Generating gridded fields of extreme precipitation for large domains with a Bayesian hierarchical model

Cameron Bracken
Department of Civil, Environmental and Architectural Engineering
University of Colorado at Boulder

Vannes, France
May 19, 2016

Coauthors

- Balaji Rajagopalan - University of Colorado
- Will Kleiber - University of Colorado
- Linyin Cheng - NOAA
- Subhrendu Gangopadhyay - Bureau of Reclamation

This talk

- Motivation - Link to weather generation
- Bayesian hierarchical model (BHM) for spatial extremes
- Application to the Western US
- Simulation procedure using the BHM

Motivation - Extremes in the western US?

- From 1983-2000 the Western States (WA, OR, CA, ID, NV, UT, AZ, MT, WY, CO, NM, ND, SD, NE, KS, OK, and TX) experienced **$24.7 billion in flood damages**, $1.5 billion annually.
- California, Washington, and Oregon alone accounted for $10.6 billion (46 percent) [Ralph et al., 2014, Pielke et al., 2002]

Boulder Flood, 2013

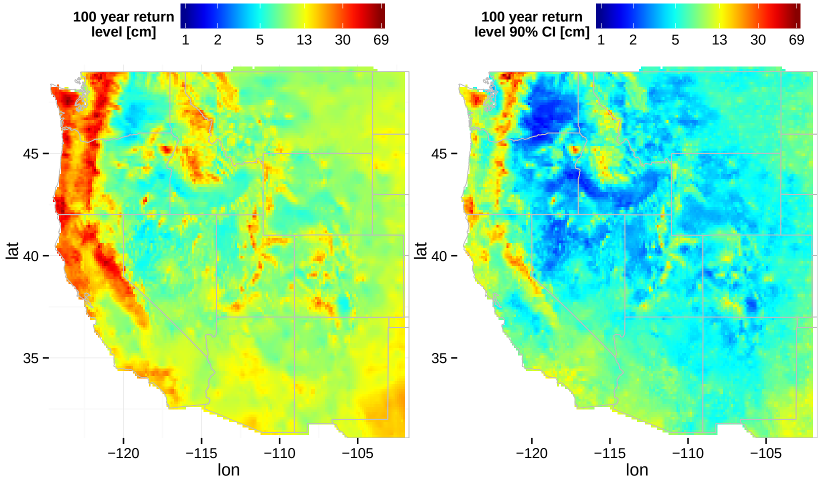
Research into hydroclimate extremes can provide more accurate and localized estimates of flood risk.

More Motivation

- How are simulated fields of extremes linked to weather generation?
- May not need a full weather generator – extremes may be the only quantity of interest
- Preserve spatial dependence of simulated extremes
- Downscaling

What are BHM Spatial Extremes models typically used for?

- Return Levels



Background on Bayesian Statistics

From Bayes' rule:

$$\underbrace{p(\theta|y,x)}_{\text{Posterior}} \propto \underbrace{p(y|\theta,x)}_{\text{Likelihood}} \underbrace{p(\theta,x)}_{\text{Prior}}$$

θ Parameters

y Dependent data (response)

x Independent data (covariates/predictors/constants)

Posterior: The answer, probability distributions of parameters

Likelihood: A computable function of the parameters, model specific

Prior: Probability distribution, incorporates existing knowledge of the system, model specific

Bayesian Hierarchical Model

In a hierarchical Bayesian model, expand terms using conditional distributions where $\boldsymbol{\theta} = (\theta_1, \theta_2)$:

$$\underbrace{p(\boldsymbol{\theta}\,|y)}_{\text{Posterior}} \propto \underbrace{p(y|\theta_1)}_{\text{Data Likelihood}} \underbrace{p(\theta_1|\theta_2)}_{\text{Process Liklihood}} \underbrace{p(\theta_2)}_{\text{Prior}}$$

Data Likelihood Relates observed data to distribution parameters

Process Likelihood Relates distribution parameters of the to each other

Statistics of Extremes

Given daily data, if we select the maximum value in each year, those data follow a generalized extreme value (GEV) distribution:

