Modeling the Frequency and Magnitude of Extreme Events

Piyas Chakraborty [*] Hao Zhang [†]

Abstract

In this work, we extend the Poisson point process approach to modeling extreme values to separately model the frequency and the magnitude of extreme events. We assume that extremes occur over time according to an inhomogeneous Poisson process with a parametric intensity function to be estimated. This intensity reflects the frequency and the magnitude of extreme events. We demonstrate through an earthquake dataset that the new model may significantly outperform the common existing models for extreme values.

Key Words: Extreme values, Block maxima, Generalized extreme value distribution, Threshold exceedances, Poisson process, Intensity function

1. Introduction

An extreme event by definition occurs with a small probability, and is often represented by a maximum or minimum. For example, a record high temperature in a region would be the maximum temperature the region ever experienced whereas an extreme high temperature would be a temperature that is expected to occur once in many years. Probability theory has been well established for the maximum of a set of independent and identically distributed (i.i.d.) random variables. Let $X_1, X_2, \ldots, X_n$ be a set of i.i.d. random variables, which for example, may represent daily temperatures. It is well known that the maximum $M_n = \max\{X_1, \ldots, X_n\}$ has approximately a Generalized Extreme Value (GEV) distribution for sufficiently large n. The conclusion still holds if the variables X_i's are identically distributed but dependent. That is why GEV distributions have been applied to model extreme events and various inferential methods have been developed, which will be briefly reviewed in the next section. We refer to Coles [2001] and de Haan and Ferreira [2006] for introduction to extreme value theory and methods.

We usually describe extreme events over time in terms of both magnitude and frequency. In situations where there is a trend, the magnitude and frequency are dynamic rather than static. For example, the IPCC 2013 report projected that the weather would become more extreme in both magnitude and frequency: "It is virtually certain that there will be more hot and fewer cold extremes as global temperature increases" [Stocker et al., 2013, p.1065]. While it is often true that extremes events become more extreme in magnitude and more frequent, it may also be possible that they become less frequent with elevated magnitudes. The latter requires new statistical models as most of the existing models imply possible correlation between the magnitude and frequency of extreme events.

The objective of this paper is to develop a new statistical approach to model the magnitude and frequency separately. Our approach extends the point process approach to extreme events. Before we describe that approach in detail, we start with a brief overview of the existing literature on the analysis of extreme values.

[*]Department of Statistics, Purdue University, 250 N. University Street, West Lafayette, IN 47906

[†]Department of Statistics, Purdue University, 250 N. University Street, West Lafayette, IN 47906

2. The Point Process Model for extremes

Point processes provide an elegant way of modeling extreme events. Several estimation methods can be derived from this point process approach [Smith, 1989, Coles, 2001]. We now provide a brief review of different approaches leading to the point process approach and will generalize it later in this section.

2.1 Threshold exceedances

Let $X_1, X_2, \ldots$ be a series of i.i.d. random variables, with common distribution function F. Observing $X_1, \ldots, X_m$, we are interested in estimating the distribution of the n-variable maximum, $M_n = \max\{X_1, \ldots, X_n\}$, where $m = nn_p$ with $m =$ Total number of observations, $n_p =$ Total number of usual time periods (blocks) in the whole observation period and $n =$ Number of observations in each time period (block size). For simplicity, we first assume the number of periods n_p is 1 so that $m = n$. It is well-known that under mild regularity conditions on F, there exist normalizing sequences $a_n > 0$ and b_n such that $(M_n - b_n)/a_n$ converges in distribution to a nondegenerate distribution $G(x)$, i.e.,

$$P\{(M_n - b_n)/a_n \leq x\} = F^n(a_n x + b_n) \to G(x; \xi, \mu, \sigma)$$

where $G(x)$ must be the Generalized Extreme Value (GEV) distribution

$$G(x; \xi, \mu, \sigma) = \exp\{-[1 + \xi(x - \mu)/\sigma]_+^{-1/\xi}\}$$

for some parameters $\mu, \sigma > 0$ and ξ [Fisher and Tippett, 1928, Gnedenko, 1943]. Hereafter, $[x]_+$ denotes $\max\{0, x\}$. Clearly if such normalizing sequences a_n and b_n exist, they can be chosen so that the limiting GEV distribution has $\mu = 0$ and $\sigma = 1$, i.e.,

$$F^n(a_n x + b_n) \to G(x; \xi, 0, 1). \tag{1}$$

It follows that

$$P(M_n \leq z) = F^n(z) \approx G(z; \xi, b_n, a_n),$$

which says that the maximum M_n has approximately a GEV distribution with location parameter $\mu = b_n$, a scale parameter $\sigma = a_n$, and the shape parameter ξ. The parameters μ, σ and ξ can be estimated by maximizing the likelihood of the block maxima when the observations are divided into blocks each of which has length n. However, in order to improve the efficiency of estimates, estimation methods have been developed that make use of individual values x_i instead of the block maxima. For example, one such method exploits the exceedance over threshold, which is based on the theoretical result [de Haan and Ferreira, 2006] that (1) is equivalent to

$$\lim_{t \to \infty} t[1 - F(a(t)z + b(t))] = -\log G(z)$$

for any continuity point z of G such that $0 < G(z) < 1$, $a(t) = a_{[t]}, b(t) = b_{[t]}$ are step functions, $[t]$ represents the integer part of t. It follows that

$$P(X_i \geq x) = 1 - F(x) \approx \frac{1}{n} \log G\left((x - b_n)/a_n\right) = \frac{1}{n}\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-1/\xi}$$

where $\mu = b_n$ and $\sigma = a_n$ as before, which implies that for a large threshold u,

$$P(X_i > x | X_i > u) \approx \left(1 + \xi \frac{x - u}{\tilde{\sigma}}\right)_+^{-1/\xi}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$.

Therefore the excess over the threshold follows a generalized Pareto distribution with properly chosen parameters parameters $\xi, \mu = 0$ and $\sigma = \tilde{\sigma}$ [Pickands, 1975]. This property provides a method for estimating these parameters, which is called the threshold excess method [Smith, 1989]. Observing $X_1, \ldots, X_n$, we can use the excesses over the threshold as well as the censored values below the threshold to maximize the corresponding likelihood function from the generalized Pareto distribution to estimate the parameters.

2.2 Poisson process representation

A more elegant approach is to exploit the Poisson point process. Consider a point processes $\boldsymbol{N_n}$ on $[0, 1] \times (-\infty, \infty)$, $\boldsymbol{N_n} = \{(i/n, X_i) : i = 1, \cdots, n\}$. Then, for sufficiently large u, on regions of the form $(0, 1) \times [u, \infty)$, $\boldsymbol{N_n}$ converges to a Poisson process with intensity measure

$$\Lambda(A) = (t_2 - t_1) \left[1 + \xi(x - \mu)/\sigma\right]_+^{-1/\xi}$$

for any $A = (t_1, t_2) \times [x, \infty)$, $x > u$, where μ, σ and ξ are the parameters in the GEV distribution of $\max\{X_1, \ldots, X_n\}$ [Coles, 2001, Theorem 7.1.1].

The intensity function can be obtained as

$$\lambda(t, x) = -\frac{\partial^2}{\partial t \partial x} \Lambda[(s, t) \times [x, \infty)] = \frac{1}{\sigma} \left[1 + \xi(x - \mu)/\sigma\right]_+^{-1/\xi - 1}.$$

When we observe n_p periods (such as years) each of length n and we want to estimate the GEV distribution of the maxima of blocks of length n, we view $\{(i/n, X_i) : i = 1, \cdots, nn_p\}$ as a Poisson point process on region $(0, n_p) \times [u, \infty)$ with the intensity defined above. Denote by $x_{ij}, j = 1, \ldots, N_i$ the N_i sample values that exceed the threshold in the i^{th} period. The likelihood function of the Poisson point process is

$$L(\mu, \sigma, \xi) = \exp\left\{-n_p \left[1 + \xi \frac{u - \mu}{\sigma}\right]_+^{-1/\xi}\right\} \prod_{i=1}^{n_p} \prod_{j=1}^{N_i} \frac{1}{\sigma} \left[1 + \xi \left(\frac{x_{ij} - \mu}{\sigma}\right)\right]_+^{-1/\xi - 1}. \quad (2)$$

Parameters can be estimated by maximizing this likelihood. Other common approaches to analyze extreme values, e.g. the block maxima model and the threshold excess model, can be derived from this representation.

