Supplementary materials for this article are available at 10.1007/s13253-013-0161-y.

A Hierarchical Model for Serially-Dependent Extremes: A Study of Heat Waves in the Western US

Brian J. REICH, Benjamin A. SHABY, and Daniel COOLEY

Heat waves take a major toll on human populations, with negative impacts on the economy, agriculture, and human health. As a result, there is great interest in studying the changes over time in the probability and magnitude of heat waves. In this paper we propose a hierarchical Bayesian model for serially-dependent extreme temperatures. We assume the marginal temperature distribution follows the generalized Pareto distribution (GPD) above a location-specific threshold, and capture dependence between subsequent days using a transformed max-stable process. Our model allows both the parameters in the marginal GPD and the temporal dependence function to change over time. This allows Bayesian inference on the change in likelihood of a heat wave. We apply this methodology to daily high temperatures in nine cities in the western US for 1979–2010. Our analysis reveals increases in the probability of a heat wave in several US cities. This article has supplementary material online.

Key Words: Bayesian modeling; Climate change; Generalized Pareto distribution; Markov model.

1. INTRODUCTION

In the last decade, major heat waves have made headlines with alarming regularity (e.g., 2003 in western Europe, 2010 in Russia, 2011 in Texas, 2012 in the eastern US). In addition to their economic impacts, these severe events have had tremendous consequences for agriculture and human health. Several studies have attempted to quantify the impacts and projected impacts of heat waves on human mortality (Peng et al. 2011; Bobb, Dominici, and Peng 2011), agriculture (Schlenker and Roberts 2009), and economic losses (Easterling et al. 2000). Given the magnitude of the impacts and the visibility in the

Brian J. Reich (✉) is an Assistant Professor (E-mail: *brian_reich@ncsu.edu*), North Carolina State University, Raleigh, USA. Benjamin A. Shaby is an Assistant Professor (E-mail: *bas59@psu.edu*), The Pennsylvania State University, University Park, PA, USA. Daniel Cooley is an Assistant Professor, Colorado State University, Fort Collins, USA

Journal of Agricultural, Biological, and Environmental Statistics, Volume 19, Number 1, Pages 119–135
DOI: 10.1007/s13253-013-0161-y

popular press, it is natural to ask whether heat waves have increased in frequency, intensity, or duration.

There are many operational definitions of heat waves, but all include some notion of persistent extremely high temperatures. As such, questions about the changing characteristics of heat waves depend critically on the dependence structure of extreme events. Extreme value methods (Coles 2001) that consider block maxima, annual high temperatures for example, are unsuitable for analysis of heat waves because they cannot accommodate any notion of duration. The closely-related theory of exceedances over a high threshold giving rise to the generalized Pareto distribution (GPD) gets us closer because it can make use of all high temperature days rather than just annual maxima. However, in its most basic form it applies to independent data rather than serially dependent sequences like we see in the temperature record.

There are two general approaches to adapting threshold methods to dependent sequences (Chavez-Demoulin and Davison 2012, for a review). The first and most common is the peaks-over-threshold method (Davison and Smith 1990), which attempts to describe the magnitude of threshold exceedances but which does not attempt to model temporal dependence. The peaks-over-threshold method identifies clusters of exceedances and analyzes the maximum of each cluster using the GPD on the justification that the cluster peaks are separated well enough to be approximately independent. The second class of approaches is to explicitly model dependence of data above the threshold. This strategy is more challenging because it requires specification of a dependence model that is appropriate for the joint tail of the distribution. It does, however, afford several advantages. First, considering all threshold exceedances rather than just cluster maxima leads to improved inference of GPD parameters (Fawcett and Walshaw 2007). Second, complications arising from the need to identify clusters (Ferro and Segers 2003, e.g.) are completely eliminated. Most importantly, explicitly modeling dependence structures enables one to investigate parameters that control the strength and range of dependence, as well as whether those parameters are changing in time. As our application is the study of heat waves, it is necessary that we explicitly model dependence in the tail. This allows us to study whether the dependence characteristics that drive heat waves are changing in time.

Several options have been proposed for modeling serial dependence of threshold exceedances. The simplest approach is to construct a Markov chain based on known bivariate parametric models that allow tail dependence (Smith, Tawn, and Coles 1997; Fawcett and Walshaw 2006). This approach is appealing because of its simplicity but restrictive in the kinds of dependence it can represent. Alternatively, Bortot and Gaetan (2013) propose a class of models using gamma auxiliary variables that is similar in spirit to the model we propose. Their hierarchical model is flexible in its serial dependence characteristics and results in a process that has GPD margins. However, their representation applies only to heavy-tailed data (i.e. in the Fréchet domain of attraction), precluding, in particular, analysis of temperature data that exhibit short tails.

We model serial dependence above the threshold using a max-stable process that we construct hierarchically using positive stable auxiliary variables (see Stephenson 2009 Fougères, Nolan, and Rootzén 2009). The proposed model accommodates a variety of

extremal dependence characteristics that result from different ways of combining the positive stable auxiliary variables. We demonstrate two possible choices, one using a kernel weighting function (see Reich and Shaby 2013; Shaby and Reich 2013, in the spatial context) and the other using an underlying first-order Markov structure. We allow the parameters in both the GPD marginal distribution and the temporal dependence function to change over time. Crucially, our hierarchical representation allows Bayesian computing and a closed-form expression for the probability of a heat wave, that is, the probability of the temperature exceeding T degrees for h consecutive days for any (h, T). An advantage of our approach is that because it is defined for any index, the distribution of any combination of observed threshold exceedances can be accommodated. This leads to relatively straightforward Bayesian inference on the change in the probability of a heat wave. Furthermore, when performing inference the hierarchical construction of the process allows the likelihood to be evaluated at all threshold exceedances, and pairwise likelihood methods (Padoan, Ribatet, and Sisson 2010) are not needed.

We apply this model to analyze daily high temperatures in nine cities in the western US for the years 1979–2010. Several studies examine how climate change could affect temperature extremes and heat waves in this region. Examining output from both current and future runs of climate models, Meehl and Tebaldi (2004) found the greatest change in North American heat waves to be centered in the western or northwestern US, depending on the definition of heat wave used. Again, looking at climate model output, Kunkel, Liang, and Zhu (2010) found the greatest increase of the annual 3-day heat wave temperature to occur in the western and southern US. Examining observations, Frich et al. (2002) found that estimated changes in a heat wave duration index were generally positive over the western US, but few were found to be significantly different from zero. Our analysis contributes to this body of work by identifying several cities with increasing rates of heat waves.

The remainder of the paper proceeds as follows. We describe the heat wave model and study its dependence properties in Sections 2 and 3. Section 4 provides the MCMC details. In Section 5 we use the proposed method to analyze daily summer temperature in the western US to study changes in the probability of heat waves over time.

