VILNIUS UNIVERSITY

FACULTY OF MATHEMATICS AND INFORMATICS

FINANCIAL AND ACTUARIAL MATHEMATICS STUDY PROGRAMME

Master's thesis

Projections of Windstorm Damage to Households in the Baltic States Under Climate Change

Vėjo audrų nuostolių prognozės Baltijos šalių gyventojų turtui klimato kaitos kontekste

Gintarė Kraujalytė

Supervisor : Dr. Rokas Gylys

Vilnius
2025

Summary

This thesis evaluates the projected impact of climate change on windstorm-related damage to residential properties in the Baltic states, focusing on the high-emission SSP5-8.5 scenario for the mid-21st century (2041–2070). It integrates hazard and exposure data together with vulnerability functions, employing methods such as bias correction and both stochastic and deterministic approaches to model damage functions for windstorm impact estimation. The analysis isolates the effects of climate change by assuming constant exposure and vulnerability. Results indicate an increase in future damage, particularly in northern and coastal regions, with spatial maps and exceedance probability curves illustrating geographic variability and associated uncertainty.

Keywords: climate change, windstorm, Baltic states, bias correction, impact functions, damage projections, exceedance probability curves, uncertainty.

Santrauka

Šiame darbe vertinamas klimato kaitos poveikis gyventojų turto nuostoliams, susijusiems su vėjo audromis Baltijos šalyse, naudojant aukštos emisijos SSP5-8.5 scenarijų XXI amžiaus viduriui (2041–2070). Darbe integruojami pavojingų gamtinių reiškinių ir pažeidžiamo turto duomenys su įtakos funkcijomis, taikant šališkumo korekciją, bei stochastinius ir deterministinius metodus žalos modeliavimui. Analizė vertina tik klimato kaitos poveikį, darant prielaidą, kad pažeidžiamas turtas ir įtakos funkcijos išlieka pastovios. Rezultatai rodo, kad ateityje vėjo audrų nuostoliai didės, ypač šiauriniuose ir pajūrio regionuose. Regioniniai žemėlapiai iliustruoja geografinius skirtumus, o viršijimo tikimybės kreivės - su tuo susijusį neapibrėžtumą.

Raktiniai žodžiai: klimato kaita, vėjo audros, Baltijos šalys, šališkumo korekcija, įtakos funkcijos, nuostolių prognozės, viršijimo tikimybės kreivės, neapibrėžtumas.

Contents

Introduction

In recent decades, Europe has experienced a noticeable increase in the impact of natural catastrophes, particularly those influenced by climate change. Among these, windstorms stand out as a major contributor to economic and insured losses. For instance, as published by one of the world's leading reinsurance company Swiss Re [1], in 2022, a cluster of storms in northwestern Europe triggered combined claims of an estimated USD 4.1 billion, bringing the total for this category to almost double the previous 10-year average. Also, in 2021 European Insurance and Occupational Pensions Authority (EIOPA) has published paper [7], which describes how climate change is expected to increase the frequency and severity of natural catastrophes. It emphasizes the need for Solvency Capital Requirements (SCR) for natural catastrophe underwriting risk to reflect the anticipated impact of climate change.

This thesis investigates the potential impact of climate change on windstorm damage in the Baltic states. The study combines insurance claims and exposure data with climate models to assess current and future risks.

By analyzing damage projections under the high-emission SSP5-8.5 scenario for the mid-21st century (2041–2070), this research aims to: quantify anticipated changes in windstorm damage, identify geographic areas of increased risk, and provide actionable insights for the insurance industry.

The analysis focuses solely on the effects of climate change, assuming all other factors, such as building distribution, value, and resilience, remain constant. This approach isolates the influence of climate change, providing a clearer understanding of its potential impact on windstorm related damage. By addressing this issue, the thesis contributes to the research on climate change impact, providing a regional perspective for the Baltic states.

1 Damage projections modeling theory

1.1 Modeling process

The process of modeling windstorm damage under climate change involves six key steps. First, exposure data is defined by identifying insured assets, their locations and values, providing the baseline for risk assessment. Next, the exposure is linked to weather hazard data, typically derived from climate models like CMIP6, which provide information on hazard intensity, such as wind speed, across time and space. In the third step, a vulnerability function, also referred to as a damage or impact function, is applied, translating hazard intensity into expected damage based on parameters like threshold wind speed or maximum damage degree. This is followed by loss calculation, where tools like CLIMADA integrate exposed objects, hazard data, and damage functions to estimate potential damage or losses, enabling a detailed understanding of risk. In the fifth step, damage projections are created using climate change scenarios (e.g., future wind patterns) to assess how damage may evolve under different conditions. Finally, exceedance probability curves are generated to quantify the likelihood of damage exceeding certain thresholds, providing insights into risk management and decision-making under a changing climate.

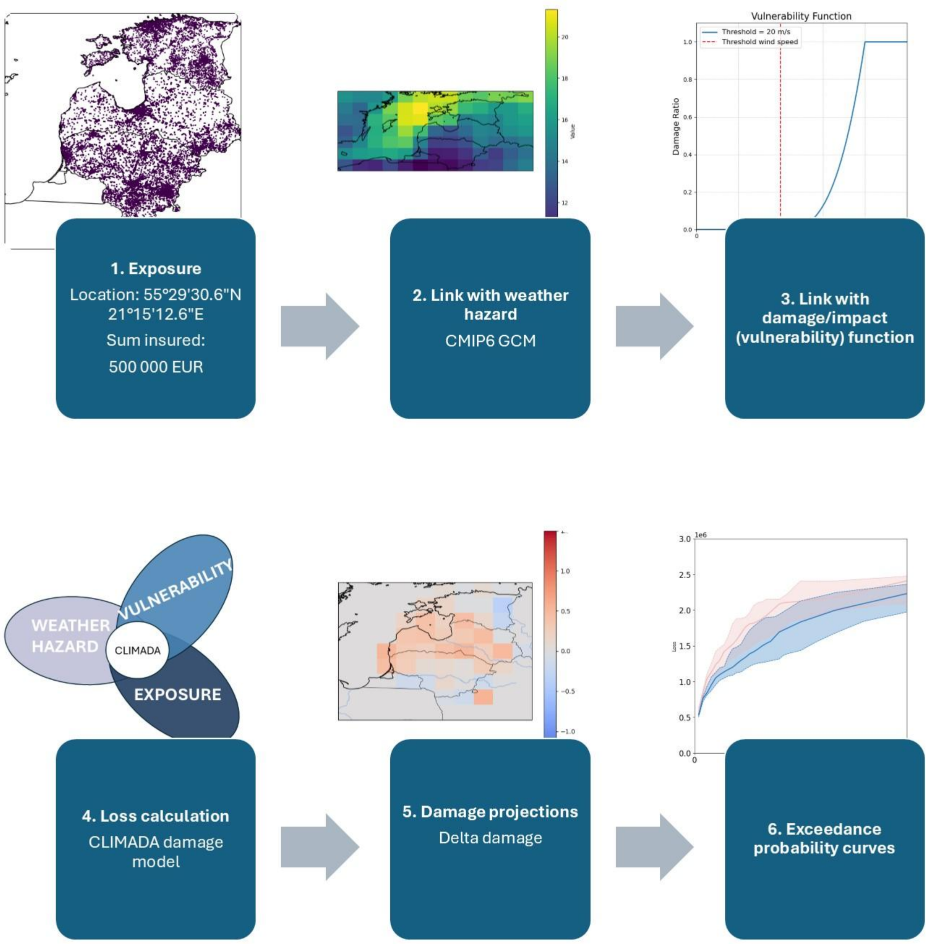


Figure 1 *Model scheme*

1.2 Bias correction

Global Climate Models (GCMs) are the primary tools for constructing climate scenarios. They provide the foundation for assessing the impact of climate change across various spatial scales, from local to global. However, impact studies rarely use GCM outputs directly because climate models exhibit systematic error (biases) due to the limited spatial resolution, simplified physics and thermodynamic processes [12].

For instance, GCMs typically divide the Earth's surface into large grid cells (e.g., 100–250 km), which smooths out fine-scale variations, leading to inaccuracies in capturing localized phenomena such as small-scale storms or temperature fluctuations. Additionally, GCMs might assume gauging methods that differ from those employed in regional observation datasets [17].

Hence, it is important to bias-correct the raw climate model outputs in order to produce climate projections that are better fit for modeling.

Bias correction involves scaling climate model outputs to account for systematic errors, ensuring better alignment with observational data. Several bias-correction approaches have been developed ranging from simple linear scaling to sophisticated quantile mapping [18].

Linear scaling method

Linear scaling (LS) is a parametric method used to correct model data using monthly correction factors. The method adjusts the data based on the differences between observed and modeled values for specific variables. LS can be applied either additively or multiplicatively, depending on the nature of the variable being corrected.

Additive correction: For variables measured on an absolute scale (temperature, pressure), LS uses an additive term. The corrected value is calculated as:

$$V_{i,m,\text{corr}}(d) = V_{i,m}(d) + \Delta V_{i,m}, \tag{1}$$

$$\Delta V_{i,m} = \mu_m(V_{\text{obs}}) - \mu_m(V_{\text{model}}), \tag{2}$$

where:

$V_{i,m}(d)$: is the modeled value for month m and day d for the climate model i,

$\Delta V_{i,m}$: is the difference between the mean modeled value and the observed value for month m,

μ_m : represents the monthly mean of daily values during the calibration period.

Multiplicative correction: Other climate variables are adjusted with a multiplier. The corrected value is calculated as:

$$V_{i,m,\text{corr}}(d) = V_{i,m}(d) \cdot c_{i,m}, \tag{3}$$

$$c_{i,m} = \frac{\mu_m(V_{\text{obs}})}{\mu_m(V_{\text{model}})}. \tag{4}$$

where:

$c_{i,m}$: is the correction factor for month m,

$\mu_m(V_{\text{model}})$: is the monthly mean of observed values during calibration period,

$\mu_m(V_{\text{obs}})$: is the monthly mean of modeled values during the calibration period.

The advantage of the linear scaling method is its simplicity and effectiveness in adjusting the average long-term weather patterns. On the other hand, this method adjusts only the mean of the distribution and fails to correct other statistical properties like variability, skewness, or extremes.

