Environmetrics 00, 1–24

DOI: 10.1002/env.XXXX

An Exponential-Gamma Mixture Model for Extreme Santa Ana Winds

Gregory P. Bopp[1]* and Benjamin A. Shaby[1]

Summary: We analyze the behavior of extreme winds occurring in Southern California during the Santa Ana wind season using a latent mixture model. This mixture representation is formulated as a hierarchical Bayesian model and fit using Markov chain Monte Carlo. The two-stage model results in generalized Pareto margins for exceedances and generates temporal dependence through a latent Markov process. This construction induces asymptotic independence in the response while allowing for dependence at extreme, but sub-asymptotic, levels. We compare this model with a frequentist analogue where inference is performed via maximum pairwise likelihood. We use interval censoring to account for data quantization, and estimate the extremal index and probabilities of multi-day occurrences of extreme Santa Ana winds over a range of high thresholds.

Keywords: Extreme value theory; Generalized Pareto distribution; Bayesian hierarchical model; Asymptotic independence; Santa Ana winds.

1. INTRODUCTION

We examine extreme winds occurring in Southern California during the Santa Ana wind season using a stochastic process model with sub-asymptotic extremal dependence in time and generalized Pareto (GPD) margins. This model is constructed as a Gamma process mixture of independent exponential random variables. Because it is easily expressed conditionally as a hierarchical model, Bayesian estimation is a natural choice. The same

[1] *Department of Statistics, Pennsylvania State University, 326 Thomas Building, University Park, PA 16801, USA*

** Correspondence to: Gregory P. Bopp, Department of Statistics, Pennsylvania State University, 326 Thomas Building, University Park, PA 16801, USA. E-mail: gpb951@psu.edu*

This paper has been submitted for consideration for publication in *Environmetrics*

modeling framework recently appeared in Bortot and Gaetan (2014) and Bortot and Gaetan (2016), who, in contrast, estimated model parameters with composite likelihoods using bivariate densities of pairs of observations. Because our implementation is fully Bayesian, it is able to make use of the full data likelihood, constructed hierarchically.

Modeling the dependence characteristics among extremes is often of interest in environmental applications. Rather than assume a Markov structure for the extremes directly, here we will assume exceedances of a high threshold are conditionally independent given an unobserved Markov process. More generally, the latent component may be any stochastic process with Gamma margins, with different choices leading to different extremal dependence characteristics in the resulting mixture process. In addition to describing dependence, the proposed model also ensures that threshold exceedances marginally follow a generalized Pareto limiting distribution.

The dry, high-speed Santa Ana winds occur in Southern California during autumn through spring (Raphael, 2003). They are caused by dry desert air masses in the high-altitude, interior Great Basin region that warm adiabatically as they sink to lower altitudes, and accelerate downslope as they pass through Sierra Nevada Mountain corridors out to the Pacific coast (Hughes and Hall, 2010). One reason they are of interest is their tendency to cause and rapidly spread wildfires near heavily populated urban areas. These high-speed winds can knock down utility poles, causing electrical arcing that can ignite wildfires. This threat has been exacerbated by the watershed management and the fire suppression strategies of the 20th-century, making Southern California one of the most susceptible environments to wildfires in the world (Westerling *et al.*, 2004). Santa Ana winds have a compound effect: in addition to increasing ignition risk, they amplify the already high endemic spread potential that is a result of Southern California's dry, Mediterranean climate and ample fuel sources consisting of low-lying shrubs and grasses. These factors make fires during the Santa Ana wind season especially difficult to contain. In contrast

to California's non-Santa Ana fires, fires during the Santa Ana season consume more of their final burn area faster (Jin *et al.*, 2015), cause appreciable property damage, and tend to result in more civilian fatalities. Between 1990 and 2009, fires during Santa Ana wind seasons were responsible for an estimated $3.1 billion in losses (Jin *et al.*, 2015). Because of the potential these winds have for igniting and spreading fires, modeling the dependence of these extreme winds may be of interest for utility pole design, fire safety regulation (Clopton, 2016), as well as understanding Santa Ana fire diffusion. To address the question of temporal dependence, we develop a model to predict the probability of consecutive multi-day extreme winds.

Two common objectives in extreme value analysis of time series data can be broadly categorized as follows: (1) to understand marginal tail distribution characteristics and (2) to describe the dependence structure among extremes. If one is willing to forgo inference on the dependence in the joint tail distribution, several methods are available for accounting for its effect on marginal quantities (Fawcett and Walshaw, 2007, e.g.), or removing it so that models for independent data can be applied. Alternatively, as we do in this paper, one may attempt to model the dependence explicitly.

For modeling peaks-over-threshold data, Davison and Smith (1990) and Coles *et al.* (1999) propose a runs method for declustering dependent data, in which only maxima of exceedances within a cluster are retained for inference. Discarding observations is generally an inefficient use of data, however. The method of Ferro and Segers (2003) provides an alternative that incorporates additional information about inter-exceedance times to decluster data. One drawback of these methods is that one typically needs to fix a declustering parameter, which can sometimes have a significant impact on the resulting inference. To minimize the effect of choosing a declustering parameter, Fawcett and Walshaw (2008) develop a Bayesian adaptation of the intervals declustering method that incorporates uncertainty in this choice.

Several methods have been proposed that explicitly model tail dependence, rather than remove it. For example, Smith *et al.* (1997) construct a class of first-order Markov models for successive observations using bivariate GPDs. Similarly, Shaby *et al.* (2016) develop a Markov switching model with two states, one for exceedances and one for the bulk of the data, with non-exceedances modeled by a Gaussian AR(1) process and exceedances by a bivariate parametric GPD family. Partitioning the data this way may reflect a belief that there is a fundamental difference in the data generating process for the bulk of the observations and the extremes. Another much broader class of models for extremal dependence can be constructed using copulas (Nelsen, 2006), which have a long history in extreme value analysis (Salvadori, 2007). The justification for copula models comes from Sklar's theorem, which states that any distribution can be decomposed into its marginal distribution and dependence structure components. A relevant example is given by Reich *et al.* (2014), who construct a Bayesian max-stable copula model with dependent random effects. In Section 3, we will compare the fit of their Markov hierarchical max-stable model to a Gaussian ARMA copula model and the latent mixture model presented here on the Santa Ana winds data.

We develop a Bayesian two-stage mixture model for exceedances over a high threshold and apply it to a dataset of daily maximum wind gusts in Southern California. The model is motivated by a hierarchical representation of the GPD as a gamma mixture of exponential distributions (Reiss and Thomas, 2007). Extremal dependence is induced through a two-stage model: (1) a conditionally independent exponential observation model and (2) a latent stochastic process model in time with Gamma margins that controls the dependence. The GPD representation as a mixture of exponential distributions applies only when the shape parameter of the GPD is positive. However, this restriction can be removed by including a marginal transformation of the data as part of the hierarchical model, which effectively uses the Gamma process mixture model as a copula. For different choices of the

latent Gamma process, either asymptotic dependence or independence is possible, with a wide range of dependence types at sub-asymptotic levels.

Choosing a Markov chain structure for the latent process, together with the conditional independence in the observation model, makes inference for this model computationally tractable. Bortot and Gaetan (2016) use a pairwise likelihood (PL) approach for inference by splitting the complete log likelihood into bivariate blocks and summing these blocks to form a composite likelihood objective function. In contrast, we perform inference based on the full likelihood, taking advantage of the hierarchical structure to construct our Markov chain Monte Carlo (MCMC) sampler.