2. TIME SERIES MODELS FOR TAIL DEPENDENCE

Let Y_t^* be the observed value for day $t = 1, \ldots, n$. For notational convenience we describe the model for single year; this is generalized in Section 3. As is common in modeling extremes above a threshold μ, we treat the data $Y_t = \max\{\mu, Y_t^*\}$ as if it were censored at μ. The threshold is chosen by standard exploratory analysis tools, and may vary by location. Above the threshold, we assume the density is well-approximated by a generalized Pareto distribution (GPD). The GPD has three parameters: lower bound μ, scale $\sigma_t > 0$, and shape ξ_t. The support is (μ, ∞) if $\xi_t > 0$ and $(\mu, \mu - \sigma_t/\xi_t)$ if $\xi_t < 0$. This gives marginal density

$$f(y_t) = \begin{cases} \pi_t & y_t = \mu, \\ [1 - \pi_t]\,\text{dGPD}[y_t|\mu, \sigma_t, \xi_t] & y_t > \mu, \end{cases} \tag{2.1}$$

where $\pi_t > 0$ is the probability below the threshold and dGPD is the GPD density. The interpretation of the scale parameter σ_t and the parameter that determines the exceedance probability π_t depend on the threshold μ. In our setting, an appropriate μ is initially chosen and fixed, and we model these remaining parameters. Our nonstationary model allows the GPD shape and scale to vary by year and day of the year to study changes in the probability of heat waves, as described below.

We account for residual dependence via a transformation of a max-stable process. Max stable processes are natural models for tail dependence as they constitute the limiting processes of the pointwise maximum of linearly-renormalized independent stochastic processes (de Haan and Ferreira 2006, Chap. 9). In particular, max-stable processes allow for asymptotic dependence (see Section 2.3), while more common stochastic processes, such as Gaussian processes, do not. Max-stable processes have typically been used to model pointwise block (e.g., annual) maximum data. For example, Padoan, Ribatet, and Sisson (2010), Ribatet, Cooley, and Davison (2011), and Reich and Shaby (2013) model annual maximum precipitation from locations scattered across a study area. Here, we use a max-stable process to model daily data that exceed a threshold. Our justification is that if (and only if) the data are asymptotically dependent, the dependence among large observations should be well approximated by the dependence structure inherent in the max-stable process. Others have similarly used both multivariate max-stable distributions and max-stable processes to model threshold exceedances. Smith, Tawn, and Coles (1997) use a bivariate max-stable model within their temporal Markov-chain approach for threshold exceedances; we note that our inference method differs from theirs because they use only bivariate representations of dependence. Also, both Huser and Davison (2013) and Jeon and Smith (2013) employ max-stable processes to spatially model threshold exceedances using pairwise likelihood methods.

Let X_t be a max-stable temporal process. It is common to describe tail dependence assuming a common marginal distribution. Since X_t is max-stable, we can assume that, without loss of generality, it has unit Fréchet marginal distribution so that $\text{Prob}[X_t < x] = \exp(-1/x)$ (Resnick 1987, Proposition 5.10). Denote by Q the quantile function (inverse distribution function) of Y_t. Then the process X_t is related to the response as $Y_t = Q(U_t)$, where $U_t = \exp(-1/X_t) \sim \text{Unif}(0,1)$ and thus Y_t has the desired marginal distribution, (2.1), by the probability inverse transform. In Section 2.1 we define our model for $\{X_t\}$.

2.1. RANDOM EFFECTS FORMULATION OF A MAX STABLE PROCESS

Dependence between the X_t's is captured using random effects $A_t > 0$. We assume that the X_t's are conditionally independent given random effects A_t. To induce dependence, the random effects are modeled as linear combinations of independent effects $B_1, \ldots, B_L$. Within this framework, we desire a model such that X_t has Fréchet marginal distributions for all $t = 1, \ldots, n$, $\mathbf{X} = (X_1, \ldots, X_n)$ is max-stable so that X_t and X_{t+h} are asymptoti-

cally dependent, and $\{X_t\}$ is stationary. To satisfy these criteria, we assume that

$$\begin{aligned} X_t|A_t &\overset{\text{indep}}{\sim} \text{GEV}\big[A_t^\alpha, \alpha A_t^\alpha, \alpha\big], \\ A_t &= \sum_{l=1}^{L} K_l(t|\gamma)^{1/\alpha} B_l, \\ B_l &\overset{\text{iid}}{\sim} \text{PS}(\alpha), \end{aligned} \tag{2.2}$$

where $X_t|A_t$ follows the generalized extreme value (GEV) distribution with location A_t^α, scale αA_t^α and shape α, B_l follows the positive stable (PS) distribution with shape parameter α and Laplace transform $\text{E}[\exp(-Bt)] = \exp(-t^\alpha)$, and $K_l(t|\gamma)$ is a weight function with α and γ controlling the strength of temporal dependence, as described below. This random effects formulation is known to lead to a marginal (over the random effects) asymmetric logistic joint distribution (Stephenson 2009), which is a common multivariate extreme value distribution. As discussed below, we model the weights $K_i(t|\gamma)$ in a way that accounts for temporal association.

For any times $\mathbf{s} = (s_1, \ldots, s_J)$ and cut-offs $\mathbf{x} = (x_1, \ldots, x_J)$ we show in the Supplementary Materials that the joint distribution is

$$P(X_{s_1} \le x_1, \ldots, X_{s_J} \le x_J) = \exp\big[-V(\mathbf{s}, \mathbf{x})\big], \tag{2.3}$$

where $V(\mathbf{s}, \mathbf{x}) = \sum_{l=1}^{L}[\sum_{j=1}^{J} x_j^{-1/\alpha} K_l(s_j|\gamma)^{1/\alpha}]^\alpha$ is the J-dimensional exponent measure function (Beirlant et al. 2004). In addition, we restrict the weights to satisfy:

1. $K_l(t|\gamma) \ge 0$;
2. $\sum_{l=1}^{L} K_l(t|\gamma) = 1$ for all t;
3. $V(\mathbf{s}, \mathbf{x}) = V(\mathbf{s} + h, \mathbf{x})$ for all $\mathbf{x} = (x_1, \ldots, x_J)$, $\mathbf{s} = (s_1, \ldots, s_J)$, and $h > 0$ so that $x_j > 0$ and $s_j \ge 1$ and $s_j + h \le n$.

Under these conditions, Supplementary Materials A.1 shows that:

1. X_t has unit Fréchet marginal distribution for all t;
2. $\mathbf{X} = (X_1, \ldots, X_n)$ is max-stable;
3. $\{X_t\}$ is strongly stationary.

In Section 2.2 we give examples of weights $K_l(t|\gamma)$ that satisfy these conditions, and in Section 2.3 we discuss the temporal dependence induced by these weights.

2.2. CHOICE OF WEIGHTS

Following Reich and Shaby (2013) for spatial data, one approach is to take the weights to be kernel functions derived by scaling a probability density function $f(t|m, s)$, where m and s are location and scale parameters, respectively. Then for a fixed grid of knots, $\mathcal{D} = \{T_1, \ldots, T_L\}$, the weights become

$$K_l(t|\gamma) = \frac{f(t|T_l, \gamma)}{\sum_{k=1}^{L} f(t|T_k, \gamma)}. \tag{2.4}$$

The intuition behind this model is that B_l controls the magnitude of the heat wave centered at day T_l, and $K_l(u|\gamma)$ determines the effect of this heat wave on previous and subsequent days.