Empirical quantile mapping

Empirical quantile mapping (EQM) method is used to correct the distribution function of simulated climate variable to be consistent with the observed distribution function, by adding a quantile-specific transfer function determined by the difference between simulated and observed empirical cumulative distribution functions (CDF). Given a specific daily model value V^d_{model} for a variable for validation and future period, EQM is trying to find the closest model quantile (CDF value) for calibration period, which is used to correct the model bias in validation and future period. The bias correction is applied using the equations:

$$V^{\text{corr},d}_{\text{model}} = V^d_{\text{model}} - \text{Bias}^{\text{closest}}_{\text{CDF(model)}-\text{CDF(obs)}}, \tag{5}$$

$$\text{Bias}^{\text{closest}}_{\text{CDF(model)}-\text{CDF(obs)}} = V^{\text{closest}}_{\text{CDF(model)}} - V^{\text{closest}}_{\text{CDF(obs)}}. \tag{6}$$

The EQM algorithm can be summarized as follows:

1. Compute and store the cumulative distribution function for simulations and observations for all grid cells during the calibration period.
2. For each grid cell, get the model value for the validation/future period (V^d_{model}) and compute the absolute differences between $V_{\text{CDF(model)}}$ and V^d_{model}.
3. Identify the percentile where the absolute difference is minimized, referred to as the closest percentile.
4. Use the bias at the closest percentile ($\text{Bias}^{\text{closest}}_{\text{CDF(model) - CDF(obs)}}$) to correct the model value V^d_{model}.

This method not only corrects the mean and variance but also corrects the whole distribution.

Figure 2 compares observed and simulated distributions at different quantiles (e.g., 0.25, 0.5, 0.75), illustrating the biases in the uncorrected simulated data. The bottom plot shows the cumulative distribution functions for both datasets, emphasizing the divergence between observed and simulated data. Quantile mapping aligns these distributions, improving the accuracy of the simulated values. The differences in the quantiles, as shown by the black arrows, highlight the necessity of bias correction.

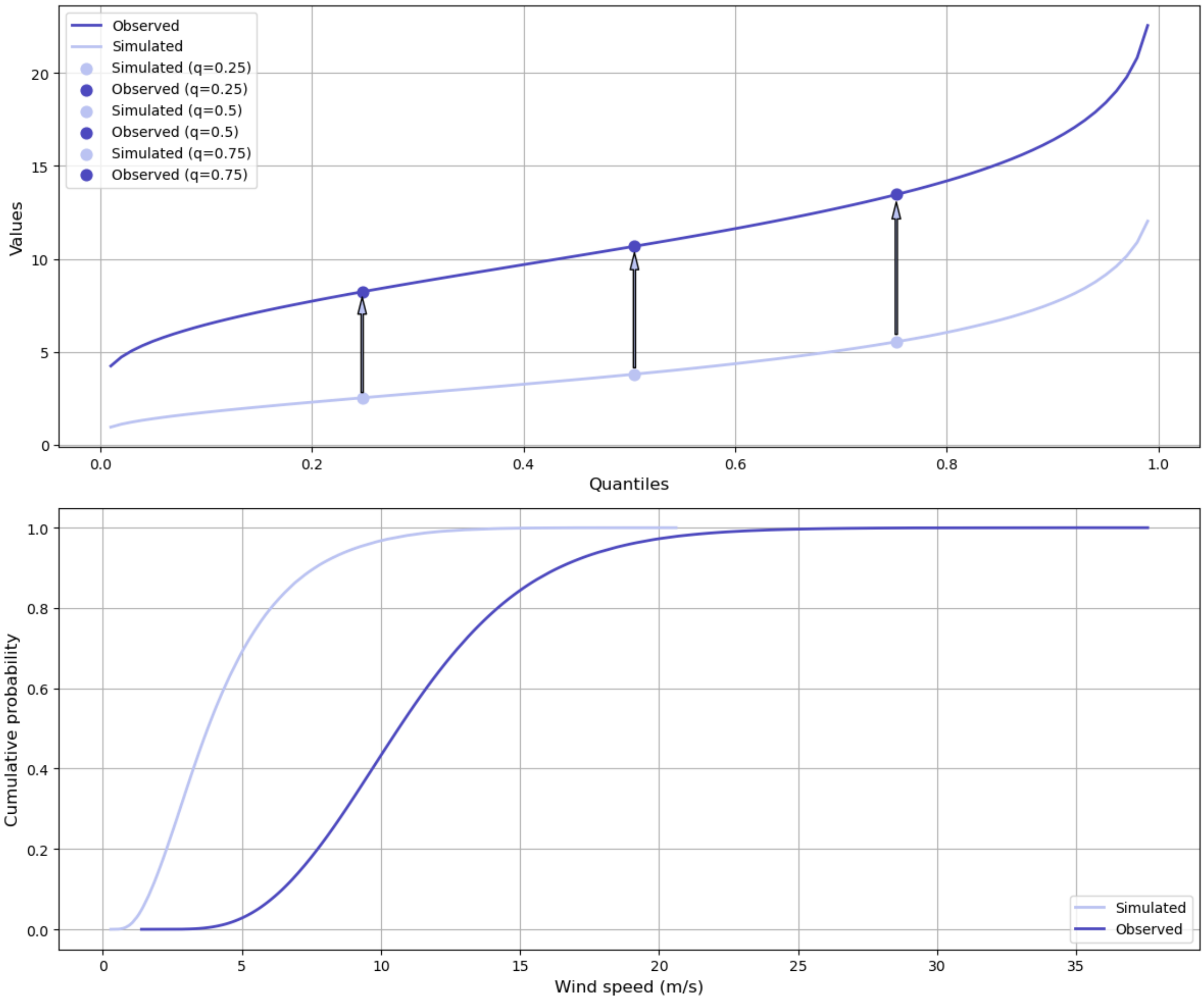


***Figure 2** Quantile mapping and cumulative distribution functions: simulated vs observed*

1.3 Vulnerability functions

Vulnerability functions relate the percentage of damage in the exposure to the hazard intensity. These functions are a crucial component in modeling windstorm impacts. Below, damage functions that have been utilized in different windstorm damage studies are discussed.

Exponential damage function

The exponential damage function assumes that losses increase exponentially with increasing wind speed [4]. It is a simple damage function with two parameters to be estimated. It is expressed as:

$$d(\nu) = e^{\alpha(\nu-\beta)}, \tag{7}$$

where:

α: is a scale parameter,

β: is a location parameter.

The loss $L(\nu)$ is related to the damage function $d(\nu)$ by the following equation:

$$L(\nu) = c \cdot d(\nu),$$

where c is a scaling constant that accounts for factors such as the monetary value of assets, exposure, or other real-world adjustments necessary to translate the damage function into actual losses.

Cubic-excess over threshold damage function

The damage function proposed by Klawa and Ulbrich [9] suggests that the loss increases cubically for wind speeds beyond a certain threshold. More precisely, the damage function is the third power of wind speed above the 98th percentile scaled by the 98th percentile. The damage function is:

$$d(\nu) = \begin{cases} \left(\dfrac{\nu - v_{98}}{v_{98}}\right)^3, & \text{if } \nu \geq v_{98}, \\ 0, & \text{if } \nu < v_{98}. \end{cases} \tag{8}$$

The total loss is then calculated using linear regression:

$$L(\nu) = \beta_0 + \beta_1 d(\nu), \tag{9}$$

where β_0 and β_1 are regression coefficients. The intercept term β_0 in the fitted linear regression can be interpreted as the base loss, which is the loss estimate for all wind speeds below the 98th percentile. However, using this loss offset for all wind speeds below the 98th percentile does not allow to address the randomness in the lower loss spectrum.

Probabilistic damage function by Prahl

The probabilistic damage function proposed by Prahl [13] employs a two-step fitting procedure estimating the occurrence probability and the loss magnitude.

The probability of damage for a given wind speed ν is modeled using a sigmoid function:

$$p(\nu) = 1 - \frac{\gamma_0}{1 + e^{\gamma_1(\nu - \gamma_2)}}, \tag{10}$$

where:

γ_1: controls the steepness of the curve,

γ_2: sets the wind speed threshold for increased probabilities,

γ_0: represents the base probability of losses.

The magnitude of non-zero losses is estimated using:

$$\tilde{M}_v \approx \sigma_0 + \left(\frac{\nu}{\sigma_2}\right)^{\sigma_1}, \tag{11}$$

where:

σ_0: is the offset loss,

σ_1: is the shape parameter,

σ_2: is the scaling parameter.

Based on the assumption that the observed losses follow a log-normal distribution, $\mathcal{LN}(\mu, \sigma^2)$, the stochastic loss magnitude is described as a random variable:

$$M_v \sim \mathcal{LN}(\ln\left(\tilde{M}_v\right), \sigma_3).$$

The location parameter of the log-normal distribution is related to the median by $\mu = \ln\left(\tilde{M}_v\right)$. The scale parameter $\sigma^2 = \sigma_3$ describes both the variability due to imprecise wind observation and the aleatory uncertainty regarding the damage caused.

The probability of damage and the magnitude of loss are treated as independent variables, conditional on wind speed. The damage function is defined as the product of the probability of damage and the magnitude of loss:

$$d(v) = I(v) \cdot M(v),$$

where $I(v)$ is an indicator random variable representing whether damage occurs at wind speed v:

$$I(v) = \begin{cases} 1, & \text{with probability } p(v), \\ 0, & \text{with probability } 1 - p(v). \end{cases}$$

Windstorm damage function by Emanuel

Emanuel [5] damage function models the fraction of property value lost $d(v)$ as:

$$d(v) = \frac{v^3}{1 + v^3},$$

where the normalized wind speed v is defined as:

$$v = \text{max}\left(\frac{V - V_{\text{thresh}}}{V_{\text{half}} - V_{\text{thresh}}}, 0\right),$$

where:

v: is the normalized wind speed above the damage threshold,

V: is the actual wind speed,

V_{thresh}: is the threshold wind speed, below which no damage occurs ($d(v) = 0$),

V_{half}: is the wind speed at which half the property value is lost ($d(v) = 0.5$).

