

Zakia Sultana | Tobias Sieg | Patric Kellermann | Meike Müller |  
Heidi Kreibich

# Assessment of Business Interruption of Flood-Affected Companies Using Random Forests

Suggested citation referring to the original publication:

Water 10 (2018) 8, 1049

DOI <https://doi.org/10.3390/w10081049>

ISSN (online) 2073-4441

Postprint archived at the Institutional Repository of the Potsdam University in:

Postprints der Universität Potsdam

Mathematisch-Naturwissenschaftliche Reihe ; 939

ISSN 1866-8372



<https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-459778>

DOI <https://doi.org/10.25932/publishup-45977>



Article

# Assessment of Business Interruption of Flood-Affected Companies Using Random Forests

Zakia Sultana <sup>1,2</sup> , Tobias Sieg <sup>1,3</sup>, Patric Kellermann <sup>1,\*</sup> , Meike Müller <sup>4</sup> and Heidi Kreibich <sup>1</sup>

<sup>1</sup> GFZ German Research Centre for Geosciences, Section 5.4 Hydrology, Telegrafenberg, 14473 Potsdam, Germany; torri\_ku@yahoo.com (Z.S.); tsieg@gfz-potsdam.de (T.S.); heidi.kreibich@gfz-potsdam.de (H.K.)

<sup>2</sup> TU-Darmstadt, Tropical Hydrogeology and Environmental Engineering, Schnittspahnstraße 9, 64287 Darmstadt, Germany

<sup>3</sup> Institute of Earth and Environmental Science, University of Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany

<sup>4</sup> Deutsche Rückversicherung Aktiengesellschaft, Hansaallee 177, 40549 Düsseldorf, Germany; meike.mueller@deutscherueck.de

\* Correspondence: patkell@gfz-potsdam.de; Tel.: +49 331 288 1721

Received: 19 June 2018; Accepted: 6 August 2018; Published: 7 August 2018



**Abstract:** Losses due to floods have dramatically increased over the past decades, and losses of companies, comprising direct and indirect losses, have a large share of the total economic losses. Thus, there is an urgent need to gain more quantitative knowledge about flood losses, particularly losses caused by business interruption, in order to mitigate the economic loss of companies. However, business interruption caused by floods is rarely assessed because of a lack of sufficiently detailed data. A survey was undertaken to explore processes influencing business interruption, which collected information on 557 companies affected by the severe flood in June 2013 in Germany. Based on this data set, the study aims to assess the business interruption of directly affected companies by means of a Random Forests model. Variables that influence the duration and costs of business interruption were identified by the variable importance measures of Random Forests. Additionally, Random Forest-based models were developed and tested for their capacity to estimate business interruption duration and associated costs. The water level was found to be the most important variable influencing the duration of business interruption. Other important variables, relating to the estimation of business interruption duration, are the warning time, perceived danger of flood recurrence and inundation duration. In contrast, the amount of business interruption costs is strongly influenced by the size of the company, as assessed by the number of employees, emergency measures undertaken by the company and the fraction of customers within a 50 km radius. These results provide useful information and methods for companies to mitigate their losses from business interruption. However, the heterogeneity of companies is relatively high, and sector-specific analyses were not possible due to the small sample size. Therefore, further sector-specific analyses on the basis of more flood loss data of companies are recommended.

**Keywords:** business interruption; floods; Random Forests; companies; variable importance

## 1. Introduction

Losses due to floods have dramatically increased over the past decades and amount to an estimated global annual average loss of US \$104 billion [1,2]. Flood events hold, with about 50%, the highest share of economic losses due to natural hazards during the last 6 decades in Germany [3]. The losses of companies, including losses due to business interruption, contribute a large portion to it [4]. To mitigate flood losses, continuous improvement in flood risk management is necessary.

For an optimal allocation of funds, cost-benefit analyses are performed, of which a central part is the estimation of benefits, i.e., averted flood losses. Cost-benefit analyses, which exclude certain loss categories, such as losses due to business interruption, lead to sub-optimal decisions [5]. Thus, quantitative knowledge about processes that determine business interruption, as well as models to estimate business interruption time and resultant losses, are necessary. However, even though business interruption losses are expected to exceed the direct flood losses of companies [6,7], little is known about it and only a few models for estimating time of, or losses due to, business interruption are available.

The objective of this study is to improve our understanding of the processes during floods that lead to business interruption and resultant losses in Germany. That is, to identify the most important variables determining business interruption duration and costs. On this basis, multivariable models for the estimation of business interruption duration and costs are developed and validated.

## 2. Literature Review

Losses due to business interruption occur in industry and commerce, in areas that are directly affected by floods. They occur due to the immediate flood impact but do not necessarily result from a physical contact between the inundation and assets, but also from the interruption of business processes, which often last much longer than the direct impact of the flood [8]. For instance, business interruptions take place if employees are not able to do their job, since their workplace is destroyed and they do not have access to an alternative working site. Business interruption losses are sometimes referred to as direct damage, as they occur due to the immediate impact of the hazard (see e.g., [9,10]) but are also sometimes referred to as primary indirect damage because the losses do not result from physical damage to property but from the interruption of economic processes (e.g., [11]). However, models to estimate losses due to business interruption, those used for direct or indirect damage, are different from both.

Common business interruption losses refer to the loss of revenue from the reduction of the flow of services, economic output or profit [6]. However, business interruption losses may be modelled as losses to stocks, e.g., when they are calculated as a fixed ratio of property losses, as conducted by the models of Anuflood [12] and Rapid Appraisal Method (RAM) [13], and as losses to flows. Stocks refer to a quantity of, e.g., money at a single point in time. Flows are defined as the outputs or services of stocks over time [14]. In most models, business interruption losses are estimated as losses of flows for a certain period of time. As a measure of the sum of flows in a company, the value added is often used [15].

Various variables are taken into account by existing models to define the susceptibility of production processes to floods. Impact variables considered are the water level (e.g., [16–18]), flood duration (e.g., [17,18]), and return period (e.g., [19,20]). For instance, [20] using four classes of return period, e.g., it is estimated that floods with a return period of 50 years lead to a period of business interruption of two months. Resistance variables taken into account are differences in the economic sectors [17–19], number of employees [16] and the value added [18,20]. For example, the model of [18] distinguishes 16 different manufacturing branches and 17 different branches in retail, distribution, office and leisure services. Unfortunately, in many studies, it remains rather unclear on which basis these variables and their quantitative influence on the business interruption were identified and quantified.

In recent years, several studies have demonstrated that machine learning approaches—particularly tree-based methods—have a good performance in relation to determining flood damage-influencing factors, achieving a more precise description of the damage processes. For instance, [21] applied bagging decision trees and regression trees to quantify the importance of various factors for the amount of flood damage to residential buildings. The study in [22] used Random Forests to calculate the variable importance for the damage estimation of various company sectors and assets. Both claimed that tree-based models are particularly suitable to the analysis of flood damage

processes, as they are able to capture nonlinear and non-monotonous dependencies between predictor and response variables, and they take interactions between the predictors into account. Thus, in this study, we also used Random Forests to identify the most important variables determining business interruption duration and costs.

Three main approaches to estimating losses due to business interruption can be distinguished [8]: (1) Applying sector-specific reference values, e.g., loss of added value per employee and day [19]; (2) comparisons of production output between flood and non-flood years (e.g., [23]); and (3) approaches that calculate production losses using a fixed share of direct damages (e.g., [12,13]). The latter two approaches are rather coarse and involve more uncertainties than the first. They are therefore particularly useful for rapid assessments in the case of, for example, emergency planning and budgeting [8].