We must select the knots $\mathcal{D}$ to satisfy the three conditions in Section 2.1. Since f is a density, $K_l(t|\gamma) \geq 0$, and the scaling in the denominator in (2.4) ensures that $\sum_{l=1}^{L} K_l(t|\gamma) = 1$ for all t. To satisfy the final condition, we take $\mathcal{D} = \{-N, -N+1, \ldots, n+N\}$, where N provides a large enough buffer so that days separated by at least N days can be considered independent and $K_l(t|\gamma)K_l(t+N|\gamma) \approx 0$ for all t. In this case, the joint distribution at times $\{t_1, \ldots, t_J\}$ is the same as the joint distribution at times $\{t_1 + h, \ldots, t_J + h\}$ and the process is stationary. We take f to be the Gaussian density and $N = 10$ days and refer to this as the "kernel model."

An alternative dependence model more reminiscent of a time series model is a first-order Markov model for the random A_t. The Markov model takes $A_1 = B_1$ and

$$A_t|A_{t-1}, \ldots, A_1 = \gamma^{1/\alpha} A_{t-1} + (1-\gamma)^{1/\alpha} B_t \tag{2.5}$$

for $t > 1$. The parameter $\gamma \in (0, 1)$ determines the strength of temporal dependence, with $\gamma = 0$ giving independence and γ near 1 corresponding to strong temporal dependence. Equivalently, this model can be written in the form of (2.2) by setting

$$K_l(t|\gamma) = \begin{cases} \gamma^{t-1} & l = 1, \\ I(t \geq l)(1-\gamma)\gamma^{t-l} & l > 1. \end{cases} \tag{2.6}$$

It can be shown that these weights satisfy the three conditions given in Section 2.1. We refer to this model for the A_t as the "Markov model."

2.3. TEMPORAL DEPENDENCE

Typical dependence summaries such as correlation measure dependence at the center of the distribution, are therefore poor at describing dependence in the joint tail. Tail dependence in a max-stable random vector is completely described by the exponent measure function V in (2.3), but this quantity is not easily interpreted. A common summary measure of bivariate tail dependence is the extremal coefficient (Schlather and Tawn 2003). Given (X_t, X_{t+h}) is max-stable, the extremal coefficient $\vartheta(h) \in [1, 2]$ can be defined implicitly by

$$\text{Prob}[X_t < x, X_{t+h} < x] = \text{Prob}[X_t < x]^{\vartheta(h)}. \tag{2.7}$$

Given that X_t has unit Fréchet margins, then $\vartheta(h) = V(\mathbf{t}_h^*, \mathbf{x}^*)$ where $\mathbf{t}_h^* = (t, t+h)$ and $\mathbf{x}^* = (1, 1)$. In the case that (X_t, X_{t+h}) are independent, then $\vartheta(h) = 2$, and when they are completely dependent, then $\vartheta = 1$.

The notion of tail dependence can be extended to random vectors that are not max-stable. Let (Z_1, Z_2) be a random vector with common marginals with upper bound z_+. The random vector is said to be asymptotically independent if

$$\chi = \lim_{z \to z_+} P(Z_1 > z | Z_2 > z) = 0, \tag{2.8}$$

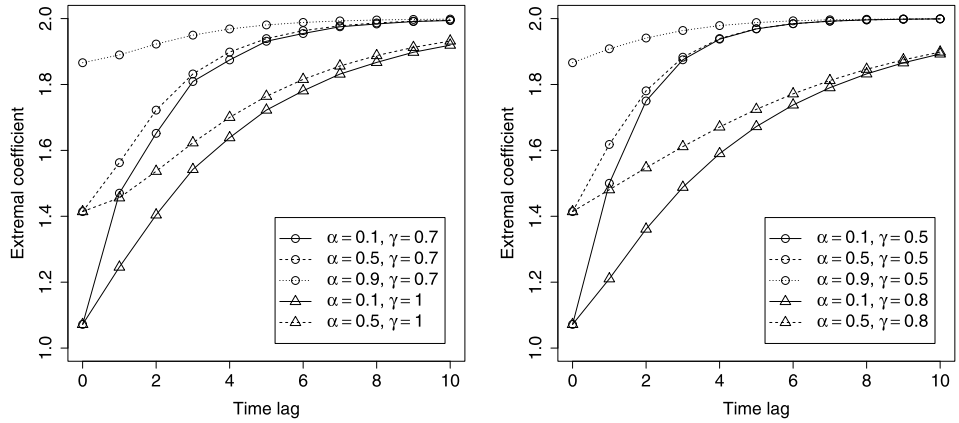

Figure 1. Extremal coefficient plots for the kernel model (left) and Markov model (right).

and asymptotically dependent otherwise (Coles, Heffernan, and Tawn 1999). As mentioned in the introduction, Gaussian-based models exhibit asymptotic independence and are therefore unable to accurately capture clusters of extremes. If (X_t, X_{t+h}) are from a max-stable process with Fréchet margins, then $\chi = 2 - \vartheta(h)$, and the process is asymptotically dependent if $\vartheta(h) < 2$.

From (2.3), we have

$$\vartheta(h) = \sum_{l=1}^{L} \left[K_l(0|\gamma)^{1/\alpha} + K_l(h|\gamma)^{1/\alpha} \right]^{\alpha}, \tag{2.9}$$

i.e., the extremal coefficient is the $L^{1/\alpha}$ norm of the kernel functions. Figure 1 plots the extremal coefficient for the kernel and Markov models. For any choice of weights, the lag-zero extremal coefficient is $\lim_{h \to 0^+} \vartheta(h) = 2^{\alpha}$. Therefore, α controls short-range dependence, with α near zero corresponding to stronger dependence. This illustrates that if $\alpha > 0$, the draws from this process are discontinuous functions of time. However, in our analysis of daily data, dependence at lags less than a day and the continuity of the process are not of primary concern.

The extremal coefficient for the Markov model has the nice form (Supplementary Materials A.2):

$$\vartheta(h) = \left(\gamma^{h/\alpha} + 1\right)^{\alpha} + 1 - \gamma^h. \tag{2.10}$$

The range of dependence is determined by γ, with $\vartheta(h) \approx 2 - \gamma^h$ for small α. The kernel model gives a similar shape, with range of dependence increasing with γ.

To measure higher order dependence, we compute the probability of a heat wave. Define

$$p(T, h) = \text{Prob}(Y_t > T, \ldots, Y_{t+h} > T) \tag{2.11}$$

as the probability of h consecutive days exceeding T degrees. Assuming common marginal probability $P(Y_t < T) = \tau$, Supplementary Materials A.3 shows that

$$p(T, h) = 1 - h\tau + \sum_{1 \le l < k \le h} \tau^{G(k,l)} - \sum_{1 \le l < k < j \le h} \tau^{G(l,k,j)} + \cdots + (-1)^h \tau^{G(1,\ldots,h)}, \tag{2.12}$$

where $G(h_1, \ldots, h_J) = \sum_{l=1}^{L} [\sum_{j=1}^{J} K_l(h_j|\gamma)^{1/\alpha}]^{\alpha}$, which can also be computed recursively as

$$p(T,h) = p(T,h-1) - \tau + \sum_{l=1}^{h-1} \tau^{G(l,h)} - \sum_{1 \le l < k < h} \tau^{G(l,k,h)} + \cdots + (-1)^h \tau^{G(1,\ldots,h)}. \tag{2.13}$$

This probability is interpreted as the probability of a heat wave occurring on an arbitrary window of h days under these climate conditions. This closed-form expression for the probability of a heat waves allows inference on differences in probability by city or time period. In the absence of a closed-from expression, estimating the posterior distribution of $p(T,h)$ would require a Monte Carlo simulation at each MCMC iteration. An accurate approximation for the small probabilities of interest here would require many Monte Carlo replications and would thus be computationally intensive.

