

Hierarchical modeling for extreme values observed over space and time

Huiyan Sang · Alan E. Gelfand

Received: 20 December 2006 / Revised: 12 December 2007
© Springer Science+Business Media, LLC 2008

Abstract We propose a hierarchical modeling approach for explaining a collection of spatially referenced time series of extreme values. We assume that the observations follow generalized extreme value (GEV) distributions whose locations and scales are jointly spatially dependent where the dependence is captured using multivariate Markov random field models specified through coregionalization. In addition, there is temporal dependence in the locations. There are various ways to provide appropriate specifications; we consider four choices. The models can be fitted using a Markov Chain Monte Carlo (MCMC) algorithm to enable inference for parameters and to provide spatio-temporal predictions. We fit the models to a set of gridded interpolated precipitation data collected over a 50-year period for the Cape Floristic Region in South Africa, summarizing results for what appears to be the best choice of model.

Keywords Coregionalization · Generalized extreme value distribution · Markov random field · Precipitation surfaces · Spatial random effects

1 Introduction

Extreme value analysis finds wide application in areas such as environmental science (e.g., [Thompson et al. 2001](#)), financial strategy of risk management (e.g., [Dahan and Mendelson 2001](#)) and biomedical data processing ([Roberts 2000](#)). In this article, we focus on climate extremes, in particular on precipitation events. Our underlying driver is the challenging ecological problem of trying to characterize the effect of changes

H. Sang (✉) · A. E. Gelfand
Institute of Statistics and Decision Sciences, Durham, NC, USA
e-mail: hs37@stat.duke.edu

A. E. Gelfand
e-mail: alan@stat.duke.edu

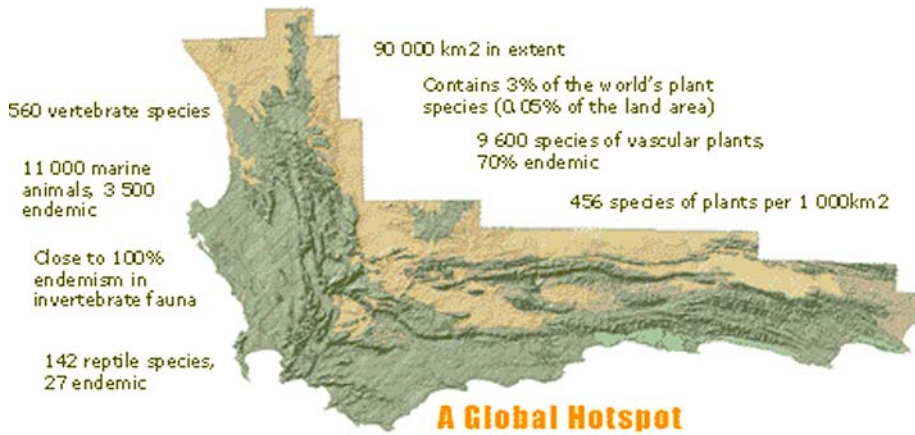


Fig. 1 Map of Cape Floristic Region in South Africa [Image from <http://www.capeaction.org.za>]

in climates on the distribution and abundance of species. Particularly, for plants, it is suggested that extreme climate events, such as drought, heavy rainfall and very high or low temperatures, might be significant factors in explaining plant performance. In fact, it is plausible that trends in climate extremes are more informative with regard to survival, growth, reproductivity, etc., than trends in the mean climate.

The motivating dataset here is derived from precipitation surfaces in the Cape Floristic Region (CFR) in South Africa. Figure 1 shows a map of the CFR in South Africa. The CFR is the smallest but, arguably, the richest of the world's six floristic kingdoms encompassing a region of roughly 90,000 km² in southwestern South Africa. The region is highly diverse and endemic across spatial scales; it includes about 9,000 plant species, 69% of which are found nowhere else. The daily precipitation surfaces we employ arise via interpolation to grid cells at 10 km resolution based on records reported by up to 3,000 stations across South Africa during the period from 1950 to 1999. Figure 2 displays the derived surface of annual maxima of daily rainfalls for 1,332 grid cells in 1999. In fact, we have 50 such surfaces dating back to 1950. The high resolution grid-aggregated rainfalls are derived from a conditional interpolation technique of Hewitson and Crane (2005) which is especially tailored for precipitation data. The technique uses a distance-decay weighting function, modified by the station-specific relation of a station to its surrounding areas. It is widely used in the meteorology community and we make no attempt here to evaluate its performance; rather, we take the output as our rainfall data.

It is generally accepted that there are, primarily, two distinct patterns generating precipitation in the CFR. There are fronts from the Atlantic Ocean affecting the western part of the region with, typically, more hitting the southwest (Capetown and vicinity), fewer hitting the northwest. There tends to be greater interannual variability in the northwest with most of the rainfall occurring in the 2–3 winter months. Similarly, most of the rainfall around Capetown occurs in the winter months. As we move east, rainfall tends to arise from Indian Ocean storms and humidity which can come at any time but are more likely in the summer. Finally, in the northeastern part of the region,

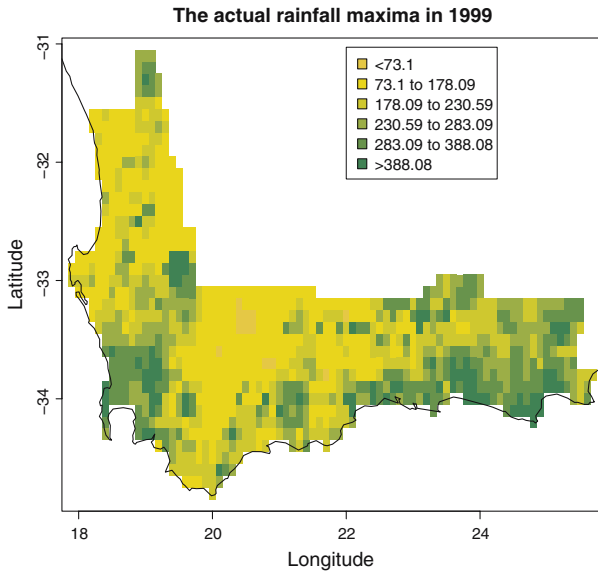


Fig. 2 Image plot of the rainfall maxima over the gridded CFR in 1999

spring and fall rain occurs and there is little seasonality in the rainfall. Figure 3a shows the week of maximum precipitation over all locations and all years, which indicates great spatial and interannual variability in the extreme rainfall occurrence. Figure 3b shows this for the western part of the CFR and Fig. 3c for the eastern part. They support the foregoing description; maximum rainfall is not confined to a specific part of the year. Finally, the climate for a grid cell is expected to depend on altitude so we introduce elevation as a covariate in the modeling. However, we find little effect on precipitation (but would anticipate a stronger effect if we were studying temperature surfaces).

By now there is an enormous literature on the modeling of extremes. At present, the standard approach utilizes generalized extreme value (GEV) distribution families. Alternatively, daily precipitation exceedances for a given threshold are modelled with Generalized Pareto Distribution (GPD). The book by Coles (2001) provides an introduction and bibliography through 2000. More recent work in the area of rainfall extremes includes Durman et al. (2001), Kharin and Zwiers (2005), Huerta and Sansó (2007) and Cooley et al. (2007).

