Daniel Cooley, Brett D. Hunter, Richard L. Smith

Univariate and Multivariate Extremes for the Environmental Sciences

0.1 Extremes and Environmental Studies

The underlying philosophy for the statistical analysis of extreme values is to "let the tail speak for itself." In practice, this is often done by analyzing only data that are considered to be extreme. Additionally, the goal of an extreme value (EV) analysis is often extrapolation; one may have data with a record length of 50 years, but need to estimate the magnitude of an event which has an annual exceedance probability (AEP) of .01; that is the so-called '100-year event'. Because extreme events are rare, EV methods generally result in large uncertainties being associated with estimated quantities.

If the philosophy is to let the tail speak, guidance for practice is to use methods specifically designed for extremes. Fortunately, there are well-developed probability results that provide a solid foundation for extremes-specific statistical methods. Most importantly, these results are quite general and imply that one does not need to know the distribution from which the data arise to be able to characterize the distribution's (joint) tail.

Extremes are of great interest in environmental sciences, as extreme environmental events often have tremendous societal and economic consequences. Much of extreme value theory can trace its roots to applications in hydrology, where researchers sought to use stream and river flow records to estimate the magnitudes of '100-year floods' and other rare events. Extremes studies are now relatively common in atmospheric science, where they have been used to study extreme precipitation, air pollution, heat waves, and other events with societal implications. Our presentation here will present some current techniques for analyzing extreme events. We provide enough background for the unfamiliar reader to follow our presentation, but we do not aim to give a complete overview of theory and methods as there exist very good sources for such material such as Coles (2001), Beirlant et al. (2004), and de Haan and Ferreira (2006). In Section 0.2, we will introduce the ideas that underlie EV analysis of univariate data, concluding with an application accounting for possible nonstationarity. In Section 0.3, we will discuss multivariate extremes, including how dependence in the tail can be described and modeled.

0.2 Univariate Extremes

0.2.1 Theoretical underpinnings

The foundation of EV theory are results which characterize the limiting distribution of sample maxima. Let Y_t be an iid sequence of random variables, and let $M_n = \bigvee_{t=1}^{n} Y_t$, where $\bigvee$ denotes maximum. If there exist sequences $\{a_n\} > 0$ and $\{b_n\}$ and non-degenerate distribution G such that

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} G \tag{1}$$

as $n \to \infty$, then G is said to be an extreme value distribution (EVD). The three-types theorem (Fisher and Tippett, 1928; Gnedenko, 1943) states EVDs must be a location-scale family of the Frechét with cdf $G(y) = \exp(-y^{-1/\xi})$ for $y \geq 0, \xi > 0$, Gumbel with cdf $G(y) = \exp\{-\exp(-y)\}$ for $y \in \mathbb{R}$, or (reverse) Weibull with cdf $G(y) = \exp\{(-y)^{-1/\xi}\}$ for $y < 0, \xi < 0$. These three types can be characterized by the

single representation

$$G(y) = \exp\left\{-(1+\xi y)^{-1/\xi}\right\}, \tag{2}$$

for $1+\xi y > 0$ and where the right-hand side is interpreted as the limit in the case when $\xi = 0$. The extreme value distributions are equivalent to the max-stable distributions, i.e., the distributions for which there exist sequences such that $G^n(a_n y + b_n) = G(y)$ for all positive integers n and all $y \in \mathbb{R}$.

If Y_t has distribution F and (1) holds, then F is said to be in the max-domain of attraction of G. The distributions in the Frechét domain of attraction are the regularly varying distributions, roughly speaking distributions whose tails behave like power functions such as the t, Pareto, and Cauchy distributions. Distributions in the Gumbel domain of attraction have exponentially-decaying tails and include the normal, gamma, and lognormal distributions. The Weibull domain of attraction includes distributions with bounded tails like the beta distribution. In practice, estimates of the parameter ξ indicate to which domain of attraction the underlying distribution belongs. The $\xi > 0$, $\xi = 0$, and $\xi < 0$ cases are respectively termed the heavy-, light-, and bounded-tailed cases. In the environmental sciences, precipitation and stream flows are often found to have heavy-tailed distributions, whereas temperatures tend to have bounded-tailed distributions.

In the environmental sciences, rarely is an iid assumption justified. When Y_t is not iid but is stationary, the class of distributions which satisfy (1) are still the EVDs, so long as a certain mixing condition is met (Leadbetter et al., 1983). While the mixing condition is difficult to explain, it is not overly restrictive, and environmental processes are often assumed to meet this condition. However, dependence in the sequence $\{Y_t\}$ often implies that the effective block size is less than n, and thus the rate of convergence of (1) to an EVD is likely slower than if Y_t were iid.

0.2.2 Modeling Block Maxima

The convergence of renormalized block maxima to a known class of distributions suggests a strategy for modeling block maxima which does not require knowledge or estimation of the underlying distribution of Y_t. From (1) and (2) if (fixed) n is large, then

$$P\left(\frac{M_n - b_n}{a_n} \le y\right) \approx \exp\left\{-(1+\xi y)^{-1/\xi}\right\}.$$

With n fixed, a_n and b_n can be moved to the right-hand side of the approximate equality and treated as parameters (respectively scale parameter τ and location parameter η), thereby obtaining the generalized EV distribution (GEV)

$$P(M_n \le y) \approx \exp\left[-\left\{1+\xi\left(\frac{y-\eta}{\tau}\right)\right\}^{-1/\xi}\right], \tag{3}$$

for y such that $1+\xi\left(\frac{y-\eta}{\tau}\right) > 0$. In addition to being theoretically justified, the three-parameter GEV is quite flexible, and the parameter ξ provides information about the nature of the tail of Y_t. More familiar distributions such as the normal or lognormal lack both theoretical justification and the GEV's flexibility for modeling tail behavior.

For a selected block size n, assume there are B blocks of stationary observations $y_{b,t}, b = 1, \ldots, B; t = 1, \ldots, n$. Let $m_b = \bigvee_{t=1}^{n} y_{b,t}$; that is, m_b is the maximum of the observations in the bth block (we have suppressed the dependence of m_b on the block size n). A sensible method for estimating the distribution of M_n is to use $m_1, \ldots, m_B$ to estimate the three parameters of the GEV distribution. Later, we will use numerical maximum-likelihood estimation, but other moment-based methods (Hosking, 1990) are also widely used.

The fact that the modeling procedure extracts block maxima and estimates the distribution of these maxima

agrees with EV philosophy. However, the GEV parameter estimates themselves are rarely of primary interest (although $\hat{\xi}$ does convey useful information about the nature of the tail). Often the quantity of interest is a high quantile such as a particular AEP level, which can be calculated from the GEV parameters. The AEP(p) level is the quantile which the annual maximum exceeds with probability p. A more familiar term is the r-year return level; that is, the level for which the expected time between exceedances of this level is r years, assuming stationarity. Thus, under stationarity an AEP(.01) level corresponds to a 100-year return level. We use AEP level as it more clearly conveys the probabilistic nature of an exceedance and remains interpretable under non-stationarity due to changing conditions such as climate change or urban development. The asymptotic arguments leading to the GEV provide arguments for its use when estimating these high quantiles requires extrapolation.

Modeling block maxima is also appealing in environmental science because often environmental data exhibit seasonality. In the presence of seasonality, one can no longer assume the data are identically distributed. In practice, when the annual maximum occurs in a particular season (e.g., temperature in summer), one can assume that the annual maximum arises from a season during which the process is approximately stationary, but one should recognize that the effective block length is not $n = 365$. The usual argument against modeling block maxima is that by retaining only the maximum of each block, other data which are useful for characterizing extreme behavior is not utilized. Confidence intervals of GEV parameters are often found to be quite wide, resulting in large uncertainties for quantities such as the AEP(p) level. Ferreira and de Haan (2015) examine efficiency of the block maximum method in the iid case.

0.2.3 Threshold exceedances

An alternative to modeling block maxima which still follows the EV philosophy is to model exceedances over a high threshold. As with block maxima, asymptotic results suggest a model for exceedances over a suitably high threshold. Balkema and de Haan (1974) and Pickands (1975) showed that if Y_t is in the domain of attraction of an EVD, then the distribution of exceedances over an increasing threshold converge to a generalized Pareto distribution (GPD). A somewhat heuristic justification of modeling the distribution of a random variable conditioned on exceeding a high threshold follows. Assuming $Y_1, Y_2, \ldots$ are iid and starting from (3) which assumes n is fixed and large, then for any t

$$\begin{aligned} P^n(Y_t \leq y) &\approx \exp\left[-\left\{1+\xi\left(\frac{y-\eta}{\tau}\right)\right\}^{-1/\xi}\right] \\ \Rightarrow n\log[1-P(Y_t>y)] &\approx -\left\{1+\xi\left(\frac{y-\eta}{\tau}\right)\right\}^{-1/\xi} \\ \Rightarrow nP(Y_t>y) &\approx \left\{1+\xi\left(\frac{y-\eta}{\tau}\right)\right\}^{-1/\xi}, \end{aligned}$$

where the last line assumes y is large enough such that $\log[1-P(Y_t>y)] \approx -P(Y_t>y)$. For u sufficiently large and $y > u$,

$$\begin{aligned} P(Y_t > y \mid Y_t > u) &= \frac{nP(Y>y)}{nP(Y>u)} \\ &\approx \frac{\left\{1+\xi\left(\frac{y-\eta}{\tau}\right)\right\}^{-1/\xi}}{\left\{1+\xi\left(\frac{u-\eta}{\tau}\right)\right\}^{-1/\xi}} \\ &= \left\{1+\xi\left(\frac{y-u}{\psi_u}\right)\right\}^{-1/\xi}, \end{aligned}$$

where $\psi_u := \tau + \xi(u-\eta)$ and the final expression is the GPD survival function. The parameter ξ continues to indicate the type of tail and the tail weight, and ψ_u is a scale parameter which depends on the threshold u.

Threshold	$u = 0$	$u = 0.5$	$u = 2$	$u = 3$	$u = 5$	Truth
$\hat{\xi}$	-.002(.010)	.062(.016)	.192(.048)	.207(.084)	.349 (.224)	.25
$\hat{q}_{.9999}$	8.41(.30)	9.46(.52)	12.42(1.59)	12.75(2.05)	13.60 (3.32)	13.03
$P(T > u)$	.500	.322	.058	.020	.004	–
# Exc	4987	3221	571	193	38	–

TABLE 0.1: Results from fitting a GPD to threshold exceedances of 10000 simulated random variables with a t-distribution with 4 df. $\hat{q}_{.9999}$ is the estimated .9999 quantile and standard errors are in parentheses. When low thresholds such as 0 or 0.5 are selected, biased estimates of ξ and the .9999 quantile result. When high thresholds such as 5 are selected, large uncertainties are associated with parameter estimates.

To estimate high quantiles of Y_t's distribution, an estimate of $\zeta_u = P(Y_t > u)$ is also required. Importantly, the only assumption on the distribution of Y_t is that it is in the domain of attraction of a EVD.