2. BAYESIAN LATENT TEMPORAL MODEL FOR EXTREMES

In this section we present a Bayesian hierarchical model for threshold exceedance time series, where the extremal dependence is modeled through a first-order latent Markov process with Gamma margins. This hierarchical approach allows us to account for serial tail dependence while maintaining GPD margins for the exceedances. The model is motivated by the following representation for the GPD: for $\xi > 0$, the distribution function $G(y; \sigma, \xi)$ can be expressed as a mixture of exponential distribution functions, $H(y; \lambda)$, with respect to a Gamma$(1/\xi, \sigma/\xi)$ mixing density, with shape $1/\xi$ and rate σ/ξ (Reiss and Thomas, 2007). In particular, if $G(y; \sigma, \xi)$ is the distribution function of a GPD with scale σ and shape ξ,

$$G(y; \sigma, \xi) = 1 - (1 + \frac{\xi y}{\sigma})^{-1/\xi}, \qquad \text{for } y > 0 \text{ and } 1 + \frac{\xi y}{\sigma} > 0.$$

the representation theorem says that

$$G(y; \sigma, \xi) = \int_0^\infty H(y; \lambda) f_\lambda(\lambda; 1/\xi, \sigma/\xi) d\lambda \tag{1}$$

for exponential distribution function $H(y;\lambda) = 1 - e^{-\lambda y}$, $y > 0$ and Gamma density $f_\lambda(\lambda; 1/\xi, \sigma/\xi) = \lambda^{1/\xi - 1}(\sigma/\xi)^{1/\xi} e^{-\lambda\sigma/\xi}/\Gamma(1/\xi)$. This theorem shows that if

$$Y|\Lambda \sim \text{Exp}(\Lambda)$$
$$\Lambda|\sigma, \xi \sim \text{Gamma}(1/\xi, \sigma/\xi)$$

then by Equation 1, marginally $Y \sim \text{GPD}(\sigma, \xi)$.

Because the parameters of the Gamma distribution are positive, this mixture representation requires that $\xi > 0$, which corresponds to the heavy-tailed case of the GPD. In many environmental applications, however, the distribution of the data may not possess heavy-tails. To relax this heavy-tail constraint, the probability integral transform can be used to put Y on a scale with arbitrary shape parameter. For $Y \sim \text{GPD}(\sigma, \xi)$, $\sigma > 0$, $\xi > 0$, applying the transformation $Y' = G^{-1}(G(Y; \sigma, \xi); \sigma', \xi')$, $\sigma' > 0$, $\xi' \in \mathbb{R}$, gives $Y' \sim \text{GPD}(\sigma', \xi')$ with unconstrained shape parameter.

Now, let $\{X_t, t \geq 1\}$ be a stationary discrete time random process. Take $u \in \mathbb{R}$ to be a high threshold, and define a threshold-censored sequence of exceedances of u by $Y'_t = \max(0, X_t - u)$. Under relatively weak regularity conditions on the distribution of $\{X_t, t \geq 1\}$, each $Y'_t | X_t > u$ is approximately generalized Pareto distributed (Leadbetter *et al.*, 1983). Since we want the tail to speak for itself, rather than allowing inference to be contaminated by the bulk of the distribution, we model the threshold-censored terms separately from the exceedances. To account for threshold censoring, non-exceedances are modeled by their non-exceedance probability, and exceedances are modeled as GPD.

$$dF(y; \sigma', \xi') = \begin{cases} p, & \text{for } y = 0 \\ (1-p)dG(y; \sigma', \xi') & \text{for } y > 0 \end{cases}$$

where $p = P(X_t \leq 0)$ is the probability of non-exceedance, and $dG(y; \sigma', \xi')$ is the density of a GPD with scale σ' and shape ξ'.

An ideal model accounts for the temporal dependence among $\{Y'_t, t \geq 1\}$ while maintaining that marginally each $Y'_t | X_t > u$ follows a GPD. To achieve this end, we set $Y_t = c(Y'_t)\mathbb{1}_{\{X_t > u\}}$ for

$$c(y) = G^{-1}(G(y; \sigma', \xi'); 1, 1) = (\kappa + 1)\left[\left(1 + \frac{y\xi'}{\sigma'}\right)^{1/\xi'} - 1\right], \qquad \sigma' > 0,\ \xi' \in \mathbb{R}$$

and use the mixture representation presented above to construct a model that accounts for temporal dependence where each Y_t is marginally distributed GPD(1,1). The marginally transformed exceedances $\{Y_t, t \geq 1\}$ are assumed to be conditionally independent and exponentially distributed given a latent process $\{\Lambda_t,\ t \geq 1\}$

$$Y_t | \Lambda_t, X_t > u \sim \text{Exp}(\Lambda_t).$$

Following Bortot and Gaetan (2014), we model the probability of exceedance as

$$P(X_t > u | \Lambda_t) = \exp(-\kappa \Lambda_t).$$

Finally, the latent process $\{\Lambda_t,\ t \geq 1\}$ is taken to be a stochastic process with Gamma$(1, 1)$ margins. We model the latent stage using a stationary, first-order Markov process (Warren, 1992), which we will refer to as the Warren process. In its general form, the Warren processes is defined as

$$\begin{aligned}
\Lambda_1 &\sim \text{Gamma}\,(1/\xi, \sigma/\xi) \\
\Pi_t | \Lambda_{t-1} &\sim \text{Poisson}\left(\Lambda_{t-1}\frac{\sigma\rho}{\xi(1-\rho)}\right), \qquad 0 \leq \rho < 1 \\
\Lambda_t | \Pi_t &\sim \text{Gamma}\left(\Pi_t + \frac{1}{\xi}, \frac{\sigma}{\xi(1-\rho)}\right).
\end{aligned}$$

Here, however, we take $\sigma = \xi = 1$, which gives the desired Gamma$(1, 1)$ margins for each Λ_t, while allowing for dependence among the $\{\Lambda_t,\, t \geq 1\}$. The dependence among exceedances is controlled through ρ, with higher values of ρ corresponding to stronger levels of dependence. For all fixed lags k, this process has autocorrelation function $\text{corr}(\Lambda_t, \Lambda_{t+k}) = \rho^{|k|}$.

A common measure of dependence in extremes is the tail dependence parameter, which describes the conditional probability that one random variable is extreme when the other is extreme. For two random variables Z_1 and Z_2 with distribution functions F_{Z_1} and F_{Z_2}, the tail dependence parameter at level u is defined by $\chi(u) = P(Z_1 > F_{Z_1}^{-1}(u) | Z_2 > F_{Z_2}^{-1}(u))$. The tail dependence parameter is defined by $\chi = \lim_{u \to 1} \chi(u)$. Two random variables are said to be asymptotically independent when $\chi = 0$ and dependent otherwise. Although the latent mixture model leads to dependence among the $\{Y_t',\, t \geq 1\}$ for all $\rho \neq 0$ at all finite levels, the Warren process induces asymptotic independence in the response, an assumption supported by exploratory analysis of the Santa Ana wind data (Figure 1). Different models for $\{\Lambda_t,\, t \geq 1\}$ lead to different extremal dependence characteristics, including asymptotic dependence (Bortot and Gaetan, 2014).

Defining the latent mixture model on the transformed $\{Y_t,\, t \geq 1\}$ allows the data, $\{Y_t',\, t \geq 1\}$, to be marginally GPD(σ', ξ'), $\sigma' > 0$, $\xi' \in \mathbb{R}$ with unconstrained tail parameter. This transformation also has the benefit of separating the parameters that control marginal distribution dependence characteristics. This separation makes it possible to more cleanly further model the parameters as functions of covariates.

If we let Z_t be an indicator for whether the tth observation is a threshold exceedance, $Z_t = \mathbb{1}_{\{X_t > u\}}$ and denote the parameter vector by $\theta = (\sigma', \xi', \rho, \kappa)$, the full data likelihood for this model is

$$f_{y',z}(\mathbf{y}'_{1:n}, \mathbf{z}_{1:n}|\theta) = \int \left(\prod_{t=1}^{n} [\lambda_t \exp\{-\lambda_t c(y'_t)\} \frac{d}{dy'_t} c(y'_t)]^{z_t} \prod_{t=1}^{n} [\exp\{-\kappa\lambda_t\}]^{z_t} [1 - \exp\{-\kappa\lambda_t\}]^{1-z_t} \cdot f_\lambda(\boldsymbol{\lambda}_{1:n}; \rho) \right) d\lambda_1 \dots d\lambda_n$$

where $f_\lambda(\boldsymbol{\lambda}_{1:n}; \rho)$ is the joint density of $\Lambda_1, \dots, \Lambda_n$, and

$$\frac{d}{dy'_t} c(y'_t) = \frac{\kappa + 1}{\sigma'} \left(1 + \frac{y'_t \xi'}{\sigma'}\right)^{1/\xi' - 1}.$$

To complete the model, we specify weakly informative normal and truncated normal priors for ξ', σ' and κ, and a Unif(0,1) prior for ρ. Hereafter we refer to the Bayesian latent mixture model as M_{BLM} and the PL one proposed by Bortot and Gaetan (2016) as M_{PLM}.

