ERASMUS UNIVERSITY ROTTERDAM
AND DE NEDERLANDSCHE BANK
ERASMUS SCHOOL OF ECONOMICS

MASTER THESIS ECONOMETRICS AND MANAGEMENT SCIENCE:
QUANTITATIVE FINANCE

Modelling physical risk due to storms and floods with an application to real estate

Name student:
Willemijn Ouwersloot

Student ID number:
481808

EI supervisor:
prof. dr. C. Zhou

Second assessor:
prof. dr. Ph.H.B.F. Franses

Company supervisors:
dr. J. J. Dijk
dr. J. A. C. van Toor

Abstract

In this paper, we aim to find the best model for physical risk due to storms and floods, also called hazards. More specifically, we examine which model should be used to estimate damage due to these hazards. Such models are characterized by a damage function, which maps hazard characteristics to the expected ratio of incurred damage. We use micro-scale damage functions to approximate macro-scale hazard damage for the sample period 1979 through 2019. Particularly, we investigate whether a logistic damage function or including hazard duration in the model improves model performance. We find that including hazard duration does not enhance model performance. Moreover, other damage functions than the logistic function seem more suited for modelling damage. The performance of the models could be improved if more granular data were available.

Finally, we perform a simulation and a stress test based on the July 2021 floods to show how physical risk models can be used in practice. For a Dutch financial institution we find that physical risk is an important source of risk. However, it is quantifiable. Estimates for risk metrics can be calculated. Therefore, risk mitigation measures can be taken.

Keywords: physical risk, damage functions, risk management

Date final version: August 10, 2021

Contents

List of abbreviations

API - antecedent precipitation index

EAL - expected annual loss

$\mathbf{ES_{0.98}}$ - expected shortfall at 98% level

FI - financial institution

MAPE - mean absolute percentage error

MSE - mean squared error

NUTS - Nomenclature of Territorial Units for Statistics

RMSE - root mean squared error

USD - U.S. Dollars

$\mathbf{VaR_{0.99}}$ - Value-at-Risk at 99% level

1 Introduction

Financial institutions (FIs) are becoming increasingly aware of the implications of climate change and the impact this can have on the risks facing the financial sector. For FIs, two sorts of climate risk are of significance. The first form of risk is called 'transition risk'. This risk stems from the transition to a low-carbon economy. Although transition risk is outside the scope of this research, Dirks (2021) conducts a research in parallel on the impact of transition risk. The second form of risk is called 'physical risk'. Physical risk is the risk that extreme weather events, or hazards, lead to large losses in, for example, assets (Pagliari, 2021). Hazards can be droughts, floods, tropical storms, forest fires, et cetera. ECB/ESRB Project Team on climate risk monitoring (2021) estimates that between 1980 and 2017, about 453 billion Euros of economic losses in the European Economic Area and the United Kingdom were suffered due to climate-related events. Furthermore, they state that if no risk mitigation measures are taken, the economic losses due to these events will have grown to nearly 50 billion Euros per year by the end of this century because of climate change. Therefore, estimating the physical risk for assets is important. This research focuses on physical risk.

Physical risk is relevant for FIs through its effects on asset values. For instance, FIs invest heavily in real estate. This is done through extending mortgage loans as well as by investing in real estate, both directly and through real estate funds. A hazard can lead to a sudden decrease in the value of the affected real estate. If an FI can estimate how much value is at risk due to hazards, they can take risk mitigation matters on their balance sheet (Merz, Kreibich, Schwarze, & Thieken, 2010).

To model physical risk, damage due to hazards is often modelled as a function of hazard, exposure and vulnerability (Koks & Haer, 2020). A damage function maps hazard characteristics to expected damages to the exposure. Hazard characteristics concern variables that can be used as a proxy for the severity of the hazard. In this research, we focus on two specific hazards: storms and floods. For storms, an example of a hazard characteristic is the maximum wind gust speed during the storm, and for floods it can be total precipitation in the period prior to the flood.

For both storms and floods, the research on appropriate damage functions is limited. Often,

very simplistic univariate functions are used. We propose new or extended damage functions and investigate their performance. Thus, the research question we aim to answer in our research is 'To what extent can we improve existing storm and flood damage functions by either including hazard duration in the function or by using a logistic damage function?'

We apply the best performing physical risk models to the real estate exposures of Dutch pension funds and insurers in Germany. We therefore construct our models on exposure, hazard, and damage data for Germany. Note that, on average, a significant amount of the invested real estate of Dutch pension funds and insurers that is not situated in the Netherlands is located in Germany (7‰).

Data on the actual damage for real estate due to storms are either not available or only known at (re)insurance companies. This complicates modelling climate related risks, as endorsed by Network for Greening the Financial System (2021). In order to model physical risk, we use several open-access data sources. Exposure data are obtained from Eberenz, Stocker, Röösli, & Bresch (2020). This data set contains, per country, the estimated total physical asset value per square kilometer. Total hazard damages on national level are given by the CRED / UCLouvain (2021) and Paprotny, Morales Napoles, & Jonkman (2018) data sets. These data sets also contain the start and end dates of the hazards. The sample period in this research is 1979 through 2019. In Germany, there are 35 recorded damaging storms and 36 recorded damaging floods during this period. Finally, hazard characteristics such as hourly maximum wind gust speed and total hourly precipitation are obtained from Climate Data Store (2021). This data set contains hourly estimates of these climate variables per 30 kilometers.

We construct micro-economic damage functions based on the aforementioned publicly available data. In existing literature, using macro-economic damage data to calibrate micro-economic functions is, to the best of our knowledge, not yet done. This approach is relevant to FIs that do not have access to micro-economic damage data, as it would allow them to model climate risk in their balance sheet.

We model the damage to physical assets as follows. For each grid cell in the entire country grid, we model the expected damage ratio using some micro-scale damage function and the hazard characteristics for that cell. The damage ratio denotes the percentage of damage to the

physical assets located on that grid cell. Physical assets are assets which are susceptible to damage due to extreme weather events. Multiplying the damage ratio with the physical asset exposure in that cell gives an estimate of damage to assets in that cell. Aggregating the damages of all cells results in an estimate for national damage due to a specific hazard. We subsequently explore which damage function best approximates the true damage on national level. We do this for both storms and floods.

As the samples for storms and floods are both small, we evaluate the accuracy of the estimated parameters by applying bootstrap. Moreover, we evaluate the models by investigating the mean absolute percentage error (MAPE) and the root mean squared error (RMSE). Furthermore, we test for significant differences in damage function accuracy in a pairwise manner by applying the Friedman test. We combine this sequential pairwise testing with the Holm step-down procedure to control the family-wise error rate (Derrac, García, Molina, & Herrera, 2011).

Finally, we apply the best performing storm and flood damage functions to show how these functions could be used to estimate and model the physical risk for FIs. We perform these applications for the German real estate portfolio of a Dutch pension fund. In the storm application we simulate storms. We then calculate the damage to the real estate portfolio for each storm. Using these values, we calculate the values for several risk metrics such as the expected annual loss (EAL), the Value-at-Risk(99%) ($\text{VaR}_{0.99}$), and the expected shortfall at 98% level ($\text{ES}_{0.98}$). For floods, we show how damage functions can be applied for stress testing. We do this by using the July 2021 floods in Germany as a stress test.

We find that for storms, almost all parameters in the models are not significantly different from zero. Moreover, the bivariate variants of the models, including a variable on storm duration, never outperform the univariate variant. This implies that including storm duration does not enhance the performance of the model. Moreover, the logistic damage function does not seem to be the best damage function for storms. Rather, the damage function based on the one proposed by Heneka & Ruck (2008) should be used.

For floods, none of the models are particularly successful in modelling flood damage either. The difference in performance between the bivariate and univariate functions is not significant. This indicates that including flood duration in the models does not improve the flood damage

predictions. Moreover, none of the models investigated in this research outperform the benchmark model. This implies that the logistic damage function is not suited well for modelling flood damage. Therefore, the benchmark damage function should be used.

We conclude that including hazard duration in the damage function does not improve the damage model. Furthermore, the logistic damage function is not the best performing damage function for storms. For floods, other damage functions may be more appropriate, as well. However, the applications show that with a reliable damage model, good insights can be obtained with regard to the physical risk that FIs face. The simulation results in very interpretable estimates of the aforementioned risk metrics. Moreover, the stress test displays how our damage models can be applied to estimate damage in worst-case scenarios. Therefore, with a correct damage function, a meaningful physical risk assessment could be done.

We start our paper by first giving an overview of relevant existing literature in Section 2. In Section 3 we describe the data sets used in this research. Then, in Section 4, we elaborate on the econometric methods applied in this research and we describe how we execute the applications. The results of this are presented in Section 5. Finally, Section 6 concludes and offers a discussion on limitations and the recommendations for further research.

2 Literature review

In this research, we aim to model the exposure to physical risk. Physical risk is the risk of direct losses due to climate events (Pagliari, 2021). Climate risks can be classified as chronic risks, such as increasing temperatures, or as risk of occurrence of extreme weather events. Both wind storms and floods are extreme weather events and are among the main factors that cause economic losses (European Environment Agency, 2020). As these hazards often have large economic impacts, we model both types of physical risk in this research.

As mentioned by Koks & Haer (2020), physical risk is often modelled as a function of hazard, exposure and vulnerability. Hazard data, described in Sections 3.2 and 3.3, are given by historical storm and flood data and concern hazard characteristics and damage. Exposure refers to the values of the assets that are subject to the hazard. The data on exposure are described in Section 3.1. The vulnerability part of the function maps hazard characteristics to the expected

damages to the exposed assets. This can be done through damage functions, as is explained in the section below.

2.1 Damage functions

Damage functions can be distinguished based on the scale on which they are applied. Prahl, Rybski, Burghoff, & Kropp (2015) state that damage functions can be applied on micro-, meso-, or macro-scale. Micro-scale damage functions typically relate the severity of a hazard to the expected fraction of damage to one single asset, for instance to one building. Macro-scale damage functions model the total damage caused by a certain hazard. Meso-scale damage functions are somewhere in between. These functions can be used to estimate the damage to a portfolio of assets, for instance. As micro-scale damage data are not publicly available, we model macro-economic losses using micro-scale damage functions. We aggregate the estimated micro-scale damages to approximate macro-scale loss.

Micro-scale damage functions should be applicable for multiple assets. However, not all assets have equal value. This makes directly modelling monetary damage to individual assets infeasible. For that reason, micro-scale damage functions often do not directly model the monetary damage. Instead, the output is a so-called loss ratio. The loss ratio represents the fraction of damage incurred relative to the reconstruction costs or total value of the asset exposed to the hazard. Whenever a micro-scale damage function is discussed in this research, the output of such a function is the loss ratio. Different types of micro-scale damage functions exist for different hazards. We provide an overview of existing literature on damage functions for storms and floods in Sections 2.1.1 and 2.1.2, respectively.

2.1.1 Storms

Prahl et al. (2015) state that windstorms account for nearly 40% of world-wide economic losses in the period 1980-2011. As storms account for such a large fraction of damages, assessing the physical risk due to these storms is relevant to anyone with an interest in assets exposed to storms. In this research, we are interested in finding the best micro-scale damage function to approximate damages due to storms. Prahl et al. (2015) compare several such storm damage

functions. They relate the severity of a storm, measured by maximum wind gust speed during the storm, to the loss ratio.

Prahl et al. (2015) investigate, among others, the exponential damage function, which relates wind speeds to expected loss ratio in an exponential function. This is a deterministic function that can be applied both on micro-scale (Prahl et al., 2015) and on macro-scale (Dorland, Tol, & Palutikof, 1999). This relationship is endorsed by for example Murnane & Elsner (2012). They find that this damage function is empirically supported for hurricanes in the United States. In this research, we apply the exponential damage function to approximate damages due to storms in Germany.

Alternatively, Prahl, Rybski, Kropp, Burghoff, & Held (2012) propose a stochastic power law loss ratio function. It consists of two parts. First, Prahl et al. (2012) model the fitted occurrence probability. This represents the probability that there is indeed a loss, given a certain wind speed. If, according to the fitted occurrence probability function, damage is indeed incurred, the expected loss ratio is given by a power law function. Here, the loss ratio is related to a baseline parameter plus the power of the scaled wind gust speed variable. In this research, only damages due to storms are known. Therefore, we do not incorporate an occurrence probability function. However, we use the power law function as the damage function.

There are also damage functions that incorporate a threshold wind speed. Such a threshold is based on the assumption that only above a certain wind speed level, damage is incurred. Klawa & Ulbrich (2003) argue that at any location, storm damages are assumed to occur at 2% of all days. Therefore, the damage threshold at a certain location is equal to the 98^{th} percentile of the maximum wind gust speed at that particular location. We discuss three proposed excess-over-threshold damage functions.

First, Klawa & Ulbrich (2003) propose a regression model, in which damages are regressed on a transformed wind speed variable. This transformed variable is the cube of normalised wind gust intensity in excess of the threshold. They choose to use the cube as it is empirically supported.

Another damage function based on threshold wind speed is the one proposed by Emanuel (2011). Like Klawa & Ulbrich (2003), Emanuel (2011) transforms the wind gust speed variable

and takes the cube of this variable as estimate for the damage ratio. Apart from a deterministic damage wind speed threshold, this transformation also depends on a parameter that indicates for which wind speed a damage ratio of 50% is achieved.

On the other hand, Heneka & Ruck (2008) present an excess-over-threshold model in which no assumption is made regarding to which power the transformed wind speed variable should be taken. Moreover, in contrast to the 98% threshold, Heneka & Ruck (2008) assume an underlying distribution for the damage threshold. Thus, the threshold above which damage is incurred, may differ for different locations in a stochastic way. However, as we do not possess micro-scale damage data, fitting a distribution for the damage threshold is infeasible. Therefore, we apply the model proposed by Heneka & Ruck (2008), but we maintain the deterministic 98% threshold.

All storm damage functions discussed above only take the maximum wind gust speed during the storm as input in the model. However, other storm characteristics may also influence the total damages, such as the duration of the storm (Donat, Pardowitz, Leckebusch, Ulbrich, & Burghoff, 2011). In practice, to the best of our knowledge, storm duration is not taken into account in storm damage functions. For that reason, we extend the storm damage functions currently given in existing literature by incorporating storm duration into our models. Moreover, we propose a storm damage function that has the form of a logistic function, as this type of function automatically transforms the input variables to a value between zero and one. We examine whether this property is useful in modelling the damage ratio.

2.1.2 Floods

ECB/ESRB Project Team on climate risk monitoring (2021) state that since 1995, about 40% of economic losses due to extreme weather events in the European Economic Area and the United Kingdom are due to floods. This makes modelling flood damage relevant for institutions with assets that are prone to this type of damage. In existing literature, the main focus in flood risk modelling lies on modelling the probability of a flood hazard. The literature is still lacking with regard to flood damage functions (Merz et al., 2010). The existing flood damage functions in literature map inundation depth to damage. Other factors, such as flow velocity and duration of the flood, are rarely taken into account.