Examples for the first approach are the following: [16] Present a probabilistic model for estimating the business interruption loss of industrial sectors caused by urban flooding in Japan. They use functional fragility curves and accelerated failure time models to estimate the extent of damage to production capacity and production-affected time, including stagnation and recovery time. The study in [24] developed a semi-quantitative framework to assess the entrepreneurial and regional-economic flood impacts of one specific production facility. Their approach relies mainly on a quantitative flood hazard modelling, resulting in a detailed inundation area and water level maps for the commercial area, as well as a rather qualitative vulnerability assessment based on co-development with the company. The US model, HAZUS, calculated monetary business interruption losses in summing up relocation expenses, which “include the cost of shifting and transferring, and the rental of temporary space”, rental income losses and capital-related losses, which reflect the income losses of the proprietor of the company [17]. All these loss types depend on business interruption duration, as well as on several “cost per day and area” factors, such as the proprietors’ income loss per day and square foot. These examples explain the conclusion of [8], i.e., that business interruption loss models are diverse, usually rather simplistic, and often non-transparent and unvalidated.

### 3. Data and Methods

#### 3.1. Flood Event data

##### 3.1.1. Description of Flood in June 2013

In June 2013, large-scale flooding occurred in many Central European countries, i.e., in Switzerland, Austria, the Czech Republic, Slovakia, Poland, Hungary, Croatia, Serbia and, particularly, in Germany, where almost all main river basins were affected [25–27].

The event of 2013 was especially characterized by extraordinarily high antecedent moisture. During the second half of May 2013, exceptional rainfall amounts, due to a quasi-stationary upper-level trough over Central Europe, had been witnessed. This circulation pattern triggered a sequence of surface lows on its eastern side that repeatedly transported warm and humid air from South-eastern Europe to Central Europe [27]. By the end of May, rainfall was at 200% of the average monthly amount (1961–1990) in large areas of Germany. Regionally, more than 300% was reached [28]. The intense and widespread precipitation that finally triggered the June 2013 flood occurred at the end of May and beginning of June. It was caused by a cut-off low that slowly moved, with its center, from France (29th May) over Northern Italy (30th May) to Eastern Europe (1st June). Overall, a combination of large-scale lifting, orographically-induced lifting and embedded convection resulted in persistent and widespread rainfall. The most intense precipitation occurred in the Danube catchment in the alpine areas of Southern Bavaria and Northern Austria [27]. For example, at the weather station Aschau-Stein of the German Weather Service in the Chiemgau Alps, a rainfall total of 405.1 mm within 96 h (the 30th of May to the 2nd of June) was registered [28].

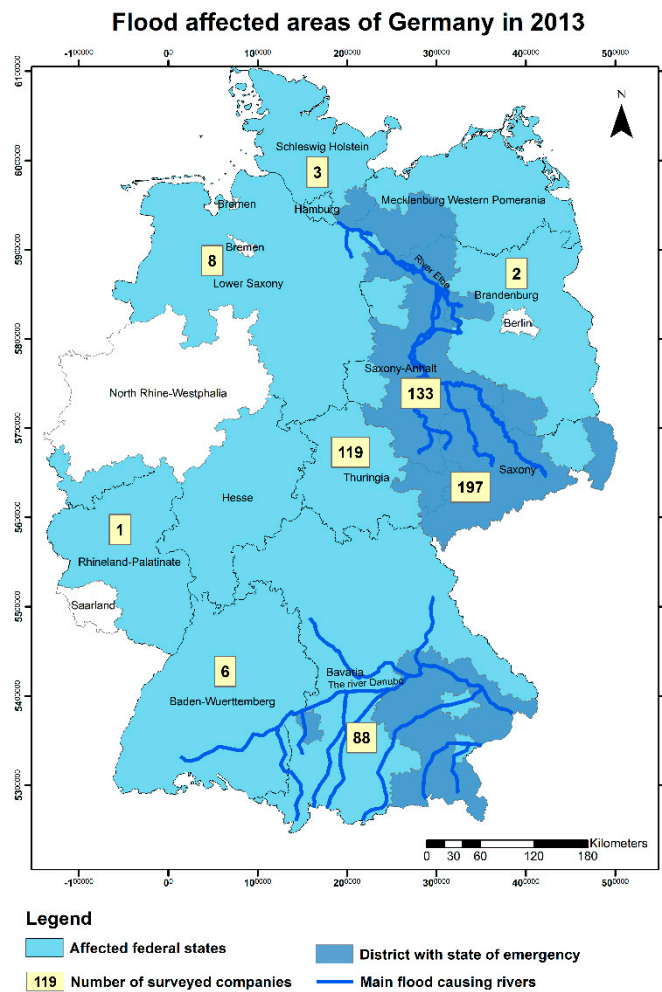
The spatially-extended and intense—but not extraordinarily intense—precipitation from the end of May until the beginning of June, in combination with the high antecedent catchment wetness, was

the main driver of the June 2013 flood [26,27]. Severe flooding occurred, especially along the Elbe River and its tributaries, Saale and Mulde, in the federal states of Saxony, Thuringia and Saxony-Anhalt, and along the Danube River in the federal state of Bavaria. Return periods of peak discharge exceeded 100 years at many gauges along the Elbe River, from Dresden to Lenzen, as well as in the Mulde and Saale catchments. Along a reach of 350 km down the Elbe River, between Coswig and the weir at Geesthacht, as well as down the Saale river, record-breaking water levels were registered [26,29]. In the Danube catchment, return periods of discharges of more than 100 years were observed along the Danube river, downstream of Regensburg and along the rivers, Inn and Salzach. In Passau, the highest water level since 1501, due to the superposition of the flood waves from the Inn and Danube rivers, was observed [30]. Using an adapted method from [31], which determines and assesses large-scale flooding based on discharge data from 162 gauges from all over the country, the flood of June 2013 can be regarded—in hydrological terms—as the most severe flood in Germany for at least the past 60 years [26]. At several locations, embankments were unable to withstand the floodwater, resulting in dike breaches and the inundation of the hinterland, e.g., 5 breaches in the Saxon part of the river Elbe and 24 failures along the river Mulde [31]. Three of these breaches had dramatic dimensions, with large-scale inundations: Near Deggendorf at the Danube River, near Groß Rosenberg at the confluence of the Saale and Elbe rivers and near Fischbeck at the Elbe River [26].

As a result, 12 out of the 16 federal states were affected by the flood, of which 8 declared a state of emergency (see Figure 1 for a geographic overview). The most affected federal states, where together more than 90% of the economic losses occurred, were Saxony, Saxony-Anhalt, Bavaria and Thuringia [4]. The flood caused 14 fatalities, 128 people were injured, 600,000 people were affected and the total direct losses amounted up to EUR 8 billion [4]. The German insurance industry paid EUR 1.65 billion in compensation [32].