The above result suggests a strategy for modeling threshold exceedances by first selecting a threshold u above which the distribution's tail is well approximated by a GPD, and then using the observations which exceed u to estimate the parameters ψ_u and ξ. A challenge with this approach is finding a suitable threshold. Threshold selection involves a bias versus variance tradeoff often found in statistics. If the chosen threshold is too low the GPD approximation will be poor, resulting in bad estimates for ψ_u and ξ, and consequently bad estimates for the high quantiles which are of interest. If the threshold is too high, there will be few exceedances, resulting in large uncertainties associated with the parameters and high quantiles. These qualities are illustrated in Table 0.1, which gives results from fitting a GPD to the exceedances above various thresholds to 10000 realizations of a t-distribution with 4 df. Table 0.1 shows that for this distribution, thresholds of 0 or 0.5 are too low resulting in estimates for ξ and $q_{.9999}$ which are biased low, and the threshold of 5 is too high resulting in large uncertainty associated with the parameter estimates. Thresholds are often selected by resorting to diagnostic plots such as the mean residual life (alternatively mean excess) plot, or plots of $\hat{\xi}$ (Coles, 2001, Sections 4.3.1 and 4.3.4). However, these plots often do not yield clear-cut values for adequate thresholds (Resnick, 2007, Section 4.4.2). After the model is fit, QQ plots or return level plots (Coles, 2001, Section 4.3.5) can be used to assess model adequacy. Due to the difficulty in selecting a threshold above which a GPD is suitable, it is tempting to instead estimate such a threshold. Several modeling approaches have attempted to estimate a threshold (Scarrott and MacDonald (2012) provide a review), and most propose some sort of model for the data in the 'bulk' of the distribution (i.e., those which lie below the threshold). Recently, Naveau et al. (2016) proposed an extended-GPD model and used it to model the entire distribution of non-zero rainfall amounts.

As with the GEV, estimation of the GPD parameters can be done via numerical maximum likelihood or by moment-based approaches. A likelihood specified as the product of densities evaluated at each data point assumes independence of the data used for estimation, and moment-based estimation methods similarly assume independence. For many environmental processes, threshold exceedances occur in clusters, contradicting the independence assumption. If estimating the AEP or some other quantity of interest which can be calculated from the marginal distribution, this dependence does not need to be explicitly modeled, but inference should account for it. Peaks-over-threshold approaches (c.f., Chavez-Demoulin and Davison, 2012, and references therein) identify clusters, and use only the maximum observation of each cluster for inference, thereby alleviating data dependence. Alternatively, Fawcett and Walshaw (2007) advocate using a simple likelihood incorporating all exceedances and appropriately adjusting uncertainty estimates arising from the misspecified likelihood (Davison, 2003, Section 4.6). Another way to obtain appropriate parameter estimate uncertainty is to use a block-bootstrap approach to account for temporal dependence. However, if one's interest involves quantities not given by the marginal, such as the sum of multiple days' measurements, then the dependence must be modeled. An approach taken by Smith et al. (1997) was to assume the data arise from a Markov chain, thereby allowing the dependence to be specified by a bivariate relationship between Y_t and Y_{t+1}. Importantly, Smith et al. (1997) employ a bivariate dependence structure appropriate for modeling tail dependence, as we will discuss in Section 0.3.

Modeling block maxima with the GEV or modeling threshold exceedances with the GPD are the most prevalent EV modeling methods, but they are not the only models suggested by EV theory. Sometimes interest is in neither the block maximum nor exceedances of a some threshold level. For example, the current ozone air quality standard set by the US Environmental Protection Agency is based on the fourth-highest observation in a year.[1] The theory which leads to the GEV being the limiting distribution for the block maximum can be extended to give the limiting distribution of the kth-largest order statistic (Coles, 2001, Section 3.5), and this approach was used by Berrocal et al. (2014) to model the fourth-largest annual ozone measurement in the Eastern US. Extreme value theory can be connected to point-processes (Coles, 2001, Chapter 7), and the GEV can be seen the Poisson-derived probability of observing no points above the level of the distribution's argument. Via the point-process formulation, an alternative parametrization can be given for threshold exceedances (Coles, 2001, p. 132). This parametrization can be described by the parameters η, τ, and ξ of the GEV distribution, and has the advantage that its parameters are independent of the threshold u, unlike the GPD parametrization's scale and exceedance rate parameters ψ_u and ζ_u.

0.2.4 Regression models for extremes

A major theme throughout statistics is constructing regression-like models to relate a response Y to covariates $\boldsymbol{x}$. The underlying argument is that the value of $\boldsymbol{x}$ induces a distributional change on Y. In extremes studies, and environmental extremes studies in particular, there have been many examples of building regression extremes models where the parameters of the GEV or GPD are modeled as functions of covariate values, or alternatively functions of time in order to model trends. Importantly, the standard arguments that lead to modeling block maxima with a GEV or threshold exceedances with a GPD require that the block maximum or threshold exceedances are taken from a sequence of identically distributed data. Here we briefly review arguments leading to regression extremes models.

Let $y_{b,t}, t = 1, \ldots, n; b = 1, \ldots, B$ be observations, where n denotes the block size and b is used to index the block. Assume these observations arise from corresponding random variables $Y_{b,t}$ and further that given covariate value $\boldsymbol{x}_b$, $Y_{b,t} \mid \boldsymbol{x}_b$ are identically distributed for $t = 1, \ldots, n$. Then, from results in Sections 0.2.1 and 0.2.2, if n is large enough, the distribution of $M_{b,n} = \bigvee_{t=1}^{n} Y_{b,t} \mid \boldsymbol{x}_b$ should be well approximated by a GEV. Given that $\boldsymbol{x}_b$ influences the distribution of $Y_{b,t}$, a natural modeling approach is to assume that the GEV parameters are functions of the covariate; that is

$$[M_{b,n} \mid \boldsymbol{x}_b] \dot{\sim} GEV(\eta = f_\eta(\boldsymbol{x}_b), \tau = f_\tau(\boldsymbol{x}_b), \xi = f_\xi(\boldsymbol{x}_b)).$$

One can also formulate regression approaches for threshold exceedances by allowing the GPD parameters ψ_u and ξ to vary with the covariate. Since the probability of exceeding a threshold also likely depends on the covariate, ζ_u and/or the threshold u itself could be modeled in terms of the covariate. As both ψ_u and ζ_u are themselves functions of the threshold, a perhaps cleaner formulation for threshold exceedances is via the point process representation and GEV parameters as was done in Smith (1989). Often the functions relating EV distribution parameters to covariates are simple linear or perhaps polynomial functions, although Chavez-Demoulin and Davison (2005) employed spline-based functions.

The above argument suggests that any regression extremes model should only employ a covariate that varies more slowly than the response variable, so that the block maximum or threshold exceedances can be assumed to arise from a larger set of observations that can be assumed identically distributed given the covariate. Thus, building a regression model where extreme summer temperatures are regressed on an annual value for the El Niño Southern Oscillation index seems reasonable. Likewise, modeling trends due to climate change by regressing extreme values on year seems reasonable, given that nonstationarity induced by climate change is negligible within a year. There is a desire to relate extreme behavior to variables which vary on the same timescale as the response variable. One approach is to preprocess the data before

[1] https://www.epa.gov/ozone-pollution/2015-national-ambient-air-quality-standards-naaqs-ozone#rule-summary

extracting extreme observations as is done by Eastoe and Tawn (2009). Russell et al. (2016) uses the bivariate extremes framework presented in Section 0.3.2 to relate daily extreme ozone pollution measurements to daily meteorological data.

0.2.5 Application: Fitting a time-varying GEV model to climate model output

GCMs (formally 'general circulation models', but colloquially 'global climate models') are tools for understanding Earth's climate via simulation. As it is impossible to perform experiments on the Earth itself, GCMs are the best tools available for learning how different forcings, such as increased greenhouse gas concentrations, affect climate. GCMs are deterministic models which employ discretized solutions of the differential equations governing atmospheric circulation and other Earth processes. Coupled models, like the one whose output we will study, combine models for the atmosphere, ocean, land-surface, and sea ice to allow these different Earth processes to interact and influence one another. Running the most state-of-the-art GCMs requires massive computing capability and is done by a number of large institutions around the globe. GCMs can be run with historical forcings in order to compare the GCM-produced climate to historical observations, and also can be run with hypothesized future forcings to produce projections for future climate.

GCMs produce weather-like output for hundreds of meteorological variables on a discretized grid at regular time intervals. GCM output can be treated like data and analyzed in order to understand how well the GCM mimics Earth's processes or to understand how changes in forcings could alter Earth's climate. Although GCM output is indexed by date, GCMs should not be thought of weather prediction models. Rather, it is convenient to think of the produced output as a possible draw from the distribution of weather given the climate of the indexed time. That is, the output from a perfect GCM would not correspond to observed weather, but the distributions of output and observations would match. Interpretation of GCM output and comparison to weather-station observations can be tricky, as the spatial resolution of GCM output corresponds to the grid cell resolution and not the point-referenced information recorded at weather stations. Differences between GCM output and weather station observations seem to be exacerbated when considering extreme behavior; for instance Figure 1 of Fix et al. (2016) shows a clear difference in the distribution of annual maximum precipitation for weather station observations and GCM output. Despite differences between GCM output and observations, the GCM output can provide useful information about how climate is likely to change in response to changes in forcings, and downscaling methods can be used to relate GCM output to observations.

The output we study is from the Community Earth System Model 1 (CESM1) produced by the National Center for Atmospheric Research. CESM1 is a Coupled Model Intercomparison Project 5 (Taylor et al., 2012) model and has a spatial resolution of approximately 1°. The output we study is somewhat unique in that it is from an initial condition ensemble experiment (Kay et al., 2015). Because GCMs are deterministic, multiple simulations can be produced by perturbing the conditions used to initialize the model run. Initial condition ensembles are used to study 'internal climate variability'. From a statistics perspective, internal climate variability is analogous to sample variability: any individual simulation produces an incomplete picture of the climate and multiple simulations (ensemble members) provide an idea of the extent of the differences between members. We have 30 ensemble members, each of which can be viewed as an independent realization of the output produced by this GCM under the given forcings. We will restrict our attention to output produced using historical forcings for the 86-year period 1920-2005, although the ensemble experiment was extended for the period 2005-2080 for two climate projections: RCP4.5 (a projection where emissions peak mid-century) and RCP8.5 (emissions continue to grow through the century) (Moss et al., 2010; Stephenson, 2018). We aim to illustrate a univariate regression model by investigating whether extreme precipitation shows a significant time trend at a particular location. Further, we use the ensemble to investigate internal variability and discuss borrowing strength across locations.

0.2.5.1 Analysis of individual ensembles and all data

We study annual maximum daily precipitation produced by the GCM. The independent replicates of the ensemble provide us with an artificially large data set where for each year we have 30 annual maximum observations. For the time being, we will model each ensemble member's time series of annual maximum precipitation individually to get an idea of how the GCM's interval variability affects statistical models of extreme precipitation. We initially restrict attention to a single grid cell, centered at 105 W longitude and 41 N latitude, which is the grid cell whose center is closest to Fort Collins, Colorado.