3. TAILORING THE MODEL TO THE WESTERN US TEMPERATURE DATA

To illustrate this method, we analyze heat waves for nine cities in the western US: Albuquerque, Denver, Los Angeles, Las Vegas, Phoenix, Salt Lake City, San Diego, San Francisco, and Tucson. Daily maximum temperature data were downloaded from the NCEP North American Regional Reanalysis website (*http://www.esrl.noaa.gov/psd/data/gridded/data.narr.html*). These data are available from 1979–2010, and have been extensively studied and inspected for quality, and have no missing values. We note that the NARR data are generally lower than temperatures recorded by weather monitoring stations. However, they are strongly correlated with observed monitor data (Supplementary Materials Section D), and therefore appropriate for testing for trends over time. The NARR data are preferred to monitor data for this study for their completeness and stability. Although a few extremes occur in the spring and fall, especially for coastal cities, for consistency we restrict our analysis to June, July and August for all cities.

For each day, we take the response to be the average of nine grid cells surrounding the city center (Figure 2). Figure 3 plots the raw data for each city. Extremely hot days clearly cluster into heat waves lasting up to a few days. There is also evidence that the extreme events are more likely to occur in July than June or August. Figure 4 plots the sample extremal coefficient (Coles, Heffernan, and Tawn 1999) for two cities and also shows extremal dependence out to a few days.

We analyze each city separately. Let Y_{it} be the observation at day $t = 1, \ldots, 92$ of year $i = 1, \ldots, 32$. We assume the marginal distribution of Y_{it} has π_{it} mass below the threshold and the GPD(σ_{it}, ξ_i) above the threshold. The X_{it} are independent across years following the model described in Section 2.1, and with dependence parameters α_i and γ_i in year i.

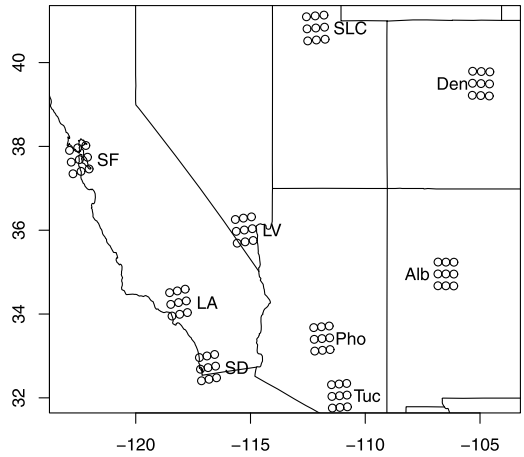

Figure 2. Locations of the nine NARR grids cell used to compute each city's spatial average of daily high temperature.

For the kernel model, we fit

$$\begin{aligned}
\text{logit}(\pi_{it}) &= \beta_0^\pi + \beta_1^\pi Z(i) + \beta_2^\pi z(t) + \beta_3^\pi z(t)^2, \\
\log(\sigma_{it}) &= \beta_0^\sigma + \beta_1^\sigma Z(i) + \beta_2^\sigma z(t) + \beta_3^\sigma z(t)^2, \\
\xi_i &= \beta_0^\xi , \\
\text{logit}(\alpha_i) &= \beta_0^\alpha + \beta_1^\alpha Z(i), \\
\log(\gamma_i) &= \beta_0^\gamma + \beta_1^\gamma Z(i),
\end{aligned} \tag{3.1}$$

where $Z(i) = (i - 16)/10$ and $z(t) = (t - 46)/30$ are standardized covariates controlling trends across and within years, respectively. The Markov model is identical except that $\text{logit}(\gamma_i) = \beta_0^\gamma + \beta_1^\gamma Z(i)$. Exploratory analysis suggests that the likelihood and intensity of extreme temperatures varies throughout the summer, and so we allow the probability below the threshold and GPD scale to vary within year following a quadratic trend. The remaining parameters ξ, α and γ are generally difficult to estimate, and so we select simpler models for these parameters to avoid poor identifiability. Of course, more general models are possible but we elect to fit this simple model and check for fit using goodness-of-fit diagnostics discussed below. We choose independent $\text{N}(0, 10^2)$ priors for all regression coefficients β_j^k.

In this model the marginal distribution varies over years and days within a year. To summarize changes in heat wave frequency, we define $p_i(T, h) = \text{Prob}(Y_{i,46} > T, \ldots, Y_{i,46+h} > T)$, where the marginal distribution parameters π and σ are assumed to be the values for day 46 (the middle of the summer) and year i for all h days. To summarize the change in this probability over time, we inspect the posterior of