Often, flood damage functions are not based on any underlying statistical relationship between inundation depth and the damage ratio. The damage functions proposed by Huizinga, De Moel, Szewczyk, et al. (2017) are an example of this. Merz & Thieken (2009) compare different depth-damage functions for the city of Cologne. Three of these functions are constructed based on data obtained from surveys. The other three functions are very simple, deterministic micro-scale functions that relate inundation depth to the damage ratio for a certain building.

We do not possess historical data regarding the inundation depth corresponding to a particular flood. Instead, we use the cumulative precipitation in the period prior to the flood as a proxy for the damage due to that flood. In literature, precipitation is already used in modelling flood occurrence. For example, Froidevaux, Schwanbeck, Weingartner, Chevalier, & Martius (2015) investigate the period that should be used to calculate cumulative precipitation in order to correctly predict floods in Switzerland. They model the occurrence of a flood based on a logistic regression in which the variable is based on the accumulated precipitation prior to the flood. They conclude that both the precipitation sum of the two days before the flood day (0-1 days before the flood) and the precipitation sum of days 2 and 3 before the flood can significantly predict the occurrence of floods. We investigate whether accumulated precipitation can also serve as a proxy for the severity of the flood, and not just as a proxy for flood occurrence. To the best of our knowledge, this is not yet done in literature.

Moreover, Froidevaux et al. (2015) investigate whether antecedent precipitation indices (APIs), as proposed by Baillifard, Jaboyedoff, & Sartori (2003), also serve as good predictors for flood occurrence. APIs are not simple precipitation sums. Rather, the API is a weighted sum of precipitation during a certain period. The precipitation of several days before the flood does not influence the value of the API as much as the precipitation of the day prior to the flood. In their research, Froidevaux et al. (2015) include all days since measurements began in the calculation of their API. However, we propose to use a fixed period for the calculation of the API, namely the seven days prior to the flood. In this manner, the number of observations used to calculate the API does not differ for observations in the sample space that are done at different times.

As Froidevaux et al. (2015) use a logistic regression to model flood occurrence, we investigate

whether the logistic damage function can be used to approximate flood severity. We include variables on accumulated precipitation and flood duration in this function. The accumulated precipitation is given by either the cumulative sum of the precipitation prior to the flood or by the API.

2.2 Comparing damage models

Merz et al. (2010) state that one of the main pitfalls in damage modelling is the lack of model validation. After constructing the model, statistically validating whether the model works is often foregone. In part, this is due to the lack of historical hazard and damage data. Prahl et al. (2015) do compare damage models based on historical data. For instance, they evaluate the mean percentage error and MAPE of the models. Furthermore, they check parameter robustness by applying a jackknife procedure. Finally, they compare all models by means of a pairwise statistical test.

We evaluate model performance by evaluating the MAPE and the RMSE of the models. Moreover, we evaluate the robustness of the parameters in the models with a bootstrap procedure. Furthermore, like Prahl et al. (2015), we compare our damage functions in a pairwise manner. However, Derrac et al. (2011) state that comparing all models in a pairwise manner leads to loss of control on the family-wise error rate. Hence, for comparing multiple models, they suggest to adjust the p-values that are obtained with the pair-wise tests according to the Holm step-down procedure. We adopt this multiple comparison approach in our research.

3 Data

In this research, we eventually apply the calibrated damage functions to estimate the physical risk for a real estate portfolio due to storm and flood damage. This real estate portfolio belongs to a Dutch pension fund or insurer. Figure 1 shows an overview of the average distribution of the real estate portfolios of Dutch pension funds and insurers per country. The percentages indicate the relative average amount of real estate in the portfolio located in each country. Only countries with a relative average exposure of over 5% are shown in the legend.

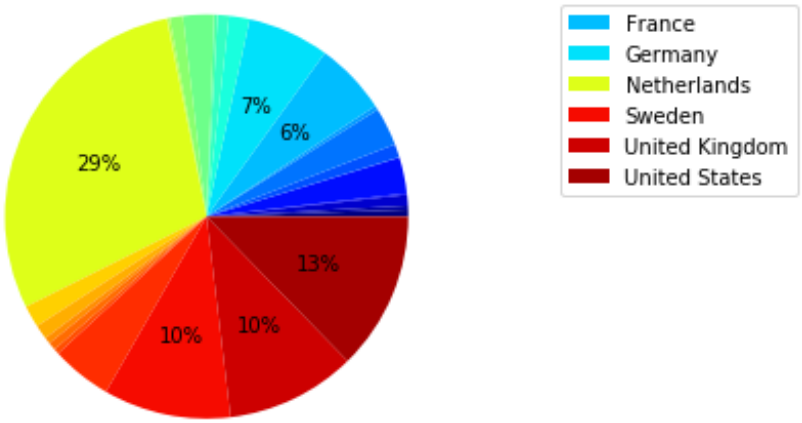


Figure 1: Average distribution of the value of real estate portfolios of Dutch FIs (i.e. pension funds and insurers) per country in the world. Only countries with a relative exposure of larger than 5% are represented in the legend.

Figure 1 shows that most real estate in the portfolios is located in the Netherlands. From the foreign located real estate of Dutch pension funds and insurers, a significant amount is situated in the United States, United Kingdom, Sweden, Germany, and France, respectively. As there are sufficient data on hazards available for Germany, we perform our research using data on hazards and exposure for Germany.

Several data sets are used in this research. Exposure data are discussed in Section 3.1. Historical damage data are discussed in Section 3.2. Finally, hazard characteristics data are elaborated upon in Section 3.3. All data preprocessing as well as parts of the modelling are done in Spyder: The Scientific Python Development Environment (2019). Part of the modelling, especially the optimisation procedure, is executed in MATLAB (2020). The toolbox that is needed for this is the Optimization Toolbox, as provided by The MathWorks (2020). Descriptions of the programming codes that are used in this research are given in Appendix D.

3.1 Asset exposures

We use a data set first composed by Eberenz et al. (2020). The data concern total physical asset exposure data per country. Physical assets are assets susceptible to physical risk. The estimated total physical asset exposure in 2014 U.S. Dollars (USD) is given for each grid cell with a spatial resolution of 30 arc-seconds. This corresponds to grid cells of a square kilometre, approximately. For Germany, there are in total 661,392 grid cells for which the value of the

physical exposure is given. There are no missing values. Some descriptive statistics are shown in Table 1.

Table 1: Descriptive statistics of physical asset exposures in Germany. All values are given in millions 2014 USD.

Statistic	**Value**
Mean	23.392
Standard deviation	114.422
Minimum	0.000
Maximum	3168.629

From the table, we observe that the asset exposure can differ extremely between different locations. This is reflected by the high standard deviation compared to the mean exposure, as well as the wide range between the minimum and maximum asset values. Figure 2 shows a heat map of the physical exposure value in Germany.

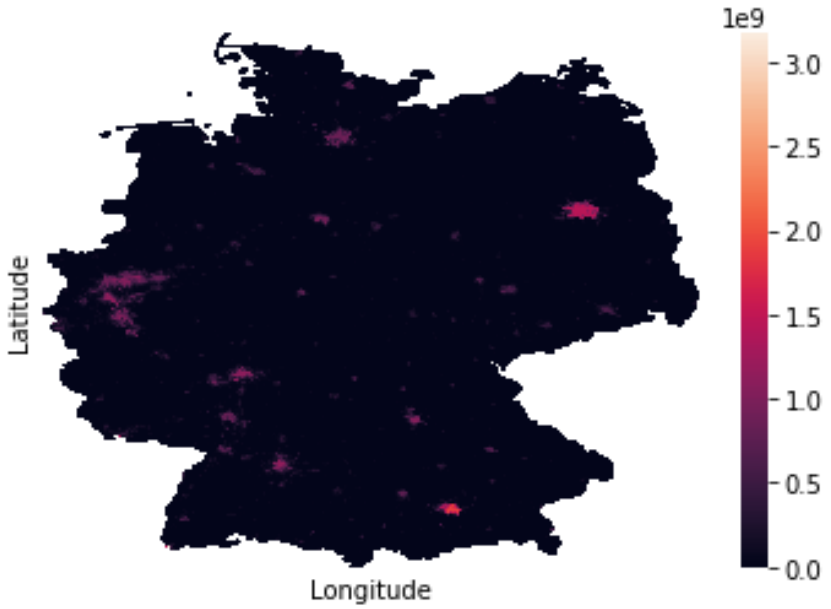


Figure 2: A heat map of the physical exposure value per square kilometre in Germany. The colouring indicates the value of the exposure in 2014 USD.

The figure endorses the observation that exposure value can differ extremely between locations. Urban ares have a lighter colouring and thus have a higher exposure value per square kilometre than rural areas.

3.2 Damages

In this research, we use historical macro-economic damage data to calibrate micro-economic damage functions. Two data sets on historical damages are used to this end.

First, macro-economic damage data due to storms are obtained from CRED / UCLouvain (2021). This data set contains data on damaging storms in the period 1979 through 2019. In this research, we include storms in Germany for which the total economic damage is known. This value is given in thousands USD in the value of the year of the storm occurrence. Moreover, for each storm it is known what the start and end date was of the storm. There are 35 storms in the data set. They are listed, along with the corresponding monetary damage, in Table 7 in Appendix A.

Macro-economic damage data due to floods are obtained from Paprotny et al. (2018). We use data on 36 damaging floods in Germany in the period 1979 through 2016. For each flood, the total damage in millions of Euros, 2011 value, is known, as well as the start and end date of the flood. Moreover, it is known which regions were affected by the flood. The regions follow the Nomenclature of Territorial Units for Statistics (NUTS) classification from 2010, as given by Eurostat (2010). The list of the 36 storms, the affected regions, and the total monetary damage can be found in Table 8 in Appendix A.

3.3 Hazard characteristics

As described in Section 2.1.1 and 2.1.2, the storm and flood damage functions model damage as a function of wind gust speed and precipitation, respectively. These data can be obtained from Climate Data Store (2021). This reanalysis data set contains hourly estimates of wind gust speed and total precipitation at a 900 arcseconds spatial resolution. This corresponds to about 30 kilometres. The wind gust speed is measured ten metres above the surface of the Earth. The maximum wind gust speed is taken as the maximum of the wind speed averaged over three second intervals (World Meteorological Organization, 2021) and is given in metres per second. The total precipitation is given in metres.

In total, there are 1184 locations in Germany for which hourly wind gust speed and total

precipitation are known. These values are known for the entire period 1979 through 2019 and there are no missing values. In order to get this data set on the same spatial resolution as the exposure data set, we perform geospatial interpolation. This is described in Section 4.1.2.

As mentioned in Sections 2.1.1 and 2.1.2, we also incorporate a hazard duration variable in our damage functions. For floods, we simply take the reported start and end date of the flood and calculate the number of hours in that period. We define that to be the flood duration. For storms, we determine the 98^{th} percentile wind gust speed for each location i in the data set, $v_{i,98}$. We take this value as a threshold above which we speak of gusts with potentially damaging speeds, as argued in Section 2.1.1. For the reported duration of the storm, we determine for each location i how many hours the maximum wind gust speed was above $v_{i,98}$. This amount of hours is the value for the storm duration variable at location i.

4 Methodology

In this section, the methodology applied in this research is deliberated upon. The data processing procedure is described in Section 4.1. The general hazard damage model is given in Section 4.2. The storm and flood damage functions are given in Sections 4.3 and 4.4, respectively. Model evaluation is discussed in Section 4.5. Finally, two applications of the research are discussed in Section 4.6.

4.1 Data processing

4.1.1 Monetary value

The different data sources report monetary values based on different reference years before 2019 and/or currencies. We convert all monetary values to the same base year, 2019. Thus, all monetary values are converted to the value they would have had in 2019. To this end, for the values reported in USD we use the annual inflation for the USA. For values reported in Euros we use the annual inflation for Germany. Both are provided by The World Bank Group (2019).

Once all monetary values are known in 2019 value, we convert the values given in USD to Euros. For this we use the average 2019 Euro - USD exchange rate as given by Macrotrends

(2021). The average exchange rate in 2019 was 1.12 Euro per USD.

4.1.2 Spatial resolution

The exposure and hazard characteristics data sets are not on the same spatial resolution. The exposure data set contains data per square kilometre approximately. The hazard characteristics data set contains observations per roughly 30 kilometres. To translate hazard characteristics at a particular location to asset damages on that location, it is necessary that the hazard characteristics data set and exposure data set contain observations on the same spatial grid. As the exposure data set has a higher resolution, we choose to perform geospatial interpolation on the hazard characteristics data in order to get this data on the same level of granularity.

We achieve this by means of inverse distance weighting. An estimate for data for unsampled locations is constructed as a weighted average of observations at other locations. The weights are inversely proportional to the distance from the unsampled location to the observed locations. Therefore, the further away observed data points are from the unsampled location, the lower the influence of those observations on the estimate for the unsampled location. The location is defined by latitude/longitude coordinates and the distance can be measured in kilometres.

Denote the number of observations, or locations, in the asset exposure data set as N. Let M denote the number of locations with an observation in the hazard characteristics data set. Then, for a hazard characteristic c, the geospatial interpolation is done as follows.

$$\hat{c}_{i\psi} = \frac{\sum_{j=1}^{M} c_{j\psi}/d_{i,j}^{\alpha}}{\sum_{j=1}^{M} 1/d_{i,j}^{\alpha}}, \qquad \forall i \in \{1,\ldots,N\}, \forall \psi \in \{1,\ldots,\Psi\}, \tag{1}$$

where $\hat{c}_{i\psi}$ is the estimate for the characteristic at location i at time ψ and $c_{j\psi}$ is the observed value of the characteristic at location j at time ψ. Ψ denotes the number of hourly observations for the measurements. $d_{i,j}$ denotes the distance in kilometres between locations i and j. For any two points, we calculate the distance between two given latitude/longitude pairs. This can be done by using the haversine formula given by

$$\mathrm{d}_{i,j} = 2r\arcsin\left(\sqrt{\sin^2\left(\frac{\mathrm{lat}_j-\mathrm{lat}_i}{2}\right) + \cos(\mathrm{lat}_i)\cos(\mathrm{lat}_j)\sin^2\left(\frac{\mathrm{lon}_j-\mathrm{lon}_i}{2}\right)}\right), \forall i \in \{1,\ldots,N\}, \forall j \in \{1,\ldots,M\},$$

where the latitude (lat) and longitude (lon) coordinates are given in radians. Moreover, r is the radius of the Earth, which is 6367 kilometers.

Finally, in Equation (1), α is a parameter which determines how quickly the influence of observations at further locations decreases. There is no theoretical foundation for the value of α. In practice, it is often set at two. In this research, we apply a jackknife method in order to find the best value for the parameter α.

With a jackknife procedure, for a particular value of α, we repeatedly remove one location with observations from the data set with all observations $\boldsymbol{c}_j$, $j \in \{1, \ldots, M\}$. We use the remaining observations to interpolate the value for the measurement at the removed location for all times $\psi \in \{1, \ldots, \Psi\}$ using Equation (1). Due to computational reasons, we do not take Ψ to include all hourly observations from 1979 through 2019. Rather, we calibrate the value for α based on the hourly measurement observations in 2019. This results in $\Psi = 8760$. We compute the RMSE of the estimated measurements of the entire sample as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{M} \sum_{\psi=1}^{\Psi} (\hat{c}_{j\psi}(\alpha) - c_{j\psi})^2}{M\Psi}}, \tag{2}$$

where $\hat{c}_{j\psi}(\alpha)$ and $c_{j\psi}$ denote the estimated and observed hazard characteristic at location j at time ψ, respectively.