This function ensures that no damage occurs for wind speeds below V_{thresh}, and damage increases cubically with wind speed, approaching unity as V becomes much greater than V_{half}:

$$d(v) \to 1, \text{ as } v \to \infty.$$

1.4 Damage projections

Damage projections are derived from m simulations for each GCM and both scenarios (historical and future) through the following steps:

1. **Input data:** Daily wind speed data from a selected climate model scenario.
2. **Bias correction:** The empirical quantile mapping bias correction algorithm is applied to align the modeled wind speed distribution with the observed distribution in the Baltic states region. For each grid cell, the bias-corrected wind speed v_{corr} is calculated as:

   $$v_{\text{corr}}(q) = v_{\text{model}}(q) - \left[v_{\text{model-cal}}(q) - v_{\text{obs}}(q)\right],$$

   where $v_{\text{model}}(q)$ is model value at quantile q, $v_{\text{model-cal}}(q)$ is model value at quantile q from the calibration period, $v_{\text{obs}}(q)$ is the observed value at quantile q.

   This method ensures that the corrected wind speeds v_{corr} are consistent with the observed statistical distribution of wind speeds, improving the reliability of projections for impact modeling (see 1.2).
3. **Storm day detection:** Days are selected when winds surpassing minimum wind intensity threshold and 99th percentile of historical daily wind intensity cover an area greater than 1.5×10^5 km^2 (see 2.3).
4. **Affected exposure selection:** For each detected storm day, affected exposures are stochastically selected using $I(v)$ indicator random variable equal 1 with probability $p(v)$ (see 1.3).
5. **Damage calculation:** Predefined damage functions (see 2.4) are applied to each affected exposure object, estimating damage based on wind intensity, exposure object value and damage

function:

$$\text{damage} = \text{exposure} \times f_{\text{imp}}(\text{hazard intensity}).$$

6. **Event level aggregation:** Damage from all affected exposure objects and all centroids is aggregated to calculate the total damage for each storm event.

7. **Damage projections:** m simulations are used to derive damage distribution for specific metric, such as damage associated with desired return period or the largest annual loss. These distributions offer insights into the probabilistic range of damage and the uncertainties surrounding the projections.

1.5 Risk assessment metrics

Risk metrics for insurance damage are key indicators used to assess and quantify the financial and operational risks associated with insured events, such as windstorms. These metrics help insurers evaluate potential exposures, set premiums, prepare for future claims and enter into appropriate reinsurance agreements. Below common measures are defined [8].

Exceedance probability

The *Exceedance probability* (EP) is a widely used measure in catastrophe modeling. It represents the probability that a certain loss value will be exceeded within a defined future time frame. This metric is crucial for planning responses to potential hazards and offering property owners an evaluation of their risk.

Let $D_1, D_2, \dots$ be a set of natural disasters. Let p_i and X_i be an annual probability of occurrence and a corresponding total loss associated with a natural disaster D_i. Thus, D_i is a Bernoulli random variable with:

$$\mathbb{P}(D_i \text{ occurs}) = p_i$$

$$\mathbb{P}(D_i \text{ does not occur}) = 1 - p_i.$$

If an event D_i does not occur, the loss is zero. The expected loss for a given event D_i in a given year is $\mathbb{E}[L] = p_i \mathbb{E}[X_i]$.

The exceedance probability is the probability that a loss random variable exceeds a certain amount of loss. This probability is sometimes denoted as $EP(x)$ and is called the *exceedance probability curve*. Let X be a loss random variable. Then:

$$EP(x) = \mathbb{P}(X > x) = 1 - \mathbb{P}(X \leq x).$$

Using probabilistic terminology, $EP(x)$ is the survival function of X.

In particular, if $x = X_i$, which is a loss associated with a disaster D_i, then:

$$EP(X_i) = \mathbb{P}(X > X_i) = 1 - \mathbb{P}(X \leq X_i) = 1 - \prod_{j=1}^{i} (1 - p_j),$$

where $D_1, D_2, \ldots, D_i$ are the events with higher level of losses such that $X_1 \geq X_2 \geq \cdots \geq X_i$.

A characteristic sometimes associated with the exceedance probability is the *return period* or the *loss return period* (RP) of a natural disaster. It is calculated as a reciprocal of the EP:

$$RP = \frac{1}{EP}.$$

Occurrence exceedance probability

The *occurrence exceedance probability* (OEP) is the probability that the largest loss in a year exceeds a certain amount of loss. This probability is sometimes denoted as $O(x)$ and is called the *occurrence exceedance probability curve*.

Let $X_1, X_2, \ldots, X_N$ be independent and identically distributed losses in a given year. Then

$$O(x) = \mathbb{P}(\max_{1 \leq i \leq N}(X_i) > x) = 1 - \mathbb{P}(\max_{1 \leq i \leq N}(X_i) \leq x) = 1 - \prod_{i=1}^{N} \mathbb{P}(X_i \leq x).$$

Using probabilistic terminology, if $X_{(1)}, X_{(2)}, \ldots, X_{(N)}$ is the ordered statistic with $X_{(N)} = \max_{1 \leq i \leq N} X_{(i)}$, then $O(x)$ is the survival function of $X_{(N)}$.

Let $F(x)$ be the cumulative distribution function (CDF) of X. Then for a fixed N, the OEP is:

$$O(x) = 1 - (F_X(x))^N .$$

If N is the random claim count with the probability mass function (PMF) P_N, then by the law of total probability,

$$O(x) = \sum_{n=0}^{\infty} \mathbb{P}(\max_{1 \leq i \leq n}(X_i) > x \mid N = n)\mathbb{P}(N = n).$$

This simplifies to:

$$O(x) = 1 - \sum_{n=0}^{\infty} \mathbb{P}(\max_{1 \leq i \leq n}(X_i) \leq x \mid N = n)\mathbb{P}(N = n).$$

$$O(x) = 1 - \sum_{n=0}^{\infty} \left(\prod_{i=1}^{n} \mathbb{P}(X_i \leq x) \right) \mathbb{P}(N = n) = 1 - \sum_{n=0}^{\infty} (F_X(x))^n \, \mathbb{P}(N = n).$$

$$O(x) = 1 - \mathbb{E}_N \left((F_X(x))^N \right) = 1 - PGF\,(F_X(x))\,,$$

where $PGF(x)$ is the probability generating function for N defined as:

$$PGF(t) = \mathbb{E}(t^N) = \sum_{n=0}^{\infty} t^n \cdot \mathbb{P}(N = n).$$

Thus,

$$O(x) = 1 - PGF(F_X(x)).$$

The expected value of $X_{(N)}$ is by definition:

$$\mathbb{E}[X_{(N)}] = \int_0^\infty O(x)\,dx.$$

These curves are particularly important for reinsurance because they help quantify the likelihood of large scale loss events that can significantly impact the financial stability of insurers. By focusing on the maximum annual loss, OEP curves allow insurers to better understand their exposure to catastrophic events. This insight is important for designing reinsurance treaties, such as excess of loss, quota share or catastrophe reinsurance, which are specifically tailored to protect insurers against high severity, low frequency events.

2 Study

This study assesses the financial losses caused by windstorms in the Baltic states under changing climate conditions. The damage is modeled as the impact of intense surface wind gusts on residential assets. The modeling framework integrates hazard, exposure, and vulnerability functions to calculate impact and risk metrics in a probabilistic and geographically explicit manner.

The influence of climate change is analyzed by comparing damage estimates for a future period to those from a historical reference period, while keeping exposure and vulnerability constant over time. The results are expressed as the percentage change in damage (ΔD%), calculated using the formula:

$$\Delta D\% = \frac{D_{\text{future}} - D_{\text{historical}}}{D_{\text{historical}}} \times 100$$

where:

D_{future}: is the damage computed for the future period,

$D_{\text{historical}}$: is the damage computed for the historical reference period.

This approach highlights the effect of climate change on windstorm related damage, isolating it from potential changes in exposure (e.g., asset distribution) or vulnerability (e.g., building resilience). The following sections provide a detailed description of the model components and the data used in the analysis.

2.1 Data

Hazard

Daily surface wind maximum outputs from 7 global circulation models (GCM) participating in Coupled Model Intercomparison Project 6th phase (CMIP6) were used to represent windstorm hazard. Model data was kept on the original model grids, as provided by the modeling centers. The selection of general circulation models was based on two criteria:

1. The GCM provides a simulation for both the historical period and the future period.
2. The GCM has a resolution of $1^\circ \times 1^\circ$.

Historical experiments show how the GCMs perform for the past climate and are used as a reference period for comparison with scenario runs for the future. For future projections, the high-emission scenario Shared Socioeconomic Pathway (SSP) 5-8.5, which assumes continued increases in CO_2 emissions throughout the 21st century, was chosen [3]. Table 1 summarizes the climate models used for the damage projections of this study. The historical scenarios cover a 30-year period from 1985 to 2014, while the future scenarios focus on the mid-century period, spanning from 2041 to 2070. Each climate model was considered separately to compute damage projections. Considering the climate models separately allows to investigate the climate model uncertainty in the projections.

Table 1 *Climate models used in the study*

Climate Model	Research country
BCC-CSM2-MR	China
CESM2	USA
CESM2-WACCM	USA
NorESM2-MM	Norway
TaiESM1	Taiwan
GFDL-ESM4	USA
CMCC-CM2-SR5	Italy

ERA5

ERA5 is a comprehensive global fifth generation reanalysis dataset developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). Reanalysis combines model data with observations from across the world into a globally complete and consistent dataset using the laws of physics. Unlike regular analysis, which focuses on current conditions for weather forecasting, reanalysis reconstructs past weather and climate conditions by integrating historical observations and advanced numerical models, filling gaps in data to provide a seamless record of the Earth's system over time [2].

ERA5 provides hourly climate data with a resolution of $0.25° \times 0.25°$ for atmospheric variables. The datasets are available in a gridded format and are updated on a daily basis.

Regridded at $1° \times 1°$ and resampled to a daily resolution ERA5 dataset was used as an observational data to bias correct GCMs using quantile mapping approach described in 1.2 subsection.