### 3.1.2. Description of the Company Survey

The data result from a survey, conducted after the June 2013 flood in Germany [4]. Computer Aided Telephone Interviews (CATI) were carried out approximately one year after the flood, between May and June 2014, by a pollster (SOKO Institute, Bielefeld). In total, 557 interviews were taken on the basis of lists of affected streets of the whole flood-affected area (see Figure 1). These lists were compiled on the basis of information from affected districts or municipalities, flood reports and press releases, as well as with the help of flood masks derived from satellite data (DLR, Center for Satellite Based Crisis information, <https://www.zki.dlr.de/>). The telephone numbers were generally retrieved from the commercial telephone directory (yellow pages), and all researched telephone numbers were contacted. About 90 questions with the following topics were asked in the survey: Flood impact parameters (e.g., contamination and water level), early warning, emergency measures, precautionary measures, company characteristics, flood damage (direct losses and business interruption), and flood experience. As not all questions were applicable in all cases and, for many questions, lists of possible answers were given (with either a single answer or multiple answers possible), the interviews took only 34 minutes on average. At the beginning of the telephone call, the person on the phone was asked who in the company has the best knowledge about the flood event and the incurred losses. Then, the interview was undertaken with this person, and, in most cases, this was a member of the management board. In total, 557 interviews were completed. For further details about the survey and the data processing, see [4,33,34].



**Figure 1.** Affected federal states, as well as districts, declared states of emergency and the number of surveyed companies after the 2013 flood (own representation, following [4]).

### 3.1.3. Description of Collected Data and Developed Indicators

The 13 variables, used as potential predictors for business interruption duration or cost (predictor variables) and the two response variables, which were all derived from the data set, are shown in Table 1. They were selected according to their potential to influence company flood damage, as indicated in previous studies [4,15,22,33].

Flood impact variables are the water level, the contamination indicator and the inundation duration. The water level and the inundation duration were given from the interviewees in cm above ground and in hours or days, respectively. The contamination indicator is the weighted sum of contaminants, such as heating oil, sewage water and chemical substances. These contaminants are weighted according to their damage potential [34]. The perceived danger of another disastrous flood event was evaluated from the interviewed people on a rank scale from 1 (=very unlikely) to 6 (=very likely).

**Table 1.** Important variables compilation related to business interruption.

	Variable Name	Abbreviation	Values (Scale and Range)
Predictor Variable			
Flood impact	Water level	wl	C: 0 cm to 960 cm above ground
	Contamination indicator	coni	O: 0 = no contamination to 6 = heavy contamination
	Inundation duration	d	C: 0 to 1440 h
Danger of recurrence	Perceived danger of flood event recurrence	pror	O: 1 = very unlikely to 6 = very likely
Precaution	Adaptation ratio	adapr	O: 0.25 = low adaptation to 1 = high adaptation
	Insurance	ins	O: 0 = no flood insurance 1 = flood insurance
Warning	Warning lead time	wt	C: 1 to 336 h
	Emergency indicator	emer	O: 0 to 4 emergency measures undertaken
	Sector	sec	N: 11 = Agriculture, forestry, fishing; 12 = Manufacturing; 13 = Commercial; 14 = Financial and 15 = Service
	Number of employees	size	C: 1 to 500 employees
Company characteristics	Spatial situation	spats	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 and 4 = less than one floor in an externally used building
	Suppliers within a 50 km radius	s50	C: 0 to 100% suppliers
	Customers within a 50 km radius	c50	C: 0 to 100% customers
Response Variable			
Business Interruption	Interruption duration	bid	C: 0 to 400 days
	Absolute damage business interruption	bic	C: 0 to 10,000,000 Euro

(C: Continuous; O: Ordinal; N: Nominal).

Variables that describe the companies' precaution are the adaptation ratio and the availability of flood insurance. These variables were derived from questions about the long-term precautionary measures the company had undertaken before the 2013 flood event (checklist with different measures, e.g., the availability of flood insurance, adapted use of the flood-prone area, with multiple possible answers). The measures, entitled the "adapted use of flood-prone area", "relocation of susceptible equipment" and "relocation of dangerous substances", are classified as adaptation measures. The adaptation ratio corresponds to the share of the implemented measures compared to all of the relevant or possible measures for the specific company. The relevance of the respective measures was a question posed to the interviewees. For more details, see [22]. The warning-related variables are the warning lead time and the emergency indicator. The warning lead time is the time in hours or days between the time when the company became aware of the upcoming flood (due to, e.g., an official warning, warning by employees, own observation) and the time when the inundation of the business premises



occurred. With the emergency indicator, the undertaken emergency measures, before and during the flood, are counted. Eight different measures (e.g., the availability of an emergency plan, conducting emergency exercises every year, installation of water barriers and installation of water pumps, as indicated in [22]) were named in the surveys, and a maximum of four were conducted. These measures can also be classified into (and should be part of) a so-called Business Continuity Strategy (BCS). The BCS is a comprehensive framework on disaster preparedness, response and recovery, which aims to ensure the continuity of business in the case of any form of internal or external impacts of catastrophic events, such as technological, man-made or natural disasters.

The variables that describe the company characteristics are the sector, the number of employees, the spatial situation, and the share of suppliers and customers within a 50 km radius. The sectors were assigned, according to Nomenclature statistique des activités économiques dans la Communauté européenne (NACE) Rev. 2 [35], into the following classes: Agricultural sector, manufacturing sector, commercial sector, financial sector and service sector. The agricultural sector could not be used for the analyses because only few observations were made. The variable spatial situation describes the company site, e.g., the business premises with more than one building and less than one floor in an externally used building. The shares of suppliers and customers within a 50 km radius were asked to get information about the local interdependence of companies within the affected region. The business interruption duration and the monetary business interruption damage were given by the interviewees.

Some basic analyses and plots are made to describe the correlation structures in the data set, i.e., scatter plots with linear correlation and correlation matrix based on Spearman's rank correlation coefficient. A correlation matrix can be used to depict significant and insignificant correlations between the variables to facilitate further analysis [21,22,36].

### 3.2. Methods

#### 3.2.1. Random Forests

As already introduced in Section 2, previous studies showed the suitability of the application of tree-based models in flood damage modeling [21,22,37,38]. This paper gives only a brief overview of the functionality of Random Forests, and we refer to [39] for a more in-depth introduction of the method.

A Random Forest is, when used for regression, an ensemble of many regression trees which are organized into different nodes, namely, root nodes, split nodes and leaf nodes. The purpose of the trees is to subdivide a data set into less heterogeneous subsets, with regard to a response variable, by means of predictor variables. The data set is subdivided at the split nodes until a stop criterion is fulfilled, and the stop criteria vary with different algorithms. Hence, the data set chunks end up in leaf nodes containing only data points whose variables meet the threshold values of the split nodes of the tree. A prediction of a single tree for a new data point is usually given by the mean value of all data points present in the leaf node, in which the new data point ends up. In this case, the prediction of the forest for a new data point is the mean of the predictions made by the single trees.

Random Forest algorithms apply an internal bagging to split the input data set into two samples for the construction of single trees. One sample usually consists of about two thirds of the input data set. This sample is used for the construction of the tree. The remaining third of the input data set is called Out-of-Bag observations (OOB). The OOB observations are used internally to estimate the accuracy of the resultant model.

The Classification and Regression Tree (CART) algorithm is the most widely used algorithm to construct a Random Forest. Some studies, however, recognized a bias, with respect to variable selection, toward variables with different scales and many possible splits within the CART algorithm [40–45]. Hence, the Conditional Inference Tree (CIT) algorithm was developed to overcome this bias and improve the interpretability of the trees [46].

