Comparison and Assessment of the Extremes of Different Types of Climate Model Simulations

Mickey Warner and Bruno Sansó*

Abstract

Climate models predict an intensification of extreme weather patterns, thus it is important to assess how similar the extreme behavior in climate model simulations is to that of observations. We fit a Bayesian hierarchical threshold exceedance model to simulations from the climate model CanCM4. Three simulation classes are analyzed and compared—decadal, historical, and pre-industrial control—as well as an observation product. To assess the extremes of the series considered we fit a generalized Pareto model to the exceedances over a threshold. Our method includes a likelihood-based hierarchical model for declustering. We find that in most domains, the distributions of the simulations have a tail behavior in agreement among themselves and with the observations. In order to study the joint tail behavior of simulations and observations, we perform a bivariate extreme value analysis using simple Pareto processes in conjunction with a Bayesian non-parametric model of an angular measure. The results show weak to moderate tail dependence in nearly every comparison made.

1 Introduction

Climate models predict an intensification of extreme activity in weather patterns (Easterling et al., 2000; Tebaldi et al., 2006). Extreme weather has social, ecological, and political impacts, necessitating the need to gauge the differences and similarities in extreme behavior between climate simulations and observations. The fifth phase of the Coupled Model Intercomparison Project (CMIP5) brings together a variety of climate models under an experimental design framework (Taylor et al., 2012; Hibbard et al., 2007). Some of the stated goals of CMIP5 is to examine the variability and predictability of climate models and to explore why some models differ in their responses. The experimental design of CMIP5 is made up of both long-term and short-term simulations, as well as very long simulations of an idealized climate in steady state, that are run as controls. The long-term simulations are initialized from a state randomly chosen from a control run. They are obtained under external forcings that emulate

*University of California Santa Cruz, Department of Applied Mathematics and Statistics

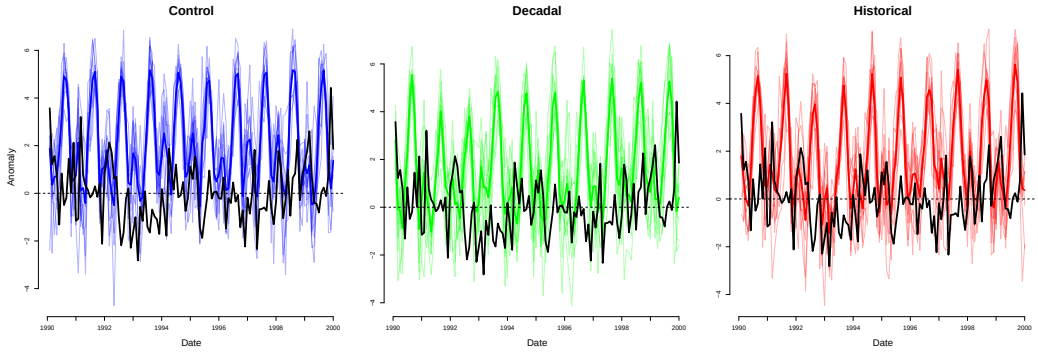

Figure 1: Global mean surface temperature monthly anomalies (w.r.t. observed climatology) from simulations of the climate model CanCM4. Left: Control run. Center: Decadal run. Right: Historical run. The light colored lines correspond to ten replicates. The solid line to the average. The dark line corresponds to the observations.

the historical conditions of the 20th Century. They are then run far into the future, usually the end of the 21st Century, under a number of emission scenarios. Short-term simulations, called decadal, are typically executed for 10- and 30-year forecasts. These decadal runs are initialized with the conditions of the ocean at a specific point in time. They aim at improving the prediction skill of climate models through time-evolving regional climate conditions and external forcings (Meehl et al., 2009). To visualize the difference between these three types of simulations, Figure 1 presents the monthly anomalies of the global mean temperature corresponding to control runs, historical simulations and decadal simulations. By monthly anomalies we mean the differences between value for a given month and the average for that month obtained using the observational record. For our analysis we consider a specific climate model, CanCM4 (described in more detail in section 2). The innovative experimental design of CMIP5 allows for assessing the prediction skill and internal variability of these decadal runs (Kim et al., 2012). Should decadal simulations come to the forefront of climate model predictions, it becomes necessary to look at all predictive aspects of this simulation class. Climate modelers are interested in comparing the properties of decadal simulations to those of control and historical runs. In particular, as discussed in this paper, it is important to assess their extremal behavior.

The study of climate extremes based on climate model simulations has attracted a great deal of attention in the climate community. Some recent work based on the theory of extreme values appears in Fix et al. (2016), where the authors fit a spatial GEV model on grid cells over the contiguous United States, using global mean temperature as a covariate. Their results on the changes over time of the 1% annual exceedance probability (loosely, the 100-year return level) were consistent with global mean temperature increases found by Kharin et al. (2013) and Allen and Ingram (2002). Weller et al. (2012) fit a parametric bivariate extreme model on pseudo-polar coordinates to assess the significance of asymptotic dependence between sea-level pressure fields and precipitation from the pineapple express phenomenon. Weller et al. (2013) model precip-

itation footprints near the Pacific coast with the GPD. Under pseudo-polar coordinates made from joining observations with six NARCCAP models, they found evidence for asymptotic independence for summer precipitation but not for winter which had moderate asymptotic dependence.

The goal of this paper is to address two questions: 1) Does using the same climate model in different ways produce similar extreme behavior? and 2) How well do the simulations agree with the observations? Answers to these questions would inform us on whether different initializations affect the extreme behavior of climate models, and if climate models provide reasonable representation of observed extremes. In tackling these problems we use some of the tools of extreme value analysis that have been become traditional, (see, for example Coles, 2001). But the peculiarities of this application, including the need to consider replicates and to make inference for bivariate extremes, required to introduce a number of novel developments that are illustrated in the paper. These can serve to enhance the toolbox of statistical methods for extreme value analysis.

It can be shown that, after standardization, the block maxima of independent random variables have a distribution that belongs to the generalized extreme value (GEV) family of distributions. Such an approach is at the core of most extreme value analysis, but it naturally requires omitting a large portion of the data. This can be remedied by using a threshold exceedance model, which consists of selecting some large threshold and fitting the exceedances to the generalized Pareto distribution (GPD). In our case study, we consider CanCM4 climate simulations that are run at several different initial conditions, producing multiple replicates. Dealing with replicates is not standard in extreme value analysis, but is very naturally handled with Bayesian hierarchical models. This is the approach considered in this paper, where we consider hierarchical models, not only for the exceedances, but also for estimating the extremal index, a parameter which describes cluster sizes of exceedances in the limit. Such models allow us to infer the extreme behavior of the underlying climate model process, as simulated by the climate model. We then focus on the extreme behavior that is inferred from the observations. A univariate analysis would be sufficient if we desire a marginal comparison only. Nevertheless, to understand how several random variables interact in the extremes, we need a multivariate extremal model. A climate simulation may act similarly to observations in the margin, but their respective extreme values may occur at drastically different times. A multivariate extreme value model allows us explore joint tail dependence. Here we introduce a novel approach for inference on the tails of a bivariate distribution. This uses a simple Pareto processes (Ferreira and de Haan, 2014) which involves the estimation of an angular measure, as considered in Goix et al. (2015). In this paper, this is tackled using Bayesian non-parametric methods that use Bernstein polynomials (Petrone, 1999).