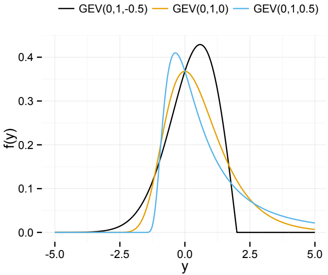
$$\mathrm{GEV}(y;\mu,\sigma,\xi) = \frac{1}{\sigma} b^{(-1/\xi)-1} \exp\left\{-b^{-1/\xi}\right\}$$

$b = 1 + \xi\left(\frac{x-\mu}{\sigma}\right)$, μ: Location, σ: Scale, ξ: Shape.

Return Level (quantile function):

$$z_r = \mu + \frac{\sigma}{\xi}\left[(-\log(1-1/r))^{-\xi} - 1\right]$$

Where r is the return period in years (100 years for example).



Hierarchical spatial model

$$
\begin{aligned}
(Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t)) &\sim C_g[\Sigma; \{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}] \\
Y(\mathbf{s}, t) &\sim GEV[\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})]
\end{aligned}
$$

Hierarchical spatial model

$$
\begin{aligned}
(Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t)) &\sim C_g[\Sigma; \{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}] \\
Y(\mathbf{s}, t) &\sim GEV[\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})] \\
\mu(\mathbf{s}) &= \beta_{\mu,0} + \mathbf{x}_\mu^T(\mathbf{s})\boldsymbol{\beta}_\mu(\mathbf{s}) + w_\mu(\mathbf{s}) \\
\sigma(\mathbf{s}) &= \beta_{\sigma,0} + \mathbf{x}_\sigma^T(\mathbf{s})\boldsymbol{\beta}_\sigma(\mathbf{s}) + w_\sigma(\mathbf{s}) \\
\xi(\mathbf{s}) &= \beta_{\xi,0} + \mathbf{x}_\xi^T(\mathbf{s})\boldsymbol{\beta}_\xi(\mathbf{s}) + w_\xi(\mathbf{s})
\end{aligned}
$$

Hierarchical spatial model

$$
\begin{aligned}
(Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t)) &\sim C_g[\Sigma; \{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}] \\
Y(\mathbf{s}, t) &\sim GEV[\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})] \\
\mu(\mathbf{s}) &= \beta_{\mu,0} + \mathbf{x}_\mu^T(\mathbf{s})\boldsymbol{\beta}_\mu(\mathbf{s}) + w_\mu(\mathbf{s}) \\
\sigma(\mathbf{s}) &= \beta_{\sigma,0} + \mathbf{x}_\sigma^T(\mathbf{s})\boldsymbol{\beta}_\sigma(\mathbf{s}) + w_\sigma(\mathbf{s}) \\
\xi(\mathbf{s}) &= \beta_{\xi,0} + \mathbf{x}_\xi^T(\mathbf{s})\boldsymbol{\beta}_\xi(\mathbf{s}) + w_\xi(\mathbf{s}) \\
w_\mu(\mathbf{s}) &\sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\mu)) \\
w_\sigma(\mathbf{s}) &\sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\sigma)) \\
w_\xi(\mathbf{s}) &\sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\xi))
\end{aligned}
$$

Hierarchical spatial model

$$
\begin{aligned}
(Y(\mathbf{s}_1,t),\ldots,Y(\mathbf{s}_n,t)) &\sim C_g[\Sigma;\{\mu(\mathbf{s}),\sigma(\mathbf{s}),\xi(\mathbf{s})\}] \\
Y(\mathbf{s},t) &\sim GEV[\mu(\mathbf{s}),\sigma(\mathbf{s}),\xi(\mathbf{s})] \\
\mu(\mathbf{s}) &= \beta_{\mu,0} + \mathbf{x}_\mu^T(\mathbf{s})\boldsymbol{\beta}_\mu(\mathbf{s}) + w_\mu(\mathbf{s}) \\
\sigma(\mathbf{s}) &= \beta_{\sigma,0} + \mathbf{x}_\sigma^T(\mathbf{s})\boldsymbol{\beta}_\sigma(\mathbf{s}) + w_\sigma(\mathbf{s}) \\
\xi(\mathbf{s}) &= \beta_{\xi,0} + \mathbf{x}_\xi^T(\mathbf{s})\boldsymbol{\beta}_\xi(\mathbf{s}) + w_\xi(\mathbf{s}) \\
w_\mu(\mathbf{s}) &\sim GP(\mathbf{0},C(\boldsymbol{\theta}_\mu)) \\
w_\sigma(\mathbf{s}) &\sim GP(\mathbf{0},C(\boldsymbol{\theta}_\sigma)) \\
w_\xi(\mathbf{s}) &\sim GP(\mathbf{0},C(\boldsymbol{\theta}_\xi))
\end{aligned}
$$