The regridding process uses bilinear interpolation to map data from the old grid to the new grid. For each new grid cell (shown in green in Figure 3), the value is calculated as a weighted average of the nearest four old grid cells (highlighted in red). The weights are based on the distances between the new grid cell and the old grid cells, ensuring that closer grid cells contribute more to the final value.

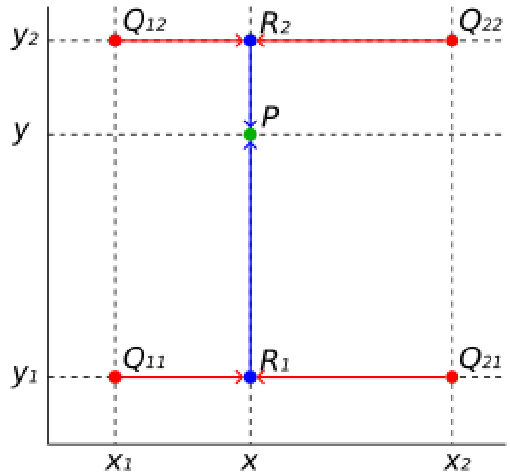


Figure 3 *Bilinear interpolation scheme*

Exposure

This study uses exposure data from Insurance Company, which operates within the Baltic region in Estonia, Latvia, and Lithuania. The dataset includes insured household objects as of August 31, 2024, covering residential houses, apartments, and other privately owned buildings.

Insurance loss data

To fit the vulnerability function parameters and calibrate the model for the Baltic states, historical claims data from the same Insurance Company portfolio was used. The dataset covers a decade-long period of loss data. To ensure consistency and comparability over time, the losses were adjusted for inflation. This adjustment was performed using a weighted index, consisting of 50% harmonized index of consumer prices (HICP) - overall index and 50% HICP - materials for the maintenance and repair of the dwelling index, calculated separately for each country [6].

Also, the year 2023 was set aside for backtesting and therefore excluded from the training set.

2.2 CLIMADA module

CLIMADA is a climate risk assessment tool which integrates exposure, hazard, and vulnerability data to calculate risk. It is an open source platform built in Python, providing a comprehensive framework for analyzing climate related impacts.

This study utilized CLIMADA primarily for two key steps:

1. **Mapping exposure objects to hazard centroids (grid points):** Exposure coordinates must be assigned to their nearest centroids. Number of centroids and their location is defined in GCM. Figure 4 illustrates this process, where each exposure point is matched to its closest hazard centroid (illustrative randomly generated exposure data sample).

2. **Impact calculation:** This step combines exposure, hazard, and vulnerability data to quantify the potential impacts of climate hazards. The severity of impact is calculated as:

$$\text{severity} = f(\text{hazard intensity, exposure, vulnerability}) = \text{exposure} \times f_{\text{imp}}(\text{hazard intensity}),$$

where f_{imp} is the impact function, which defines the extent to which an exposure is affected by a specific hazard.

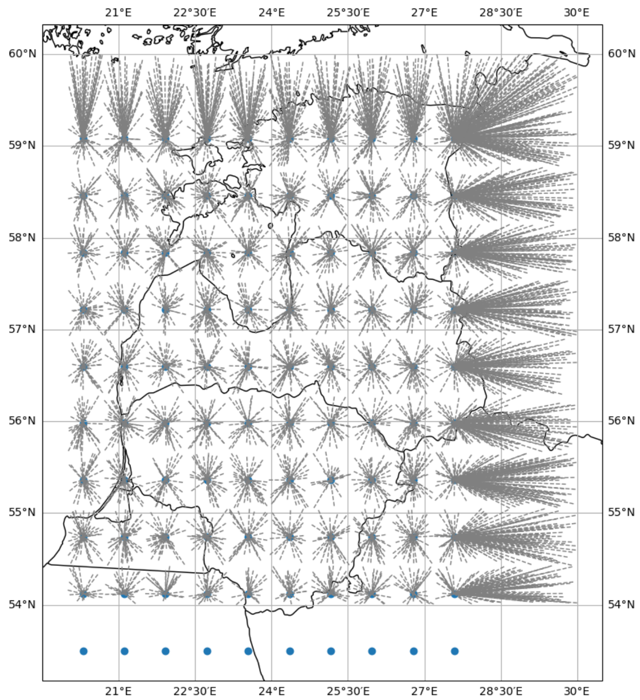


Figure 4 *Process of exposure object assignment to hazard centroids*

2.3 Storm days selection criteria

Storm days are selected from the bias-corrected daily GCM data based on two conditions applied locally at each grid point. First, a grid cell in the GCM is considered stormy if the daily maximum surface wind intensity exceeds the local 99th percentile value, calculated from the model's historical simulation. Previous studies [9, 15] have used the 98th percentile of local wind intensity as the threshold for storm-related damage. However, backtesting with empirical data indicates that the 99th percentile is more accurate for identifying storms in the Baltics. The selected daily wind speeds are further refined to ensure they correspond to intensities capable of causing actual damage, by requiring that daily wind speed exceeds 22 m/s to be classified as stormy.

The histogram (Figure 5) shows an increase in claim frequency starting from wind speeds of approximately 22 m/s, indicating that lower wind speeds are less likely to cause damage resulting in claims. This threshold is supported by the observation that 80% of damage occur at wind speeds above 22 m/s, reducing the influence of outliers from lower wind speed ranges that might not represent typical damage-causing events, ensuring the analysis remains focused on impactful storms.

Also, the total area of the stormy grid cells on a particular day must amount to a minimum A_{min} for the day to be considered stormy. Stormy grid cells do not have to be contiguous to be included in the total stormy area required for a storm day. A value of 1.5×10^5 km^2 is chosen for A_{min}, which is representative of the typical area of the wind footprint of an mid-latitude storm [10]. Therefore, storm day is defined in the following formula [16]:

$$\text{Stormy day}_t \iff \sum_i \Big(a_i \cdot \big[(v_{i,t} \geq v_{i,99}) \wedge (v_{i,t} \geq v_{\text{threshold}}) \big] \Big) \geq A_{\text{min}},$$

where:

$v_{i,t}$: daily maximum wind intensity at grid cell i on day t,

$v_{i,99}$: 99th percentile of historical daily wind intensity at grid cell i,

$v_{\text{threshold}}$: minimum wind intensity threshold (22 m/s),

a_i: area of grid cell i,

A_{min}: minimum total affected area threshold (e.g., 1.5×10^5 km^2).

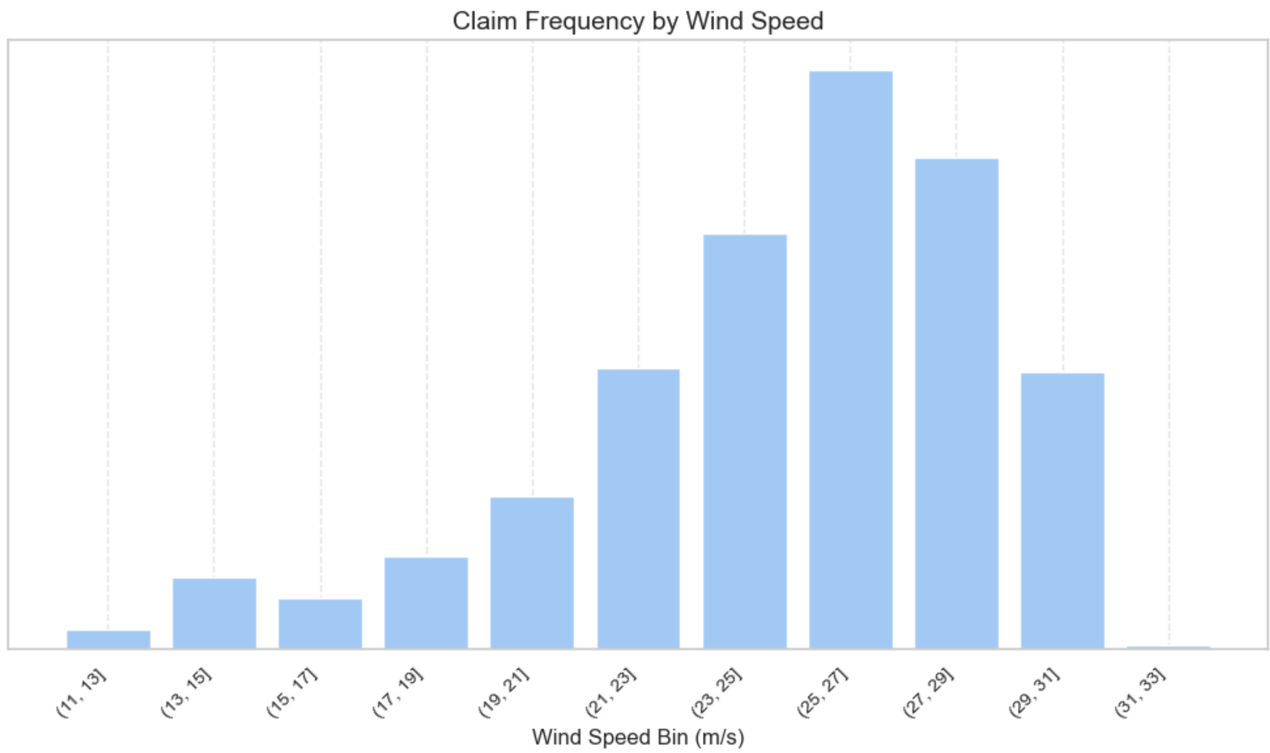


***Figure 5** Empirical claims frequency*

2.4 Vulnerability functions

In this study, three damage functions were utilized: a deterministic step function, generalized linear model (GLM) and a stochastic damage function.

Deterministic step function

Deterministic step vulnerability function was constructed to project storm damage for both historical and future climate scenarios. This function characterizes the mean damage degree (MDD) - the percentage of an exposure value that is damaged at a given wind intensity. Step function was directly derived from an empirical claims data statistics:

$$\text{MDD}(v) = \begin{cases} 0, & \text{if } v < v_{\text{threshold}}, \\ \text{MDD}_{\text{optimal}}, & \text{if } v \geq v_{\text{threshold}}, \end{cases} \tag{12}$$

where:

- v: maximum daily wind speed,
- $v_{\text{threshold}} = 22\,\text{m/s}$,
- $\text{MDD}_{\text{optimal}} = 0.018041918$.

