# Extreme value comparison of different climate model simulations and observations

Mickey Warner, Bruno Sansó

**Abstract.** Climate models predict the intensity 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. We extend the comparison to an observation product. To assess the extremes of the series considered we fit a generalized Pareto model to the exceedances over a threshold. This model is applied to data over California and the contiguous United States. Comparisons for relevant distributions that result from the analysis are made visually with posterior parameter intervals and numerically using the Bhattacharyya distance between probability densities. We find that in some domains, the simulations are in agreement among themselves and with the observations, but in others they are quite different. 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. The results show weak to moderate tail dependence in nearly every comparison made [which are simulations to observations].

# 1 Introduction

The main focus of this paper is to compare the extreme values of an observation product with those of CanCM4 climate simulations. We address at least 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 whether the climate model is a reasonable representation of observed extremes.

The classic approach to analyzing extreme values is to model block maxima (e.g. annual maxima). It can be shown that under certain conditions the block maxima of independent random variables have a distribution which belongs to the generalized extreme value (GEV) family of distributions (Coles, 2001). Such an approach naturally requires omitting a large portion of the data. This can be remedied by using a threshold exceedance model which involves selecting some large threshold and fitting the exceedances to the generalized Pareto distribution (GPD). See Coles (2001) for an excellent introduction to these and other approaches in extreme value analysis.

We take a Bayesian approach to modeling threshold exceedances (section 3.1). Under this framework, taking into account the replicated experiments is naturally addressed with a hierarchical model.

We attempt to answer the questions posed earlier by comparing posterior intervals for statistical model parameters and other quantities such as return level (section 3.3) and

Bhattacharyya distance (section 3.4). Additionally, the bivariate analysis (section 4) allows us to measure the strength of tail dependence between simulations and observations.

# 2 Data

## 2.1 Climate model simulations

The Fourth Generation Coupled Global Climate Model (CanCM4) from the Canadian Centre for Climate Modeling and Analysis (CCCma) is made up 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 variety of variables on a roughly 2.5° degree grid over the globe (see Merryfield et al. (2013)). Two variables will be analyzed: precipitation (labled `pr`, in meters) and maximum temperature (labeled `tasmax`, in Kelvin). We further restrict our attention to analyzing two seasons—summer and winter—and two regions—California and the U.S.

Three experimental classes that are of particular interest are decadal, historical, and pre-industrial control runs. The decadal simulations provide climate estimates for ten years into the future, after conditioning on the state of the ocean at the starting time of the simulation. We consider two decades in this analysis: 1962–1971 and 1990–1999, which are conditioned on ocean states in 1961 and 1989, respectively. Historical simulations are obtained for the years 1961–2005 and are noted for including events that affect the climate such as volcanoes. The pre-industrial control, or simply control, simulations begin at 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.

## 2.2 Observations

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). To make the observations comparable to the climate simulations, we take weighted sums or averages of the climate simulations and just sums or averages of the observations. See section 2 for details, along with other changes made to the data in preparation for analysis.

## 2.3 Aggregation

We will analyze precipitation and temperature over both California and the United States. In each case, we take the climate moodel grid cell locations and create non-overlapping cells, or rectangles, such that each location is roughly in the center of the cell (see Figure 1). 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

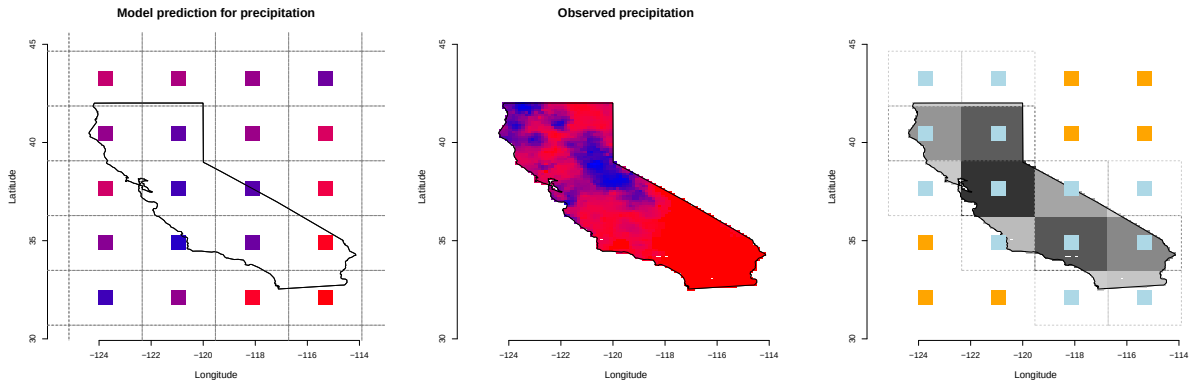


Figure 1: Left: CanCM4 simulation grid cells. Center: Observation locations. Right: method for computing weighted sum or average for CanCM4 to make values comparable with observations; the lighter gray points mean less weight is applied to the climate simulations and the darker gray means more weight. The data shown are from a single day in January.

(the right plot in Figure 1 shows which climate simulation locations have non-zero weight). For precipitation, we take a weighted sum and for temperature a weighted average. 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.

## 2.4 De-trending

Climate data are often non-stationary series characterized by complicated trends and cycles. As such, these present problems when studying extremes. 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). We will review some basic concepts for DLMs, see Prado and West (2010) chapter 4 for more details.

A normal DLM is specified by the quadruple $\{\mathbf{F}_t, v_t, \mathbf{G}_t, \mathbf{W}_t\}$ which determine how a univariate time series $y_1, \ldots, y_T$ is modeled over time. We assume

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

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

An advantage to model (1) is its capability in yielding 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 $\mathbf{F}_t^\top \mathbf{a}_t$ (we refer the reader to Prado and West (2010) for the algorithmic details).

Our DLM is finalized in the following way. We construct $\mathbf{F}_t$ and $\mathbf{G}_t$ such that the evolution of $\boldsymbol{\theta}_t$ has annual and semi-annual periods, i.e. the first and second harmonics.

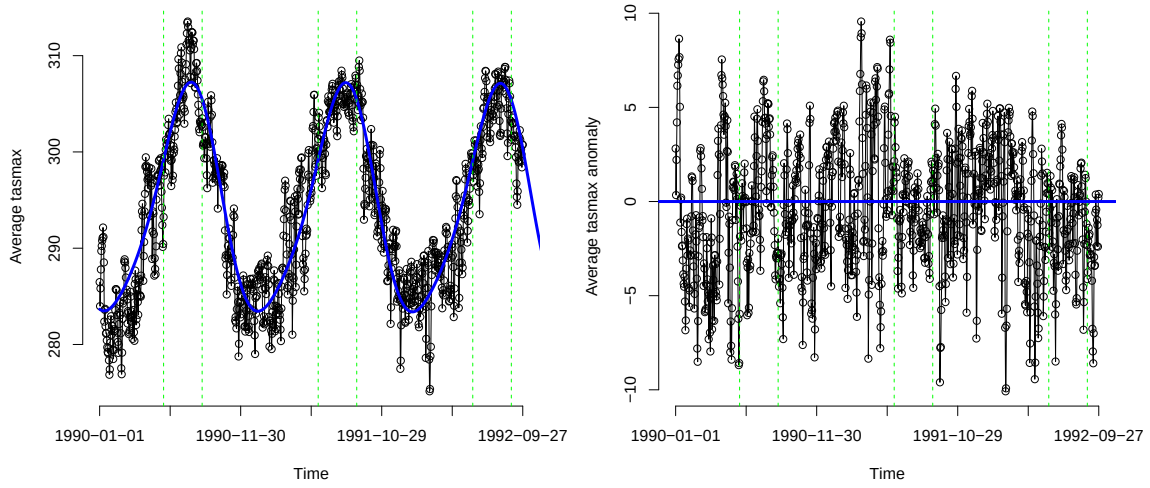


Figure 2: One of the DLMs used to calculate the anomalies. Shown is one of the decadal replicates of average `tasmax` in California for about the first two and one-half years of the time-series. The green dashed lines mark the beginning and the end of the summer months.

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.