The CIT algorithm is used in this study to model the impacts of business interruption, since the data sets used contain variables with different scales and many possibilities of splitting. The data analysis was conducted with the statistical programming language and environment, “R” (R Foundation, Vienna, Austria, version 3.3.3) [47]. The package, “party” (R Foundation, Vienna, Austria, version 1.2), was used to compute the Random Forests [45,46,48]. Each Random Forest consists of 1000 trees (ntree), and 3 variables were randomly chosen as candidate variables at each node for splitting (mtry). Each leaf node consists of at least 7 observations. In this study, the number of predictor variables to grow the Random Forest is 13. These variables are listed in Table 1. Splitting of the data set for model validation is described in Section 3.2.4.

### 3.2.2. Stage-Damage-Function

For method comparison, stage damage functions (SDF) are also used to predict business interruption costs and duration. SDFs use the water level as the only predictor variable. In Germany, stage-damage functions, in the form of square-root functions, are widely used to estimate the damage to companies [49].

$$D = a_D + b_D \times \sqrt{h} \quad (1)$$

where

$D$ —the damage (can be either business interruption costs or business interruption duration)

$a, b$ —parameters (subscript indicates the related case)

$h$ —water level above the ground surface in cm

The parameters  $a$  and  $b$  of the square-root function are fitted to the respective training data set, which is also used to train the Random Forest. For business interruption costs, the fitting resulted in the parameter values  $a_C = -121795.4$  and  $b_C = 293649$ , and, for business interruption duration, the parameter values amount to  $a_D = 7.41$  and  $b_D = 46.42$ .

### 3.2.3. Variable Importance

In this study, we also use Random Forests to assess the individual relevance of a predictor variable from a set of input predictor variables in order to estimate business interruption costs and duration. Random Forests estimate the so-called variable importance by randomly permuting the predictor variable values to simulate the absence of the respective variable [22]. Subsequently, by comparing the OOB prediction accuracy resulting from the predictions based on original and permuted values, the importance of the particular predictor variable is derived [22]. In other words, the amount that the prediction accuracy decreases as a consequence of the permutation of the predictor variable values is used as a measure for variable importance.

### 3.2.4. Model Validation

A splitting of the input data sets is applied to allow for a comparison between the model results. The sampling method used is the Jackknife, which was developed to assess the stability of estimates [50]. Of the input data set, 75% is used for the training of the Random Forest and the SDFs, while the remaining 25% serve as a basis for the validation of both models using three different error measures:

The Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |est - obs| \quad (2)$$

The Root Mean Square Error (RMSE):

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

The Mean Bias Error (MBE):

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

where

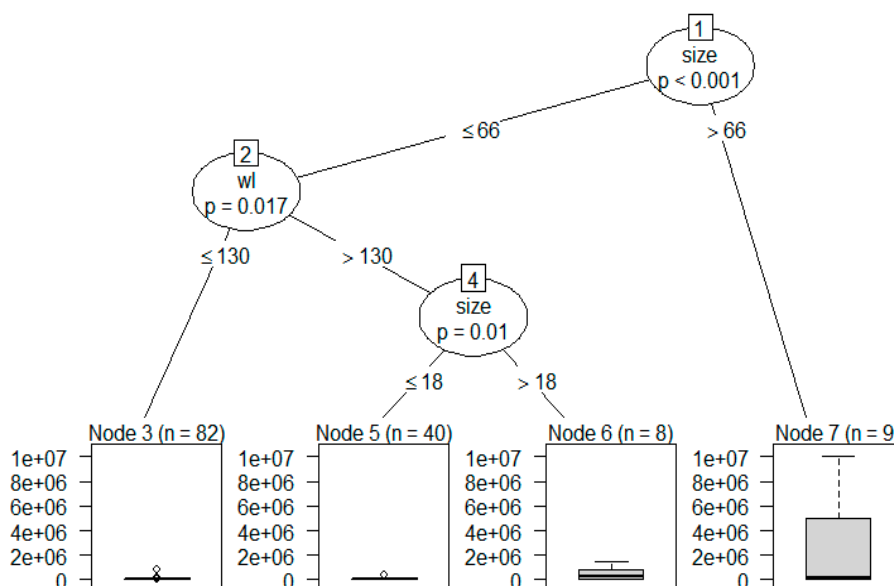
*est*—estimated value

*obs*—observed value

*n*—number of observations

The MAE measures the mean deviation from the predicted values to the observed values, the square root of the average of squared errors is considered by the RMSE, and systematic overestimation or underestimation of the models is captured by the MBE.

An exemplary single tree of a random forest, with the response variable business interruption cost, is shown in Figure 2. This tree consists of one root node, two split nodes or decision nodes and four leaf nodes, with a minimum of 8 observations.

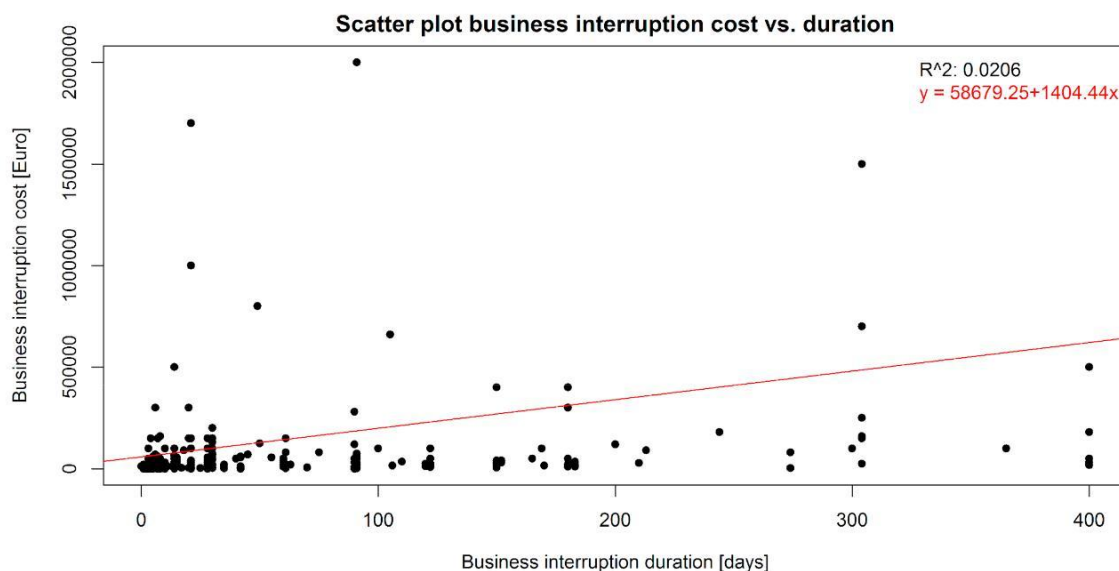


**Figure 2.** Exemplary representation of a single tree of a random forest for concept visualization, with the response variable business interruption costs in Euro.

## 4. Results and Discussion

### 4.1. Correlations with Business Interruption

The scatter plot in Figure 3 shows the linear correlation between business interruption cost *bic* and business interruption duration *bid* in the observation data. The low value of the coefficient of determination  $R^2$  (0.0206) indicates a relatively weak linear relationship between these two variables. Additionally, the influence of outliers seems to be high. This can be explained by the high heterogeneity of companies with respect to business processes and volume, size and sector. It is quite understandable that, e.g., one week of business interruption will lead to much higher costs for a manufacturing company with 100 employees than it would be in the case of a service company with 2 employees. It was revealed in the foregoing that there are significant differences between the sectors in nearly all phases of flood management [51]. For instance, the manufacturing sector was shown to have comparatively the best preparedness and precaution status but, due to the high assets and business volumes, also the highest total direct damage [51].

