5th International Workshop on Climate Informatics
September 24-25, 2015
Hosted by the National Center for Atmospheric Research in Boulder, CO

# EFFICIENT BAYESIAN HIERARCHICAL MODELING OF SPATIAL PRECIPITATION EXTREMES

Cameron Bracken[1,2], Balaji Rajagopalan[1,3], Linyin Cheng[3], Subhrendu Gangopadhyay[2]

***Abstract*—An efficient Bayesian hierarchical model for spatial extremes on a large domain is proposed. In the data layer a Gaussian elliptical copula having generalized extreme value (GEV) marginals is applied. Spatial dependence in the GEV parameters are captured with a latent spatial regression. Using a composite likelihood approach and a method for incorporating stations with missing data, we are able to efficiently incorporate a large precipitation dataset. The model is demonstrated by application to seasonal precipitation extremes at approximately 2800 stations covering the western United States, -125E – -100E longitude and 30N to 50N latitude. The hierarchical model provides parameters on a 1/8th degree grid and consequently maps of return levels and associated uncertainty for each season. The model results indicate that return levels vary coherently both spatially and across seasons, providing valuable information about the space-time variations of risk of extreme precipitation in the western US, helpful for infrastructure planning.**

Engineering design of infrastructure such as flood protection, dams, etc. and management of water supply and flood control require robust estimates of return levels and associated errors of precipitation extremes. Spatial modeling of precipitation extremes not only can capture spatial dependence between stations but also reduce the overall uncertainty in at-site return level estimates by borrowing strength across spatial locations [1]. Hierarchical Bayesian modeling of extremes precipitation was first introduced by [1] and since has been widely discussed in the literature [2], [3]. Hierarchical modeling is an alternative to regional frequency analysis providing gridded or pointwise estimates of return levels within a study region [4].

Bayesian hierarchical models for spatial extremes have typically been limited to small geographic regions that include on the order 100 stations covering areas on the order of 100,000 km$^2$. Large geographic regions with many stations present a computational challenge for hierarchical Bayesian models, especially when computing the likelihood of Gaussian processes (GPs), which for $n$ data points, requires inverting an $n \times n$ matrix, an $O(n^3)$ operation. Several approaches exist for speeding up GP likelihood computations such as low-rank approximations [5] in which the GP is approximated at a small number of knots and composite likelihood methods [6] where the likelihood computation is broken into groups containing a small number of stations. The use of a composite likelihood approach is explored here because we not only wish to estimate covariance parameters but to also produce maps of return levels with small credible intervals.

Some attempts have been made to model extremes in large regions and large datasets in a Bayesian hierarchical context. [7] use a hierarchical max-stable model with climate model output in the east coast to examine spatially varying GEV parameters, with a max stable process for the data dependence level. [8] model gridded precipitation data over the entire US, for annual maxima at a 5x5 degree resolution (43 grid cells) and copula for data dependence, incorporating spatial dependence directly in a spatial model on the data, not parameters. [9] and [10] model over 1000 grid cells of climate model output using spatial autoregressive models which take advantage of data on a regular lattice to simplify computations.

The research contributions of this study are as follows. A Bayesian hierarchical model is proposed which is capable of incorporating thousands of observation locations by utilizing a composite likelihood method. The GEV shape parameter is modeled spatially in order to capture the detailed behavior of extremes in the western US. In addition the model is capable of incorporating stations with missing data with little additional computational overhead. The model is applied to observed precipitation extremes in each season, providing estimated seasonal return levels for the western US.

## I. MODEL STRUCTURE

The joint distribution of the $m$ data in each year is modeled as a realization from a Gaussian elliptical copula with generalized extreme value (GEV) distribution marginals. The copula is characterized by pair-

Corresponding author: C Bracken, University of Colorado, Boulder CO cameron.bracken@colorado.edu [1] Bureau of Reclamation [2] Cooperative Institute for Research in Environmental Sciences [3]

wise dependence matrix $\Sigma$. Spatial dependence is further captured through spatial processes on the location $\mu(\mathbf{s})$, scale $\sigma(\mathbf{s})$ and $\xi(\mathbf{s})$ parameters. We assume the parameters can be described through a latent spatial regression where the residual component $w_\gamma(\mathbf{s})$ follows a mean 0, stationary, isotropic Gaussian process (GP) with covariance function $C_\gamma(\mathbf{s}, \mathbf{s}')$. The corresponding covariance matrix is $C_\gamma(\boldsymbol{\theta}_\gamma) = [C_\gamma(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}_\gamma)]_{i,j=1}^m$, where $\gamma$ represents any GEV parameter ($\mu$, $\sigma$, $\xi$) and $\boldsymbol{\theta}_\gamma$ represents the covariance parameters. The hierarchical model structure is:

$$Y(\mathbf{s}, t) \sim Gcop_m[\Sigma; \mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})] \tag{1}$$

$$\mu(\mathbf{s}) = \mathbf{x}_\mu^T(\mathbf{s})\beta_\mu + w_\mu(\mathbf{s}) \tag{2}$$

$$\sigma(\mathbf{s}) = \mathbf{x}_\sigma^T(\mathbf{s})\beta_\sigma + w_\sigma(\mathbf{s}) \tag{3}$$

$$\xi(\mathbf{s}) = \mathbf{x}_\xi^T(\mathbf{s})\beta_\xi + w_\xi(\mathbf{s}) \tag{4}$$

where $Y(\mathbf{s}, t)$ is the response at site s and time $t$ and $Gcop_m$ stands for “m-dimensional Gaussian elliptical copula” with dependence matrix $\Sigma$. The spatial data layer processes in each year are assumed independent and identically distributed. Alternatives to using a copula to construct the joint distribution are an assumption of conditional independence [1] and max-stability [11].