In the end, we are left with the residuals. See Figure 2. The blue line in the left plot is the mean of $y_t$, $\mathbf{F}_t^\top \mathbf{a}_t$, given 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 anoamlies. So, for example in winter, 28 February is followed immediately by 1 December.

# 3 Univariate analysis

In this section we describe our method for analyzing each of our four data sources (the three climate simulation types and the observations). For each source we have four factors having two levels each. This provides us with $4 \times 4^2 = 64$ total data sets to which we apply the models and methods of this section.

## 3.1 Threshold exceedance model

### 3.1.1 Univariate

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 paremeter $\xi \in \mathbb{R}$ 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. Re-ordering the exceedances so $i = 1 \ldots, k$, the likelihood of $(\xi, \sigma)$ is derived from (2) as

$$L(y_1, \ldots, y_k; \sigma, \xi) = \sigma^{-k} \sum_{i=1}^{k} \left(1 + \frac{\xi y_i}{\sigma}\right)_+^{-1/\xi - 1} \tag{3}$$

where $z_+ = \max(z, 0)$.

### 3.1.2 Hierarchical model

Suppose we have $R$ replicates or computer simulations, each with $n_i$ observations, for $i = 1, \ldots, R$. Let $X_{ij}$ denote the $j$th observation in replicate $i$. We assume

$$X_{ij} \sim F_i, \qquad i = 1, \ldots, R, \qquad j = 1, \ldots, n_i$$

and 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

$$\begin{aligned} L(\mathbf{y}; \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\zeta}) &= \prod_{i=1}^{R} f_{Y_i}(\mathbf{y}_i | \sigma_i, \xi_i, \zeta_i) \\ &\approx \prod_{i=1}^{R} \left[ (1 - \zeta_i)^{n_i - k_i} \zeta_i^{k_i} \prod_{j \in A_i^c} \frac{1}{\sigma_i} \left(1 + \xi_i \frac{y_{ij}}{\sigma_i}\right)_+^{-1/\xi_i - 1} \right] \end{aligned} \tag{4}$$

Note that the parameters describing the tail of $F_i$ (i.e. $\xi_i, \sigma_i$) depend only on those observations which exceed $u$. The parameter $\zeta_i = P(X_{ij} > u)$, which is necessary for calculating return levels (section 3.3), is based only on the number of exceedances.

We complete the hierarchical model formulation by specifying the following priors:

$$\begin{aligned}
\xi_i|\xi,\tau^2 &\sim Normal(\xi,\tau^2)\\
\sigma_i|\alpha,\beta &\sim Gamma(\alpha,\beta)\\
\zeta_i|\zeta,\eta &\sim Beta(\zeta\eta,(1-\zeta)\eta)
\end{aligned} \tag{5}$$

$$\begin{aligned}
\xi &\sim Normal(m,s^2) & \tau^2 &\sim InvGamma(a_\tau,b_\tau)\\
\alpha &\sim Gamma(a_\alpha,b_\alpha) & \beta &\sim Gamma(a_\beta,b_\beta)\\
\zeta &\sim Beta(a_\zeta,b_\zeta) & \eta &\sim Gamma(a_\eta,b_\eta)
\end{aligned}$$

By combining (4) and (5) we obtain the full posterior distribution. Samples are obtained via MCMC. The inverse gamma is parametrized to have mean $b/(a-1)$ and the gammas have mean $a/b$.

## 3.2 Extremal Index

The threshold exceedance model described in section 3.1 relies on an assumption of independence 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 $\theta$ if for each $\tau > 0$ there is a sequence $\{u_n\}_{n\geq 1}$ such that,

$$\begin{aligned}
\lim_{n\to\infty} n(1-F(u_n)) &\to \tau \text{ and}\\
\lim_{n\to\infty} P(\max(X_1,\ldots,X_n)\leq u_n) &\to \exp(-\theta\tau).
\end{aligned}$$

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 $T_\theta$ is the random variable for interexceedance times in the limit, then $T_\theta$ is distributed according to the mixture

$$(1-\theta)\epsilon_0 + \theta\mu_\theta \tag{6}$$

where $\epsilon_0$ is the degenerate probability distribution at 0 and $\mu_\theta$ is an exponential distribution with mean $\theta^{-1}$. This means that the role of $\theta$ 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 $\theta$ since is it impossible to observe an interexceedance time of zero in practice.

We next describe a hierarchical model used to esimate $\theta$. This is distinct from the threshold exceedance model and is used only in getting a single estimate for $\theta$, which is used to decluster the exceedances and to calculate return levels.

### 3.2.1 Estimation

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

$$\begin{aligned} l(\theta, p; \mathbf{T}) = m_1 \log(1 - \theta p^\theta) + (N - 1 - m_1)\{\log(\theta) + \log(1 - p^\theta)\} \\ + \theta \log(p) \sum_{i=1}^{N-1} (T_i - 1) \end{aligned} \tag{7}$$

where $p$ is the probability of not exceeding the threshold. We require that this likelihood be used in a hierarchical model.

Suppose 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 $\theta_1, \ldots, \theta_R$. However, since these all come from the same climate model, we may wish to assume that the $\theta_i$ 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 (7), 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 choosing priors for $\theta$, $\nu$, $p$, and $\tau$—the latter two parameters being required only for model (7). We assume

$$\begin{aligned} \theta &\sim Beta(a_\theta, b_\theta) \\ \nu &\sim Gamma(a_\nu, b_\nu) \\ p &\sim Beta(a_p, b_p) \\ \tau &\sim Gamma(a_\tau, b_\tau) \end{aligned}$$

with the hyperparameters chosen to be

$$\begin{array}{lll} \theta\text{:} & a_\theta = 1 & b_\theta = 1/2 \\ \nu\text{:} & a_\nu = 1 & b_\nu = 1/10 \\ p\text{:} & a_p = 100\hat{F} & b_p = 100(1 - \hat{F}) \\ \tau\text{:} & a_\tau = 1 & b_\tau = 1/10 \end{array}$$

where $\hat{F} = \sum_{i=1}^{R} \sum_{j=1}^{n} \mathbb{1}(X_{i,j} \leq u)$. Our parametrization for the gamma random variables are such that $X \sim Gamma(\alpha, \beta)$ has mean $\alpha/\beta$. The prior values for $\theta$ attempt to mitigate some of the issues surrounding model (7)

By assuming conditional indepedence between the simulations, we can construct the following log-likelihood

$$L = \sum_{i=1}^{R} l(\theta_i, p_i; \mathbf{T}^{(i)}) \tag{8}$$

where $\mathbf{T}^{(i)}$ is the vector of interexceedance times for replicate $i$ having length $N_i$. In the univariate setting for the observation product, only model (7) is needed.

Süveges (2007) proposed on 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 that given in (7). In a separate simulation study, both likelihoods performed very similarly, with some preference to model (7), within the hierarchical setting.

### 3.2.2 Declustering

Declustering is done as given 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 $C_i$th 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.

## 3.3 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

$$G(x_m) = 1 - \frac{1}{m} \tag{9}$$

and has the convenient interpretation as the quantity that is exceeded on average once every $m$ observations.

When working with the generalized Pareto model (2), it can be shown that the $m$-observation return level is

$$x_m = u + \frac{\sigma}{\xi}\left[(m\zeta\theta)^{\xi} - 1\right] \tag{10}$$

where the terms $\zeta$ and $\theta$ account for the probability of exceeding $u$ and being within a cluster, respectively. We can obtain a distribution for $x_m$ using MCMC samples for $(\xi, \sigma, \zeta)$. Posterior samples for $\theta$ are obtained separately from $(\xi, \sigma, \zeta)$, and so we choose to use the posterior mean for $\theta$ when computing return levels. The intention here is avoid possible complications due to the fact we do not have samples of the joint vector $(\xi, \sigma, \zeta, \theta)$.

## 3.4 Bhattacharyya distance

Since our focus is on comparing different climate summaries, we must assess differences between a variety of posterior distributions. A naive approach may be to simply determine

whether posterior intervals overlap. Though we make use of visuals in our comparison, we desire a more quantitative approach. Here, we make use of Bhattacharyya distance.