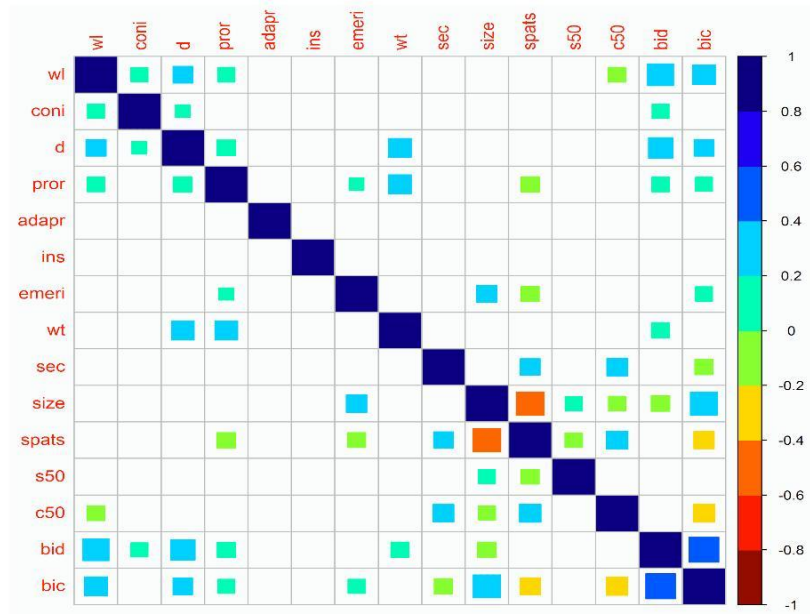


**Figure 3.** Scatter plot of business interruption cost vs. business interruption duration, with the regression value.

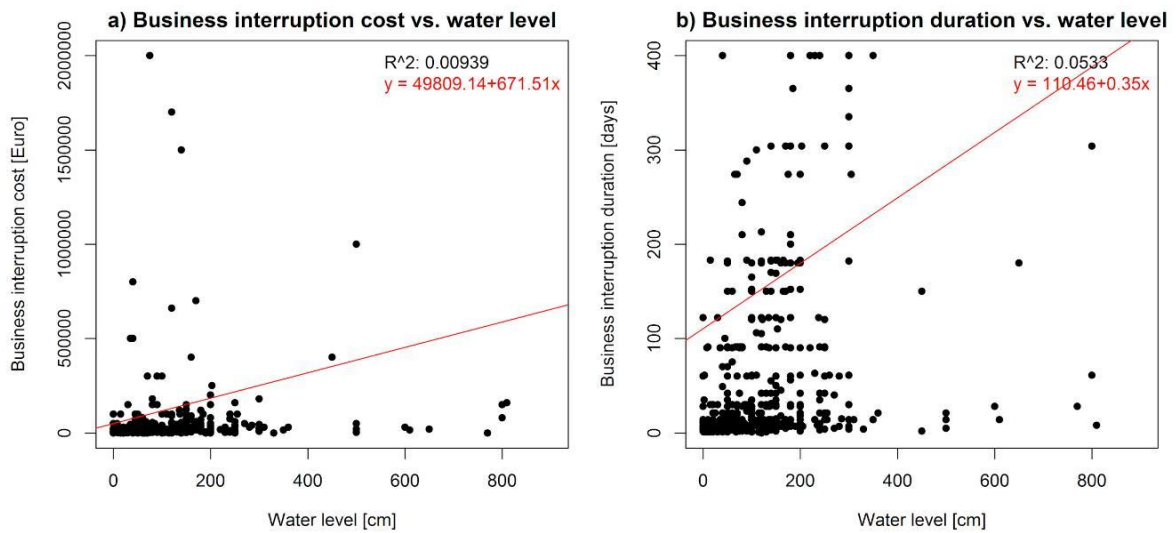
The correlation matrix of the 15 variables, comprising the 13 predictor and the two response variables, as described in Table 1, is shown in Figure 4. It is based on the Spearman's rank correlation, which is nonparametric, relatively robust to outliers and does not rely on the linearity of the statistical dependence of the variables involved. Each color of the correlation matrix represents a correlation coefficient interval, with a size of 0.2. The sizes of the colored boxes indicate the strength of the respective correlation. The white empty boxes indicate a non-significant correlation between variables. The correlation coefficients are relatively weak and range from  $-0.41$  to  $0.46$ , and a large number of coefficients are close to zero. Despite the heterogeneity of the companies, business interruption cost *bic* has the highest positive and significant correlation with the business interruption duration *bid* ( $0.46$ ). Moreover, business interruption cost is also significantly correlated with the company *size* ( $0.39$ ), water level *wl* ( $0.28$ ), inundation duration *d* ( $0.2$ ), perceived danger of flood event recurrence *pror* ( $0.14$ ) and the emergency indicator *emeri* ( $0.15$ ). In contrast, the variables customers within a 50 km radius *c50* ( $-0.24$ ), spatial situation *spats* ( $-0.22$ ) and sector *sec* ( $-0.18$ ) have a negative correlation with the business interruption cost.

The water level *wl* shows the highest positive and significant correlation with the business interruption duration *bid* ( $0.35$ ). The inundation duration *d* ( $0.31$ ), contamination *coni* ( $0.15$ ), perceived danger of flood event recurrence *pror* ( $0.16$ ) and warning time *wt* ( $0.17$ ) positively and significantly correlate with the business interruption duration *bid*. The only significant negative correlation with business interruption duration *bid* is given by the company *size* variable ( $-0.18$ ).

The scatter plots of business interruption costs *bic* in relation to the water level *wl* are shown in Figure 5a, and business interruption duration *bid* in relation to the water level *wl* is shown in Figure 5b. The water level *wl* shows relatively high correlations with business interruption costs *bic* and duration *bid* (Figure 3), and it is found to be the most important variable for flood damage in a variety of studies [15,21,22,52,53]. The value of the response variable (y-axis) increases with the increase of the water level (x-axis) in both cases. In comparison, as already indicated by the Spearman correlation, the water level seems to be more important for determining the business interruption duration than for determining business interruption costs (Figures 4 and 5). However, according to the values of the correlation coefficients  $R^2$ , linear correlations are, additionally, rather weak, compared to what is probably the high influence of outliers.



