Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **72**, 809–828
https://doi.org/10.1093/jrsssc/qlad043
Advance access publication 23 June 2023
Discussion Paper

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

Assessing present and future risk of water damage using building attributes, meteorology, and topography*

Claudio Heinrich-Mertsching[1], Jens Christian Wahl[1], Alba Ordoñez[1], Marita Stien[2], John Elvsborg[2], Ola Haug[1] and Thordis L. Thorarinsdottir[1]

[1]Norwegian Computing Center, Oslo, Norway
[2]Gjensidige Forsikring ASA, Oslo, Norway

Address for correspondence: Claudio Heinrich-Mertsching, Norwegian Computing Center, Norsk Regnesentral, Postboks 114, Blindern, NO-0314 Oslo, Norway. Email: claudio@nr.no

*Read before The Royal Statistical Society at the first meeting on 'Statistical aspects of climate change' held at the Society's 2022 annual conference in Aberdeen on Wednesday, 14 September 2022, the President, Professor Sylvia Richardson, in the Chair.

Abstract

Weather-related risk makes the insurance industry inevitably concerned with climate and climate change. Buildings hit by pluvial flooding is a key manifestation of this risk, giving rise to compensations for the induced physical damages and business interruptions. In this work, we establish a nationwide, building-specific risk score for water damage associated with pluvial flooding in Norway. We fit a generalised additive model that relates the number of water damages to a wide range of explanatory variables that can be categorised into building attributes, climatological variables, and topographical characteristics. The model assigns a risk score to every location in Norway, based on local topography and climate, which is not only useful for insurance companies but also for city planning. Combining our model with an ensemble of climate projections allows us to project the (spatially varying) impacts of climate change on the risk of pluvial flooding towards the middle and end of the 21st century.

Keywords: climate change risk, generalised additive model, non-life insurance, pluvial flooding, topographical risk

1 Introduction

A recent report by the World Meteorological Organization found that floods were the most common of weather-, climate-, and water-related disaster types recorded in the period 1970–2019 (Douris et al., 2021). While single events of large fluvial (river) floods can cause damages worth billions of Euros (Barredo, 2007), a large proportion of overall flood damages is caused by pluvial flooding—surface water flooding resulting from heavy rainfall—due to the far greater reach of these events (Houston et al., 2011; Spekkers et al., 2011). For instance, Houston et al. (2011) assess that around 2 million people in the UK are at risk from pluvial flooding. Pluvial floods are commonly considered an invisible hazard, as they can strike with little warning in areas with no recent record of flooding (Netzel et al., 2021), and the risk of pluvial flooding may increase in the future due to a combination of climate change, urbanisation and lack of investment in sewer and drainage infrastructure (Skougaard Kaspersen et al., 2017).

A building's exposure to pluvial flood risk depends on a range of factors such as the building's attributes, local weather, and topography. The extent to which these factors influence the risk exposure is commonly assessed based on insurance claims data on reported flood damages, see Gradeci et al. (2019) for a systematic review of the use of insurance claims data in analysing pluvial flood events. A critical challenge when assessing flood impact is the lack of good quality flood impact data (Hammond et al., 2015). One specific challenge is that insurance claims data may not

Received: October 29, 2021. **Revised:** May 31, 2022. **Accepted:** July 5, 2022

separate between fluvial and pluvial flood damage (Bernet et al., 2017). In Norway, however, fluvial flood damage is covered by a compulsory natural perils insurance linked to fire insurance and managed by the Norwegian Natural Perils Pool, while pluvial flood and other rainfall-induced damage is covered by a private insurance and managed directly by the primary insurer. In the following, we do not separate between pluvial flood and other rainfall-induced damage, and, for simplicity, we refer to these as water damages.

The study of water damage and its relationship to meteorological, hydrological and topographical variables is nicely summarised in Lyubchich et al. (2019), and Gradeci et al. (2019). A commonality among these studies is that the number of claims and the claim size is aggregated in space [see Table 6 in Gradeci et al. (2019) and Table 3 in Lyubchich et al. (2019)], for example at the level of municipality or postal code. Many papers also model daily claim frequency or severity and use meteorological and hydrological variables that are associated with the specific daily event, see for example Spekkers et al. (2014) and Haug et al. (2011). For assessing the risk of specific buildings, or risk on an annual basis (for example, for pricing) there are two drawbacks of using a daily model on spatially aggregated data. Firstly, spatially aggregated data disguise building-specific information, which makes it hard to assess the risk of a specific building at a specific location. Variables such as topographical indices that are available at a high spatial resolution will also lose their accuracy and thus potentially explanatory power if spatially aggregated. A second challenge of a daily model that uses daily weather variables is that predicting future claims beyond approximately two weeks is challenging due to the high uncertainty of possible weather outcomes and a lack of skilful long-range weather predictions for this time resolution (van Straaten et al., 2020).

The goal of this study is twofold. Our first goal is to provide an estimation of current, or near-term, water damage risk for any building or potential building site in Norway. Our second goal is to project potential changes in water damage risk in a future climate. To this aim, we employ detailed topographical information at a 20×20 m resolution over Norway and a more general quarterly summary of local weather statistics as detailed information of future weather is unlikely to be robustly projected for a future climate. Insurance data from the insurance company Gjensidige, including information on building attributes—to account for building-specific risk—and reported water damages are available for 729,031 unique locations in Norway within the time period 2009–2021. These data cover approximately a quarter of the national market. The output of our analysis is the risk of water damages for a property located anywhere in Norway. Here, we compare the use of a generalised additive model with either a Poisson or a negative binomial likelihood. The parametric structure of the models is such that the individual risk components related to topography, weather, and building characteristics can be extracted and assessed individually or combined.

2 Data

Our model incorporates several different data sets. Claims data from the insurance company Gjensidige are combined with topography data derived from a digital elevation model (DEM) and historical meteorological data provided by the Norwegian Meteorological Institute. Moreover, regional climate projections provided by the EURO-CORDEX initiative are used to project claim frequency for future climate scenarios. A brief description of these data sets follows.

2.1 Insurance data

The insurance data were provided by the Norwegian insurance company Gjensidige and contain customer information from 1 January 2009 to 23 April 2021. The data set contains insurance information for private houses, apartments, cabins, agricultural, and industrial buildings, as well as apartment complexes with a single coverage for the entire complex, located in Norway. For each insured property, we have information on the exact location, the insurance period, the value of the property, whether the building is used as a rental property and building characteristics such as size, age, type of roof and whether the building has a basement. In our analysis every building with its unique combination of characteristics constitute one observation. Modification of any property attribute, e.g. its size, over the building's insurance period gives rise to separate observations with associated coverage lengths. The data set includes 32,534 water damage claims, corrsponding to an average of 0.007 claims per insured year. Overall, the buildings in the data set have between 0 and 15 claims. Out of 1, 740, 915 observations, 1, 712, 157 (99.5%) contain zero claims.

2.2 Topography data

Topographical information is obtained from the Norwegian Mapping Authority's DEM corresponding to the bare-Earth surface where all natural and built features are removed. The DEM is generated from data primarily collected via airborne laser-scanning equipment and organises 10 m gridded elevation data into non-overlapping 50 km square blocks. Rather than constructing one huge national data set, we assemble blocks into regional rectangles of manageable size, where each region typically covers 1–3 counties (there are a total of 11 counties in Norway). The rectangular area is chosen large enough so that it encompasses the full hydrological catchment area for every single location within any of the counties that it represents. To further facilitate processing of the regional DEM data, a coarser spatial resolution of 20 m is established by averaging the elevation of the four 10 m cells underlying every 20 m cell.

From the 20 m gridded and regionalised DEM data, three topographical indices of particular interest to precipitation-induced damage are calculated: *Height Above Nearest Drainage* (HAND) (Nobre et al., 2011), which specifies for each grid cell in the DEM the relative elevation between the cell and the nearest waterway cell (for example, river or ocean) that it drains into. One interpretation of the HAND index is that it judges the risk of a DEM cell being affected if its associated waterway overruns its banks. The second index is the *slope*, β, of a DEM cell which specifies the local angle of inclination in the water flow direction out of the cell and measures the capability that a grid cell has to drain water away. A similar concept that also takes into account the amount of water potentially flowing into the DEM cell is the *Topographical Wetness Index* (TWI) developed by Beven and Kirkby (1979). This is defined as $\text{TWI} = \log(a/\tan(\beta))$, where a is the size of the upslope contributing area and β is the slope. Grid cells with a high TWI therefore indicate locally flat terrain or a large upslope area, both of which increase the likelihood of accumulating water.

2.3 Meteorological data

Historical meteorological information is derived from the gridded data product seNorge version 2018 (Lussana, 2020). Using statistical interpolation of station observations, seNorge contains estimates of daily near-surface air temperature and precipitation on a grid with resolution 1 km covering all of Norway from 1957 to the present day. The seNorge data set is used to derive climatological indices on the same resolution, defined as quarterly means of daily temperature and precipitation for the time period 1991–2020. In addition to these climatological indices, several alternative climatological indices were considered in an explorative stage of our research. This included high empirical quantiles of the quarterly distributions of daily temperature, daily precipitation, and multi-day precipitation as well as high return period values from intensity–duration–frequency curves for daily precipitation. However, as these did not improve the predictive performance of our models and yielded less interpretable models, we have chosen to focus on the quarterly means of precipitation and temperature, which are more robust climatological indices.

2.4 Climate projections

Projections of the climatological indices for a future climate are provided by the EURO-CORDEX initiative (Jacob et al., 2020). EURO-CORDEX provides a multi-model ensemble of regional climate projections for Europe at a spatial resolution of 12 km, obtained by running a limited-area regional climate model (RCM) using the output of a global general circulation model (GCM) as boundary conditions. In addition, the RCM output has been bias-corrected using a cumulative distribution function transformation (Michelangeli et al., 2009; Vrac et al., 2012) or distribution-based scaling (Yang et al., 2010), using data from the regional reanalysis MESAN (Häggmark et al., 2000; Landelius et al., 2016) or the observation-based gridded data product E-OBS (Cornes et al., 2018) as calibration data.

Specifically, we consider 12 different combinations of GCMs, RCMs, and bias-correction approaches, as listed in Table 1. This multi-model ensemble of climate projections includes two different GCMs (EC-EARTH Hazeleger et al., 2012 and MPI-ESM-LR Giorgetta et al., 2013), four different RCMs (CCLM4-8-17, HIRHAM5, RACMO22E, and RCA4, see Jacob et al., 2014 for

Table 1. Overview of the EURO-CORDEX climate projection ensemble used in this paper

	GCM	RCM	Bias-correction method
1	EC-EARTH	CCLM4-8-17	CDFT22s-MESAN-1989–2005
2	EC-EARTH	CCLM4-8-17	DBS45-MESAN-1989–2010
3	EC-EARTH	HIRHAM5	CDFT22s-MESAN-1989–2005
4	EC-EARTH	HIRHAM5	DBS45-MESAN-1989–2010
5	EC-EARTH	HIRHAM5	CDFt-EOBS10-1971–2005
6	EC-EARTH	RACMO22E	CDFT22s-MESAN-1989–2005
7	EC-EARTH	RACMO22E	CDFt-EOBS10-1971–2005
8	EC-EARTH	RCA4	DBS45-MESAN-1989–2010
9	MPI-ESM-LR	CCLM4-8-17	CDFT22s-MESAN-1989–2005
10	MPI-ESM-LR	CCLM4-8-17	DBS45-MESAN-1989–2010
11	MPI-ESM-LR	RCA4	CDFT22s-MESAN-1989–2005
12	MPI-ESM-LR	RCA4	DBS45-MESAN-1989–2010

Note. For each ensemble member, the general circulation model (GCM), the regional climate model (RCM), and the bias-correction method (combination of method, data product, and data period) is listed.

details), as well as three different versions of the bias-correction approaches described above. Considering such a variety of combinations helps us better account for the uncertainties associated with each layer of modelling. Additionally, we consider two representative carbon pathways (RCPs): RCP 4.5 is an intermediate scenario where emissions peak around 2040 and decline afterwards, while RCP 8.5 presents a worst-case scenario where emissions continue to rise throughout the entire 21st century (Jacob et al., 2014).

An ensemble of future climatological indices for two future periods, 2031–2060 and 2071–2100, are obtained by calculating the projected differences of the climatological index between the future period and the historical period 1991–2020 from each climate model. These projected differences are then added to the historical index, derived from the seNorge data. Considering differences rather than absolute projections removes potential (constant) biases of the climate models. The projected future indices are therefore derived on the high-resolution spatial scale of 1 km provided by seNorge, but for all grid cells lying within the same 12×12 km RCM grid cell the same changes apply.

3 Methods

3.1 Statistical modelling framework

The aim of the statistical modelling is to predict the number of claims, N_i, for contract i. To this aim, we model the distribution of N_i as a function of various covariates and use the length of the contract, l_i, as offset. Our covariates can be grouped in four separate classes: (1) property-specific characteristics, (2) topographical information at the property location, (3) climatological information at the property location, and (4) a fixed effect of which county the property is located in and a random effect over municipalities to increase the flexibility of the model. This way, municipalities with little information get regularised towards the mean of the county, which is important since many municipalities contain only few observations. Currently, there are 11 counties and 356 municipalities in Norway.

For the climatological indices, we compare annual models based on annual averages of precipitation and surface temperature to quarterly models based on quarterly averages. For the latter, each contract i is split into (up to) four sub-contracts. For example, a contract from 1 January 2019 to 20 January 2020 would be split into four contracts, accounting for the respective overlaps: a contract of 3 months and 20 days in quarter 1 (Q1), and three contracts of 3 months each in Q2–Q4. The claims of the original contract are then assigned to the sub-contracts of the appropriate quarter. This increases the temporal resolution of the climate statistics, enabling the model to pick up on

quarterly differences in claims connected to seasonal differences in weather. Note that we estimate a single model over all quarters. That is, we assume the same effect of the climate statistics throughout the year, only their value may vary by season.

For the distribution of N_i, we compare a Poisson and a negative binomial model, denoted by $N_i \sim \text{Po}(\mu_i)$ and $N_i \sim \text{NB}(\mu_i, \theta)$, respectively. For both distributions, μ_i is the expected number of claims for the ith contract. While the distribution of the Poisson model is fully determined by its mean, the negative binomial model has an additional parameter θ, specified by $\text{Var}(N_i) = \mu_i + \mu_i^2/\theta$. In particular, the negative binomial model exhibits larger variance than the Poisson model, and the parameter θ controls for overdispersion.

For both models we employ a generalised additive model (GAM; Hastie & Tibshirani, 1990). This allows for non-linear relationships between the covariates and the number of claims without needing to specify the functional form of these relationships, while still being highly interpretable. The interpretability aspect is crucial for understanding and explaining the risk structure to both potential insurance holders and decision-makers in the context of climate change adaptation. We use a logarithmic link function, such that the relationship between the expected number of claims and the available covariates is specified by

$$\log \mu_i = z_i^T \gamma + \sum_{j=1}^{J} f_j(x_{ij}) + u_{R[i]} + \log(l_i) + \log(v_i). \tag{1}$$

Here, z_i^T represents the vector of categorical variables for contract i with parameter vector γ and f_j represents a smooth function modelling the effect of the jth continuous covariate x_{ij}. The variable $u_{R[i]}$ is a random effect assigned to each policy i belonging to municipality $R[i] \in \{1, \ldots, K\}$ with $K = 355$ in the current setting as one municipality has no weather data available. The variables l_i and v_i represent the length of the ith contract and the value of the building, respectively, which are used as offsets in our model. The final model contains five discrete and $J = 7$ continuous covariates. The discrete variables include a fixed effect for the county and are otherwise property specific, e.g. whether a property has a basement and the quality of the building. The continuous variables include climatological, topographical, and property-specific covariates, see Section 4.1 for details.

We assume that the effects, $u_1, \ldots, u_K$, for all municipalities are independent and normally distributed with a common variance σ_u^2, that is, $u_k \overset{\text{iid}}{\sim} N(0, \sigma_u^2)$. Normally distributed random effects fit nicely into the GAM framework, as they can be expressed as a smooth spline with a penalty matrix equal the identity matrix. They can therefore be estimated simultaneously with the smooth splines in equation (1) (Wood, 2017). The offset $\log(l_i)$ accounts for the increase in expected claims over time as long as the contract is active. The use of the property value, v_i, as offset is convenient in insurance risk modelling, where it is more natural to model the number of claims per time and insured value, which is more closely connected to the expected payout. Since it would be unreasonable to assume μ_i to increase linearly in v_i, we additionally consider v_i as a property-specific covariate.

Each f_j is represented as a weighted sum of basis functions,

$$f_j(x_{ij}) = \sum_{k=1}^{K} \beta_{jk} b_{jk}(x_{ij}),$$

where the β_{jk} are unknown parameters to be estimated from the data, and $b_{j1}, \ldots, b_{jK}$ are known basis functions, here given by the truncated thin plate regression basis (Wood, 2003). It should be noted that, specifically for precipitation, a natural alternative to the truncated thin plate basis would be a spline basis that enforces monotonicity, see Meyer (2018). This would ensure that the risk of water damage increases in the precipitation amount. This methodology has been investigated using the package cgam (Liao & Meyer, 2019), but is too computationally demanding given the size of our data and does, moreover, not support the negative binomial distribution.

3.2 Risk assessment

We define the *risk* of the ith contract as $r_i := \frac{\mu_i}{l_i v_i}$, that is, the expected number of claims per year and value of insured property. The structure of the GAM facilitates that the risk r_i can be decomposed into the product of partial risks corresponding to different groups of covariates. For example, the non-standardised *climatological* risk for the ith contract is defined as

$$\log \tilde{r}_i^{\text{clim}} := \sum_{j \in \text{clim}} f_j(x_{ij}),$$

where the sum runs over the climatological covariates only (i.e. average precipitation and surface temperature). The partial climatological risk is then obtained by standardisation

$$\tilde{r}_i^{\text{clim}} := \frac{\tilde{r}_i^{\text{clim}}}{\frac{1}{N}\sum_{l=1}^{N} \tilde{r}_l^{\text{clim}}}, \tag{2}$$

where the average in the denominator is taken with respect to all contracts. This standardisation achieves that the average climatological risk equals 1, making it easy to see whether a given contract has above- or below-average risk due to local climatology. Similarly, we can define the partial *topographical* risk r_i^{topo} and the partial *building-specific* risk $r_i^{\text{b-s}}$, by restricting the right-hand side of equation (1) to only the summands corresponding to covariates of these groups. The covariates not belonging to either of these three groups are the county fixed effect and the municipiality random effect, which we group into the partial *regional* risk r_i^{reg}. This yields a complete decomposition of the total risk of the ith contract as

$$r_i = r_0 r_i^{\text{clim}} r_i^{\text{topo}} r_i^{\text{b-s}} r_i^{\text{reg}}, \quad \text{where} \quad r_0 := \left(\frac{1}{N}\sum_{l=1}^{N} \tilde{r}_l^{\text{clim}}\right) \dots \left(\frac{1}{N}\sum_{l=1}^{N} \tilde{r}_l^{\text{reg}}\right).$$

This decomposition facilitates interpretation of the results. The factor r_0 is the same for all contracts and, therefore, represents base risk. Each of the four other factors averages to 1, making it easy to identify prevailing risk drivers (or risk reductors) for a specific property. The decomposition also allows visualisation of spatially varying risk, by plotting maps of the partial risks, see Section 4.2. Note that, after fitting a model, these risk factors can be calculated for all locations for which the corresponding covariate information is available yielding national risk maps.

3.3 Inference

In order to avoid overfitting, we employ a penalised log-likelihood given by

$$l(\boldsymbol{\phi}) - \frac{1}{2}\sum_{j=1}^{J+1} \lambda_j \boldsymbol{\beta}_j^T \boldsymbol{S}_j \boldsymbol{\beta}_j, \tag{3}$$

where $l(\boldsymbol{\phi})$ is the log-likelihood, $\boldsymbol{\phi} = (\boldsymbol{\gamma}, \boldsymbol{\beta})$ are the model parameters, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_J, \lambda_u)$ are the smoothing parameters controlling the smoothness of the f_js and the precision $(1/\sigma_u)$ of the random effect, and $\boldsymbol{S}_j$ is a matrix where the klth element equals $\int b''_{jk}(x) b''_{jl}(x)\, \mathrm{d}x$. The model parameters and smoothing parameters are estimated in two steps. First, equation (3) is optimised w.r.t $\boldsymbol{\phi}$ holding $\boldsymbol{\lambda}$ fixed using penalised iteratively reweighted least square (PIRLS). Second, the smoothing parameters are estimated using Laplace approximated restricted maximum likelihood, holding $\boldsymbol{\phi}$ fixed. For the negative binomial model, the dispersion parameter θ is estimated alongside the smoothing parameters.

All computations are performed with R version 4.0.5 (R Core Team, 2021) using the package `mgcv` version 1.8-26 (Wood, 2011) for parameter estimation of the GAMs. As our data set contains millions of observations and hundreds of covariates (when all levels are one-hot encoded), we

use the methodology proposed for large data sets in Wood et al. (2017), Wood et al. (2014), and Li and Wood (2020), as implemented in the `bam` function in `mgcv`.

3.4 Model evaluation

We evaluate and compare competing models using two proper scoring rules (Gneiting & Raftery, 2007), the mean square error (MSE) and the Brier score (Brier, 1950), where a smaller value equals a better performance.

The MSE is defined as

$$\text{MSE} := \frac{1}{N}\sum_{i=1}^{N}\frac{1}{l_i}(N_i - \hat{\mu}_i)^2,$$

where N_i is the observed number of claims for the ith contract, l_i is the length of the contract (in years), and $\hat{\mu}_i$ is the predicted expected number of claims for N_i. The scaling factor $\frac{1}{l_i}$ counteracts the fact that the variance of N_i grows with the length of the contract. It can be shown that under this scaling, the rank of competing models is equal for different time units in the contract length.

The distribution of the claims data is heavily skewed. While for the vast majority of contracts we have $N_i = 0$, the data also contains many contracts with two or more claims, some contracts even reaching more than 10 claims. The MSE is sensitive to such outliers (Thorarinsdottir & Schuhen, 2018), and an evaluation using the MSE therefore puts specific emphasis on the predictive performance for these outliers. To complement the picture, we therefore consider the Brier score as a second performance metric. Specifically, we consider the mean Brier score for the event of observing at least one claim,

$$\text{BS} := \frac{1}{N}\sum_{i=1}^{N}(\mathbb{1}\{N_i \geq 1\} - \hat{p}_i)^2,$$

where $\mathbb{1}$ denotes the indicator function and $\hat{p}_i$ is the predicted probability that the ith contract files at least one claim. This metric is insensitive to outliers and shifts the focus to predicting whether a given house will have a claim or not.

3.5 Model selection methodology

For assessing out-of-sample predictive performance, we perform a 10-fold cross-validation where 10% of the contracts are removed at random during model fitting, and the fitted model is used to predict the claims for the withheld 10%. This process is repeated 10 times with each contract left out exactly once during the model training.

We compare using either a Poisson or a negative binomial target distribution for the regression model as well as four alternative configurations of the regression model in equation (1). The simplest baseline model includes only building-specific covariates. Additionally, we consider a model including topographical indices (TWI, slope, and HAND), a model including climatological indices (mean temperature and mean precipitation) and a model including both topographical and climatological indices.

For the climatological information, we consider both a quarterly and an annual model as discussed in Section 3.1 above. The splitting into quarterly sub-contracts results in a different set of contracts than for an annual model, which makes it difficult to directly compare the MSE and the Brier score for these models. We address this by reversing the quarterly split before evaluation. Specifically, for the quarterly model, the ith contract is split into (up to) four contracts $i_1, \ldots, i_4$ during fitting, with corresponding numbers of claims $N_{i_1}, \ldots, N_{i_4}$. After fitting the model, we obtain four different predicted means $\hat{\mu}_{i_1}, \ldots, \hat{\mu}_{i_4}$. For evaluation, we recompute $\hat{\mu}_i = \hat{\mu}_{i_1} + \ldots + \hat{\mu}_{i_4}$ and consider this quantity a prediction for N_i. Similarly, for the Brier score we obtain four probabilities $\hat{p}_{i_1}, \ldots, \hat{p}_{i_4}$ and can retrieve $\hat{p}_i$ as $\hat{p}_i = 1 - \prod_{j=1}^{4}(1 - \hat{p}_{i_j})$, assuming conditional independence of the number of claims across different quarters.

Under the Poisson model, the sum of the four quarterly predictions is again Poisson distributed. The only difference is that the quarterly model resolves seasonal variation in the climatological indices. However, for the negative binomial model, a change in model resolution yields a different model distribution (Diggle & Milne, 1983), so that the annualised quarterly model is not directly comparable to the annual model.

3.6 Model diagnostic

After selecting a model as described above, we assess the fit of the final model using conditional expectation diagrams. Such diagrams are obtained by sorting the prediction–observation pairs $(N_i, \mu_i)_{i=1,\ldots,n}$ by their predicted expectation μ_i, and thereafter dividing into k sub-groups of equal size. The conditional expectation diagram then plots a single point for each group, its x-coordinate being the mean predicted expectation, or mean μ_i, for the group and its y-coordinate being the mean conditional expectation, or mean N_i, for the group. Since the predictions were sorted before grouping, the predicted expectation gradually increases in the group index from 1 to k. For example, the point with the highest x-value corresponds to the group of n/k highest predicted risks. For a perfectly calibrated prediction model, the k points are located on or near the diagonal. Systematic deviations from the diagonal reveal conditional biases of the predictive model, see Section 4.1 for details. These diagrams constitute a straightforward generalisation of reliability diagrams which are widely used as a diagnostic tool for prediction of categorical variables, see Wilks (2011).

4 Results

Here, we present the results of our analysis of Norwegian water damage data. In a cross-validation study, we compare risk assessment models with different sets of covariates at two temporal resolutions and under two different distributional assumptions. We then present estimates of annual water damage risk under current and future climate.

4.1 Model selection

Figure 1 shows the out-of-sample predictive performance of 16 different risk assessment models based on a cross-validation study, where the models differ in distributional assumptions, covariate selection, and temporal resolution. A baseline model that only includes property-specific information is outperformed by models that include topographical and/or climatological variables. Further, as expected, including topographical information with high spatial precision gives a greater improvement in performance than including more general climatological information. Concerning model distribution, we see clear improvement in performance for the negative binomial model if the estimation is performed at a quarterly temporal resolution, even for the baseline model where all the covariates are identical. In the negative binomial model, the mean and the θ parameter are estimated jointly, and there is a non-linear relationship between the mean and the variance. If the estimation is performed at a quarterly time resolution, the variance of the resulting annualised model tends to be smaller than that of the original annual model. This, in turn, affects the mean estimate and leads to slightly larger annual mean estimates, in particular for contracts with large expected number of claims. Consequently, the predictive errors are smaller for these contracts yielding an overall smaller MSE.

While proper scoring rules favour the true data generating process by construction, they may evaluate different aspects of competing models and thus not always yield identical rankings when none of the models represent the true data generating process. The best performance of each distributional model is obtained when both topographical and climatological information is included in the model. However, the model rankings for this setting are not consistent across the two types of evaluation in Figure 1. The quarterly Poisson model ranks first under MSE and third under the Brier score, while the quarterly negative binomial model ranks second under MSE and first under the Brier score. As the overall highest ranked model, we continue our analysis with the negative binomial model estimated on a quarterly basis with both topographical and climatological covariates.

