

A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling

Erin M. Schliep · Daniel Cooley · Stephan R. Sain · Jennifer A. Hoeting

Received: 15 April 2009 / Revised: 15 November 2009 / Accepted: 26 November 2009 / Published online: 18 December 2009
© Springer Science+Business Media, LLC 2009

Abstract We analyze output from six regional climate models (RCMs) via a spatial Bayesian hierarchical model. The primary advantage of this approach is that the statistical model naturally borrows strength across locations via a spatial model on the parameters of the generalized extreme value distribution. This is especially important in this application as the RCM output we analyze have extensive spatial coverage, but have a relatively short temporal record for characterizing extreme behavior. The hierarchical model we employ is also designed to be computationally efficient as we analyze RCM output for nearly 12000 locations. The aim of this analysis is to compare the extreme precipitation as generated by these RCMs. Our results show that, although the RCMs produce similar spatial patterns for the 100-year return level, their characterizations of extreme precipitation are quite different. Additionally, we examine the spatial behavior of the extreme value index and find differing spatial patterns for the point estimates for the RCMs. However, these differences may not be significant given the uncertainty associated with estimating this parameter.

E. M. Schliep · D. Cooley (✉) · J. A. Hoeting
Department of Statistics, Colorado State University,
Fort Collins, CO, USA
e-mail: cooleyd@stat.colostate.edu

S. R. Sain
Geophysical Statistics Project, National Center
for Atmospheric Research, Boulder, CO, USA

Keywords Generalized extreme value distribution · Climate change impacts · Intrinsic autoregressive model · NARCCAP · Reanalysis-driven simulations

AMS 2000 Subject Classification 62-Statistics

1 Introduction

The motivation for this work comes from concern about impacts of climate change. The summary for policymakers of the Intergovernmental Panel on Climate Change's (IPCC) 2007: Synthesis Report (IPCC 2007a) begins with the statement: "Warming of the climate system is unequivocal, as is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice and rising global average sea level." As attention turns toward assessing potential impacts of climate change and toward possible approaches for climate change mitigation, many new questions arise. Policymakers will need to not only have scientists' best estimates of the effects of climate change, but also accurate quantifications of the uncertainty of these estimates as they weigh the costs of mitigating climate change or the costs of addressing the impacts of a changed climate.

In terms of impacts, it is often rare but extreme weather events that are the most costly in economic or human terms. There is currently much interest in assessing possible changes in weather extremes due to climate change, and a recent summary of the state of the science and findings for the United States can be found in CCSP (2008). Because it is impossible to observe weather under a changed climate until the climate change actually occurs, much of our knowledge about the impacts of climate change come from climate model simulations. In this work, we compare the maximum precipitation data generated by a number of climate models to begin to understand and quantify how extreme precipitation may differ between the climate models.

Extreme value analyses can be challenging due to the lack of informative data. Even when given a long time series of data, because extreme events are rare, there still can remain considerable uncertainty in the parameter estimates that describe the tail behavior of the modeled distribution. When given data from multiple locations, one may wish to sensibly "trade space for time" to better estimate the parameters at all locations.

There have been several approaches of varying complexity that have attempted to borrow strength across location. The methodology described in this paper is motivated by the data sets which we analyze. We investigate extreme precipitation from six regional climate model (RCM) simulations. The RCM output are spatially rich in that we have daily precipitation amounts at every location in our study domain. At the same time, the RCM output are temporally poor as we have only 20 year simulations, and this is a relatively short period to determine extreme value behavior. Additionally, the data analysis is computationally challenging as we have nearly 12000 locations that we need to model simultaneously. To understand the nature of extreme precipitation as produced by these RCMs, we construct a Bayesian

hierarchical model which sensibly borrows strength across location and is computationally feasible. This is done by constructing a multivariate spatial model on the parameters of the extreme value distribution.

1.1 Climate models and the NARCCAP project

Climate models are deterministic models that produce simulated weather. These models are able to capture the known physics of the Earth's climate system by producing a discretized solution to the differential equations that describe the fluid dynamics of the Earth's atmosphere and oceans. The weather from these models is simulated at discrete time intervals on a given spatial grid. Climate models are usually extremely computationally intensive, expensive to run, and run on supercomputers. For an overview of climate models and their evaluation see IPCC (2007b, ch. 8). The power of climate models is that they allow scientists to not only simulate weather under current atmospheric and ocean conditions, but also under possible altered conditions due to natural and/or anthropogenic causes.

The output of a climate model run are a vast collection of simulated weather data. The model records numerous variables (e.g., temperature, precipitation, humidity, barometric pressure to name just a few) at all grid locations and at every time interval. As is common practice in the climate literature, we will refer to the weather 'data' produced by the climate model as 'model output'.

There are different types of climate models that are used for different purposes. An atmosphere-ocean general circulation model (AOGCM) is used to model synoptic (large-scale) phenomena. These models are run on a spatial domain that spans the globe and thus are often referred to as global climate models or GCMs. GCMs have grid scales on the order of 100's of kilometers. Therefore, GCMs are able to capture climate phenomena on the continental or sub-continental scale; however their spatial resolution leaves them unable to model climate patterns or changes at a more local level.

GCMs differ from weather forecast models. The purpose of a GCM is not to mimic weather at a specific time. That is, the GCM-simulated weather corresponding to a particular "day" most likely will not resemble the actual observed weather for that day. However, the weather produced by a GCM simulation will span decades of (simulated) time. The climatological characteristics of this simulated weather should capture that of actual weather which would occur given the general atmospheric conditions that characterize the GCM run.

While the understanding of synoptic climate phenomena is an important facet of climate change research, it is often desirable to investigate localized climate patterns. RCMs are a different type of climate model that simulate weather at a higher resolution than GCMs. The grid scale for RCMs are on the order of 10's of kilometers, and therefore, RCMs are better able to answer questions about local impacts of climate change. RCMs do not simulate weather over the entire globe, but rather the spatial domain of a RCM is restricted to a study area of interest. Because local weather is largely determined by synoptic scale phenomena, RCMs need to be driven by something which provides the large-scale weather patterns. That is, data capturing the synoptic climate phenomena is needed as input for the RCM.

It is natural to, therefore, drive a RCM with GCM model output. This allows researchers to compare different climate scenarios at the local level. It is also possible, however, to drive RCMs with reanalysis model output. Reanalysis models are retrospective models of the atmosphere that take observed weather data and upscale to produce model output on a regular grid and discretized time intervals. The reanalysis model's dynamic upscaling method takes into account the known physics of the Earth system, like a GCM. Reanalysis model output can be thought of as data that a GCM would produce if the model were to know the actual weather, and therefore a fundamental difference of reanalysis output is that, unlike GCM output, it should reflect the observed weather from a specific time. Reanalysis model output also differs from observed weather data in its spatial and temporal scale. Reanalysis driven RCMs are beneficial because they make it possible to compare the RCMs to actual weather observations.

Multiple institutions are currently producing RCMs to gain knowledge of climate patterns over specified regions. With such an abundance of models, it is currently unclear as to the similarities and differences between the climate patterns being projected by the models. The North American Regional Climate Change Assessment Program (NARCCAP) is a project that is currently producing a suite of simulations from different RCMs (Mearns et al. 2009). NARCCAP is studying six different RCMs and these models are being driven by a set of four AOGCMs and, additionally, a reanalysis model. The AOGCMs are run according to the A2 scenario (Nakicenovic and Swart 2000) while the reanalysis models are, as described above, based on observed data. The study region for all of these RCM simulations spans North America. A primary goal of NARCCAP is to investigate variability between these simulations and to attempt to attribute this variability to differences in both the GCMs and the RCMs.

1.2 Hierarchical models for extremes

The purpose of this work is to compare the extreme precipitation simulated by each of the six RCMs. We will focus only on the RCM simulations being driven by reanalysis output. Because each of the RCMs share a common driver, any differences are attributable to the RCMs themselves.

