Journal of Environmental Statistics
September 2019, Volume 9, Issue 3. http://www.jenvstat.org

Spatial Semi-parametric Spectral Density Estimation for Multivariate Extremes, with Application to Fire Threat

Mauricio Nascimento
Department of Statistics
Penn State University

Benjamin A. Shaby
Department of Statistics
Colorado State University

Abstract

We analyze the joint tail of two variables related to fire threat associated with Santa Ana Winds in Southern California. To do this, we apply a flexible model for the joint tail of asymptotically dependent multivariate distributions, when samples are taken at several locations across space. We use a spatial prior on the underlying multivariate extremal dependence structure, which enables us to borrow strength across space while still allowing for different joint tail distributions at different spatial locations, and to predict the joint tail of the distribution at un-observed locations. A simulation shows that this model is able to capture complex dependence structures well.

Keywords: angular distribution, Bayesian, Dirichlet mixture.

1. Introduction

We present a flexible model for the joint tail of asymptotically dependent multivariate distributions, when samples are taken at several locations across space. We use a spatial prior on the underlying multivariate extremal dependence structure, which enables us to borrow strength across space while still allowing for different joint tail distributions at different spatial locations, and to predict the joint tail of the distribution at un-observed locations. We apply our multivariate spatial model to joint extremes of two variables that, taken together, are informative about extreme wildfire threat.

California suffers from fires that burn more than 172,000 acres of land annually. These fires translate into loss of habitat, infrastructure, and life. In 2015, for example, wildfires caused damage estimated to be more than $3 billion, with the biggest losses caused by fires started from downed electrical power lines (CalFire 2015). As a consequence, it is important to un-

derstand the conditions that allow wildfires to ignite and spread. The California Department of Forestry and Fire Protection (CalFire) is responsible, among other things, for responding to active fires and for forming risk mitigation strategies in the state. In order to efficiently allocate resources and formulate preventative policy proposals, CalFire needs to know where conditions conducive to fire ignition and spread are likely to occur. Of particular interest are the conditions that represent the most extreme threat.

Conditions that are conducive to wildfire ignition and spread are the subject of intense study and have drawn keen interest of regulators and infrastructure planners. This interest has reached new levels in the wake of the worst wildfire season in California history in 2018. Furthermore, the tails of the distribution of fire threat characteristics warrant particular attention because fire size and damage distributions have been shown to be sensitive to underlying tail assumptions (Moritz, Morais, Summerell, Carlson, and Doyle 2005; Bowman, Balch, Artaxo, Bond, Carlson, Cochrane, D'Antonio, DeFries, Doyle, Harrison, Johnston, Keeley, Krawchuk, Kull, Marston, Moritz, Prentice, Roos, Scott, Swetnam, van der Werf, and Pyne 2009). Here, we study the joint tail of two variables that represent the risk due to meteorological conditions of wildfire ignition and spread potential.

One phenomenon that plays a key role in wildfire threat in California is the Santa Ana winds. Santa Ana winds are an atmospheric condition in Southern California that produces hot, dry, high-speed winds that originate in inland desert regions. This phenomenon is particularly active during the period of October through March. Because of these characteristics, and because they occur in a region of high development and population density, these winds are responsible for initiating particularly damaging wildfires (Westerling, Cayan, Brown, Hall, and Riddle 2004). We therefore focus our attention to the region of Southern California that experiences Santa Ana winds.

A common ignition mechanism in the Southern California is electrical arcing resulting from utility poles downed in high winds. We therefore take wind speed to be the key meteorological variable that serves as a proxy for fire ignition potential. The composite Fosberg Fire Weather Index (FFWI) combines relative humidity, temperature, and wind measurements into a single measure of fire weather that determines fire spread potential (Fosberg 1978). Moritz, Moody, Krawchuk, Hughes, and Hall (2010) reconstructed the weather at South California during the Santa Ana wind season, and found that high level of FFWI was associated with higher frequency of wild fires. We therefore study the joint tail of wind speed and FFWI, which represents the region in the space of meteorological variables where the probability of ignition and the potential for spread is concurrently extreme.

There have been previous approaches to modeling multivariate extreme observations. Coles and Tawn (1991) described 6 parametric distributions to model multivariate extremal dependence: the asymmetric and negative asymmetric logistic models, the Dirichlet model, the biologistic model, the nested logistic model, and the time series logistic model. More recent work includes Cooley, Davis, and Naveau (2010), who proposed the pairwise beta distribution, which defines a full joint distribution for multivariate extremes through their pairwise relationships, similar to a covariance matrix for Gaussian data. Vettori, Huser, Segers, and Genton (2017) used Bayesian model averaging with the nested logistic model defined in Coles and Tawn (1991) to probabilistically create clusters of exchangeable dependent variables, with dependence allowed to differ among clusters. They used reversible jump methods to average across the random number of clusters.

Several nonparametric approaches have been proposed as well, mostly limited to the bivariate case. Most focus on estimating the spectral measure, which we define in Section 2. Einmahl, de Haan, and Piterbarg (2001) estimated the bivariate spectral measure nonparmetrically using the ranks of the data, and proved consistency and asymptotic normality. Einmahl and Segers (2009) also estimated the bivariate spectral measure, this time using an empirical likelihood approach. de Carvalho, Oumow, Segers, and Warchoł (2013) proposed a simplified version of the Einmahl and Segers (2009) bivariate estimator using Euclidean likelihood approach, and showed that it has the same asymptotic behavior. Guillotte, Perron, and Segers (2011) proposed a nonparametric Bayesian scheme to estimate the bivariate spectral measure. The infinite-dimensional prior on the spectral measure was shown to be dense in the space of valid spectral measures. Inference was performed using a trans-dimensional metropolis Hastings algorithm.

The problem is considerably more difficult in dimension greater than two. Marcon, Padoan, Naveau, Muliere, and Segers (2017) used Bernstein polynomials to estimate the Pickands dependence function, which is an alternative way of characterizing an extreme-value dependence. This approach in principle applies to any dimension, but is difficult to scale to more than two dimensions. Most closely related to the current work is Boldi and Davison (2007), who proposed a nonparametric mixture of Dirichlet distributions to model the dependence structure between arbitrarily many variables through the spectral distribution, after standardization of the margins to unit Fréchet. A critical moment restriction resulted in very poor mixing of the MCMC, however, and limited the practicality of the approach. Sabourin and Naveau (2014) fit an identical model using a re-parametrization of the Dirichlet mixture that avoids the awkward constraints and associated mixing problems.

An important aspect of our model is that it borrows strength across space to improve estimation of the multivariate tail dependence. However, we make no attempt to model the spatial dependence of the extreme events themselves. In this sense, our model inherits from Cooley, Nychka, and Naveau (2007), which considers a univariate response and places spatial Gaussian process priors on marginal generalized Pareto parameters. In contrast, a great deal of recent effort has been made to model the spatial dependence in the extreme events themselves (for a review, see Davison, Huser, and Thibaud 2013). Models that attempt to simultaneous capture both multivariate dependence and spatial dependence of the extreme events include multivariate max-stable processes (Genton, Padoan, and Sang 2015; Reich and Shaby 2018). These models are more realistic than our model in the sense that when limiting spatial dependence in the response variables is strong, multivariate max-stable processes can account for it. However, the dependence that they allow between variables is much more restrictive than our model.

