Extreme Value Modelling of Water-related Insurance Claims

Christian Rohrbeck[1]*, Emma F. Eastoe[2], Arnoldo Frigessi[3], Jonathan A. Tawn[2]

[1]STOR-i Centre of Doctoral Training, Lancaster University, [2]Department of Mathematics and Statistics, Lancaster University, [3]Oslo Centre for Biostatistics and Epidemiology

Summary: This paper considers the dependence between weather events, e.g., rainfall or snow-melt, and the number of water-related property insurance claims. Weather events which cause severe damages are of general interest, decision makers want to take efficient actions against them while the insurance companies want to set adequate premiums. The modelling is challenging since the underlying dynamics vary across geographical regions due to differences in topology, construction designs and climate. We develop new methodology to improve the existing models which fail to model high numbers of claims. The statistical framework is based on both mixture and extremal mixture modelling, with the latter being based on a discretized generalized Pareto distribution. Furthermore, we propose a temporal clustering algorithm and derive new covariates which lead to a better understanding of the association between claims and weather events. The modelling of the claims, conditional on the locally observed weather events, both fits the marginal distributions well and captures the spatial dependence between locations. Our methodology is applied to three cities across Norway to demonstrate its benefits.

Keywords: Covariates; Extremal dependence; Extremal mixture; Insurance claims; Mixture modelling; Poisson hurdle model; Spatio-temporal modelling

1 Introduction

Since large parts of society and the economy are weather-sensitive, insurances against undesirable weather events have become an important economical factor. Mills (2005) state that the payout by insurance companies for weather related disasters in developing countries is three times higher than the international aid. In order to set premiums correctly, the insurance companies require accurate models. Thus, it is necessary to understand which characteristics of weather events are responsible for damages. While natural disasters such as Hurricane Katrina, which caused damages

*Address for correspondence: Christian Rohrbeck, STOR-i CDT, Science and Technology Building, Lancaster University, Lancaster, LA1 4YF, UNITED KINGDOM, Email: c.rohrbeck@lancaster.ac.uk

of over \$100 billion in 2005 (Knabb et al., 2005), lead to large monetary losses, the majority of insured losses are related to small scale weather events (Mills, 2005; Botzen and Van Den Bergh, 2008). Damages caused by precipitation are, for instance, studied by Schuster et al. (2006) and Kubilay et al. (2013). In this paper, interest lies in the impact of small-scale weather events, e.g., heavy rain or snow-melt.

Traditionally, in the actuarial literature, the distribution of the total money claimed for a weather event is derived from a model for the joint distribution of the number of claims N and the average claim size S per property affected in the event. Klugman et al. (2012) model this joint distribution, conditional on weather covariates $\mathbf{X}$, as

$$[(N, S) \mid \mathbf{X}] = [S \mid N] \times [N \mid \mathbf{X}] .$$

Hence, the average claim size S is considered to be conditionally independent of the weather effects $\mathbf{X}$, given the number of the claims N. The justification for this assumption is that the severity of the insurance claims depends on various factors, including the wealth of the population, the type of construction, building and repair standards, the age of structures and general economic factors (Department for Environment, Food and Rural Affairs, 2004). A Gamma model is often assumed for the claim size with covariate N (Frees and Valdez, 2008; Haug et al., 2011), although mixed Gamma models have also been suggested (Yip and Yau, 2005).

The most critical part when modelling the distribution of $[(N, S) \mid \mathbf{X}]$ is the distribution of $N \mid \mathbf{X}$ due to the complex and strong effect of the covariates (Scheel et al., 2013). So we focus our study on capturing the relationship between weather covariates and the number of claims. We have a particular focus on the high numbers of claims, as these are the most critical for the insurance industry as this influences re-insurance strategies. We also derive the marginal distribution of N from this model by integrating out the effect of the covariates over their distribution $\pi(\mathbf{X})$, that is,

$$\mathbb{P}(N = n) = \int \mathbb{P}(N = n \mid \mathbf{X} = \mathbf{x}) \times \pi(\mathbf{x})\, d\mathbf{x}, \quad \text{for } n \geq 0. \tag{1}$$

This component of our model is fundamental to any assessment of the impact of climate change for the insurance industry as $\pi(\cdot)$ varies with climate change so the associated marginal for N can be derived for any future period (Sanders and Phillipson, 2003; Jenkins et al., 2008; Botzen and van den Bergh, 2012).

We consider the insurance and weather data used by Haug et al. (2011) and Scheel et al. (2013). The insurance data provide the daily number of claims caused by either precipitation, surface water, snow melt, undermined drainage, sewage back-flow or blocked pipes for all Norwegian municipalities between 1997 and 2006. Let $N_{k,t}$ denote the number of claims on day t for municipality k. Table 1 details the set of meteorological and hydrological covariates $\mathbf{X}_{k,t}$ which are either empirical or model generated with a single value for each covariate for day t and munic-

Table 1: Weather covariates $\mathbf{X}_{k,t}$ provided by the Norwegian Meteorological Institute and the Norwegian Water Resources and Energy Directorate.

Variable	Description	Unit
$R_{k,t}$	Total amount of precipitation in day t (Between 6am on day t to 6am on day $t+1$)	mm
$C_{k,t}$	Mean temperature in day t	°C
$D_{k,t}$	Drainage run-off in day t	mm
$S_{k,t}$	Snow-water equivalent in day t (Amount of water in form of snow)	mm

ipality k. The weather data are derived by spatial interpolation, weighted proportionally to the population density within the municipality. Norway's climate varies spatially due to the country's large geographical extent and the input of the Gulf Stream. For instance, western coastal areas observe relatively mild temperatures and large amounts of rainfall while central (inland) areas, such as Oslo, are drier and have more of a continental climate. These differences are likely to lead to a spatial variation of the claim dynamics and have to be accounted for in the modelling framework.

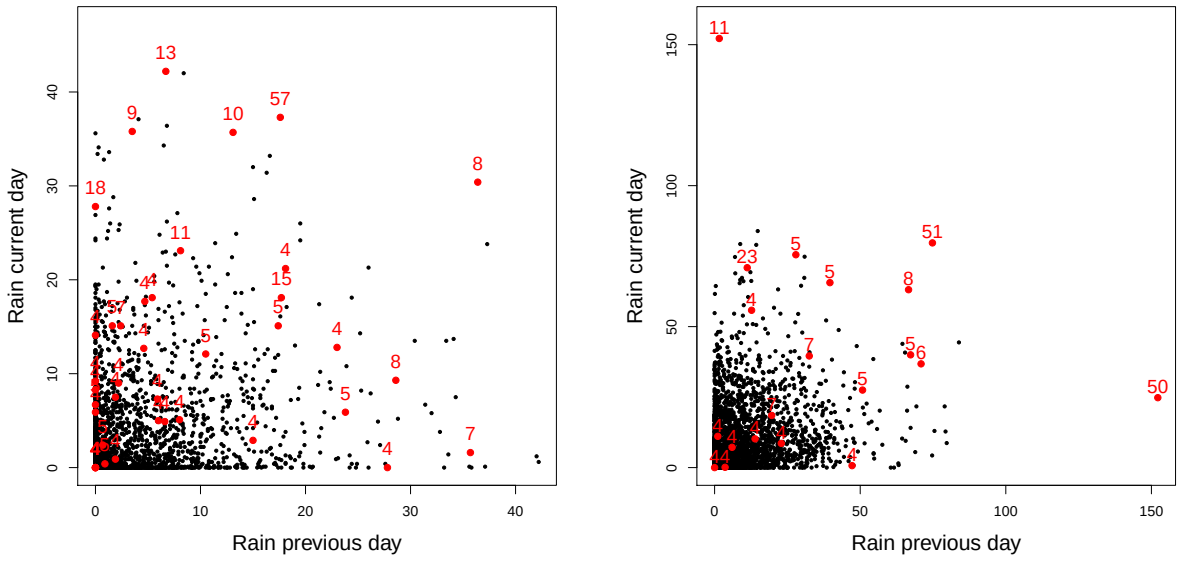
Scheel et al. (2013) propose a Bayesian Poisson hurdle (BPH) model for $N_{k,t} \mid \mathbf{X}_{k,t}$ to account for the frequency of zero claims, $N_{k,t} = 0$, being larger than a standard Poisson model would suggest, and the covariate mechanisms leading to any claim being potentially different from the covariate effects for the number of claims given damage occurred. They also derive additional simple covariates from the covariates in Table 1. Formally, their probability model is then given by

$$\mathbb{P}\left(N_{k,t} = n \mid \mathbf{X}_{k,t}\right) = \begin{cases} \alpha_{k,t} & \text{if } n = 0 \\ (1-\alpha_{k,t}) \, \dfrac{\lambda_{k,t}^n}{n! \left[\exp(\lambda_{k,t}) - 1\right]} & \text{if } n > 0, \end{cases} \tag{2}$$

where both $\lambda_{k,t} > 0$ and $\alpha_{k,t} \in [0,1]$ depend on the covariates $\mathbf{X}_{k,t}$. According to distribution (2), $\alpha_{k,t}$ corresponds to the frequency of zero claims while $\lambda_{k,t}$ is the rate of a zero-truncated Poisson distribution for the number of claims, given at least one claim is reported.

Scheel et al. (2013) assess the predictive performance of their BPH model on a weekly basis and the results are generally positive. Table 2 in Scheel et al. (2013) indicates, however, that their BPH model substantially underestimates the most important feature of the distribution, the high numbers of claims, and hence underpredicts the impact of high precipitation levels, especially for Oslo.

Figure 1 provides some insight into the causes of the lack of model fit for the BPH model by Scheel et al. (2013). Firstly, observations $N_{k,t} > 3$ for Oslo or Bergen are not always associated with high amounts of precipitation on either the claim day t or the previous day $t-1$. While some



(a) Oslo (b) Bergen

Figure 1: Observed covariate values for $R_{k,t}$ and $R_{k,t-1}$ for the original data by Scheel et al. (2013) for (a) Oslo and (b) Bergen. Days with $N_{k,t} > 3$ are highlighted, giving the value of $N_{k,t}$.

claims are linked to weak rainfall coinciding with snow-melt, others occur over periods of mild and dry weather. The latter may be caused by localized weather events which are not recorded by any measurement station. Further, claims caused by blocked pipes or sewage back-flow are not necessarily related to the recent weather. Ignoring such effects may influence the estimated model and lead to biased estimation of the covariate effects. Finally, while claim numbers for Oslo lie between zero and three claims on about 97% of days, much higher numbers occur and these are generally related to high precipitation levels, sometimes in combination with snow-melt. A Poisson distribution is incapable of fitting these extremes while accounting for the high frequency of lower claims.

This paper introduces several new methods in order to improve the model fit which have generic relevance to the modelling of insurance claim data. Interest lies, in particular, in the days with high numbers of claims. We extend the zero-truncated Poisson component in the BPH approach of Scheel et al. (2013) using discrete extreme value and mixture models. Extreme value models such as the generalized Pareto distribution (GPD) are widely applied to estimate the tail of a random variable (Holmes and Moriarty, 1999; Coles, 2001; Li et al., 2005). Here, a discretized analogue of the GPD is used since $N_{k,t}$ takes non-negative integer values only. There are only very limited previous examples of discrete extreme value models (Prieto et al., 2014; Buddana and Kozubowski, 2014) or mixture models (Smith and Goodman, 2000; Bottolo et al., 2003) used in extreme value modelling, and none of these cover cases where both are relevant and non-extreme values are simultaneously modelled, i.e., what we require for modelling the claims data.

In addition to advancing the statistical model, the input data are transformed following an exploratory analysis of the data for Oslo in Figure 1(a). This leads to the derivation of new daily covariates which exploit temporal and spatial patterns in $\mathbf{X}_{k,t}$. Furthermore, we introduce a temporal clustering algorithm to obtain periods of consecutive days which are exposed to the same weather event for each municipality. The distribution of clustered claims, conditional on a set of cluster covariates, is then modelled municipality-wise. Specifically, our methodology is applied to the data for the municipalities of Oslo (Figure 1(a)), Bergen (Figure 1(b)) and Bærum. The model estimates are used to assess dependence of claims over different municipalities, conditionally on the covariates, and to derive the marginal distribution in expression (1) to predict the frequency of extreme claim numbers, under the assumption of no climate change. We find that the clustered claims are spatially independent, given the covariates, indicating that our model has captured the key meteorological factors that explain water-related insurance claims in Norway.

The remainder of this paper is organized as follows: Section 2 details our extensions of the zero-truncated Poisson distribution and introduces an approach to optimize tail dependence for additional covariates. Section 3 defines the new daily covariates and Section 4 introduces the temporal clustering algorithm. Our extended model is then applied to the three Norwegian municipalities in Section 5 where conditional, marginal and spatial properties of the claim process are estimated. The paper concludes with a summary and discussion in Section 6.

2 Extension of the Bayesian Poisson hurdle model

This section details our extensions to the Poisson hurdle model in expression (2). Specifically, we focus on the zero-truncated Poisson component to obtain a better model for claim occurrences, that is, $N_{k,t} \mid (\mathbf{X}_{k,t}, N_{k,t} > 0)$. Since $\lambda_{k,t}$ and $\alpha_{k,t}$ are conditionally independent, given the data (Scheel et al., 2013), any change in this component does not affect $\alpha_{k,t}$. For notational simplicity, the indexes k and t are dropped in the following. Section 2.1 introduces a mixture model while Section 2.2 presents an integer-valued GPD and combines it with the zero-truncated Poisson distribution via an extremal mixture model. The marginal distribution for the number of claims in an event is then derived in Section 2.3. Section 2.4 details a general methodology to optimize the tail dependence between a response and a family of predictors which is later applied in Section 4.

2.1 Mixture modelling

Figure 1, coupled with exploratory analysis, indicates that claim dynamics are mainly driven by the observed precipitation and snow-melt levels. However, some claims are reported over periods which exhibit mild and dry weather, implying the occurrence of unobserved claim processes. Information on the precise cause of damage, e.g., snow-melt or sewage back-flow, may allow the fit of a separate model for each cause but these data are not available.

We propose a two-component mixture distribution with discrete positive-valued random variables Y and Z for $N \mid (\mathbf{X}, N > 0)$ to accommodate for the varying weather-dependence of these claim types. The first component Y captures the dependence of N on the weather covariates $\mathbf{X}$ while the second component Z considers the claims which are caused by unobserved processes. All claims on a day are assumed to come from exactly one of the two components. The probability mass function for $N \mid (\mathbf{X}, N > 0)$ is then formally given by

$$\mathbb{P}\left(N = n \mid \mathbf{X}, N > 0\right) \;\; = p\, \mathbb{P}\left(Y = n \mid \mathbf{X}\right) \;\; + \;\; (1-p)\, \mathbb{P}\left(Z = n\right), \; n \geq 1, \tag{3}$$

where p denotes the probability of $N \mid (\mathbf{X}, N > 0)$ being distributed according to $Y \mid \mathbf{X}$, with the modelling of this distribution discussed in Section 2.2. We assume that Z has a zero-truncated Poisson distribution with rate parameter $\kappa > 0$, with

$$\mathbb{P}\left(Z = n\right) = \frac{\kappa^n}{n!\left[\exp(\kappa) - 1\right]}, \; n \geq 1. \tag{4}$$

Note, the case $p = 0$ in distribution (3) corresponds to the BPH model in (2) without covariate structure and $p = 1$ gives exactly the BPH model if $Y \mid \mathbf{X}$ is a zero-truncated Poisson distribution. The choice of only two components is due to parsimony and the results in Section 5 indicate that this number appears to be sufficient.

