A Nonparametric Mixture Modeling Framework for Extreme Value Analysis

Ziwei Wang, Abel Rodríguez and Athanasios Kottas *

Abstract: We develop Bayesian nonparametric modeling and inference methods for the analysis of extremes of stochastic processes. We use a point process approach under which the pairwise observations, comprising the time of excesses and the exceedances over a high threshold, are assumed to arise from a non-homogeneous Poisson process. To understand and capture the behavior of rare events, we propose a nonparametric Dirichlet process mixture model for the point process intensity. Particular emphasis is placed on the choice of the mixture kernel to ensure desirable results for the implied tail behavior of the marginal extreme value distribution. At the same time, the mixture nature of the nonparametric model for the intensity of extremes enables more general inferences than traditional parametric methods, which capture temporal heterogeneity for the occurrence of extremes. In particular, the modeling framework yields flexible inference for the joint intensity of extremes, for the marginal intensity over time, and for different types of return level curves. The methodology is illustrated with a simulated data example, and with data involving returns of the Dow Jones index over a five year period.

KEY WORDS: Bayesian nonparametrics; Dirichlet process mixture model; Generalized Pareto distribution; Non-homogeneous Poisson process; Return level functions.

*Z. Wang is Ph.D. student, A. Rodríguez is Associate Professor, and A. Kottas is Associate Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064. (E-mail: *zwang4@ams.ucsc.edu* (Z. Wang); *abel@ams.ucsc.edu* (A. Rodríguez); and *thanos@ams.ucsc.edu* (A. Kottas).) This research was supported in part by the National Science Foundation under awards SES 1024484 and DMS 0915272, and by funds granted by SIGFIRM at the University of California, Santa Cruz.

1 Introduction

Extreme value analysis, which focuses on the study of the tail behavior of a stochastic processes, plays a key role in a number of fields, such as environmental sciences and finance. The literature on extreme value analysis for independent and identically distributed observations is well developed. One popular approach is to model blockwise maxima using the generalized extreme value distribution (Fisher and Tippett, 1928; Gnedenko, 1943). Alternative approaches include modeling the exceedances over a given threshold using a generalized Pareto distribution (Pickands, 1975; Davison and Smith, 1990), or joint modeling of exceedances and the time of their occurrence using a non-homogeneous Poisson process (Smith, 1989; Coles and Tawn, 1996). For a detailed review of statistical models for univariate extremes see, for example, Kotz and Nadarajah (2000) and Coles (2001).

The literature on modeling extremes from more general stochastic processes is less well developed. A simple extension of the blockwise maxima approach can be obtained by assuming that observations are conditionally independent according to a generalized extreme value distribution and introducing dependence across space and/or time by modeling its parameters, using Gaussian processes and/or dynamic linear models (Huerta and Sansó, 2007; Cooley et al., 2007; Sang and Gelfand, 2009). Alternatively, one may consider max-stable processes (Smith, 1990; Schlather, 2002), which provide another natural generalization of the generalized extreme value distribution. Although this approach is appealing from a theoretical perspective, likelihood-based inference for max-stable processes is difficult, since in most cases no closed-form expression for the likelihood is available. Hence, most inference procedures for max-stable processes employ composite likelihoods (Padoan et al., 2010; Genton et al., 2011), which are unappealing from a Bayesian perspective.

This paper pursues a different approach to modeling the extremes of inhomogeneous temporal processes. We consider an extension of the point process approach discussed in Coles and Tawn (1996) that allows for a more general structure for the intensity function of the underlying Poisson point process. More specifically, we use nonparametric mixtures of bivariate kernels to model the intensity function associated with the times and values of the exceedances over a given threshold. A related approach was discussed in Kottas and Sansó (2007) where mixtures of bivariate beta kernels were used to model the

intensity function of the point process. Here, we provide a more scientifically relevant modeling framework for extremes by considering alternative types of mixtures for the Poisson process intensity which ensure that the marginal distributions of the underlying process belong to the Fréchet domain of attraction. Moreover, we develop inference for important extreme value analysis functionals, including different types of return level functions. Our approach also shares some similarities with that of Coles et al. (1994), which extends the point process approach to allow for temporal dependence across locations. Nonparametric mixture models have also been applied to create more flexible models for the extremes of independent and identically distributed observations by using mixtures of Pareto distributions to model the exceedances over a given threshold (Tressou, 2008).

The rest of the article is organized as follows. In Section 2, we present the proposed nonparametric mixture modeling approach, including discussion of relevant background, details of the model formulation, theoretical results, and definition of return level functions. In Section 3, we provide details on model implementation, including methods for prior specification and posterior simulation. Section 4 illustrates the methodology using a simulated data example and a real data set of daily returns of the Dow Jones index over a five year period. Finally, Section 5 concludes with a summary and discussion of possible extensions.

2 Nonparametric point process modeling for analysis of extremes

2.1 Background and Motivation

Let $X_1, \cdots, X_r$ be a sequence of independent random variables with common distribution function F_0. Consider modeling jointly the time and value of exceedances as a point process (e.g., Joe et al., 1992), that is, for regularly spaced observations, we consider the ordered pairs $\{(j, X_j) : j = 1, \ldots, r\}$, where the first entry denotes the period over which each observation is collected. If we restrict attention to those observations that fall above a given threshold u, then our sample is thinned to the pairs $\{(Z_i, Y_i) : i = 1, \ldots, n\}$, where $n \leq r$, Y_i is the value of the i-th exceedance and Z_i is the time at which the i-th exceedance occurred.

The pairs $\{(Z_i, Y_i) : i = 1, \ldots, n\}$ can be regarded as arising from a two-dimensional point process $\{N(A) : A \subset \mathcal{A} = \{1, \ldots, r\} \times [u, \infty)\}$. Pickands (1971) showed that the limiting form of this point process as $u \to \infty$ is a bivariate non-homogeneous Poisson process with intensity function

$$\frac{1}{\sigma}\left\{1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right\}_+^{-1/\xi-1} \tag{1}$$

where $z_+ = \max\{z, 0\}$. Here, the shape parameter ξ is determined by the tail behavior of F_0. In particular, if F_0 has polynomial tails then $\xi > 0$, in which case we say that F_0 is in the Fréchet domain of attraction. On the other hand, if F_0 has bounded support then $\xi < 0$, while $\xi \to 0$ corresponds to the limit when F_0 has exponential tails. Note that this specification implies that the distribution of the blockwise maximum is given by

$$P(y) = \text{pr}(\max\{X_1, \ldots, X_r\} > y) = 1 - \exp\left[-\left\{1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right\}_+^{-1/\xi}\right], \tag{2}$$

which corresponds to the generalized extreme value distribution. Similarly, we can compute the conditional distribution of the exceedances over the threshold u as

$$\text{pr}(X \leq y \mid X > u) = 1 - \left\{1 + \frac{\xi(y-u)}{\sigma + \xi(u-\mu)}\right\}_+^{-1/\xi}, \; y \geq u \tag{3}$$

which is the generalized Pareto distribution. Bayesian inference for this class of models, including elicitation of informative priors from experts, is discussed in Coles and Tawn (1996).

This paper is concerned with extending the point process approach for modeling the tails of a general stochastic process $\{X_t : t \in [0, T]\}$. In the sequel, we denote by F_t the marginal distribution function for X_t, which is related to the conditional distribution of the exceedances over a threshold u at time t through $\text{pr}(X_t \leq y \mid X_t > u) = \{F_t(y) - F_t(u)\}/\{1 - F_t(u)\}$, for $y \geq u$. Similarly, $P_t(y)$, the distribution of the blockwise maxima at time t, can be defined by considering an imaginary sample $X_{t,1}, \ldots, X_{t,r}$, independent and identically distributed from F_t, such that $P_t(y) = \text{pr}(\max_{i \leq r}\{X_{t,i}\} > y)$. In our approach, the pairs $\{(Z_i, Y_i) : i = 1, \ldots, n\}$ are again treated as a realization from a non-homogenous Poisson process, here, on $\mathcal{A} = [0, T] \times [u, \infty)$, with intensity function $\lambda(t, y)$, so that $N(A) \sim \text{Poi}\{\int_A \lambda(t, y)\text{d}t\text{d}y\}$ for any

measurable set $A \subset \mathcal{A}$.

We focus on modeling the intensity function of extremes to provide more flexible inference than the limiting parametric intensity form in (1), which, for instance, is restricted by time homogeneity. To this end, we formulate a mixture model for $\lambda(t,y)$ by exploiting the connection of the Poisson process intensity with a density function. We build on a modeling approach originally developed in Kottas and Sansó (2007), which has also been applied to analysis of immunological studies (Ji et al., 2009) and neuronal data analysis (Kottas and Behseta, 2010; Kottas et al., 2012). Our objective is to develop a flexible inferential framework for extreme value analysis. Hence, in contrast to this earlier work, we seek more structured modeling for the kernel in the mixture representation for $\lambda(t,y)$ to achieve a balance between desirable theoretical properties for the tail behavior of the marginal distribution F_t, and general inference for key extreme value analysis functionals.

2.2 The Modeling Approach

To generate a flexible model for extreme value analysis under the point process approach, we aim at estimating nonparametrically the intensity function, $\lambda(t,y)$, over time and exceedance values. The key observation underlying our modeling approach is that the Poisson process intensity function can be decomposed as $\lambda(\cdot) = \gamma\, f(\cdot)$, where $\gamma \equiv \Lambda(\mathcal{A}) = \int_{\mathcal{A}} \lambda(t,y)\mathrm{d}t\mathrm{d}y$ is the total intensity of exceedances, and $f(\cdot) = \lambda(\cdot)/\Lambda(\mathcal{A})$ is a density function on $\mathcal{A}$, which fully controls the shape of the intensity function. The implicit assumption is that $\Lambda(\mathcal{A}) < \infty$, which can be justified by noting that the Poisson process definition implies that $\exp\{-\Lambda(\mathcal{A})\} = \mathrm{pr}(\{X_t < u : t \in [0,T]\})$. Hence, provided the threshold u and the underlying stochastic process are such that $\mathrm{pr}(\{X_t < u : t \in [0,T]\}) > 0$, the previous identity implies that $\Lambda(\mathcal{A}) < \infty$. For example, this condition is satisfied if $\{X_t : t \in [0,T]\}$ is a Brownian motion; in general, continuous sample paths would likely be needed if one seeks more specific conditions on the underlying stochastic process such that $\mathrm{pr}(\{X_t < u : t \in [0,T]\}) > 0$ holds true.