2.1. Data Quantization

Data quantization, or rounding, are common in environmental data and can have an effect on tail inference. Deidda and Puliga (2009) investigated the performance of several GPD scale and shape estimators under varying degrees of data quantization and demonstrated increased bias for pronounced cases. One way to account for data quantization is to model threshold exceedances as interval censored realizations of continuous GPD random variables. The M_{BLM} can easily be modified to incorporate measurement uncertainty caused by data quantization. Due to the hierarchical construction, only the form of the observation model changes, wherein the exponential density is replaced by an integral over the censoring interval.

Suppose that $\{W'_t,\, t \geq 1\}$ is an interval censored version of $\{Y'_t,\, t \geq 1\}$ such that the censoring occurs in evenly spaced intervals of $2\epsilon > 0$, where each interval centered at values

on the support of the distribution of W'_t, $\{2\epsilon, 4\epsilon, \ldots\}$. Then denoting

$$w_t^+ = G^{-1}(G(w'_t + \epsilon; \sigma', \xi'); 1, 1) \text{ and } w_t^- = G^{-1}(G(w'_t - \epsilon; \sigma', \xi'); 1, 1), \qquad \text{for } w'_t > 0$$

the conditional observation model becomes

$$\begin{aligned} P(W'_t = w'_t | X_t > u + \epsilon, \Lambda_t) &= P(w'_t - \epsilon < Y'_t \leq w'_t + \epsilon | X_t > u + \epsilon, \Lambda_t) \\ &= P(w_t^- < Y_t \leq w_t^+ | X_t > u + \epsilon, \Lambda_t) \\ &= H(w_t^+ | \Lambda_t) - H(w_t^- | \Lambda_t) \\ &= \exp\{-\Lambda_t w_t^-\} - \exp\{-\Lambda_t w_t^+\} \end{aligned}$$

where H is an exponential distribution function with rate λ_t. Additionally, censoring at the threshold is accounted for by modeling the conditional exceedance probability as

$$P(X_t > u + \epsilon | \Lambda_t) = \exp(-\kappa \Lambda_t).$$

In this case setting $Z_t = \mathbb{1}_{\{X_t > u+\epsilon\}}$, the corresponding full data likelihood is

$$\begin{aligned} f_{w',z}(\mathbf{w}'_{1:n}, \mathbf{z}_{1:n} | \theta) = \int \Bigg(&\prod_{t=1}^{n} [\exp\{-\lambda_t w_t^-\} - \exp\{-\lambda_t w_t^+\}]^{z_t} \prod_{t=1}^{n} [\exp\{-\kappa \lambda_t\}]^{z_t} [1 - \exp\{-\kappa \lambda_t\}]^{1 - z_t} \\ &\cdot f_\lambda(\boldsymbol{\lambda}_{1:n}; \rho) \Bigg) d\lambda_1 \ldots d\lambda_n \end{aligned} \tag{2}$$

The PL approach to inference on the frequentist interpretation of this model (i.e. up to the exclusion of the priors and interpretation of θ, see Bortot and Gaetan (2016)) can also easily be modified to account for data quantization. Rather than integrating over the latent parameters, the PL approach replaces the full likelihood given in Equation 2 by a PL, which is simpler to evaluate and still captures some of the dependence characteristics of the model. Using the same interval censoring approach, the maximum PL estimator is

found by maximizing

$$PL_{IC}(\theta) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{\min(i+\Delta,n)} f(\tilde{w}_i, \tilde{w}_j; \theta)$$

where $f(\tilde{w}_i, \tilde{w}_j; \theta)$ is the joint density of (W_i', Z_i, W_j', Z_j) and Δ is a tuning parameter for the maximum lag between observations. These pairwise joint densities can be expressed in terms of Laplace transforms of $\{\Lambda_t, t \geq 1\}$ (see Appendix A for details). The interval censored Bayesian and PL models are hereafter referred to as M_{BLMC} and M_{PLMC} respectively.

2.2. A Comparison of Pairwise Likelihood and Bayesian Estimators

In this section we assess the performance of Bayesian posterior mean and maximum PL estimators for the continuous observation models, M_{BLM} and M_{PLM}. For details on how the PL estimator and confidence intervals are computed, see Bortot and Gaetan (2014) and Bortot and Gaetan (2016). The Bayesian model is fit using Markov chain Monte Carlo, which has the advantage of using the full data likelihood but comes with a substantial computational burden, whereas the maximum PL is quick to compute but sacrifices information about the parameters that is not explained by pairs of observations. For this comparison, 1,000 datasets are simulated according to the continuous model described in Section 2 with $n = 1,000$ observations and parameter settings $\sigma' = 2.5, \xi' = -0.15, \rho = 0.7$, and $\kappa = 9$ (MCMC details described in Appendix B). The PL is calculated by taking the product of the joint density of pairs of observations up to a maximum lag of 4 time steps apart. For the Bayesian model, the following priors were used $\sigma' \sim \text{TN}_+(0, 10^3)$, $\xi' \sim \text{N}(0, 1)$, $\rho \sim \text{Unif}(0, 1)$, and $\kappa \sim \text{TN}_+(0, 10^3)$, where $\text{TN}_+(\mu, \tau)$ is a truncated normal distribution with location $\mu \in \mathbb{R}$, scale $\tau > 0$, and support $\mathbb{R}_+$. Bayesian 95% HPD credible intervals and PL 95% confidence intervals are compared based on their coverage and

interval scores (Gneiting and Raftery, 2007) and parameter estimates are compared based on their bias and RMSE (Table 1). An interval score is a proper scoring rule that rewards narrower intervals but penalizes those that miss the true parameter.

[Table 1 about here.]

The simulation shows comparable performance between the two methods in terms of bias and RMSE. However, the coverages of the Bayesian HPD credible intervals are consistently closer to the nominal level than the PL confidence intervals, and the interval scores are smaller for the Bayesian credible intervals than the PL confidence intervals. If one is only concerned with accurate point estimates, the additional computational burden associated with MCMC may not be justifiable. However, for uncertainty quantification, the Bayesian approach gives more accurately calibrated intervals.

3. APPLICATION TO SANTA ANA WINDS DATA

The data for this application come from the Hadley Center (available at http://www.metoffice.gov.uk/hadobs/hadisd/index.html). The observations consist of daily maximum wind speeds at the March Air Reserve Base weather station in Perris, California during the years 1974-2014, which have been rounded to the nearest 1/2 m/sec. Because the Santa Ana winds tend to prevail between October and April, we restrict our analysis to this period. Furthermore, since consecutive Santa Ana seasons are separated by 5 months, we assume seasons to be independent. To illustrate the pattern of daily maximum wind speeds, a subset of the data from October, 2008 through April, 2012 are plotted in Figure 1. Although the bulk of the data show clear seasonal patterns, with the most day-to-day variation in wind speed occurring during the winter, the plotted time series supports the assumption of stationarity for exceedances during the Santa Ana seasons, which we make for the remainder of the paper. Because the latent Warren process model

induces asymptotic independence, we check empirical estimates of the lag-1 tail dependence parameter but here as a function of increasing thresholds u since the data are assumed to have common margins (i.e. $\chi_1(u) = P(X_{t+1} > u | X_t > u)$). As the empirical estimate of $\chi_1(u)$ appears to be decreasing with higher wind speeds (Figure 1), plausibly converging towards 0, the assumption of asymptotic independence appears to be reasonable. The step-like behavior of the empirical estimate is due to the quantization of recorded wind speeds.

[Figure 1 about here.]