2.2 Extremal mixture modelling

Defining the mixture component $Y \mid \mathbf{X}$ in model (3) as a zero-truncated Poisson distribution leads to a poor fit of the extreme claim numbers for Oslo and Bergen in Figure 1. Hence, we extend the model in order to allow for a more flexible tail behaviour. In particular, the lower claim numbers are still modelled as being distributed according to a zero-truncated Poisson model but the highest observations are modelled using an extreme value tail model. First, a distribution for the extremes of a discrete random variable is presented without the consideration of covariates. The zero-truncated Poisson model is then combined with this distribution and covariates are included.

Consider the modelling of $Y \mid Y > u$, where $u \in \mathbb{R}$ is a high threshold. We view that the best modelling approach for the discrete variable Y is to consider it as $Y = \lceil H \rceil$, where H is a continuous random variable. In an extreme value modelling framework, the distribution of H above a high threshold u is generally modelled by a GPD with scale parameter σ_u and shape parameter ξ (Coles, 2001). For a large enough choice of u, the distribution of $H \mid H > u$ is then approximately given by

$$\mathbb{P}\left(H \leq h + u \mid H > u\right) = 1 - \left(1 + \frac{\xi h}{\sigma_u}\right)_+^{-\frac{1}{\xi}}, \; h > 0, \tag{5}$$

where $x_+ = \max(x, 0)$, $\sigma_u > 0$ and $\xi \in \mathbb{R}$, with the value for $\xi = 0$ interpreted as the limit as $\xi \to 0$. We then derive a discretized GPD to model $Y \mid Y > u$ via a GPD for H above threshold $\lfloor u \rfloor$. The probability mass function for $Y \mid Y > u$, for $n > u$, is formally given by

$$\begin{aligned} \mathbb{P}(Y = n \mid Y > u) &= \mathbb{P}(H \leq n \mid H > \lfloor u \rfloor) - \mathbb{P}(H \leq n-1 \mid H > \lfloor u \rfloor) \\ &= \begin{cases} \left[1 + \frac{\xi(n-1)}{\sigma_u}\right]_+^{-\frac{1}{\xi}} - \left[1 + \frac{\xi n}{\sigma_u}\right]_+^{-\frac{1}{\xi}} & \xi \neq 0 \\ \exp\left(-\frac{n-1}{\sigma_u}\right) - \exp\left(-\frac{n}{\sigma_u}\right) & \xi = 0. \end{cases} \end{aligned} \tag{6}$$

In the following, the distribution (6) is termed an *integer-valued Generalized Pareto distribution*, $\text{IGPD}(\sigma_u, \xi, u)$, above threshold u with scale σ_u and shape ξ. Interpretation of the shape parameter ξ is equivalent to that of the GPD: a negative shape parameter $\xi < 0$ corresponds to the distribution being short-tailed, with upper bound. Conversely, $\xi > 0$ indicates a power-law tail, much heavier than a Poisson distribution.

It is interesting to examine how the properties of this distribution vary with the threshold, i.e., how the distribution changes as the threshold is increased to $v > u$. The GPD has a threshold stability property, that is, if $H - u \mid H > u \sim \text{GPD}(\sigma_u, \xi)$, then for any higher threshold $v > u$,

$$H - v \mid H > v \sim \text{GPD}(\sigma_u + \xi(v - u), \xi).$$

As such, ξ is constant with increasing threshold while the scale parameter σ_u is not. An equivalent property also holds for the defined IGPD. In particular, if $Y \mid Y > u \sim \text{IGPD}(\sigma_u, \xi, u)$, then for any $v > u$,

$$Y \mid Y > v \sim \text{IGPD}(\sigma_u + \xi(\lfloor v \rfloor - \lfloor u \rfloor), \xi, v); \tag{7}$$

see Appendix A for the proof. This is important since it allows the selection of a threshold u for the IGPD via a threshold stability property, the same technique as applied for a GPD (Coles, 2001).

Prieto et al. (2014) and Hitz (2017) consider a similar formulation to expression (6). The GPD has an asymptotic justification for its form given by limit results of Pickands (1971) as the threshold tends to the upper endpoint of the distribution. Similar limit results hold for discrete random variables (Anderson, 1970, 1980; Shimura, 2012) but these only hold for $\xi \geq 0$ and are unable to provide non-degenerate limits for the Poisson distribution as the tail decays too quickly. Thus, these limit results do not provide flexible tail models for discrete random variables above non-limit thresholds.

The Poisson distribution does not follow an IGPD exactly above any high threshold u for any value of ξ. However, Anderson et al. (1997) show that asymptotically the distribution of the excesses of the threshold of a Poisson variable GPD limit, with $\xi = 0$, if the threshold and the

Poisson mean parameter tend to infinity at appropriate rates. Therefore, an estimate of ξ that is statistically significantly different from zero for the IGPD indicates that the tail of the underlying distribution is not Poisson.

The IGPD in expression (6) is combined with the zero-truncated Poisson distribution to form an extremal mixture distribution, i.e., a distribution with different forms below and above a threshold u. Such mixtures have been widely studied in a continuous variable setting (Coles and Tawn, 1991; Frigessi et al., 2002; Behrens et al., 2004; Carreau and Bengio, 2009; MacDonald et al., 2011) and the estimation of the threshold u is considered too. Here, observations smaller than or equal to u are modelled as being zero-truncated Poisson distributed while being IGPD otherwise. The probability mass function for $Y \mid \mathbf{X}$ is then given by

$$\mathbb{P}\left(Y = n \mid \mathbf{X}\right) = \begin{cases} \dfrac{\lambda^n}{n!\left[\exp(\lambda) - 1\right]} & 1 \leq n \leq u \\ C_u \, \mathbb{P}\left(Y = n \mid \mathbf{X}, Y > u\right) & n > u, \end{cases} \tag{8}$$

where C_u denotes the probability of a zero-truncated Poisson distribution with parameter λ exceeding u and $\mathbb{P}\left(Y = n \mid \mathbf{X}, Y > u\right)$ is given by model (6). The parameters λ and σ_u both vary with the covariates $\mathbf{X}$, with ξ constant, a standard modelling assumption in extreme value modelling (Coles and Tawn, 1996). Following Eastoe and Tawn (2009), from distribution (7), σ_u needs to be linear in ξ to ensure that the structural form of the model is invariant to the precise choice of threshold. However, σ_u is typically modelled in applications with a log-linear model (Davison and Smith, 1990) which is not of the required form for threshold invariance. To overcome this weakness, we propose taking

$$\sigma_u = \zeta + \exp\left(\beta_0 + \boldsymbol{\beta}^T \mathbf{X}\right), \tag{9}$$

with $\zeta > 0$, $\beta_0 \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^q$, where q denotes the number of covariates.

The model defined via the expressions (3), (4) and (6) leads to $N \mid (\mathbf{X}, N > u)$ being a mixture of an integer-valued GPD, i.e., $Y \mid Y > u \sim \text{IGPD}(\sigma_u, \xi, u)$, and a truncated Poisson distribution, i.e., $Z \mid Z > u \sim \text{tPois}(\kappa, u)$, with mixing probability p_u given by

$$p_u = \frac{p \, \mathbb{P}(Y > u)}{p \, \mathbb{P}(Y > u) + (1 - p) \, \mathbb{P}(Z > u)}. \tag{10}$$

Critically for model and threshold selection is the property of threshold-stability. For any $v > u$ and with $v < u - \sigma_u/\xi$ for $\xi < 0$, the distribution of $N \mid (\mathbf{X}, N > v)$ is also a mixture of $\text{IGPD}(\sigma_v, \xi, v)$, and $\text{tPois}(\kappa, v)$ variables with mixing probability given by expression (10) with $u = v$, and where $\sigma_v = \sigma_u + \xi(\lfloor v \rfloor - \lfloor u \rfloor)$, with σ_u given by representation (7); see Appendix B for the proof. Thus both the distribution and structure of the covariate effect in our model is not a function of the threshold u, provided a sufficiently high threshold is selected.

2.3 Marginal distribution of claims

In addition to wanting to know about how covariates lead to the largest number of claims, we are also interested in estimating the marginal distribution of the number of claims, as explained in the introduction. Although it is possible to estimate $\mathbb{P}(N = n)$ directly with some new statistical model, it is likely to be complex due to the strong effects of the covariates, and it is unlikely to be self-consistent with the conditional distribution of $N \mid \mathbf{X}$. It is more natural to estimate $\mathbb{P}(N = n)$ using the estimated conditional distribution $\mathbb{P}(N = n \mid \mathbf{X})$ described in Section 2.2 since the weather covariates $\mathbf{X}$ describe the key sources of variation of N. Specifically, we can write the marginal survivor function for $v > 0$ as

$$\begin{aligned} \mathbb{P}(N > v) &= \mathbb{P}(N > v \mid N > 0)\ \mathbb{P}(N > 0) \\ &= \int_{\mathbf{x}} \mathbb{P}(N > v \mid \mathbf{x}, N > 0)\, \pi(\mathbf{x})\, \mathrm{d}\mathbf{x}\ \mathbb{P}(N > 0), \end{aligned} \tag{11}$$

where π is the joint density of $\mathbf{X}$ given that $N > 0$. The benefits of conditioning on $N > 0$ first are that we only have to model the distribution of covariates when they lead to a claim and we avoid the need to model $\mathbb{P}(N > 0 \mid \mathbf{X})$. For our model, the term $\mathbb{P}(N > v \mid \mathbf{x}, N > 0)$ is given by expression (3); $\mathbb{P}(N > 0)$ is estimated empirically and the estimation of $\pi(\cdot)$ is discussed below.

From expressions (3) and (11), the distribution of N is given by the following mixture model

$$\mathbb{P}(N > v \mid N > 0) = p \int_{\mathbf{x}} \mathbb{P}(Y > v \mid \mathbf{x})\, \pi(\mathbf{x})\, \mathrm{d}\mathbf{x} + (1 - p)\ \mathbb{P}(Z > v), \tag{12}$$

where Z does not depend on the covariates $\mathbf{X}$ but Y does. We then model the probability $\mathbb{P}(Y > v \mid \mathbf{x})$ using the extremal-truncated Poisson-IGPD mixture model (8) with threshold u. It is generally sufficient to estimate the integral in expression (12) with the empirical distribution of $\mathbf{X} \mid N > 0$ being sufficient for estimating π. However our exploratory analysis has shown that events of the form $Y > v$, where $v \gg u$, i.e., a large marginal quantile of Y, can only be achieved when one of our weather covariates is in the upper tail of its distribution, so we cannot simply use the empirical estimate of π in this case. For this case, we propose a univariate parametric tail model which is applied to the relevant marginal component of $\mathbf{X}$; details are explained in Section 5.3.

2.4 Optimizing tail dependence to develop new covariates

The generalized linear modelling framework by Scheel et al. (2013) has limited ability to account for the interaction effect of multiple risk factors; e.g., snow-melt and rainfall. This is due to a range of reasons, these include: simple interaction terms not capturing the non-linearity of the known physical properties of the relationship, parsimony, and a lack of weight given to extreme events, which is when the signal to noise ratio is at its greatest. These weaknesses motivate our

approach to construct an additional covariate, based upon $\mathbf{X}$, which overcomes these deficiencies and is tuned using the extreme number of claims data.

Specifically, a new covariate X^* is derived non-linearly from $\mathbf{X}$, as $X^* = f(\mathbf{X}, \boldsymbol{\theta})$, with unknown parameters $\boldsymbol{\theta}$ and the function f is selected based on the context of the problem. Since X^* is motivated by the extreme claim numbers, $\boldsymbol{\theta}$ should be selected such that the tail dependence between X^* and N is maximized. As the dependence structure between X^* and N is invariant to the marginal distributions (Nelsen, 2007), we transform the observations of X^* and N to common marginals. Furthermore, as our interest is in extreme values, we use a distribution with a heavy tail to emphasize the extremes and, therefore we map (X^*, N) to Fréchet margins. Specifically, we use the probability integral transformation with the distributions of X^* and N each being estimated empirically; see Section 4.3 for details. Let $(V_1(\boldsymbol{\theta}), V_2)$ denote the transformed variables. Although N is discrete, the approximation by a continuous random variable is reasonable as the focus is on the upper tail which has considerably variability. We adapt the approach by Russell et al. (2016) for maximizing covariate combinations for extreme value analysis to estimate $\boldsymbol{\theta}$. For notational simplicity, we write V_1 instead of $V_1(\boldsymbol{\theta})$ in the following paragraphs and we will return to this notation at the end of the section.

The approach of Russell et al. (2016) is based on the properties of bivariate regular variation (Resnick, 2013), which is a weak assumption. For (V_1, V_2) identically distributed random variables with unit Fréchet margins, bivariate regular variation means that for any Borel set $B \subset [0, 1]$ and $v \geq 1$

$$\lim_{s \to \infty} \mathbb{P}(V_1 + V_2 > sv,\ V_1/(V_1 + V_2) \in B \mid V_1 + V_2 > s) = v^{-1}\Psi(\{B\}), \tag{13}$$

where Ψ is known as the spectral distribution, corresponding to the distribution function of a $[0, 1]$ random variable with mean $\frac{1}{2}$. Critically, bivariate regular variation implies that $V_1 + V_2$ and $V_1/(V_1 + V_2)$ are asymptotically conditionally independent. The weakest tail behaviour between V_1 and V_2 occurs when $\Psi(\{0\}) = \Psi(\{1\}) = 1/2$ and the strongest when $\Psi(\{\frac{1}{2}\}) = 1$, the former and latter corresponding to asymptotic independence (Ledford and Tawn, 1996) and perfect dependence, respectively. The greater the mass that the spectral measure places close to $\frac{1}{2}$, the stronger the tail dependence.

To apply the asymptotic property of bivariate regular variation in practice, we need to be able to estimate Ψ. In practice, we assume that the limit (13) holds for a large finite value of s, i.e., for $0 \leq w \leq 1$, with

$$\mathbb{P}(V_1/(V_1 + V_2) \leq w \mid V_1 + V_2 > \tilde{s}) = \Psi(w),\ \forall \tilde{s} > s. \tag{14}$$

Given observations $\{(V_{1,i}, V_{2,i})\}_{i=1}^m$, Ψ can then be estimated by

$$\tilde{\Psi}_s(w) = \frac{1}{|Q_s|} \sum_{i=1}^{m} \mathbb{1}(V_{1,i} + V_{2,i} > s,\ V_{1,i}/(V_{1,i} + V_{2,i}) \leq w), \tag{15}$$

where Q_s denotes the set of points $(V_{1,i}, V_{2,i})$ with $V_{1,i}+V_{2,i} > s$ and $\mathbb{1}$ corresponds to the indicator function. Empirical estimators of the spectral distribution of this form are widely used (Einmahl et al., 1997). Note, more recent approaches impose a constraint on the mean (Boldi and Davison, 2007; Einmahl and Segers, 2009; de Carvalho and Davison, 2014; Hanson et al., 2017) and may be considered alternatively.

