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

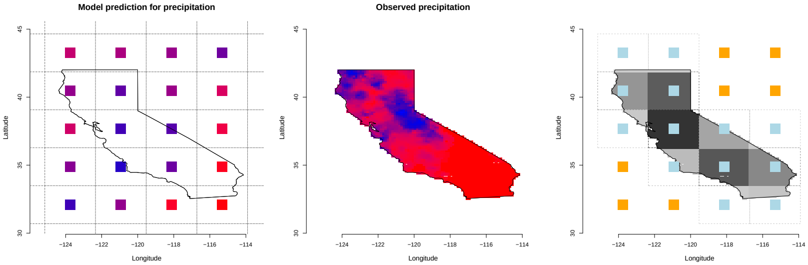
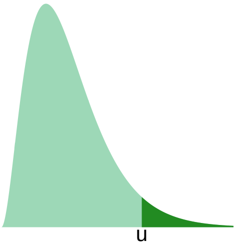


Figure: Left: CanCM4 simulation grid cells. Center: Observation locations. Right: method for computing weighted sum or average for CanCM4 to make values comparable with observations.

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

Two objectives before performing the analysis:

1. Make climate simulations comparable to observations
2. Get near-independent random variables for model fitting

These are accomplished by

1. Taking weighted sums (`pr`) or weighted averages (`tasmax`)
2. Computing anomalies based on DLMs, and
3. Declustering

Weighted sum or average

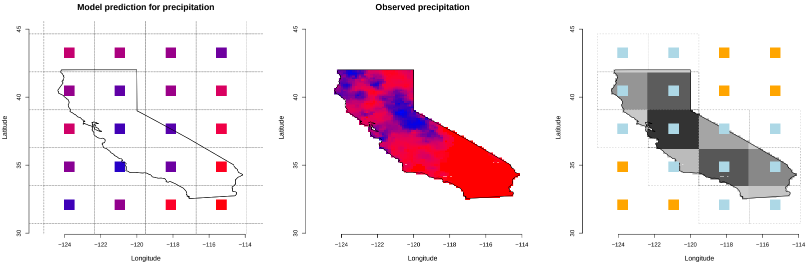
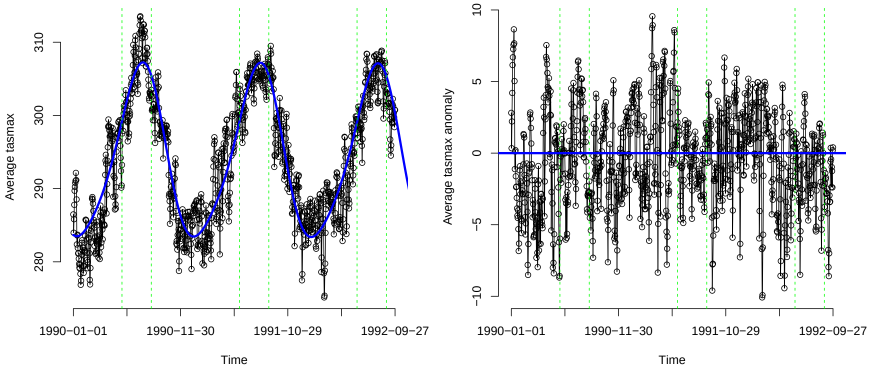


Figure: Left: CanCM4 simulation grid cells. Center: Observation locations. Right: method for computing weighted sum or average for CanCM4 to make values comparable with observations.

DLM-based anomaly



Extremal index (declustering)

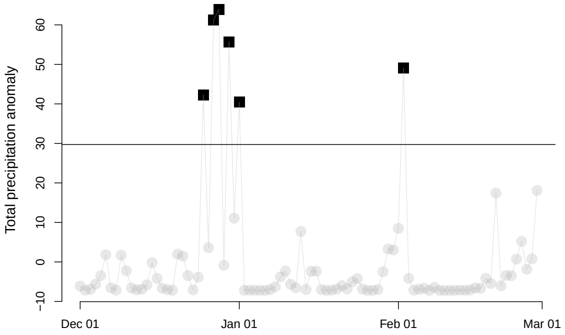
The extremal index θ is the inverse of the limiting mean cluster size

It can be estimated using interexceedance times, $T_i = S_{i+1} - S_i$, with a log-likelihood of

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

p is the probability of not exceeding the threshold

Declustering



Likelihood

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

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

n_i is the number of X_{ij}'s
k_i is the number of Y_{ij}'s
ζ_i is the probability of exceeding the threshold

Priors

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

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

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

Return level

For a distribution G, the return level x_m is the solution to

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

The value x_m is exceeded on average once every m observations.