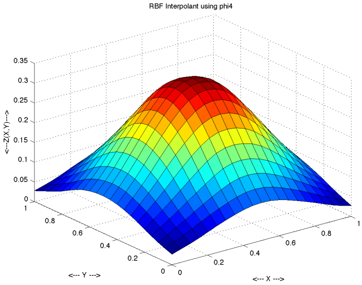
Covariates: $x^T(\mathrm{s}) =$ (Elevation, Mean Seasonal Precip)^T
C: Stationary, isotropic, exponential covariance model

Spatially varying regression coefficients

$$\beta_\mu(\mathbf{s}) = \sum_{i=1}^{k} c_i^\mu \eta_i^\mu(\mathbf{s})$$

$$\beta_\sigma(\mathbf{s}) = \sum_{i=1}^{k} c_i^\sigma \eta_i^\sigma(\mathbf{s})$$

$$\beta_\xi(\mathbf{s}) = \sum_{i=1}^{k} c_i^\xi \eta_i^\xi(\mathbf{s})$$



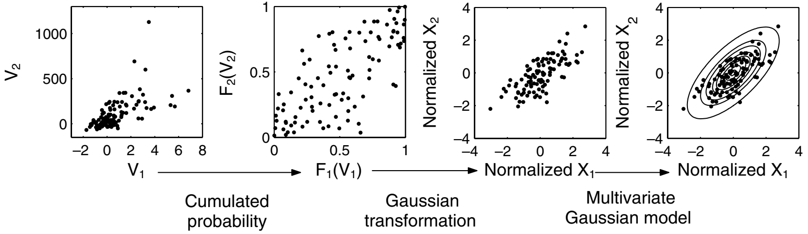
$$\eta_i(\mathbf{s}) = \exp(d_i/\phi_i)$$

Elliptical copula for data dependence

The Gaussian elliptical copula constructs the joint cdf of of a random vector $(V_1, \ldots, V_n)$ as

$$F_C(v_1, \ldots, v_n) = \Phi_\Sigma(u_1, \ldots, u_n) \tag{1}$$

where $\Phi_\Sigma(\cdot)$ is the joint cdf of an n-dimensional multivariate normal distribution with covariance matrix Σ, $u_i = \phi^{-1}(F_i[v_i])$, ϕ^{-1} is the inverse cdf (quantile function) of the standard normal distribution and $F_i(\cdot)$ is the marginal GEV cdf of variable i.



Elliptical copula for data dependence

The corresponding joint pdf is

$$f_C(y_1,\ldots,y_m) = \frac{\prod_{i=1}^{m} f_i[y_i]}{\prod_{i=1}^{m} \psi[u_i]} \Psi_\Sigma(u_1,....u_m) \tag{2}$$

where f_i is the marginal GEV pdf at site i, ψ is the standard normal pdf and Ψ_Σ is the joint pdf of an m-dimensional multivariate normal distribution.

Exponential dependence matrix for spatial data:

$$\Sigma(i,j) = \exp(-||\, \mathrm{s}_i - \mathrm{s}_j \,||/a_0) \tag{3}$$

a_0 is the copula range parameter. Termed the **dependogram** since values are not covariances [Renard, 2011].