For this study we compare Bayesian and PL models with and without interval censoring to account for the possible effect of data quantization. Since the data are rounded to the nearest 1/2 m/sec, censoring intervals are taken to be centered at the recorded observation and of length 1/2. Each model is fit over a range of high thresholds (0.9, 0.925, and 0.95 empirical quantiles) to assess parameter stability and the plausibility of the GPD assumption. Higher thresholds beyond the 0.95 quantile are not considered so as to preserve enough data to inform the latent process. To fit M_{BLM} and M_{BLMC}, we use variable-at-a-time Metropolis-Hastings algorithms with adaptively chosen proposal variances (Shaby and Wells, 2010) (MCMC details described in Appendix B). The following priors were specified for the remaining parameters: $\sigma' \sim \text{TN}_+(0, 10^3)$, $\xi' \sim \text{N}(0, 1)$, and $\rho \sim \text{Unif}(0, 1)$. To improve the mixing of the sampler, κ is fixed at its empirical quantity, which is a function of the marginal probability of exceedance

$$1 - p = P(X_t > u) = E[e^{-\kappa \Lambda_t}] = 1/(1 + \kappa)$$

For example, in the case of $\hat{p} = 0.9$, we fix $\hat{\kappa} = (1/(1 - \hat{p}) - 1) = 9$ (replace $P(X_t > u)$ by $P(X_t > u + \epsilon)$ for the interval censored model).

Estimates for the PL models are found by maximizing the objective function consisting of

all pairwise densities for pairs y'_t and $y'_{t'}$ (or w'_t and $w'_{t'}$) within $|t - t'| \leq \Delta = 4$ as described in Bortot and Gaetan (2016). Other choices of this tuning parameter give essentially the same results. Confidence intervals are constructed using the Godambe information approximation also described therein. For the sake of comparison with the Bayesian models, κ is also fixed at its empirical quantity.

Bayesian posterior means and 95% HPD credible intervals, and maximum PL estimates and 95% confidence intervals are reported in Table 2. The parameter estimates are consistent between Bayesian and PL models. Moreover, data quantization does not appear to have a strong affect on inference in this case, as parameter estimates for interval censored and continuous models are very similar. Since the fits are relatively stable across thresholds, the M_{BLM} and M_{PLM} fit using the 0.9 empirical quantile threshold (7.5 m/sec) are used for the remainder of the analysis. For all fits the shape parameter estimate is less than zero, corresponding to the light tailed regime of the GPD family. The observed light tailed behavior agrees with an extreme wind analysis performed by Fawcett and Walshaw (2006) over a region of central and northern England. The estimated latent process autocorrelation parameter $\hat{\rho}$ is fairly high, indicating fairly strong dependence among successive λ_t and, which we will see, corresponds to moderate dependence among extremes at sub-asymptotic levels.

[Table 2 about here.]

To visualize the relationship between the latent process and the wind exceedances, posterior means and 95% credible intervals of $c^{-1}(1/\lambda_t)$ during the 2011-2012 Santa Ana season, overlaid with the observed exceedances, are plotted in Figure 2. Since $1/\lambda_t$ correspond to the conditional means of $Y_t|\lambda_t$, applying this marginal transformation puts the latent component on the same scale as the observations $\{Y'_t, t \geq 1\}$. As expected, the λ_t corresponding to long periods without exceedances exhibit greater variation than those near exceedances, and spikes in the transformed latent mean correspond to large

exceedances.

[Figure 2 about here.]

To assess the marginal fit of M_{BLM}, we examine a QQ-plot for exceedances (Figure 3). The points lie very close to the line of unit slope, indicating a very good marginal fit to the data. We also examine the dependence characteristics of the Bayesian and PL fits and compare them to two other Bayesian models for extremal time series data: (1) the Markov hierarchical max-stable model of Reich *et al.* (2014), hereafter referred to as M_{HMS}, which is similar in that it also uses a mixture representation and also incorporates marginal transformations, but qualitatively different in that it produces asymptotic dependence, and (2) a Bayesian Gaussian ARMA(1,1) copula model, hereafter referred to as M_{GAC}, formulated as follows:

Let $\{X_t, t \geq 1\}$ be a stationary discrete time random process. Take $u \in \mathbb{R}$ to be a high threshold, and define a threshold-censored sequence of exceedances of u by $Y_t' = \max(0, X_t - u)$. When $X_t > u$ set $Y_t = \Phi^{-1}(G(Y_t'; \sigma', \xi'))$, where Φ is the quantile function of a standard normal distribution, model $\{Y_t, t \geq 1\}$ as an ARMA(1,1) process with $N(0,1)$ innovations, so that the Y_t are latent when $X_t \leq u$. That is, for $t = 0, \pm 1, \pm 2, \ldots$

$$Y_t = \phi Y_{t-1} + \epsilon_t + \theta \epsilon_{t-1}, \qquad \phi \neq 0,\ \theta \neq 0$$
$$\epsilon_t \overset{iid}{\sim} N(0,1)$$

Finally, a Unif(0,1) is placed on $p = P(X_t > u)$, the probability of exceeding the threshold u, and $\sigma' \sim \mathrm{TN}_+(0, 10^3)$, $\xi' \sim \mathrm{N}(0,1)$ priors are used for the marginal GPD scale and shape parameters. It is physically justifiable that winds have positive serial autocorrelation. To ensure that the ARMA(1,1) model is stationary ($|\phi| < 1$), invertible ($|\theta| < 1$), and has non-negative lag-1 autocorrelation ($\theta \leq \phi$), the following prior is used for ϕ and θ: $\pi(\phi, \theta) = (1/2)\mathbb{1}_{\{-1 < \theta \leq \phi < 1\}}$ (Shumway and Stoffer, 2017). Unlike the hierarchical max-stable

model, the Gaussian ARMA copula model induces asymptotic independence among the $\{Y'_t, t \geq 1\}$ (Sibuya, 1960).

[Figure 3 about here.]

For a graphical assessment of model fit, we compare the concordance between empirical and estimates of dependence measures. In the case of the M_{PLM}, confidence intervals for dependence measures are constructed by the applying the block bootstrap to the data, where each block consists of an entire Santa Ana season. For each bootstrapped sample, the PL model is fit and a time series of observations are simulated from the model using the estimated parameters. Dependence measures are calculated for each sample of simulated data to produce a sampling distribution from which bootstrap confidence intervals can be calculated. Estimates are based on 1,000 bootstrap replicates. For the M_{BLM}, M_{HMS}, and M_{GAC} models, posterior predictive estimates of dependence measures are constructed as follows: for each MCMC sample $\theta^{(m)}$, a time series of observations $Y_{1:n}^{(m)}$ are simulated conditional on $\theta^{(m)}$. The collection of dependence measures calculated for each sample $Y_{1:n}^{(m)}$ of draws from the posterior predictive distribution form a reference distribution from which credible intervals can be constructed. If the empirical estimate of some dependence summary falls within a posterior credible set, that is an indication that the model adequately captures the particular aspect of the data that the summary measure elucidates (Gelman *et al.*, 1996).

First, estimates of $\chi_1(u)$ for these four models are compared with the empirical estimate (Figure 3). The credible and confidence intervals for M_{BLM} and M_{PLM} overlap with the 95% confidence intervals based on the empirical estimator over the range of the data. The reported empirical confidence intervals for the conditional probability $\chi_1(u)$ are the usual confidence intervals for a proportion based on the normal approximation to the binomial distribution (i.e. $\hat{\chi}_1(u) \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\chi}_1(u)(1-\hat{\chi}_1(u))/n_u}$ where n_u is the number of observations above u). The intervals for all four models decay to 0 beyond the range of the

data. This is somewhat unexpected for M_{HMS}, since it induces asymptotic dependence. However this behavior may be explained by the fact that the sub-asymptotic dependence visually appears to be weak among wind extremes, causing the model to infer a small but non-zero value for $\chi_1(u)$.

Next, we consider the behavior of the extremal index (Leadbetter *et al.*, 1983), which corresponds to the limiting reciprocal mean cluster size of exceedances, for increasing thresholds. Applying the extRemes R package implementation of the runs method (Coles *et al.*, 1999) at lengths of 2 and 5, we estimate the extremal index on the original time series over the 0.9 to 0.99 wind speed quantiles (Figure 4). As mentioned in Section 2, the Warren process induces asymptotic independence among extremes, and as expected, for high quantiles the credible interval under this model for the extremal index includes 1, corresponding to no clustering. While all four models cover roughly the same region at the upper bound of the empirical distribution, the M_{BLM} and M_{PLM} models appear to capture the trajectory of the empirical extremal index at sub-asymptotic levels, whereas the M_{HMS} and M_{GAC} models slightly overshoot the empirical estimates of the extremal index.

