WATER RESOURCES RESEARCH, VOL. 46, W02510, doi:10.1029/2009WR007757, 2010

Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series

Emma F. Eastoe[1] and Jonathan A. Tawn[1]

Received 20 January 2009; revised 24 August 2009; accepted 21 September 2009; published 12 February 2010.

[1] In a peaks over threshold analysis of a series of river flows, a sufficiently high threshold is used to extract the peaks of independent flood events. This paper reviews existing, and proposes new, statistical models for both the annual counts of such events and the process of event peak times. The most common existing model for the process of event times is a homogeneous Poisson process. This model is motivated by asymptotic theory. However, empirical evidence suggests that it is not the most appropriate model, since it implies that the mean and variance of the annual counts are the same, whereas the counts appear to be overdispersed, i.e., have a larger variance than mean. This paper describes how the homogeneous Poisson process can be extended to incorporate time variation in the rate at which events occur and so help to account for overdispersion in annual counts through the use of regression and mixed models. The implications of these new models on the implied probability distribution of the annual maxima are also discussed. The models are illustrated using a historical flow series from the River Thames at Kingston.

Citation: Eastoe, E. F., and J. A. Tawn (2010), Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series, *Water Resour. Res.*, *46*, W02510, doi:10.1029/2009WR007757.

1. Introduction

[2] Given an observed flow series, several questions naturally arise regarding the extremal aspects of the series. For example, "how frequently do flood events occur?" and "what, for large n, is the n-year return level of the annual maxima?". The answers to such questions are of interest to, amongst others, hydrologists, engineers, planners and insurance companies. Recent attention has also focused on the issues of whether the frequency (and size) of flood events is changing over time, perhaps as a consequence of climate change [*Reynard et al.*, 2001; *Mudelsee et al.*, 2003; *Prudhomme et al.*, 2003] and how to pool data across multiple sites to improve the estimation of flood event characteristics [*Cunderlik and Burn*, 2003; *Ribatet et al.*, 2007].

[3] The most common statistical approach to modeling event times and sizes is the peaks over threshold (POT) analysis. *Davison and Smith* [1990] discuss POT from a statistical perspective, while *Lang et al.* [1999] provide a good overview of the application of POT to river flow data and the particular practical challenges posed by such data. The POT analysis requires a full daily, or hourly, series from which a high threshold is used to identify independent flood events. Typically flood events, which contain multiple highly dependent threshold exceedances, are characterized by the size of their peak (above the threshold) and the time at which the peak occurs [*Lang et al.*, 1999; *Robson and Reed*, 1999]. This is a theoretically justified approach [*Leadbetter*, 1991]. We assume throughout that events can be identified; that is, for a given series and a high threshold u the times and peaks of the flooding events are known.

[4] Under the assumption of stationarity, *Leadbetter et al.* [1983] use asymptotic results to suggest that the process of event times is modeled by a one-parameter homogeneous Poisson process and the sizes of peaks Y by the two-parameter generalized Pareto (GP) distribution. The GP distribution has the following conditional distribution; for $y > u$,

$$\Pr[Y \leq y | Y > u] = \begin{cases} 1 - \left[1 + \xi\left(\dfrac{y-u}{\psi}\right)\right]_+^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\dfrac{y-u}{\psi}\right) & \text{if } \xi = 0, \end{cases}$$

where ψ and ξ are scale and shape parameters, respectively, and $z_+ = \max(z, 0)$. This model implies that the annual maxima are independent and identically distributed (IID) from year to year, and follow the generalized extreme value (GEV) distribution [*Davison and Smith*, 1990; *Engeland et al.*, 2004]. In this paper, we will work with water rather than calendar years. Thus, a year runs from October to September.

[5] A consequence of the homogeneous Poisson process is that the number of events N in a year follow a Poisson distribution and so has index of dispersion, defined by

$$D = \frac{\mathrm{Var}(N)}{\mathbb{E}[N]},$$

equal to 1. Furthermore, event counts in different years are IID. Due to the Poisson process model, these features hold regardless of the time period on which the count N is made.

[6] For many rivers, the number of flood events per year are known to be overdispersed, i.e., $D > 1$ [*Lang*, 1999]. This

[1]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK.

0043-1397/10/2009WR007757$09.00

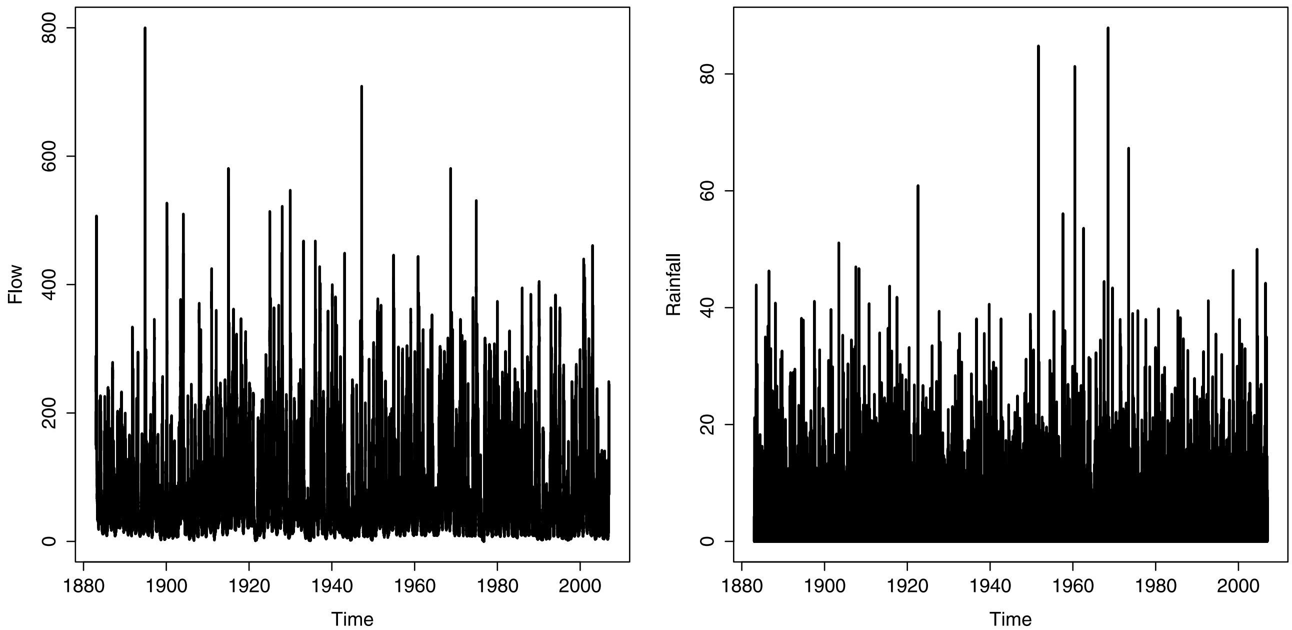

Figure 1. (left) Daily mean flow (m^3 s^{-1}) of River Thames at Kingston and (right) total daily rainfall (mm) measured in Oxford. Both data sets are for the period 1 January 1883 until 31 December 2006.

suggests that the homogeneous Poisson process model for event times might not be entirely appropriate. In this case, *Lang* [1999, and references therein] retain the IID assumption for *N*, but replace the Poisson distribution with the negative binomial distribution which exhibits overdispersion. Such a model is also attractive since, a special case of it, when combined with the GP model for the size of peaks, implies a generalized logistic (GL) distribution for the annual maximum flow, which is the preferred distribution used to model UK annual maxima flow series [*Robson and Reed*, 1999]. However, the assumption that the annual maxima are IID over different years is retained.

[7] The three main problems with using the negative binomial model for the number of events in a year are that it does not have the theoretical justification of the Poisson model, it is not as easily extended to a model for the full process of events and it assumes that the number of events and annual maxima are IID from year to year. In this paper, we discuss a new model for the full process which is motivated by an alternative formulation of the negative binomial model. This full process model retains the asymptotically motivated Poisson process, but assumes that the process is inhomogeneous with a time varying rate parameter. It allows overdispersion in the counts *N*, regardless of the time period over which the count is made and no longer requires that the counts and annual maxima be IID. The inhomogeneous Poisson process model is more realistic than the homogeneous model since it is well known that both flow series themselves and the processes which cause flood events, such as heavy rainfall and soil moisture, are time varying.