This is not the first study whose aim is to characterize extreme values of climate model output. Most previous studies have used a point-by-point analysis and have not tried to borrow strength across locations (e.g., Kharin et al. 2007; Frei et al. 2006). This point-by-point approach results in considerable parameter uncertainty as it is difficult to characterize extreme behavior given a relatively short data record.

One approach for trading space for time is regional frequency analysis (RFA). This approach is summarized in the monograph by Hosking and Wallis (1997) and traces its roots back to the pioneering work of Dalrymple (1960). RFA pools data among predetermined homogenous geographic regions. Spatial hierarchical modeling of extremes is a recent alternative to RFA and has only become feasible as a result of increased computational capabilities and the statistical methodologies that have been developed to utilize these capabilities. Like RFA, a hierarchical model

combines information across locations effectively increasing the sample size at each location. The hierarchical approach differs from RFA in that it constructs a spatial model for the parameters that describe the marginal distributions' tails.

Much of the early work in spatial hierarchical modeling was done in the context of epidemiology; the book by Banerjee et al. (2004) gives an excellent overview of the general practice and many of the examples within reflect this origin. Only recently has the technique been applied to extreme value analyses. Casson and Coles (1999) produced an early work that modeled simulated hurricane wind speeds with a one-dimensional spatial model. Cooley et al. (2007) used a hierarchical approach to model extreme precipitation data from weather stations and then used their hierarchical model to interpolate over the study region. Sang and Gelfand (2009) modeled annual maximum rainfall over a region of South Africa. In their work, Sang and Gelfand use a spatial autoregressive model in their hierarchy as their data occur on a regular lattice. The spatial model that we employ here is similar. This work builds on that of Cooley and Sain (2008), who model extreme precipitation from the output of a single RCM. As their study region covered the western one-third of the contiguous United States, Cooley and Sain showed that it is beneficial to spatially model the shape parameter of the extreme value distribution; previous studies had assumed this parameter to be constant over their study regions.

The work here differs from that of Cooley and Sain (2008) in a number of ways. The goal of Cooley and Sain was to compare the extreme precipitation produced by a control (current climate conditions) run to that of a run under a future climate scenario where both runs come from a common RCM driven by a GCM. Here, our goal is to compare extreme precipitation from six different RCMs that are being driven by the same reanalysis output. Another difference is that in this work we model annual maxima for each season via a model based on the generalized extreme value (GEV) distribution whereas Cooley and Sain used a point process model for threshold exceedances. Additionally, the spatial domain of our study area is much larger, covering all of North America, and this leads to increased computational difficulty. Like Cooley and Sain, we model all three (location, scale, and shape) parameters of the extreme value distribution over the study region. Of particular interest is the shape parameter (also called the extreme value index) which determines the tail behavior of the distribution and which is often difficult to estimate.

The remainder of the article is structured as follows. In Section 2, we provide further information about the RCMs that provide the output and the reanalysis model that was used to drive the RCMs. In Section 3, we explain the hierarchical model that we construct and our implementation method for obtaining inference. We present a summary of results for the six RCMs in Section 4. We conclude with a brief discussion and summary.

2 Description of regional climate model output

Table 1 lists the six RCMs being compared in this work as well as the corresponding modeling group. We henceforth refer to the RCMs by their four letter acronym in

Table 1 RCMs used in the comparison study

<i>i</i>	Model	Full name	Modeling group
1.	CRCM	Canadian Regional Climate Model	OURANOS / UQAM
2.	ECPC	Experimental Climate Prediction Center	UC San Diego / Scripps
3.	HRM3	Hadley Regional Model 3	Hadley Centre
4.	MM5I	MM5 - PSU/NCAR mesoscale model	Iowa State University
5.	RCM3	Regional Climate Model version 3	UC Santa Cruz
6.	WRFP	Weather Research & Forecasting model	Pacific Northwest Nat'l Lab

the first column of the table. Because they were developed independently, the RCMs differ in their parameterizations; the major characteristics of the six RCMs can be found at <http://narccap.ucar.edu/data/rcm-characteristics.html>. We will be comparing the results of these six RCMs, each being driven by the same reanalysis output. The RCM simulation we investigate begins in December 1981 and terminates in November 2002. The weather simulated by the RCMs is output every 3 hours, however we choose to analyze the daily precipitation totals as this is a familiar summary measure and many other data sets are summarized this way. While the spatial grid for the six RCMs vary somewhat, the daily precipitation values for each are interpolated to a common 98×120 grid with 50 km resolution, yielding 11760 grid cells covering nearly all of North America.

Each of these RCM simulations is driven by the NCEP Reanalysis 2 output (Kanamitsu et al. 2002) which comes from the National Oceanic & Atmospheric Administration National Centers for Environmental Predictions. The NCEP Reanalysis 2 project is ongoing. It has produced a reanalysis dating back to 1979 and currently has generated output through December of 2005. NCEP uses a global T-62 Gaussian grid resulting in a spatial grid of 192 (longitude) \times 94 (latitude) which covers the globe.

Seasonality has a large effect on precipitation characteristics and climatologists are interested in extreme precipitation for all seasons; therefore, we analyze each season's precipitation separately. For this work, we focus on the winter season and we analyze the seasonal maximum daily precipitation amounts produced by the RCMs. An alternative to separately analyzing each season would be to incorporate seasonality as a covariate, however this would add another level of complexity which we choose not to address in this work. We define the winter season as the months of December, January and February, and retain the daily maximum precipitation amount at each location during this 90-day period. Because the RCM runs are only 20 years in length, we have only 20 observations at each location with which to characterize the upper tail. We analyze six separate data sets of seasonal maxima, one for each RCM. Although all of the results found in Section 4 are for the winter season only, similar analyses will be done for each 3-month season as part of the NARCCAP project.

3 IAR-based spatial Bayesian hierarchical model

Our focus is in modeling the marginal behavior of the extreme precipitation at each location. We construct a hierarchical model to analyze seasonal maxima in order to compare the six RCMs. The hierarchical model has three levels: data, process, and prior. The borrowing of strength comes from a spatial component in the process level of the statistical model.

3.1 Data level

The GEV distribution comprises the foundation of our hierarchical model. Let d be the number of cells in our region, $D = \{1, \dots, d\}$, and let j in D be the index of the cell. The spatial domain of the NARCCAP region is a 98 by 120 grid, thus, $d = 11760$. Further, let Z_{ijt} be the random variable representing the maximum precipitation value simulated for the winter season by RCM run i , (where $i = 1, 2, \dots, 6$, see Table 1), at grid cell j , during the year t , where $t = (1, \dots, 20)$.

We assume the Z_{ijt} follow the GEV distribution with parameters μ_{ij} , σ_{ij} , and ξ_{ij} , that is:

$$\mathbb{P}(Z_{ijt} \leq z) = \exp \left[- \left(1 + \xi_{ij} \frac{z - \mu_{ij}}{\sigma_{ij}} \right)^{-1/\xi_{ij}} \right],$$

for values of z such that $(1 + \xi_{ij} \frac{z - \mu_{ij}}{\sigma_{ij}}) > 0$ and where μ_{ij} , σ_{ij} , and ξ_{ij} are the location, scale, and shape parameters for RCM i and grid cell j respectively.

In our hierarchical model, it is necessary to employ a likelihood at the data level. It is also known that GEV likelihood-based estimates for return levels are often poor when the number of observations is few. Hosking et al. (1985) show that probability weighted moment (PWM) estimation outperforms maximum likelihood for small sample sizes. Coles and Dixon (1999) show the poor ML estimates of high quantiles result from poor estimation of the shape parameter ξ , and that a likelihood approach which penalized for high values of ξ performed similarly to PWM estimation. Martins and Stedinger (2000) take a similar approach and create a penalized likelihood specifically designed for precipitation data; they find the parameter estimates that maximize the posterior distribution resulting from a GEV likelihood and a prior density that incorporates well-accepted knowledge about the tail of the precipitation distribution. The Martins and Stedinger prior consists of a shifted beta density function with support on the interval $[-0.5, 0.5]$ and mean of 0.1. Since our RCM simulations are only 20 years in length, we found that it was necessary to penalize high estimates for ξ in order to obtain sensible estimates for the shape parameters in our MCMC inference (described in Section 3.4). We assess the statistical model’s sensitivity to the penalty in Section 4.2.