Risk models for the insurance industry are often perceived as not performing particularly well at predicting high numbers of claims, see Scheel et al. (2013) for an example. In our case, out of the 6 million contracts, 99.5% have zero claims and 138 contracts have between 3 and 7 claims. The mean predicted number of claims for these extreme observations is 0.192, due to many

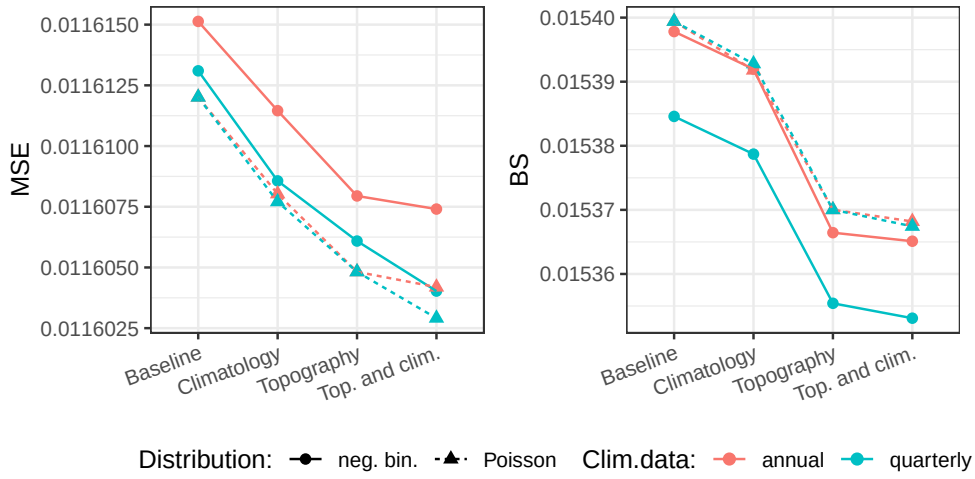

Figure 1. Comparison of models for water damage risk in Norway under a Poisson and a negative binomial likelihood with property-specific covariates only (baseline), additional climatological or topographical information, or both. Annual risk is estimated based on four quarterly models, or a single annual model. We assess predicted number of annual damage claims using MSE (left) and predicted probability of seeing one or more claims per year using the Brier score (right). Scores are given as mean scores over a 10-fold cross-validation of the entire data set.

observations with similar covariate values having zero claims. While this is substantially higher than the overall mean of 0.005, it nevertheless seems like a gross underestimation. For the most extreme observation with 7 claims the model predicts a chance of just under 0.05% of observing 7 or more claims, and similarly small probabilities are achieved for almost all observations with 3 or more claims. Unintuitively, however, this does not indicate a bad predictive fit in the context of the full data set. In fact, out of 6 million observed contracts, roughly 3000 contracts should fall above the 99.95th percentile of their predictive distribution. This highlights the importance of perceiving the predictions as probabilistic, rather than expecting the mean of the prediction to be always close to the observed number of contracts.

In order to avoid this fallacy associated with evaluating predictive performance of a model based on only the most extreme observations (Lerch et al., 2017), we consider conditional expectation diagrams. Figure 2 shows out-of-sample conditional expectation diagrams for this model. Here, we consider the contracts on the quarterly time resolution, which results in roughly 6 million contracts. The left diagram considers $k = 100$ sub-groups with 60,000 values in each. We observe an overall very good fit of the model. The right diagram considers only the 100,000 contracts with the highest predicted number of claims. That is, it considers the contracts approximately corresponding to the two rightmost points in the left diagram. The diagrams show that, even for the high-risk customers, the model is well calibrated and does not exhibit conditional bias. The shown diagrams are constructed from the cross-validation results and therefore assess the model fit out-of-sample, as is appropriate for our prediction setting.

Figure 3 shows the estimated effect of the topographical and climatological covariates for the quarterly negative binomial model. Specifically, we show the multiplicative effect of these variables by taking the exponential of the additive effect. All response panels share the same scale for the y-axis, but the range is suppressed to prevent adaptation of the fitted model by direct competitors of Gjensidige. The top row shows the effect of the three topographical indices. The TWI has a clear linear effect. The higher the wetness index the higher the expected number of claims. The effect of the slope index is shaped as a parabola with a minimum at approximately 10 degrees and a higher risk for both a lower and a higher slope. The effect of the slope index is difficult to interpret independently of the TWI which directly depends on the slope. The estimated spline for slope having a clearly pronounced shape gives evidence that information is added by considering both variables. The HAND index has the largest effect when its value is between 0 and 100 m. This is intuitive since, at a certain point, it becomes irrelevant to move even higher above the nearest drainage point. The effect of precipitation is clearly non-linear; for low values the effect

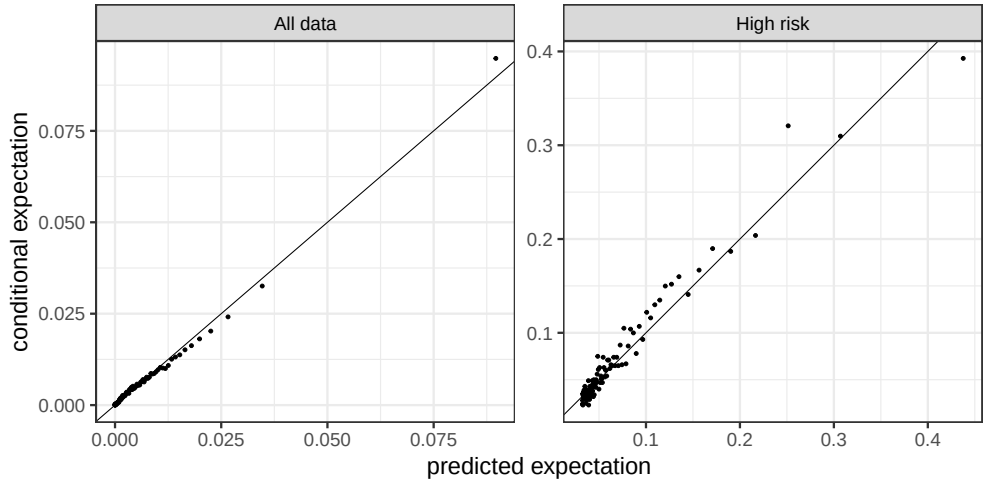

Figure 2. Conditional expectation diagrams for the quarterly negative binomial model using both topographic and climatological covariates. The right-hand side diagram is constructed from the 100,000 contracts with the highest predicted number of claims μ_i.

is flat, then close to linear between 3 and 9 mm of average daily precipitation in a quarter. The decreasing effect for precipitation values above 9 mm is somewhat counterintuitive and might be the result of little data in the upper tail, or signal a localised adaptation to current climate in very wet areas. For temperature, the effect is small for average quarterly temperature of less than 10°C, while it is highly non-linear for higher values. See Section 5 for a further discussion of the climatological effects. Lastly, we show a QQ-plot of the random effect of municipality which shows that the Gaussian assumption is fulfilled.

During model selection, we also investigated various subsets of property attributes for the building risk component. Of the considered building characteristics, only the variable 'type of roof' did not yield predictive power and was removed from the analysis. A basement variable is highly significant, properties with basement exhibiting a substantially higher risk than those without. Moreover, privately inhabited properties such as houses and cabins are at a higher risk of experiencing water damage than industrial or agricultural buildings by approximately an order of magnitude. Finally, the risk contribution from the construction year peaks in the early 1960s. This indicates both that building regulations have improved and that older properties are well capable of withstanding today's climatic exposure.

4.2 Risk assessment for a stationary climate

As described in Section 3.2, we can decompose our model into several risk factors, corresponding to the groups of covariates. In particular, the model assigns a specific topographical risk r_i^{topo} and a specific climatological risk r_i^{clim} to each contract, see equation (2). These factors depend only on the location and not on any contract-specific characteristics. They can therefore be calculated for any location for which topographical/climatological indices are available. Figure 4 shows maps of the corresponding sub-risks, as well as a combined map showing the product $r_i^{\text{topo}} r_i^{\text{clim}}$ for two example areas of 60 × 60 km, around the Norwegian cities of Bergen and Tromsø that obtain different risk profiles. The greater Bergen area, famous for heavy rainfall, gets assigned an above-average (that is, >1) climatological risk, resulting in an above average combined risk. For Tromsø, the climatological risk and subsequently the combined risk is more varied within the considered region.

4.3 Projected risk development due to climate change

We employ 12 regional climate projections under both RCP 4.5 and 8.5 for the two future periods 2031–2060 and 2071–2100 in order to assess risk development due to climate change. It can be observed in Figure 3 that the estimated splines for the climatological variables exhibit implausible

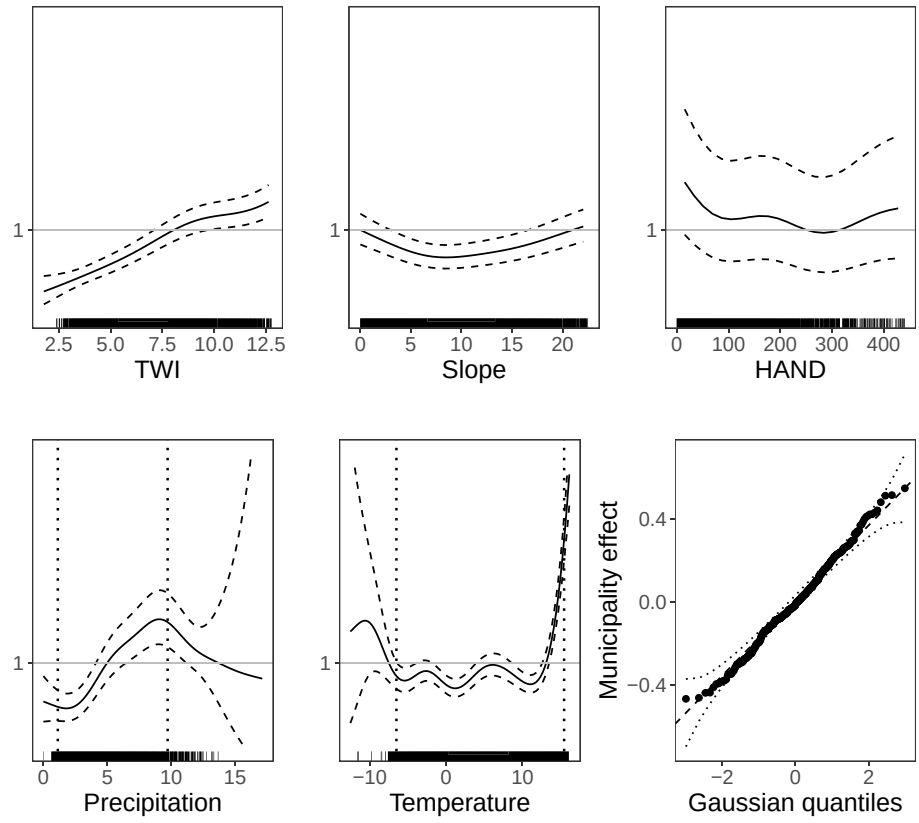

Figure 3. Multiplicative response effects of topographical (first row) and climatological (second row) covariates on water damage risk under the quarterly negative binomial model indicated by the mean effect (solid line) with a 95% confidence band (dashed lines), all on a joint *y*-axis scale. In-sample covariate values are indicated by black tick marks along the *x*-axes. For temperature and precipitation, dotted vertical lines show the 1st and 99th percentile, cf. Section 4.3. Bottom right: QQ-plot of the municipality random effect.

behaviour towards the tail of the distribution of the training data. For example, the spline for average daily precipitation decreases starting at 9.5 mm/day. Simultaneously, the uncertainty of the spline increases dramatically, suggesting that this effect likely constitutes an artefact of the statistical model rather than reflecting a causal relationship. This does not have much effect for the risk assessment in the current climate, since only few data points are located within the affected range. In climate projections for future time periods, however, the distribution of the climatological variables is shifted. As a result, the tail estimates of the splines can have much higher impact. To counteract an implausible extrapolation of the model beyond the range of the training data, we regularise the model by freezing the two climatological splines before the 1st percentile and past the 99th percentile of the training data, indicated as vertical dotted lines in Figure 3. This extrapolation issue and associated model assumptions are further discussed in Section 5.

To visualise future projected changes in risk, we consider the ratio of future and present climatological risk factors. Figure 5 shows the 10th, 50th, and 90th percentile of projected risk ratios based on the 12 different climate models. The projections show a significant increase in risk for the west coast up to and including the Lofoten archipelago, as well as for the southeastern part of the country around the capital, Oslo. Towards the end of the century, both RCP scenarios project large areas with increased risk of 25% and higher. Under RCP 8.5, an increase of more than 35% is projected for some areas. None of the considered scenarios project a decrease in risk for any part of the country.

The maps shown in Figure 5 are exclusively based on the projected changes in mean quarterly precipitation and surface temperature. Alternatively, we can use our full model to project the

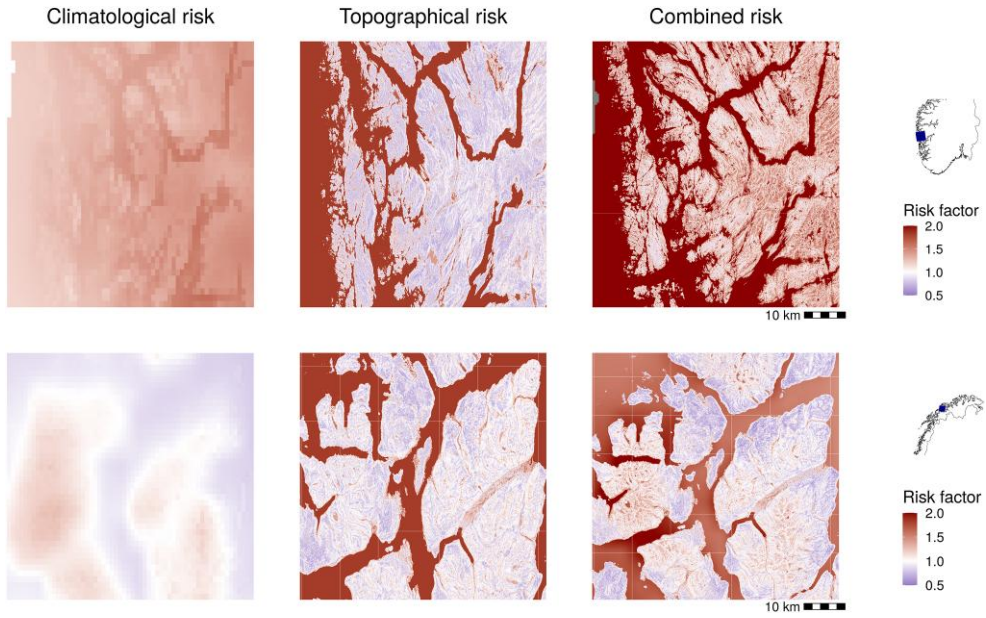

Figure 4. Climatological, topographical, and combined risk factor for the greater Bergen area (top row) and greater Tromsø area (bottom row), see the small maps on the right-hand side for the location of Bergen and Tromsø, respectively. The colourscale is centred at white, corresponding to the average risk, with below-average risk in blue and above-average risk in red.

change in number of claims per municipality. These projections incorporate topography and building-specific characteristics of the contracts within each municipality and are thus more directly connected to Gjensidige's portfolio. The resulting projections are shown in Figure 6. These overall ratios are somewhat higher than for the climatological risk in Figure 5, in that a larger overall area has a projected relative change of more than 15% in Figure 6 compared to Figure 5. This effect can partly be explained by the uneven spatial distribution of buildings within the considered municipalities: The majority of Norway's population lives in coastal areas, where a higher increase of risk is expected according to Figure 5.

Figure 7 shows densities of projected relative climatological risk factors across all grid cells covering Norway for the different climate model configurations. The figure shows that these marginal distributions are rather similar between the different climate models for each projection period and RCP scenario. All densities for the future period 2031–2060 as well as densities for 2071–2100 under RCP 4.5 have a mode within the interval [1, 1.05], while the projections for 2071–2100 under RCP 8.5 are more diffuse with modes within the interval [1, 1.3]. Furthermore, the densities are skewed with a heavy upper tail and an increasing skewness for a more pessimistic RCP and/or a time period further in the future.

We compare our findings to the results of Haug et al. (2011), who modelled the number of claims on the level of municipalities at a daily temporal resolution. That is, they consider a spatially aggregated insurance portfolio and meteorological information at a higher temporal resolution compared to our analysis. This renders use of building-specific covariates impossible, and even if they include hydrological variables like runoff, they do not use topography in their model. Further, Haug et al. (2011) consider a single climate model and are thus only able to capture model uncertainty and not uncertainty due to differences between different climate models. We extend this by considering both uncertainty in model parameters and variability resulting from different climate model configurations (see Table 1). In Figure 8, we show projected future change in the number of claims for three counties considered in Haug et al. (2011).[1]

[1] Haug et al. (2011) compare Akershus, Buskerud, and Hordaland, counties that no longer exist due to a restructuring of counties in 2020. For comparison, we derive predictions for these old counties by averaging predictions for all properties lying within the old county borders.

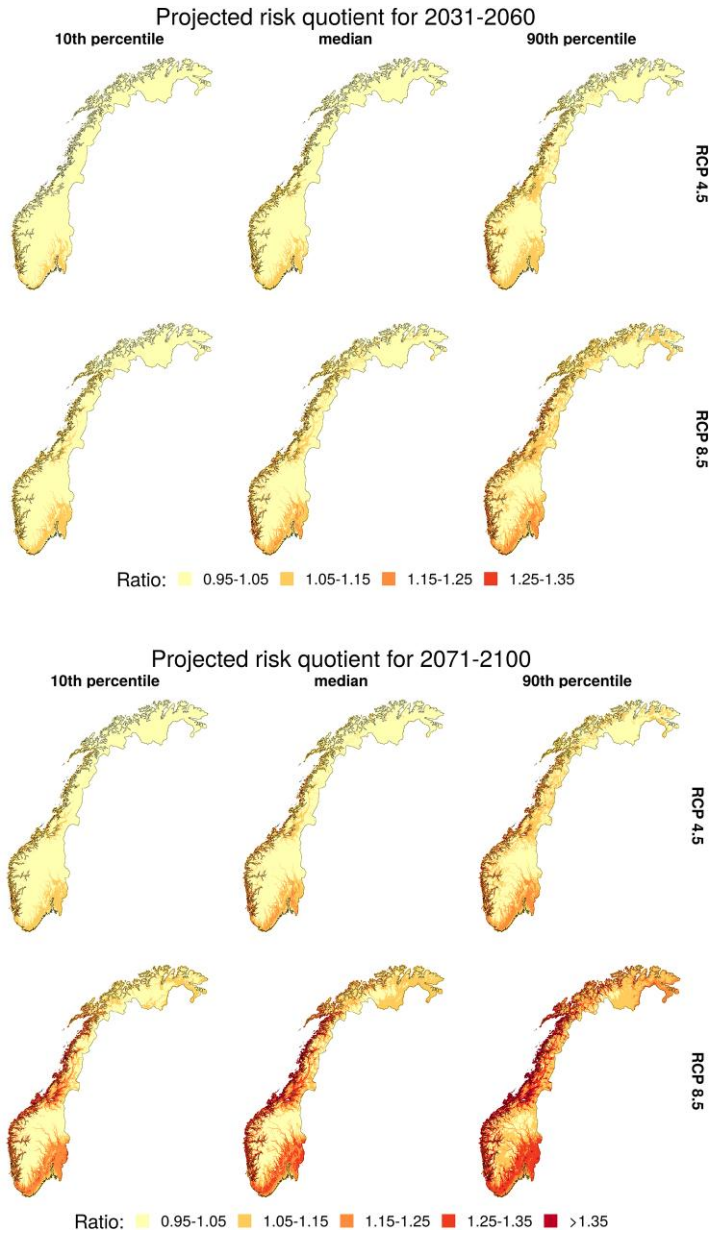

Figure 5. Ratio of projected future climatological risk for 2031–2060 (top) and 2071–2100 (bottom), and historical 1991–2020 climatological risk, see Section 3.2 for a definition. Each column shows pointwise 10th, 50th, and 90th percentiles of the probabilistic projections based on 12 climate models.

Figure 8 displays the point estimate and approximate 95% confidence intervals for the ratio of expected number of claims in the historical time period compared to 2071–2100. The estimates are based on 50 simulations from a multivariate normal approximation of the posterior distribution of the model parameters. For all counties the variation between the climate model configurations is greater than the variation due to parameter uncertainty, showing that this is an essential part of the overall uncertainty quantification. Our results are not directly comparable to those in Haug et al. (2011) as they use different climate models, emission scenarios, statistical models, and reference period, but some comments are justified. While our projected ratios for Akershus are approximately the same on average as the results reported in Haug et al. (2011), Buskerud, and

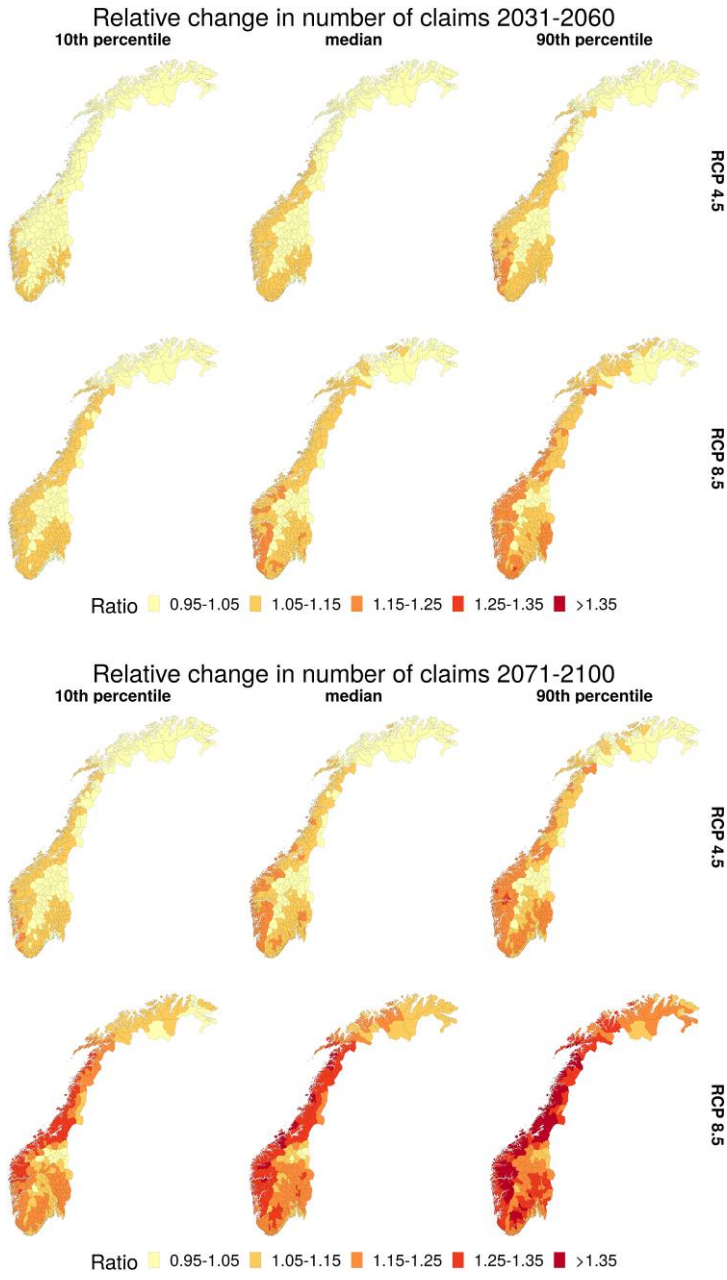

Figure 6. Projected relative change in number of claims for all municipalities in Norway based on climate projections for 2031–2060 (top) and 2071–2100 (bottom) relative to estimates for 2009–2021 based on 1991–2020 climatology. For each future period, the first row is based on climate projections under RCP 4.5 and the second row under RCP 8.5. The columns shows the 10th, 50th, and the 90th percentile (for each municipality) based on 12 climate models.

especially Hordaland, have significantly higher projected ratios in the current setting than previously reported.

5 Discussion

The projection of water damage risk to a future climate involves extrapolating the model in equation (3) beyond the range of the climatological indices for the reference period 1991–2020.

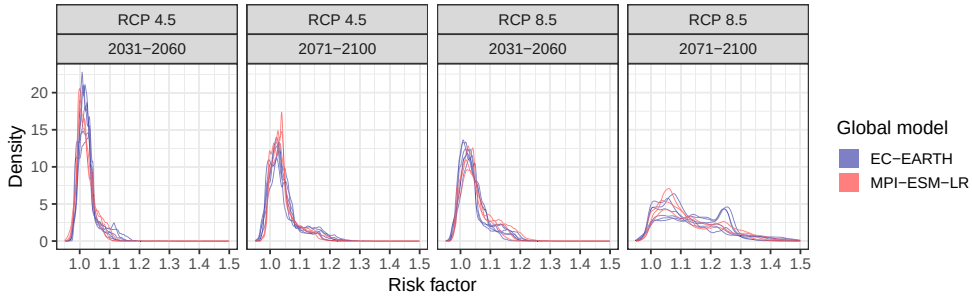

Figure 7. Densities for relative risk factors derived from the climate projections compared to the reference period 1991–2020. Each density represents one climate model configuration (see Table 1), and shows the distribution of projected risk factors across all grid cells covering Norway. The colours of the densities highlight the two different driving GCMs.

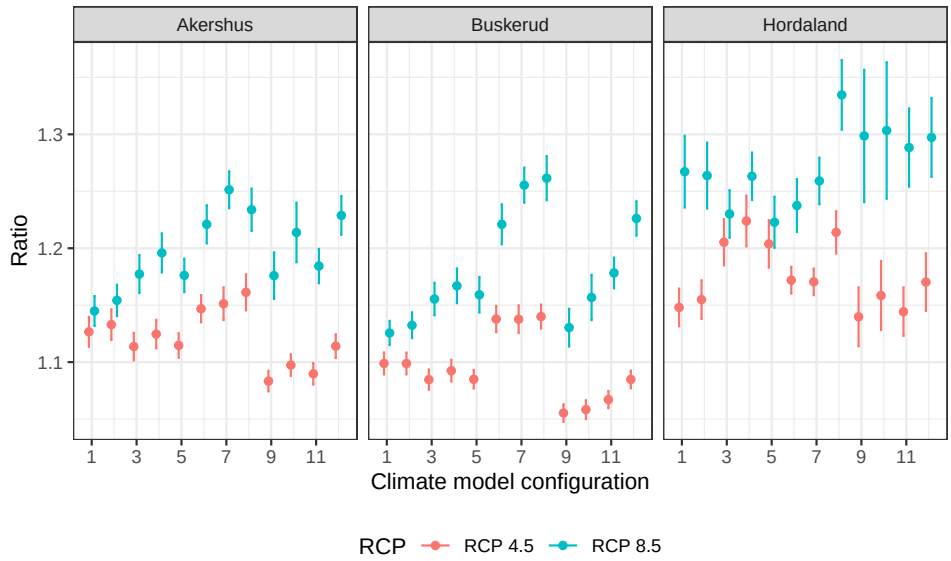

Figure 8. Ratios of change in the number of claims due to climate change under RCP 8.5 and RCP 4.5 for the period 2071–2100 for three counties in Norway. The climate model configuration numbers on the *x*-axis correspond to the ordering in Table 1. The uncertainty bands are given by twice the standard deviation based on 50 samples from the posterior distribution of the model parameters.

As Norwegian climate is expected to get warmer and wetter (Hanssen-Bauer et al., 2009), the extrapolation primarily involves an extrapolation beyond the upper limit of the in-sample values, see Figure 3. The projections presented in Section 4.3 are based on splines truncated at the 1st and 99th percentile of the in-sample distribution of covariate values. In order to assess the effect of this modelling choice, we have also formed water damage risk projections using a linear model for the climatological covariate effect applied with or without truncation, as well as models that only use a precipitation index rather than both precipitation and temperature indices. A model comparison is shown in Figure 9. Figure 9 shows that the modelling choice presented in Section 4.3 is the least conservative of the considered options. In particular, we see that the projected increase in water damage risk is, to a large extent, driven by the inclusion of a temperature index. However, a comparison of the predictive performance of these models in the current climate through a cross-validation study as presented in Section 4.1 shows that the non-linear splines

with both temperature- and precipitation-based indices significantly outperforms the other modelling options (results not shown).