$$\log_{10}\left[\frac{p_{32}(T, h)}{p_1(T, h)}\right] \tag{3.2}$$

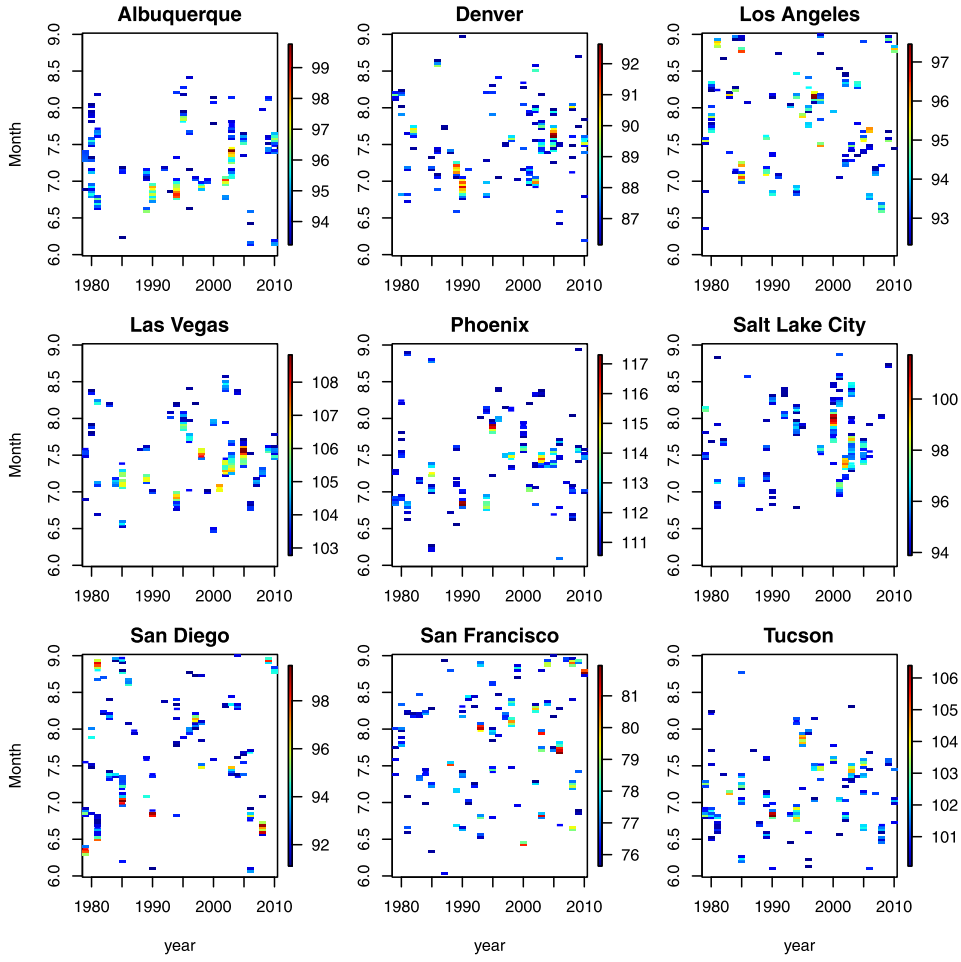

Figure 3. Raw data for each city with values below the city's 0.95 quantile shaded white. Month ranges: from 6 for June 1 to 9 for September 1.

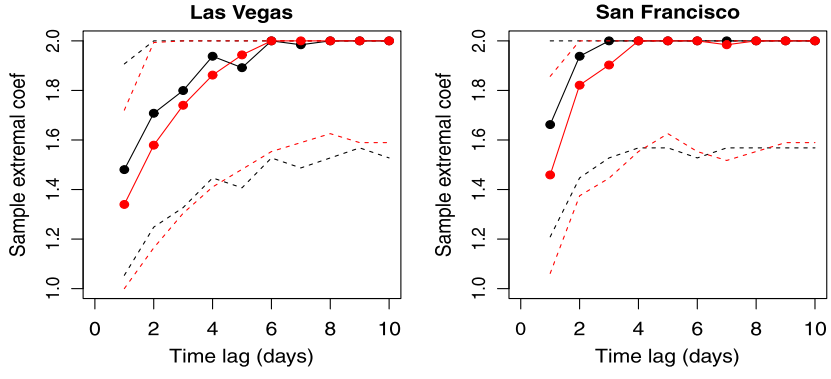

Figure 4. Sample extremal coefficient (computed using exceedances of the city's 95th percentile) and 90 % interval computed separately using data before (black) and after (red) 1994 (Color figure online).

for various combinations of T and h, where $i = 1$ and $i = 32$ are the first and last years in the data set, respectively.

4. COMPUTATIONAL DETAILS

For computational purposes, we require the joint distribution of the responses. Fortunately, the random effects representation in (2.2) allows a Bayesian analysis. Conditioned on the random effects A_{it} via $\mathbf{B} = \{B_{il}\}$ and parameters $\boldsymbol{\beta} = \{\beta_j^k | j = 0, \ldots, 3; k = \pi, \sigma, \xi, \alpha, \gamma\}$, the observations are independent with

$$\text{Prob}[Y_{it} \le y_{it} | \mathbf{B}, \boldsymbol{\beta}] = \begin{cases} \exp\{-A_{it} \log[1/\pi_{it}]^{1/\alpha_{it}}\} & y_{it} = \mu, \\ \exp\{-A_{it} L(y_{it})^{1/\alpha_{it}}\} & y_{it} > \mu, \end{cases} \tag{4.1}$$

where $A_{it} = \sum_{l=1}^{L} K_l(t|\gamma_i)^{1/\alpha_i} B_{il}$. Using $s(y_{it}) = [\xi_{it}(y_{it} - \mu)/\sigma_{it} + 1]^{-1/\xi_{it}} \in (0, 1)$, $l(y_{it}) = 1 - [1 - \pi_{it}]s(y_{it})$, and $L(y_{it}) = \log[1/l(y_{it})]$, and differentiating (4.1) gives the density

$$f(y_{it}|\mathbf{B}, \boldsymbol{\beta}) = \begin{cases} \exp\{-A_{it} \log[1/\pi_{it}]^{1/\alpha_{it}}\} & y_{it} = \mu, \\ \exp\{-A_{it} L(y_{it})^{1/\alpha_{it}}\} \frac{A_{it} L(y_{it})^{1/\alpha_{it}-1} [1-\pi_{it}] s(y_{it})^{1+\xi_{it}}}{\alpha_{it}\sigma_{it} l(y_{it})} & y_{it} > \mu. \end{cases} \tag{4.2}$$

The full likelihood of all observations is simply a product of terms of this form across time.

We also require the positive stable density function, which has the awkward form

$$f(B|\alpha) = \int_0^1 \frac{\alpha B^{-1/(1-\alpha)}}{1-\alpha} c(D) \exp\left[-c(D) B^{-\alpha/(1-\alpha)}\right] dD, \tag{4.3}$$

where $c(D) = \frac{\sin(\alpha\pi D)}{\sin(\pi D)}^{1/(1-\alpha)} \frac{\sin[(1-\alpha)\pi D]}{\sin(\alpha\pi D)}$. For computing, we approximate this as a sum over 100 equally-spaced points covering the unit interval. Alternatively, Stephenson (2009) introduces D as an auxiliary variable to sample from the exact posterior. However, we find our approximation leads to slightly faster convergence.

The posterior then is simply

$$f(\mathbf{B}, \boldsymbol{\beta}|\mathbf{y}) \propto \prod_{it} f(Y_{it}|\mathbf{B}, \boldsymbol{\beta}) \prod_{il} f(B_{il}|\alpha_i) f(\boldsymbol{\beta}). \tag{4.4}$$

To sample from this posterior, we use Metropolis–Hastings sampling. We use log-normal candidates for the elements of $\mathbf{B}$ and Gaussian candidates for the elements of $\boldsymbol{\beta}$. The candidate distributions are tuned to have acceptance rate near 0.4. We generate 100,000 samples and discard the first 10,000 as burn-in. Convergence is monitored using trace plots of several representative parameters.