For the GPD, the return level is given by

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

Bhattacharyya distance

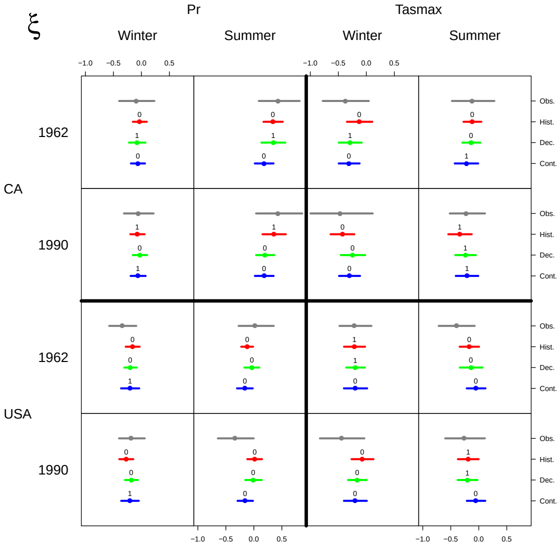
Bhattacharyya coefficient

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

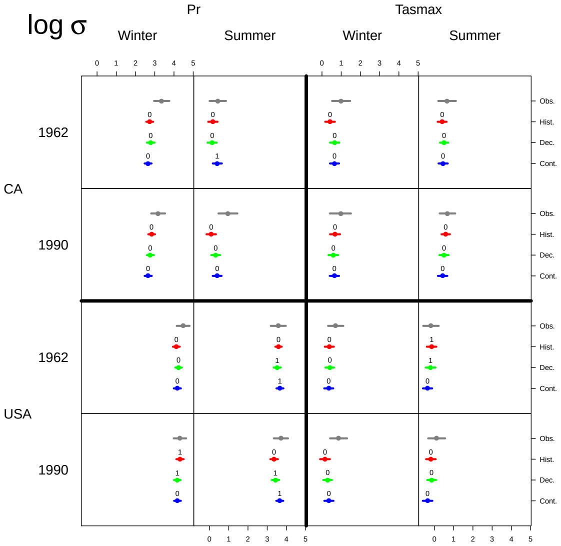
Bhattacharyya distance

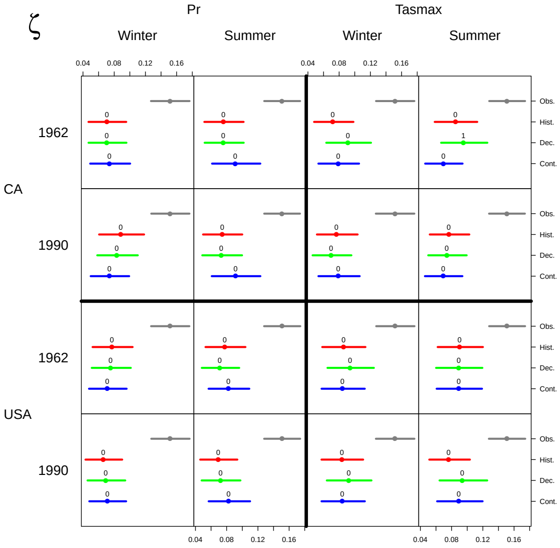
$$D_B(p,q) = -\log BC(p,q).$$

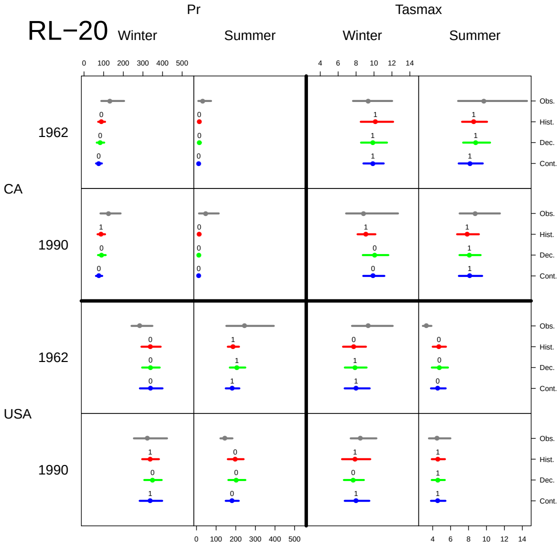
D_B is computed between parameters in the replicates (and observations) and parameters in the hierarchy.