Under this formulation for the intensity of extremes, we can express the Poisson process likelihood function as

$$L(\gamma, f(\cdot); \{(t_i,y_i) : i = 1,...,n\}) \propto \exp(-\gamma)\gamma^n \prod_{i=1}^{n} f(t_i,y_i). \tag{4}$$

Hence, the problem of estimating the intensity function for the point process of exceedances can be broken down into two independent problems, namely, estimating the total intensity of the Poisson process, and estimating the probability density associated with the distribution of exceedances over the region $\mathcal{A}$.

To generate a rich prior for the Poisson process density, we consider a nonparametric mixture model, $f(t,y) \equiv f(t,y;G) = \int k(t,y \mid \theta)\mathrm{d}G(\theta)$, where $k(t,y \mid \theta)$ is a parametric density on $\mathcal{A}$ indexed by parameter vector θ, and G is a random mixing distribution. Placing a Dirichlet process prior $\mathsf{DP}(\alpha, G_0)$ (Ferguson, 1973) on G results in a Dirichlet process mixture model for $f(t,y;G)$ (Lo, 1984; Escobar and West, 1995). Here, G_0 is the centering distribution of the Dirichlet process, and α controls how close the realization G is to G_0; large values of α result in small variability in Dirichlet process realizations. To study model properties as well as for posterior simulation, we will make use of the Dirichlet process stick-breaking definition (Sethuraman, 1994). According to this definition, the Dirichlet process prior implies that G admits an almost sure representation of the form $G(\cdot) = \sum_{l=1}^{\infty} w_l \delta_{\vartheta_l}$, where $\{\vartheta_1, \vartheta_2, \ldots\}$ is an independent and identically distributed sample from G_0, $w_1 = v_1$ and for $l \geq 2$, $w_l = v_l \prod_{s<l}(1-v_s)$ with $\{v_1, v_2, \ldots\}$ another independent and identically distributed sample from a $\mathsf{Beta}(1,\alpha)$ distribution.

This specification leads to the following mixture model for the intensity of extremes,

$$\lambda(t,y) \equiv \lambda(t,y;G,\gamma) = \gamma\, f(t,y;G) = \gamma \int k(t,y \mid \theta)\mathrm{d}G(\theta), \qquad G \mid \alpha, \psi \sim \mathsf{DP}(\alpha, G_0), \tag{5}$$

where ψ collects the parameters of the centering distribution G_0; as discussed in Section 3, the full Bayesian model involves priors for hyperparameters α and ψ.

Since the purpose of studying extreme values is often to extrapolate the tail behavior of the distribution beyond the observed range of exceedances, and accurate extrapolation in this setting heavily depends on properties of the tail of the density $f(t,y)$, the choice of the mixture kernel $k(t,y \mid \theta)$ is a critical aspect of the model formulation. Indeed, note that, unlike other applications of Poisson processes to spatial modeling, in this problem the nature of the argument of the dimensions associated with $\mathcal{A} = [0,T] \times [u,\infty)$ is very different. Hence, in specifying the mixture kernel density for model (5), we consider a product form

$$k(t,y \mid \theta) \equiv k(t,y \mid \theta_1, \theta_2) = k_1(t \mid \theta_1) k_2(y \mid \theta_2), \tag{6}$$

that is, kernel components k_1 and k_2 are independent before mixing. However, after mixing over the random G, dependence is induced to the resulting mixture model $f(t,y;G)$ for the Poisson process density.

A key objective of our modeling approach is to remove the restriction of time homogeneity implied by (1), and thus for the intensity in the time direction, we seek as general a specification as possible. Because of its flexibility, a (rescaled) beta distribution emerges as a natural choice for the kernel component over time,

$$k_1(t \mid \theta_1) = \frac{\Gamma(\tau)}{T^{\tau-1}\Gamma(T^{-1}\kappa\tau)\Gamma\{\tau(1-T^{-1}\kappa)\}} t^{T^{-1}\kappa\tau-1}(T-t)^{\tau(1-T^{-1}\kappa)-1}, \quad t \in (0,T) \tag{7}$$

where $\theta_1 = (\kappa,\tau)$, $\kappa \in (0,T)$ is the mean of the beta distribution, and $\tau > 0$ is a scale parameter.

The choice of kernel component $k_2(y \mid \theta_2)$ is more delicate. The previous attempt in Kottas and Sansó (2007) to nonparametric mixture modeling for extremes under the point process approach utilized kernels with bounded support, defined by a bivariate beta distribution, and thus the underlying F_t was implicitly assumed to have compact support. This is restrictive for many applications. Moreover, kernel k_2 is used to capture through mixing the tail behavior of the underlying distribution where we do not expect, for instance, multimodalities. Hence, applying the mixture model with a beta density for k_2 may lead to overfitting, which is especially damaging for extrapolation. In contrast, the asymptotic theory for extremes suggests what the tail behavior is, and we can make use of that information to improve inference under the mixture model.

Hence, we build the intensity function in the exceedances direction from a kernel defined through a special case of the generalized Pareto distribution,

$$k_2(y \mid \theta_2) = \frac{1}{\sigma}\left(1 + \frac{\xi(y-u)}{\sigma}\right)^{-1/\xi-1}, \quad y \geq u \tag{8}$$

such that $\theta_2 = (\sigma,\xi)$ with $\sigma > 0$ and $\xi > 0$. The location parameter is set to the specified threshold value u to ensure that the resulting mixture kernel $k(t,y \mid \theta_1,\theta_2)$ has support on $\mathcal{A}$. Moreover, we focus on the $\xi > 0$ range for the shape parameter, which ensures that the corresponding marginal distributions belong to the Fréchet maximum domain of attraction, that is, we are modeling an underlying stochastic process

with heavy tailed behavior. Specifically, the distribution for X_t is in the Fréchet domain of attraction if, for sufficiently large x, $\text{pr}(X_t > x) \approx Cx^{-\rho}L(x)$, where C and ρ are non-negative quantities, which are constants in x, and $L(x)$ is a slowly varying function, that is, $L(x)$ satisfies $\lim_{x\to\infty} L(sx)/L(x) = 1$, for all $s > 0$ (Embrechts et al., 1999). The tail index parameter ρ has a useful interpretation as a risk indicator – larger values of ρ^{-1} lead to larger probability of exceeding the specified level x – and its estimation has been considered extensively in the literature; see, for instance, the related discussion and references in Tressou (2008).

The specific result under our modeling approach is formulated below as a theorem, whose proof can be found in Appendix A.

THEOREM 1. *Assume a non-homogenous Poisson process model on $\mathcal{A} = [0,T] \times [u,\infty)$ for the times and values of the exceedances, given a fixed threshold u, of a stochastic process $\{X_t : t \in [0,T]\}$ with right-continuous sample paths. Consider the mixture model defined by (5) – (8) for the Poisson process intensity function. Then, the marginal distributions of the process, $pr(X_t > x)$, belong to the Fréchet maximum domain of attraction.*

Key to the proof of the theorem is a representation of the tail probability for the underlying process marginals at any specific time point in terms of the conditional Poisson process density at that time point. This result is of independent interest and is thus given as a lemma with the proof also included in Appendix A.

LEMMA 1. *Consider a stochastic process $\{X_t : t \in [0,T]\}$ with right-continuous sample paths, and the point process whose points comprise the time, t, and value, y, of exceedances of process $\{X_t : t \in [0,T]\}$ above a given threshold u. Assume a non-homogenous Poisson model for the point process with intensity function $\lambda(t,y) = \gamma f(t,y)$, for $(t,y) \in [0,T] \times [u,\infty)$, where $\gamma = \int_{\mathcal{A}} \lambda(t,y)dtdy$. Then, for any specified time point t_0,*

$$pr(X_{t_0} > x \mid X_{t_0} > u) = \int_x^\infty f(y \mid t_0)dy = \int_x^\infty f(t_0,y)dy/\{f(t_0)\}, \quad x > u.$$

The practical utility of the lemma is that it enables time-dependent inference for tail probabilities of

the marginal distributions of the underlying process – which is observed only through its exceedances above the given threshold – based on the nonparametric mixture model for the point process density. Note that the result of Lemma 1 is not specific to the particular modeling approach as its proof utilizes only the $\lambda(t,y) = \gamma f(t,y)$ formulation for the Poisson process intensity. However, the rest of the proof for Theorem 1 uses the mixture representation for the Poisson process density and the specific mixture kernel built from (7) and (8). In fact, the argument relies on a truncation approximation to the Dirichlet process representation, which is also used in the posterior simulation approach; see Section 3.1. The assumption of right-continuous sample paths for the underlying process is needed for the proof of the lemma. From a theoretical point of view, this assumption is not restrictive given the availability of results on existence of versions of stochastic processes with right-continuous sample paths.

An appealing feature of the mixture model formulation in (5) – (8) is that the model for the intensity of extremes can be interpreted as accommodating time inhomogeneities through local adaptive fitting of generalized Pareto distributions, where the mode of the beta kernel associated with each distinct mixture component serves to localize the effect of the generalized Pareto kernel in time. This feature provides flexibility with respect to capturing non-standard underlying intensity shapes as they are suggested by the data. In addition, note that if $\alpha \to 0^+$, which results in a single mixture component, and if the beta kernel component is reduced to a uniform, we recover as a special case the parametric model for the intensity function in (1) with $\mu = u$.

