service sector. Similar to the results of the building damage, the lack of identified damage drivers is assumed to be due to the small sample sizes and/or hidden variables.

3.2.3. Goods and Stock

Random forests trained with data from all sectors to estimate the damage to goods and stock identify water level, contamination, mitigation ratio, and size as the most important variables. This is the most diverse outcome compared to the other assets. Calculations of the variable importance of the damage to goods and stock for the financial and the service sectors were not possible due to the low number of data points.

The damage to goods and stock of manufacturing companies is mostly affected by contamination and the number of employees. Explaining the impact of company size is not straightforward. It might not have a direct influence on the damage, but rather an indirect influence on several damage-driving and -preventing characteristics. Thus, Figure 4 reveals a correlation between company size and the spatial situation, as well as between company size and the adaptation ratio. Further, a slight, but not significant, positive correlation between size and water level is indicated considering the correlation matrix of goods and stock, which is even evaluated to be significant considering the correlation matrix of equipment. The correlation might be explained by location preferences of the companies depending on the size. Capturing information about several damage predictors, the company size might be preferred by the tree growing algorithm as a splitting criterion, rather than having split points for each of the correlated variables. Consequently, the correlated variables, such as water level, would be used less for data splitting, which explains their relative small variable importances.

The damage to goods and stock of companies from the commercial sector is mostly influenced by the water level. The degree of contamination has an influence as well, yet this effect is lower compared to the manufacturing companies. One reason for this could be that manufacturing companies are more likely to have hazardous material or liquids on their business premises which can potentially contaminate the water and consequently the other stocks. The emergency measures taken and the mitigation ratio show a relative high variable importance as well. Thus, a more efficient protection of goods and stock seems to be possible through emergency measures for companies from the commercial sector than from the manufacturing sector. This could be due to the characteristics of commercial companies' stock. The goods might be easier to relocate within a short time period than to rearrange the warehouse of a manufacturing company.

3.2.4. Discussion of the Variable Importances

In the previous subsections, the variable importance measure of random forests is used to identify relevant predictor variables for modeling companies' flood damage. Due to a limited data availability and heterogeneity in both company and flooding characteristics, general statements, despite the obvious impact of the water level, are hard to determine. Nevertheless, a distinction between different company sectors reduces the heterogeneity of the data and reveals sector-specific damage predictors not found in the joint consideration of all sectors. Thus, not only are new potential predictor variables for flood damage estimation recognized, but also different damaging processes are displayed for the different company sectors and assets. The provided variable importances give indications about the different damage processes, but the variance for some predictors is still large and possible conclusions should be considered carefully.

The robustness of the results might be increased for a larger data set that is less prone to deviations caused by outliers. Moreover, the heterogeneity within in the data could be decreased by further separations into smaller sub-sectors, which requires an increasing size of the total data set as well.

Furthermore, important variables describing damage processes seem to be missing (not recorded), especially for companies from the financial and service sectors. Many of the mitigation and adaptation measures included in this study are most likely not relevant for the damage processes of financial and service companies. An identification of these missing variables is hardly possible within the scope of this study, but a further differentiation of the company sectors during future data collections with a subsequent analysis could reveal additional factors that need to be considered.

Overall, the separation of the company sectors leads to a better insight into companies' flood damage processes. The benefit of considering different company sectors for prediction purposes as well as the effect of available training data is considered in the following section 3.3.

3.3. Flood Damage Model Performance

This section presents and discusses the results of the validation of the random forests predicting flood damage separated by sectors. A validation of the models trained with the maximum number of available training data points is presented in Figure 6. In Figure 7 prediction errors of models built with different data set sizes are shown.

3.3.1. Model Validation

Figure 6 shows boxplots of the three validation measures RMSE, MAE, and MBE for sets of 1000 random forests predicting the damage to buildings, equipment, and goods. Random forests trained with data from only one sector (sector-specific) and trained with data from all sectors (sector-unspecific) are validated with independently sampled validation data sets and compared with each other. The validation is carried out with data from the same company sector that was used to train the model.

3.3.1.1. Buildings

The RMSEs of random forests trained with data from all company sectors have a mean value of approximately 0.25. *Merz et al*. [2013] estimated the building damage suffered by private households with bagging decision trees and regression trees. RMSEs around 0.1 were estimated, which is lower than the RMSEs estimated in this study. However, the data availability for private households is better than for companies, thus 1103 records were used to build the trees in *Merz et al*. [2013], while only 430 records were used in this study. This could lead to a better model performance. Furthermore, it can be assumed that the heterogeneity of the flood damage data for companies is higher than of the data for private households.