[8] In the inhomogeneous Poisson process model, we explain time variation in the rate parameter using a regression component and a random component. The former is useful if some (or all) of the necessary explanatory variables (covariates) are available and the latter if some (or all) of these required covariates are unavailable. Such a model reflects the underlying physical processes more realistically, so should improve the model fit and thus increase our confidence in the return levels estimated by subsequent extrapolation. Further, the inclusion of any available observed covariates allows an assessment to be made of how the frequency of events varies not just in time, but also in response to the covariates. The inclusion of regression and random components does increase model complexity, but we believe that this is outweighed by the benefits described.

[9] The data are introduced in section 2. Statistical models for the annual counts of flood events are discussed in section 3 and for the full process of events in section 4. In both cases, the implication on the distribution for the annual maximum is discussed. Methods of statistical inference for the count and process models are presented in section 5. To illustrate the methods, a full model fit of the data is given in section 6. Conclusions and suggestions for further extensions to the model, including allowing the GP parameters to vary in time, are given in section 7.

[10] Many of the models to be discussed use statistical methods which are commonly used in other applications by the statistical community; however, some of them are novel to either (or both of) the analysis of extreme events and/or flood frequency modeling. We try to indicate where this is the case.

2. Exploratory Analysis of Case Study Data

[11] The case study is on the River Thames at Kingston, since a long historical data set of the daily mean flow (m^3 s^{-1}) is available at this site (1 January 1883 to 31 December 2007). This data set, shown in Figure 1, was obtained from the UK National River Flow Archive at http://192.171.153.213/data/nrfa/index.html. Event identification was carried out using the following algorithm. A high threshold for the

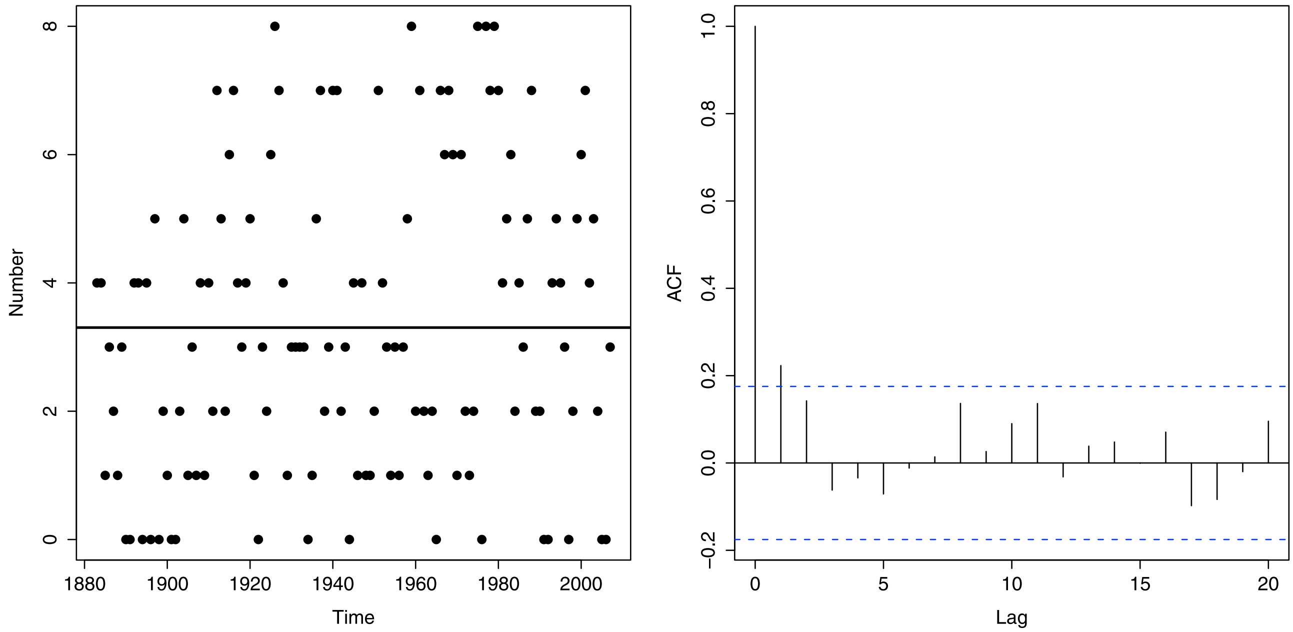

Figure 2. (left) Observed number of flow events per year for Kingston data and (right) autocorrelation function for these counts. The sample mean of 3.3 is shown by a horizontal dashed line on the observed flows, whereas the dashed lines on the autocorrelation plot show approximate 95% confidence intervals for the null hypothesis that $\rho = 0$.

flow data was first specified and threshold exceedances extracted. Any exceedances separated by more than m days of consecutive nonexceedances were assumed to come from independent events. Conversely, any exceedances separated by at most m consecutive days of nonexceedances were assumed to come from the same events. The times of events are defined as the day on which the peak flow in the event occurred. For the Kingston data, following this procedure with $m = 2$ days and a threshold of 200 $m^3 s^{-1}$ produced an average of 3.3 events per year. This follows standard practice in the analysis of river flow series in which typically thresholds with an average of either 1, 3 or 10 exceedances are used [*Robson and Reed*, 1999].

[12] The variance in the number of events per year is 5.7, which implies that the number of events are indeed overdispersed with $D = 1.71$. A number of parametric tests exist for testing the null hypothesis that $D = 1$ [*Karlis and Xekalaki*, 2000; *Lang et al.*, 1999]. Instead, we considered a Monte Carlo experiment in which a large number of homogeneous Poisson processes, of the same length of the observed process and also having a mean number of 3.3 events per year, were simulated. A sampling distribution of the index of dispersion was thus derived under the null hypothesis of $D = 1$. From this sampling distribution, a 95% tolerance interval for $D = 1$ was found to be (0.757, 1.243), suggesting that the observed data are inconsistent with a homogeneous Poisson process.

[13] Figure 2 shows the number of events observed in each of the years and the associated autocorrelation function of these counts. These suggest that there may be some year-to-year dependence in the number of events, particularly since runs of consecutive years have similar numbers of events. This is inconsistent with both the homogeneous Poisson process and IID negative binomial models. Further, since the events themselves are assumed to be independent, we do not believe that the numbers of events should be dependent from year to year. Instead, we suggest that such dependence is induced by the dependence structures of the underlying covariates, rather than being a genuine feature of the counts themselves.

[14] To try to ascertain whether or not this is the case, in addition to the flow data, we also have daily total rainfall (mm) at a site in Oxford. These data, also shown in Figure 1, will be used as a covariate in the regression component of our models. In this context, catchment averaged rainfall would have been a preferable covariate; however, it was unavailable for the time period required. The rainfall data were extracted from the UK Meteorological Office MIDAS Land Surface Observation data set through the British Atmospheric Data Centre; see http://badc.nerc.ac.uk/data/ukmo-midas/ for details.

3. Existing Models for Annual Counts

[15] In this section, we review two models for the number of events per year and show that each of these models, when combined with the GP distribution model for the peak event sizes, can be used to imply a model for the annual maximum flow. While these results are already well established [e.g., *Lang*, 1999], a brief review serves to introduce notation and motivate the models introduced in the following section.

[16] Let the random variables N_i and M_i represent the number of events and the annual maximum flow in year i, respectively. Both models in this section involve the assumption that the counts N_i are an IID sample from a probability distribution and that the event peaks are an IID sample from a GP distribution. These models imply that the annual maxima M_i are an IID sample.

Table 1. Expectations, Variances, Index of Dispersion, and Implied Distribution for the Annual Maximum M_i for Poisson and Negative Binomial Models for Annual Counts

Model for Counts	$\mathbb{E}[N_i]$	$\mathrm{Var}(N_i)$	D	M_i
Poisson	λ	λ	1	GEV
Negative binomial	λ	$\lambda(1+\lambda\alpha)$	$1+\lambda\alpha$	Extended GL

[17] In the first model, the counts N_i are assumed to be an IID sample from a Poisson distribution with mean $\lambda > 0$. The Poisson distribution has a probability mass function

$$\Pr[N_i = n] = \frac{\lambda^n \exp\{-\lambda\}}{n!}, n = 0, 1, 2, \dots. \quad (1)$$

The expectation, variance, and index of dispersion for N_i under this model are given in Table 1. A consequence of the Poisson and GP models is that the annual maximum has a GEV distribution [*Davison and Smith*, 1990; *Engeland et al.*, 2004], for $x > u$,

$$G(x) = \Pr[M_i \le x] = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{x-a}{b}\right)\right]^{-1/\xi}\right\} & \text{if } \xi \ne 0, \\ \exp\left\{-\exp\left[-\left(\frac{x-a}{b}\right)\right]\right\} & \text{if } \xi = 0, \end{cases} \quad (2)$$

where $a = u + \psi(\lambda^\xi - 1)/\xi$, $b = \psi\lambda^\xi$ and, for the case $\xi = 0$, using the limits $a \to \psi\log\lambda$ and $b \to \psi$ (as $\xi \to 0$). See section A1 for details.