Traditionally, one of the key goals of extreme value theory is to estimate the return level function of the process, which is strongly connected to the intensity function; see, for example, Coles (2001). In the case of non-homogeneous processes, we can define two different types of return level functions. For a given point $t_0 \in (0,T)$ and a small $\varepsilon > 0$, we define the ε-conditional return level curve as given by the solution to the equation $\mathrm{pr}(\{X_t > x_m : t \in [t_0 - \varepsilon, t_0 + \varepsilon]\}) = m^{-1}$, for different values of m. Under the nonparametric mixture model, for any $x > u$,

$$\mathrm{pr}(\{X_t > x : t \in [t_0 - \varepsilon, t_0 + \varepsilon]\}; G, \gamma) = 1 - \exp\left[-\gamma \int \{1 - K_2(x \mid \sigma, \xi)\} \{K_1(t_0 + \varepsilon \mid \kappa, \tau) - K_1(t_0 - \varepsilon \mid \kappa, \tau)\} \, \mathrm{d}G(\kappa, \tau, \sigma, \xi)\right], \quad (9)$$

where K_1 and K_2 denote the distribution functions for the beta and generalized Pareto kernel components, respectively. Hence, the ε-conditional return level x_m at time t_0 corresponds approximately to a realization of the process that would be exceeded only once in every m periods if future draws were to be taken according to the underlying F_{t_0}.

We can also define a marginal return level curve through the average intensity function $\tilde{\Lambda}([x,\infty)) = T^{-1}\int_0^T \int_x^\infty \lambda(t,y)\mathrm{d}t\mathrm{d}y$, for $x > u$. Proceeding as before, we define the marginal return level curve as the solution to $\mathrm{pr}(\tilde{X} > x_m) = m^{-1}$, where

$$\mathrm{pr}(\tilde{X} > x; G, \gamma) = 1 - \exp\left[-T^{-1}\gamma \int \{1 - K_2(x \mid \sigma, \xi)\}\, \mathrm{d}G(\sigma, \xi)\right], \tag{10}$$

and $\tilde{X}$ corresponds to the outcome associated with an "average" period. Hence, unlike the conditional return level curve, which provides information about the likelihood of extremes at a specific time point t_0, the marginal return level curve provides an average over all $t \in [0, T]$.

To understand the relationship between conditional and marginal return level curves, it is useful to compare equations (9) and (10) with the tail probability obtained by modeling $\lambda(t,y)$ using (1). Since in that case the intensity function is time homogeneous, the marginal and conditional tail probabilities agree for the traditional parametric model, and correspond to the one generated from (2). For non-homogeneous processes the marginal and conditional return level curves provide important and distinct insights into the behavior of the underlying stochastic process. While marginal return level curves can be used to assess what extremes look like on a "normal" period, the conditional return level curves can be used to examine specific past dates, providing insights about the behavior of the underlying process on a particularly "good" or "bad" period.

3 Implementation Details

3.1 Posterior Simulation and Inference

Based on the form of the Poisson process likelihood in (4), the marginal posterior distribution for γ is analytically available as a gamma distribution under a gamma prior or the marginal reference prior, which

is given by $p(\gamma) \propto \gamma^{-1} 1(\gamma > 0)$ (Kottas and Behseta, 2010). In particular, under the latter prior, the joint posterior distribution is proper, and $p(\gamma \mid \text{data})$ is simply a gamma$(n, 1)$ distribution.

Inference for the Poisson process density requires the computation of the posterior distribution for the random mixing distribution G and the Dirichlet process prior hyperparameters. Full posterior inference under Dirichlet process mixture models can be obtained by using a truncated version of G,

$$G^N(\cdot) = \sum_{l=1}^{N} p_l \delta_{\zeta_l}(\cdot),$$

where the ζ_l are independent draws from the base distribution G_0 and $p_1, \ldots, p_N$ are the associated weights defined using a stick-breaking construction under the constraint $p_N = 1 - \sum_{l=1}^{N-1} p_l$. Introducing configuration variables $\mathbf{L} = (L_1, \cdots, L_n)$, where $L_i = l$ if and only if the mixing parameter corresponding to observation (t_i, y_i) is given by ζ_l, the hierarchical model for the data is written as:

$$\begin{aligned}
(t_i, y_i) \mid \kappa_{L_i}, \tau_{L_i}, \sigma_{L_i}, \xi_{L_i} &\overset{ind.}{\sim} k_1(t_i \mid \kappa_{L_i}, \tau_{L_i}) k_2(y_i \mid \sigma_{L_i}, \xi_{L_i}), \quad i = 1, \ldots, n \\
L_i \mid \mathbf{p} &\overset{ind.}{\sim} \sum_{l=1}^{N} p_l \delta_l(L_i), \qquad i = 1, \ldots, n \\
\mathbf{p} \mid \alpha &\sim f(\mathbf{p} \mid \alpha) \\
\zeta_l = (\kappa_l, \tau_l, \sigma_l, \xi_l) &\overset{ind.}{\sim} G_0(\zeta_l \mid \psi), \qquad l = 1, \ldots, N
\end{aligned}$$

where the induced prior $f(\mathbf{p} \mid \alpha)$ for the vector of weights $\mathbf{p} = (p_1, \ldots, p_N)$, given α, is given by a generalized Dirichlet distribution (Ishwaran and James, 2001). The structure of the centering distribution, G_0, and its hyperparameters, ψ, is discussed in Section 3.2, where we also discuss the priors for α and ψ which complete the full Bayesian model.

We employ a blocked Gibbs sampler (Ishwaran and Zarepour, 2000; Ishwaran and James, 2001) to obtain samples from the full posterior distribution $p(\sigma, \xi, \kappa, \tau, \mathbf{L}, \mathbf{p}, \alpha, \psi \mid \text{data})$. Details of the posterior simulation algorithm are provided in Appendix B.

Using the posterior samples for $G^N \equiv \{(p_l, \kappa_l, \tau_l, \sigma_l, \xi_l) : l = 1, ..., N\}$, we can obtain full inference for the joint intensity of extremes, $\gamma \sum_{l=1}^{N} p_l \; k_1(t \mid \kappa_l, \tau_l) k_2(y \mid \sigma_l, \xi_l)$, the marginal density of exceedance

times, $\sum_{l=1}^{N} p_l \, k_1(t \mid \kappa_l, \tau_l)$, and for tail probabilities of the underlying process based on Lemma 1. Similarly, approximate inferences for marginal and conditional return level curves can be obtained by replacing G in (9) and (10) with its truncation approximation G^N.

3.2 Prior Specification

We assume that the different components of the centering distribution G_0 are independent, that is, $G_0(\kappa, \tau, \sigma, \xi) = G_0^{\kappa}(\kappa) G_0^{\tau}(\tau) G_0^{\sigma}(\sigma) G_0^{\xi}(\xi)$. For G_0^{σ}, we use an inverse-gamma distribution with fixed shape parameter $a_\sigma = 2$, which implies infinite prior variance, and random mean parameter b_σ, which is assigned an exponential prior with mean d_σ. Hence, the specific choice of d_σ allows us to control the prior mean value for σ while being relatively non-infomative about this choice. For G_0^{ξ}, we take an exponential distribution with mean b_ξ; an inverse-gamma prior with shape parameter 2 and mean d_ξ is placed on b_ξ. We suggest that the values of d_σ and d_ξ are selected to reflect the scale of the data under a single component of the mixture model. In particular, with a prior guess at the mean and variance for the exceedance values, we can numerically solve for σ and ξ from the equation of the mean and variance of the generalized Pareto distribution. Then, we set the solutions to d_σ and d_ξ, which are the prior means for σ and ξ, respectively.

Regarding the parameters of the beta kernel component, for G_0^{κ}, we work with a beta distribution, with fixed parameters, for the scaled mean κ/T. For G_0^{τ}, we take an inverse-gamma distribution with shape parameter equal to 2 and mean parameter b_τ to which we place an exponential prior with mean d_τ. To specify d_τ and the parameters for G_0^{κ}, we study the implied prior for the marginal density of exceedance times. Based on the connection between the parametric and nonparametric formulations for the point process model, a non-informative specification may be built from a uniform prior mean for this marginal density. On the other hand, for some applications we may wish to encourage priors that favor clustering of extreme values, and this can also be achieved through appropriate specification of G_0^{κ} and G_0^{τ}. We provide illustrations of both scenarios with the data examples of Section 4.

Finally, we use a gamma prior for α, and the reference prior, discussed in Section 3.1, for γ.

4 Illustrations

4.1 Simulation Study

To illustrate our modeling approach, we first consider a simulated data set where observations were generated according to a non-linear regression model, $X_t = \mu(t) + We_t$, with mean function

$$\mu(t) = -0.5 + 1.6(t/T) + 0.5\sin(-5.4 + 10.8(t/T)) + 1.1\{1 + 4(2(t/T) - 1)^2\}^{-1}.$$

Here, the noise terms e_t are independently distributed according to a Student t distribution with 3 degrees of freedom for every t, and $W = 0.32$. This choice implies that $\text{var}\{We_t\} = 0.3$ for all t. The raw dataset contains $T = 10,000$ observations equally spaced in the interval $[0,T]$; to assemble the final data set, we retain observations that are larger than the threshold $u = 2.1$, which results in $n = 525$ extreme observations.

We assign a gamma prior to the precision parameter α with mean 5 and variance 2.5. Also, following the approach discussed in Section 3.2, we set $d_\sigma = 0.296$ and $d_\xi = 0.257$. Moreover, we set $d_\tau = 300$, and consider two prior choices for G_0^κ. The first is a beta distribution for κ/T with mean 0.5 and variance $1/28$, whereas the second prior is based on a uniform distribution for κ/T. The effect of these two choices on the implied prior for the marginal density of exceedances over time is illustrated in the first two rows of Figure 1. The first row plots 10 prior realizations for this marginal density, while the second row shows the prior mean along with 95% pointwise credible intervals. Both prior specifications induce a large degree of variability for the marginal density of exceedance times, with individual realizations being highly multimodal. However, the first prior choice tends to favor exceedances located in the middle of the time interval, while the second prior implies a more uniform distribution of exceedances.

The algorithm discussed in Section 3.1 was used to fit our model. A total of 4,000 posterior samples were used for all inferences. These samples were obtained after thinning a sample of 200,000 from which 40,000 iterations were discarded as burn-in. Posterior mean estimates of the joint intensity function $\lambda(t,y)$ are presented in the third row of Figure 1, while the last row shows the posterior mean and 95% credible intervals for the marginal density of exceedance times. Note that posterior inference is quite

robust to the specific prior choices.