Let z_{ij} be a vector of the 20 years of maximum precipitation values for run i at location j , and let $z_i = (z_{i1}^T, \dots, z_{id}^T)^T$. Furthermore let $\mu_i = (\mu_{i1}, \dots, \mu_{id})$ and define σ_i and ξ_i analogously. We assume that z_{ijt} given μ_{ij} , σ_{ij} , and ξ_{ij} is

conditionally independent of z_{ikt} given μ_{ik} , σ_{ik} , and ξ_{ik} for $j \neq k$. This conditional independence assumption is common in hierarchical modeling (e.g. Banerjee et al. 2004) and this assumption is discussed in Section 5. The penalized likelihood function we employ at the data level of the model to compare the six RCMs is:

$$\begin{aligned} \pi [z_i | (\mu_{ij}, \sigma_{ij}, \xi_{ij})] &= K \prod_{j=1}^d \prod_{t=1}^{20} \exp \left\{ - \left[1 + \xi_{ij} \left(\frac{z_{ijt} - \mu_{ij}}{\sigma_{ij}} \right) \right]^{-1/\xi_{ij}} \right\} \\ &\times \frac{1}{\sigma_{ij}} \left[1 + \xi_{ij} \left(\frac{z_{ijt} - \mu_{ij}}{\sigma_{ij}} \right) \right]^{-1/\xi_{ij}-1} \\ &\times \frac{\Gamma(15)}{\Gamma(9)\Gamma(6)} (.5 + \xi_{ij})^8 (.5 - \xi_{ij})^5. \end{aligned} \tag{1}$$

The K in Eq. 1 comes from the fact that the likelihood does not result from a proper density due to the incorporation of the Martins and Stedinger prior. Although K is not known, this is unimportant as this normalizing constant disappears when drawing inference via MCMC.

3.2 Process level

The spatial component of the statistical model occurs at the process level. We employ a multivariate spatial random effect which introduces dependence between the GEV parameters at each location. As noted in Section 1.2, we spatially model all three GEV parameters because our study region is large. We assume

$$\begin{aligned} \mu_{ij} &\sim N(\mathbf{X}_j^T \boldsymbol{\beta}_{i\mu} + U_{ij\mu}, 1/\tau_\mu^2) \\ \log(\sigma_{ij}) &\sim N(\mathbf{X}_j^T \boldsymbol{\beta}_{i\sigma} + U_{ij\sigma}, 1/\tau_\sigma^2) \\ \xi_{ij} &\sim N(\mathbf{X}_j^T \boldsymbol{\beta}_{i\xi} + U_{ij\xi}, 1/\tau_\xi^2) \end{aligned} \tag{2}$$

where \mathbf{X}_j is the vector of covariate information for location j , $\boldsymbol{\beta}_{i\theta}$ is a vector of regression coefficients for RCM i , τ_θ^2 is a fixed precision common to all RCMs, $\mathbf{U}_i = (\mathbf{U}_{i\mu}, \mathbf{U}_{i\sigma}, \mathbf{U}_{i\xi})^T$ is a spatial random effect where $\mathbf{U}_{i\theta} = (U_{i1\theta}, \dots, U_{id\theta})$, and θ is used generically to stand in for μ, σ , and ξ . For this work, the covariate information \mathbf{X}_j consists of latitude and longitude at location j . The values for $(\tau_\mu, \tau_\sigma, \tau_\xi) = (4, 200, 2000)$. These values reflect the difference in the scales of the three GEV parameters, but were also chosen so that most of the variability would have to be explained by the spatial effect \mathbf{U}_i .

The relationship between the three GEV parameters at different locations occurs via \mathbf{U}_i . We employ a multivariate intrinsic autoregressive (IAR) model (Banerjee et al. 2004, Sec. 3.3) as our spatial model. This model was chosen because it is computationally feasible for the size of the spatial region and because it is a simple way to borrow strength across locations. Schliep (2009) employed a geostatistical model with a Matérn covariance structure (Schabenberger and Gotway 2005, p. 143)

to model a subregion of the NARCCAP domain, but found that the model was too computationally challenging for the entire domain.

The IAR model is an improper formulation of the spatial conditional autoregressive (CAR) model, Schabenberger and Gotway (2005, Sec. 6.2.2.2). A CAR model begins with specified conditional distributions at each location and from these constructs a valid joint distribution. Usually one assumes a Gaussian formulation, and additionally one assumes the conditional relationships have a Markov property, in which case the resulting CAR is a Gaussian Markov random field (GMRF, Rue and Held 2005). GMRFs are most easily described via their precision matrices Q where $Q = \Sigma^{-1}$, and Σ is the covariance matrix of the model. Q is often a function an adjacency or proximity matrix W , which is determined by the spatial relationships in the studied locations. Because their conditional relationships are explicit, CARs are particularly well suited to Gibbs sampling methods.

Here, we illustrate construction of a simple one-dimensional zero-mean CAR model following (Banerjee et al. 2004, pp. 79–80). Assume the conditional relationship is given by $[Y_j|y_k, k \neq j] \sim N(\sum_{k=1}^d b_{jk}y_k, 1/t_j^2)$. Here, b_{jk} are the spatial dependence parameters ($b_{jj} = 0$), and t_j^2 is the conditional precision. Then the joint distribution of $Y = (Y_1, \dots, Y_d)^T$ is normal with mean zero and precision matrix $Q = D(I - B)$ where D is diagonal with $d_{jj} = t_j^2$, and B is the matrix $\{b_{jk}\}$. Given a symmetric proximity matrix W with non-negative elements, if we let $b_{jk} = \rho w_{jk} (\sum_{k=1}^d w_{jk})^{-1}$ and $t_j^2 = t^2 \sum_{k=1}^d w_{jk}$ where ρ and t^2 are global dependence and scale parameters, it can be shown that $Q = t^2(D_W - \rho W)$ where D_W is diagonal with $d_{W jj} = \sum_{k=1}^d w_{jk}$. Further, it can be shown that Q is positive definite if and only if $|\rho| < 1$.

If one assumes a CAR model as the likelihood for the data and allows $\rho = 1$, it is easy to see that $t^2(D_W - \rho W)\mathbf{1} = \mathbf{0}$, and the model is improper. However, such a model can be employed as a prior in a Bayesian setting and can lead to a proper posterior distribution. This is the so-called IAR model. An advantage of this IAR model is that there is no dependence parameter to estimate. For this project, we view the IAR model as a very simple spatial model which achieves our desired goal of borrowing strength across locations.

Our multivariate IAR is constructed by employing a first order proximity matrix; that is, $w_{jk} = 1$ if cells j and k share a border and is 0 otherwise. We let $Q_1 = D_W - W$ where D_W is defined as above and the subscript 1 denotes that Q_1 is univariate. To achieve a multivariate model for $(U_{i\mu}, U_{i\sigma}, U_{i\xi})$ we assume separability, thus $Q = T \otimes Q_1$. T here is a 3×3 precision matrix which contains the dependence information between $(U_{i\mu}, U_{i\sigma}, U_{i\xi})$. The matrix Q can be very large; for our example, the dimension of Q is 35280×35280 .