As we analyze block maxima, our model is a time-varying GEV. Let $M_b^{(j)}(\boldsymbol{s}) = \bigvee_{t=1}^{n} Y_{b,t}^{(j)}(\boldsymbol{s})$, where $Y_{b,t}^{(j)}(\boldsymbol{s})$ denotes the precipitation for day t in year $b = 1920, \ldots, 2005$ at location $\boldsymbol{s}$ for ensemble member j. We assume

$$M_b^{(j)}(\boldsymbol{s}) \sim GEV(\eta_b(\boldsymbol{s}) = \beta_0(\boldsymbol{s}) + \beta_1(\boldsymbol{s})(b - 1919), \tau(\boldsymbol{s}), \xi(\boldsymbol{s})). \tag{4}$$

We choose to capture nonstationary behavior in only the location parameter, as increasing the model complexity by adding trends to the scale or shape parameters did not seem warranted.

Parameters were estimated by numerical maximum likelihood and standard errors were obtained from the numerically computed Hessian of the likelihood surface at the ML estimates. The left panel of Figure 1 shows the 30 estimates of β_1 along with normal-based 95% confidence intervals, and the right panel shows the same for ξ. The main message from these figures is that there are very large uncertainties associated with these parameter estimates. Most ensembles do not find $\hat{\beta}_1$ to be significantly different from zero. Most of the point estimates for $\hat{\xi}$ are positive, which is generally what is found for weather-station precipitation data, so this gives some confidence that the GCM can produce heavy-tailed behavior. However, more than half of the CIs include zero and individual ensemble point estimates for $\hat{\xi}$ range from -.11 to .33 Distributions with these values of ξ have dramatically different tail behaviors with -.11 corresponding to a bounded tail and .33 a quite heavy tail ($\xi \geq 1/3$ implies an infinite third moment).

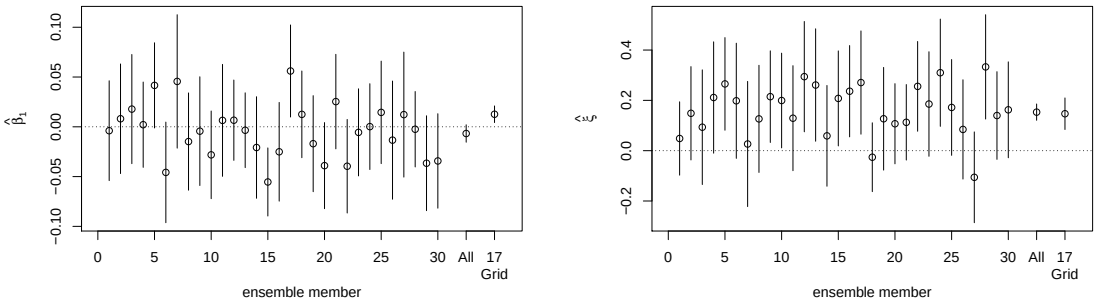

FIGURE 1: Estimates for β_1 and ξ for the Fort Collins grid cell from individual ensemble members. "All" on horizontal axes indicates when all ensemble members were used. "17 Grid" indicates only data from ensemble member 17 was used and the method to borrow strength across location in Section 0.2.5.2 was implemented.

Given that ensemble member from a GCM are effectively iid realizations, we have the luxury of utilizing data from all ensemble members to estimate the parameters of model (4). ML estimates (standard errors) are $\hat{\beta}_0 - 1919 = 24.3(0.24), \hat{\beta}_1 = -.0067(.0045), \widehat{\log}(\tau) = 1.82(.018)$, and $\hat{\xi} = 0.15(.017)$. Not surprisingly, as shown in Figure 1, the uncertainty is dramatically reduced when block maxima from the entire ensemble are used. The 95% CI for β_1 continues to include zero.

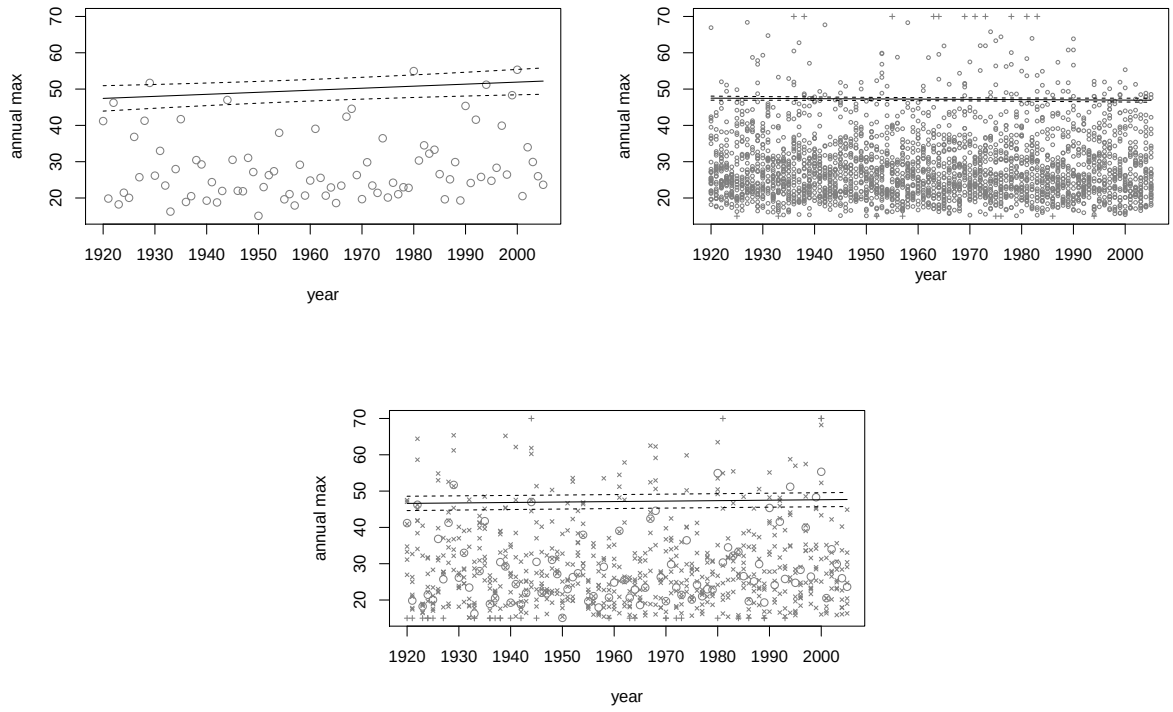

FIGURE 2: Top left: Circles denote annual maximum daily precipitation (mm) for Fort Collins grid cell from ensemble member 17. Top right: Same but for all ensembles. Bottom center shows ensemble 17's annual maximum precipitation for Fort Collins grid cell with circles and annual maxima from eight surrounding grid cells with X's. Plotted in each figure is the point estimate for the .95 quantile (equivalently AEP(.05) level) along with 95% confidence intervals. Point estimates for the top left, top right, and bottom center respectively use only ensemble 17's data for the Fort Collins grid cell, all ensembles' data for the Fort Collins grid cell, and the borrowing strength estimate described in Section 0.2.5.2. In the right and bottom plots, values greater than 70 mm or less than 15 mm fall outside the plotting window and have been denoted by crosses.

0.2.5.2 Borrowing strength across locations

When analyzing observational data, and even in most cases when analyzing climate model output, the practitioner does not have the luxury of utilizing multiple ensembles to reduce uncertainty in the parameter estimates. In this section we will focus on analyzing only ensemble member 17, which we select specifically because it has the highest point estimate $\hat{\beta}_1$ and whose 95% CI for this parameter does not include zero. Of course, if data were generated from model (4) with no trend, it would not be surprising that one of 30 ensembles would lead one to erroneously conclude that there was in fact a positive trend. Still, if one only had the output from ensemble 17, one would likely reach different conclusions about the behavior of extreme precipitation for this grid cell. The top left panel of Figure 2 shows ensemble 17's annual maximum data and the estimate .95 quantile of the annual maximum (i.e., .05 AEP level) along with a 95% confidence interval generated by the delta method. The top right panel of Figure 2 shows the annual maximum data from all ensembles and the estimated .05 AEP level using the data from all ensembles.

Given only output from ensemble 17, we can still consider ways in which parameter uncertainty can be reduced and estimation can be improved. One way to likely decrease parameter uncertainty would be to

utilize more data by employing a threshold exceedance approach rather than a block maximum approach. However, threshold exceedance approaches include the additional challenges of determining a threshold and dealing with short-term temporal dependence.

It is somewhat silly to think of this grid cell in isolation, as there is additional information contained in nearby grid cells which could be employed to aid estimation. Extremes studies in environmental sciences have a long history of borrowing strength across locations, dating at least to Dalrymple (1960). A widely-used method with a long history is regional frequency analysis, which defines regions over which data are pooled to estimate parameters (Hosking and Wallis, 1997). Another approach is to construct hierarchical models which employ statistical spatial models on EV parameters (e.g., Cooley et al., 2007; Sang and Gelfand, 2010; Dyrrdal et al., 2015). Bayesian inference is often employed for these hierarchical models.

Looking at the behavior of $\hat{\beta}_1(\boldsymbol{s})$ for adjacent locations to the Fort Collins grid cell is likely to lead one to be suspicious of the strength of the trend found at the Fort Collins grid cell when analyzing only ensemble member 17. Figure 3 shows the point estimate for $\hat{\beta}_1(\boldsymbol{s})$ and the lower and upper bounds of the 95% CIs. The point estimates for the eastern cells east are negative, while those for the central and western cells are positive. It seems unlikely that extreme precipitation would have opposite trends over such a short spatial distance, even given the fact that Fort Collins likes at the eastern boundary of the Rocky Mountains. Furthermore, values for the lower CI bound show that the Fort Collins cell and cell immediately north would have significant positive trends, and the upper CI bound values show that the cell immediately east of Fort Collins would have a significantly negative trend.

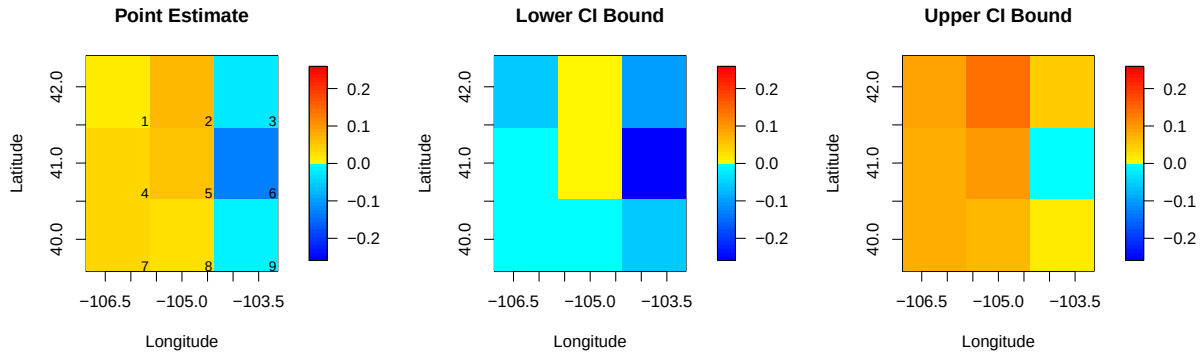

FIGURE 3: Point estimates (left), lower 95% CI bound (center), and upper 95% CI bound (right) for $\beta_1(\boldsymbol{s})$ (mm/year). Yellow-to-red colors indicate positive values while turqoise-to-blue indicate negative values.