In addition to providing estimates of the intensity function, we are interested in investigating the ability of the model to estimate the tails of the stochastic process, and in comparing its performance with the parametric model discussed in Section 2.1. For this purpose, we present in Figure 2 the true conditional return level function corresponding to four time points, along with posterior estimates generated under the parametric model for the intensity function given in (1), as well as under the nonparametric model, using the two prior choices discussed above. For the parametric model, we utilize a normal prior on the location parameter μ with mean 3.23 and variance 10; an inverse-gamma prior on the scale parameter σ with shape parameter 2 and mean 0.43; and an exponential prior on the shape parameter ξ with mean 0.048. The parametric model is fitted using a Gaussian random walk Metropolis algorithm that samples jointly the three parameters on an appropriately transformed scale for each parameter.

The right column of Figure 2 shows that the parametric model performs poorly at capturing the true return level curve at all four time points. Moreover, the credible intervals are very narrow, which suggests that the parametric model dramatically underestimates the uncertainty in situations where the process is not homogeneous in time. On the other hand, the estimates generated by the nonparametric model are quite accurate, particularly for time points $t_0 = 5100$ and $t_0 = 6500$ around which a relatively large number of exceedances are concentrated. Again, posterior inference is robust to the two prior choices. Finally, note that the nonparametric mixture model estimates for the marginal return level curve are almost identical to those generated by the parametric model, involving only a minimal increase in posterior uncertainty.

4.2 Dow Jones Data

Here, we discuss the analysis of extremes of the daily returns for the Dow Jones index between September 11, 1995 and September 7, 2000; a previous analysis of this data set is presented in Coles (2001). Modeling the lower tail of the distribution for returns of financial assets is critical to compute risk measures such as the Value at Risk or the Expected Shortfall.

In the sequel we work with the negative log returns of the index, $y_t = -\log(x_t/x_{t-1})$, where x_t is

the closing price at day t. Note that, in this case, drops in the index correspond to positive values of y_t, while increases correspond to negative values. Figure 3 shows the $n = 82$ values of y_t above the threshold $u = 1.5$. The vertical dashed lines mark the dates at which three financial crises started: the mini-crash on October 27, 1997, the Russian financial crisis on August 17, 1998, and the bursting of the dot-com bubble on March 10, 2000. We see that the three biggest drops in the index align well with these three financial crises, and that a large number of exceedances concentrate around those dates.

For the analysis of this data set we used priors that are similar to the second prior specification discussed in Section 4.1. Figure 4 shows 10 prior realizations for the marginal density of exceedance times, as well as prior mean and 95% interval estimates for this density. Given that the time period under study includes multiple crises around which extreme values may cluster, a prior choice that favors multimodal exceedance time densities is arguably justified. At the same time, the corresponding prior mean is fairly close to a uniform density with wide uncertainty bands. Hence, even though our prior favors the clustering of extreme values, we make no prior assumption about the location of such clusters.

Figure 4 includes posterior inference results for the bivariate intensity function, and for the marginal density of exceedance times. Note that the nonparametric model captures reasonably well the localized characteristics of the raw data and cyclical nature of the business cycle. Interestingly, the estimates also suggest an increasing risk of extreme loses over the time period under study. Both of these features are captured by the model even though it does not contain any explicit term to account for them.

Next, we report on inference for ε-conditional return level curves, obtained using equation (9) with $\varepsilon = 0.5$, that is, daily conditional return level curves. Based on the nonparametric mixture model, the left column of Figure 5 shows posterior mean and 95% interval estimates for the conditional return level curve at four specific dates, September 26, 1996, October 27, 1997, August 17, 1998, and July 20, 1999. Note that the posterior mean estimates of the return level curve at October 27, 1997 and August 17, 1998 are uniformly above the ones at September 26, 1996 and July 20, 1999. This is consistent with the fact that the former two dates fall within periods of financial distress, while the latter do not. In addition, we compare the nonparametric model estimates against those generated by the parametric model; see the right column of Figure 5. Since the parametric model is unable to capture the time inhomogeneity

in the data, it produces the same point estimate at all dates with much narrower posterior uncertainty bands. Finally, Figure 6 shows posterior point and interval estimates for the marginal return level curve under both the parametric and nonparametric models. We note that the point estimates are similar, but the uncertainty levels associated with the nonparametric model are higher.

To assess prior sensitivity for the nonparanetric model, Figures 7 and 8 provide inference results under three different prior choices for the parameters of the beta distribution for G_0^{κ}, for the exponential prior, with mean d_τ, for the mean parameter b_τ of the inverse-gamma distribution for G_0^{τ}, and for the parameters of the gamma prior for α. Specifically, the left column of Figures 7 and 8 corresponds to the first prior specification with a uniform distribution for κ/T, $d_\tau = 40$, and $\mathsf{E}(\alpha) = 5$, $\text{var}(\alpha) = 2.5$; the middle column to the second prior choice using a beta distribution for κ/T with $\mathsf{E}(\kappa/T) = 0.5$ and $\text{var}(\kappa/T) = 1/28$, $d_\tau = 300$, and $\mathsf{E}(\alpha) = 5$, $\text{var}(\alpha) = 2.5$; and the right column to the third prior choice involving a uniform distribution for κ/T, $d_\tau = 2$, and $\mathsf{E}(\alpha) = 1$, $\text{var}(\alpha) = 1$. As in earlier examples, Figure 7 shows prior realizations for the marginal density of exceedance times, and prior mean and interval estimates for this density. The first prior favors unimodal densities and results in a slightly U-shaped prior mean density. Although the second prior, which is the same as the first prior considered in Section 4.1, encourages multimodal density realizations, it yields a unimodal prior mean density. Finally, the third prior is chosen to strongly favor U-shaped density realizations, including a U-shaped prior mean density with a relatively low level of uncertainty associated with it; clearly, this is a prior choice that would not be recommended for this particular problem.

Figure 7 plots posterior estimates for the intensity function and for the density of exceedance times. The estimates under the first two priors show features that are similar to those we obtained under the original prior. The estimates under the third prior, which strongly supports the absence of localized features, capture the increasing trend in risk, but not the clustering of extremes. Figure 8 presents inference for the daily conditional return level curves at the same time points we considered above, along with estimates of the marginal return level curve. Again, posterior inference under the first two priors is similar to that obtained under the original prior. However, results under the third prior differ, particularly in terms of the uncertainties that the model attaches to the posterior mean estimates.

5 Discussion and Future Work

We have presented a Bayesian nonparametric model for the analysis of extremes under a generalization of the point process approach. Our model is built to relax the time homogeneity restriction through flexible mixture modeling for the intensity function of extremes. Particular emphasis has been placed on the mixture model formulation for this intensity function to obtain desirable properties for the tail behavior of the underlying process whose extremes are recorded. Our empirical results suggest that the model is quite robust to the choice of priors when sample sizes are moderately large, as in our simulation study, but might be affected by strongly informative priors when sample sizes are relatively small, as in the Dow Jones data application. As a general strategy we suggest the use of priors similar to the ones discussed in Section 4.2, which allow for clustering of extreme values when such a feature is suspected in the data.

The starting point of the mixture model formulation is the product kernel specification in (6), motivated by the different nature of the arguments that comprise the support of the bivariate Poisson process. A consequence of this specification is that the implied tail index parameter does not depend on time; see the proof of Theorem 1. From a practical point of view, this is arguably not a serious limitation, since the typically small number of exceedances will likely not suffice to inform temporally dependent tail index indicators. Nevertheless, this methodological extension can be developed through choice of an appropriate bivariate mixture kernel $k(t, y \mid \theta)$. For instance, a possible modification of the form in (6) involves the same beta kernel component for $k_1(t)$ with a conditional Pareto distribution for $k_2(y \mid t)$ defined by extending the shape parameter ξ in (8) to a parametric function $\xi(t)$. Then, the same argument as in the proof of Theorem 1 yields a temporally dependent tail index indicator. In particular, the choice $\xi(t) = \exp(\beta_0 + \beta_1 t)$, with Dirichlet process mixing on the real-valued parameters β_0 and β_1, leads to a mixture model that includes the model of Section 2.2 as a special case.

A different direction for elaboration of the modeling framework involves inference and prediction for extremes recorded over time and space. For instance, for environmental processes observed at multiple monitoring stations over time, the mixture model for the intensity of extremes can be extended to a spatio-temporal model by extending the mixing distribution G to a random spatial surface $G_{\mathcal{S}} = \{G_s : s \in \mathcal{S}\}$.

The literature includes several nonparametric prior models to capture the spatial dependence of random distributions over an appropriate space $\mathcal{S}$. In particular, the spatial Dirichlet process (Gelfand et al., 2005; Kottas et al., 2008) offers one possible approach by inducing the spatial dependence in the atoms of the random mixing distributions G_s. An alternative modeling strategy is to introduce the spatial dependence on the associated weights in the mixing distribution; see, for instance, Rodriguez and Dunson (2011). Fully nonparametric modeling for spatio-temporal extremes along these lines will be reported in a future manuscript.

Appendix A: Proofs

Proof of Lemma 1. Consider a generic $x > u$, where u is the given threshold, and fix a time point t_0. Using continuity of probability measure and the right-continuity of the sample paths of the underlying stochastic process $\{X_t : t \in [0,T]\}$, we have $\text{pr}(X_{t_0} > x) = \lim_{\Delta t \to 0} \text{pr}\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\}\right)$.

The same argument applies to $\text{pr}(X_{t_0} > u) = \lim_{\Delta t \to 0} \text{pr}\left(\{X_t > u : t \in [t_0, t_0 + \Delta t)\}\right)$, resulting in

$$\text{pr}(X_{t_0} > u) = \lim_{\Delta t \to 0} \left\{ 1 - \exp\left(- \int_u^{\infty} \int_{t_0}^{t_0 + \Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \right\} \tag{A.1}$$

based on the Poisson assumption for the point process of exceedances.