**Figure 4.** Spearman correlation of the 15 variables. Colored boxes: Correlation is significant at the 1% level.

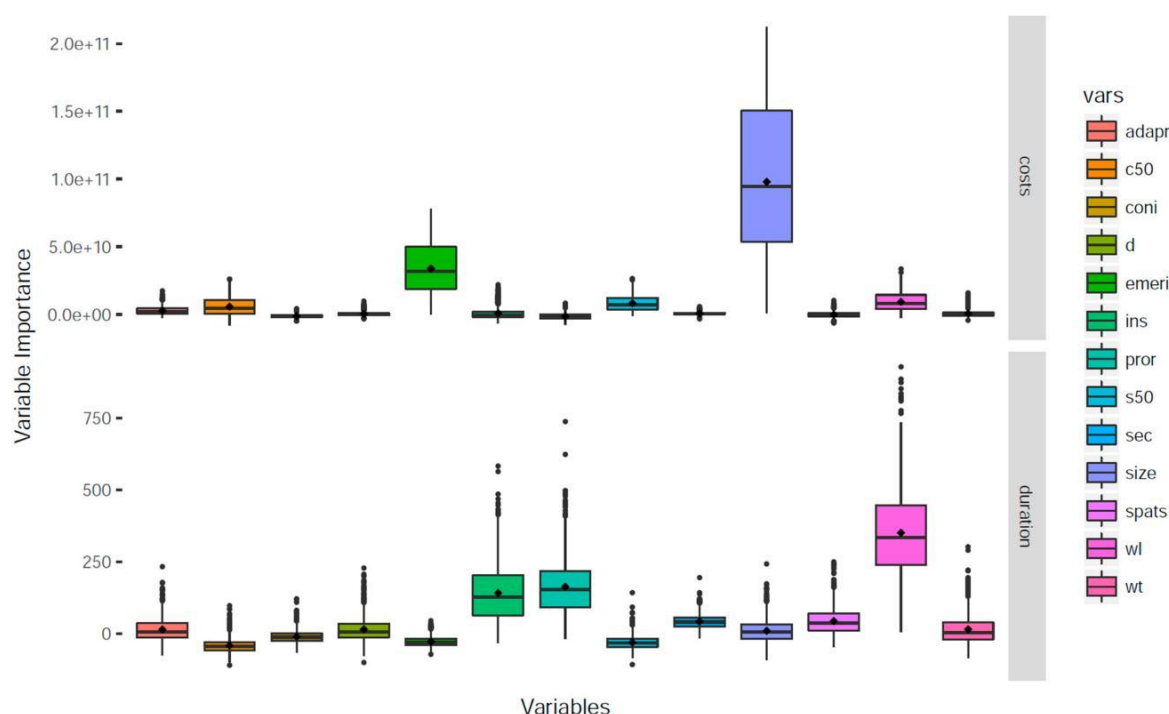


**Figure 5.** Scatter plots of (a) business interruption cost vs. water level and (b) business interruption duration vs. water level, separately with a regression value.

#### 4.2. Important Variables Determining Business Interruption

In general, correlation coefficients can only reflect pairwise and monotonic relationships of the variables, whereas Random Forests are capable of assessing non-monotonic and multivariate relationships. In other words, variable importance measures using Random Forests can capture the influences of variables on business interruption-related flood consequences, which are not detected by traditional correlation coefficients [22]. Additionally, it has been shown before, that multivariate algorithms are better suited to describing the complex damage processes during (and after) flooding [21,22,39]. Therefore, the variable importance measure based on a Random Forest is applied, and the results are used for the development of the Random Forest model.

The variable importance analysis based on the Random Forest algorithm reveals that the company size is by far the most important variable for predicting business interruption costs *bic* (Figure 6). This result confirms the findings of [54], wherein the size of the company was found to be one of the major predictors for estimating business interruption after the Nisqually earthquake in 2001 in Seattle, USA. The variable emergency indicator, which is positively correlated with the size of the company (Figure 4), seems to be also of considerable relevance for predicting business interruption costs. Other variables, namely, the water level *wl* and number of customers/suppliers within a radius of 50 km *c50/s50* are by far less important for determining business interruption costs. Since the water level *wl* seems not very important for determining business interruption costs *bic*, it appears quite critical to use the water level as a predictor for business interruption costs, e.g., using stage-damage functions.



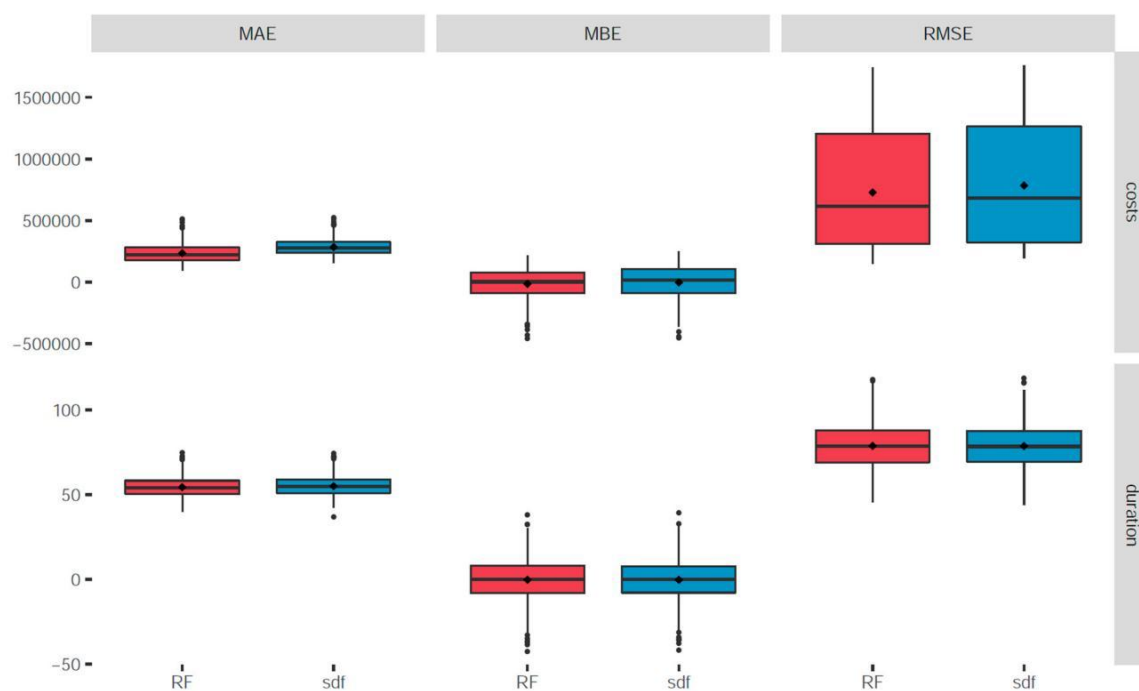
**Figure 6.** Box plots of predictor variable importance. The variable importance distributions are derived from 1000 Random Forests trained with individually sampled data.

The most important variable for predicting business interruption duration *bid* is the water level *wl* (see Figure 6). Many previous studies designate the water level likewise as the most important variable for the estimation of direct flood damage as well as business interruption [15–18,21,22,53,54]. Thus, in this case, stage-damage functions might be useful tools for estimating business interruption duration. The variables perceived, such as the danger of flood event recurrence and insurance coverage, are also relatively important for predicting business interruption duration.

A normalization of the variables *bic* and *bid* could enable a better comparison between different companies. For example, the variable *bic* could be normalized with the annual turnover of the companies. However, these kinds of data were not available for this study. Future studies, especially with a focus on modelling, should consider a normalization of the response variables.

#### 4.3. Models Estimating Business Interruption

The results of the random forest model and the SDF validation reveal very high errors and hardly any difference between the two models (Figure 7).



**Figure 7.** Error comparison of the Random Forest model and the stage damage function.

The validation of the Random Forest model for predicting business interruption costs resulted in a median MAE of EUR 222,515, a median MBE of EUR 2675 and a median RMSE of EUR 617,010 (Figure 7). The error measures of the SDF predicting business interruption costs were slightly worse, with a median MAE of EUR 275,756, a median MBE of EUR 18,050 and a median RMSE of EUR 683,185. The validation of both models predicting business interruption duration revealed almost no difference between the Random Forest model and the SDF with respect to error statistics: The median MAEs were 54.08 days and 54.67 days, the median MBEs were  $-0.01$  days and 0.04 days and the median RMSEs were 78.73 days and 78.38 days, respectively (Figure 7). In comparison with the empirical median and mean business interruption costs and durations of all companies in our sample, which are EUR 15,000 and EUR 173,356, and 15 days and 54 days, respectively, it must be concluded that both models are not able to provide reasonably accurate estimates.