This paper will describe a model for multiple variables while taking into account the spatial dependence in the multivariate relationships. We describe the model in the general case of d variables, although we later restrict our attention to the special case of $d = 2$ for wind speed and FFWI. Section 2 describes the model originally proposed by Boldi and Davison (2007) and re-parameterized by Sabourin and Naveau (2014), and our extension from a single multivariate sample to a collection of multivariate samples observed at several spatial locations. Section 3 describes our simulation study to assess the performance of our model. Finally, Section 4 contains the joint analysis of wind speed and FFWI in southern California.

2. Model

Our model for the joint tail of wind and FFWI builds upon classical extreme value theory. We will define the model for d-dimensional random vectors, and proceed with the analysis for the special case of $d = 2$, which pertains to our application. Let $\mathbf{Y}$ be a random vector of observations in $\mathbb{R}^d$ with joint distribution function $\mathbf{F}$ and marginal distribution F_i for $1 \leq i \leq d$. Define a new variable of transformed observations $\mathbf{X}$ as

$$\mathbf{X} = (-1/F_1(Y_1), ..., -1/F_d(Y_d))$$

so that each X_i has a unit Fréchet marginal distribution, i.e. $P(X_i \leq x) = e^{-1/x}$. This transformation allows all variables to have a common marginal distribution that is convenient for defining the dependence structure among them in the far joint tail. To model the dependence in the far tail of a random vector whose margins are unit Fréchet, it is useful make a further transformation from the original coordinates to pseudo-polar coordinates. First, define the radial component to be

$$R = \sum_{i=1}^{d} X_i \tag{1}$$

and the angular component to be

$$\mathbf{W} = \frac{\mathbf{X}}{R}, \tag{2}$$

where $\mathbf{W} \in \mathbb{S}_d$ and $\mathbb{S}_d$ is the unit simplex defined by $\left\{\mathbf{w} : w_i \geq 0, \sum_{i=1}^{d} x_i = 1\right\}$. This transformation us useful because, as long as $\mathbf{F}$ is in the multivariate maximum domain of attraction of a max-stable random vector, $\mathbf{F}$ can be expressed in terms of its angular and radial components, which are independent in the limit (Resnick 1987). That is, for a large radial threshold r_0,

$$P(R > r, \mathbf{W} \in A \mid R > r_0) = \frac{r}{r_0} H(A)$$

as $r_0 \to \infty$, for $r > r_0$, $A \in \mathbb{S}_d$, and H a probability distribution usually referred to as the *angular distribution* or the *angular measure*. This decomposition makes it clear that the dependence structure of the random vector is completely described by the angular probability measure H, independent of the radial component. As a consequence, to define the model for dependence in the joint tail, we need only to consider the angular measure H, independently of the distribution of the radial component.

A probability measure H on $\mathbb{S}_d$ is a valid angular probability measure if and only if it satisfies the moment condition

$$\int_{\mathbb{S}_d} w_i dH(\mathbf{w}) = \frac{1}{d} \qquad \text{for all } i = 1, \ldots, d. \tag{3}$$

Several parametric families have been proposed that can be used to model H (Cooley, Davis, and Naveau 2012, e.g.), but in general, no parametric family includes the entire space of

valid angular measures because, as long as the constraint (3) is satisfied, H is otherwise un-restricted.

Tail dependence between a pair of variables is often succinctly summarized using the quantity known as χ, defined as a limit as the quantile q approaches 1 as

$$\chi = \lim_{q \to 1} P\left(Y_i > F_i^{-1}(q) \mid Y_j > F_j^{-1}(q)\right).$$

When the tail dependence parameter χ has a limiting value greater than zero, we say the two variables are asymptotically dependent. The interpretation of asymptotic dependence is that regardless of how far in the tails of the distribution components of the random vector are, the components remain dependent on one another. When the probability mass of the angular distribution H is concentrated in the interior of the simplex, the result is asymptotic dependence among the variables (Resnick 1987). Conversely, when the limiting value of χ is zero, we say the two variables are asymptotically independent, with the interpretation that if one goes far enough in the tails of the distribution, components become independent of one another. If any probability mass of H lies on the vertices or edges of the simplex, the corresponding variables are asymptotically independent.

In the context of wind and FFWI, asymptotic dependence would imply that far in the tail of he distribution, extreme values wind speed and extreme FFWI could occur simultaneously. In contrast, assuming asymptotic independence would mean that in the far in the tail, extreme values of wind speed and extreme values of FFWI would never occur concurrently. For the scope of this paper we will assume asymptotic dependence between wind speed and FFWI. This assumption is natural because wind speed is one of the ingredients of FFWI, so we would expect that extreme values of wind speed would be strongly associated with extreme values of FFWI. We therefore model the dependence in the joint tail by considering angular measures that have mass concentrated in the interior of the simplex.

2.1. Semi-parametric representation of multivariate dependence

Since multivariate dependence in the far tail is entirely described by a distribution H on the unit simplex, we focus our efforts of specifying a flexible representation for H that can be extended to the multivariate spatial context. The best-known distribution with support on the unit simplex is the Dirichlet distribution. A sensible approach then to constructing flexible models for angular distributions H would be to use Dirichlet distributions as building blocks. Boldi and Davison (2007) proposed a semi-parametric model for H using a mixture of Dirichlet distributions. The density of this mixture model is defined as

$$h(\boldsymbol{w}) = \sum_{k=1}^{K} p_k \frac{\Gamma(\nu_m)}{\prod_{j=1}^{J} \Gamma(\nu_m \mu_j^{(m)})} \prod_{j=1}^{J} w_j^{\nu \mu_j - 1}, \tag{4}$$

where p_k is the mixture weight of the kth component, with $p_k \geq 0, \sum_{k=1}^{K} p_k = 1$, $\boldsymbol{\mu}_k \in \mathbb{S}_d$ is the location parameter of of the kth component, and $\nu_k \in \mathbb{R}^+$ is the concentration parameter of the kth component. This model is attractive because it is very flexible compared to other parametric models. The modeler can choose the number of components K to be as large as needed estimate the angular density $h(\boldsymbol{w})$ well, including shapes of $h(\boldsymbol{w})$ that

are not symmetric and have otherwise unusual shapes. In fact, as $K \to \infty$, this family of distributions is dense in the space of distributions on the interior of the simplex, which in a sense makes this model as flexible as one would ever want. The difficulty with this model arises from the moment restriction (3). This restriction is enforced by requiring that

$$\sum_{i=1}^{k} p_i \mu_i = (1/d, ..., 1/d). \tag{5}$$

This constraint on the parameters forces the model to satisfy (3), but because it is not a hyper-rectangle in the parameter space, it makes estimation difficult. Specifically, when estimating the parameters using Markov Chain Monte Carlo (MCMC) methods, the constraint (5) makes the components of the chain highly dependent, leading to catastrophically poor mixing (Boldi and Davison 2007).