[18] For the second model, the N_i are assumed to be an IID sample from the two-parameter negative binomial distribution. The negative binomial distribution is often derived from a Poisson distribution with a rate λ_i, which is a random variable. In particular, the following two models for N_i are equivalent.

$$N_i \sim \text{Poisson}(\lambda_i), \lambda_i = \lambda\gamma_i \text{ and } \gamma_i \sim \text{Gamma}(1/\alpha, 1/\alpha)$$

and

$$N_i \sim \text{Negative binomial}(1/\alpha, 1/(1+\lambda\alpha)), \quad (3)$$

where $\lambda > 0$ and $\alpha > 0$ [*Poortema*, 1999]. The parameter α accounts for the dispersion in the expected number of events from year to year. In the case in which $\alpha \to 0$, the variance of the γ_i tends to zero, so that all $\gamma_i \to 1$, and the model reduces to the Poisson model described in equation (1).

[19] Both formulations of this second model assume that the N_i are independent of year to year. In the first, this follows by assuming that the γ_i are independent from year to year. Here the gamma probability density function and negative binomial probability mass function are, respectively,

$$f(x) = \frac{(1/\alpha)^{1/\alpha}}{\Gamma(1/\alpha)} x^{-1/\alpha - 1} \exp\{-x/\alpha\}, x > 0 \quad (4)$$

and

$$\Pr[N_i = n] = \frac{\Gamma(n + 1/\alpha)}{n!\Gamma(1/\alpha)} \times \left(\frac{1}{1+\lambda\alpha}\right)^{1/\alpha}\left(1 - \frac{1}{1+\lambda\alpha}\right)^n, n = 0, 1, 2, \dots,$$

where $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}\,\mathrm{d}x$ is the gamma function. The expectation, mean, and index of dispersion of N_i under the negative binomial model are given in Table 1.

[20] As a consequence of the negative binomial model for N_i and the GP model for the event peaks, the annual maxima have the following distribution, for $x > u$.

$$H_r(x) = \begin{cases} p^r\left\{1-(1-p)\left(1-\left[1+\xi\left(\frac{x-u}{\psi}\right)\right]_+^{-1/\xi}\right)\right\}^{-r} & \text{if } \xi \ne 0, \\ p^r\left\{1-(1-p)\left(1-\exp\left(-\frac{x-u}{\psi}\right)\right)\right\}^{-r} & \text{if } \xi = 0, \end{cases} \quad (5)$$

where $p = 1/(1 + \lambda\alpha)$ and $r = 1/\alpha$ (see section A2 for details). This is a four-parameter extension of the three-parameter GL distribution, which occurs as a special case when $\alpha = 1$. Note that the negative binomial distribution with $\alpha = 1$ is referred to as the geometric distribution.

4. Models for Processes

4.1. Introduction

[21] The models in this section all attempt to explain how the rate at which events occur varies over time. All of these models are based on an inhomogeneous Poisson process, parameterized by a rate $\lambda(t) \ge 0$ which varies with time (and/or additional explanatory variables). An important consequence of the Poisson process model, which we will make use of and was hinted at in section 1, concerns the random variable $N(A)$ which represents the number of events in the time interval A. Assuming an inhomogeneous Poisson process the distribution of $N(A)$ is Poisson with mean $\int_A \lambda(t)\,\mathrm{d}t$. In our example, the time at which an event occurs is known only to the nearest day, so the integral is approximated by a sum over the rate parameter for the days in the interval A.

[22] A second consequence of the Poisson process model is that the integrated interevent times are IID with a standard exponential distribution. We will use this result for model diagnostics. Further theoretical details on Poisson processes are given by *Cox and Isham* [1980] and *Kingman* [1993].

[23] We also require the following notation. Let Y_{ij} represent the average flow on day j of year $i \in \{1, \dots, k\}$, where k is the number of years for which data are available. Define the indicator function δ_{ij} to denote the occurrence of an event peak on day j of year i, so that

$$\delta_{ij} = \begin{cases} 1 & \text{if there is an event peak on day } j \text{ in year } i, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

4.2. Regression Models

[24] The most natural model for the sequence $\{\delta_{ij}\}$ is as a sequence of Bernoulli variables, with the probability of an

event $\lambda_{ij} = \Pr[\delta_{ij} = 1]$ taken to be a function of a $p \times 1$ vector of covariates $\mathbf{x}_{ij}$. The covariates may themselves be time varying. Further, they may vary at different temporal scales. For example, some covariates may only vary between years, some between months, and others between days. The covariates may also include seasonal indicators or functions (e.g., seasonal sinusoids).

[25] Since by definition events are rare, within any year i, the probability of an event λ_{ij} will be close to zero for most j. Consequently, the expected number of events per year $\sum_{j=1}^{365}\lambda_{ij}$ will be small and so by the Poisson approximation to the Binomial distribution, an appropriate model for the number of events per year is $N_i \sim \text{Poisson}(\sum_{j=1}^{365}\lambda_{ij})$.

[26] For the same reason, we can approximate the sequence of Bernoulli trials by a Poisson process with a rate parameter λ_{ij} on day j in year i. This is the asymptotically motivated model for the rate of extreme events [*Davison and Smith*, 1990]. Although this model formerly implies that two or more events could occur per day, the probability of this is of order λ_{ij}^2 and is therefore very small. Then a standard model for λ_{ij} as a function of covariates is

$$\lambda_{ij} = g(\boldsymbol{\beta}'\mathbf{x}_{ij}), \tag{7}$$

where the function g is chosen to ensure that $\lambda_{ij} \geq 0$ and $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression coefficients. A popular choice of g is the exponential function, $g(x) = \exp(x)$; such a model is referred to as either as a Poisson regression model or as a generalized linear model with log-link function [*McCullagh and Nelder*, 1989].

[27] This model implies that the annual number of events is independent from year to year, but is not identically distributed with

$$N_i \sim \text{Poisson}(\lambda_i), \text{where } \lambda_i = \sum_{j=1}^{365}\lambda_{ij}. \tag{8}$$

From this model for the annual counts N_i, it follows that the annual maxima M_i are independent but nonidentically distributed, following a GEV distribution with location and scale parameters which change on a year-to-year basis. In particular, in equation (2), the parameters a and b become $a = u + \psi(\lambda_i^\xi - 1)/\xi$ and $b = \psi\lambda_i^\xi$, where λ_i is given by equation (8).

[28] A special case of the regression model described in equation (7) is when the covariates available only vary between years, i.e., $\mathbf{x}_{ij} = \mathbf{x}_i$ for all j in a given year i. If this is the case, then the covariate model alone cannot capture any within-year variation and $\lambda_{ij} = g(\boldsymbol{\beta}'\mathbf{x}_i)$ for all j in a given year i. Consequently, the distribution of the annual counts N_i remains the same as in equation (8), except that the mean parameter λ_i becomes $\lambda_i = \sum_{j=1}^{365}\lambda_{ij} = 365g(\boldsymbol{\beta}'\mathbf{x}_i)$.

4.3. Mixed Models

[29] It is often found that not all of the observed variability in the rate is accounted for by the regression model just described. One way to account for some of this extra variation is to include what is often referred to as a "random effect" term into the model for λ_{ij}. We illustrate such a model by including an *annual* random effect, although random effects could be included at any sensible temporal scale, e.g., decadal, seasonal or monthly scales.

[30] Under the annual random effects model, we have a Poisson process with a rate parameter which takes the form

$$\lambda_{ij} = \gamma_i g(\boldsymbol{\beta}'\mathbf{x}_{ij}),\ \gamma_i \sim \text{Gamma}(1/\alpha, 1/\alpha), \tag{9}$$

where the random variable γ_i is often referred to as a random effect. The random effects are taken to have the Gamma$(1/\alpha, 1/\alpha)$ distribution in analogous to the alternative representation of the negative binomial model in equation (4). The parameter α quantifies any extra annual variability in the rate which is not explained by the regression part of the model and so the γ_i can be interpreted as covariates, on the annual scale, which are unobserved. The simplest model is to take the γ_i to be independent from year to year. However, a more flexible model, in particular given their interpretation as covariates, is for the random effects to retain the Gamma$(1/\alpha, 1/\alpha)$ marginal distribution, but to introduce some year-to-year dependence. As a consequence of allowing dependence in the random effects, the distributions of both the annual counts N_i and the annual maxima M_i will be neither independent nor identically distributed.