The mean value of the MAEs is approximately 0.18. These are lower errors compared to the validation results of the Flood Loss EstimationMOdel for the commercial sector (FLEMOcs) of *Seifert et al*. [2010], who observed an MAE of 0.23. The results of the random forests seem to be more precise, although the applied

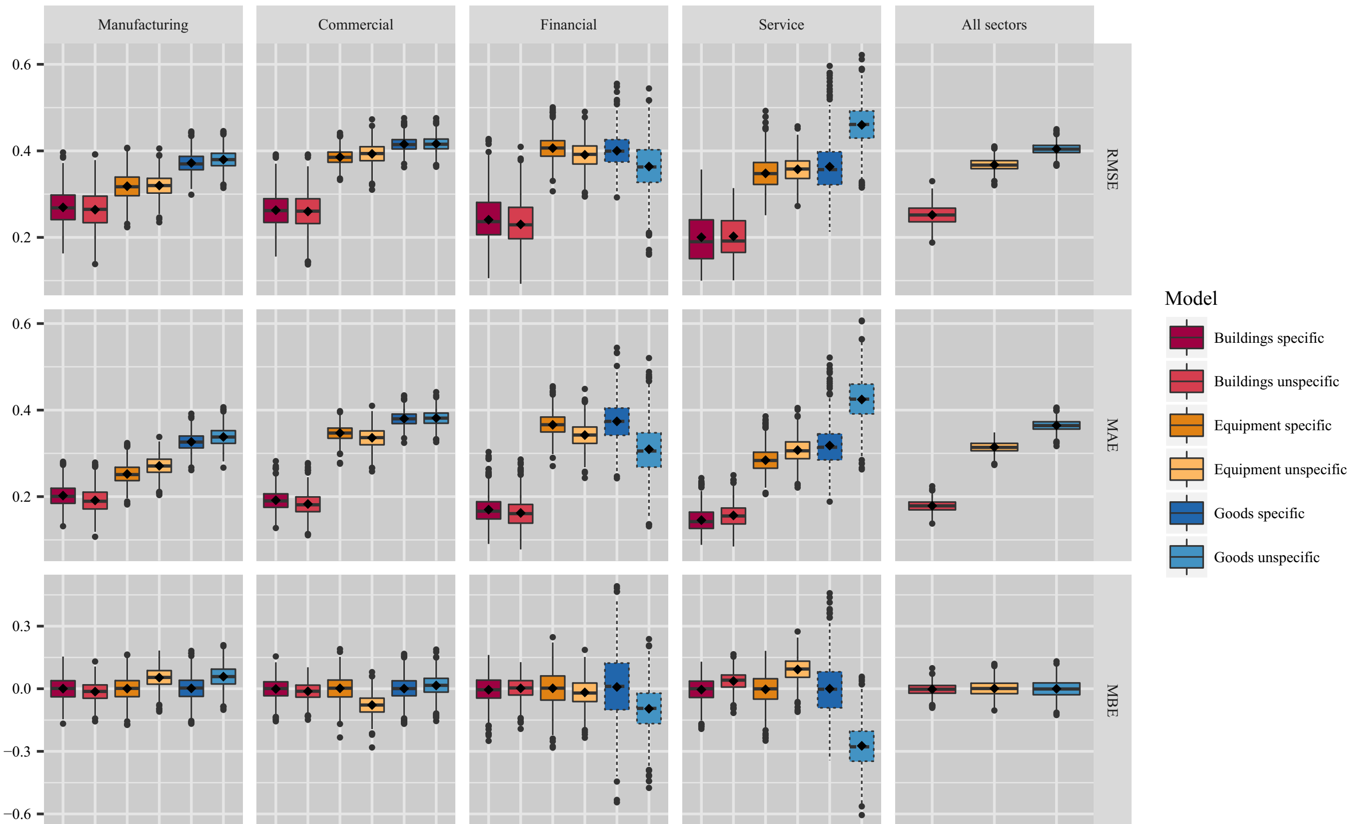

Figure 6. Validation of flood damage to buildings (red), equipment (orange), and goods and stock (blue) estimated by 1000 individually trained random forests. Brighter colors indicate that the models were trained with sector-unspecific data. Measures used for validation are the root mean square error (RMSE, top), the mean absolute error (MAE, middle) and the mean bias error (MBE, bottom). Models trained with sector-specific data were trained with less data. Note that boxplots with dashed lines were generated with only a small data sample.

validation methods are slightly different. The values of the MBE are around 0, which was also observed by *Merz et al*. [2013] and *Seifert et al*. [2010].

The mean values of the results of the validation with data from the individual sectors are not much different from the validation with data from all sectors. Yet, the variation of the results is higher. This could mainly be due to the number of data points used for validation, which is lower for the validation of the individual sectors. *Chai and Draxler* [2014] noted that the robustness of the RMSE and MAE is lower, if only a few data points are available for the calculation of the measures. Random forests trained with data points from individual sectors show similar values to those trained with data points from all sectors, when evaluated with samples from the respective company sector. Nevertheless, random forests built with data from all sectors were trained with more data.

The distributions of the MBE reveals that models trained with data from all sectors and models trained with data from a specific sector over- and underestimate the building damage in equal parts resulting in a mean MBE of around 0. Hence, there is no systematic over or underestimation of the models for the manufacturing, commercial and financial sector. However, random forests trained with data from all sectors predicting damage to buildings from the service sector show a higher mean MBE than 0 indicating a systematic overestimation.

