Non-stationarity in peaks-over-threshold river flows: a regional random effects model

Emma Eastoe

Department of Mathematics and Statistics

Fylde College

Lancaster University

Lancaster

LA1 4YF

UK

e.eastoe@lancaster.ac.uk

March 28, 2018

Abstract

Extreme events of environmental data sets often display temporal non-stationarity. In this paper, models are developed for peaks-over-threshold (POT) river flow data sets in which the sizes of the events are non-stationary. Consider POT data from a single site; for event sizes which are assumed to be stationary over time, the generalised Pareto distribution provides an appropriate asymptotically-motivated statistical model. However, the assumption of stationarity is generally invalid, since the stochastic behaviour of event sizes varies across years under to the influence of other climatological processes, eg. precipitation. If observations on these underlying processes are available, parametric and semi-parametric regression methods can be used to model the non-stationarity. Unfortunately these underlying processes are rarely observed and, even if they are, it

is often not clear which process, or combination of processes, should be included in the model. We develop a regional random effects model which accounts for non-stationarity in event sizes without the need for measurements on any underlying processes. The proposed model can be used to predict both unconditional extreme events such as the m-year maximum, as well as extreme events that are conditional on the value of the random effect. Further the random effects can be used to learn about likely candidates for the underlying climatological processes which cause non-stationarity in the flood process. The model is applied to UK flood data from 817 stations split which are between 81 hydrometric regions.

Key words non-stationarity, flooding, random effects, spatial pooling, return levels

1 Introduction

Extreme events extracted from environmental data sets (river floods, wind speeds, precipitation, droughts etc) often display non-stationarity in the form of one or more of: seasonal or decadal cycles, long term trends, inter-year variability and spatial dependence (Cox et al., 2002; Eastoe and Tawn, 2009; Katz, 2010; Chavez-Demoulin and Davison, 2012; Cooley et al., 2012). In this paper we focus solely on modelling temporal non-stationarity. Such non-stationarity is often due to the variables of interest having dependence on, and interactions with, other climate variables which themselves may interact with each other through highly complex and non-linear systems. Non-stationarity may also be caused by human intervention, eg. changes in land-use, or larger-scale effects of climate change. The challenge in extreme value analysis is to make predictions of events even more extreme than those already observed whilst accounting for this non-stationarity. The work presented here was developed specifically to address non-stationarity in extreme river flow events, the methodology should be generally applicable to other environmental data sets.

Changes in frequencies and magnitudes of flood events are a global concern (Easterling et al., 2000; Merz et al., 2010) and within the UK a number of destructive flooding events have been experienced in recent years. For example, flooding which occurred in the aftermath of Storm Desmond in December 2015 caused £400-500 million of damage, mostly in the North West of England and South West of Scotland. Prolonged, intense rainfall led to river flooding which caused flooding of thousands of homes and businesses, evacuations, power-cuts to tens of thousands of properties and long-term damage to transport infrastructure. As this example demonstrates, river flooding in the UK is mostly caused by heavy and prolonged rainfall events; non-stationarity in the duration, frequency, intensity of these rainfall events make it unreasonable to model extreme flooding events using traditional models which assume stationarity. Further, flooding patterns may see step-changes after alterations in land-use (Reynard et al., 2001) eg. building on previously green land or de-forestation. However whilst there has been recent interest in accounting for non-stationarity in flood frequency modelling (Milly et al., 2002; Salas and Obeysekera, 2013; Serinaldi and Kilsby, 2015) there is as yet no universally accepted way to do this.

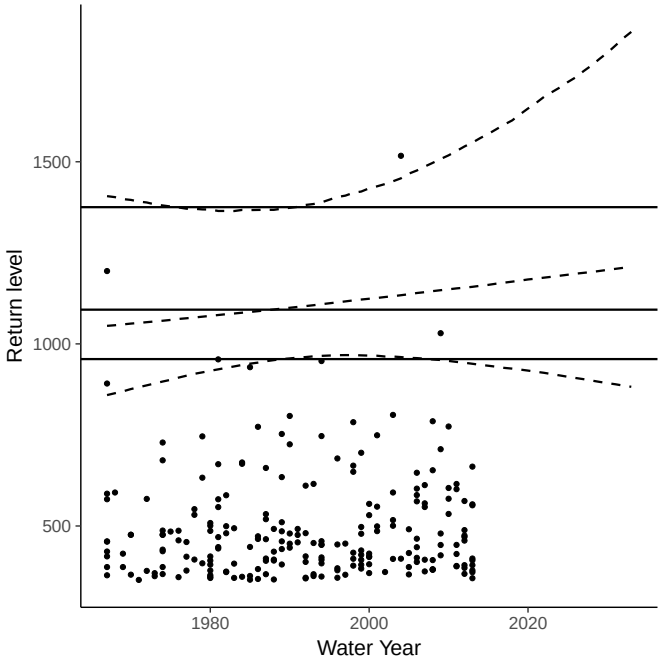

Figure 1: Peaks-over-threshold (POT) data for the River Eden at Sheepmount (dots) with estimated 99% quantiles for the POT sizes under stationary (central full line) and regression (central dashed line) models. The outer lines show 95% credibility intervals for the stationary (full) and regression (dashed) models respectively.

The common approach to modelling river floods is to use peaks-over-threshold (POT) data (Lang, 1999). Exceedances of a high threshold are extracted from daily, or hourly, river flow observations. Exceedances which occur within short time periods are grouped into events, with the time interval between events chosen so that events are approximately independent. The POT data set used in this paper was collated, quality controlled and archived by the National River Flow Archive (NRFA)[1]. Figure 1 shows an example of the data available at a single site, the River Eden at Sheepmount (in Carlisle) with details of the full data set given in Section 5. Sheepmount is an urban site in the North-West of England which was chosen for illustrative purposes as it was affected by major flooding in both January 2005 and December 2015. Data are available for 46 years (1967–2013), with data missing for 1995 only and an average of 4.8 events per year. Figure 1 plots the peak flows as excesses over the event identification threshold of 351.6$m^3 s^{-1}$.

Modelling POT data of this type requires models for both the frequency and size of events. Whereas

[1]Details of how POT extraction was carried out for the data that we use can be found at `http://nrfa.ceh.ac.uk/peaks-over-threshold`

Eastoe and Tawn (2010) examined models for event *frequencies*, here we develop models for event *sizes* where size is the peak flow of an event. To illustrate the difference between stationary and non-stationary models for event sizes, Figure 1 shows two estimates for the 99% quantile. One estimate comes from a model that assumes stationarity, and the other from a regression model which assumes a yearly trend in one of the model parameters (further details are given in Section 3). The regression-based estimate has been extrapolated for 20 years beyond the observation period. During the observation period there is reasonable consistency between the two point estimates, however the quantile estimated from the regression model continues to increase over time, gradually moving further away from the estimate under the stationary model.

The regression-based method demonstrated in Figure 1 is limited since it is only suitable when measurements on covariates (eg. precipitation) are available; alternative approaches are needed when this is not the case. We propose, as one possible model for this scenario, an extension of the random effects model for event frequencies described by Eastoe and Tawn (2010). More detail on the frequency model is given in Section 2, but the central concept is to first specify a probability distribution to model the count data (the 'data model') and then to model the parameters of this distribution as a function of an unobserved random effect (Laird and Ware, 1982), which may change value from year to year. The random effect, which must be estimated, can be viewed as an approximation to any unobserved processes (eg. precipitation) which alter the expected event count for a given year. In a regression model, the parameter of the data model is instead assumed to be a linear function of known covariates. Whilst regression models have lower dimension, since only the coefficients for each covariate must be estimated, the random effects models do not require measurements on covariates, nor do they impose as rigid a structure on the way that the parameters change with these covariates.

To develop equivalent non-stationary models for the sizes of the POT events, a model for the data must first be selected. Pickands (1975) showed that for a random variables Y with common distribution F which has upper end-point y^F, then in the limit as $u \to y^F$ the excess $Y - u | Y > u$ follows a generalised Pareto (GP) distribution. The GP has conditional distribution function given, for

$y > 0$, by

$$\Pr[Y - u \leq y | Y - u \geq 0] = 1 - \left[1 - \frac{k}{\psi} y\right]_+^{1/k}, \tag{1}$$

where $s_+ = \max(0, s)$ and $\psi > 0$ and k are scale and shape parameters respectively. This result holds for almost all distributions F and the value of k is determined by the rate at which the tail of F decays. Following Davison (1984), the simplest model for POT sizes is to assume that all event peaks are independent draws from a GP distribution. We refer to this as the stationary model. This model, and the non-stationary extensions introduced later, are all based on the assumption that the limiting result in equation (1) can be assumed to hold exactly for exceedances of a 'high threshold' u. In the case of the UK POT data, site-specific thresholds were selected to identify the events, so we use these thresholds for our statistical models. In many cases a threshold is not pre-specified and must be selected by the analyst; Scarrott and MacDonald (2012) provide a comprehensive review of ways to do this.

Smith (1989), Davison and Smith (1990) and many subsequent authors, have investigated extensions to the stationary model in which one, or both, of the GP parameters are modelled as functions of covariates, using either linear or smooth functions. Semi- or non-parametric smoothing models, such as locally weighted regression (Hall and Tajvidi, 2000), generalised additive models (Chavez-Demoulin and Davison, 2005) and vector generalised additive models (Yee and Stephenson, 2007), are less restrictive than linear models (Davison and Smith, 1990) since the functional form of the relationship between covariates and model parameter has many more degrees of freedom. However whilst smoothers describe well the *current* behaviour of the process, they cannot be used for extrapolation into the *future*. A viable alternative to both approaches is to model the GP parameters instead as functions of annual random effects. Such a model should give similar within-sample results to the semi-parametric smoothers but, as it is fully parametric, also allows extrapolation. Further, random effects models should be more flexible than fully parametric regression models since they do not impose any structural constraint on the relationship between parameters and covariates. Finally covariates need not be observed, or even

known in order for a random effects model to be fitted.

The rest of the paper is outlined as follows. Section 2 reviews random effects methodology for event frequencies, with new models for the POT event sizes introduced in Section 3. Model fitting procedures are described in Section 4. Due to the hierarchical structure of the model, a Bayesian inference approach is most natural, however this does necessitate the use of computationally intensive Markov Chain Monte Carlo methods to draw from the posterior distributions. A detailed analysis of the UK POT data is given in Section 5. This includes predictions made on future extreme events using the proposed model. Finally, some observations and thoughts for future work are given in Section 6.

2 Background

2.1 Random effects model for counts

The following overview of the random effects model which Eastoe and Tawn (2010) used to model event frequencies provides an introduction to the concept of random effects modelling which is fundamental to the non-stationary GP models developed in Section 3. Let $\{N_i : i = 1, \ldots, n_y\}$ denote a sequence of annual flood counts where n_y is the number of years. The simplest model for count data assumes that the N_i are independent and identically distributed (IID) with

$$N_i \sim \text{Poisson}(\lambda), \quad \lambda > 0. \tag{2}$$

The rate parameter λ is constant over time following the assumption that the counts are identically distributed. A generalised linear modelling approach can be taken if it is thought that the mean behaviour of the count distribution varies in response to some observed covariates. Let $\mathbf{x}_1, \ldots, \mathbf{x}_{n_y}$ be vectors of covariates which are observed on an annual scale e.g. annual mean precipitation. Then

model (2) can be extended by setting

$$\begin{aligned} N_i &\sim \text{Poisson}(\lambda_i) \\ \log \lambda_i &= \boldsymbol{\lambda}' \mathbf{x}_i, \end{aligned} \tag{3}$$

where $\boldsymbol{\lambda}$ is a vector of coefficients each of which describes the effect of a covariate on the Poisson rate. Note that the N_i are still assumed to be independent, and the log link function ensures that λ_i is always positive.