Threshold of 22 m/s is already described in subsection 2.3. MDD is selected as an average damage ratio from empirical claims data.

Generalized linear model

Damage function was defined using a generalized linear model with a Gamma distribution and an inverse link function. The model predicts MDD based on wind speed (v) and categorical variables representing building characteristics (the construction period and the building type):

$$MDD(v) = \frac{1}{\alpha + \beta v + \gamma \cdot \text{C}(\text{Construction year}) + \delta \cdot \text{C}(\text{Building group})},$$

where

α: is the coefficient representing the intercept of the model,

β: is the coefficient for the effect of wind speed v,

γ: is the coefficient accounting for the construction year,

δ: is the coefficient accounting for the building group.

Here $C(\cdot)$ denotes the categorical nature of the variable.

The Gamma family distribution is specified, along with the inverse link function, to appropriately model the skewness and variability in the MDD.

Stochastic function

A more complex stochastic vulnerability function was developed, utilizing the power law-based approach introduced by Prahl [14] as a foundation to describe the magnitude of losses:

$$\tilde{MDD}(v) \approx \left(\frac{v}{b}\right)^c + a,$$

where:

v: is the maximum daily wind speed,

a: is the offset loss,

b: is the constant to scale the wind speed,

c: is the shape parameter.

Since the building type and year of construction influence the extent of storm damage to properties, these factors were also incorporated into the vulnerability function. Different categories are presented in the Table 2 For each building group i and construction year j:

$$\tilde{MDD}(v) \approx \left(\frac{v}{b_{i,j}}\right)^{c_{i,j}} + a_{i,j}.$$

To assess MDD, vulnerability function was calibrated in respect of parameters a, b and c for each building group and construction year combination. The process used Bayesian optimization [11], iteratively comparing the modeled MDD against observed values while minimizing the mean squared error.

The group "Before 1970s - Other" was identified as the most vulnerable, resulting in the highest MDD at a given wind speed compared to other exposure groups.

Table 2 *Building type and construction year groups*

Building group	Construction year
Apartments	Before 1970s
Residential houses	Before 1970s
Other	Before 1970s
Apartments	1970s 1980s
Residential houses	1970s 1980s
Other	1970s 1980s
Apartments	1990s 2000s
Residential houses	1990s 2000s
Other	1990s 2000s
Apartments	2010s present
Residential houses	2010s present
Other	2010s present

In accordance with the observation by Prahl, the residuals between empirical data and modeled data are assumed to follow a log-normal distribution, denoted as $\mathcal{LN}(\mu, \sigma^2)$. This assumption holds in this study as well. Histogram (Figure 6) shows that log-transformed residuals follow a symmetric distribution centered around zero. QQ plot (Figure 7) indicates that normal distribution is a close fit with slight deviation in the tails. Consequently, the mean damage degree is described as a stochastic variable:

$$MDD(v) \sim \mathcal{LN}(\ln\Big(\tilde{MDD}(v)\Big), \sigma^2).$$

The variance σ^2 is assumed to be constant and is derived from the total residuals dataset. It is computed after applying a natural logarithm transformation to the observed and modeled MDD values.

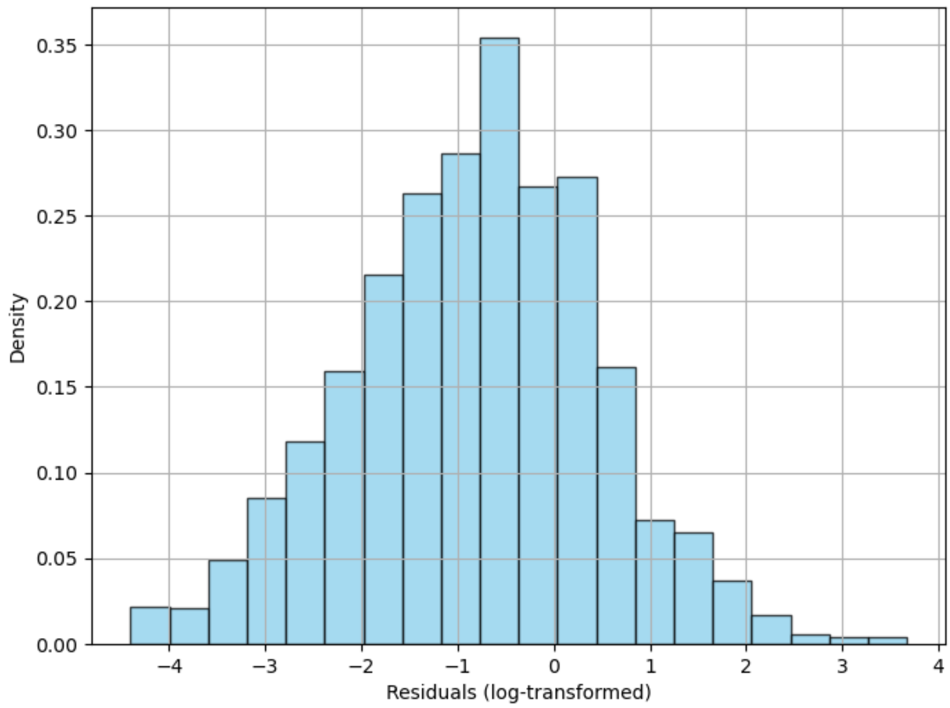


Figure 6 *Histogram of log-transformed residuals*

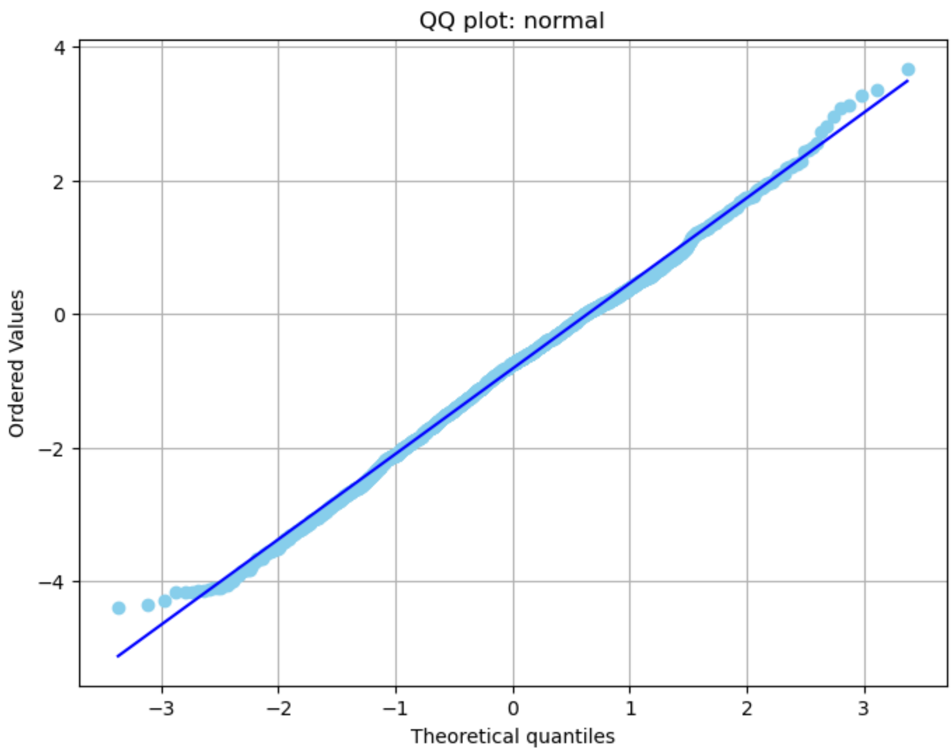


Figure 7 *QQ plot of log-transformed residuals*

2.5 Claims frequency

So far the study accounted for storm days frequency and damage magnitude when loss event happens. However, it is also important to estimate the proportion of exposed objects that experience damage during a storm. Binomial distribution was used to model the count of damaged exposures:

$$E(X) = \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} = np,$$

where n is the number of exposed objects and p is the probability of loss.

To estimate claim occurance probability, binned empirical claims data was fit with a probability function [14]:

$$p(v) = 1 - \frac{\alpha}{1 + e^{\gamma(v-\beta)}},$$

with base probability $(1-\alpha)$, shift β, and slope γ.

The parameters α, β, and γ of the $p(v)$ curve were estimated using a non-linear least squares method, which minimizes the differences between the observed data and the modeled values. Initial guesses for the parameters were selected to assist the fitting process.

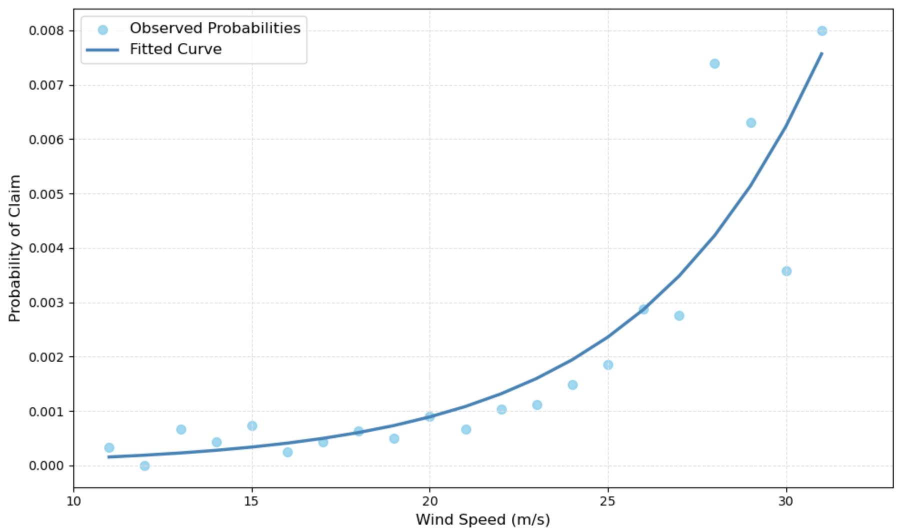


Figure 8 *Claims frequency: observed vs fitted*

It was noted that in the Baltic states, wind speed has a stronger correlation with the frequency of claims count rather than the claims magnitude.