3.3.1.2. Equipment

The validation of random forests trained with data from all sectors shows a mean RMSE of 0.37 and a mean MAE of 0.31. The performance is similar to FLEMOcs, which was validated with a RMSE of 0.37 and a MAE of 0.30 [*Seifert et al*., 2010].

The validation results for all sectors and for each company sector differ more clearly. The MAE and RMSE values are higher for the commercial and financial sectors, while the values for the manufacturing and service sectors tend to be lower. This could partly be explained by higher variances and mean values for rloss

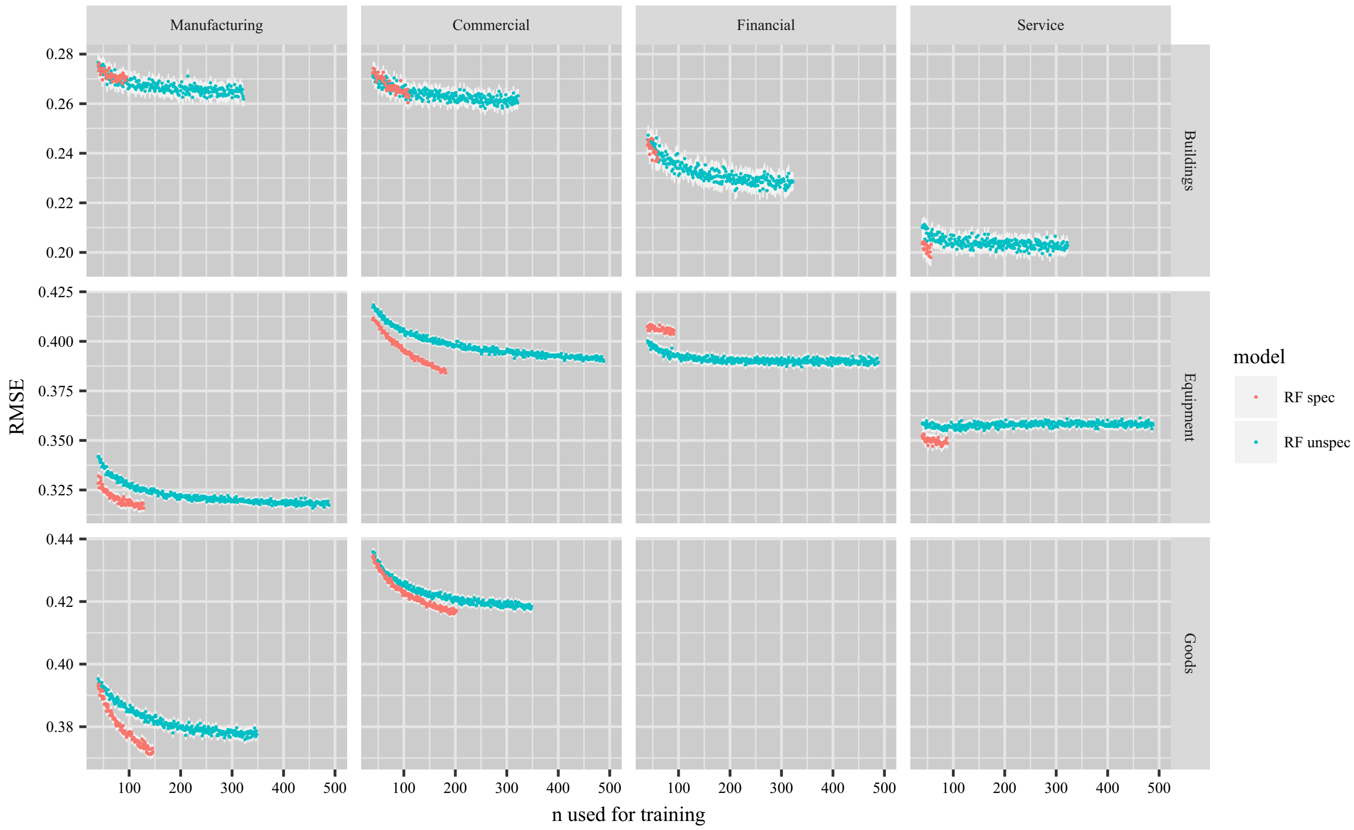

Figure 7. Mean RMSE of sector specific (red) and sector unspecific (blue) random forests trained with differently sized data sets with a 95% confidence interval (light gray).

in the commercial and financial sectors. Random forests trained with data from one sector only perform slightly better than random forests built with data from all sectors, considering the manufacturing and service sectors. The opposite can be found when considering the commercial and financial sectors. The values of the MBE indicate a systematic overestimation of random forests trained with data from all sectors predicting the damage to equipment of manufacturing and service companies and a underestimation of the damage to equipment of commercial companies. The average estimations of models trained with data from single sectors are unbiased. Better results from random forests trained with data from one single sector support the assumption that different damage processes occur in the individual company sectors. However, particularly the model performance for the financial sector seems to profit more from additional data points than from sector-specific data.