## 5. Conclusions

This study improves the understanding of flood-related business interruption duration and costs in Germany, based on analyses of empirical data from 557 companies affected by the June 2013 flood. Due to the high heterogeneity, with respect to the business processes and volume, as well as the size and sector, of the companies, there is only a relatively weak linear relationship between business interruption costs and duration. The processes leading to business interruption costs and duration are complex, with various variables influencing these flood consequences.

The results of the variable importance identification show that water depth is the most important flood impact parameter for the magnitude of business interruption duration, whereas business interruption costs are mainly driven by the size of the affected company. That is to say, business interruption duration seems to be mainly driven by the hazard severity characteristics, and business interruption costs appear to be rather determined by company characteristics. We could further show that other variables, such as the perceived danger of flood event recurrence, insurance coverage and emergency indicators, can also have a significant influence on business interruption-related flood consequences.

SDF, but also the developed Random Forest-based loss model, are not able to estimate business interruption costs or duration with reasonable accuracy. Thus, the data analyses and estimation attempts can only partly explain the effects that determine business interruption duration and resultant costs, although the empirical data used are unique with regard to both the data volume and level of detail. One reason for this might be the large heterogeneity of commercial companies. A sufficient representation of the processes leading to the occurrence and magnitude of business interruption seems to require an even larger and more comprehensive database. Consequently, more empirical damage data on the commercial (sub-) sector level need to be collected to further facilitate the development of reliable business interruption loss models as well as a more in-depth understanding of the processes involved. Once such a database is available, future research should, on the one hand, strive for a reduction of uncertainties in the estimation of business interruption and, on the other hand, analyze potential changes in a company's vulnerability with respect to business interruption.

**Author Contributions:** Conceptualization, Z.S., T.S. and H.K.; Formal analysis, Z.S. and T.S.; Supervision, H.K.; Validation, T.S.; Writing—original draft, Z.S., T.S., P.K., M.M. and H.K.

**Funding:** This study 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). Additionally, the research was partly supported by the European Union's Horizon 2020 funded IMPREX project (Grant Agreement number 641811). The survey was conducted through a joint venture between the University of Potsdam, the German research center for Geosciences and the Deutsche Rückversicherung, and was funded by the German Ministry for Education and Research (BMBF) as part of the Flood 2013 project (13N13017).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. IPCC. *Climate Change 2014*; Synthesis Report, Tech. Report; IPCC: Geneva, Switzerland, 2014. [CrossRef]
2. United Nations Office for Disaster Risk Reduction (UNISDR). Making Development Sustainable: The Future of Disaster Risk Management. *Global Assessment Report on Disaster Risk Reduction 2015*. 2015. Available online: [www.unisdr.org/we/inform/publications/42809](http://www.unisdr.org/we/inform/publications/42809) (accessed on 04 September 2017).
3. Kreibich, H.; Bubeck, P.; Kunz, M.; Mahlke, H.; Parolai, S.; Khazai, B.; Daniell, J.; Lakes, T.; Schröter, K. A review of multiple natural hazards and risks in Germany. *Nat. Hazards* **2014**, *74*, 2279–2304. [CrossRef]
4. Thieken, A.H.; Bessel, T.; Kienzler, S.; Kreibich, H.; Müller, M.; Pisi, S.; Schröter, K. The flood of June 2013 in Germany: How much do we know about its impacts? *Nat. Hazards Earth Syst. Sci.* **2016**, *16*, 1519–1540. [CrossRef]
5. Kreibich, H.; van den Bergh, J.C.J.M.; Bouwer, L.M.; Bubeck, P.; Ciavola, P.; Green, C.; Hallegatte, S.; Logar, I.; Meyer, V.; Schwarze, R.; et al. Costing natural hazards. *Nat. Clim. Chang.* **2014**, *4*, 303–306. [CrossRef]
6. Rose, A.; Huyck, C.K. Improving Catastrophe Modelling for Business Interruption Insurance needs. *Risk Anal.* **2016**, *36*, 1896–1915. [CrossRef] [PubMed]
7. ICPR. *Atlas of Flood Danger and Potential Damage due to Extreme Floods of the Rhine*; International Commission for the Protection of the Rhine: Koblenz, Germany, 2001.
8. Meyer, V.; Becker, N.; Markantonis, V.; Schwarze, R.; van den Bergh, J.C.J.M.; Bouwer, L.M.; Bubeck, P.; Ciavola, P.; Genovese, E.; Green, C.; et al. Review article: Assessing the costs of natural hazards—State of the art and knowledge gaps. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 1351–1373. [CrossRef]
9. Wilhite, D.A.; Svoboda, M.D.; Hayes, M.J. Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water Resour. Manag.* **2007**, *21*, 763–774. [CrossRef]
10. Kok, M.; Huizinga, H.J.; Vrouwenfelder, A.C.W.M.; Barendregt, A. Standard Method 2004. In *Damage and Casualties Caused by Flooding*; Client Highway and Hydraulic Engineering Department: The Hague, The Netherlands, 2004; 66p.
11. Smith, K.; Ward, R. *Floods: Physical Processes and Human Impacts*; John Wiley & Sons: Chichester, UK, 1998.
12. NR&M (Department of Natural Resources and Mines, Queensland Government). *Guidance on the Assessment of Tangible Flood Damages*; NR&M: Queensland, Australia, 2002.
13. NRE (Victorian Department of Natural Resources and Environment, Victoria). *Rapid Appraisal Method (RAM) for Floodplain Management, Report prepared by Read Sturgess and Associates*; NRE: Melbourne, Australia, 2000.