The contribution of this paper is to develop models for rainfall that reflect dependence in space and in time. In particular, the GEV is characterized by a location, a scale and a shape parameter. Conceptually, these could all vary in space and time and they could be mutually dependent. Through exploratory data analysis we clarify the nature of this variation and then introduce it appropriately into a hierarchical model. In fact, we offer several such models, fitting them and comparing them. We present interpretation of the data analysis including the nature of spatial association and the nature of trend at grid cell resolution over the 50 year period from 1950 to 1999. In fact, we also hold out the last year of data to implement forecasting with validation studied through the hold out data. As a result, we extend the recent work of

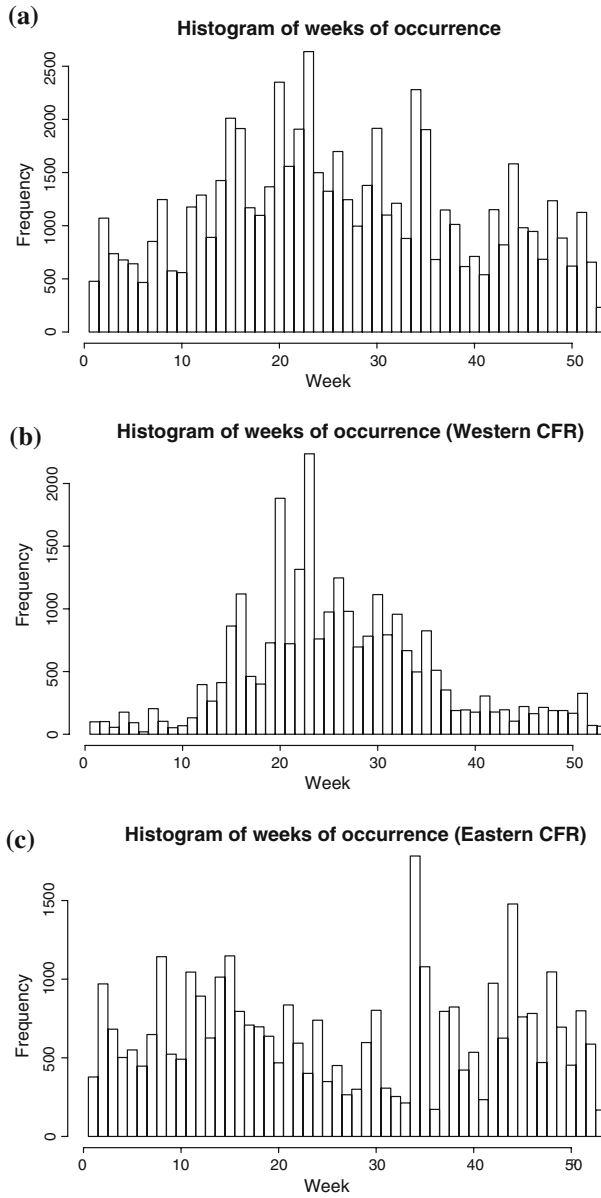


Fig. 3 (a) Histogram of weeks when annual rainfall maxima occurred in CFR. (b) Histogram of weeks when annual rainfall maxima occurred in western CFR. (c) Histogram of weeks when annual rainfall maxima occurred in eastern CFR

Cooley et al. (2007) to introduce a dynamic aspect, in particular, to model the location spatio-temporally, to introduce spatial dependence between scale and location and, through Markov random fields, to handle a considerably larger dataset.

The format of the paper is as follows. In Sect. 2 extreme value theory is proposed to describe the conditional distribution of annual maxima at individual sites leading to grid cell level modeling. In Sect. 3 we present exploratory analysis to enable us to connect the grid level models in space and time. In Sect. 4 we offer various hierarchical Bayesian models which provide appropriate connections based on the inference about latent parameter structure in Sect. 3. Model implementation, employing MCMC methods and model comparison for the annual rainfall data in the CFR are illustrated in Sect. 5. In Sect. 6 we conclude with a brief summary and suggestions for future work.

2 Review of extreme value theory

Extreme value theory begins with a sequence Y_1, Y_2, \dots of independent and identically distributed random variables and, for a given n asks about parametric models for $M_n = \max\{Y_1, \dots, Y_n\}$. If the distribution of the Y_i is specified, the exact distribution of M_n is known. In the absence of such specification, extreme value theory considers the existence of $\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq y\right) \equiv F(y)$ for two sequences of real numbers $a_n > 0, b_n$. If $F(y)$ is a non-degenerate distribution function, it belongs to either the Gumbel, the Frechet or the Weibull class of distributions, which can all be usefully expressed under the umbrella of the GEV.

$$G(y; \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \xi \left(\frac{y - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \tag{1}$$

for $y : 1 + \xi(y - \mu)/\sigma > 0$. In (1), $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and $\xi \in \mathbb{R}$ is the shape parameter.

The GEV distribution is heavy-tailed and its probability density function decreases at a slow rate when the shape parameter ξ is positive. On the other hand, the GEV distribution has a bounded upper tail for a negative shape parameter. Note that n is not specified; the GEV is viewed as an approximate distribution to model the maximum of a sufficiently long sequence of random variables. Here, we will assign daily precipitation into annual blocks and then assume the maxima are conditionally independent (but not identically distributed) across years given their respective, parametrically modelled, μ, σ and ξ . That is, each maximum is assumed to follow its own GEV.

A further remark is appropriate here. The annual blocks provide roughly 365 daily precipitation measurements. Many regions in the CFR are essentially deserts, resulting in a very high proportion of “zeroes”. (In some grid cells we see 90% zeroes.) However, it can be claimed that the GEV is still an appropriate model for the maximum in this case. The argument is merely to envision the daily observations as $Y_t = Y_t^* 1(Y_t^* > 0)$ where Y_t^* has support \mathbb{R}^1 . Then, if $\max Y_t^*$ is assumed to follow a GEV, since $\max Y_t = \max Y_t^*$ with very high probability, we can use a GEV model for $\max Y_t$.

A useful feature of (1) is the return level z_p associated with return period $1/p$, which is derived from the inverse of the GEV cumulative distribution function. Defining $x_p = -\log(1 - p)$, we have

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} [1 - x_p^{-\xi}] & \xi \neq 0 \\ \mu - \sigma \log(x_p) & \xi = 0 \end{cases} \tag{2}$$

Evidently, the return level z_p provides the threshold which is exceeded by the extreme value with probability p . Equivalently, z_p can be viewed as a threshold which is such that we expect an exceedance once every $1/p$ years.

We introduce the GEV distribution as a first stage model for annual precipitation maxima, specifying μ, σ and ξ at the second stage to reflect underlying spatio-temporal structure. In particular, let $Y_{i,t}$ denote the annual maximum of daily rainfall at location i in year t . We assume the $Y_{i,t}$ follow a GEV distribution with parameters $\mu_{i,t}, \sigma_{i,t}$ and $\xi_{i,t}$, respectively. Again, we assume the $Y_{i,t}$ are conditionally independent given their μ 's, σ 's and ξ 's. Attention focuses on specification of the model for $\mu_{i,t}, \sigma_{i,t}$ and $\xi_{i,t}$.

3 Exploratory analysis

To facilitate modeling the $\mu_{i,t}, \sigma_{i,t}$ and $\xi_{i,t}$, we conduct some exploratory analysis with two objectives—learning about temporal trend in these parameters and learning about spatial dependence in these parameters. First, it is known that estimation of ξ is challenging in GEV models (Cooley et al. 2007) and it is unlikely that we will be able to reliably discern space-time pattern in a set of $\xi_{i,t}$. Furthermore, model fitting with spatially or temporally structured $\xi_{i,t}$ will be very difficult. So, we assume ξ is unknown but constant and focus on learning about $\mu_{i,t}$ and $\sigma_{i,t}$.