[31] The mathematical details of one possible dependence model, in which the sequence of random effects are taken to have a first-order autoregressive structure, is given in section A3. This model requires only one additional parameter $-1 \leq \rho \leq 1$, which describes the association between consecutive random effects. If $\rho = 0$, the random effects are independent, whereas if $\rho > 0$ ($\rho < 0$) the random effects are positively (negatively) associated.

[32] The distributions of the annual counts N_i and annual maxima M_i implied by the mixed model are both dependent and nonidentically distributed with

$$N_i \sim \text{Negative binomial}\left(1/\alpha, 1/\left(1 + \alpha\sum_{j=1}^{365} g(\boldsymbol{\beta}'\mathbf{x}_{ij})\right)\right). \tag{10}$$

Consequently the annual maxima follow the extended GL distribution of equation (5) with $p = 1/(1 + \lambda\alpha)$ replaced by $p = 1/(1 + \alpha\sum_{j=1}^{365} g(\boldsymbol{\beta}'\mathbf{x}_{ij}))$. Dependence in the random effects also affects the distribution of the maximum of m-consecutive years $M_i^{(m)} = \max\{M_i, \ldots, M_{i+m-1}\}$, although we do not discuss this further.

[33] A special case of the mixed model occurs when no covariates are observed. In this case, the mixed model can only capture variation in the rate parameter at the same time scale as the random effects. For example, with annual random effects, we can only account for year-to-year variation in the rate. In fact, the negative binomial model for N_i discussed in section 3 is a consequence of the mixed model with no covariates and IID gamma distributed random effects.

4.4. Comments

[34] In the literature, there already exist examples of covariate models being used to model temporal variability in the rate of extreme flow events [*Katz et al.*, 2002; *Chavez-Demoulin and Davison*, 2005]. Motivated in part by the alternative formulation for the existing negative binomial model for overdispersion in annual counts of flood events, we have introduced the concept of a mixed model, in which the rate parameter is modeled as a function of both covariates and

random effects. Both the Poisson and negative binomial models for N_i can be derived as special cases of this model.

[35] The mixed model is an attempt to remedy the three main disadvantages of the negative binomial model which were discussed in section 1. The mixed model provides a consistent Poisson process framework in which to model the time varying rate, regardless of the time scale at which covariates are observed and even whether or not they *are* observed. It also has, as a special case, the existing negative binomial model. Even if no covariates are observed, temporal variation and dependence can be accounted for by including (dependent) random effects at the necessary temporal scales.

[36] A further advantage of the mixed model is the interpretation of the random effects, as approximating additional, but unobserved, covariates. For example, annual random effects can be thought of as unobserved covariates which only vary from year to year. As such, the estimated random effects can be used to help select suitable covariates for the model (by looking for covariates which display similar structure to the estimated random effects) and also to decide when sufficient covariates have been included in the model (when this is the case, the random effects will show little variability in time). This interpretation helped to motivate the use of dependent random effects, since we believe that the number of events from year to year should be independent (since the events themselves are), and therefore any dependence in the counts is due to dependence in the underlying physical processes which influence the rate at which extreme events occur.

5. Inference

5.1. Selection of Inference Method

[37] A range of inference methods are used for extreme value modeling: probability weighted moments (PWM), maximum likelihood (ML), and Bayesian methods. The PWM approach by *Greenwood et al.* [1979] is widely used when the variables are independent and identically distributed as it has good small sample size properties, but it cannot be used to fit models with random effects or covariates. ML and Bayesian methods both have the flexibility to fit regression models, but for complex models containing random effects, Bayesian methods tend to be most widely used in all other statistical applications. The fundamental reason for this is that Bayesian inference treats random effects as additional unknown parameters and so no difference in methodology is required by their inclusion and they are estimated simultaneously with the model parameters. In contrast, the ML approach treats the random effects as fixed unknowns.

[38] When estimating models containing random effects, ML requires two separate maximizations: one being the likelihood obtained by integrating out the random effects, which is maximized over the model parameters; the other is the likelihood of the random effects given the estimated model parameters. When closed form expressions from integrating out the random effects are not possible, numerical methods are required for the integration [*Breslow and Clayton*, 1993]. In particular, for our model with dependent random effects, a k-dimensional integral is required, which is computationally demanding and liable to be highly sensitive to the method of numerical integration that is used. However, even when closed form integration is possible, the ML estimates will not reflect the full uncertainty of the random effects. Therefore, for consistency of inference over the different models, a Bayesian approach is adopted throughout for parameter estimation, as the same approach can be used for all models and the uncertainty is accounted for appropriately.

5.2. Bayesian Inference

[39] In Bayesian inference, model parameters are assumed to be unknown and to have an underlying probability distribution (as opposed to being unknown and fixed as in ML). Bayes theorem is used to combine a prior probability distribution for these parameters with the likelihood function for the model (given the observed data) to obtain a posterior probability distribution for the parameters. The estimated posterior distribution is often summarized using the posterior median and $100(1 - \tau)$% posterior intervals in analogy with the ML estimate and confidence interval output of a ML fit.

[40] When the form of the statistical model is unknown, model selection is required. As all the models we consider are nested; i.e., a special case of one model leads to another model, a standard method of model selection is to use 95% posterior intervals to see if the estimated parameters are consistent with the simpler model structure. This is the approach we will take. However, formal methods of Bayesian model selection exist such as Bayes factors [*Kass and Raftery*, 1995; *Chib and Jeliazkov*, 2001, and *Sinharay and Stern*, 2005] and DIC [*Spiegelhalter et al.*, 2002], though neither is widely used due to computational problems and theoretical objections, respectively. More generally, Bayesian methods allow for the uncertainty of the model selection to be accounted for in the resulting inferences by treating a range of models as feasible and allowing their ability to fit the data to weight their contribution to the overall inference [*Green*, 1995].

[41] The main disadvantages of using the Bayesian approach are that it requires a prior distribution for the parameters; the posterior distribution may not have a closed form; and it may be hard to obtain the marginal posterior for a given parameter from the joint posterior for all parameters (particularly for high dimensional models). However there are a wide variety of tools available to simulate from the posterior distribution when it is not possible to sample from it directly. These are collectively known as Markov Chain Monte Carlo (MCMC) methods; two of the simplest being the Gibbs sampler and the Metropolis–Hastings random walk [*Gamerman and Lopes*, 2006]. Although not discussed further, a recent series of papers [*Rue and Martino*, 2007 and *Rue et al.*, 2009] suggest alternatives to MCMC. In addition, for a general introduction to Bayesian inference, see *Gilks et al.* [1996] and *Gelman et al.* [2004]. Examples of the use of Bayesian inference in a hydrological context are *Reis Jr. and Stedinger* [2005] and *Ribatet et al.* [2007].

[42] The Bayesian approach can also be used to obtain a predictive distribution for some future observation. The predictive distribution accounts for parameter uncertainty by averaging over the distribution of the posterior parameters (see *Coles and Tawn* [1996], for an application in extremes). For example, a predictive distribution could be obtained for N_i or M_i in future or unobserved years (this is not discussed further in this paper).

5.3. Priors

[43] As mentioned above, in order to undertake a Bayesian analysis, prior distributions must be defined for all the parameters of the model. All parameters in all models are assumed, a priori, to be independent of each other. We then use the following prior distributions.

$$\begin{aligned}\lambda &\sim \mathrm{Gamma}(a_1, a_2),\ a_1 > 0, a_2 > 0\\ \alpha &\sim \mathrm{Gamma}(b_1, b_2),\ b_1 > 0, b_2 > 0\\ \boldsymbol{\beta} &\sim \mathrm{MVN}(\boldsymbol{\beta}_0, \sigma I),\ \sigma > 0\\ r = (1+\rho)/2 &\sim \mathrm{Beta}(c_1, c_2),\ c_1 > 0, c_2 > 0.\end{aligned}$$

The parameters a_i, b_i, c_i (i = 1, 2), and σ are referred to as hyperparameters. Values of these are fixed prior to the analysis. The reparameterization of ρ to r was necessary to improve the convergence features of the MCMC. The remaining priors were chosen either because they were conjugate, or to ensure that parameters took values only on the correct range. Note that we do not need to specify priors for the random effects, since these are implied (where necessary) by the model.