[Figure 4 about here.]

Sustained high-speed winds over several days pose a greater risk for spreading fires quickly. Figure 5 displays the empirical and model 95% credible intervals for successive multi-day exceedances of a high threshold inferred by the M_{BLM}. The probability of extreme wind for more than three days appears to be negligible at all high thresholds. The probabilities for 2 and 3 day exceedances also decay relatively quickly for increasing thresholds, but the probability of multi-day exceedances at sub-asymptotic levels may be of interest for developing fire suppression strategies or fire safety regulations. The credible intervals show good coverage of the empirical estimates for 1-3 day exceedance probabilities.

[Figure 5 about here.]

In addition to graphical assessments, we also compute two numerical model criteria, the Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002) and Log Pseudo-Marginal Likelihood (LPML) (Gelfand and Dey, 1994). Both are calculated for the three Bayesian models. Lower values of DIC and higher values of LPML correspond to better models. M_{BLM} shows the best fit in terms of both model fit criteria, followed by M_{GAC} (Table 3).

[Table 3 about here.]

4. DISCUSSION

Understanding the marginal and dependence characteristics among temporal extremes is often central to the design of infrastructure and regulations. In this paper we present a Bayesian hierarchical model for temporal dependence among extremes that incorporates both model uncertainty as well as variability in the data generating process. The model is constructed using a Gamma mixture representation of the GPD. While it induces asymptotic independence among extremes, a broad class of dependence types are possible at sub-asymptotic levels. Marginal transformations make it possible to model both light and heavy tailed data and to further model marginal parameters as functions of covariates. This model does not, however, allow for a smooth transition between asymptotic independence and dependence cases. It may be possible to extend this model to the spatial setting by replacing the latent Gamma process with a Gamma random field (Wolpert and Ickstadt, 1998). However, considering that the latent process mixes relatively slowly when considering only a single, temporal dimension, such a model may be difficult to fit when extended to higher order, spatial dimensions. Although the computational burden associated with fitting the Bayesian model is much greater than that of the PL model, a simulation study shows similar estimation performance between the two models, with

slightly better calibrated uncertainty intervals for the Bayesian model. We apply the Bayesian and PL models to a dataset of daily maximum wind gusts in Southern California during the Santa Ana season. The fitted model quantiles and dependence characteristics appear to agree well with the empirical data. Fits show a rapidly decaying probability of consecutive days of extreme winds for increasing thresholds.

Acknowledgements

We gratefully acknowledge support from NOAA NA16NWS4620043 and DOE DE-AC02-05CH11231, which made this work possible. We thank David Sapsis of the California Department of Forestry and Fire Protection for suggesting the Santa Ana analysis.

REFERENCES

Bortot P, Gaetan C, 2014. A latent process model for temporal extremes. *Scand. J. Stat.* **41**(3): 606–621.

Bortot P, Gaetan C, 2016. Latent process modelling of threshold exceedances in hourly rainfall series. *J. Agric. Biol. Environ. Stat.* **21**(3): 531–547.

Clopton AKV, 2016. Order instituting rulemaking to develop and adopt fire-threat maps and fire-safety regulations. State of California Public Utilities Commission, docket # R.15-05-006.

Coles S, Heffernan J, Tawn J, 1999. Dependence measures for extreme value analyses. *Extremes* **2**(4): 339.

Davison AC, Smith RL, 1990. Models for exceedances over high thresholds. *J. Roy. Statist. Soc. Ser. B* **52**(3): 393–442, with discussion and a reply by the authors.

Deidda R, Puliga M, 2009. Performances of some parameter estimators of the generalized Pareto distribution over rounded-ff samples. *Phys. Chem. Earth* **34**(10): 626–634.

Fawcett L, Walshaw D, 2006. A hierarchical model for extreme wind speeds. *J. Roy. Statist. Soc. Ser. C* **55**(5): 631–646.

Fawcett L, Walshaw D, 2007. Improved estimation for temporally clustered extremes. *Environmetrics* **18**(2): 173–188.

Fawcett L, Walshaw D, 2008. Bayesian inference for clustered extremes. *Extremes* **11**(3): 217–233.

Ferro CAT, Segers J, 2003. Inference for clusters of extreme values. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**(2): 545–556.

Gelfand AE, Dey DK, 1994. Bayesian model choice: asymptotics and exact calculations. *J. Roy. Statist. Soc. Ser. B* **56**(3): 501–514.

Gelman A, Meng XL, Stern H, 1996. Posterior predictive assessment of model fitness via realized discrepancies. *Statist. Sinica* **6**(4): 733–807, with comments and a rejoinder by the authors.

Gneiting T, Raftery AE, 2007. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102**(477): 359–378.

Hughes M, Hall A, 2010. Local and synoptic mechanisms causing southern California's Santa Ana winds. *Clim. Dynam.* .

Jin Y, Goulden M, Faivre N, Veraverbeke S, Sun F, Hall A, Hand M, Hook S, Randerson J, 2015. Identification of two distinct fire regimes in southern california: implications for economic impact and future change. *Environ. Res. Lett.* **10**(9).

Leadbetter MR, Lindgren G, Rootzén H, 1983. *Extremes and related properties of random sequences and processes*. Springer Series in Statistics, Springer-Verlag, New York-Berlin, xii+336 pp.

Nelsen RB, 2006. *An introduction to copulas*. Springer Series in Statistics, Springer, New York, second edition, xiv+269 pp.

Raphael MN, 2003. The Santa Ana winds of California. *Earth Interact.* **7**(8): 1–13.

Reich BJ, Shaby BA, Cooley D, 2014. A hierarchical model for serially-dependent extremes: a study of heat waves in the western US. *J. Agric. Biol. Environ. Stat.* **19**(1): 119–135.

Reiss RD, Thomas M, 2007. *Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields*. Birkhäuser Verlag, Basel, third edition, xviii+511 pp., with 1 CD-ROM (Windows).

Salvadori G, 2007. *Extremes in nature: an approach using copulas*. Springer, Dordrecht, Netherlands.

Shaby B, Wells M, 2010. Exploring an adaptive Metropolis algorithm. Technical Report 1011-14, Duke University Department of Stastical Science.

Shaby BA, Reich BJ, Cooley D, Kaufman CG, 2016. A Markov-switching model for heat waves. *Ann. Appl. Stat.* **10**(1): 74–93.

Shumway RH, Stoffer DS, 2017. *Time series analysis and its applications*. Springer Texts in Statistics, Springer, Cham, fourth edition, xiii+562 pp., with R examples.

Sibuya M, 1960. Bivariate extreme statistics. I. *Ann. Inst. Statist. Math. Tokyo* **11**: 195–210.

Smith RL, Tawn JA, Coles SG, 1997. Markov chain models for threshold exceedances. *Biometrika* **84**(2): 249–268.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A, 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**(4): 583–639.

Warren D, 1992. A multivariate gamma distribution arising from a Markov model. *Stoch. Hydrol. Hydraul.* **6**(3): 183–190.

Westerling AL, Cayan DR, Brown TJ, Hall BL, Riddle LG, 2004. Climate, Santa Ana winds and autumn wildfires in southern California. *Eos, Trans. Am. Geophys. Union.* **85**(31): 289.

Wolpert RL, Ickstadt K, 1998. Poisson/gamma random field models for spatial statistics. *Biometrika* **85**(2): 251–267.