Pros and Cons of Gaussian elliptical copulas

Benefits:

- Easy to implement
- Few parameters
- Flexible (can use any marginal distribution)

Requires checking two conditions in application:

- Asymptotic independence (dependence is hard to find in practice)
- Multivariate normality after transformation (usually satisfied)

Composite Likelihood (CL)

- Evaluating the full GP likelihood for 2500 stations is infeasible in a Bayesian model
  - One inversion (or Cholesky decomposition) of the covariance matrix takes seconds
- CL is an approximation of the full likelihood
- Stations are broken up into G groups each with n_g stations

$$L_c(\boldsymbol{\theta}|\mathbf{y}_1, \ldots, \mathbf{y}_G) = \prod_{g=1}^{G} L_g(\boldsymbol{\theta}|\mathbf{y}_g) \tag{4}$$

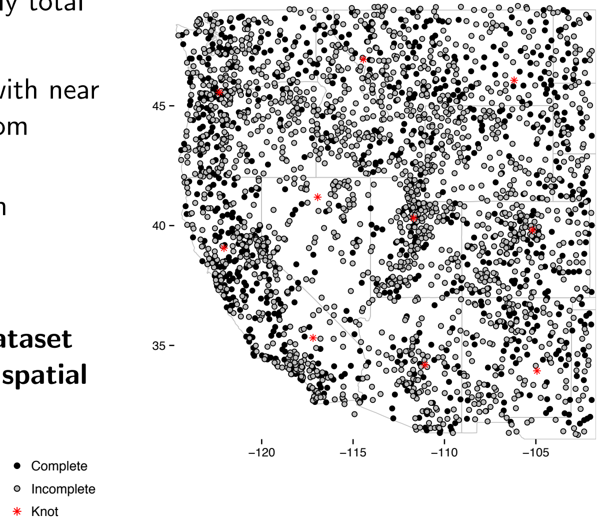
- CL Requires $O(Gn_g^3)$ computations as opposed to $O(n^3)$.
- This approximation is applied to the copula as well as each of the GEV parameter residuals.

Precipitation Data

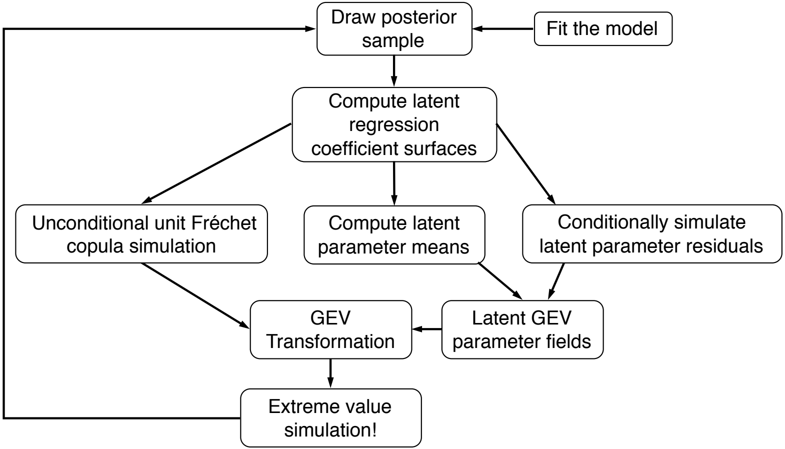
Global Historical Climatology Network (GHCN), daily total precip data

- ∼2500 stations with near complete data from 1948-2013
- 3 day aggregation window
- Fall maxima