The regression model in equation (3) can only be fitted when all covariates are known and observed. Further it enforces a fairly restrictive form on the relationship between covariate and parameter, and hence also between covariate and expected annual count. An alternative approach is to model the rate as a function of an unknown random effect γ_i which varies across years:

$$\begin{aligned} N_i &\sim \text{Poisson}(\lambda\gamma_i), \quad \lambda > 0, \\ \gamma_i &\sim \text{Gamma}(1/\alpha, 1/\alpha), \quad \alpha > 0. \end{aligned} \tag{4}$$

The unknown random effects $\{\gamma_i\}$ can be viewed as an approximation to any unobserved annual-scale covariates (eg. mean annual precipitation) which influence the flooding process. Since the values of the γ_i are unknown, they must be estimated from the data, for example by specifying an additional parametric model for the random effects. The simplest such model is to assume that the γ_i are IID random variables from some pre-defined distribution, which in the above definition is the Gamma$(1/\alpha, 1/\alpha)$ distribution, and the parameter(s) of this distribution must also be estimated during the model fit. For further details of this model, including generalisation to a point process model for event occurrences, we refer the reader to Eastoe and Tawn (2010).

2.2 Stationary model for event sizes

Turning now to event sizes, recall from Section 1 that a suitable model for the sizes of POT events is the two-parameter GP distribution. For data sets of independent and stationary event sizes, numerous methods have been proposed to estimate the GP parameters, including: method of moments, probability weighted moments (Hosking and Wallis, 1987), maximum likelihood (Davison, 1984), likelihood-moments (Zhang, 2007), elemental percentile (Castillo and Hadi, 1997), Bayes (Coles and Tawn, 1996; Castellanos and Cabras, 2007) and empirical Bayes (Zhang and Stephens, 2009) approaches. Each method has advantages and disadvantages, with some being more robust when applied to small data sets and others being more suited to heavy tailed data sets. Mackay et al. (2011) provide a thorough review of these and other methods, as well as a comparison via a simulation study. Given the range of choice, we choose the Bayesian framework since moments-based inference does not easily extent to non-stationary models, and inference for random effects is more natural random effects in the Bayesian setting than the likelihood one (see discussion in Section 4).

Let $\mathbf{y} = (y_1, \ldots, y_n)$ denote event sizes from a POT data set. Assuming that the data are independent draws from a GP(ψ, k) distribution, the posterior distribution for the parameters (ψ, k) is found using Bayes theorem,

$$\pi(\psi, k|\mathbf{y}) = L(\mathbf{y}|\psi, k)\pi(\psi, k)$$

where L is the likelihood function

$$L(\mathbf{y}|\psi, k) = \psi^{-n} \prod_{i=1}^{n} \left[1 - \frac{k}{\psi} y_i\right]_+^{1/k-1},$$

and $\pi(\psi, k)$ is a prior distribution on the parameters. No conjugate prior exists, so uninformative priors are usually specified and a sampling method, such as Markov Chain Monte Carlo (MCMC) sampling, is used to estimate the posterior distribution. Expert information about the parameters can be incorporated via the prior distribution (Coles and Tawn, 1996); since extreme events are scarce by

definition, this can be extremely beneficial. However it is also rare for such information to be available.

Since the GP is fitted conditional on observing a threshold excess, this conditioning must then be undone to estimate the full marginal distribution for the tail of Y. Let $\phi = \Pr[Y > u]$ then, for $y > 0$,

$$\Pr[Y \leq u + y] = 1 - \Pr[Y - u > y | Y - u > 0] = 1 - \phi \left[1 - k\frac{y}{\psi}\right]_+^{1/k}. \tag{5}$$

Consequently, there is now a third parameter ϕ which must be estimated. Assuming that events occurrence is stationary in time, ϕ is estimated using the Poisson model of Section 2.1. If λ is the expected number of extreme events per year and n_{py} is the total number of observations per year, then $\phi = \lambda / n_{py}$, and the probability of exceeding any level above u follows from equation (5). For stationary data it is also common to estimate the return level r_N, the level exceeded on average once every N years. Since it follows that r_N has probability of exceedance $p = 1/(N n_{py})$, equation (5) can be inverted to give

$$r_p = \frac{\psi}{k}\left[1 - \left(\frac{1-p}{\phi}\right)\right]^k.$$

As we discuss below, it is less obvious what this quantity means in the context of non-stationarity.

3 Methods for non-stationarity

3.1 Review of existing methods

There is a substantial literature on the use of parametric and semi-parametric regression models for extremes (see Section 1). We shall work within the parametric regression framework, first proposed in the context of extreme value modelling by Smith (1989) and Davison and Smith (1990). These models account for non-stationarity in the events sizes $\mathbf{y}$ by allowing one or both of the GP parameters to be a linear function of covariate(s). Let $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,p})$ denote a vector of p covariates corresponding to event peak y_t and let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_p)$ and $\mathbf{k} = (k_1, \ldots, k_p)$ denote vectors of unknown regression

coefficients. Then for $t = 1, \ldots, n$,

$$\begin{aligned} Y_t - u | Y_t > u &\sim \text{GP}(\psi(\mathbf{x}_t), k(\mathbf{x}_t)) \\ \log \psi(\mathbf{x}_t) &= \boldsymbol{\psi}' \mathbf{x}_t \\ k(\mathbf{x}_t) &= \mathbf{k}' \mathbf{x}_t. \end{aligned} \tag{6}$$

We refer to this as the GP regression model. For river flooding, ideal covariates would reflect both meteorological conditions and the recent state of the catchment, eg. precipitation over the last week and soil moisture deficit integrated over recent months.

Before considering interpretation of the model, there are several things to note about the model implementation. Firstly, a log link function is used to ensure that the scale parameter is always positive. Secondly, the two parameters need not both include all p parameters and covariate selection, eg. comparison of likelihood ratio statistics or Bayes factors, should be used to select the significant covariates. Covariate effects in the shape parameter can be hard to identify and it is often assumed that $k(\mathbf{x}_t) = k$. As with the stationary GP model, the parameters (regression coefficients) can be estimated using either maximum likelihood or Bayesian inference. Recent variations on this model which we do not investigate further include using a covariate-dependent threshold or pre-processing the data prior to fitting the GP model. For the former: Kyselỳ et al. (2010) and Northrop and Jonathan (2011) obtain such a threshold using quantile regression, whereas Sigauke and Bere (2017) use a cubic smoothing spline. See Eastoe and Tawn (2009) for an implementation of pre-processing.

Now consider interpretation of the model. The regression coefficient ψ_j (k_j) measures the effect of the jth covariate on the scale (shape) parameter. By construction, the effect size is the same across the covariate domain, so it is unwise to extrapolate the model beyond the observed range of the covariate. To see why, suppose there is a single covariate x_t which is the year in which observation t occurred. Whilst it might be reasonable to assume that the scale parameter increases linearly across years during the observation period, it is unlikely that this rate of increase will carry on indefinitely over, say, the next century. This makes prediction of future extreme events difficult and is a distinct disadvantage of

regression models, further motivating the need for an alternative approach.

3.2 Random effects models

Random effects have been used to capture spatial heterogeneity in the GP parameters when modelling precipitation (Cooley et al., 2007; Cooley and Sain, 2010; Sang and Gelfand, 2010) and wildfires (Turkman et al., 2010). We refer the reader to Section 4 of Davison et al. (2012) for a full review of these methods. Our model differs from these since the goal is to capture temporal, rather than spatial, trends. We start by introducing the basic separate-site GP random effects model. Once this has been defined, we motivate a regional model which uses a spatial pooling approach, similar to that used in regional frequency analysis (RFA) (Hosking and Wallis, 1997; Fowler and Kilsby, 2003), to gain power in estimating the time-varying random effects for sites with short record lengths. However, unlike in RFA in which all data within a pre-defined region is pooled, our model has a time-varying random effect which is shared across all sites within a region whilst all other parameters in the model remain site-specific.

Let $\{Y_{ji} : j = 1, \ldots, n_y; i = 1, \ldots, n_j\}$ denote a sequence of random variables representing the POT event sizes for a particular site. The variable Y_{ji} corresponds to event i in year j, n_y denotes the number of years in the observation period and n_j denotes the number of points in year j. The GP random effects model is then defined for $j = 1, \ldots, n_y$ and $i = 1, \ldots, n_j$ as

$$\begin{aligned} Y_{ji} - u | Y_{ji} > u &\sim \mathrm{GP}(\psi_j, k) \\ \log \psi_j &= \psi_0 + \zeta_j \\ \zeta_j &\overset{IID}{\sim} \mathrm{N}(0, \tau), \quad \tau > 0 \end{aligned} \tag{7}$$

Further, the excesses are assumed to be independent conditional on the random effects. In this model, the shape parameter is constant in time and the scale parameter varies randomly over years since it is a function of the random effects $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{n_y})$. Inter-year variability is controlled by the random effects parameter τ, with larger τ corresponding to greater variability. The link function ensures that

the scale is always positive without needing constraints on either the intercept ψ_0 or the random effects. As in the Poisson model, the random effects $\boldsymbol{\zeta}$ can be viewed as a proxy for unobserved climate-related processes. There are $n_y + 3$ parameters to estimate: $\boldsymbol{\theta} = (\tau, \psi_0, k)$ and $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_{n_y})$.

The UK POT data sets used in Section 5 typically have an average of 5 events per year, with sites having between 10 and 100 years of data. Initial implementations of model (7) showed that it was hard to identify the random effects in many of these data sets due to relatively small sample sizes and at sites with short record lengths any signal in the estimated random effects was eclipsed by the sampling and parameter uncertainty. This led to the development of a *regional random effects* model. For the UK POT data, *region* refers to areas known as hydrometric regions. These regions are mostly integral river catchments which share one or more common outlets, or occasionally multiple adjacent catchments which have separate outlets but share similar topography. It can therefore be assumed that any climate-related processes which influence flood sizes will act on a spatial scale that encompasses all sites within a particular region, and so a common annual random effect for all sites within one region is reasonable.

Let Y^h_{sji} define the ith event peak in year j at site s in hydrometric region h. Let n_h be the number of sites in hydrometric region h then, for $s = 1, \ldots, n_h$,

$$\begin{aligned} Y^h_{sji} - u_s | Y^h_{sji} > u_s &\sim \text{GP}(\psi^h_{sj}, k_s) \\ \log \psi^h_{sj} &= \psi_{s,0} + \zeta^h_j \\ \zeta^h_j &\overset{IID}{\sim} \text{Normal}(0, \tau^h), \quad \tau^h > 0. \end{aligned} \tag{8}$$

Conditional on the annual random effects, the event peaks are assumed to be independent in both time and space. The scale parameter intercept $\psi_{s,0}$ and the shape parameter k_s vary freely across sites whereas the annual random effects $\boldsymbol{\zeta}^h = (\zeta^h_1, \ldots, \zeta^h_{n_h})$ and the random effects parameter τ^h are common to all sites. For the UK POT data, pooling the random effects in this way increased their identifiability, thus increasing confidence in the model fit and predictive abilities (see Section 5). The random effects are now a proxy for any climatological processes which act at a regional scale. For any

hydrometric region, the model has $2n_h + n_y + 1$ parameters. Further not all sites in the region will contribute to the estimate of the random effect for a given year; if there were no events observed at that site in that year then there is no information on the random effect.