5. ANALYSIS OF HEAT WAVES IN THE WESTERN US

To illustrate the proposed methods, we analyze the daily temperature data described in Section 3. An essential step in any threshold exceedance analysis is to select the threshold, μ. Figure 5 plots the mean residual lifetime (MRL) for each city with data pooled across month and year. If the data follow the GPD above μ, then the MRL plot should

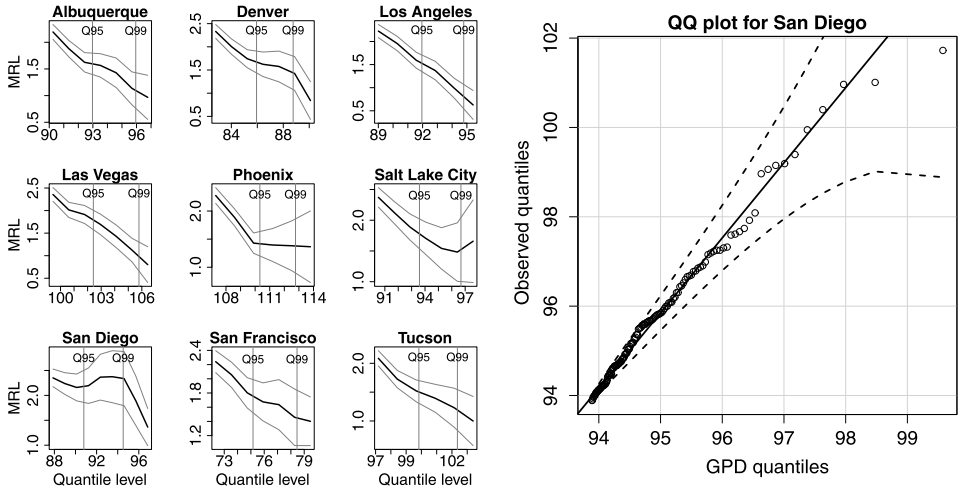

Figure 5. Mean residual lifetime (and 95 % confidence intervals; the vertical lines labeled "Q95" and "Q99" are the 0.95 and 0.99 quantiles) for each city (left), and a qq-plot (with 95 % intervals) comparing the data above the 0.95 quantile and fitted (using maximum likelihood estimation) GPD quantile function for San Diego (right).

be linear above μ. The MRL plot is linear above the 0.95 quantile for most of the cities. San Diego is the least linear. Figure 3 (right panel) compares the sample quantiles with the quantile function of the GPD with parameters fit using maximum likelihood estimation for San Diego. We see that the GPD distribution fits fairly well except for a few very extreme temperatures. Therefore, we proceed with the 0.95 quantile as the threshold for all cities. This MRL analysis assumes stationarity over time, which is of course suspect. We use this as a guide to selecting an overall threshold, and still allow for nonstationarity both in the GDP parameters and threshold exceedance probability to capture changes over time. The Supplementary Materials (Section D) also include results using the more conservative 0.97 quantile as the threshold. The results are qualitatively the same as those presented here for the 0.95 quantile. This provides further evidence that the 0.95 quantile is sufficiently large.

For each city we fit both the Markov and kernel model. Since the kernel and Markov models have the same number of parameters and differ only by the shape of the dependence function, we compare models using simply the posterior mean of the deviance (minus two times the log likelihood). The deviance is minimized by the kernel model for all but one city (San Francisco), so we focus primarily on the kernel model. However, the two models give similar results, as discussed below.

Figure 6 plots the data and fitted quantiles for Las Vegas and San Francisco. The quadratic model seems to capture the seasonal trends in both cities. Extreme events are more concentrated in July in Las Vegas than San Francisco, where extreme temperatures occasionally occur in the spring and fall. In both cities, the 95th and 99th percentiles increase by 2–7 degrees over the course of the study depending on the city and time of year. In addition to changes in the marginal distribution, our model can detect changes in the dependence structure. For example, Figure 7 shows the posterior of the extremal coefficients for Las Vegas and San Francisco. Las Vegas generally has lower extremal coefficient, and thus longer periods of high temperature. In both cases, the extremal coefficients are lower

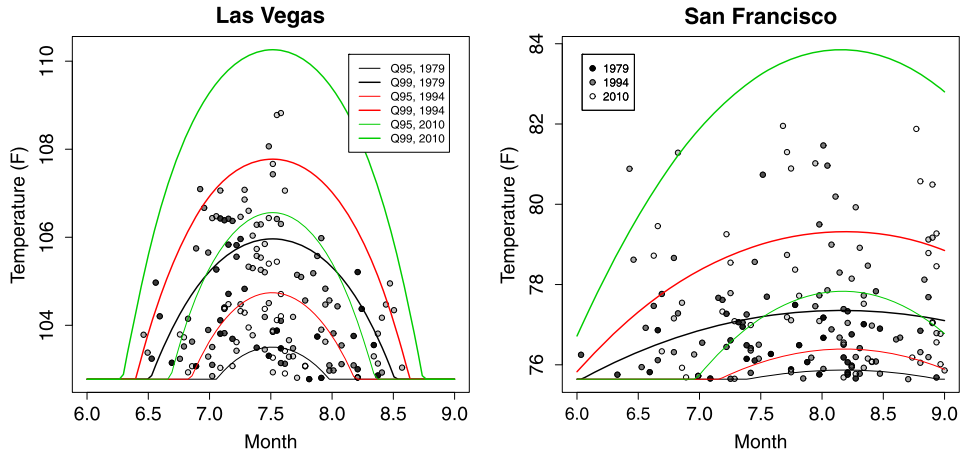

Figure 6. Raw data and fitted quantiles for Las Vegas and San Francisco. The exceedances of the 0.95 quantile are shaded according to the year, and plotted along with the posterior median (from the kernel model) of the 0.95 and 0.99 quantile of the marginal distribution. Month ranges: from 6 for June 1 to 9 for September 1.

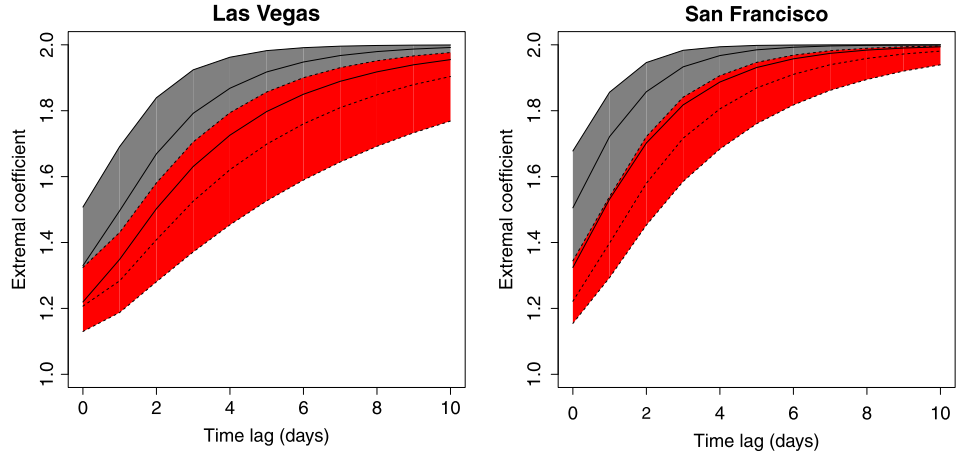

Figure 7. Posterior median and 90 % interval of the extremal coefficients for the kernel model in 1979 (gray) and 2010 (red) (Color figure online).

in 2010 for all lags, indicating longer periods of hot temperatures. Although both have wide interval estimates, these posterior extremal coefficients generally coincide with the sample extremal coefficients in Figure 4.

For a more formal assessment of model adequacy, we use the Bayesian p-value approach (Gelman et al. 2004). We generate a replicate data set from our model at each iteration of the MCMC chain conditioned on the current iteration of π_{it}, σ_{it}, ξ_i, α_i and γ_i (but not the latent A_{il}). For the simulated data set at iteration number $s = 1, \ldots, N$ we then compute summary statistics $D_1, \ldots, D_N$. This distribution is compared to the summary statistic of the observed data, D_0, using the Bayesian p-value $\sum_{s=1}^{N} I(D_s > D_0)/N$. The p-values near zero or one both suggest model inadequacy.