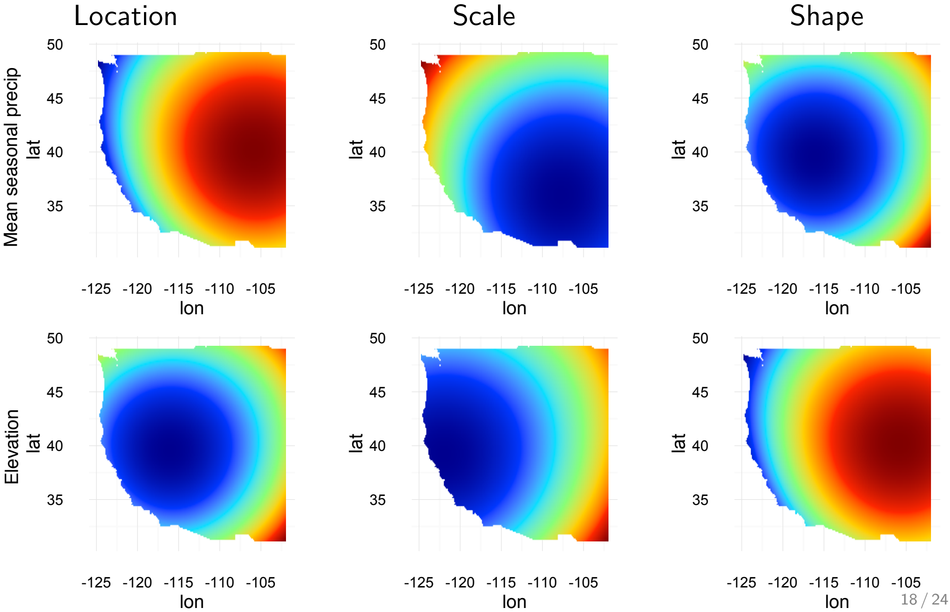
Very large region/dataset for typical Bayesian spatial model



Simulation procedure

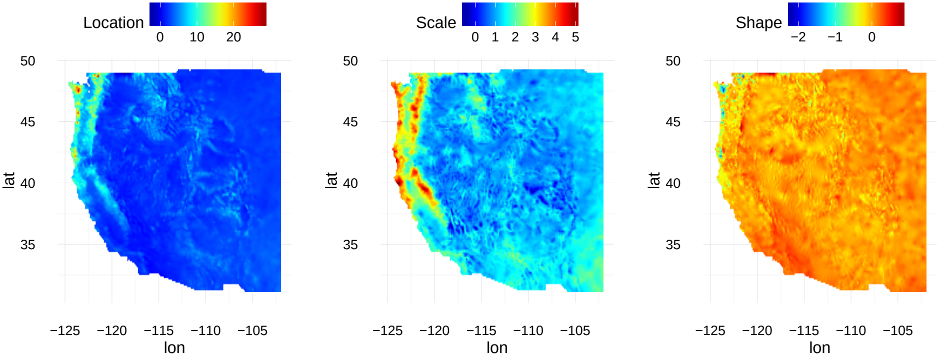


Regression coefficient surfaces - one posterior sample

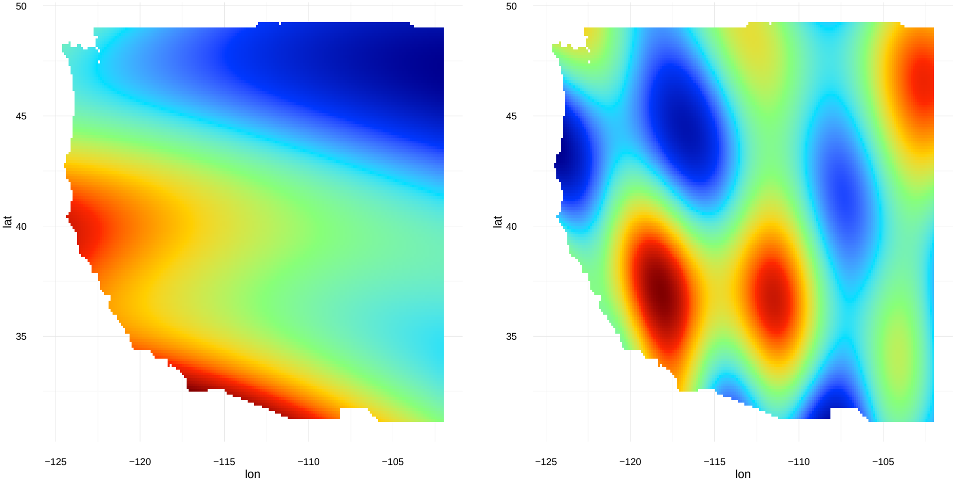


GEV Parameter simulation = conditional sim + mean (covariates + betas)

$$\mu(\mathbf{s}) = \beta_{\mu,0} + \mathbf{x}_{\mu}^{T}(\mathbf{s})\boldsymbol{\beta}_{\mu}(\mathbf{s}) + w_{\mu}(\mathbf{s})$$