APPENDIX

A. INTERVAL CENSORED PAIRWISE LIKELIHOOD DENSITIES

In this section we derive the form of the bivariate densities for the interval censored PL model. Since we have fixed $\sigma = \xi = 1$ in the Warren process, the univariate Laplace transform for Λ_t is

$$LP^{(1)}(s) = 1/(1+s)$$

and the bivariate Laplace transform of Λ_t and $\Lambda_{t+\Delta}$ is

$$LP^{(2)}_{\Delta}(s_1, s_2) = \frac{1 + s_2\rho^{\Delta}}{(1+s_2)(1+s_1+s_2\rho^{\Delta})}$$

The four cases of the pairwise joint densities are given by

- $z_i = z_j = 1$

$$\begin{aligned} f(\tilde{w}_i, \tilde{w}_j; \theta) &= \int [\exp\{-\lambda_i(w_i^- + \kappa)\} - \exp\{-\lambda_i(w_i^+ + \kappa)\}] \\ &\qquad \cdot [\exp\{-\lambda_j(w_j^- + \kappa)\} - \exp\{-\lambda_j(w_j^+ + \kappa)\}] g(\lambda_i, \lambda_j; \rho) d\lambda_i d\lambda_j \\ &= LP^{(2)}_{j-i}(w_i^- + \kappa, w_j^- + \kappa) + LP^{(2)}_{j-i}(w_i^+ + \kappa, w_j^+ + \kappa) \\ &\qquad - LP^{(2)}_{j-i}(w_i^- + \kappa, w_j^+ + \kappa) - LP^{(2)}_{j-i}(w_i^+ + \kappa, w_j^- + \kappa) \end{aligned}$$

- $z_i = 1, z_j = 0$

$$\begin{aligned} f(\tilde{w}_i, \tilde{w}_j; \theta) &= \int [\exp\{-\lambda_i(w_i^- + \kappa)\} - \exp\{-\lambda_i(w_i^+ + \kappa)\}] \\ &\qquad \cdot [1 - \exp\{-\lambda_j \kappa\}] g(\lambda_i, \lambda_j; \rho) d\lambda_i d\lambda_j \\ &= LP^{(1)}(w_i^- + \kappa) - LP^{(1)}(w_i^+ + \kappa) \\ &\qquad - LP^{(2)}_{j-i}(w_i^- + \kappa, \kappa) + LP^{(2)}_{j-i}(w_i^+ + \kappa, \kappa) \end{aligned}$$

- $z_i = 0, z_j = 1$

$$\begin{aligned} f(\tilde{w}_i, \tilde{w}_j; \theta) &= \int [1 - \exp\{-\lambda_i \kappa\}] \cdot \\ &\qquad [\exp\{-\lambda_j(w_j^- + \kappa)\} - \exp\{-\lambda_j(w_j^+ + \kappa)\}] g(\lambda_i, \lambda_j; \rho) d\lambda_i d\lambda_j \\ &= LP^{(1)}(w_j^- + \kappa) - LP^{(1)}(w_j^+ + \kappa) \\ &\qquad - LP^{(2)}_{j-i}(\kappa, w_j^- + \kappa) + LP^{(2)}_{j-i}(\kappa, w_j^+ + \kappa) \end{aligned}$$

- $z_i = 0, z_j = 0$

$$\begin{aligned} f(\tilde{w}_i, \tilde{w}_j; \theta) &= \int [1 - \exp\{-\lambda_i \kappa\}] \cdot [1 - \exp\{-\lambda_j \kappa\}] \\ &= 1 - 2LP^{(1)}(\kappa) + LP^{(2)}_{j-i}(\kappa, \kappa) \end{aligned}$$

B. MCMC DETAILS

We perform MCMC for the model in Section 2 using a variable-at-a-time Metropolis algorithm to draw posterior samples. For the simulation study the parameters consist of $\theta = (\sigma', \xi', \rho, \kappa)$ in addition to the latent variables $\Lambda_{1:n}$ and $\Pi_{2:n}$, where n is the number of observations in the time series (for the simulation study $n = 1,000$, and for the data application $n = 8,700$). In the case of the data application, κ is fixed at its empirical value to improve the mixing of the sampler. Sampling is done by initializing each parameter, and conditionally on the remaining parameters, updating each in turn. All updates are done using a Gaussian random walk proposal distribution. For example, to update σ', a candidate sample $\sigma'^{(*)} \sim \mathrm{N}(\sigma'^{(m-1)}, s^2_{\sigma'})$, where $\sigma'^{(m-1)}$ is the value of the $(m-1)^{th}$ MCMC iteration, and $s^2_{\sigma'}$ is the proposal variance tuning parameter. For observations $(\mathbf{y'}_{1:n}, \mathbf{z}_{1:n})$, denote the observation model density for M_{BLM} as

$$f(\mathbf{y'}_{1:n}, \mathbf{z}_{1:n}|\sigma', \xi', \kappa, \lambda_{1:n}) = \prod_{t=1}^{n}[\lambda_t \exp\{-\lambda_t c(y'_t)\}\frac{d}{dy'_t}c(y'_t)]^{z_t} \prod_{t=1}^{n}[\exp\{-\kappa\lambda_t\}]^{z_t}[1-\exp\{-\kappa\lambda_t\}]^{1-z_t}$$

the conditional density for the latent components

$$f(\boldsymbol{\lambda}_{1:n}, \boldsymbol{\pi}_{2:n}|\rho) = \exp\{-\lambda_1\}\prod_{t=2}^{n}[\frac{1}{\pi_t}\left(\frac{\lambda_{t-1}\rho}{1-\rho}\right)^{\pi_t}\exp\{-\frac{\lambda_{t-1}\rho}{1-\rho}\}][\frac{\lambda_t^{\pi_t}}{\Gamma(\pi_t+1)(1-\rho)^{\pi_t+1}}\exp\{\frac{-\lambda_t}{1-\rho}\}]$$

and densities for priors given in Sections 2 and 3 as $p(\sigma'), p(\xi'), p(\rho)$, and $p(\kappa)$. For the interval censored model (M_{BLMC}) replace $f(\mathbf{y'}_{1:n}, \mathbf{z}_{1:n}|\sigma', \xi', \kappa, \lambda_{1:n})$ with $f(\mathbf{w'}_{1:n}, \mathbf{z}_{1:n}|\sigma', \xi', \kappa, \lambda_{1:n})$ throughout

$$f(\mathbf{w'}_{1:n}, \mathbf{z}_{1:n}|\sigma', \xi', \kappa, \lambda_{1:n}) = \prod_{t=1}^{n}[\exp\{-\lambda_t w_t^-\} - \exp\{-\lambda_t w_t^+\}]^{z_t} \prod_{t=1}^{n}[\exp\{-\kappa\lambda_t\}]^{z_t}[1-\exp\{-\kappa\lambda_t\}]^{1-z_t}$$

Denoting the sample from the $r-1$ iteration by $\theta^{(r-1)}$, the acceptance probability for the candidate $\sigma'^{(*)}$ is $\min(R,1)$ at the m^{th} iteration for acceptance ratio

$$R = \left\{ \frac{f(\mathbf{y}'_{1:n}, \mathbf{z}_{1:n} | \sigma'^{(*)}, \xi'^{(m-1)}, \rho^{(m-1)}, \kappa^{(m-1)}, \boldsymbol{\lambda}_{1:n}^{(m-1)}) p(\sigma'^{(*)})}{f(\mathbf{y}'_{1:n}, \mathbf{z}_{1:n} | \sigma'^{(m-1)}, \xi'^{(m-1)}, \rho^{(m-1)}, \kappa^{(m-1)}, \boldsymbol{\lambda}_{1:n}^{(m-1)}) p(\sigma'^{(m-1)})} \right\}$$

Similarly, the acceptance ratio for the λ_t update is

$$R = \left\{ \frac{f(\mathbf{y}'_{1:n}, \mathbf{z}_{1:n} | \sigma^{(m-1)}, \xi'^{(m-1)}, \rho^{(m-1)}, \kappa^{(m-1)}, \boldsymbol{\lambda}_{-t}^{(m-1)} \boldsymbol{\lambda}_{t}^{(*)}) f(\boldsymbol{\lambda}_{-t}^{(m-1)} \boldsymbol{\lambda}_{t}^{(*)}, \boldsymbol{\pi}_{2:n}^{(m-1)} | \rho^{(m-1)}))}{f(\mathbf{y}'_{1:n}, \mathbf{z}_{1:n} | \sigma'^{(m-1)}, \xi'^{(m-1)}, \rho^{(m-1)}, \kappa^{(m-1)}, \boldsymbol{\lambda}_{1:n}^{(m-1)}) f(\boldsymbol{\lambda}_{1:n}^{(m-1)}, \boldsymbol{\pi}_{2:n}^{(m-1)} | \rho^{(m-1)})} \right\}$$