3.3.1.3. Goods and Stock

Random forests trained with data from all sectors have mean values of 0.4 for RMSE and 0.37 for MAE. These errors are larger than those reported for the validation of goods and stock from FLEMOcs with a RMSE of 0.35 and MAE of 0.31 [*Seifert et al*., 2010]. Thus, the performance of the random forests is lower compared to the estimation of rloss of goods and stock by FLEMOcs. The data sets used to derive the random forests are partly similar to those used for the derivation of FLEMOcs. FLEMOcs focuses on the event of 2002, while the RFs focus on both events from 2002 and 2013. *Schröter et al*. [2014] show that models trained with data from one event have a limited transferability to other events. This indicates event differences which are not captured within the models. Capturing two events leads to an increase in the data variability which might lead to higher validation errors.

Random forests trained with data from one sector perform either equally well or slightly better than those trained with all sectors. In contrast, random forests trained with data from all sectors show a systematic overestimation when predicting the damage to goods and stock of manufacturing companies, while models predicting the damage to goods and stock of financial and service companies show a underestimation. This cannot be observed for models built with sector-specific data, leading to the conclusion that models built with sector-specific data should be preferred.

3.3.2. Effect of Different Training Data Sets

Figure 7 shows the mean values for RMSEs for random forests predicting the flood damage with a 95% confidence interval. Every point represents the mean RMSE of 1000 random forests, each built with an individually sampled training data set of size n. The size of the training data sets is stepwise increased to evaluate the effect of additional data points used for the training of the models. Although random forests can deal with large data sets, the method is also capable to provide reasonable results when trained with small data sets [*Strobl et al.*, 2007]. The smallest training data sets contain 40 data points, while the maximum is 75% of the size of the entire data set for the respective asset and sector (see Table 3 for the absolute numbers of data points). The random forests are divided into two groups: one was trained with data from one sector only (sector-specific random forests) and the other group was trained with data from all sectors (sector-unspecific random forests). The validation is always done with data sets from one sector only.

For almost all sectors and assets, a decrease in the mean RMSE with an increase of the training data set size can be observed. Table 3 compares the prediction performance of sector-specific and sector-unspecific models trained with the same amount of data (75% of the sector-specific data), as well as sector-unspecific models trained with 75% of the entire asset-related data set. For each model considered, the percentages of the improvement in the RMSE compared to the worst performing model of the corresponding asset and sector are provided.

Considering the building damage, the distinction between sector-specific and sector-unspecific random forests hardly influences the model's prediction performance. Sector-specific random forests show a lower prediction error only in the service sector, while the errors in the other sectors are comparable. This result is plausible, since Figure 3 shows a similar distribution of the relative building damage over all sectors, which is in contrast to the large differences between the sectors for damage caused to equipment and goods. Further, it is reasonable that differences between the buildings of different companies are to a certain extend captured by the variable spatial situation, which receives the second highest importance in the random forests trained on all data. Hence, models trained with data from all sectors predict the building damage for any sector as precisely as models trained with data from the respective sector.

The mean prediction errors of sector-specific and sector-unspecific random forests estimating the equipment damage are significantly different from each other. Sector-specific models perform better than sector-unspecific models for the manufacturing, commercial and service sectors, when built with the same amount of training data points. This indicates that models trained with sector-specific data have a higher capability to model the damage accurately.

For the model performance analysis of random forests predicting damage to goods and stock, only companies from the manufacturing and commercial sectors were taken into account. The mean prediction errors

Table 3. Performance Improvement of Random Forests Trained With Sector-Specific (spec) and Unspecific (unspec) Data Sets[a]

	Manufacturing		Commercial		Financial		Service	
	n specific	n total	n specific	n total	n specific	n total	n specific	n total
Buildings	n = 95	n = 322	n = 110	n = 322	n = 60	n = 322	n = 57	n = 322
spec	2.14%		3.93%		3.94%		3.81%	
unspec	2.33%	5.34%	3.45%	4.04%	2.67%	7.67%	1.53%	4.04%
Equipment	n = 128	n = 488	n = 181	n = 488	n = 88	n = 488	n = 91	n = 488
spec	7.60%		8.20%		0.75%		3.55%	
unspec	4.63%	7.19%	4.50%	6.81%	3.70%	4.64%	1.47%	0.98%
Goods and Stock	n = 146	n = 348	n = 202	n = 348				
spec	5.91%		4.31%					
unspec	3.41%	4.50%	3.37%	4.00%				

[a]The improvement is shown as the relative decrease of the prediction error of random forests trained with all training data points available for the respective model compared to the highest prediction error of random forests predicting damage to the respective asset and sector. The columns "n specific" allow for a comparison between the performance of sector-specific and sector-unspecific models trained with the same amount of data points, whereas the columns "n total" present the improvement of sector-unspecific models trained with all available training data points.

of the sector-specific models are lower than those from the sector-unspecific models when trained with the same amount of data.