Next, we illuminate the need for spatial modeling. At each grid cell, we fitted MLE's for μ_i and σ_i , treating the 50 measurements in time as independent, using the S-Plus package *isnev* (Coles 2001). We then computed customary residuals and studied the dependence in these residuals between pairs of adjacent grid cells. Since these residuals are roughly from a $GEV(0, 1, \xi)$, moments need not exist so it is not appropriate to compute product moment correlations. Instead we computed Spearman rank correlations. In fact, there are a very large number of pairs of adjacent grid cells, hence a very large number of correlations that can be computed. We created a histogram of these correlations (which is not presented in the interest of space) finding that the median of these correlations is roughly 0.6. Evidently there is strong residual spatial dependence, justifying the spatial random effects models we develop in Sect. 4.

Returning to $\mu_{i,t}$ and $\sigma_{i,t}$, we initially consider the forms

$$\mu_{i,t} = \mu_i + \alpha_{i,\mu} t \tag{3}$$

$$\log(\sigma_{i,t}) = \log(\sigma_i) + \alpha_{i,\sigma} t \tag{4}$$

In (3) and (4), μ_i and $\log \sigma_i$ can be viewed as baselines at grid cell i . The $\alpha_{i,\mu}$ and $\alpha_{i,\sigma}$ denote the site-level *trend* coefficients for the location and scale parameters.

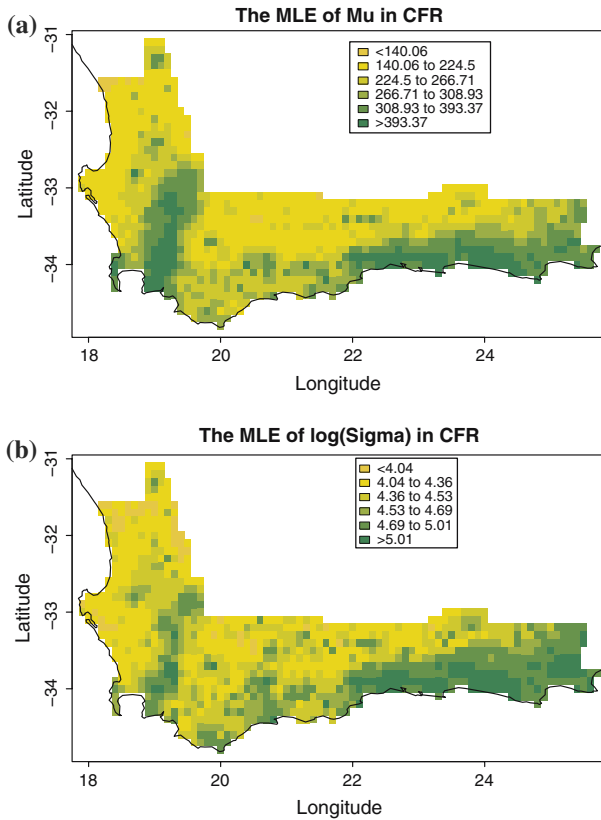


Fig. 4 MLE's of the μ_i and $\log \sigma_i$ for the exploratory models

(Nonzero) slopes inform about temporal dependence and the intercepts inform about spatial dependence. In exploratory mode, we assume independence across sites and obtain maximum likelihood estimators for these parameters along with ξ , modifying open source code for the S-Plus package ismev (Coles 2001).

In Fig. 4 we provide two maps showing the μ_i and $\log \sigma_i$. We clearly see evidence of spatial pattern in both, leading us to model the locations and scales with spatial structure. In fact, using the grid cells themselves with, say usual first order neighbor proximity, we can run Moran's I and Geary's C randomization tests for spatial association (Banerjee et al. 2004) for the μ_i and the $\log \sigma_i$. For the μ_i we obtain $I = 0.82$ ($P < 0.01$) and $C = 0.14$ ($P < 0.01$). For the $\log \sigma_i$ we obtain $I = 0.74$ ($P < 0.01$) and $C = 0.23$ ($P < 0.01$). So in both cases, there seems to be evidence of significant spatial dependence. The implication is that extreme rainfall distributions are anticipated to be similar for sites near to each other (though nothing in this analysis says that the day of occurrence of an extreme rainfall at one site coincides with the day of extreme rainfall at an adjacent site).

There is also evidence of temporal trend in the location parameters; roughly 95% of the interval estimates for $\alpha_{i,\mu}$ are significant. There is much weaker evidence of

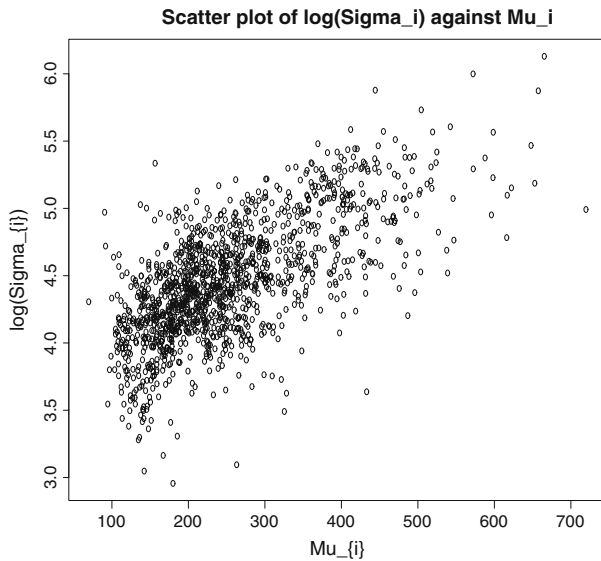


Fig. 5 Scatter plot of the estimated μ_i and $\log \sigma_i$ for the exploratory models

temporal trend in the scale parameters; far fewer of the $\alpha_{i,\sigma}$ are significant. Moreover, fitting models with both space and time specification in the scale parameters is difficult and, in any event, we view our modeling as primarily illustrative. Therefore, in the formal space–time models proposed in Sect. 4, we assume there is temporal dependence for the μ ’s but not for the σ ’s. Of course, analogous versions of models A–D below could be offered for the $\log \sigma$ ’s.

We might anticipate dependence between μ_i and $\log \sigma_i$, i.e., often the scale varies with the center. If so, we would employ a bivariate spatial model. The scatter plot in Fig. 5 for each pair of μ_i and $\log \sigma_i$ shows evidence of high correlation between these two parameters, indeed suggesting the need for modeling dependence between them. In Sect. 4 we build formal multivariate spatial models to account for the findings above.

4 Formal hierarchical space–time models

Let $Y_{i,t}$ denote the annual maximum precipitation at grid cell i at time t with conditionally independent first stage specification given by $y_{i,t} | (\mu_{i,t}, \sigma_{i,t}, \xi_{i,t}) \sim \text{GEV}(\mu_{i,t}, \sigma_{i,t}, \xi_{i,t})$. The conditional independence assumption is interpreted as interest in smoothing the precipitation surfaces around which the interpolated data is centered rather than smoothing the data surface itself. As a formal assumption, it is defensible in time since the annual maxima at a site likely occur with sufficient time between them to be assumed independent. In space, we would expect small scale dependence in the data at a given time. However, as noted in the Introduction, our observations are assigned to grid cells at 10km resolution. The exploratory analysis in the previous section

motivates modeling residual spatial dependence at this scale. However, with areal unit data at such scale, we have no tools to learn about finer scale dependence.