All other parameter and latent variable updates are analogous. During the burn-in period, the proposal variances for each candidate distribution is adaptively tuned to give acceptance rates near 0.4. After the burn-in period, the candidate proposal variances are fixed, to satisfy the necessary mixing conditions of a stationary Markov chain. Each chain in the simulation study is run until all parameters achieve an effective sample size (ESS) of 100, after thinning by a factor of 100 and discarding the first 10,000 samples. The Bayesian model parameter estimates reported in Table 2 are based on 2,000,000 MCMC iterations, thinned by a factor of 100, after discarding the first 20,000 burn-in samples. The ESS and effective samples/sec (ES/sec) are reported in the supplementary material. The main bottleneck is performing inference on the dependence parameter, ρ, which is slow to mix since it is informed indirectly through the $\{\lambda_t,\, t \geq 1\}$ process. When a smaller proportion of observations are threshold exceedances, ρ mixes more slowly. Trace and autocorrelation plots are used to monitor convergence. All computing was done using R (https://www.R-project.org/).

FIGURES

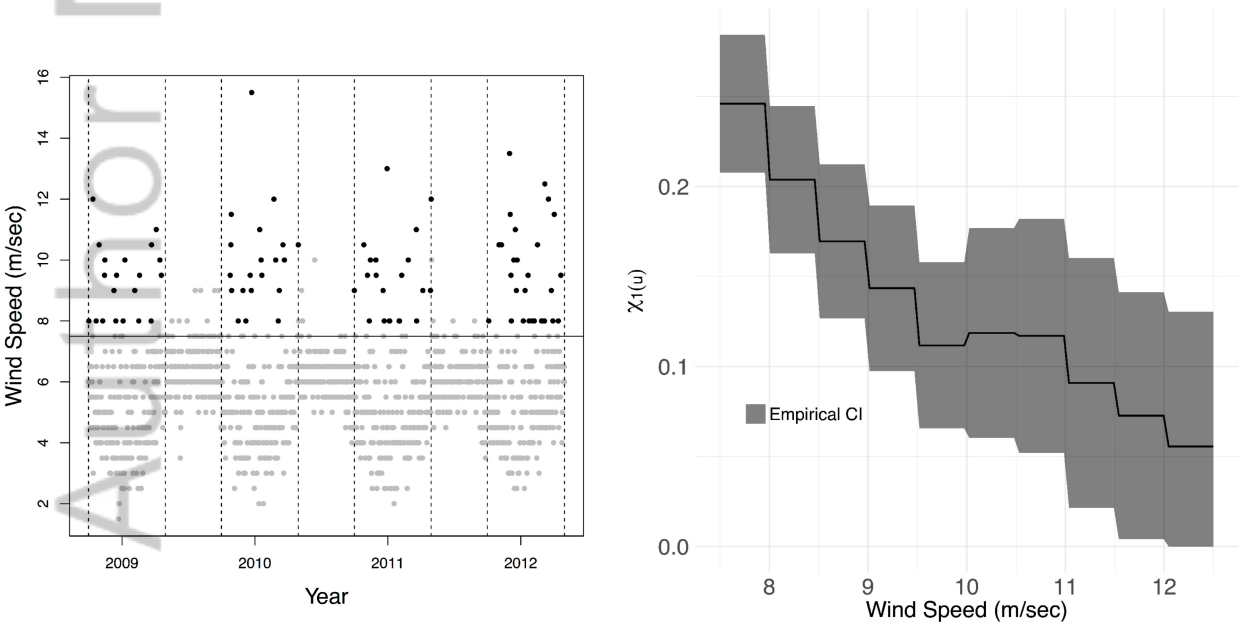

Figure 1. Left: Subset of daily maximum wind gusts from October, 2008 through April, 2012. Exceedances of the threshold $u = 7.5$ m/sec during the Santa Ana season are shown in black. Right: Empirical estimate and pointwise 95% CIs for $\chi_1(u) = P(X_{t+1} > u | X_t > u)$.

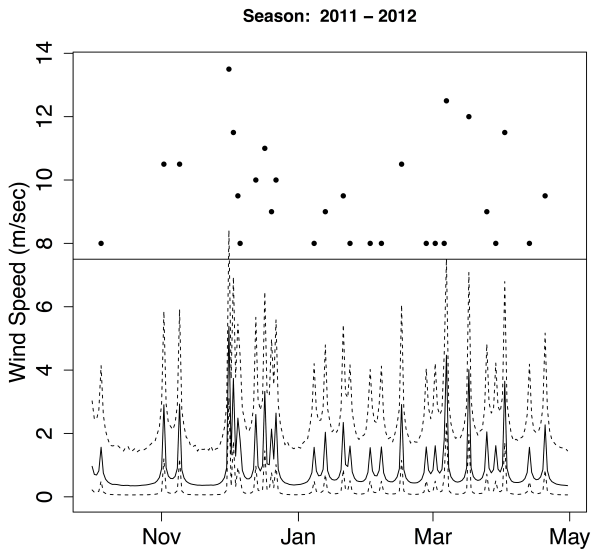

Figure 2. Exceedances (points) of a high threshold (solid horizontal line) and posterior means of $c^{-1}(1/\lambda_t)$ (solid lines) with 95% point-wise credible intervals (dashed lines) for the 2011-2012 season.

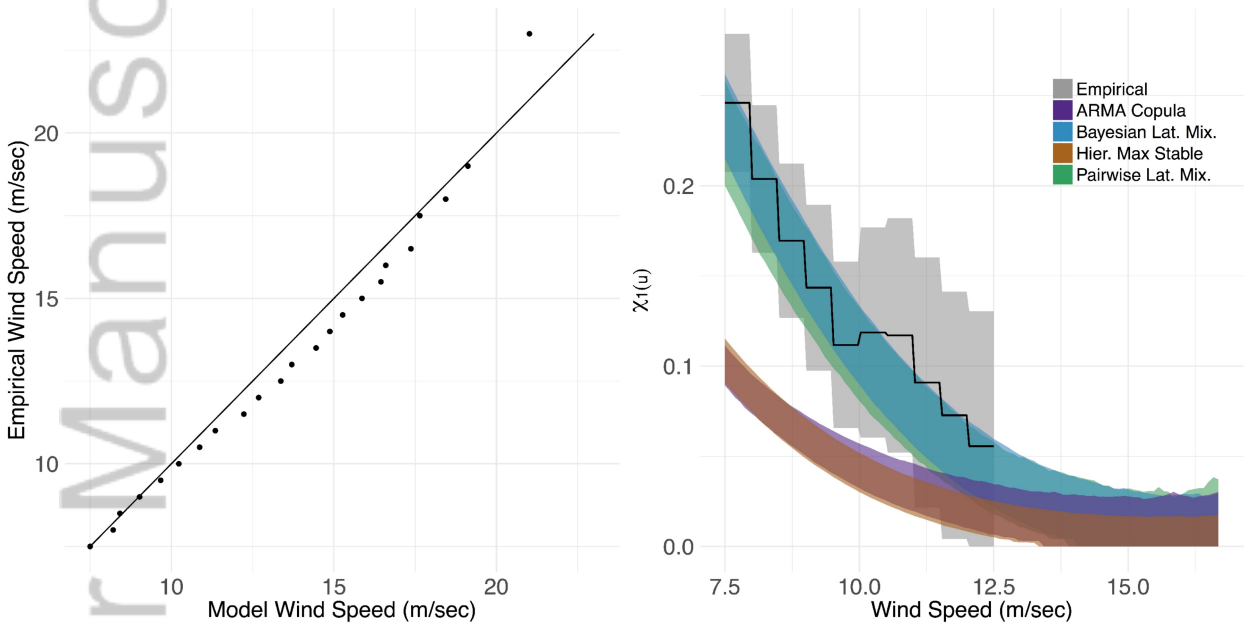

Figure 3. Left: Q-Q plot of empirical and Bayesian latent mixture model quantiles. Right: Estimates and 95% confidence/credible intervals of $\chi_1(u)$: empirical (grey), pairwise likelihood latent mixture model (green), Bayesian latent mixture model (blue), Bayesian Gaussian ARMA copula model (purple), and Bayesian Markov hierarchical max-stable model (brown).