Re-parametrization of the mixture model

To mitigate the difficulties caused by the constraint (5), Sabourin and Naveau (2014) re-parametrized the mixture model (4) in a way that replaces the dependent constraint (5) with independent box constraints on a transformed set of parameters. The transformation is made from the original parameters to the new parameters such that $T(p_1, \ldots, p_k, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K) \Longrightarrow (\epsilon_1, \ldots, \epsilon_{K-1}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{K-1})$, where $\epsilon_m \in (0, 1), m = 1, \ldots, K-1$ (See Appendix A for details). This transformation replaces the original mixture weights $p_1, \ldots, p_K$ and the Kth location parameter $\boldsymbol{\mu}_K$ and replaces them with a collection of $K-1$ "eccentricity" parameters $\epsilon_1, \ldots, \epsilon_{K-1}$. The last location parameter $\boldsymbol{\mu}_K$ may be recovered as a deterministic function of the first $K-1$ locations and the eccentricities, with the reduced degree of freedom taking the place of the awkward constraint (5). The key is that under the Sabourin and Naveau (2014) parameterization, the moment constraint (3) is satisfied even though all parameters ϵ_i and $\boldsymbol{\mu}_i, 1 \leq i \leq K-1$, are independent of each other in the prior. Hence, the parameter space becomes a hyper-rectangle, and the complicated cross dependencies that hindered MCMC mixing in the original parameterization are completely removed.

2.2. Spatial prior

Our main goal is to use spatial information to both improve estimation of the flexible multivariate tail dependence model and be able to predict the dependence structure at locations where no observations are sampled. Since the tail dependence is completely described by the mixture defined by equation (4), allowing the (transformed) parameters in (4) to vary smoothly in space is equivalent to allowing the dependence structure to vary smoothly in space. We therefore allow each parameter in the mixture to follow a smooth process so that locations close together will have similar values, while locations farther apart will have values that are independent from each other.

To induce this spatial smoothing in the underlying dependence parameters, we assign them Gaussian process priors according to

$$\begin{aligned} \log(\nu_k(s)) &\sim \text{GP}(\boldsymbol{X}(s)\boldsymbol{\beta}_{\nu_k}, C_{\nu_k}) \quad \text{for } k \in 1, ..., K \\ \text{logit}(\epsilon_k(s)) &\sim \text{GP}(\boldsymbol{X}(s)\boldsymbol{\beta}_{\epsilon_k}, C_{\epsilon_k}) \quad \text{for } k \in 1, ..., K-1 \\ \boldsymbol{\mu}_k &\sim \text{Dirichlet}(\boldsymbol{\mu}_0, \nu_0) \quad \text{for } k \in 1, ..., K-1, \end{aligned} \tag{6}$$

where the notation GP$(m(s), C)$ refers a Gaussian process with mean function $m(s)$ and covariance function C, and $\boldsymbol{X}(s)$ is a collection of spatially-varying covariates. The prior specification in (6) also includes link functions necessary to transform the support of the parameters to the real line. For example, each concentration parameter $\nu(s_0)$ at a location s_0 is a value in $\mathbb{R}^+$, so taking its logarithm allows us to assign it a Gaussian process prior. Similarly, we transform each eccentricity parameter $\epsilon(s_0)$ at location s_0 from $(0, 1)$ to the real line using the logit function. Finally, the first $K - 1$ Dirichlet location parameters $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{K-1}$ are shared across spatial locations to avoid over-parameterization (see Lock and Dunson 2015), and have independent Dirichlet priors. Because of the lost degree of freedom from the Sabourin and Naveau (2014) parameterization, the remaining Dirichlet location parameter $\boldsymbol{\mu}_K(s)$ is a function of the other location parameters and the spatially-varying eccentricities, so it is itself spatially varying, even though it is not assigned a prior distribution.

For simplicity, all covariance functions are assumed to be stationary and isotropic, with $C_{\nu_k}(s_i, s_j) = \sigma^2_{\nu_k} \exp\{-\|s_i - s_j\|/\rho_{\nu_k}\}$ independently for $k \in 1, ..., K$ for the log concentration parameters and $C_{\epsilon_k}(s_i, s_j) = \sigma^2_{\epsilon_k} \exp\{-\|s_i - s_j\|/\rho_{\epsilon_k}\}$ independently for $k \in 1, ..., K - 1$ for the logit eccentricity parameters. To complete the model, the scale and range parameters $\sigma_{\nu_1}, \ldots, \sigma_{\nu_K}, \rho_{\nu_1}, \ldots, \rho_{\nu_K}, \sigma_{\epsilon_1}, \ldots, \sigma_{\epsilon_{K-1}}$, and $\rho_{\epsilon_1}, \ldots, \rho_{\epsilon_{K-1}}$ are all assigned vague zero-centered positive half normal priors, and the regression coefficients $\boldsymbol{\beta}_{\nu_1}, \ldots, \boldsymbol{\beta}_{\nu_K}$ and $\boldsymbol{\beta}_{\epsilon_1}, \ldots, \boldsymbol{\beta}_{\epsilon_{K-1}}$ are assigned vague zero-centered normal priors.

2.3. MCMC

We fit the model using Markov Chain Monte Carlo (MCMC), for a fixed number of mixture components K, and then use model diagnostics to choose the value of K (see Section 3). All parameter updates are performed using Metropolis-Hastings (MH), except the regression coefficients, which have closed-form full conditional distributions. All MH proposals are random walk normal proposals, except for $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{K-1}$, which live on the unit simplex and get Dirichlet proposals.

3. Simulation

We performed a simulation study to assess the performance of the model. We created a grid of 20 spatial locations on the unit square and sampled 100 replications from our model at those locations. We constructed $h(\boldsymbol{w})$ according to (4) using three mixture components. The spatial range parameters $\rho_{\nu_1}, \rho_{\nu_2}, \rho_{\nu_3}, \rho_{\epsilon_1}$ and ρ_{ϵ_2} were all set to 0.2, which results in moderate spatial correlation for the Dirichlet parameters across the spatial domain. The spatially varying covariate matrix $\boldsymbol{X}(s)$ was a column vector of ones for simplicity, specifying an intercept-only model. The regression coefficients $\boldsymbol{\beta}$ corresponding to the concentration parameters were set to 3, and those corresponding to the eccentricity parameters were set to -1.791 for the first component and 0.619 for the second component. The first 2 Dirichlet location parameters $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ were set to $(0.4, 0.6)$ and $(0.2, 0.8)$, respectively.