3.3 Prior level

For computational purposes, conjugate prior distributions are used in our work for the parameters β and matrix T . The prior for each intercept term $\beta_{i\theta 0}$ (where $\theta = \mu, \sigma,$ or ξ) is an independent Gaussian distribution with mean set to the mean of the cell-wise maximum likelihood estimate of the parameter and variance 100. The regression

coefficients $\beta_{i\theta 1}$ and $\beta_{i\theta 2}$ have a zero-mean Gaussian with variance 10. A Wishart prior with 3 degrees of freedom (the minimum allowed) is assigned to the precision matrix T . The mean is given by the diagonal matrix with main diagonal $(.02, 4, 40)^T$. These values reflect the variability found in each location's MLE parameter estimates during preliminary analysis. This empirical Bayesian approach was used as we have little prior knowledge about the variability inherent in the parameter estimates for the GEV parameters for RCM data, but these values simply reflect the different scales in the precision of these estimates.

3.4 MCMC algorithm

The model is implemented using a Gibbs sampler (Robert and Casella 1999, ch. 7). The parameters of the GEV distribution will be drawn using a Metropolis-Hastings step. Starting values for $(\mu_{ij}, \sigma_{ij}, \xi_{ij})$ are the corresponding maximum likelihood estimates for RCM i at location j plus some random error noise. Candidates for the three parameters, $(\mu_{ij}, \sigma_{ij}, \xi_{ij})$ are drawn in a block for each location j using a uniform random walk.

Instead of applying the spatial model directly to the GEV parameters, in Eq. 2 the spatial process is modeled as a random effect in the mean of the normally-distributed parameters. This allows the entire parameter vector U , which is a normal-IAR likelihood-prior combination, to be drawn at once. The vector can be corrected to meet the linear constraints of the IAR model using a method in Rue and Held (2005, pp. 89–90). The Kronecker form of Q and the fact that W is a very sparse matrix allow for very quick matrix operations. As Q is already defined as a precision matrix, no large matrix inversions are necessary. We employ the R package `spam` for sparse matrix operations (Furrer 2008; Furrer and Sain 2008).

The other parameters can also be drawn directly from their respective distributions. The conjugate relationship standard normal-normal is used to draw the β 's. The T^{-1} matrix can be drawn directly using a result from Banerjee et al. (2000) which shows the conjugate relationship of an inverse-Wishart prior in a Kronecker product.

4 Results of comparison of RCM winter extreme precipitation

4.1 Parameter and return level estimates

For each of the six RCM output sets, the statistical model was run for 15000 MCMC iterations and the first 5000 were discarded to allow for burn-in. To reduce dependence, each parameter was retained only every 10th iteration leaving us with a sample of $n = 1000$ draws from the posterior distribution of each of the model parameters. A chain of 5000 iterations takes a bit less than 12 h to run.

Four parallel chains were run to assess convergence. Convergence was assessed by viewing trace plots of all of the higher level parameters β , T and randomly selected locations for the GEV parameters μ_{ij} , σ_{ij} , ξ_{ij} and spatial random effects $U_{ij\mu}$, $U_{ij\sigma}$, $U_{ij\xi}$. Not surprisingly, trace plots showed rapid convergence of the GEV

parameters, and more slow convergence for the higher-level parameters, but all appeared to have converged by the 5000th iteration.

Model fit was assessed using quantile-quantile plots where the 20 observed seasonal maximum measurements at a location were plotted versus the quantiles of the GEV distribution with the posterior mean values used for μ , σ , and ξ . Figure 1 shows the qq-plot at four randomly selected locations, and at each location the models appear to fit the RCM output well. Note that at the Pacific Coast location, the difference between the RCMs is quite noticeable, for instance the HRM3 model's seasonal maxima are all lower than those of the other RCMs.

Our MCMC provides us with draws from the posterior distributions of μ , σ , ξ , U , β , and T for each of the six RCMs. Maps of the means and variances of the posterior distributions for μ , σ , and ξ can be found at www.stat.colostate.edu/~cooleyD/Papers/additionalFig.pdf.

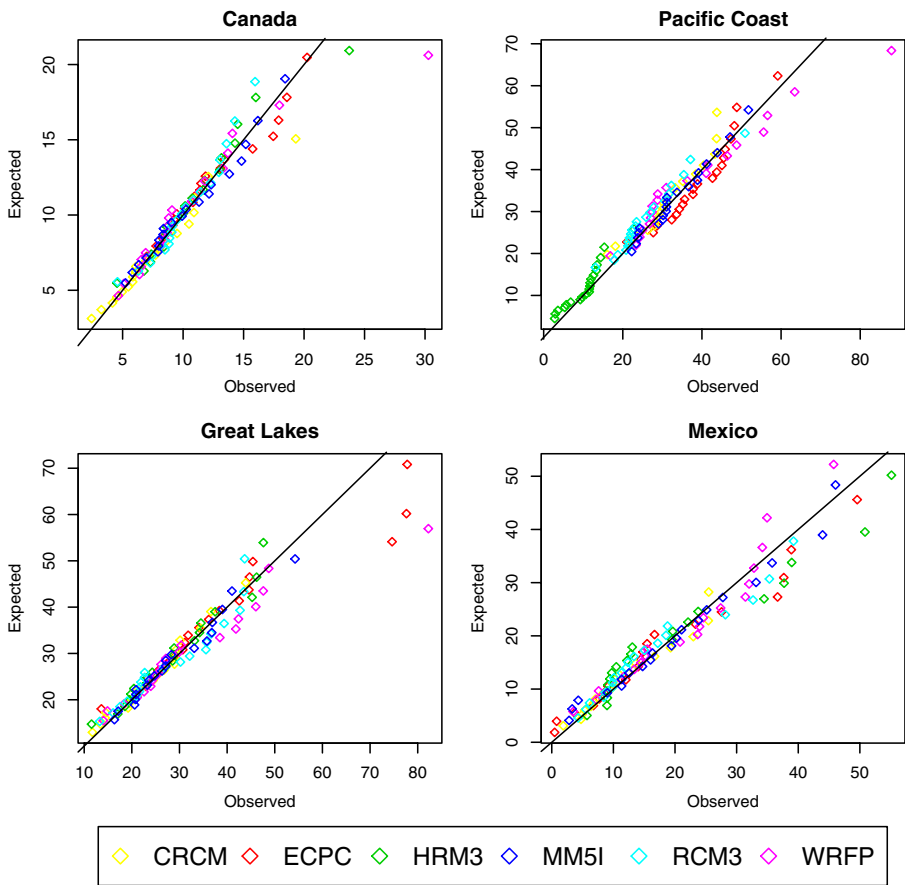


Fig. 1 Quantile-quantile plots for four locations within our region. The values are given in mm/day and the scales vary based on the precipitation values at each location.

We summarize our findings by examining the 100-year return levels, that is, the $(1 - 1/100)$ th quantile of the annual maximum precipitation event at each location. Since we only analyze winter precipitation in this work, a 100-year return level is the maximum amount of precipitation we expect to observe once every 100 years during the winter months. The return level is simply a function of the GEV parameters. As we have draws $(\mu_{ijs}, \sigma_{ijs}, \xi_{ijs})$ from the posterior distribution of each of these for RCM $i = 1, \dots, 6$, location $j = 1, \dots, d$, and iteration $s = 1, \dots, n$, we can obtain draws from the posterior distribution of the 100 year return level:

$$r_{ijs} = \mu_{ijs} + \frac{\sigma_{ijs}}{\xi_{ijs}} [(-\log(1 - 1/100))^{-\xi_{ijs}} - 1].$$

In providing 100-year return level estimates, notice we extrapolate beyond the range of the data as we have only 20 years of precipitation values. Of course, extrapolation increases the amount of uncertainty of these estimates.

Figure 2 shows the posterior mean of the 100 year winter precipitation event for the six RCMs. The return values are reported in units mm/day. All six RCMs show higher return levels along the Pacific Coast and also in the southeast United States and all show lower values for the 100 year return level for central Canada and the north-central United States. Additionally, all show isolated areas of higher return levels in the Rocky Mountain Region.