2.3 Nonstationarity

There are situations where the block maxima may not have identical distributions but present some trend. For example, in the presence of global warming, annual maximum temperatures might have increasing means. Then it is interesting to see how the maximum of X_i's vary over time. One common approach is to assume the parameters μ, σ and ξ for each block maximum change over time. For example, these parameters may be polynomial functions of the time t, and the coefficients can be estimated by maximizing (2). If we write the parameters of GEV distribution of the ith block maximum as

$$\mu_i = \mu(i, \theta),\ \xi_i = \xi(i, \theta),\ \sigma_i = \sigma(i, \theta),$$

Then θ can be estimated by maximizing the following likelihood function [Smith, 1989]

$$L(\theta) = \prod_{i=1}^{n_p} \exp\left\{-\left[1 + \xi_i \left(\frac{u - \mu_i}{\sigma_i}\right)\right]_+^{-1/\xi_i}\right\} \prod_{j=1}^{N_i} \frac{1}{\sigma_i} \left[1 + \xi_i \left(\frac{x_{ij} - \mu_i}{\sigma_i}\right)\right]_+^{-1/\xi_i - 1}. \quad (3)$$

This method has been applied to many studies on extremes [Jaruskova, 2009, Acero et al., 2011, Mannshardt-Shamseldin et al., 2010]

2.4 Our model: introducing nonuniformity in frequency

One implication of the above model formulations is that the marginal intensity along the time domain depends on the time component through the GEV parameters only. As a result the frequency of extreme events over time ends up having a similar pattern with the magnitude of extreme events which are governed by the GEV parameters. Moreover the marginal intensity along the time domain is a constant when the GEV parameters are taken to be constants over time. Equivalently, this point process then implies that each block maxima exceeds the threshold u with equal probability.

Hence the usual model for nonstationarity suggests that extremes occur with the same frequency over time, or at the most with a frequency that follows a similar pattern as the magnitudes, which may not be always true. For example, extreme temperatures seems to occur with increasing frequency and elevated magnitude over the years but in case of extreme earthquakes more extreme magnitudes imply reduced frequency and vice versa.

Technically, one could choose a time varying threshold $u(t)$ so that the exceedance over the threshold occurs with a constant probability over time. Such a function of threshold $u(t)$ is hard to choose except in some special cases, e.g., with seasonal patterns. In this work, we retain the constant threshold and propose a new point process model to accommodate the time varying probability of exceedance.

In the new model, we assume that the two-dimensional point processes $\boldsymbol{N_n} = \{(t_{ij}, X_{ij}) : i = 1, \cdots, n_p; j = 1, \cdots, N_i\}$ is an inhomogeneous Poisson point process on $(0, 1] \times [u, \infty)$ for some large threshold u with an intensity function

$$\lambda'(t, x) = n_p \lambda(t) \frac{1}{\sigma(t)} \left[1 + \xi(t) \frac{x - \mu(t)}{\sigma(t)}\right]_+^{-1/\xi(t)-1}$$

where the functions $\mu(t)$, $\xi(t)$ and $\sigma(t)$ are step functions as before and remain constants on intervals $t \in (\frac{i-1}{n_p}, \frac{i}{n_p}], i = 1, \ldots, n_p$. Specifically, for $t \in (\frac{i-1}{n_p}, \frac{i}{n_p}]$,

$$\mu(t) = \mu_i, \xi(t) = \xi_i, \text{ and } \sigma(t) = \sigma_i.$$

In addition, we assume that $\mu_i = \mu(i, \theta)$ for some parametric function $\mu(\cdot, \theta)$ and make similar assumptions on ξ_i and σ_i. Hence this intensity function is a natural extension of the one in the previous subsection. We are essentially introducing a multiplicative exceedance rate $\lambda(\cdot)$ to the bivariate intensity function to model the frequency of the exceedances.

Note that for most examples in this paper, we assume a quadratic parametric form for λ, namely $\lambda(t) = at^2 + bt + c, t \in (0, 1]$ and since we need to have the constraint $\int_0^1 \lambda(t)dt = 1$ i.e. $\lambda(t)$ is a density on $(0, 1]$ to have identifiability of the parameters, only 2 parameters out of a, b and c can be considered to be free parameters. Moreover the constant multiplier n_p with the density $\lambda(\cdot)$ ensures that the parameters $(\mu(t), \sigma(t), \xi(t))$ correspond to the approximate GEV distribution of the block (annual) maxima at any time point t. The likelihood function now becomes

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n_p} \exp\left\{-n_p \int_{\frac{i-1}{n_p}}^{\frac{i}{n_p}} \lambda(t)dt \left[1 + \xi_i \left(\frac{u - \mu_i}{\sigma_i}\right)\right]_+^{-1/\xi_i}\right\} \\ &\quad \times \prod_{j=1}^{N_i} \frac{\lambda(t_{ij})}{\sigma_i} \left[1 + \xi_i \left(\frac{x_{ij} - \mu_i}{\sigma_i}\right)\right]_+^{-1/\xi_i - 1}. \end{aligned} \tag{4}$$

When $\lambda \equiv 1$ for all i (i.e., exceedance frequency remains a constant over time), the likelihood in (4) becomes (3). In the next section, we will consider examples where ξ_i and σ_i are constants (not depending on i) and

$$\mu_i = \beta_0 + \beta_1 \frac{i}{n_p} + \beta_2 \left(\frac{i}{n_p}\right)^2, i = 1, \ldots, n_p.$$

The complete vector of parameters is $\theta = (\beta_0, \beta_1, \beta_2, a, b, \sigma, \xi)$. $\beta_0, \beta_1, \beta_2, \xi$ and σ model the fluctuation in the magnitudes of threshold exceedances whereas a and b model the fluctuation in the frequencies of threshold exceedances.

From now on, we will drop the subscript '+' from the expressions with the tacit assumption that if the quantities within the square brackets are negative, they are replaced by zero.

Let us assume that $N(A)$ is the number of observations from the Poisson process for any $A = (t_1, t_2) \times [x, \infty)$ with $x > u$ so that $N(A)$ has a Poisson distribution with mean $\Lambda(A)$. Then for the process X observed at time $t \in (0, 1)$, we can write:

$$\begin{aligned}
P(X(t) > x + y | X(t) > x) &= \frac{P(X(t) > x + y)}{P(X(t) > x)} \\
&= \lim_{\Delta t \to 0} \frac{1 - P(N((t - \Delta t, t + \Delta t) \times (x + y, \infty)) = 0)}{1 - P(N((t - \Delta t, t + \Delta t) \times (x, \infty)) = 0)} \\
&= \lim_{\Delta t \to 0} \frac{1 - e^{-\Lambda((t-\Delta t, t+\Delta t) \times (x+y, \infty)))}}{1 - e^{-\Lambda((t-\Delta t, t+\Delta t) \times (x, \infty))}} \\
&= \lim_{\Delta t \to 0} \frac{\Lambda((t - \Delta t, t + \Delta t) \times (x + y, \infty)))}{\Lambda((t - \Delta t, t + \Delta t) \times (x, \infty))} \\
&= \lim_{\Delta t \to 0} \frac{\int_{t-\Delta t}^{t+\Delta t} \int_{x+y}^{\infty} \lambda'(t', x') dx' dt'}{\int_{t-\Delta t}^{t+\Delta t} \int_{x}^{\infty} \lambda'(t', x') dx' dt'} \\
&= \lim_{\Delta t \to 0} \frac{n_p\{2\Delta t \lambda(t)\} \left[1 + \xi(t) \frac{x+y-\mu(t)}{\sigma(t)}\right]^{-1/\xi(t)}}{n_p\{2\Delta t \lambda(t)\} \left[1 + \xi(t) \frac{x-\mu(t)}{\sigma(t)}\right]^{-1/\xi(t)}} \\
&= \lim_{\Delta t \to 0} \left[\frac{1 + \xi(t) \frac{x+y-\mu(t)}{\sigma(t)}}{1 + \xi(t) \frac{x-\mu(t)}{\sigma(t)}}\right]^{-1/\xi(t)} \\
&= \left[1 + \xi(t) \frac{y/\sigma(t)}{1 + \xi(t) \frac{x-\mu(t)}{\sigma(t)}}\right]^{-1/\xi(t)} \\
&= \left[1 + \xi(t) \frac{y}{\tilde{\sigma}(t)}\right]^{-1/\xi(t)}
\end{aligned}$$

with $\tilde{\sigma}(t) = \sigma(t) + \xi(t)(x - \mu(t))$ where $x \geq u$ and $y > 0$ are such that all the terms within the square brackets in the expressions above are positive. The conditional probability is similar to the one in the case of uniform frequency over time that we have already discussed. This is intuitive as the above expression depends only on the magnitude of the process $X(t)$ and not on its frequency and hence should be free of $\lambda(t)$.