Bhattacharyya (1943) proposed a means for measuring the degree of similarity between two probability distributions. For two continuous random variables on support $\mathcal{X}$ with densities $p$ and $q$, the Bhattacharyya coefficient is defined as

$$BC(p, q) = \int_{\mathcal{X}} \sqrt{p(x)q(x)}dx \tag{11}$$

and the Bhattacharyya distance is

$$D_B(p, q) = -\log BC(p, q). \tag{12}$$

We use kernel density estimation to calculate $p$ and $q$ along a grid of the support and then approximate the integral in (11). If the support is different for the two random variables (as will typically be the case when comparing random variables whose parameters determine the support such as the generalized Pareto), we will integrate over the intersection of the supports.

Our approach is to compute distances from the replicates to their mean and determine whether the observations could have reasonably come from the climate model. Taking the shape parameter as an example, from the hierarchical model in 3.1.2 we have posterior samples for $\xi_1^c, \ldots, \xi_R^c$ for, say, some decadal simulations. We also have posterior samples for the mean $\xi^c$. Using the kernel density estimation mentioned earlier, we obtain $R$ Monte Carlo estimates $D_B(\xi_i^c, \xi^c)$, for $i = 1, \ldots, R$. From the univariate model 3.1.1 we have the shape parameter $\xi^o$ for the observation product. Finally, we calculate $D_B(\xi^o, \xi^c)$ and ask whether this quantity falls within the range of $D_B(\xi_i^c, \xi^c)$. When this occurs, we say $\xi^o$ is "similar" to the $\xi_i^c$ since the observation differs from the mean climate model in a similar way as the replicates differ from the mean.

# 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, though all the issues associated with univarite extremes persist (threshold selection, too few data points) and a handful of new problems arise.

Coles (2001) chapter 8 goes into the math behind the distribution of bivarate maxima

Goix et al. (2015) talk about some really complicated crap for modeling multivariate extremes

de Haan and Resnick (1977) more complicated crap, limit theory for multiextremes

Coles and Tawn (1991) provide several families for modeling bivariate extremes

Coles et al. (1999) discuss means of estimating tail dependence

Dependence in the tails cannot be measured using a typical correlation which measures dependence in the center of the distribution. This gives rise to a statistic which measures asymptotic tail dependence discussed in section 4.2.

Extremes are rare by definition, so increasing the number of dimensions exacerbates the issue of working with too few data points. For a threshold exceedance model, it is unclear how to appropriately select a multivariate threshold.

We perform several bivariate analyses using Pareto processes to model the joint tail behavior.

## 4.1 Simple Pareto processes

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

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

with $A \in \mathcal{B}(C_1^+(S))$, $P(\partial A) = 0$, $W$ some simple Pareto process (see Appendix A.2), and

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

This theorem provides us with the means of transforming our data to a Pareto process, which we will in turn use to describe asymptotic tail dependence.

We assume that $t$ is large enough that the theorem applies (implying $u = u_t$ and $\sigma = \sigma_t$). Being interested in the bivariate case, we have data at only two fixed locations $s_1, s_2 \in S$ for each bivariate comparison. The particular values of $s_1$ and $s_2$ are irrelevant since we are comparing climate simulations to observations which have no quantitative meaning as far as their position in $S$ is concerned; we only require that the labels for the observations be appropriately distinguished.

We further assume that the parameters $\xi$ and $\sigma$ are indexed by $s \in S$, so that our transformation is

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

The first stage of our analysis involves estimating $\xi(s)$ and $\sigma(s)$ marginally, which is accomplished by selecting a high threshold $u(s)$ and fitting the generalized Pareto distribution to the excesses $X(s) - u(s)$. The posterior means $\hat{\xi}(s)$ and $\hat{\sigma}(s)$ are then used for the transformation from $X(s)$ to $W(s)$.

Every observation of $X(s)$, say $X_1(s), \ldots, X_{n(s)}(s)$ is transformed with

$$W_i(s) = T_t X_i(s) = \left(1 + \hat{\xi}(s) \frac{X_i(s) - u_t(s)}{\hat{\sigma}(s)}\right)_+^{1/\hat{\xi}(s)}, \qquad i = 1, \ldots, n(s) \tag{13}$$

forming the vector $\mathbf{W}(s) = (W_1(s), \ldots, W_{n(s)}(s))^\top$. After performing this transformation, two components are combined to form a joint vector $(W(s_1), W(s_2))$ having realizations $\mathbf{W}_{12} = (\mathbf{W}(s_1), \mathbf{W}(s_2))$, an $n(s) \times 2$ matrix. Note that when we perform the bivariate analysis, we guarantee that $n(s_1) = n(s_2) = n(s)$.

By Theorem 3.2, $\mathbf{W}_{12}$ has rows that are realizations of a simple Pareto process. By the constructive definition of a simple Pareto process (Appendix A.2), we can write $\mathbf{W}_{12}$ as

$$\mathbf{W}_{12} = \begin{pmatrix} Y_1 V_1(s_1) & Y_1 V_1(s_2) \\ Y_2 V_2(s_1) & Y_2 V_2(s_2) \\ & \vdots \\ Y_n V_n(s_1) & Y_n V_n(s_2) \end{pmatrix}$$

where $Y_i$ is a standard Pareto random variable, $V_i(s_j) \geq 0$ $(j = 1, 2)$, and $V_i(s_1) \vee V_i(s_2) = 1$, for $i = 1, \ldots, n = n(s)$. This is easily obtained by

$$Y_i = W_i(s_1) \vee W_i(s_2), \quad \text{and} \quad V_i(s_j) = W_i(s_j)/Y_i \quad (j = 1, 2), \qquad \text{for } i = 1, \ldots, n,$$

where $a \vee b = \max(a, b)$. The points $(V_i(s_1), V_i(s_2))$ fall along the curve of the non-negative unit sphere with supremum norm $\{(v_1, v_2) : ||(v_1, v_2)||_\infty = 1, v_1 \geq 0, v_2 \geq 0\}$ which is thus one dimensional. An alternative representation is to specify $(V_i(s_1), V_i(s_2))$ in terms of a scaled angle

$$\phi_i = \frac{2}{\pi} \arctan\left(\frac{V_i(s_2)}{V_i(s_1)}\right) \in [0, 1].$$

We scale $\phi_i$ to be in $[0, 1]$ so we can model the density of the angles using a Bernstein-Dirichlet prior (BDP) (Petrone, 1999). Since we will be fitting models to many sets of $\phi_i$'s, we desire a model that is flexible at capturing a large variety of possible distributions. The BDP is fit using the `R` package `DPpackage`.

Since Theorem 3.2 holds when $\sup_{s \in S} T_t X(s) > 1$ we need only those $\phi_i$ for which $Y_i > 1$. This corresponds to only using the angles that are associated with a threshold exceedance. Also, because the BDP has difficulty when too much mass is on the edges, we make the following adjustment. The $\phi_i \in [0, 0.005]$ are treated as zero and those in $[0.995, 1]$ are treated as one. The remainder are used in the BDP model. We find that $k_{max} = 300$, which determines the maximum degree of Bernstein polynomials the model is to allow for, was acceptable when applied to the angles.

## 4.2 Asymptotic tail dependence

We wish to characterize the strength of tail dependence for climate simulation and observation pairs. This is typically done with the following statistic. Suppose $X$ and $Y$ share a common marginal distribution. Then

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

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

Under the simple Pareto process from section 4.1, the distribution function for $W_i \equiv W(s_i)$ is

$$F_{W_i}(w) = 1 - \frac{E(V_i)}{w}$$

where $V_i \equiv V(s_i)$ and $w > 1$. Using this fact, we can standardize $W_i$ to be uniform and compute $\chi$ in this way

$$\begin{aligned}\chi &= \lim_{u \to 1} P(F_{W_1}(W_1) > u | F_{W_2}(W_2) > u) \\ &= E\left(\frac{V_1}{E(V_1)} \wedge \frac{V_2}{E(V_2)}\right), \end{aligned} \tag{14}$$

where $a \wedge b = \min(a, b)$. The major downside to the simple Pareto process is that we do not allow for asymptotic independence unless $P(V_1 \wedge V_2 = 0) = 1$, which would require conditioning on a set of zero probability.