2.6 Backtest

To validate the reliability of the developed climate change impact model, a backtest was performed using ERA5 data for the year 2023. Notably, historical claims from storms that occurred in 2023 were excluded from the calibration process to ensure the independence of the validation data.

Backtesting was conducted for all vulnerability functions separately. For stochastic function the differences between the modeled and the realized 2023 damage ranged from -10% to +10%, with the mean difference being very close to the actual 2023 damage. For the step function backtesting showed higher modeled results compared to the realized loss, with deviations reaching up to +14%, depending on the simulation. The GLM showed a tendency to lightly underestimate damage, with deviations ranging from -15% to +2%.

Overall, the results demonstrated that each vulnerability function has distinct strengths, with deviations falling within a reasonable range. The stochastic function provides balanced estimates while capturing a greater degree of variability, the step function effectively captures upper bounds, and the GLM slightly underestimates damage, with minimal risk of overestimation. Together, these functions highlight the value of combining deterministic and stochastic approaches in estimating windstorm impact.

3 Results

3.1 Damage projections

This section presents results from three vulnerability functions applied to windstorm scenarios. Tables below present damage projections for a 10-year return period, comparing historical and future scenarios across seven different climate models. Changes (%) between the historical and future damage estimates are also analyzed to highlight potential trends in windstorm impacts due to climate change.

All impact functions indicate an increase in windstorm damage for most GCMs, signaling that climate change could intensify windstorm impact in the future.

The stochastic vulnerability function produces higher increases in damage compared to the step function, particularly for models like TaiESM1, CESM2-WACCM, and BCC-CSM2-MR. This suggests that the stochastic approach may better capture potential tail risks.

The projections highlight variability between GCMs, with some models (e.g., NorESM2-MM) even predicting decreases in damage. This variability underscores the importance of using multiple climate models to account for uncertainties in windstorm projections under climate change.

The findings provide strong evidence that climate change will worsen windstorm damage in the future, although the magnitude varies depending on the model and impact function.

GCM	Historical Damage	Future Damage	Change (%)
BCC-CSM2-MR	1,031,067	1,251,006	21%
CESM2	1,300,487	1,421,123	9%
CESM2-WACCM	1,273,500	1,486,730	17%
NorESM2-MM	1,206,865	1,052,310	-13%
TaiESM1	1,135,140	1,683,455	48%
GFDL-ESM4	1,255,113	1,340,847	7%
CMCC-CM2-SR5	1,295,116	1,463,716	13%

Table 3 *Historical and future damage estimates using step vulnerability function*

GCM	Historical Damage	Future Damage	Change (%)
BCC-CSM2-MR	894,919	1,059,556	18%
CESM2	1,149,230	1,140,201	-1%
CESM2-WACCM	958,239	1,172,972	22%
NorESM2-MM	1,004,380	913,960	-9%
TaiESM1	950,163	1,315,353	38%
GFDL-ESM4	1,089,391	1,091,137	0%
CMCC-CM2-SR5	1,060,724	1,130,377	7%

Table 4 *Historical and future damage estimates using GLM*

GCM	Historical Damage	Future Damage	Change (%)
BCC-CSM2-MR	909,773	1,181,490	30%
CESM2	1,260,838	1,227,820	-3%
CESM2-WACCM	1,012,082	1,334,189	32%
NorESM2-MM	1,113,237	972,639	-13%
TaiESM1	1,016,107	1,492,165	47%
GFDL-ESM4	1,225,954	1,158,552	-5%
CMCC-CM2-SR5	1,169,111	1,238,152	6%

Table 5 *Historical and future damage estimates using stochastic vulnerability function*

3.2 Spatial maps

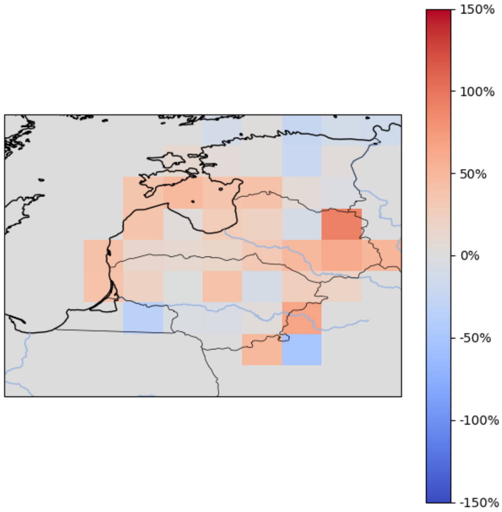
Spatial maps in Figure 9 illustrate the patterns of damage change under climate change scenarios. These maps, generated using different GCMs, depict the relative change in expected annual damage.

Expected annual damage is calculated at each exposure, then aggregated to grids (presented in spatial maps). Calculation is performed separately for each model and both scenarios: historical and future. Finally, for every grid cell, relative change in damage between future and historical expected annual damage is calculated:

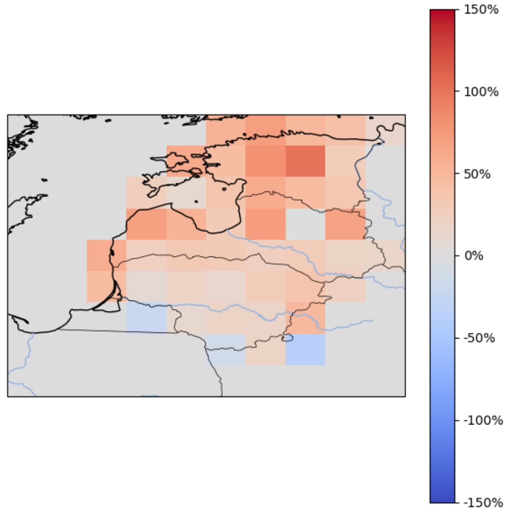
$$\Delta D\% = \frac{D_{\text{future}} - D_{\text{historical}}}{D_{\text{historical}}} \times 100.$$

Across all maps, there is a positive trend in the northern regions, with an increase in damage reaching up to 100% under future climate conditions. This increase is most pronounced in coastal areas.

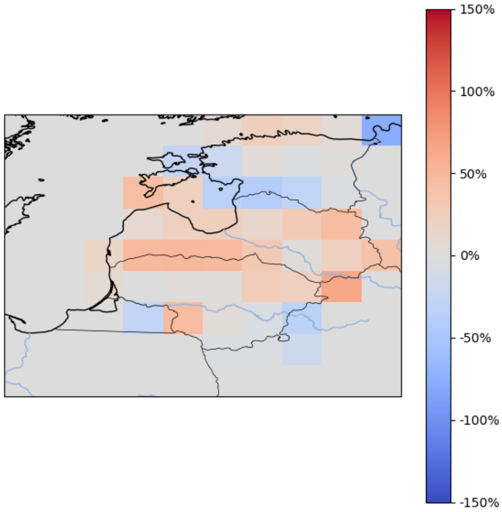
In contrast, the south regions show a trend of negative change, where annual damage is expected to decline. This spatial disparity highlights how climate change impact varies geographically, with northern and coastal areas near the Baltic Sea being more exposed to increasing risks compared to inland or southern regions.



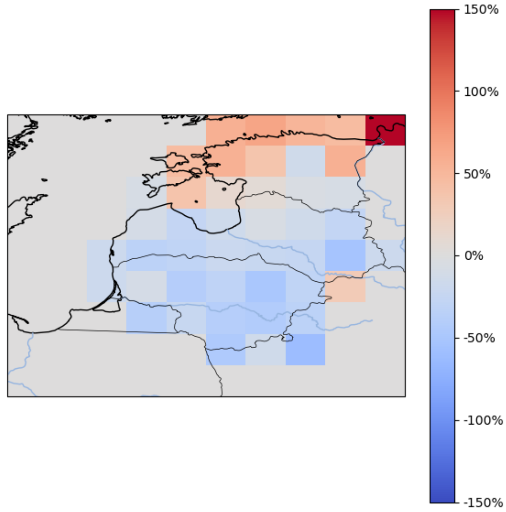
(a) *CESM2 GCM*



(b) *CESM2-WACCM GCM*



(c) *GFDL-ESM4 GCM*



(d) *NorESM2-MM GCM*

Figure 9 *Regional changes in expected annual damage (ΔD%)*

3.3 Exceedance probability curves

This section examines how damage events with long return periods are expected to change under future climate conditions. To achieve this, collection of models and simulations were analyzed, which provided insights into the uncertainty and variability inherent in climate projections.

EP curves separately for historical and future scenarios were created following these steps:

1. **Damage projection:** Damage projections were calculated using algorithm defined in 1.4. For EP curves only results from stochastic damage function were used.

2. **Sampling storm days:** For each GCM, $\frac{2}{3}$ of storm days were randomly selected, allowing repeats (sampling with replacement).

3. **Combining data:** Randomly sampled storm days from 7 GCMs were combined into single dataset.

4. **Calculating EP curve:** Exceedance probability for a single event loss was estimated as defined in 1.5.

5. **Bootstrapping for uncertainty:** 300 random selections of the EP curves were generated by repeatedly resampling the storm days using bootstrap resampling. This approach allowed for the generation of multiple EP curves, which were then used to estimate the median, 5th and 95th percentiles of the EP curve distribution.

This bootstrapping approach using a random subsampling allowed to estimate the combined effects of internal variability and climate model uncertainty on the projections.

Figure 10 presents the exceedance probability curves for both historical conditions and the future SSP5-8.5 scenario, highlighting the differences in damage frequency and intensity over time. Solid lines represent the medians and dashed lines the 5th and 95th percentiles of the EP curves distributions.

The average difference between the median values of the EP curves for the future and historical climate distributions indicates an increase in storm damage intensity under future climate conditions. This increase in intensity suggests, for instance, that damage with a return period of 50 years under the current climate could occur with a reduced return period of approximately 26 years under future climate conditions.

Although future projections exhibit narrower variability for longer return periods, they emphasize an increasing trend in damage amount.