While doing RFA or constructing a hierarchical model lies outside the scope of this review, we wish to illustrate how borrowing strength across locations can both reduce uncertainty and aid comprehension. We consider only the Fort Collins grid cell and the eight neighboring cells, but a more thorough analysis would employ many more over a study region of interest. Furthermore, a more thorough analysis would likely employ a spatial model on the parameters rather than the admittedly simple approach we use below.

Let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_9$ denote the grid cell locations as shown in the left panel of Figure 3. We will assume that local differences in the distribution of annual maxima can be captured in the intercept term of the location parameter $\beta_0(\boldsymbol{s}_i)$ and the scale parameter $\tau(\boldsymbol{s}_i)$. We will assume that $\beta_1(\boldsymbol{s}_i) = \beta_1$ and that $\xi(\boldsymbol{s}_i) = \xi$. That is, we will assume that the temporal trend is shared by these nine grid cells, as we believe that it seems reasonable that even if there are local differences in the distributions of annual maxima, the way these distributions change due to changes in forcings will likely be regional. We also assume a common shape parameter for the region, as we believe the fundamental nature of the tail will be nearly constant over such a small region. Furthermore, ξ is difficult to estimate and so borrowing strength to estimate this parameter is

advantageous. Let $m_b(\boldsymbol{s}_i)$ be the annual maximum precipitation measurement for block (year) b at location $\boldsymbol{s}_i$. Our likelihood is

$$\ell(\boldsymbol{\beta}_0, \beta_1, \boldsymbol{\tau}, \xi) = \prod_{i=1}^{9} \prod_{b=1920}^{2005} g(m_b(\boldsymbol{s}_i); \eta_b(\boldsymbol{s}_i) = \beta_0(\boldsymbol{s}_i) + \beta_1(b - 1919), \tau(\boldsymbol{s}_i), \xi),$$

where g denotes the GEV density function, $\boldsymbol{\beta}_0 = (\beta_0(\boldsymbol{s}_1), \ldots, \beta_0(\boldsymbol{s}_9))^T$, and $\boldsymbol{\tau} = (\tau(\boldsymbol{s}_1), \ldots, \tau(\boldsymbol{s}_9))^T$.

The estimated AEP(.05) level is shown in the bottom panel of Figure 2. The uncertainty in this quantile estimate has been reduced and the trend is less pronounced than when only output for the Fort Collins grid cell was used. This reduction in trend is confirmed by the point estimate, which appears in the left panel of Figure 1 at the location marked "17 Grid". Although the magnitude of the estimated trend is reduced, normal-based confidence intervals would still lead to the conclusion of a signifcant positive trend. The right panel of Figure 1 shows a noteworthy reduction in the width of the CI for $\hat{\xi}$.

The idea of borrowing strength across locations involves a bias/variance tradeoff different from the one associated with threshold selection mentioned earlier. When estimating the parameters associated with a particular location, using data from other locations leads to bias, and one needs to think carefully about how one's model is pooling information from other locations. In our simple example, we relied on intuition about parameter behavior to decide for which parameters we would borrow strength. Both RFA and hierarchical models deal with data-pooling issues, but in different ways. RFA uses diagnostic tools to pre-define regions over which to pool data, while hierarchical modeling typically allows the spatial model to dictate how pooling will be accomplished. We note that borrowing strength across locations for the purpose of better estimating marginal parameters is fundamentally different both in approach and aim from spatially modeling dependence in the data, and these differences are examined in Davison et al. (2012).

When we used the entire ensemble to estimate the trend for the Fort Collins grid cell, we found no significant trend and the point estimate was negative, contrary to the expected behavior of extreme precipitation in a warming climate. Fix et al. (2016) studied this same GCM output more extensively. There are two important differences between the analysis in Fix et al. (2016) and that done here. First, Fix et al. (2016) more sensibly regressed the GEV location and log-scale parameters on global mean temperature rather than year, and this allows the trend in extreme precipitation to be non-linear in time. Second, Fix et al. (2016) modeled the output for the period 1920-2080, using both the historical forcings we studied for the period 1920-2005, and additionally output forced by future projections (RCP4.5 and RCP8.5). Fix et al. (2016) found a significant positive relationship between the GEV parameters and increasing global mean temperature, and much of the change in model-produced extreme precipitation behavior occurred in the latter portion of the studied period due to accelerated warming.

0.3 Multivariate Extremes

We now turn our attention to modeling multivariate extreme values. We will restrict our attention to the bivariate case as this is most easily discussed and visualized. The ideas presented herein can be extended to dimensions greater than two, but modeling extremes of even moderate dimension (e.g., 4 or 5) remains a challenging problem.

Our underlying guiding principle continues to be "let the tail speak for itself," but now we must accurately account for the *dependence* which is found in the joint tail of the distribution. A full accounting of dependence requires one to know the entire joint distribution. In the absence of such knowledge, correlation is widely used

as a summary metric of bivariate dependence. However, correlation alone does not typically yield complete information about dependence unless the distribution is assumed to be Gaussian. Correlation will not be useful for our purpose in characterizing dependence in the tail, as its very definition implies that it measures dependence at the center of a distribution.

We will present a framework which will asymptotically characterize the dependence in the tail of a bivariate distribution. This framework arises from asymptotic results, but is also motivated by practical modeling goals which could require extrapolation. Before we develop this framework it is useful to present a summary metric of bivariate tail dependence. Suppose $\boldsymbol{Y} = (Y_1, Y_2)$ is a bivariate vector with marginal cdfs $F_1(y)$ and $F_2(y)$. Coles et al. (1999) define

$$\chi = \lim_{u \to 1} P(F_1(Y_1) > u \mid F_2(Y_2) > u).$$

Similar to correlation, χ is a useful summary metric, but does not give complete information about dependence in the tail. There are several other related bivariate metrics of tail dependence such as the extremal coefficient (Schlather and Tawn, 2003; Davison et al., 2018) and the madogram (Cooley et al., 2006). An advantage of χ is that its definition is readily interpretable and clearly focuses on dependence in the joint tail. We note that similar to correlation, the definition of χ requires one to account for differences in marginal behavior (here by applying F_i) before quantifying dependence.

When $\chi = 0$, the components Y_1 and Y_2 are termed asymptotically independent. Asymptotic independence is a degenerate case as many bivariate distributions with different dependence structures have this property. In this review, we focus on the asymptotic dependence case ($\chi > 0$) for two reasons. First, positive values of χ indicate that largest values of Y_1 and Y_2 can occur concurrently, and we are particularly interested in accurately modeling phenomena where risk is exacerbated by joint occurrences of large values. Second, modeling extreme behavior under asymptotic independence is much more difficult, though it continues to be an area of interest (e.g., Ledford and Tawn, 1996; Resnick, 2002; Weller and Cooley, 2013).

0.3.1 Multivariate EVDs and componentwise block maxima

The classical approach for multivariate extremes is to describe the class of limiting distributions of componentwise maxima. Let $\boldsymbol{Y}_t = (Y_{t,1}, Y_{t,2})$ be an iid sequence of random vectors, and let $\boldsymbol{M}_n = (\bigvee_{t=1}^n Y_{t,1}, \bigvee_{t=1}^n Y_{t,2})$. If there exist sequences $\boldsymbol{a}_n > 0, \boldsymbol{b}_n \in \mathbb{R}^2$ such that

$$\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \xrightarrow{d} G,$$

then G is said to be a multivariate extreme value distribution (MVEVD) (arithmetic operations on vectors are applied componentwise throughout). The MVEVDs can be shown to be max-stable, that is $G^n(\boldsymbol{a}_n \boldsymbol{y} + \boldsymbol{b}_n) = G(\boldsymbol{y})$ for all $n \in \mathbb{Z}_+$. As in the univariate case, if $\boldsymbol{Y}_t$ is not iid but is stationary and mixes sufficiently, the limit remains a MVEVD (Beirlant et al., 2004, Section 10.5).

Because convergence of multivariate distributions requires convergence of the univariate marginals, the univariate marginals of a MVEVD G must be EVDs. However, marginal convergence to EVDs does not guarantee that a multivariate distribution belongs to the class of MVEVDs. For example, the distribution $F(\boldsymbol{y}) = \exp\{-(y_1^{-1} + y_1^{-1}y_2^{-1} + y_2^{-1})\}$ has unit Frechét marginals, and hence $\boldsymbol{a}_n = (n, n)^T$ and $\boldsymbol{b}_n = \boldsymbol{0}$. However, $F^n(\boldsymbol{a}_n \boldsymbol{y} + \boldsymbol{b}_n) = \exp\{-(y_1^{-1} + n^{-1}y_1^{-1}y_2^{-1} + y_2^{-1})\} \neq F(\boldsymbol{y})$. F is in the domain of attraction of the MVEVD $G(\boldsymbol{y}) = \exp\{-(y_1^{-1} + y_2^{-1})\}$.

Characterizing the class of MVEVDs is more difficult than the univariate case because it cannot be reduced to a simple parametric form. Generally, the family is characterized by assuming some canonical form for the marginal distributions. We will assume that the marginals of G are unit Frechét, which is sometimes referred

to as the 'simple' case (de Haan and Ferreira, 2006, p. 217). One can show that the class of simple MVEVDs has the form

$$G(\boldsymbol{y}) = \exp\{-V(\boldsymbol{y})\}, \tag{5}$$

where the exponent measure function V has the property $V(s\boldsymbol{y}) = s^{-1}V(\boldsymbol{y})$ for any $s > 0$ (Beirlant et al. (2004) gives a full development). With this property, G is easily shown to be max-stable. Furthermore, the tail dependence metric $\chi = 2 - V(1, 1)$.

Given that n is large enough, the distribution of $\boldsymbol{M}_n$ should be well-approximated by a MVEVD. Let n denote a selected block length and let $\boldsymbol{y}_{b,t} = (y_{b,t,1}, y_{b,t,2}), b = 1, \ldots, B; t = 1, \ldots, n$ be a sequence of observations from a stationary process. Let $\boldsymbol{m}_b = (\bigvee_{t=1}^{n} y_{b,t,1}, \bigvee_{t=1}^{n} y_{b,t,2})$ be the B componentwise block maximum observations. One method for estimating the distribution of $\boldsymbol{M}_n$ is to fit a parametric model for a MVEVD to the block maxima data $\boldsymbol{m}_1, \ldots \boldsymbol{m}_B$. Several bivariate parametric models for $V(\boldsymbol{x})$ have been suggested (Kotz and Nadarajah, 2000, Section 3.4).