# 5 Results

For 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
4. Region — California or U.S.A.

There are then 16 combinations of the factors to be made. For each combination, the hierarchial model described in section 3.1.2 is fit to the decadal, historical, and control runs; the univariate model in section 3.1.1 is fit to the observation product since this data source does not have replicates.

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 seection 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 values of the thresholds themselves are not too important since different thresholds may produce similar return levels.

For the simple Pareto process of section 4, we make the comparison between the observations and each climate simulation. Specifically, for each of the 16 factor combinations mentioned earlier, we fit the BDP to the angles resulting from taking the observations to each replicate of a particular climate simulation (say, decadal) and then compute $\chi$ from (14). This would result in $R = 10$ estimates for $\chi$, one for each replicate. We also fit the BDP to all 10 sets of angles together to get a "overall" measure for asymptotic dependence between the observations and the simulation.

Figures 3 through 9 show posterior parameters and other quantities of interest from the univariate analysis. For the hierarchical model, we show the results of the *mean* process. For example, in Figure 3 the parameter shown is the posterior for $\xi$, the mean of $\xi_1, \ldots, \xi_R$. This is in opposition to inference on an unknown replicate which would require sampling, among

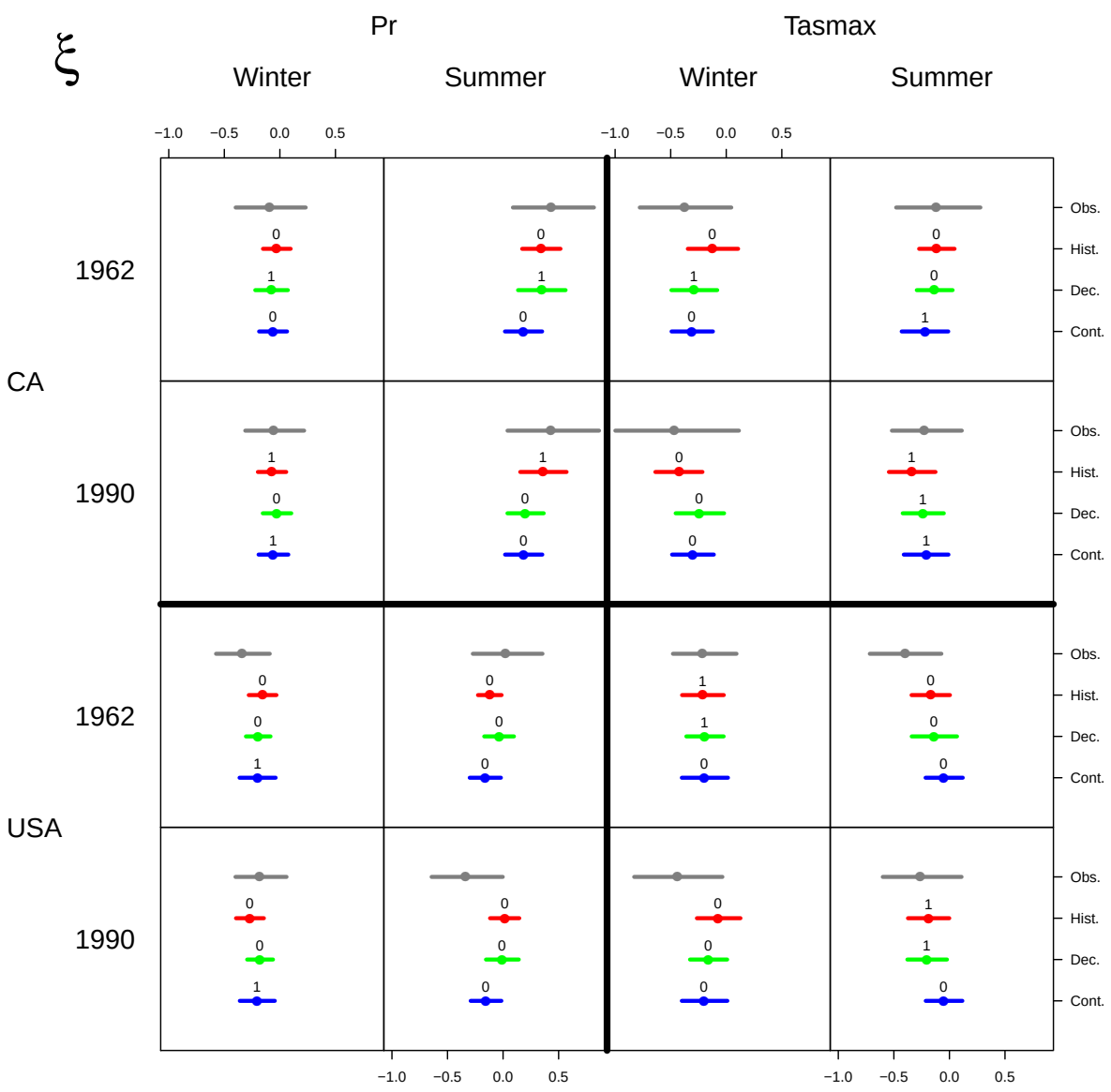


Figure 3: Posterior shape parameter, $\xi$, under each domain and each of the four data types. The points are the means and the lines mark the 95% h.p.d. intervals. Note: The $x$-axes are the same for every plot. The $y$-axes (for this and all subsequent figures) denote only the data type and thus hold no quantitative meaning.