In Figure 3, we see a steep increase in the spline for high temperatures. Physically, there is no immediate explanation why high temperatures should generate more claims. Thus, these temperatures most likely act as a proxy for some other phenomenon expressed in the data. One hypothesis would be that high surface temperatures are associated with more frequent cloud bursts typically seen during the warm season and in particular on hot days (local, convective precipitation). In urban areas, such events are known to bring potential risk of vast water damage due to a high proportion of non-porous surfaces, often in combination with limited capacity of the sewer system. Claims generated from such conditions will possibly ascribe high values of the temperature variable as the key source for their occurrence. Rural areas also experience hot days with cloud bursts, on average to the same extent as urban areas. Buildings in rural areas are less vulnerable to these events and, for these areas, we would expect the spline to have less of a steep increase for high temperatures. However, as the majority of the buildings in the insurance portfolio are situated in urban areas, the model estimates are formed primarily from the pattern of urban claims.

In a study of water damages in Switzerland, Bernet et al. (2019) show that critical thresholds for precipitation events vary substantially across complex terrain. The models presented in Section 3 assume that a certain level of quarterly precipitation and temperature has the same effect on damage risk everywhere in Norway, cf. Figure 3. Similar to Switzerland, there are substantial country-wide differences in climate. In particular, the amount and type of precipitation observed along the west coast is quite different from that observed, e.g. in the Oslo region on the east side (Hanssen-Bauer et al., 2009). To account for this, building traditions and wastewater collection systems vary across the country. We thus investigated models with spatially varying effects of precipitation and temperature. However, we found that this added model complexity yields poorer predictive performance which we believe to be due to overfitting. The climatic variability (and its impact on building and infrastructure traditions) is one of several effects that are implicitly accounted for by the county and municipality effects. Infrastructure adaptation to local climate might also partially explain why the response effect of precipitation is not monotonically increasing, cf. Figure 3. It should be noted that future projections for risk are made conditional on the current infrastructure in each municipality or county.

The majority of studies that model the effect of weather on insurance risk consider spatially aggregated claims data, see introduction for references. We take a different approach, modelling each property and location separately, but aggregating weather variables over longer time spans. This has the major advantage that it allows to incorporate building-specific covariates, such as whether the property has a basement or not. Moreover, a spatially disaggregated model is required to include the effect of local topography, since variables such as slope vary rapidly in space and quickly lose accuracy when they are spatially averaged. Our analysis shows that including both topographic and building-specific covariates substantially improves the predictive skill of the model. Moreover, risk models applied by insurance companies generally include building-specific information. This makes it challenging in practice to include information from spatially aggregated models into operational risk models.

On the other hand, using temporally aggregated weather data comes with the disadvantage that the model cannot learn from clusters of claims that are associated with heavy rainfall events. However, the aggregation of weather data in time is appropriate for predicting changes in risk due to climate change. Climate models generally aim to project changes in the distribution of weather rather than to provide accurate predictions of daily future weather. We did explore models putting more emphasis on aggregate measures of weather extremes rather than average weather variables (results not shown). For example, we fitted models considering the 95th and 99th percentile of local rainfall. Such models generally did not perform better than models based on average precipitation. The likely reason is that such indices are generally strongly correlated with average precipitation (in our data correlations were typically around $\rho \approx 0.99$).

During our research, we also experimented with models that neither aggregate in space nor time, i.e. targeting the binary response whether a claim happened for a specific property on a given day. While such models are in principle feasible for the data sets considered here, they are much more computationally costly. Moreover, the resulting models tended to be sensitive to outliers in the weather. This resulted in occasionally unreasonably high probabilities for a claim of 90% or

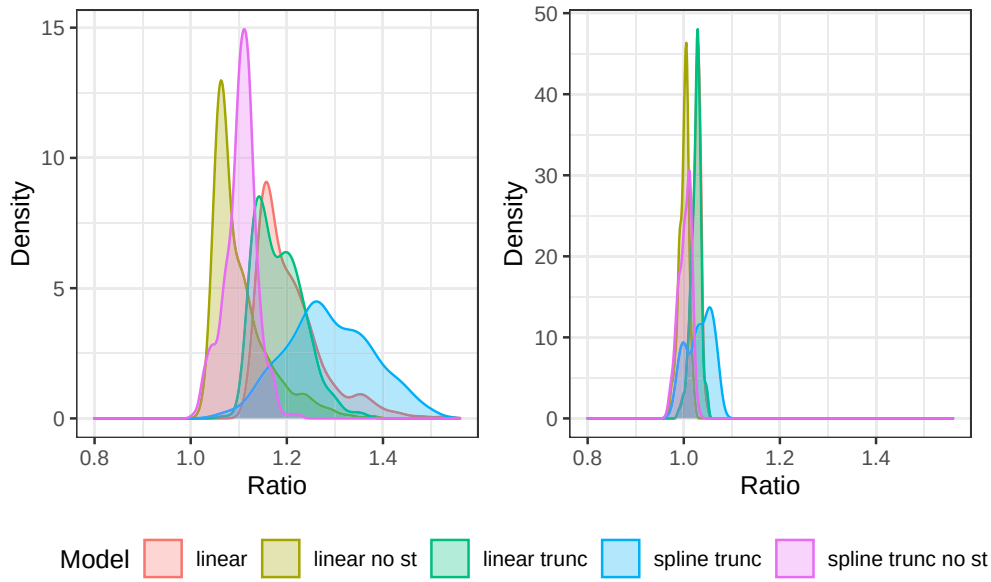

Figure 9. Left: Densities of 90th percentile of relative change in number of claims due to climate change under RCP 8.5 for the period 2071–2100 across all gridpoints in Norway. Right: Densities of 10th percentile of relative change in number of claims due to climate change under RCP 4.5 for the period 2031–2060 across all gridpoints in Norway. Each colour represents projections from different regression models: *spline* represents the model described in Section 3, *linear* means that the effect of precipitation and surface temperature is linear on log scale, *trunc* means precipitation and surface temperature is truncated at the 1st and 99th percentile of the training data, and *no st* means that surface temperature is excluded from the model.

higher for single properties at days with extreme precipitation, in particular when future climate projections were considered.

6 Conclusions

This paper proposes a nationwide model for risk of rainfall-induced water damage in Norway. The model incorporates local topographical and climatological information as well as contract-specific property data. The overall risk assessment can be decomposed into several factors which, in particular, allows for the derivation of topographical and climatological risk maps covering the entire country. Climate projections yield projections for changes in water damage risk in future decades due to climate change. Based on a multi-model ensemble of regional climate projections, we project a spatially varying increase in risk of as much as 25% for large areas towards the end of this century.

Insurance companies' risk models materialise in tariffs that price the individual customer and the individual object. Historical damages and losses—including those caused by pluvial flooding—form the basis for the tariffs, and trends together with knowledge of future changes make the tariffs predictive. The results of this study improve the pricing quality in multiple ways. Gjensidige will gain knowledge about future developments in the overall level of losses, the geographical loss distribution related to present and future climate and the risk level of individual buildings due to their location in the terrain. These are aspects that expand and challenge current risk assessment practice in the insurance sector. The model provides not only precise risk information for Gjensidige's existing customers but also any potential new object—built or planned—in Gjensidige's future portfolio. All of this is crucial information for pricing according to risk at both portfolio and customer level.

Risk associated with climate and climate change plays a special role compared to other risks due to global efforts for climate change adaptation and mitigation. The risk models that have been developed here provide Gjensidige with a basis for redistribution of premiums at building level (topography) and future redistribution and adaptation at geographical level. More importantly, the

new knowledge can be utilised to reduce risk for both municipalities and individual customers, supporting and furthering a sustainable business strategy. Municipalities can gain valuable insights regarding investment priorities such as upgrades of sewerage networks or planning of new residential or commercial areas. Gjensidige will also seek to use the results to develop preventive measures that customers can implement themselves.

After an earlier trial period, Norwegian insurance companies will from 2022 provide claims data related to weather and climate to the Norwegian authorities. The purpose of this programme is for the municipalities to get an overview of areas with high-risk exposure so that measures can be taken where they are expected to have the greatest effect. The results of the current study provide valuable knowledge regarding expected future changes in this context. Models of the type presented here can help ensure that measures are prioritised to provide the greatest possible societal benefit, both by preventing economic and social costs and because every damage entails a CO_2 cost. The most effective climate change adaptation and mitigation strategy for an insurance company is to prevent damage occurrence.

Acknowledgments

We thank Andreas Lura for extracting the insurance data, Magne Aldrin for helpful discussions and Annabelle Redelmeier for help with the manuscript. Moreover, we would like to thank the associate editor and an anonymous reviewer for their suggestions which helped to improve the content of the paper.

Data availability

Our analysis utilizes four different datasets, the details of which can be found in Section 2. The insurance data is not publicly available due to its confidential nature. The topography data is generated from a freely accessible Digital Elevation Model, generated by the Norwegian Mapping Authorities (Kartverket), and is available at https://hoydedata.no. The meteorological data (seNorge version 2018), produced by the Norwegian Meteorological Service, is available at https://thredds.met.no/thredds/catalog/senorge/seNorge_2018/. Finally, the regional climate projections we use are provided by the EURO-CORDEX initiative and freely available. For information on how to access this data, please visit https://cordex.org/.

References

Barredo J. I. (2007). Major flood disasters in Europe: 1950–2005. *Natural Hazards*, *42*(1), 125–148. https://doi.org/10.1007/s11069-006-9065-2

Bernet D. B., Prasuhn V., & Weingartner R. (2017). Surface water floods in Switzerland: What insurance claim records tell us about the damage in space and time. *Natural Hazards and Earth System Sciences*, *17*(9), 1659–1682. https://doi.org/10.5194/nhess-17-1659-2017

Bernet D. B., Trefalt S., Martius O., Weingartner R., Mosimann M., Röthlisberger V., & Zischg A. P. (2019). Characterizing precipitation events leading to surface water flood damage over large regions of complex terrain. *Environmental Research Letters*, *14*(6), 064010. https://doi.org/10.1088/1748-9326/ab127c

Beven K. J., & Kirkby M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, *24*(1), 43–69. https://doi.org/10.1080/02626667909491834

Brier G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Cornes R. C., van der Schrier G., van den Besselaar E. J. M., & Jones P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, *123*(17), 9391–9409. https://doi.org/10.1029/2017JD028200

Diggle P. J., & Milne R. K. (1983). Negative binomial quadrat counts and point processes. *Scandinavian Journal of Statistics*, *10*, 257–267.

Douris J., Kim G., Baddour O., Abrahams J., Lapitan J. M., Shumake-Guillemot J., Green H., Murray V., Bhattacharjee S., Palm E., Sengupta R., Stevens D., & Zommers Z. (2021). *Atlas of Mortality and Economic Losses from Weather, Climate, and Water Extremes (1970–2019)*. World Meteorological Organization.

Giorgetta M. A., Jungclaus J., Reick C. H., Legutke S., Bader J., Böttinger M., Brovkin V., Crueger T., Esch M., Fieg K., & Glushak K. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597. https://doi.org/10.1002/jame.20038

Gneiting T., & Raftery A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Gradeci K., Labonnote N., Sivertsen E., & Time B. (2019). The use of insurance data in the analysis of surface water flood events–a systematic review. *Journal of Hydrology*, *568*, 194–206. https://doi.org/10.1016/j.jhydrol.2018.10.060

Häggmark L., Ivarsson K.-I., Gollvik S., & Olofsson P.-O. (2000). MESAN, an operational mesoscale analysis system. *Tellus A: Dynamic Meteorology and Oceanography*, *52*(1), 2–20. https://doi.org/10.3402/tellusa.v52i1.12250

Hammond M. J., Chen A. S., Djordjević S., Butler D., & Mark O. (2015). Urban flood impact assessment: A state-of-the-art review. *Urban Water Journal*, *12*(1), 14–29. https://doi.org/10.1080/1573062X.2013.857421

Hanssen-Bauer I., Drange H., Førland E. J., Roald L. A., Børsheim K. Y., Hisdal H., Lawrence D., Nesje A., Sandven S., Sorteberg A., & Sundby S. (2009). Climate in Norway 2100. *Background information to NOU Climate adaptation (In Norwegian: Klima i Norge 2100. Bakgrunnsmateriale til NOU Klimatilplassing)*. Norsk klimasenter.

Hastie T. J., & Tibshirani R. J. (1990). *Generalized additive models*. Chapman & Hall.

Haug O., Dimakos X. K., Vårdal J. F., Aldrin M., & Meze-Hausken E. (2011). Future building water loss projections posed by climate change. *Scandinavian Actuarial Journal*, *2011*(1), 1–20. https://doi.org/10.1080/03461230903266533

Hazeleger W., Wang X., Severijns C., Ştefănescu S., Bintanja R., Sterl A., Wyser K., Semmler T., Yang S., Van den Hurk B., & Van Noije T. (2012). EC-Earth V2. 2: Description and validation of a new seamless earth system prediction model. *Climate Dynamics*, *39*(11), 2611–2629. https://doi.org/10.1007/s00382-011-1228-5

Houston D., Werrity A., Bassett D., Geddes A., Hoolachan A., & McMillan M. (2011). *Pluvial (rain-related) flooding in urban areas: The invisible hazard*. https://eprints.gla.ac.uk/162145/7/162145.pdf.

Jacob D., Petersen J., Eggert B., Alias A., Christensen O. B., Bouwer L. M., Braun A., Colette A., Déqué M., Georgievski G., & Georgopoulou E. (2014). EURO-CORDEX: New high-resolution climate change projections for European impact research. *Regional Environmental Change*, *14*(2), 563–578. https://doi.org/10.1007/s10113-013-0499-2

Jacob D., Teichmann C., Sobolowski S., Katragkou E., Anders I., Belda M., Benestad R., Boberg F., Buonomo E., Cardoso R. M., & Casanueva A. (2020). Regional climate downscaling over Europe: Perspectives from the EURO-CORDEX community. *Regional Environmental Change*, *20*(2), 1–20. https://doi.org/10.1007/s10113-020-01606-9

Landelius T., Dahlgren P., Gollvik S., Jansson A., & Olsson E. (2016). A high-resolution regional reanalysis for Europe. Part 2: 2D analysis of surface temperature, precipitation and wind. *Quarterly Journal of the Royal Meteorological Society*, *142*(698), 2132–2142. https://doi.org/10.1002/qj.2813

Lerch S., Thorarinsdottir T. L., Ravazzolo F., & Gneiting T. (2017). Forecaster's dilemma: Extreme events and forecast evaluation. *Statistical Science*, *32*(1), 106–127. https://doi.org/10.1214/16-STS588

Li Z., & Wood S. N. (2020). Faster model matrix crossproducts for large generalized linear models with discretized covariates. *Statistics and Computing*, *30*(1), 19–25. https://doi.org/10.1007/s11222-019-09864-2

Liao X., & Meyer M. C. (2019). cgam: An R package for the constrained generalized additive model. *Journal of Statistical Software*, *89*(5), 1–24. https://doi.org/10.18637/jss.v089.i05

Lussana C. (2020). 'seNorge observational gridded datasets', seNorge_2018, version 20.05, arXiv, arXiv:2008.02021, preprint: not peer reviewed.

Lyubchich V., Newlands N. K., Ghahari A., Mahdi T., & Gel Y. R. (2019). Insurance risk assessment in the face of climate change: Integrating data science and statistics. *WIREs Computational Statistics*, *11*(4), e1462. https://doi.org/10.1002/wics.1462

Meyer M. C. (2018). A framework for estimation and inference in generalized additive models with shape and order restrictions. *Statistical Science*, *33*(4), 595–614. https://doi.org/10.1214/18-sts671

Michelangeli P.-A., Vrac M., & Loukos H. (2009). Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophysical Research Letters*, *36*(11), L11708. https://doi.org/10.1029/2009GL038401

Netzel L. M., Heldt S., Engler S., & Denecke M. (2021). The importance of public risk perception for the effective management of pluvial floods in urban areas: A case study from Germany. *Journal of Flood Risk Management*, *14*(2), e12688. https://doi.org/10.1111/jfr3.12688

Nobre A. D., Cuartas L. A., Hodnett M., Rennó C. D., Rodrigues G., Silveira A., & Saleska S. (2011). Height Above the Nearest Drainage – A hydrologically relevant new terrain model. *Journal of Hydrology*, *404*(1–2), 13–29. https://doi.org/10.1016/j.jhydrol.2011.03.051

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Scheel I., Ferkingstad E., Frigessi A., Haug O., Hinnerichsen M., & Meze-Hausken E. (2013). A Bayesian hierarchical model with spatial variable selection: The effect of weather on insurance claims. *Journal of the*

Royal Statistical Society: Series C (Applied Statistics), *62*(1), 85–100. https://doi.org/10.1111/j.1467-9876.2012.01039.x

Skougaard Kaspersen P., Høegh Ravn N., Arnbjerg-Nielsen K., Madsen H., & Drews M. (2017). Comparison of the impacts of urban development and climate change on exposing European cities to pluvial flooding. *Hydrology and Earth System Sciences*, *21*(8), 4131–4147. https://doi.org/10.5194/hess-21-4131-2017

Spekkers M., Ten Veldhuis J., Kok M., & Clemens F. (2011). Analysis of pluvial flood damage based on data from insurance companies in the Netherlands. In G. Zenz, & R. Hornich, (Eds.), *Proceedings International Symposium Urban Flood Risk Management, UFRIM, 2011, September 21–23, Graz, Austria*. Technische Universitat Graz.

Spekkers M. H., Kok M., Clemens F. H. L. R., & ten Veldhuis J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Natural Hazards and Earth System Sciences*, *14*(9), 2531–2547. https://doi.org/10.5194/nhess-14-2531-2014

Thorarinsdottir T. L., & Schuhen N. (2018). Verification: Assessment of calibration and accuracy. In S. Vannitsem, D. S. Wilks, & J. Messner (Eds.), *Statistical Postprocessing of Ensemble Forecasts* (pp. 155–186). Elsevier.

van Straaten C., Whan K., Coumou D., van den Hurk B., & Schmeits M. (2020). The influence of aggregation and statistical post-processing on the subseasonal predictability of European temperatures. *Quarterly Journal of the Royal Meteorological Society*, *146*(731), 2654–2670. https://doi.org/10.1002/qj.3810

Vrac M., Drobinski P., Merlo A., Herrmann M., Lavaysse C., Li L., & Somot S. (2012). Dynamical and statistical downscaling of the French Mediterranean climate: Uncertainty assessment. *Natural Hazards and Earth System Sciences*, *12*(9), 2769–2784. https://doi.org/10.5194/nhess-12-2769-2012

Wilks D. S. (2011). *Statistical methods in the atmospheric sciences* (Vol. *100*). Academic Press.

Wood S. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Chapman and Hall/CRC.

Wood S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society (B)*, *65*(1), 95–114. https://doi.org/10.1111/1467-9868.00374

Wood S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood S. N., Goude Y., & Shaw S. (2014). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *64*(1), 139–155. https://doi.org/10.1111/rssc.12068

Wood S. N., Li Z., Shaddick G., & Augustin N. H. (2017). Generalized additive models for gigadata: Modeling the U.K. black smoke network daily data. *Journal of the American Statistical Association*, *112*(519), 1199–1210. https://doi.org/10.1080/01621459.2016.1195744

Yang W., Andréasson J., Phil Graham L., Olsson J., Rosberg J., & Wetterhall F. (2010). Distribution-based scaling to improve usability of regional climate model projections for hydrological climate change impacts studies. *Hydrology Research*, *41*(3–4), 211–229. https://doi.org/10.2166/nh.2010.004

Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **72**, 829–843
https://doi.org/10.1093/jrsssc/qlad020
Advance access publication 16 February 2023
Discussion Paper

ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS
Journal of the Statistics Society
Series C
Applied Statistics

The importance of context in extreme value analysis with application to extreme temperatures in the U.S. and Greenland

Daniel Clarkson[1], Emma Eastoe[1] and Amber Leeson[2]

[1]Department of Mathematics and Statistics, Lancaster University, Lancaster, Lancashire, UK
[2]Data Science Institute, Lancaster University, Lancaster, Lancashire, UK

Address for correspondence: Daniel Clarkson, Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK. Email: d.clarkson@lancaster.ac.uk

*Read before The Royal Statistical Society at the first meeting on 'Statistical aspects of climate change' held at the Society's 2022 annual conference in Aberdeen on Wednesday, 14 September 2022, the President, Professor Sylvia Richardson, in the Chair.

Abstract

Statistical extreme value models allow estimation of the frequency, magnitude, and spatio-temporal extent of extreme temperature events in the presence of climate change. Unfortunately, the assumptions of many standard methods are not valid for complex environmental data sets, with a realistic statistical model requiring appropriate incorporation of scientific context. We examine two case studies in which the application of routine extreme value methods result in inappropriate models and inaccurate predictions. In the first scenario, incorporating random effects reflects shifts in unobserved climatic drivers that led to record-breaking US temperatures in 2021, permitting greater accuracy in return period prediction. In scenario two, a Gaussian mixture model fit to ice surface temperatures in Greenland improves fit and predictive abilities, especially in the poorly-defined upper tail around 0°C.

1 Introduction

The roots of the statistical modelling of extreme events lie in the theoretical work of Von Mises (1936) and Gnedenko (1943) who sought to understand the asymptotic behaviour of sample maxima. From the 1950s, this probability theory began to be translated into the statistical modelling framework now known as the block maxima method (Gumbel, 1958). Both theory and modelling framework were later extended to a Peaks over threshold (PoT) approach (Davison & Smith, 1990; Pickands, 1975). It was recognized early on that both block maxima and PoT frameworks could make major contributions to the study of natural hazards, leading to a long-standing relationship with the environmental sciences and civil engineering. Specific application areas include hydrology (Jonathan & Ewans, 2013; Katz et al., 2002), air pollution (Eastoe & Tawn, 2009; Gouldsborough et al., 2021; Reich et al., 2013), temperatures (Acero et al., 2014; Winter et al., 2016), precipitation (Katz, 1999), drought (Burke et al., 2010), wind speeds (Fawcett & Walshaw, 2006), statistical downscaling (Friederichs, 2010; Towe et al., 2017), space weather (Rogers et al., 2020; Thomson et al., 2011), and climate change problems (Cheng et al., 2014; Cooley, 2009; Sterl et al., 2008).

Both theory and models have evolved considerably since the mid-twentieth century. In particular, there have been massive advances in methodology for multivariate extreme events (Heffernan & Tawn, 2004; Ramos & Ledford, 2009; Rootzén et al., 2018) and the related areas of spatial (Cooley et al., 2012; Huser & Wadsworth, 2019), temporal (Ledford & Tawn, 2003; Winter

Received: October 31, 2021. **Revised:** December 23, 2022. **Accepted:** December 23, 2022

& Tawn, 2017), and spatio-temporal (Economou et al., 2014; Simpson & Wadsworth, 2021) extreme value modelling. These developments are in part a response to the increasing availability of large spatio-temporal environmental data sets obtained, for example, from satellite images, high-resolution process-based models and online measurement tools.

Nevertheless, the continued development of methods for univariate extremes remains a crucial research topic, not least due to the absence of a unified best practice methodology for modelling univariate extremes in the presence of complex trends, such as those seen in the presence of climate change. There is also still no single standard to inform the quantification and communication of risk in the presence of trends, whether or not this is climate change driven. While such a standard may ultimately prove to be unobtainable, the commonly employed approach of reporting a single risk measure to cover all scenarios, regardless of long-term trends or other changes, can, and should, be improved upon. Lastly, specification of a model that describes the marginal model well is an essential first step in the very large number of copula-based multivariate extreme value models (Joe, 1994).

This paper examines extreme temperature data for two regions with contrasting climates and very different geophysical features. The first case study assesses the heatwave experienced in the western U.S. during the summer of 2021 to ascertain whether the record-breaking temperatures from this event are consistent with historical temperature extremes. This study was motivated by a recent analysis of ERA5 reanalysis model output (Philip et al., 2021) which used a statistical extreme value model fitted to pre-2021 temperature extremes to estimate the upper bound of the temperature distribution. They found that the 2021 observations exceeded this upper bound at multiple locations. We propose some adaptations to their analysis to better investigate if this is the case and, if so, to understand why. The second case study analyses spatio-temporal records of ice surface temperature for the Greenland ice sheet, motivated by the need to find an appropriate univariate model for the measurements in each cell as the first stage of a spatial extreme value model. Given the large number of cells (1139), the marginal model should ideally be designed to permit automatic fitting, however the characteristics of the upper tail of the temperature distribution differ considerably across the ice sheet and so careful consideration is required to achieve this objective.

A feature common to both case studies is that the tail behaviour of the univariate distribution of interest cannot be described using out-of-the-box extreme value methods. Instead, by applying a pragmatic, contextualized and data-driven approach, much improved models can be developed while avoiding the need to incorporate huge numbers of parameters or highly complex structures. This approach has several advantages. Firstly, since the models are computationally both cheap and stable, analysis of large data sets is entirely feasible. Secondly, simple models are less prone to errors in fitting and interpretation, making them particularly useful for those without statistical training. Lastly, a simple model results in straightforward estimates of risk which makes it easier for the end-user to put these estimates to use.

2 Extreme value analysis

This section provides background on univariate extreme value modelling and introduces the random effects model that is subsequently used in Section 3. Univariate extreme value methods encompass models for both block maxima and PoT data sets. Central to this paper is the PoT approach which characterizes the tail behaviour of the underlying process by modelling only those observations in the data set that lie above a preselected high threshold. In contrast, block maxima methods describe the distribution of the maxima of predefined and nonoverlapping blocks; for environmental studies this block is usually a year. Each method has a theoretically motivated probability distribution at its core: the generalized extreme value distribution (block maxima) and the generalized Pareto distribution (PoT). While theoretical justification of these distributions is beyond the scope of the paper, note that, as with the central limit theorem, the justification relies on asymptotic arguments. For finite sample sizes, the distributions therefore only approximate the tail behaviour and are not guaranteed to describe the data exactly.

2.1 Vanilla block maxima and PoT models

The theory on which the vanilla versions of both block maxima and PoT models are based starts with the assumption of a sequence $Y_1, \ldots, Y_m$ of independent replicates of a random variable Y, with distribution $F(y) = \Pr[Y \leq y]$. Under some nonrestrictive conditions on F, Von Mises (1936)

and Gnedenko (1943) showed that in the limit as $m \to \infty$, the distribution of the normalized block maximum $(M_m - b_m)/a_m = (\max\{Y_1, \ldots, Y_m\} - b_m)/a_m$ is one of the three types of max-stable distribution: Weibull, Fréchet, or Gumbel. While both the rate of convergence to the limiting distribution and the sequences $\{a_m > 0\}$ and $\{b_m\}$ are determined by the underlying distribution F which, in a modelling context is unknown, this result is nevertheless used to justify the use of the generalized extreme-value distribution (GEV) to model the maxima of large but finite blocks.

The GEV is a class of distributions combining the three types of max-stable distribution. The resulting cumulative distribution function is,

$$G(z) = \exp\left\{-\left[1 + \frac{\xi}{\sigma}(z - \mu)\right]^{-1/\xi}\right\}, \quad -\infty < z < y^+$$

where $y^+ < \infty$ in the case $\xi < 0$. This distribution has location μ, scale $\sigma > 0$ and shape ξ parameters, with the Weibull, Fréchet, and Gumbel distributions corresponding to $\xi < 0$, $\xi > 0$ and $\xi \to 0$, respectively. As a model the GEV has the advantage of increased flexibility over choosing one of the three types a priori.