In the next step, we construct an objective function to assess the closeness of the spectral density Ψ to $\frac{1}{2}$. A classic way of measuring dependence in extremes is via the coefficient of asymptotic dependence, χ, defined as

$$\chi = \lim_{s\to\infty} \mathbb{P}(V_2 > s \mid V_1 > s), \tag{16}$$

with $\chi = 0$ corresponding to asymptotic independence, $\chi = 1$ to perfect dependence, and larger values of χ corresponding to stronger asymptotic dependence (Coles et al., 1999). In terms of Ψ we can write

$$\chi = \int_0^1 2\min(w, 1-w)\ \mathrm{d}\Psi(w). \tag{17}$$

Here, the term $2\min(w, 1-w)$ can be viewed as a weighting term, which downweights any departure of Ψ from $\{\frac{1}{2}\}$ as the weighting gives the value 1 at $w = \frac{1}{2}$ and the weighting effect on χ decreases linearly away from this point. In empirical studies, using χ as an objective function to maximize over $\boldsymbol{\theta}$, we found that the χ measure does not downweight strongly enough values of $\boldsymbol{\theta}$ that lead to $\Psi_s(w)$ putting mass near 0 and 1 (i.e., very weak dependence) and, hence, results in poor inference for $\boldsymbol{\theta}$. Part of the reason is that the threshold s is finite, so mass that should be at 0 and 1 in the limit as $s \to \infty$ is still away from these values, implying stronger dependence than is really present.

Instead of χ, we want a functional of $\Psi(w)$ which downweights large departures of $V_1/(V_1+V_2)$ from $\frac{1}{2}$ more strongly, in particular giving them zero weight if $V_1/(V_1+V_2)$ is within ϵ, $0 < \epsilon < \frac{1}{2}$ of 0 or 1 to overcome the sub-asymptotic choice of the threshold s in practice. For fixed ϵ, we propose the functional

$$D_\epsilon = \int_0^1 \left[1 - \min\left\{ \left| \frac{\log\left(\frac{w}{1-w}\right)}{\log\left(\frac{\epsilon}{1-\epsilon}\right)} \right|, 1 \right\}\right] d\Psi(w). \tag{18}$$

This functional, with a "tent-like" weighting function, has some similar properties to χ, such as $D_\epsilon = 0$ and 1 for asymptotic independence and perfect dependence, respectively, and increasing values indicate stronger asymptotic dependence. However, the key differences between χ and D_ϵ are that if $\Psi(w)$ puts all its mass within ϵ distance of 0 and 1, then $D_\epsilon = 0$ but $\chi > 0$ and also that D_ϵ does not weights small departures of $V_1/(V_1+V_2)$ from $\frac{1}{2}$ as much as χ does. Thus, there are advantages of using D_ϵ over χ for estimating $\boldsymbol{\theta}$ in order to give a strong relationship between X^* and N in their extremes.

The defined functional D_ϵ can then be used to estimate the set of parameters $\boldsymbol{\theta}$, with each

value of $\boldsymbol{\theta}$ providing a different estimate $\{\tilde{\Psi}_s(w;\boldsymbol{\theta}) : 0 \leq w \leq 1\}$. Here, the dependence measure $D_\epsilon = D_\epsilon(\boldsymbol{\theta})$ is estimated using

$$\begin{aligned}\tilde{D}_{\epsilon,s}(\boldsymbol{\theta}) &= \int_0^1 \left[1 - \min\left\{ \left| \frac{\log\left(\frac{w}{1-w}\right)}{\log\left(\frac{\epsilon}{1-\epsilon}\right)} \right|, 1 \right\}\right] d\tilde{\Psi}_s(w;\boldsymbol{\theta}) \\ &= 1 - \frac{1}{|Q_s|}\sum_{i=1}^{m} \min\left\{ \left| \frac{\log\left(\frac{V_{1,i}(\boldsymbol{\theta})}{V_{2,i}}\right)}{\log\left(\frac{\epsilon}{1-\epsilon}\right)} \right|, 1 \right\} \mathbb{1}(V_{1,i}(\boldsymbol{\theta}) + V_{2,i} > s).\end{aligned} \tag{19}$$

We select $\boldsymbol{\theta}^*$ as $\boldsymbol{\theta}^* = \text{argmax}_{\boldsymbol{\theta}}\ \tilde{D}_{\epsilon,s}(\boldsymbol{\theta})$ and set $X^* = f(\mathbf{X}, \boldsymbol{\theta}^*)$. The choice of ϵ depends primarily on sample size and on $\tilde{\Psi}$, with the larger the sample size and the more concentrated $\tilde{\Psi}_s(w;\boldsymbol{\theta})$ about $\frac{1}{2}$ leading to smaller and larger ϵ, respectively.

3 Defining new daily covariates

The original covariates $\mathbf{X}_{k,t}$ in Table 1 mainly summarize the daily weather conditions. However, the daily resolution and the derivation of the covariate values via weighted spatial interpolation induce inaccuracy in the input variables as critical information, such as the maximum rainfall intensity, is smoothed out over time and space. Consequently, weather events which induce a substantial difference in the claim risk may appear similar in terms of $\mathbf{X}_{k,t}$ and, hence, lead to an underestimation of the effect of severe rainfall events. With a view to reducing this uncertainty, we analyze the spatial and temporal structure in $\mathbf{X}_{k,t}$ for the highest daily claim numbers in Oslo. This analysis motivates the introduction of three new physically/topologically motivated daily covariates for each municipality which exploit the spatial and temporal patterns in $\mathbf{X}_{k,t}$ as an additional source of information. The generic relevance of these covariates is demonstrated by applying them to both Bergen and Bærum in Section 5.

The absence of more detailed weather data excludes the possibility of a more structured construction of covariates using physical rainfall-runoff models. Similarly, using machine learning approaches to derive empirical relationships proved unsuccessful relative to our approach of covariate construction, as it fails to account for our knowledge of physical/topological and neighbourhood structures between municipalities.

Section 3.1 introduces a covariate to capture the amount of snow-melt affecting the properties. Sections 3.2 and 3.3 then define covariates associated to the temporal and spatial rainfall patterns, respectively. In the following, the notation $k' \sim k$ refers to municipalities k and k' being adjacent.

3.1 Snow-melt

Long periods of snow-melt, or rapid melts of large volumes of snow, can give flood levels that are comparable to large rainfall events. Hence, periods of high temperatures or rain, conditional on

snow being on the ground, may affect the claim dynamics and induce a higher risk for property damages. Information on the level of snow-melt is derived via the daily observed mean temperature $C_{k,t}$ and the snow-water equivalent $S_{k,t}$. Scheel et al. (2013) consider the difference in the snow-water equivalent over a day, that is, $S_{k,t-1} - S_{k,t}$. Positive values then represent an additional source of water for properties to deal with while negative values correspond to a rise of the amount of snow on the ground.

We argue that $S_{k,t-1} - S_{k,t}$ is limited in its capability to capture the risk induced by snow-melt. Firstly, an explanatory analysis concluded that a negative difference does not affect the claim dynamics on the day. Secondly, positive values of $S_{k,t-1} - S_{k,t}$ only approximate the true amount of snow-melt in municipality k. Certain topological factors are likely to be ignored since observations are weighted according to the population density. Consider a city which lies at the foot of a mountain range. Properties are then affected by the snow-melt both within the city and on higher ground while $S_{k,t-1} - S_{k,t}$ captures the former only.

We use the observations for the adjacent municipalities to introduce a new snow-melt covariate $\Delta S_{k,t}$ as a spatially weighted average. In particular, our formulation for $\Delta S_{k,t}$ varies from Scheel et al. (2013) as $\Delta S_{k,t} > S_{k,t-1} - S_{k,t}$ if an adjacent municipality exhibits higher levels of snow-melt. Formally, $\Delta S_{k,t}$ is defined by

$$\Delta S_{k,t} = \frac{1}{1+\omega_k^S}\left[S_{k,t-1} - S_{k,t} + \omega_k^S \max_{m\in\{k,k'\sim k\}} (S_{m,t-1} - S_{m,t})\right] \mathbb{1}_{\{C_{k,t}\geq 0\}}, \tag{20}$$

with weight $\omega_k^S \geq 0$. The maximum term in (20) is derived over the set of adjacent municipalities $k' \sim k$ and k itself. Note, $\Delta S_{k,t} = S_{k,t-1} - S_{k,t}$ if the snow-melt in municipality k exceeds snow-melt in its neighbours and $C_{k,t} > 0$, or if $\omega_k^S = 0$. The indicator function is included in order to ensure that no snow-melt occurs for temperatures $C_{k,t}$ below 0°C.

3.2 Surface water

An increased claim risk is induced by the interaction of multiple weather events or the duration of one event over consecutive days. Scheel et al. (2013) attempt to account for such processes via the values of two covariates: the drainage run-off $D_{k,t}$ and the aggregated rain on the previous three days, denoted by $R_{k,3t}$. Their results indicate that both $R_{k,3t}$ and $D_{k,t}$ have a small effect on the distribution of $N_{k,t} \mid N_{k,t} > 0$. However, $R_{k,3t}$ and $D_{k,t}$ are limited in their potential to explain interaction effects. Values for $D_{k,t}$ change very slowly from day to day, that is, $D_{k,t}$ may be high despite the last rain being several days ago. Further, $R_{k,3t}$ cannot distinguish whether high amounts of rainfall were recorded two or three days ago. The derivation of new covariates appears advisable.

To help our construction of a new covariate, we consider a highly idealized model of the ability of infrastructure to handle surface water. Assume that a maximum c_k mm of water drains off

within a day. Here, the value c_k may correspond to a certain quantile of the observed rain and be linked to the capacity of the drainage system. The amount of water left in the system on day t, $W_{k,t}$, is then given by

$$W_{k,t} = \left(W_{k,t-1} + R_{k,t-1} + \Delta S_{k,t-1} - c_k\right)_+ . \tag{21}$$

A value of $W_{k,t}$ greater than 0 implies that the previous weather events affect the risk induced by the weather on day t, for instance, in the form of surface water. Further, $W_{k,t}$ is assumed to influence the claim dynamics if, and only if, $R_{k,t} + \Delta S_{k,t} > c_k$ since the value of $W_{k,t}$ in (21) decreases otherwise, implying that no additional properties are threatened by surface water. This results in the definition of a new *amplifier* covariate,

$$A_{k,t} = W_{k,t} \mathbb{1}_{\left\{R_{k,t} + \Delta S_{k,t} > c_k\right\}}, \tag{22}$$

which captures the risk induced by heavy rainfall in combination with high surface water levels.

3.3 Rainfall intensity

Since $R_{k,t}$ corresponds to the aggregated precipitation measurements over 24 hours, it provides little insight into the peak-daily intensity. High values of $R_{k,t}$ can be due to either short-term intense or longer-term moderate rainfall but the former is likely to induce a higher risk for property flooding. We attempt to derive additional information from the spatial variation of $\{R_{k,t}\}$ on day t. To achieve this, we assume that the intensity correlates with the difference in the precipitation levels of adjacent municipalities. Further, an intense rainfall within a municipality is also taken to affect the claim dynamics of the adjacent municipalities, though on a smaller scale.

These considerations result in our definition of the covariate *intensity*, $I_{k,t}$, which is based on the spatial pattern of $\{R_{k,t}\}$ at day t. Let $\widetilde{k}$ be the municipality, adjacent to municipality k, with the highest level of precipitation, i.e.,

$$\widetilde{k} = \underset{k' \sim k}{\operatorname{argmax}} \; R_{k',t}.$$

If $R_{k,t} \geq R_{\widetilde{k},t}$, the centre of the rainfall event lies within municipality k and, hence, may be rather intense. Similarly, if $R_{\widetilde{k},t} > R_{k,t}$, we consider the adjacent municipalities $k' \sim \widetilde{k}$ to explore whether the rainfall event leads to the highest precipitation levels in municipality $\widetilde{k}$. In order to represent the impact of a rainfall event at municipality $\widetilde{k}$ for municipality k, we introduce a weight $\omega_k^R \in [0, 1]$ to downscale the intensity. Finally, if the rainfall is centred in neither of these municipalities, the

rainfall is considered as not intense. The covariate value $I_{k,t}$ is then defined as

$$I_{k,t} = \begin{cases} R_{k,t} - R_{\widetilde{k},t} & \text{if } R_{k,t} > R_{\widetilde{k},t} \\ \omega_k^R \left(R_{\widetilde{k},t} - \max\limits_{k' \sim \widetilde{k}} \; R_{k',t} \right) & \text{if } R_{\widetilde{k},t} > \max\limits_{k' \sim \widetilde{k}} \; R_{k',t} \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

Note, the last case in (23) corresponds to the municipalities k and $\widetilde{k}$ observing lower precipitation levels than at least one of their adjacent neighbours. The upper bound for ω_k^R is justified since $I_{k,t}$ should not be higher than $I_{k',t}$ if the highest precipitation levels are recorded for municipality k'. Similarly to $A_{k,t}$, $I_{k,t}$ only affects the claim dynamics for high rainfall levels, $R_{k,t} > c_k$, since the intensity of the rainfall is presumably not important for the claim dynamics otherwise.

4 Clustering approach

We introduce an algorithm to obtain clusters of consecutive days which are exposed to the same severe weather event. This approach is motivated by the observed dependence between $N_{k,t}$ and $N_{k,t+1}$, in particular, for their large values. For instance, the highest rainfall level in Figure 1(b) results in observations of 11 and 50 claims on consecutive days. From a practical perspective, these observations are mainly due to two processes. Firstly, the recording process is lagged as some policy holders report a damage the same day while others do so the following day. Secondly, the daily resolution potentially splits a weather event across two or more observations. Hence, it is desirable to derive time periods, such that each of them covers a weather event and the subsequent period of elevated claim risk, where the latter may correspond to no claims arising.

Section 4.1 details our cluster algorithm which derives such periods of consecutive days, based on the covariates, and thus reduces the effects of claim lag in the recording process. Covariates summarizing the weather events over the cluster periods are defined in Section 4.2. The event-based covariates are then tuned to increase their ability to describe the occurrence of the largest numbers of claims in Section 4.3. We conclude by defining a probability model for the association between the clustered number of claims and weather covariates in Section 4.4, using the approaches introduced in Section 2.

4.1 Derivation of cluster periods

Interest lies in the derivation of J_k clustered weather periods for municipality k, $\{(\alpha_{k,j}, \beta_{k,j})\}_{j=1}^{J_k}$, based upon $\mathbf{X}_{k,t}$, where $\alpha_{k,j}$ and $\beta_{k,j}$ represent the start and end point, respectively, of the jth cluster. While the daily claims within a cluster period $[\alpha_{k,j}, \beta_{k,j}]$ are assumed to depend on the same weather event, the claims in two different clusters are considered as temporally independent.

In particular, the claim dynamics on day $\alpha_{k,j}$ are solely dependent on the weather events on the same day, irrespective of the weather on day $\beta_{k,j-1}$.