We fit the model four times, with the number of components ranging from 2 to 5. For comparison with a parametric model for the angular distribution, we also fitted a logistic model (Coles 2001) to the data, using a spatial Gaussian process prior for the logistic dependence parameter α. Since mixture models are susceptible to label switching, convergence diagnostics based on MCMC parameter trace plots are ineffective. Instead, we monitored conver-

gence by plotting the posterior log likelihood of the data. To choose the number of components, we used 5-fold cross validation, using the log likelihood as the criterion. To do this, we randomly divided the 20 locations into 5 groups of 4 and ran the model 5 times, each time using 16 locations for model fitting and the remaining 4 locations for evaluation. The average log likelihood across the 5 test sets, for the varying number of model components as well as the logistic model, are shown in table 1.

Table 1: Cross validation results for the dependence model with 2 to 5 components, as well as the parametric logistic model. The cross validation procedure chose 2 mixture components, although the data was simulated using 3 components.

# Components	CV log likelihood
Logistic	-105.18
2	52.53
3	-92.49
4	-181.77
5	-176.29

Even though the data was simulated using 3 mixture components, the cross validation procedure found that the model with 2 components to be the best fit. Figure 1 shows the true simulated angular density $h(\boldsymbol{w})$ in black at four arbitrarily-selected locations, along with the predicted posterior angular density using the best-fitting spatial Dirichlet mixture with 2 components (blue curve) and corresponding pointwise 95% credible region (blue region). The fitted spatial logistic model and corresponding pointwise 95% credible region is shown in red. It is evident from Figure 1 that the simple parametric model is not capable of accurately representing the somewhat complex tail dependence in the simulated data, as evidenced by the credible region failing to include the true curve over much of the domain. In contrast, the mixture model with 2 components does a qualitatively good job capturing the important features, even though the true density was simulated with three components. From Figure 1 then, it is not surprising that cross validation selected 2 components rather than 3.

4. Analysis of concurrent extreme wind speed and FFWI

The data we used for this analysis consists of daily maximum wind speeds, along with other variables needed to construct the Fosberg Fire Weather Index, from 20 weather stations in a region of Southern California that is susceptible to the Santa Ana Winds. The data was downloaded from the Hadley Center website (available at `http://www.metoffice.gov.uk/hadobs/hadisd/index.html`). We extracted the data from October through March, when the Santa Ana Winds are most active, from 1973–2015, and calculated FFWI using the formula in Fosberg (1978). Figure 2 shows the locations of the weather stations we considered, which lie inside a 22,500 km^2 area in Southern California which includes large population centers like Los Angeles and San Diego. This location is of particular interest because because of its vulnerability to large wildfires and its proximity to so much at-risk population and infrastructure.

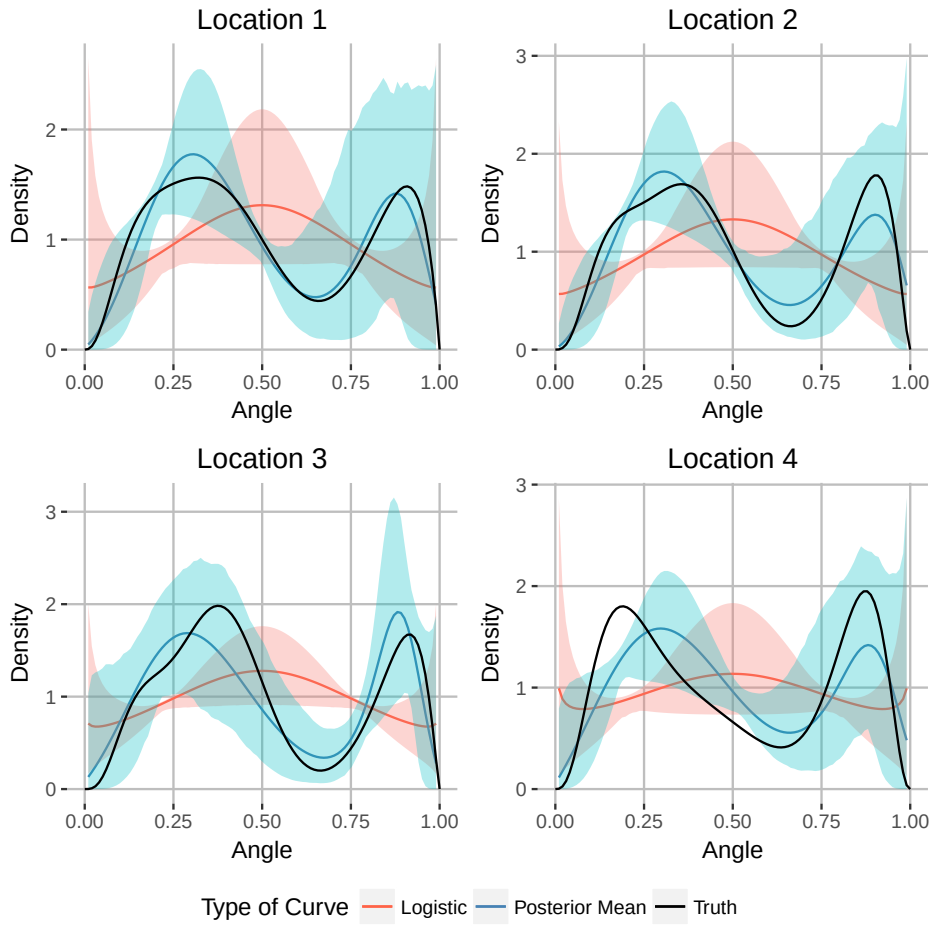


Figure 1: Predicted and true simulated angular densities for 4 arbitrarily-chosen spatial locations. The solid black curve is the true angular density $h(\boldsymbol{w})$. The solid red curve is the posterior mean angular density estimated using the logistic model, and the red band is the corresponding pointwise 95% credible region. The blue curve is the posterior mean angular density estimated using the mixture of Dirichlets from (4), and the blue band is the corresponding pointwise 95% credible region.

4.1. Marginal model

Our tail dependence model assumes unit Fréchet marginal distributions for exceedances of high thresholds, for all components. We therefore have to choose thresholds and transform wind speed and FFWI to unit Fréchet. The standard technique for transforming to unit Fréchet is to use a nonparametric rank transformation. We did not do this for two reasons. First, we wanted to use spatial information to borrow strength in estimating the marginal distributions. Second, applying a rank transformation to data at observation locations would have left us no way to predict at un-observed locations. For these reasons, we used a spatial model to estimate the marginal distributions. To do this, we fit spatial generalized extreme value (GEV) distribution models separately to annual maxima of the wind speed and FFWI variables. This model uses Gaussian process priors for GEV location and (log) scale parameters, and assumes conditionally independent responses, similar to the model in Cooley *et al.*

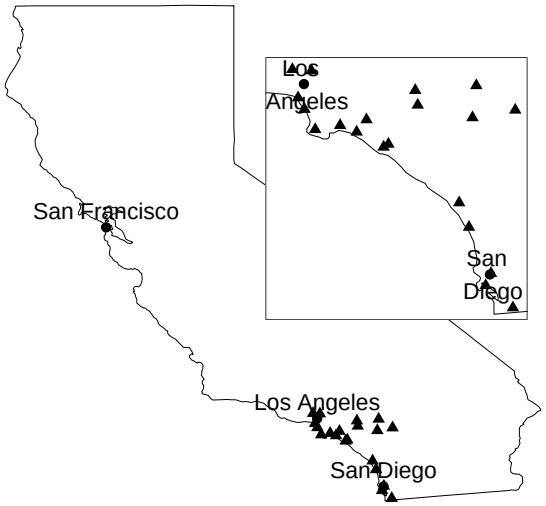


Figure 2: Location of weather stations in California.