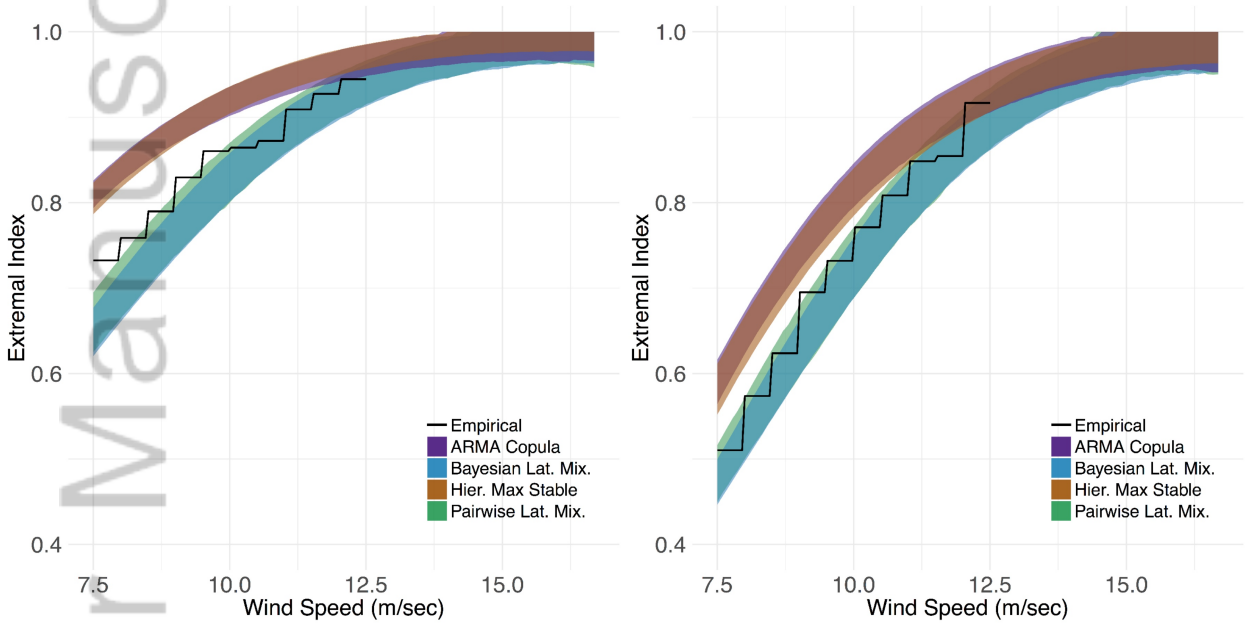

Figure 4. Estimates and 95% confidence/credible intervals of the extremal index: empirical (line), pairwise likelihood latent mixture model (green), Bayesian latent mixture model (blue), Bayesian Gaussian ARMA copula model (purple), and Bayesian Markov hierarchical max-stable model (brown) (Coles *et al.*, 1999). Each are calculated for run lengths 2 (left) and 5 (right).

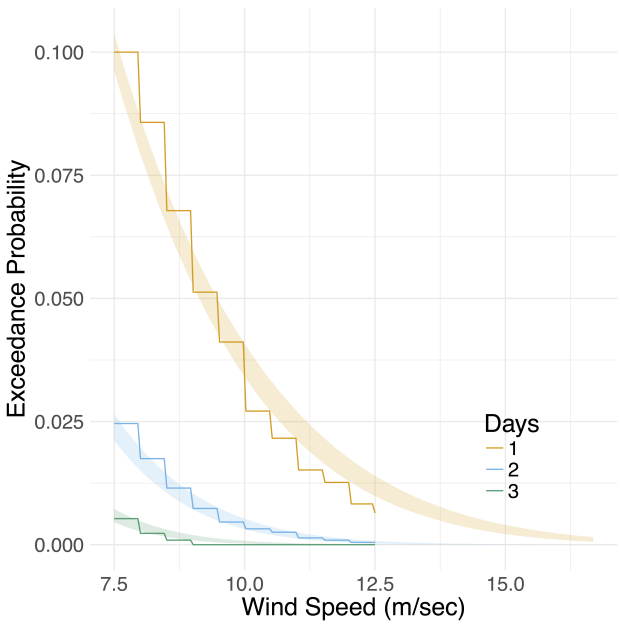

Figure 5. Right: Empirical estimates (lines) and Bayesian latent mixture model 95% credible intervals (bands) of exceedance probabilities for 1 (orange), 2 (blue), and 3 (green) consecutive days.

TABLES

	σ'		ξ'		ρ		κ	
	B	PL	B	PL	B	PL	B	PL
Bias	**0.044**	0.099	**0.015**	−0.035	−0.065	**−0.017**	0.322	**0.173**
RMSE	**0.334**	0.387	**0.089**	0.109	0.134	**0.085**	1.408	**1.300**
Coverage	**0.957**	0.919	**0.968**	0.894	**0.954**	0.980	**0.941**	0.986
Int. Score	**1.523**	2.653	**0.409**	0.898	**0.385**	0.455	**6.381**	6.881

Table 1. Bayesian (B) 95% HPD credible intervals and PL confidence intervals are compared based on interval coverages and interval scores, and point estimates are compared on bias and RMSE for the M_{BLM} and M_{PLM} models. The better value for each criteria is given in bold.

	σ'		ξ'		ρ	
Model	Est.	95% CI	Est.	95% CI	Est.	95% CI
M_{BLM}						
90	2.727	(2.532, 2.932)	-0.144	(-0.171, -0.109)	0.717	(0.671, 0.757)
92.5	2.834	(2.612, 3.067)	-0.160	(-0.188, -0.124)	0.718	(0.669, 0.761)
95	2.306	(2.081, 2.548)	-0.106	(-0.153, -0.048)	0.791	(0.747, 0.828)
M_{PLM}						
90	2.735	(2.450, 3.020)	-0.150	(-0.201, -0.099)	0.700	(0.602, 0.797)
92.5	2.854	(2.531, 3.176)	-0.167	(-0.216, -0.118)	0.725	(0.630, 0.821)
95	2.326	(1.918, 2.734)	-0.117	(-0.213, -0.022)	0.787	(0.697, 0.876)
M_{BLMC}						
90	2.324	(2.024, 2.623)	-0.117	(-0.186, -0.049)	0.797	(0.732, 0.863)
92.5	2.798	(2.578, 3.031)	-0.161	(-0.188, -0.126)	0.723	(0.682, 0.761)
95	2.260	(2.036, 2.496)	-0.106	(-0.153, -0.046)	0.793	(0.748, 0.837)
M_{PLMC}						
90	2.726	(2.440, 3.012)	-0.149	(-0.201, -0.097)	0.701	(0.604, 0.798)
92.5	2.845	(2.522, 3.168)	-0.166	(-0.217, -0.116)	0.726	(0.630, 0.821)
95	2.312	(1.900, 2.724)	-0.116	(-0.214, -0.017)	0.787	(0.697, 0.876)

Table 2. Bayesian posterior means (95% HPD credible intervals), and maximum PL estimates (95% confidence intervals) for continuous and interval censored models using three different thresholds, the empirical percentiles 90%, 92.5%, and 95%.

Model	DIC	LPML
M_{BLM}	**4480.73**	**−2237.05**
M_{GAC}	4493.54	−2280.61
M_{HMS}	4567.28	−2341.83

Table 3. Two measures of model fit are estimated for the latent mixture model, Gaussian ARMA Copula model, and Markov hierarchical max-stable model: LPML (higher indicates better fit) and DIC (lower indicates better fit). Values of criteria indicating the best fit are in bold. Both DIC and LPML indicate the the latent mixture model has the best fit.