2.4.1 Calculation of Return Levels

We can use similar calculations to find the return levels of the process $X(t)$ across time t. Return levels or quantiles are usually used to conveniently interpret extreme value models. The M-observation return level $z_M(t)$ at time t (usually for large M) is defined by:

$$P(X(t) > z_M(t)) = \frac{1}{M}.$$

In order to avoid singularity, we replace the above definition with:

$$P(N((t-\Delta t, t+\Delta t) \times (z_M(t), \infty)) = 0) = 1 - \frac{1}{M}$$

with a sufficiently small Δt so that we can assume that the probability of 2 or more observations on $(t-\Delta t, t+\Delta t) \times [u, \infty)$ is negligible which is an usual assumption about Poisson processes. Moreover we usually decluster the observations by retaining only the most extreme observations from each of the clusters of observations along the time domain. Indeed for all of our examples we have predefined the cluster width as 1 day and only retained the daily extrema to form the observed data points. Hence for such examples $2\Delta(t)$ may be taken as 1 day on the re-scaled time axis. Also, $\mu(t)$, $\xi(t)$ and $\sigma(t)$ are step functions so they can be assumed to be constants on $(t-\Delta t, t+\Delta t)$ for suitably small Δt which will lead to easier computations.

Now from the above equation, we get:

$$\begin{aligned}
& e^{-\Lambda((t-\Delta t, t+\Delta t)\times(z_M(t),\infty))} = 1 - \frac{1}{M} \\
& \Leftrightarrow \int_{t-\Delta t}^{t+\Delta t} \int_{z_M(t)}^{\infty} \lambda'(t', x')dt'dx' = -\log(1 - \frac{1}{M}) \\
& \Leftrightarrow n_p \int_{t-\Delta t}^{t+\Delta t} \lambda(t') \int_{z_M(t)}^{\infty} \frac{1}{\sigma(t)} \left[1 + \xi(t)\frac{x' - \mu(t)}{\sigma(t)}\right]^{-1/\xi(t)-1} dx'dt' = -\log(1 - \frac{1}{M}) \\
& \Leftrightarrow n_p \{ \int_{t-\Delta t}^{t+\Delta t} \lambda(t')dt' \} \left[1 + \xi(t)\frac{z_M(t) - \mu(t)}{\sigma(t)}\right]^{-1/\xi(t)} = \log(\frac{M}{M-1}) \\
& \Leftrightarrow z_M(t) = \mu(t) + \frac{\sigma(t)}{\xi(t)} \left(\phi(t)^{\xi(t)} - 1\right)
\end{aligned}$$

where $\phi(t) = n_p \frac{\int_{t-\Delta t}^{t+\Delta t} \lambda(t')dt'}{\log(\frac{M}{M-1})} \approx n_p M\{2\Delta t \lambda(t)\}$ if M is sufficiently large and Δt is sufficiently small. Moreover if $2\Delta(t)$ is taken as 1 day where data has been collected for m days, then $\phi(t)$ can be further approximated as $\phi(t) \approx n_p M \frac{\lambda(t)}{m}$.

One problem with the above computation is that often the density $\lambda(t)$ becomes 0 or very close to 0 at or around some $t \in (0, 1)$ for some data. Hence $\int_{t-\Delta t}^{t+\Delta t} \lambda(t')dt' \approx 0$ for such t leading to numerical instabilities in the above computation. In such cases, we can calculate return levels for wider intervals so that the integral of the density $\lambda(t)$ over such intervals are not too close to 0.

Since $\mu(t), \sigma(t)$ and $\xi(t)$ are usually assumed to be step functions which are constants on pre-defined time periods or blocks (eg. years), i.e. on intervals $t \in (\frac{i-1}{n_p}, \frac{i}{n_p}], i = 1, \cdots, n_p$, a natural option is to calculate return levels for each time period or block (e.g. yearly return levels). When each block is fixed as 1 year (as in all of our examples), the M-year return level $y_M(i)$ for the i^{th} year is defined by:

$$P\left(X(t) \leq y_M(i); t \in (\frac{i-1}{n_p}, \frac{i}{n_p}]\right) = 1 - \frac{1}{M}$$
$$\Leftrightarrow P\left(N((\frac{i-1}{n_p}, \frac{i}{n_p}] \times (y_M(i), \infty)) = 0\right) = 1 - \frac{1}{M}$$

Calculations similar to the above one will lead to:

$$y_M(i) = \mu_i + \frac{\sigma_i}{\xi_i}\left(\phi_i^{\xi_i} - 1\right)$$

where $\phi_i = n_p\{\log(\frac{M}{M-1})\}^{-1} \int_{\frac{i-1}{n_p}}^{\frac{i}{n_p}} \lambda(t)dt \approx n_p M \int_{\frac{i-1}{n_p}}^{\frac{i}{n_p}} \lambda(t)dt$, if M is sufficiently large. We expect the process $X(t)$ to exceed these M-year return levels, i.e. the step function $\tilde{z}_M(t) = y_M(i)$, if $t \in (\frac{i-1}{n_p}, \frac{i}{n_p}], i = 1, \cdots, n_p$ once in every M years on an average.

Sometimes the calculation for blockwise return levels can also run into numerical instabilities if $\lambda(t)$ is very close to zero on one or more blocks. In such cases, we may have to do the calculations for wider time intervals. For example, if we calculate a single return level over the entire time interval from the i^{th} block to the j^{th} one ($1 \leq i \leq j \leq n_p$), the corresponding return level $\tilde{y}_M(i,j)$ satisfies:

$$P\left(X(t) \leq \tilde{y}_M(i,j); t \in (\frac{i-1}{n_p}, \frac{j}{n_p}]\right) = 1 - \frac{1}{M}$$
$$\Leftrightarrow P\left(N((\frac{i-1}{n_p}, \frac{j}{n_p}] \times (\tilde{y}_M(i,j), \infty)) = 0\right) = 1 - \frac{1}{M}$$
$$\Leftrightarrow \int_{\frac{i-1}{n_p}}^{\frac{j}{n_p}} \int_{\tilde{y}_M(i,j)}^{\infty} \lambda'(t', x')dt'dx' = -\log(1 - \frac{1}{M})$$
$$\Leftrightarrow n_p \sum_{k=i}^{j} \int_{\frac{k-1}{n_p}}^{\frac{k}{n_p}} \lambda(t') \int_{\tilde{y}_M(i,j)}^{\infty} \frac{1}{\sigma(t)}\left[1 + \xi(t)\frac{x' - \mu(t)}{\sigma(t)}\right]^{-1/\xi(t)-1} dx'dt' = \log(\frac{M}{M-1})$$
$$\Leftrightarrow n_p \sum_{k=i}^{j} \int_{\frac{k-1}{n_p}}^{\frac{k}{n_p}} \lambda(t') \left(\int_{\tilde{y}_M(i,j)}^{\infty} \frac{1}{\sigma_k}\left[1 + \xi_k\frac{x' - \mu_k}{\sigma_k}\right]^{-1/\xi_k-1} dx'\right) dt' = \log(\frac{M}{M-1})$$
$$\Leftrightarrow \sum_{k=i}^{j} \{n_p \int_{\frac{k-1}{n_p}}^{\frac{k}{n_p}} \lambda(t')dt'\} \left[1 + \xi_k\frac{\tilde{y}_M(i,j) - \mu_k}{\sigma_k}\right]^{-1/\xi_k} - \log(\frac{M}{M-1}) = 0.$$

$\tilde{y}_M(i,j)$ is thus obtained by solving the following equation for x:

$$\sum_{k=i}^{j}\{n_p \int_{\frac{k-1}{n_p}}^{\frac{k}{n_p}} \lambda(t)dt\}\left[1+\xi_k\frac{x-\mu_k}{\sigma_k}\right]^{-1/\xi_k} - \log(\frac{M}{M-1}) = 0.$$

This doesn't have an easy closed form solution unless $i = j$ where we have a closed form solution, namely $\tilde{y}_M(i,i) = y_M(i)$ obtained earlier for $i = 1, \cdots, n_p$. This is easy to see from their definitions as well. When $j > i$, i.e. when we calculate return levels over two or more blocks, we can obtain $\tilde{y}_M(i,j)$ by solving the above equation by standard numerical methods for nonlinear equations. There shouldn't be numerical issues unless $\int_{\frac{k-1}{n_p}}^{\frac{k}{n_p}} \lambda(t)dt \approx 0$ for all $k = i, \cdots, j$ in which case we have to select an wider interval.