The posterior standard deviations of the 100 year precipitation events for the winter season for the six RCMs are shown in Fig. 3. These standard deviations are not

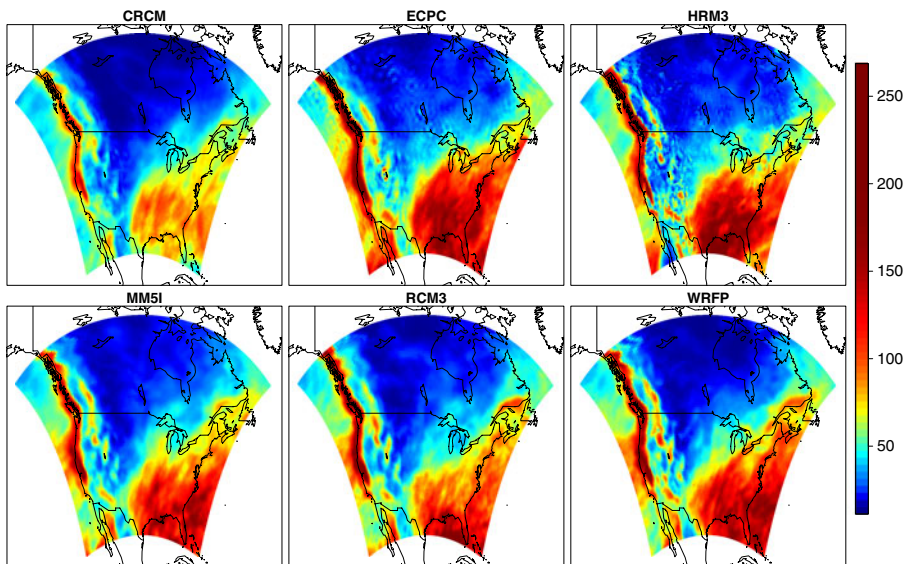


Fig. 2 Mean of the posterior distribution of 100 year return levels for daily winter precipitation (in mm) for each of the six RCMs. All show a pattern of higher return levels for the Pacific Northwest and the Southeast United States and a much lower 100 year return level for central Canada. Precipitation is given in mm/day.

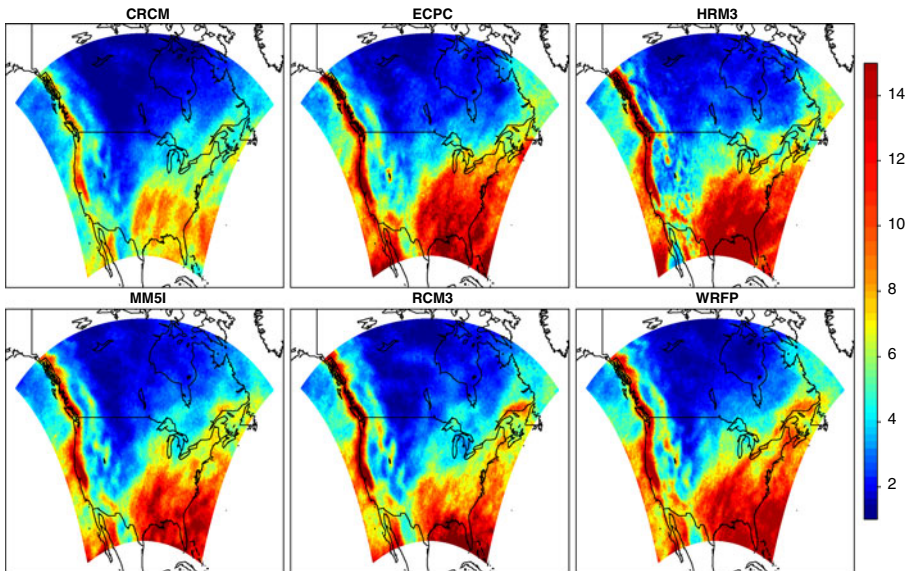


Fig. 3 Standard deviations of the posterior distribution of 100 year return levels for daily winter precipitation for each of the six RCMs. Locations of high variation correspond to locations that show higher mean precipitations in Fig. 2. Values were cut off at 15 to better contrast the lower values.

obtained from asymptotic theory but rather directly from the posterior draws of the return level distributions. Each of the six RCMs reports larger standard deviations for locations with greater posterior means as seen in Fig. 2. Therefore the underlying patterns of the 100 year return levels for precipitation are similar for the posterior means and standard deviations.

Despite the general similarities in patterns, it is clear that each RCM tells a slightly different story about extreme precipitation. The RCMs differ as to their return level point estimates; for instance, the CRCM model estimates the 100 year return level for the southeast US to be in the range of 75–120 mm, whereas the HRM3 model has estimates for the same region of over 150 mm. The RCMs also differ as to where they set their boundaries for the increased return levels for the Southeast United States; the ECPC model extends the increased levels up to include the state of Michigan and the Great Lakes while most of the other RCMs do not show higher levels this far north. The RCMs differ considerably about the level of extreme precipitation for the west coast of Mexico and the Pacific Ocean in the extreme southwest of the study region. The isolated areas in the Rockies are a bit more broadly defined in the WRFP model than in the MM5I or the CRCM models. Quite interestingly, the HRM3 model seems to have a less smooth surface for its estimate of the 100 year return level than do the other RCMs, even though all the RCMs are analyzed using the same spatial statistical model.

To more easily compare the differences, Fig. 4 shows the difference between each RCM's posterior mean and the grand mean calculated across RCMs (i.e., the mean of the posterior means). From these plots it is quite clear that despite the similar patterns, there are substantial differences between the return level estimates of the six RCMs.

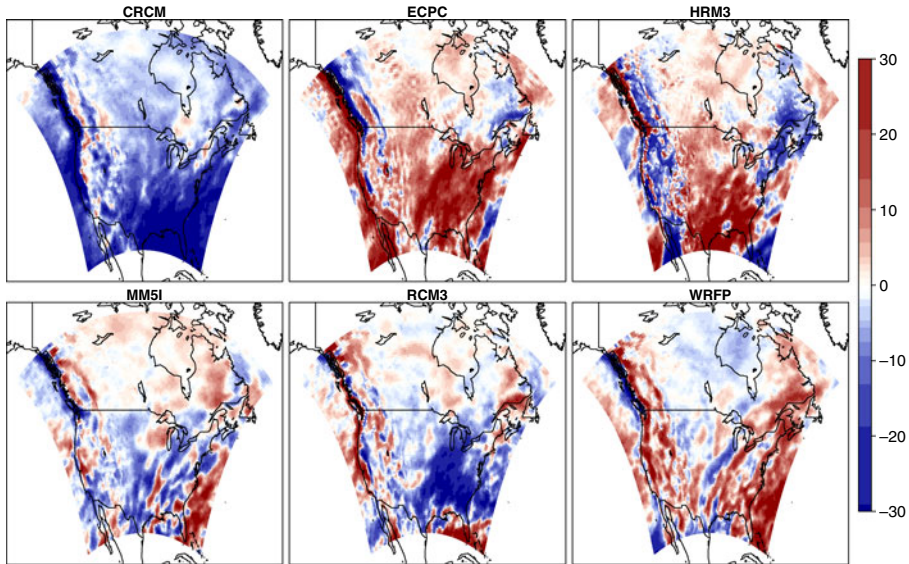


Fig. 4 Difference between the posterior mean of 100 year return levels for each model and the overall posterior mean of return levels across models. Despite the similar patterns shown in Fig. 2, this shows that there are considerable differences in the estimated return levels from model to model. The difference in values are given in mm/day. Values were cut off at -30 and 30 to better present the values near 0 .

The CRCM model appears to generally have lower estimates for the 100-year return level across most of the study region. The RCM3 model has lower estimates for the Southeast United States. The ECPC and the HRM3 models generally have higher return level estimates, particularly in the Southeast.