Our approach to identify cluster start points $\alpha_{k,j}$ is based upon two pre-specified trigger events which affect the claim dynamics on subsequent days: rain on the current day exceeding c_k, $R_{k,t} > c_k$, and snow-melt occurring, $\Delta S_{k,t} > 0$. The first trigger event is motivated by the discussion in Section 3.2 while the second trigger reflects our expectation that snow-melt in combination with rainfall induces a high claim risk over several days. These events then initialize clusters of length greater than or equal to one day. The main criterion for the end of a cluster considers the change in the drainage run-off, i.e., $\Delta D_{k,t} = D_{k,t} - D_{k,t-1}$. In particular, a cluster period ends if $\Delta D_{k,t}$ drops below a threshold d_k. Additionally, clusters triggered by snow-melt also end if no snow is left on the ground. The cluster approach described above results in Algorithm 1.

Algorithm 1 Derive clusters for municipality k

Require: Weather covariates $\Delta S_{k,t}$, $\Delta D_{k,t}$, $R_{k,t}$, and thresholds c_k and d_k

Go to first time point $t = 1$
while Unclustered observations left **do**
    if $\Delta S_{k,t} > 0$ **then**
        Set start point $\alpha = t$ and initial end point $\beta = t + 1$
        while $\Delta D_{k,\beta} > d_k$ AND $\Delta S_{k,\beta} > 0$ **do**
            Shift end point $\beta \leftarrow \beta + 1$
    else if $R_{k,t} > c_k$ **then**
        Set start point $\alpha = t$ and initial end point $\beta = t + 1$
        while $\Delta D_{k,\beta} > d_k$ **do**
            Shift end point $\beta \leftarrow \beta + 1$
    else
        Set start and end point to $\alpha = \beta = t$
    Store start and end points of cluster period (α, β)
    Go to next time point $t = \beta + 1$
return Cluster periods

4.2 Cluster data

The daily data have to be adapted to the cluster periods derived by Algorithm 1. Consider the jth cluster period for municipality k with start and end point $\alpha_{k,j}$ and $\beta_{k,j}$, respectively. Instead of the daily numbers of claims, interest lies the aggregated number of claims over the jth cluster period which is given as

$$\widetilde{N}_{k,j} = \sum_{t=\alpha_{k,j}}^{\beta_{k,j}} N_{k,t}. \tag{24}$$

While adaption of the original response $N_{k,t}$ to the cluster periods is straightforward, more care is required for the explanatory variables. Scheel et al. (2013) find that the amount of rainfall is

correlated with $N_{k,t}$ for Oslo, Bergen and Bærum in terms of the Poisson component of the hurdle model. Further, the results also suggest that snow-melt is informative for Bergen. Our analysis also revealed that snow-melt is informative for Oslo when accounting for spatial patterns. We thus derive cluster covariates which capture information related to these events. As the daily amount of snow-melt does not take very high values, snow-melt for cluster j is summarized via one covariate, the accumulated snow-melt over the cluster period

$$\Delta S^{\Sigma}_{k,j} = \sum_{t=\alpha_{k,j}}^{\beta_{k,j}} \Delta S_{k,t}, \tag{25}$$

where $\Delta S_{k,t}$ is defined via expression (20).

Considering rainfall, intense rainfall on a day and longer-term rainfall scenarios have to be accounted for. To capture these characteristics, we define two covariates $R^{\max}_{k,j}$ and $R^{\Sigma}_{k,j}$, respectively. While $R^{\max}_{k,j}$ focuses on a single day over the cluster period, $R^{\Sigma}_{k,j}$, takes the amount of precipitation over all days into account. Let γ_j denote the day with highest value $R_{k,t}$ over the period $\alpha_{k,j}$ to $\beta_{k,j}$, i.e., $\alpha_j \leq \gamma_j \leq \beta_j$. Then

$$R^{\max}_{k,j} = \eta_k \; A_{k,\gamma_j} + R_{k,\gamma_j} \exp\left(\rho_k \; I_{k,\gamma_j}\right), \tag{26}$$

where A_{k,γ_j} and I_{k,γ_j} are defined as in (22) and (23), respectively. The parameters η_k and ρ_k are selected to optimize the tail dependence of $R^{\max}$ and $\widetilde{N}$, details are given in Section 4.3. The non-linear structure of expression (26) aims to account for two separate claim processes which are associated to rainfall. In particular, the first additive component accounts for the risk in terms of surface water induced by previous rainfall events while the second component considers the rainfall on the day. The impact of the rainfall on the day for claims depends on both the rainfall and its intensity. Our arguments for the construction of the covariates $A_{k,t}$ and $I_{k,t}$ suggests that $\eta_k \in [0, 1]$ and $\rho_k \geq 0$. Covariate $R^{\Sigma}_{k,j}$ is

$$R^{\Sigma}_{k,j} = \sum_{t=\alpha_{k,j}}^{\beta_{k,j}} R_{k,t} - R_{k,\gamma_j}, \tag{27}$$

i.e., the aggregation of the rainfall, except for the highest day, in the cluster. Note, $R^{\Sigma}_{k,j}$ takes the value zero if the jth cluster is of length 1.

4.3 Selection of parameter values

The covariates introduced in this work depend on several parameters whose tuning is considered in this section. First, the parameter ω^S_k in (20) is selected based upon a simple generalized linear model fit for the original daily data for municipality k. The parameter ω^S_k has to be estimated

prior to the cluster algorithm since it is important to gain insight into whether $\omega_k^S = 0$ or not. The maximum likelihood estimator of ω_k^S is found using the model

$$N_{k,t} \sim \text{Poisson}\left(\exp[\phi_0 + \phi_1 \Delta S_{k,t}(\omega_k^S)] \right).$$

The parameter may be estimated again after the clustering algorithm but the results in Section 5 are obtained without this additional step.

The vector of covariate observations of the maximum rainfall covariate in expression (26), $\mathbf{R}_k^{\max}$, depends on the parameters ρ_k, η_k and also on the weight ω_k^R via $I_{k,t}$. Since $I_{k,t}$ and $A_{k,t}$ are predominately designed with respect to the high numbers of claims, ρ_k, η_k and ω_k^R are selected such that the tail dependence between $\mathbf{R}_k^{\max}$ and $\widetilde{\mathbf{N}}_k$ in expression (24) is maximized. Here, we adapt the approach detailed in Section 2.4 with $X^* = f(\mathbf{X}, \boldsymbol{\theta}_k) = R_k^{\max}$ given by expression (26) and the optimization is over a set of candidates for $\boldsymbol{\theta}_k = \left(\eta_k,\ \rho_k, \omega_k^R\right)$. This involves first transforming the data to Fréchet margins, selecting a threshold s above which the conditional independence property (13) holds, then estimating $\tilde{\Psi}_s(w; \boldsymbol{\theta}_k)$ and finally deriving the distance measure $\tilde{D}_{\epsilon,s}(\boldsymbol{\theta}_k)$ for each candidate. Combining these ideas leads to the following selection process for the optimal candidate:

1. Derive the covariate values $\mathbf{R}_k^{\max}(\boldsymbol{\theta}_k)$ for each candidate $\boldsymbol{\theta}_k$ on a grid.

2. Use the empirical distribution functions and the probability integral transform to transform $\mathbf{R}_k^{\max}$ and $\widetilde{\mathbf{N}}_k$ to have Fréchet margins

$$\mathbf{N}^* = -\left\{ \log\left[\frac{rank\left(\widetilde{\mathbf{N}}_k\right)}{m+1} \right] \right\}^{-1} \quad \text{and} \quad \mathbf{R}^* = -\left\{ \log\left[\frac{rank\left(\mathbf{R}_k^{\max}\right)}{m+1} \right] \right\}^{-1}$$

3. The threshold s in (15) is chosen as a 99.5% quantile of the set $\{\mathbf{N}^* + \mathbf{R}^*\}$. Further set

$$Q_s = \{i = 1, \ldots, m \ : \ N_i^* + R_i^* > s := q_{0.995}\left(\mathbf{N}^* + \mathbf{R}^*\right)\}.$$

4. Derive the distance measure as outlined in Section 2.4 by substituting $V_{1,i} = N_i^*$ and $V_{2,i} = R_i^*$ into the distance measure in (19), where we used $\epsilon = 5 \times 10^{-4}$.

5. The optimal set of parameters $\boldsymbol{\theta}_k^*$ is then the one which provides the maximum of $\tilde{D}_{\epsilon,s}(\boldsymbol{\theta}_k)$.

4.4 Statistical model and inference for clustered claims

We consider the association between the response $\widetilde{N}_{k,j}$ and the covariates $\widetilde{\mathbf{X}}_{k,j} = \left(R_{k,j}^{\Sigma}, \Delta S_{k,j}^{\Sigma}, R_{k,j}^{\max}\right)$ in expressions (24) through (27). Specifically, only cluster periods with at least one claim are considered, that is, the distribution $\widetilde{N}_{k,j} \mid \left(\widetilde{\mathbf{X}}_{k,j}\widetilde{N}_{k,j} > 0\right)$. The extremal mixture model introduced

in Section 2 is applied and so $\widetilde{N}_{k,j} \mid \left(\widetilde{\mathbf{X}}_{k,j}\widetilde{N}_{k,j} > 0\right)$ is modelled via a mixture of two random variables $\widetilde{Y}_k$ and $\widetilde{Z}_k$ with mixture probability p_k. Here, $\widetilde{Y}_k$ is distributed according to expression (8) with the scale $\sigma_{k,u}$ and the rate λ_k varying in the covariates while the shape ξ_k is constant. Formally, we define

$$\begin{aligned} \log(\sigma_{k,u} - \zeta_k) &= \beta_{k,0} + \beta_{k,1} R^{\Sigma}_{k,j} + \beta_{k,2} \Delta S^{\Sigma}_{k,j} + \beta_{k,3} R^{\max}_{k,j} \\ \log \lambda_k &= \delta_{k,0} + \delta_{k,1} R^{\Sigma}_{k,j} + \delta_{k,2} \Delta S^{\Sigma}_{k,j} + \delta_{k,3} R^{\max}_{k,j}. \end{aligned} \tag{28}$$

The component $\widetilde{Z}$ is defined as a zero-truncated Poisson distribution with rate κ_k. As noted in Section 2.2 this model for $\widetilde{N}_{k,j}$ is stable in its distribution and covariate representational forms for all choices of the threshold u_k.

For the data considered in the following Section 5, we found strong evidence, by using the Bayesian Information Criterion (BIC) (Schwarz, 1978), that for our chosen thresholds $\zeta_k = 0$ could be taken without loss of efficiency. This has the benefit of parsimony (reducing the number of parameters to 11) but removes the threshold-stability of the covariate model (28) for the scale parameter of the IGPD. Conclusions of the statistical analysis are approximately unchanged by our choice, but we note that others may prefer to have retained the ζ_k parameter in the inference.

The selected statistical model is, thus, specified by 11 parameters which are estimated via Bayesian inference. Specifically, a Metropolis-Gibbs algorithm is used which updates each parameter values individually in turn; see Appendix C for details. Alternatively, estimates may be obtained via an Expectation-Maximization algorithm. However, we found that this led to poor estimates since the support of $\widetilde{Y}$ varies in the shape parameter ξ_k, in particular, in case $\xi_k < 0$.

5 Application to the insurance data

We apply the methodology developed in Sections 2–4 to address the features of the insurance claims data we identified in the introduction. In this section, we present results for the municipalities of Oslo, Bærum and Bergen, where the first two are adjacent and the latter is approximately 300 miles away from them. Oslo and Bergen were chosen since they have both the highest number of policies and the largest average number of claims per day. Bærum was selected as it records the highest daily claim number over the 10-year period for Norway.

Section 5.1 considers the derivation of the cluster weather periods and Section 5.2 explores the model estimates. The marginal distribution for N is then derived in Section 5.3 and we illustrate its use in predicting the frequency of very large claims. As our covariate selection in Section 3 involves some parameters that were either chosen or estimated, we conduct a sensitivity analysis in Section 5.4 to illustrate that the uncertainty of this stage of the analysis does not lead to any major changes in our overall model fit. Finally, Section 5.5 investigates the extent to which the fitted covariate model captures the spatial dependence between claims for the adjacent municipalities of

Table 2: Occurrence of cluster lengths for the three Norwegian municipalities considered in Section 5.

Cluster length	1	2	3	4	5	6	> 6
Oslo	2091	254	57	98	43	23	17
Bærum	2453	105	43	92	46	19	18
Bergen	1868	340	55	131	39	23	11

Oslo and Bærum.

5.1 Derivation of the cluster data

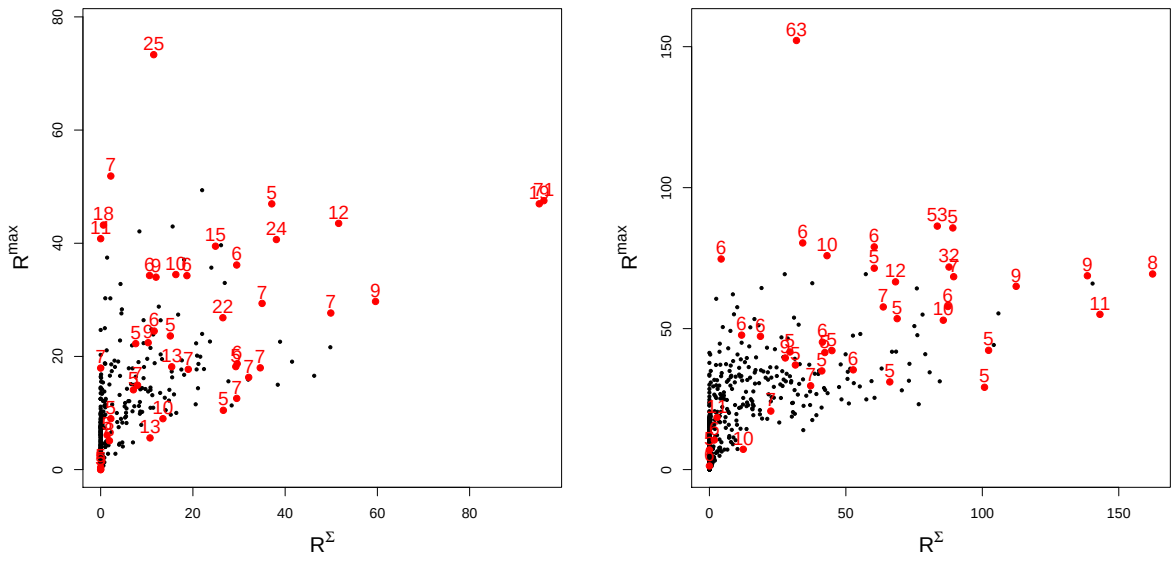
The first step to deriving the cluster periods for each of the three municipalities, via Algorithm 1, is the estimation of the weight ω_k^S in expression (20), as this is required for the snow-melt covariate $\Delta S_{k,t}$. We do this using the method in Section 4.3 with ω_k^S being found to be positive for Oslo and Bergen, cities which are both located at the foot of mountain ranges and may, hence, be exposed to snow-melt on higher ground. Further, we need to select the thresholds c_k and d_k for surface water and drainage run-off, respectively. An explanatory analysis for Oslo indicates that daily rainfall levels exceeding the 80% quantile induce periods of higher claim numbers. Similarly, an increased claim risk is found for the following days, as long as the change in drainage levels exceeds the 80% quantile. Hence, we set $c_k = q_{0.8}\left(R_{k,t} \mid R_{k,t} > 0\right)$ and $d_k = q_{0.8}\left(\Delta D_{k,t}\right)$ for each municipality individually.