It is worth noting that if we divide the entire time domain into disjoint intervals comprising of k blocks ($k \geq 1$) each and calculate return levels for each interval as above, we expect the process $X(t)$ to exceed those return levels once in every kM blocks on an average. Choosing k to be too small (e.g. $k = 1$) may lead to numerical instabilities depending on the shape of $\lambda(\cdot)$. On the other hand, choosing k to be too large will lead to the loss of significant information about the local features pertaining to the time varying components of the model, especially $\lambda(\cdot)$. In particular, choosing $k = n_p$ will produce one single return level for the entire time domain by ignoring the fluctuations in the time varying parameters to some extent.

3. Methodology

As mentioned before, μ, σ and ξ can be assumed to be time-varying. In all our examples, only μ will be assumed to be time-dependent. As we have already discussed, μ is assumed to be constant over each usual time period or block (e.g. one month or year). Assuming that the data has been collected over n_p such time periods, the value of μ during the i^{th} time period is modeled as $\mu_i = \beta_0 + \beta_1 y_i + \beta_2 z_i$ for $i = 1, \cdots, n_p$. Here y_i's are obtained by standardizing the series $\{\frac{i}{n_p} : i = 1, \cdots, n_p\}$ and z_i's are obtained by standardizing the series $\{y_i^2 : i = 1, \cdots, n_p\}$ in order to get rid of multicollinearity. This is based on the following facts which are easy to prove.

If the two vectors $\mathbf{y}$ *and* $\mathbf{z}$ *are obtained by standardizing the series* $\{\frac{i}{T} : i = 1, \cdots, T\}$ *and* $\{\frac{i^2}{T^2} : i = 1, \cdots, T\}$ *respectively, then the correlation between* $\mathbf{y}$ *and* $\mathbf{z}$ *approaches* $\sqrt{\frac{15}{16}}$ *as* $T \to \infty$.

Hence the correlation is nearly perfect for a large enough dataset. On the other hand, we also have:

If $\mathbf{z}$ *is redefined as the standardized version of the series* $\{y_i^2 : i = 1, \cdots, T\}$ *instead, then the correlation between* $\mathbf{y}$ *and* $\mathbf{z}$ *is* 0 *for any value of* T.

We assume as before that the number of threshold exceedances within the i^{th} time period is N_i with the occurrence times (scaled) and exceedance magnitudes in that period being $\{(t_{ij}, x_{ij}) : j = 1, \cdots, N_i\}$. We model $\lambda(\cdot)$ as $\lambda(t) = at^2 + bt + c$ on $t \in (0, 1]$. Since $\lambda(\cdot)$ is a density, $\frac{a}{3} + \frac{b}{2} + c = 1$ and we consider only a and b to be free or independent parameters. c is automatically determined from a and b as $c = 1 - \frac{a}{3} - \frac{b}{2}$. We have also

tried out the Beta density to model $\lambda(\cdot)$ which is another 2-parameter family: $\lambda(t) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1}$ on $t \in (0,1]$. But compared to the Beta family, the quadratic family described above seem to produce better fit more often and perform much better in terms of computation almost always. Hence we will mostly work with the quadratic family for $\lambda(\cdot)$ in our examples.

So the likelihood for our model becomes

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n_p} \exp\{-n_p \int_{\frac{i-1}{n_p}}^{\frac{i}{n_p}} \lambda(t)dt \left[1+\xi(u-\mu_i)/\sigma\right]^{-1/\xi}\} \\ &\quad \times \prod_{j=1}^{N_i} \frac{1}{\sigma}\lambda(t_{ij}) \left[1+\xi(x_{ij}-\mu_i)/\sigma\right]^{-1/\xi-1}. \end{aligned}$$

Here μ_i's and $\lambda(\cdot)$ involve the parameters $\beta_0, \beta_1, \beta_2, a$ and b. Interestingly, if we assume $\lambda \equiv 1$, we have

$$L(\theta | a=0, b=0) = \prod_{i=1}^{n_p} \exp\{-\left[1+\xi(u-\mu_i)/\sigma\right]^{-1/\xi}\} \prod_{j=1}^{N_i} \frac{1}{\sigma}\left[1+\xi(x_{ij}-\mu_i)/\sigma\right]^{-1/\xi-1}$$

which is the conventional likelihood with only μ showing nonstationarity as discussed in the previous section.

On the other hand, if we assume μ to be a constant function of time t i.e. $\mu_i = \beta_0, \forall i$, we have

$$\begin{aligned} L(\theta | \beta_1 = 0, \beta_2 = 0) &= \exp\{-n_p \left[1+\xi(u-\beta_0)/\sigma\right]^{-1/\xi}\} \\ &\quad \times \prod_{i=1}^{n_p}\prod_{j=1}^{N_i} \frac{1}{\sigma}\lambda(t_{ij}) \left[1+\xi(x_{ij}-\beta_0)/\sigma\right]^{-1/\xi-1} \\ &= \left[\exp\{-n_p \left[1+\xi(u-\beta_0)/\sigma\right]^{-1/\xi}\}\right. \\ &\quad \left. \times \prod_{i=1}^{n_p}\prod_{j=1}^{N_i} \frac{1}{\sigma}\left[1+\xi(x_{ij}-\beta_0)/\sigma\right]^{-1/\xi-1}\right] \left[\prod_{i=1}^{n_p}\prod_{j=1}^{N_i} \lambda(t_{ij})\right] \end{aligned}$$

which is separable in terms of (μ, σ, ξ) and λ and can be maximized by maximizing each component within the square brackets separately. We will refer to this as the separable model.

We can maximize $\log L(\theta)$ using standard optimization techniques to obtain the maximum likelihood estimates of $\theta = (\beta_0, \beta_1, \beta_2, a, b, \sigma, \xi)$. To facilitate the optimization, we re-parameterize $\lambda(\cdot)$ in terms of $\lambda(0) = c$ and $\lambda(1) = a + b + c$. This is because $\lambda(\cdot)$ shows similar shape when a and b are changed keeping $a + b$ fixed. For example, Figure 1 shows that the densities with $a = 2, b = -2$ and $a = 4, b = -4$ have pretty similar shapes whereas the densities with $a = 2, b = -4$ and $a = 4, b = -2$ have widely different shapes.

As a result, parameterizing using a and b tends to produce ridges on the log-likelihood surface along different contours of $a + b = K$ for a varying constant K. These ridges are present even near the global maxima whereas re-parameterizing in terms of c and $a + b + c$ should produce a much more prominent peak. To illustrate this, let us look at Figure 2 that shows the profile log-likelihood surface of $\lambda(\cdot)$ for a maximum temperature data from China that we will discuss in the next section. The left panel shows the surface when we parameterize using a and b and the right one shows the surface when we re-parameterize

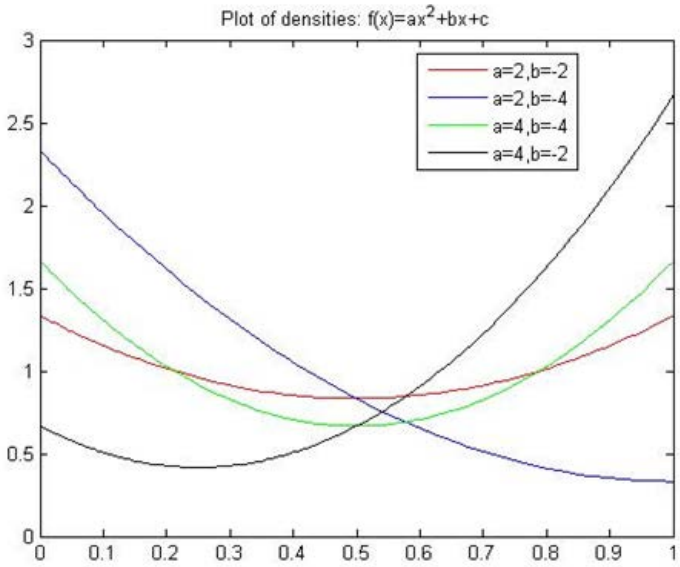

Figure 1: Shapes of quadratic $\lambda(\cdot)$ for different values of a and b

in terms of c and $a + b + c$. It is evident that the usual optimization techniques would get closer to the global maxima more quickly in the second case, in particular if the search is initiated far from the global maxima.