The remainder of the paper is outlined as follows. In section 2 we describe the data in more detail as well as the adjustments required to make the sim-

ulations comparable to the observations. The univariate model, including the hierarchical models for the exceedances and the extremal index, and quantities useful for comparison are given in section 3. Section 4 describes our bivariate approach. Results are presented in section 5 and we conclude with a discussion in section 6.

2 Data

The Fourth Generation Coupled Global Climate Model (CanCM4) has been produced by the Canadian Centre for Climate Modeling and Analysis (CCCma). It consists of an atmospheric component, CanAM4 (von Salzen et al., 2013), and an ocean component, CanOM4. The two components are coupled daily to produce climate predictions of a number of variables on a roughly 2.5° degree grid over the globe (see Merryfield et al. (2013)). Two variables will be analyzed: precipitation (labeled `pr`, in meters) and maximum temperature (labeled `tasmax`, in degrees Kelvin). We further restrict our attention to analyzing two seasons—summer and winter—and two regions—California and the continental U.S.

For decadal simulations we focus on two decades in this analysis: 1962–1971 and 1990–1999, which are initialized on the state of the ocean in 1961 and 1989, respectively. Historical simulations are obtained for the years 1961–2005. The pre-industrial control, or simply control, simulations correspond to climate conditions comparable to those preceding the industrial revolution, and are run over a thousand years. The purpose of the control runs is to provide some measure of internal variability of the climate system. Decadal and historical simulations are run at $R = 10$ different initial conditions. To obtain $R = 10$ "replicates" for the control simulations, we randomly select ten non-overlapping 10-year periods. An observation product is obtained from Maurer et al. (2002). The observations are based on daily measurements from weather stations throughout the United States and are interpolated onto a fine grid (about 1/8° degree spacing, middle plot of Figure 2).

As noted above, the climate models are run on a different grid than the observational data set. The two grids are shown in Figure 2 for the area over California. In order to make the simulations comparable to the observations, we adjust the data in the following manner. We take the climate model grid cell locations and create non-overlapping cells, or rectangles, such that each location is roughly in the center of the cell (left plot of Figure 2). Then we count the number of locations from the observation product that are contained within each cell. The number of locations within the cells are used to weight the climate simulations (right plot of Figure 2). For the simulations, we take a weighted sum for precipitation and a weighted average for temperature. No weighting is used for the observations. Instead, a straight sum or average of all locations within our region of interest (either California or U.S.) is used. This method places the simulations and the observations on the same scale and yields daily time-series.

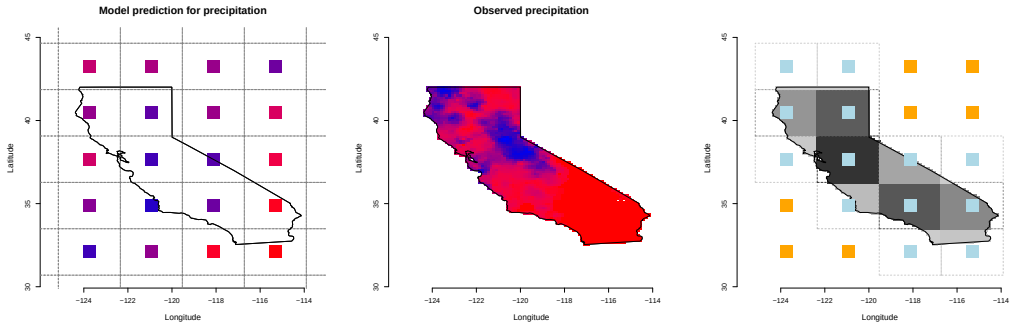

Figure 2: Left: CanCM4 simulation grid cells. Center: Observation locations. Right: method for computing weighted sum or average for CanCM4 to make values comparable with observations; the lighter gray points mean less weight is applied to the climate simulations and the darker gray means more weight. The data shown (left and center) are precipitations from a single day in January, ranging from low (red) to high (blue) precipitation.

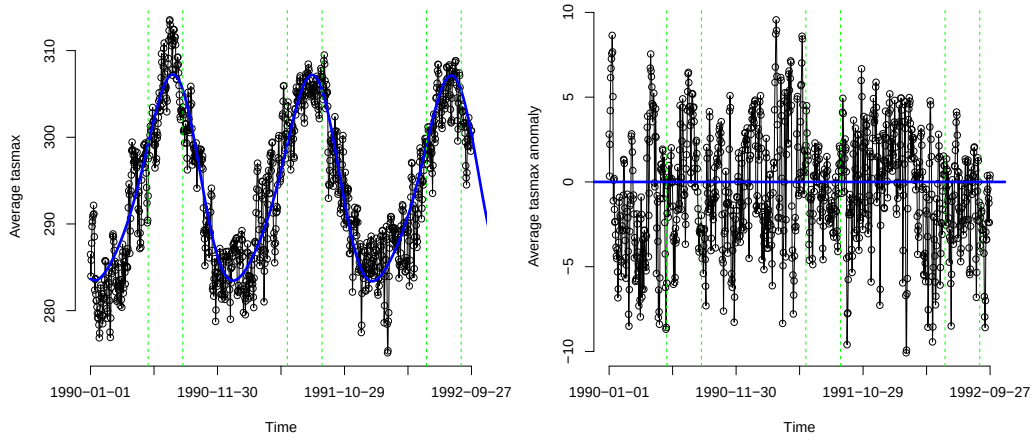

Figure 3: One of the DLMs used to calculate the anomalies. Shown is one of the decadal replicates of average `tasmax` in California for about the first two and one-half years of the time-series. The green dashed lines mark the beginning and the end of the summer months.