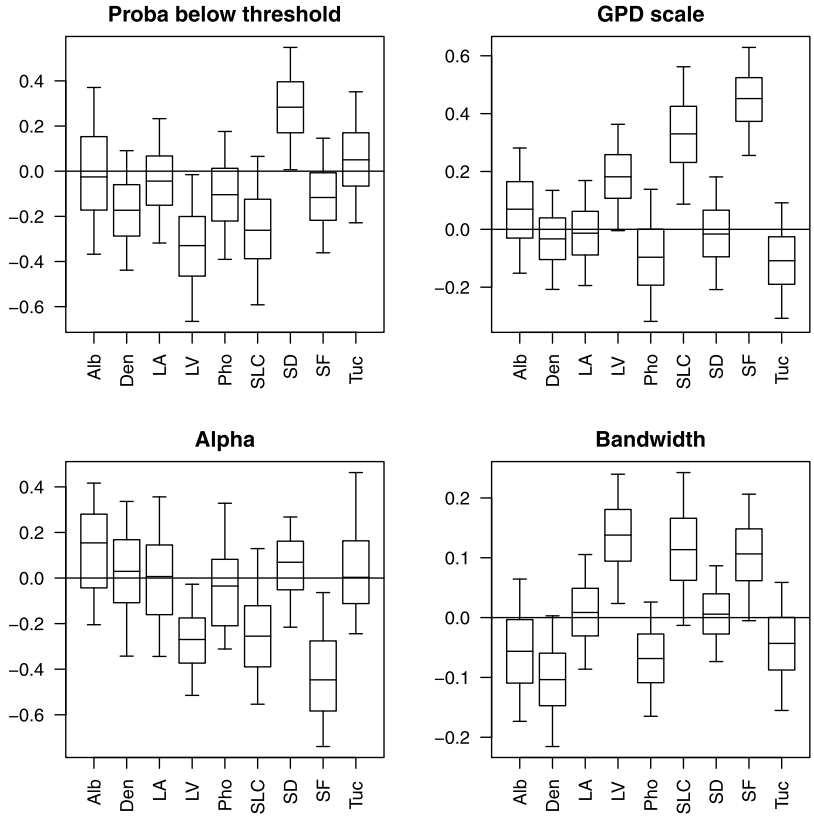

Figure 8. Posterior distribution (horizontal lines give the 0.05, 0.25, 0.50, 0.75, and 0.95 posterior quantiles) for the kernel model by city for the time trends in the probability below the threshold, β_1^{π}, the GPD log scale, β_1^{σ}, dependence parameter, β_1^{α}, and log bandwidth, β_1^{γ}.

The p-values are computed separately for the nine city, and by six time periods: (1) all months and years; (2) 1979–1994 (all seasons); (3) 1995–2010; (4) June (all years); (5) July; and (6) August. We consider four summary statistics: (1) the largest observation; (2) the second largest observation; (3) the number of observations above the threshold; and (4) the number of pairs of consecutive observations above the threshold. The first two are designed to study marginal tail behavior, the third is designed to study the threshold probability, and the fourth studies temporal extremal dependence. This gives a total of $9 * 6 * 4 = 216$ p-values. Of these, only 11 (5.1 %) are less than 0.05 or greater than 0.95. The most extreme p-value (0.98) corresponds to an overestimation July maximum temperature in Phoenix. However, the relatively small number of extreme p-values supports the fit of our model.

Figure 8 shows the posterior distributions of the year effects β_1^{π}, β_1^{σ}, β_1^{α}, and β_1^{γ} in (3.1), which control the evolution of model parameters over time. There are few significant changes over time for six of the nine cities: Albuquerque, Denver, Los Angeles, Phoenix, San Diego, and Tucson. For these cities the most significant effects are a slight increase

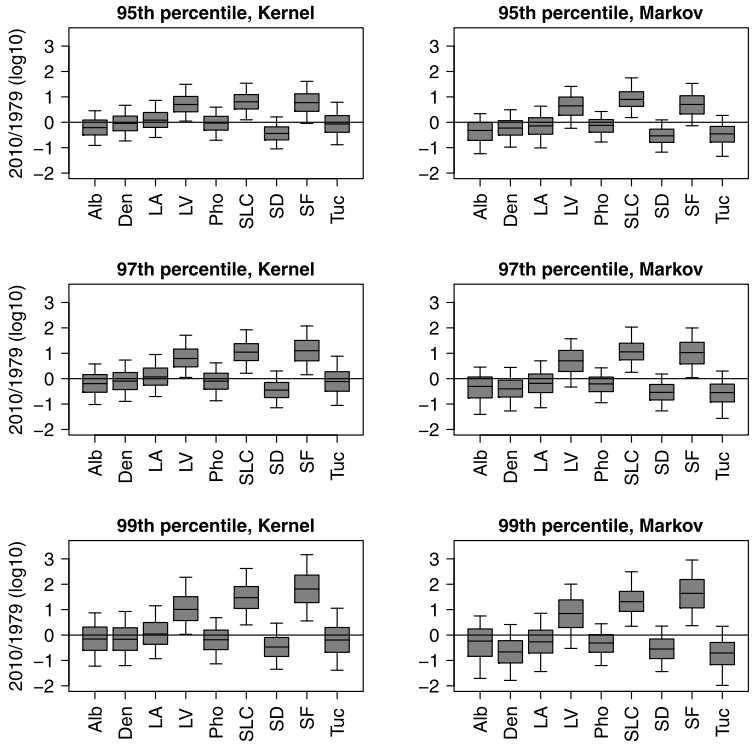

Figure 9. Posterior distribution (horizontal lines give the 0.05, 0.25, 0.50, 0.75, and 0.95 posterior quantiles) by city for ratio of the probability of a heat wave $p(T, h)$ for 2010 relative to 1970. The probabilities are given for $h = 4$ consecutive days above T degrees, where T is taken to be the city's 95th, 97th, and 99th percentiles. Results are presented separately for the kernel and Markov model, and are given on the log base 10 scale.

in the proportion of days below the threshold in San Diego and a decrease in the kernel bandwidth (and thus length of a typical heat wave) in Denver.

Las Vegas, Salt Lake City, and San Francisco each have increases in both extreme marginal quantiles and extremal dependence. Las Vegas shows a decrease in the probability below the threshold, and all three cities show an increase in the GPD scale. In San Francisco, the posterior 90 % interval for time trend for the GPD scale (β_1^{σ}) is (0.221, 0.618), which corresponds to 90 % interval of (25.8, 85.4) percent change per decade. These changes can be seen in Figure 3, where there appears to be an increase over time in the magnitude of the exceedances. In addition to changes in the marginal GPD parameters, these three cities show decreases in smoothness parameter α and increases in the kernel bandwidth γ, both corresponding to stronger extremal dependence.