On a technical note, because the scale and shape parameters in the GP distribution are negatively correlated, Chavez-Demoulin and Davison (2005) show that reparameterising the scale parameter in the GP distribution to be $\nu = \psi/(1-k)$ leads to asymptotically uncorrelated parameters and consequently a likelihood (posterior) surface which is easier to explore. The same re-parameterisation seemed to help considerably in inference for the random effects model. We therefore re-parameterise the GP distribution in equation (8) to be in terms of (ν^h_{sj}, k_s), placing the regional random effects directly on the logarithm of ν^h_{sj}.

3.3 Mixture model

The GP regression model defined in equation (6) and the regional random effects model in equation (8) can be combined to create a regional mixed effects model. Let $\mathbf{x}_{sji}$ denote the vector of p covariates corresponding to the ith event in year j at site s, and let $\boldsymbol{\psi}_s$ by a vector of p regression coefficients which are specific to site s, then the regional mixed effects model is defined as follows:

$$\begin{aligned} Y^h_{sji} - u_s | Y^h_{sji} > u_s &\sim \mathrm{GP}(\psi^h_{sji}, k^h_s) \\ \log \psi^h_{sji} &= \psi_{s,0} + \boldsymbol{\psi}'_s \mathbf{x}_{sji} + \zeta^h_j \\ \zeta^h_j &\overset{IID}{\sim} \mathrm{Normal}(0, \tau^h). \end{aligned} \tag{9}$$

The regression coefficients and shape parameter are site specific, with only the random effects shared across the hydrometric region h, although this could be altered so that regression coefficients are also shared over the region. Covariates are not restricted to be at the same spatial-temporal resolution as the random effects, thus we could include covariates that vary within a year alongside annual random effects and because both covariates and regression coefficients are site-specific, the random effects account for any inter-year variability that is not already accounted for at the site-level by the covariates

(precipitation, soil moisture etc).

For the UK POT data (see Section 5), no physical covariates are available and a simplified version of this model is used with a single site-specific covariate x_{sji} taken to be the year of observation j. The scale parameter of this model has three components: intercept, site-specific annual linear trend, and regional annual random effect. Comparison of this model with the GP regional random effects model, allows assessment of the contributions to the inter-year variability from the monotonic linear trend and the unobserved climate processes respectively.

4 Inference

In a likelihood framework, estimation of parameters and random effects is complicated by the fact that the former are treated as fixed but unknown values, whereas the latter are random variables. In a Bayesian framework both parameters and random effects are treated as random variables, which conceptually simplifies the inference. In addition, the resulting posterior distributions provide a natural basis for prediction (Section 4.2). Writing $\boldsymbol{\nu}_0 = (\nu_{0,1}, \ldots, \nu_{0,n_s})$ and $\mathbf{k} = (k_1, \ldots, k_{n_s})$ as the vectors of site-specific scale intercept and shape parameters respectively, the full joint posterior for the GP regional random effects model is

$$\pi(\boldsymbol{\nu}_0, \mathbf{k}, \tau^h, \boldsymbol{\zeta}^h | \mathbf{y}) \quad \propto \quad L(\mathbf{y}|\boldsymbol{\nu}_0, \mathbf{k}, \boldsymbol{\zeta}^h) f(\boldsymbol{\zeta}^h|\tau^h) \pi(\boldsymbol{\nu}_0) \pi(\mathbf{k}) \pi(\tau^h). \tag{10}$$

This posterior distribution combines the likelihood function for the data,

$$L(\mathbf{y}|\boldsymbol{\nu}_0, \mathbf{k}, \boldsymbol{\zeta}^h) \quad = \quad \prod_{s=1}^{n_h} \prod_{j=1}^{n_y} \left\{ \prod_{i=1}^{n_{sj}} {\psi_{sj}^h}^{-1} \left[1 - \frac{k_s}{\psi_{sj}} y_{sji} \right]_+^{1/k_s - 1} \right\}^{\mathbb{I}_{sj}}$$

with the conditional prior for the random effects,

$$f(\boldsymbol{\zeta}^h|\tau^h) \quad = \quad \prod_{j=1}^{n_y} \phi(\zeta_j^h; 0, \tau^h),$$

and prior distributions for each of the parameters, $\pi(\boldsymbol{\nu}_0)$, $\pi(\mathbf{k})$ and $\pi(\tau^h)$. In the above expressions, n_{sj} is the non-zero number of events in year j at site s, $\mathbb{I}_{sj}$ is an indicator function taking the value 1 is there is at least one event at site s in year j and zero otherwise, and $\phi(\cdot;\mu,\sigma)$ denotes the probability density function of the Normal(μ,σ) distribution. In this paper it is assumed a priori that the parameters $\boldsymbol{\nu}_0$, $\mathbf{k}$ and τ^h are independent, with the components of each of $\boldsymbol{\nu}_0$ and $\mathbf{k}$ also assumed to be mutually independent. Flat Normal priors are then place on each component of $\boldsymbol{\nu}_0$ and $\mathbf{k}$, and on τ^h.

For the mixture model, the posterior is very similar, except that a prior for the regression coefficients must be specified. The analysis conducted in this paper takes a simple uninformative prior, with all coefficients across all sites assumed to be mutually independent, and a flat Normal prior placed on each of the site-specific coefficients $\boldsymbol{\nu}_s$. Let $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{n_s})$, then the posterior distribution for the mixed model is

$$\pi(\boldsymbol{\nu}_0, \boldsymbol{\nu}, \mathbf{k}, \tau^h, \boldsymbol{\zeta}^h|\mathbf{y}) \quad \propto \quad L(\mathbf{y}|\boldsymbol{\nu}_0, \boldsymbol{\nu}, \mathbf{k}, \boldsymbol{\zeta}^h) f(\boldsymbol{\zeta}^h|\tau^h)\pi(\boldsymbol{\nu}_0)\pi(\boldsymbol{\nu})\pi(\mathbf{k})\pi(\tau^h),$$

where the likelihood for the data and the conditional prior for the random effects can both be derived from the model specification given in equation (9).

4.1 MCMC scheme

For some random effects models, including the Poisson model in equation (4), it is possible to integrate the random effects out of the likelihood so that inference for the parameters is carried out independently to inference for the random effects. This is not possible for either of the GP regional random effects models. Further, for both models neither the joint nor conditional posterior distributions are available in closed form, and so these must be estimated by sampling. Since the joint posterior distributions for both random effects and mixture models are only known up to an unknown normalising constant, sampling methods from the MCMC toolkit must be used (Gilks et al., 1995). In what follows, 'parameters' refers to both model parameters and random effects.

MCMC sampling provides iterative schemes designed to converge to the target posterior distribution

after an initial 'burn-in' period. These algorithms must be tuned to ensure both convergence, since the length of the burn-in period is unknown, and adequate mixing (Brooks and Roberts, 1998), where mixing means that the posterior parameter space is fully explored. Since the random effects models have a large number of parameters, tuning is a time-consuming process and so we use an adaptive MCMC algorithm (Andrieu and Thoms, 2008; Roberts and Rosenthal, 2009) which automatically tunes the sampling mechanism to ensure mixing and enables inference to be automated to a greater degree than with a non-adaptive algorithm. For us, this makes it easier to fit the model to all hydrometric regions without needing to manually tune the MCMC algorithm each time. Despite the automation, 'spot-checks' of the trace plots of the parameter draws are still made for randomly chosen regions and sites.

We considered three adaptive algorithms. The first was proposed by Haario et al. (2001) and the second is an adaptation of the first by Roberts and Rosenthal (2009). Both algorithms allow joint updating of all parameters, however both have additional tuning parameters which we found made both algorithms difficult to automate. The third algorithm, an adaptive Metropolis-Within-Gibbs scheme proposed by Roberts and Rosenthal (2009), worked much better. This algorithm updates each parameter separately via a Metropolis-Hastings Random Walk (MHRW) step, allowing the scaling parameter used for these updates to evolve and eventually to converge to an optimum value, resulting in an optimal acceptance rate for each parameter.

Consider a generic parameter vector $\boldsymbol{\theta}$, with ith component θ_i, data $\mathbf{y}$ and a posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$. For a *non-adaptive* componentwise MHRW with Normal update, a scaling parameter $\sigma_{\theta,i}$ is fixed in advance so that at the jth iteration of the sampling scheme, a proposed value for the parameter θ_i^p is simulated from the current value of the chain θ_i^{j-1} using the equation

$$\theta_i^p = \theta_i^{j-1} + \sigma_{\theta,i} Z_i, \quad Z_i \sim \text{Normal}(0,1).$$

The proposal is accepted with probability proportional to the ratio of the posterior density at each of the proposed and current values; if the proposal is rejected, then θ_i^j is set equal to θ_i^{j-1}. The optimal value

of the scaling parameter $\sigma_{\theta,i}$, ie. the value which gives the optimal acceptance rate (Roberts et al., 2001), depends on the unknown parameter space. In practice, $\sigma_{\theta,i}$ is selected by running multiple chains each with a different value for $\sigma_{\theta,i}$ and comparing diagnostics, eg. trace plots and effective sample size. For multi-parameter models, components of the parameter vector can be updated either jointly or sequentially at each iteration of the algorithm. For further details see Chib and Greenberg (1995).

In the *adaptive* version given by Roberts and Rosenthal (2009), the components of the parameter vector are updated sequentially at each iteration, but now the component-wise scaling parameters are adjusted by a pre-defined amount every B iterations. Adjustment depends on (i) the component-wise acceptance rate within the preceding block of B iterations and (ii) how many blocks of B iterations have already run. For (i), if the acceptance rate is lower (higher) than the pre-defined optimal rate, then the scaling parameter is decreased (increased). For (ii), changes in the absolute value of the scaling parameter decrease as the algorithm progresses. We follow Roberts and Rosenthal (2009), taking $B = 50$, aiming for a component-wise optimal acceptance rate of 0.44 and setting the scaling parameter change as $\min(0.01, n_B^{-1/2})$ where n_B is the current number of blocks of B iterations. We found the algorithm to be relatively slow, since components are updated separately, however the increased computational burden was off-set by the speed-up in convergence due to automatic selection of the scaling parameter.

4.2 Prediction

One common output from an analysis of extreme events is the return level curve. However under non-stationarity, the traditional concept of a return level may be somewhat misleading. For example, consider a process $\{Y_t\}$ with N-year return level r_N. From the definition of a return level in Section 2, the expected waiting time between exceedances of r_N is N years. Letting n_{py} be the number of observations per year, eg. $n_{py} = 365$ if data are observed daily, it follows that for all t,

$$\Pr[Y_t > r_N] = \frac{1}{N n_{py}}.$$

That is, for any t, the probability of the N-year return level being breached is $(Nn_{py})^{-1}$. If the process is non-stationary it is not reasonable to assume the existence of a value r_N for which the exceedance probability is constant in time. For example, suppose the distribution of flood event sizes changes according to a climate regime which, for any given year, will be in one of two unknown and randomly selected states. Intuitively, we should first define an N-year return level for each state in the climate regime; these are referred to as 'conditional' return levels by Eastoe and Tawn (2009). The overall (or marginal) return level should then account for the frequencies of the climate states. This becomes increasingly complicated if the climate state space is continuous not discrete, there are multiple climatic or geophysical processes involved and/or the process(es) are unobserved.