3.3.3. Discussion of the Flood Damage Model Performance

The validation of flood damage estimations by random forests shows reasonable results compared to other recently published models [*Seifert et al.*, 2010; *Merz et al.*, 2013]. Yet the prediction errors are still relatively large, due to the high variation of damage values. Especially the damage to equipment of commercial companies as well as the damage to goods and stock of manufacturing and commercial companies follows a bimodal distribution, with the highest probabilities at the domain boundaries (Figure 3). A separation of both modes based on the predictors considered is only realized to a certain extent. Subsequently, the derived models that aim to minimize the mean squared error, provide estimates between the two peaks of the distribution, that differ more or less strongly from the observed values.

The variation in model performance is quite large and strongly depends on the data sampled for model training and validation. With respect to that high variation, the performance improvement of sector-specific compared to sector-unspecific random forests is rather small. Yet the trend shows that models trained with specific data may outperform models trained with more, but unspecific, data.

The variation of the response variable that cannot be explained by the considered predictors indicates the existence of further predictor variables that have not been yet considered. Consequently, future flood damage studies should aim for the identification of as yet hidden damage-driving or -preventing factors, instead of merely increasing the amount of data. The observed improvement in sector-specific modeling suggests a stronger focus on variables that characterize individual companies and their assets. These variables could be e.g. information about the type of equipment, details about warehouses or specific characteristics about the companies' spatial situation. A further differentiation of the company sectors into subsectors could facilitate the specification. To support the higher model complexity that arises with additional predictors and to provide a representative data sample, an extended amount of data is suggested as well.

4. Conclusion

The objective of this paper was to study the flood damage caused to company buildings, equipment, and goods and stock with respect to two aspects. The first aspect is the identification of damage-influencing variables for company assets in general, as well as identifying different damage drivers for specific company sectors. The second aspect is the analysis of the flood damage model performance with respect to a sector-specific or unspecific consideration and with regard to the size of the available data set.

The most important variables identified are the water level, contamination, precautionary measures adopted, and the number of company employees. Differences between the sectors and assets can be found in terms of the identified important variables. For instance, adaptation measures taken are an important predictor variable for the building damage sustained by manufacturing companies, whereas the estimation of the building damage to commercial companies is influenced by the adopted emergency measures and the spatial situation of the company. These findings indicate that damage processes are different between the company sectors. The water level is identified as the most important variable. However, other variables are important as well, especially with regard to the damage to equipment and goods and stock. This supports the conclusion, drawn by other studies already, that water level is not sufficient to estimate the company damage caused by flooding.

For the analysis of the flood damage model performance, random forests trained with data from all sectors and those trained with data from only one sector were validated and compared with one another. Furthermore, the effect of different training data set sizes was investigated. Sector-specific models predicting damage to equipment and goods and stock mainly showed a lower prediction error when trained with the same amount of data points. Even with more training data, the performance of the sector-unspecific models was either equal to or lower than the performance of sector-specific models. Subsequently, models trained with more, but sector-unspecific, data do not necessarily result in more precise predictions. Future data collections should consequently focus on a sector-specific, detailed and reliable data acquisition to allow for a consideration of company-specific characteristics.

It can be concluded that the identification of damage drivers and processes remains difficult, not least because of the limited data. Yet a sector-specific consideration reduces the heterogeneity in the data and helps to reveal new predictor variables. A sector-specific adaptation of damage models improves the prediction quality of the flood-related damage to all company assets considered.

Acknowledgments
The present work was developed within the framework of the Research Training Group "Natural Hazards and Risks in a Changing World" (NatRiskChange; GRK 2043/1) funded by the Deutsche Forschungsgemeinschaft (DFG). The surveys were conducted through a joint venture between the GeoForschungsZentrum Potsdam, the Deutsche Rückversicherung AG, Düsseldorf, and the University of Potsdam. Besides the partners' own resources, additional funds were provided by the German Ministry for Education and Research (BMBF) as part of the following research projects: DFNK 01SFR9969/5, and Flood 2013 13N13017. The data sets of the 2002 flood are available via the German flood damage database HOWAS21 (http://howas21.gfz-potsdam.de/howas21/). The data sets of the 2013 flood will be implemented into HOWAS21 by the end of 2017. The authors thank Annegret Thieken for her valuable comments on the manuscript. Furthermore, we thank the Editor and the reviewers for their constructive comments on the manuscript.

References

Apel, H., G. T. Aronica, H. Kreibich, and A. H. Thieken (2009), Flood risk analyses—How detailed do we need to be?, *Nat. Hazards*, *49*(1), 79–98, doi:10.1007/s11069-008-9277-8.

Breiman, L. (2001), Random forests, *Machine Learning*, *45*(1), 5–32, doi:10.1023/A:1010933404324.

Bubeck, P., and H. Kreibich (2011), Natural hazards: Direct costs and losses due to the disruption of production processes, technical report, ConHaz WP1 Final report, Potsdam. [Available at https://www.ufz.de/export/data/2/122161_CONHAZ%20REPORT%20WP01_2.pdf.]