Figure 9 illustrates the change in probability of a heat wave of four consecutive days above the city's 0.95, 0.97 and 0.99 quantile for the nine cities. The kernel and Markov models produce roughly the same results, and we therefore focus discussion on the kernel model. There are increases in Las Vegas, Salt Lake City, and San Francisco. For Las Vegas, the main changes over time are the probability above the threshold and the extremal coefficient. These changes lead to a change in the probability of a heat wave for all three

quantile levels. For San Francisco, the more dramatic change is in the GPD scale parameter, and therefore the change in probability is larger for the very extreme 0.99 quantile than the 0.95 quantile. The median (posterior 90 % interval) of the ratio of the probability of a heat wave of four days above the 0.97 quantile for 2010 versus 1979 are 9.3 (1.6, 50.3) for Las Vegas, 7.6 (1.1, 63.4) for Salt Lake City, and 14.5 (1.3, 133.5) for San Francisco. These correspond to a dramatic changes in the return period for these events (defined as the inverse of its probability). For San Francisco, the posterior median of the return period decreases from 875.7 years (161.9, 5521.4) in 1979 to 60.3 years (27.4, 170.3) in 2010.

6. CONCLUSIONS

In this paper we have proposed a new framework for analyzing serially-dependent extremes. Our model has nice extremal dependence properties including a closed-form expression for the probability of heat wave, and allows inference on changes over time in the marginal distribution, the extremal dependence function, and the probability of a heat wave. We apply our model to nine cities in the western US, and find significant increases in the probability of a heat wave in Las Vegas, Salt Lake City, and San Francisco.

A potential limitation of our approach is the somewhat restrictive definition of a heat wave as h consecutive days above T degrees. There are certainly other plausible definitions of a heat wave that do not have this form, for example, a period of h consecutive days with average temperature over T degrees. The probability of this event does not have a closed form under our model. Therefore, to estimate probabilities (or changes in the probability) of a heat wave of this form, we would resort to a weather generator, that is, we would estimate the probability of a heat wave by sampling many time series from the model and computing the sample proportion with a heat wave.

A possible step towards improving the proposed temporal model is to include spatial dependence. Extreme values are by definition rare, and thus borrowing information across nearby cities could lead to substantial gains in power for detecting increasing trends. Conceivably, the GPD parameters could be treated as spatial processes, and the residual spatiotemporal dependence could be captured using three-dimensional kernel functions. This would present a computational challenge, but is a promising area of future work.

[Received December 2012. Accepted September 2013. Published Online September 2013.]

REFERENCES

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., Waal, D. D., and Ferro, C. (2004), *Statistics of Extremes: Theory and Applications*, New York: Wiley.

Bobb, J. F., Dominici, F., and Peng, R. D. (2011), "A Bayesian Model Averaging Approach for Estimating the Relative Risk of Mortality Associated with Heat Waves in 105 U.S. Cities," *Biometrics*, 67, 1605–1616.

Bortot, P., and Gaetan, C. (2013, accepted), "A Latent Process Model for Temporal Extremes." *Scandinavian Journal of Statistics*.

Chavez-Demoulin, V., and Davison, A. C. (2012), "Modelling Time Series Extremes," *REVSTAT*, 10, 109–133.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*. *Springer Series in Statistics*, London: Springer.

Coles, S. G., Heffernan, J., and Tawn, J. (1999), "Dependence Measures for Extreme Value Analysis," *Extremes*, 2, 339–365.

Davison, A., and Smith, R. (1990), "Models for Exceedances Over High Thresholds," *Journal of the Royal Statistical Society: Series B*, 393–442.

Easterling, D., Meehl, G., Parmesan, C., Changnon, S., Karl, T., and Mearns, L. (2000), "Climate Extremes: Observations, Modeling, and Impacts," *Science*, 289, 2068–2074.

Fawcett, L., and Walshaw, D. (2006), "Markov Chain Models for Extreme Wind Speeds," *EnvironMetrics*, 17, 795–809.

——— (2007), "Improved Estimation for Temporally Clustered Extremes," *EnvironMetrics*, 18, 173–188.

Ferro, C. A. T., and Segers, J. (2003), "Inference for Clusters of Extreme Values," *Journal of the Royal Statistical Society: Series B*, 65, 545–556.

Fougères, A.-L., Nolan, J. P., and Rootzén, H. (2009), "Models for Dependent Extremes Using Stable Mixtures," *Scandinavian Journal of Statistics*, 36, 42–59.

Frich, P., Alexander, L., Della-Marta, P., Gleason, B., Haylock, M., Klein Tank, A., and Peterson, T. (2002), "Observed Coherent Changes in Climatic Extremes During the Second Half of the Twentieth Century," *Climate Research*, 19, 193–212.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Boca Raton: CRC Press.

de Haan, L., and Ferreira, A. (2006), *Extreme Value Theory. Springer Series in Operations Research and Financial Engineering*, New York: Springer.

Huser, R., and Davison, A. (2013), "Space-Time Modelling of Extreme Events," *Journal of the Royal Statistical Society: Series B*. doi:10.1111/rssb.12035.

Jeon, S., and Smith, R. (2013, submitted), "Dependence Structure of Spatial Extremes Using Threshold Approach." arXiv:1209.6344

Kunkel, K. E., Liang, X.-Z., and Zhu, J. (2010), "Regional Climate Model Projections and Uncertainties of Us Summer Heat Waves" *Journal of Climate*, 23, 4447–4458.

Meehl, G. A., and Tebaldi, C. (2004), "More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century," *Science*, 305, 994–997.

Padoan, S., Ribatet, M., and Sisson, S. (2010), "Likelihood-Based Inference for Max-Stable Processes," *Journal of the American Statistical Association*, 105, 263–277.

Peng, R., Bobb, J., Tebaldi, C., McDaniel, L., Bell, M., and Dominici, F. (2011), "Toward a Quantitative Estimate of Future Heat Wave Mortality Under Global Climate Change," *Environmental Health Perspectives*, 119, 701.

Reich, B., and Shaby, B. (2013), "A Hierarchical Max-Stable Spatial Model from Extreme Precipitation," *The Annals of Applied Statistics*, 6, 1430–1451.

Resnick, S. (1987), *Extreme Values, Regular Variation, and Point Processes*, New York: Springer.

Ribatet, M., Cooley, D., and Davison, A. C. (2011), "Bayesian Inference for Composite Likelihood Models and an Application to Spatial Extremes," *Statistica Sinica*, 22, 813–846.

Schlather, M., and Tawn, J. (2003), "A Dependence Measure for Multivariate and Spatial Extreme Values: Properties and Inference," *Biometrika*, 90, 139–156.

Schlenker, W., and Roberts, M. (2009), "Nonlinear Temperature Effects Indicate Severe Damages to Us Crop Yields Under Climate Change," *Proceedings of the National Academy of Sciences*, 106, 15594–15598.

Shaby, B. A., and Reich, B. J. (2013), "Bayesian Spatial Extreme Value Analysis to Assess the Changing Risk of Concurrent Extremely High Temperatures Across Large Portions of European Cropland," *EnvironMetrics*. doi:10.1002/env.2178.

Smith, R., Tawn, J., and Coles, S. (1997), "Markov Chain Models for Threshold Exceedances," *Biometrika*, 84, 249.

Stephenson, A. G. (2009), "High-Dimensional Parametric Modelling of Multivariate Extreme Events," *Australian & New Zealand Journal of Statistics*, 51, 77–88.