One solution is to use simulation to predict extreme events such as the m-year maximum, ie. the largest event in m years of concurrent data. This has several advantages: simulated values can be fed directly into numerical risk models, simulation of extreme events over different time intervals is trivial, and interpretation of such events is no longer in terms of probabilities. By combining the GP random effects model with a model for annual counts, eg. the Poisson random effects model of Section 2.1, it is extremely straightforward to simulate m years of POT data. First the number of events for each of the m years is simulated using the counts model. For the event peaks, m years of regional random effects $\zeta_1, \ldots, \zeta_m$ are then drawn from the Normal$(0, \tau^h)$ distribution. Given both the simulated counts and the regional random effects, event peaks are then simulated from the GP(ψ_{sj}, k_s) distribution described in equation (8), and any summary statistic W, eg. the m-year maximum, is extracted directly. Carrying out this procedure for multiple draws from the posterior distribution results in both a point estimate and a credibility interval for W.

5 Analysis of UK POT data

The UK POT data set introduced in Section 1 contains POT events for 826 sites over 90 hydrometric regions. Of these regions, 9 have only one site and so are excluded from the anaylsis, leaving 817 sites over 81 hydrometric regions. The median number of sites in a region is 7, with lower and upper

quartiles of 3 and 12 respectively. Due to the fact that not all sites in the network were established at the same time, the number of years of data per site varies considerably across sites. The longest record is 159 years, with two further sites having over 100 years of data; 113 sites have over 50 years of data, and 604 sites have 20-50 years. There is no spatial clustering evident in the record lengths, though there is very little data in the Highlands of Scotland - possibly because the region is sparsely populated and flooding there would have minimal impact. Some sites have missing data but since we have no information to the contrary, we treat this as missing at random and therefore ignorable.

The adaptive MCMC algorithm discussed in Section 4 was used to sample from the joint posterior distribution of the parameters and random effects in the GP regional random effects model. For each region, the algorithm was run with a burn-in of 53000 iterations, after which the chains were thinned to reduce auto-correlation by keeping only every 10th iteration. Inspection of trace plots for a randomly selected subset of the region suggested that some could have had a far shorter burn-in. A smaller number may have benefitted from slightly longer chains and a longer thinning lag to reduce further the sample auto-correlation. The thinning lag selected was a compromise between computational time and an adequate sample size.

5.1 Single regional analysis

To enable a detailed investigation of the model, and to compare the separate-site stationary, separate-site regression, separate-site random effects, regional random effects and regional mixed models, we first present results for a single hydrometric region. Region 76 was chosen because it contains the city of Carlisle which lies on the River Eden (see Section 1). Carlisle experienced severe flooding in both January 2005 and December 2015 and although it escaped the 2011 floods which affected many nearby areas, nearby transport infrastructure was badly affected. The aftermath of the 2005 event has been well documented in the hydrology (Neal et al., 2009; Roberts et al., 2009; Horritt et al., 2010), public health and well-being (Carroll et al., 2009, 2010) and planning (Crichton et al., 2009) literatures.

Estimates of the shape parameter k, from the separate-site stationary GP and both separate-site and regional random effects GP models, are shown in Table 1 for each site in the region. For most

sites the tail becomes lighter once non-stationarity is accounted for. For this region not properly accounting for inter-year variability in the scale parameter results in this variability being absorbed into the shape parameter leading to a heavier tailed distribution. Uncertainty in the estimates is largest for the separate-site random effects model. Whilst the regional regional random effects model might be expected to have the narrowest credibility intervals due to the sharing of information across sites, there is a trade-off between information gained and the increased number of model parameters, so this is not always the case.

Site ID	Separate-site (stationary)	Separate-site (random effects)	Regional
76001	-0.158 (-0.450,0.043)	0.051 (-0.279,0.350)	-0.114 (-0.409,0.105)
76002	-0.080 (-0.324,0.095)	-0.060 (-0.315,0.150)	0.123 (-0.112, 0.309)
76003	-0.076 (-0.289,0.081)	-0.052 (-0.278,0.130)	0.131 (-0.065,0.296)
76004	-0.087 (-0.2240.027)	-0.038 (-0.184,0.094)	0.085 (-0.051,0.195)
76005	-0.230 (-0.455,-0.064)	-0.151 (-0.415,0.137)	0.007 (-0.205,0.177)
76007	-0.068 (-0.229,0.047)	-0.045 (-0.214,0.108)	0.155 (-0.010,0.291)
76008	0.019 (-0.120,0.135)	0.049 (-0.094,0.181)	0.104 (-0.024,0.212)
76010	-0.065 (-0.238,0.052)	0.030 (-0.184,0.338)	0.014 (-0.151,0.146)
76011	-0.053 (-0.195,0.037)	-0.037 (-0.184,0.086)	-0.023 (-0.158,0.065)
76014	0.133 (-0.057,0.286)	0.138 (-0.058,0.296)	0.157 (-0.038,0.308)
76015	-0.101 (-0.303,0.023)	-0.086 (-0.294,0.088)	0.011 (-0.195,0.158)
76017	-0.119 (-0.354,0.029)	-0.040 (-0.311,0.247)	0.155 (-0.084,0.355)
76019	-0.025 (-0.276,0.138)	0.301 (-0.15,0.841)	0.045 (-0.206,0.212)
76806	-0.232 (-0.809,0.150)	-0.190 (-0.788,0.247	-0.104 (-0.649,0.235)
76809	-0.039 (-0.344,0.155)	-0.008 (-0.322,0.259)	-0.029 (-0.329,0.169)
76811	-0.166 (-0.602,0.095)	0.097 (-0.480,0.858)	-0.110 (-0.535,0.132)

Table 1: Shape parameter estimates (posterior median) for separate-site stationary, separate-site random effects and regional random effects GP models. Numbers in brackets are 95% credibility intervals.

The posterior density for the regional random effects parameter τ^h is shown in Figure 2. The posterior median (95% credibility interval) for this parameter is $0.37\ (0.29, 0.47)$. Also shown are the individual site random effects parameters from the separate-site random effects models, for which point estimates lie between 0.11 to 0.52, with a cross-site median of 0.21. Sharing random effects between sites results in an increase in the inter-year variability of the random effects and a decrease in the uncertainty of this variability. These changes may be attributed to increased information from pooling across sites which increases the information on the unobserved process.

Figure 3 shows estimates of the time-varying scale parameter $\psi_{s,j}$ for each of the sites in region 76; the plot is shaded to show variability with site latitude. It is clear that the scale parameter varies

considerably across years and this is particularly evident at sites with larger scale parameters. There appears to be some North-South trend, with northerly sites generally having larger scale parameters. No similar East-West trend was evident, and it is possible that the North-South pattern simply reflects the fact that the catchments to the North are larger and therefore see bigger floods (see comment on this below). A similar analysis of the shape parameter estimates (not shown) showed that these had neither a North-South nor an East-West trend. Also shown in Figure 3 are the separate-site and regional random effects model estimates of the scale parameter $\psi_{s,j}$ for the River Eden at Sheepmount (see Section 1). The stationary fit at this site gives $\hat{\psi} = 136.6\ (113.0, 164.3)$. The plots show clearly that there is more power to describe inter-year variability when the additional assumption of shared regional random effects is made; the regional model estimates are significantly different from each other in many years, and from the stationary estimate in most years. The uncertainty in the separate-site model (not shown) is far greater than that in either the stationary model or the regional model. Similar results were seen across the remaining sites.

Separate-site regression and regional mixed models with a linear time trend were also fitted. In the mixed model only the random effects are shared across sites, and the linear trend is site-specific. Comparison of the scale parameters for the mixture model and the regional random effects model show negligible differences both in terms of point estimates and credibility intervals, suggesting that the linear trend is very weak in comparison to that of the random effects. Table 2 shows the linear trend under both models with a regression component. Under the mixed model only three sites have a credibility interval which does not contain zero, and in each case zero lies just below the interval and the remaining 13 sites have no significant trend. Under the separate-site regression model, only two of the sites have a significant trend, although the actual coefficients are similar in value to those estimated under the mixed effects model. A comparison of the scale parameters from the regional random effects and separate-site regression models for the River Eden at Sheepmount is shown in Figure 3. The trend seen in the regression model is reflected in the more flexible random effects model, even though a trend is not explicitly defined. The random effects show greater variability around the longer-term trend since they are not restricted to change monotonically.

Model fit was assessed through a visual inspection of quantile-quantile (QQ) plots (see Figure 10 and following section for discussion). These QQ plots suggest that all models were a good fit to the data, with the random effects models showing the best fit due to their smoothing effect and so we move to prediction of extreme events. Figure 4 shows, for each year in the observational period, model-based estimates of (i) the conditional probability of exceeding the marginal 90% quantile of the observations and (ii) the conditional 99% quantile of the peak sizes, for the River Eden at Sheepmount. Quantities (i) and (ii) condition both on an event being above the threshold and on an event being in a given year. Under the separate-site stationary model, the estimated exceedance probability is 0.11 (0.080,0.14) and the estimated quantile is 742.3 (606.7,1024.8). Results for the regional mixed model are not shown since they were almost identical to results from the regional random effects model.

There are considerable differences between estimates from the three models. Consider the separate-site regression and random effects models; the probabilities estimated using the former appear to be a smoothed version of those from the latter model. On the other hand, the quantiles estimated from the random effects model do not appear to have the same increasing trend as those predicted by the regression model, yet the magnitudes of the quantiles estimated under the two models are not dissimilar. In contrast, estimates of both extreme probabilities and quantiles from the regional random effects model show far more inter-year variability than the estimates from either separate-site model. It can be concluded that both the likelihood of an extreme event, and the size of an extreme event with a given waiting time, vary hugely from year to year; inspection of the credibility intervals (final plot in Figure 4) suggests that these changes are not just down to random variability.