Büchele, B., H. Kreibich, A. Kron, A. Thieken, J. Ihringer, P. Oberle, B. Merz, and F. Nestmann (2006), Flood-risk mapping: Contributions towards an enhanced assessment of extreme events and associated risks, *Nat. Hazards Earth Syst. Sci.*, *6*(4), 483–503, doi:10.5194/nhess-6-485-2006.

Bundesministerium des Innern (2013), Bericht zur flutkatastrophe 2013: Katastrophenhilfe, entschädigung, wiederaufbau [in German], technical report, Bundesministerium des Innern, Berlin. [Available at http://www.bmi.bund.de/SharedDocs/Downloads/DE/Broschueren/2013/kabinettbericht-fluthilfe.html.]

Chai, T., and R. R. Draxler (2014), Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature, *Geosci. Model Dev.*, *7*, 1247–1250, doi:10.5194/gmd-7-1247-2014.

de Moel, H., B. Jongman, H. Kreibich, B. Merz, E. Penning-Rowsell, and P. J. Ward (2015), Flood risk assessments at different spatial scales, *Mitigation Adapt. Strategies Global Change*, *20*(6), 865–890, doi:10.1007/s11027-015-9654-z.

Emschergenossenschaft and Hydrotec (2004), Hochwasser-Aktionsplan Emscher, Kapitel 1: Methodik der schadensermittlung [in German], technical report, Emschergenossenschaft/Hydrotec, Essen. [Available at http://www.eglv.de/fileadmin/Medien/Dokumente/PDF/Anlagen/anlagen_methodik_schadensermittlung.pdf.]

Eurostat (2008), Statistical classification of economic activities in the European Community, NACE Rev. 2. [Available at http://ec.europa.eu/eurostat/ramon/.]

Falter, D., K. Schröter, N. V. Dung, S. Vorogushyn, H. Kreibich, Y. Hundecha, H. Apel, and B. Merz (2015), Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain, *J. Hydrol.*, *524*, 182–193, doi:10.1016/j.jhydrol.2015.02.021.

Gerl, T., M. Bochow, and H. Kreibich (2014), Flood damage modeling on the basis of urban structure mapping using high-resolution remote sensing data, *Water*, *6*(8), 2367–2393, doi:10.3390/w6082367.

Grigg, N. S., and O. J. Helweg (1975), State of the art of estimating flood damage in urban areas, *Water Resour. Bull.*, *11*(2), 379–390, doi:10.1111/j.1752-1688.1975.tb00689.x.

Hasanzadeh Nafari, R., T. Ngo, and W. Lehman (2016a), Calibration and validation of FLFA rs—A new flood loss function for Australian residential structures, *Nat. Hazards Earth Syst. Sci.*, *16*, 15–27, doi:10.5194/nhess-16-15-2016.

Hasanzadeh Nafari, R., T. Ngo, and W. Lehman (2016b), Development and evaluation of FLFAcs—A new flood loss function for Australian commercial structures, *Int. J. Disaster Risk Reduction*, *17*, 13–23, doi:10.1016/j.ijdrr.2016.03.007.

Hasanzadeh Nafari, R., T. Ngo, and P. Mendis (2016c), An assessment of the effectiveness of tree-based models for multi-variate flood damage assessment, *Water*, *8*, 282, doi:10.3390/w8070282.

Hothorn, T., K. Hornik, and A. Zeileis (2006), Unbiased recursive partitioning: A conditional inference framework, *J. Comput. Graphical Stat.*, *15*(3), 651–674, doi:10.1198/106186006X133933.

Hothorn, T., P. Bühlmann, S. Dudoit, A. Molinaro, and M. J. Van Der Laan (2006), Survival Ensembles, *Biostatistics*, *7*, 3, 355–373, doi:10.1093/biostatistics/kxj011.

Hothorn, T., K. Hornik, and A. Zeileis (2015), party: A Laboratory for Recursive Partytioning, in R Package Version 0.9–0. [Available at http://CRAN.R-project.org.]

Hudson, P., W. J. W. Botzen, H. Kreibich, P. Bubeck, and J. C. J. H. Aerts (2014), Evaluating the effectiveness of flood damage mitigation measures by the application of propensity score matching, *Nat. Hazards Earth Syst. Sci. Discuss.*, *2*, 1731–1747, doi:10.5194/nhessd-2-681-2014.

Huttenlau, M., J. Stöffel, and H. Stiefelmeyer (2010), Risk-based damage potential and loss estimation of extreme flooding scenarios in the Austrian Federal Province of Tyrol, *Nat. Hazards Earth Syst. Sci.*, *10*, 2451–2473, doi:10.5194/nhess-10-2451-2010.

Jensen, D. D., and P. R. Cohen (2000), Multiple comparisons in induction algorithms, *Machine Learning*, *38*, 309–338, doi:10.1023/A:1007631014630.