We do this procedure for both wind gust speed and total precipitation for $\alpha \in \{0, 1, 2, 3, 4, 5\}$. We choose these values as 2 is often chosen in practice, and including the values 1 and 3 introduces some spread in possible values of the decay parameter. Moreover, note that a decay parameter of 0 results in equal weights across all observations, regardless of the distance between the observed and unobserved hazard characteristics. The values 4 and 5 are included in case the interpolation should be done very locally.

As the value of the decay parameter increases, the influence of observations located further away decreases. To account for the case where the interpolation should be done extremely locally, we investigate a nearest neighbours approach. With this approach, the estimate for the unobserved hazard characteristic takes the value of the closest located observation. In the case that several observations are located equally far away, we take the average of those observations.

As with the inverse distance weighting interpolation, we apply a jackknife procedure on the hourly measurement observations in 2019. We then compute the RMSE of the estimated measurements of the sample.

We select either the nearest neighbours approach or the values of α that result in the lowest RMSE for wind gust speed and total precipitation. We then interpolate the values for the hazard characteristics for each latitude/longitude pair in the asset exposure data set.

4.2 Damage modelling

For each hazard, the total economic loss for the entire country is known. The total estimated damage due to a hazard can be calculated by aggregating all location-specific estimated damages. The location-specific damage, L_{it}, can be found by multiplying the damage ratio for location $i \in \{1, \ldots, N\}$ due to hazard $t \in \{1, \ldots, T\}$ with total asset exposure at that location, e_i. As above, N denotes the number of locations for which asset exposure is known. T denotes the number of damaging hazards. The unobserved damage ratio for location i due to hazard t is denoted by y_{it}. Thus,

$$L_{it} = e_i y_{it}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}. \tag{3}$$

As the true value of the damage ratio y_{it} is unobserved, we assume it is given by a deterministic relationship, a damage function. All damage functions, for storms and floods alike, can be represented by the following formula.

$$y_{it} = \min\{\max\{g(\boldsymbol{x}_{it}; \boldsymbol{\theta}), 0\}, 1\}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}, \tag{4}$$

where y_{it} is the damage ratio and $g(\boldsymbol{x}_{it}; \boldsymbol{\theta})$ is a function that takes the independent variable vector $\boldsymbol{x}_{it}$ as input and depends on the parameter vector $\boldsymbol{\theta}$. $\boldsymbol{x}_{it}$ contains information on hazard characteristics. The variables included in the vector $\boldsymbol{x}_{it}$ depend on whether or not hazard duration is included as variable, and on whether we model storm or flood damages. The minimum and maximum functions ensure that the damage ratio takes a value between zero and one. Several functions can be applied for $g(\cdot)$. For storms, multiple damage functions are described in Section

4.3. For floods, the damage functions applied in this research are described in Section 4.4.

The total damage due to a hazard can be calculated by aggregating all location-specific damages. The entire procedure for modelling total loss from location-specific losses is given by the following formulas.

$$L_t = \sum_{i=1}^{N} L_{it} + \varepsilon_t, \qquad \forall t \in \{1, \ldots, T\}, \text{ where} \tag{5}$$

$$L_{it} = e_i y_{it}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}, \text{ where} \tag{6}$$

$$y_{it} = \min\{\max\{g(\boldsymbol{x}_{it}; \boldsymbol{\theta}), 0\}, 1\}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}, \tag{7}$$

where L_t denotes the total damage due to hazard t, and L_{it} denotes the location-specific damage due to hazard t at location i. ε_t is the residual for hazard t.

We aim to fit the function $g(\cdot)$ such that we best approximate the observed total loss. That is, we aim to find the best estimate for $\boldsymbol{\theta}$, such that the mean squared error (MSE) is minimised. The MSE is defined as the mean of the squared differences between true total losses, L_t, and the corresponding estimated losses, $\hat{L}_t$. Mathematically, this can be represented as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \frac{1}{T} \sum_{t=1}^{T} \left(L_t - \left(\sum_{i=1}^{N} e_i \min\{\max\{g(\boldsymbol{x}_{it}; \boldsymbol{\theta}), 0\}, 1\} \right) \right)^2. \tag{8}$$

For each damage function $g(\cdot)$ proposed in this research, we optimise the expression in Equation (8) in a numerical way. As there is no analytical solution, an optimal solution is not guaranteed. Therefore, the optimisation process stops when the decrease in MSE between iterations is smaller than 10^{-6}. A graphical representation of the procedure can be found in Figure 12 in Appendix B.

4.3 Storm damage functions

In this section, we discuss the damage functions for storms. First, we introduce some notation. v_{it} denotes the maximum wind gust speed at location i during hazard t. Moreover, we define d_{it} to be a duration variable. This variable denotes how many hours during hazard t the maximum wind gust speed at location i exceeds the local 98$^{\text{th}}$ percentile of observed wind gust speeds,

$v_{i,98}$. We define $\boldsymbol{x}_{it}$ to be the $2x1$ vector constructed as $[v_{it}, d_{it}]'$.

For each damage function discussed in this section, we follow the optimisation procedure described in Section 4.2.

4.3.1 Benchmark

In existing literature, there is not one storm damage function that generally serves as a benchmark. Therefore, in this research, we define

$$g(v_{it}; \boldsymbol{\theta}) = \alpha + \beta v_{it}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\} \tag{9}$$

as the benchmark model. Here, α and β are scalars.

4.3.2 Exponential and logistic damage function

The exponential damage function is based on the one compared in Prahl et al. (2015) and is given by the following equation.

$$g(v_{it}; \boldsymbol{\theta}) = e^{\alpha + \beta v_{it}}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}, \tag{10}$$

where α and β are scalars. We extend the function by also including duration as an independent variable. This leads to the following.

$$g(\boldsymbol{x}_{it}; \boldsymbol{\theta}) = e^{\alpha + \boldsymbol{x}_{it}' \boldsymbol{\beta}}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}, \tag{11}$$

where $\boldsymbol{\beta}$ now represents a $2x1$ vector of parameters.

Both Equations (10) and (11) are unbounded. The dependent variable can take any value in the range (0,∞). However, the damage ratio takes a value between zero and one. Therefore, we propose a logistic function to estimate the damage ratio. The logistic function automatically maps the independent variables to a value between zero and one, and is given by

$$g(v_{it}; \boldsymbol{\theta}) = \frac{1}{1 + e^{\alpha + \beta v_{it}}}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}. \tag{12}$$

Here, α and β are again scalars. When we take storm duration into account, the logistic function is as follows.

$$g(\boldsymbol{x}_{it};\boldsymbol{\theta}) = \frac{1}{1+e^{\alpha+\boldsymbol{x}'_{it}\boldsymbol{\beta}}}, \qquad \forall i \in \{1,\ldots,N\}, \forall t \in \{1,\ldots,T\}, \tag{13}$$

where $\boldsymbol{\beta}$ is again a $2x1$ vector of parameters.

4.3.3 Power law function

The power law damage function is based on the one presented by Prahl et al. (2012). It is given by

$$g(v_{it};\boldsymbol{\theta}) = \alpha + \left(\frac{v_{it}}{\beta}\right)^{\gamma}, \qquad \forall i \in \{1,\ldots,N\}, \forall t \in \{1,\ldots,T\}, \tag{14}$$

where α, β, and γ are scalars. If we incorporate duration into the model, we obtain the following.

$$g(\boldsymbol{x}_{it};\boldsymbol{\theta}) = \alpha + \left(\frac{v_{it}}{\beta_1}\right)^{\gamma_1} + \left(\frac{d_{it}}{\beta_2}\right)^{\gamma_2}, \qquad \forall i \in \{1,\ldots,N\}, \forall t \in \{1,\ldots,T\}. \tag{15}$$

4.3.4 Excess-over-threshold functions

There are three excess-over threshold damage functions that we apply in this research. Following Klawa & Ulbrich (2003), we assume the threshold above which damage at a location is incurred, is the 98^{th} percentile of the locally measured wind gust speed, $v_{i,98}$.

The first excess-over-threshold damage function is based on the one proposed by Klawa & Ulbrich (2003). We first transform the variable v_{it} by

$$v_{\text{new},it} = \max\left\{\left(\frac{v_{it}-v_{i,98}}{v_{i,98}}\right)^3, 0\right\}, \qquad \forall i \in \{1,\ldots,N\}, \forall t \in \{1,\ldots,T\}. \tag{16}$$

Here, the cube in the function is due to the physical reason that the cube of wind speed is proportional to the advection of kinetic energy.

Subsequently, the loss ratio is related to this transformed wind gust speed variable in a linear way,

$$g(v_{\text{new},it};\boldsymbol{\theta}) = \alpha + \beta v_{\text{new},it}, \qquad \forall i \in \{1,\ldots,N\}, \forall t \in \{1,\ldots,T\}, \tag{17}$$

where α and β are scalars. Here, α acts like a baseline parameter: this is the minimum damage incurred. This parameter should be non-negative.

We extend the function by including the duration variable in the linear regression. We define $\boldsymbol{x}_{\text{new},it}$ as the $2x1$ vector $[v_{\text{new},it}, d_{it}]'$ and obtain

$$g(\boldsymbol{x}_{\text{new},it}; \boldsymbol{\theta}) = \alpha + \boldsymbol{x}'_{\text{new},it}\boldsymbol{\beta}, \qquad \forall i \in \{1, \dots, N\}, \forall t \in \{1, \dots, T\}, \tag{18}$$

where $\boldsymbol{\beta}$ is now a $2x1$ parameter vector.

Another threshold model is based on Emanuel (2011). They state that as the damage ratio can take values between zero and one, the function should be bounded. Therefore, the function is defined as follows.

$$g(v_{it}; \boldsymbol{\theta}) = \frac{f(v_{it}; \boldsymbol{\theta})}{1 + f(v_{it}; \boldsymbol{\theta})}, \text{ where} \tag{19}$$

$$f(v_{it}; \boldsymbol{\theta}) = \left(\frac{\max\{v_{it} - v_{i,98}, 0\}}{\alpha - v_{i,98}} \right)^3. \tag{20}$$

Here, α is a scalar. It can be interpreted as the wind gust speed that, if observed, would result in a damage ratio of 50%.

We incorporate hazard duration in the damage function, while still guaranteeing the bounds for the damage ratio. This results in the following function.

$$g(\boldsymbol{x}_{it}; \boldsymbol{\theta}) = \frac{f(\boldsymbol{x}_{it}; \boldsymbol{\theta})}{1 + f(\boldsymbol{x}_{it}; \boldsymbol{\theta})}, \text{ where} \tag{21}$$

$$f(\boldsymbol{x}_{it}; \boldsymbol{\theta}) = \left(\frac{\max\{v_{it} - v_{i,98}, 0\}}{\alpha - v_{i,98}} \right)^3 + \beta d_{it}. \tag{22}$$

The third excess-over-threshold model is based on Heneka & Ruck (2008). They let go of the assumption that damage increases with the cube of wind gust speed in excess of the threshold.

Rather, the power in the function is left as a parameter. The function is as follows.

$$g(v_{it};\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } v_{it} < v_{i,98} \\ \left(\frac{v_{it}-v_{i,98}}{v_{\text{tot}}-v_{i,98}}\right)^{\alpha}, & \text{if } v_{i,98} \leq v_{it} \leq v_{\text{tot}} \\ 1, & \text{if } v_{\text{tot}} < v_{it} \end{cases}, \tag{23}$$

where v_{tot} represents the wind gust speed above which the damage ratio is equal to one. Thus, the parameters in this equation are v_{tot} and α.

4.4 Flood damage functions

In this section, we discuss the damage functions we apply for floods. As described in Section 3.2, for each flood it is known which NUTS regions were affected by the flood. Thus, it is known in which regions damages to assets could have been sustained. We incorporate this in the flood damage functions. In these functions, only exposure at a location affected by a flood is allowed to have a non-zero damage ratio. To this end, we slightly alter Equation (7). For flood t, let N_t denote the set of locations in the NUTS regions affected by the flood. We use the software QGIS Development Team (2021) to determine in which NUTS region each exposure location falls. With this information, we can determine N_t for each flood. Then the expression for the damage ratio can be given as

$$y_{it} = \mathbb{1}\{i \in N_t\}\min\{\max\{g(\boldsymbol{x}_{it};\boldsymbol{\theta}), 0\}, 1\}, \qquad \forall i \in \{1,\dots,N\}, \forall t \in \{1,\dots T\}, \tag{24}$$

where the composition of $\boldsymbol{x}$ may differ based on which variables are included in the function.

Aside from that, the notation largely corresponds to the notation we maintained for the storm damage functions in Section 4.3. However, T now denotes the number of unique damaging flood hazards. p_{it} denotes the accumulated precipitation 0-3 days prior to the flood event.

Moreover, as mentioned in Section 2.1.2, we also investigate the effect on the model of using an API based on the week prior to the flood. Let τ denote the number of days before flooding and let $r_{i,t-\tau}$ denote the total precipitation at location i on day τ before flooding t. Then the API

can be calculated as

$$\text{API}_{it} = \sum_{\tau=0}^{7} K^{\tau} r_{i,t-\tau}, \qquad \forall i \in \{1,\dots,N\}, \forall t \in \{1,\dots,T\}, \tag{25}$$

where K is a decay factor. It is a proxy for the fact that not all precipitation leads to flooding. Some precipitation gets stored in, for instance, ground water. Froidevaux et al. (2015) state that for Switzerland a K of 0.8 is sufficient. We assume this to hold true for Germany as well. Therefore, we set K equal to 0.8.

Finally, d_{it} denotes the duration variable. For floods, this is simply the number of hours between the reported start and end date of the flood.

We define $\boldsymbol{x}_{\text{precip},it}$ to be the $2x1$ vector constructed as $[p_{it}, d_{it}]'$. Moreover, $\boldsymbol{x}_{\text{API},it}$ is the $2x1$ vector $[\text{API}_{it}, d_{it}]'$

The damage functions $g(\cdot)$ are given in Sections 4.4.1 and 4.4.2. For each mentioned damage function, the damage ratio is calculated according to Equation (24), and the model is optimised according to the optimisation procedure described in Section 4.2.

4.4.1 Benchmark

The current literature is lacking with regard to flood damage functions that do not depend on inundation depth. No universal benchmark is defined. Therefore, in this research, we define

$$g(p_{it}; \boldsymbol{\theta}) = \alpha + \beta p_{it}, \qquad \forall i \in \{1,\dots,N\}, \forall t \in \{1,\dots,T\} \tag{26}$$

as the benchmark damage function we aim to beat. Here, α and β are scalars.