Finally, we remove the water year 2004 from the data set to assess model sensitivity to the exceptional event of January 2005 which resulted in major flooding to parts of this region. Separate-site stationary, regional random effects and separate-site regression models were re-fitted, with time-trend coefficients for the latter displayed in Table 2. With this year excluded, there was a decrease in time trend estimates at almost all sites, and at Sheepmount the coefficient is less than a fifth of what it was with 2004 included. Figure 5 shows the conditional 99% quantile of the POT sizes at Sheepmount, estimated from separate-site regression and regional random effects models. These results suggests

Site ID	Regression Separate-Site	Mixed	Regression w/o 2004 Separate-Site
76001	1.11 (-0.117,2.23)	0.765 (-0.573,2.07)	1.12 (-0.977,2.24)
76002	0.66 (-0.311,1.57)	1.10 (0.223,1.90)	0.619 (-0.339,1.56)
76003	0.549 (0.0247,1.06)	0.740 (0.17,1.33)	0.467 (-0.0355,1.00)
76004	-0.153 (-0.572,0.262)	0.0679 (-0.452,0.588)	-0.190 (-0.611,0.299)
76005	0.530 (-0.204,1.29)	0.709 (-0.0368,1.45)	0.364 (-0.346,1.10)
76007	0.173 (-0.403,0.761)	0.250 (-0.377,0.903)	0.0478 (-0.488,0.613)
76008	0.238 (-0.181,0.683)	0.401 (-0.171,0.932)	0.209 (-0.209,0.646)
76010	1.07 (0.377,1.81)	0.937 (0.136,1.73)	0.902 (0.220,1.61)
76011	-0.478 (-1.07,0.11)	-0.518 (-1.20,0.18)	-0.535 (-1.11,0.0609)
76014	0.175 (-0.343,0.725)	0.0453 (-0.608,0.720)	0.119 (-0.399,0.677)
76015	0.261 (-0.660,1.12)	0.0201 (-0.905,0.953)	0.200 (-0.724,1.05)
76017	0.226 (-2.26,2.80)	0.104 (-1.82,2.24)	0.367 (-1.78,2.50)
76019	0.465 (-2.30,3.11)	0.747 (-1.71,3.23)	1.17 (-1.41,3.55)
76806	-0.269 (-5.72,4.46)	0.703 (-3.05,4.15)	0.657 (-4.92,5.48)
76809	2.38 (-0.329,4.86)	2.24 (-0.29,4.55)	2.80 (0.104,5.11)
76811	-3.09 (-6.77,1.77)	-1.27 (-4.56,2.93)	-2.45 (-6.51,2.56)

Table 2: Coefficients for linear trend in the separate-site regression and regional mixed models. Years were standardised by the observed range prior to fitting the model to improve stability in the model fit. Third column shows the same coefficient for the separate-site models, but with water year 2004 removed.

that the regression model is most sensitive to the exceptionally large event in January 2005; once this is excluded the predicted return level is approximately stationary over the observational period. By contrast, predictions under the regional random effects model for the remaining years are unchanged by the exclusion of 2004.

5.2 Full UK analysis

We now fit the regional random effects model to all regions in mainland UK. The mixed model is not implemented as the above case study suggests that, at least when no climate-related variables are available, it performs no differently to the random effects model. Plots of the random effects parameter τ^h are given in Figure 6 show evidence for inter-region variability in the parameter and, although there is no evidence for spatial trends, many neighbouring sites do seem to have similar values for τ_h. Figure 7 shows the site-specific scale intercept $\nu_{s,0}$ increasing with increased DPLBAR (a measure of the catchment size), reflecting the fact that larger catchments will generally have capacity for larger floods. There is also some evidence that sites in the North and West of the country have a higher $\nu_{0,s}$ than sites with an equivalent catchment size in the South and East; this is not unexpected since

Northern and Western areas of the UK suffer most from winter storms originating over the Atlantic. Similar plots for the shape parameter k_s (not shown) suggested no relationship with either catchment size or latitude/longitude. Impact on the shape parameter of the random effects can be seen in Figure 8. In contrast with region 76, generally inclusion of random effects seems to result in a heavier tail.

Estimated random effects are shown in Figure 9 for 1967, 2000 and 2006. These years were chosen as extensive flooding occurred at some point during the year in at least one region. High (low) random effects tend to cluster across neighbouring regions, again illustrating the flexibility of the model: although there is no term in the model to enforce this clustering, it can still reflect that in some years the scale parameter is larger (smaller) across multiple hydrometric regions. The advantage of not incorporating larger-scale spatial homogeneity is that the extent of this homogeneity is not pre-defined; the disadvantage is that we cannot make use of it for predictive purposes.

Standardised QQ plots were inspected for model fit at a random selection of sites with four such plots are shown in Figure 10, for sites in north-west England, south-west Scotland, eastern England and south-west England. The fit is good at all sites, with the possible exception of the Scottish site for which the model over-predicts the highest value. This observation corresponds to an event in December 1994 which caused the worst flooding ever seen in this region, which may explain the poorer fit. The smoothing effect of the model is reflected in the smoothness of the QQ plots; QQ plots from the stationary model are far more jagged.

Estimated exceedance probabilities at all sites are shown in Figure 11 for 1969, 2000 and 2005. As in the previous section, these are the model-based probabilities of exceeding the site-specific 90% quantile v calculated from the observed POT data, and are conditional on being above the POT modelling threshold. High (low) exceedance probabilities cluster spatially, with this clustering often extending over multiple hydrometric regions. The plots show that in 2006 the most extreme events were relatively localised occurring mostly in Yorkshire and Lincolnshire, whereas in 1967 and 2000 extreme events were seen over a much wider area. Similar plots (not shown) for higher quantiles v show that more extreme events tend to be more localised.

Finally, the posterior distributions of the m-year maxima were simulated for $m = 10, 50, 100$. These

maxima correspond to the size of the largest event that might be expected in m years, and are not the same as m-year return levels. The distribution for the m-year maxima was obtained by simulating m years of POT events from each of 10000 draws from the posterior distribution of the parameters. Using a different seed for each simulation enables integration over simulation uncertainty, which should anyway be small relative to data and parameter uncertainty. Since this simulation requires a model for event frequencies, the Poisson random effects model given in equation (4) was used.

Posterior distributions of the 10-, 50- and 100-year maxima for the River Eden at Sheepmount are shown in Figure 12. Point estimates for each of these quantities are 937 (675,1562), 1196 (899,1929) and 1306 (992,2098). Both point estimates and plots are consistent with the data. The largest observed data point over the 47 year observation period is 1164, which lies just under the estimated 50-year maxima, and the five largest observations lie above the 10-year maximum with the sixth largest lying just below it. As expected the uncertainty in the estimates increases with the size of the block, reflecting the increase in extrapolation, and the posterior distributions for all the maxima have heavy upper tails, again because of extrapolation uncertainty. Such plots allow visualisation of the kinds of extreme event that might occur once every m years after allowing for inter-year variability.

6 Discussion

We have introduced a GP-based model for non-stationary POT event sizes by assuming that inter-year variability in the scale parameter of the GP distribution varies is due to dependence on one or more unobserved climate-related process. Existing methods based on parametric or semi-parametric regression models cannot be used since the process(es) are unobserved. Instead both the process itself and its effect on the scale parameter are estimated within a random effects modelling framework. Estimation is possible by assuming that the unobserved process(es) can be approximated by an annual random effect which is modelled as an independent sample from some underlying probability distribution. To help identification of these random effects, it is further assumed that, for any given year, all sites within a particular hydrometric region share a common random effect. Such a model is physically justified, since

the sizes of these regions relative to the scale of the surface pressure systems which generate storm events allow the assumption that sites within a region also share a common climatology.

Unlike regression models, the estimated random effects, which can be viewed as a proxy for annual precipitation or some other climate index, can be used to help scientists to look for suitable physical processes to use in a regression model. The random effects model does not place as strong an assumption on the functional form of the model parameters as a fully parametric regression model, allowing greater modelling flexibility, and unlike semi-parametric regression models, can be used to simulate forwards in time hence enabling predictions of the future.

Under the random effects model, it is straightforward to produce estimates of tail probabilities and quantiles for the conditional distribution of POT sizes, where conditioning is on being above the POT modelling threshold. To obtain the equivalent unconditional estimates, a model for event frequencies, such as those in Eastoe and Tawn (2010), is required. Predictions for unconditional extreme events can then be obtained by simulation (Sections 4.2 and 5). When making predictions using a parametric regression model it is possible to make predictions for a *particular* year if covariate values for that year are known. A conditional prediction can *only* be made using the random effects model if we are willing to assume what the value of the random effect will be for that year; it is not possible to obtain conditional predictions automatically, since the model has no way of predicting whether the unobserved process (random effect) will be high or low in that year. Conversely, it is much easier under the random effects model to predict unconditional events such as the m-year maximum since the distribution of the (unobserved) process follows from the model.

We now outline possible extensions to the current work, all of which are beyond the current scope, but are worthwhile future projects. The random effects model presented here makes two strong assumptions: (i) that the distribution of the random effects remains the same over time, and (ii) that the random effects are uncorrelated over time. For many climate-related processes the former is unlikely since many processes show decadal cycles or long term trends. It is also possible that the process may be correlated, at least at short time lags. Building non-stationarity into the random effects distribution is non-trivial, and requires extra assumptions. The simplest approach would be to incorporate decadal

cycles, change points or time-trends into the mean of the random effects distribution using linear regression. An alternative is to model the mean as a function of a Gaussian process or a dynamic linear model (Huerta and Sansó, 2007). The Gaussian process would explicitly allow for autocorrelation in the random effects, whereas the simpler approach of including a decadal cycle would explain autocorrelation only if the dependence were caused by similarities in the random effects of the model.

The model as proposed does not fully exploit the spatial structure of the data since it is assumed that, conditional on the regional random effects, the sites are mutually independent. This is unlikely to be the case, particularly for sites in the same catchment. Extensions could model spatial structure in the GP parameters using Gaussian processes (Cooley et al., 2007; Cooley and Sain, 2010; Sang and Gelfand, 2010; Sharkey and Winter, 2017), or in the regional random effects parameters using a conditional autoregressive (CAR) prior (Besag et al., 1991). Alternatively, it may be possible to model the extremal dependence in the data directly using a multivariate extreme value model. One difficulty with the latter approach is that extreme events do not always occur simultaneously across neighbouring sites. A third possibility would be to extend the 'magnitude adjustment' of the independent log-likelihood (Ribatet et al., 2012; Sharkey and Winter, 2017) to the case of temporally non-stationary data.

Acknowledgements

The author would like to thank Ilaria Prosdocimi (University of Bath) for an initial cleaning of the UK POT data set used here, as well as for discussions which helped start this project. She would also like to thank Rob Lamb (JBA Trust) and Jonathan Tawn (Lancaster University) for further highly useful discussions.

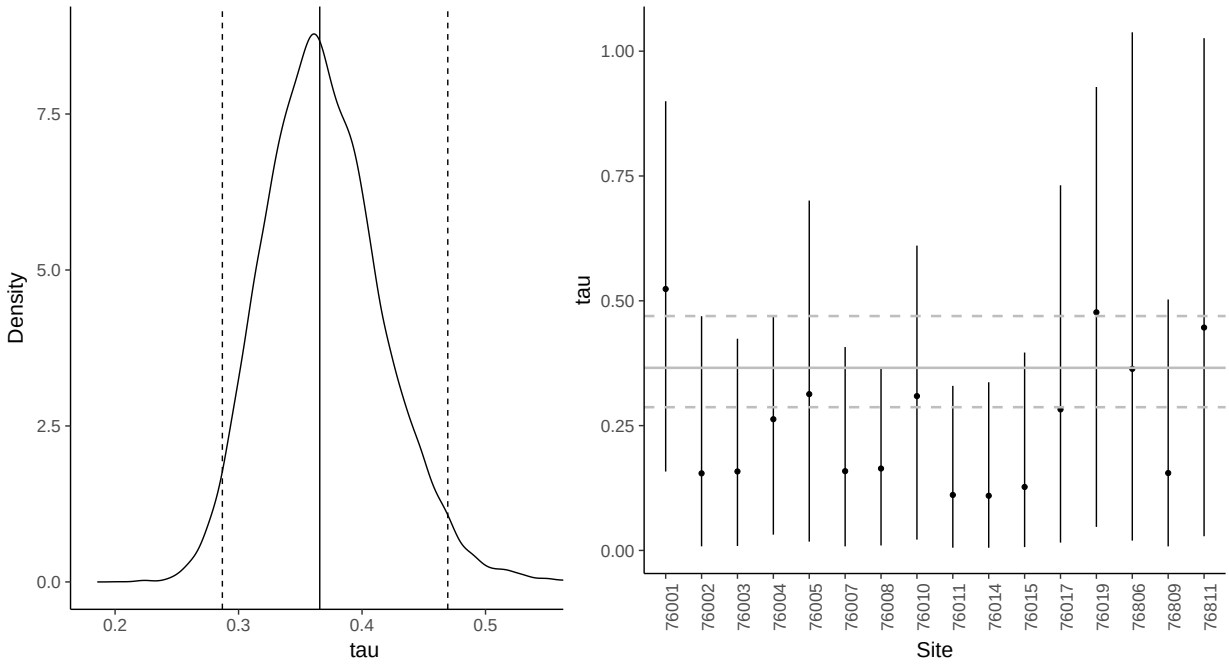

Figure 2: Regional random effects parameter τ^h for region 76: posterior density (left) and comparison of regional estimate with separate-site estimates (right). Vertical lines on the posterior for τ^h show posterior median (centre) and 95% credibility interval. On the right-hand plot, vertical lines show 95% credibility intervals for the separate-site models. Horizontal lines show estimate (full line) and 95% credibility intervals (dashed lines) for the regional model.