Next, define M_0 as the number of exceedances in time interval $[t_0, t_0 + \Delta t)$. Then, we can write

$$\text{pr}\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\}\right) = \sum\nolimits_{m \geq 1} \text{pr}\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\} \cap \{M_0 = m\}\right).$$

Using the Poisson process assumption, $\lim_{\Delta t \to 0} (\Delta t)^{-1} \text{pr}(M_0 = m) = 0$, for $m \geq 2$, resulting in $\text{pr}(X_{t_0} > x) = \lim_{\Delta t \to 0} \text{pr}\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\} \cap \{M_0 = 1\}\right)$. Based again on the Poisson process structure, $\text{pr}\left(\{X_t > x : t \in [t_0, t_0 + \Delta t)\} \cap \{M_0 = 1\}\right) = \text{pr}(1 \text{ event in } [t_0, t_0 + \Delta t) \times [x, \infty)) \times \text{pr}(0 \text{ events in } [t_0, t_0 + \Delta t) \times [u, x)) = \left\{ \exp\left(- \int_x^{\infty} \int_{t_0}^{t_0 + \Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \left(\int_x^{\infty} \int_{t_0}^{t_0 + \Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \right\} \times \left\{ \exp\left(- \int_u^{x} \int_{t_0}^{t_0 + \Delta t} \lambda(t,y) \mathrm{d}t \mathrm{d}y \right) \right\},$

and thus,

$$\text{pr}(X_{t_0} > x) = \lim_{\Delta t \to 0} \exp\left(-\int_u^\infty \int_{t_0}^{t_0+\Delta t} \lambda(t,y)\mathrm{d}t\mathrm{d}y\right)\left(\int_x^\infty \int_{t_0}^{t_0+\Delta t} \lambda(t,y)\mathrm{d}t\mathrm{d}y\right) = \lim_{\Delta t \to 0} \int_x^\infty \int_{t_0}^{t_0+\Delta t} \lambda(t,y)\mathrm{d}t\mathrm{d}y. \tag{A.2}$$

Combining (A.1) and (A.2), we obtain for any $x > u$,

$$\text{pr}(X_{t_0} > x \mid X_{t_0} > u) = \frac{\text{pr}(X_{t_0} > x)}{\text{pr}(X_{t_0} > u)} = \frac{\lim_{\Delta t \to 0}(\Delta t)^{-1} \int_x^\infty \int_{t_0}^{t_0+\Delta t} \lambda(t,y)\mathrm{d}t\mathrm{d}y}{\lim_{\Delta t \to 0}(\Delta t)^{-1}\left\{1 - \exp\left(-\int_u^\infty \int_{t_0}^{t_0+\Delta t} \lambda(t,y)\mathrm{d}t\mathrm{d}y\right)\right\}}.$$

The limit in the numerator yields $\int_x^\infty \lambda(t_0,y)\mathrm{d}y$. Regarding the denominator, we use a first-order Maclaurin series expansion of the function $g(\Delta t) = \exp\left(-\int_u^\infty \int_{t_0}^{t_0+\Delta t} \lambda(t,y)\mathrm{d}t\mathrm{d}y\right)$ to obtain $g(\Delta t) = g(0) + (\Delta t)g'(0) + R_2(\Delta t) = 1 - (\Delta t)\int_u^\infty \lambda(t_0,y)\mathrm{d}y + R_2(\Delta t)$, where $\lim_{\Delta t \to 0}(\Delta t)^{-1} R_2(\Delta t) = 0$. Hence,

$$\lim_{\Delta t \to 0}(\Delta t)^{-1}\left\{1 - \exp\left(-\int_u^\infty \int_{t_0}^{t_0+\Delta t} \lambda(t,y)\mathrm{d}t\mathrm{d}y\right)\right\} = \int_u^\infty \lambda(t_0,y)\mathrm{d}y = \gamma f(t_0),$$

and thus finally,

$$\text{pr}(X_{t_0} > x \mid X_{t_0} > u) = \frac{\int_x^\infty \lambda(t_0,y)\mathrm{d}y}{\gamma f(t_0)} = \frac{\gamma \int_x^\infty f(t_0,y)\mathrm{d}y}{\gamma f(t_0)} = \int_x^\infty f(y \mid t_0)\mathrm{d}y.$$

Proof of Theorem 1. Consider a sufficiently large, generic $x > u$, where u is the given threshold. We seek to prove that the marginal distributions of the underlying process satisfy $\text{pr}(X_t > x) \approx Cx^{-\rho}L(x)$, where ρ is the tail index parameter and $L(x)$ is a slowly varying function.

First, note that, since $1 - F_t(u)$ is a positive constant in x, $\text{pr}(X_t > x)$ belongs to the Fréchet maximum domain of attraction if and only if $\text{pr}(X_t > x \mid X_t > u) = \{1 - F_t(x)\}/\{1 - F_t(u)\}$ does, and importantly, both distributions have the same tail index parameter. Hence, it suffices to work with conditional distribution $\text{pr}(X_t > x \mid X_t > u)$ at a specific time point t.

To complete the proof, we employ the truncation approximation G^N to the Dirichlet process representation for mixing distribution G, which as discussed in Section 3.1, provides the version of the mixture

model applied to the data. Then based on Lemma 1, we obtain

$$\text{pr}(X_t > x \mid X_t > u) = \int_x^\infty f(y \mid t)\text{d}y = \frac{\sum_{l=1}^N p_l k_1(t \mid \kappa_l, \tau_l)\left\{1+\sigma_l^{-1}\xi_l(x-u)\right\}^{-1/\xi_l}}{\sum_{l=1}^N p_l k_1(t \mid \kappa_l, \tau_l)} = \sum_{l=1}^N \omega_l (A_l + B_l x)^{-1/\xi_l}$$

where $\omega_l = p_l k_1(t|\kappa_l,\tau_l)/\{\sum_{l=1}^N p_l k_1(t|\kappa_l,\tau_l)\}$, $A_l = 1-\sigma_l^{-1}\xi_l u$, and $B_l = \sigma_l^{-1}\xi_l$. Note that the weights ω_l depend on the specified time point t, but not on level x. Next, letting $\rho = \min\{\xi_l^{-1} : l = 1,...,N\}$ and $l^* = \arg\min\{\xi_l^{-1} : l = 1,...,N\}$, we can write

$$\text{pr}(X_t > x \mid X_t > u) = \sum_{l=1}^N \omega_l (A_l + B_l x)^{-1/\xi_l} = x^{-\rho} \sum_{l=1}^N \omega_l \left(\frac{A_l}{x} + B_l\right)^{-\frac{1}{\xi_l}} x^{-\frac{1}{\xi_l}+\rho} = x^{-\rho} L(x).$$

Now,

$$\lim_{x\to\infty} \left(\frac{A_l}{x} + B_l\right)^{-\frac{1}{\xi_l}} x^{-\frac{1}{\xi_l}+\rho} = \begin{cases} B_l^{-1/\xi_l} & \text{if } l = l^* \\ 0 & \text{otherwise} \end{cases}$$

and therefore $\lim_{x\to\infty} L(sx)/L(x) = 1$, for any $s > 0$, which completes the argument.

Appendix B: Details for the posterior simulation algorithm

Here, we provide the details for the simulation from the full posterior distribution of the hierarchical model for the Poisson process density discussed in Section 3.1. The Markov chain Monte Carlo algorithm iteratively updates model parameters according to the following steps:

Updating L_i, $i = 1,\ldots,n$: each draw of L_i is from a discrete distribution, $\sum_{l=1}^N \tilde{p}_{l,i}\delta_l(L_i)$, where the vector of revised weights, $\tilde{p}_l \propto (p_1 k(t_i,y_i|\zeta_1),\ldots,p_N k(t_i,y_i|\zeta_N))$.

Updating α and $\mathbf{p}$: The updates for these parameters are generic for any choice of kernel in the Dirichlet process mixture model; details are given in Ishwaran and Zarepour (2000).

Updating the centering distribution parameters: based on the conditionally conjugate priors used for b_σ, b_ξ, and b_τ, each of the corresponding posterior full conditionals is available in closed form. In particular, for b_σ this is a gamma distribution with shape parameter $1 + Na_\sigma$ and scale parameter $\left(d_\sigma^{-1} + \sum_{l=1}^N \sigma_l^{-1}\right)^{-1}$. Moreover, b_ξ has an inverse-gamma posterior full conditional with shape parameter

$2+Na_\xi$ and scale parameter $d_\xi+\sum_{l=1}^{N}\xi_l$. Finally, the full conditional for b_τ is a gamma distribution with shape parameter $1+Na_\tau$ and scale parameter $\left(d_\tau^{-1}+\sum_{l=1}^{N}\tau_l^{-1}\right)^{-1}$.

Updating $(\kappa_l,\tau_l,\sigma_l,\xi_l)$, for $l=1,...,N$: Let n^* be the number of distinct clusters in vector $\mathbf{L}$, and $L_1^*,\cdots L_{n^*}^*$ the distinct values in vector $\mathbf{L}$. Then if $l\notin\{L_j^*:j=1,\ldots,n^*\}$, $(\kappa_l,\tau_l,\sigma_l,\xi_l)$ is drawn from the centering distribution G_0. If $l\in\{L_j^*:j=1,\ldots,n^*\}$, the posterior full conditional for each of the four mixing parameters is given by:

$$p(\sigma_l|\cdots,\text{data})\propto \mathrm{d}G_0^{\sigma}(\sigma_l)\prod_{\{i:L_i=l\}}k_2(y_i\mid\sigma_l,\xi_l)$$

$$p(\xi_l|\cdots,\text{data})\propto \mathrm{d}G_0^{\xi}(\xi_l)\prod_{\{i:L_i=l\}}k_2(y_i\mid\sigma_l,\xi_l)$$

$$p(\kappa_l|\cdots,\text{data})\propto \mathrm{d}G_0^{\kappa}(\kappa_l)\prod_{\{i:L_i=l\}}k_1(t_i\mid\kappa_l,\tau_l)$$

$$p(\tau_l|\cdots,\text{data})\propto \mathrm{d}G_0^{\tau}(\tau_l)\prod_{\{i:L_i=l\}}k_1(t_i\mid\kappa_l,\tau_l)$$

where $k_1(t\mid\kappa_l,\tau_l)$ and $k_2(y\mid\sigma_l,\xi_l)$ is given by (7) and (8), respectively, and the centering distributions $G_0^{\sigma},G_0^{\xi},G_0^{\kappa},G_0^{\tau}$ are defined in Section 3.2. Since no direct sampler is available for these distributions, we employ separate Gaussian random walk Metropolis steps on appropriately transformed versions of the parameters, that is, logarithmic transformations for σ_l, ξ_l and τ_l, and a logit transformation for κ_l. In all cases, the variances of the Gaussian proposals were tuned to obtain acceptance rates of around 20% to 30%.