With the clusters now identified, we use the optimization approach in Section 4.3 to estimate the parameters $\left(\eta_k, \rho_k, \omega_k^R\right)$ of $R_{k,j}^{\max}$ in expression (26) to help us derive the key covariate for our analysis. The optimization yields values of $\eta_k > 0$ for all municipalities while $\rho_k = 0$ for Bærum and Bergen. Since Bergen is surrounded by mountain ranges, values of the rainfall intensity of the event covariate, $I_{k,t}$ of expression (23), may be potentially high but uninformative. The sensitivity of the overall inference to these selected clusters and covariate parameters is explored in Section 5.4.

Under the cluster identification described above, Table 2 shows that about one third of the days are allocated to clusters of a length greater than 1. Further, clusters are almost always less than 7 days, which is the window that the insurance industry typically treats as a single event for re-insurance purposes. Figure 2 illustrates that, post clustering, most of the high number of claims coincide with high values for $R^{\max}$ and R^{Σ}, suggesting that our methods of Section 3 for constructing justifiable covariates and their relationship to claims has been successful.

5.2 Model estimates

The IGPD threshold u_k is set to be 4,2 and 4 for Oslo, Bærum and Bergen, respectively, as these appear to correspond to levels which only can arise due to weather induced claim sizes. For the



(a) Oslo (b) Bergen

Figure 2: Dependence between the aggregated rain R^{Σ} and the maximum rain within a day $R^{\max}$ for (a) Oslo and (b) Bergen. Periods with $\widetilde{N} > 4$ are highlighted.

model in Section 4.4, we define uninformative priors for all model parameters and run a MCMC algorithm for 100,000 iterations with every 50th sample being stored for analysis after a burn-in period of 25,000 to generate a sample from the posterior distribution. Convergence to the posterior distribution is checked via trace plots and Brooks–Gelman–Rubin diagnostics (Brooks and Gelman, 1998) with three sampled chains. Our `R` implementation took about 20 minutes per chain on a 2.80-GHz Intel Core i7 processor. In the following, when the municipalities are considered individually, the indexes are dropped for notational simplicity.

Summaries of the marginal posterior distributions of the 11 model parameters are presented in Table 3. The posterior distributions of p indicate that $80 - 90\%$ of the observations with $\widetilde{N} > 0$ are estimated to be related to the weather covariates $\widetilde{\mathbf{X}}$. Furthermore, $\xi = 0$ is contained in the 90% credibility interval for only 2 of the 3 municipalities. Hence, there is evidence that the tail behaviour of $\widetilde{Y}$ is not of a Poisson form for Bergen. The covariate effects $(\beta_i, \delta_i \ : \ i = 1, \ldots, 3)$ are generally lower for Bergen than for Oslo and Bærum. Since Bergen exhibits higher precipitation levels than Oslo and Bærum, the buildings are presumably designed to withstand more severe rainfall events than the ones in Oslo. The posterior estimates further show that covariate effects are non-negative except for β_1, which measures the effect of R^{Σ}, i.e., the accumulated effect of rainfall of the event excluding the maximum daily rainfall. Hence, the increased risk induced by larger values of R^{Σ} is mainly captured via δ_1. Collectively, this indicates that an increase in R^{Σ} results in more claims above 4 in Bergen but a reduction in the variability of these claims over 4.

The municipalities of Oslo and Bærum exhibit similar covariate effects for each of the covariates

Table 3: Posterior mean estimates and central 90% credibility interval of the model parameters for the municipalities of Oslo, Bærum and Bergen with thresholds $u_k = 4, 2$ and 4, respectively.

City	Statistic	p	β_0	β_1	β_2	β_3	ξ	δ_0	δ_1	δ_2	δ_3	κ
Oslo	Mean	0.90	0.12	0.21	0.23	0.76	-0.32	-0.16	0.42	0.32	0.71	2.21
	5% quantile	0.83	-0.61	0.09	0.10	0.46	-0.76	-0.30	0.29	0.22	0.57	1.65
	95% quantile	0.96	0.81	0.33	0.35	1.03	0.15	-0.03	0.56	0.42	0.86	2.93
Bærum	Mean	0.83	-1.80	0.15	0.35	1.31	0.16	-0.87	0.44	0.33	0.89	1.14
	5% quantile	0.67	-2.92	-0.01	0.18	0.87	-0.40	-1.31	0.18	0.20	0.58	0.70
	95% quantile	0.95	-0.90	0.34	0.53	1.70	0.82	-0.56	0.88	0.50	1.30	1.79
Bergen	Mean	0.88	-0.61	-0.03	0.19	0.37	0.53	-0.52	0.15	0.13	0.41	1.23
	5% quantile	0.79	-1.64	-0.17	0.05	0.17	0.10	-0.74	0.09	0.07	0.33	0.65
	95% quantile	0.95	0.30	0.12	0.33	0.57	1.10	-0.35	0.20	0.20	0.49	1.99

R^{Σ} and ΔS^{Σ}, which is unsurprising given their spatial proximity. Further, the estimates for the non-weather related rate κ differ by a factor of 2 for Oslo and Bærum, which is consistent with the number of policies in Oslo being about twice that of Bærum. The large difference of the β_3 posteriors (i.e., the effect of $R^{\max}$) between Oslo and Bærum is mainly driven by one large observation of 143 claims. Indeed, β_3 has much more similar posterior means of 0.75 and 0.81 for Oslo and Bærum, respectively when leaving out each of the their highest number of claims.

At each municipality, the estimated behaviour of $\widetilde{N} \mid \left(\widetilde{X}, \widetilde{N} > 0\right)$ is further investigated in Figure 3 which shows the changes in the estimated frequency for a set of $\widetilde{N}$ events for each covariate whilst fixing the other covariate values. In general, the probability of a high number of claims increases with increasing values for each of the three covariates, with $R^{\max}$ being the main risk factor for high number of claims. Further, the risk for very high numbers of claims increases more strongly for Oslo and Bærum than for Bergen. For instance, a covariate value of $R^{\max} = 50$ results in a probability of 0.6 for observing more than 6 claims in Oslo while it is only ~ 0.1 in Bergen. These findings are consistent with previous arguments that properties in Bergen are likely to be designed to withstand higher precipitation levels than in Oslo.

Table 4 assesses the fit of the estimated overall model for each possible value less than or equal to the threshold u and for a pooled estimate for above u. This assessment is derived for each of three non-overlapping ranges of the covariates. In particular, observations are split into three subsets with respect to $R^{\max}$: zero values, and below and above the median of the covariate given $R^{\max} > 0$ for which empirical and model-based frequencies are estimated. The estimated frequencies are derived from the marginal posterior predictive probabilities. For instance for the first case $R^{\max} = 0$, we derive the predictive frequency via expression (12) and set $\pi(\mathbf{x})$ as the product of empirical distributions of $(R^{\Sigma}, R^{\max} = 0)$ and ΔS^{Σ}, that is, rainfall and snow-melt are assumed to be independent. Table 4 illustrates that the model-based estimated frequency for $\widetilde{N}$

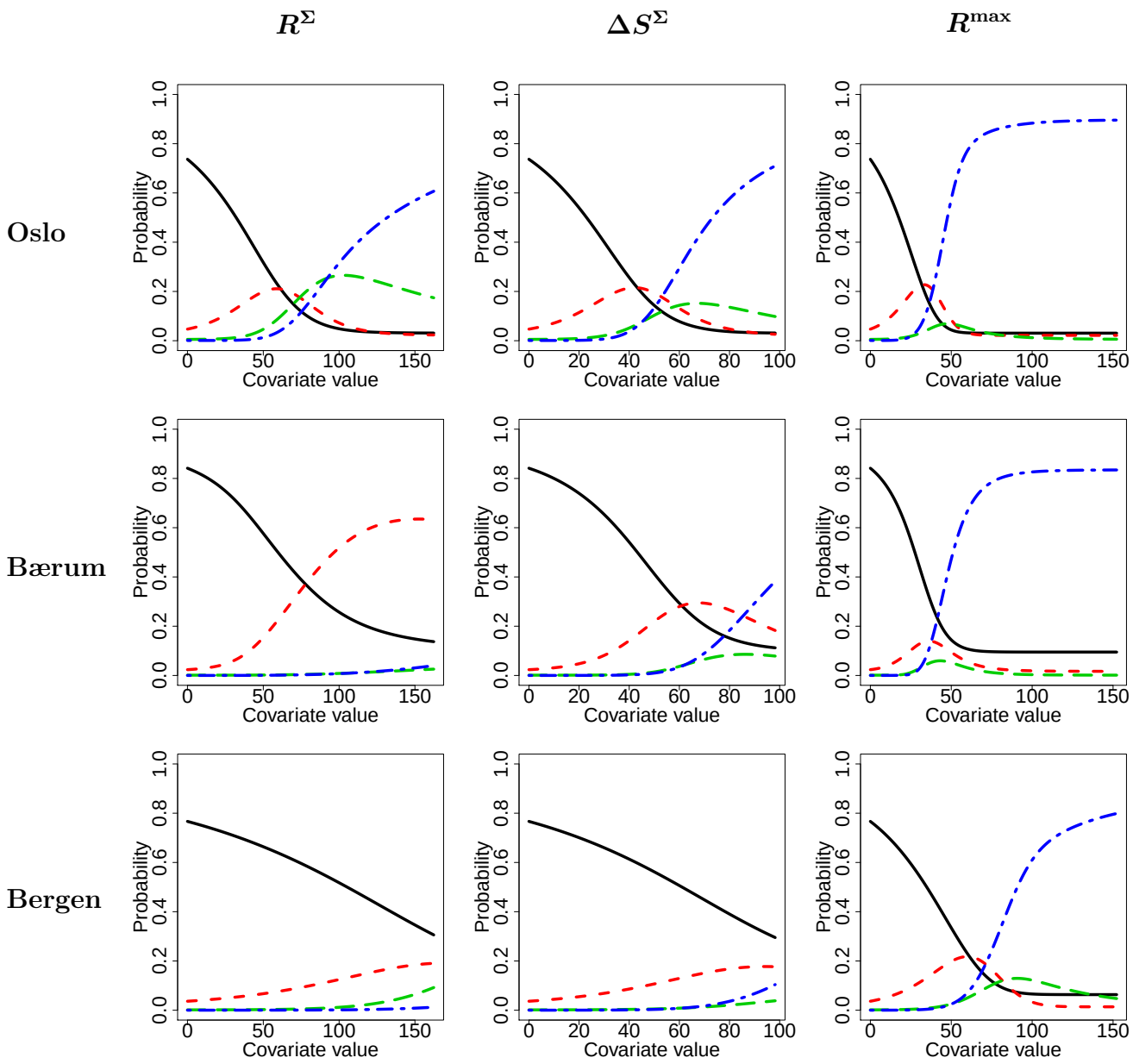


Figure 3: Probability for certain events of $\widetilde{N} \mid (\widetilde{X}, \widetilde{N} > 0)$ for Oslo, Bærum and Bergen varying with each of the covariates R^{Σ}, ΔS^{Σ} and $R^{\max}$. The events are $\widetilde{N} = 1$ (——), $\widetilde{N} = 3$ (- - - -), $\widetilde{N} = 5$ (— — —) and $\widetilde{N} > 6$ (— · —). In Column 1, the probability is considered with respect to R^{Σ} while the remaining covariates are fixed at their minimum value. Equivalently, Column 2 and 3 consider ΔS^{Σ} and $R^{\max}$, respectively.

lies within the empirical 95% confidence interval in all cases, where the confidence intervals are obtained by considering observations as realizations of a multinomial distribution with 5 possible outcomes for Oslo and Bergen and 3 for Bærum.

To conclude our analysis on the estimated model for $\widetilde{N} \mid \left(\widetilde{\mathbf{X}}, \widetilde{N} > 0\right)$, we compare the full model to three less-complex alternatives: (i) a zero-truncated Poisson as in (2), (ii) a Poisson-

Table 4: Posterior mean and empirical frequencies both times $\times 10^2$ for the number of claims between 1 and 4 for different rainfall settings for Oslo, Bærum and Bergen. For the empirical frequency, central 95% confidence intervals are given in parentheses. The rainfall settings are (1) $R^{\max} = 0$, (2) $0 < R^{\max} \leq q_{0.5}\,(R^{\max}|R^{\max} > 0)$ and (3) $R^{\max} > q_{0.5}\,(R^{\max}|R^{\max} > 0)$.

$\widetilde{N}$	$R^{\max}$	Oslo		Bærum		Bergen	
		estimated	empirical	estimated	empirical	estimated	empirical
1	(1)	72	75 (72,79)	83	84 (80,87)	77	79 (74,83)
	(2)	69	74 (69,80)	81	83 (78,90)	74	76 (72,81)
	(3)	40	34 (28,41)	49	47 (39,57)	44	45 (40,51)
2	(1)	20	18 (14,21)	13	13 (9,17)	19	16 (12,21)
	(2)	22	17 (12,23)	15	14 (9,21)	21	18 (13,22)
	(3)	25	24 (17,30)	25	22 (14,31)	26	24 (18,30)
3	(1)	5	5 (1, 8)			4	4 (0,9)
	(2)	6	7 (2,14)			4	3 (0,8)
	(3)	14	12 (6,19)			13	13 (7,19)
4	(1)	2	2 (0, 6)			1	0 (0,5)
	(2)	2	0 (0, 6)			1	2 (0,6)
	(3)	9	14 (7,21)			7	8 (2,13)
$> u_k$	(1)	1	0 (0, 4)	3	4 (0,8)	0	0 (0,4)
	(2)	1	1 (0, 7)	4	2 (0,9)	0	1 (0,5)
	(3)	13	16 (10,23)	25	31 (23,40)	10	11 (6,17)

mixture without the extremal mixture model for $\widetilde{Y}$ and (iii) an extremal mixture model without the component $\widetilde{Z}$. Table 5 gives the BIC averaged over all posterior samples and results indicate that our full model fits the data better than the competing models. Using the deviance information criterion (Spiegelhalter et al., 2002) largely supports this conclusion. The municipalities are similar in showing evidence that the additional flexibility offered by both our mixture and tail modelling components leads to substantial improvements.