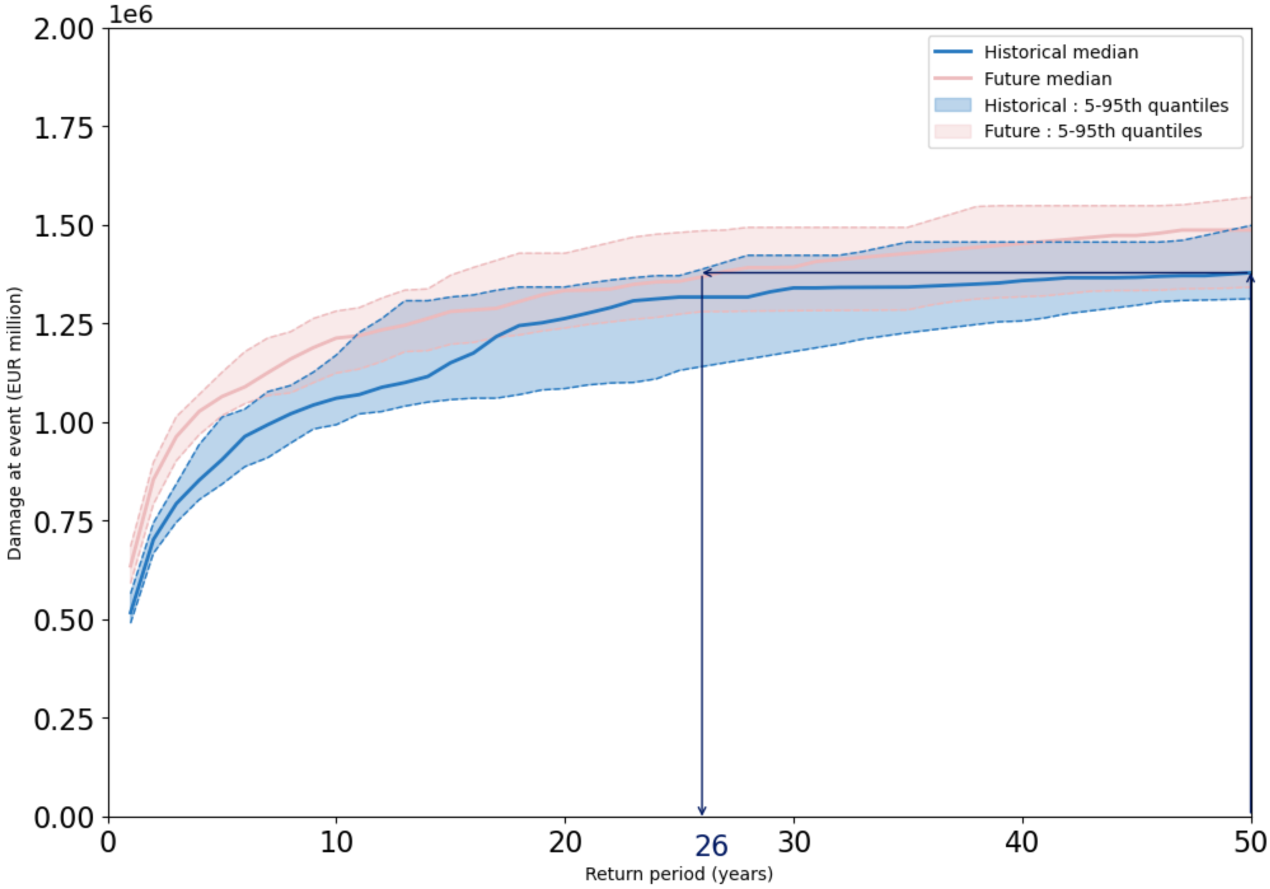


Figure 10 *Exceedance probability curves*

3.4 Occurrence exceedance probability curves

Occurrence exceedance probability curves were generated using the same steps as exceedance probability curves, but with one key difference: only the largest losses in each year were considered.

The exceedance probability curves (Figure 10) represent the probability of any event exceeding a given damage threshold within a year, capturing all possible events. In contrast, the occurrence exceedance probability curves (Figure 11) focus on the largest annual event, highlighting the most extreme losses expected per year. The future projections in both figures reveal an upward shift in damage due to climate change, but the gap between historical and future scenarios is more pronounced in the occurrence exceedance curve, emphasizing the greater impact of intensifying extreme events.

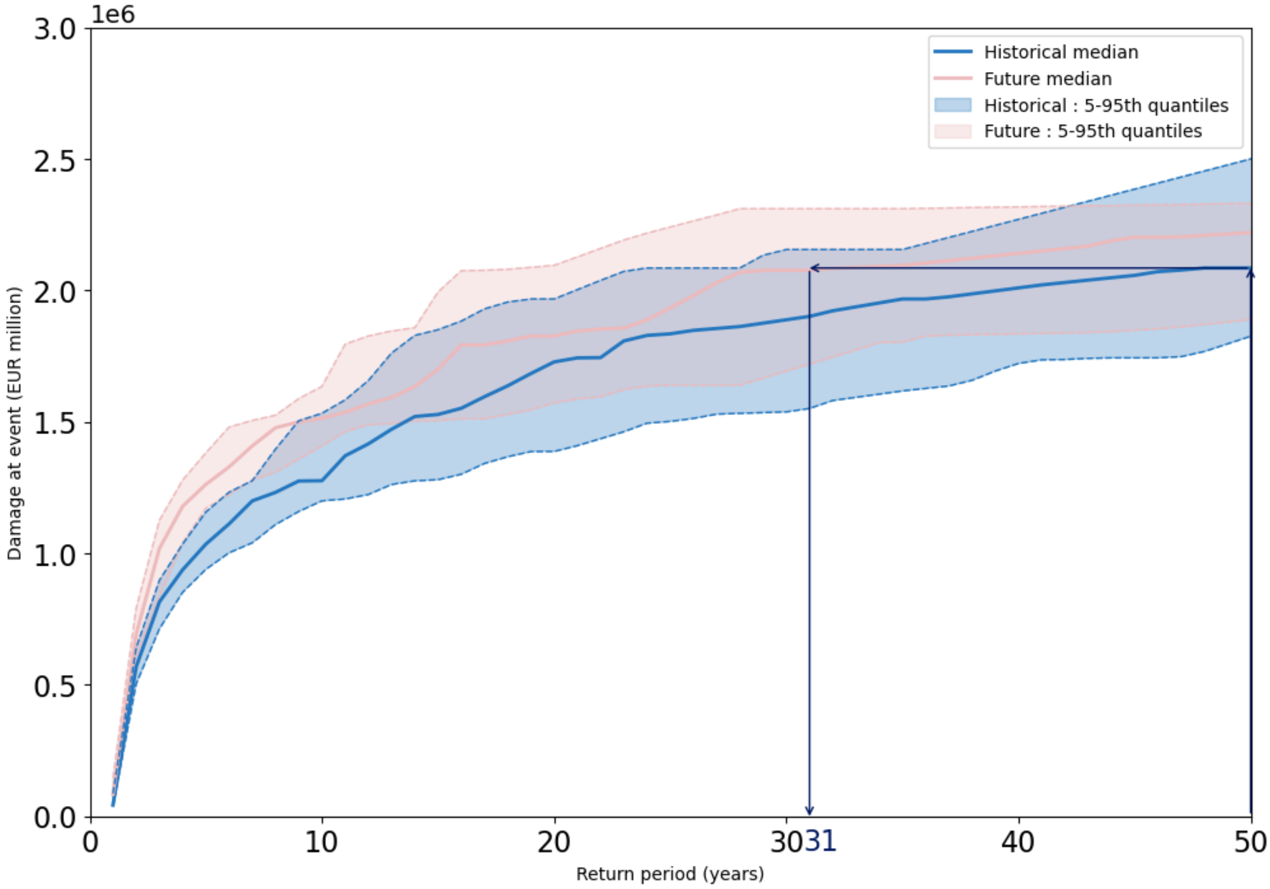


Figure 11 *Occurrence exceedance probability curves*

4 Results and conclusions

Damage projections indicate an increase in windstorm-related damage under future climate conditions, although variability exists across different global circulation models and vulnerability functions. Key results include:

- All vulnerability functions projected an increase in damage for most GCMs, with the stochastic function producing the largest increase, particularly for models such as TaiESM1 and CESM2-WACCM, where damage increases reached up to 48%.
- Variability among GCMs was observed, with some models (e.g., NorESM2-MM) predicting decreases in damage, emphasizing the importance of multi-model analyses to address uncertainties.

Spatial map analysis demonstrated notable geographical disparities in projected damage:

- Northern and coastal regions are expected to experience the most pronounced increase in damage, with some regions showing increase exceeding 100%.
- Conversely, southern regions exhibit little change or a potential decline in damage, reflecting varying regional responses to climate change.

The study employed exceedance probability and occurrence exceedance probability curves to quantify changes in damage likelihood and intensity:

- EP curves indicated a reduction in return periods for significant damage events.
- OEP curves highlighted an upward shift in the most extreme annual losses, highlighting the increasing financial risks posed by intensifying extreme events.

In conclusion, this study underscores the escalating risks of windstorm damage in the Baltic states due to climate change, providing insights to policymakers, insurers or researchers.

References and sources

[1] C. Banerjee, L. Bevere, J. Finucane, R. Lechner. “A Perfect Storm: Natural Catastrophes and Inflation in 2022.” In: *Swiss Re Institute - Sigma Research* (2023).

[2] *Climate Copernicus, What Is the Copernicus Climate Change Service’s ERA5 Reanalysis Dataset and What Can It Do for You?* URL: https://climate.copernicus.eu/what-copernicus-climate-change-services-era5-reanalysis-dataset-and-what-can-it-do-you (viewed 2024-12-23).

[3] *Climate Data Store, CMIP6 climate projections*. URL: https://cds.climate.copernicus.eu/datasets/projections-cmip6?tab=overview (viewed 2024-09-15).

[4] C. Dorland, R. S. J. Tol, J. P. Palutikof. “Vulnerability of the Netherlands and Northwest Europe to Storm Damage under Climate Change.” In: *Climatic Change* 43 (1999), pages 513–535. https://doi.org/10.1023/A:1005492126814.