Unit Fréchet Copula Simulations



GEV from unit Fréchet

If Y is a random variable with a GEV distribution with location μ, scale σ and shape ξ. Then,

$$Z = [1 + \xi(Y - \mu)/\sigma]^{1/\xi}$$

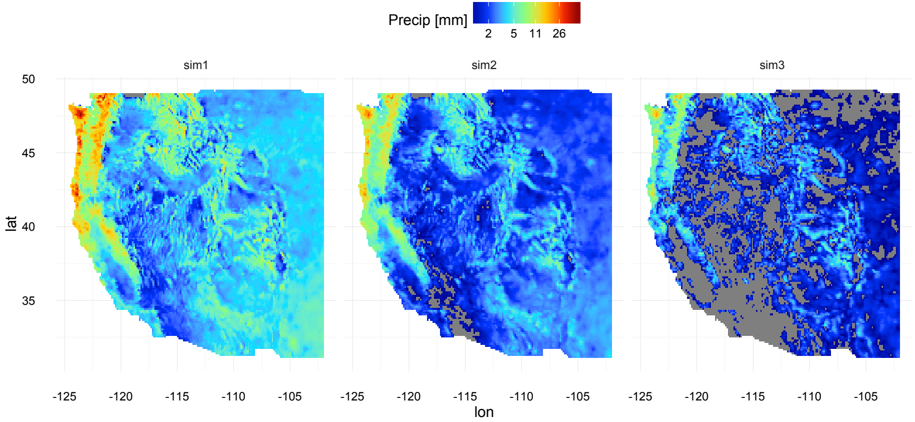
is unit Fréchet distributed i.e. $Z \sim$ GEV(1,1,1).

If Z is a unit Fréchet random variable. Then,

$$Y = \mu + \sigma(Z^{\xi} - 1)/\xi$$

is unit GEV distributed with location, scale and shape parameters equal to μ, σ and ξ respectively.

Extreme precipitation simulations



Threshold of 1 cm.

Discussion and Contributions

Discussion:

- MCMC sampling takes 3+ days!
- Conditional simulation is expensive
- Gaussian elliptical copula is an alternative to a max stable process when some conditions hold

Discussion and Contributions

Discussion:

- MCMC sampling takes 3+ days!
- Conditional simulation is expensive
- Gaussian elliptical copula is an alternative to a max stable process when some conditions hold

Conclusions:

- Bayesian spatial extremes model for return levels
- Can be used for simulations of extremes on a grid at arbitrary resolution
- Composite likelihood and spatially varying regression coefficients make it feasible for larger regions

Discussion and Contributions

Discussion:

- MCMC sampling takes 3+ days!
- Conditional simulation is expensive
- Gaussian elliptical copula is an alternative to a max stable process when some conditions hold

Conclusions:

- Bayesian spatial extremes model for return levels
- Can be used for simulations of extremes on a grid at arbitrary resolution
- Composite likelihood and spatially varying regression coefficients make it feasible for larger regions

Thanks!

References I

R A Pielke, M W Downton, and JZB Miller. Flood damage in the United States, 1926-2000: a reanalysis of National Weather Service estimates, 2002.

F M Ralph, M Dettinger, A White, D Reynolds, D Cayan, T Schneider, R Cifelli, K Redmond, M Anderson, F Gherke, J Jones, K Mahoney, L Johnson, S Gutman, V Chandrasekar, J Lundquist, N Molotch, L Brekke, R Pulwarty, J Horel, L Schick, A Edman, P Mote, J Abatzoglou, R Pierce, and G Wick. A Vision for Future Observations for Western U.S. Extreme Precipitation and Flooding. *Journal of Contemporary Water Research & Education*, 153(1):16–32, April 2014.

B Renard. A Bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, 47 (W11513):1–21, 2011.