5.4. Existing Models for Annual Counts

[44] We use the Poisson model for annual counts introduced in section 3 as a baseline model for comparison with the more sophisticated models. This model has a single parameter λ. Let $\mathbf{n} = (n_1, \ldots, n_k)$ be a vector containing the observed numbers of events in each of k years. Using the priors in section 5.3, the posterior distribution for λ is also gamma but with parameters $(a_1 + \sum_{j=1}^{k} n_i,\ a_2 + k)$.

[45] The negative binomial model (equivalently IID annual random effects with no covariates) has two parameters, α and λ. Using the priors in section 5.3 and integrating out the random effects, the joint posterior for (λ, α) is proportional to

$$p(\lambda, \alpha|\mathbf{n}, \mathbf{a}, \mathbf{b}) \propto \lambda^{\sum_{i=1}^{k} n_i + a_1 - 1} \exp\{-a_2\lambda\}\alpha^{b_1-1}\exp\{-b_2\alpha\} \cdot \prod_{i=1}^{k} \frac{\Gamma(n_i + 1/\alpha)}{(\lambda + 1/\alpha)^{n_i + 1/\alpha}},$$

where $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$. It is straightforward to show that, given the implicit Gamma($1/\alpha$, $1/\alpha$) prior, λ, α, and n_i, the random effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)$ are independent with posterior distributions Gamma($1/\alpha + n_i$, $\lambda + 1/\alpha$).

5.5. Process Models

[46] The regression model has p parameters (the regression coefficients). Using the prior for $\boldsymbol{\beta}$ defined in section 5.3, the posterior for $\boldsymbol{\beta}$ is given by

$$p(\beta|\boldsymbol{\delta}, \mathbf{x}, \boldsymbol{\beta}_0, \sigma) \propto \exp\left\{-\sum_{i=1}^{k}\sum_{j=1}^{365} g(\boldsymbol{\beta}'\mathbf{x}_{ij}) - \frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\} \cdot \prod_{i=1}^{k}\prod_{j=1}^{365} g(\boldsymbol{\beta}'\mathbf{x}_{ij})^{\delta_{ij}},$$

where $\boldsymbol{\delta} = (\delta_{1,1}, \delta_{1,2}, \ldots, \delta_{k,365})$ is the vector of indicators defined in equation (6) and $\mathbf{x}$ is the $365k \times p$ matrix of covariates in which row $365(i-1)+j$ corresponds to the covariates for day j in year i.

[47] The mixed model with IID random effects has $p + 1$ parameters, p regression coefficients, and the dispersion parameter α. The joint posterior for $(\boldsymbol{\beta}, \alpha)$ is found by integrating out the random effects to obtain

$$p(\boldsymbol{\beta}, \alpha|\boldsymbol{\delta}, \mathbf{x}, \boldsymbol{\beta}_0, \sigma, \mathbf{b}) \propto \frac{(1/\alpha)^{k/\alpha}}{\Gamma(1/\alpha)^k}\prod_{i=1}^{k}\left[\frac{\Gamma(n_i + 1/\alpha)}{\left(\sum_{j=1}^{365} g(\boldsymbol{\beta}'\mathbf{x}_{ij}) + 1/\alpha\right)^{n_i+1/\alpha}} \times \prod_{j=1}^{365} g(\boldsymbol{\beta}'\mathbf{x}_{ij})^{\delta_{ij}}\right] \times \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\}\alpha^{b_1-1}\exp\{-b_2\alpha\}.$$

The posterior distribution for the random effects again turns out to be gamma, i.e., $\gamma_i|\boldsymbol{\delta}_i, \mathbf{x}_i, \boldsymbol{\beta}, \alpha \sim$ Gamma($n_i + 1/\alpha$, $\sum_{j=1}^{365} g(\boldsymbol{\beta}'\mathbf{x}_{ij}) + 1/\alpha$).

[48] The mixed model with dependent annual random effects has $p + 2$ parameters having in addition the dependence parameter ρ. The random effects cannot be integrated out of this model, so the joint posterior for the parameters $\boldsymbol{\beta}$, α, and r (ρ) and the random effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)$ is given by

$$\begin{aligned}&p(\boldsymbol{\beta}, \alpha, r, \boldsymbol{\gamma}|\boldsymbol{\delta}, \mathbf{x}, \boldsymbol{\beta}_0, \sigma, \mathbf{b}, \mathbf{c})\\ &\propto \prod_{i=1}^{k}\left(\exp\left\{-\sum_{j=1}^{365}\gamma_i g(\boldsymbol{\beta}'\mathbf{x}_{ij})\right\}\prod_{j=1}^{365}\left[\gamma_i g(\boldsymbol{\beta}'\mathbf{x}_{ij})\right]^{\delta_{ij}}\right)\\ &\times \prod_{i=1}^{k} f(\gamma_i|\alpha)\prod_{i=2}^{k}\phi_2(z_{i-1}, z_i)\prod_{i=2}^{k-1}\phi(z_i)^{-2}\phi(z_1)^{-1}\phi(z_k)^{-1}\\ &\times \alpha^{b_1-1}e^{-b_2\alpha}r^{c_1-1}(1-r)^{c_2-1}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)'(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\},\end{aligned}$$

where f, ϕ, and ϕ_2 are probability density functions of the Gamma($1/\alpha$, $1/\alpha$), see equation (4), standard normal and bivariate normal distributions respectively, and $z_i = \Phi^{-1}(F(\gamma_i))$, where Φ and F are distribution functions associated with ϕ and f respectively. The bivariate normal distribution, in this case has mean $\mathbf{0}$, unit variances and correlation ρ.

[49] In all cases, including the independent random effects, a Metropolis–Hastings random walk algorithm is required to draw a sample from the joint posterior distribution for the parameters, via the conditional posterior distributions. Only the regression parameters could be updated using a block update.

6. Case Study Data Analysis

6.1. Introduction

[50] Since we saw in section 2 that there is significant evidence for overdispersion in the annual counts, we start with the negative binomial model for annual counts (Model 1). We then fit the mixed model with dependent random effects but no covariates (Model 2), the regression model (Model 3) and finally the mixed model with either independent (Model 4) or dependent random effects (Model 5). As an example of what we mean by nested models, Model 1 is nested inside Model 2.

[51] For each of these models we present a range of diagnostic summaries. We estimate parameters using the posterior median (PM) and parameter uncertainty using a