2.1 De-trending

Climate data are often non-stationary series characterized by complicated trends and cycles. Since we are interested in the behavior of the extremes, each time-series is "de-trended" prior to parameter estimation. This is accomplished through the use of dynamic linear models (DLMs) (Prado and West, 2010). Consider a univariate time series $y_1, \ldots, y_T$. We assume that

$$\begin{aligned} y_t &= \mathbf{F}_t^\top \boldsymbol{\theta}_t + \nu_t, & \nu_t &\sim N(0, v_t) \\ \boldsymbol{\theta}_t &= \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t & \mathbf{w}_t &\sim N(\mathbf{0}, \mathbf{W}_t) \end{aligned} \tag{1}$$

where $\boldsymbol{\theta}_t$ is the length-p state vector, $\mathbf{F}_t$ is a length p vector of known constants are regressors, ν_t is observation noise, $\mathbf{G}_t$ is the known $p \times p$ state evolution matrix, and $\mathbf{w}_t$ is the state evolution noise. Note that ν_t and $\mathbf{w}_t$ are independent

in t, and mutually independent.

From Model (1) we obtain a smooth and flexible mean across time. After conditioning on the data up to time T, we extrapolate back over time to obtain the posterior distributions $p(\boldsymbol{\theta}_t|D_T)$ for all $t < T$, which have mean $\mathbf{a}_t$. Using these distributions, and given $\mathbf{F}_t$, the mean of y_t is simply $\mathbf{F}_t^\top \mathbf{a}_t$ (see Prado and West, 2010, for the algorithmic details). We construct $\mathbf{F}_t$ and $\mathbf{G}_t$ such that the evolution of $\boldsymbol{\theta}_t$ has annual and semi-annual cycles, i.e. the first and second harmonics. Higher harmonics did not seem to make significant contributions in modeling the time-series. A discount factor of $\delta = 0.9999$ was chosen, signifying low systematic variance. We assume the prior for v_t is an inverse gamma having sensible shape and scale parameters.

An example of fitting the model to one of the time series from the decadal simulations is shown in Figure 3. The blue line in the left plot is the mean of y_t, $\mathbf{F}_t^\top \mathbf{a}_t$, conditional on the whole time series. The interior of the vertical green lines mark the summer months. The right plot is the result of subtracting the observation y_t with the mean from the DLM, which produces a roughly stationary sequence. Thus, in our extreme value analysis we fit our model to the residuals, or anomalies. For each time-series to be analyzed, we fit a DLM having the characteristics described above to obtain the anomalies. When working within a specific season, either winter (December, January, February) or summer (June, July, August), we concatenate across years to form a single time series of seasonal anomalies. So, for the winter data, the anomaly corresponding to 28 February is followed immediately by that of 1 December of the following year.

3 Univariate analysis

Under some mild assumptions, for random variable X and for large enough u, the distribution of $X - u$ (the exceedance), conditional on $X > u$ is approximately

$$P(X - u \leq y | X > u) \approx H(y) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi} \tag{2}$$

defined on $\{y : y > 0 \text{ and } (1 + \xi y/\sigma) > 0\}$. $H(y)$ is the distribution function for a generalized Pareto random variable with shape parameter $\xi \in \mathbb{R}$ and scale $\sigma > 0$. Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables and u be a high threshold. Define $Y_i = X_i - u$ for $X_i > u$ be the k exceedances. By taking derivatives in Equation (2) we obtain a likelihood for ξ and σ, based on the k exceedances.

In order to consider the replicates of the climate model simulations we extend the above model in a hierarchical fashion. Suppose we have R replicates or computer simulations, each with n_i observations, for $i = 1, \ldots, R$. Let X_{ij} denote the jth observation in replicate i. We assume that $X_{ij} \sim H_i$, for $i = 1, \ldots, R, j = 1, \ldots, n_i$, and that all X_{ij} are mutually conditionally

independent. For a fixed u and each i, define the following sets:

$$A_i = \{j : x_{ij} \leq u\}, \quad A_i^c = \{j : x_{ij} > u\}$$

where $|A_i| = n_i - k_i$ and $|A_i^c| = k_i$ with k_i being the number of exceedances in replicate i. We define our exceedances as $y_{ij} = (x_{ij} - u) \cdot \mathbb{1}_{(j \in A_i^c)}$, so that all observations not exceeding u are marked as 0. Let $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,n_i})^\top$ and $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_R^\top)^\top$.

The likelihood is given by

$$\begin{aligned} L(\mathbf{y}; \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\zeta}) &= \prod_{i=1}^{R} h_{Y_i}(\mathbf{y}_i | \sigma_i, \xi_i, \zeta_i) \\ &\approx \prod_{i=1}^{R} \left[(1-\zeta_i)^{n_i - k_i} \zeta_i^{k_i} \prod_{j \in A_i^c} \frac{1}{\sigma_i} \left(1 + \xi_i \frac{y_{ij}}{\sigma_i}\right)_+^{-1/\xi_i - 1} \right]. \end{aligned} \tag{3}$$

Note that ξ_i, σ_i are informed only by those observations which exceed u. The parameter $\zeta_i = P(X_{ij} > u)$, which is necessary for calculating return levels (section 3.2), depends only on the number of exceedances. We complete the hierarchical model formulation by specifying the priors: $\xi_i | \xi, \tau^2 \sim Normal(\xi, \tau^2), \sigma_i | \alpha, \beta \sim Gamma(\alpha, \beta)$ and $\zeta_i | \zeta, \eta \sim Beta(\zeta\eta, (1-\zeta)\eta)$. For the hyperparameters that are used to specify the priors we have that $\xi \sim Normal(m, s^2), \tau^2 \sim InvGamma(a_\tau, b_\tau)$, α, β and η follow gamma distributions with parameters $a_\alpha, b_\alpha, a_\beta, b_\beta$ and a_η, b_η respectively, and $\zeta \sim Beta(a_\zeta, b_\zeta)$. The inverse gamma is parametrized to have mean $b/(a-1)$ and the gammas have mean a/b. These prior distributions are combined with the likelihood to obtain a joint posterior distributions that is explored using Markov chain Monte Carlo methods.

3.1 A hierarchical model to estimate the extremal index

The threshold exceedance model described above relies on an assumption of independence between observations, which is unrealistic for a time-series. When there is dependence between the random variables, the extremes are related according to the so-called extremal index (Leadbetter, 1983), denoted by $\theta \in [0, 1]$, which arises in the following way, as summarized in Ferro and Segers (2003). For $\{X_n\}_{n \geq 1}$ a strictly stationary sequence of random variables with marginal distribution F, the sequence has extremal index θ if for each $\tau > 0$ there is a sequence $\{u_n\}_{n \geq 1}$ such that, $\lim_{n \to \infty} n(1 - F(u_n)) \to \tau$ and $\lim_{n \to \infty} P(\max(X_1, \ldots, X_n) \leq u_n) \to \exp(-\theta\tau)$. The extremal index describes the behavior of exceedances in the limit and can be loosely interpreted as $\theta = (\text{limiting mean cluster size})^{-1}$. As an example, suppose $\theta = 0.5$, then we would expect exceedances of a large threshold to occur in pairs; for $\theta = 0.33$, in groups of 3.