The vanilla PoT model is motivated by a result of Pickands (1975) who showed that, under some nonrestrictive regularity conditions, as $u \to y^+$ for $y > u$,

$$F(y) = 1 - \Pr[Y > u]\Pr[Y > y \mid Y > u] \approx 1 - \phi\bar{G}(y) = 1 - \phi\left[1 + \frac{\xi}{\psi}(y - u)\right]_+^{-1/\xi} \tag{1}$$

where $a_+ = \max\{a, 0\}$, $0 \le \phi \le 1$, $\psi > 0$ and $\bar{G}(y)$ is the survival function of the generalized Pareto distribution. The parameters ϕ, ψ, and ξ are known as the rate, scale and shape, respectively; the rate $\phi = \Pr[Y > u]$ is sometimes called the exceedance probability. The value of the shape parameter is determined by the rate at which $F(y) \to 1$ as $y \to y^+$; exponential decay corresponds to $\xi = 0$, and heavy (light) decay to $\xi > 0$ ($\xi < 0$). For $\xi \ge 0$, there is no finite end-point, while $y^+ = u - \sigma/\xi < \infty$ if $\xi < 0$.

By assuming that this limit result is a suitable approximation to the behaviour of the exceedances of a finite, but high, level u, it can be used to motivate a model in which the magnitudes of the threshold exceedances $\{y_i - u : y_i > u, \ i = 1, \ldots, n\}$ are a random sample from the generalized Pareto distribution and the exceedance times a random sample from a homogeneous Poisson process. While not essential an assumption of independence is often made about the exceedances. If there is evidence of serial dependence, a routine and theoretically justified approach is to identify all clusters of extremes and model only the frequency and magnitude of the cluster maxima (Davis et al., 2013; Drees, 2008; Laurini & Tawn, 2003; Smith & Weissman, 1994). In either case, the parameters (ϕ, ψ, ξ) are estimated with standard statistical inference procedures, e.g maximum likelihood or Bayesian inference.

A crucial modelling consideration is the choice of threshold u, which requires a compromise between retaining a sufficiently large number of exceedances such that estimation of the model parameters is possible, while still ensuring that the assumption of the limiting approximation is plausible. Various tools are available to aid threshold selection. Most of these use two key threshold stability properties of the generalized Pareto distribution to identify the minimum plausible threshold above which the distribution is a suitable description of the data.

2.2 Random effects PoT model

The most common extension of the vanilla PoT model is to model one or more of the distribution parameters (ϕ, ψ, ξ) as a function of covariates. The motivation for this is that most physical processes display changes in their behaviour over time. Such changes may be cyclic, smooth or sudden and often arise in response to changes in one or more of climate, weather conditions or other geophysical processes. While many commonly used regression modelling frameworks, both parametric (Davison & Smith, 1990; Eastoe & Tawn, 2009; Smith, 1989) and semi-parametric (Hall & Tajvidi, 2000; Jones et al., 2016; Yee & Stephenson, 2007; Youngman, 2019), have been incorporated into the PoT framework, there are several barriers that prevent their successful implementation. Firstly, since most environmental processes are driven by multiple interacting factors,

appropriate drivers will often be either unknown or unobserved, or both. Secondly, PoT data sets are small by definition and there is frequently insufficient information to identify the often subtle covariate relationships. Lastly, while the motivating extreme value theory provides a justification to extrapolate into the underlying tail distribution given values of the covariates, there is no similar justification to extrapolate the covariate-response relationship beyond the measurement period. This is a serious limitation for climate change forecasts.

A plausible alternative is the PoT random effects model which overcomes the above limitations for cases where changes in extremal behaviour occur between, but not within, years. Previously applied to hydrological data by Eastoe (2019) and Towe et al. (2020), inter-year variability in both the magnitudes and frequency of the extremes is modelled by latent variables (or 'random effects') which vary between years. The advantage of this framework is that, unlike a parametric regression, it requires no prior specification of the functional form for the inter-year variability. In this sense, the model is a compromise between the rigid fully parametric regression model and the flexible non-parametric regression. It can be used to identify long-term trends, unusually extreme years, and groupings of years in which the extremal behaviour is similar. The model also has the capability to make predictions about extreme episodes over a period of time (e.g., a number of decades) but not to do so for a specific out-of-sample year. Although it is beyond the scope of the current paper, such predictions could be achieved by incorporating serial dependence in the random effects.

For generality, we define a threshold u_i for year i. This allows for cases in which it is appropriate for the threshold, and hence the definition of an extreme observation, to vary between years. Let Y_{ij} be the jth observation in year i then, for $y > u_i$,

$$\Pr[Y_{ij} \leq y \mid Y_{ij} > u_i] = 1 - \phi_i \left[1 + \frac{\xi_i}{\psi_i}(y - u_i) \right]_+^{-1/\xi_i} \tag{2}$$

where

$$\psi_i \sim F_\psi(\psi_i \mid \theta_\psi), \quad \xi_i \sim F_\xi(\xi_i \mid \theta_\xi), \quad \phi_i \sim F_\phi(\phi_i \mid \theta_\phi) \tag{3}$$

and F_ψ, F_ξ, and F_ϕ are prespecified distributions with parameters θ_ψ, θ_ξ, and θ_ϕ, respectively. In hierarchical modelling terminology, equations (2) and (3) are data and latent process models, respectively, and (ψ_i, ξ_i, ϕ_i) are random effects for year i. We assume throughout that the responses Y_{ij} are independent given the random effects and that the random effects are mutually independent and serially uncorrelated.

Let y denote the full vector of data, N be the number of years in the observation period and n_i be the number of observations in each year. Denote the vectors of random effects by $\psi = (\psi_1, \ldots, \psi_N)$, $\xi = (\xi_1, \ldots, \xi_N)$, $\phi = (\phi_1, \ldots, \phi_N)$. The starting point for inference is the joint likelihood function

$$L(y \mid \psi, \xi, \phi, \theta_\psi, \theta_\xi, \theta_\phi) = L(y \mid \psi, \xi, \phi) f(\psi \mid \theta_\psi) f(\xi \mid \theta_\xi) f(\phi \mid \theta_\phi),$$

where

$$L(y \mid \psi, \xi, \phi) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} \left\{ \phi_i \psi_i^{-1} \left[1 + \frac{\xi_i}{\psi_i}(y_{ij} - u_i) \right]_+^{-1/\xi_i - 1} \right\}^{\mathbb{I}[y_{ij} > u_i]} (1 - \phi_i)^{\mathbb{I}[y_{ij} \leq u_i]} \tag{4}$$

and $f(\psi \mid \theta_\psi)$, $f(\xi \mid \theta_\xi)$, and $f(\phi \mid \theta_\phi)$ are each the product of the relevant distribution taken from equation (3), for example $f(\psi \mid \theta_\psi) = \prod_{i=1}^{N} f_\psi(\psi_i \mid \theta_\psi)$ where f_ψ is the density associated with F_ψ. Note that the dimension of the full parameter vector is $3N + |\theta_\psi| + |\theta_\xi| + |\theta_\phi|$.

Since the random effects cannot be integrated out of the joint likelihood function, maximum likelihood parameter estimates must be obtained through numerical optimization of the very high-dimensional log-likelihood function. A more pragmatic approach is to use Bayesian inference; to do so, prior distributions $\pi_\psi(\psi)$, $\pi_\xi(\xi)$, and $\pi_\phi(\phi)$ must now be specified for θ_ψ, θ_ξ, and θ_ϕ,

respectively. We recommend that these are chosen to be as uninformative as possible, e.g., flat Gaussian distributions, in order to avoid undue impact on the analysis. Combining the prior distributions with the likelihood function (4), the joint posterior is

$$\pi(\theta_\psi, \theta_\xi, \theta_\phi, \psi, \xi, \phi) = L(y \mid \psi, \xi, \phi) f(\psi \mid \theta_\psi) f(\xi \mid \theta_\xi) f(\phi \mid \theta_\phi) \pi_\psi(\psi) \pi_\xi(\xi) \pi_\phi(\phi).$$

Since closed forms cannot be found for the joint and marginal posterior distributions, the posterior distribution is estimated using a sampling algorithm. We use an adaptive Markov chain Monte Carlo (MCMC) algorithm to enable automatic tuning of the proposal step-sizes and achieve fast convergence. Our algorithm is based on the adaptive Metropolis–Within–Gibbs scheme proposed by Roberts and Rosenthal (2009). Each parameter is updated separately via a Metropolis-Hastings random walk (MHRW) step, allowing separate evolution of the MHRW scaling parameter associated with each of the model parameters. Earlier work by Eastoe (2019) showed that of a number of possible alternatives, this algorithm worked the best for this particular model.

3 Case study 1: US air temperatures

This case study is motivated by the 2021 heatwave which affected three western states in the United States. An analysis of climate model annual maxima by Philip et al. (2021) showed that the observed 2021 annual maximum were inconsistent with historical extremes. Further, this finding recurred across a number of stations and held despite adjustments made to the data to account for long-term trends in global temperature. Our objectives are to examine whether (a) an analysis of observational annual maxima is consistent with the ERA5 analysis of Philip et al. (2021), (b) an analysis of PoT data points to the same conclusions as an analysis of annual maxima, and (c) a more flexible model for long-term trends captures better the 2021 event.

The three regions most affected by the heatwave were Oregon, Washington, and British Columbia. For these regions, we obtain maximum daily temperatures from Version 3 of the Global Historical Climatology Network (Menne et al., 2012). Stations are included only if they came online in or before 1950, were still online at the time of the 2021 heatwave, and have at most 20% of observations missing. This leaves the majority of stations in Oregon (62) and Washington (61). While there are fewer in British Columbia (17), the majority are situated in the region most affected by the heatwave, i.e., the south of the province close to the Washington border.

The heatwave epicentre is clearly shown in Figure 1. Areas including Portland and The Dalles recorded some of the highest temperatures, with the highest observation across all stations being 47.8°C at The Dalles Municipal Airport (lat 45.62, long −121.16) on 28 June 2021. This date appears to correspond to the peak of the heatwave, with station-wide average temperature of 39.6°C. While temperatures at coastal stations are lower than at the epicentre, extremity is defined relative to local climate norms and so measurements from these stations are included in the analysis when the temperature exceeds the station-specific threshold. All temperatures are adjusted by subtracting the average of Global mean surface temperature (GMST) from the previous 4 years (Lenssen et al., 2019). GMST represents temperature deviations from the 1951 to 1980 means, and is used to remove the effects of global climate change.

3.1 Annual maxima analysis

We start by considering annual maxima. The GEV is fitted to the adjusted maxima using maximum-likelihood estimation. Following Philip et al. (2021), we fit to pre-2021 data only and use the resulting model to assess the likelihood of the 2021 temperatures being predicted prior to the heatwave.

From the GEV model fits, it is immediately clear why there is concern: 41% of the stations have a 2021 annual maximum which exceeds the finite upper end-point implied by the GEV model fitted to pre-2021 data, i.e., these 2021 observations are apparently infeasible. Even more concerning is that this occurs despite the adjustments for GMST to account for global climate variation—without this adjustment, 48% of stations exhibit the same behaviour. Note that these are relatively coarse statistics—the upper limit can only be estimated at locations for which $\xi < 0$ and the

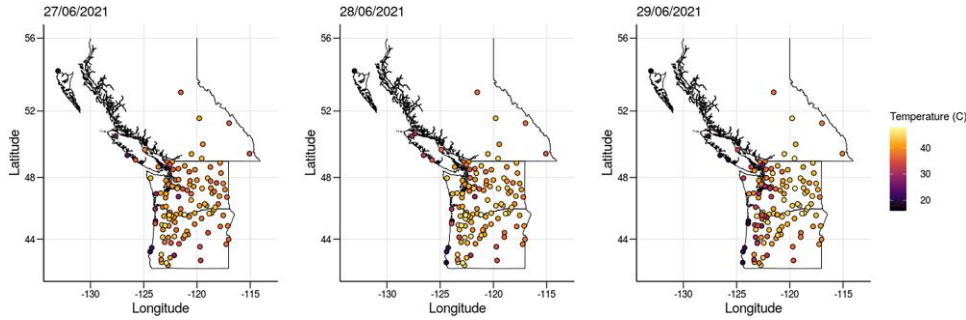

Figure 1. Maximum temperature observations for weather stations in Oregon (South), Washington (Centre), British Columbia (North) on 27 June 2021, 28 June 2021, and 29 June 2021. Stations with missingvalue(s) on these days are omitted.

uncertainty in the estimated limit has not been taken into consideration—however they suggest that climate variation, at least as measured by GMST, does not fully explain the variation in annual maxima, and indeed may play only a small role. Further, annual maxima only capture the temperature on a single day and are not therefore capable of informing a broader understanding of shifts in frequency and magnitude of extreme heatwaves. Such events are likely to be influenced by many more factors than simply the overall climate trends of the region. Capturing all such factors is made difficult by the complexity of meteorological and climate processes and incomplete recording of relevant physical variables.

3.2 Peaks over threshold analysis

Progressing to the PoT model, GMST adjusted data is used with station-specific 97.5% quantile to give year- and station-specific thresholds. This results in between 419 and 639 pre-2021 threshold exceedances at each station—considerably more data than the 71 annual maxima. To assess the extremity of the 2021 event under the PoT model fitted to pre-2021 data, Figure 2 shows predicted return periods associated with the 2021 annual maxima. The estimated period is finite at all stations, with a maximum of 559 years (Estevan Point, British Columbia). The return period is greater than 5 (10) years for 24.2% (8.3%) of the stations. Thus the PoT analysis suggests that the 2021 maximum temperatures are extreme, but not as extreme as the GEV fit implies. However, while the return periods determine the rarity of a single observation, they do not quantify whether the heatwave itself was exceptionally extreme.

The PoT random effects model uses the same thresholds as the vanilla PoT model. As discussed earlier, the model cannot predict the extremal behaviour in a specific year and therefore, in contrast to the vanilla fit, we include the 2021 data when estimating the parameters. A simplified version of the model described in equations (2) and (3) was used with only the scale parameter allowed to vary between years. The justification for a constant shape parameter is that estimation of a varying shape parameter requires so much information that it is common practice to model it as either constant or piece-wise constant. A varying rate parameter was found to be unnecessary due to the year-specific threshold. Each site was modelled separately.

To ensure that the scale parameter estimates are within the parameter space, we use a log link function:

$$\tau_i = \log \psi_i \sim \text{Normal}(\theta_{\psi,0}, \theta_{\psi_1}).$$

We used independent Normal(0, 20) priors for ξ, $\theta_{\psi,0}$, and $\theta_{\psi,1}$. The adaptive MCMC algorithm was run for 100,000 iterations at each site with a burn-in of 10,000. The output was thinned by retaining only every 50th point in each chain. Trace plots and posterior distributions at a random selection of sites were examined for mixing and convergence. An example of these plots is shown for the shape and scale parameter estimates for the Cedar Lake station in Figure A.1 of the appendix.

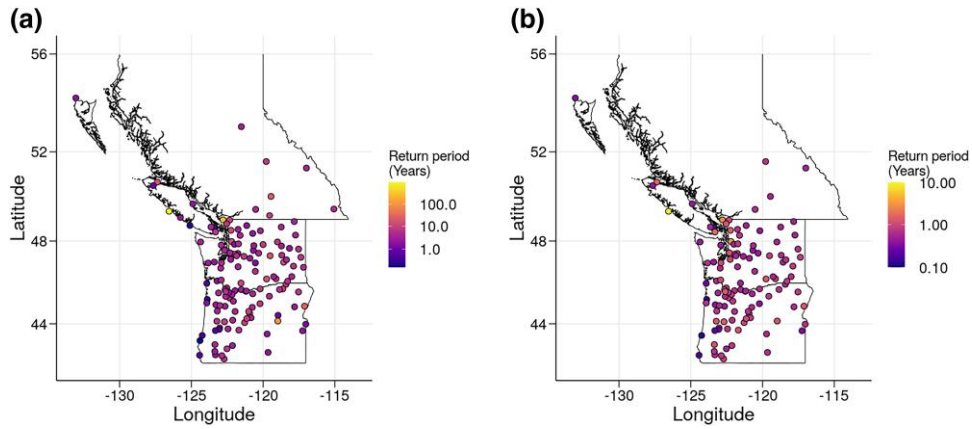

Figure 2. (a) Return periods of the 2021 maxima at each weather station using the fitted eneralized Pareto distributions (GPDs), and (b) Return periods of the 2021 maxima at each weather station using the fitted random effects PoT model. Sixteen stations are not included as their 2021 maxima are below their 97.5% quantile.

Figure 3 shows the estimated scale parameter random effects by station and by year, highlighting years with relatively high extreme temperatures. Note that the inter-year variability in the random effects does not show evidence of a long-term trend, suggesting that the unidentified drivers which cause this variability are unlikely to themselves have long-term trends that are substantially different to that seen in GMST. The random effects in a given year are reasonably consistent across stations, with a maximum cross-station standard deviation of 0.28. The overall range is (− 1.301, 1.073) with 17 of the top 20 estimates occurring in 2021.

The random effects give a much clearer picture of the relative magnitude of the 2021 heatwave. Considering inter-year variability much reduces the return periods of each 2021 maxima for those that exceed the threshold (Figure 3), and in particular, the cross-station mean of the random effects is highest in 2021 (0.468) compared with the next highest (0.250) in 1961. The highest variance in random effects (0.270) also occurs in 2021, perhaps due to different parts of the region experiencing the heatwave to varying degrees, with extremely high relative temperatures at some stations and more typical temperatures at others, and while previous years have isolated random effects at individual stations that reach the same levels as 2021, consistency of the high random effects in 2021 distinguishes it from previous warm years, e.g., 1961.

4 Case Study 2: Greenland ice surface temperatures

One impact of the rise in global temperatures caused by climate change has been an increased volume of melt water generated by the world's ice sheets. This has led to a sea level rise that has already increased the risk of flooding in low-lying areas and will continue to do so while temperatures continue to rise. The two largest ice sheets, Greenland and the Antarctic, have had an accelerating contribution to sea level rise since the early 1990s (Rignot et al., 2011), and are likely to continue to play a key role in the near future (Rahmstorf et al., 2017). Since melt occurs when the ice surface temperature exceeds 0°C, understanding the spatial extent, frequency, magnitude and trends of such events is vital to quantify melt risk both now and in the future, not least because the consequences of increased melt are likely to persist for years to come (Kulp & Strauss, 2019).

Spatial extreme value models provide an ideal tool to analyse the widespread melt events that are of most concern. This case study developed from what was initially intended as a routine first step in fitting such a model to a spatio-temporal Ice Surface Temperature (IST) data set from a multilayer Greenland Moderate Resolution Imaging Spectroradiometer (MODIS)-based product (Hall et al., 2018), the MODIS/Terra Sea Ice Extent 5-Min L2 Swath 1 km, version 6 (MOD29). The spatial resolution of the data is 0.78 km × 0.78 km over the period 01 January 2001 to 31 December 2019, and the temporal resolution is daily observations. The IST data product uses a cloud mask to filter observations that occur in cloudy conditions since water vapour in the clouds

can interfere with infrared radiation leading to measurement inaccuracies. Since cloudy days are generally warmer than clear days (Koenig & Hall, 2010), the data is missing not at random. Consequently, we do not attempt imputation of missing values and instead limit the inferences from our model to clear days only. The data are unseasonalized throughout the analysis.

Since melt occurs when the IST exceeds 0°C, positive temperatures are of most concern. For the majority of cells, particularly inland and/or at higher altitudes, there are too few positive observations to take 0°C as the PoT modelling threshold. However, when taking a modelling threshold below 0°C, there are problems with the PoT fit, as seen in Figure 4. Many cells have a secondary mode corresponding to a large mass of observations close to 0°C and a much lower-weighted tail covering small positive temperature values. The most likely explanation of this is that once the temperature exceeds 0°C the ice melts. This change in state from ice to melt water, which is no longer considered the surface of the ice sheet, creates a soft upper limit close to 0°C on ISTs. We cannot determine from the data which of the positive observations are water temperatures and which are due to measurement or recording uncertainty (±1°C Hall et al., 2018). Regardless of the reason, the soft upper limit results in a distribution mode close to 0°C, rendering the generalized Pareto distribution unsuitable. Lastly, we note that the shape of the tail distribution, and especially the behaviour at or near 0°C, varies primarily in response to altitude and proximity to the coast.

The objective now is to build a model to reflect these insights and establish how the melt threshold impacts the distribution of ISTs. To facilitate ease of application at all locations, we propose a truncated Gaussian mixture model for the full distribution of ISTs. This circumvents the need to define a tail region—as discussed previously, for this particular data set such a definition is non-trivial due to the varying cross-site behaviour of the upper tail. Utilizing information on behaviour in sub-tail regions should help better identify behaviour close to the melt threshold. Furthermore, a mixture model allows joint modelling of the ice and melt observations; this is critical since we have no information on which category each observation belongs to.

The truncated Gaussian mixture model has separate model components for ice and melt temperatures. Observations are not preassigned to a component, rather the fit provides an estimate of the probability with which each observation belongs to each component. For n_I ice components and a single melt component, let ϕ_i be the weight associated with ice component i and ϕ_M be the weight associated with the melt component, such that $\sum_{i=1}^{n_I} \phi_i + \phi_M = 1$. Let $f_i(x)$ and $f_M(x)$ be the probability density functions of the ice $i \in \{1, \ldots, n_I\}$ and melt components, respectively, such that for each i, $X \sim TN(\mu_i, \sigma_i^2, a_i, b_i)$, where μ_i is the mean, $\sigma_i > 0$ the standard deviation, and a_i and b_i the lower and upper truncation points. Then the mixture model probability density function of IST x is:

$$p(x) = \sum_{i=1}^{n_I} \phi_i f_i(x) + \phi_M f_M(x). \tag{5}$$

Ice temperature components are bounded above by $b = 0$, since these should not exceed 0°C, and have no lower bound since the only physical lower limit is absolute zero, which is far below the range of the data. For the melt component, there is no upper bound as ISTs are not feasibly physically limited on the ice sheet. The lower bound is set to −1.65 as this has previously been estimated as the melting point of saline ice and thereby acts as a conservative minimum estimate for the ice sheet. It follows that data in the range $[-1.65, 0]$ can be modelled by either the ice or melt components, capturing the uncertainty in the true nature of the ice sheet when these observations were measured. The probability of melt $\rho(x)$ for given IST x is defined as the ratio of the densities of the ice and melt components, such that:

$$\rho(x) = \frac{\phi_M f_M(x)}{\phi_M f_M(x) + \sum_{i=1}^{n_I} \phi_i f_i(x)}. \tag{6}$$

As a result of the model truncation points, $\rho(x) = 0$ below −1.65°C and $\rho(x) = 1$ above 0°C. ISTs between these values vary smoothly within this range. Small discontinuities in the melt probability may arise at −1.65°C and 0°C due to the truncation of the mixture components at these points; the

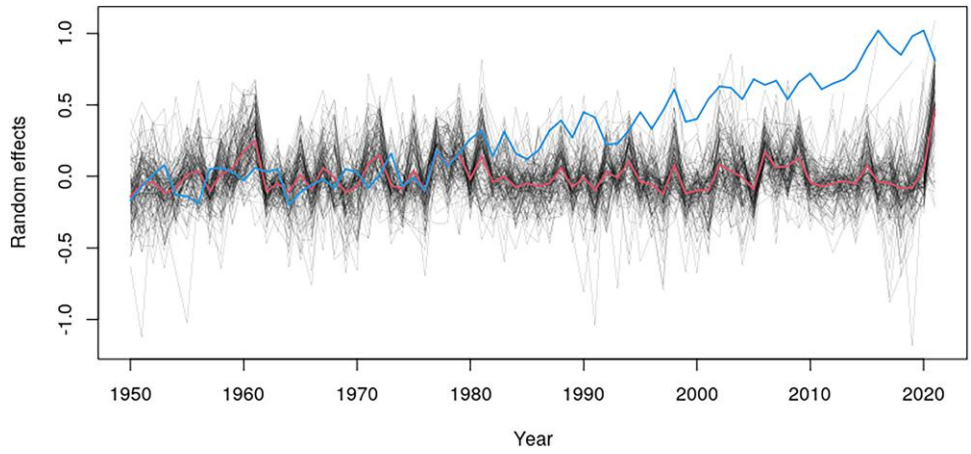

Figure 3. Estimated scale parameter annual random effects for each weather station (thin black lines), the annual mean for all stations (central bold red line), and GMST (bold blue line, in degrees). Results cover the period 1950 to 2021 inclusive.

magnitude of the discontinuity depends on the severity of the truncation. To find the most appropriate number of ice components for the model and allow for the diversity of distributions across the ice sheet, we fit the model with n_I from 3 to 6 and compare the fits using Bayesian information criterion (BIC). In contrast, a single melt component is used since, in theory, the melt temperatures should show much less cross-cell variability than the ice temperatures.

We use the Expectation-maximization (EM) algorithm to estimate the parameters ϕ, μ and σ^2 for each model component. We found 800 iterations of the algorithm optimal in terms of both convergence and computational time. To improve the stability of the algorithm, particularly for cells with few or no observations above -1.65°C, we set additional restrictions on two parameters: $\phi \geq 0.0005$ and $\sigma \geq 0.35$. This allows a melt component—however small—to be fitted even if there are no observations in the melt temperature range, and avoids melt components with $\sigma \approx 0$ since this results in a point-mass density for cells with only a single observation above -1.65°C. We apply the model to 1,139 cells across the ice sheet; this represents a subsample of 1 in every 50 available cells in both horizontal and vertical directions. The cells are equally spaced in distance and are taken from those cells identified as ice by the MODIS ice mask. The number and spread of the sub-sampled cells accurately reflects the variety in glaciological and climatological conditions across the ice sheet. A range of average temperatures, surface conditions, latitudes, longitudes, elevations, and distances to the coast are all observed within the sample.

The impact of the soft upper limit on ISTs can be seen from the number of ice components selected by the BIC to model the data at different cells. Figure 5 shows the number of ice components in the best-fitting model to be higher at cells close to the coastline than those closer to the centre of the ice sheet. There are also trends within coastal areas, with cells on the south west coast generally having the highest number of model components and areas in the north west coast having fewer. Cells closer to the coast generally have higher average temperatures, experience more melt and therefore have more frequently truncated observations. Furthermore, melt estimates from the model agree with empirical estimates of melt well using previously established methods of identifying melt with a single threshold (Clarkson et al., 2022). This coastal temperature trend is consistent with local climate trends of higher temperatures occurring closer to the coast and demonstrates the impact of the upper limit from the truncated observations. This also demonstrates the difficulty mentioned above of finding a single distribution to describe the data at all locations of the ice sheet.

While there is value in examining how each parameter varies across the ice sheet, to understand melt it is most useful to consider variability and trends in the weight of the melt component. This gives an overview of melt observed across the ice sheet, and reflects broad climate trends and the coastal effect. Figure 5 shows coastal areas having more of their distribution associated with melt temperatures, with cells in the south having a slightly higher weight than those in the north. Melt

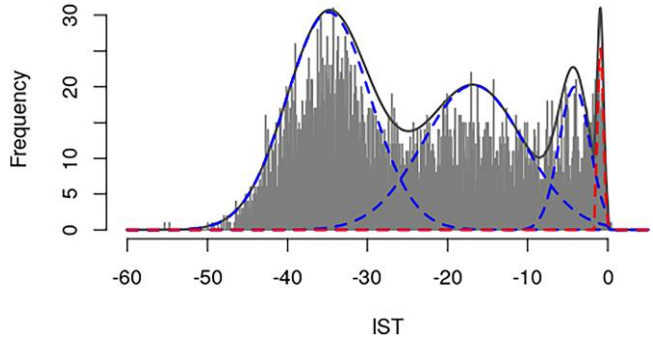

Figure 4. Example mixture model fit for a single cell (82.47N, −37.50E). Histogram of the data with 3 ice component densities (dotted blue lines), melt component density (dotted red line), and the overall density (bold black line). Density function has been scaled to the maximum value of the histogram data.

probability (6) is estimated for each observation and used to calculate the expected number of melt observations at each cell. The cross-cell average is then used to estimate the number of melt observations across the ice sheet in each year. These values are likely to be under-estimates since the expected melt at each cell would be higher at almost all cells if it were possible to also include cloudy days in the analysis. Figure 6 shows the majority (90.7%) of cells have a nontrivial expectation of melt (>0.001) in an average year. Although melt is more common on the coasts, it is clear from the expected melt estimates that this is not exclusive with some inland cells also experiencing melt with relatively high frequency, and the main factors that appear to determine the melt extent are distance to the coast and elevation.