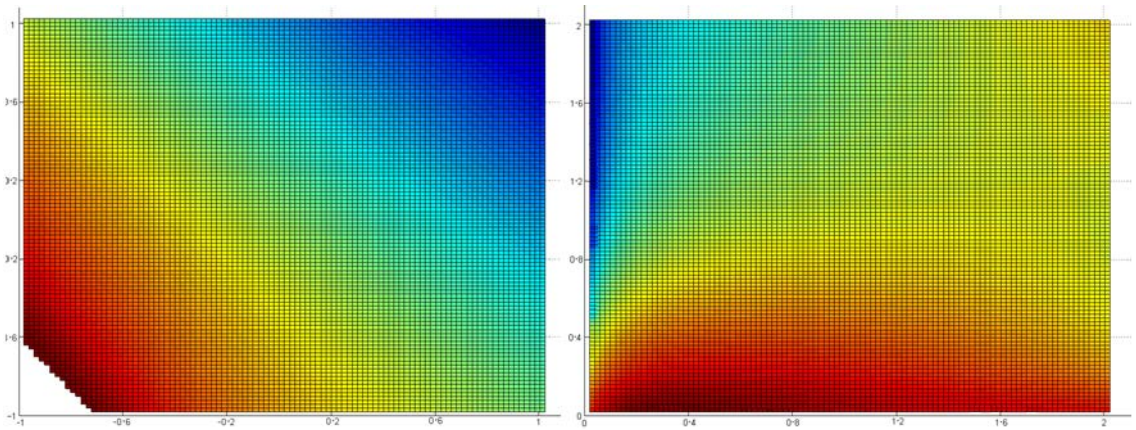

Figure 2: Contour plots of the profile log-likelihood for different parameterizations of $\lambda(\cdot)$

Upon obtaining the maximum likelihood estimates of all the parameters, we also need to check how well the model with the estimated parameters fits the data. To test the usefulness of introducing separate sets of parameters to describe μ and λ for any specific data, we use the usual likelihood ratio tests for $H_0 : \beta_1 = 0, \beta_2 = 0$ and $H_0 : a = 0, b = 0$ respectively. Note that the two hypotheses stipulate that μ and λ respectively are constant functions of time t.

To carry out these tests, we fit and compare four models: a model with constant $\lambda(t)$ and constant $\mu(t)$, a model with a constant $\lambda(t)$ and a quadratic $\mu(t)$, a model with a quadratic $\lambda(t)$ and a constant $\mu(t)$, and a model with a quadratic $\lambda(t)$ and a quadratic $\mu(t)$. While the first model is static as in (2) and not very useful for any of the datasets we have analyzed (because of the obvious nonstationarity present in them), the second one is the usual nonstationary model from (3) used extensively in literature. The third one

is the separable model which we have already described. The last one is our proposed nonstationary model from (4).

We also develop a goodness-of-fit test based on the Poisson process representation of the data over the two-dimensional region $(0, 1) \times [u, \infty)$. We first divide the region into rectangular pieces with roughly the same number of observed points. To achieve that, we first divide the whole region along the median of the observed values in the time domain. Then we divide each smaller region along the median of the observed values within that region in the magnitude domain. We continue doing this until we have sufficiently many smaller subregions each with at least 5 observed points. We can calculate the expected frequency of points within each subregion based on the maximum likelihood estimates of our model parameters. These let us perform an usual chi-square goodness-of-fit test. We assume the degrees of freedom in a conservative manner by not taking account of the estimated model parameters. It should be noted that a high p-value in this test indicates a good fit for the corresponding model.

4. Data Analysis

4.1 Maximum temperature data

We first analyzed a data of daily maximum temperatures in the month of July over the years 1954 - 2011 obtained from one particular weather station in the Fujian province of China. An extreme temperature was assumed to occur if the daily maximum temperature was more than $33\,^{\circ}\mathrm{C}$ so that there were 83 days when the daily maximum went above this threshold. We then fitted our model using both families of densities to model the frequency of threshold exceedances - Beta and quadratic. The fitted μ function that models the magnitude of the threshold exceedances has been shown in Figure 3 along with the actual observed magnitudes of the exceedances. We observe that the μ function mimics the observed fluctuations in the exceedance magnitude irrespective of the family chosen to model λ.

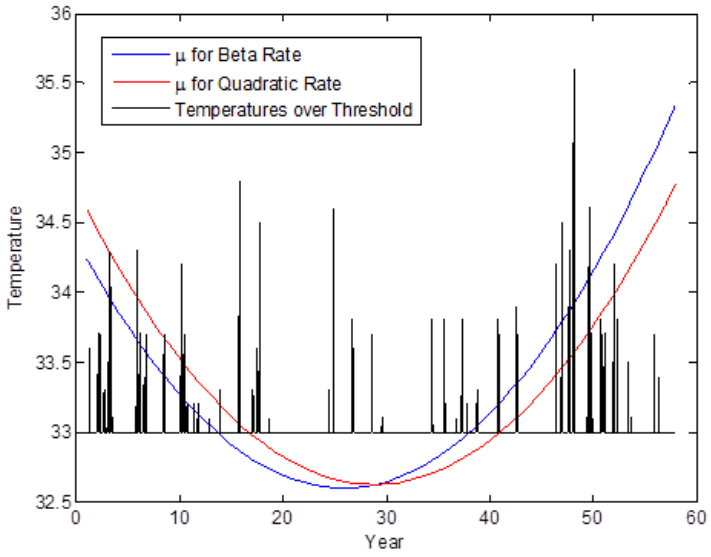

Figure 3: Fitted location parameter $\mu(t)$ and the observed extreme temperatures

Figure 4 shows the yearly frequency of threshold exceedances over the observation period of 58 years in the left panel. This fluctuation in the frequency is modeled by the

fitted density function λ on $(0, 1)$ shown on the right panel. It is evident from the figure that the λ function fails to capture the observed pattern for both Beta and quadratic families of densities. On the contrary, it presents an almost opposite shape.

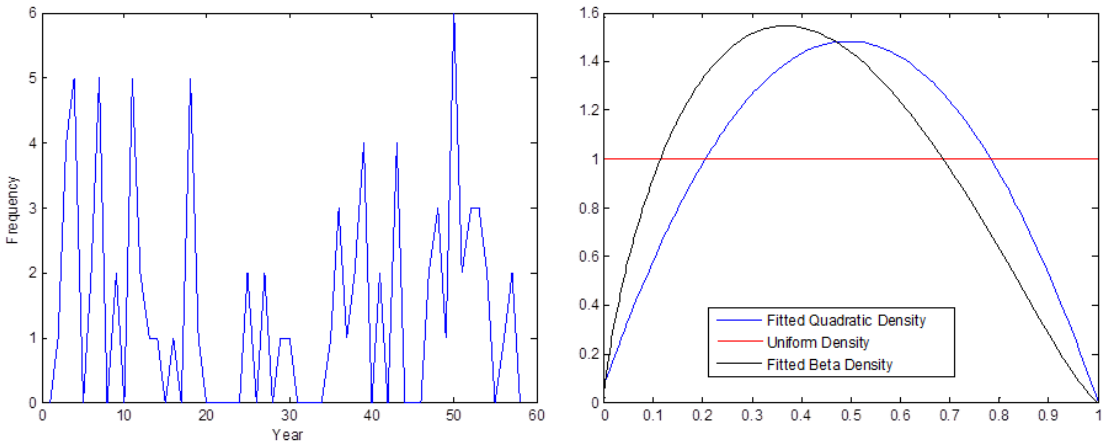

Figure 4: Fitted density $\lambda(t)$ and the observed frequency of extreme temperatures

When we keep μ fixed over time, we have seen that the model becomes separable in the sense that the likelihood is a product of two components - one involving μ and the other involving λ and λ is estimated based only on the times of the exceedances whereas μ, σ and ξ are estimated based only on the exceedance magnitudes. This separable model always produces a reliable estimate of λ since it captures the fluctuations in the frequency based only on the times of the exceedances. But it cannot capture the nonstationarity in the exceedance magnitudes through μ, σ and ξ which are constants. Figure 5 shows that the fitted λ function from the separable model indeed mimics the observed fluctuation in the frequency of threshold exceedances closely, for both families of $\lambda(\cdot)$ - Beta and quadratic.