Jongman, B., H. C. Winsemius, J. C. J. H. Aerts, E. Coughlan, D. Perez, and M. K. V. Aalst (2015), Declining vulnerability to river floods and the global benefits of adaptation, *Proc. Natl. Acad. Sci. U. S. A.*, *112*, 2271–2280, doi:10.1073/pnas.1414439112.

Kass, G. V. (1980), An exploratory technique for investigating large quantities of categorical data, *Appl. Stat.*, *29*(2), 119–127, doi:10.2307/2986296.

Kim, Y.-O., S. B. Seo, and O.-J. Jang (2012), Flood risk assessment using regional regression analysis, *Nat. Hazards*, *63*(2), 1203–1217, doi:10.1007/s11069-012-0221-6.

Koks, E. E., H. de Moel, J. C. J. H. Aerts, and L. M. Bouwer (2014a), Effect of spatial adaptation measures on flood risk: Study of coastal floods in Belgium, *Reg. Environ. Change*, *14*, 413–425, doi:10.1007/s10113-013-0514-7.

Koks, E. E., B. Jongman, T. G. Husby, and W. J. W. Botzen (2014b), Combining hazard, exposure and social vulnerability to provide lessons for flood risk management, *Environ. Sci. Policy*, *47*, 42–52, doi:10.1016/j.envsci.2014.10.013.

Kreibich, H., A. H. Thieken, and B. Merz (2005), Precautionary measures reduce flood losses of households and companies—Insights from the 2002 flood in Saxony, Germany, in *Floods, from Defence to Management, edited by* E. Van Beek and M. Taal, pp. 851–860, Taylor and Francis, London.

Kreibich, H., M. Müller, A. H. Thieken, and B. Merz (2007), Flood precaution of companies and their ability to cope with the flood in August 2002 in Saxony, Germany, *Water Resour. Res.*, *43*, W03408, doi:10.1029/2005WR004691.

Kreibich, H., I. Seifert, B. Merz, and A. H. Thieken (2010), Development of FLEMOcs—A new model for the estimation of flood losses in the commercial sector, *Hydrol. Sci. J.*, *55*(8), 1302–1314, doi:10.1080/02626667.2010.529815.

Kreibich, H., A. Botto, B. Merz, and K. Schroeter (2016), Probabilistic, multivariable flood loss modeling on the mesoscale with BT-FLEMO, *Risk Analysis*, *37*, 774–787, doi:10.1111/risa.12650.

Mechler, R., and L. M. Bouwer (2014), Understanding trends and projections of disaster losses and climate change: Is vulnerability the missing link?, *Clim. Change*, *133*, 23–35, doi:10.1007/s10584-014-1141-0.

Merz, B., H. Kreibich, R. Schwarze, and A. Thieken (2010), Assessment of economic flood damage, *Nat. Hazards Earth Syst. Sci.*, *10*(8), 1697–1724, doi:10.5194/nhess-10-1697-2010.

Merz, B., H. Kreibich, and U. Lall (2013), Multi-variate flood damage assessment: A tree-based data-mining approach, *Nat. Hazards Earth Syst. Sci.*, *13*(1), 53–64, doi:10.5194/nhess-13-53-2013.

Merz, B., F. Elmer, M. Kunz, B. Mühr, K. Schröter, and S. Uhlemann-Elmer (2014), The extreme flood in June 2013 in Germany, *La Houille Blanche*, *1*, 5–10, doi:10.1051/lhb/2014001.

Meyer, V., et al. (2013), Assessing the costs of natural hazards-state of the art and knowledge gaps, *Nat. Hazards Earth Syst. Sci.*, *13*(5), 1351–1373, doi:10.5194/nhess-13-1351-2013.

Molinari, D., S. Menoni, G. T. Aronica, F. Ballio, N. Berni, C. Pandolfo, M. Stelluti, and G. Minucci (2014), Ex post damage assessment: An Italian experience, *Nat. Hazards Earth Syst. Sci.*, *14*(4), 901–916, doi:10.5194/nhess-14-901-2014.

MURL (Ministerium für Umwelt, Raumordnung und Landwirtschaft des Landes Nordrhein-Westfalen) (2000), Potentielle Hochwasserschäden am Rhein in Nordrhein-Westfalen, Düsseldorf [in German], technical report, MURL, Düsseldorf. [Available at http://nr-feldmann.de/Doku/Behoerden/Stresstest%203.2-2000.pdf.]

Penning-Rowsell, E. C. and E. B. Chatterton (1977), *The Benefits of Flood Alleviation: A Manual of Assessment Techniques*, Gower Tech. Press, Aldershot, U. K.

Penning-Rowsell, E., C. Johnson, S. Tunstall, S. Tapsell, J. Morris, J. Chatterton, and C. Green (2005), *The Benefits of Flood and Coastal Erosion Risk Management: A Handbook of Assessment Techniques*, Middlesex Univ. Press, London.

Poussin, J. K., W. J. W. Botzen, and J. C. J. H. Aerts (2015), Effectiveness of flood damage mitigation measures: Empirical evidence from French flood disasters, *Global Environ. Change*, *31*, 74–84, doi:10.1016/j.gloenvcha.2014.12.007.