(2007). We included latitude, longitude and elevation as covariates in the marginal models for both wind speed and FFWI. We fit this model using the implementation in the R package **SpatialExtremes** (Ribatet 2018). Figure 3 shows the 50 and 100-year return level surface calculated from the fitted models for the wind speed (Figure 3b and 3d) and FFWI (Figure 3a and 3c) marginal components.

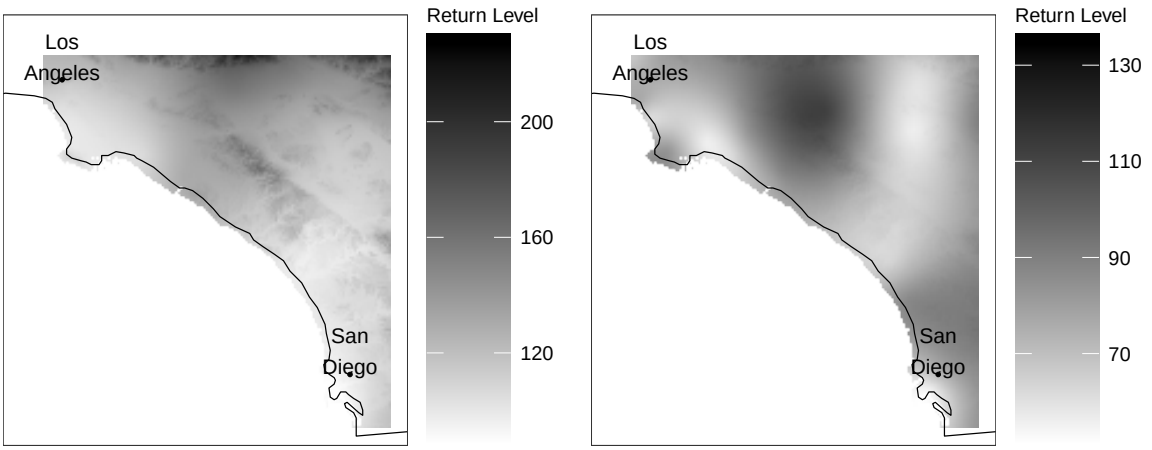
Return values of FFWI are smaller near the coast compared to values in the northeastern portion of the study region. In contrast, areas along the coast tend to have the strongest extreme wind, with the northeastern portion tending to have less extreme wind speeds. Over almost the entire region, the 100-year return level is greater than 100, which is sometimes considered the largest feasible value of FFWI.

We selected days when both FFWI and wind speed exceeded their marginal 90% empirical quantiles (35.15 and 18.34mph for FFWI and wind speed, respectively), where the quantiles were computed by pooling data across all observed locations. Out of the original 127,802 and 143,162 daily observations from FFWI and wind speed, respectively, 10,484 and 12,027 exceeded their respective marginal thresholds, and 6,228 did so simultaneously. Setting the threshold at 90% empirical quantiles was a compromise between being far enough in the tail for the asymptotic model to be a good approximation on one hand, and on the other hand having enough data to fit the dependence model.

Based on the posterior sample of the marginal GEV parameter fields obtained from the spatial GEV model, we marginally transformed the exceedances to unit Fréchet using the posterior mean GEV parameters at the observation locations.

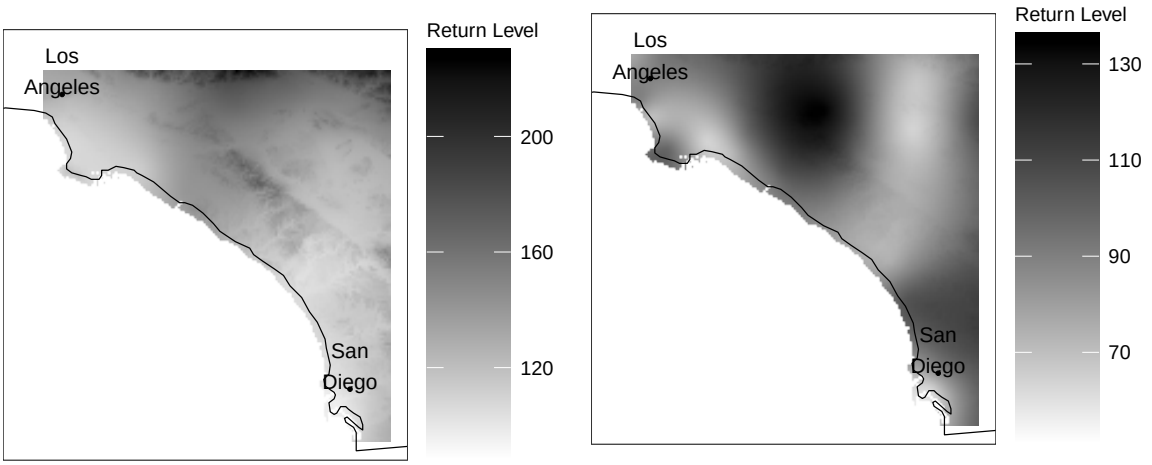
4.2. Dependence model

After transforming the exceedances to unit unit Fréchet, we further transformed them into pseudo-polar coordinates using equations (1) and (2). We then fit our spatial multivariate extremal dependence model to the derived angles.



(a) FFWI, 50 year return level.

(b) Wind speed (miles/hour), 50 year return level.



(c) FFWI, 100 year return level.

(d) Wind speed (miles/hour), 100 year return level.

Figure 3: Posterior mean 50 and 100 year return levels for FFWI and wind speed components.

In order to choose the number of components to include, we used cross validation, randomly partitioning the data into 5 folds by location, just as in the simulation in Section 3. This scheme of cross validation selects for the ability to predict the joint tail at un-observed locations, preventing over-fitting at any individual location, which helps to regularize the estimation of the flexible tail model. Table 2 shows the cross validation results for the FFWI and wind speed data. The cross validation suggests that the 4-component model provides the best fit. Therefore, we proceed with this analysis using 4 mixture components. For com-

parison, we also evaluated a spatial logistic model the same cross validation scheme. As Table 2 shows, the logistic model is not competitive with any of the mixture models, and is unable to fit the dependence well.

Table 2: Cross validation results for the extremal dependence between FFWI and wind speed, for a spatial logistic model and spatial mixture models with 2 to 5 mixture components. The righ-hand column shows predictive log likelihoods when all replications at 80% of the locations are used for fitting and the remaining 20% of the locations are used for testing. Larger log likelihoods are preferred.