Following Sect. 3, we assume $\xi_{i,t} = \xi$ for all i and t , and $\sigma_{i,t} = \sigma_i$ for all t , where σ_i will be modelled using spatial random effects. Modeling interest therefore focuses on the $\mu_{i,t}$. But, in addition, we want the $\mu_{i,t}$ and σ_i to be co-varying at the same site. This requires the introduction of an association model for a collection of spatially co-varying parameters over a collection of grid cells. Gelfand and Vounatsou (2002) and Jin et al. (2005) developed models for multivariate spatial random effects with areal data. We also do this but adopt the coregionalization method (Gelfand et al. 2004) as discussed below to offer a constructive specification of the joint distribution through transformation of simpler conditional and marginal models. The gridding in the precipitation data suggests the use of conditionally autoregressive (CAR) models; they also greatly reduce the computational burden in fitting our high hierarchical models.

4.1 Modeling the $\mu_{i,t}$

As a result, we propose the specification, $p(\mu_{i,t}|\boldsymbol{\beta}, W_{i,t}, \tau^2) = N(\mathbf{X}'_i\boldsymbol{\beta} + W_{i,t}, \tau^2)$. Again, $\mu_{i,t}$ is the location parameters at location i at year t , $i = 1, 2, \dots, S$ and $t = 1, 2, \dots, T$. \mathbf{X}_i is the site-specific vector of potential explanatory variables for extreme rainfall, such as geographic coordinates using a trend surface, altitude of the sites and local temperature, with the corresponding coefficient vector $\boldsymbol{\beta}$. $W_{i,t}$ is a spatial–temporal random effect.

Possibilities for modeling $W_{i,t}$ include:

- An additive form:
 - Model A: $W_{i,t} = \psi_i + \delta_t$, $\delta_t = \phi\delta_{t-1} + \omega_t$, where $\omega_t \sim N(0, W_0^2)$ *i.i.d*
- A linear temporal component with spatial random effects:
 - Model B: $W_{i,t} = \psi_i + \rho(t - t_0)$
 - Model C: $W_{i,t} = \psi_i + (\rho + \rho_i)(t - t_0)$
- A multiplicative form in space and time:
 - Model D: $W_{i,t} = \psi_i\delta_t$, $\delta_t = \phi\delta_{t-1} + \omega_t$, where $\omega_t \sim N(0, W_0^2)$ *i.i.d*

The additive form in Model A might appear to over-simplify spatial–temporal structure. However, the data may not be rich enough to find space–time interaction in the $\mu_{i,t}$. Models B and C provide evaluations of temporal trends in terms of global and local assessments, respectively. The coefficient $\rho + \rho_i$ in Model C represents the spatial trend in location parameters, where ρ could be interpreted as the global change level in the CFR per year. Finally, Model D provides a multiplicative representation of $W_{i,t}$, similar in spirit to the recent work of Huerta and Sansó (2007). Models A and D yield special cases of a dynamic linear model (West and Harrison 1997). More general versions can be built in the $W_{i,t}$, e.g., $W_{i,t} = \phi W_{i,t-1} + \eta_{i,t}$ where the $\eta_{i,t}$ are modeled as independent over t but dependent across i . (See, e.g., Banerjee et al. 2004, Chap. 8 in this regard.) We do not follow this path here since, from Sect. 3, we want to model the dependence between location and scale parameters in the GEV model. In Models A, B, and D, we do this by specifying $\log \sigma_i$ and ψ_i to be dependent. In Model C, we specify $\log \sigma_i$, ψ_i and ρ_i to be dependent.

The nugget term $\tau^2\epsilon_{i,t}$ is introduced for computational convenience. By fixing a small value for variance τ^2 we enjoy the benefit of a Gaussian framework in sampling the parameters in the second stage specification with conditionally independent $\mu_{i,t}$. Of course, the $Y_{i,t}$'s continue to be conditionally independent as above. Details of the sampling schemes for MCMC model fitting are provided in Appendix.

Again, we will have grid cell level scale parameters, σ_i . Following the findings of Sect. 3, we need to jointly model $\log \sigma_i$ and the ψ_i (regardless of which of the above models we choose). This is the topic of the next subsection.

4.2 The coregionalized CAR model

Recall the univariate CAR model (Besag 1974), frequently used to model spatial random effects (Banerjee et al. 2004). Suppose $V = (V_1, V_2, \dots, V_S)'$ is a vector of spatial random effects which is defined at areal sites from 1 to S . The joint distribution of V is defined through conditional Gaussian specifications at each site, $p(V_i|V_j, j \neq i) = N(\sum_j w_{i,j}/w_{i+}V_j, \lambda/w_{i+})$ where λ is a scale parameter. The CAR is an improper distribution. We use the usual neighbor-based proximities, $w_{i,j} = 1$ if i and j share any common boundary, $w_{i,j} = 0$ otherwise.

When we have multivariate spatial random effects over the same region that have dependence among themselves as well as spatial dependence across sites, a multivariate CAR model is then desired. Gelfand and Vounatsou (2002) and Jin et al. (2005) developed models for multivariate areal data by extending multivariate CAR ideas building on the work of Mardia (1988). Here, we propose a coregionalization model in the spirit of Gelfand et al. (2004).

Let $U_i = (U_{1,i}, U_{2,i}, \dots, U_{P,i})'$ denote the $P \times 1$ vector of spatial random effects for the study region. Introducing a lower triangular transformation matrix A with elements $a_{i,j}$, let $U_i = AV_i$, where $V_i = (V_{1,i}, V_{2,i}, \dots, V_{P,i})'$. Denote $V^j = (V_{j,1}, V_{j,2}, \dots, V_{j,S})$, for $j = 1, \dots, P$. For the V^j , we assume independent univariate CAR models as we mentioned above. The CAR models can have scale parameter 1 because the entries in A provide the scaling. We further assume that proximity matrix W remains constant for the P univariate CAR models. Also, each V^j has an improper distribution, hence the distribution of the U^j is improper. However, as with customary univariate CAR priors, we introduce centering (on the fly, i.e., after each model-fitting iteration, Besag et al. 1995). This converts the improper CAR specification to a proper one while still enabling us to take advantage of the convenient full conditional distributions associated with the CAR model. Now, with proper priors, we are ensured that the posterior will be proper.

We work with the log transformation of σ_i , i.e., $\sigma_i = \sigma_0 \exp \lambda_i$. λ_i is now centered at 0. The coregionalization CAR model is used for λ_i along with ψ in Models A, B and D, i.e., $(\lambda_i, \psi_i)' = A(V_{1,i}, V_{2,i})'$, where

$$A = \begin{pmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{pmatrix}$$

and V_1 and V_2 are two independent univariate CAR models.

For Model C, ρ_i in the temporal linear component is spatially varying. So here we extend the multivariate CAR model to incorporate three spatial components, leading to $(\lambda_i, \psi_i, \rho_i)' = \mathbf{A}(V_{1,i}, V_{2,i}, V_{3,i})'$, where now \mathbf{A} is a 3×3 lower triangular matrix.

4.3 Bayesian implementation

Posterior inference for the model parameters is implemented by model fitting with Gibbs samplers (Gelfand and Smith 1990) and Metropolis Hasting updating (Gelman 2003). Details of the full conditional distribution for each parameter are given in Appendix.

In our model, a vague normal prior is assigned to the shape parameter, ξ . We assign inverse gamma priors for positive value σ_0 . Coefficients β and ρ have normal priors centered at the exploratory estimates with large variances. In Models A and D, we follow the dynamic linear model prior setting to update $\delta_1, \delta_2, \dots, \delta_T, \phi$ and W_0 . For the coregionalization matrix \mathbf{A} , we assign truncated normal priors with positive value support for the diagonal entries, and normal priors for the other entries. Altogether, our priors are very weak suggesting little concern with regard to sensitivity analysis. However, we did implement some sensitivity study, primarily on the various prior uncertainties, which confirmed this. Details are available from the authors.