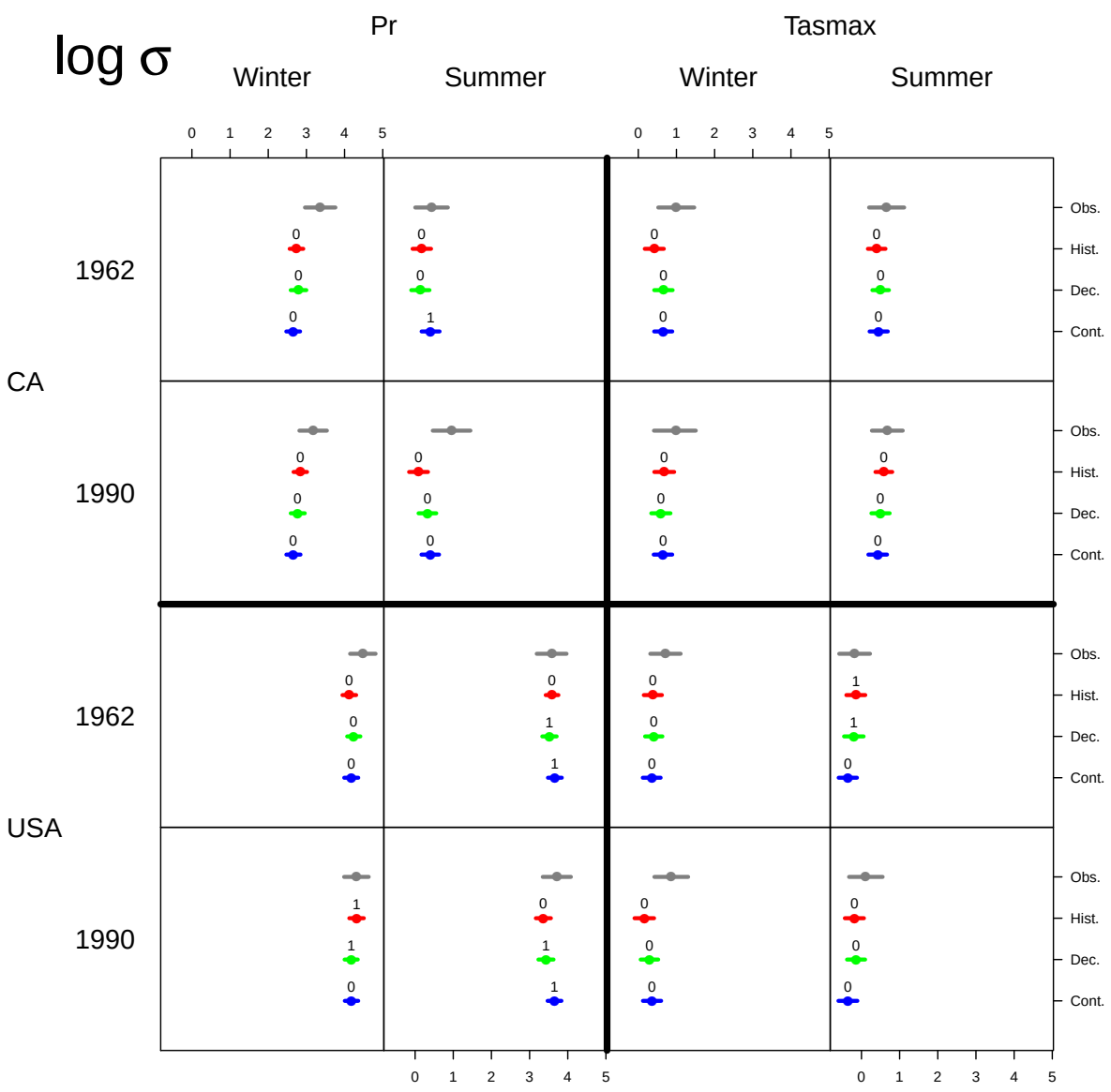


Figure 4: Natural logarithm of the posterior scale. For the CanCM4 simulations, the parameter shown is $\log(\alpha/\beta)$ (the mean scale) because $\sigma_i$ follows a Gamma distribution with mean $\alpha/\beta$. No change of variables is necessary for the observations. Note: The $x$-axes are the same for every plot.

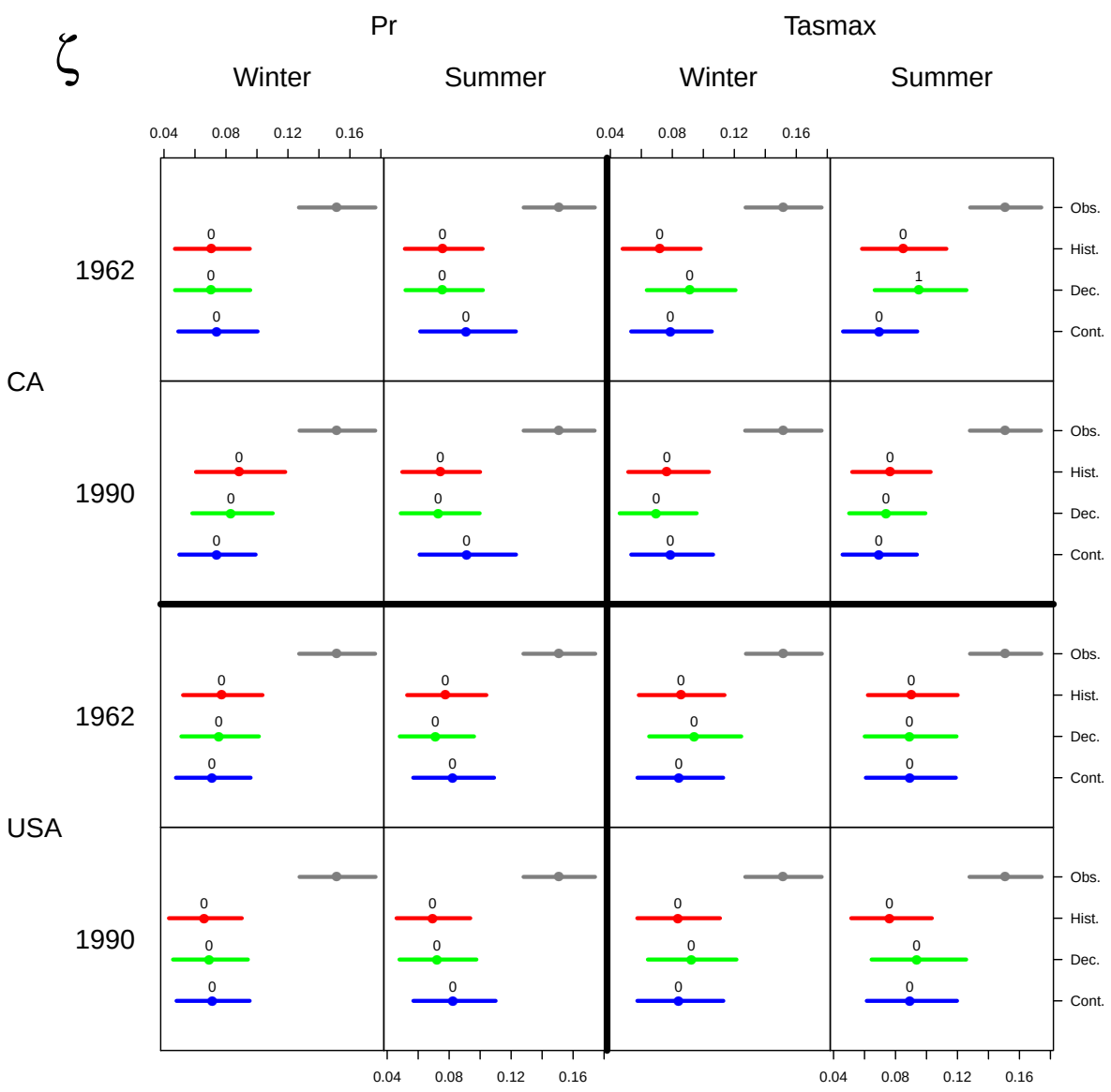


Figure 5: The probability of exceeding the threshold. These parameters are closely tied to the threshold, a user-specified quantity. Since we chose thresholds as the 0.95 quantile for the climate simulations and the 0.85 quantile for the observations, we do not expect there to be much overlap between these posteriors.