5 Discussion

Extreme value analysis has a long-established link with hazard modelling, risk management and climate change analysis. While historically focus has been on the tail behaviour of univariate responses, more recently considerable attention has been given to advancing extreme value theory and models for multivariate responses, due to (a) an increasing awareness of the potential for impact from multiple hazards and (b) an explosion in the availability of high-dimensional spatio-temporal data products, e.g., satellite images and numerical climate model output. Analysis of the extreme events of such data sets raises many challenges, from defining and identifying jointly extreme events to developing both statistical models and inference tools and the probability theory which underpins them. In this paper, we step back to assess ongoing challenges in univariate extreme value modelling. Such challenges are not restricted to model development, and it is our opinion that for them to be met successfully requires sustained dialogue between statisticians and fellow scientists. Evaluating impacts of climate change on natural hazards requires an understanding of the physical behaviour of both the hazard and any driving processes. Only once this is appropriately incorporated into the statistical model can attribution and prediction of the effects of climate change begin.

The two analyses presented in this paper demonstrate the necessity for novel, bespoke, context-driven models to address questions driven by climate change. While IST and air temperature data sets share some similarities—both measure the same physical variable on or near the Earth's surface—the most suitable modelling approach for each is quite different due to differences in the physical behaviour of the processes and their drivers, the data collection methods and the underlying research questions. For ISTs, the physical properties of both the ice sheet and the melt process drive the model. Seemingly straightforward tasks—such as defining melt—are nontrivial and, despite representing the tail of the sample, the distribution of temperatures around the melt point does not comply with the assumptions necessary for an out-of-the-box extreme value analysis. While the proposed Gaussian mixture model is not consistent with more routine methods for modelling tail data, it does incorporate an understanding of the melt process that cannot be included in a PoT model. In contrast, the core challenge in the US air temperature study arises from the 2021

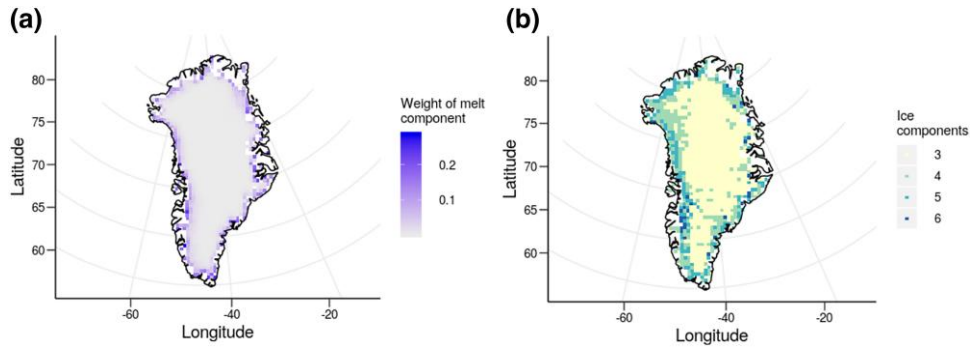

Figure 5. (a) Estimated weight $\hat{\phi}$ of the melt component of each fitted mixture model. Due to restrictions during the fitting process the minimum possible value of this estimate is 0.0005. (b) The number of ice components in the best-fitting mixture model at each cell in our study area.

heatwave which, despite accounting for global temperature trends, cannot be explained through an extreme value analysis of historical data. To improve our understanding of the extremity of the 2021 event, a PoT random effects model is used to give a better indication of the relative size of the 2021 heatwave in the context of previous extreme temperatures, climate trends, and inter-year variability due to unobserved climate drivers.

All statistical models are a compromise between model parsimony and functionality. Further, the scope for generalization from any statistical model is entirely dependent on the quality and quantity of the data to which the model is fitted. With ever more powerful computing power at our disposal it can be tempting to develop high-dimensional models that aim to mimic their process-based counterparts, regardless of the size and quality of the data set. It is our belief that parsimony is in fact to be preferred, and that a key advantage of statistical models is their comparative cheapness to fit and ease of interpretation. At the same time, some compromise may be desirable; where incorporation of additional model features significantly improves interpretation, predictive abilities or our understanding of the physical process, then such features should be included where the information in the data supports this. Ultimately the utility of the model should be assessed by its ability to describe the data set, with both the most suitable modelling strategy and the necessary level of complexity judged against this benchmark.

The models proposed here are no exception, having both advantages and limitations. The random effects model provides a more flexible description of the data than permitted under a parametric regression model. However, unlike a parametric regression model, it does not permit estimation of the extreme behaviour in a specific year. It further assumes that, conditional on the annual random effects, the observations are independent. In practice, there is likely to be some short-range serial dependence in the largest values. The model could be extended to incorporate serial dependence in either, or both, of the random effects and observations; the consequences of this would be an increased number of model parameters and a much more complex inference process. The mixture model could be criticized for diverging from the traditional approach of analysing only the tail data. Preliminary investigations using this approach provided such a poor fit to the data that it was immediately clear that an alternative approach—taking into account the unusual shape of the tail distribution—was required. The resulting model provides a much better description of the upper tail and, indeed, is a good fit to the full range of values. Neither the random effects nor the mixture model accounts for the inherent spatial structure in the physical processes and while this is beyond the scope of the current research it is the subject of ongoing research by the authors.

We end by discussing some broader open questions for the application of extreme value analysis in climate change scenarios. The first of these concerns guidance for scientists on the most appropriate way to incorporate change in either univariate or multivariate models. We conjecture that it is unlikely that there will ever be a definitive way to do this due to the diverse physical properties exhibited by different environmental hazards. Instead, statisticians should strive for an increased

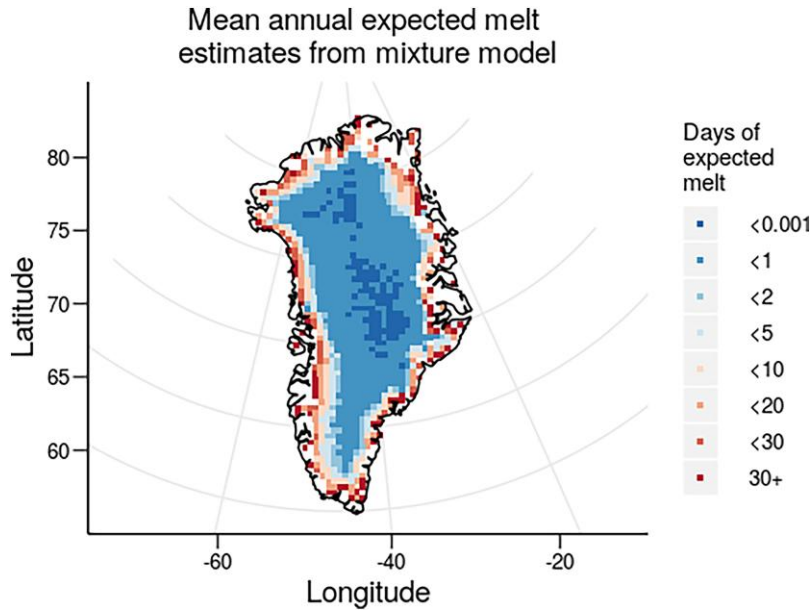

Figure 6. Expected melt calculated from the fitted mixture models. The probability of each observation representing a melt temperature is calculated then averaged to the expectation in a single year.

awareness of the appropriateness, advantages and limitations of the available methods through direct comparisons of these different strategies. Such comparisons should also support an increased understanding of the benefits of trialling multiple modelling approaches.

Secondly, extreme value models are asymptotically justified in the sense of extrapolation into the tail of an underlying distribution. Most existing approaches to incorporating change assume the model parameters vary in time and/or in response to driving physical processes. Such regression-based models, whether semi- or fully parametric, have no similar justification for the extrapolation of trends into the future. Further, the predictions of future events will be very sensitive to the form assumed for the change. Again, this is strong motivation to search for more reliable, robust methods.

Thirdly, most routine multivariate, spatio- and temporal extreme value models require a separate model for the univariate marginal distributions. It is therefore critical that these marginal models reflect as faithfully as possible both the empirical properties of the data and physical properties of the underlying processes. As in the case of the ice surface temperatures discussed in Section 4, this might mean reconsidering the use of an extreme value model for the marginal distributions.

Finally, much work remains to identify risk measures which still have an interpretation under climate change; return levels (periods) make less sense if there is a long-term trend—what is deemed extreme now may no longer be extreme in 100 years. Return levels and periods which condition on a time period or covariates (Eastoe & Tawn, 2009) are more realistic but harder to interpret. Related to this is the problem of engaging with fellow scientists and the general public on the topic of changes in the risk of extreme events. It is vital to emphasize that acknowledgement of such change should not imply a loss of credibility in the predictions, but rather a more accurate reflection of a changing world.

Acknowledgments

We thank Paul Young for an introduction to the US temperature data, and David Leslie and Rachael Duncan for helpful comments on the draft. D.C., E.E. and A.L. were supported by the EPSRC-funded Data Science of the Natural Environment project (EP/R01860X/1).

Conflict of interest: None declared.

Appendix. MCMC trace plots

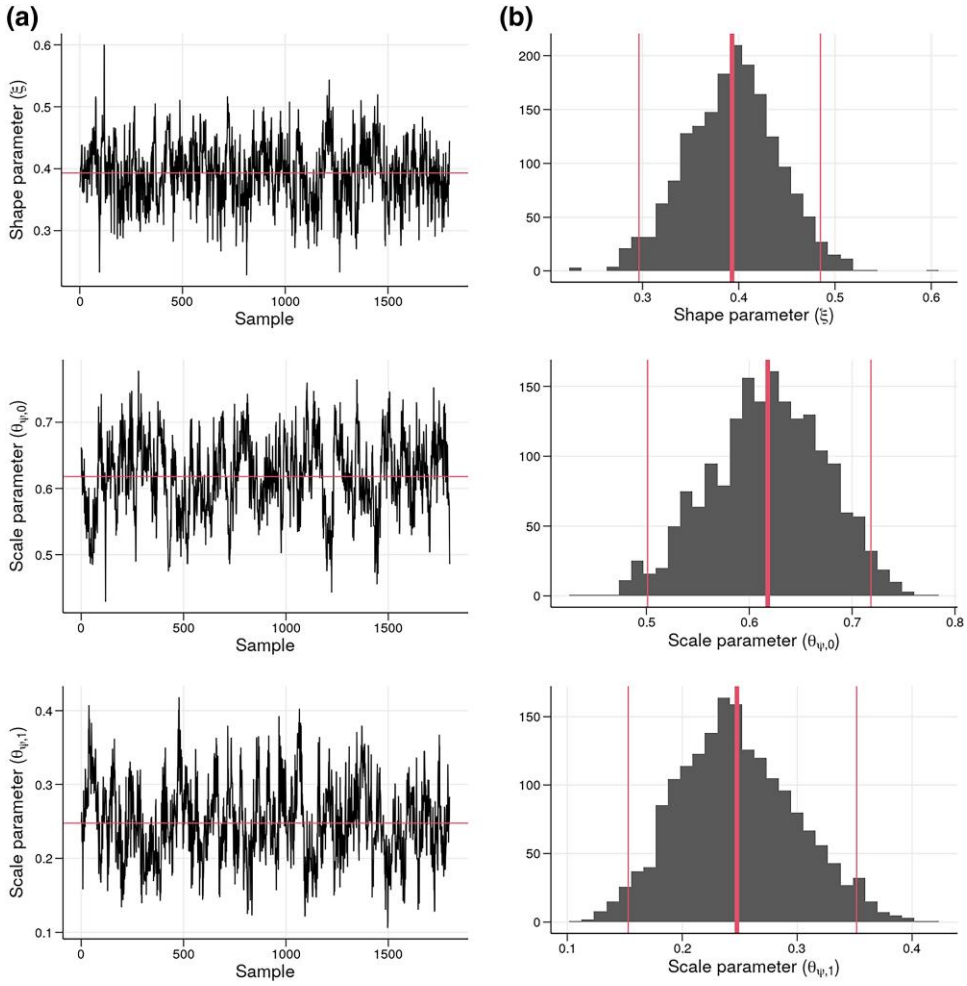

Figure A.1. (a) Trace plot of the shape parameter for weather station 'Cedar Tree' with the mean of the estimates (red horizontal line). (b) Histogram of the sampled shape parameters from the same MCMC chain as the trace plot, with the mean (bold red line) and 2.5% and 97.5% quantiles (normal red lines).

References

Acero F., García J. A., Gallego M. C., Parey S., & Dacunha-Castelle D. (2014). Trends in summer extreme temperatures over the Iberian Peninsula using nonurban station data. *Journal of Geophysical Research: Atmospheres*, *119*(1), 39–53. https://doi.org/10.1002/2013JD020590

Burke E. J., Perry R. H., & Brown S. J. (2010). An extreme value analysis of uk drought and projections of change in the future. *Journal of Hydrology*, *388*(1–2), 131–143. https://doi.org/10.1016/j.jhydrol.2010.04.035

Cheng L., AghaKouchak A., Gilleland E., & Katz R. W. (2014). Non-stationary extreme value analysis in a changing climate. *Climatic change*, *127*(2), 353–369. https://doi.org/10.1007/s10584-014-1254-5

Clarkson D., Eastoe E., & Leeson A. (2022). Melt probabilities and surface temperature trends on the Greenland ice sheet using a Gaussian mixture model. *The Cryosphere*, *16*(5), 1597–1607. https://doi.org/10.5194/tc-16-1597-2022

Cooley D. (2009). Extreme value analysis and the study of climate change. *Climatic Change*, *97*(1), 77–83. https://doi.org/10.1007/s10584-009-9627-x

Cooley D., Cisewski J., Erhardt R. J., Jeon S., Mannshardt E., Omolo B. O., & Sun Y. (2012). A survey of spatial extremes: Measuring spatial dependence and modeling spatial effects. *Revstat*, *10*(1), 135–165. https://doi.org/10.57805/revstat.v10i1.114

Davis R. A., Mikosch T., & Zhao Y. (2013). Measures of serial extremal dependence and their estimation. *Stochastic Processes and their Applications*, *123*(7), 2575–2602. https://doi.org/10.1016/j.spa.2013.03.014
Davison A. C., & Smith R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, *52*(3), 393–425. https://doi.org/10.1111/j.2517-6161.1990.tb01796.x
Drees H. (2008). Some aspects of extreme value statistics under serial dependence. *Extremes*, *11*(1), 35–53. https://doi.org/10.1007/s10687-007-0051-1
Eastoe E. F. (2019). Nonstationarity in peaks-over-threshold river flows: A regional random effects model. *Environmetrics*, *30*(5), e2560. https://doi.org/10.1002/env.2560
Eastoe E. F., & Tawn J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *58*(1), 25–45. https://doi.org/10.1111/j.1467-9876.2008.00638.x
Economou T., Stephenson D. B., & Ferro C. A. (2014). Spatio-temporal modelling of extreme storms. *The Annals of Applied Statistics*, *8*(4), 2223–2246. https://doi.org/10.1214/14-AOAS766
Fawcett L., & Walshaw D. (2006). A hierarchical model for extreme wind speeds. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *55*(5), 631–646. https://doi.org/10.1111/j.1467-9876.2006.00557.x
Friederichs P. (2010). Statistical downscaling of extreme precipitation events using extreme value theory. *Extremes*, *13*(2), 109–132. https://doi.org/10.1007/s10687-010-0107-5
Gnedenko B. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, *44*(3), 423–453. https://doi.org/10.2307/1968974
Gouldsborough L., Hossain R., Eastoe E., & Young P. (2021). A temperature dependent extreme value analysis of UK surface ozone, 1980–2019. Submitted.
Gumbel E. J. (1958). *Statistics of extremes*. Columbia University Press.
Hall D. K., Cullather R. I., DiGirolamo N. E., Comiso J. C., Medley B. C., & Nowicki S. M. (2018). A multilayer surface temperature, surface albedo, and water vapor product of greenland from modis. *Remote Sensing*, *10*(4), *555* https://doi.org/10.3390/rs10040555
Hall P., & Tajvidi N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, *15*(2), 153–167. https://doi.org/10.1214/ss/1009212755
Heffernan J. E., & Tawn J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x
Huser R., & Wadsworth J. L. (2019). Modeling spatial processes with unknown extremal dependence class. *Journal of the American Statistical Association*, *114*(525), 434–444. https://doi.org/10.1080/01621459.2017.1411813
Joe H. (1994). Multivariate extreme-value distributions with applications to environmental data. *Canadian Journal of Statistics*, *22*(1), 47–64. https://doi.org/10.2307/3315822
Jonathan P., & Ewans K. (2013). Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering*, *62(1)*, 91–109. https://doi.org/10.1016/j.oceaneng.2013.01.004
Jones M., Randell D., Ewans K., & Jonathan P. (2016). Statistics of extreme ocean environments: Non-stationary inference for directionality and other covariate effects. *Ocean Engineering*, *119(1)*, 30–46. https://doi.org/10.1016/j.oceaneng.2016.04.010
Katz R. W. (1999). Extreme value theory for precipitation: Sensitivity analysis for climate change. *Advances in Water Resources*, *23*(2), 133–139. https://doi.org/10.1016/S0309-1708(99)00017-2
Katz R. W., Parlange M. B., & Naveau P. (2002). Statistics of extremes in hydrology. *Advances in water resources*, *25*(8–12), 1287–1304. https://doi.org/10.1016/S0309-1708(02)00056-8
Koenig L. S., & Hall D. K. (2010). Comparison of satellite, thermochron and air temperatures at Summit, Greenland, during the winter of 2008/09. *Journal of Glaciology*, *56*(198), 735–741. https://doi.org/10.3189/002214310793146269
Kulp S. A., & Strauss B. H. (2019). New elevation data triple estimates of global vulnerability to sea-level rise and coastal flooding. *Nature Communications*, *10*(1), 4844. https://doi.org/10.1038/s41467-019-12808-z
Laurini F., & Tawn J. A. (2003). New estimators for the extremal index and other cluster characteristics. *Extremes*, *6*(3), 189–211. https://doi.org/10.1023/B:EXTR.0000031179.49454.90
Ledford A. W., & Tawn J. A. (2003). Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *65*(2), 521–543. https://doi.org/10.1111/1467-9868.00400
Lenssen N., Schmidt G., Hansen J., Menne M., Persin A., Ruedy R., & Zyss D. (2019). Improvements in the gistemp uncertainty model. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6307–6326. https://doi.org/10.1029/2018JD029522
Menne M. J., Durre I., Vose R. S., Gleason B. E., & Houston T. G. (2012). An overview of the global historical climatology network-daily database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897–910. https://doi.org/10.1175/JTECH-D-11-00103.1

Philip S. Y., Kew S. F., van Oldenborgh G. J., Yang W., Vecchi G. A., Anslow F. S., Li S., Seneviratne S. I., Luu L. N., Arrighi J., Singh R., van Aalst Hauser, Marghidan C. P., Ebi K. L., Bonnet R., Vautard R., Tradowsky J., Coumou D., Lehner F., … Otto F. E. L. (2022). Rapid attribution analysis of the extraordinary heatwave on the Pacific Coast of the US and Canada June 2021. Earth System Dynamics, *13(4)*, 37. https://doi.org/10.5194/esd-13-1689-2022

Pickands J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*(1), 119–131. https://doi.org/10.1214/aos/1176343003

Rahmstorf S., Foster G., & Cahill N. (2017). Global temperature evolution: Recent trends and some pitfalls. *Environmental Research Letters*, *12*(5), 054001. https://doi.org/10.1088/1748-9326/aa6825

Ramos A., & Ledford A. (2009). A new class of models for bivariate joint tails. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *71*(1), 219–241. https://doi.org/10.1111/j.1467-9868.2008.00684.x

Reich B., Cooley D., Foley K., Napelenok S., & Shaby B. (2013). Extreme value analysis for evaluating ozone control strategies. *The Annals of Applied Statistics*, *7*(2), 739. https://doi.org/10.1214/13-AOAS628

Rignot E., Velicogna I., Monaghan A., & Lenaerts J. T. M. (2011). Acceleration of the contribution of the Greenland and Antarctic ice sheets to sea level rise. *Geophysical Research Letters*, *38*(5), L05503. https://doi.org/10.1029/2011GL047109

Roberts G. O., & Rosenthal J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, *18*(2), 349–367. https://doi.org/10.1198/jcgs.2009.06134

Rogers N. C., Wild J. A., Eastoe E. F., Gjerloev J. W., & Thomson A. W. (2020). A global climatological model of extreme geomagnetic field fluctuations. *Journal of Space Weather and Space Climate*, *10*(1), 5. https://doi.org/10.1051/swsc/2020008

Rootzén H., Segers J., & Wadsworth J. L. (2018). Multivariate peaks over thresholds models. *Extremes*, *21*(1), 115–145. https://doi.org/10.1007/s10687-017-0294-4

Simpson E. S., & Wadsworth J. L. (2021). Conditional modelling of spatio-temporal extremes for Red Sea surface temperatures. *Spatial Statistics*, *41*(1), 100482. https://doi.org/10.1016/j.spasta.2020.100482

Smith R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, *4*(4), 367–377. https://doi.org/10.1214/ss/1177012400

Smith R. L., & Weissman I. (1994). Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(3), 515–528. https://doi.org/10.1111/j.2517-6161.1994.tb01997.x

Sterl A., Severijns C., Dijkstra H., Hazeleger W., van Oldenborgh G. J., Burgers G., van Leeuwen P. J., & van Velthoven P. (2008). When can we expect extremely high surface temperatures? *Geophysical Research Letters*, *35*(14), 1-5. https://doi.org/10.1029/2008GL034071

Thomson A. W., Dawson E. B., & Reay S. J. (2011). Quantifying extreme behavior in geomagnetic activity. *Space Weather*, *9*(10), S10001. https://doi.org/10.1029/2011SW000696

Towe R., Eastoe E., Tawn J., & Jonathan P. (2017). Statistical downscaling for future extreme wave heights in the North Sea. *The Annals of Applied Statistics*, *11*(4), 2375–2403. https://doi.org/10.1214/17-AOAS1084

Towe R., Tawn J., Eastoe E., & Lamb R. (2020). Modelling the clustering of extreme events for short-term risk assessment. *Journal of Agricultural, Biological and Environmental Statistics*, *25*(1), 32–53. https://doi.org/10.1007/s13253-019-00376-0

Von Mises R. (1964). La distribution de la plus grande de n valeurs. In Ph. Frank, S. Goldstein, M. Kac, W. Prager, G. Szegö, & G. Birkhoff (Eds), *Selected papers of Richard von Mises: Volume II. Probability and statistics, general*. Providence, RI: American Mathematical Society.

Winter H. C., & Tawn J. A. (2017). kth-order Markov extremal models for assessing heatwave risks. *Extremes*, *20*(2), 393–415. https://doi.org/10.1007/s10687-016-0275-z

Winter H. C., Tawn J. A., & Brown S. J. (2016). Modelling the effect of the El Nino-Southern Oscillation on extreme spatial temperature events over Australia. *The Annals of Applied Statistics*, *10*(4), 2075–2101. https://doi.org/10.1214/16-AOAS965

Yee T. W., & Stephenson A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes*, *10*(1), 1–19. https://doi.org/10.1007/s10687-007-0032-4

Youngman B. D. (2019). Generalized additive models for exceedances of high thresholds with an application to return level estimation for US wind gusts. *Journal of the American Statistical Association*, *114*(528), 1865–1879. https://doi.org/10.1080/01621459.2018.1529596

Journal of the Royal Statistical Society Series C: Applied Statistics, 2023, **72**, 844–871

Journal of the Statistics Society
Series C
Applied Statistics

Discussion Paper Contribution

Proposer of the vote of thanks and contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Adrian E. Raftery

Department of Statistics, University of Washington, Seattle, USA

Address for correspondence: Adrian E. Raftery. Department of Statistics, Box 354322, University of Washington, Seattle, WA 98195-4322, USA. Email: raftery@uw.edu

Professor Adrian E. Raftery (University of Washington): It is a pleasure to congratulate the authors of these two papers on the important topic of statistical aspects of climate change. I am happy that Climate Change and the Environment is one of the Royal Statistical Society's campaigning priorities for 2022, both because of the importance of the topic to society, and the vital role that statistics has played and can play in addressing it.

When climate change became a major issue in the 1980s, statistics was not a big part of research on it. However, since then statistics has become vital to climate change research, thanks to the efforts of an extraordinary founding generation of statisticians, including Mark Berliner, Noel Cressie, Peter Guttorp, Douglas Nychka, Richard Smith, Michael Stein, Claudia Tebaldi, their students and collaborators, and others. They developed new statistical methods for climate problems ranging from the integrated assessment of future global climate change to the detailed modelling of granular atmospheric processes.

Some things I learned from this literature include the importance of clearly defining the atmospheric question being asked and choosing the right temporal and spatial scales to answer it, the usefulness of extant physical models even though they are often deterministic, that purely statistical models often are not enough on their own, the value of combining physical models with statistical methods, and the importance of accounting for spatial correlation when the problem demands it. These insights provide a framework for considering the interesting papers we have heard today.

The paper 'Assessing present and future risk of water damage using building attributes, meteorology and topography,' exemplifies the lessons from this earlier statistical climate change literature. The purpose is clear, to provide probabilistic forecasts of future water damage for any building in Norway, both with current climate and accounting for future climate change, for insurance purposes. The quarterly timescale is well chosen, since it corresponds better to the purpose than the often used daily interval. Using a building-specific spatial scale rather than a more aggregated one, also responds well to the needs of the problem.

The approach uses physical models for future climate and statistical models for finer-level processes, which is a good balance. The model represents spatial correlation at best crudely, through the random effects for municipalities. Undoubtedly, more refined spatial modelling would be possible, but it seems unlikely it would contribute much to answer the specific questions at hand, regarding marginal distributions for individual buildings.

The approach is probabilistic, with the main sources of uncertainty well identified and taken into account, and the model well validated. I liked the use of the conditional expectation diagram to assess the model. This is related to reliability diagrams and rank histograms, and all three

Received: June 9, 2023. **Accepted:** June 12, 2023

methods trace their origins to the atmospheric science literature. They provide an appealing alternative approach to the residual-based methods more commonly used in statistics, and statisticians should use them more often. Overall, the approach is 'sophisticatedly simple' in the phrase of Arnold Zellner, and its interpretability and ease of use are major advantages, sometimes undervalued in the statistics profession, which can place a premium on complicated approaches.

All that said, improvements are possible. The approach is conditional on choosing one of the IPCC's RCP deterministic emissions scenarios (which have since been replaced by the SSPs Intergovernmental Panel on Climate Change, 2021). The interpretation of these is unclear, in particular how plausible they are, which is most likely, and to what extent they cover the range of likely outcomes. Fully probabilistic projections of carbon emissions and hence global and local climate change have been developed since (Chen et al., 2022; Liu & Raftery, 2021; Raftery et al., 2017; Rennert et al., 2021; Srikrishnan et al., 2022) and could easily be incorporated into the framework. This is worth considering.