Under the conditional independence assumptions in the first stage settings, $\mu_{i,t}$ are sampled independently for each pair of (i, t) . Given $\mu_{i,t}$, we can directly sample β, ρ, V_2 and off-diagonal entries in \mathbf{A} by adopting conjugate normal priors. Parameters without closed form full conditional distributions are updated using Metropolis-Hastings. $\xi, \mu_{i,t}$ and V_1 are sampled by the random walk Metropolis-Hastings with Gaussian proposals. The proposal distributions for σ_0 and diagonal entries in \mathbf{A} are truncated normals centered at the current samples.

5 Model comparison and results

Again, the dataset consists of annual maximum records at 1,078 grid cells from 1950 to 1999. We fit the four hierarchical space–time models with latitude/longitude and elevation at the cell centroid entering linearly as covariates. Models are fitted using two parallel chains. Model D takes the longest to run, completing roughly three iterations per minute using R code with dual 2.8 GHz Xeon CPUs and 12 GB memory. We ran 10,000 iterations to collect posterior samples after a burn in period of 2,000 iterations, thinning using every fifth iteration. Trace plots of parameters indicate good convergence of the respective marginal distributions.

In the models proposed in Sect. 4, temporal evolution in the extreme rainfalls is taken into account in the model for $W_{i,t}$. More precisely, each of the models enables prediction for any grid cell for any year. Annual maximum rainfalls in year $T + 1$ could be simply obtained by updating samples from:

$$f(\mathbf{Y}_{T+1} | \mathbf{Y}_1, \dots, \mathbf{Y}_T, \mathbf{X}) = \int f(\mathbf{Y}_{T+1} | \theta, \mathbf{X}) f(\theta | \mathbf{Y}_1, \dots, \mathbf{Y}_T, \mathbf{X}) d\theta \quad (5)$$

Table 1 DIC, AAPE and $\hat{\rho}$ for Models A–D

| Models | \bar{D} | p_D | DIC | AAPE | AAD | $\hat{\rho}$ |
|--------|-----------|-------|--------|------|-------|--------------|
| A | 638212 | 7614 | 645826 | 84.5 | 110.3 | 0.985 |
| B | 639835 | 7332 | 647167 | 83.7 | 104.5 | 0.944 |
| C | 638845 | 7814 | 646659 | 77.9 | 103.3 | 0.945 |
| D | 653321 | 7802 | 661123 | 81.7 | 118.5 | 0.966 |

In fact, we held out the annual maximum rainfalls in 1999 for validation purposes, in order to compare models in terms of predictive performance. Posterior medians are adopted as the point estimates of the predicted annual maxima in 1999 because of the skewness of the predictive distribution for Y_{T+1} . We check the predictive performance by computing the averaged absolute predictive errors (AAPE) for each model. Given the true value of $Y_{i,1999}$ in the hold out dataset, AAPE is computed by

$$AAPE = \frac{1}{S} \sum_{i=1}^S |\hat{Y}_{i,1999} - Y_{i,1999}| \tag{6}$$

where $\hat{Y}_{i,1999}$ is the median of the posterior samples $\{\hat{Y}_{i,1999}^{(b)}\}$ for $b = 1, 2, \dots, B$ with the averaged absolute deviance (AAD) from the $\hat{Y}_{i,1999}$ estimated by

$$AAD = \frac{1}{BS} \sum_{b=1}^B \sum_{i=1}^S |\hat{Y}_{i,1999}^{(b)} - \hat{Y}_{i,1999}| \tag{7}$$

A second model comparison is to study, for each model, the proportion of the true annual maximum rainfalls in 1999 which lie in the associated estimated 95% credible intervals, denoted by $\hat{\rho}$, is computed as in Table 1. A third model selection criterion which is easily calculated from the posterior samples is the deviance information criterion (DIC) (Spiegelhalter et al. 2002).

Table 1 summarizes the DIC score, $\hat{\rho}$, AAPE and the associated AAD for each predictive posterior samples by each models. Model A wins among the four models under the DIC criterion. However, Model C has the lowest prediction error in terms of AAPE and AAD and is second in DIC. In addition, the $\hat{\rho}$ for Model C is quite close to 0.95, as desired. As a result, we summarize results based on Model C.

Table 2 provides the posterior means for the parameters and the corresponding 95% credible intervals under Model C. The resulting estimates for the coefficients β_{lat} , β_{lon} and β_{elev} in the regression part are insignificant; over the CFR we find no linear trend in these variables in explaining extreme rainfalls. This finding agrees with the results in Coles and Tawn (1996) for a different rainfall study. The posterior mean of the ξ takes value of 0.122 with 95% credible interval (0.112, 0.129), indicating the GEV distributions of annual maximum rainfalls in CFR have heavy upper tails. The estimate of ρ in Table 2 is nearly significant and suggests that, over the whole CFR, annual maximum rainfall is increasing, on average, by 0.015 mm per year over the past 50 years.

Table 2 Posterior sample means of parameters and the corresponding 95% credible intervals for Model C.

| | Mean | 95% CI |
|----------------|-------------|---------------------|
| β_0 | 249.4 | (236.7, 255.2) |
| β_{lat} | $-1.64e-03$ | ($-0.054, 0.044$) |
| β_{lon} | $2.03e-03$ | ($-0.015, 0.017$) |
| β_{elev} | $-1.23e-05$ | ($-0.001, 0.001$) |
| σ_0 | 90.15 | (88.42, 92.23) |
| ξ | 0.122 | (0.112, 0.129) |
| ρ | 0.150 | ($-0.06, 0.623$) |

β_{lat} , β_{lon} and β_{elev} are the coefficients of longitude, latitude and elevation; ξ the shape parameter of GEV distribution; σ_0 the center of spatially varying scale parameters; ρ the global trend in location parameters for Model C

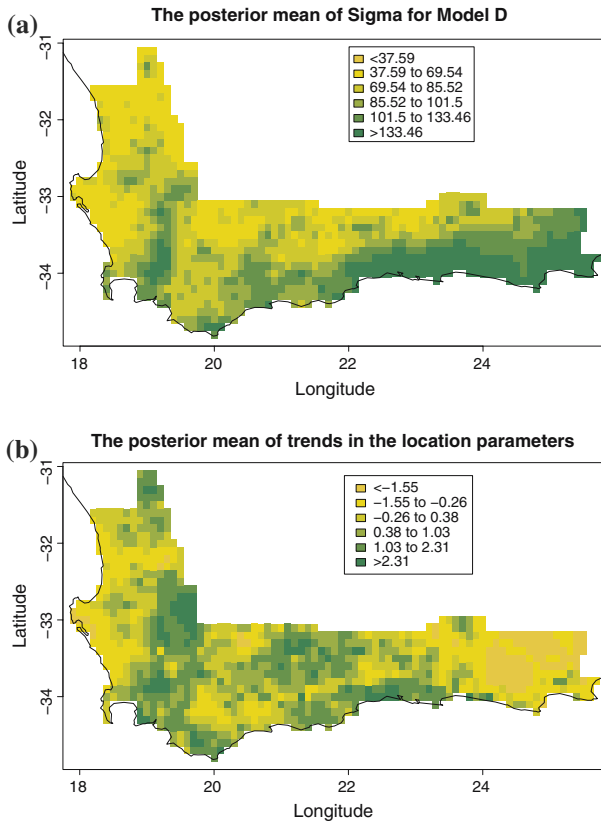


Fig. 6 Posterior sample means of scale parameters and local trends obtained from Model C

Model C allows for a further examination of the site-level trends. Figure 6 displays the posterior means of the site-level trends and the σ_i 's. The magnitudes of the estimates are represented through the grey scale plot and suggest spatial pattern in the ρ_i 's and the σ_i 's. Specifically, despite an overall tendency towards increasing extreme

rainfalls in the CFR, some small areas show tendency towards decreasing extreme rainfalls.