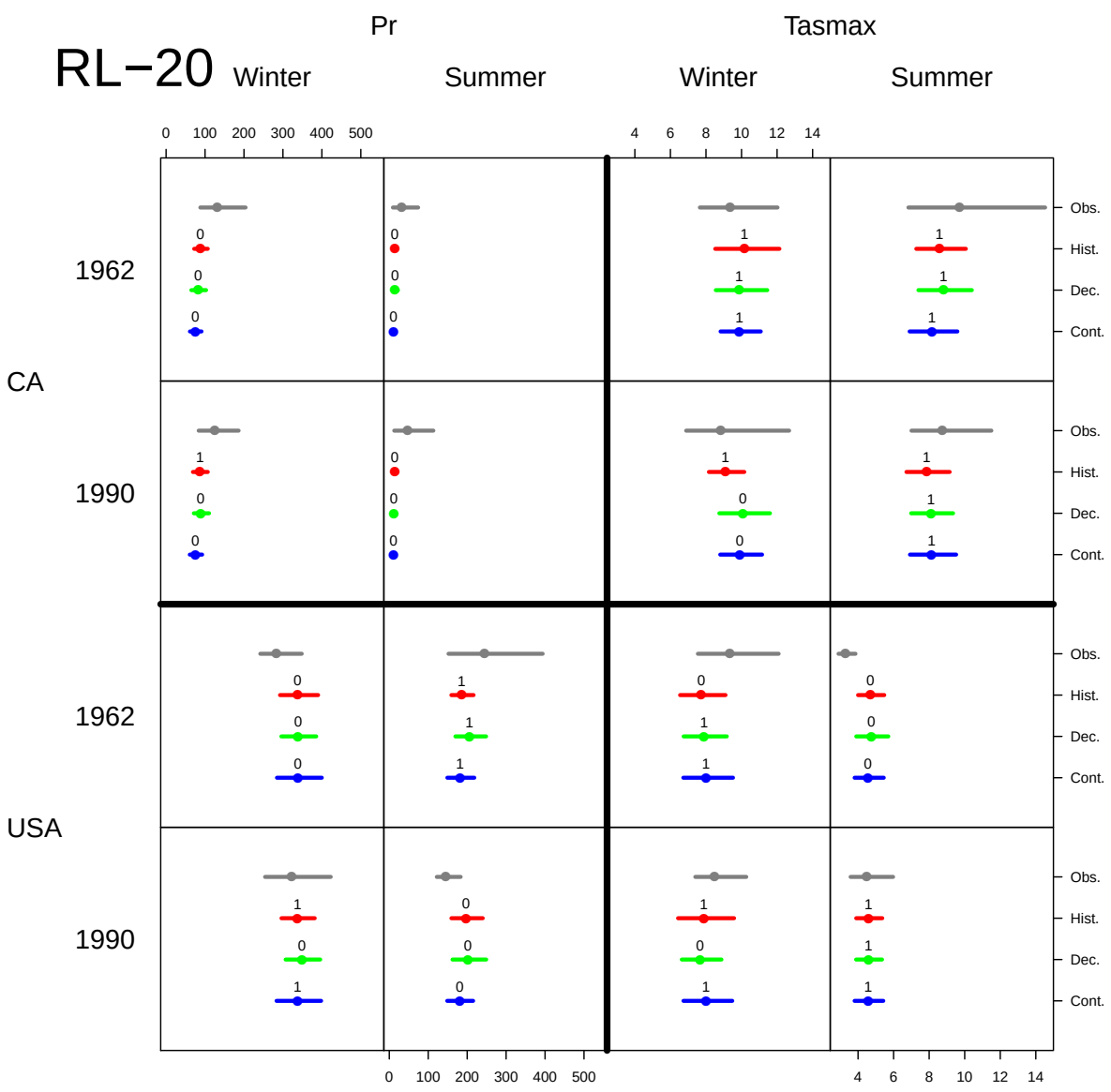


Figure 6: 20-year return levels. Note: The left two columns have the same $x$-axes, which are different than those in the right two columns, which have the same.

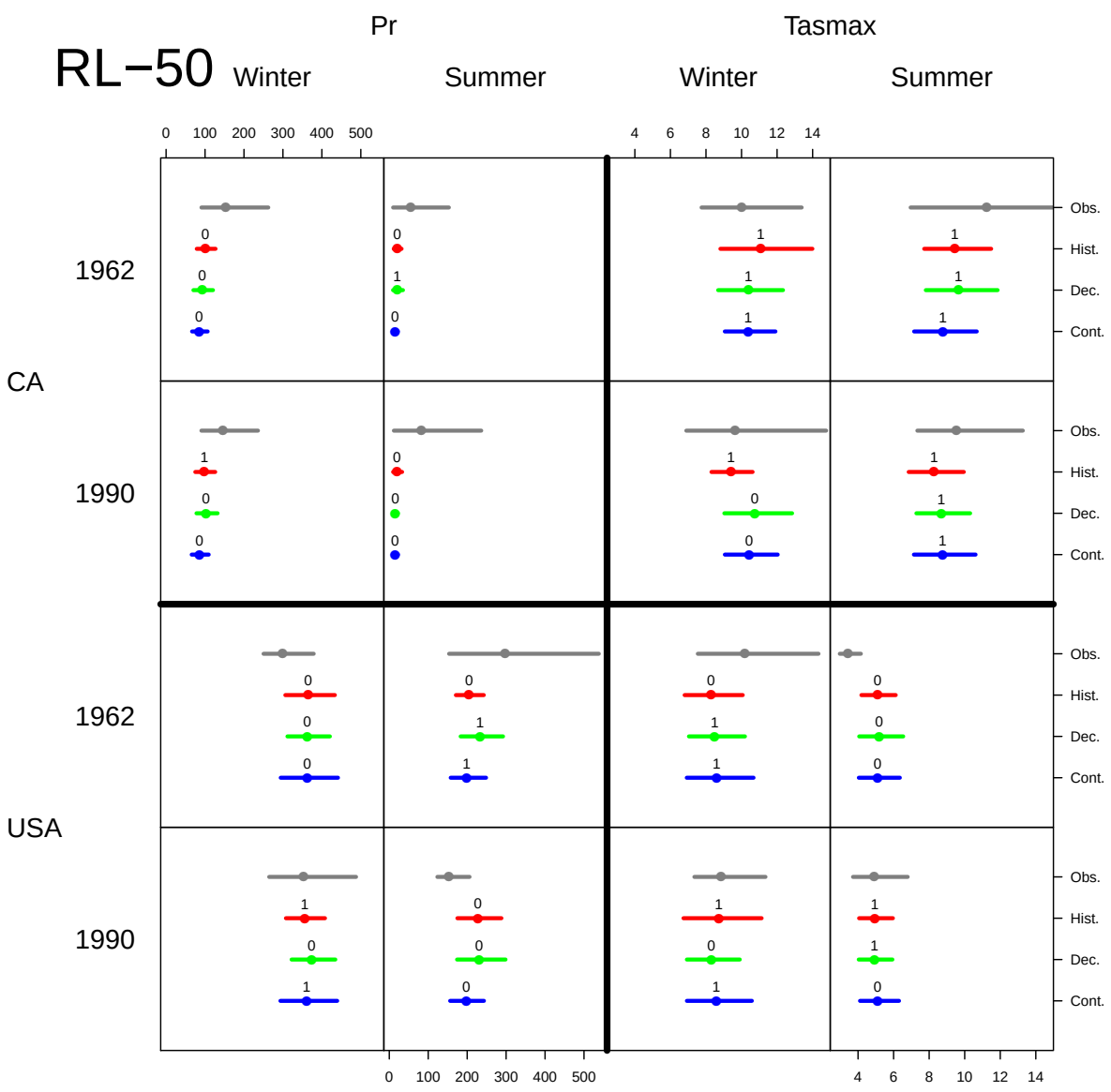


Figure 7: 50-year return levels. The $x$-axes are the same as those in Figure 6.