# Components	CV log likelihood
Logistic	-6195.596
2	-6075.282
3	-5537.285
4	-4794.236
5	-5075.750

Proceeding with a 4-component mixture, we re-fit the model using the complete data. Figure 4 shows the fitted angular densities for 4 locations in the study region. Solid lines represent pointwise posterior means, and shaded regions represent pointwise 95% credible regions. Observed angles are shown as ticks on the x-axes. These four locations show considerable heterogeneity in their joint tail characteristics between FFWI and wind speed. At all observed locations, the tick marks coincide with high posterior density, which suggests that the model is doing a good job capturing the dependence, and that imposing the restriction that the Dirichlet location parameters must be shared across space was not too restrictive. In each angular density plot, the angular mass is concentrated far in the interior of the simplex, suggesting that the two variables are indeed asymptotically dependent in this geographical region.

In addition to the spatial prior enabling borrowing strength across locations, it also enables spatial prediction of the joint tail at un-observed locations. Figure 5a shows the predicted angular density at one such un-observed location, which was burned by the disastrous Lilac Fire in 2017. At this particular location, the model predicts strong dependence between extreme FFWI and wind speed, with the highest posterior density at around 0.3. This result indicates that joint extremes of FFWI and wind speed are likely to occur simultaneously, exacerbating fire threat due to each variable being individually extreme.

To make this result more interpretable, we back-transformed the two variables from pseudo-polar coordinates to the unit Fréchet scale and then to the scale of the original data, and calculated joint exceedance probabilities implied by the angular density shown in Figure 5a. The tail of the joint survivor function of FFWI and wind speed predicted at this location is shown in Figure 5b. Each contour represents a set of constant joint exceedance probability. That is, at every (y_1, y_2) point along the contour of constant probability p, the joint probability of FFWI exceeding y_1 and wind speed exceeding y_2 is p. The joint survivor function might be of direct use for risk management. For example, if safety standards require consideration of events that have probability 0.00005 (i.e. events which occur on average once every 100 years, since our study period is about 200 days in each year), these events in the joint space of FFWI and wind speed are described by the contour in Figure 5b labeled $5\text{e}{-05}$.

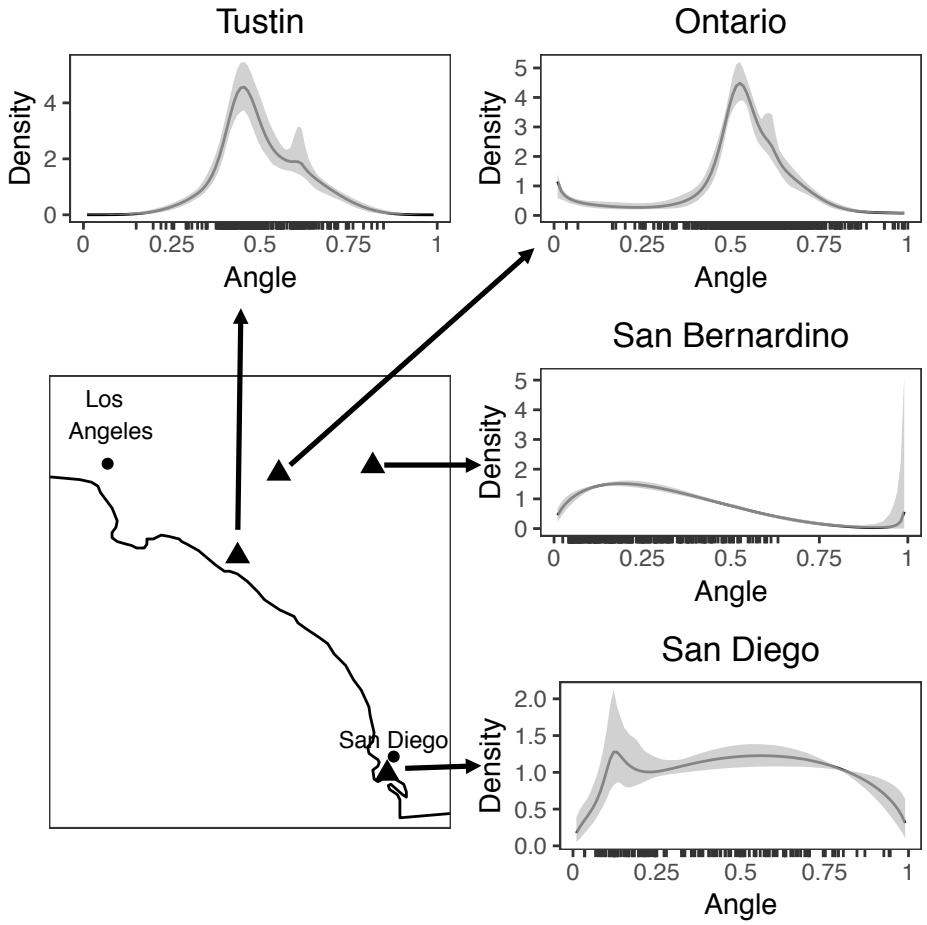
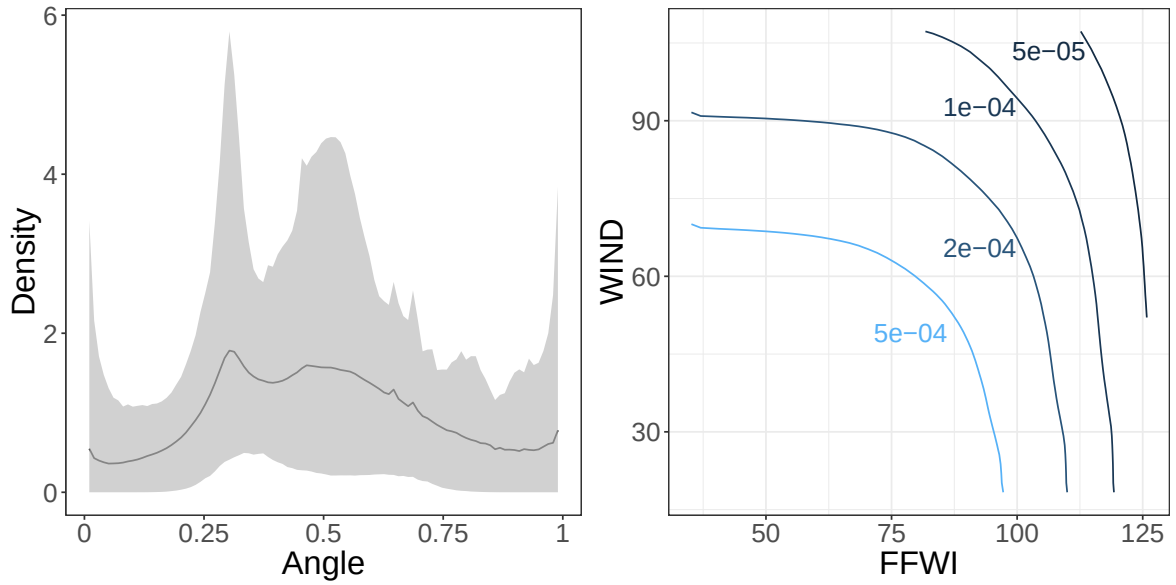


Figure 4: Estimated posterior densities for 4 locations across Southern California. The black curves are the pointwise posterior mean densities, the gray bands are the 95% pointwise credible bands, and the ticks on the x-axes are the observed angles.