The cdf in (5) assumes the marginals are unit Frechét, which is generally not true in practice. It may be useful to envision a two-step process, where one first fits GEVs to the univariate marginals, transforms the data to have unit Frechét margins, then fits a parametric form for $V(\boldsymbol{x})$. A marginal transformation to Frechét can be defended by Proposition 5.10 of Resnick (1987), which states that the domain of attraction is unchanged by monotone transformations of the marginals. This result can be interpreted as an extremes result of Sklar's theorem about copulas (Nelsen, 2006). In practice, inference for the marginals and dependence structure can be done all-at-once, allowing one to better capture the full uncertainty than a two-step method. Coles (2001) provides an illustrative example where the annual sea levels at two Australian locations are modeled by a bivariate EVD with a logistic (Gumbel, 1960) dependence structure. The `R` package `evd` (Stephenson, 2002) provides a tool to perform bivariate maximum likelihood estimation.

Although the aforementioned statistical approach is nicely justified by the argument that the MVEVDs are the limiting distribution of a vector of block maxima, one might ask why one would want to model such data. Constructing a vector of componentwise maxima may seem unnatural, as the constructed block-maximum $\boldsymbol{m}_b$ likely does not correspond to any observation $\boldsymbol{y}_{b,t}, t \in \{1, \ldots, n\}$. When confronted with such an approach, our experience is that cooperating scientists are at first dubious. There are questions of interest which are naturally answered by an analysis of block maxima. The sea level example of Coles (2001) could be motivated by an insurance application where the company's reserves are calculated on an annual basis, and one needs to assess the probability that the maximum sea level in any given year could exceed some damage threshold resulting in claims from both of the two cities. More importantly, a vector of componentwise maxima which occur at different times still retains information about tail dependence. Heuristically, this can be understood as follows. Suppose the first component of block b's maximum, $m_{b,1}$ occurs at time t^* and is unusually large for a block maximum. If the components of $\boldsymbol{Y}_t$ exhibit asymptotic dependence, then we would expect $y_{b,t^*,2}$ to be quite large. Since $m_{b,2} > y_{b,t^*,2}$, even if $m_{b,2}$ does not occur at time t^* we would still expect it to be larger than usual. Another reason which could motivate the analysis of componentwise block maxima is that extreme observations due to the same event might not be recorded contemporaneously (Johndrow and Wolpert, 2016); for example, a single storm might affect two locations on different days.

0.3.2 Multivariate threshold exceedances

In the univariate case, the GPD used to model threshold exceedances was intricately tied to results for block maxima. In the multivariate setting, we would like any modeling approach to likewise be related to the MVEVDs. From a modeling point-of-view, we desire an approach which utilizes only observations deemed extreme, and which can exhibit asymptotic dependence. This second requirement precludes most familiar multivariate models, including any model which exhibits Gaussian dependence (e.g., a Gaussian copula) (Sibuya, 1960).

To model multivariate threshold exceedances, one must first define what is meant by a multivariate exceedance. One approach is to set threshold values for each univariate marginal distribution which naturally leads one to describe events in terms of Cartesian coordinates. The multivariate GPD of Rootzen and Tajvidi (2006) is defined in terms of Cartesian coordinates, and can be tied to the exponent measure function $V(\boldsymbol{x})$ in the $\xi = 1$ case. In the bivariate case with thresholds for each marginal, an observation can have one component exceed the threshold while other does not. Smith et al. (1997) proposed using a censored likelihood approach for such settings. Below, we will define a threshold exceedance in terms of the norm of the vector, and we will model via the framework of regular variation, which again can be related to the representation for MVEVDs in (5). Heffernan and Tawn (2004) give an alternative, conditional approach to modeling multivariate extreme values. This approach was related to regular variation in Heffernan and Resnick (2007), and while we do not discuss it here, the conditional approach has been shown to be particularly useful for modeling the joint tail in the case of asymptotic independence.

Regular variation implies that the tail of the distribution behaves like a power function. A comprehensive treatment of regular variation is given by Resnick (2007). The MV regularly varying distributions classify the family of distributions which are in the domain of attraction of a MVEVD with heavy-tailed marginals with a common tail index $\xi > 0$. Fundamental to our approach is that regular variation is easily described in terms of polar coordinates.

Before we formally define regular variation, let us investigate behavior via a specific example. $\boldsymbol{Z} = (Z_1, Z_2)$ has a bivariate logistic distribution (Gumbel, 1960) if $P(\boldsymbol{Z} \leq \boldsymbol{z}) = \exp\{-(z_1^{-1/\beta} + z_2^{-1/\beta})^{\beta}\}$, for $z_1, z_2 > 0$ and some $\beta \in (0, 1]$. This distribution has unit Fréchet marginals and also happens to be max-stable, although only the fact that it is regularly varying is important for our purposes here. We construct iid replicates $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ of $\boldsymbol{Z}$ where $\beta = .5$, and in the left and center panels of Figure 4 we plot $\boldsymbol{z}_i/n$ for $n = 500$ and $n = 5000$. The plots appear similar because the realizations have been normalized by n, and although the right panel has ten times as many points as the left, the normalization causes most of these points to pile up near the origin.

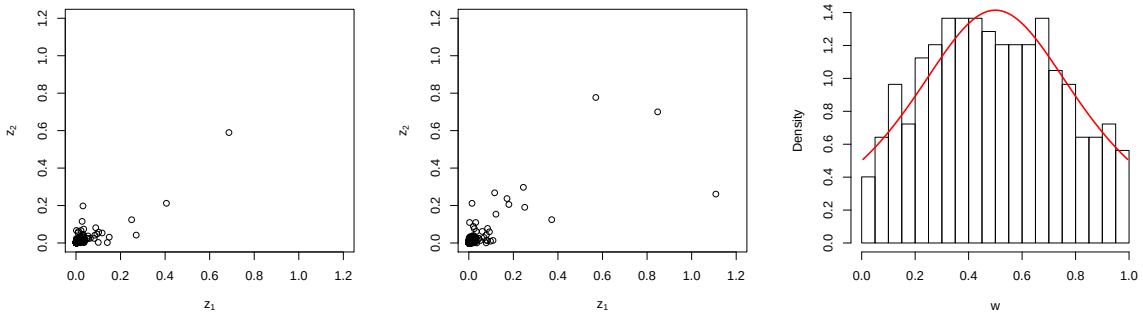

FIGURE 4: Scatterplots of $\boldsymbol{z}_i/n$, where $\boldsymbol{z}_i$ are realizations from a bivariate logistic random vector and $n = 500$ (left) and $n = 5000$ (center). The right panel shows the angular density h along with a histogram of the angular components of the largest 5% of the realizations.

The definition of regular variation involves a convergence statement and is not straightforward, but our example above helps explain. A nonnegative random vector $\boldsymbol{Z} = (Z_1, Z_2)$ is said to be regularly varying if there exists a sequence a_n such that

$$nP\left(\frac{\boldsymbol{Z}}{a_n} \in A\right) \xrightarrow{v} \nu(A), \tag{6}$$

where ν is a Radon measure on $\mathcal{C} = [0, \infty]^2 \setminus \{\boldsymbol{0}\}$, A is any ν-continuity set in $\mathcal{C}$, and v denotes vague convergence (a type of weak convergence) for measures (Resnick, 2007, p.49). For the logistic simulation

the normalizing sequence $a_n = n$. The left-hand side of (6) reminds one of convergence of a binomial to a Poisson, and the measure ν on the right-hand side yields the expected number of points falling in a set A. Although we do not develop it here, there is an equivalent point-process representation of regular variation, and Davison et al. (2018) use a similar argument to develop spatial extremes models in Chapter XX. The origin is excluded from $\mathcal{C}$ because the limit measure ν would be infinite for any set which included it, as more and more points pile up at the origin as $n \to \infty$.

It can be shown that for any scalar c, the measure ν has the property

$$\nu(cA) = c^{-1/\xi}\nu(A), \tag{7}$$

where $1/\xi > 0$ is the index of regular variation and from which we see the aforementioned power law tail behavior. For our logistic example, $\xi = 1$.

The radial scaling property (7) gives rise to an equivalent definition to (6), but in terms of polar coordinates. For any norm, let $R = \|\boldsymbol{Z}\|$ and $\boldsymbol{W} = \boldsymbol{Z}/R$,

$$nP\left(R/a_n > r, \boldsymbol{W} \in B\right) \xrightarrow{v} r^{-1/\xi}H(B), \tag{8}$$

where H is some Radon measure on $\mathcal{S} = \{\boldsymbol{x} \in \mathcal{C} : \|\boldsymbol{x}\| = 1\}$, and B is any H-continuity set in $\mathcal{S}$. Critically for modeling, (8) implies that the radial component R becomes independent of the angular component $\boldsymbol{W}$ as the random vector becomes large (in terms of the radial component). The measure H describes the dependence found in the joint tail of $\boldsymbol{Z}$ and is termed the spectral or angular measure; we prefer the lesser-used 'angular' terminology so as not to lead to confusion with Fourier-based methods. More mass of H toward the center of $\mathcal{S}$ implies stronger dependence and that Z_1 and Z_2 are more likely to be large at the same time. The special case of asymptotic independence corresponds to H consisting of point masses at the points (0,1) and (1,0). If H is continuously differentiable, let $h(\boldsymbol{w})$ denote this derivative. Depending on the choice of $\{a_n\}$, H may or may not be a probability measure; for our logistic example with $a_n = n$, $H(\mathcal{S}) = 2$.

Expression (8) implies that the large values of $\boldsymbol{Z}$ can be well modeled by a univariate radial component R whose (large) values follow a power law described by ξ and an independent angular component $\boldsymbol{W}$ whose behavior is determined by H. In the $\xi = 1$ case, it is convenient to use the L_1 norm: $\|\boldsymbol{Z}\|_1 = Z_1 + Z_2$ to define these polar coordinates. Furthermore in the bivariate case, $\mathcal{S}$ is isomorphic to $[0, 1]$. For our logistic example, the form of h is known. Letting $r_i = \|\boldsymbol{z}_i\|_1$ and $w_i = z_{i,1}/r_i$, the right panel of Figure 4 shows a histogram of the angular components w_i that correspond to the largest 5% of the radial components along with the density given by (normalized) h.

To tie regular variation to the MVEVDs, assume $\boldsymbol{Y}_t$ (introduced in Section 0.3.1) is in the domain of attraction of a simple MVEVD. Then $\boldsymbol{Y}_t$ is regularly varying with $\xi = 1$ and limiting measure ν (alternatively angular measure H) such that the exponent measure function

$$V(\boldsymbol{y}) = \nu([\boldsymbol{0}, \boldsymbol{y}]^c) = \int_{(\boldsymbol{w},r)\in[\boldsymbol{0},\boldsymbol{y}]^c} r^{-2}dH(\boldsymbol{w}) = \int_{\mathcal{S}} \max\left(\frac{w}{y_1}, \frac{1-w}{y_2}\right) dH(\boldsymbol{w}).$$

If the regularly varying random vector has common marginals, then there exists a balance condition on the angular measure:

$$\int_{\mathcal{S}} w dH(w) = \int_{\mathcal{S}} (1-w) dH(w). \tag{9}$$

If h exists and is symmetric about 0.5, then this balance condition will be met, but symmetry is not required.