References

Coles, S., Tawn, J., and Smith, R. L. (1994), "A seasonal Markov model for extremely low temperatures," *Environmetrics*, 5, 221–239.

Coles, S. G. (2001), *An introduction to statistical modeling of extreme values*, Springer-Verlag, New York.

Coles, S. G. and Tawn, J. (1996), "A Bayesian analysis of extreme rainfall data," *Applied Statistics*, 45, 463–478.

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

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

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1999), *Modeling extremal events for insurance and finance*, no. 369, Berlin: Springer-Verlag.

Escobar, M. and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian analysis of some nonparametric problems," *The Annals of Statistics*, 1, 209–230.

Fisher, R. A. and Tippett, L. H. C. (1928), "Limiting forms of the frequency distributions of the largest or smallest member of a sample." *Proceedings of the Cambridge Philosophical Society*, 24, 180–190.

Gelfand, A. E., Kottas, A. E., and MacEachern, S. N. (2005), "Bayesian nonparametric spatial modeling with Dirichlet process mixing." *Journal of the American Statistical Association*, 100, 1021–1035.

Genton, M. G., Ma, Y., and Sang, H. (2011), "On the likelihood function of Gaussian max-stable processes," *Biometrika*, **98**, 481–488.

Gnedenko, B. V. (1943), "Sur la distribution limite du terme maximum d'une série aléatoire," *Ann. Math.*, 44, 423–453.

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

Ishwaran, H. and James, L. F. (2001), "Gibbs sampling methods for stick-breaking priors." *Journal of the American Statistical Association*, 96, 161–173.

Ishwaran, H. and Zarepour, M. (2000), "Markov Chain Monte Carlo in approximate Dirichlet and Beta two-parameter process hierarchical models." *Biometrika*, 87, 371–390.

Ji, C., Merl, D., Kepler, T. B., and West, M. (2009), "Spatial mixture modelling for unobserved point processes: Examples in immunofluorescence histology." *Bayesian Analysis*, 4, 297–316.

Joe, H., Smith, R. L., and Weissman, I. (1992), "Bivariate threshold methods for extremes," *Journal of the Royal Statistical Society, Series B*, 54, 171–183.

Kottas, A. and Behseta, S. (2010), "Bayesian Nonparametric Modeling for Comparison of Single-Neuron Firing Intensities." *Biometrics*, 66, 277–286.

Kottas, A., Behseta, S., Moorman, D. E., Poynor, V., and Olson, C. R. (2012), "Bayesian nonparametric analysis of neuronal intensity rates." *Journal of Neuroscience Methods*, 203, 241–253.

Kottas, A., Duan, J. A., and Gelfand, A. E. (2008), "Modeling disease incidence data with spatial and spatio-temporal Dirichlet process mixtures." *Biometrical Journal*, 50, 29–42.

Kottas, A. and Sansó, B. (2007), "Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis." *Jounal of Statistical Planning and Inference*, 137, 3151–3163.

Kotz, S. and Nadarajah, S. (2000), *Extreme Value Distributions - Theory and Applications*, Imperial College Press.

Lo, A. (1984), "On a Class of Bayesian Nonparametric Estimates: I. Density Estimates"," *The Annals of Statistics*, 12, 351–357.

Padoan, S. A., Ribatet, M., and Sisson, S. A. (2010), "Likelihood-based inference for max-stable processes," *Journal of the American Statistical Association*, **105**, 263–277.

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

— (1975), "Statistical inference using extreme order statistics." *The Annals of Statistics*, 3, 119–131.

Rodriguez, A. and Dunson, D. B. (2011), "Nonparametric Bayesian models through probit stick-breaking processes." *Bayesian Analysis*, **6**, 145–178.

Sang, H. and Gelfand, A. E. (2009), "Hierarchical modeling for extreme values observed over space and time," *Environmental and Ecological Statistics*, 16, 407–426.

Schlather, M. (2002), "Models for stationary max-stable random fields," *Extremes*, **5**, 33–44.

Sethuraman, J. (1994), "A constructive definition of Dirichlet priors," *Statistica Sinica*, 4, 639–650.

Smith, R. (1989), "Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone," *Statistical Science*, 4, 367–393.

Smith, R. L. (1990), "Max-stable processes and spatial extremes," Tech. rep., University of North Carolina at Chapel Hill.

Tressou, J. (2008), "Bayesian nonparametrics for heavy tailed distribution. Application to food risk assessment," *Bayesian Analysis*, 3, 367–392.

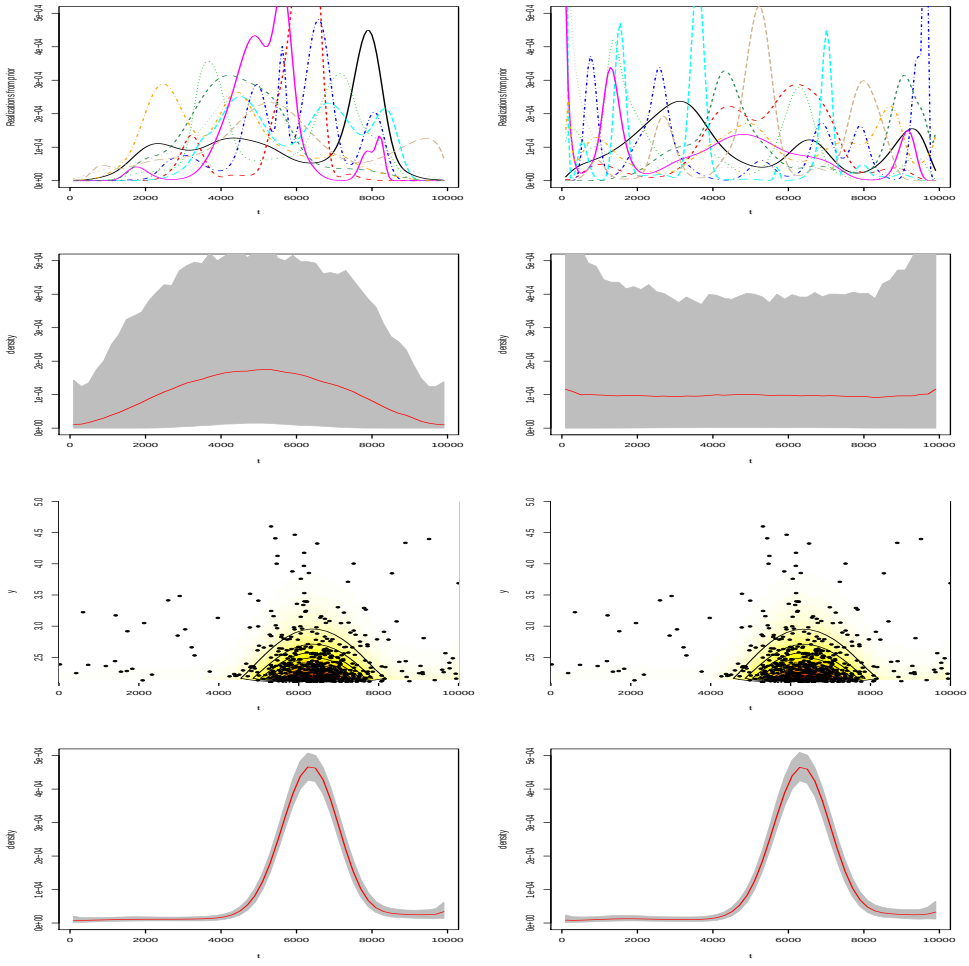
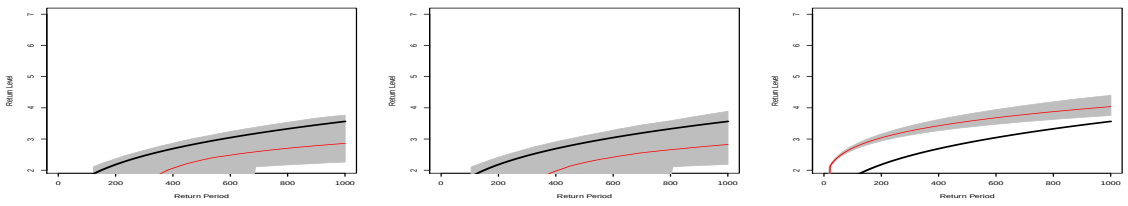
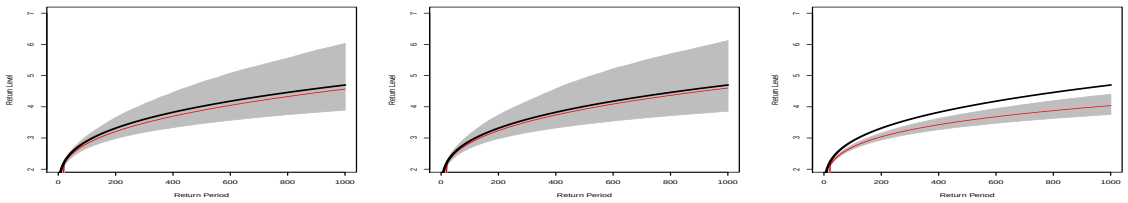


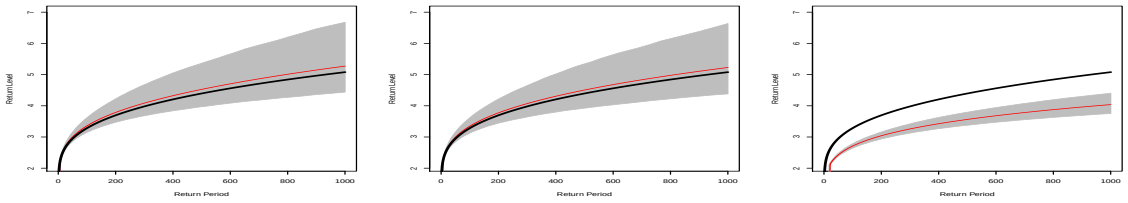
Figure 1: Simulation study. From top to bottom, 10 prior realizations of the marginal density of exceedance times; prior mean (red line) and 95% intervals (grey bands) of the marginal density of exceedance times; posterior mean for the bivariate intensity function; and posterior mean and 95% intervals for the marginal density of exceedance times. The left column corresponds to the prior choice involving a beta distribution for κ/T with mean 0.5 and variance $1/28$, and the right column to the prior based on a uniform distribution for κ/T.