Ferro and Segers (2003) show that the extremal index arises in the limiting distribution of the times between exceedances of a threshold. If T_θ is the

random variable for interexceedance times in the limit, then T_θ is distributed according to the mixture $(1-\theta)\epsilon_0 + \theta\mu_\theta$ where ϵ_0 is the degenerate probability distribution at 0 and μ_θ is an exponential distribution with mean θ^{-1}. This means that the role of θ is two-fold: it is both the proportion of non-zero interexceedance times and the inverse mean of non-zero interexceedance times. This poses a challenge when estimating θ since is it impossible to observe an interexceedance time of zero in practice. We next describe a hierarchical model used to estimate θ. This is an extension the likelihood-based threshold exceedance model in Ferro and Segers (2003) and is used to get an estimate for θ, which is used to decluster the exceedances and to calculate return levels.

Suppose we have observations $X_1, \ldots, X_n$. For a threshold u, the N exceedances $Y_i = X_i - u$ given $X_i > u$ occur at times $1 \leq j_1 < \cdots < j_N \leq n$. The observed interexceedance times are given by $T_i = j_{i+1} - j_i$ for $i = 1, \ldots, N-1$. Ferro and Segers (2003) provide the following log-likelihood

$$\begin{aligned} l(\theta, p; \mathbf{T}) = m_1 \log(1 - \theta p^\theta) + (N - 1 - m_1)\{\log(\theta) + \log(1 - p^\theta)\} \\ + \theta \log(p) \sum_{i=1}^{N-1} (T_i - 1) \,, \end{aligned} \tag{4}$$

where p is the probability of not exceeding the threshold. Süveges (2007) proposed an alternative likelihood for estimating the extremal index which dealt with some of the issues noted in Ferro and Segers (2003). This likelihood was extended in Süveges and Davison (2010). Though there are advantages to the alternative likelihood, we prefer to use the one given in (4). In fact, in a simulation study, both likelihoods performed very similarly, with some preference to model (4), in the context of the hierarchical setting that we describe below.

In our application, we have R replicates from a climate model with values from replicate i denoted $X_{i,1}, \ldots, X_{i,n}$. If we assume these simulations are independent from each other, then we expect there to be R unique extremal indices $\theta_1, \ldots, \theta_R$. By assuming conditional independence between the simulations, we can construct the log-likelihood $L = \sum_{i=1}^{R} l(\theta_i, p_i; \mathbf{T}^{(i)})$ where $\mathbf{T}^{(i)}$ is the vector of interexceedance times for replicate i having length N_i. However, since all replicates come from the same climate model, we may wish to assume that the θ_i come from a common distribution. $\theta_i \overset{iid}{\sim} Beta\left(\theta\nu, (1-\theta)\nu\right)$, having mean $\theta\nu/(\theta\nu + (1-\theta)\nu) = \theta$. Under model (4), we place a similar prior on the p_i, $p_i \overset{iid}{\sim} Beta\left(p\tau, (1-p)\tau\right)$. The model is completed by assuming that θ and p follow a beta distributions with parameters a_θ, b_θ and a_p, b_p respectively, and ν and τ follow gamma distributions with parameters a_ν, b_ν and a_τ, b_τ respectively.

Using the inference on θ_i, we perform declustering as in Ferro and Segers (2003). Each replicate is declustered separately. Let $\hat{\theta}_i$ be the posterior mean of the extremal index of each replicate. Calculate $C_i = \lfloor \hat{\theta}_i N_i \rfloor + 1$, the estimated number of independent clusters. Let T_{C_i} be the C_ith largest interexceedance time in $\mathbf{T}^{(i)}$. In the case of ties, decrement C_i by one until T_{C_i+1} is strictly

greater than T_{C_i}. Clusters are formed by grouping the exceedances that are separated in time by no more than T_{C_i}. In other words, two exceedances are in the same cluster if their interexceedance time is less than or equal to T_{C_i}. The C_i clusters that will be formed using the above scheme are assumed to be independent. For each cluster we compute the cluster maximum, this being the ultimate quantity used in our inference.

3.2 Return levels

A most useful quantity in an extreme value analysis is the return level. Generally, for a distribution G, the return level x_m is the solution to $G(x_m) = 1-1/m$ which is interpreted as the quantity that is exceeded on average once every m observations. For our model the return level is

$$x_m = u + \frac{\sigma}{\xi}\left[(m\zeta\theta)^{\xi} - 1\right] \tag{5}$$

where the terms ζ and θ account for the probability of exceeding u and being within a cluster, respectively.

4 Bivariate analysis

The univariate analysis described in section 3 may reveal comparable extremal behavior among some of the simulations and observations, but it is insufficient to describe any tail dependence. For this, we work under the framework of multivariate extremes. Univariate extreme value analysis can be generalized to a multivariate setting, wherein the limiting model for joint maxima is obtained. The model is comparable to the GEV distribution in the univariate case and further allows the modeling of tail dependence. Chapter 8 of Coles (2001) provides an introduction to multivariate extremes, specifically in the bivariate case. Some possible families of distributions for modeling bivariate extremes are offered in Coles and Tawn (1991).

Weller et al. (2012) perform a bivariate extreme value analysis in two stages. The first, which we also do, is to fit each marginal with the GPD. The second stage involves transforming the marginal data to standard Frechet random variables using the GPD parameter estimates. A bivariate vector is formed and converted into pseudo-polar coordinates. From here, they perform a hypothesis test on the presence of asymptotic dependence and then fit parametric models to the angles. Our approach differs in that we use a nonparametric model on the angles obtained from a simple Pareto process.

For exceedance over a threshold in multivariate settings, the work of Rootzen and Tajvidi (2006) defines the multivariate generalized Pareto distribution. Further analysis of these classes of distributions is presented in Falk and Guillou (2008). In this context, Michel (2008) provides a detailed discussion of different inferential approaches. In this paper we propose a novel approach that is based on a constructive definition of a Pareto process presented in

(Ferreira and de Haan, 2014) that decomposes the vector of interest into a length component and an angular component.

The primary result we use is Theorem 3.2 from Ferreira and de Haan (2014). Let $C(S)$ be the space of continuous real functions on S, equipped with the supremum norm, where S is a compact subset of $\mathbb{R}^d$. Let X be from $C(S)$. Then the conditions of their Theorem 3.1 imply

$$\lim_{t\to\infty} P\left(T_t X(s) \in A \middle| \sup_{s\in S} T_t X(s) > 1\right) = P(W \in A) \tag{6}$$

where W is a simple Pareto process (see Appendix A.2), and

$$T_t X(s) = \left(1 + \xi(s)\frac{X(s) - u_t(s)}{\sigma_t(s)}\right)_+^{1/\xi(s)} \tag{7}$$

is the standardized version of $X(s)$, with u_t, σ_t and ξ are continuous functions in S, and $\sigma_t(s) > 0, \forall s$. The condition that $\sup_{s\in S} T_t X(s) > 1$ is equivalent to $\sup_{s\in S} X(s) > u_t(s)$. Thus, $u_t(s)$ can be seen as a sequence of thresholds.