Return levels for the occurrence of extreme events are of practical interest in climate studies. Given posterior samples of $\mu_{i,t}$, σ_i and ξ under the model, posterior samples of return levels with return period 25 years are derived by Eq. 2. A map of estimated $z_{1/25}$ in 1998 is presented in Fig. 7a. Figure 7b compares return level $z_{1/25,1998}$ and $z_{1/25,1950}$ obtained from Model C. Though, under our models, the return levels change with time (so they are not useful as actual expected return times), it is informative to see how they vary over time. We see slight temporal changes in $z_{1/25}$ over the past 50 years. In the southwestern part of the CFR, return levels $z_{1/25}$ decreased at least 2% from 1950 to 1998, suggesting a decrease in the intensity of heavy rainfalls in CFR over the past 50 years. In contrast, more than half of the regions in CFR have increasing return levels $z_{1/25}$, suggesting the shift of the distributions of annual maxima rainfall towards larger heavy rainfalls. Finally, the posterior samples provide a natural way to evaluate the corresponding uncertainty of return level estimates as shown in Fig. 8.

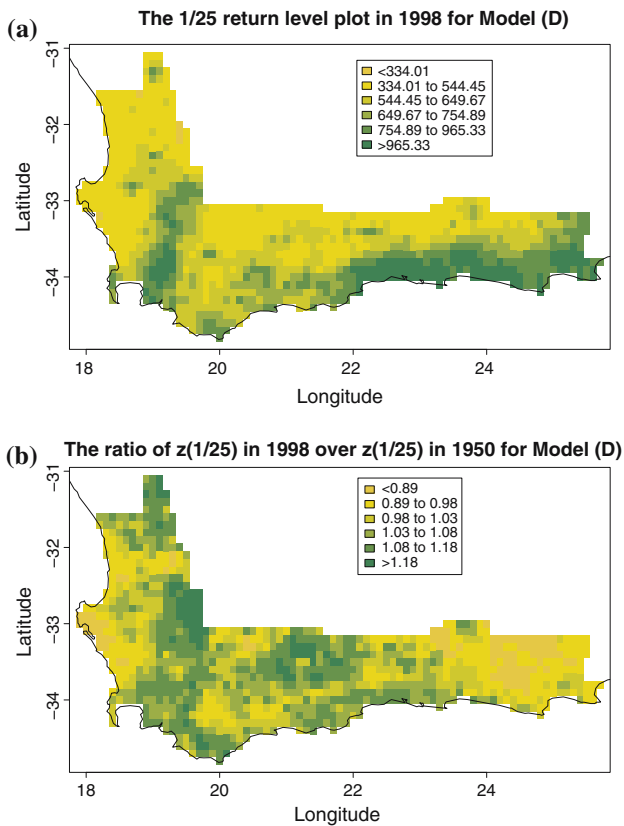


Fig. 7 (a) Posterior sample means of the 1/25 return levels in 1998. (b) Posterior ratio of return level $z_{1/25}$ in 1998 over $z_{1/25}$ in 1950 obtained from Model C

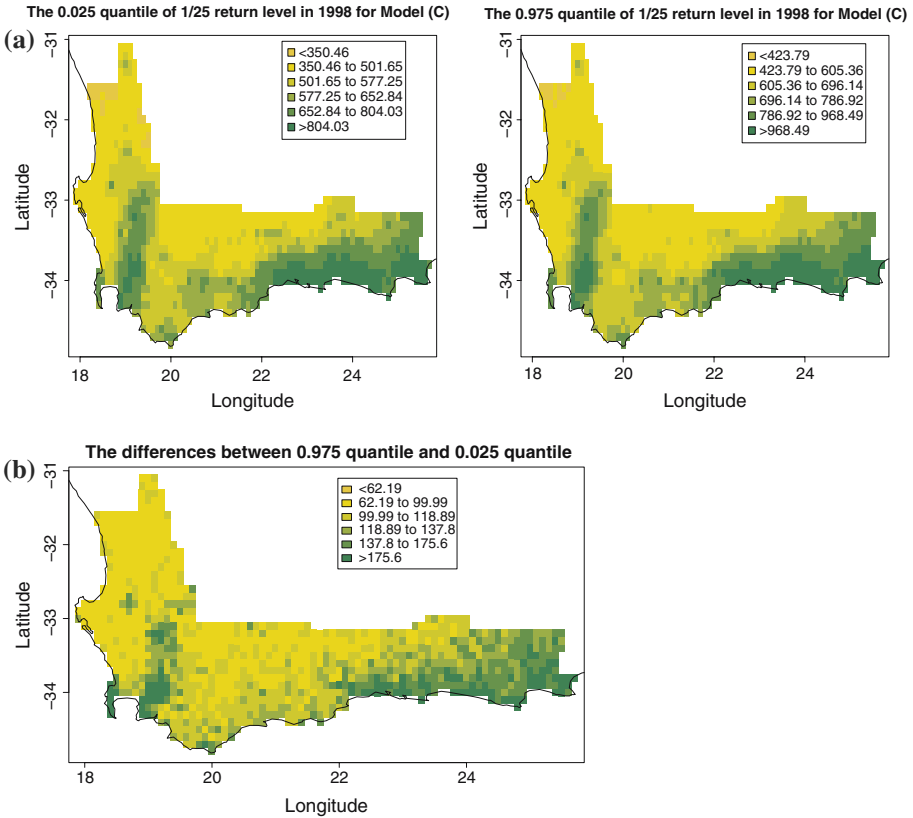


Fig. 8 (a) The lower and upper 0.025 quantiles of the 1/25 return levels in 1998 obtained from Model C. (b) The lengths of 95% credible intervals of the 1/25 return levels in 1998 obtained from Model C

The lower and upper 0.025 quantiles vary considerably over the region and the lengths of the associated interval estimates do as well.

6 Discussion and future work

We have presented flexible hierarchical models for the dynamic and spatial change in annual maximum rainfalls collected at more than 1,300 areal regions over 50 years. Our hierarchical models are based on the GEV distribution to describe the asymptotic behavior of maxima taken from a time series of daily records. We consider space and time-varying location parameters in the GEV distributions. In addition, we assume spatially varying scale parameters. A coregionalized CAR model is then introduced to capture the underlying spatial dependence between these parameters at site level. Inference for the space–time return level becomes straightforward under the MCMC model fitting approach.

Various extensions for this work are possible. Besides annual maxima of rainfalls, we have interest in similar modeling for annual temperature extremes. A further, and

more demanding, challenge will take us to bivariate space–time modeling for precipitation and temperature extremes (see, e.g., [Heffernan and Tawn 2003](#)).

In this model, the interpolation technique which yields the gridded data is not a point-to-point interpolation, but rather generates grid cell area integrals. The interpolation technique fits a surface to best estimate the area integral and, thus, inherently suppresses the peaks. Therefore, it is expected that interpolation of extremes will underestimate true extremes over the same area, i.e., the maximum of an average is at most the average of maximums. It is an open challenge to account for this effect in modeling interpolated extreme values.