As Figs. 2 and 4 only give point estimates and do not quantify the uncertainty associated with the return level estimates, we wanted to assess the significance of the differences. Figure 5 gives an admittedly crude way of assessing significance. Plotted in the left panel is the ratio of between-RCM variance $s_{j, btwn}^2$ to within-RCM variance $s_{j, wi}^2$ of the posterior return level estimates. The between-RCM variance was calculated in the usual way, $s_{j, btwn}^2 = 1/5 \sum_{i=1}^6 (r_{ij} - r_{.j})^2$, where r_{ij} is the posterior mean for the i th RCM at the j th location and $r_{.j}$ is the grand mean. The within-RCM variance $s_{j, wi}^2$ is a pooled variance among the six RCMs at each location. One can view Fig. 5 as a pointwise F-statistic as would be calculated in an ANOVA test. The 0.95 quantile of a $F_{5, 5995}$ distribution is 2.22, so locations above this level might be viewed as indicating a significant difference between RCMs. One should be cautious treating this as a hypothesis test however, as one would need to account for the multiple testing issues of performing nearly 12000 ANOVA tests as well as the spatial dependence in the return level estimates. These issues are non-trivial, and we do not attempt to address them here, for reference please see Ventura et al. (2004). Nevertheless, as much of the region has a value for this ratio greater than 2.22, there is pretty strong evidence for significant differences in the return level estimates between the RCMs.

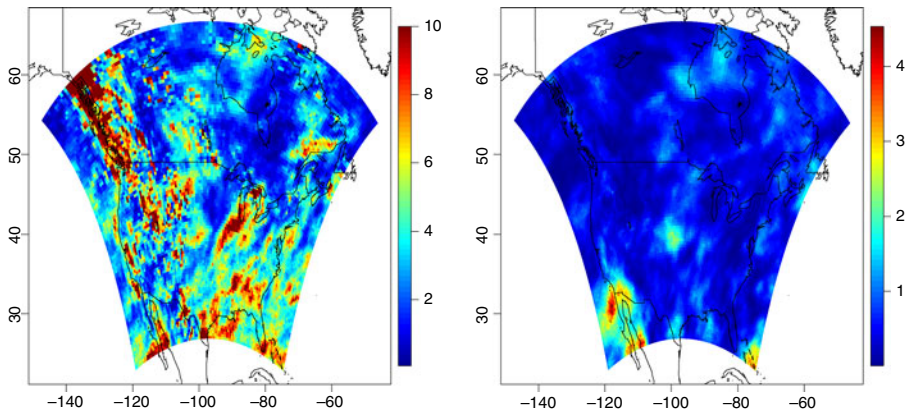


Fig. 5 Left Panel: The pointwise F-statistic for the return level comparison, the between-model variance divided by the within-model variance. The nominal 95% quantile of the $F_{5,5995}$ is 2.22. As many locations show an F-statistic larger than this value, one can see that the differences found in Fig. 4 appear to be significant, although performing an actual hypothesis test is complicated due to both multiple testing concerns and spatial dependence. Values were cut off at 10 so as to better contrast the lower values of the scale. Right panel: The pointwise F-statistic for the model-to-model comparison for ξ . Despite the differences seen in the point estimates in Fig. 6, few areas show an F-statistic greater than the 95% quantile of 2.22 due to the considerable uncertainty associated with the estimation of ξ .

Of particular interest in extremes is the shape parameter ξ . Figure 6 gives the posterior mean for this parameter for each of the six RCMs. A brief inspection of these plots reveals that aside from some general patterns of high values for the shape parameter in the southern and central areas of the study region and low values along the Pacific Coast, the patterns for the point estimate of this parameter differ considerably. For example, all six of the RCMs have an area of higher values of ξ somewhere in south-central to southeast Canada, but the location for this area with an apparent heavier tail varies in location, extent, and intensity for all RCMs. However, an inspection of the F-statistic plot for this parameter reveals that things might not differ as much as they first appear in Fig. 6. The right panel of Fig. 5 gives the same between-RCM to within-RCM variance ratio as was discussed earlier, but now for the extreme value index ξ . Notice that only in an area of Mexico and off its Pacific coast and perhaps a small area in the extreme southeast of our study region is this ratio clearly above the 0.95 quantile of 2.22. It is well known that there is considerable uncertainty when estimating the GEV shape parameter, and it appears that the differences in the estimates that appear in Fig. 6 are not so great when the uncertainty is taken into account.

Figure 7 shows what is gained by the hierarchical approach. Penalized maximum likelihood estimates were obtained for the GEV parameters μ , σ , ξ for the WRFP model output using the approach of Martins and Stedinger (2000). These estimates were used to produce maps of the 100-year return level point estimate and also the point estimate for the shape parameter ξ . Comparing these plots to Figs. 2 and 6 shows that the point-by-point approach yields estimates that are less smooth, particularly for ξ . Comparison of the scales also shows a difference. The return level point estimates now exceed 300 mm/day in some locations, whereas the maximum under

the hierarchical model was around 250 mm/day. This is likely due to a localized extreme event which occurred during the 20 year simulation, causing the point-by-point method to over-estimate the 100 year return level at these locations. Likewise, the point estimates for ξ take range of values from $(-0.1, 0.3)$ which is much larger than the range given by the hierarchical approach. Typically extreme value index values of $(0, 0.2)$ are found for precipitation measurements, but it is not unusual to have poor estimates of the tail parameter (even using the penalized likelihood approach) with such short data records.

Finally, we summarize information about the parameter estimates for β and T which appear in the process level of the statistical model. We found that the estimates for the β parameters were consistent between the different RCMs. The 95% credible intervals for the $\beta_\mu, \beta_\sigma,$ and β_ξ corresponding to longitude and latitude did not include zero indicating that these coefficients were significant. However, it is clear from the point estimate maps for $\mu, \sigma,$ and ξ that a simple trend surface would be inadequate to explain the complex spatial behavior of these parameters and that the spatial random effects $U_\mu, U_\sigma,$ and U_ξ are necessary.

The estimates for T were also quite consistent across RCMs. Since T is a precision matrix, it is difficult to interpret, so we invert and standardize to create a correlation matrix R . For the WRFP model,

$$\hat{R}_{WRFP} = \begin{bmatrix} 1.00 & 0.73 & -0.19 \\ 0.73 & 1.00 & -0.11 \\ -0.19 & -0.11 & 1.00 \end{bmatrix}.$$

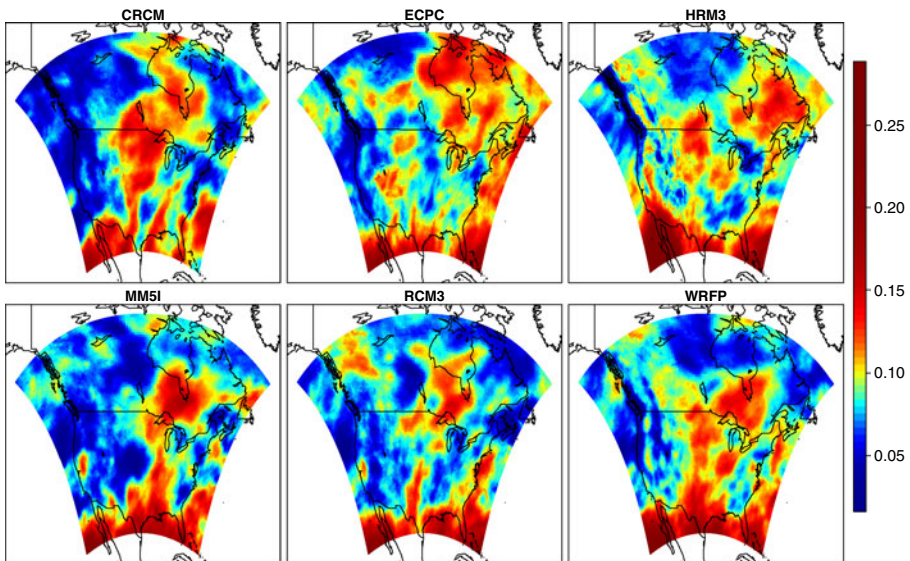


Fig. 6 Mean of the posterior distribution for the extreme value index ξ . While there are some general similarities in the point estimates such as heavier tails in the center of the continent, there is considerable variability in the patterns of the point estimates from model to model.