4.4.2 Logistic function

As Froidevaux et al. (2015) model flood occurrence with a logistic regression, we investigate whether a logistic function can approximate flood severity as well. We propose the following flood damage function.

$$g(p_{it}; \boldsymbol{\theta}) = \frac{1}{1 + e^{\alpha + \beta p_{it}}}, \qquad \forall i \in \{1,\dots,N\}, \forall t \in \{1,\dots,T\}, \tag{27}$$

where α and β are scalars. We evaluate the performance of the same damage function, but with the API as precipitation proxy rather than the cumulative precipitation. This results in the function

$$g(\text{API}_{it}; \boldsymbol{\theta}) = \frac{1}{1 + e^{\alpha + \beta \text{API}_{it}}}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}. \tag{28}$$

Moreover, we extend the functions for both precipitation proxies with the duration variable. This results in

$$g(\boldsymbol{x}_{\text{precip},it}; \boldsymbol{\theta}) = \frac{1}{1 + e^{\alpha + \boldsymbol{x}'_{\text{precip},it}\boldsymbol{\beta}}}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}, \tag{29}$$

and

$$g(\boldsymbol{x}_{\text{API},it}; \boldsymbol{\theta}) = \frac{1}{1 + e^{\alpha + \boldsymbol{x}'_{\text{API},it}\boldsymbol{\beta}}}, \qquad \forall i \in \{1, \ldots, N\}, \forall t \in \{1, \ldots, T\}, \tag{30}$$

respectively. Here, $\boldsymbol{\beta}$ is a $2x1$ parameter vector.

4.5 Model evaluation

The models are evaluated based on several measures. First, parameter robustness is investigated with bootstrap. This is further explained in Section 4.5.1. Moreover, model accuracy is evaluated by computing the mean MAPE and mean RMSE of each model. This is discussed in Section 4.5.2. Finally, the models are compared by applying a test. This is further elaborated upon in Section 4.5.3.

4.5.1 Bootstrap

Instead of dividing the total sample space with all hazard records in a train and test set once, we apply bootstrap for storms and floods separately. The number of observations on total economic loss due to a hazard is T. With bootstrap, we sample observations from this set with replacement. For B iterations, we randomly select T instances of total economic loss and corresponding hazard characteristics. We ensure the total number of different hazards in the sampled set is somewhere between $\lceil 0.4T \rceil$ and $\lfloor 0.8T \rfloor$.

In each iteration, the hazards that are included in the sampled set are used to calibrate the models according to Equation (8). The other, unsampled, hazards, are used to test the trained

model. For both storms and floods, we therefore train and test all models B times. In this research, we take $B = 25$ due to computational reasons.

For each model, the resulting 25 parameter estimates result in a mean and standard deviation of the estimate. Using the standard deviation, we can examine the significance of the parameters. To ensure replicability of the bootstrapping procedure, we set the random seed to 481808.

4.5.2 Evaluation measures

In each of the B bootstrap iterations we have a train set and a test set. For the observations in the test set, we obtain estimates of the total economic loss. Comparing these estimates with the true total loss, we can calculate loss metrics. We apply two loss metrics in this research: the MAPE and the RMSE.

For a particular model and a particular iteration b in the bootstrap method, the MAPE and RMSE can be calculated as

$$\text{MAPE}_b = \frac{1}{T_{\text{test},b}} \sum_{t=1}^{T_{\text{test},b}} \left| \frac{L_t - \hat{L}_t}{L_t} \right|, \qquad \forall b \in \{1, \ldots, B\}, \text{ and} \tag{31}$$

$$\text{RMSE}_b = \sqrt{\frac{\sum_{t=1}^{T_{\text{test},b}} \left(L_t - \hat{L}_t\right)^2}{T_{\text{test},b}}}, \qquad \forall b \in \{1, \ldots, B\}, \tag{32}$$

where $T_{\text{test},b}$ denotes the number of hazards in the test set of bootstrap iteration b. L_t and $\hat{L}_t$ denote the true and estimated value of total damages, respectively.

For each iteration $b \in \{1, \ldots, B\}$, we obtain a value for the MAPE and RMSE according to Equations (31) and (32). We average these values to obtain final model accuracy estimates for any particular model,

$$\overline{\text{MAPE}} = \frac{\sum_{b=1}^{B} \text{MAPE}_b}{B}, \qquad \text{and} \tag{33}$$

$$\overline{\text{RMSE}} = \frac{\sum_{b=1}^{B} \text{RMSE}_b}{B}. \tag{34}$$

4.5.3 Model comparison

For both the storm and flood models, we perform a multiple comparison test to determine the best model. In this test, we investigate all possible pairwise comparisons. First, we apply the Friedman test to determine whether the performance between model significantly differs. The null hypothesis of the Friedman test is equal medians between all populations. Equal medians imply similar behaviour between models. Here, a population consists of the MAPEs of all bootstrap iterations of the model. The Friedman test is a two-sided test.

For each bootstrap iteration, we calculate the MAPE of the test set. We do this for all models. Let the set of models for a particular hazard type be H, with model $h \in \{1, \ldots, H\}$. This then results in a collection of HxB MAPEs. For each bootstrap iteration $b \in \{1, \ldots, B\}$, we rank the models from 1 to H. Rank 1 is assigned to the model with the lowest MAPE, rank H is assigned to the model with the highest MAPE. This leads to ranks $r_b^h, \forall b \in \{1, \ldots, B\}, \forall h \in \{1, \ldots, H\}$, where r_b^h denotes the rank of model h for bootstrap iteration b. Thus, the rank of each model is determined for each bootstrap iteration separately.

For each model h, a final rank is calculated as the average over the ranks for the individual bootstrap iterations. Thus,

$$R_h = \frac{1}{B} \sum_{b=1}^{B} r_b^h, \qquad \forall h \in \{1, \ldots, H\}, \tag{35}$$

where B is the number of bootstrap iterations. In this manner, all models get assigned an individual rank.

Under the null hypothesis, we have

$$F = \frac{12B}{H(H+1)} \left(\sum_{h=1}^{H} R_h^2 - \frac{H(H+1)^2}{4} \right) \xrightarrow{d} \chi^2(H-1), \tag{36}$$

where F is the Friedman test statistic. The distribution of F converges to a χ^2 distribution with $H-1$ degrees of freedom, provided that both B and H are large enough. Typically, $B > 13$ or $H > 5$ are enough to apply the approximation. Otherwise, exact critical values have been computed in literature. For both storms and floods, B equals 25. Therefore, we can apply the

χ^2 approximation.

If the null hypothesis of equal performance is rejected, we investigate the separate pairwise comparisons. Here the null and alternative hypothesis are unchanged. They apply, however, to two models, instead of H. For each pair of models we compute the following statistic

$$z_{hl} = (R_h - R_l)/\sqrt{\frac{H(H+1)}{6B}}, \qquad \forall h, l \in \{1, \ldots, H\}, h \neq l, \tag{37}$$

where R_h and R_l are the average rankings by the Friedman test for models h and l, as given in Equation (35). We may approximate the distribution for z_{hl} by the standard normal. For each hypothesis in the family, we can then compute the corresponding p-value.

The separate comparisons represent different hypotheses that all belong to the same family of hypotheses. These hypotheses are, therefore, logically interrelated. A consequence of this, is that the probability of making a Type 1 error over the entire test is larger than the significance level for a single comparison, α. A Type 1 error corresponds with rejecting the null hypothesis, while it is true.

For instance, the probability of not making a Type 1 error in a single comparison is $1 - \alpha$. Assume one wants to perform $G > 1$ pairwise comparisons. Then, the probability of not making a Type 1 error over the entire set of comparisons is $(1 - \alpha)^G$. As $1 - \alpha < 1$, we get that the probability of making at least one Type 1 error is

$$1 - (1 - \alpha)^G > 1 - (1 - \alpha) = \alpha. \tag{38}$$

For this reason, we cannot compare the p-values for the corresponding z_{hl} to the significance level α, as the true probability of making a Type 1 error is higher than that. Instead, we use adjusted p-values by applying the Holm step-down procedure.

Let $m = \binom{H}{2} = \frac{H(H-1)}{2}$ be the total number of pairwise comparisons. Order the p-values from smallest to largest, such that $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$. The corresponding hypotheses are $\text{Hyp}_{(1)}$, $\text{Hyp}_{(2)}, \ldots,$ $\text{Hyp}_{(m)}$. The Holm procedure adjusts the p-values in a step-down manner. We start with the smallest, most significant, p-value, $p_{(1)}$, and compare it to $\frac{\alpha}{m}$. If $p_{(1)} < \frac{\alpha}{m}$, we reject $\text{Hyp}_{(1)}$. We then examine whether $p_{(2)} < \frac{\alpha}{m-1}$. As long as the hypotheses are rejected,

we continue down the list of p-values. In general, we compare $p_{(k)}$ to $\frac{\alpha}{m-k+1}$. When a null hypothesis is not rejected, all remaining hypotheses are not rejected, as well.

This entire procedure allows us to compare the damage models for storms and floods, separately. We can then conclude what damage function is best used for which hazard type.

4.6 Methodology of applications

In this section, we describe the methodology of two applications of our research. One application deals with how to calculate risk metrics using a damage function. This type of application is useful in determining the physical risk for an institution. When physical risk is expressed in a risk metric such as the $\text{VaR}_{0.99}$, it becomes easier for companies to take risk-mitigating measures on their balance sheet. The other application shows how to use the models in this research for stress testing.

4.6.1 Storms

In this section, we illustrate how the best performing damage functions can be used in physical risk assessment for storms. We calculate risk metrics such as the EAL, the $\text{VaR}_{0.99}$ and the $\text{ES}_{0.98}$ for FI f with a relatively large exposure in Germany. The values of these metrics indicate the physical risk the FI is under, and aid in determining the amount of capital that should be held in order to account for the risk. For instance, the Basel Committee prescribes that capital up to the value of the $\text{VaR}_{0.99}$ should be held in order to account for market risk (Sharma, 2012). In this research, we use the risk metrics to calculate the amount of capital that should be held to account for the physical risk of the portfolio.

Let Q denote the number of storms in each year. For the sample period 1979-2019, we know the realisations of Q. We assume Q follows a Poisson distribution with parameter λ equal to the mean number of yearly storms. We check this assumption by performing the chi-square goodness-of-fit test as described by Wackerly, Mendenhall III, & Scheaffer (2008).

Furthermore, let D_t denote the total duration of storm $t \in \{1, \ldots, T\}$, where T denotes the total number of historical storms. The total duration is the number of hours during the storm for which somewhere in Germany the local maximum wind gust speed exceeds the corresponding

local threshold value.

The exponential distribution can be used to model the time between events. As we model the duration of the storm, we model the time between the start and end of the storm. As this concerns a time interval, we assume that D follows an exponential distribution. The scale parameter of this distribution is equal to the observed average storm duration, $\hat{\mu}_D$. The rate parameter, on which the exponential distribution depends, is the inverse of the scale parameter. Thus, we assume $D \sim \exp\left(\frac{1}{\hat{\mu}_D}\right)$. We examine whether this assumption is reasonble by applying the Kolmogorov-Smirnov test as described by Massey Jr. (1951).

We simulate $S = 1000$ years of storms in the following way. First, we fix the random seed at the value 481808. For each simulation year $s \in \{1, \ldots, S\}$, we draw the number of storms that are simulated during year s by randomly drawing an observation from the distribution of Q. For each storm j in simulated year s, we simulate total storm duration, $D_{s,j}$, from the exponential distribution of D. We round the value $D_{s,j}$ down to the nearest integer. Then, from the historical storms that have a total duration of at least $D_{s,j}$, we randomly select one. From this selected storm, we draw $D_{s,j}$ subsequent hourly observations, starting at a random point during the storm.

It can happen that $D_{s,j}$ is larger than any observed D_t. In the simulation, we therefore impose that there should always be at least three historical storms to choose from. Then, even if $D_{s,j}$ is larger than the third largest observed D_t, we select one of the three longest historical storms and draw $D_{s,j}$ subsequent hourly observations from that storm.

For each simulated storm, we calculate the damage ratio for real estate at a certain location using the best storm damage function. We multiply this ratio with the value of the underlying real estate to obtain the estimated damage to that particular instance of real estate. We calculate the total loss of the fund's real estate portfolio by aggregating over all individual losses to real estate. For storms occurring in the same year, we aggregate to total yearly losses. We obtain the yearly losses for pension fund f for the simulated S years of storms, denoted by $L_s, \forall s \in \{1, \ldots, S\}$.

We calculate estimates of three risk metrics. First, we estimate the EAL as

$$\widehat{\text{EAL}} = \frac{\sum_{s=1}^{S} L_s}{S}. \tag{39}$$

Moreover, we compute estimates for the $\text{VaR}_{0.99}$ and $\text{ES}_{0.98}$. Theoretically, the VaR_α is defined as the smallest number l, such that the probability that the loss L exceeds this threshold is smaller than or equal to $1-\alpha$. ES_α is the expected loss given that the loss exceeds VaR_α. Mathematically, these metrics are given by

$$\text{VaR}_\alpha = \inf\{l : \Pr(L > l) \leq 1-\alpha\}, \qquad \text{and} \tag{40}$$

$$\text{ES}_\alpha = \mathbb{E}\left(L | L \geq \text{VaR}_\alpha\right). \tag{41}$$

The $\text{VaR}_{0.99}$ and $\text{ES}_{0.98}$ can be estimated through historical simulation. Order all yearly losses, $L_1, \ldots L_S$, from smallest to largest losses, $L_{(1)}, \ldots L_{(S)}$. The historical simulation estimates for $\text{VaR}_{0.99}$ and $\text{ES}_{0.98}$ are then

$$\widehat{\text{VaR}}_{0.99} = L_{(990)}, \qquad \text{and} \tag{42}$$

$$\widehat{\text{ES}}_{0.98} = \frac{\sum_{k=980}^{1,000} L_{(k)}}{21}. \tag{43}$$

4.6.2 Floods

In July 2021, northwestern Europe was hit by severe floodings. At the time of writing, the value of the total damage is still unknown. It would be of interest to insurance companies if they could already estimate how high the damage claims are going to be. We perform an application to show that the models in this research can be used to get such an estimate.

Moreover, our models can be used for stress testing. We can calculate the damage to an FIs portfolio of real estate for a certain forecast severe weather event. Stress testing is important to FIs as it grants them the possibility to hedge against the risk of those damages.

We show an application of how flood damage functions can be used in stress testing. One of the regions that incurred the most damage due to the July 2021 floods is Ahrweiler, Germany. For this region, we estimate the damage ratio as well as the losses for each exposure location in our exposure data set. We do this using the best performing flood damage function. The necessary precipitation data is obtained from Climate Data Store (2021).

5 Results

In this section, we evaluate the performance of the storm and flood damage models. The suffix '-U' after a model name indicates that the model is univariate. In that case only the maximum wind gust speed or a precipitation variable is used in the model. The suffix '-B' indicates a bivariate model. Such a model contains both the hazard characteristic variable that is included in the corresponding '-U' variant of the model, as well as a hazard duration variable. All significance tests in this research are conducted at a 5% significance level.