5. Conclusion

We analyzed the joint tail of two variables related to fire threat associated with Santa Ana Winds in Southern California. We wanted a flexible model that could accommodate the complicated shape of the angular distribution that determines extremal dependence between the two variables. We used spatial priors to regularize estimation of the flexible angular distribution model, which also enabled us to predict the joint tail at un-observed locations. Do do this, we used a mixture of Dirichlet distributions, combined with the new parametrization proposed by Sabourin and Naveau (2014). The re-parametrization enabled the mixture parameters to be independent in the prior for each location, which enabled specification of Gaussian process priors to borrow strength between locations. Our simulation analysis showed that the model is capable of recovering complicated shapes of tail dependence, and 5-fold fold cross validation did a reasonable job of choosing a good number of mixture components. Although we restricted our attention to the bivariate case due to the structure of the fire threat application, the spatial mixture of Dirichlet distributions is straightforwardly generalizable to higher dimensions.



(a) (b)

Figure 5: Panel (a) shows the predicted angular density at an un-observed location which was near the center of the Lilac Fire in 2017. The solid line represents the pointwise posterior mean, and the shaded region represents the pointwise 95% credible region. Panel (b) shows the estimated joint survivor function of FFWI and wind speed, calculated by transforming pseudo-polar coordinates with the angular density in panel (a) first to unit Fréchet vectors and then to the original scale of the data. Each contour represents a set of constant joint exceedance probability. For example, at every (y_1, y_2) point along the contour labeled $5\text{e}{-04}$, the joint probability of FFWI exceeding y_1 and wind speed exceeding y_2 is $5\text{e}{-04}$.

The joint analysis of extreme Fosberg Fire Weather Index and wind speed was able to capture the tail dependence between the two variables, as well as the spatial correlation in the angular densities across locations. One possible drawback to our modeling approach is that it is unable to capture spatial tail dependence in the extreme events themselves (as opposed to spatial dependence in the underlying joint tail distribution), so that if one wished to evaluate areal joint exceedance probabilities, that would be impossible. For example, if practitioners needed to know the probability that an area of at least size A would simultaneously experience a joint exceedance of X FFWI and Y mph wind, they would not be able to do that with our model.

A second drawback of our approach is that we estimated marginal surfaces and transformed marginally to unit Fréchet as a pre-processing step. Done this way, the uncertainty in the marginal transformation is not propagated to the final analysis. It might be possible to do both simultaneously, as Sabourin (2015) did in the non-spatial context. We made a sustained attempt to adapt the Sabourin (2015) model to the spatial context so that we could properly account for uncertainty in the marginal estimation, but we were unable to make it work.

One limitation of our result is that FFWI only takes into account wind speed, temperature, and humidity, and it assumes that the fuels are extremely fine with high moisture of extinction, a condition most suited to grasslands. In particular, FFWI does not take into account

precipitation. This means that the index assumes a constant fuel moisture and equilibrium moisture content. This is a known limitation for operational applications. A modified version of FFWI was proposed by Goodrick (2002) to include information about precipitation and fuel availability, and could be used instead of the original FFWI in a similar analysis to the one presented here.

Acknowledgements

The authors wish to thank Davis Sapsis of CalFIRE for suggesting a joint tail analysis of FFWI and wind speed. The authors gratefully acknowledge the support of DOE grant DE-AC02-05CH11231 and NSF grant DMS-1752280. Computations for this research were performed on the Pennsylvania State University's Institute for CyberScience Advanced CyberInfrastructure (ICS-ACI). This content is solely the responsibility of the authors and does not necessarily represent the views of the Institute for CyberScience.

A. Parameter transformation

The main goal of the transformation proposed by Sabourin and Naveau (2014) is to remove the constraint (5) which induces strong prior dependence in the parameters and causes disastrously poor MCMC mixing. After transformation, new parameter space becomes a rectangular subset of $\mathbb{S}_d^{k-1} \times (0,1)^{K-1} \times (\mathbb{R}^+)^K$, removing the prior dependence and dramatically improving MCMC mixing. Intuitively, we replace the weight vector $\mathbf{p}$ and the last mean vector $\boldsymbol{\mu}_K$ by new parameters $\epsilon_1, \ldots, \epsilon_{K-1}$. Where each $\epsilon_m \in (0,1)$ is referred to as an *eccentricity*, defined to indicate departure from centrality induced by decreasing the subsets of mixture components. The idea is that we decrease the number of parameters by one, with the decreased degrees of freedom playing the role of the constraint in the original parameterization. The transformation is defined by a series of recursive equations.

First, given the original parameters $p_1, \ldots, p_K$ and $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$, each eccentricity parameter ϵ_m is defined as

$$\epsilon_m = \frac{\|\boldsymbol{\gamma}_m - \boldsymbol{\gamma}_{m-1}\|}{\|\boldsymbol{I}_m - \boldsymbol{\gamma}_{m-1}\|}. \tag{7}$$

This transformation uses intermediate variables $\boldsymbol{I}_m$, $\boldsymbol{\gamma}_m$, and $\boldsymbol{\gamma}_{m-1}$. Each $\boldsymbol{\gamma}_m$ can be interpreted as the the center of mass of the mth mixture component. Each $\boldsymbol{\gamma}_m$ is then defined as

$$\boldsymbol{\gamma}_m = \rho_m^{-1} \sum_{j=m+1}^{K} p_j \boldsymbol{\mu}_j, \tag{8}$$

which depends on a collection of variables $\rho_0, \ldots, \rho_{K-1}$. The first $\rho_0 = 1$ and the remaining $\rho_1, \ldots, \rho_{K-1}$ are defined recursively as

$$\rho_m = \rho_{m-1} - p_m. \tag{9}$$

Finally, each $\mathbf{I}_m$ in (7) is set according to

$$\mathbf{I}_m = \boldsymbol{\gamma}_{m-1} + T_m(\boldsymbol{\gamma}_{m-1} - \boldsymbol{\mu}_m), \tag{10}$$

where $T_m = \sup\{t \geq 0 : \boldsymbol{\gamma}_{m-1} + t(\boldsymbol{\gamma}_{m-1} - \boldsymbol{\mu}_m) \in \mathbb{S}_d\}$
The result is that the transformed parameters live in rectangular region and still result in mixtures that satisfy the moment constraint (3). See Sabourin and Naveau (2014) for additional details.

Bibliography

Boldi MO, Davison AC (2007). "A mixture model for multivariate extremes." *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **69**(2), 217–229. ISSN 1369-7412. URL https://doi.org/10.1111/j.1467-9868.2007.00585.x.