Also, there has been considerable work in the U.S. by government and industry on evaluating the risk of flooding, water damage and wildfire for all individual properties, some leading to online products (Federal Emergency Management Administration, 2022; Porter et al., 2021; Risk Factor, 2022). It would be interesting to compare these with the paper's methodology. This raises the question of whether the present methodology could be adapted to assess the risk of wildfires, which may, admittedly, be less of a concern in Norway.

The second paper, 'The importance of context in extreme value analysis with application to extreme temperatures in the U.S. and Greenland' describes two case studies where standard extreme value analysis seems inadequate, and proposes alternatives. The first is the 2021 heatwave in the Pacific Northwest. As it happens, I was there at the time, and both from the data and personal experience, there is no doubt that the temperatures were far beyond anything previously experienced. The authors propose adding station-specific random effects to the Peaks-over-Thresholds model and show that it gives a marginal predictive distribution of the 2021 extrema that is more in line with observations.

A difficulty in considering this analysis is that the question being asked is not made clear. It is about modelling the marginal distributions of maximum temperatures, but for what purpose? For forecasting? For insurance purposes? For climate adaptation policy-making? Each of these will lead to different modelling considerations.

One issue is that this is a marginal analysis, and spatial correlation is not considered. This may be reasonable for some purposes, such as assessing the risk of high temperatures at an individual location. For other purposes, such as those that involve responses beyond the level of an individual location, accounting for spatial correlation would seem important, and the statistical climatology literature is rich in ways of doing so. Could the proposed methodology be adapted to account for spatial correlation?

For forecasting, numerical weather prediction models did forecast the 2021 heatwave quite successfully even though it was outside the range of all previous experience. The random effects model does not do that. The present analysis largely ignores the physics and physical models, and so the results may illustrate the limitations of purely statistical modelling overall. It is not clear that modifying the statistical model without including the physics is the best way to proceed.

The second case study models Greenland ice surface temperatures, and finds that, again, standard extreme value modelling does not perform well. In particular, the distribution has multiple modes, including one just below 0°C. It is proposed to replace standard extreme value modelling by a Gaussian mixture model, which appears to fit the data well.

Again, it was hard to comment in depth on this case study, because the underlying questions were not made explicit. I was puzzled by the decision to model ice surface temperatures at all. In the sea ice literature, it is more standard to model ice thickness (Gao et al., 2021). This is more relevant to practical questions that arise, such as choosing shipping routes (Melia et al., 2017). More generally, why not just model the binary ice/no ice variable? The main interest here seems to be whether the temperature is above or below freezing, rather than specific temperature. This would seem to largely avoid the modelling complications that this paper tries to solve in the first place.

The decision to ignore spatial correlation seems even harder to justify here than in the first case study. It is hard to think of practical questions where the spatial aspect is unimportant. Sea ice typically forms and melts in ice floes and ice sheets, which are inherently spatial. The decision to take a

purely statistical approach and ignore physical models also seems questionable, as there is now a literature in this area and several physical models (Blanchard-Wrigglesworth et al., 2015; Bushuk et al., 2017; Guemas et al., 2016; Zampieri et al., 2018).

There are several statistical approaches to modelling sea ice that do take account of both spatial correlation and physical aspects (Chang et al., 2016; Director et al., 2017, 2021; Dirkson et al., 2019; Lee et al., 2020; Zhang & Cressie, 2019, 2020). It would be interesting to see if they could be combined with the present approach.

Overall, these two papers provide interesting contributions to statistical solutions to current problems in climate science, and it is a pleasure to propose the vote of thanks.

References

Blanchard-Wrigglesworth E., Cullather R. I., Wang W., Zhang J., & Bitz C. M. (2015). Model forecast skill and sensitivity to initial conditions in the seasonal Sea Ice Outlook. *Geophysical Research Letters*, *42*(19), 8042–8048. https://doi.org/10.1002/2015GL065860

Bushuk M., Msadek R., Winton M., Vecchi G. A., Gudgel R., Rosati A., & Yang X. (2017). Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophysical Research Letters*, *44*(10), 4953–4964. https://doi.org/10.1002/2017GL073155

Chang W., Haran M., Applegate P., & Pollard D. (2016). Calibrating an ice sheet model using high-dimensional binary spatial data. *Journal of the American Statistical Association*, *111*(513), 57–72. https://doi.org/10.1080/01621459.2015.1108199

Chen X., Raftery A. E., Battisti D. S., & Liu P. R. (2023). Long-term probabilistic temperature projections for all locations. *Climate Dynamics*, *60*, 2303-2314. https://doi.org/10.1007/s00382-022-06441-8

Director H. M., Raftery A. E., & Bitz C. M. (2017). Improved sea ice forecasting through spatiotemporal bias correction. *Journal of Climate*, *30*(23), 9493–9510. https://doi.org/10.1175/JCLI-D-17-0185.1

Director H. M., Raftery A. E., & Bitz C. M. (2021). Probabilistic forecasting of the Arctic sea ice edge with contour modeling. *The Annals of Applied Statistics*, *15*(2), 711–726. https://doi.org/10.1214/20-AOAS1405

Dirkson A., Merryfield W. J., & Monahan A. H. (2019). Calibrated probabilistic forecasts of Arctic sea ice concentration. *Journal of Climate*, *32*(4), 1251–1271. https://doi.org/10.1175/JCLI-D-18-0224.1

Federal Emergency Management Administration (2022). *Flood resilience for homeowners, renters and business owners*. https://www.fema.gov/flood-maps/products-tools/know-your-risk/homeowners-renters#understand. Downloaded August 13, 2022.

Gao P. A., Director H. M., Bitz C. M., & Raftery A. E. (2021). Probabilistic forecasts of Arctic sea ice thickness. *Journal of Agricultural, Biological and Environmental Statistics*, *27*(2), 280–302. https://doi.org/10.1007/s13253-021-00480-0

Guemas V., Blanchard-Wrigglesworth E., Chevallier M., Day J. J., Déqué M., Doblas-Reyes F. J., Fučkar N. S., Germe A., Hawkins E., Keeley S., & Koenigk T. (2016). A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. *Quarterly Journal of the Royal Meteorological Society*, *142*(695), 546–561. https://doi.org/10.1002/qj.2401

Intergovernmental Panel on Climate Change (2021). *Climate change 2021: The physical science basis. Contribution of Working Group 209 I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Lee B. S., Haran M., Fuller R. W., Pollard D., & Keller K. (2020). A fast particle-based approach for calibrating a 3-D model of the Antarctic ice sheet. *The Annals of Applied Statistics*, *14*(2), 605–634. https://doi.org/10.1214/19-AOAS1305

Liu P. R., & Raftery A. E. (2021). Country-based rate of emissions reductions should increase by 80% beyond nationally determined contributions to meet the 2°C target. *Communications Earth & Environment*, *2*(1), 1–10. https://doi.org/10.1038/s43247-021-00097-8

Melia N., Haines K., Hawkins E., & Day J. J. (2017). Towards seasonal Arctic shipping route predictions. *Environmental Research Letters*, *12*(8), Article 084005. https://doi.org/10.1088/1748-9326/aa7a60

Porter J. R., Shu E. G., Amodeo M. F., Freeman N., Almufti I., Ackerson M., Mehta J., Buyco K., Moussa H., Smith T., Barnes K., Hogan J., Djurovic J., Bouvier N., Bakare K., Olley J., & Birmingham R. (2021). *Using a high-precision flood risk assessment tool to understand commercial building and market impacts in the united states*. https://ssrn.com/abstract=3981118. Downloaded August 13, 2022.

Raftery A. E., Zimmer A., Frierson D. M. W., Startz R., & Liu P. (2017). Less than 2°C warming by 2100 unlikely. *Nature Climate Change*, *7*(9), 637–641. https://doi.org/10.1038/nclimate3352

Rennert K., Prest B. C., Pizer W. A., Newell R., Anthoff D., Kingdon C., Rennels L., Cooke R., Raftery A. E., Ševčíková H., & Errickson F. (2021). The social cost of carbon: Advances in long-term probabilistic projections of population, GDP, emissions, and discount rate. *Brookings Papers on Economic Activity*, *2021*(Fall), 223–305.

Risk Factor (2022). *Risk factor: Find your home's risk factors*. https://riskfactor.com. Downloaded August 13, 2022.
Srikrishnan V., Guan Y. W., Tol R. S. J., & Keller K. (2022). Probabilistic projections of baseline twenty-first century CO2 emissions using a simple calibrated integrated assessment model. *Climatic Change*, *170*(37). https://doi.org/10.1007/s10584-021-03279-7
Zampieri L., Goessling H. F., & Jung T. (2018). Bright prospects for Arctic sea ice prediction on subseasonal time scales. *Geophysical Research Letters*, *45*(18), 9731–9738. https://doi.org/10.1029/2018GL079394
Zhang B., & Cressie N. (2019). Estimating spatial changes over time of Arctic sea ice using hidden 2× 2 tables. *Journal of Time Series Analysis*, *40*(3), 288–311. https://doi.org/10.1111/jtsa.12425
Zhang B., & Cressie N. (2020). Bayesian inference of spatio-temporal changes of Arctic sea ice. *Bayesian Analysis*, *15*(2), 605–631. https://doi.org/10.1214/20-BA1209

https://doi.org/10.1093/jrsssc/qlad044
Advance access publication 16 June 2023

Seconder of the vote of thanks and contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Jonathan Rougier

University of Bristol, Bristol, UK; and Rougier Consulting Ltd.

Address for correspondence: Jonathan Rougier, School of Mathematics, Fry Building, Woodland Road, Bristol BS8 1UG, UK. Email: j.c.rougier@bristol.ac.uk

I would like to congratulate both sets of authors for their papers, which are substantial pieces of work on important and challenging topics. In the interests of brevity, I will focus on *Assessing present and future risk of water damage* by Claudio Heinrich-Mertsching et al. This article has a more client-focused analysis, something that is pertinent to me, now working as a consulting statistician rather than an academic statistician. In particular, I would like to consider how we, as statisticians, help the client to shape the research question, and how we demonstrate that, though our statistical efforts, we add value to the client, and justify our fees.

Let me jump straight to the authors' Figure 1, shown in the left-hand panel of my Figure 1. The 'Baseline' is a model using only property covariates. This figure appears to show that adding climate and topography covariates improves the Brier Score substantially over property alone, and certainly this is what we would expect, for a model to predict claims on insurance contracts for water damage. But close inspection of this figure reveals a highly compressed vertical axis. The Brier Score can go down to zero, and this seems like a reasonable endpoint for the vertical axis, given that almost all contracts have no claims. The right-hand panel of my Figure 1 shows the same information, with the extended vertical axis—ignore the dashed line saying 'Baseline' for now. It does appear, on this scale, that the improvements in the Brier score are insubstantial. This is odd because it confounds our expectations, and is therefore worth exploring further.

I have a rather simple view of the 'baseline' in a client setting, which is the performance the client can get *without hiring a statistician (e.g. me)*. I can say, 'Look, here's where you are. What do you think would be an operationally meaningful improvement?'. That way we all agree, at the outset, on what a successful outcome looks like. I have worked on some prediction applications which are very challenging, where the baseline error rate is about 30%, and where 5 percentage points would be an operationally meaningful reduction. But the authors are considering a different type of

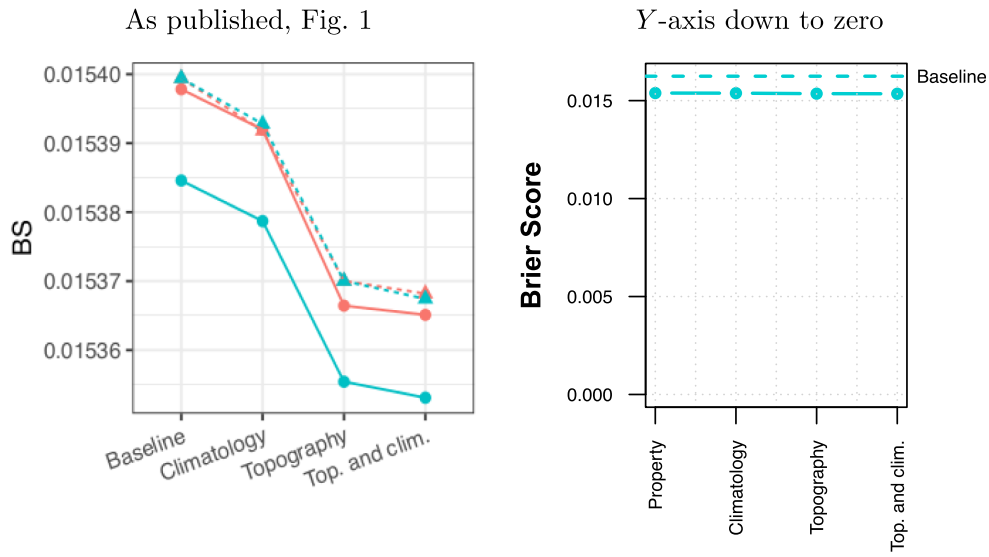

Figure 1. The authors' Figure 1, left-hand panel. My version of their Figure 1, right-hand panel. In my version, the vertical axis goes down to zero, 'Baseline' on the horizontal axis has been relabelled as 'Property', and there is a new dashed horizontal line labelled 'Baseline'; see the text.

application. Most of their contracts do not have claims, and therefore the baseline error rate ought to be very low, already close to zero in the Brier Score, and a meaningful reduction will be much smaller.

The authors report that out of 1,740,915 observations, 1,712,157 contain zero claims (end of their Section 2.1). So approximately 1.7% of contracts contain one or more claims. In the Brier Score, we can use this gross value instead of the contract-specific value $\hat{p}_i$. In other words, we completely ignore all information about the property, the climate, the topography. The Brier score with this baseline is

$$\mathrm{BS}_{\mathrm{baseline}} = 1.7\% \times (1 - 1.7\%) \approx 0.0162,$$

using the exact figures. So, the authors have added value if they can reduce the Brier score to below 0.0162. At this point, I would have a discussion with the client about how big a reduction was operationally meaningful. Perhaps the authors can comment on this in their rejoinder.

The right-hand panel of my Figure 1 has the baseline Brier score added, as a horizontal dashed line. We immediately see two things:

1. The authors have added about 0.001 of value (10 basis points).
2. Climate and topography appear not to add meaningful value over property alone.

On the basis of the second point above, I think it would be risky to interpret the fitted climate and topography values in the model. Something is not right: we expect climate and topography covariates to add meaningful value, but apparently they have not. So what is going on?

I think the root of the issue can be found in the authors' choice of model. The authors fit an 'interpretable' Generalised Additive Model (GAM) of the form

$$\log \mathbb{E}(N_i/\ell_i) = f(\mathrm{property}_i) + g(\mathrm{climate}_i) + h(\mathrm{topography}_i),$$

where N_i is the number of claims from contract i, and ℓ_i is the duration of contract i. In other words, they enforce a *'No Interactions Condition (NIC)'*:

The effect of changing from one property type to another does not depend on climate or topography.

I suspect that most of us would question the NIC, for this application. For example, putting in a basement incurs a higher risk in a property at the bottom of a slope in a rainy catchment. This is a three-way interaction between property, climate, and topography. Since the goal of the article was to predict present and future water damage risk, it seems strange to constrain the model, from the start, not to allow any interactions of this type. I wonder if this constraint is the reason why including climate and topography covariates does not seem to improve the performance of the model.

It is not the technology that restricts the choice of model. My first choice, rather than a GAM, would be Random Forest regression; see e.g. Hastie et al. (2009, chapter 15) and Efron and Hastie (2016, chapter 17). Random Forests are all about interactions! I would fit

$$y_i = \sqrt{\frac{N_i}{\ell_i}} = \text{RanFor}(\text{building}_i, \text{climate}_i, \text{topography}_i),$$

using a Poisson variance-stabilising transform on the left-hand side. For prediction, I would set

$$\hat{\mu}_i = \ell_i \cdot \max\{0, \hat{y}_i\}^2$$

in a Poisson model for the number of claims. I would hope that with this model, including climate and topography covariates would produce a meaningful reduction in the Brier score over property alone: say at least 10 basis points, so that we could see it on my Figure 1.

It is always annoying when another statistician says 'I would have modelled that differently'. But in this application there seems to be a plausible case that the authors' decision to exclude interactions in their model has compromised its performance, and I would like to find out more about how this decision was reached.

References

Efron B., & Hastie T. (2016). *Computer age statistical inference*. Cambridge University Press.

Hastie T., Tibshirani R., & Friedman J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

The vote of thanks was passed by acclamation.

https://doi.org/10.1093/jrsssc/qlad045
Advance access publication 13 June 2023

Richard L. Smiths contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Richard L. Smith

Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, USA

Address for correspondence: Richard L. Smith, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC 27599-3260, USA. Email: rls@email.unc.edu

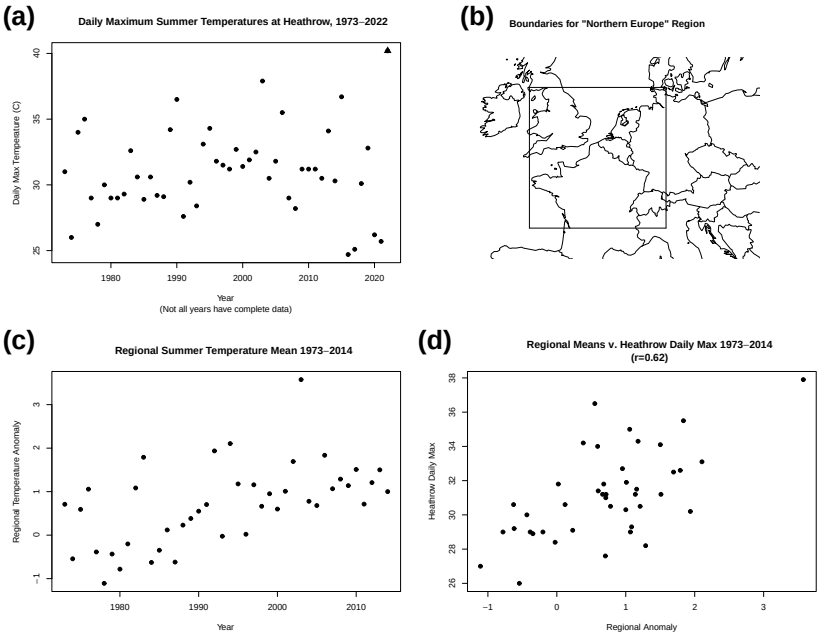

Figure 1. (a) Annual maximum temperatures at Heathrow; triangle represents 2022 value. (b) Region of Northern Europe used to define regional summer means. (c) Regional summer means by year. (d) Regional summer means plotted against Heathrow annual maxima.

I found both papers very interesting but would like particularly to discuss the one by Clarkson and co-authors.

One particular focus is the 'zero probability problem', that with a negative shape parameter, it is possible for some observed event to be assigned probability 0 by an extreme value analysis. This problem has been known for some time but has been given particular focus by papers of the 'World Weather Attribution' group, see e.g. (Philip et al., 2022; Zachariah et al., 2022). The authors' solution, a combination of peaks over threshold analysis and random parameters, is perfectly reasonable, but an alternative is simply to switch to a Bayesian approach as a more realistic recipe for handling uncertainties in the parameter estimates. As far as I know, Coles and Powell (1996) were the first authors to make this argument for events of very low probability.

Going beyond their analysis, I would like to comment on the broader climate context. In July 2022, many parts of the UK suffered extreme heat. To take a specific data point, on July 19, Heathrow Airport recorded a daily high temperature of 40.2°C (104.4° F). Much of the subsequent discussion has concerned the probability of such an extreme event, in the present climate but also in the past and future.

The approach I am currently developing involves an associated "climate" variable—in this case, summer mean temperature anomalies from a region of western Europe (Figure 1). Climate model runs (CMIP6) were used to characterise the distribution of this regional mean from 1850 to 2100, and then combined with an extreme value analysis to estimate the probability of an annual maximum temperature at Heathrow exceeding 40°C (Figure 2). The results show that the present probability of this event is about seven times larger than it was in the 1950s; under any scenario, it will increase sharply in the future, but there is a big difference between essentially uncontrolled greenhouse gases (ssp585) and the kind of emissions that might result with realistic controls (ssp245). This is the kind of result I would choose to discuss with our political leaders.

On a more technical level, I believe the extreme value analysis could be improved by taking advantage of spatial dependence among stations (Casson & Coles, 1999; Risser, 2020; Russell et al., 2020).

I thank the RSS for the opportunity to contribute this discussion.

Conflicts of interest: The author has no conflicts of interest.

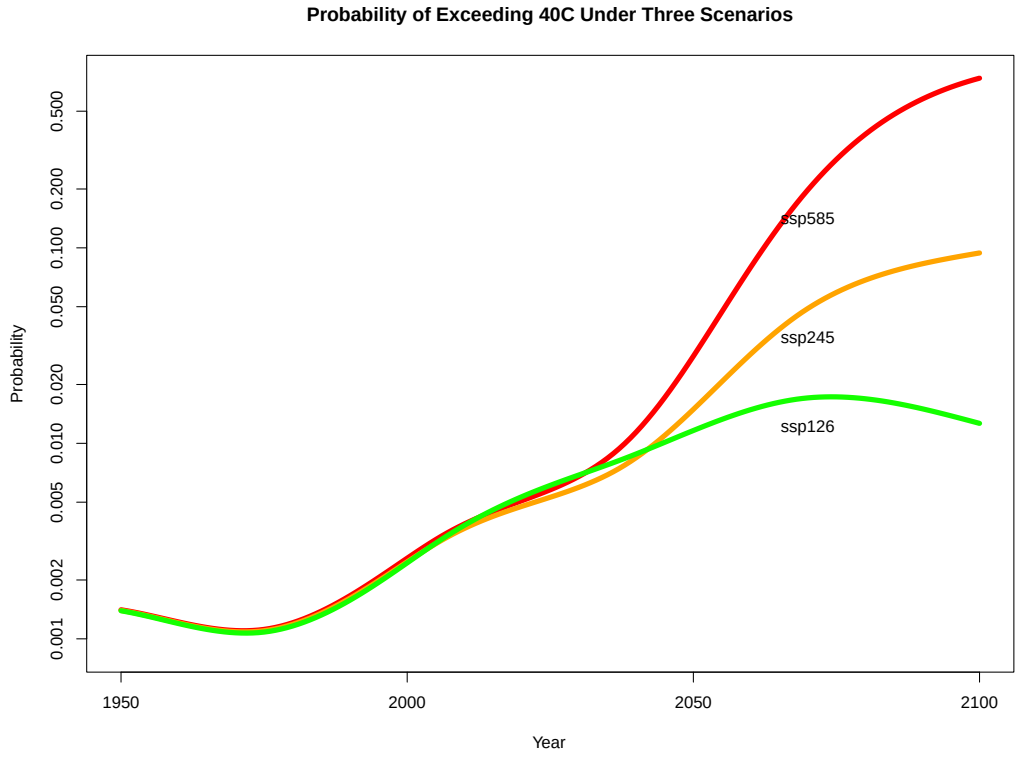

Figure 2. Estimated probability that the annual maximum temperature at Heathrow exceeds 40°C, calculated from historical model runs (1950–2014) and from three emissions scenarios (2015–2100).

Data availability

The data are all from public sources and may be accessed directly at http://rls.sites.oasis.unc.edu/ClimExt/intro.html.

References

Casson E., & Coles S. G. (1999). Spatial regression models for extremes. *Extremes*, *1*(4), 449–468. https://doi.org/10.1023/A:1009931222386

Coles S. G., & Powell E. A. (1996). Bayesian methods in extreme value modelling: A review and new developments. *International Statistical Review*, *64*(1), 119–136. https://doi.org/10.2307/1403426

Philip S. Y., Kew S. F., Jan van Oldenborgh G., Yang W., Vecchi G. A., Anslow F. S., Li S., Seneviratne S. I., Luu L. N., Arrighi J., Singh R., van Aalst M., Hauser M., Schumacher D. L., Pereira C., Kristie L Ebi M., Bonnet R., Vautard R., Tradowsky J., Coumou D., Lehner F., Wehner M., Rodell C., Stull R., Howard R., Gillett N.,& Otto F. E. L. (2022). Rapid attribution analysis of the extraordinary heatwave on the Pacific Coast of the US and Canada June 2021. *Earth System Dynamics*, *13*, 1689–1713. https://doi.org/10.5194/esd-13-1689-2022

Risser Mark (2020). *Nonstationary Bayesian modeling for a large data set of derived surface temperature return values*. arXiv, https://arxiv.org/abs/2005.03658, preprint: not peer reviewed.

Russell B., Risser M., Smith R. L., & Kunkel K. E. (2020). Investigating the association between late spring Gulf of Mexico sea surface temperatures and US Gulf Coast precipitation extremes with focus on Hurricane Harvey. *Environmetrics*, *31*(2), e2595. https://doi.org/10.1002/env.2595

Zachariah M., Vautard R., Schumacher D. L., Vahlberg M., Heinrich D., Raju E., Thalheimer L., Arrighi J., Singh R., Li S., Sun J., Vecchi G., Yang W., Seneviratne S. I., Tett S. F. B., Harrington L. J., Wolski P., Lott F. C., McCarthy M., Tradowsky J. S., & Otto F. E. L. (2022). *Without human-caused climate change temperatures of 40°c in the UK would have been extremely unlikely*. https://www.worldweatherattribution.org/wp-content/uploads/uk-heat-scientific-report.pdf (posted July 28, 2022; reference checked October 7, 2022).

https://doi.org/10.1093/jrsssc/qlad046
Advance access publication 22 June 2023

Christian Rohrbeck's contribution to the discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Christian Rohrbeck

Department of Mathematical Sciences, University of Bath, Bath, UK

Address for correspondence: Christian Rohrbeck, Department of Mathematical Sciences, University of Bath, Claverton Down, BA2 7AY, UK. Email: cr777@bath.ac.uk

Christian Rohrbeck (University of Bath): The authors are to be commended for their interesting contribution to the statistical analysis of weather-related property insurance claims. My contribution focuses on the temporal aggregation and climate indices used in the modelling framework:

There exists a variety of statistical models for spatially aggregated property insurance claims, such as Scheel et al. (2013), which aimed to identify the weather patterns causing high number of claims over an area. However, the application of these methods to predict claims in a changing climate is challenging because it requires generating future weather patterns. The temporal aggregation proposed by the authors removes this requirement in an elegant fashion as it allows to focus on climatological indices. However, it is unclear whether the considered covariates can capture the weather patterns causing pluvial flooding identified in previous publications.

Rohrbeck et al. (2018) found that aggregated and intense rainfall and prolonged periods of snowmelt are key drivers for insurance claims due to pluvial flooding in Bærum, Bergen, and Oslo. The topographical covariates in the model may capture some of these effects, but this aspect is not discussed in the article. Figure 1 in the presented paper further shows only a relatively modest improvement in terms of predictive performance when climatological and topographical covariates are added to the baseline model only using property-specific covariates. It may be interesting to explore whether this relative improvement is consistent across Norway by producing the graphs in Figure 1 for different regions in Norway. This assessment would allow the authors to explore whether the model estimates are mainly driven by the cities with large number of insurance policies, or whether the framework can adequately capture regional differences in vulnerability to pluvial flooding across Norway. Similarly, the authors could compare predictive performance for the different quarters considered in the analysis to explore whether their model captures the different weather dynamics across seasons. Did the authors explore these aspects in their analysis?

Finally, in a changing climate, the weather events mentioned above may become more, or less, frequent and severe. The considered climate indices may be insufficient to capture such trends adequately as they only capture differences in a subset of the events causing pluvial flooding. Do the authors have any intuition regarding this aspect?

Conflict of interests: None declared.

Data availability

No new data were created or analyzed as part of this contribution.