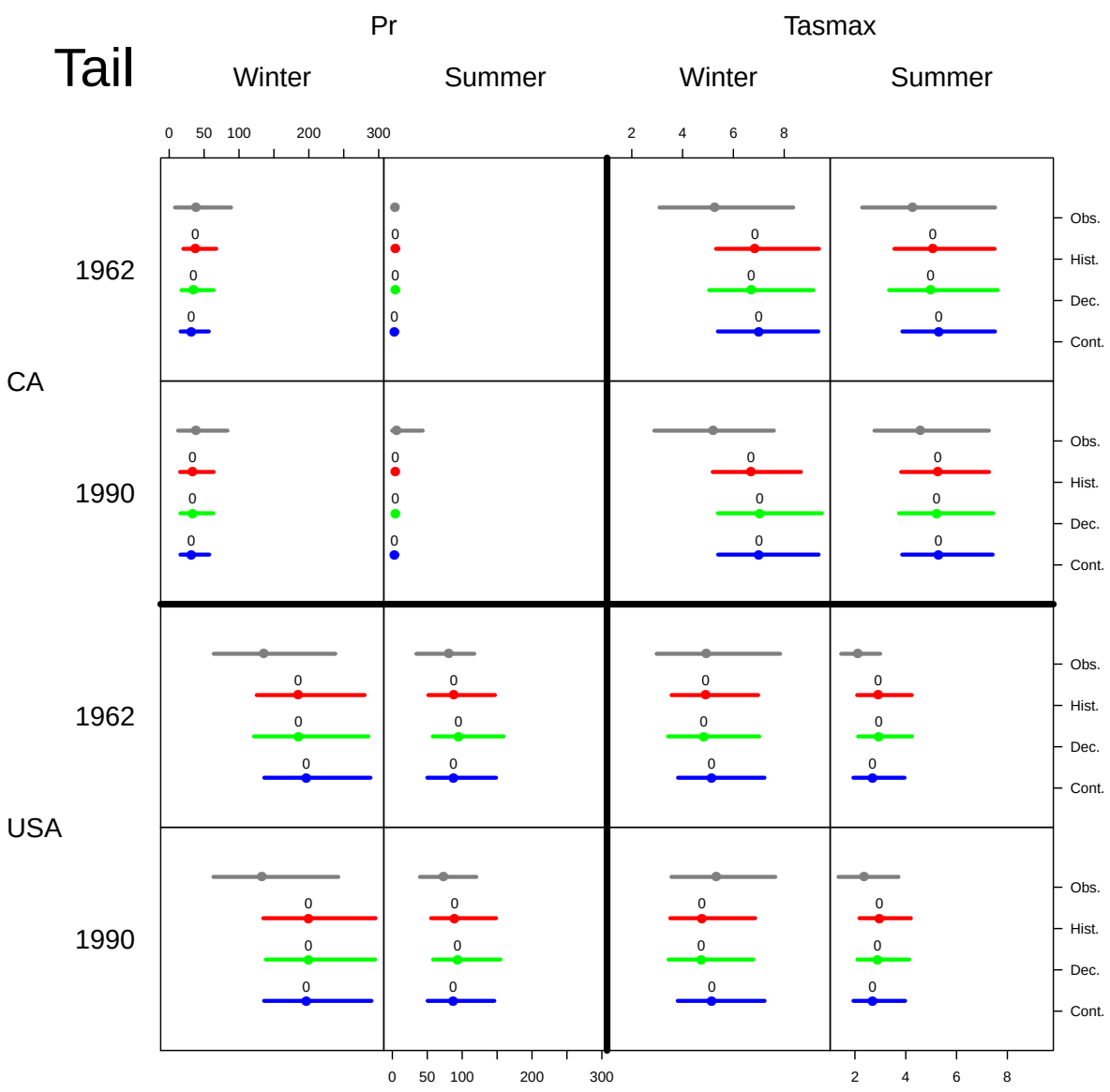


Figure 8: Mean and 95% h.p.d. for the upper tail (i.e. the generalized Pareto) of the ensemble average. As in Figures 6 and 7, the left two columns have the same $x$-axes and the right two columns have the same $x$-axes.

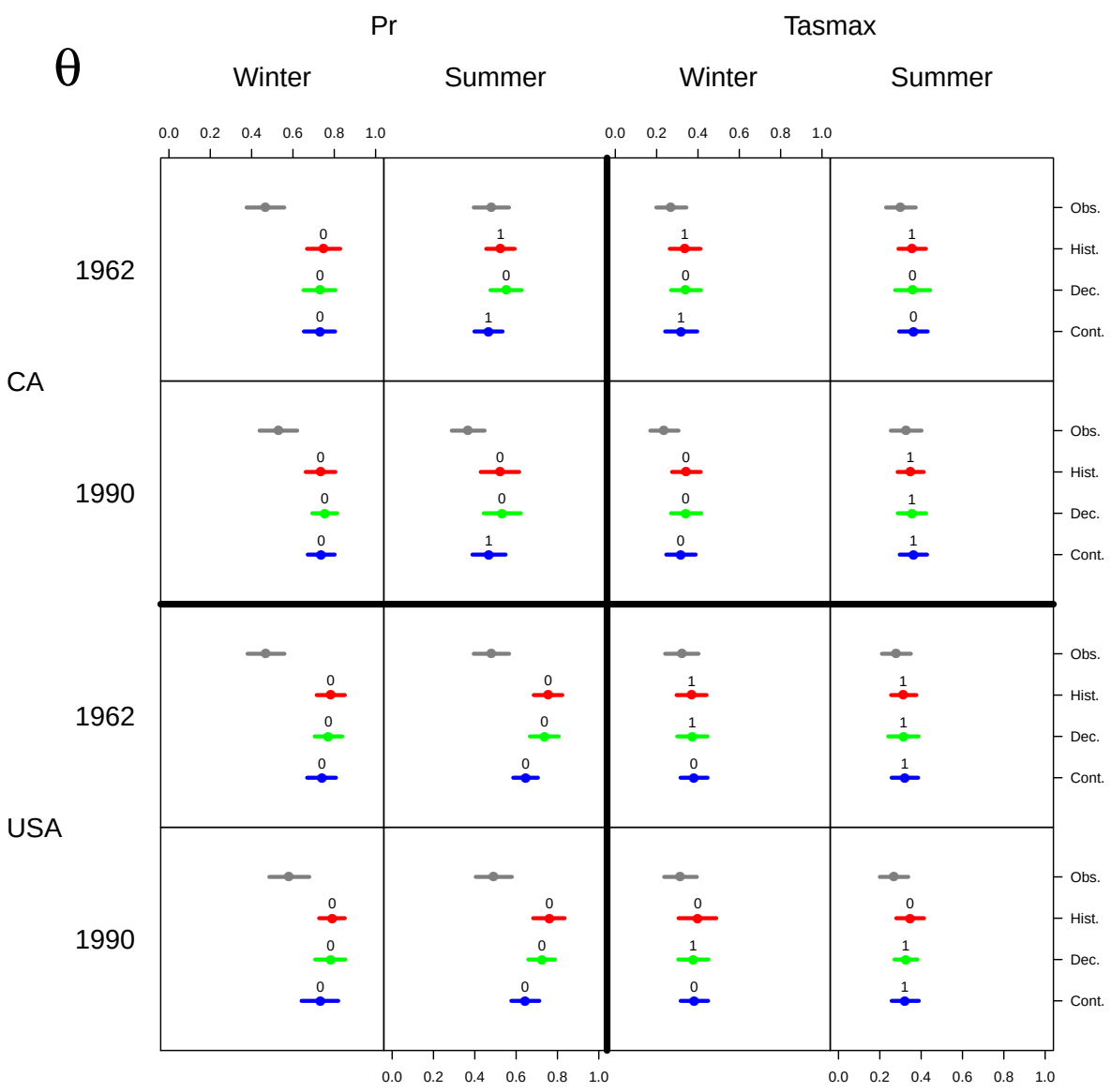


Figure 9: The mean extremal index. Like the parameters shown in Figures 3 and 4, the hierarchical mean is shown for the CanCM4 simulations.