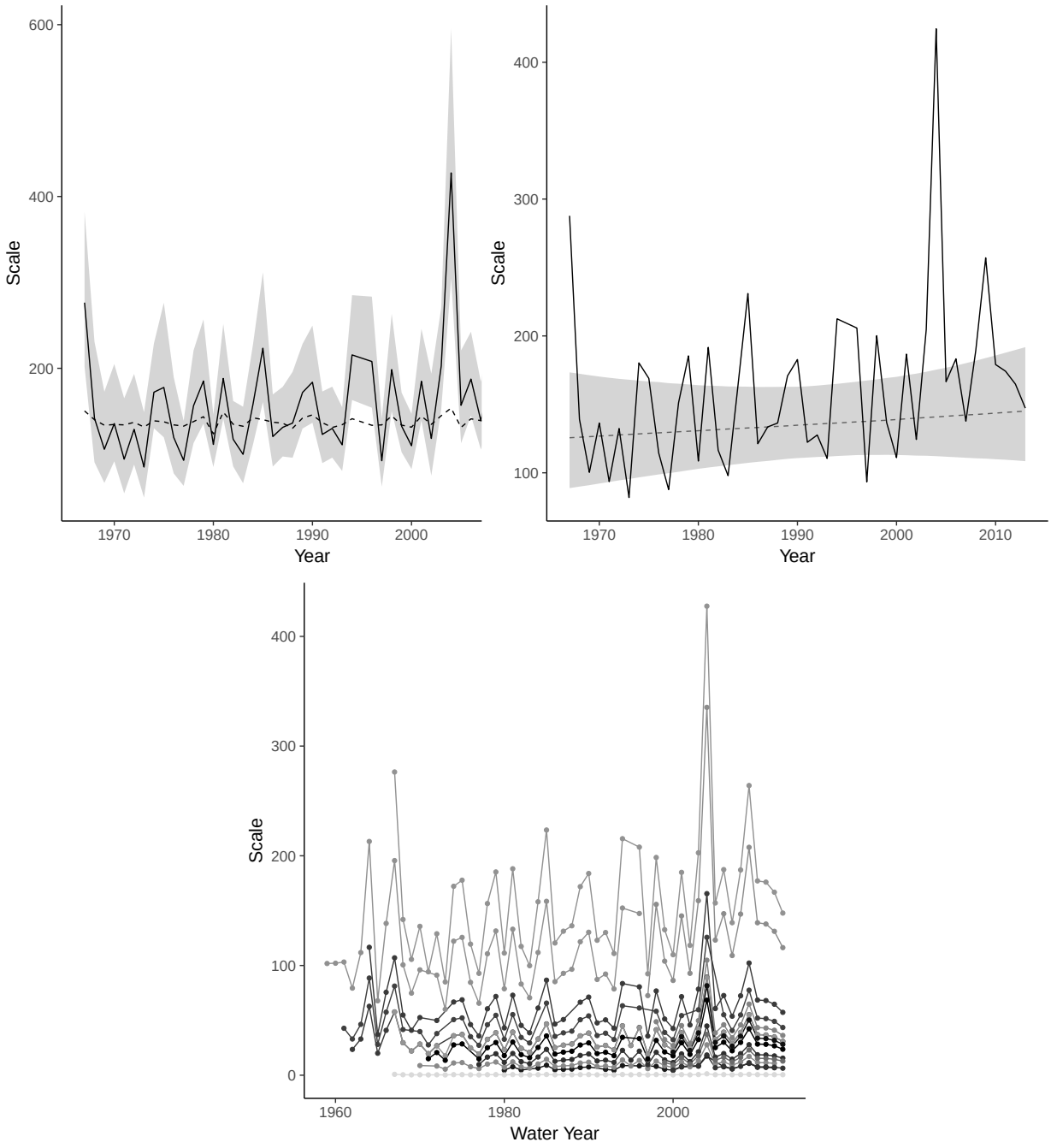

Figure 3: Top: scale parameters for the River Eden at Sheepmount from the regional (full line) and separate-site (dashed line) random effects models (left), and the regional random effects (full line) and separate-site regression (dashed line) models (right). Shaded regions indicate 95% credibility intervals for the regional model (left) and the separate-site regression model (right). Bottom: Scale parameters $\psi_{s,j}$ as a function of water year for each of the sites in region 76 estimated using the GP regional random effects model. Shading indicates latitude, with more northerly sites given a lighter shading.

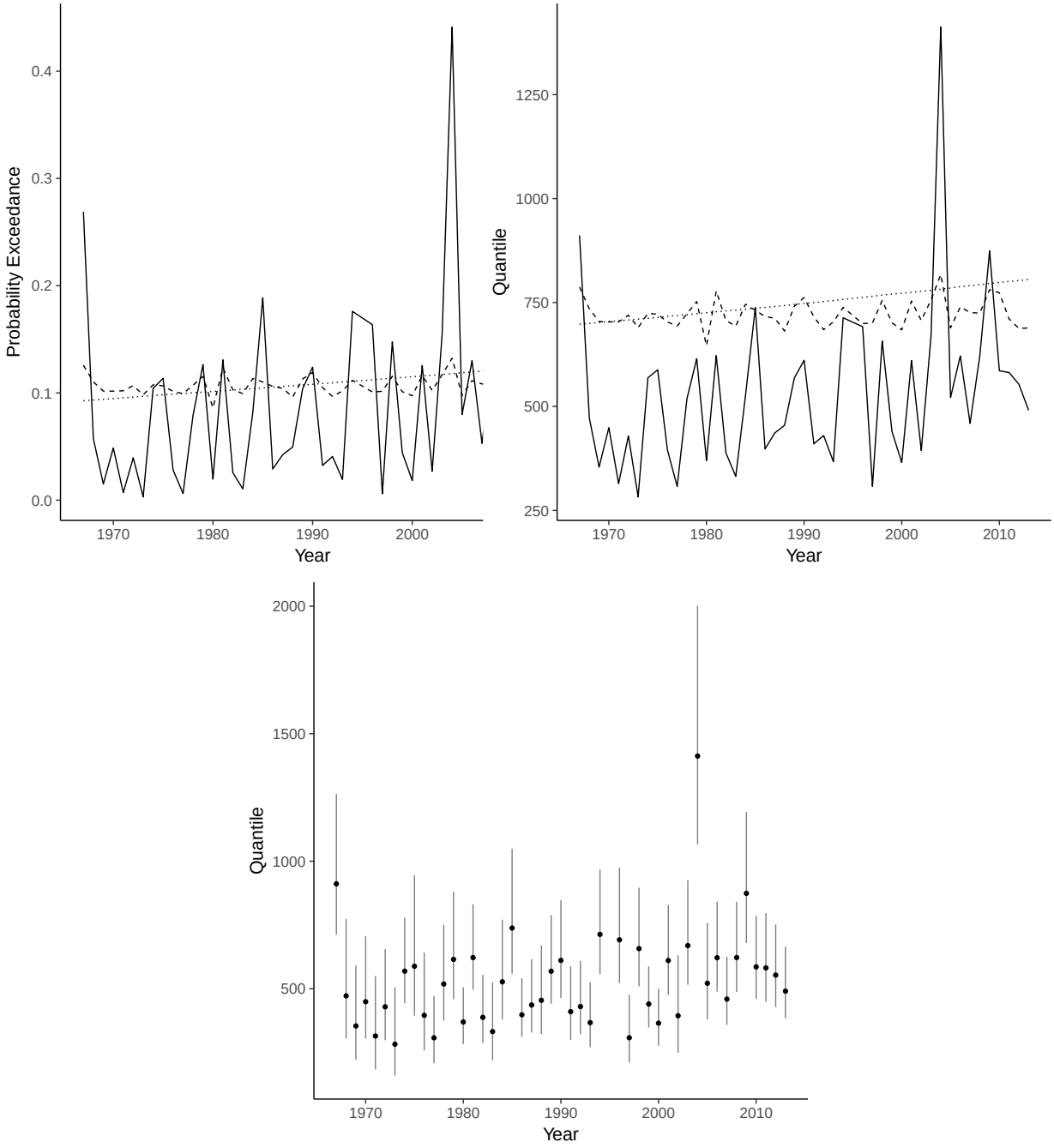

Figure 4: Model-based estimates for Sheepmount of the probability of a POT event exceeding $328.8 m^3 s^{-1}$, ie. the 90% quantile of the observed POT data (top left) and the 99% quantile of the POT event sizes (top right). Estimates from: separate-site time-trend (dotted line), separate-site random effects (dash-dot line) and regional random effects (full line) models. Bottom plot shows the quantile estimates from the regional random effects model with 95% credibility intervals.

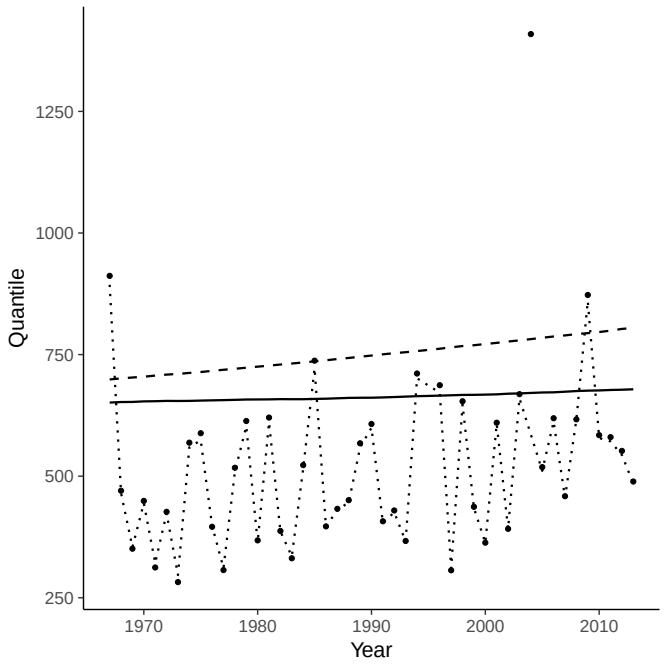

Figure 5: Estimated 99% quantiles for the POT sizes for models fitted with and without the 2004 water year. Separate-site regression models: with 2004 (dashed line) and without 2004 (full line). Regional random effects models: with 2004 (circles) and without 2004 (dotted line).