We assume t is large enough that the theorem applies (implying $u = u_t$ and $\sigma = \sigma_t$). In this paper we are interested in the bivariate case, thus we can think of S as a set containing two elements only, s_1 and s_2. According to Equation (6) the distribution of a bivariate random vector, after the transformation in (7), and given that the maximum of the transformed vector is above a high enough threshold, is equal to that of a simple Pareto vector, say W. Using the definition of a Pareto process in Appendix A.2, we have $W = YV = Y(V_1, V_2)$, Y is a Pareto random variable and $V_1 \vee V_2 = \max(V_1, V_2) = 1$. Thus, as the points (V_1, V_2) fall along the curve of the non-negative unit sphere with supremum norm $\{(v_1, v_2) : ||(v_1, v_2)||_\infty = 1, v_1 \geq 0, v_2 \geq 0\}$, and each of them is associated with a unique direction, Y determines the length of the vector W and V its direction. Thus we obtain an alternative representation of V specifying (V_1, V_2) in terms of a scaled angle

$$\phi = \frac{2}{\pi}\arctan\left(\frac{V_2}{V_1}\right) \in [0, 1]. \tag{8}$$

Our modeling approach consists of building a very flexible model for V. This is achieved by modeling the distribution of ϕ using a Bernstein-Dirichlet prior (BDP) (Petrone, 1999). More specifically, we model the density of ϕ as

$$p(\phi) = \sum_{m=1}^{M} \omega_m \, be(\phi \mid m, M - m + 1), \quad \phi \in [0, 1], \tag{9}$$

where $be(\cdot \mid a, b)$ denotes the density of a Beta distribution with shape parameters a and b (such that the mean is $a/(a + b)$). The mixture weights are defined through increments of a cumulative distribution function F with support on $[0, 1]$, such that $\omega_m = F(m/M) - F((m - 1)/M)$, for $m = 1, \ldots, M$. If F can take flexible shapes, then the mixture weights can adapt to selecting the appropriate beta basis densities to achieve general NHPP density, and

thus intensity, shapes. This motivates assigning a nonparametric prior to F, such as a Dirichlet process prior (Ferguson, 1973, 1974) with a Beta centering distribution.

4.1 Asymptotic tail dependence

A measure that is used to characterize the strength of the dependence, in the tail, for two random variables X and Y that share a common marginal distribution, is given by

$$\chi = \lim_{z \to z^*} P(X > z | Y > z),$$

where z^* is the (possibly infinite) right end-point (Coles, 2001). χ informs us of the distribution of extremes for the variable X given that Y is very large. When $\chi > 0$, X and Y are said to be asymptotically dependent, otherwise they are asymptotically independent.

For a simple Pareto process the distribution function for $W_i \equiv W(s_i)$ is $F_{W_i}(w) = 1 - E(V_i)/w$, where $V_i \equiv V(s_i)$ and $w > 1$. Using this fact, we can standardize W_i to be uniform and compute χ as

$$\chi = \lim_{q \to 1} P(F_{W_1}(W_1) > q | F_{W_2}(W_2) > q) = E\left(\frac{V_1}{E(V_1)} \wedge \frac{V_2}{E(V_2)} \right), \qquad (10)$$

where $a \wedge b = \min(a, b)$. Notice that the tail dependence in a Pareto random vector depends only on the distribution of the angular component. Thus, by considering a flexible enough model for ϕ we are able to capture the whole range of tail dependence. Asymptotic independence, though, can not be achieved with a model based on Pareto processes.

5 Results

Our analysis considers each of the four data sources (i.e. the three climate simulation classes and the observation produce). There are four factors with two levels each. The factors, with their levels, are: (1) Variable – precipitation or maximum temperature; (2) Season – winter or summer; (3) Decade – 1962–1971 or 1990–1999; and (4) Region – California or U.S.A. For each of the 16 combinations, we de-trended and de-clustered the data. Then we used the hierarchical model described in Section 3 for the decadal, historical, and control runs, and a univariate threshold model for observational product.

Thresholds are chosen to be the 0.95 quantile for the climate simulations and 0.85 for the observations. These quantiles can be justified in part by looking at mean residual life plots (not shown), see section 4.3.1 of Coles (2001). Such plots indicate that the generalized Pareto approximation (2) is valid for exceedances of the selected thresholds. We also have to consider the sample size of the exceedances, and these quantiles give us enough data to accurately fit the models. The chosen thresholds are given in Tables 1 and 2.

Factor combination			Control	Decadal	Historical	Observed
1962	CA	Summer	1.04	2.42	2.44	0.76
1962	CA	Winter	18.08	19.43	22.22	11.61
1962	USA	Summer	53.83	61.98	55.32	42.65
1962	USA	Winter	141.88	127.03	130.85	69.13
1990	CA	Summer	1.04	2.51	2.42	0.49
1990	CA	Winter	18.08	17.75	17.61	15.49
1990	USA	Summer	53.83	62.73	59.73	42.33
1990	USA	Winter	141.88	144.45	140.18	69.42

Table 1: Precipitation thresholds for climate simulations and observations under each scenario.

Factor combination			Control	Decadal	Historical	Observed
1962	CA	Summer	4.00	3.53	3.72	2.49
1962	CA	Winter	5.52	5.19	5.47	3.26
1962	USA	Summer	2.02	2.22	2.17	1.51
1962	USA	Winter	3.93	3.57	3.69	3.23
1990	CA	Summer	4.00	3.87	3.91	2.94
1990	CA	Winter	5.52	5.56	5.29	3.31
1990	USA	Summer	2.02	2.17	2.26	1.46
1990	USA	Winter	3.93	3.57	3.65	3.66

Table 2: Temperature thresholds for climate simulations and observations under each scenario.