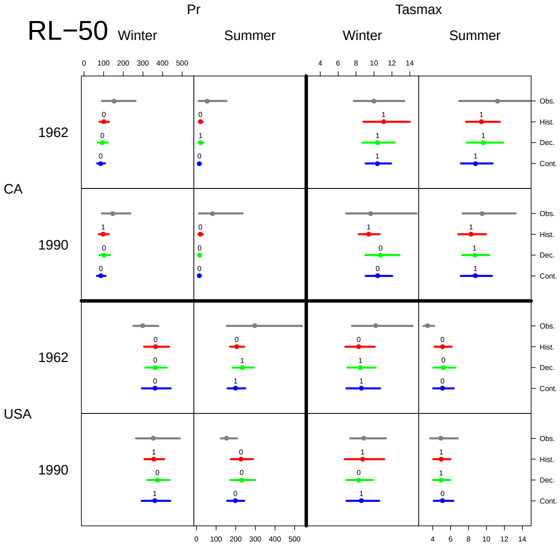
ξ
Pr
Tasmax
Winter
Summer
Winter
Summer
1962
1990
1962
1990
CA
USA
Obs.
Hist.
Dec.
Cont.
−1.0
−0.5
0.0
0.5


log σ
Pr
Tasmax
Winter
Summer
Winter
Summer
CA
USA
1962
1990
1962
1990
Obs.
Hist.
Dec.
Cont.
0 1 2 3 4 5


ζ
Pr
Tasmax
Winter
Summer
Winter
Summer
0.04
0.08
0.12
0.16
1962
1990
CA
1962
1990
USA
Obs.
Hist.
Dec.
Cont.

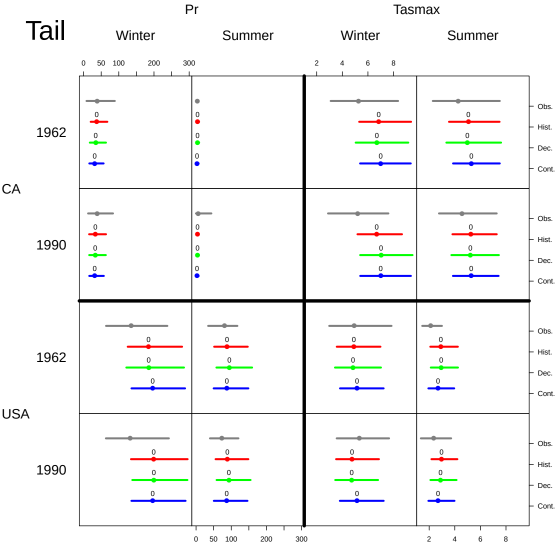


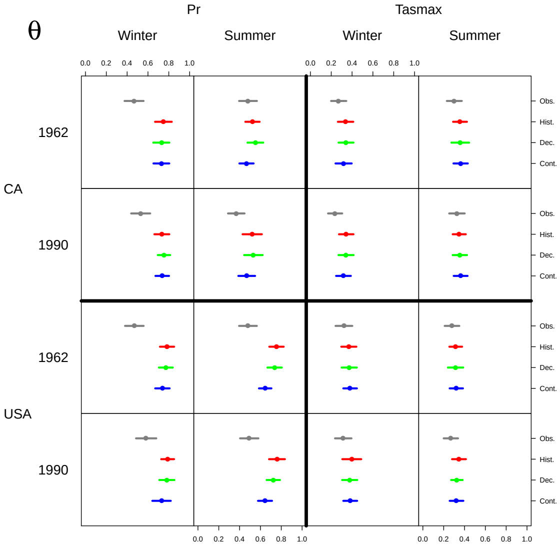
RL−20
Pr
Tasmax
Winter
Summer
Winter
Summer
CA
USA
1962
1990
1962
1990
Obs.
Hist.
Dec.
Cont.



RL−50
Pr
Tasmax
Winter
Summer
Winter
Summer
CA
1962
1990
USA
1962
1990
Obs.
Hist.
Dec.
Cont.
0 100 200 300 400 500
4 6 8 10 12 14

Tail


Pr
Tasmax
Winter
Summer
Winter
Summer
0 50 100 200 300
2 4 6 8
CA
USA
1962
1990
1962
1990
Obs.
Hist.
Dec.
Cont.


θ
Pr
Tasmax
Winter
Summer
Winter
Summer
0.0 0.2 0.4 0.6 0.8 1.0
CA
USA
1962
1990
1962
1990
Obs.
Hist.
Dec.
Cont.

Next

Bivariate analysis using simple Pareto processes