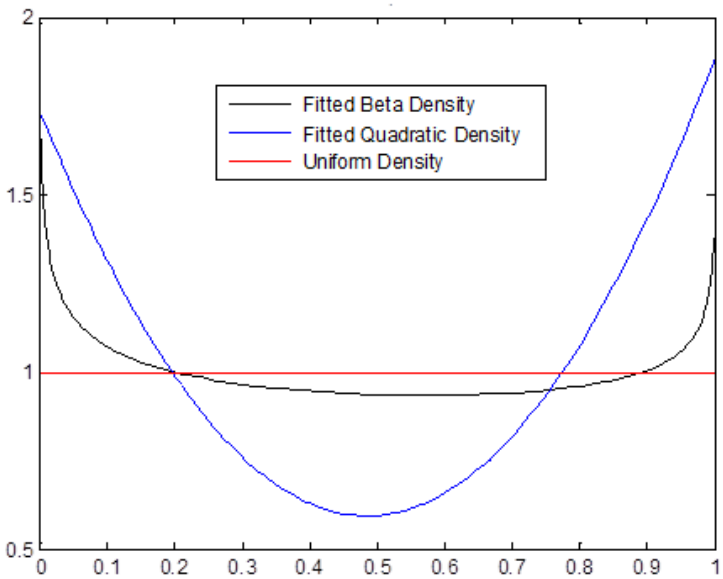

Figure 5: Fitted density $\lambda(t)$ from the separable model for the maximum temperature data

It is clear from the above figures that the patterns of change in the magnitude and the frequency of the threshold exceedances over the course of 58 years have been very similar

(both being U-shaped). We also know that the usual nonstationary model with $\lambda \equiv 1$ is adequate in explaining similar patterns in the frequency and magnitude of extremes over time. In other words, μ can play the role of λ in explaining the frequency for those kinds of data of extreme values. In that sense, the usual nonstationary model cannot be described as inadequate for such data. On the other hand, introducing λ through our new model leads to overparameterization which in turn leads to the wrong estimates for λ rendering our proposed model not much effective. This becomes more evident from the maximized log-likelihood values for each of the fitted models displayed in Table 1.

Table 1: Maximized log-likelihood values for Maximum temperature data

Model	Max. log-likelihood
Constant μ, $\lambda \equiv 1$	-91.7573
Quadratic μ, $\lambda \equiv 1$	-88.6387
Constant μ, Quadratic λ	-87.6028
Quadratic μ, Quadratic λ	-84.2725

As a result, the p-values corresponding to the likelihood ratio tests for $H_0 : \beta_1 = 0, \beta_2 = 0$ and $H_0 : a = 0, b = 0$ turn out to be 0.0358 and 0.0127 respectively, both being insignificant at 1% level. It clearly shows that making both μ and σ time-varying doesn't give us any particular advantage.

This gives us an insight that our proposed model will work much better for a data where the patterns of the magnitude and the frequency of the threshold exceedances are not similar and if possible, in conflict with each other. To achieve this we simulate a dataset with such properties.

We have also noticed that Beta and quadratic families for $\lambda(\cdot)$ have produced similar results in this example which has been generally observed in many examples. But the quadratic family performs much better in terms of convergence of the optimization algorithm. So we will only talk about the quadratic family from now on.

4.2 A simulation study

We simulate 60 years of data in the same format as the previous dataset in such a manner that there are 36 threshold excesses in the first 10 years, 25 excesses in the next 10 years and so on so that there is only one threshold excess in the last 10 years. The occurrence times were chosen uniformly within each 10 year period so that there are 91 exceedances in total. The simulation has been done keeping in mind that we intend to fit λ as a quadratic function. Next we assume $\mu(t) = 36 + 0.75y(t) + 0.25y^2(t)$ where $y(t)$ is the year corresponding to time $t \in (0, 1)$ such that any time $t \in (\frac{i-1}{60}, \frac{i}{60}]$ corresponds to the i^{th} year for $i = 1, \cdots, 60$. We have standardized $y(t)$ and $y^2(t)$ values to remove collinearity. Now for each simulated exceedance time t, we simulate the magnitude of exceedance from the generalized Pareto distribution with $\xi = -0.3, \sigma = 1$ and $\mu = \mu(t)$. These parameters were chosen in such a way that the data values look like actual summer temperatures. For this data, the magnitudes of threshold excesses are increasing with time whereas the frequency of exceedances is decreasing. This is evident from Figure 6 which shows the Poisson process representation of our simulated data.

Figure 7 shows that the fitted function λ from our model closely mimics the simulated fluctuation in the frequency of exceedances and is also close to the reliable estimate of the λ function obtained from the separable model.

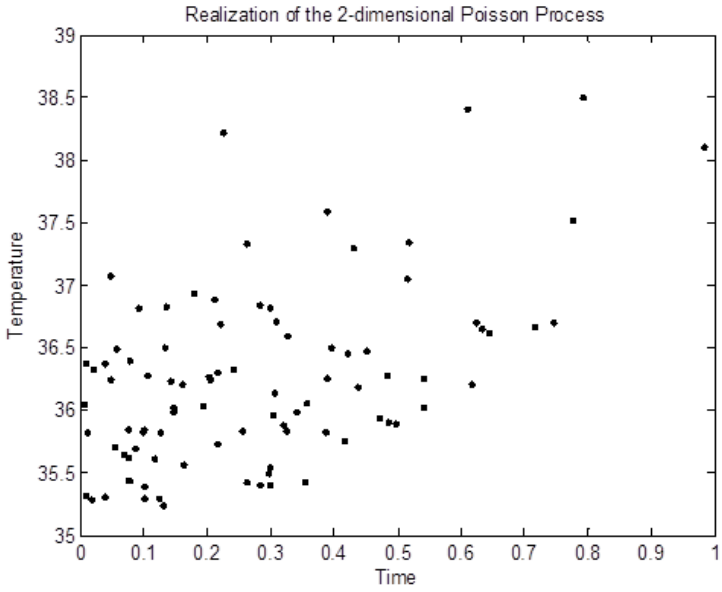

Figure 6: Simulated threshold exceedances

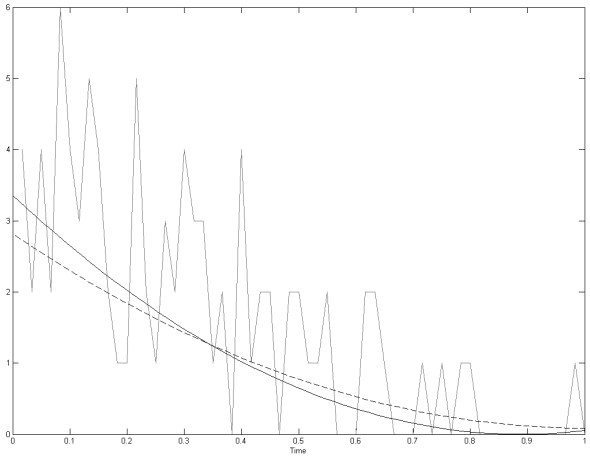

Figure 7: Frequency of simulated exceedances (grey line) and fitted $\lambda(t)$ from the separable model (dashed line) and the proposed model (solid line)

We also see from Figure 8 that the fitted μ function from our model mimics the simulated fluctuation in the magnitude of exceedances. More importantly, if we fit the usual nonstationary model assuming $\lambda \equiv 1$ as in (3), the fitted μ now shows a completely opposite pattern.

This illustrates that the usual nonstationary model is clearly inadequate here and also demonstrates the usefulness of our approach in modeling the magnitudes and frequencies using μ and λ separately. This becomes more evident from the maximized log-likelihood values provided in Table 2 for each of the fitted models.

Using these values, we perform likelihood ratio tests for the null hypotheses $H_0 : \beta_1 = 0, \beta_2 = 0$ and $H_0 : a = 0, b = 0$. The p-values turn out to be 9.54×10^{-4} and 3.97×10^{-6} respectively. This evidently shows the merit of modeling μ and λ separately.

Then we perform the goodness-of-fit test as described in the previous section. Modeling

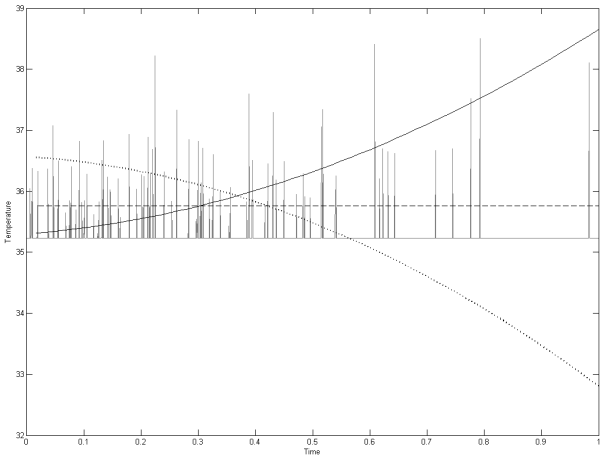

Figure 8: Magnitudes of simulated exceedances (grey bars) and fitted $\mu(t)$ from the usual nonstationary model (dotted line), the separable model (dashed line) and the proposed model (solid line)

Table 2: Maximized log-likelihood values for Simulated data

Model	Max. log-likelihood
Constant μ, $\lambda \equiv 1$	-140.12
Quadratic μ, $\lambda \equiv 1$	-116.04
Constant μ, Quadratic λ	-110.56
Quadratic μ, Quadratic λ	-103.60

$\lambda(t)$ as a quadratic function changes the p-value of the chi-square test to 0.3254 from 0.0179 which is the p-value under $\lambda \equiv 1$. The two-dimensional region $(0, 1) \times [u, \infty)$ is divided into 16 subregions for this test so that each has at least 5 observed points as displayed in Figure 9.