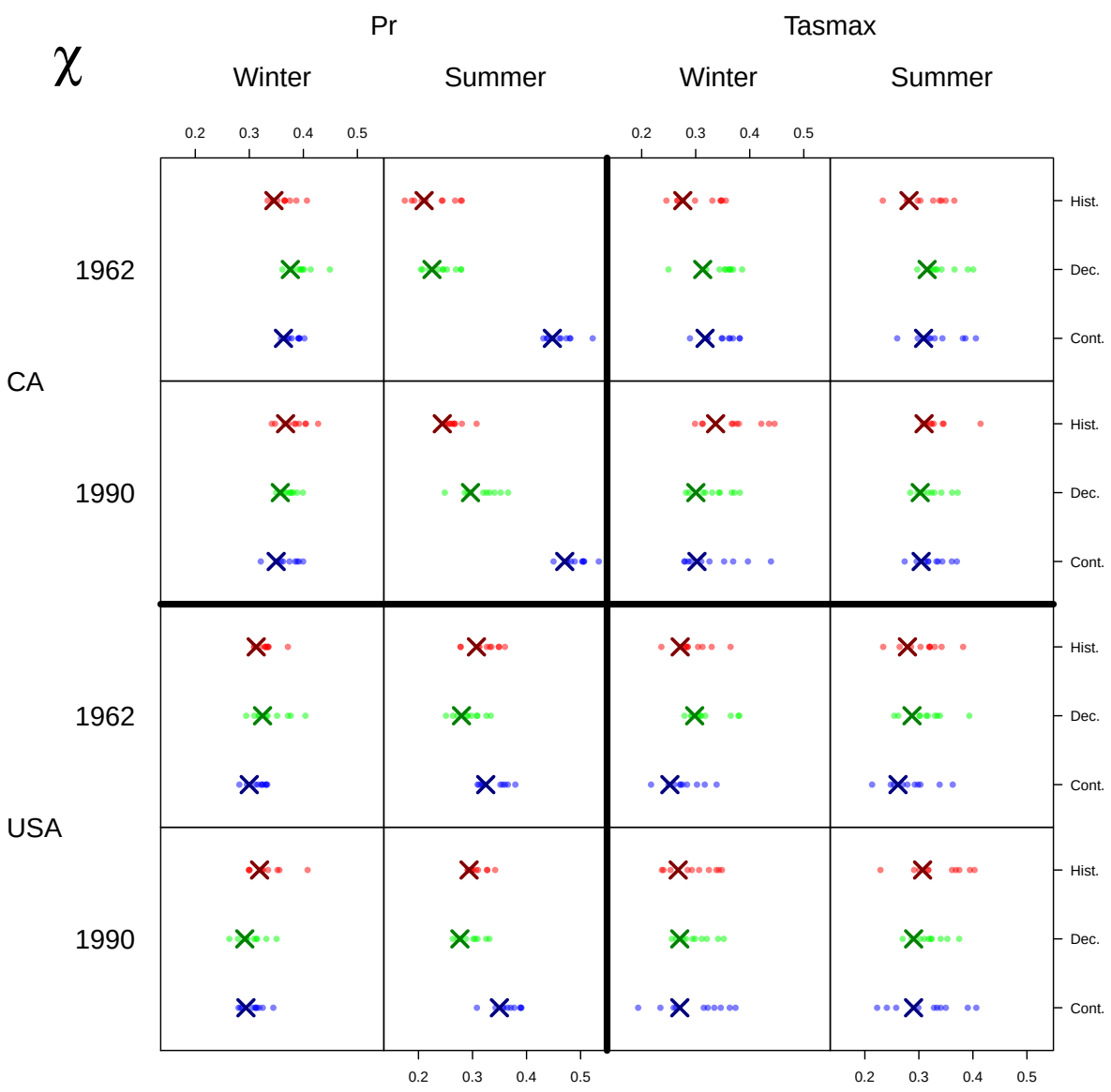


Figure 10: Measure of asymptotic dependence, calculated using results from the simple Pareto process. The dots mark the $\chi$ for each replicate against the observations. The X marks the value of $\xi$ when all replicates are taken together.

other things, a new shape parameter $\xi^*$. Therefore, the intervals are more narrow than if we looked at the posterior predictive distribution for a new replicate, but the parameters will be comparable to those from the univariate model with the observations and give us a sense of how the climate simulation performs on average.

The posterior shape parameters in Figure 3 show overlapping bounds in many cases, but in some combinations of factors we can see some departure from the observations. The numbers shown above the lines are indicators for whether the posterior from the observations is similar to the posterior of the replicates—in the sense of Bhattacharyya distance described in section 3.4—for a particular simulation class.

Figure 4 shows the logarithm of the mean scale parameter, $\log(\sigma)$ for the observations and $\log(\alpha/\beta)$ for the simulations, and the posteriors for $\zeta$, the probability of exceeding the threshold, are given in Figure 5.

Figures 6 and 7 give the 20- and 50-year return levels, respectively. In these figures, we have the same $x$-axes for the columns in total precipiation and for the columns under average maximum temperature. Some intervals are difficult to see given the scale, but we can still inspect how the return levels from the observations differ from those of the climate siulations. There is evidence of similarity between the data sources for maximum temperature (except for the 1962–1971 decade in the United States). The simulations struggle to find agreement with the observations when considering precipitation, only having two domains where the observations are similar to at least two of the climate sources (1990 USA Winter and 1962 USA Summer).

Figure 8 shows posterior predictive samples (based on the mean parameters for the hierarchical model) drawn from the generalized Pareto distribution, conditioned on the random variables being greater than the given threshold. With respect to the 95% posterior intervals, there is significant overlap of the simulations with the observations, but this is not the case with Bhattacharyya distance. This is because nearly all replicates were very close to their mean, having a distance of roughly 0.02, while the observations were different enough to have a distance of about 0.2 from the mean process. Despite all this, we saw in Figures 6 and 7 that there is still some similarity between the important return level quantities, which account for $\zeta$ and $\theta$.

The posterior mean and 95% highest posterior density intervals for the extremal index are shown in Figure 9. The climate simulations seem to be consistent with the observations for the temperature, but not so for precipiation where they tend to overestimate $\theta$. The Bhattacharyya distance was not computed for these parameters.

Estimates for asymptotic dependence are given in Figure 10. Many of them are in the region $[0.2, 0.4]$ indicating weak to moderate dependence. In only two instances do we have $\chi$ close to 0.5. And in these cases the comparison is made with the control runs, which should be 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.

# 6 Conclusions

We have proposed a hierarchical threshold model to handle replicates of climate simulations. The model was applied to a variety of factor combinations and compared to a univariate threshold model for observations. A handful of similarities and differences were found to exist between simulations and the observation product.

We have accounted for trends in the time-series by subtracting out the mean of dynamic linear models, thus yielding anomalies. This poses issues of interpretability and practicality. Our extreme value analysis are in terms of the anomalies and this may not be terribly useful when making comparisons. Should we be interested in the precipitation or temperature extremes in real terms, we must add back the (possibly unknown) mean. However, with a sufficient model on the mean, it would not be unreasonable to use our analysis for projecting extremes in the future.

Some improvement on our use of the Bhattacharyya distance can be made. In this paper we decided that the observations were "similar" enough to the climate simulations if the Bhattacharyya distance fell within the distances from the replicates to their mean. A better approach may be to compute a bootstrap sample of the distances and then calculate the proportion of times this exceeds the distance from the observations.

Our univariate analysis for each data source reveal similarities in many aspects (i.e., similar tail behavior and return levels). However, when we consider the joint distribution of the time series, we find there to be little in common. Meaning there would be some difficulty in using the results from a climate simulation to make a prediction for future extreme observations.

# 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 $I\!R^d 4$. Let $W$ be a stochastic process in $C^+(S)$ and $\omega_0$ a positive constant. When $W$ 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, $P(Y \leq y) = 1 - 1/y$, $y > 1$,

(c) $Y$ and $V$ are independent.

then $W$ is called a simple Pareto process.

## A.3 Measure of asymptotic dependence for simple Pareto process

Provided $V_1 > 0$ and $V_2 > 0$,

$$
\begin{aligned}
\chi &= \lim_{u\to 1} P(F_{W_1}(W_1) > u | F_{W_2}(W_2) > u) \\
&= \lim_{u\to 1} P\left(1 - \frac{E(V_1)}{W_1} > u \middle| 1 - \frac{E(V_2)}{W_2} > u\right) \\
&= \lim_{u\to 1} P\left(W_1 > \frac{E(V_1)}{1-u} \middle| W_2 > \frac{E(V_2)}{1-u}\right) \\
&= \lim_{u\to 1} \frac{P\left(W_1 > \frac{E(V_1)}{1-u}, W_2 > \frac{E(V_2)}{1-u}\right)}{P\left(W_2 > \frac{E(V_2)}{1-u}\right)} \\
&= \lim_{u\to 1} \frac{P\left(YV_1 > \frac{E(V_1)}{1-u}, YV_2 > \frac{E(V_2)}{1-u}\right)}{1-u} \\
&= \lim_{u\to 1} \frac{1}{1-u} P\left(Y > \frac{E(V_1)}{(1-u)V_1}, Y > \frac{E(V_2)}{(1-u)V_2}\right) \\
&= \lim_{u\to 1} \frac{1}{1-u} P\left(Y > \frac{E(V_1)}{(1-u)V_1} \vee \frac{E(V_2)}{(1-u)V_2}\right) \\
&= \lim_{u\to 1} \frac{1}{1-u} P\left(Y > \frac{1}{1-u}\left(\frac{E(V_1)}{V_1} \vee \frac{E(V_2)}{V_2}\right)\right) \\
&= \lim_{u\to 1} \frac{1}{1-u} (1-u) E\left[\left(\frac{E(V_1)}{V_1} \vee \frac{E(V_2)}{V_2}\right)^{-1}\right] \\
&= E\left(\frac{V_1}{E(V_1)} \wedge \frac{V_2}{E(V_2)}\right)
\end{aligned}
$$

# References


Bhattacharyya, A. (1943), "On a measure of divergence between two statistical populations defined by their probability distribution," *Bull. Calcutta Math. Soc.*

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

Coles, S., Heffernan, J., and Tawn, J. (1999), "Dependence measures for extreme value analyses," *Extremes*, 2, 339–365.

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

de Haan, L. and Resnick, S. I. (1977), "Limit theory for multivariate sample extremes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40, 317–337.

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.

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

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.

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.

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.

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.

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.