[5] K. Emanuel. “Global Warming Effects on U.S. Hurricane Damage.” In: *Weather, Climate, and Society* 3 (2001), pages 261–268.

[6] *European Central Bank Data Portal*. URL: https://data.ecb.europa.eu/ (viewed 2024-09-10).

[7] European Insurance and Occupational Pensions Authority (EIOPA). “Methodological Paper on Potential Inclusion of Climate Change in the Nat Cat Standard Formula.” In: (2021).

[8] “Exceedance Probability in Catastrophe Modeling.” In: *Casualty Actuarial Society E-Forum, Winter 2021* (2020), pages 1–61.

[9] M. Klawa, U. Ulbrich. “A Model for the Estimation of Storm Losses and the Identification of Severe Winter Storms in Germany.” In: *Natural Hazards and Earth System Sciences* 3 (2003), pages 725–732. https://doi.org/10.5194/nhess-3-725-2003.

[10] T. Kruschke. “Winter Wind Storms: Identification, Verification of Decadal Predictions, and Regionalization.” In: (2014). URL: https://www.researchgate.net/publication/281206536_Winter_wind_storms_Identification_verification_of_decadal_predictions_and_regionalization.

[11] G. Louppe, M. Kumar. *Bayesian Optimization with Skopt*. URL: https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html (viewed 2024-11-12).

[12] C. Navarro-Racines, J. Tarapues-Montenegro, J. Ramírez-Villegas. “Bias-Correction in the CCAFS-Climate Portal: A Description of Methodologies.” In: *Decision and Policy Analysis (DAPA) Research Area. International Center for Tropical Agriculture (CIAT)* (2015).

[13] B. F. Prahl, D. Rybski, J. P. Kropp, O. Burghoff. “Performance of Storm Damage Functions: a Sectoral Impact Model Intercomparison.” In: *Natural Hazards and Earth System Sciences Discussions* 2 (2014), pages 5835–5887.

[14] B. F. Prahl, D. Rybski, J. P. Kropp, O. Burghoff, H. Held. “Applying Stochastic Small-Scale Damage Functions to German Winter Storms.” In: *Geophysical Research Letters* 39 (2012). https://doi.org/10.1029/2012GL050961.

[15] C. Schwierz, P. Köllner-Heck, E. Z. Mutter, D. N. Bresch, P. L. Vidale, M. Wild, C. Schär. “Modelling European Winter Wind Storm Losses in Current and Future Climate.” In: *Climatic Change* 101 (2010), pages 485–514. https://doi.org/10.1007/s10584-009-9712-1.

[16] L. G. Severino, C. M. Kropf, H. Afargan-Gerstman, C. Fairless, A. J. de Vries, D. I. V. Domeisen, D. N. Bresch. “Projections and Uncertainties of Winter windstorm Damage in Europe in a Changing Climate.” In: *Natural Hazards and Earth System Sciences* 24 (2024), pages 1555–1578.

[17] E. Soriano, L. Mediero, C. Garijo. “Selection of Bias Correction Methods to Assess the Impact of Climate Change on Flood Frequency Curves.” In: *Water* 11 (2019). https://doi.org/10.3390/w11112266.

[18] H. Zhang, S. Chapman, R. Trancoso, N. Toombs, J. Sykts. “Assessing the Impact of Bias Correction Approaches on Climate Extremes and the Climate Change Signal.” In: *Meteorological Applications* 31 (2024). https://doi.org/10.1002/met.2204.

5 Annex

The annex provides selected examples of Python scripts used for calculations.

Mapping exposures to centroids

```python
import numpy as np
from climada.util.coordinates import match_coordinates
import pandas as pd
from climada.entity import Exposures

location_data = Exposures.from_hdf5('Exposure/exposure_data.h5')
assigned = match_coordinates(
        np.stack([location_data.gdf['latitude'].values,
                  location_data.gdf['longitude'].values], axis=1),
        haz.centroids.coord, distance="euclidean", threshold=100,
    )
location_data.gdf['centroids'] = assigned
```

Stochastic affected exposure selection

```python
from climada.engine import Impact, ImpactCalc
import numpy as np
from scipy.sparse import csr_matrix

alpha = 0.9990
beta = 56.3576
gamma = 0.1921

class ImpactCalcStochastic(ImpactCalc):
    def impact_matrix(self, exp_values, cent_idx, impf):
        mat = super().impact_matrix(exp_values, cent_idx, impf)

        haz = self.hazard

        [date_count, centroid_count] = haz.intensity.shape
        exposure_count_per_centroid = self.
            _get_exposure_count_per_centroid(centroid_count, cent_idx)

        date_x_exposure_count_per_centroid = np.apply_along_axis(lambda
            row: exposure_count_per_centroid, axis=1, arr=haz.intensity.
            toarray())
```

```
        intensity_fv = np.clip(1 - alpha / (1 + np.exp(gamma * (haz.intensity.toarray() - beta))), 0, 1)
        intensity_binomial = np.random.binomial(n=date_x_exposure_count_per_centroid, p=intensity_fv, size=intensity_fv.shape)
        intensity_x_exposure_binomial_mask = np.zeros(mat.shape)
        exposure_indexes = np.arange(len(cent_idx))
        for bins, exps in zip(intensity_binomial, intensity_x_exposure_binomial_mask):
            for centroid, n in enumerate(bins):
                if n > 0:
                    choices = np.random.choice(exposure_indexes[cent_idx == centroid], n, replace = False)
                    exps[choices] = 1
        masked_mat = mat.toarray() * intensity_x_exposure_binomial_mask
        return csr_matrix(masked_mat)

    def _get_exposure_count_per_centroid(self, centroid_count, cent_idx):
        [unique_values, counts] = np.unique(cent_idx, return_counts=True)
        output = np.zeros(centroid_count)
        output[unique_values] = counts
        return output.astype(int)
```

Storm day selection

```
import xarray as xr

def detect_storm_days(ds: xr.Dataset, qt_ds: xr.Dataset):

    qt99 = qt_ds["sfcWind"].quantile(0.99, dim='time')

    wind = ds["sfcWind"]
    wind_mask = wind.where(wind>qt99)
    wind_mask = wind_mask.where(wind_mask>=22)

    min_stormy_grid_points_per_storm_day = 15
    wind_mask = wind_mask.dropna(dim="time",thresh=min_stormy_grid_points_per_storm_day)

    wind_mask = wind_mask.unstack().fillna(0)

    return wind_mask
```

GLM

```python
import pandas as pd
import statsmodels.api as sm
import statsmodels.formula.api as smf

claims_data = pd.read_csv('claims_with_fg10.csv')

claims_data['built_group'] = claims_data['built_group'].replace(
    {'1940s 1960s': 'Before 1970s', 'Pre 1940s': 'Before 1970s'}
)

# gamma GLM with inverse link function
model = smf.glm(
    formula="MDD ~ fg10 + C(built_group) + C(building_group)",
    data=claims_data,
    family=sm.families.Gamma(link=sm.families.links.inverse_power())
)

result = model.fit()

print(result.summary())
```

Bias correction

```python
import xarray as xr

def bias_correct_at_centroid(dat_mod, dat_obs, dat_mod_all, minq=0.001,
   maxq=1.000, incq=0.001):
    assert len(dat_mod.lat) == len(dat_mod_all.lat), f"{len(dat_mod.lat)}
        != {len(dat_mod_all.lat)}"
    assert len(dat_mod.lon) == len(dat_mod_all.lon), f"{len(dat_mod.lon)}
        != {len(dat_mod_all.lon)}"

    q = np.arange(minq, maxq, incq)
    mod_vals = dat_mod.values.copy()
    obs_vals = dat_obs.values.copy()
    mod_all_vals = dat_mod_all.values.copy()

    mod_lat = dat_mod.lat.values
    mod_lon = dat_mod.lon.values

    lat_space = mod_lat[1] - mod_lat[0]
```

```
    lon_space = mod_lon[1] - mod_lon[0]

    obs_lat = dat_obs.latitude.values
    obs_lon = dat_obs.longitude.values

    for lat_idx in range(len(dat_mod.lat)):
        lat_min = mod_lat[lat_idx] - lat_space / 2
        lat_max = mod_lat[lat_idx] + lat_space / 2

        for lon_idx in range(len(dat_mod.lon)):
            lon_min = mod_lon[lon_idx] - lon_space / 2
            lon_max = mod_lon[lon_idx] + lon_space / 2

            obs_lat_idx = np.where((obs_lat > lat_min) & (obs_lat < lat_max))
            obs_lon_idx = np.where((obs_lon > lon_min) & (obs_lon < lon_max))

            obs_vals_at_centr = obs_vals[:,:,obs_lon_idx][:,obs_lat_idx,:].flatten()
            mod_vals_at_centr = mod_vals[:,lat_idx,lon_idx]
            mod_all_vals_at_centr = mod_all_vals[:,lat_idx,lon_idx]

            cdf_obs = _calc_cdf_xr(obs_vals_at_centr, q)
            cdf_mod = _calc_cdf_xr(mod_vals_at_centr, q)

            assert cdf_mod.shape == cdf_obs.shape

            cdf_dif = cdf_mod - cdf_obs
            dat_mod_all_corrected = _map_quantile(mod_all_vals_at_centr, cdf_mod, cdf_dif, q)
            mod_all_vals[:,lat_idx,lon_idx] = dat_mod_all_corrected

    return xr.DataArray(mod_all_vals, dims=dat_mod_all.dims, coords=dat_mod_all.coords, name='sfcWind')

def _map_quantile(mod_all_flat_values, cdf_mod, cdf_dif, q):
    perc_mod = np.interp(mod_all_flat_values, cdf_mod, q)
    cor_term = np.interp(perc_mod, q, cdf_dif)
    dat_mod_adj = mod_all_flat_values - cor_term

    return dat_mod_adj

```

```python
def _calc_cdf_xr(data, q):
    dat_sorted = np.sort(data)
    p = 1. * np.arange(len(dat_sorted)) / (len(dat_sorted) - 1)

    cdf = np.interp(q, p, dat_sorted)

    return cdf
```