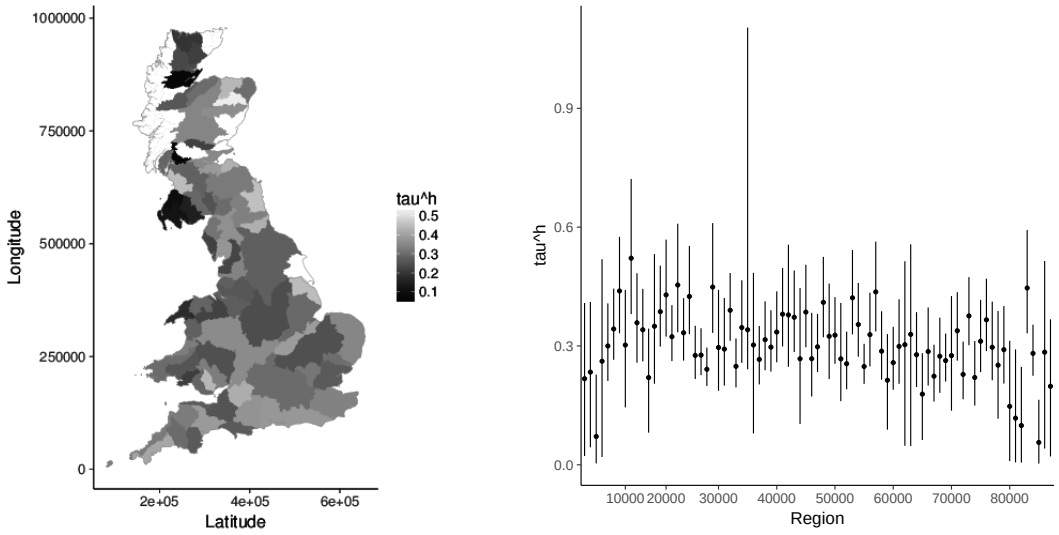

Figure 6: GP regional random effects parameter τ: Posterior medians (left); Posterior medians, points, and 95% credibility intervals, vertical lines, (right). Left-hand plot: regions coloured white had fewer than 2 stations and were excluded from the analysis.

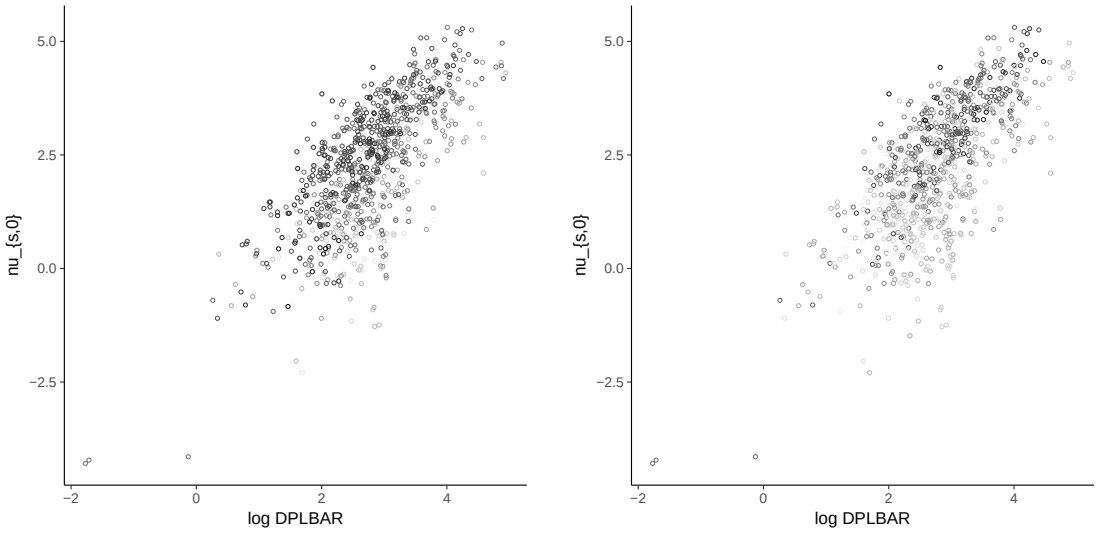

Figure 7: GP site specific scale intercept $\nu_{s,0}$ plotted against the logarithm ($\ln$) of DPLBAR, a measure of catchment size. Points shaded according to Easting (left) and Northing (right); darker points are further West and North respectively.

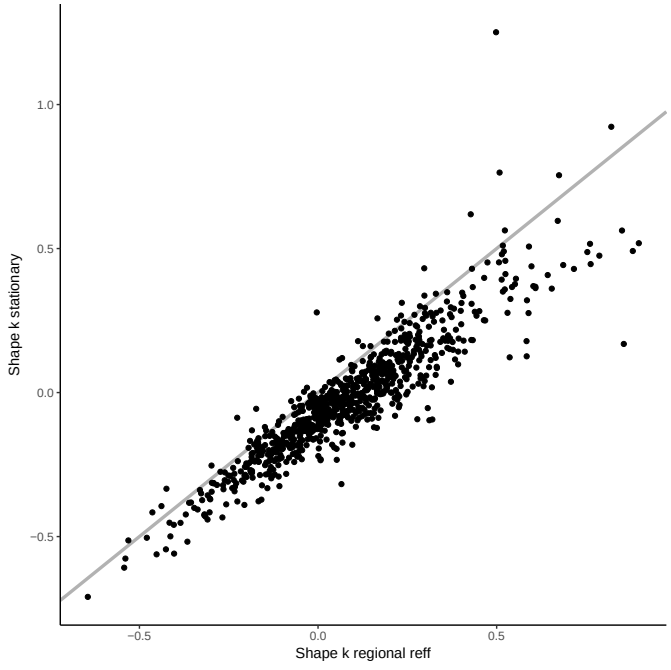

Figure 8: GP shape parameter for the regional random effects and separate site stationary models. Gray line shows the line $y = x$.

Figure 9: Regional random effects for 1967 (top left), 2000 (top right) and 2006 (bottom). Darker (lighter) regions have larger negative (positive) random effects. White regions had no observations that year.

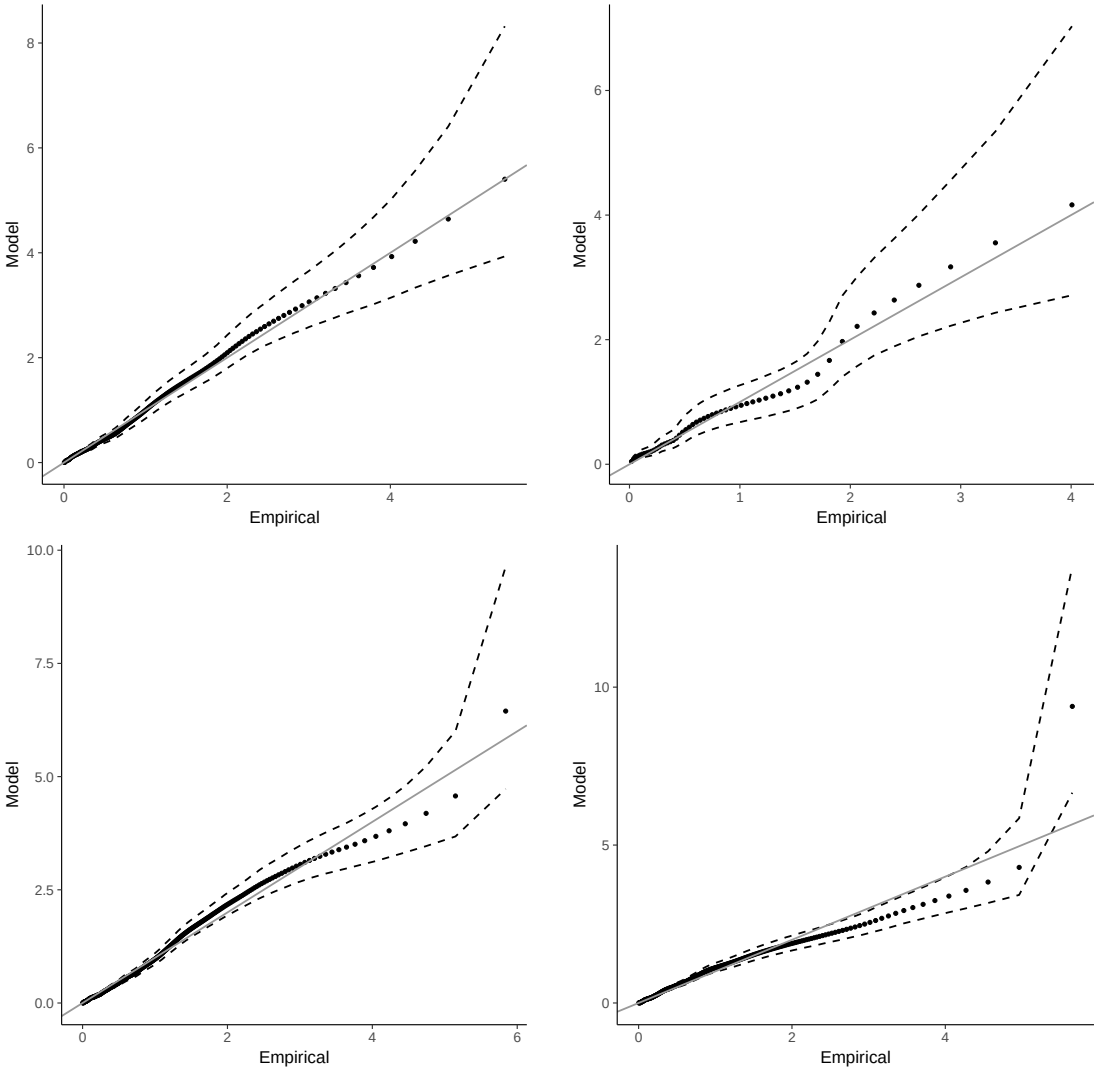

Figure 10: QQ plots for the regional random effects model for four stations. Clockwise from top left: R. Eden at Sheepmount (NW England), R. Camel at Camelford (SW England), R. Don at Doncaster (E England) and R. Kelvin at Killermont (SW Scotland). Dashed lines show 95% credibility intervals and gray line shows $y = x$.