5.3 Marginal distribution of clustered claims

The posterior distribution of marginal distribution of $\widetilde{N}$ is derived as the product of the posterior distributions of $\mathbb{P}\left(\widetilde{N} > 0\right)$ and $\mathbb{P}\left(\widetilde{N} > v \mid \widetilde{N} > 0\right)$. The former probability is straightforward to obtain by assuming that the occurrence of $\widetilde{N} > 0$ is Bernoulli distributed with a uniform prior. Table 6 (Column 4) provides the posterior mean and central 90% credibility intervals. The posterior probability for $\mathbb{P}\left(\widetilde{N} > v \mid \widetilde{N} > 0\right)$ is more complex to derive as it requires Monte Carlo integration over the weather covariates using expression (12) by replacing $\pi(\mathbf{x})$ by its the empirical estimate $\tilde{\pi}(\mathbf{x})$ for $\mathbf{X} \mid \widetilde{N} > 0$. For the posterior this needs evaluating for each of the J posterior

Table 5: Average Bayesian Information Criterion (ABIC) and Deviance Information criterion (DIC) for several competing models considering the distribution of $\widetilde{N} \mid (\widetilde{\mathbf{X}}, \widetilde{N} > 0)$ for Oslo, Bærum and Bergen. The best model fit for each municipality is highlighted.

Model	City	ABIC	DIC
Poisson	Oslo	2158	4.06
	Bærum	1079	3.96
	Bergen	2005	**3.92**
Poisson-Mixture	Oslo	2137	5.74
	Bærum	963	6.21
	Bergen	1977	5.63
Poison-IGPD	Oslo	2088	8.26
	Bærum	939	8.69
	Bergen	1937	8.75
Poisson-IGPD-Mixture	Oslo	**1779**	**3.18**
	Bærum	**596**	**−0.62**
	Bergen	**1632**	4.72

samples $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(J)}$ obtained by the MCMC algorithm in Section 5.2.

We use the posterior distribution of $\widetilde{N}$ to assess the model fit in Section 5.2 in terms of the marginal distribution of $\widetilde{N}$. An individual QQ plot is derived for each sample $\boldsymbol{\theta}^{(j)},\ j = 1, \ldots, J$, from the posterior distribution and collectively these give the posterior intervals for the QQ plot. Figure 4 (Column 1) shows that our model fits the whole distribution very well as the diagonal line lies within the 95% credibility interval for each municipality. The fit is at its weakest for Oslo around 20 claims which is due to the occurrence of three claim periods with 22-25 claims and two with 16-21 claims. For Bærum, the highest observation is not fitted ideally due to it being by far the highest observation over the 10-year period, however it is still consistent with our model when uncertainty is accounted for.

Using this marginal assessment of fit we can illustrate clearly how our clustering approach improves upon an analysis of the daily data. We fit the model of Section 2 to the original daily data with the covariates being the precipitation on the previous and current day, R_{t-1} and R_t, respectively, and the difference in the snow-water equivalent $S_{t-1} - S_t$. We set the IGPD threshold to $u = 3$ for Oslo and Bergen while u remains unchanged for Bærum. The modification of the threshold is required since the frequency of higher number of claims is lower in the daily data than in the clustered data. Figure 4 (Column 2) shows a much worse model fit for the daily data, in particular, for the medium to large claim numbers.

Interest lies in estimating the probability of extreme numbers of claims since it appears that the marginal distribution of $\widetilde{N}$ has a heavy tail, e.g., the largest claim event of 143 for Bærum substantially exceeds all events with other large numbers of claims in this municipality. Hence we

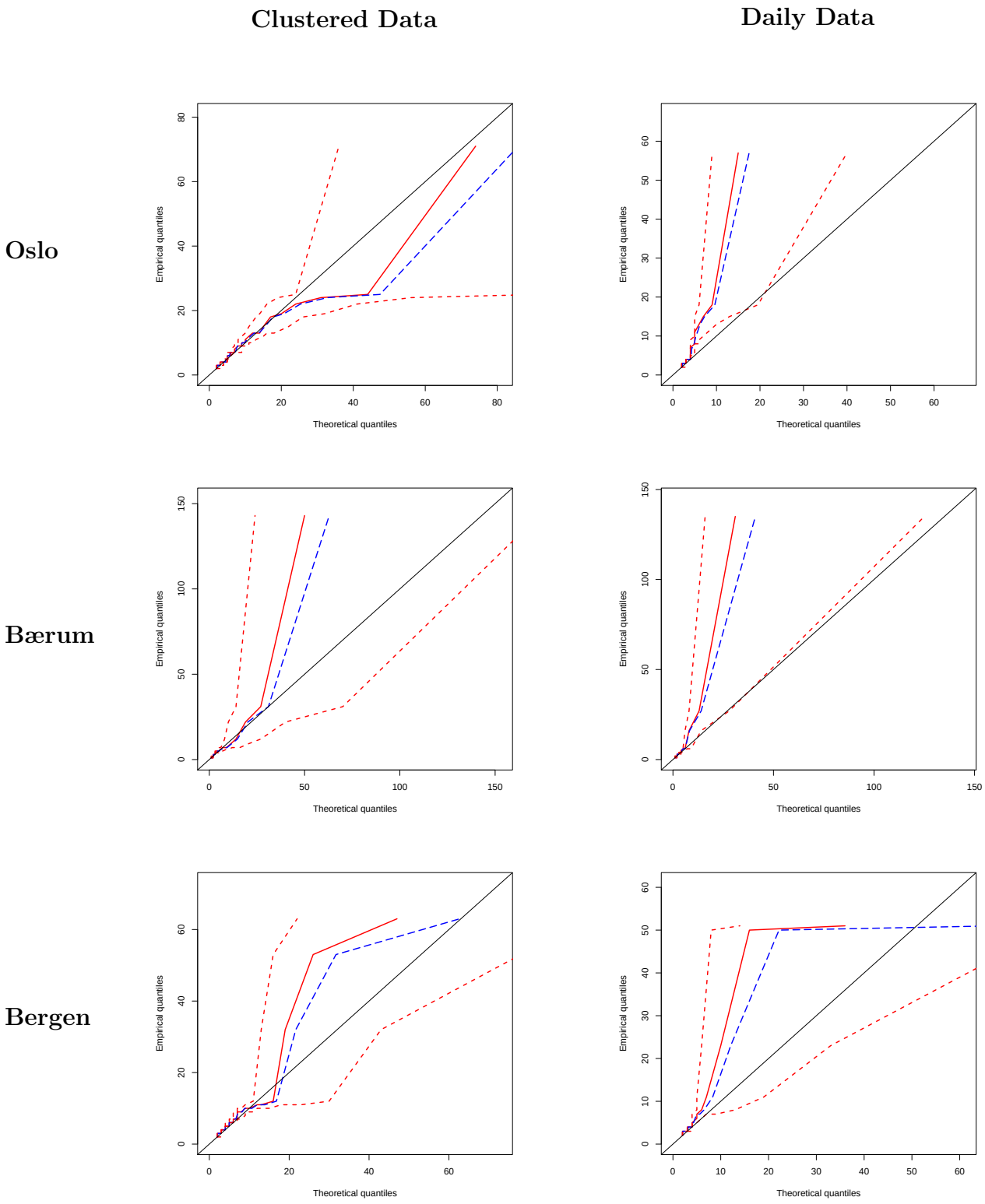


Figure 4: Posterior Quantile-Quantile plots for Oslo, Bærum and Bergen obtained by the full model. Column 1 provides the results for the clustered data while Column 2 considers the original daily data. The lines in each plot represent (– – –) Posterior mean, (——) Posterior median and (- - - - -) Central 95% posterior interval.

Table 6: Estimated scale ν and shape η for the distribution $R^{\max}|\widetilde{Y} > u \sim \text{GPD}(\nu, \eta)$ and standard errors. Column 3 provides the posterior mean and central 90% credibility intervals of the probability that $\widetilde{N}$ exceeds 100 conditional on $\widetilde{N} > 0$. Column 4 gives the empirical maximum likelihood estimate and central 90% confidence intervals of the frequency for $\widetilde{N} > 0$.

Municipality	ν	η	$\mathbb{P}\left(\widetilde{N} > 100 \mid \widetilde{N} > 0\right)$	$\mathbb{P}\left(\widetilde{N} > 0\right)$
Oslo	37.6	-0.48	0.00029	0.391
	(6.7)	(0.11)	$(6.3 \times 10^{-7}, 0.00096)$	(0.376, 0.407)
Bærum	27.73	-0.47	0.00044	0.209
	(4.8)	(0.11)	$(5.1 \times 10^{-5}, 0.00122)$	(0.197, 0.222)
Bergen	67.34	-0.40	0.00052	0.393
	(12.0)	(0.10)	$(4.8 \times 10^{-5}, 0.00148)$	(0.377, 0.409)

want to estimate $\mathbb{P}\left(\widetilde{N} > v\right)$ for large v, with $v \gg u$. Figure 3 shows that our conditional model indicates that extreme claims are strongly associated with extreme values of $R^{\max}$ but that the other covariates have limited association. Hence the use of empirical estimate $\tilde{\pi}(\mathbf{x})$ in expression (12) is likely to lead to underestimation of $\mathbb{P}\left(\widetilde{N} > v\right)$ since this limits $R^{\max}$ to the observed sample. Hence a parametric model is required to enable extrapolation for the distribution of $R^{\max}$, but we do not need to be concerned with the other covariates.

To help motivate our approach note that the first term on the right hand side of expression (12) is $\mathbb{P}\left(\widetilde{Y} > v\right)$, and that for $v > u$,

$$\begin{aligned} \mathbb{P}\left(\widetilde{Y} > v\right) &= \mathbb{P}\left(\widetilde{Y} > v \mid \widetilde{Y} > u\right) \mathbb{P}\left(\widetilde{Y} > u\right) \\ &= \int_{\mathbf{x}} \mathbb{P}\left(\widetilde{Y} > v \mid \mathbf{x}, \widetilde{Y} > u\right) \pi\left(\mathbf{x} \mid \widetilde{Y} > u\right) \mathrm{d}\mathbf{x} \;\times\; \mathbb{P}\left(\widetilde{Y} > u\right). \end{aligned} \tag{29}$$

The probabilities $\mathbb{P}\left(\widetilde{Y} > u\right)$ and $\mathbb{P}\left(\widetilde{Y} > v \mid \mathbf{x}, \widetilde{Y} > u\right)$ are estimated as described as above, with the latter entirely determined by the IGPD. However for $v \gg u$ we used a semi-parametric model-based estimate of $\pi\left(\mathbf{x} \mid \widetilde{Y} > u\right)$. Specifically, marginal exceedances of $R^{\max} \mid \widetilde{Y} > u$ over some threshold u_R are modelled by a GPD(ν, η) model, with tail probability λ, i.e., $\lambda = \mathbb{P}(R^{\max} > u_R \mid \widetilde{Y} > u)$ and the other covariates $(R^{\Sigma}, \Delta S^{\Sigma}) \mid R^{\max} > u_R, Y > u)$ are fixed at their average observed values (μ_1, μ_2) from this empirical conditional distribution. Hence we have a model for $\pi\left(\mathbf{x} \mid \widetilde{Y} > u\right)$ of

$$\widehat{\pi}\left(\mathbf{x} \mid \widetilde{Y} > u\right) = \begin{cases} \widetilde{\pi}\left(\mathbf{x} \mid \widetilde{Y} > u\right) & \text{for } r^{\max} < u_R \\ \frac{\lambda}{\nu}\left(1 + \eta(r^{\max} - u_E)/\nu\right)_+^{-1-1/\eta} \mathbb{1}(r^{\Sigma} = \mu_1, \Delta s^{\Sigma} = \mu_2) & \text{for } r^{\max} \geq u_R. \end{cases}$$

The mean residual life plots in Figure 5 are used to select the threshold u_R. A threshold of

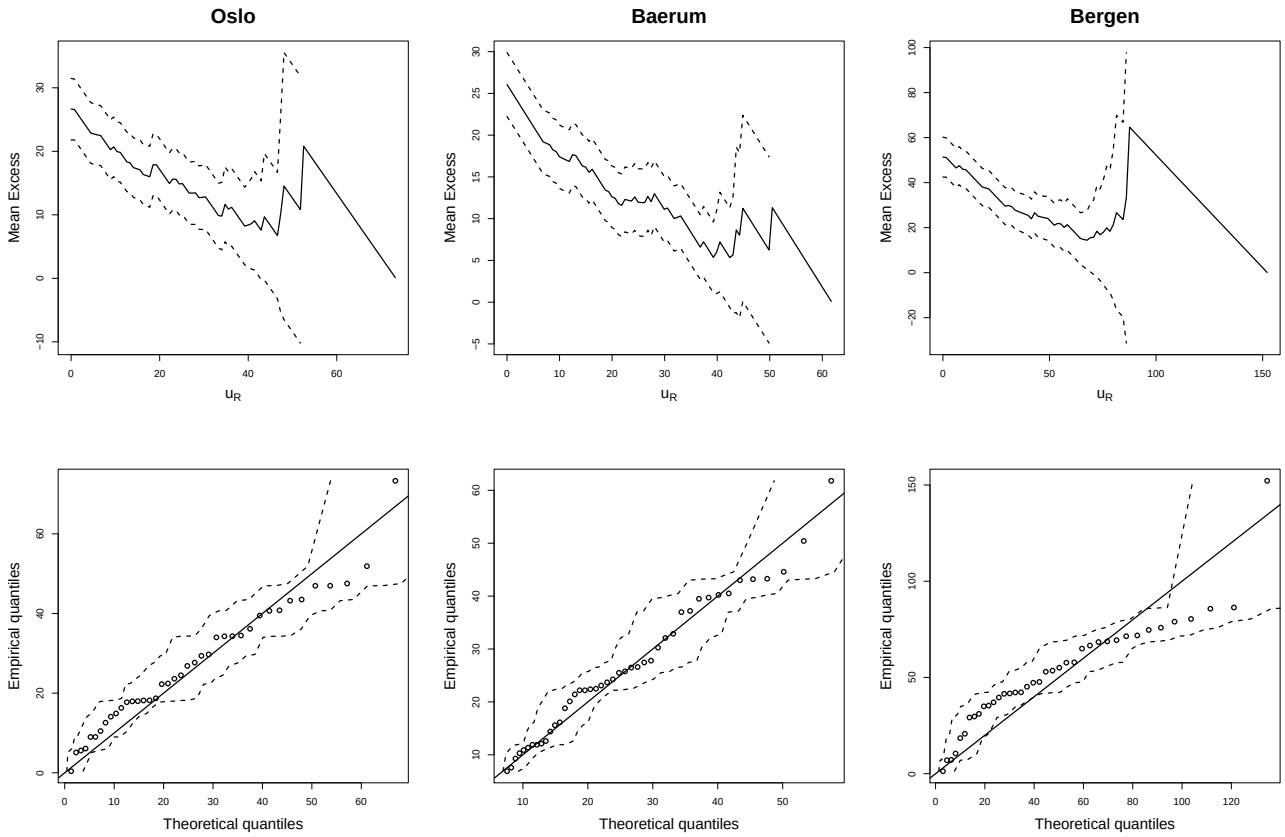


Figure 5: Mean residual life plots (Row 1) and Quantile-Quantile plots with central 95% confidence intervals of the fitted GPD distribution (Row 2) for $\pi\left(R^{\max} \mid \widetilde{Y} > u\right)$ for the municipalities of Oslo, Bærum and Bergen.