In the process layer (Equations 2-4), $\mathbf{x}_\gamma^T(\mathbf{s}_i)$ is a vector of $p_\gamma$ spatially varying predictors (prepended with a 1) and $\boldsymbol{\beta}_\gamma$ is a vector of $p_\gamma + 1$ regression coefficients (including an intercept term).

A composite likelihood approximation is applied to each layer of the model (including the copula layer) enabling the efficient incorporation of over 2800 precipitation stations.

## II. Example Results

Figure 1 shows some example results from the model, 100-year return levels for winter.

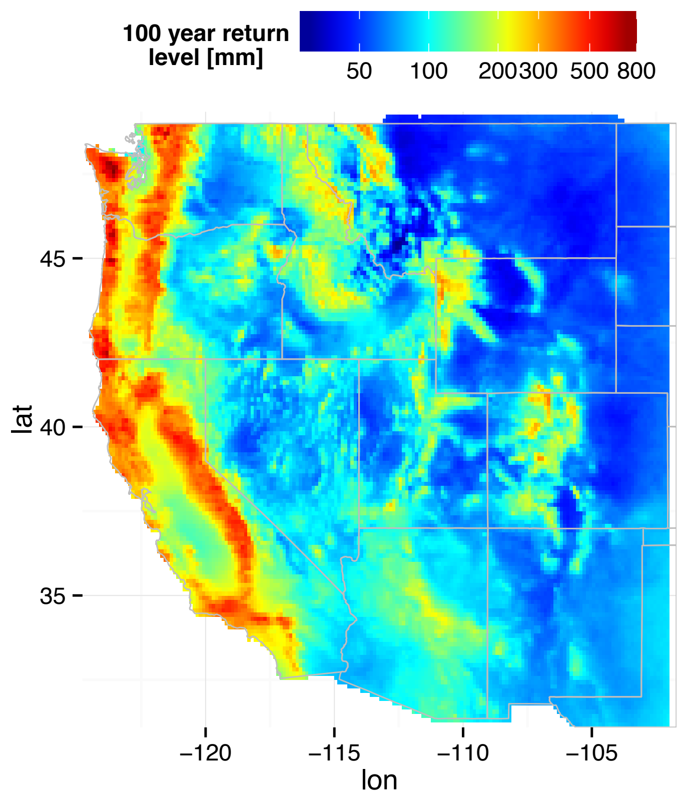


Fig. 1. 100-year precipitation return levels for the western US

## Acknowledgments

Funding for this research by a Science and Technology grant from Bureau of Reclamation is gratefully acknowledged. This work utilized the Janus supercomputer at the University of Colorado.

## References

[1] D. Cooley, D. Nychka, and P. Naveau, “Bayesian spatial modeling of extreme precipitation return levels,” *Journal of the American Statistical Association*, 2007.

[2] A. C. Davison, S. A. Padoan, and M. Ribatet, “Statistical Modeling of Spatial Extremes,” *Statistical Science*, vol. 27, pp. 161–186, May 2012.

[3] H. Sang and A. E. Gelfand, “Continuous Spatial Process Models for Spatial Extreme Values,” *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 15, pp. 49–65, Jan. 2010.

[4] B. Renard, “A Bayesian hierarchical approach to regional frequency analysis,” *Water Resources Research*, 2011.

[5] S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang, “Gaussian predictive process models for large spatial data sets,” *Journal of the Royal Statistical Society*, 2008.

[6] P. C. Caragea and R. L. Smith, “Asymptotic properties of computationally efficient alternative estimators for a class of multivariate normal models,” *Journal of Multivariate Analysis*, vol. 98, pp. 1417–1440, Aug. 2007.

[7] B. J. Reich and B. Shaby, “A hierarchical max-stable spatial model for extreme precipitation,” *The annals of applied statistics*, vol. 6, no. 4, pp. 1430–1451, 2012.

[8] S. Ghosh and B. K. Mallick, “A hierarchical Bayesian spatio-temporal model for extreme precipitation events,” *Environmetrics*, vol. 22, pp. 192–204, Mar. 2011.

[9] D. Cooley and S. R. Sain, “Spatial Hierarchical Modeling of Precipitation Extremes From a Regional Climate Model,” *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 15, pp. 381–402, Mar. 2010.

[10] H. Sang and A. E. Gelfand, “Hierarchical modeling for extreme values observed over space and time,” *Environmental and Ecological Statistics*, vol. 16, pp. 407–426, Sept. 2009.

[11] S. A. Padoan, M. Ribatet, and S. A. Sisson, “Likelihood-Based Inference for Max-Stable Processes,” *dx.doi.org*, vol. 105, pp. 263–277, Jan. 2010.