Factor combination				Observed	Historical	Probability
1962	CA	Summer	Pr	2.06	2.44	0.213
1962	CA	Winter	Pr	22.69	22.22	0.347
1962	USA	Summer	Pr	51.16	55.32	0.307
1962	USA	Winter	Pr	84.62	130.85	0.482
1990	CA	Summer	Pr	2.22	2.42	0.244
1990	CA	Winter	Pr	21.04	17.61	0.371
1990	USA	Summer	Pr	46.06	59.73	0.577
1990	USA	Winter	Pr	105.54	140.18	0.318
1962	CA	Summer	Temp	3.26	3.72	0.281
1962	CA	Winter	Temp	4.76	5.47	0.273
1962	USA	Summer	Temp	1.87	2.17	0.281
1962	USA	Winter	Temp	4.13	3.69	0.270
1990	CA	Summer	Temp	3.97	3.91	0.307
1990	CA	Winter	Temp	4.27	5.29	0.340
1990	USA	Summer	Temp	1.93	2.26	0.304
1990	USA	Winter	Temp	4.68	3.65	0.264

Table 3: The probability that an observational anomaly exceeds the given value, conditioned on the historical anomaly exceeding its threshold. The values are computed using the second equation in A.3 with $q = 1 - E(V_2)$.

The posterior distributions for ξ and σ from the climate simulations showed, with few exceptions, considerable overlap with those of the observations. Except for summer precipitation in California, posterior shapes for each data source were centered around zero with slightly more mass below zero. This suggests only weak evidence for anomaly extremes having an upper bound.

The extremal indexes, which measure limiting mean cluster size in the exceedances, under maximum temperature were all very similar. For each data source and every factor combination (excluding precipitation), we estimate extremal indexes to be around 0.2 to 0.4. This suggests that particularly hot days come in groups of around 2 to 5 days, regardless of season, location, and time. The story with precipitation is a little different. In this case, we typically had a larger extremal index for the simulations (around 0.6 to 0.8) than for the observations (0.4 to 0.6). Hence, the extremal index for observations would suggest that especially rainy days come in pairs, but simulations would have an average cluster size of about 1.5. An exception here was observed for California summer during 1962–1971 which had an agreement between simulations and observations and showed limiting mean cluster sizes of about 2.

The comparison between simulations and observations can be characterized well with the 20-year return levels shown in Figure 4. Despite using different quantiles for thresholds (and having different posterior distributions), we can still observe similarities in the return levels. These return levels suggest that for any time period, but not on any given day, we expect to see similar extreme anomalies.

We use the simple Pareto process of section 4 to compare each type of climate model simulation to the observations. Specifically, for each replicate in the simulations and its corresponding observational record we perform the transformation (7). From the two resulting time series we compute daily scaled angles via (8). The `DPPackage` in `R` is employed to fit a BDP model to these angles. For the BDP we use a uniform centering distribution with prior precision as a gamma random variable having mean and standard deviation 10. We set the maximum number of beta mixtures to be $M_{max} = 300$. The BDP from the package returns a density estimate of the angles along a fine grid on the unit interval. Finally, a bootstrap sample of the grid using the normalized density estimate produces our posterior distribution of the angles. It is from this posterior distribution in which we compute χ with (10) as a measure of asymptotic dependence.

The above process is utilized for each of the three climate simulation classes under all 16 factor combinations. Along with the $R = 10$ χ's computed for each replicate, an "overall" χ was computed by pooling together the angles from each replicate under a particular scenario. In total, a BDP model was fit $(10 + 1) \times 3 \times 16 = 528$ times. The results are shown in Figure 5. The larger X marks the "overall' measure and the smaller dots mark each replicate. Most estimates for asymptotic dependence are in the region $[0.2, 0.4]$ indicating weak to moderate dependence. In only two instances do we have χ close to 0.5. And in these cases the comparison is made with the control runs, which should be

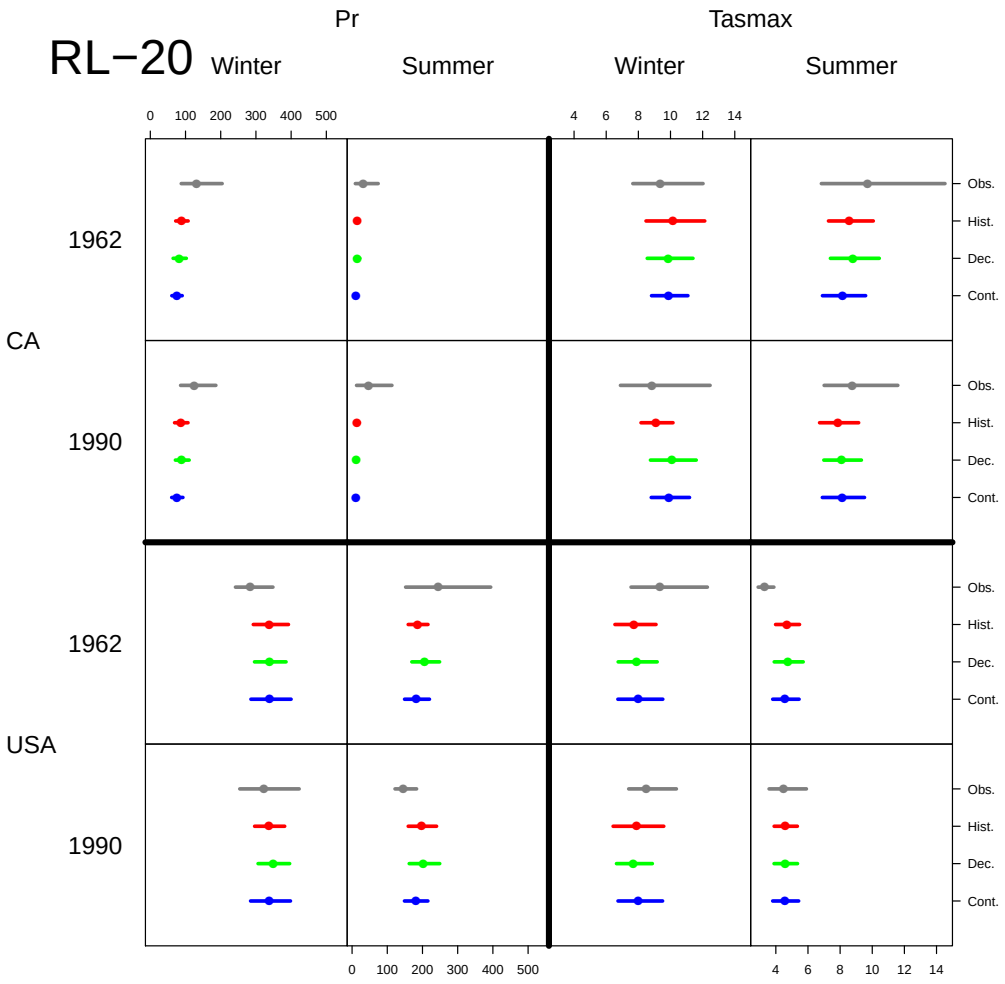

Figure 4: 20-year return levels from the mean process under each of the 16 factor combinations.

some cause of concern since the times for the control runs do not have the same meaning as those from the observations. In every other case we see relatively the same strength of dependence from the three climate simulations, though the dependence is weak.