References

Rohrbeck C., Eastoe E. F., Frigessi A., & Tawn J. A. (2018). Extreme value modelling of water-related insurance claims. *The Annals of Applied Statistics*, *12*(1), 246–282. https://doi.org/10.1214/17-AOAS1081

Scheel I., Ferkingstad E., Frigessi A., Haug O., Hinnerichsen M., & Meze-Hausken E. (2013). A Bayesian hierarchical model with spatial variable selection: The effect of weather on insurance claims. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(1), 85–100. https://doi.org/10.1111/j.1467-9876.2012.01039.x

https://doi.org/10.1093/jrsssc/qlad047
Advance access publication 16 June 2023

Miguel de Carvalho, Alina Kumukova, and Vianey Palacios Ramirezs contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Miguel de Carvalho, Alina Kumukova and Vianey Palacios Ramírez

School of Mathematics, University of Edinburgh, Edinburgh, UK

Address for correspondence: Miguel de Carvalho, School of Mathematics, University of Edinburgh, Edinburgh, UK.
Email: miguel.decarvalho@ed.ac.uk

We congratulate the authors for a fresh, thought-provocative perspective on applied modelling of extremes. While we may not fully agree with some philosophical details on the approaches taken, we absolutely concur with the rising need for going beyond standard EVT models. In particular, we have been advocating with others (Lugrin et al., 2016; Hanson et al., 2017; de Carvalho et al., 2022; Padoan & Rizzelli, 2022) the need for semi- and nonparametric Bayesian approaches in extremes—and this article indirectly provides another case for it. Many nonparametric Bayesian approaches are an extension of standard parametric methods in the sense that they are centred a priori around a parametric model, $\{G_{0,\theta} : \theta \in \Theta \subseteq \mathbb{R}^q\}$ but assign positive mass to a variety of alternatives. Clearly, nonparametric Bayesian EVT modelling is not as straightforward as simply centring a Dirichlet process (DP) on a GEV or a POT—as it is well-known that the tails of the DP are exponentially lighter than those of the baseline (Ghosal & Van der Vaart, 2015, Section 4.2.3). Recent developments in Palacios Ramírez et al. (2022) suggest however that this issue can be mitigated by resorting, for example, to stable law processes—and obviously related infinite mixture models may be handy for situations such as those covered in Section 4.

On a different note, we particularly appreciate the random effect approach in Section 2.2. Still, one wonders about why not using Extended Generalized Pareto Distribution-based methods (e.g., Papastathopoulos & Tawn, 2013; Naveau et al., 2016; de Carvalho et al., 2022) that would model the full set of observations rather than the exceedances alone and would mitigate the need for threshold selection? Indeed, in Section 4, all the data are used to learn about the targets of interest but not in Section 3.

Finally, we conclude with a comment on nonstationary multivariate extreme-valued models (e.g., de Carvalho, 2016; Castro et al., 2018; Escobar-Bach et al., 2018; Mhalla et al., 2019)—that index the angular measure with a covariate, H_x—or another related functional. It would seem natural to devise random effects versions for the latter context.

Conflict of interest: None declared.

References

Castro D., de Carvalho M., & Wadsworth J. L. (2018). Time-varying extreme value dependence with application to leading European stock markets. *Annals of Applied Statistics*, *12*(1), 283–309.

de Carvalho M. (2016). Statistics of extremes: Challenges and opportunities. In F. Longin (Ed.), *Extreme events in finance: A handbook of extreme value theory and its applications*. Wiley.

de Carvalho M., Kumukova A., & Dos Reis G. (2022). Regression-type analysis for block maxima on block maxima. *Extremes*, *25*, 595–622.

Escobar-Bach M., Goegebeur Y., & Guillou A. (2018). Local robust estimation of the pickands function. *The Annals of Statistics*, *46*(6A), 2806–2843. https://doi.org/10.1214/17-AOS1640

Ghosal S., & Van der Vaart A. W. (2015). *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press.

Hanson T. E., de Carvalho M., & Chen Y. (2017). Bernstein polynomial angular densities of multivariate extreme value distributions. *Statistics and Probability Letters*, *128*, 60–66. https://doi.org/10.1016/j.spl.2017.03.030

Lugrin T., Davison A. C., & Tawn J. A. (2016). Bayesian uncertainty management in temporal dependence of extremes. *Extremes*, *19*(3), 491–515. https://doi.org/10.1007/s10687-016-0258-0
Mhalla L., de Carvalho M., & Chavez-Demoulin V. (2019). Regression type models for extremal dependence. *Scandinavian Journal of Statistics*, *46*(4), 1141–1167. https://doi.org/10.1111/sjos.12388
Naveau P., Huser R., Ribereau P., & Hannart A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*.
Padoan S. A., & Rizzelli S. (2022). Consistency of Bayesian inference for multivariate max-stable distributions. *Annals of Statistics*, *50*(3), 1490–1518.
Palacios Ramírez V., de Carvalho M., & Gutiérrez Inostroza L. (2022). 'Heavy-tailed NGG mixture models,' arXiv, arXiv:2211.00867.
Papastathopoulos I., & Tawn J. A. (2013). Extended generalised pareto models for tail estimation. *Journal of Statistical Planning and Inference*, *143*(1), 131–143. https://doi.org/10.1016/j.jspi.2012.07.001

https://doi.org/10.1093/jrsssc/qlad048
Advance access publication 14 June 2023

Christine P Chai's contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Christine P. Chai

Microsoft Corporation, Redmond, WA, USA

Address for correspondence: Christine P. Chai, Microsoft Corporation, Redmond, WA, USA. Email: cpchai21@gmail.com

Statistical aspects of climate change

Paper 1: Assessing present and future risk of water damage using building attributes, meteorology and topography
Authors: Claudio Heinrich-Mertsching*, Jens Christian Wahl*, Alba Ordonez*, Marita Stien#, John Elvsborg#, Ola Haug*, Thordis L. Thorarinsdottir*
*Norwegian Computing Center, Oslo, Norway
#Gjensidige Forsikring ASA, Oslo, Norway

Paper 2: The importance of context in extreme value analysis with application to extreme temperatures in the USA and Greenland
Authors: Daniel Clarkson, Emma Eastoe, Amber Leeson, University of Lancaster, UK

Disclaimer: The opinions and views expressed here are those of the author and do not necessarily state or reflect those of Microsoft.

Discussion

Climate change has been an emerging research field given its tremendous effect on human beings, and I am happy to see discussions like (Clarkson et al., 2023; Heinrich-Mertsching et al., 2022). To increase awareness of modelling climate change, we also need to inoculate the public against online misinformation (Treen et al., 2020; Van der Linden et al., 2017). Neither statistical model is perfect, and their limitations can easily be attacked by researchers with a vested interest in the fossil fuels industry. In terms of environmental science, there is still much work to do in both the scientific side and the education side.

In the second paper (Clarkson et al., 2023), I appreciate the introduction to the history of extreme value theory and applications. With the increasing availability of spatio-temporal image data, spatial statistics has become multidimensional and more complex than ever. Sometimes a parsimonious model is sufficient, and I am pleasantly surprised to see a relatively simple method to model extreme temperatures over time. In the extreme value analysis with random effects, this article is focused on the upper tail (high temperature). Can the Gaussian mixture model fit be applied to the lower tail as well? Unexpectedly cold winters also cause damage to human beings and property.

Conflict of interest: The author is employed at Microsoft Corporation, and she completed a PhD in statistical science from Duke University.

Data availability

No new data were generated or analyzed in support of this research.

References

Clarkson D., Eastoe E., & Leeson A. (2023). The importance of context in extreme value analysis with application to extreme temperatures in the USA and Greenland. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *72*(4), 829–843. https://doi.org/10.1093/jrsssc/qlad020.

Heinrich-Mertsching C., Wahl J. C., Stien M., Elvsborg J., Haug O., & Thorarinsdottir T. L. (2022). Assessing present and future risk of water damage using building attributes, meteorology and topography. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 1–37.

Treen K. M. D. I., Williams H. T. P, & O'Neill S. J. (2020). Online misinformation about climate change. *Wiley Interdisciplinary Reviews: Climate Change*, *11*(5), e665. https://doi.org/10.1002/wcc.665.

Van der Linden S., Leiserowitz A., Rosenthal S., & Maibach E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, *1*(2), 1600008. https://doi.org/10.1002/gch2.201600008

https://doi.org/10.1093/jrsssc/qlad049
Advance access publication 19 June 2023

Anna Choi and Tze Leung Lai's contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Anna Choi and Tze Leung Lai

Center for Innovative Study Design, Stanford University, Stanford, CA, USA

Address for correspondence: Anna Choi, Center for Innovative Design, Stanford University, Stanford, CA 94305, USA.
Email: annahcchoi@gmail.com

Authors and Papers under Discussion:

1. 'Assessing present and future risk of water damage using building attributes, meterology and topograph' Heinrich-Mertsching C, Wahl JC, Ordonez A, Stien M, Elvsborg J, Haug O, and Thorarinsdottir TL.
2. 'The importance of context in extreme value analysis with application to extreme temperatures in the USA and Greenland' Clarkson D, Eastoe E, and Leeson A.

Although the two papers under discussion are based on different methods and involve different illustrative applications, they are of great appeal to us who are involved in developing courses on environmental health in Stanford's new Doerr School of Sustainability and Climate Change. In fact, September 29, 2022 marks the historic opening of the School led by Dean Arun Majumdar. The paper 'Assessing present and future risk of water damage using building attributes, meteorology and topography' by Heinrich-Mertsching et al. from Norway establishes a 'nationwide, building-specific risk score for water damages associated with pluvial flooding in Norway', and the basic methodology is to 'fit a generalised additive model that relates the number of water damages to a wide range of explanatory variables' (or covariates). Combining the model with an ensemble of climate projections enables projection of the (spatially varying) impacts of climate change on the 'risk of pluvial flooding towards the middle and end of the 21st century.' The paper 'The importance of context in extreme value analysis with application to extreme temperatures in the USA and Greenland' by Clarkson et al. of Lancaster University in the UK uses extreme value theory to analyse this problem and applies a random effects Peaks over Threshold approach to the case studies in Sections 3 and 4 (for air temperatures in three western US states and ice surface temperatures in Greenland, respectively). Data science, as manifested in these 2 papers of different approaches and motivations, is very useful to develop for students in the Doerr School to learn.

Conflicts of interest: None declared.

https://doi.org/10.1093/jrsssc/qlad050
Advance access publication 22 June 2023

Valérie Chavez-Demoulin, Anthony C Davison and Erwan Koch's contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

V. Chavez-Demoulin[1], A. C. Davison[2] and E. Koch[2,*]

[1]Faculty of Business and Economics, University of Lausanne, CH-1015 Lausanne, Switzerland
[2]Institute of Mathematics, EPFL, CH-1015 Lausanne, Switzerland

* Present address: Expertise Center for Climate Extremes (ECCE), University of Lausanne, CH-1015 Lausanne, Switzerland

Address for correspondence: A. C. Davison, Institute of Mathematics, EPFL, CH-1015 Lausanne, Switzerland. Email: anthony.davison@epfl.ch

This interesting paper poses many questions on the use of extreme value theory for climate modelling.

An alternative to defining separate annual thresholds for the peaks over thresholds model would be to use the upper r order statistics for each year. This extends the use of annual maxima but requires only a single value r to be chosen rather than many thresholds. Moreover, this approach has the advantage of retaining the parameters of the generalized extreme-value (GEV) distribution, which typically have a more stable interpretation than do those of the generalized Pareto distribution; when the threshold changes, so do the exceedance probability and the scale parameter. Since threshold stability is a key property of the generalized Pareto model, it seems

unfortunate that the joint distribution of the random effects will change if the threshold is varied. While the prior distribution appears to be chosen purely for computational convenience and as an approach to penalisation, comparison of results for different thresholds would entail appropriate transformations of their respective priors. It would be interesting to see if the Markov chain Monte Carlo output was more readily interpretable when transformed to the GEV scale. As the authors point out, the random effects model can be seen as adjusting for variation not captured by observed explanatory variables, but if phenomena appear that have been rarely observed in the past, such as so-called 'heat domes', then the basis of the mixture model, exchangeability of the unobserved variation from year to year, becomes moot.

On a more computational note, the scale and shape parameters of both models are negatively correlated, because higher observations could be explained by increasing either, so it may be preferable to impose the priors and perform the sampling after some form of parameter orthogonalisation.

We assume that the Global Mean Surface Temperature is somehow localised. A broad-brush covariate such as this should presumably be adapted for application in a rather small region, perhaps using similar local covariates or some form of smoothing.

The term 'return level' is well-anchored in the literature, but nevertheless it seems preferable to avoid it, because of its underlying assumption of stationarity. Moreover the appearance of so-called '100-year events' every few years is poor publicity for statistics and it seems better to avoid the term.

Conflict of interest: None declared.

https://doi.org/10.1093/jrsssc/qlad051
Advance access publication 13 June 2023

Ankur Dutta's contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Ankur Dutta

Indian School of Business, Hyderabad, Telangana, India

Address for correspondence: Ankur Dutta, 66 Garfa Main Road, Pratapgarh More, Jadavpur, Kolkata-700075, West Bengal, India. Email: ankurdutta43@gmail.com

The authors of this work has provided many insights into the alarming situations caused by the recent climatic changes and as statisticians, we appreciate all efforts in this regard. This paper at its core provided valuable methods and new ideas in the light of extreme value theory. The authors of this article deserve praise for their efforts.

In case study 1, for the US temperature, the authors have used both the methods namely the Block Maxima (BM) analysis and the Peaks over Threshold (POT). Now considering the fact that in the BM method the main assumption is the maxima of all the blocks (i.e. data divided into suitable blocks) follow a suitable well defined Generalised Extreme Value (GEV) distribution. However, it misses on some of the high observations as when divided into blocks (here local climate condition) the maxima of one block might be lower than some of the lower observations in other blocks so eventually some lower observations get retained and the rest are ruled out. Therefore, one might say that the POT method seems more suitable, since it marks a threshold and any value over that threshold is considered. Hence, it would have been helpful if a thorough justification for choosing the POT approach over the BM method is provided.

Several papers all based on simulated data, have examined the relative merits of POT and BM, according to Cunnane (1973) for $\xi = 0$ (shape parameter of the GEV distribution) and for maximum likelihood estimators, the POT estimate for a high quantile is better only if the number of exceedances is larger than 1.65 times the number of blocks. Wang (1991) writes that POT is as efficient as the BM model for high quantiles, based on Probability Weighted Moment estimators. Again Madsen and Rosbjerg with Pearson (1997) and with Rasmussen (1997) write that POT is better for $\xi > 0$, whereas for $\xi < 0$, BM is more effective with the number of exceedances larger than the number of blocks. So, for locations where $\xi < 0$, if a justification for choosing one of the two methods is provided then it will help.

In case study 2, Greenland ice surface temperatures the melt threshold impact on the distribution of Ice Surface Temperatures (IST) is observed that is under the consideration that melt occurs when the IST exceeds 0°C. This accounts for observing the positive temperatures, but as the work mentions that there are too few positive observations to take 0° as the POT threshold. Hence, one might ask the question why not use the BM method here? Therefore, an explanation would have been helpful in this regard.

Conflict of interest: None declared.

References

Cunnane C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, *18*(3–4), 257–271. https://doi.org/10.1016/0022-1694(73)90051-6

Madsen H., Pearson C. P., & Rosbjerg D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 2. Regional modeling. *Water Resources Research*, *33*(4), 759–769. https://doi.org/10.1029/96WR03849

Madsen H., Rasmussen P. F., & Rosbjerg D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events 1. At-site modeling. *Water Resources Research*, *33*(4), 747–757. https://doi.org/10.1029/96WR03848

Wang Q. J. (1991). The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology*, *129*(1–4), 263–280. https://doi.org/10.1016/0022-1694(91)90054-L

https://doi.org/10.1093/jrsssc/qlad052
Advance access publication 16 August 2023

Juliette Legrand and Thomas Opitz's contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Juliette Legrand and Thomas Opitz

Biostatistics and Spatial Processes, INRAE, Avignon, France

Address for correspondence: Thomas Opitz, 228 Route de l'Aérodrome, 84914 Avignon, France. Email: thomas.opitz@inrae.fr

We congratulate the authors for their carefully constructed spatial model of risk occurrences caused by pluvial flooding. It astutely incorporates meteorological and topographic risk drivers and accounts for other, unknown factors of spatial variability, making it reliable

for short-term prediction and long-term projection under climate change. We emphasise that many difficulties identified by the authors arise more generally for various other environmental risks.

Our research group working on wildfires at INRAE faces similar challenges regarding the modelling of their occurrences and sizes (e.g. Koh et al., 2023; Pimont et al., 2021). The problem of a *covariate shift* in prediction with new data is typical for meteorological covariates simulated by climate models in change scenarios. Prediction often suffers from noisy behaviour of semiparametric predictor components with basis function representation (e.g. splines, as in the discussed paper). Different choices of knots and shapes for basis functions may produce quite different predictions conditional to the tails of the covariate distribution.

Regarding wildfires, useful meteorological covariates are the Fire Weather Index (FWI) and its subindices, which aggregate common weather variables (precipitation, humidity, temperature, wind speed) into biophysical variables providing a rating of fire danger components (e.g. ignition, spread). In the Bayesian setting of the *Firelihood* model of Pimont et al. (2021), the influence of FWI on counts and burnt areas of wildfires is robustly modelled through a step function (i.e. a zero-degree spline, with around 50 knots) and Gaussian first order random walk prior. Similar to the discussed paper, the spline function is *frozen* for new (i.e. projected) FWI values beyond the range of historical observations and is extrapolated as a constant. Future projections of FWI often highly exceed its historical maxima (e.g. Varela et al., 2019).

We currently explore avenues towards improved modelling with covariate shifts in the tails of the covariate distribution. A promising idea is to infer an appropriate parametric predictor in the tail region of the covariate, requiring a hybridisation of a semiparametric bulk model with a parametric tail model. Moreover, insights can be gained from the conditional extremes framework (Heffernan & Tawn, 2004) by conditioning the response variable on threshold exceedances of the covariate. While this approach applies to continuous variables, it could be extended to count-valued responses using extreme-value theory for discrete variables (Hitz et al., 2017).

References

Heffernan J. E., & Tawn J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x

Hitz A., Davis R., & Samorodnitsky G. (2017). 'Discrete extremes', arXiv, arXiv:1707.05033, preprint: not peer reviewed.

Koh J., Pimont F., Dupuy J.-L., & Opitz T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *Annals of Applied Statistics*, *17*(1), 560-582.

Pimont F., Fargeon H., Opitz T., Ruffault J., Barbero R., Martin-StPaul N., Rigolot E., Rivière M., & Dupuy J.-L. (2021). Prediction of regional wildfire activity in the probabilistic Bayesian framework of Firelihood. *Ecological Applications*, *31*(5), e02316. https://doi.org/10.1002/eap.2316

Varela V., Vlachogiannis D., Sfetsos A., Karozis S., Politi N., & Giroud F. (2019). Projection of forest fire danger due to climate change in the French Mediterranean region. *Sustainability*, *11*(16), 4284.

https://doi.org/10.1093/jrsssc/qlad054
Advance access publication 16 June 2023

Saralees Nadarajah's contribution to the Discussion of 'The First Discussion Meeting on Statistical aspects of climate change'

Saralees Nadarajah

Department of Mathematics, University of Manchester, Manchester, UK

Address for correspondence: Saralees Nadarajah, Department of Mathematics, University of Manchester, Manchester M13 9PL, UK. Email: saralees.nadarajah@Manchester.ac.uk

Clarkson et al. (2022) is an interesting paper. But, I have questions about the motivation and methodology.

The paper is clearly motivated by climate change. The biggest culprits for the climate change are the UK and the U.S. The industrial revolutions in both these countries contributed much to climate change. Now other countries especially African countries are being asked to take measures to avoid further climate change. African countries contributed negligibly to climate change.

Most of the countries worst affected by climate change are in the developing world or the underdeveloped world (e.g. Afghanistan, Bangladesh, Fiji, India, Myanmar, Nepal, Pakistan, Sri Lanka, the Philippines, Thailand, Chad, Kenya, Madagascar, Malawi, Mozambique, Niger, Nigeria, Rwanda, Somalia, Sudan, Yemen, Zimbabwe, Bolivia, Haiti, Puerto Rico, and the Bahamas). Instead of focusing on some of these countries, Clarkson et al. (2022) focus on two countries in the developed world. One of the two countries considered by Clarkson et al. (2022) is the U.S., the biggest ever contributor to climate change!

The methodology in Clarkson et al. (2022) involves adding random effects to the parameters of the generalised Pareto distribution. The idea of adding random effects to parameters of Pareto distributions is not new at all. Nadarajah (University of Nebraska, Lincoln, USA, unpublished manuscript) constructed models for the Pareto type II distribution (also known as the Lomax distribution) by letting its scale parameter to be random. Let X be a Pareto type II random variable with probability density function

$$f_X(x) = \frac{a\sigma^a}{(x+\sigma)^{a+1}}$$

for $x > 0$, $a > 0$, and $\sigma > 0$. Letting σ to be a random variable (independent of X) with probability density function $g(\cdot)$, Nadarajah derived closed form expressions for the unconditional probability density function of X:

$$\int_0^\infty \frac{a\sigma^a}{(x+\sigma)^{a+1}} g(\sigma)\mathrm{d}\sigma. \tag{1}$$

Thirteen possible forms for $g(\cdot)$ were considered. The closed form expressions for (1) involved known special functions like the incomplete gamma function, the Appell function of the first kind, the generalised hypergeometric function and the Kummer function.

The extension of (1) is straightforward if X is a generalised Pareto random variable. Let X be a generalised Pareto random variable with probability density function

$$f_X(x) = \frac{\phi}{\sigma}\left[1 + \frac{\xi(x-u)}{\sigma}\right]^{-\frac{1}{\xi}-1}$$

for $x > u$, $-\infty < \xi < \infty$, and $\sigma > 0$. If we take σ to be an exponential random variable, e.g. independent of X, calculations similar to those in the unpublished manuscript show that the unconditional probability density function of X is

$$\int_0^\infty \frac{\phi}{\sigma}\left[1+\frac{\xi(x-u)}{\sigma}\right]^{-\frac{1}{\xi}-1}\frac{1}{\lambda}\exp\left(-\frac{\sigma}{\lambda}\right)\mathrm{d}\sigma=\frac{\phi}{\lambda}\Gamma\left(\frac{1}{\xi}+1\right)\Psi\left(\frac{1}{\xi}+1,\,1;\,\frac{\xi(x-u)}{\lambda}\right),$$

where $\Psi(a, b; x)$ denotes the Kummer function defined by

$$\Psi(a,b;x)=\frac{\Gamma(1-b)}{\Gamma(1+a-b)}\,{}_1F_1(a,b;x)+\frac{\Gamma(b-1)}{\Gamma(a)}x^{1-b}\,{}_1F_1(1+a-b,2-b;x),$$

where ${}_1F_1(a, b; x)$ denotes the confluent hypergeometric function defined by

$${}_1F_1(a;b;x)=\sum_{k=0}^{\infty}\frac{(a)_k}{(b)_k}\frac{x^k}{k!},$$

where $(f)_k = f(f+1)\cdots(f+k-1)$ denotes the ascending factorial. The properties of the these special functions can be found in Prudnikov et al. (1986) and Gradshteyn and Ryzhik (2000). In-built functions for computing these functions are available in packages like Maple, Matlab, and Mathematica.

Conflicts of interest: The author declares no conflicts of interest.

Funding

No funding was received for this study.

Data availability

Not applicable.

Code availability

Not applicable.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

The author gave explicit consent to publish this manuscript.

References

Clarkson D., Eastoe E., & Leeson A. (2022). The importance of context in extreme value analysis with application to extreme temperatures in the USA and Greenland (with discussion). *Journal of the Royal Statistical Society C*, to appear.

Gradshteyn I. S., & Ryzhik I. M. (2000). *Table of integrals, series, and products* (6th ed.). Academic Press.

Prudnikov A. P., Brychkov Y. A., & Marichev O. I. (1986). *Integrals and series, volumes 1, 2 and 3*. Gordon and Breach Science Publishers.

The authors replied later, in writing, as follows:

https://doi.org/10.1093/jrsssc/qlad053
Advance access publication 13 June 2023

Authors' reply to the Discussion of 'Assessing present and future risk of water damage using building attributes, meteorology and topography' at the first meeting on 'Statistical aspects of climate change'

Claudio Heinrich-Mertsching[1], Jens Christian Wahl[1], Alba Ordoñez[1], Marita Stien[2], John Elvsborg[2], Ola Haug[1] and Thordis Thorarinsdottir[1]

[1]Norwegian Computing Center, Postboks 114, Blindern, NO-0314 Oslo, Norway
[2]Gjensidige Forsikring ASA, Postboks 700 Sentrum, NO-0160 Oslo, Norway

Address for correspondence: Claudio Heinrich-Mertsching, Norwegian Computing Center, Postboks 114, Blindern, Oslo NO-0314, Norway. Email: claudio@nr.no

We are deeply grateful to our colleagues, and in particular to the main discussants Adrian E. Raftery and Jonathan Rougier, for their insightful thoughts and comments. For our response, we have grouped our comments according to four overarching themes, namely

(a) weather and climate information,
(b) model specification,
(c) model verification, and
(d) relation to other work.

1 Weather and climate information

The future risk projections presented in the paper are, as Raftery correctly states in his comment, conditional on a given deterministic representative concentration pathway emission scenario. Uncertainty regarding future emissions is thus not accounted for in the overall uncertainty assessment and subjective choices have been made regarding the selection of a given scenario. While this has long been the available approach for climate impact studies of the type presented here, the recent series of work listed in Raftery's comment is changing the state-of-the-art. We agree that probabilistic projections of carbon emissions should be considered if possible. However, such projections are, to the best of our knowledge, not yet available for the bivariate projections of temperature and precipitation required in our analysis. Furthermore, given the complex orography of Norway, we have chosen to use the dynamically downscaled projections provided by regional climate models (RCMs) rather than the output of global circulation models (GCMs) directly. At the time of our analysis, RCM projections for Norway based on the more recent shared socioeconomic pathway scenarios were not yet available.

Rohrbeck discusses alternative modelling strategies for property insurance claims that focus on high temporal resolution rather than high spatial resolution (Rohrbeck et al., 2018). We agree that a full-resolution claims model (that is of high temporal and spatial resolution) would better describe the fine-grained relationship between water damages and specific weather events, and thus potentially provide improved risk estimates. However, it is important to keep in mind the goal of our risk model, namely to price an insurance contract. This goal is not necessarily equivalent to predicting the number of claims on a specific day and, importantly, requires a robust risk assessment. When investigating the daily versions of the weather and climate data used in our analysis and claims models at our high spatial resolution combined with daily temporal resolution, we found that these data products include outliers that significantly affect both risk estimates and predictive performance. These results combined with our findings shown in Figure 1 that the data exhibit a high correlation between quarterly mean precipitation and high quantiles of the daily precipitation distribution, led us to the modelling choices presented in the paper. Alternative applications such as a warning system designed to issue alerts ahead of upcoming rainstorm incidents would require another approach and likely benefit from a high temporal resolution of weather data.