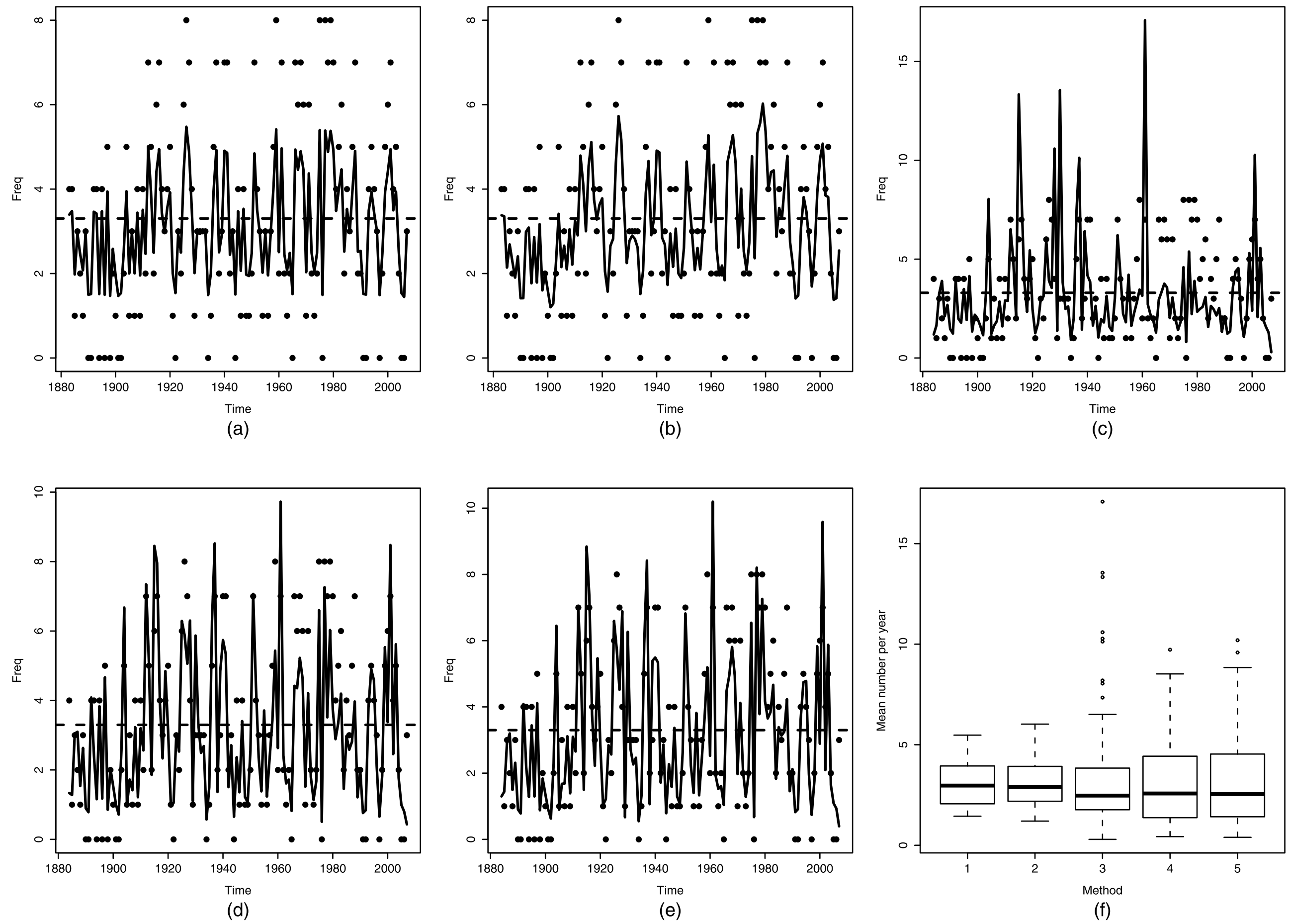

Figure 3. (a–e) Estimated expected annual counts (lines) for all five fitted models. Observed numbers of events are shown by dots, and the overall sample mean is shown by a horizontal dashed line. (f) Box plots of these estimated expected annual counts.

95% posterior interval. A plot of the PM estimate of the expected number of events in each observed years is provided as a goodness-of-fit diagnostic. These expected numbers are estimated, where appropriate, conditionally on the estimated random effects and the observed covariates. Goodness of fit for the regression and mixed models is also assessed though a quantile-quantile (QQ) plot of the posterior integrated intensity, with a good model fit indicated by points lying close to the 45° line. The basis for these QQ plots is that, as mentioned in section 4, the estimated intensities integrated between consecutive events should be a random sample from the exponential distribution with rate 1 [*Cox and Isham*, 1980].

6.2. Results

[52] Model 1 has two parameters, λ and α. It was assumed a priori that $\lambda \sim$ Gamma(1, 1) and $\alpha \sim$ Gamma(1, 1). The model fit did not appear to be sensitive to the use of alternative prior parameters. The PM's (95% posterior intervals) for the two parameters are $\lambda = 3.26$ (2.85, 3.75) and $\alpha = 0.300$ (0.153, 0.509). Note that α is significantly greater than 0 but significantly less than 1. The former result confirms that there is evidence of overdispersion in the data; hence, this model is preferable to assuming that the annual counts are IID Poisson. The latter result suggests that the counts do not follow a geometric distribution and therefore the GL distribution is inappropriate as a model for the annual maxima.

[53] A plot of the estimated mean number of events per year under this model is given in Figure 3a. This shows that the random effects shrink the estimated number of events in each year toward the overall across-years mean λ. The shrinkage effect is perhaps better illustrated by the box plot of the estimated mean numbers of events in a year shown in Figure 3f.

[54] Model 2 has three parameters, λ, α, and ρ. The prior distributions for λ and α were the same as for Model 1, and we took, a priori, $\frac{\rho+1}{2} \sim$ Beta(3, 3). The PM's (95% posterior intervals) for the parameters are λ = 3.24 (2.74, 3.85), α = 0.307 (0.171, 0.518), and $\rho = 0.376$ (0.0474, 0.657). Again α is significantly greater than 0 (evidence for overdispersion) and less than 1 (evidence against the geometric distribution). Also ρ is significantly greater than zero (evidence of a positive correlation between the random effects). This is consistent with the observation from Figure 2 that consecutive years have runs of similar numbers of events. Figures 3b and 3f show that allowing for dependence on the random effects results in slightly less shrinkage toward the overall mean annual count in the estimated annual counts, in comparison to Model 1.

[55] We now fit Models 3–5, each of which involves a regression component. We use baseflow (calculated using linear interpolation of local minima of the observed data, see *Gustard et al.* [1992]) and a moving average of the rainfall

Table 2. Estimates of Regression Coefficients for Models 3–5[a]

Covariate	Model 3	Model 4	Model 5
Intercept	−6.82 (−6.60, −7.06)	−7.25 (−6.96, −7.53)	−7.20 (−6.95, −7.41)
Three-month average rainfall	0.664 (0.496, 0.823)	0.793 (0.587, 0.984)	0.786 (0.593, 0.962)
Baseflow	0.0132 (0.0112, 0.0150)	0.0160 (0.0137, 0.0184)	0.0157 (0.0134, 0.0181)

[a]A log-link was used in all cases.

data introduced in section 2 (with a window of the last three months). In all models, the regression coefficients are assumed a priori to be independent with Normal(0, 1000) marginal distributions. Where appropriate, prior distributions for the random effects parameters α and ρ were taken to be the same as already discussed. A log-link function was applied throughout.

[56] For all three models, Table 2 shows that the estimates of the regression coefficients are very similar. In particular, in all models the effects of both covariates are significantly above zero at the 2.5% level. This confirms that the covariates are useful in explaining some of the variability in the rate at which flood events occur. The estimates suggest that increases in both baseflow and the average rainfall over the past three months cause increases in the rate of events.

[57] For Model 4, the estimate of the random effects parameter α is 0.229 (0.105, 0.389), whereas for Model 5 the random effect parameters α and ρ are estimated as α = 0.217 (0.109, 0.386) and ρ = 0.436 (−0.0682, 0.754). In both models, α is still significantly greater than 0, but less than 1. However, for Model 5, there is not enough evidence to suggest positive correlation between the random effects. This is probably because the significant correlation estimated in the random effects in Model 3 is now being accounted for by the covariates (rainfall and baseflow).

[58] Figures 3c–3e show plots of the estimated mean number of events in each year for Models 3–5. These estimates vary considerably more from year to year than the estimates produced under Models 1 and 2. The shrinkage effect of the random effects is reduced by the additional information in the covariates. This is further illustrated in Figure 3f, which shows box plots of the estimated mean number of events per year for all fitted models. Model 3 overestimates quite considerably the mean number of events, especially in those years in which the observed number of events was high.

[59] Figure 4 shows QQ plots of the interevent integrated intensities under each of the three process models. Models 4 and 5 appear to have a better fit than the simpler Model 3, since the points in the QQ plots lie almost exactly on a straight line under these two models. Further autocorrelation functions (not shown) of the interarrival times for these models showed them to be independent.

[60] Following these investigations, we suggest that Model 4 provides the best model for this data set. Finally, Figure 5 shows a plot of the estimated random effects under this best model. There appears to be some pattern to these estimates; in particular the random effects in the period 1960–1990 seem to behave differently to those in the rest of the study period. This is probably because we are missing a covariate which would explain the change in behavior between 1960 and 1990, although we are not sure what this would be.

7. Conclusions

[61] We have reviewed two types of model for annual counts of flood events, one based on the Poisson distribution and the other on the negative binomial distribution. It turns out that the both of these models are a consequence of the asymptotically motivated Poisson process model for the event times. The Poisson distribution is a consequence of assuming that the Poisson process is homogeneous. The negative binomial distribution is a consequence of a Poisson process in which the rate parameter is allowed to vary stochastically from year to year.