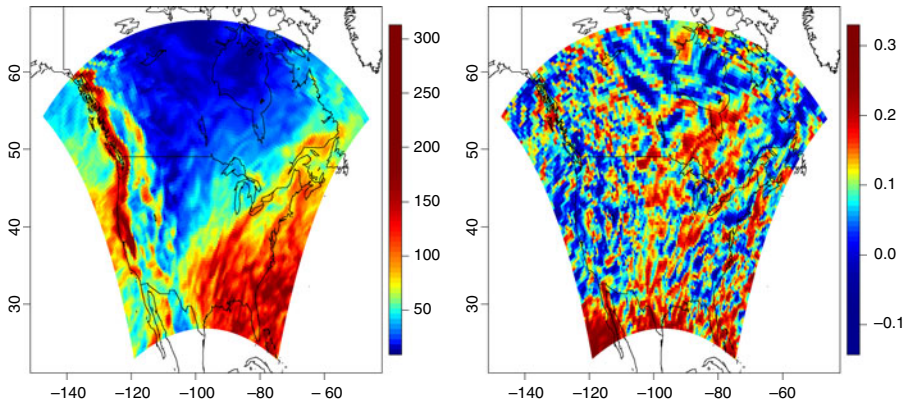


Fig. 7 Left panel shows the estimated 100-year return level for the WRF model as obtained via maximum likelihood estimates for μ , σ , ξ done via a point-by-point analysis. Right panel shows the estimated shape parameter ξ as obtained via penalized maximum likelihood using the Stedinger prior, also done via a point-by-point analysis. In addition to being less smooth, both show an increased range of values compared to Figs. 2 and 6.

In extreme value analyses, it is typical to find that the scale and shape parameters are negatively correlated, and we see a negative correlation between the vectors U_σ and U_ξ . Additionally, we see a strong positive correlation between U_μ and U_σ , and this results from the fact that locations that had higher estimates for μ also tended to have larger estimates for σ .

4.2 Sensitivity analysis

To assess the sensitivity to the priors and the Martins and Stedinger (2000) penalty, we ran several additional MCMC simulations on the WRF model output. In runs where we doubled the variance of the normal distributions that serve as the priors for the regression coefficients, we saw no substantive change in the posterior distributions for the β nor for the GEV parameter estimates. Likewise, in runs where the mean of the Wishart prior for the precision matrix T was halved, we saw no substantive change in its posterior distribution nor for the GEV estimates.

The assumption that is probably most suspect is the penalty term applied to the shape parameter in the data level of the model. Martins and Stedinger (2000) propose the penalty specifically for precipitation studies, as most studies find the shape parameter (as it is parameterized here) as slightly positive, and usually taking on values in the range of $[0.0, 0.2]$. Taken in this context, having (in effect) a beta prior with support on $[-0.5, 0.5]$ does not seem overly informative. However, one should keep in mind that we are not analyzing observed precipitation data, but RCM output of precipitation.

To assess sensitivity to the penalty term, we replaced the Martins and Stedinger penalty with three alternatives: (1) no penalty term, (2) a uniform distribution on $[-0.5, 0.5]$, and (3) a normal distribution with mean 0.1 and variance 0.16. Maps,

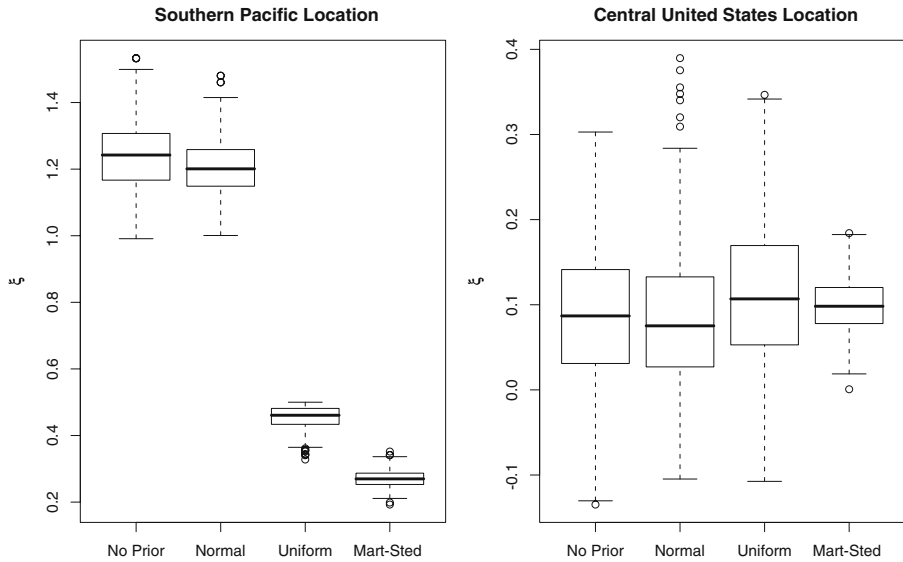


Fig. 8 Describes the sensitivity of ξ to the (Martins and Stedinger 2000) penalty. The left plot summarizes the posterior distributions for a location in the Pacific Ocean near the extreme southwest corner of the study region. Here the penalty greatly affects the point estimate for ξ with less informative priors yielding estimates for ξ greater than 1. The right plot shows a location where the point estimate is largely unaffected, but its uncertainty is greatly reduced by using the Martins and Stedinger penalty.

similar to Fig. 6, were produced for the posterior mean of ξ for each of these runs.¹ For the most part, the point estimates were very similar, with the notable exception that for both the runs with no penalty and with a normal-distribution penalty, point estimates for ξ were much higher in the extreme southwest portion of the study region which corresponds Pacific locations west of Baja California. In some of these locations, the shape parameter is estimated to be greater than 1, implying that that distribution would not have a finite mean. Whether such an estimate is sensible, even for RCM output, is questionable. Figure 8 gives boxplots summarizing the posterior distribution of ξ at two locations: one off the coast of Baja California and the other for a location in Eastern Nebraska. The figure clearly shows the effect of the different penalties applied to the model. For the Nebraska location, although the point estimate for ξ is largely unaffected, the uncertainty associated with this estimate is reduced by applying the Martins and Stedinger penalty. Whether this reduction is appropriate is also questionable.

It is not surprising that the penalty has a significant affect on the posterior distribution of ξ (and consequently the return level), especially given that we have only 20 seasonal maxima with which to estimate this difficult-to-estimate parameter. Other hierarchical model studies in extremes (e.g., Cooley et al. 2007) have found that for data with longer records, no penalty is needed to obtain sensible estimates. Although ξ is obviously sensitive to the type of penalty imposed, we do not feel that this affects

¹ www.stat.colostate.edu/~cooleyd/Papers/additionalFig.pdf

our conclusions in Section 4.1 regarding the differences found between the extreme precipitation as modeled by the six RCMs

In summary, given a data record of short length, it may be necessary to incorporate a penalty to obtain sensible estimates for ξ when employing a likelihood method. However, the researcher should recognize that this a priori information will have a large effect on estimation of the shape parameter.

5 Conclusions and discussion

There are two natural audiences for this work. The first are statisticians. This work occurs at an interesting intersection of extreme value analysis, spatial statistics, and computational methods. Although this project is applied in the sense that it was motivated by a particular data set, the methods developed herein are novel. Furthermore, statisticians working in extremes are rarely given multiple data sets that are so closely related as the data from the six RCMs that are analyzed in this work. These data gives us a unique opportunity to compare estimates of the extreme value index ξ and return levels and assess, in some sense, their sensitivity.