5.1 Storms

For the storm damage models, the investigated damage functions are the benchmark function (Bench), the exponential function (Ex), the logistic function (L), the power law function (P), the function based on the one proposed by Klawa & Ulbrich (2003) (K), the function based on the one proposed by Emanuel (2011) (Em), and the function based on the one proposed by Heneka & Ruck (2008) (H).

5.1.1 Spatial interpolation

We first determine the best manner of interpolating the maximum wind gust speed. Inverse distance weighting with several options for the decay parameter, the α in Equation (1), are investigated as well as the nearest neighbours approach. Table 2 presents the results of the jackknife procedure described in Section 4.1.2.

Table 2: RMSEs corresponding to different decay parameters in inverse distance weighting interpolation as well as the nearest neighbours interpolation of the maximum wind gust speed. The RMSEs are given in metres per second. The best method, resulting in the lowest RMSE, is indicated in bold.

Decay parameter	RMSE
0	2.650
1	1.882
2	0.920
3	0.433
4	0.285
5	**0.235**
Nearest neighbours	0.276

The table shows that the inverse distance weighting with a decay parameter of five results in the lowest RMSE in the jackknife procedure. Thus, a decay parameter of five results in the most accurate interpolated wind gust speeds.

As the value of the decay parameter grows, the influence on the interpolation by observations that are located further away, decreases. A large decay parameter therefore corresponds to a very local interpolation. The table seems to indicate that the larger the value of the decay parameter, the more accurate the interpolated wind gust speed becomes. This implies that the wind gust speed at unobserved locations depends relatively heavily on observations done nearby. However, the nearest neighbours approach is not the best approach. Therefore, we can conclude that the wind gust speed is still somewhat influenced by observations further away and does not solely depend on the observation done most nearby.

5.1.2 Models

In this section we compare the performances of different storm damage functions. As mentioned in Section 4.5, we use bootstrap to train and test each model B times. This results in B estimates of the model parameters. For each model, the mean parameter estimates along with the standard deviation are shown in Table 3.

Table 3: Mean estimates and standard deviation of the parameters in the storm damage functions. The dependent variable in these functions is the damage ratio. The standard deviation is shown in parentheses behind the parameter estimate. Parameters that are significantly different from zero on a 5% significance level are indicated with an asterisk. Note that parameters may have a different meaning across models.

	α	β or β_1	β_2	γ_1	γ_2	v_{tot}
Bench	-0.68 (1.616)	0.026 (0.055)	-	-	-	-
Ex-U	-6.664 (5.947)	-0.385 (0.944)	-	-	-	-
Ex-B	-4.884 (4.339)	-0.438 (0.905)	-2.88 (6.593)	-	-	-
L-U	13.386 (8.985)	-0.382 (0.314)	-	-	-	-
L-B	6.961 (6.148)	-0.176 (0.236)	7.7 (8.404)	-	-	-
P-U	$-2.627*10^5$ ($6.542*10^5$)	$-9.754*10^5$ ($24.214*10^5$)	-	0.236 (0.329)	-	-
P-B	$-6.445*10^6$ ($32.101*10^6$)	$-0.268*10^6$ ($0.937*10^6$)	$0.017*10^6$ ($0.660*10^6$)	-2.644 (22.085)	0.107 (0.181)	-
K-U	0.034 (0.019)	0.185* (0.079)	-	-	-	
K-B	0.056* (0.023)	0.391* (0.115)	-0.006* (0.003)	-	-	
Em-U	38.962* (1.837)	-	-	-	-	
Em-B	29.627* (3.071)	-0.035 (0.045)	-	-	-	
H-U	6.738 (3.968)	-	-	-	-	51.399 (65.439)

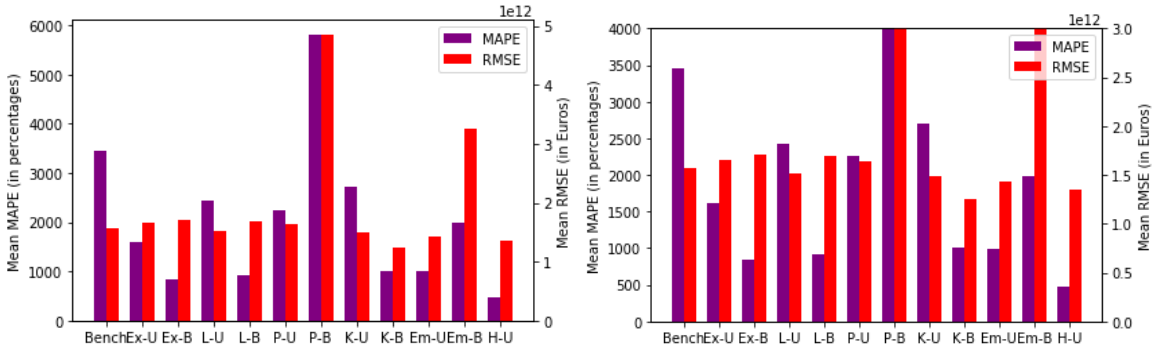
We observe that most parameters are not significantly different from zero at a 5% significance level. The only parameters that are significant are those in K-U, K-B, Em-U and Em-B.

For K-U and K-B, the parameter for the transformed wind gust variable (β or β_1) is significant.

Above the threshold, a higher wind speed thus results in a significantly higher damage ratio. In K-B, the parameter for the duration variable (β_2) is significant as well. This is the only model for which including the duration significantly improves the model. However, the estimate is negative, implying that a longer storm duration leads to less damage. Note that the constants in the K models (α) are non-negative. As mentioned in Section 4.3.4, this should be the case as they act as a baseline parameter.

The results for Em-U and Em-B indicate that the parameter that corresponds to α is significantly different from zero. Note that this parameter represents the wind speed for which a damage ratio of 50% occurs. Thus, from a physical perspective this parameter should be significantly different from zero.

Most investigated models imply that the value of the maximum wind gust speed as well as storm duration do not affect the damage ratio significantly. However, we further investigate the relative performance of the models by evaluating the B test MAPEs and RMSEs that result from the bootstrap procedure. For each model, the means of these two metrics are shown in Figure 3.



(a) The values of the mean MAPE and mean RMSE are shown for all models.

(b) As Figure 3a, except the y-axes are cut off at certain values.

Figure 3: Mean of the test MAPEs and RMSEs of all storm damage models. The purple bars denote the mean MAPE and correspond to the left y-axis. The red bars denote the mean RMSE and correspond to the right y-axis.

The figure shows that, in general, all mean MAPE and RMSE values are quite high. This could be due to the fact that macro-economic damage data are used to fit micro-scale damage functions. Another possibility is that it could be due to an insufficient amount of data to train the models.

The mean MAPE and RMSE for P-B and Em-B are extremely high compared to the other models. P-B has relatively many parameters in the model compared to the amount of available data. This could cause the model to perform poorly.

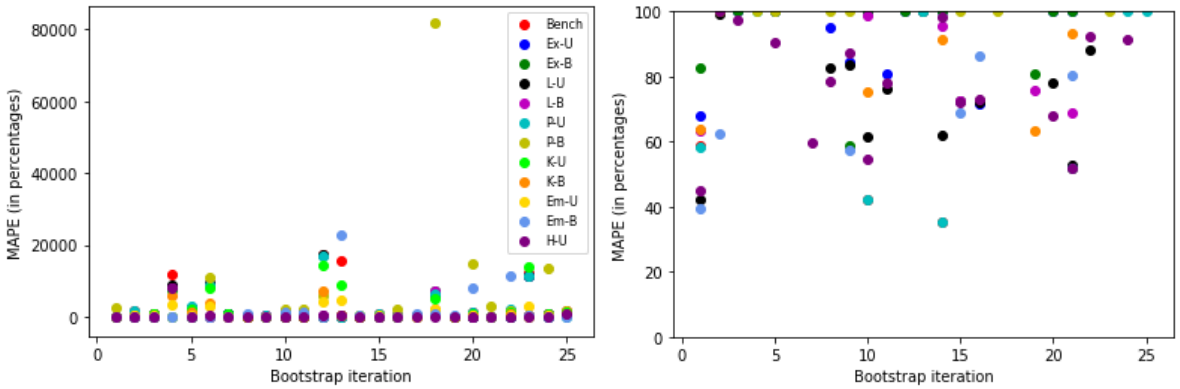
For most models, the '-B' variants result in lower mean MAPEs and RMSEs than their '-U' counterparts. This contradicts the observation following Table 3, which was that storm duration was rarely significant in the models.

Moreover, we observe that especially Ex-B, L-B, K-B, Em-U and H-U have a low mean MAPE. We investigate whether these models indeed perform significantly better, and whether the '-B' models sometimes outperform their '-U' counterparts by performing the Friedman test as described in Section 4.5.3.

First, we test whether there is any significant difference in performance between models at all. We obtain a test statistic of $F = 55.269$ with a corresponding p-value of 0.000. Thus, we reject the null hypothesis of equal population medians and can continue pairwise comparing the performance of the different models. The results of these comparisons are given in Table 9 in Appendix C.

From the table we observe that Ex-U, L-U and H-U outperform the benchmark function. Moreover, H-U outperforms the P models as well as K-U and Em-U. All other pairwise comparisons do not result in a test statistic that is significantly different from zero. The L-U model thus does outperform the benchmark model, but it is not the best damage function for modeling storm damage. Finally, the '-B' version of a damage model never performs significantly different than the '-U' version for all damage functions. Therefore, including storm duration in damage models does not lead to better damage predictions. Based on the significance test we conclude that H-U performs best, followed by Ex-U and L-U.

The Friedman test is based on the ranking of the model performances in each bootstrap iteration. We further investigate the conclusions made above by examining the model performance in each bootstrap iteration. Figure 4 presents an overview of each model's test MAPE in each bootstrap iteration.



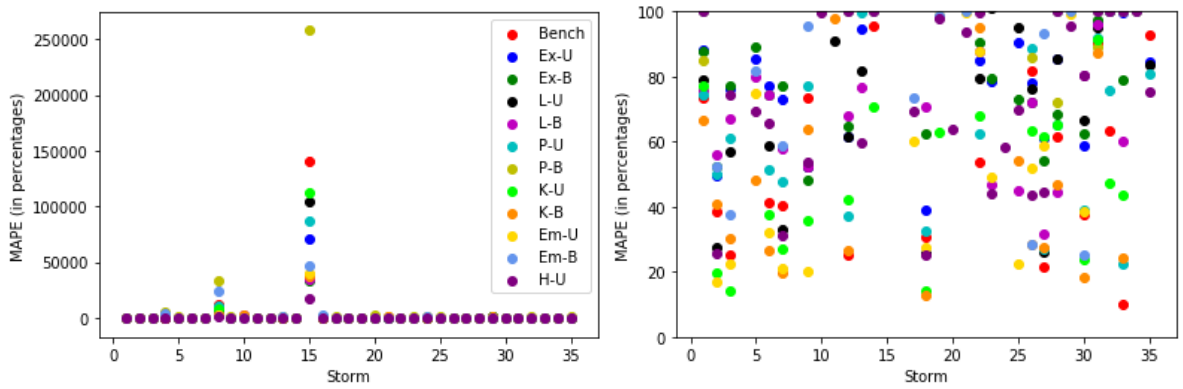
(a) All MAPEs are shown.

(b) Only MAPEs with a value of maximum 100 are shown. The same legend applies as in Figure 4a.

Figure 4: Scatter plot of each storm damage model's test MAPE in each bootstrap iteration.

We observe from the figure that model performance can differ extremely between different test sets, as we can see from Figure 4a. This figure shows, however, that in general the H-U function is stable in performance. Only for bootstrap iteration 4 does this model have a relatively high test MAPE. Figure 4b shows only the MAPEs with a value of maximum 100. Here we can see that the H-U function is frequently among the best-performing models. This supports the conclusion following the Friedman test, that the H-U model is, on average, the best performing damage model.

Note that Figure 4a shows that there are some bootstrap iterations with a test set for which the model MAPEs are, on average, very high. In each bootstrap iteration, different storms are included in the test set. Perhaps some storm damages are more difficult to approximate than others. Therefore, we also examine the test MAPE for each storm for each model. Over the B bootstrap iterations, for each storm we calculate the MAPE for a model based on the instances that the storm was included in the test set. Thus, for each model this results in T MAPEs. An overview of these MAPEs can be found in Figure 5.



(a) All MAPEs are shown.

(b) Only MAPEs with a value of maximum 100 are shown. The same legend applies as in Figure 5a.

Figure 5: Scatter plot of each model's MAPE for each storm. The MAPE for a particular storm and model is based on the instances this storm is included in the test set of the bootstrap iterations.

From the figure we observe that especially storm 15 is very hard to model correctly. For all models, the MAPEs for this storm are extremely high compared to the other storms. Storm 8 also shows high MAPEs for several models, though to a lesser extent than storm 15. Looking back at Figure 4, perhaps the iterations in which the model MAPEs are relatively high are those that include storm 8 or 15.

Figure 5b shows that even though H-U is not often the model that results in the lowest storm MAPE, it shows a relatively stable performance and is often among the best performing models. Furthermore, in instances that storm damages are hard to model correctly, such as for storms 8 and 15, H-U is the model that results in the lowest MAPE. Therefore, we conclude that it is the most robust models to outliers. This robustness of the H-U model could be due to the fact that it is a threshold model. Only wind speeds exceeding a certain level lead to damages. This could stabilize the results of the model.

From Figure 3 we concluded that comparing all models, H-U results in the lowest test mean MAPE. That, combined with the overall good performance according to the Friedman test and robustness to outliers, leads us to conclude that H-U should be used to model damages due to storms in Germany. Therefore, H-U is the model we use in the simulation as described in Section 4.6.1.

5.2 Floods

We now turn to evaluating the flood damage models. The investigated damage functions are the benchmark function (Bench) and the logistic function (L). 'cp' and 'api' indicate whether the cumulative precipitation or the API are used as variable in the function.

5.2.1 Spatial interpolation

Again, we first determine the best manner of interpolating the total precipitation: inverse distance weighting with several options for the decay parameter or the nearest neighbours approach. Table 4 shows the results.

Table 4: RMSEs corresponding to different decay parameters in the inverse distance weighting interpolation and the nearest neighbours interpolation of the total precipitation. The RMSEs are given in metres. The best method, resulting in the lowest RMSE, is indicated in bold.

Decay parameter	RMSE
0	$27.451*10^{-5}$
1	$22.645*10^{-5}$
2	$14.180*10^{-5}$
3	$8.192*10^{-5}$
4	$5.913*10^{-5}$
5	$\mathbf{5.095*10^{-5}}$
Nearest neighbours	$5.601*10^{-5}$

According to the table, the value five again results in the lowest RMSE in the jackknife procedure. This value should therefore be used to obtain the most accurate interpolated total precipitation. As described in Section 5.1.1, the decreasing RMSE as the decay parameter increases, indicates that the estimates for total precipitation can best be based on observations done nearby. However, the fact that nearest neighbours is not the best method shows that the interpolation does depend on locations that are located further away.