[62] A number of models have also been proposed for the process of event times. In order to incorporate the time variability in the rate parameter, these are all based on an inhomogeneous Poisson process. The simplest such model is a regression model, in which the (log of the) Poisson mean is taken to be a linear function of covariates. A more sophisticated family of models, referred to as mixed models, allows the inclusion of unobserved covariates. These are modeled as random variables and referred to as random effects. The time scale of these unobserved covariates must be specified. In this paper we have illustrated the methods using annual random effects, mostly because we are interested in obtaining improved estimates for the annual counts in order to estimate the distribution of the size of the annual maximum flood. In

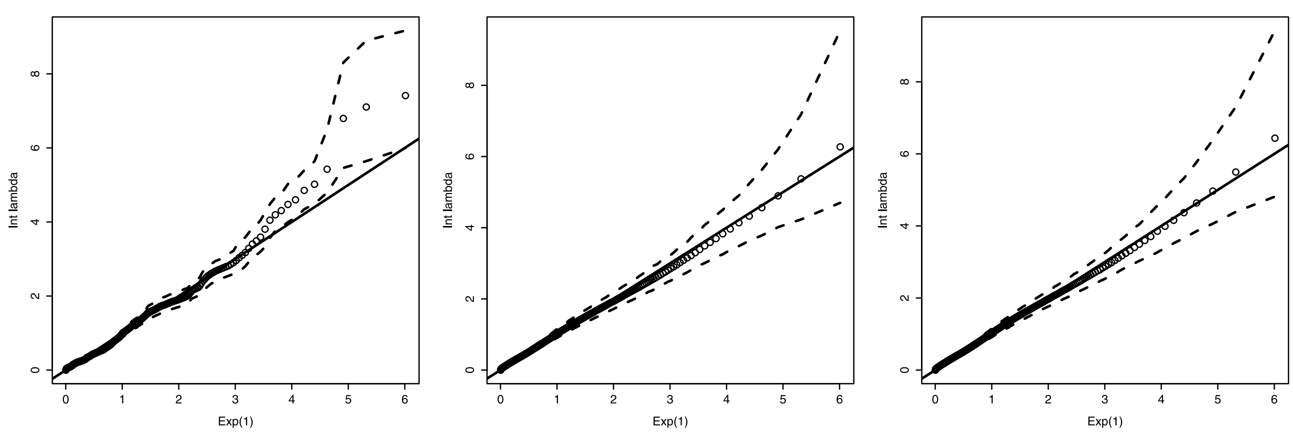

Figure 4. QQ plots for integrated intensities from Models (left) 3, (middle) 4, and (right) 5.

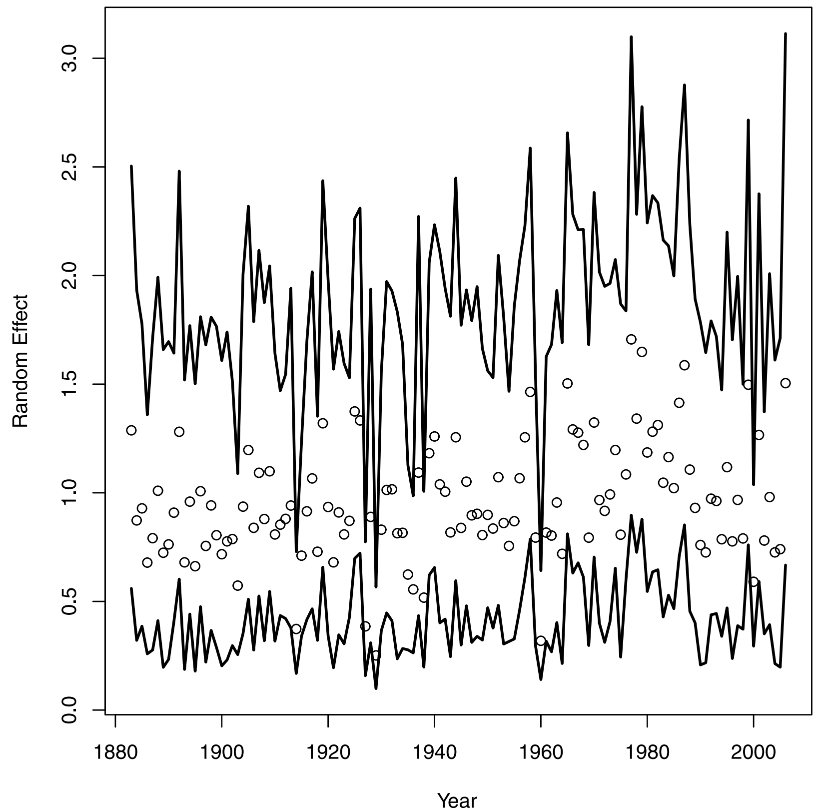

Figure 5. Plot of estimated random effects from Model 4. Circles show pointwise PMs and lines pointwise 95% posterior intervals.

theory, random effects can be included at multiple time scales, within the same model.

[63] The negative binomial model is a special case of the mixed model with IID random effects and no covariates. We have also suggested that it might be beneficial to allow year-to-year dependence in the random effects. This has a similar effect on the random effects as a nonparametric smoother, but has the benefit that it can be used for prediction.

[64] The various models have been illustrated by a case study using data from the River Thames at Kingston. Catchment averaged precipitation was unavailable for the duration of the series and so baseflow and a three-month average point rainfall were used as covariates. Parameter estimates, estimated annual counts, and QQ plots of integrated intensities were used to compare the fit of five models. Evidence was found of a significant positive relationship between both covariates and the rate at which flood events were observed to occur. From the estimates of α, there was evidence to support the inclusion of random effects and against the use of the geometric distribution for N_i. The estimates of ρ provided evidence of a first-order dependence structure in the random effects only when covariates were not included in the model. The mixed model with independent random effects was selected as the best model for the data, although it was noted that an extra covariate might be necessary to explain a change in the behavior of the estimate random effects after 1960.

[65] A natural extension of this work would be to use similar techniques to model the sizes of the event flow peaks. As noted in section 2 while the sizes of the event peaks appear to be independent, at least in the case study shown here, this does not imply that they are identically distributed. Additional information and predictive power could be gained by allowing random effects for size and rate of events to be dependent. This seems a reasonable step given that the rate and size of events are determined by similar covariates (for example, intensity and duration of rainfall and soil conditions).

Appendix A

A1. Derivation of the GEV Distribution

[66] The derivation of the implied distribution for the annual maximum under the Poisson model for annual counts of equation (1) is as follows. For $x > u$,

$$\begin{aligned}
G(x) &= \Pr[M_i \le x] \\
&= \sum_{n=0}^{\infty} \Pr[M_i \le x | N_i = n] \Pr[N_i = n] \\
&= \sum_{n=0}^{\infty} \left(\prod_{j=1}^{n} \Pr\left[Y_{ij} \le x | Y_{ij} > u\right] \right) \Pr[N_i = n] \\
&= \sum_{n=0}^{\infty} \Pr[Y_{i1} \le x | Y_{i1} > u]^n \Pr[N_i = n] \\
&= \sum_{n=0}^{\infty} \left(1 - \left[1 + \xi \left(\frac{x-u}{\psi} \right) \right]_+^{-1/\xi} \right)^n \frac{\exp\{-\lambda\}\lambda^n}{n!} \\
&= \begin{cases} \exp\left\{ - \left[1 + \xi \left(\dfrac{x - u - \psi(\lambda^\xi - 1)/\xi}{\psi \lambda^\xi} \right) \right]^{-1/\xi} \right\} & \text{if } \xi \ne 0, \\ \exp\left\{ - \exp\left[- \left(\dfrac{x - u - \psi \log \lambda}{\psi} \right) \right] \right\} & \text{if } \xi = 0. \end{cases}
\end{aligned}$$

The idea here is to average overall possible values for the annual number of events N_i. To get from the first to the second line, note that if the annual maximum M_i is below the level x, then all the event peaks Y_{ij} $j = 1, \ldots, n$ must also fall below this level and are also independent. The third line is a simplification of the second line under the assumption that the size of the event peaks is identically distributed. The fourth line uses the assumed distributions (GP for event peak sizes and Poisson for annual counts). To get the final result merely requires algebraic manipulation.