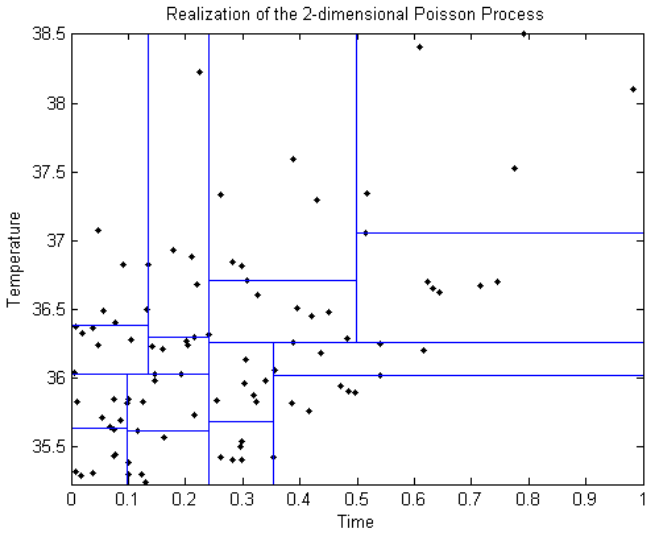

Figure 9: Subregions for the goodness-of-fit test for the simulated data

Both the tests reveal very strong statistical evidence that the intensity $\lambda(t)$ is not constant over time which shows the merit of our proposed model in analyzing this simulated data.

Let us now look at the return level calculations for this data. Due to no exceedance in the interval $(0.85, 0.95)$ in the rescaled time domain, the fitted $\lambda(\cdot)$ is very close to 0 on that interval, creating numerical instabilities for daily or yearly calculation of return levels on the interval $(0.85, 0.95)$. So we calculate the 60 year-return levels computed over blocks of 3 years using our model and compare it with similar return levels calculated from the usual nonstationary model in Figure 10. Note that we expect both of them to be exceeded just once in the entire time domain. It is clear that the return levels from our model make much more sense, both around the initial part of the time domain where the usual model makes the return levels too high and around $(0.85, 0.95)$ where it is expected that the return levels should go down due to the exceedance frequency becoming too low.

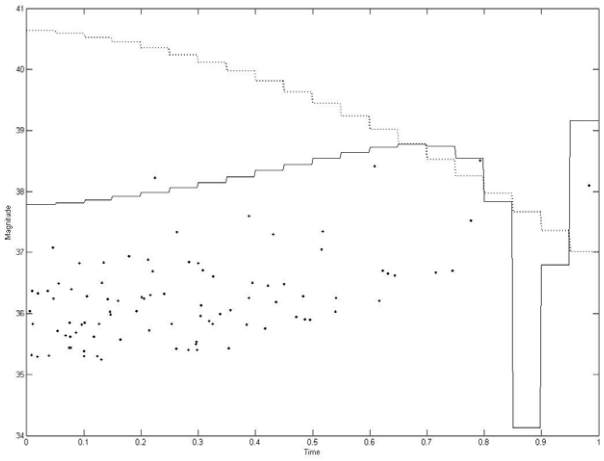

Figure 10: Return levels from the proposed model (solid line) and the usual nonstationary model (dotted line) for the simulated data

We now demonstrate the predictive performance of our model. For prediction purposes, we have left out the last 10 years of the data and estimate all parameters based on the rest of the observations and use them to estimate the intensity function $\lambda'(t, x)$ over the prediction region $[5/6, 1) \times [u, \infty)$. We divide the prediction region into a 20×50 grid and calculate the predicted intensity function over each small rectangle. We now display in Figure 11 the difference between the predicted intensities using our model from (4) and the usual model from (3) using a color contour over the prediction region. Here reddish and bluish hues signify positive and negative differences respectively. Clearly our model predicts higher intensity than the usual model in the vicinity of the actual exceedance (shown as the black circle) in the prediction region for the simulated data.

4.3 California Earthquake Data

We now analyze the daily earthquake data in the state of California. Earthquakes are a common occurrence in California and about $37,000$ earthquakes are recorded annually. From the website of U.S. Geological Survey, we obtained the daily earthquake data from 1973 to 2012 which had epicenters inside a rectangular region that encompasses the entire California and includes some portions of Oregon, Nevada, Arizona, Mexico and the Pacific

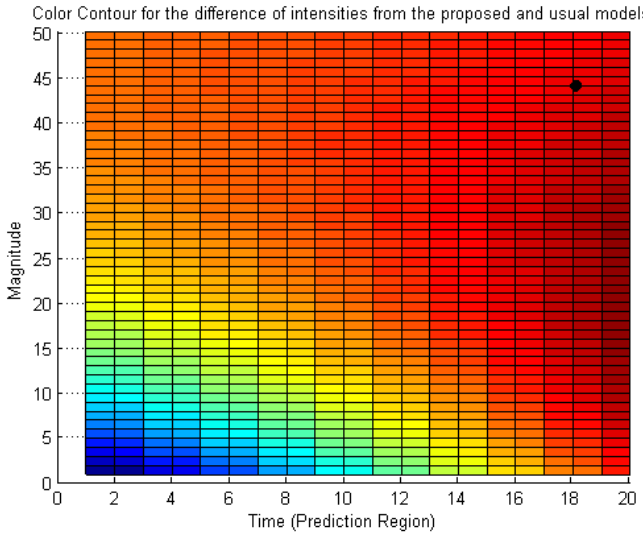

Figure 11: Contour plot for the difference of predicted intensities from the proposed model and the usual nonstationary model for the simulated data

Ocean. As a declustering measure only the earthquake with maximum intensity is considered to be an extreme event in case there were multiple significant earthquakes on a single day. The threshold was chosen to be 5.7 in the moment magnitude scale which is somewhat consistent with the notion of extreme earthquakes in terms of the amount of damage they can cause. We show the threshold exceedances in their Poisson process representation in Figure 12.

It is also expected that after a massive earthquake another such earthquake will not take place in the next several days, even months, with the exception of immediate aftershocks. We can disregard the aftershocks here as they will most probably be eliminated by the declustering process. On the other hand, a long period without any significant seismic activity probably indicates the build-up for the next massive earthquake. Therefore we expect a conflicting relation between the frequency and magnitude of significant earthquakes in a particular region. This is observed quite clearly in Figure 12. Around 1973, a high frequency of extreme earthquakes was observed though they were not 'very' extreme (none of them exceeded 6.5 in the moment magnitude scale). On the other hand, in recent years significantly fewer extreme earthquakes were observed but several of them exceeded 6.5 in the moment magnitude scale.

We again fit and compare the four models that we fitted in the earlier example. Figure 13 shows that the fitted function λ from our model closely mimics the observed fluctuation in the frequency of exceedances and is very close to the reliable estimate of the λ function obtained from the separable model.

The fitted functions $\mu(t)$ are plotted in Figure 14. We see that when $\lambda(t)$ is assumed to be a constant in the usual nonstationary model, the fitted $\mu(t)$ does not capture the trend in the exceedance magnitudes. On the other hand, when $\lambda(t)$ is modeled as a quadratic function by our model, the fitted $\mu(t)$ captures the trend reasonably well. This again illustrates the fact that the usual nonstationary model may be inadequate while displaying the utility of our approach.

The maximized log-likelihood for the fitted models are provided in Table 3. Using these values, we perform likelihood ratio tests for the null hypotheses $H_0 : \beta_1 = 0, \beta_2 = 0$ and $H_0 : a = 0, b = 0$. The p-values turn out to be 0.0173 and 0.0041 respectively.