5.2.2 Models

In this section we evaluate the performances of the flood damage models. Again, we apply the bootstrap procedure described in Section 4.5.1 to train and test the models B times, resulting in B parameter estimates. The means and standard deviations of these estimates are shown in

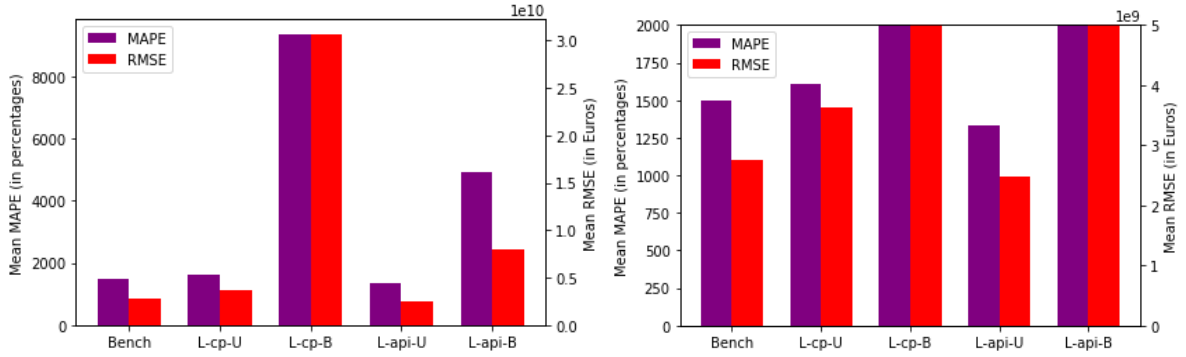
Table 5.

Table 5: Mean estimates and standard deviation of the parameters in the flood damage functions. The dependent variable in these functions is the damage ratio. The standard deviation is shown in parentheses behind the parameter estimate. Parameters that are significantly different from zero on a 5% significance level are indicated with an asterisk.

	α	β or β_1	β_2
Bench	0.001 (0.003)	-0.009 (0.097)	-
L-cp-U	7.147* (1.322)	-0.134 (38.581)	-
L-cp-B	11.946 (19.952)	25.326 (115.999)	-0.012 (0.04)
L-api-U	7.287* (1.131)	-7.265 (22.148)	-
L-api-B	8.825 (6.613)	-0.554 (38.713)	-0.005 (0.012)

From the table follows that for all models neither the total precipitation variables nor the flood duration significantly influence the damage ratio. The only significant parameters are the constants in the L-cp-U and L-api-U models. These results imply that for all flood damage models investigated in this research, given that a flood strikes a location, one can best estimate a constant damage ratio for that location.

We further investigate relative model performance by looking at the means of the B test MAPEs and RMSEs. For each flood damage model, these metrics are shown in Figure 6.



(a) The values of the mean MAPE and mean RMSE are shown for all models.

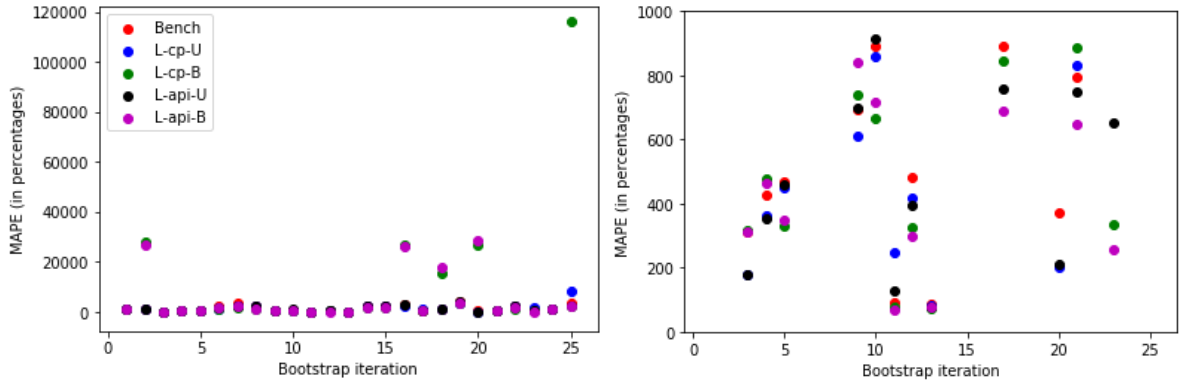
(b) As Figure 6a, except the y-axes are cut off at certain values.

Figure 6: Mean of the test MAPEs and RMSEs of all flood damage models. The purple bars denote the mean MAPE and correspond to the left y-axis. The red bars denote the mean RMSE and correspond to the right y-axis.

First of all, the figure shows that the mean MAPE and RMSE are very high for all models. Thus, in general, the models do not perform very well in explaining flood damage. Similar to the results for the storm models, this could be due to a lack of sufficient data or the fact that we use micro-scale damage functions to approximate macro-economic damage data.

Moreover, the figure shows that the '-B' models perform worse than their '-U' counterparts. Especially the mean MAPEs for the '-B' models are extremely high. Furthermore, the mean MAPE and RMSE for the functions based on the API are lower than the corresponding models that are based on the cumulative precipitation. From all models in the figure, L-api-U is the one that has both the lowest mean MAPE as well as the lowest mean RMSE.

We investigate whether L-api-U should indeed be used to estimate flood damage by performing the Friedman test as described in Section 4.5.3. Again, we first test whether there is any significant difference in model performance. This test results in a test statistic of $F = 6.720$ with a corresponding p-value of 0.151. Thus, we do not reject the null hypothesis of equal population medians. Therefore, we do not continue pairwise testing the models. However, we still investigate the model performances in each bootstrap iteration in order to obtain an overview of the relative model performance. Figure 7 presents each model's test MAPE for each bootstrap iteration.



(a) All MAPEs are shown.

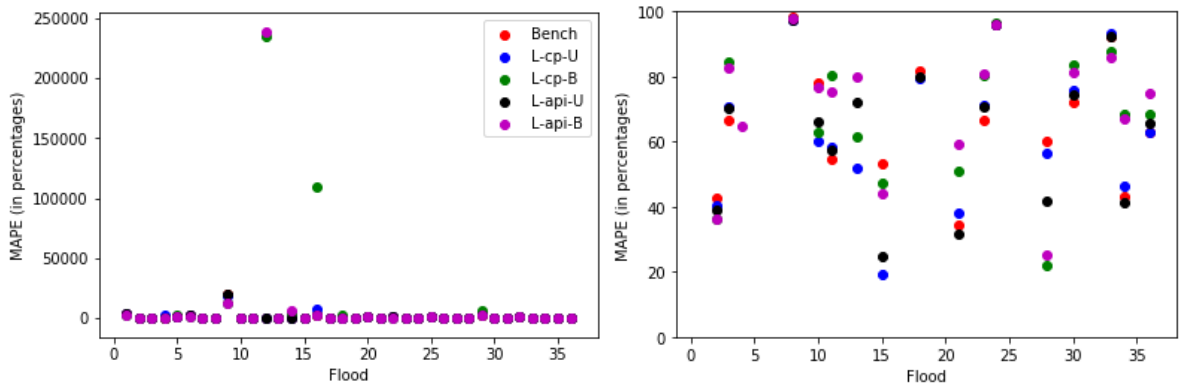
(b) Only MAPEs with a value of maximum 1000 are shown. The same legend applies as in Figure 7a.

Figure 7: Scatter plot of each flood damage model's test MAPE in each bootstrap iteration.

Again, the figure shows that model performance in terms of test MAPEs differs greatly between bootstrap iterations. Figure 7b shows only the MAPEs with a value of maximum 1000. This figure shows more clearly that the ranking of the model performance also differs greatly between bootstrap iterations. Not one model steadily outperforms the other models.

Figure 7a shows that for floods, as well, there are some bootstrap iterations for which the test MAPEs are very high. Different floods are included in the test set of each bootstrap iteration. We therefore examine whether some floods may be more difficult to approximate. As we did for

storms, for the B bootstrap iterations and for each flood, we calculate the model MAPE based on the instances that the flood was included in the test set of the bootstrap iteration. For each model, the MAPEs for each flood are shown in Figure 8.



(a) All MAPEs are shown.

(b) Only MAPEs with a value of maximum 100 are shown. The same legend applies as in Figure 8a.

Figure 8: Scatter plot of each model's MAPE for each flood. The MAPE for a particular flood and model is based on the instances this flood is included in the test set of the bootstrap iterations.

The figure shows that there there is no model which consequently results in lower flood MAPEs than other models. Furthermore, the damages of floods 9, 12, and 16 seem relatively hard to model, as multiple models result in large MAPEs for these hazards. It could be that the bootstrap iterations where one or more of these floods is included in the test set, are those iterations in Figure 7 with relatively high model MAPEs. From Figure 8a we observe that perhaps the benchmark model is most robust to outliers, as this model does not have relatively high MAPEs for floods 9, 12, and 16.

To conclude, as mentioned above, Table 5 indicates that none of the flood damage models approximate the observed damage well. Figure 6 shows that L-api-U results in the lowest mean MAPE and RMSE, although this is not significant according to the Friedman test. The benchmark function is most robust to outliers according to Figure 8. The flood application concerns a stress test. In stress tests, severe circumstances are generated in order to model the outcomes under these circumstances. Therefore, we choose to estimate the damages in the application with the most robust damage function, which is the benchmark damage function.

5.3 Applications

In this section we present the results of the storm and flood applications. The results of these applications are given in Sections 5.3.1 and 5.3.2, respectively.

5.3.1 Storms

In this section we perform an application of the research as described in Section 4.6.1. As mentioned there, we assume that the number of storms per year is Poisson distributed with parameter λ. We estimate λ as the average number of storms per year for the years in the data set. This results in $\hat{\lambda} = 0.854$. Figure 9 shows a histogram of the number of storms per year, along with the fitted Poisson distribution.

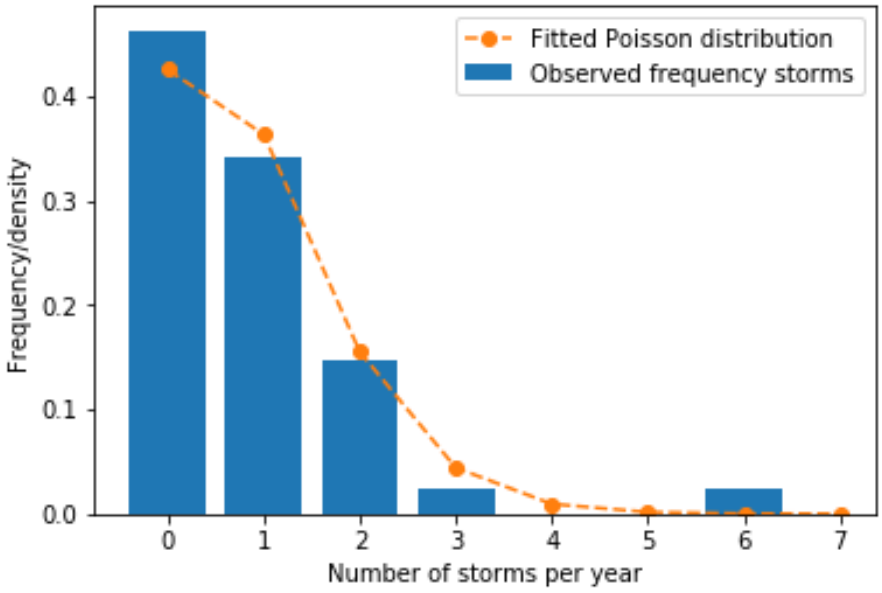


Figure 9: The blue bars are a histogram of the observed number of storms per year. The orange dotted line represents the corresponding fitted Poisson distribution. The line is dotted as the Poisson distribution is a discrete distribution.

From the figure, we observe that the Poisson distributions fits the observed data reasonably well. However, we further investigate the fit of the distribution by applying the chi-square goodness-of-fit test as mentioned in Section 4.6.1. This results in a test statistic of $\chi^2 = 1.307$ with a corresponding p-value of 0.520. We therefore do not reject the Poisson distribution to model the number of storms per year.

Furthermore, in Section 4.6.1 we assume that storm duration can be approximated by

an exponential distribution with rate parameter $\frac{1}{\mu_D}$, where μ_D is the mean storm duration. We estimate μ_D as the average storm duration of the storms in the sample, which results in $\hat{\mu}_D = 47.286$. Thus, the rate parameter is estimated at $\frac{1}{\hat{\mu}_D} = \frac{1}{47.286}$. Figure 10 presents a histogram of the storm duration, along with the fitted exponential distribution.

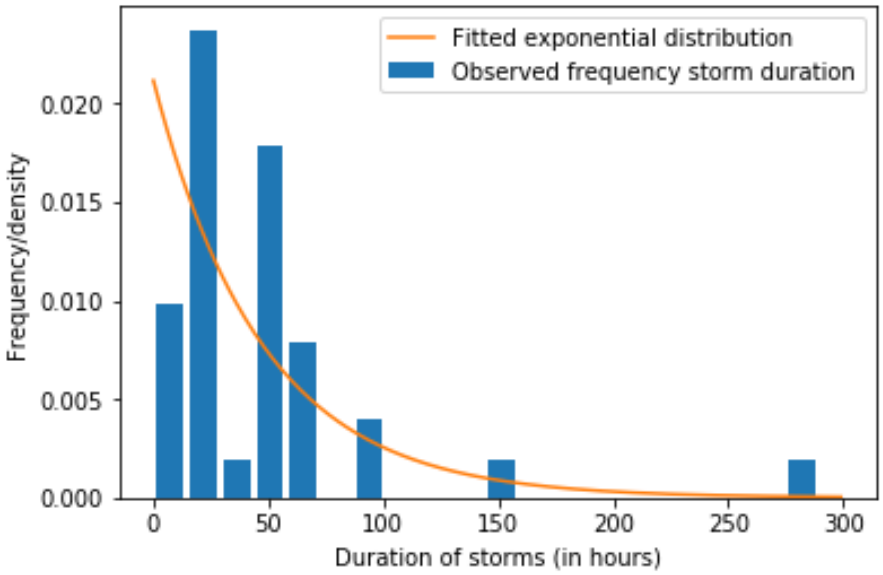


Figure 10: The blue bars are a histogram of the observed storm duration. The orange line represents the corresponding fitted exponential distribution.

The exponential distribution approximates the observed storm duration to some extent. Following Section 4.6.1, we investigate whether it is reasonable to assume this distribution for storm duration by applying the Kolmogorov-Smirnov test. This results in a test statistic of 0.150, with a corresponding p-value of 0.372. Thus, we do not reject the null hypothesis that the distribution of storm duration is the exponential distribution.

As both distributional assumptions seem reasonable, we simulate S storms using the Poisson and exponential distributions for the number of storms and their duration, respectively. Following the results in Section 5.1.2, we use the H-U damage function to calculate the simulated annual portfolio loss of the German real estate portfolio of a Dutch pension fund. These values are then used to calculate the estimates for the EAL, $\text{VaR}_{0.99}$ and $\text{ES}_{0.98}$. The values for these estimates are given in percentages of the total portfolio value and are presented in Table 6.