Acknowledgements The authors thank Gabi Hegerl, Andrew Latimer, Anthony Rebelo and John Silander, Jr. for valuable conversations along with Bruce Hewitson and Dawn Woodard for help in the development of the dataset. They are also thankful to the editor and the referees for valuable comments. The work of the authors was supported in part by NSF DEB 0516198.

Appendix

Monte carlo sampling procedure

The hierarchical models we proposed are implemented by MCMC algorithm. We draw samples of parameters from their full conditional distribution, respectively. We illustrate the MCMC sampling procedures for Models A and C below.

Let $\beta = (\beta_0, \beta_{lat}, \beta_{lon})$, $\delta = (\delta_1, \delta_2, \dots, \delta_T)$, $\mu = (\mu_{1,1}, \mu_{1,2}, \dots, \mu_{1,T}, \mu_{2,1}, \dots, \mu_{S,T})'$, $\mu_t = (\mu_{1,t}, \mu_{2,t}, \dots, \mu_{S,t})'$, $\psi = (\psi_1, \psi_2, \dots, \psi_T)$, $\delta = (\delta_1, \delta_2, \dots, \delta_T)'$, $t = (1, 2, \dots, T)'$, $X = (X_1, X_2, \dots, X_S)'$, $X_i = (1, Lat_i, Lon_i, Elev_i)'$, $W = (W_{1,1}, W_{1,2}, \dots, W_{1,T}, W_{2,1}, \dots, W_{S,T})'$. Let $g(\cdot)$ be the probability density function of the GEV distribution. Let $N(\cdot)$ denotes the probability density function of normal distribution. Below we use θ_- generically to denote parameters other than the target parameter.

MCMC for the implementation of Model A

- **Updating μ :** Sample $\mu_{i,t}$ independently via Metropolis-Hasting algorithm for each i and t .

$$f(\mu_{i,t}|\theta_-) \propto g(Y_{i,t}; \mu_{i,t}, \sigma_i, \xi)N(\mu_{i,t}; X_i'\beta + W_{i,t}, \tau^2)$$

- **Updating ξ :** The Metropolis-Hastings step can be applied to sampling ξ based on its full conditional distribution.

$$f(\xi|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T g(Y_{i,t}; \mu_{i,t}, \sigma_i, \xi)N(\xi; \mu_{\xi}^0, \sigma_{\xi}^{02})$$

- **Updating β :** Sample β directly from the full conditional distribution, which is

$$f(\boldsymbol{\beta}|\boldsymbol{\theta}_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + W_{i,t}, \tau^2) N(\boldsymbol{\beta}; \boldsymbol{\mu}_\beta^0, \Sigma_\beta^0) \\ \sim N(Bb, B)$$

where $B^{-1} = \mathbf{X}^T \mathbf{X} + \Sigma_\beta^0$, $b = \mathbf{X}^T (\boldsymbol{\mu} - \mathbf{W}) + \Sigma_\beta^0 \boldsymbol{\mu}_\beta^0$

- **Updating σ_0 :** Sample σ_0 from its full conditional distribution

$$f(\sigma_0|\boldsymbol{\theta}_-) \propto \prod_{i=1}^S \prod_{t=1}^T g(Y_{i,t}; \mu_{i,t}, \sigma_0 \exp(a_{11} V_{1,i}), \xi) IG(\sigma_0; 2, \beta_{\sigma_0}^0)$$

Update $\boldsymbol{\sigma} = \sigma_0 \exp(a_{11} \mathbf{V}_1)$

- **Updating matrix \mathbf{A} :** \mathbf{A} is a 2×2 lower triangular matrix. Sample a_{11} via Metropolis-Hastings step. a_{12} and a_{22} are updated from their posterior normal distributions by adopting normal priors.

$$f(a_{11}|\boldsymbol{\theta}_-) \propto \prod_{i=1}^S \prod_{t=1}^T g(Y_{i,t}; \mu_{i,t}, \sigma_0 \exp(a_{11} V_{1,i}), \xi) N_+(a_{11}; \mu_{11}^0, \sigma_{11}^0{}^2)$$

$$f(a_{12}|\boldsymbol{\theta}_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + a_{12} V_{1,i} + a_{22} V_{2,i} + \delta_t, \tau^2) N(a_{12}; \mu_{12}^0, \sigma_{12}^0{}^2)$$

$$f(a_{22}|\boldsymbol{\theta}_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + a_{12} V_{1,i} + a_{22} V_{2,i} + \delta_t, \tau^2) N_+(a_{22}; \mu_{22}^0, \sigma_{22}^0{}^2)$$

- **Updating $\boldsymbol{\delta}$:** Conditional on $\mu_{i,t}$, $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$, $\boldsymbol{\mu}_t - \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\psi} = \delta_t \mathbf{1} + \boldsymbol{\epsilon}_t$; $\boldsymbol{\epsilon}_t \sim N(0, \tau^2 \mathbf{I})$, $\delta_t = \phi \delta_{t-1} + \omega_t$; $\omega \sim N(0, W_0)$. This is a specific dynamic linear model $DLM(Y; F, G, V, W)$ (West and Harrison 1997; Huerta and Sansó 2007) with known observation covariance matrix $\tau^2 \mathbf{I}$. Sample the state vector $\boldsymbol{\delta}$ using the Forward Filtering Backward Sampling algorithm (FFBS). Forward in time, sequentially sample δ_t from $\delta_t | (Y_1, \dots, Y_t, W_0, \phi)$ for $t = 1, 2, \dots, T$. Backwards in time, smooth the samples of δ_t by updating $\delta_t | (\delta_{t+1}, (y_1, \dots, y_T), W_0, \phi)$ for each time step. The evolution variance W_0 can be directly sampled from an Inverse-Wishart distribution if we adopt $IG(a_{W_0}^0, \beta_{W_0}^0)$ as priors for W_0 , i.e.,

$$f(W_0|\boldsymbol{\theta}_-) \propto IG\left(a_{W_0}^0 + \frac{T}{2}, \beta_{W_0}^0 + \frac{\sum_{t=1}^T (\delta_t - \phi \delta_{t-1})^2}{2}\right),$$

where $\delta_0 = 0$.

Finally, sample ϕ from its conditional distribution where we use normal conjugate prior, i.e., $f(\phi|\boldsymbol{\delta}, W_0) \propto \sum_{t=2}^T N(\delta_t; \phi \delta_{t-1}, W_0) N(\phi; \mu_\phi^0, \sigma_\phi^0{}^2)$

- **Updating \mathbf{V}_1** by Metropolis-Hastings step: The conditional distribution for $V_{1,i}$ is:

$$f(V_{1,i}|V_{1,j}, j \neq i, V_{2,i}, \theta_-) \propto \prod_{t=1}^T g(Y_{i,t}; \mu_{i,t}, \sigma_0 \exp(a_{11} V_{1,i}), \xi) f(V_{1,i}; \mu_{v_{1i}}, \sigma_{v_{1i}}^2)$$

- **Updating V_2 :** The full conditional distribution for V_2 is a Gaussian. For each $i = 1, 2, \dots, S$, update $V_{2,i}$ directly from

$$f(V_{2,i}|V_{2,j}, j \neq i, V_{1,i}, \theta_-) \propto N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + a_{12} V_{1,i} a_{22} V_{2,i} + \delta_t, \tau^2) \times N\left(\sum_j w_{ij} V_{2,j} / w_{i,+}, w_{i,+}\right)$$

Finally, Update $W_{i,t} = \psi_i + \delta_t$.

MCMC for Model C: Follow the procedures of implementing Model A to update $\boldsymbol{\mu}$, ξ , $\boldsymbol{\beta}$ and σ_0 .