Returning to the modeling objectives listed at the beginning of this section, MV regular variation is a suitable modeling framework for multivariate threshold exceedances as its definition only describes the tail of the distribution, it is a framework that can accommodate asymptotic dependence, and it has natural connections to the MVEVDs. Similar to modeling with parametric MVEVD models, one can envision a two step

process where the marginals are transformed to a convenient heavy-tailed distribution, and then dependence is described by modeling H. Proposition 5.10 of Resnick (1987) continues to imply that this marginal transformation does not affect the fundamental nature of dependence. This two step procedure is similar in spirit to copula methods, but with important differences: (1) only a small subset of large observations (defined in terms of the radial component after transformation) are used in the analysis, and (2) a model specifically suited for describing tail dependence is used. Coles and Tawn (1991) introduced a likelihood-based inference method for exceedances based on a point-process representation of regular variation.

0.3.3 Application: Santa Ana winds and dryness

Risk for wildfires is related to many conditions, among which are air temperature, windspeed, humidity, and supply of fuel. An atmospheric regime commonly referred to as the Santa Ana winds is known to be related to wildfire occurrence in southern California. The Santa Ana winds are actually a multivariate meteorological regime. Because the regime originates in the desert east of southern California, they are associated with warm temperatures, low humidity, and high winds. This regime mostly occurs in autumn and winter, but we restrict our attention to autumn (September, October, and November) as these months have historically had more fires than the winter months. Two of the most destructive fires on record in California, the Cedar Fire in 2003 and the Witch Fire in 2007, were associated with Santa Ana conditions..

We obtain data from the March AFB station[2] from the HadISD dataset (Dunn et al., 2012). The station was chosen because of its location (Riverside County CA, longitude -117.3, latitude 33.9, elevation 468 m), its length and relative completeness of record, and because this station's variables appeared to show a signal with known Santa Ana events. The data encompass the years 1973-2015, and our analysis will assume the data are temporally stationary.

We wish to have a bivariate time series that conveys a daily risk of fire, and we focus on meterological variables which summarize windspeed and the dryness of the air. As is often the case in environmental studies, the data we analyze result from some preprocessing. The HadISD data are recorded hourly, and provide a windspeed (m/s) and dewpoint measurement. We choose to work with relative humidity (%) rather than dewpoint, as dewpoint is directly affected by air temperature and displays a decreasing trend over the course of the autumn season. Relative humidity is calculated as a function of temperature and dewpoint via a commonly used formula employing constants suggested by Alduchov and Eskridge (1996). Figure 5 shows the time series of hourly windspeeds and relative humidity for 10 days in October 2003, including the day of occurrence of the Cedar Fire. Both time series show clear diurnal cycles: windspeeds tend to be largest during the day and relative humidity increases at night. On October 25, we see the pattern is disrupted by the Santa Ana event which corresponded with the Cedar Fire. Windspeeds are stronger than on previous days, and the diurnal cycle of relative humidity is disrupted with very little increase in humidity that night. Based on exploratory analysis of this and several other known Santa Ana events, we construct our daily summaries as follows. Our daily windspeed measurement is taken to be the mean of the four maximum hourly windspeeds during the 24 hour period beginning at midnight the day of recording. Taking the mean of the four maxima not only helps to provide a daily windspeed summary but also helps to create daily data that are more continuous than the hourly data. Our daily dryness measurement is taken to be the mean of the six relative humidity measurements beginning at 11 pm the day of recording and ending at 4 am the following day, which is then negated so that the greatest values correspond to the driest conditions. Although our daily dryness measure includes several relative humidity measurements taken from the first few hours of the following day, our hope is that this signal is associated with the recorded day's dryness, and we believe this signal is clearer than one which would employ daytime relative humidity measurements which are often quite dry at this location regardless of whether in the Santa Ana regime or not.

[2]Station number 722860-23119.

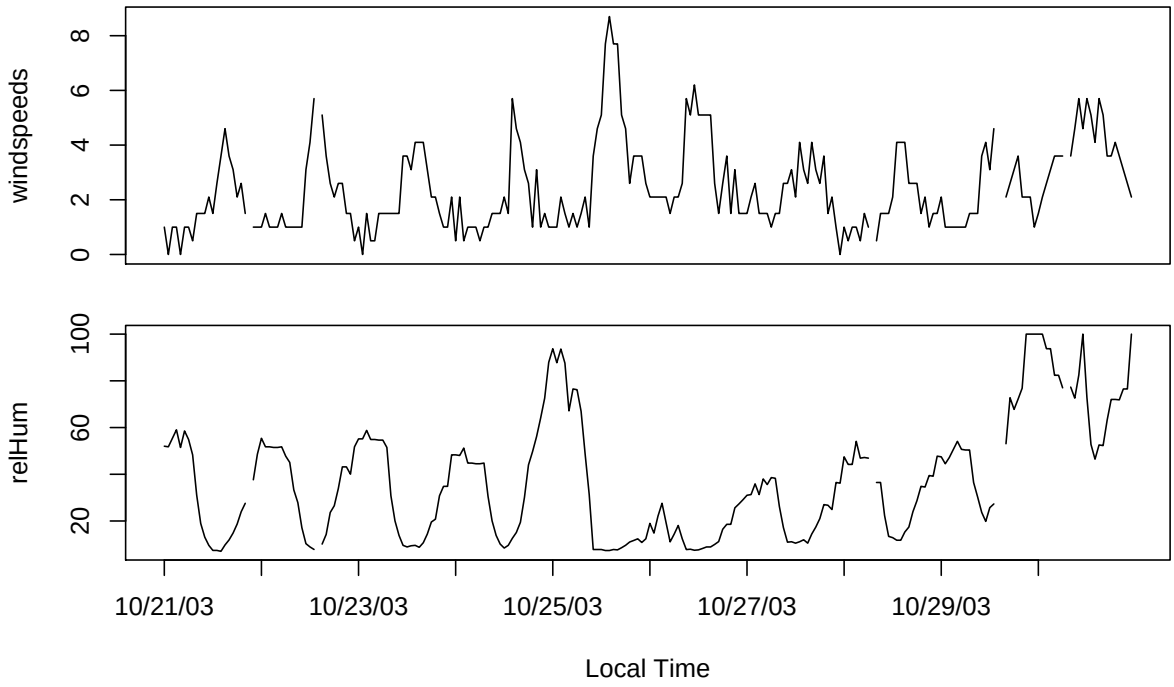

FIGURE 5: Time series of hourly windspeed (m/s) and relative humidity (%) for 10/21/2003 through 10/31/2003. The Cedar fire occurred on 10/25/2003.

The left panel of Figure 6 shows our daily summary values for windspeed and dryness. The scatterplot shows that high windspeed does not have to occur when the air is particularly dry, nor do dry conditions only occur when the windspeed is high. However, the plot also shows that very high windspeed and very dry conditions can occur simultaneously, and putting these two conditions together leads to the explosive fire conditions about which we are concerned. The red dot and orange dot respectively indicate October 25, 2003 and October 21, 2007, the initial days of the Cedar and Witch Fires. Both are days which experienced high winds and dry conditions (the Witch Fire in fact corresponds to the day with the driest conditions in our record), but a number of days show similar dryness and higher windspeed. Of course, high windspeeds and dry air do not themselves cause large fires to happen; a wildfire requires an ignition source, the size of a fire depends on the immediacy of response and accessibility of location, and fire risk is not fully captured by these two variables.

0.3.3.1 Assessing tail dependence

We wish to assess the tail dependence between our daily and dryness variables, and ultimately to assess how often these high-risk conditions occur simultaneously. The center panel of Figure 6 shows $\hat{\chi}$ for u increasing from 0.70 to almost 1. If we attribute the drop at the end of the plot to very low sample size (which seems justified by the uncertainty shown by the approximate 95% confidence intervals), the plot shows that estimates for χ seem to level off at about 0.3 for high values of u. This can loosely be interpreted by saying that when windspeeds are at their highest levels, our dryness measure is also at its highest levels about 30% of the time. Based on the plot of $\hat{\chi}$ and the scatterplot, we conclude that our data exhibit asymptotic dependence.

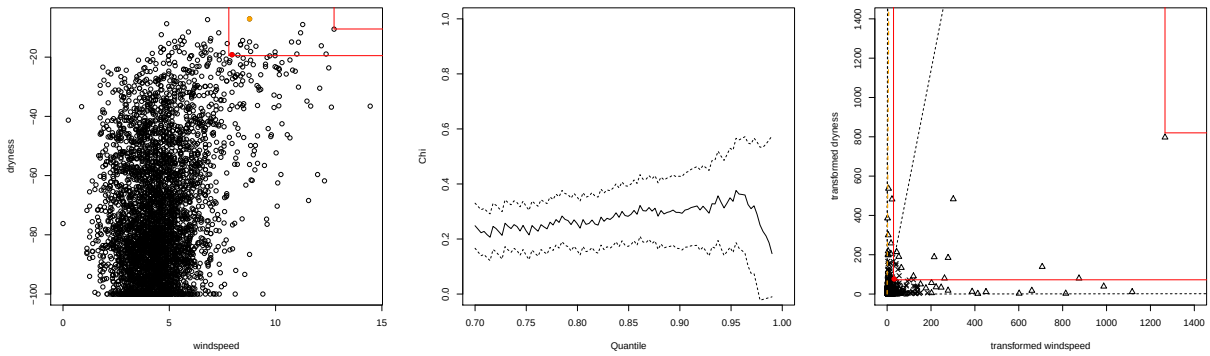

FIGURE 6: Left: scatterplot of daily summary values for windspeed and dryness. Center: plot of $\hat{\chi}$. Right: scatterplot of daily summaries after marginal transformation to unit Frechét. The red solid lines region define risk regions $\mathcal{R}_1$ and $\mathcal{R}_2$ in the original scale and $\mathcal{R}_1^*$ and $\mathcal{R}_2^*$ in the Frechét scale. The red dot corresponds to the day of the Cedar Fire, and the orange dot corresponds to the day of the Witch Fire. Dashed rays in the Frechét scale plot point to events which lie outside the plot window.

Let $\boldsymbol{Y}_t = (Y_{t,1}, Y_{t,2})$ be the random vector whose variates denote our windspeed and dryness measures. We will estimate the probability of an event occurring in two different risk regions. Based somewhat on the values associated with the Cedar Fire, we define $\mathcal{R}_1 = \{\boldsymbol{y} \in \mathbb{R}^2 : y_1 > 7.8, y_2 > -19.5\}$. The risk region $\mathcal{R}_1$ is shown in the left panel of Figure 6. As the Cedar Fire is in $\mathcal{R}_1$, we know that conditions in this region are conducive to large fires. We define $\mathcal{R}_2 = \{\boldsymbol{y} \in \mathbb{R}^2 : y_1 > 12.75, y_2 > -10.5\}$. This region, also shown in the left panel of Figure 6, is very extreme and contains no observations in our data record. We aim to obtain estimates of $p_1 = P(\boldsymbol{Y}_t \in \mathcal{R}_1)$ and $p_2 = P(\boldsymbol{Y}_t \in \mathcal{R}_2)$. It is clear that it will be necessary to accurately account for the dependence in the joint tail to assess p_1 and p_2.