Quinlan, J. R. (1986), Induction of decision trees, *Machine Learning*, *1*, 81–106.

R Core Team (2016), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna.

Rodgers, J. L. (1999), The bootstrap, the jackknife, and the randomization test: A sampling taxonomy, *Multivariate Behavioral Res.*, *34*(4), 441–456, doi:10.1207/S15327906MBR3404.

Scawthorn, C., et al. (2006), HAZUS-MH flood loss estimation methodology. II. Damage and loss assessment, *Nat. Hazards Rev.*, *7*(2), 72–81, doi:10.1061/(ASCE)1527-6988(2006)7:2(72).

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

Schröter, K., M. Kunz, F. Elmer, B. Mühr, and B. Merz (2015), What made the June 2013 flood in Germany an exceptional event? A hydrometeorological evaluation, *Hydrol. Earth Syst. Sci.*, *19*(1), 309–327, doi:10.5194/hess-19-309-2015.

Schröter, K., S. Lüdtke, K. Vogel, H. Kreibich, and B. Merz (2016), Tracing the value of data for flood loss modelling, in *3rd European Conference on Flood Risk Management (FLOODRISK 2016)*, Issue E3S Web Conf., 7, 05005, 1-5, EDP Sciences - Web of Conferences, Les Ulis Cedex A, France, doi:10.1051/e3sconf/20160705005.

Segal, M. R. (1988), Regression trees for censored data, *Biometrics*, *44*(1), 35–47.

Seifert, I., H. Kreibich, B. Merz, and A. H. Thieken (2010), Application and validation of FLEMOcs—A flood-loss estimation model for the commercial sector, *Hydrol. Sci. J.*, *55*(8), 1315–1324, doi:10.1080/02626667.2010.529815.

Shih, Y.-S. (2004), A note on split selection bias in classification trees, *Comput. Stat. Data Anal.*, *45*, 457–466.

Smith, D. (1994), Flood damage estimation—A review of urban stage-damage curves and loss functions, *Water SA*, *20*(3), 231–238.

Strobl, C., A.-L. Boulesteix, A. Zeileis, and T. Hothorn (2007), Bias in random forest variable importance measures: Illustrations, sources and a solution, *BMC Bioinf.*, *8*, 25, doi:10.1186/1471-2105-8-25.

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

Thieken, A. H., A. Olschewski, H. Kreibich, S. Kobsch, and B. Merz (2008), Development and evaluation of FLEMOps—A new flood loss estimation model for the private sector, *WIT Trans. Ecol. Environ.*, *118*, 315–324, doi:10.2495/FRIAR080301.

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

Thieken, A., H. Kreibich, M. Müller, and J. Lamond (2017), Data collection for a better understanding of what causes flood damage—Experiences with telephone surveys, in *Flood Damage Survey and Assessment: New Insights From Research and Practice*, vol. *228*, *Geophys. Monogr. Ser.*, edited by D. Molinari, S. Menoni, F. Ballio, AGU, Washington, D. C.

Vogel, K., C. Riggelsen, B. Merz, H. Kreibich, and F. Scherbaum (2012), Flood damage and influencing factors: A Bayesian network perspective, in *6th European Workshop on Probabilistic Graphical Models (PGM 2012)*, Univ. of Granada, Granada, Spain.

Ward, P. J., B. Jongman, F. S. Weiland, A. Bouwman, R. V. Beek, M. F. P. Bierkens, W. Ligtvoet, and H. Winsemius (2013), Assessing flood risk at the global scale: Model setup, results, and sensitivity, *Environ. Res. Lett.*, *8*, 1–10, doi:10.1088/1748-9326/8/4/044019.

Wei, P., Z. Lu, and J. Song (2015), Variable importance analysis: A comprehensive review, *Reliability Eng. Syst. Safety*, *142*, 399–432, doi:10.1016/j.ress.2015.05.018.

White, A. P., and W. Z. Liu (1994), Bias in information-based measures in decision tree induction, *Machine Learning*, *15*, 321–329.

Willmott, C. J., and K. Matsuura (2005), Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, *Clim. Res.*, *30*, 79–82.

Willmott, C. J., K. Matsuura, and S. M. Robeson (2009), Ambiguities inherent in sums-of-squares-based error statistics, *Atmos. Environ.*, *43*, 749–752, doi:10.1016/j.atmosenv.2008.10.005.

Winsemius, H., L. V. Beek, B. Jongman, P. J. Ward, and A. Bouwman (2013), A framework for global river flood risk assessments, *Hydrol. Earth Syst. Sci.*, *17*, 1871–1892, doi:10.5194/hess-17-1871-2013.

Zhai, G., T. Fukuzono, S. Ikeda, Z. Guofang, T. Fukuzono, and S. Ikeda (2005), Modeling flood damage: Case of Tokai flood 2000, *J. Am. Water Resour. Assoc.*, *41*, 77–92, doi:10.1111/j.1752-1688.2005.tb03719.x.