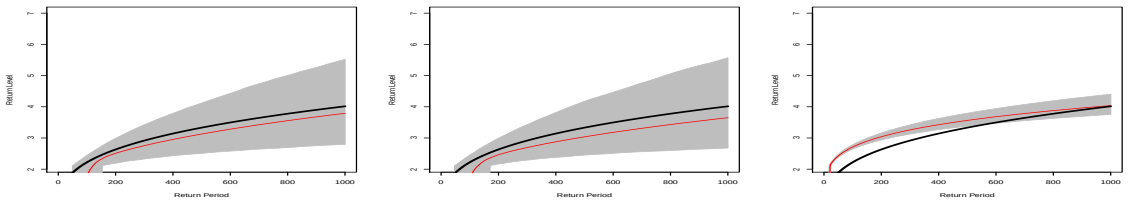
(a) $t_0 = 500$



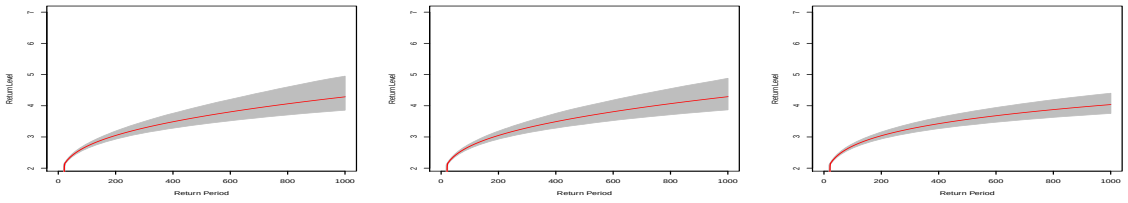
(b) $t_0 = 5100$



(c) $t_0 = 6500$



(d) $t_0 = 9500$



(e) Average Return Levels

Figure 2: Simulation study. The top four rows plot the posterior mean (red line) and 95% intervals (grey bands), and the true 1000-observation return level (black line) at four time points. The bottom row includes the posterior mean and 95% interval estimates of the marginal return levels. Results are shown for the nonparametric model under the beta distribution for κ/T with mean 0.5 and variance $1/28$ (left column), the nonparametric model under the uniform distribution for κ/T (middle column), and the parametric model (right column).

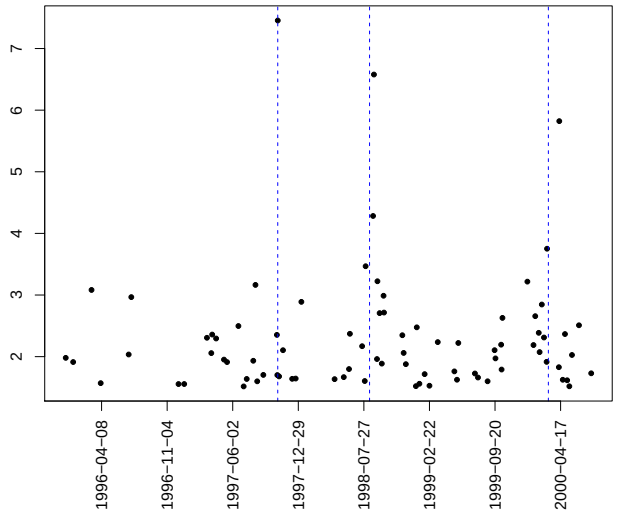


Figure 3: Dow Jones data. Plot of the extreme negative log returns for the Dow Jones index above threshold $u = 1.5$, from September 11, 1995 to September 7, 2000. The blue dashed lines indicate the time points of three financial crises: the mini-crash on October 27, 1997, the Russian financial crisis on August 17, 1998, and the bursting of the dot-com bubble on March 10, 2000, respectively.

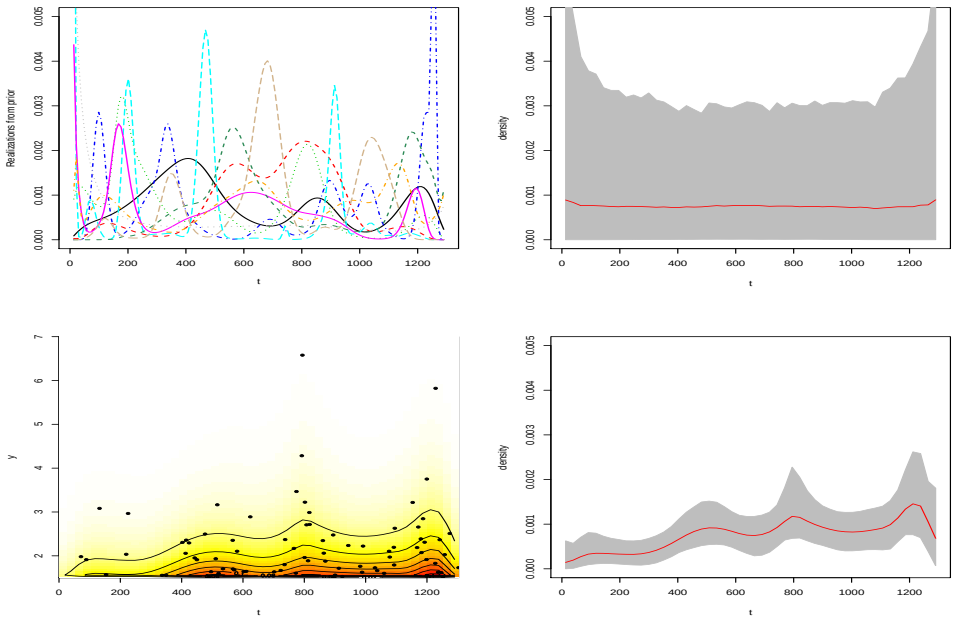
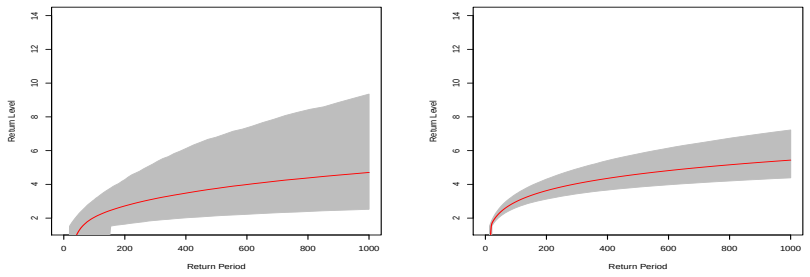
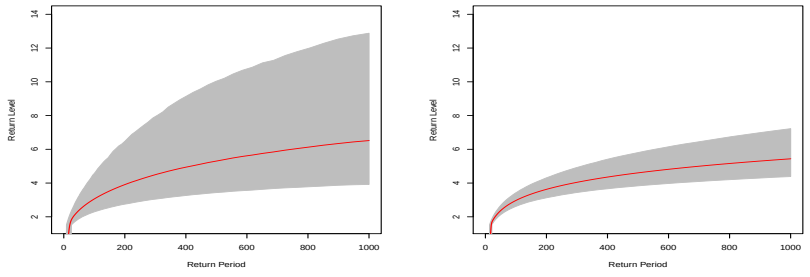


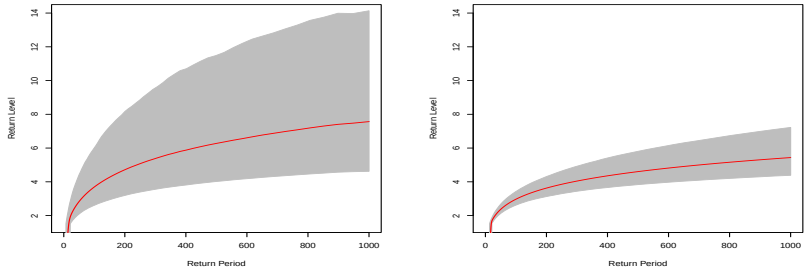
Figure 4: Dow Jones data. Clockwise from top left, 10 prior realizations of the marginal density of exceedance times; prior mean (red line) and 95% intervals (grey bands) of the marginal density of exceedance times; posterior mean and 95% interval estimates for the marginal density of exceedance times; and posterior mean of the bivariate intensity function.



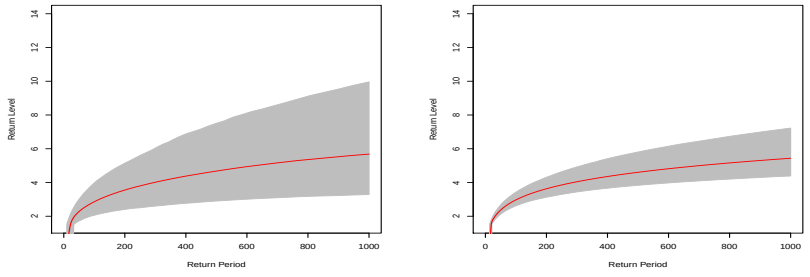
(a) September 26, 1996 (b) September 26, 1996



(c) October 27, 1997 (d) October 27, 1997



(e) August 17, 1998 (f) August 17, 1998



(g) July 20, 1999 (h) July 20, 1999

Figure 5: Dow Jones data. The posterior mean (red line) and 95% interval estimates (grey bands) of the 1000-day conditional return level curves at four different dates under the nonparametric model (left column) and the parametric model (right column).