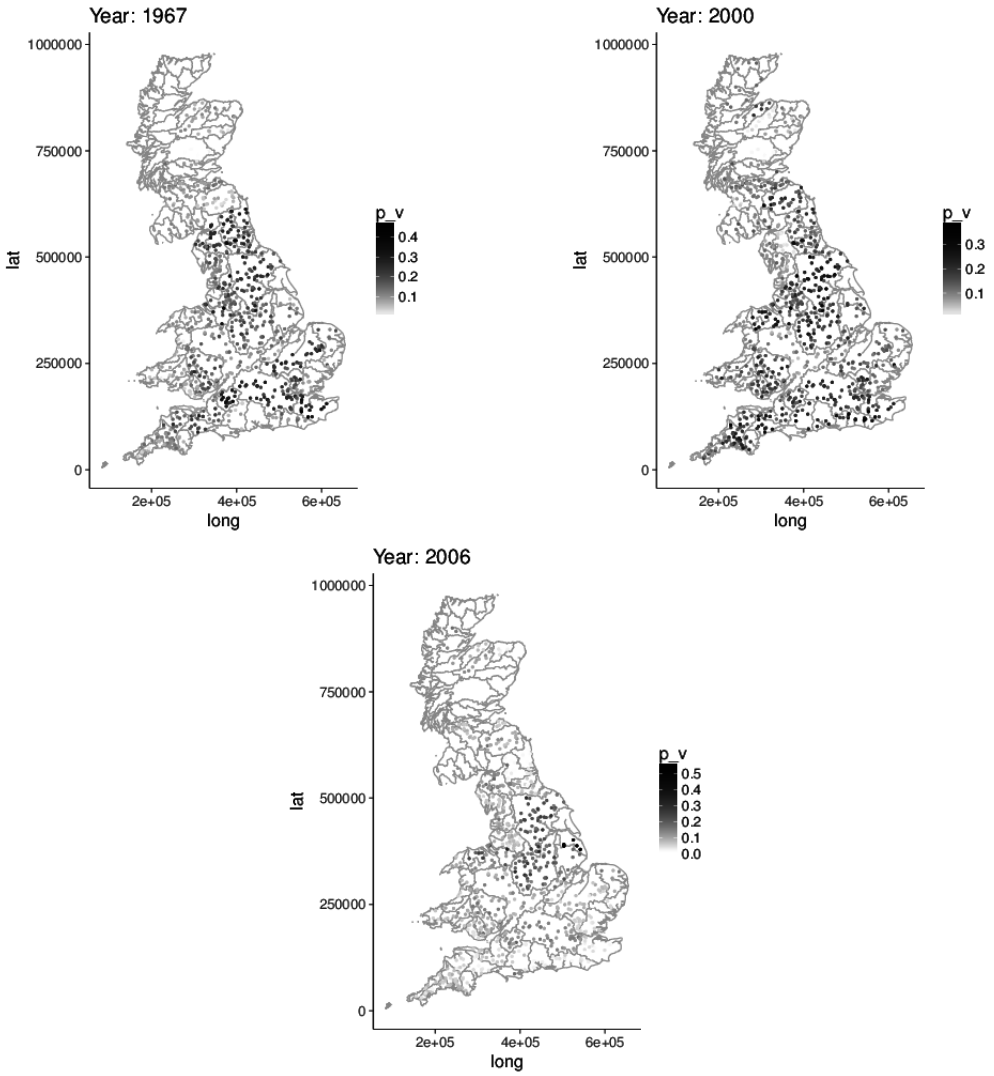

Figure 11: Conditional probabilities p_v of exceeding the site-specific empirical 90% quantile v in 1967 (top left), 2000 (top right) and 2006 (bottom). Darker shading indicates a higher probability. In all cases, probabilities are conditional on exceeding the POT modelling threshold and the 99% quantile is calculated using all POT data at the site assuming stationarity.

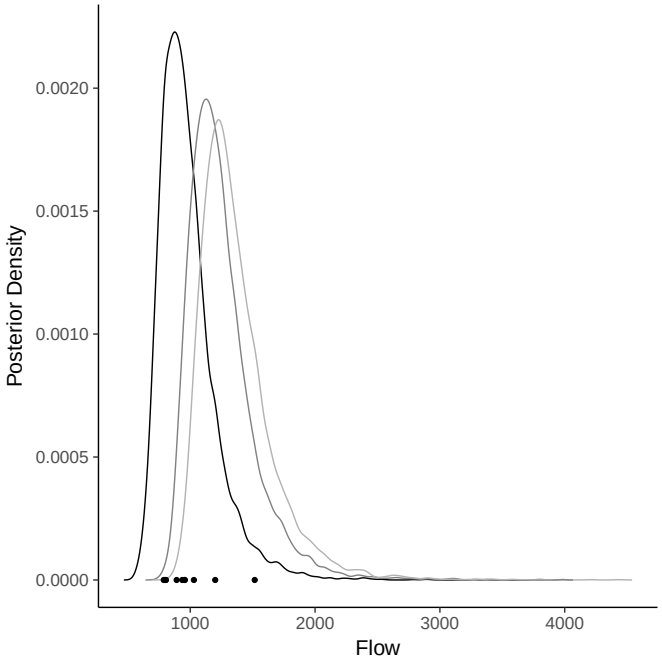

Figure 12: Posterior distribution of the 10- (black), 50- (dark gray) and 100-year (light gray) maxima for the River Eden at Sheepmount. Points show the largest 10 events over the observation period (1967–2013, with one missing year).

References

Andrieu, C. and Thoms, J. (2008). A tutorial on adaptive MCMC. *Statistics and Computing*, 18(4):343–373.

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.

Brooks, S. P. and Roberts, G. O. (1998). Convergence assessment techniques for Markov Chain Monte Carlo. *Statistics and Computing*, 8(4):319–335.

Carroll, B., Balogh, R., Morbey, H., and Araoz, G. (2010). Health and social impacts of a flood disaster: responding to needs and implications for practice. *Disasters*, 34(4):1045–1063.

Carroll, B., Morbey, H., Balogh, R., and Araoz, G. (2009). Flooded homes, broken bonds, the meaning of home, psychological processes and their impact on psychological health in a disaster. *Health & Place*, 15(2):540–547.

Castellanos, M. E. and Cabras, S. (2007). A default bayesian procedure for the generalized Pareto distribution. *Journal of Statistical Planning and Inference*, 137(2):473–483.

Castillo, E. and Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, 92(440):1609–1620.

Chavez-Demoulin, V. and Davison, A. (2012). Modelling time series extremes. *REVSTAT-Statistical Journal*, 10(EPFL-ARTICLE-180506):109–133.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1):207–222.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335.

Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data. *Applied Statistics*, pages 463–478.

Cooley, D., Cisewski, J., Erhardt, R. J., Jeon, S., Mannshardt, E., Omolo, B. O., and Sun, Y. (2012). A survey of spatial extremes: measuring spatial dependence and modeling spatial effects. *Revstat*, 10(1):135–165.

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

Cooley, D. and Sain, S. R. (2010). Spatial hierarchical modeling of precipitation extremes from a regional climate model. *Journal of Agricultural, Biological, and Environmental statistics*, 15(3):381–402.

Cox, D., Isham, V., and Northrop, P. (2002). Floods: some probabilistic and statistical approaches. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 360(1796):1389–1408.

Crichton, M. T., Ramsay, C. G., and Kelly, T. (2009). Enhancing organizational resilience through emergency planning: learnings from cross-sectoral lessons. *Journal of Contingencies and Crisis Management*, 17(1):24–37.

Davison, A. C. (1984). Modelling excesses over high thresholds, with an application. In *Statistical Extremes and Applications*, pages 461–482. Springer.

Davison, A. C., Padoan, S. A., Ribatet, M., et al. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

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

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O. (2000). Climate extremes: observations, modeling, and impacts. *Science*, 289(5487):2068–2074.

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.

Eastoe, E. F. and Tawn, J. A. (2010). Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. *Water Resources Research*, 46(2). W02510.

Fowler, H. and Kilsby, C. (2003). A regional frequency analysis of United Kingdom extreme rainfall from 1961 to 2000. *International Journal of Climatology*, 23(11):1313–1334.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). *Markov Chain Monte Carlo in practice*. CRC press.

Haario, H., Saksman, E., Tamminen, J., et al. (2001). An adaptive Metropolis algorithm. *Bernoulli*, 7(2):223–242.

Hall, P. and Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Statistical Science*, pages 153–167.

Horritt, M. S., Bates, P. D., Fewtrell, T. J., Mason, D. C., and Wilson, M. D. (2010). Modelling the hydraulics of the Carlisle 2005 flood event. In *Proceedings of the Institution of Civil Engineers-Water Management*, volume 163, pages 273–281. Thomas Telford Ltd.

Hosking, J. and Wallis, J. R. (1997). *Regional Frequency Analysis*. Cambridge University Press, April 1997.

Hosking, J. R. and Wallis, J. R. (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3):339–349.

Huerta, G. and Sansó, B. (2007). Time-varying models for extreme values. *Environmental and Ecological Statistics*, 14(3):285–299.

Katz, R. W. (2010). Statistics of extremes in climate change. *Climatic Change*, 100(1):71–76.

Kyselỳ, J., Picek, J., and Beranová, R. (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Global and Planetary Change*, 72(1-2):55–68.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, pages 963–974.

Lang, M. (1999). Theoretical discussion and monte-carlo simulations for a negative binomial process paradox. *Stochastic environmental research and risk assessment*, 13(3):183–200.

Mackay, E. B., Challenor, P. G., and Bahaj, A. S. (2011). A comparison of estimators for the generalised Pareto distribution. *Ocean Engineering*, 38(11-12):1338–1346.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A. (2010). Review article 'assessment of economic flood damage'. *Natural Hazards and Earth System Sciences*, 10(8):1697.

Milly, P. C. D., Wetherald, R. T., Dunne, K., and Delworth, T. L. (2002). Increasing risk of great floods in a changing climate. *Nature*, 415(6871):514.

Neal, J. C., Bates, P. D., Fewtrell, T. J., Hunter, N. M., Wilson, M. D., and Horritt, M. S. (2009). Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. *Journal of Hydrology*, 368(1-4):42–55.

Northrop, P. J. and Jonathan, P. (2011). Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. *Environmetrics*, 22(7):799–809.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131.

Reynard, N. S., Prudhomme, C., and Crooks, S. M. (2001). The flood characteristics of large UK rivers: potential effects of changing climate and land use. *Climatic Change*, 48(2-3):343–359.

Ribatet, M., Cooley, D., and Davison, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statistica Sinica*, pages 813–845.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Roberts, G. O., Rosenthal, J. S., et al. (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367.

Roberts, N. M., Cole, S. J., Forbes, R. M., Moore, R. J., and Boswell, D. (2009). Use of high-resolution nwp rainfall and river flow forecasts for advance warning of the Carlisle flood, north-west England. *Meteorological Applications*, 16(1):23–34.

Salas, J. D. and Obeysekera, J. (2013). Revisiting the concepts of return period and risk for nonstationary hydrologic extreme events. *Journal of Hydrologic Engineering*, 19(3):554–568.

Sang, H. and Gelfand, A. E. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental statistics*, 15(1):49–65.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Serinaldi, F. and Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77:17–36.

Sharkey, P. and Winter, H. C. (2017). A Bayesian spatial hierarchical model for extreme precipitation in Great Britain. *ArXiv e-prints*.

Sigauke, C. and Bere, A. (2017). Modelling non-stationary time series using a peaks over threshold distribution with time varying covariates and threshold: An application to peak electricity demand. *Energy*, 119:152–166.

Smith, R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, pages 367–377.

Turkman, K. F., Turkman, M. A., and Pereira, J. (2010). Asymptotic models and inference for extremes of spatio-temporal data. *Extremes*, 13(4):375–397.

Yee, T. W. and Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes*, 10(1-2):1–19.

Zhang, J. (2007). Likelihood moment estimation for the generalized Pareto distribution. *Australian & New Zealand Journal of Statistics*, 49(1):69–77.

Zhang, J. and Stephens, M. A. (2009). A new and efficient estimation method for the generalized Pareto distribution. *Technometrics*, 51(3):316–325.