$u_R = 0.1mm$ seems suitable as the plot is approximately linear above this level once uncertainty is accounted for. The level corresponds to the smallest positive amount of rainfall. The GPD is fitted separately for each municipality via maximum likelihood and estimates and standard errors for the scale parameter ν and shape parameter η for the GPD, as in expression (5), are provided in Table 6. The estimated shape parameter η is negative for all three municipalities, that is, the associated GPD is short-tailed with a finite upper end point. Figure 5 shows that the tail fit for $R^{\max} \mid \widetilde{Y} > u$ is good for Oslo and Bærum while being slightly off for Bergen.

Focusing on $v = 100$, Table 6 shows posterior summaries for the $\mathbb{P}\left(\widetilde{N} > 100 \mid \widetilde{N} > 0\right)$ for the three municipalities. The results indicate that about 1 in 5000 events for Bergen will cause more than 100 claims. Considering that about 2,500 events were observed over a 10 year horizon, that corresponds to one occurrence every 20 years on average. The same approach implies that such an event happens every 30–40 years for Oslo and Bærum. Hence, the observation of 143 claims for Bærum is a very rare event.

5.4 Sensitivity analysis

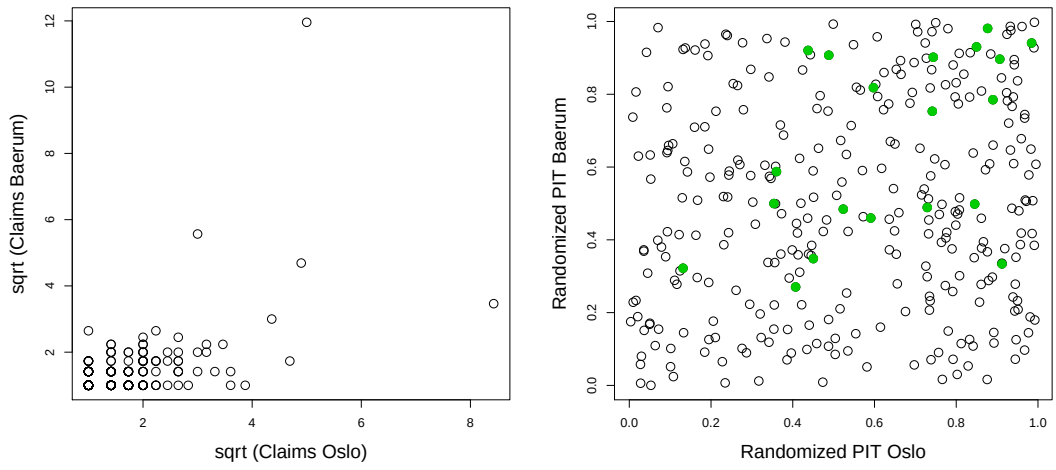
Several thresholds were fixed in Section 5.1 to derive the cluster periods, as well as, the parameters ω^R and (η, ρ) in expressions (23) and (26), respectively. Specifically, c and d in Algorithm 1 were set to the 80% quantile of the observed rainfall and difference in drainage, respectively, while the threshold t in expression (19) was fixed to the 99.5% quantile in Section 4.3. Here, interest lies in exploring the sensitivity of the results in Section 5.2 with respect to these settings for c, d and t.

We start by considering the threshold t. If t is the 98.5% quantile, instead of the 99.5% quantile, then the estimated ρ is now positive for all three municipalities. To assess sensitivity in terms of model fit, the full parametric model (28) is estimated with the resulting new covariate values. By comparing the BIC and DIC of this estimated model to the one in Table 5, we find little, to no, change in the BIC. In particular, the largest difference is found for Bærum with an increase in BIC of 7. With respect to the estimated covariate effects, Oslo and Bergen are very similar while some larger changes are found for Bærum. The latter is related to the cluster period with the highest number of claims as it is the period of rainfall which is both most 'intense' and contains the largest daily accumulation. Consequently, while different threshold choices for t affect the estimated parameters, and potentially the covariate effects, little sensitivity is found in terms of model fit and subsequent inferences.

To assess the sensitivity on c and d, we consider the QQ plots for $\widetilde{N}$, considered in Section 5.3, of the estimated models rather than comparing the BIC and DIC. This is due to the clustered data being dependent on these thresholds, affecting the interpretability of the BIC and DIC measures. We take d as the 75% and 85% quantile while keeping c fixed to the original 80% quantile and vice-versa. The QQ plots illustrate that these models fit the clustered data essentially as well as the original model in Section 5.2. There is a slightly poor fit of the highest claim numbers for Oslo and Bærum when c or d, corresponds to the 85% quantile. Since higher values c or d imply, on average, shorter cluster periods, taking c or d too high leads to some claims across days being classified as independent although they are related to the same severe weather event. For instance, in the case of Bærum, a higher c leads to a lower estimated p and a lighter tail which provides a poorer fit of the extremes.

5.5 Examination of the conditional spatial claim dependence

Neighbouring municipalities tend to have dependent numbers of aggregated claims from the same weather event. This dependence is illustrated in the left panel of Figure 6, which shows positive dependence of the claim numbers for the adjacent municipalities of Oslo and Bærum, with there being a particularly strong dependence in the extreme values. The plot is presented after the use of a square root transformation since the marginal claim numbers distribution is heavy tailed. The estimated Kendall's τ has a central 95% confidence interval of $(0.26, 0.44)$ and this interval is invariant to the square root, or any monotone, marginal transformation.



(a) (b)

Figure 6: Plots of simultaneous (a) clustered claims for Oslo and Bærum and (b) the randomized probability integral transformed samples using the estimated conditional distributions of claims given weather at each municipality. Observations for which simultaneously more than 4 and 2 claims are observed for Oslo and Bærum, respectively, are highlighted.

The only possible cause for this dependence in claims is through the spatial dependence of the weather covariates, as claims in one region are not directly related to those in a different area. Specifically, for any weather event at time t, it is logically reasonable that conditional independence of claim numbers given weather conditions holds, i.e., that

$$\left[\left(\widetilde{N}_{1,t}, \widetilde{N}_{2,t}\right) \mid \left(\widetilde{\mathbf{X}}_{1,t}, \widetilde{\mathbf{X}}_{2,t}\right)\right] = \left[\widetilde{N}_{1,t} \mid \widetilde{\mathbf{X}}_{1,t}\right] \times \left[\widetilde{N}_{2,t} \mid \widetilde{\mathbf{X}}_{2,t}\right],$$

where here the municipalities of Oslo and Bærum are numbered 1 and 2, respectively. Thus a good test of the predictive ability of our selected weather covariates is to test whether the conditional variables $\widetilde{N}_{1,t} \mid \widetilde{\mathbf{X}}_{1,t}$ and $\widetilde{N}_{2,t} \mid \widetilde{\mathbf{X}}_{2,t}$ are independent or not.

The complexities of this assessment relates to the cluster periods of events at the two locations not having identical start and end times and the discrete nature of $\widetilde{N}_{k,t}$ are discussed below. Ignoring these issues for the moment, in the right panel of Figure 6 we show model-based estimates of $\mathbb{P}\left(\widetilde{N}_{i,t} \leq \widetilde{n}_{i,t} \mid \widetilde{\mathbf{X}}_{i,t} = \widetilde{\mathbf{x}}_{i,t}\right)$, where $\{(\widetilde{n}_{i,t}, \widetilde{\mathbf{x}}_{i,t}) \ : \ t = 1, \ldots, m\}$ denote the set of cluster periods data derived for municipality i, $(i = 1, 2)$. These two conditional distributions are evaluated using the fitted model in Section 5.2. If the model is a good fit then, marginally, each variable should be Uniform$(0, 1)$. This aspect of fit for each individual municipality was assessed in Section 5.3. The points also appear to be relatively uniformly distributed over $(0, 1)^2$, indicating independence. Kendall's τ for this joint sample has a central 95% confidence interval of $(0.00, 0.14)$, showing that the dependence has been much reduced relative to the unconditional joint distribution. Further-

more, as independence, corresponding to $\tau = 0$, is in this interval, it is supportive of the hypothesis that our selected covariates $\widetilde{\mathbf{X}}$ capture all the important weather features related to claim numbers.

First consider the issue of cluster periods not being identical for the two municipalities. Many of the weather event clusters that are identified by the weather cluster extraction scheme of Section 5.1 start at the same time, but some have non-overlapping periods. Between the jth and $(j+1)$th occurrence when weather cluster starts at exactly the same time for both municipalities there are $l_{i,j} \geq 0$ weather events for municipality i $(i = 1, 2)$. Of these $l_{i,j}$ events, we select separately the weather event giving maximum number of claims at each municipality and the associated weather covariates from that event are recorded. In case of ties, that is, two or more of the $l_{i,j}$ clusters give the same maximum number of claims, the first of these clusters is selected. We treat these events as joint spatial events even though their start times do not necessarily always match up. Given that the dependence in weather covariates is so strong between municipalities, in practice this joint event definition process retains approximately 90% of the weather cluster periods identified previously using municipality specific selection methods. As we are interested in cases where $\widetilde{N}_{i,t} > 0$ for $i = 1$ and 2, then events which fail to achieve this condition are discarded, leaving 50% and 70% of the total claims for Oslo and Bærum respectively. These are the data shown in Figure 6 and analysed subsequently.

To account for the discrete nature of the clustered claims $\widetilde{N}$ when evaluating the model-based estimates of $\mathbb{P}\left(\widetilde{N}_{i,t} \leq \widetilde{n}_{i,t} \mid \widetilde{\mathbf{X}}_{i,t} = \widetilde{\mathbf{x}}_{i,t}\right)$ we use the randomized probability integral transform (Smith, 1985; Brockwell, 2007). Specifically we replace this conditional probability by a

$$\text{Uniform}\left[\mathbb{P}\left(\widetilde{N} \leq \widetilde{n}_i - 1 \mid \widetilde{\mathbf{x}}_i, \widetilde{N} > 0\right), \mathbb{P}\left(\widetilde{N} \leq \widetilde{n}_i \mid \widetilde{\mathbf{x}}_i, \widetilde{N} > 0\right)\right] \tag{30}$$

values, where the probabilities in expression (30) are set to their posterior means.

6 Discussion

We extended the modelling framework by Haug et al. (2011) and Scheel et al. (2013) in order to improve the model fit for higher number of claims. Additional information was gained by analysing the spatial and temporal patterns with respect to snow-melt and precipitation. A temporal cluster algorithm, based solely on the observed weather covariates, was introduced in order to reduce the effects of potential lags in the recording process and to account for weather events which affect the claim dynamics on consecutive days. The original daily data were then adapted to the respective cluster periods and one covariate was tuned to maximize its relevance to large claims.

A mixture model with an extremal mixture component was applied to model the number of claims over the cluster periods. Results have shown good performance for lower as well as higher marginal and conditional numbers of claims. Furthermore, the spatial dependence between claims in different municipalities appears to be accounted for by the derived weather covariates.

The derived model can also be applied to assess the impact of climate change. Haug et al. (2011) use the daily data and perform an effect study, subject to the insurance portfolio of properties of future periods being close in value and quality to the one of the model fitting period. Their results indicate an increase in the claim frequency for all municipalities. In order to perform a similar study with our new model, it is necessary to simulate weather observations for cluster periods rather than single days.

There are various way to extend the model presented in this paper. Firstly in the model fitting of the extremal mixture model for claims, the distribution can be restricted to a uni-modal form by excluding parameter settings which induce

$$\mathbb{P}\left(\widetilde{Y} = \lfloor u \rfloor - 1 \mid \widetilde{\mathbf{X}}\right) > \mathbb{P}\left(\widetilde{Y} = \lfloor u \rfloor \mid \widetilde{\mathbf{X}}\right) < \mathbb{P}\left(\widetilde{Y} = \lfloor u \rfloor + 1 \mid \widetilde{\mathbf{X}}\right).$$

This set of inequalities imposes additional constraints on the parameters λ, σ_u and ξ. This paper focused on the periods with $\widetilde{N} > 0$ but there is interest for all periods. We considered a Poisson-IGPD mixture with the same parameter values as for the zero-truncated Poisson-IGPD mixture in Section 4 and found that the model underpredicts the frequency of periods with zero claims $\widetilde{N} = 0$. Hence, the model could be extended via a hurdle component as in the BPH. Furthermore, Figure 3 shows that the event $\widetilde{N} = 1$ has a probability of about 0.10 even for very high values of $R^{\max}$ due to the non-weather related mixture component. One may argue that such predictions are unrealistic since extreme precipitation levels over a day should lead to large damages, regardless of their intensity. Therefore, the mixture probability p could be modelled as a function of the covariate $R^{\max}$.

Further research can also be undertaken from a spatial perspective. Spatial dependence of the parameters of the conditional distribution of $\widetilde{N} \mid \left(\widetilde{X}, \widetilde{N} > 0\right)$ may be introduced to allow for a better model fit similarly to Scheel et al. (2013). For instance, the threshold $u = 2$ for Bærum may be too low for the extremal mixture model but there are not enough observations to raise it to $u = 3$. Additional information may be borrowed from the adjacent municipalities, in particular Oslo, in order to achieve this. Spatial dependence could be modelled via a conditional autoregressive prior (Besag, 1974; Besag et al., 1991) on $(\beta_1, \beta_2, \beta_3)$ in (28).

Acknowledgements

Rohrbeck gratefully acknowledges funding of the EPSRC funded STOR-i Centre for Doctoral Training (grant number EP/H023151/1). This research was also financially supported by the Norwegian Research Council. We also thank Ida Scheel for providing access to the Norwegian insurance and weather data and the editors and referees for very helpful suggestions that substantially improved the presentation of the work.