14. Rose, A.; Lim, D. Business interruption losses from natural hazards: Conceptual and methodological issues in the case of the Northridge earthquake. *Environ. Hazards* **2000**, *4*, 1–14.
15. Penning-Rowsell, E.; Johnson, C.; Tunstall, S.; Tapsell, S.; Morris, J.; Chatterton, J.; Green, C. *The Benefits of Flood and Coastal Risk Management: A Handbook of Assessment Techniques*; Middlesex University Press: London, UK, 2005; p. 89.
16. Yang, L.; Kajitani, Y.; Tatano, H.; Jiang, X. A methodology for estimating business interruption loss caused by flood disasters: Insights from business surveys after Tokai Heavy Rain in Japan. *Nat. Hazards* **2016**, *84*, 411–430. [[CrossRef](#)]
17. FEMA. *Hazus®-mh mr5. Flood Model*; Technical manual; Federal Emergency Management Agency: Washington, DC, USA, 2011.
18. Parker, D.J.; Green, C.H.; Thompson, P.M. *Urban Flood Protection Benefits: A project Appraisal Guide*; Gower Technical Press: Aldershot, UK, 1987.
19. MURL (Ministerium für Umwelt, Raumordnung und Landwirtschaft des Landes Nordrhein-Westfalen). *Potentielle Hochwasserschäden am Rhein in NRW*; MURL: Düsseldorf, Germany, 2000.
20. Booyens, H.J.; Viljoen, M.F.; de Villiers, G.T. Methodology for the calculation of industrial flood damage and its application to an industry in Vereeniging. *Water* **1999**, *25*, 41–46.
21. Merz, B.; Kreibich, H.; Lall, U. Multi-variate flood damage assessment: A tree-based data-mining approach. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 53–64. [[CrossRef](#)]
22. Sieg, T.; Vogel, K.; Merz, B.; Kreibich, H. Tree-based flood damage modeling of companies: Damage processes and model performance. *Water Resour. Res.* **2017**, *53*, 1–19. [[CrossRef](#)]
23. SLF—Eidgenössisches Institut für Schnee und Lawinenforschung. *Der Lawinenwinter 1999; Ereignisanalyse*; Eidgenössisches Institut für Schnee und Lawinenforschung (SLF): Davos, Switzerland, 2000; p. 588.
24. Pfurtscheller, C.; Vetter, M. Assessing entrepreneurial and regional-economic flood impacts on a globalized production facility. *J. Flood Risk Manag.* **2015**, *8*, 329–342. [[CrossRef](#)]
25. Merz, B.; Elmer, F.; Kunz, M.; Mühr, B.; Schröter, K.; Uhlemann-Elmer, S. The extreme flood in June 2013 in Germany. *Houille Blanche* **2014**, 5–10. [[CrossRef](#)]
26. Schröter, K.; Kunz, M.; Elmer, F.; Mühr, B.; Merz, B. What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 309–327. [[CrossRef](#)]
27. DWD—Deutscher Wetterdienst (Ed.) *Das Hochwasser an Elbe und Donau im Juni 2013*; Berichte des Deutschen Wetterdienstes 242; DWD: Offenbach, Germany, 2013. (In German)
28. BfG. *Das Hochwasserextrem des Jahres 2013 in Deutschland: Dokumentation und Analyse*; Mitteilung 31; BfG: Koblenz, Germany, 2014. (In German)
29. Blöschl, G.; Nester, T.; Komma, J.; Parajka, J.; Perdigão, R.A.P. The June 2013 flood in the Upper Danube Basin, and comparisons with the 2002, 1954 and 1899 floods. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 5197–5212. [[CrossRef](#)]
30. Uhlemann, S.; Thielen, A.H.; Merz, B. A consistent set of trans-basin floods in Germany between 1952–2002. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1277–1295. [[CrossRef](#)]
31. DKKV (Ed.) *Das Hochwasser im Juni 2013—Bewährungsprobe für das Hochwasserrisikomanagement in Deutschland*; DKKV Schriftenreihe 53; DKKV: Bonn, Germany, 2015. (In German)
32. GDV—Gesamtverband der Deutschen Versicherungswirtschaft e.V. *Naturgefahrenreport 2017*. Available online: <https://www.gdv.de/resource/blob/11662/d69427a10ce7f276ff35b21901f9648c/publikation---naturgefahrenreport-2017-data.pdf> (accessed on 12 June 2018). (In German)
33. Kreibich, H.; Seifert, I.; Merz, B.; Thielen, A.H. Development of FLEMOcs—A new model for the estimation of flood losses in companies. *Hydrol. Sci. J.* **2010**, *55*, 1302–1314. [[CrossRef](#)]
34. Büchele, B.; Kreibich, H.; Kron, A.; Thielen, A.H.; Ihringer, J.; Oberle, P.; Merz, B.; Nestmann, F. Flood-risk mapping: Contributions towards an enhanced assessment of extreme events and associated risks. *Nat. Hazards Earth Syst. Sci.* **2006**, *6*, 485–503. [[CrossRef](#)]
35. Eurostat. *Statistical Classification of Economic Activities in the European Community, NACE Rev. 2. 2008*. Available online: <http://ec.europa.eu/eurostat/ramon/> (accessed on 4 September 2017).
36. Chinh, D.T.; Gain, A.K.; Dung, N.V.; Haase, D.; Kreibich, H. Multivariate analyses of flood loss in Can Tho city, Mekong delta. *Water* **2016**, *8*, 6. [[CrossRef](#)]
37. Schröter, K.; Kreibich, H.; Vogel, K.; Riggelsen, C.; Scherbaum, F.; Merz, B. How useful are complex flood damage models? *Water Resour. Res.* **2014**, *50*, 3378–3395. [[CrossRef](#)]

38. Kreibich, H.; Botto, A.; Merz, B.; Schröter, K. Probabilistic, Multivariable Flood Loss Modeling on the Mesoscale with BT-FLEMO. *Risk Anal.* **2016**, *37*, 774–787. [[CrossRef](#)] [[PubMed](#)]
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
40. Kass, G.V. An exploratory technique for investigating large quantities of categorical data. *Appl. Stat.* **1980**, *29*, 119–127. [[CrossRef](#)]
41. Segal, M.R. Regression trees for censored data. *Biometrics* **1988**, *44*, 35–47. [[CrossRef](#)]
42. White, A.P.; Liu, W.Z. Technical Note: Bias in Information-Based Measures in Decision Tree Induction. *Mach. Learn.* **1994**, *15*, 321–329. [[CrossRef](#)]
43. Jensen, D.D.; Cohen, P.R. Multiple comparisons in induction algorithms. *Mach. Learn.* **2000**, *38*, 309–338. [[CrossRef](#)]
44. Shih, Y.-S. A note on split selection bias in classification trees. *Comput. Stat. Data Anal.* **2004**, *45*, 457–466. [[CrossRef](#)]
45. Strobl, C.; Boulesteix, A.-L.; Zeileis, A.; Hothorn, T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform.* **2007**, *8*, 25. [[CrossRef](#)] [[PubMed](#)]
46. Hothorn, T.; Hornik, K.; Zeileis, A. Unbiased recursive partitioning: A conditional inference framework. *J. Comput. Graph. Stat.* **2006**, *15*, 651–674. [[CrossRef](#)]
47. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2016.
48. Hothorn, T.; Bühlmann, P.; Dudoit, S.; Molinaro, A.; van der Laan, M.J. Survival Ensembles. *Biostatistics* **2006**, *7*, 355–373. [[CrossRef](#)] [[PubMed](#)]
49. Emschergenossenschaft. *Hochwasser-Aktionsplan Emscher, Kapitel 1: Methodik der Schadensermittlung*; Emschergenossenschaft: Essen, Germany, 2004.
50. Rodgers, J.R. The bootstrap, the jackknife, and the randomization test: A sampling taxonomy. *Multivar. Behav. Res.* **1999**, *34*, 441–456. [[CrossRef](#)] [[PubMed](#)]
51. Kreibich, H.; Müller, M.; Thieken, A.H.; Merz, B. Flood precaution of companies and their ability to cope with the flood in August 2002 in Saxony, Germany. *Water Resour. Res.* **2007**, *43*. [[CrossRef](#)]
52. Gerl, T.; Bochow, M.; Kreibich, H. Flood Damage Modeling on the Basis of Urban Structure Mapping Using High-Resolution Remote Sensing Data. *Water* **2014**, *6*, 2367–2393. [[CrossRef](#)]
53. Hasanzadeh Nafari, R.; Ngo, T.; Mendis, P. An assessment of the effectiveness of tree-based models for multi-variate flood damage assessment in Australia. *Water* **2016**, *8*, 282. [[CrossRef](#)]
54. Chang, S.E.; Falit-Baiamonte, A. Disaster vulnerability of businesses in the 2001 Nisqually earthquake. *Environ. Hazards* **2002**, *4*, 59–71. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).