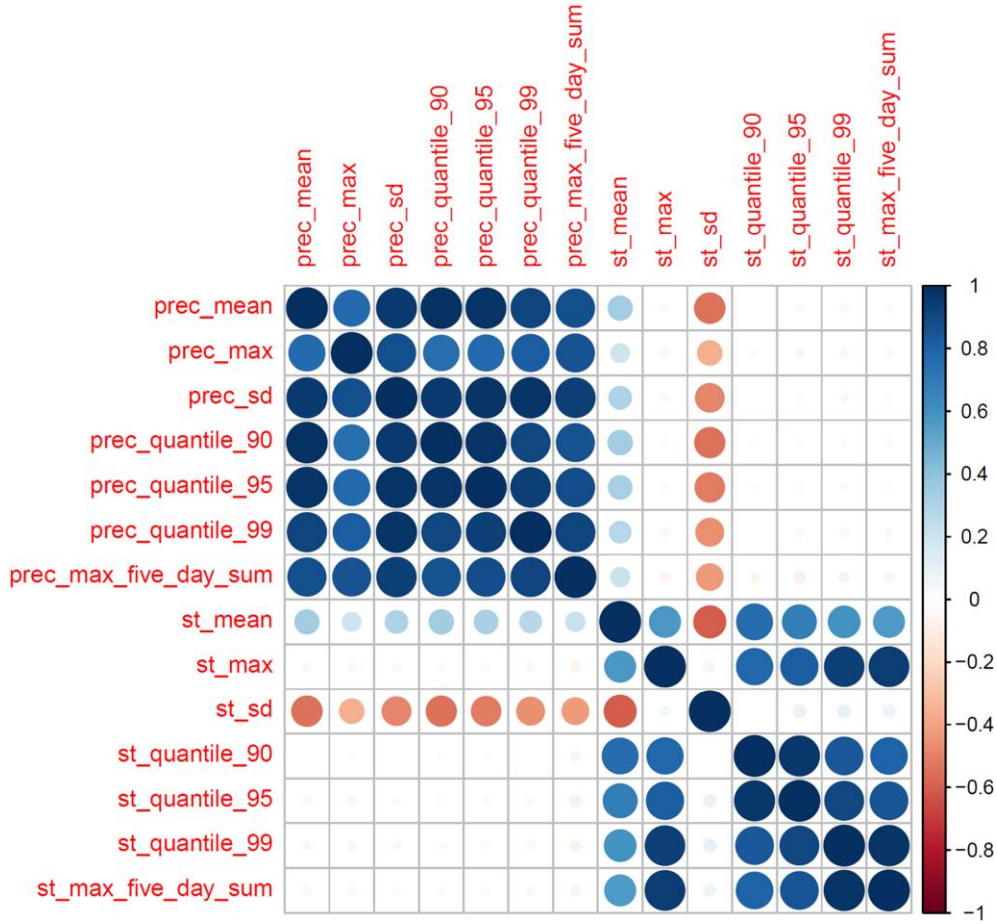

Figure 1. Correlation matrix for a range of climatological indices. The indices are calculated per season and per insured property, from daily values derived from the seNorge data product. The two indices max_five_day_sum consider the highest amount of precipitation (highest sum of daily surface temperatures) over five consecutive dates.

2 Model specification

Several discussion contributions point out that our model does not include interaction terms. Such terms seem natural to include when thinking about the effects of topography, precipitation, and building-specific properties, such as having a cellar. One might even consider a different class of a supervised learning model (such as random forests) that is better suited for a large number of interaction terms than generalised additive models (GAMs). During the model development, we investigated the inclusion of interaction terms. Figure 2 shows the predictive performance of the models analysed in the paper and two GAMs with interactions. The model *t-m int.* includes all climatological and topographical information, and an interaction term between seasonal precipitation and the topographical wetness index (TWI). The model *cellar int.* includes two interaction terms, one between the cellar variable[1] and TWI, and one between the cellar variable and precipitation. As the figure shows, the former model performs worse than the models presented in the paper, both in terms of mean squared error (MSE) and Brier score. While the latter model leads to slightly improved performance in terms of MSE, the model without interaction terms outperforms it slightly under the Brier score.

The inclusion of interaction terms prevents the decomposition into partial risk factors described in Section 3.1 of the paper. This decomposition relies on a model that is additive at the log scale and

[1] Factor-variable specifying whether the property has a cellar or not, or whether this is not known.

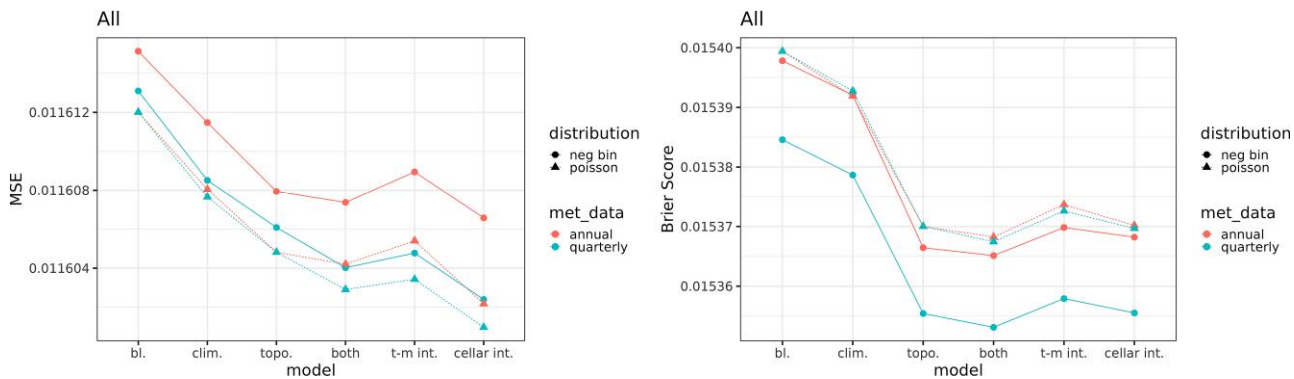

Figure 2. Predictive model performance in terms of mean squared error (MSE) and Brier score (cf. Figure 1 in the paper). The results include two models that allow for interactions: The model t-m int. includes an interaction term between precipitation and Topographical Wetness Index (TWI), while the model cellar int. includes pairwise interaction terms between both precipitation and the cellar variable, as well as TWI and the cellar variable.

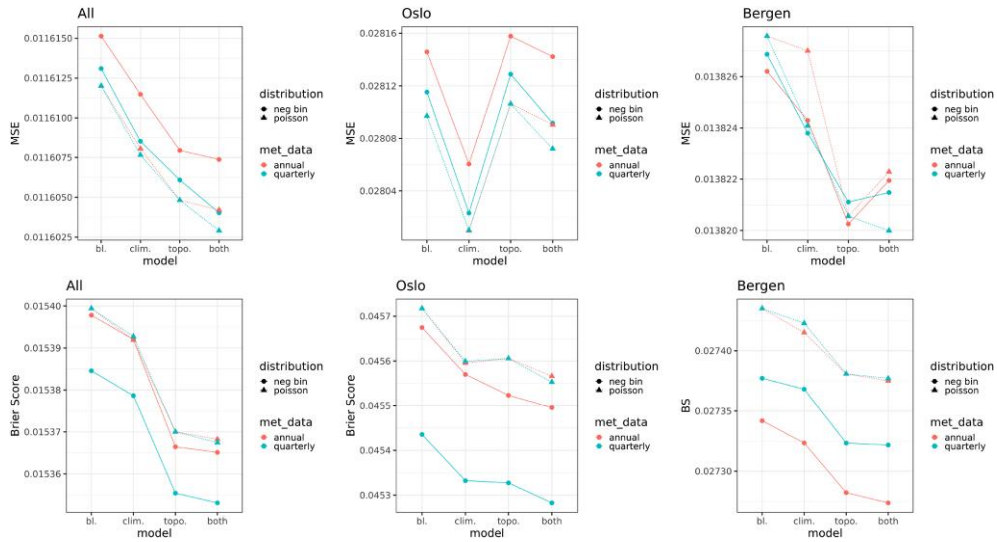

Figure 3. Mean squared error (MSE, top row) and Brier score (bottom row) for evaluation based on all data (left), restricted to Oslo municipality (center) and Bergen municipality (right). The model is estimated using the entire data set, only evaluation is restricted to a subset. Scores are calculated using cross-validation, see the paper for details.

does not include interactions across the partial risk factors (topography, climatology, region, and building-specific). Maps of partial risks, as presented in the paper, are only possible under this risk decomposition, including maps showing future climate risk (Figure 5 in the paper). For example, an interaction term between climate covariates and the cellar variable would result in a separate climate risk map for each level of the cellar variable (including the level "unknown"). Such a separation would make the model output less intuitive to use for decision-makers. Combined with the performance results shown in Figure 2, this led us to exclude interaction terms in the final model.

More generally, the choice of a GAM results in a highly interpretable model. The spline effects shown in Figure 3 of the paper are easy to understand and do not require expert background in statistics or machine learning. We feel that this is particularly important in the context of a study like ours, which is relevant for decision-makers both in insurance and politics, but also of interest to the general public concerned with the impacts of climate change. As mentioned by Legrand and Optiz, the issue of a covariate shift, where the prediction involves extrapolating beyond the support of the covariate distribution for the training data, is typical in climate change impact studies. While methods such as random forests do not allow for flexibility in how covariate shift is handled, GAMs are highly flexible. We are pleased to learn that there is on-going work investigating this more formally than was possible within the scope of our paper, and we look forward to see the results of these efforts.

3 Model verification

Several discussion contributions mention the seemingly small improvements in performance when including climatological and topographical information, referring to the small differences in the scores shown in Figure 1 of the paper (cf. left-hand side plots in Figure 3 of this reply). Specifically, Rougier argues that the model improvements become essentially invisible when the y-scale is expanded to 0, which would correspond to the score of a perfect prediction. While this expansion may seem intuitive at first glance, we believe that it is unjustified, since a perfect prediction is unattainable. Proper scores such as the Brier score or the MSE are designed to rank competing models. However, comparing the scale of improvement to the score of a perfect model is misleading in most cases. An example highlighting this is when predictions are made for real-valued quantities and are evaluated by the logarithmic score which is given by the negative log-likelihood of the prediction model evaluated in the observation. A perfect prediction would here be a point mass in the outcome, which achieves a score of $-\infty$. Clearly, no model improvement would look relevant when the y-scale is extended to $-\infty$.

Table 1. *p*-values of permutation tests comparing the different models in Figure 1 of the paper

		MSE								Brier score					
		Poisson			neg. bin.					Poisson			neg. bin.		
		bl	clim	topo	bl	clim	topo			bl	clim	topo	bl	clim	topo
Quarterly	clim	0	–	–	0	–	–	Quarterly	clim	0	–	–	0	–	–
	topo	0	0.124	–	0	0.14	–		topo	0	0	–	0	0	–
	both	0	0	0	0	0	0		both	0	0	0	0	0	0
Annual	clim	0	–	–	0	–	–	Annual	clim	0	–	–	0	–	–
	topo	0.002	0.088	–	0	0.05	–		topo	0	0	–	0	0	–
	both	0	0	0	0	0	0		both	0	0	0	0	0	0

Note. The tests are based on 500 permutation resamples. Here, bl stands for baseline. The better-performing model is specified by the row, the worse-performing model by the column. *p*-values of exactly zero are estimated when the unpermuted score difference is smaller than the score difference of the permuted resample, for all 500 resamples.

For formally assessing whether the model improvements reported in Figure 1 of the paper are relevant, we can test the significance of the score differences. Specifically, we apply permutation tests as discussed in Good (2013). Table 1 shows p-values of permutation tests comparing the distributions of score values for models with or without topographical and meteorological information. Only models of the same type are compared here, i.e. a model assuming a Poisson distribution based on quarterly data are only compared to other models of this type (corresponding to the upper left quadrant of the table). The table shows that the improvement gained by including climatological and/or topographical information compared to using the baseline model is always highly significant (p-value of 0.2% or lower). The only score differences that are non-significant are the differences in MSE between the model including only climatological information and the model including only topographical information.

Moreover, we can give a rough estimate of the monetary value associated with model improvements on the scale reported in Figure 1 of the paper, highlighting the utility of the improvement for Gjensidige. Utility is more difficult to measure rigorously than significance, since it typically depends on the way the forecast is used in decision-making. Figure 1 shows a decrease in MSE of roughly 0.00001. Since this is on squared scale, we first have to replace the MSE by the mean absolute error, which increases the score difference to roughly 0.0004, between the baseline annual Poisson model and annual Poisson model with climatology and topography information. Since we evaluate predictions for the expected number of damages for single customers, we multiply this value by the number of Gjensidige's customers (more than 500,000), and by the average payout per claim (more than 75,000 NOK). Correspondingly, a drop of 0.0004 in absolute error corresponds to a monetary value of about 17.4 million NOK (about 1.57 million USD) for Gjensidige. While this estimation is crude, it highlights the utility of the observed small error improvements by including topography and meteorology. The massive scaling involved here is the main reason why the y-scale in Figure 1 of the paper should not be interpreted intuitively.

Another important question raised by Rohrbeck concerns to what extent our results are robust when attention is focused on sub-regions. Figure 3 shows the scores for the full data (left column, corresponding to Figure 1 in the paper), as well as the mean scores for the municipalities of Oslo (center column) and Bergen (right column). Oslo and Bergen are Norway's largest urban areas. While it would be interesting to compare to a mostly rural municipality as well, this is difficult in practice, since the scores are less robust in such areas due to the generally low numbers of insured properties. The MSE and the Brier score evaluate slightly different aspects of the predictions in that for the MSE, we compare the predicted expected number of claims to the observed number of claims while for the Brier score, the predicted probability of one or more claims is compared against the occurrence of one or more claims.

For the Brier score, Figure 3 (bottom row) shows that while the average scores are overall somewhat higher in Oslo and Bergen than for the national average, the model rankings regarding inclusion of covariates are similar for the two subsets as they are over the full data set. However, we see that for the negative binomial model the annual model ranks better than the quarterly model in Bergen, contrary to the results for both Oslo and the full data set. This corresponds well with Bergen having a maritime climate with less seasonal variation in precipitation while the more continental climate of Oslo is characterised, among other things, by intense convective precipitation events in summer and early fall that are generally associated with a higher risk of water-related damage. That is, the quarterly model is able to pick up different dynamics across quarters in locations with a strong seasonality in claim frequencies.

For the MSE, cf. Figure 3 (top row), the model rankings are similar for Bergen and the full data set, while the results for Oslo show a reduction in predictive performance if topographical information is included in the model. We believe this effect to be caused by outliers in the Oslo data which particularly affect the MSE rather than the Brier score. The Oslo data are partly of slightly different character than the data for other parts of the country. For Oslo, the data include several housing associations that consist de facto of several apartments, while being listed as a single contract in Gjensidige's data. For housing associations, it is common to observe multiple claims at once while this is rare for other types of housing units. Thus, a small reduction in predicted risk can lead to a large difference in squared error if a claim occurs, since the predicted expected number of claims is usually much smaller than 1. However, we decided not to remove these data points from the data set and, furthermore, it is not straightforward to treat such contracts as multiple contracts since information on the number of units within each housing association is usually

not available in the data set. These considerations are, in fact, a key motivation for us to focus attention on the Brier score as discussed in the paper. Overall, Figure 3 shows that our conclusion of improved model performance through topographical and climatological information is quite robust, despite substantial variability of model performance across different regions.

4 Relation to other work

As mentioned by Raftery, Legrand, and Optiz, the modelling challenges we face here are neither limited to Norway nor the specific application of water damages. Opitz and his collaborators use a more sophisticated spatiotemporal model to jointly model the occurrence intensity and size distribution of wildfires at a daily temporal resolution (Koh et al., 2023; Pimont et al., 2021). Their modelling framework seems well suited for the task, and specifically, it seems appropriate to jointly model the occurrence risk and the expected size of a wildfire. For water damage claims on the other hand, claim sizes depend on external factors such as real estate valuations and construction costs which are challenging to project far into the future. It thus seems appropriate to focus on the occurrence risk of water damages when considering projections into the far future. For modelling the wildfire occurrence intensity, Pimont et al. (2021) and Koh et al. (2023) apply a log-Gaussian Cox process (LGCP) model which is closely linked to our Poisson model, since estimating the LGCP corresponds to performing a Poisson regression with a logarithmic link function. Similarly, as discussed in e.g. Jullum et al. (2020), the negative binomial model may also be linked to certain types of point process models.

Several, free or subscription-based, products exist that inform users in certain regions of local risks from natural hazards such as flooding, wildfires, heat, and wind. Raftery mentions, for example, the free online tool Risk Factor[2] which provides such information for locations in the United States. The methodology behind their assessment of fluvial, pluvial, and coastal flood risk is described in Bates et al. (2021). As Bates et al. (2021) describe, the tool provides a detailed and impressive assessment of water levels. However, it appears that the water level estimates are not associated with building-specific information or damage data. We believe that this is one of the main strengths of our analysis, namely that we are able to construct a building-specific risk model that includes both environmental and building attributes and that is based on building-specific damage data, the key ingredient that facilitates precise individual risk discrimination.

References

Bates P. D., Quinn N., Sampson C., Smith A., Wing O., Sosa J., Savage J., Olcese G., Neal J., Schumann G., Giustarini L., Coxon G., Porter J. R., Amodeo M. F., Chu Z., Lewis-Gruss S., Freeman N. B., Houser T., Delgado M., … Bolliger I. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, *57*(2), e2020WR028673. https://doi.org/10.1029/2020WR028673

Good P. (2013). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. Springer Science & Business Media.

Jullum M., Thorarinsdottir T., & Bachl F. E. (2020). Estimating seal pup production in the Greenland sea by using Bayesian hierarchical modelling. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *69*(2), 327–352. https://doi.org/10.1111/rssc.12397

Koh J., Pimont F., Dupuy J.-L., & Opitz T. (2023). Spatiotemporal wildfire modeling through point processes with moderate and extreme marks. *The Annals of Applied Statistics*, *17*(1), 560–582. https://doi.org/10.1214/22-AOAS1642

Pimont F., Fargeon H., Opitz T., Ruffault J., Barbero R., Martin-StPaul N., Rigolot E., RiviÉre M., & Dupuy J.-L. (2021). Prediction of regional wildfire activity in the probabilistic Bayesian framework of firelihood. *Ecological Applications*, *31*(5), e02316. https://doi.org/10.1002/eap.2316

Rohrbeck C., Eastoe E. F., Frigessi A., & Tawn J. A. (2018). Extreme value modelling of water-related insurance claims. *The Annals of Applied Statistics*, *12*(1), 246–282. https://doi.org/10.1214/17-AOAS1081

https://doi.org/10.1093/jrsssc/qlad067
Advance access publication 16 August 2023

[2] www.riskfactor.com, accessed on July 1, 2023.

Authors' reply to the Discussion of 'The importance of context in extreme value analysis with application to extreme temperatures in the USA and Greenland' at the first meeting on 'Statistical aspects of climate change'

Daniel Clarkson[1], Emma Eastoe[1] and Amber Leeson[2]

[1]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK
[2]Data Science Institute, Lancaster University, Lancaster, UK

Address for correspondence: Daniel Clarkson, Department of Mathematics and Statistics, Lancaster University, Lancaster, UK. Email: d.clarkson@lancaster.ac.uk

We thank the discussants for their engaging comments and contributions to the discussion. The paper and the following discussion have been extremely useful in allowing us to hear a range of opinions on this important area of research, both from a methodological and application perspective. We also thank the authors of the other discussion paper from this discussion meeting, titled 'Assessing present and future risk of water damage using building attributes, meteorology and topography' (Heinrich-Mertsching et al., 2021), for their interesting and well-motivated research, and for contributing to an excellent discussion session in person.

Our aim was to present a short analysis of two different climate change data sets that both required adaptation of standard methodologies in order to conduct an extreme value analysis and thereby answer the research questions posed. The two case studies were intended to demonstrate that useful extreme value analysis could be conducted on a wide range of topics even when standard methodologies are not applicable. Within the limit posed by the short format of paper, we hope to have motivated further applied extreme value analysis that is flexible in methodology depending on the type of data that is being analysed.

The decision not to incorporate spatial dependence into the models was raised in many of the discussants' responses. We decided not to include spatial dependence in the work for several reasons. We agree with the points made by various discussants around the importance of spatial model considerations, and the decision not to include our spatial analysis within this paper was primarily made to emphasise the message of the paper—the importance of using context—than it was to produce the most realistic model. Spatial dependence could certainly be used in both cases and our ongoing work (yet to be published) has developed such a model for the Greenland ice sheet data. Models without spatial considerations allowed us to display at least two applications of extreme value methods to climate change data sets within a short paper whilst maintaining the contextual approach that we were aiming to present.

Furthermore, we feel that in recent years, innovative marginal modelling has been overlooked, yet it is crucial for accurate analysis not least as it is an integral part of any spatial model. By focusing on marginal models, we were able to display the benefits of contextual, parsimonious models, showing that potential modelling solutions to climate change problems do not necessarily require complexity in order to be useful.

Professor Richard L. Smith comments on the zero probability problem were greatly appreciated for providing some additional context on the setting of the work. The analysis of The World Weather Attribution group (Philip et al., 2022)—as Professor Smith mentioned—was a key motivator for our paper. Our aim was to suggest a reasonably simple alternative to the modelling approaches provided by this group that more accurately captured the extent of extreme temperature events. A Bayesian approach could also have achieved this, and is a potential future direction for the work if we get the opportunity to develop it further. The climate variable approach is an

interesting development of climate change-based extreme value analysis. The point made around communication of extreme temperatures and their probabilities is well made and of great importance when considering the impact of this type of analysis. Displaying the probability of extreme temperatures given different scenarios potentially allows for more informed discussions on future policies, and could be useful in communicating the results of extreme value analysis to those outside of the topic area.

We thank Professor Adrian E. Raftery for their thorough discussion of the model choices, and the benefits of both physical considerations and physical models. As mentioned earlier, we agree with the points around the potential of spatial structure to improve the models, especially for some research purposes. The discussion of the research question being asked is well made. The results of the US heatwave analysis could certainly be extended or re-framed depending on the research questions under consideration; our focus was more on the methodology than on the implications of the results. Other models, such as numerical weather prediction models, can be capable of forecasting extreme temperatures such as the 2021 heatwave, and it is reasonable to suggest that the inclusion of physical details within the model structure could improve future analyses. This was easier to display concisely for the Greenland ice sheet case study, as the expertise of the authors was more suited to the physics of ice sheets than of land surface temperatures.

The point made by Professor Raftery on the choice to use ice surface temperatures rather than binary melt data is another interesting point of discussion. This was considered at length early on in our work. We chose to use ice surface temperatures because of the extra level of detail that can be gained from modelling these over binary melt observations. There are clear challenges of working with this type of data, but if we are able to model temperatures accurately we can observe the entire distribution of ice surface temperatures, allowing us to gain insight into broader temperature trends across the ice sheet and across temperatures of all magnitudes, distributional properties of temperatures, and modelled probabilities of melt/specific temperatures in addition to answering the same melt-based questions that can be answered using binary melt data. Temperatures outside of typical melt zones can illustrate larger-scale temperature trends, as temperatures in colder areas may be increasing at different rates to those in melt zones. The motivation for spatial considerations is well made and subsequent work has included spatial model components, and the examples provided by Professor Raftery demonstrate the benefits of this approach.

Dr. Saralees Nadarajah provided interesting comments on the motivation of the work and the use of random effects. We appreciate the highlighting of the countries studied within this paper and we agree that the countries that will feel the greatest immediate impacts of climate change are in the developing world. A major motivation for this study was to demonstrate how conclusions can differ radically depending on the model assumptions. The regions were chosen since data were easy to access, allowing us to focus on modelling assumptions rather than on data challenges. It is a good point that other countries may be at greater risk from the impacts of climate change, and therefore data sets in other areas of the world where impacts are more immediate or of greater risk will be considered when deciding on study regions for future work if they are available. It is also worth noting that a motivation for modelling ice sheet melt is to understand sea level rise, which will impact low-lying, underdeveloped countries—such as Bangladesh—the most. We also appreciate the additional detail of the Pareto type II distribution.

Ankur Dutta's explanation of the Block Maxima approach fits well with our choice to use the alternative Peak over Threshold approach for the US case study, as we wanted to use as much of the data as possible whilst ensuring that the asymptotic assumptions of the approach remained true. Particularly within the context of the heatwave occurring over multiple days, having a threshold made more sense than taking block maxima as it allowed all extreme value to contribute to the model inference. For the Greenland case study, we aimed to understand both the properties of melt temperatures and their prevalence. Initial testing showed standard threshold and Block Maxima approaches to be inconsistent, and no fixed or quantile threshold method could be consistently applied across the ice sheet due to the diversity of surface conditions and the complexity of the melt process.

We thank Dr. Christine P. Chai for their suggestion of applying the Gaussian mixture model to the lower tail in addition to the upper tail. Although it is not the focus of this paper, there is potential for examining trends in the lower tail since the mixture model captures the entire temperature distribution rather than just the upper tail. We previously considered some initial research questions based on lower quantiles of the distribution or negatively extreme temperatures. On the ice sheet, this

could provide evidence of warmer winters, and in other land areas the suggested impacts of such temperatures on humans and properties could be interesting to examine in more detail.

Professor Valérie Chavez-Demoulin, Professor Anthony C. Davison, and Dr. Erwan Koch's description of using r order statistics for each year makes a good case for using such a method when the selection of a threshold is difficult, in addition to providing a potentially more stable interpretation of the model parameters. Consideration of phenomena such as heat-domes would also be interesting to consider, as we chose to examine data only based on local or global temperatures rather than on specific phenomena that could influence temperatures. The notes on the computational aspect of the model are also appreciated, as is the suggestion of further local covariates. In further analysis of the Greenland ice surface temperatures, we used more localised covariates to look at spatial trends in temperatures, and this could similarly be done for the US work.

Professor Miguel de Carvalho, Ms Alina Kumukova, and Vianey Palacios Ramírez's comments on the approach taken in our analysis and other potential modelling considerations suggest other methods that are also valued within a contextual approach to extreme value modelling. The need for Bayesian approaches both semi- and non-parametric is clearly described and thought-provoking, and we appreciate the efforts to use methods beyond standard modelling techniques. It is encouraging to hear the shared enthusiasm for other modelling techniques, and from this and other discussants' comments, the need for Bayesian approaches is highlighted as a useful area for further work. The points around the Extended Generalised Pareto Distribution and non-stationary multivariate extreme value models are also very appreciated. The non-stationary multivariate extreme value models in particular are of interest for future work, and although they were beyond the scope of this paper, adding random effects into these contexts could aid spatial analysis similarly to the other suggestions of adding spatial considerations to our modelling approaches.

We really appreciated the discussion, comments, and feedback on the work throughout the discussion paper process. It was extremely useful to hear so many opinions and comments on our work and that our approach towards extreme value analysis is shared by other researchers. We will aim to take on board as many of the comments as we can, and hope that our work may motivate further contextual work within an extreme value analysis in the near future.

Conflicts of interest: None declared.

Data availability

The MODIS data is available online (https://doi.org/10.5067/7THUWT9NMPDK) from the Multilayer Greenland Ice Surface Temperature, Surface Albedo, and Water Vapor from MODIS, Version 1 dataset. The AWS station data is available to download online (https://doi.org/10.22008/promice/data/aws) from PROMICE. The US weather station temperature data is available to download online (https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily) from The Global Historical Climatology Network daily database.

References

Heinrich-Mertsching, C., Wahl, J. C., Ordonez, A., Stien, M., Elvsborg, J., Haug, O., & Thorarinsdottir, T. L. (2021). Assessing present and future risk of water damage using building attributes, meteorology and topography.

Philip, S. Y., Kew, S. F., van Oldenborgh, G. J., Anslow, F. S., Seneviratne, S. I., Vau-tard, R., Coumou, D., Ebi, K. L., Arrighi, J., Singh, R., van Aalst, M., Pereira Marghi-dan, C., Wehner, M., Yang, W., Li, S., Schumacher, D. L., Hauser, M., Bonnet, R., Luu, L. N., … Otto, F. E. L. (2022). Rapid attribution analysis of the extraordinary heat wave on the pacific coast of the US and Canada in June 2021. *Earth System Dynamics*, *13*(4), 1689–1713. https://doi.org/10.5194/esd-13-1689-2022

https://doi.org/10.1093/jrsssc/qlad059
Advance access publication 28 June 2023