A Threshold-stability of the IGPD

Lemma 1. *Let N be an integer-valued random variable with $N \mid N > u \sim IGPD(\sigma_u, \xi, u),\ u \in \mathbb{R}$. Then for any $u < v < u - \frac{\sigma_u}{\xi}$, $N \mid N > v \sim IGPD\left(\sigma_u + \xi\left(\lfloor v \rfloor - \lfloor u \rfloor\right), \xi, v\right)$ for any $\sigma_u > 0$ and $\xi \in \mathbb{R}$.*

Proof. We prove the lemma via the survival function $\mathbb{P}\left(N > n \mid N > v\right)$, where n is integer with $n > v$. By applying conditional probabilities, $\mathbb{P}\left(N > n \mid N > v\right)$ can be expressed by

$$\begin{aligned}\mathbb{P}\left(N > n \mid N > v\right) &= \frac{\mathbb{P}\left(N > n \mid N > u\right)}{\mathbb{P}\left(N > v \mid N > u\right)} \\ &= \frac{\mathbb{P}\left(H > \ n \ - \lfloor u \rfloor\right)}{\mathbb{P}\left(H > \lfloor v \rfloor - \lfloor u \rfloor\right)}\end{aligned}$$

where $N = \lceil H \rceil$, with H a GPD with parameters σ_u and ξ. It follows that

$$\begin{aligned}\mathbb{P}\left(N > n \mid N > v\right) &= \frac{\left[1 + \frac{\xi(\ n \ - \lfloor u \rfloor)}{\sigma_u}\right]_+^{-\frac{1}{\xi}}}{\left[1 + \frac{\xi(\lfloor v \rfloor - \lfloor u \rfloor)}{\sigma_u}\right]_+^{-\frac{1}{\xi}}} \\ &= \left[\frac{\sigma_u + \xi\left(\ n \ - \lfloor u \rfloor\right)}{\sigma_u + \xi\left(\lfloor v \rfloor - \lfloor u \rfloor\right)}\right]_+^{-\frac{1}{\xi}} \\ &= \left[\frac{\sigma_u + \xi\left(\ n \ - \lfloor v \rfloor + \lfloor v \rfloor - \lfloor u \rfloor\right)}{\sigma_u + \xi\left(\lfloor v \rfloor - \lfloor u \rfloor\right)}\right]_+^{-\frac{1}{\xi}} \\ &= \left[1 + \frac{\xi\left(\ n \ - \lfloor v \rfloor\right)}{\sigma_u + \xi\left(\lfloor v \rfloor - \lfloor u \rfloor\right)}\right]_+^{-\frac{1}{\xi}},\end{aligned}$$

which is the survival function of a IGPD above threshold v with scale parameter $\sigma_u + \xi\left(\lfloor v \rfloor - \lfloor u \rfloor\right) > 0$ and shape parameter ξ. □

B Threshold-stability of the mixture tail

Lemma 2. *Let N be an integer-valued random variable with $N \mid N > u$ having distribution function*

$$\mathbb{P}\left(N = n \mid N > u\right) = p\ \mathbb{P}(Y = n) + (1 - p)\ \mathbb{P}(Z = n)$$

where $Y \sim IGPD(\sigma_u, \xi, u)$ and Z being a truncated Poisson above threshold u with parameter κ. Then for any $v > u$, the random variable $N \mid N > v$, is distributed according to a mixture of an $IGPD(\sigma_u + \xi\left(\lfloor v \rfloor - \lfloor u \rfloor\right), \xi, v)$ and a truncated Poisson above v with rate parameter κ and mixture

probability

$$p_v = \frac{p\ \mathbb{P}(Y > v)}{p\ \mathbb{P}(Y > v) + (1-p)\ \mathbb{P}(Z > v)}.$$

Proof. Consider any combination $n > v > u$. Then, based on conditional probabilities,

$$\begin{aligned}
&\mathbb{P}(N > n \mid N > v) \\
&\qquad = \frac{\mathbb{P}(N > n \mid N > u)}{\mathbb{P}(N > v \mid N > u)} \\
&\qquad = \frac{p\ \mathbb{P}(Y > n) + (1-p)\ \mathbb{P}(Z > n)}{p\ \mathbb{P}(Y > v) + (1-p)\ \mathbb{P}(Z > v)} \\
&\qquad = \frac{p\ \mathbb{P}(Y > n \mid Y > v)\ \mathbb{P}(Y > v) + (1-p)\ \mathbb{P}(Z > n | Z > v)\ \mathbb{P}(Z > v)}{p\ \mathbb{P}(Y > v) + (1-p)\ \mathbb{P}(Z > v)}
\end{aligned}$$

By defining

$$p_v = \frac{p\ \mathbb{P}(Y > v)}{p\ \mathbb{P}(Y > v) + (1-p)\ \mathbb{P}(Z > v)},$$

we obtain

$$\mathbb{P}(N > n \mid N > v) = p_v\ \mathbb{P}(Y > n \mid Y > v) + (1 - p_v)\ \mathbb{P}(Z > n \mid Z > v).$$

Based on the threshold-stability in Appendix A, $Y \mid Y > v \sim \text{IGPD}(\sigma_u + \xi(\lfloor v \rfloor - \lfloor u \rfloor), \xi, v)$. Further, $Z \mid Z > v$ is a truncated Poisson above v with rate κ. Hence, $N \mid N > v$ is distributed according to a mixture of an IGPD and a truncated Poisson. □

C Details of the MCMC algorithm

Let $\mathcal{D} = \{(\widetilde{n}_i, \widetilde{\mathbf{x}}_i)\,,\ i = 1, \ldots, m\}$ denote the set of observed claim numbers and covariates effects. Further, a latent binary variable v_i is introduced for each observation $\widetilde{n}_i$ which is defined by

$$v_i = \begin{cases} 1 & \text{if } \widetilde{n}_i \text{ is a realization from the distribution } \widetilde{Y} \\ 0 & \text{otherwise.} \end{cases}$$

We set a Beta$(1, 1)$ prior $\pi(p)$ on the mixing probability p and an improper prior on the remaining parameters, $\pi(\boldsymbol{\beta}, \boldsymbol{\delta}, \xi, \kappa) \propto 1$. Hence, the posterior distribution $\pi(p, \boldsymbol{\beta}, \xi, \boldsymbol{\delta}, \kappa, v_1, \ldots, v_I | \mathcal{D})$ is proportional to

$$\prod_{i=1}^{m} \left\{ \left[p\ \mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}, \xi, \boldsymbol{\delta}, \widetilde{\mathbf{x}} \right) \right]^{v_i} \ \left[(1-p)\ \mathbb{P}\left(\widetilde{Z} = \widetilde{n}_i \mid \kappa \right) \right]^{1-v_i} \right\} \ \pi(p)$$

Realizations from this posterior distribution are sampled by a Metropolis-within-Gibbs algorithm which runs for a fixed number of iterations J. Let $p^{(0)}, \boldsymbol{\beta}^{(0)}, \xi^{(0)}, \boldsymbol{\delta}^{(0)}$ and $\lambda^{(0)}$ denote the initial parameter values. The update procedure for all parameters within one iteration step $j = 1, \ldots, J$ is as follows:

At the start of iteration step j, the latent variables $v_1^{(j)}, \ldots, v_m^{(j)}$ are sampled from a Bernoulli distribution

$$v_i^{(j)} \sim \text{Bernoulli}\left(w_i^{(j)}\right).$$

The probability of observation $\widetilde{n}_i$ being sampled from the covariate-driven component $\widetilde{Y}$, $w_i^{(j)}$, is given by

$$w_i^{(j)} = \frac{p^{(j-1)}\ \mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^{(j-1)}, \xi^{(j-1)}, \boldsymbol{\delta}^{(j-1)}, \widetilde{\mathbf{x}}_i\right)}{p^{(j-1)}\ \mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^{(j-1)}, \xi^{(j-1)}, \boldsymbol{\delta}^{(j-1)}, \widetilde{\mathbf{x}}_i\right) + (1 - p^{(j-1)})\ \mathbb{P}\left(\widetilde{Z} = \widetilde{n}_i\right)}.$$

Since we placed a conjugate Beta prior on p, the parameter value is updated by sampling from the full-conditional Beta posterior

$$p^{(j)} \sim \text{Beta}\left(\sum_{i=1}^{I} v_i^{(j)} + 1,\ I - \sum_{i=1}^{I} v_i^{(j)} + 1\right).$$

The model parameters $\boldsymbol{\beta}$, ξ and $\boldsymbol{\delta}$ are updated separately via Random-Walk-Metropolis with Gaussian proposal. For the covariate effects $\boldsymbol{\beta}$, the proposal $\boldsymbol{\beta}^*$ is accepted with probability

$$\min\left\{1, \prod_{v_i^{(j)}=1,\ \widetilde{n}_i > u} \frac{\mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^*, \xi^{(j-1)}, \boldsymbol{\delta}^{(j-1)}, \widetilde{\mathbf{x}}_i\right)}{\mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^{(j-1)}, \xi^{(j-1)}, \boldsymbol{\delta}^{(j-1)}, \widetilde{\mathbf{x}}_i\right)}\right\},$$

whilst the proposal ξ^* has acceptance probability

$$\min\left\{1, \prod_{v_i^{(j)}=1,\ \widetilde{n}_i > u} \frac{\mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^{(j)}, \xi^*, \boldsymbol{\delta}^{(j-1)}, \widetilde{\mathbf{x}}_i\right)}{\mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^{(j)}, \xi^{(j-1)}, \boldsymbol{\delta}^{(j-1)}, \widetilde{\mathbf{x}}_i\right)}\right\}.$$

Note, the likelihood needs only to be evaluated for the observations with latent variable $v_i^{(j)} = 1$ and the number of observations $\widetilde{n}_i$ greater than the threshold. Next, the covariate effects for the rate parameter κ are updated. Here, the likelihood has to be evaluated for all observations with $v_i^{(j)} = 1$ as $\boldsymbol{\delta}$ effects the threshold exceedance model. The acceptance ratio is thus given by

$$\min\left\{1, \prod_{v_i^{(j)}=1} \frac{\mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^{(j)}, \xi^{(j)}, \boldsymbol{\delta}^*, \widetilde{\mathbf{x}}_i\right)}{\mathbb{P}\left(\widetilde{Y} = \widetilde{n}_i \mid \boldsymbol{\beta}^{(j)}, \xi^{(j)}, \boldsymbol{\delta}^{(j-1)}, \widetilde{\mathbf{x}}_i\right)}\right\}.$$

Finally, the rate parameter κ is updated via an independence sampler with uniform proposal distribution. The acceptance probability then yields to

$$\min\left\{1,\ \prod_{v_i^{(j)}=0}\frac{\mathbb{P}\left(\widetilde{Z}=\widetilde{n}_i\mid\kappa^*\right)}{\mathbb{P}\left(\widetilde{Z}=\widetilde{n}_i\mid\kappa^{(j-1)}\right)}\right\}.$$

References

Anderson, C. W. (1970). Extreme value theory for a class of discrete distributions with applications to some stochastic processes. *Journal of Applied Probability*, 7(1):99–113.

Anderson, C. W. (1980). Local limit theorems for the maxima of discrete random variables. *Mathematical Proceedings of the Cambridge Philosophical Society*, 88:161–165.

Anderson, C. W., Coles, S. G., and Hüsler, J. (1997). Maxima of Poisson-like variables and related triangular arrays. *The Annals of Applied Probability*, 7(4):953–971.

Behrens, C. N., Lopes, H. F., and Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3):227–244.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pages 192–236.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Boldi, M.-O. and Davison, A. C. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):217–229.

Bottolo, L., Consonni, G., Dellaportas, P., and Lijoi, A. (2003). Bayesian analysis of extreme values by mixture modeling. *Extremes*, 6(1):25–47.

Botzen, W. J. and Van Den Bergh, J. C. (2008). Insurance against climate change and flooding in the Netherlands: present, future, and comparison with other countries. *Risk Analysis*, 28(2):413–426.

Botzen, W. J. and van den Bergh, J. C. (2012). Risk attitudes to low-probability climate change risks: Wtp for flood insurance. *Journal of Economic Behavior & Organization*, 82(1):151–166.

Brockwell, A. E. (2007). Universal residuals: A multivariate transformation. *Statistics and Probability Letters*, 77(14):1473–1478.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

Buddana, A. and Kozubowski, T. J. (2014). Discrete Pareto distributions. *Economic Quality Control*, 29(2):143–156.

Carreau, J. and Bengio, Y. (2009). A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes*, 12(1):53–76.

Coles, S., Heffernan, J., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):377–392.

Coles, S. G. and Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 329–347.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–442.

de Carvalho, M. and Davison, A. C. (2014). Spectral density ratio models for multivariate extremes. *Journal of the American Statistical Association*, 109(506):764–776.

Department for Environment, Food and Rural Affairs (2004). Review of UK climate change indicators. Available from http://www.ecn.ac.uk/iccuk/.

Eastoe, E. F. and 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.

Einmahl, J. H., de Haan, L., and Sinha, A. K. (1997). Estimating the spectral measure of an extreme value distribution. *Stochastic Processes and their Applications*, 70(2):143 – 171.

Einmahl, J. H. J. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *Ann. Statist.*, 37(5B):2953–2989.

Frees, E. W. and Valdez, E. A. (2008). Hierarchical insurance claims modeling. *Journal of the American Statistical Association*, 103(484):1457–1469.

Frigessi, A., Haug, O., and Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes*, 5(3):219–235.

Hanson, T. E., de Carvalho, M., and Chen, Y. (2017). Bernstein polynomial angular densities of multivariate extreme value distributions. *Statistics and Probability Letters*, 128:60 – 66.

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

Hitz, A. (2017). Modelling of Extremes. Unpublished Oxford University PhD thesis.

Holmes, J. and Moriarty, W. (1999). Application of the generalized Pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 83(1):1–10.

Jenkins, G. J., Perry, M. C., and Prior, M. J. (2008). The climate of the United Kingdom and recent trends. Met Office Hadley Centre, Exeter, UK.

Klugman, S. A., Panjer, H. H., and Willmot, G. E. (2012). *Loss Models: from Data to Decisions*. John Wiley & Sons, 4 edition.

Knabb, R. D., Rhome, J. R., and Brown, D. P. (2005). Tropical Cyclone Report: Hurricane Katrina: 23-30 August 2005. Technical report, National Hurricane Center.

Kubilay, A., Derome, D., Blocken, B., and Carmeliet, J. (2013). Cfd simulation and validation of wind-driven rain on a building facade with an Eulerian multiphase model. *Building and Environment*, 61:69–81.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187.

Li, Y., Cai, W., and Campbell, E. (2005). Statistical modeling of extreme rainfall in southwest western Australia. *Journal of Climate*, 18(6):852–863.

MacDonald, A., Scarrott, C. J., Lee, D., Darlow, B., Reale, M., and Russell, G. (2011). A flexible extreme value mixture model. *Computational Statistics & Data Analysis*, 55(6):2137–2157.

Mills, E. (2005). Insurance industry in a climate of change. *Science*, 309(5737):1040–1044.

Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science & Business Media.

Pickands, J. (1971). The two-dimensional Poisson process and extremal processes. *Journal of Applied Probability*, 8(4):745–756.

Prieto, F., Gómez-Déniz, E., and Sarabia, J. M. (2014). Modelling road accident blackspots data with the discrete generalized Pareto distribution. *Accident Analysis & Prevention*, 71:38–49.

Resnick, S. I. (2013). *Extreme Values, Regular Variation and Point Processes*. Springer, New York.

Russell, B. T., Cooley, D., Porter, W. C., Reich, B. J., and Heald, C. L. (2016). Data mining to investigate the meteorological drivers for extreme ground level ozone events. *Annals of Applied Statistics*, 10(3):1673–1698.

Sanders, C. H. and Phillipson, M. C. (2003). UK adaption strategy and technical measures: the impacts of climate change on buildings. *Building Research & Information*, 31(3–4):210–221.

Scheel, I., Ferkingstad, E., Frigessi, A., Haug, O., Hinnerichsen, M., and 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.

Schuster, S. S., Blong, R. J., and McAneney, K. J. (2006). Relationship between radar-derived hail kinetic energy and damage to insured buildings for severe hailstorms in eastern Australia. *Atmospheric Research*, 81(3):215–235.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shimura, T. (2012). Discretization of distributions in the maximum domain of attraction. *Extremes*, 15(3):299–317.

Smith, J. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting*, 4(3):283–291.

Smith, R. L. and Goodman, D. J. (2000). Bayesian risk analysis. In Embrechts, P., editor, *Extremes and Integrated Risk Management*, pages 235–251. Risk Books, London.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Yip, K. C. and Yau, K. K. (2005). On modeling claim frequency data in general insurance with extra zeros. *Insurance: Mathematics and Economics*, 36(2):153–163.