Let $\boldsymbol{y}_t = (y_{t1}, y_{t2}), t = 1, \ldots, n$ denote our bivariate windspeed and dryness observations. Based on mean residual life plots (Coles, 2001, Section 4.3), GPD's were fit to the upper 5% of the data of each marginal. Estimates of ξ were -0.15 (se = 0.07) for windspeed and -0.41 (0.05) for dryness. Thus, the univariate marginals appear to have bounded tails, and these fitted GPD's yield point estimates of the upper endpoints for the respective distributions of 17.8 and -5.3. Bounded tails agree with basic intuition and knowledge about these variates; for instance, it would be impossible for our dryness variate to return a positive value.

Using the regular variation framework to model tail dependence requires a transformation of the marginal distributions. We choose to use a combination of parametric and nonparametric methods to transform the marginal distributions to be approximately unit Frechét. Let $\hat{u}_i, \hat{\psi}_{u_i}$, and $\hat{\xi}_i$ denote the empirical .95 quantile, and GPD scale and shape estimates for the ith marginal, $i = 1, 2$. We let

$$\hat{F}_i(y) = \begin{cases} (n+1)^{-1} \sum_{t=1}^{n} \mathbb{I}(y_{t,i} \leq y) & \text{for } y \leq \hat{u}_i, \\ 1 - .05\big(1 + \hat{\xi}_i(y - \hat{u}_i)/\psi_{ui}\big)^{-1/\hat{\xi}_i} & \text{for } y > \hat{u}_i. \end{cases} \tag{10}$$

Then, letting $t(\boldsymbol{y}) = (t_1(y_1), t_2(y_2))$ where $t_i(y) = G^{-1}(\hat{F}_i(y))$ and G is the cdf of the unit Frechét, we construct transformed observations $\boldsymbol{z}_t = t(\boldsymbol{y}_t)$.We also produce the polar transformed observations $r_t = z_{t,1} + z_{t,2}$ and $w_t = z_{t,1}/r_t$. Additionally, we can transform the risk regions to the new space: $\mathcal{R}_i^* = t(\mathcal{R}_i), i = 1, 2$.

The right panel of Figure 6 shows the data after transformation, as well as the transformed risk regions. Shown with a red dot is the day corresponding to the Cedar Fire. Plots on the Frechét scale can be hard to interpret because the transformed data are so heavy-tailed. In the figure, events that have radial components R_t larger than the .99 empirical quantile have been plotted with triangles, events with R_t between the .98

and .99 empirical quantiles are shown with X's, and all other events are plotted with circles. We see that on this Frechét scale, one's eye is naturally drawn to only the very largest events, as it is difficult to distinguish any points other than the largest 1%. Because the largest transformed points are so large, we have chosen the plot window to show only the region $[0, 1400]^2$, and five points lie outside this region. One of these points lies near the x-axis at (15381, 10.9), and another lies distant from both axes at (309, 1894). The other three points are near the y-axis at (14.8, 6977), (2.5, 2257), and (52, 9265), and this last point corresponds to the Witch Fire. Rays are shown in the figure extending to these five points; we have indicated with an orange dashed line the ray to the Witch Fire, which lies inside $\mathcal{R}_1^*$.

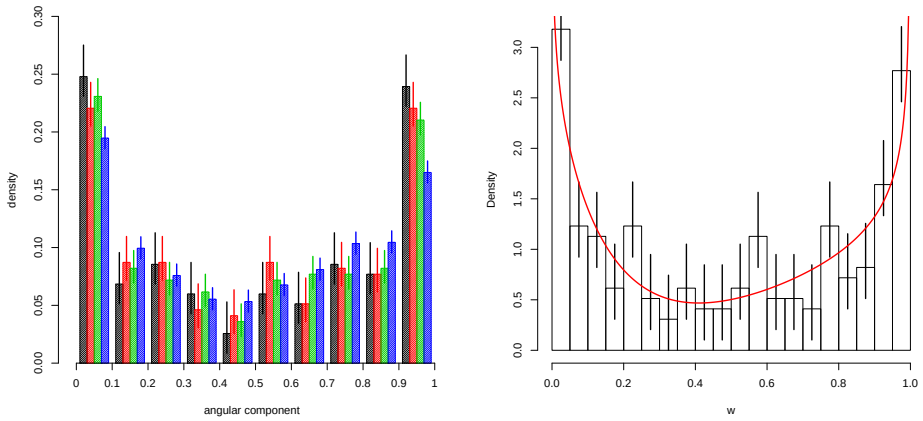

FIGURE 7: Left: histograms of the angular components of the events whose radial components exceed the .97 (black), .95 (red), .90 (green), and .75 (blue) quantiles. Right fitted parametric angular measure model of a mixture of two beta distributions.

We next investigate and model the angular measure. Histograms of angular components for events whose radial components exceed the .97, .95, .90, and .75 quantiles are shown in the left panel of Figure 7. Accounting for the uncertainty associated with these histograms (shown by .95 confidence intervals based on the multinomial distribution), it appears that the distributions of angular components whose radial components exceed the .90, .95, and .97 quantiles are very similar. Not surprisingly, one sees a noteworthy change in the distribution of angular components whose corresponding radial components only exceed the .75 quantile. We decide to model the angular measure by retaining angular components whose radial component exceeds the empirical .95 quantile, which coincides with the quantiles used for parametric modeling of the marginal tail, though there is no theoretical reason why these quantiles should match. Furthermore, as shown in the right panel of Figure 6, many points with large radial components will have either $z_{t,1}$ or $z_{t,2}$ below the marginal .95 quantile. The right panel of Figure 7 shows a histogram of the values of $w_{t,1}$ for the events whose radial component r_t exceeds the .95 empirical quantile, along with .95 confidence intervals based on the multinomial distribution.

Using the likelihood-based approach of Coles and Tawn (1991), we fit parametric models to the angular components of the large points. We fit a few angular measure models, including the angular measure corresponding to the logistic model and one given by a beta density, both of which have a single parameter and yield symmetric densities. Of the models we tested, we select a model based on a mixture of two beta densities as proposed by Boldi and Davison (2007), which allows for asymmetry and which has four parameters. Letting $\boldsymbol{\theta} = (\pi_1, \alpha_1, s_1, s_2)$,

$$h(w; \boldsymbol{\theta}) = \sum_{j=1}^{2} \pi_j \frac{\Gamma(s_j)}{\Gamma(s_j\alpha_j)\Gamma(s_j(1-\alpha_j))} w^{s_j\alpha_j - 1}(1-w)^{s_j(1-\alpha_j)-1}, \; w \in [0,1]$$

where the mixture weights $\pi_1 + \pi_2 = 1$ and $\pi_j > 0$, $s_j > 0$, and $0 \leq \alpha_j \leq 1$ all for $j = 1, 2$. The

additional constraint $\pi_1\alpha_1 + \pi_2\alpha_2 = 1/2$ assures condition (9) is met. Maximum likelihood estimates are $\hat{\boldsymbol{\theta}} = (0.39, 0.12, 7.19, 2.78)$. Choice of this model is also supported by AIC values of the three model fits. The fitted angular measure is shown in the right panel of Figure 7. While the fitted model does not match all the irregularities of the histogram, we contend that it captures much of the behavior of the angular components, and that fitting even more complex parametric models would risk over-fitting. For more flexibility, it is possible to model angular measures for bivariate vectors non-parametrically and maintain the balance condition (9) (Einmahl and Segers, 2009), but we do not pursue that here. To make sense of the asymmetry, note that if w is close to zero then the event has high/extreme dryness and non-high windspeed.

0.3.3.2 Risk region occurrence probability estimation

As p_1 and p_2 are equivalent to $P(\boldsymbol{Z}_t \in \mathcal{R}_1^*)$ and $P(\boldsymbol{Z}_t \in \mathcal{R}_2^*)$, we work in the Frechét transformed space and perform estimation of these probabilities in two different ways. First, using our parametric model fit above, for $i = 1, 2$,

$$\nu(\mathcal{R}_i^*) = \int_{(\boldsymbol{w},r)\in\mathcal{R}_i^*} r^{-2} h(\boldsymbol{w}, \boldsymbol{\theta}) d\boldsymbol{w} = \int_{w=0}^{1} \min\left(\frac{w}{z_{i,1}^*}, \frac{1-w}{z_{i,2}^*}\right) h(w, \boldsymbol{\theta}) dw, \tag{11}$$

where $(z_{1,1}^*, z_{1,2}^*) = (29.2, 72.5)$ and $(z_{2,1}^*, z_{2,2}^*) = (1266.4, 820.0)$ are the lower-left corners of $\mathcal{R}_1^*$ and $\mathcal{R}_2^*$ respectively. Plugging in $\hat{\boldsymbol{\theta}}$, we obtain $\hat{\nu}(\mathcal{R}_1^*) = 3.5 \times 10^{-3}$ and $\hat{\nu}(\mathcal{R}_2^*) = 1.7 \times 10^{-4}$. To obtain our probability estimate, we begin from (6), setting $a_n = 2n$ as this is the correct normalizing sequence for $\boldsymbol{Z}$ with Frechét marginals and h a probability density. Assuming n is fixed and large enough such that the convergence implies approximate equality, and letting $A = \mathcal{R}_i^*/2n$, we obtain

$$\begin{aligned} nP\left(\frac{\boldsymbol{Z}_t}{2n} \in \frac{\mathcal{R}_i^*}{2n}\right) &\approx \nu\left(\frac{\mathcal{R}_i^*}{2n}\right) \\ \Rightarrow P(\boldsymbol{Z}_t \in \mathcal{R}_i^*) &\approx 2\nu(\mathcal{R}_i^*). \end{aligned} \tag{12}$$

Plugging in $\hat{\nu}(\mathcal{R}_1^*)$ and $\hat{\nu}(\mathcal{R}_2^*)$ above yields estimates $\hat{p}_1 = 7.0 \times 10^{-3}$ and $\hat{p}_2 = 3.4 \times 10^{-4}$; uncertainty for these estimates is discussed below. Given that our studied season consists of 91 autumn days, the first probability estimate implies that we expect to see about 6 days in 10 seasons which fall in $\mathcal{R}_1$ (with conditions roughly as or more extreme than those associated with the Witch Fire). The second probability estimate implies that we would expect to see 3 days in 100 seasons which fall in $\mathcal{R}_2$.

Alternative to employing a parametric angular measure model, a nonparametric estimate can be obtained via the scaling property (7). We will scale by a factor of 10, and we define $\mathcal{R}_i^\dagger = \{z \in [0,\infty]^2 \setminus \{0\} : 10z \in \mathcal{R}_i\}$, for $i = 1, 2$. Via (7) and assuming the approximation in (12), one can obtain

$$P(\boldsymbol{Z}_t \in \mathcal{R}_i^*) \approx \frac{1}{10} P(\boldsymbol{Z}_t \in \mathcal{R}_i^\dagger).$$

Employing empirical estimates for $P(\boldsymbol{Z}_t \in \mathcal{R}_1^\dagger)$ and $P(\boldsymbol{Z}_t \in \mathcal{R}_2^\dagger)$ yields estimates $\tilde{p}_1 = 6.8 \times 10^{-3}$ and $\tilde{p}_2 = 1.5 \times 10^{-4}$.