The other audience for this work are climate modelers. As shown in Fig. 2, the characteristic pattern of the winter extreme precipitation produced by the six RCMs is remarkably consistent. However, as shown in Figs. 4 and 5, there appear to be significant differences between the RCMs in how they characterize extreme precipitation. The differences in return level estimates in Fig. 4 was not known or expected. In future work, we intend to apply the model to the other three seasons for these RCM runs and make similar comparisons. As the NARCCAP project continues and RCM runs are produced which are driven by different GCMs as opposed to reanalysis data, it will be important to characterize which effects are due to the GCMs and which are due to the RCMs.

Extreme events are likely to impact more than an isolated location, making the conditional independence assumption in the data level of our model incorrect. Once the marginal effects have been accounted for, one can still find residual/local dependence in the annual maxima for a given year. This residual dependence should be accounted for in the data level of the model; however, how to incorporate dependence at this level is not clear. Ideally, our data level would consist of a likelihood based in extreme value theory which modeled the dependence at all of our locations. Parametric and non-parametric multivariate extreme models thus far have been limited to a small number of dimensions (e.g., Cooley et al. 2009) which would be inadequate for most any spatial application. Max-stable process models have been suggested (e.g., Schlather 2002) but thus far, the distribution function of such models is known in closed form only for the bivariate case, and thus a likelihood relating all locations is not yet available. Alternatively, copula approaches have been suggested by Sang and Gelfand (2009) which do allow one to model residual dependence in the data. The copula approach is promising, however new difficulties arise if one wishes to reconcile a particular copula model (e.g. Gaussian) with the theory of multivariate extremes. We and many other researchers are investigating how best to capture

dependence in the data level of a hierarchical model for extremes. Although our model here does not account for this residual dependence, our aim is to only characterize the marginal effects. We believe our model is more than adequate to compare the extreme precipitation from these six RCMs.

A natural extension of this work would be to combine the multiple runs into one model with a multivariate response, rather than model them separately as we do here. Doing so might allow more detailed comparison between RCMs. This of course would bring new computational challenges. To address these, one approach might be to employ an additional matrix (6×6 in this case) and apply it via a Kronecker product to our current precision matrix Q . It is unclear if simultaneous modeling warrants the increased complexity of such a model. One wonders if a simple parallel computing strategy could be employed as the draws of the parameters $(\mu_{ij}, \sigma_{ij}, \xi_{ij})$ which are the current bottleneck in our Gibbs sampler might be done at the same time for the different model runs.

Acknowledgments We wish to thank the North American Regional Climate Change Assessment Program (NARCCAP) for providing the data used in this paper. NARCCAP is funded by the National Science Foundation (NSF), the U.S. Department of Energy (DoE), the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Environmental Protection Agency Office of Research and Development (EPA). The research of Erin Schliep was supported, in part, by the Department of Statistics at Colorado State University and through a grant from the Weather and Climate Impact Assessment Science Program at the National Center for Atmospheric Research. Dan Cooley's research is supported, in part, through NSF grant DMS-0905315. The research of Stephan Sain was supported, in part, through NSF grants ATM-0502977, ATM-0534173, and DMS-0707069. The research of Jennifer Hoeting was supported through NSF grant DEB-0717367. The National Center for Atmospheric Research is sponsored by the NSF.

References

- Banerjee, S., Gelfand, A.E., Polasek, W.: Geostatistical modelling of spatial interaction data with application to postal service performance. *J. Stat. Plan. Inference* **90**, 87–105 (2000)
- Banerjee, S., Carlin, B., Gelfand, A.: *Hierarchical Modeling and Analysis for Spatial Data*. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC, Boca Raton, FL (2004)
- Casson, E., Coles, S.: Spatial regression models for extremes. *Extremes* **1**, 449–468 (1999)
- CCSP: CCSP synthesis and assessment product 3-3, 2008: weather and climate extremes in a changing climate. Regions of focus: North America, Hawaii, Caribbean, and U.S. Pacific Islands. A Report by U.S. Climate Change Science Program and the Subcommittee on Global Change Research, Department of Commerce, NOAA's National Climatic Data Center, Washington, DC. <http://www.climate-science.gov/Library/sap/sap3-3/final-report/> (2008)
- Coles, S., Dixon, M.: Likelihood-based inference for extreme value models. *Extremes* **2**, 523 (1999)
- Cooley, D., Sain, S.R.: Spatial hierarchical modeling of precipitation extremes from a regional climate model. *JABES*. <http://www.stat.colostate.edu/~cooleyd/Papers/rcmPaper.pdf> (2008)
- Cooley, D., Nychka, D., Naveau, P.: Bayesian spatial modeling of extreme precipitation return levels. *J. Am. Stat. Assoc.* **102**, 824–840 (2007)
- Cooley, D., Davis, R.A., Naveau, P.: The pairwise beta: a flexible parametric multivariate model for extremes. <http://www.stat.colostate.edu/~cooleyd/Papers/pairwiseBeta.pdf> (2009)
- Dalrymple, T.: Flood frequency analyses. Water supply paper 1543-a, U.S. Geological Survey, Reston (1960)
- Frei, C., Scholl, R., Fukutome, S., Schmidli, J., Vidale, P.L.: Future change of precipitation extremes in Europe: intercomparison of scenarios from regional climate models. *J. Geophys. Res.* **111**, D06105 (2006)
- Furrer, R.: Spam: SPArse matrix. R package version 0.14-1 (2008)

- Furrer, R., Sain, S.R.: Spam: a sparse matrix R package with emphasis on mcmc methods for Gaussian Markov random fields. *J. Stat. Softw.* http://www.image.ucar.edu/~ssain/Steve_Sain_Webpage/Publications.html (2008, to appear)
- Hosking, J.R.M., Wallis, J.R.: *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge (1997)
- Hosking, J., Wallis, J., Wood, E.: Estimation of the generalized extreme-value distribution by the method of probability-weighted-moments. *Technometrics* **27**, 251–261 (1985)
- IPCC: *Climate Change 2007: Synthesis Report*. IPCC, Geneva. <http://www.ipcc.ch/index.htm> (2007a)
- IPCC: *IPCC Fourth Assessment Report: Working Group I Report, The Physical Science Basis*. IPCC, Geneva. <http://www.ipcc.ch/index.htm> (2007b)
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J.J., Fiorino, M., Potter, G.L.: NCEP/DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83**, 1631–1643 (2002)
- Kharin, V.V., Zwiers, F.W., Zhang, X.B., Hegerl, G.C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate* **20**, 1419–1444 (2007)
- Martins, E., Stedinger, J.: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.* **36**, 737–744 (2000)
- Mearns, L., Gutowski, W., Jones, R., Leung, R., McGinnis, S., Nunes, A., Qian, Y.: A regional climate change assessment program for North America. *EOS, Trans. Am. Geophys. Union* **90**, 311 (2009). doi: [10.1029/2009EO360002](https://doi.org/10.1029/2009EO360002)
- Nakicenovic, N., Swart, R.: *IPCC Special Report on Emission Scenarios*. Cambridge (2000)
- Robert, C.P., Casella, G.: *Monte Carlo Statistical Methods*. Springer, New York (1999)
- Rue, H., Held, L.: *Gaussian Markov Random Fields: Theory and Applications*. Monographs on Statistics and Applied Probability. Chapman and Hall, Boca Raton (2005)
- Sang, H., Gelfand, A.E.: Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* **16**, 407–426 (2009)
- Schabenberger, O., Gotway, C.A.: *Statistical Methods for Spatial Data Analysis*. Texts in Statistical Science. Chapman and Hall/CRC, Boca Raton (2005)
- Schlather, M.: Models for stationary max-stable random fields. *Extremes* **5**(1), 33–44 (2002)
- Schliep, E.: *Spatial hierarchical models to compare regional climate model output*. Master's thesis, Department of Statistics, Colorado State University, Fort Collins (2009)
- Ventura, V., Paciorek, C., Risbey, J.: Controlling the proportion of falsely-rejected hypotheses when conducting multiple tests with climatological data. *J. Climate* **17**, 4343–4356 (2004)