Bowman DMJS, Balch JK, Artaxo P, Bond WJ, Carlson JM, Cochrane MA, D'Antonio CM, DeFries RS, Doyle JC, Harrison SP, Johnston FH, Keeley JE, Krawchuk MA, Kull CA, Marston JB, Moritz MA, Prentice IC, Roos CI, Scott AC, Swetnam TW, van der Werf GR, Pyne SJ (2009). "Fire in the Earth system." *Science*, **324**(5926), 481–484.

CalFire (2015). "2015 Wildfire Activity Statistics." *Technical report*, CA: California Department of Forestry and Fire Prevention and Office of the State Fire Marshal.

Coles S (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer-Verlag London, Ltd., London. ISBN 1-85233-459-2. URL https://doi.org/10.1007/978-1-4471-3675-0.

Coles SG, Tawn JA (1991). "Modelling extreme multivariate events." *J. Roy. Statist. Soc. Ser. B*, **53**(2), 377–392. ISSN 0035-9246. URL http://links.jstor.org/sici?sici=0035-9246(1991)53:2<377:MEME>2.0.CO;2-4&origin=MSN.

Cooley D, Davis RA, Naveau P (2010). "The pairwise beta distribution: a flexible parametric multivariate model for extremes." *J. Multivariate Anal.*, **101**(9), 2103–2117. ISSN 0047-259X. URL https://doi.org/10.1016/j.jmva.2010.04.007.

Cooley D, Davis RA, Naveau P (2012). "Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data." *Ann. Appl. Stat.*, **6**(4), 1406–1429. ISSN 1932-6157. URL https://doi.org/10.1214/12-AOAS554.

Cooley D, Nychka D, Naveau P (2007). "Bayesian spatial modeling of extreme precipitation return levels." *J. Amer. Statist. Assoc.*, **102**(479), 824–840. ISSN 0162-1459. doi:10.1198/016214506000000780. URL https://doi.org/10.1198/016214506000000780.

Davison AC, Huser R, Thibaud E (2013). "Geostatistics of dependent and asymptotically independent extremes." *Math. Geosci.*, **45**(5), 511–529. ISSN 1874-8961. doi:10.1007/s11004-013-9469-y. URL https://doi.org/10.1007/s11004-013-9469-y.

de Carvalho M, Oumow B, Segers J, WarchołM (2013). "A Euclidean likelihood estimator for bivariate tail dependence." *Comm. Statist. Theory Methods*, **42**(7), 1176–1192. ISSN 0361-0926. doi:10.1080/03610926.2012.709905. URL https://doi.org/10.1080/03610926.2012.709905.

Einmahl JHJ, de Haan L, Piterbarg VI (2001). "Nonparametric estimation of the spectral measure of an extreme value distribution." *Ann. Statist.*, **29**(5), 1401–1423. ISSN 0090-5364. doi:10.1214/aos/1013203459. URL https://doi.org/10.1214/aos/1013203459.

Einmahl JHJ, Segers J (2009). "Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution." *Ann. Statist.*, **37**(5B), 2953–2989. ISSN 0090-5364. doi:10.1214/08-AOS677. URL https://doi.org/10.1214/08-AOS677.

Fosberg MA (1978). "Weather in wildland fire management: the fire weather index." In *Proceedings of the Conference on Sierra Nevada Meteorology*, pp. 1–4. American Meteorological Society, Lake Tahoe, California, USA.

Genton MG, Padoan SA, Sang H (2015). "Multivariate max-stable spatial processes." *Biometrika*, **102**(1), 215–230. ISSN 0006-3444. doi:10.1093/biomet/asu066. URL https://doi.org/10.1093/biomet/asu066.

Goodrick SL (2002). "Modification of the Fosberg fire weather index to include drought." *International Journal of Wildland Fire*, **11**(4), 205–211.

Guillotte S, Perron F, Segers J (2011). "Non-parametric Bayesian inference on bivariate extremes." *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **73**(3), 377–406. ISSN 1369-7412. doi:10.1111/j.1467-9868.2010.00770.x. URL https://doi.org/10.1111/j.1467-9868.2010.00770.x.

Lock EF, Dunson DB (2015). "Shared kernel Bayesian screening." *Biometrika*, **102**(4), 829–842. ISSN 0006-3444. doi:10.1093/biomet/asv032. URL https://doi.org/10.1093/biomet/asv032.

Marcon G, Padoan SA, Naveau P, Muliere P, Segers J (2017). "Multivariate nonparametric estimation of the Pickands dependence function using Bernstein polynomials." *J. Statist. Plann. Inference*, **183**, 1–17. ISSN 0378-3758. doi:10.1016/j.jspi.2016.10.004. URL https://doi.org/10.1016/j.jspi.2016.10.004.

Moritz MA, Moody TJ, Krawchuk MA, Hughes M, Hall A (2010). "Spatial variation in extreme winds predicts large wildfire locations in chaparral ecosystems." *Geophysical Research Letters*, **37**(4), n/a.

Moritz MA, Morais ME, Summerell LA, Carlson J, Doyle J (2005). "Wildfires, complexity, and highly optimized tolerance." *Proc. Nat. Acad. Science*, **102**(50), 17912–17917. URL https://doi.org/10.1073/pnas.0508985102.

Reich BJ, Shaby BA (2018). "Modeling of multivariate spatial extremes." *RESEARCHERS.ONE*. URL https://www.researchers.one/article/2018-09-12.

Resnick SI (1987). *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York. ISBN 0-387-96481-9. doi:10.1007/978-0-387-75953-1. URL https://doi.org/10.1007/978-0-387-75953-1.

Ribatet M (2018). *SpatialExtremes: Modelling Spatial Extremes*. R package version 2.0-7, URL https://CRAN.R-project.org/package=SpatialExtremes.

Sabourin A (2015). "Semi-parametric modeling of excesses above high multivariate thresholds with censored data." *J. Multivariate Anal.*, **136**, 126–146. ISSN 0047-259X. URL https://doi.org/10.1016/j.jmva.2015.01.014.

Sabourin A, Naveau P (2014). "Bayesian Dirichlet mixture model for multivariate extremes: a re-parametrization." *Comput. Statist. Data Anal.*, **71**, 542–567. ISSN 0167-9473. URL https://doi.org/10.1016/j.csda.2013.04.021.

Vettori S, Huser R, Segers J, Genton MG (2017). "Bayesian model averaging over tree-based dependence structures for multivariate extremes." *arXiv preprint arXiv:1705.10488*.

Westerling AL, Cayan DR, Brown TJ, Hall BL, Riddle LG (2004). "Climate, Santa Ana Winds and autumn wildfires in southern California." *Eos, Transactions American Geophysical Union*, **85**(31), 289. URL https://doi.org/10.1029/2004EO310001.

Affiliation:

Mauricio Nascimento
Department of Statistics, Penn State University
University Park, PA 16802
E-mail: mfn120@psu.edu

http://www.jenvstat.org
Submitted: 2018-10-31
Accepted: 2019-09-15