Fully accounting for uncertainty in multivariate extremes methods requires accounting for uncertainty in the marginal estimation, and in the modeling of the tail dependence structure. We proceed with a straightforward paired bootstrap method which mimics our estimation methods above. Although we are not attempting to model temporal dependence, we use a block bootstrap approach with block length of five days in order to account for the increased parameter uncertainty due to temporal dependence. Let $\boldsymbol{y}_t^{(b)}, t = 1, \ldots, n$ represent a block-resample of $\boldsymbol{y}_t, t = 1, \ldots, n$. Marginally, we fit a GPD above the .95 quantile of this resample, and transform the marginals via $\boldsymbol{z}_t^{(b)} = t^{(b)}(\boldsymbol{y}_t^{(b)})$, where $t^{(b)}$ is as before, but with the data and GPD parameter estimates from the resample. Additionally, we create $\mathcal{R}_i^{*(b)} = t^{(b)}(\mathcal{R}_i)$. For the parametric method, we retain the points whose radial components are in the largest 5%, obtain parameter estimates $\hat{\boldsymbol{\theta}}^{(b)}$ for the

mixture-of-betas angular measure model, and integrate (11). For the non-parametric method, we calculate the number of bootstrap resampled points falling in $\mathcal{R}_i^{*(b)}/10$ and scale as above. To obtain 95% confidence intervals, we repeat these procedure for $b = 1, \ldots, B = 1000$, and take the .025 and .975 quantiles of the estimates $\hat{p}_i^{(b)}$. Our parametric bootstrap-estimated 95% confidence interval for p_1 is $(5.2 \times 10^{-3}, 8.9 \times 10^{-3})$ and for p_2 is $(8.1 \times 10^{-5}, 5.5 \times 10^{-4})$; our nonparametric intervals are respectively $(5.2 \times 10^{-3}, 8.8 \times 10^{-3})$ and$(2.5 \times 10^{-5}, 4.1 \times 10^{-4})$. Standard empirical estimates are $\check{p}_1 = 4.9 \times 10^{-3}$ and $\check{p}_2 < 2.5 \times 10^{-4}$. While the empirical estimate for p_1 doesn't quite fall in the 95% confidence interval, it does fall in a 98% interval.

For comparison, we obtain an estimate for p_1 using a simple Gaussian copula approach. Via rank transform, we obtain uniform marginals, and then estimate the correlation parameter of the Gaussian copula to be $\hat{\rho} = .23$. Then by transforming $\mathcal{R}_1$ to the corresponding region in $[0, 1]^2$, we obtain an estimate of 1.4×10^{-3} for p_1, much lower than either of the estimates via the regular variation framework, and also less than the empirical estimate for p_1. The Gaussian copula model is different from the regular variation-based approach in that it is asymptotically independent. Likely more important in terms of affecting the estimate of p_1, the Gaussian copula's dependence parameter estimate was based on the entire data set rather than focusing on extreme observations.

Our example illustrates a method for modeling dependence in the tail and using this model to estimate the probabilities associated with pre-defined risk regions. Our method relies on the framework of regular variation, which is inherently linked to multivariate EV theory, and which utilizes only information found in the extreme observations. Employing this framework on this windspeed and dryness data requires marginal transformation to a heavy-tailed distribution, and the transformation we employed is rather dramatic, as seen in the right panel of Figure 6. This could be considered a drawback, as interpreting dependence in this transformed heavy-tailed data set requires familiarity with the framework. Importantly, the regular variation framework allows one to model tail dependence in the asymptotically dependent case. If data exhibit asymptotic dependence, it is particularly important to employ a model which can capture asymptotic dependence when extrapolating beyond the range of the data, as we do to estimate p_2.

Our example assumed the data were stationary, ignoring changes occurring due to climate change. It is of great interest to better understand how fire risk will change due to an altered climate. This data set is likely too short to accurately model how these variates have changed over the study period, and even if such a model were obtained, projecting a nonstationary model into the future should be done cautiously. Nevertheless, one could imagine applying a method such as this to climate model output, run under both current climate and a projected future climate. By comparing the bivariate extremes models, one could assess how extreme behavior differs between the two periods, both marginally and in its tail dependence structure.

0.4 Conclusions

In this chapter, we aimed to provide an introduction to extreme value modeling both in the univariate and bivariate cases. Although our coverage of these topics is necessarily incomplete, hopefully the philosophy of letting the tail speak for itself was made clear in both our review of methods and our illustrative examples.

In the introduction we mentioned that extremes studies are relatvely prevalent in hydrology and atmospheric science. These disciplines benefit from data records from recording stations which take measurements at regular intervals and whose records can span a large number of years. Such data records make extreme analyses feasible, since a subset of extreme data can be extracted. Other environmental areas such as ecology have seen little extremes work done, and one reason for this is likely a lack of long, sustained data records.

In our applications, we studied extreme precipitation and windspeed/humidity, the extremes of which can be viewed as exhibiting short-lived, 'spike'-like behavior. Although there has been much work describing temporal dependence in extremes, existing EV methods seem best suited for describing short-lived, spike-like phenomena. Prolonged extreme environmental events such as droughts or heat waves remain challenging to model. One can still imagine that with such phenomena, it would continue to be beneficial to focus on and model only the tail, and not allow inference about such phenomena to be contaminated by non-extreme data.

Acknowledgements: D. Cooley is partially supported by NSF grant DMS-1243102. R.L. Smith is partially supported by NSF grants DMS-1127914, DMS-1638521, and DMS-1242957.

Bibliography

Alduchov, O. A. and Eskridge, R. E. (1996). Improved Magnus form approximation of saturation vapor pressure. *Journal of Applied Meteorology*, 35(4):601–609.

Balkema, A. and de Haan, L. (1974). Residual life time at great age. *The Annals of Probability*, 2(5):792–804.

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

Berrocal, V., Gelfand, A., and Holland, D. (2014). Assessing exceedance of ozone standards: a space-time downscaler for fourth highest ozone concentrations. *Environmetrics*, 25(4):279–291.

Boldi, M.-O. and Davison, A. C. (2007). A mixture model for multivariate extremes. *Journal of the Royal Statistical Society, Series B*, 69:217–229.

Chavez-Demoulin, V. and Davison, A. C. (2005). Generalized additive models for sample extremes. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 54(1):207–222.

Chavez-Demoulin, V. and Davison, A. C. (2012). Modelling time series extremes. *REVSTAT-Statistical Journal*, 10:109–133.

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

Coles, S. G., Heffernan, J., and Tawn, J. A. (1999). Dependence measures for extreme value analysis. *Extremes*, 2:339–365.

Coles, S. G. and Tawn, J. A. (1991). Modeling multivariate extreme events. *Journal of the Royal Statistical Society, Series B*, 53:377–92.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In Bertail, P., Doukhan, P., and Soulier, P., editors, *Dependence in Probability and Statistics*, Springer Lecture Notes in Statistics. Springer, New York.

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

Dalrymple, T. (1960). Flood frequency analyses. Water supply paper 1543-a, U.S. Geological Survey, Reston, VA.

Davison, A. C. (2003). *Statistical Models*, volume 11. Cambridge University Press.

Davison, A. C., Huser, R., and Thibaud, E. (2018). Spatial extremes chapter, this volume. In Gelfand, A., editor, *The Handbook of Environmental Statistics*, pages XX–XX. Chapman and Hall. THIS ENTRY WILL NEED TO BE UPDATED TO REFLECT THE FINAL REFERENCE INFORMATION.

Davison, A. C., Padoan, S., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27(2):161–186.

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

Dunn, R. J., Willett, K. M., Thorne, P. W., Woolley, E. V., Durre, I., Dai, A., Parker, D. E., and Vose, R. S. (2012). HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Climate of the Past*, 8:1649–1679.

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., and Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in norway. *Environmetrics*, 26(2):89–106.

Eastoe, E. F. and Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 58(1):25–45.

Einmahl, J. H. and Segers, J. (2009). Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics*, pages 2953–2989.

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

Ferreira, A. and de Haan, L. (2015). On the block maxima method in extreme value theory:PWM estimators. *The Annals of Statistics*, 43(1):276–298.

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

Fix, M., Cooley, D., Sain, S., and Tebaldi, C. (2016). A comparison of US precipitation extremes under RCP8.5 and RCP4.5 with application to pattern scaling. *Climatic Change*. DOI: 10.1007/s10584-016-1656-7.

Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *The Annals of Mathematics*, 44(3):423–453.

Gumbel, E. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statis. Univ. Paris*, 9:171–173.

Heffernan, J. E. and Resnick, S. I. (2007). Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, pages 537–571.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society, Series B*, 66:497–546.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the royal statistical society, Series B*, 52:105–124.

Hosking, J. R. M. and Wallis, J. R. (1997). *Regional Frequency Analysis: An Approach based on L-Moments*. Cambridge, University Press, Cambridge., U.K.

Johndrow, J. and Wolpert, R. (2016). Tail waiting times and the extremes of stochastic processes. *arXiv:1512.07848*.

Kay, J., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., Arblaster, J., Bates, S., Danabasoglu, G., Edwards, J., et al. (2015). The community earth system model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, 96(8):1333–1349.

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

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.

Ledford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., et al. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, 463(7282):747–756.

Naveau, P., Huser, R., Ribereau, P., and Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52:27532769.

Nelsen, R. (2006). *An Introduction to Copulas, 2nd Edition*. Lecture Notes in Statistics No. 139. Springer, New York.

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

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

Resnick, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336.

Resnick, S. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

Rootzen, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12:917–930.

Russell, B. T., Cooley, D., Porter, W. C., Reich, B. J., and Heald, C. L. (2016). Data mining to investigate the meteorological drivers for extreme ground level ozone events. arXiv:1504.08080v4.

Sang, H. and Gelfand, A. E. (2010). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:49–65.

Scarrott, C. and MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT–Statistical Journal*, 10(1):33–60.

Schlather, M. and Tawn, J. (2003). A dependence measure for multivariate and spatial extreme values: Properties and inference. *Biometrika*, 90:139–156.

Sibuya, M. (1960). Bivariate extreme statistics, i. *Ann. Inst. Statist. Math.*, 11:195–210.

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

Smith, R. L., Tawn, J. A., and Coles, S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, 84:249–268.

Stephenson (2018). Climate projections reference, this volume. In Gelfand, A., editor, *The Handbook of Environmental Statistics*, pages XX–XX. Chapman and Hall. THIS ENTRY WILL NEED TO BE UPDATED TO REFLECT THE FINAL REFERENCE INFORMATION.

Stephenson, A. G. (2002). evd: Extreme value distributions. *R News*, 2(2):0.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4):485–498.

Weller, G. and Cooley, D. (2013). A sum characterization of hidden regular variation with likelihood inference via expectation-maximization. *Biometrika*, 101:17–36.