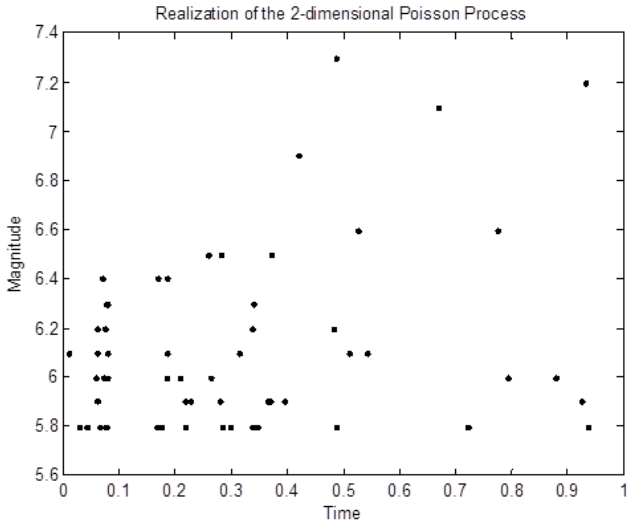

Figure 12: Earthquakes in California exceeding the threshold 5.7

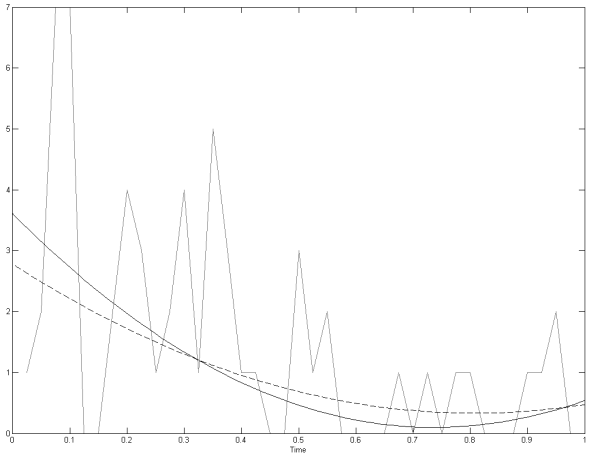

Figure 13: Frequency of exceedances (grey line) and fitted $\lambda(t)$ from the separable model (dashed line) and the proposed model (solid line) for the earthquake data

This evidently shows the merit of modeling μ and λ separately, in particular the importance of λ. The goodness-of-fit test seems quite unstable for this data as the points in the two-dimensional space piled up along horizontal lines due to the magnitudes being always recorded upto 1 decimal place and hence the test results widely vary if the boundaries of the subregions are shifted slightly.

Let us now look at the return level calculations for this data. We calculate the 40 year-return levels computed yearly using our model and compare it with similar return levels calculated from the usual nonstationary model in Figure 15. Note that we expect both of them to be exceeded just once in the entire time domain. It is clear that the return levels from our model looks more reasonable whereas those form the usual model are too high during the initial part of the time domain and too low during the later part.

Finally we demonstrate the predictive performance of our model for the earthquake data. For prediction purposes, we now leave out the last 5 (out of 40) years of the data

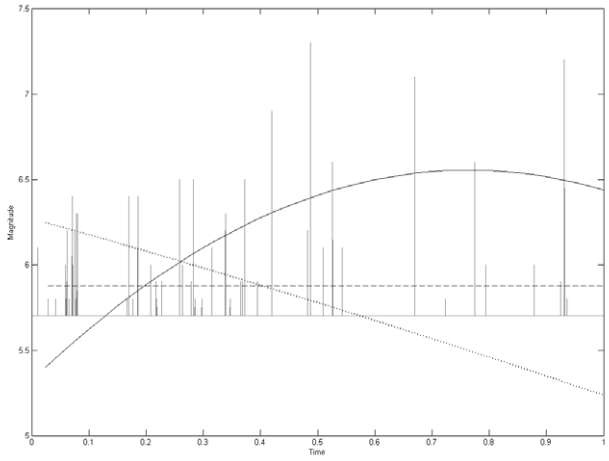

Figure 14: Magnitudes of exceedances (grey bars) and fitted $\mu(t)$ from the usual nonstationary model (dotted line), the separable model (dashed line) and the proposed model (solid line) for the earthquake data

Table 3: Maximized log-likelihood values for California earthquake data

Model	Max. log-likelihood
Constant μ, $\lambda \equiv 1$	-43.4530
Quadratic μ, $\lambda \equiv 1$	-31.8396
Constant μ, Quadratic λ	-30.4106
Quadratic μ, Quadratic λ	-26.3547

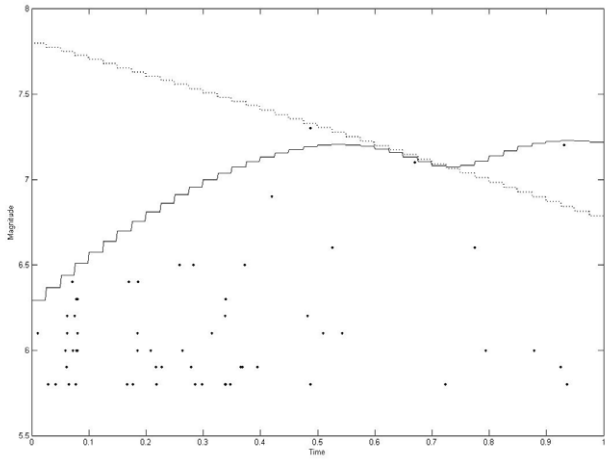

Figure 15: Return levels from the proposed model (solid line) and the usual nonstationary model (dotted line) for the earthquake data

and estimate the intensity function $\lambda'(t, x)$ over the prediction region $[7/8, 1) \times [u, \infty)$. We again divide the prediction region into a 20×50 grid and display in Figure 16 the difference between the predicted intensities using our model and the usual nonstationary model using a color contour. Reddish and bluish hues again signify positive and negative

differences respectively. Clearly our model predicts higher intensity than the usual model in the vicinity of the very extreme earthquake that happened in the prediction region whereas it has a slightly lower (or almost equal) intensity near the less extreme earthquakes. But we can argue that predicting the more devastating earthquake should be more important in this scenario.

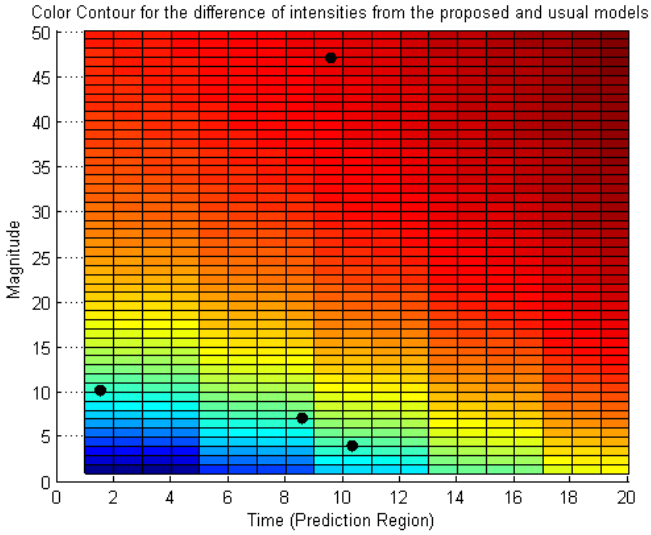

Figure 16: Contour plot for the difference of predicted intensities from the proposed model and the usual nonstationary model for the earthquake data

5. Discussion

Using these examples, we have established that the usual models for nonstationarity in extreme events seem to be inadequate if the observed data presents opposite or at least different changing patterns in frequency and magnitudes of the extreme events. We have seen that such data may not be rare in real life examples. It is worth noting that for more regular processes like temperature, frequency and magnitudes of the extremes increase or decrease together. If there is an extremely hot day, it is usually followed by several such days. This kind of trend is hard to remove from the data even by any practical declustering scheme. For such processes, the existing models should work fine or even prove to be quite adequate.

On the other hand, whenever extreme events occur in a process as a result of the sudden release of some form of energy that gradually builds up over time, any severe extreme event will usually be preceded and followed by periods of relative lull resulting in a negative relation between the magnitude and frequency of the extremes. Environmental examples like earthquakes, cyclones or tornadoes will give rise to such data. For that kind of data, we have proposed an extension of the existing models which seem to perform much better in terms of model fit, estimation of return levels and prediction of future extremes. This method can be extended into models for spatial extremes where spatial patterns of frequency and magnitudes of extreme values are conflicting. Extending this two dimensional Poisson process approach for temporal extremes to analyze spatio-temporal extremes by using Poisson processes in higher dimensional Euclidean spaces seems to be an area for further investigation.

References

F. J. Acero, J. A. Garcia, and M. C. Gallego. Peaks-over-threshold study of trends in extreme rainfall over the iberian peninsula. *Journal of Climate*, 24(4):1089–1105, 2011.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer, 2006.

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

B. V. Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44:423–453, 1943.

D. Jaruskova. Estimating high levels exceedance probabilities by point process approach with applications to northern moravia precipitation and discharges series. *Journal of Hydrology and Hydromechanics*, 57(3):162–171, 2009.

E. C. Mannshardt-Shamseldin, R. L. Smith, S. R. Sain, L. O. Mearns, and D. Cooley. Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *The Annals of Applied Statistics*, 4(1):484–502, 2010.

J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3 (1):119–131, 1975.

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

T. F. Stocker, D. Qin, G. K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. M. Midgley, editors. *IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, 2013.