We quantify the relationship between historical anomalies and observational anomalies in Table 3. The values in that table are calculated from simple Pareto processes (see A.3). Each row provides the probability that an observational anomaly exceeds a certain value given that the historical anomaly exceeds its threshold. Note the value under the observed column is not the threshold for the observations. We observe that in all but one of the cases the probability is less than 50%. The exception being summer precipitation in the USA in the 90's.

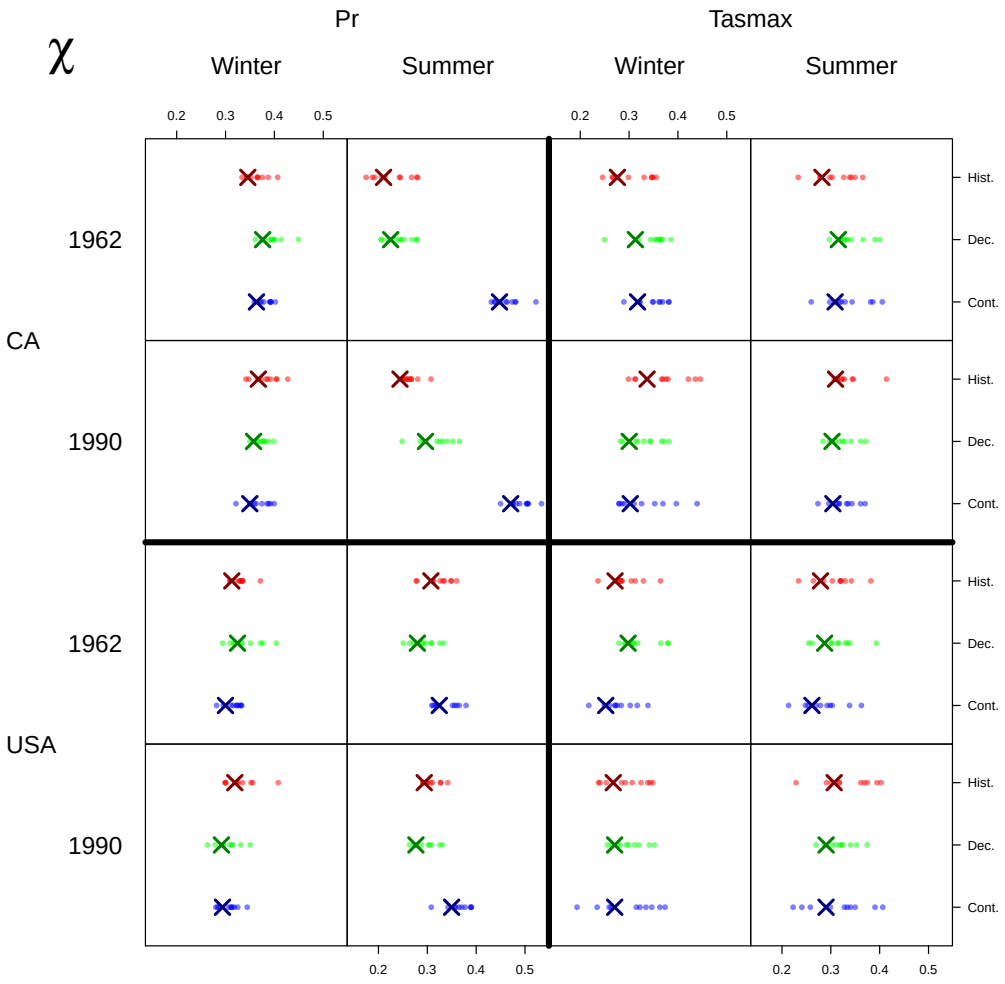

Figure 5: Measure of asymptotic dependence, calculated using results from the simple Pareto process. The dots mark the χ for each replicate paired against the observations. The X marks the value of χ when all replicates are taken together.

6 Conclusions

We have proposed a hierarchical threshold model to study the extreme behavior of environmental variables. The model was applied to a variety of factor combinations involving different runs for a climate model, and compared to a univariate threshold model for an observational record. Our application is peculiar, due to the existence of replicates, which is an uncommon feature in etreme value analysis. Our likelihood based approach is able to handle this peculiarity using a hierarchical structure. The key climatological question in this application is whether different types of climate model simulations produce simulations with comparable tail behavior. The evidence from the current case study shows that, indeed, the tails of the simulations obtained from a control run, a historical run and a decadal simulation behave similarly. In addition, when compared to the tails of an observational record, we did not detect substantial differences. Our univariate analyses for each data source reveal similarities in the tail behavior and return levels. However, when we consider the asymptotic tail distribution, jointly between observations and simulations, we find that the dependence is relatively weak. This implies that the probability of large precipitation or temperature, conditional on high simulated values, is not very high. This undermines the predictive ability of climate model extremes.

7 Acknowledgments

This research was partially funded by the National Science Foundation grant DMS-1513076.

A Appendix

A.1 Likelihood for hierarchical model

$$
\begin{aligned}
L(\mathbf{y};\boldsymbol{\sigma},\boldsymbol{\xi},\boldsymbol{\zeta}) &= \prod_{i=1}^{R} f_{Y_i}(\mathbf{y}_i|\sigma_i,\xi_i,\zeta_i) \\
&= \prod_{i=1}^{R}\left[\prod_{j\in A_i} F_{X_i}(u) \times \prod_{j\in A_i^c} f_{X_i}(y_{ij}+u)\right] \\
&\approx \prod_{i=1}^{R}\left[\prod_{j\in A_i} F_{X_i}(u) \times \prod_{j\in A_i^c} [1-F_{X_i}(u)]h(y_{ij}|\sigma_i,\xi_i)\right] \qquad \text{(approximation (2))} \\
&= \prod_{i=1}^{R}\left[\prod_{j\in A_i} (1-\zeta_i) \times \prod_{j\in A_i^c} \frac{\zeta_i}{\sigma_i}\left(1+\xi_i\frac{y_{ij}}{\sigma_i}\right)_+^{-1/\xi_i-1}\right] \qquad (\zeta_i = 1-F_{X_i}(u)) \\
&= \prod_{i=1}^{R}\left[(1-\zeta_i)^{n_i-k_i}\zeta_i^{k_i} \prod_{j\in A_i^c} \frac{1}{\sigma_i}\left(1+\xi_i\frac{y_{ij}}{\sigma_i}\right)_+^{-1/\xi_i-1}\right]
\end{aligned}
$$

A.2 Definition of a simple Pareto process

The constructive definition of a simple Pareto process is as follows (from Theorem 2.1 of Ferreira and de Haan (2014)):

Let $C^+(S)$ be the space of non-negative real continuous functions on S, with S some compact subset of $\mathbb{R}^d$. Let W be a stochastic process in $C^+(S)$ and ω_0 a positive constant. When $W(s) = YV(s)$ for all $s \in S$, for some Y and $V = \{V(s)\}_{s\in S}$, satisfies:

(a) $V \in C^+(S)$ is a stochastic process satisfying $\sup_{s\in S} V(s) = \omega_0$ almost surely, and $E[V(s)] > 0$ for all $s \in S$.