A2. Derivation of the Extended GL Distribution

[67] The derivation of the implied distribution for the annual maximum under the negative binomial model for the annual counts of equation (3) follows in a similar manner (some steps are therefore omitted). Let $p = 1/(1 + \lambda\alpha)$ and $r = 1/\alpha$, then

$$
\begin{aligned}
H_r(x) &= \Pr[M_i \le x] \\
&= \sum_{n=0}^{\infty}\left(1-\left[1+\xi\left(\frac{x-u}{\psi}\right)\right]_+^{-1/\xi}\right)^n \frac{\Gamma(r+n)}{\Gamma(r)n!}p^r(1-p)^n \\
&= \begin{cases} p^r\left\{1-(1-p)\left(1-\left[1+\xi\left(\frac{x-u}{\psi}\right)\right]^{-1/\xi}\right)\right\}^{-r} & \text{if } \xi \ne 0, \\ p^r\left\{1-(1-p)\left(1-\exp\left(-\frac{x-u}{\psi}\right)\right)\right\}^{-r} & \text{if } \xi = 0. \end{cases}
\end{aligned}
$$

A3. Mathematical Description of the Autoregressive Process

[68] The first-order autoregressive process used to model year-to-year dependence in the random effects is defined as follows. Take $Z_1 \sim \text{Normal}(0, 1)$ and then for $i = 2, \ldots, k$,

$$Z_i = \rho Z_{i-1} + \epsilon_i, \text{ where } \epsilon_i \sim \text{Normal}\left(0, 1-\rho^2\right),$$

where the residual series $\{\epsilon_i\}$ is independent and $-1 \le \rho \le 1$ is a correlation coefficient. The marginal distribution of this process is standard normal, that is $Z_i \sim \text{Normal}(0, 1)$ for $i = 1, \ldots, k$. To transform this process to have the required Gamma($1/\alpha$, $1/\alpha$) marginal distributions the probability integral transform is used, so that $\gamma_i = F^{-1}[\Phi(Z_i)]$, where $\Phi(\cdot)$ is the standard normal distribution function and $F^{-1}(\cdot)$ is the inverse of the Gamma($1/\alpha$, $1/\alpha$) distribution function.

[69] The vector of random effects $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)$ therefore has a joint distribution function which depends on both the dispersion parameter α and the dependence parameter ρ,

$$
\begin{aligned}
f_{\gamma}(\boldsymbol{\gamma}|\alpha,\rho) = &\frac{1}{(2\pi)^{k/2}|\Sigma|^{1/2}} \\
&\cdot \exp\left\{-\frac{1}{2}\Phi^{-1}(F(\boldsymbol{\gamma}|\alpha))'\Sigma^{-1}\Phi^{-1}[F(\boldsymbol{\gamma}|\alpha)]\right\} \\
&\cdot \prod_{i=1}^{k}\frac{f(\gamma_i|\alpha)}{\phi\left(\Phi^{-1}[F(\gamma_i|\alpha)]\right)},
\end{aligned}
$$

where $f(\cdot)$ and $F(\cdot)$ are the density and distribution functions associated with the Gamma($1/\alpha$, $1/\alpha$) distribution and Σ is the covariance matrix with (i, j)th entry $\rho^{|i-j|}$. The final product in the density function $f_{\gamma}(\cdot)$ comes from the Jacobian term due to the change of variables from $\mathbf{Z}$ to $\boldsymbol{\gamma}$. For the likelihood, it turns out that it is sufficient to write the joint distribution for $\boldsymbol{\gamma}$ as the product of the conditional distributions $f(\gamma_i|\gamma_{i-1}, \alpha, \rho)$ and the marginal distribution $f(\gamma_1|\alpha)$; therefore, only the bivariate analog of the above distribution is required.

[70] **Acknowledgments.** This work was carried out as part of the NERC FREE project "A data-driven exploratory study of extreme events based on joint probability analysis" (grant number NE/F001118/1). The authors are grateful for useful discussions and advice from Cecilia Svensson, David Jones, Thomas Kjeldsen and Lisa Stewart at the Centre for Ecology and Hydrology, Wallingford. They also thank the anonymous referees for suggestions which led to a considerable improvement in the presentation of the paper.

References

Breslow, N. E., and D. G. Clayton (1993), Approximate inference in generalized linear mixed models, *J. Am. Stat. Assoc.*, *88*, 9–25.
Chavez-Demoulin, V., and A. C. Davison (2005), Generalized additive modelling of sample extremes, *Appl. Stat.*, *54*, 207–222.
Chib, S., and I. Jeliazkov (2001), Marginal likelihood from the Metropolis-Hastings output, *J. Am. Stat. Assoc.*, *96*, 270–281.
Coles, S. G., and J. A. Tawn (1996), A Bayesian analysis of extreme rainfall data, *Appl. Stat.*, *45*, 463–478.
Cox, D. R., and V. Isham (1980), *Point Processes*, Chapman and Hall, London.
Cunderlik, J. M., and D. H. Burn (2003), Non-stationary pooled flood frequency analysis, *J. Hydrol.*, *276*, 210–223.
Davison, A. C., and R. L. Smith (1990), Models for exceedances over high thresholds (with discussion), *J. R. Stat. Soc., Ser. B*, *52*, 393–442.
Engeland, K., H. Hisdal, and A. Frigessi (2004), Practical extreme value modelling of hydrological floods and droughts: a case study, *Extremes*, *7*, 5–30.
Gamerman, D., and H. F. Lopes (2006), *Markov Chain Monte Carlo*, 2nd ed., CRC Press, Boca Raton, Fla.
Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004), *Bayesian Data Analysis*, 2nd ed., Chapman and Hall, London.
Gilks, W. R., S. Richardson, and D. J. Spiegelhalter (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.
Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*, 711–732.
Greenwood, J. A., J. M. Landwehr, N. C. Matalas, and J. R. Wallis (1979), Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form, *Water Resour. Res.*, *15*, 1049–1054.
Gustard, A., A. Bullock, and J. M. Dixon (1992), Low flow estimation in the United Kingdom, *Rep. 108*, Inst. of Hydrol, Wallingford, UK.
Karlis, D., and E. Xekalaki (2000), A simulation comparison of several procedures for testing the Poisson assumption, *Statistician*, *49*, 355–382.
Kass, R. E., and A. E. Raftery (1995), Bayes factors, *J. Am. Stat. Assoc.*, *90*, 773–795.
Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Res.*, *25*, 1287–1304.
Kingman, J. F. C. (1993), *Poisson Processes*, Oxford Univ. Press, New York.
Lang, M. (1999), Theoretical discussion and Monte Carlo simulations for a negative binomial process paradox, *Stochastic Environ. Res. Risk Assess.*, *13*, 183–200.
Lang, M., B. M. J. T.Ouarda, and B. Bobée (1999), Towards operational guidelines for over-threshold modelling, *J. Hydrol.*, *225*, 103–117.
Leadbetter, M. R. (1991), On a basis for "Peaks over Threshold" modeling, *Stat. Probab. Lett.*, *12*, 357–362.
Leadbetter, M. R., G. Lindgren, and H. Rootzén (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.
McCullagh, P., and J. A. Nelder (1989), *Generalized Linear Models*, 2nd ed., Chapman and Hall, London.
Mudelsee, M., M. Boröngen, G. Tetzlaff, and U. Grünewald (2003), No upward trends in the occurrence of extreme floods in central Europe, *Nature*, *425*, 166–169.
Prudhomme, C., D. Jakob, and C. Svensson (2003), Uncertainty and climate change impact on the flood regime of small UK catchments, *J. Hydrol.*, *277*, 1–23.
Poortema, K. (1999), On modelling overdispersion of counts, *Stat. Neerlandica*, *53*, 5–20.

Reis, D. S., Jr., and J. R. Stedinger (2005), Bayesian MCMC flood frequency analysis with historical information, *J. Hydrol.*, *313*, 97–116.

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

Ribatet, M., E. Sauquet, J-M. Grésillon, and T. B. M. J. Ourda (2007), A regional Bayesian POT model for flood frequency analysis, *Stochastic Environ. Res. Risk Assess.*, *21*, 327–339.

Robson, A., and D. Reed (1999), *Flood Estimation Handbook*, vol. 3, *Statistical Procedures for Flood Frequency Estimation*, Wallingford HydroSolutions, Wallingford, UK.

Rue, H., and S. Martino (2007), Approximate Bayesian inference for hierarchical Gaussian Markov random fields models, *J. Stat. Plann. Inf.*, *137*, 3177–3192.

Rue, H., S. Martino, and N. Chopin (2009), Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *J. R. Stat. Soc. B*, *71*, 319–392.

Sinharay, S., and H. S. Stern (2005), An empirical comparison of methods for computing Bayes Factors in generalized linear mixed models, *J. Am. Stat. Assoc.*, *14*, 415–435.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002), Bayesian measures of model complexity and fit, *J. R. Stat. Soc. B*, *64*, 583–639.

E. F. Eastoe and J. A. Tawn, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK. (e.eastoe@lancaster.ac.uk)