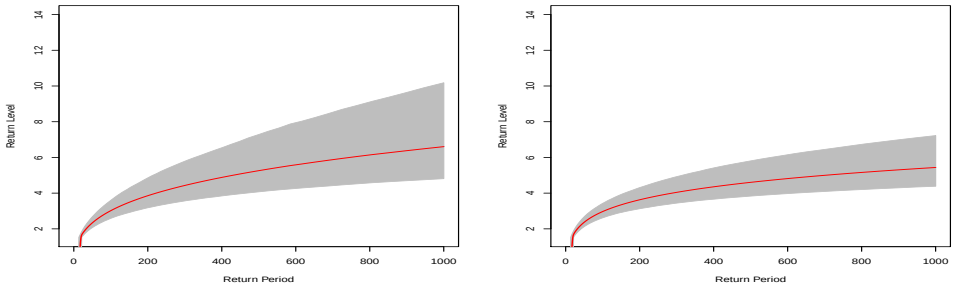


Figure 6: Dow Jones data. The posterior mean (red line) and 95% interval estimates (grey bands) of the marginal 1000-day return level curve under the nonparametric model (left panel) and the parametric model (right panel).

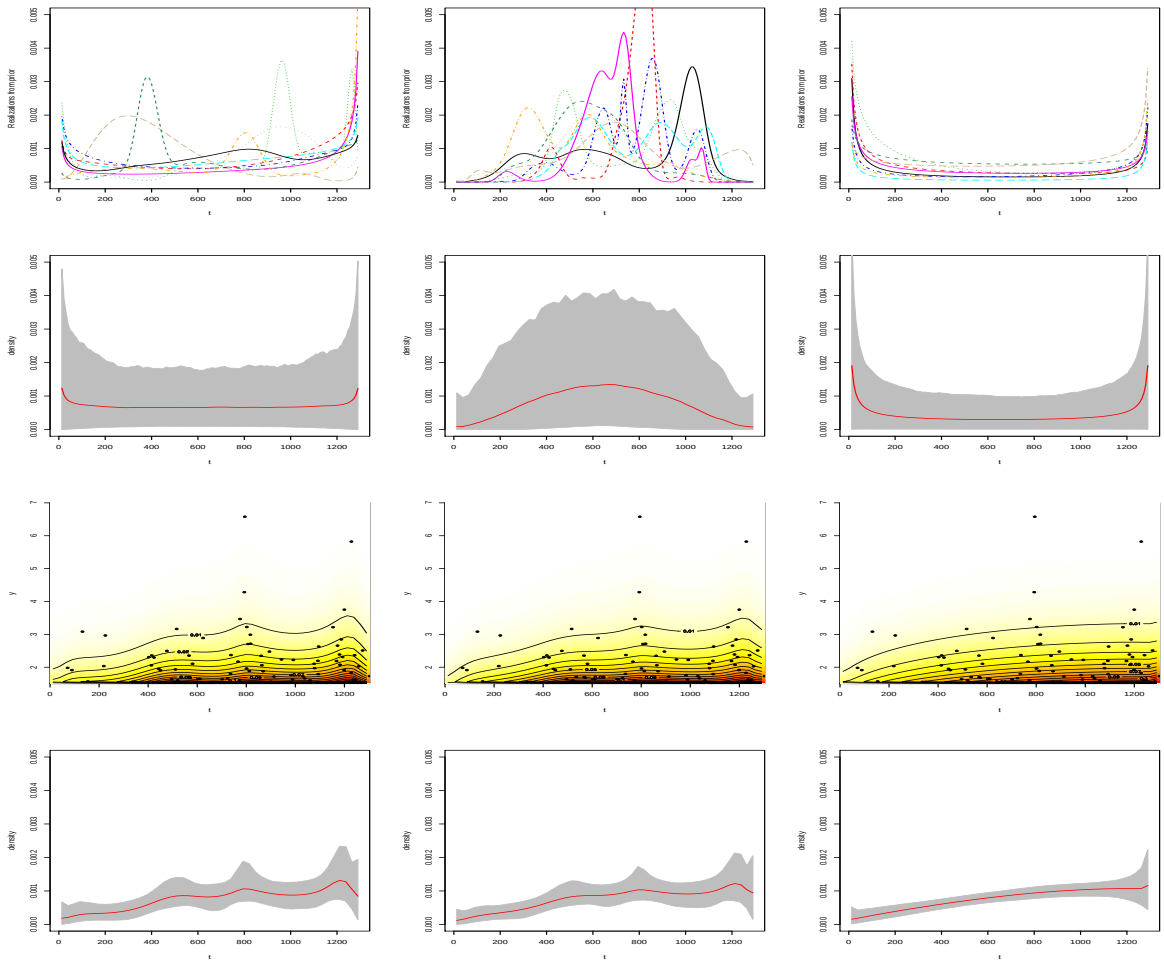
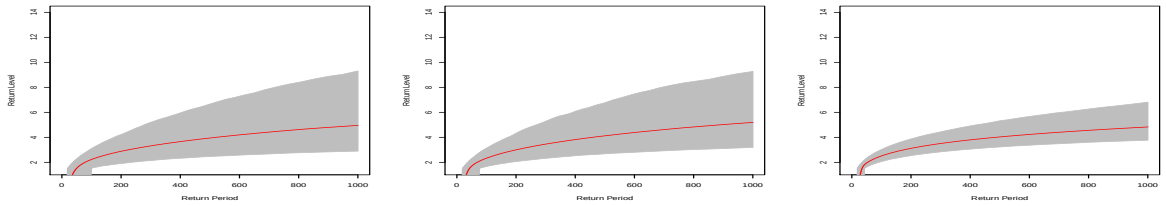
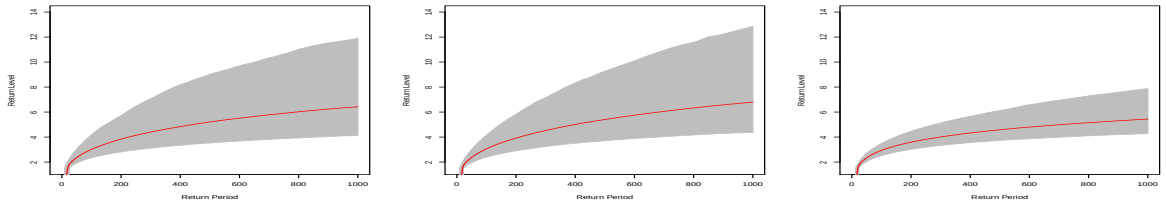


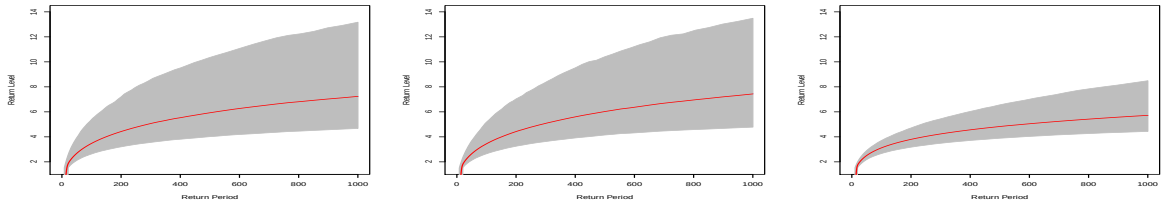
Figure 7: Prior sensitivity analysis results for the Dow Jones data; see Section 4.2 for details about the three priors corresponding to the columns of the figure. From top to bottom, 10 prior realizations for the marginal density of exceedance times; prior mean (red line) and 95% interval estimates (grey bands) of the marginal density of exceedance times; posterior mean of the bivariate intensity function; and posterior mean and 95% interval estimates of the marginal density of exceedance times.



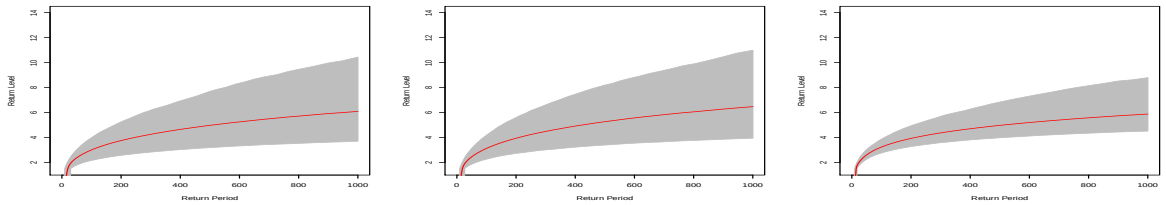
(a) September 26, 1996



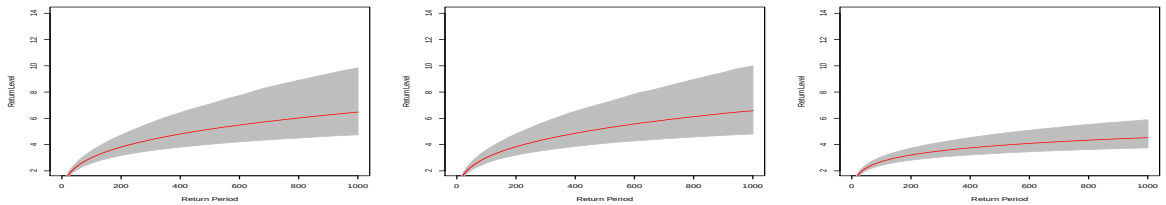
(b) October 27, 1997



(c) August 17, 1998



(d) July 20, 1999



(e) Marginal Return Levels

Figure 8: Prior sensitivity analysis results for the Dow Jones data; see Section 4.2 for details about the three priors corresponding to the columns of the figure. The top four rows include the posterior mean (red line) and 95% interval estimates (grey bands) of the 1000-day conditional return level curves at four different dates. The bottom row plots the corresponding estimates for the 1000-day marginal return level curve.