- **Updating ρ :** Sample ρ from its full conditional distribution

$$f(\rho|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + \psi_i + (\rho + \rho_i)(t - t_0), \tau^2) N(\rho; \mu_\rho^0, \sigma_\rho^2)$$

Update $\delta = \rho(t - t_0)$

- **Updating A :** For Model C, A is a 3×3 lower triangular matrix. Sample a_{11} via Metropolis-Hasting step. $a_{12}, a_{22}, a_{13}, a_{23}$ and a_{33} are updated from their posterior normal distributions by adopting conjugate normal priors.

$$f(a_{11}|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T g(Y_{i,t}; \mu_{i,t}, \sigma_0 \exp(a_{11} V_{1,i}), \xi) N_+(a_{11}; \mu_{11}^0, \sigma_{11}^2)$$

$$f(a_{12}|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + a_{12} V_{1,i} + a_{22} V_{2,i} + \delta_{i,t}, \tau^2) N(a_{12}; \mu_{12}^0, \sigma_{12}^2)$$

$$f(a_{22}|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + a_{12} V_{1,i} + a_{22} V_{2,i} + \delta_{i,t}, \tau^2) N_+(a_{22}; \mu_{22}^0, \sigma_{22}^2)$$

$$f(a_{13}|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + \psi_i + (\rho + \rho_i)(t - t_0), \tau^2) N(a_{13}; \mu_{13}^0, \sigma_{13}^2)$$

$$f(a_{23}|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + \psi_i + (\rho + \rho_i)(t - t_0), \tau^2) N(a_{23}; \mu_{23}^0, \sigma_{23}^2)$$

$$f(a_{33}|\theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + \psi_i + (\rho + \rho_i)(t - t_0), \tau^2) N_+(a_{33}; \mu_{33}^0, \sigma_{33}^2)$$

- For $i = 1, 2, \dots, S$ and $t = 1, 2, \dots, T$, update $\psi_i = a_{12} V_{1,i} + a_{22} V_{2,i}$, $\rho_i = a_{13} V_{1,i} + a_{23} V_{2,i} + a_{33} V_{3,i}$, $\sigma_i = \sigma_0 \exp(a_{11} V_{1,i})$, $W_{i,t} = \psi_i + (\rho + \rho_i)(t - t_0)$.

- **Updating V_1, V_2, V_3 :** Update V_1 by Metropolis-Hastings step. The conditional distribution for $V_{1,i}$ is

$$f(V_{1,i}|V_{1,j}, j \neq i, V_{2,i}, \theta_-) \propto \prod_{t=1}^T g(Y_{i,t}; \mu_{i,t}, \sigma_0 \exp(a_{11} V_{1,i}), \xi) N(V_{1,i}; \mu_{v_{1i}}, \sigma_{v_{1i}}^2)$$

The full conditional distribution for V_2 is a Gaussian distribution. For each $i = 1, 2, \dots, S$, update $V_{2,i}$ directly from

$$f(V_{2,i}|V_{2,j}, j \neq i, V_{1,i}, V_{3,i}, \theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + a_{12} V_{1,i} + a_{22} V_{2,i} + (\rho + a_{13} V_{1,i} + a_{23} V_{2,i} + a_{33} V_{3,i})(t - t_0), \tau^2) \times N\left(\sum_{j=1}^S w_{i,j} V_{2,j} / w_{i,+}, w_{i,+}\right)$$

The full conditional distribution for V_3 is a Gaussian distribution. For each $i = 1, 2, \dots, S$, updating $V_{3,i}$ directly from:

$$f(V_{3,i}|V_{3,j}, j \neq i, V_{1,i}, V_{2,i}, \theta_-) \propto \prod_{i=1}^S \prod_{t=1}^T N(\mu_{i,t}; \mathbf{X}'_i \boldsymbol{\beta} + a_{12} V_{1,i} + a_{22} V_{2,i} + (\rho + a_{13} V_{1,i} + a_{23} V_{2,i} + a_{33} V_{3,i})(t - t_0), \tau^2) \times N\left(\sum_{j=1}^S w_{i,j} V_{3,j} / w_{i,+}, w_{i,+}\right)$$

- Follow the updating procedures described in Model A to sample the other parameters.

References

Banerjee S, Carlin BP, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC Press,
 Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J Roy Stat soc, Ser B, (Methodol) 36:192–236
 Besag J, Green PJ, Higdon D, Mengersen K (1995) Bayesian computation and stochastic systems. Stat Sci 10(1):3–66
 Coles SG, Tawn JA (1996) Modeling extreme of the areal rainfall processes. J Roy Stat Soc, Ser B, (Methodol) 58:329–347
 Coles SG (2001) An introduction to statistical modeling of extreme values. Springer-Verlag, New York
 Cooley D, Nychka D, Naveau P (2007) Bayesian spatial modeling of extreme precipitation return levels. J Am Stat Assoc 102:824–840
 Dahan E, Mendelson H (2001) An extreme value model of concept testing. Manage Sci 47:102–116

- Durman CF, Gregory JM, Hassell DC, Jones RG, Murphy JM (2001) A comparison of extreme European daily precipitation simulated by a global and regional climate model for present and future climates. *Q J Roy Meteorol Soc* 127:573
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85(410):398–409
- Gelfand AE, Vounatsou P (2002) Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 4:11–25
- Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF (2004) Nonstationary multivariate process modelling through spatially varying coregionalization (with discussion). *Test* 13(2):1–50
- Gelman A (2003) *Bayesian data analysis*. CRC Press
- Heffernan JE, Tawn JA (2003) A conditional approach for multivariate extreme values. *J Roy Stat Soc, Ser B, (Methodol)* 66:497–546
- Hewitson BC, Crane RG (2005) Gridded area-averaged daily precipitation via conditional interpolation. *J Climate* 18:41–57
- Huerta G, Sansó B (2007) Time-varying models for extreme values. *Environ Ecol Stat* 14:285–299
- Jin X, Carlin BP, Banerjee S (2005) Generalized hierarchical multivariate CAR models for areal data. *Biometrics* 61:950–961
- Kharin VV, Zwiers FW (2005) Estimating extremes in transient climate change simulations. *J Climate* 1:1
- Mardia KV (1988) Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J Multivariate Anal* 24:265–284
- Roberts SJ (2000) Extreme value statistics for novelty detection in biomedical Signal Processing. *IEE Proc Sci Technol Meas* 47(6):363–367
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit (with discussion). *J Roy Stat Soc, Ser B, (Stat Methodol)* 64(4):583–639
- Thompson ML, Reynolds J, Cox LH, Guttorp P, Sampson PD (2001) A review of statistical methods for the meteorological adjustment of tropospheric ozone. *Atmos Environ* 35:617–630
- West M, Harrison J (1997) *Bayesian forecasting and dynamic models*. Springer Verlag, New York

Author Biographies

Huiyan Sang is a Ph.D. student of Statistical Science at Duke University in Durham, NC. She received the B.S. degree in mathematics and applied mathematics (2004) from Peking University in Beijing. Her research interests include spatial statistics, modeling extreme values and complex modeling for environmental data.

Alan E. Gelfand is the J B Duke Professor of Statistical Science at Duke University in Durham, NC. He received a B.S. from the City College of New York and an M.S. and Ph.D. from Stanford University. An early contributor to the development of computational machinery for fitting hierarchical Bayesian models, his current research focuses on spatial and spatio-temporal statistics. His primary areas of application are to problems in environmental science, ecology, and climatology.