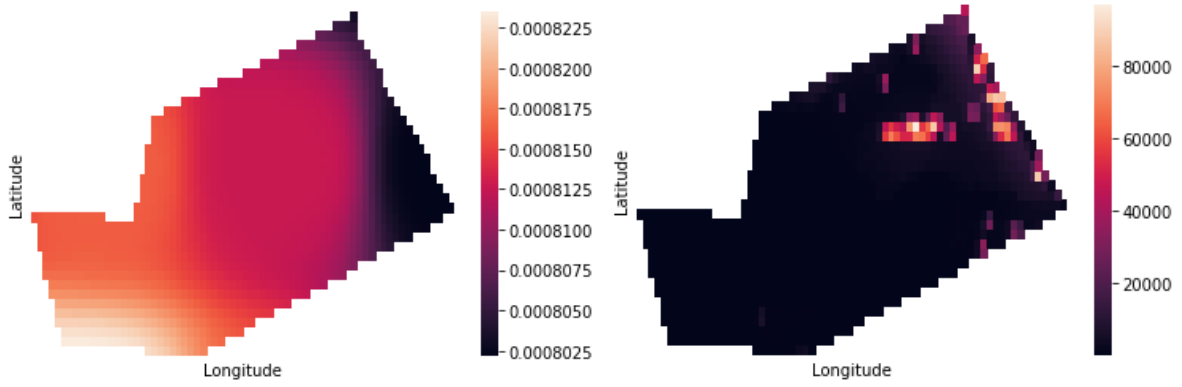
Table 6: Risk metrics and their values in percentages of the total portfolio worth. The metrics are based on a simulation of damages due to 1000 years of storms.

Risk metric	Value
$\widehat{\text{EAL}}$	1.340%
$\widehat{\text{VaR}}_{0.99}$	40.583%
$\widehat{\text{ES}}_{0.98}$	41.445%

On average, about one percent of the total portfolio worth is estimated to be lost each year due to storm damage. The $\text{VaR}_{0.99}$ and $\text{ES}_{0.98}$ are about 40% of the portfolio's worth. As the real estate portfolio can represent a substantial amount of the total asset portfolio of an FI, such values for the risk metrics indicate that FIs would do well to take the physical risk due to storms into account on their balance sheet. Therefore, the simulation shows that with a reasonable damage function, very interpretable risk metrics can be calculated and risk mitigation matters could be taken, correspondingly.

5.3.2 Floods

In this section, we perform an application of the research as described in Section 4.6.2. Figure 11 shows two heat maps of Ahrweiler, which spans 786.95 square kilometres. One heat map represents the estimated damage ratio to physical assets due to the July 2021 floods. The other heat map concerns the corresponding estimated damages.



(a) Heat map of the expected damage ratio.

(b) Heat map of the expected losses (in Euros).

Figure 11: Heat maps that show the expected damage ratio and losses in the district of Ahrweiler, Germany, due to the July 2021 floods.

Figure 11a shows that the southwest of the district is estimated to suffer relatively more damage than the northeast. However, Figure 11b shows that the expected losses are higher in the

northeast. This is due to the fact that the larger cities in the region, such as Bad Neuenahr-Ahrweiler, are located in the northeast of the area. The physical asset value per square kilometre is higher in urban areas, which explains the colouring of this heat map.

Aggregating all expected losses for the region, we obtain a total expected loss of 5.779 million Euros for Ahrweiler. As mentioned in Section 5.2.2, the flood damage functions in this research are not particularly well-suited for modelling flood damage. Nevertheless, the application shows that the models in this research can be used to get a forecast of total damages when an extreme weather event hits. With a suitable damage function, those forecasts are reliable.

The heat maps above are based on actual floodings. However, an estimate of regional losses in severe hypothetical circumstances can be obtained by applying the models investigated in this research to simulated extreme weather events. Thus, this application shows that our research is useful for stress testing.

6 Conclusion

In this research, we investigate whether we can improve existing storm and flood damage functions by including hazard duration in the function. To this end, we estimated several micro-scale damage functions to model macro-scale storm and flood damage. These damage functions map hazard characteristics to the expected fraction of damage to exposure. For storms, the hazard characteristic is the maximum wind gust speed during the storm. For floods, the hazard characteristic is either the cumulative sum of precipitation in the three days prior to the flood or the weighted sum of precipitation in the seven days before the flood, which is the so-called antecedent precipitation index or API.

We consider various storm damage functions: a benchmark function, the exponential function, logistic function, power law function, and several threshold functions, which are the ones based on Klawa & Ulbrich (2003), Emanuel (2011) and Heneka & Ruck (2008). For floods, the investigated damage functions are a benchmark function and the logistic damage function. For all functions except for the one based on Heneka & Ruck (2008) we estimate a univariate variant, including only the hazard characteristic as variable, and a bivariate variant, which also includes hazard duration.

For storms, none of the models succeed in approximating the macro-economic damage. The estimated parameters are significant in only a few instances. A significance test shows that the bivariate variant of a model never outperforms the univariate variant, indicating that including storm duration does not enhance model performance. Overall, the damage function based on the function by Heneka & Ruck (2008) is the best performing one.

Further, none of the models for floods succeed in approximating flood damage either. As with storms, in the flood damage models almost no parameters are significant. Thus, it seems that both precipitation and flood duration are not suited to model flood damage. The fact that flood duration does not seem to improve the model is endorsed by the difference in performance between the bivariate and univariate functions not being significant. In fact, not one model is able to outperform the benchmark damage function.

Overall, there does not seem to be strong evidence that including hazard duration in a damage function enhances its performance. Moreover, the logistic damage function is not the best performing damage function for storms. Furthermore, the logistic damage function fails to approximate flood damages.

In this research we also aimed at quantifying physical risk due to storms and floods through two applications: a simulation and a stress test. First, we simulated storms and estimated the values of several risk metrics. The distributional assumptions that are made in this simulation are not rejected. Therefore, if a suitable damage function could be found, the estimates for the risk metrics that result from this simulation would be quite interpretable and reliable. FIs could use this risk analysis to estimate their exposure to physical risk. This way, they can get an estimate for the expected annual loss or Value-at-Risk due to storms, for instance. Moreover, the stress test also provides interesting insights into the vulnerability of an FI to physical risk. Using these applications, risk-mitigating measures could be taken on the balance sheet of the FIs.

However, the models as investigated in this research are not yet able to approximate storm and flood damage enough to make reliable decisions on risk management. The mediocre performance of the models could be due to a lack of appropriate data. The models in this research are estimated on 35 and 36 historical records for storms and floods, respectively. This

is not a large number of observations to estimate a model on. Moreover, for storms, the damage is reported on a national level, making it impossible to distinguish regions that may be less hard hit by the storm. For floods, it is known which regions are affected by the flood, but these regions still cover a large area.

The main difficulty in quantifying and modelling physical risk thus lies in a so-called data gap. In order to properly fit the models on historical data, that data should be available. The models in this research are fit solely based on open-access data. This led to only 35 and 36 historical damage records being available for storms and floods, respectively. If, for instance, damage data on a more granular level were available, the models in this research could be estimated more precisely.

Thus, we conclude that the ability to quantify and model physical risk could improve the balance sheet of FIs. However, the models investigated in this research are not yet able to adequately model damage due to extreme weather events. Damage data on a more granular level, for example on ZIP-code level, could improve the fit of the models. We strongly suggest that such data, after being anonymised, should become more publicly available. If this would happen, we encourage that further research be done with these more granular damage data to determine whether other damage functions can better approximate storm and flood damage. Moreover, instead of constructing a model to approximate macro-economic damage, a model to approximate micro-economic damage could be estimated. This would lead to a more local model, but perhaps such a model would result in more accurate predictions of hazard damage. Another advantage of more granular damage data would be that other methods would become possible to implement, such as a fixed effect model. Such a model incorporates a location effect. Furthermore, using more granular data, it could be researched whether other hazard characteristics, for instance flow velocity for floods, improve the damage models. A final suggestion for future research would be to examine whether a proxy for climate change should be included in the damage models. These directions of research could lead to physical risk models becoming more reliable and eventually becoming a tool to predict and mitigate the risk for damage due to extreme weather events.

References

Baillifard, F., Jaboyedoff, M., & Sartori, M. (2003). Rockfall hazard mapping along a mountainous road in Switzerland using a GIS-based parameter rating approach. *Natural Hazards and Earth System Sciences*, *3*(5), 435–442.

Climate Data Store. (2021). *ERA5 hourly data on single levels from 1979 to present.* Retrieved 2021-05-03, from `https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form`

CRED / UCLouvain. (2021). *EM-DAT.* Retrieved 2021-04-07, from `https://www.emdat.be`

Derrac, J., García, S., Molina, D., & Herrera, F. (2011). A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms. *Swarm and Evolutionary Computation*, *1*(1), 3–18.

Dirks, D. (2021). *On the financial risks of the transition to a more circular economy* (Unpublished master's thesis). University of Amsterdam.

Donat, M. G., Pardowitz, T., Leckebusch, G., Ulbrich, U., & Burghoff, O. (2011). High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany. *Natural Hazards and Earth System Sciences*, *11*(10), 2821–2833.

Dorland, C., Tol, R. S., & Palutikof, J. P. (1999). Vulnerability of the Netherlands and Northwest Europe to storm damage under climate change. *Climatic Change*, *43*(3), 513–535.

Eberenz, S., Stocker, D., Röösli, T., & Bresch, D. N. (2020). Asset exposure data for global physical risk assessment. *Earth System Science Data*, *12*(2), 817–833.

ECB/ESRB Project Team on climate risk monitoring. (2021). Climate-related risk and financial stability. *European Central Bank*.

Emanuel, K. (2011). Global warming effects on US hurricane damage. *Weather, Climate, and Society*, *3*(4), 261–268.

European Environment Agency. (2020). *Disasters in Europe: more frequent and causing more damage.* Retrieved 2021-03-31, from https://www.eea.europa.eu/highlights/natural-hazards-and-technological-accidents

Eurostat. (2010). *METADATA DOWNLOAD, NUTS (Nomenclature of Territorial Units for Statistics), by regional level, version 2010 (NUTS 2010).* Retrieved 2021-05-03, from https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_CLS_DLD&StrNom=NUTS_33&StrLanguageCode=EN

Froidevaux, P., Schwanbeck, J., Weingartner, R., Chevalier, C., & Martius, O. (2015). Flood triggering in Switzerland: the role of daily to monthly preceding precipitation. *Hydrology and Earth System Sciences*, *19*(9), 3903–3924.

Heneka, P., & Ruck, B. (2008). A damage model for the assessment of storm damage to buildings. *Engineering Structures*, *30*(12), 3603–3609.

Huizinga, J., De Moel, H., Szewczyk, W., et al. (2017). Global flood depth-damage functions: Methodology and the database with guidelines.

Klawa, M., & Ulbrich, U. (2003). A model for the estimation of storm losses and the identification of severe winter storms in Germany. *Natural Hazards and Earth System Sciences*, *3*(6), 725–732.

Koks, E. E., & Haer, T. (2020). A high-resolution wind damage model for Europe. *Scientific Reports*, *10*(1), 1–11.

Macrotrends. (2021). *Euro Dollar Exchange Rate (EUR USD) - Historical Chart.* Retrieved 2021-05-04, from https://www.macrotrends.net/2548/euro-dollar-exchange-rate-historical-chart

Massey Jr., F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, *46*(253), 68–78.

MATLAB. (2020). *version 9.8 (r2020a)*. Natick, Massachusetts: The MathWorks Inc. Retrieved from https://nl.mathworks.com/downloads/

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article "Assessment of economic flood damage". *Natural Hazards and Earth System Sciences*, *10*(8), 1697–1724.

Merz, B., & Thieken, A. H. (2009). Flood risk curves and uncertainty bounds. *Natural Hazards*, *51*(3), 437–458.

Murnane, R. J., & Elsner, J. B. (2012). Maximum wind speeds and US hurricane losses. *Geophysical Research Letters*, *39*(16).

Network for Greening the Financial System. (2021). Progress report on bridging data gaps. *NGFS Publications*.

Pagliari, M. S. (2021). LSIs' exposures to climate change related risks: an approach to assess physical risks.

Paprotny, D., Morales Napoles, O., & Jonkman, B. (2018). HANZE: a pan-European database of exposure to natural hazards and damaging historical floods since 1870. *Earth System Science Data*, *10*(1), 565–581.

Prahl, B. F., Rybski, D., Burghoff, O., & Kropp, J. P. (2015). Comparison of storm damage functions and their performance. *Natural Hazards and Earth System Sciences*, *15*(4), 769–788.

Prahl, B. F., Rybski, D., Kropp, J. P., Burghoff, O., & Held, H. (2012). Applying stochastic small-scale damage functions to German winter storms. *Geophysical Research Letters*, *39*(6).

QGIS Development Team. (2021). QGIS Geographic Information System [Computer software manual]. Retrieved from `https://www.qgis.org`

Sharma, M. (2012). Evaluation of Basel III revision of quantitative standards for implementation of internal models for market risk. *IIMB Management Review*, *24*(4), 234–244.

Spyder: The Scientific Python Development Environment. (2019). *Version 4.0.1*. Python Software Foundation. Retrieved from `https://www.anaconda.com/products/individual`

The World Bank Group. (2019). *Inflation, consumer prices (annual %).* Retrieved 15-04-2021, from https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?most_recent_year_desc=true

The MathWorks, I. (2020). Matlab optimization toolbox [Computer software manual]. Natick, Massachusetts, United State. Retrieved from https://www.mathworks.com/help/optim/

Wackerly, D. D., Mendenhall III, W., & Scheaffer, R. L. (2008). Mathematical Statistics with Applications. In (pp. 714–716). Pacific Grove: Brooks/Cole Publishing Company.

World Meteorological Organization. (2021). *Variable: Wind gust.* Retrieved 12-04-2021, from https://www.wmo-sat.info/oscar/variables/view/wind_gust

A List of hazards

Table 7: List of storms in Germany in the period 1979-2019. For each storm, the start and end date are given, along with the value of the total economic loss due to the storm. This value is given in thousands of USD for the year of the hazard.

Start date	End date	Damages (in thousands USD)
7-12-1984	7-12-1984	950,000
25-1-1990	26-1-1990	1,200,000
3-2-1990	4-2-1990	600,000
7-2-1990	8-2-1990	60,000
13-2-1990	15-2-1990	180,000
25-2-1990	27-2-1990	1,200,000
28-2-1990	1-3-1990	1,200,000
5-1-1991	6-1-1991	5000
14-1-1993	14-1-1993	380,000
5-7-1993	6-7-1993	30,000
27-1-1994	29-1-1994	400,000
3-7-1994	5-7-1994	645,200
21-1-1995	2-2-1995	320,000
21-7-1995	21-7-1995	209,600
26-10-1996	30-10-1996	500
27-10-1998	2-11-1998	150,000
4-12-1999	5-12-1999	150,000
24-12-1999	27-12-1999	1,600,000
6-7-2001	7-7-2001	300,000
28-1-2002	29-1-2002	150,000
10-7-2002	10-7-2002	100,000
26-10-2002	26-10-2002	1,800,000
2-1-2003	2-1-2003	300,000

12-1-2004	13-1-2004	130,000
8-1-2005	8-1-2005	270,000
18-1-2007	18-1-2007	5,500,000
29-2-2008	2-3-2008	1,200,000
29-5-2008	2-6-2008	1,500,000
23-7-2009	24-7-2009	50,000
28-2-2010	28-2-2010	1,000,000
27-7-2013	28-7-2013	4,800,000
28-7-2014	29-7-2014	400,000
22-6-2017	23-6-2017	740,000
18-8-2017	19-8-2017	159,000
17-1-2018	18-1-2018	588,475

Table 8: List of floods in Germany in the period 1979-2016. For each flood, the start and end date are given, along with the NUTS regions affected by the flood. Moreover, the value of the total economic loss due to the flood is given in millions of Euros for the year 2011.