(b) Y is a standard Pareto random variable, $P(Y \le y) = 1 - 1/y$, $y > 1$,

(c) Y and V are independent.

then W is called a simple Pareto process.

A.3 Measure of asymptotic dependence for simple Pareto process

To obtain χ we need $\lim_{q\uparrow 1} Pr(F_{W_1}(W_1) > q | F_{W_2}(W_2) > q)$. We observe that $Pr(W_i > w) = E(1 \wedge V_i/w) = E(V_i/w)$, for a large enough w, as $0 < V_i \le 1$. Thus, for q close to 1, $F_{W_i}^{-1}(q) = EV_i/(1-q)$, $i = 1, 2$. We then have that

χ is given as a ratio where the numerator is $Pr(W_1 > w_1, W_2 > w_2)$, where $w_i = EV_i/(1-q)$. This is equal to

$$Pr\left(Y > \frac{w_1}{V_1} \vee \frac{w_2}{V_2}\right) = E\left(1 \wedge \left(\frac{w_1}{V_1} \vee \frac{w_2}{V_2}\right)^{-1}\right) = (1-q)E\left(\frac{V_1}{EV_1} \wedge \frac{V_2}{EV_2}\right),$$

for large enough q. To complete the proof we observe that the denominator in χ is given by $Pr(F_{W_2}(W_2) > q)$ which is equal to $1-q$.

Transforming from the Pareto process back to the original variables, we can obtain

$$Pr\left(X_1 > u_1 + \frac{\sigma_1}{\xi_1}\left(w_1^{\xi_1} - 1\right) \middle| X_2 > u_2 + \frac{\sigma_2}{\xi_2}\left(w_2^{\xi_2} - 1\right)\right)$$

which can be used to describe bivariate extreme behavior in concrete terms.

References

Allen, M. R. and Ingram, W. J. (2002), "Constraints on future changes in climate and the hydrologic cycle," *Nature*, 419, 224.

Coles, S. (2001), *An introduction to statistical modeling of extreme values*, vol. 208, Springer.

Coles, S. G. and Tawn, J. A. (1991), "Modelling extreme multivariate events," *Journal of the Royal Statistical Society. Series B (Methodological)*, 377–392.

Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., and Mearns, L. O. (2000), "Climate extremes: observations, modeling, and impacts," *science*, 289, 2068–2074.

Falk, M. and Guillou, A. (2008), "Peaks-over-threshold stability of multivariate generalized Pareto distributions," *Journal of Multivariate Analysis*, 99, 715–734.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

— (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629.

Ferreira, A. and de Haan, L. (2014), "The generalized Pareto process; with a view towards application and simulation," *Bernoulli*, 20, 1717–1737.

Ferro, C. A. and Segers, J. (2003), "Inference for clusters of extreme values," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65, 545–556.

Fix, M. J., Cooley, D., Sain, S. R., and Tebaldi, C. (2016), "A comparison of US precipitation extremes under RCP8. 5 and RCP4. 5 with an application of pattern scaling," *Climatic Change*, 1–13.

Goix, N., Sabourin, A., and Clémençon, S. (2015), "Sparsity in Multivariate Extremes with Applications to Anomaly Detection," *arXiv preprint arXiv:1507.05899*.

Hibbard, K. A., Meehl, G. A., Cox, P. M., and Friedlingstein, P. (2007), "A strategy for climate change stabilization experiments," *Eos, Transactions American Geophysical Union*, 88, 217–221.

Kharin, V. V., Zwiers, F., Zhang, X., and Wehner, M. (2013), "Changes in temperature and precipitation extremes in the CMIP5 ensemble," *Climatic change*, 119, 345–357.

Kim, H.-M., Webster, P. J., and Curry, J. A. (2012), "Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts," *Geophysical Research Letters*, 39.

Leadbetter, M. R. (1983), "Extremes and local dependence in stationary sequences," *Probability Theory and Related Fields*, 65, 291–306.

Maurer, E., Wood, A., Adam, J., Lettenmaier, D., and Nijssen, B. (2002), "A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States," *Journal of climate*, 15, 3237–3251.

Meehl, G. A., Goddard, L., Murphy, J., Stouffer, R. J., Boer, G., Danabasoglu, G., Dixon, K., Giorgetta, M. A., Greene, A. M., Hawkins, E., et al. (2009), "Decadal prediction: can it be skillful?" *Bulletin of the American Meteorological Society*, 90, 1467–1485.

Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., Ajayamohan, R., Fyfe, J. C., Tang, Y., and Polavarapu, S. (2013), "The Canadian seasonal to interannual prediction system. Part I: Models and initialization," *Monthly weather review*, 141, 2910–2945.

Michel, R. (2008), "Some notes on multivariate generalized Pareto distributions," *Journal of Multivariate Analysis*, 99, 1288–1301.

Petrone, S. (1999), "Bayesian density estimation using Bernstein polynomials," *Canadian Journal of Statistics*, 27, 105–126.

Prado, R. and West, M. (2010), *Time series: modeling, computation, and inference*, CRC Press.

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

Süveges, M. (2007), "Likelihood estimation of the extremal index," *Extremes*, 10, 41–55.

Süveges, M. and Davison, A. C. (2010), "Model misspecification in peaks over threshold analysis," *The Annals of Applied Statistics*, 4, 203–221.

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, 485–498.

Tebaldi, C., Hayhoe, K., Arblaster, J. M., and Meehl, G. A. (2006), "Going to the Extremes," *Climatic Change*, 79, 185–211.

von Salzen, K., Scinocca, J. F., McFarlane, N. A., Li, J., Cole, J. N., Plummer, D., Verseghy, D., Reader, M. C., Ma, X., Lazare, M., et al. (2013), "The Canadian fourth generation atmospheric global climate model (CanAM4). Part I: representation of physical processes," *Atmosphere-Ocean*, 51, 104–125.

Weller, G. B., Cooley, D., Sain, S. R., Bukovsky, M. S., and Mearns, L. O. (2013), "Two case studies on NARCCAP precipitation extremes," *Journal of Geophysical Research: Atmospheres*, 118.

Weller, G. B., Cooley, D. S., and Sain, S. R. (2012), "An investigation of the pineapple express phenomenon via bivariate extreme value theory," *Environmetrics*, 23, 420–439.