Comparison of extreme values of different climate model simulations and observations

Mickey Warner

30 Oct 2017

Introduction

CanCM4 simulation classes (with $R = 10$ replicates each):

1. Decadal
2. Historical
3. Control

Observations over U.S. interpolated from weather stations

Factors:

1. Variable — Total Precipitation (`pr`) or Average Maximum Temperature (`tasmax`)
2. Season — Winter or Summer
3. Decade — 1962–1971 or 1990–1999
4. Region — California or USA

Locations

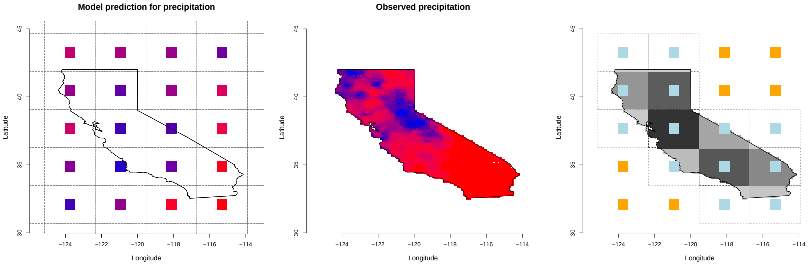
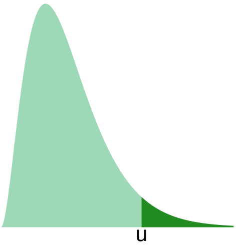


Figure: Left: CanCM4 simulation grid cells. Center: Observation locations. Right: method for computing weighted sum or average for CanCM4 to make values comparable with observations.

Extremes

For r.v. X and large threshold u, the exceedance $Y = X - u$, for $X > u$, approximately follows the generalized Pareto distribution (GPD), which has density

$$f_Y(y) = \frac{1}{\sigma}\left(1 + \xi \frac{y}{\sigma}\right)_+^{-1/\xi - 1}$$



Data processing

Goals: 1) make the data from the simulations comparable to those of the observations and 2) get near-independent random variables for model fitting

1. Weighted sums (pr) or averages (tasmax) give a time-series for each factor combination and data source

Using the GPD is not justified without independent random variables.

We try to get iid r.v.s by first obtaining the anomalies and then declustering

2. Anomalies are obtained by fitting a DLM to each time-series
3. Decluster by estimating the extremal index with interexceedance times, then perform runs declustering

Three key ideas. Why? Getting (closer to) iid rvs, another option is to use to time-varying parameters instead of the dlm, but our time

Likelihood

Replicate i, observation j, exceedances $Y_{ij} = X_{ij} - u$, and keep only those $Y_i > 0$. These have likelihood

$$L(\mathbf{y}; \boldsymbol{\sigma}, \boldsymbol{\xi}, \boldsymbol{\zeta}) = \prod_{i=1}^{R} \left[(1 - \zeta_i)^{n_i - k_i} \zeta_i^{k_i} \prod_{j=1}^{k_i} \frac{1}{\sigma_i} \left(1 + \xi_i \frac{y_{ij}}{\sigma_i} \right)_+^{-1/\xi_i - 1} \right]$$

where n_i is number of X_{ij}'s and k_i is number of Y_{ij}'s, and ζ_i is the probability of exceeding the threshold (which is chosen, so ζ_i is essentially determined by human)

Priors

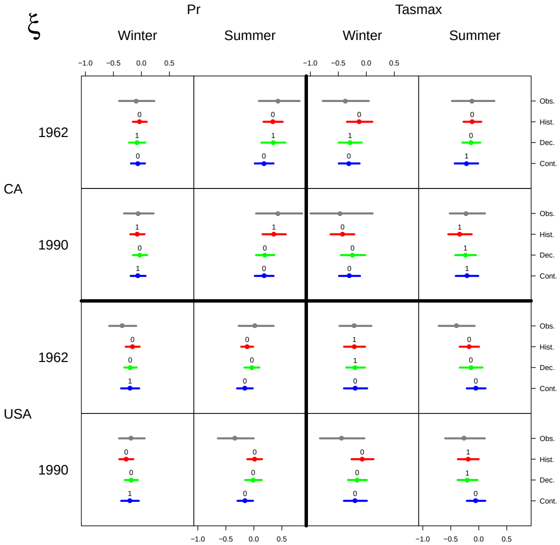
These priors complete the hierarchical model formulation. Greek letters are random variables while English letters are fixed.

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

$$\begin{aligned}
\alpha_\sigma \sim Gamma(a_\alpha, b_\alpha) &\qquad \beta_\sigma \sim Gamma(a_\beta, b_\beta)\\
\xi \sim Normal(m, s^2) &\qquad \tau^2 \sim Gamma(a_\tau, b_\tau)\\
\mu \sim Beta(a_\mu, b_\mu) &\qquad \eta \sim Gamma(a_\eta, b_\eta)
\end{aligned}$$

Bhattacharyya distance

Bivariate analysis, simple Pareto processes



ξ
Pr
Tasmax
Winter
Summer
Winter
Summer
CA
USA
1962
1990
1962
1990
Obs.
Hist.
Dec.
Cont.