Start date	**End date**	**NUTS regions affected**	**Losses (mln EUR, 2011)**
4-2-1980	7-2-1980	DE111; DE246; DEA22; DEA23; DEA46	27.105
8-3-1981	14-3-1981	DE241; DE243; DE245; DE247; DE24A; DE24B; DEG0B	86.730
4-6-1981	4-6-1981	DE737; DE915; DE918	86.730
21-7-1981	23-7-1981	DE213; DE21G; DE21K; DE222; DE228	21.682
10-4-1983	14-4-1983	DEA22; DEA23; DEA2C; DEB11; DEB16; DEB17; DEB21; DEB22; DEB25	104.865

26-5-1983	31-5-1983	DE117; DE122; DE125; DEA11; DEA22; DEA23; DEB16; DEB21; DEC01	56.466
6-2-1984	9-2-1984	DE721; DE724; DEA22; DEA2C; DEA2D; DEA57; DEA5A; DEB13; DEB16; DEB21; DEB25	171.631
21-6-1984	21-6-1984	DE11B	45.083
30-4-1986	30-4-1986	DE112; DE113; DE114	1.504
26-3-1988	4-4-1988	DE222; DE223; DE228; DE232	36.504
14-2-1990	16-2-1990	DE13A	10.295
9-5-1990	29-7-1990	DE71D	3.432
1-8-1991	5-8-1991	DE222; DE224; DE228; DE22B; DE235; DE238	56.592
20-12-1993	10-1-1994	DE125; DE254; DEA1B; DEA22; DEA2C; DEA23; DEB11; DEB14; DEB15; DEB16; DEB17; DEB1A; DEB21; DEB3J; DEC01	762.093
13-4-1994	18-4-1994	DEE02; DEE08; DEE09; DEE0A; DEE0B; DEE0C; DEG03; DEG05; DEG0G; DEG0I; DEG0J; DEG0P	178.315
27-6-1994	29-6-1994	DE148; DE273; DE300; DE40E	172.371
22-1-1995	2-2-1995	DEA11; DEA22; DEA2C; DEA23; DEB11; DEB17; DEB18	320.577
10-7-1997	9-8-1997	DE403; DE409; DE40C	373.803
30-10-1998	3-11-1998	DE911; DE914; DE915; DE916; DE918; DE925; DE926; DE927; DE929; DE931; DE938; DE93B; DE949; DEA46	155.049
19-2-1999	22-2-1999	DE138; DE147; DEB34	8.587

12-5-1999	29-5-1999	DE122; DE138; DE139; DE147; DE211; DE216; DE21D; DE21L; DE21N; DE222; DE226; DE232; DE271; DE273; DE279; DE27A; DE27E	457.948
5-8-2000	6-8-2000	DE213; DE21K; DE226; DE27B; DE27C; DE27E	6.900
6-6-2002	7-6-2002	DE274; DE276; DE27C	109.735
11-8-2002	20-8-2002	DE141; DE218; DE21M; DE228; DE232; DE407; DE80A; DE934; DED21; DED2C; DED2E; DED2F; DED41; DED42; DED43; DED44; DED45; DED51; DED52; DED53; DEE01; DEE03; DEE05; DEE06; DEE0C; DEE0E; DEF06	9951.846
2-1-2003	8-1-2003	DE118; DEE02; DEG0D	10.843
21-8-2005	26-8-2005	DE212; DE216; DE21D; DE21H; DE271; DE273; DE276; DE279; DE27E	201.458
28-3-2006	17-4-2006	DED21; DED2E; DED2F; DED53	106.268
22-6-2009	27-6-2009	DE214; DE215; DE216; DE21F; DE21M; DE222; DE228	35.643
27-6-2009	4-7-2009	DE111; DE131; DE300; DE715; DEA11; DEA13; DEA14; DEA51; DEA55; DEB32; DEB3E; DEB3E	203.673
6-8-2010	8-8-2010	DED2C; DED2D; DED2F; DED41; DED42; DED45	879.313

27-9-2010	3-10-2010	DE407; DE40B; DED2C; DED2D; DED2E; DED51; DED52; DEE0B; DEE0E	136.445
7-1-2011	10-1-2011	DE116; DE11B; DE129; DE222; DE263; DEA13; DEA22; DEA51; DEB11; DEB16; DEB21; DEE06; DEE0A; DEE0C; DEE0D; DEE0E	115.000
28-5-2013	18-6-2013	DE114; DE141; DE142; DE213; DE222; DE224; DE232; DE263; DE279; DED2F; DED43; DED45; DED52; DED53; DEE02; DEE03; DEE07; DEG02; DEG0J; DEG0L; DEG0M	6451.866
28-7-2014	29-7-2014	DE112; DE135; DE137; DE142; DEA33; DEA37	426.813
27-5-2016	30-5-2016	DE116; DE118; DE119; DE11A; DE11D; DE144; DE232; DE238; DE251; DE256; DE25A; DEA28; DEB15; DE712; DE713; DE714; DE71A; DE71B; DE71C; DE725	687.928
31-5-2016	9-6-2016	DE113; DE11D; DE125; DE12A; DE21G; DE22A; DE600; DE712; DE71B; DE94F; DEA1F; DEA22; DEA33; DEB12; DEB17; DEB23; DEC03	1651.027

B Visualisation of the optimisation process

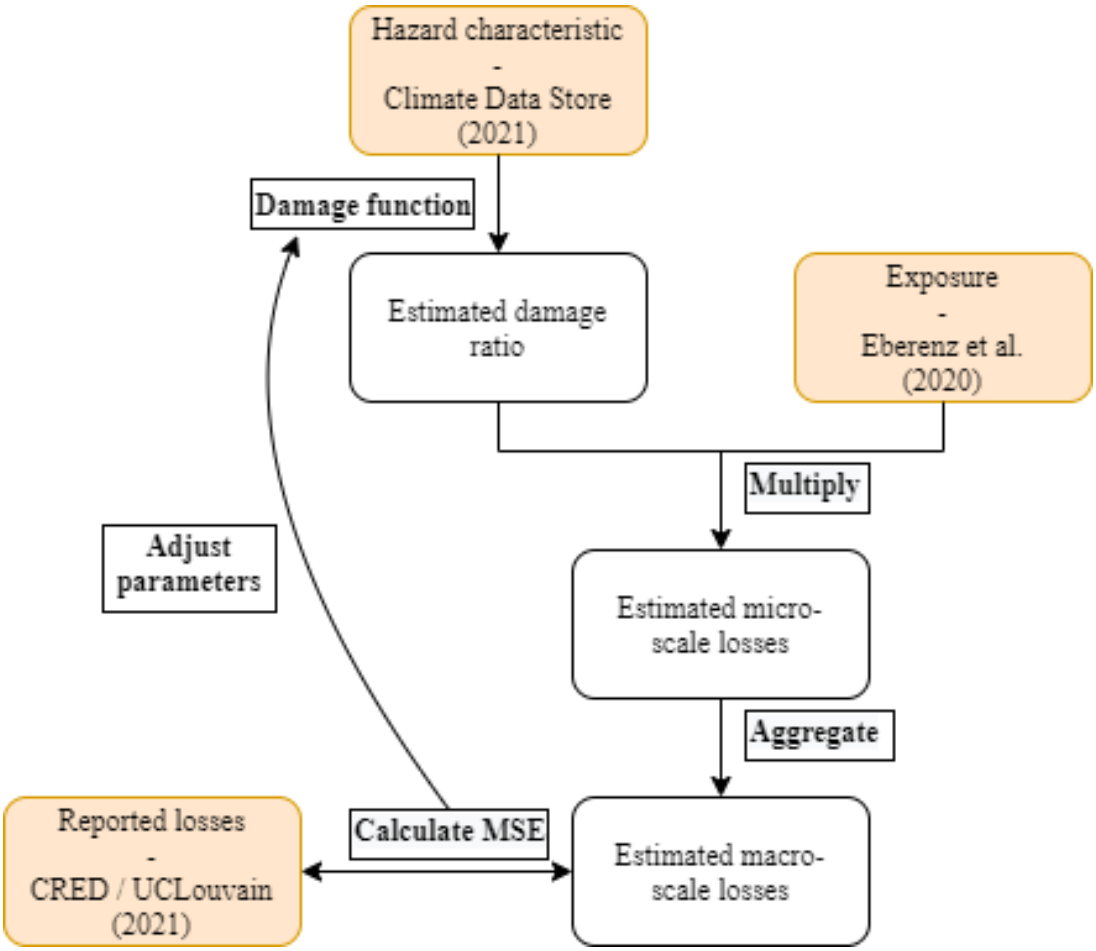


Figure 12: Visual representation of the model optimisation process described in Section 4.2. The process continues until the change in MSE between iterations is less than 10^{-6}. The input, or data sets, are shaded in orange.

C Friedman statistics table storms

Table 9: Pairwise Friedman test statistic based on the test MAPEs of all damage functions for all bootstrap iterations. The damage functions are the benchmark (Bench), the exponential function (Ex), the logistic function (L), the power law function (P), the function based on the one proposed by Klawa & Ulbrich (2003) (K), the function based on the one proposed by Emanuel (2011) (Em), and the function based on the one proposed by Heneka & Ruck (2008) (H). The '-U' and '-B' stand for univariate and bivariate and indicate a function including only maximum wind gust speed or maximum wind gust speed and storm duration, respectively. A positive test statistic indicates that the model based on the damage function mentioned in the row outperforms the model based on the damage function mentioned in the column. The value between parentheses behind the test statistic denotes the adjusted p-value of the statistic. The p-value is adjusted following the Holm step-down procedure. Only the statistics indicating significant differences are shown.

	Bench	Ex-U	Ex-B	L-U	L-B	P-U	P-B	K-U	K-B	Em-U	Em-B	H-U
Bench	-											
Ex-U	3.412 (0.039)	-										
Ex-B	-	-	-									
L-U	3.373 (0.045)	-	-	-								
L-B	-	-	-	-	-							
P-U	-	-	-	-	-	-						
P-B	-	-	-	-	-	-	-					
K-U	-	-	-	-	-	-	-	-				
K-B	-	-	-	-	-	-	-	-	-			
Em-U	-	-	-	-	-	-	-	-	-	-		
Em-B	-	-	-	-	-	-	-	-	-	-	-	
H-U	5.727 (0.0)	-	-	-	-	4.452 (0.001)	5.158 (0.0)	5.256 (0.0)	-	3.452 (0.035)	-	-

D Read me - programming codes

This section contains descriptions of the codes that are used in this research.

D.1 Python

- check_exposure_germany.py

  Get summary statistics for the Eberenz et al. (2020) exposure data set in Germany.

- correct_for_inflation.py

  Correct reported monetary values for inflation.

- process_netcdf_data.py

  For storms, read in netCDF files and create hourly wind gust speed data files for observations during storms for each decade. Moreover, for 2019 create hourly wind gust speed file and determine the threshold wind gust speed (98%) per observed location. The outputs are csv files.

- process_netcdf_floods.py

  For floods, read in netCDF files and create hourly total precipitation files for observations during the week prior to and the first day of the flood, for each decade. Moreover, for 2019 create hourly total precipitation file. The outputs are csv files.

- find_interpolation_parameter.py

  Find the decay factor that should be used in spatial interpolation. Can be used for both storms and floods.

- interpolate_for_exposure_locs.py

  For each location in the Eberenz et al. (2020) data set and for each storm, get an estimate for the maximum wind gust speed as well as the storm duration.

- interpolate_for_exposure_locs_floods.py

  For each location in the Eberenz et al. (2020) data set, obtain for each flood its cumulative precipitation as well as the antecedent precipitation index (API).

- get_duration_floods.py

  Create file that contains the duration in hours of each flood.

- make_dummy_matrix_region_affectd.py

  Returns a dummy matrix where for each exposure location (in the Eberenz et al. (2020) data set) it is shown whether the location was affected by a certain flood.

- test_storms.py and test_floods.py

  This code is used to create plots of the performance of the storm and flood damage functions, respectively. Moreover, a table of the estimated parameters along with their standard deviations are created and the Friedman test is performed.

- make_total_duration_matrix.py

  This code is used to get the total duration matrix for all storms. Total duration of a storm is defined as the number of hours that the maximum wind gust speed exceeds the local threshold value somewhere in the locations data set.

- fit_poisson.py

  This code fits a Poisson distribution to the number of storms per year and checks whether this distribution is reasonable to assume.

- fit_distribution_duration.py

  This code fits an exponential distribution to the total storm duration and checks whether this distribution is reasonable to assume.

- prepare_simulation.py

  This code is used to calculate interpolated maximum wind gust speeds for observations during the storms at the exposure locations in the German real estate portfolio of a Dutch FI.

- simulation.py

  This code simulates storms and estimates risk metrics for the German real estate portfolio of a Dutch FI.

- floods_july.py

  Returns the total precipitation and API files for the 2021 floods in Germany.

- plots_stress_test_july.py

  Shows the results of the stress test of the July 2021 floods in Ahrweiler, Germany, as heat maps.

D.2 Matlab

- main.m

  For both storms and floods, the main file is the code that needs to run in order to estimate the parameters in and get the performance of the damage functions.

- calculate_measures.m

  This function calculates several performance metrics based on the estimated damage ratios. The damage ratios are used to compute estimated losses. These are compared to the true losses to obtain the performance metrics.

- firstoutput.m

  This script returns the first output of the function this script is called on.

- Storm damage functions

  For each storm damage function, there is a script which estimates the parameters in the function for a given train set of storms. The scripts are:

  - optimize_benchmark.m
  - optimize_exp_univ.m
  - optimize_exp_biv.m
  - optimize_log_univ.m
  - optimize_log_biv.m
  - optimize_power_univ.m
  - optimize_power_biv.m

  - optimize_klawa_univ.m
  - optimize_klawa_biv.m
  - optimize_emanuel_univ.m
  - optimize_emanuel_biv.m
  - optimize_heneka_univ.m

- Flood damage functions

  For each flood damage function, there is a script which estimates the parameters in the function for a given train set of floods. The scripts are:

  - optimize_benchmark.m
  - optimize_log_univ.m
  - optimize_log_biv.m

- stress_test_july.m

  In this function we perform a stress test to forecast the damage ratios and losses due to the 2021 floods in Ahrweiler, Germany.