Understanding and modeling hydroclimate extremes in the western US

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

Doctoral Dissertation Proposal
July 2015

Motivation - Why study hydroclimate extremes?

- 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.

Extremes in Engineering

Extreme value analysis has been standard practice in civil engineering design and management dating back to [Gumbel 1941].

Extremes in Engineering

Extreme value analysis has been standard practice in civil engineering design and management dating back to [Gumbel 1941].

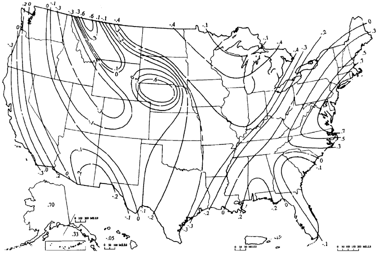
> *"In order to apply any theory we have to suppose that the data are homogeneous, i.e. that no systematical change of climate and no important change in the basin have occurred within the observation period and that no such changes will take place in the period for which extrapolations are made." [Gumbel 1941]*

Extremes in Engineering

Extreme value analysis has been standard practice in civil engineering design and management dating back to [Gumbel 1941].

> *"In order to apply any theory we have to suppose that the data are homogeneous, i.e. that no systematical change of climate and no important change in the basin have occurred within the observation period and that no such changes will take place in the period for which extrapolations are made." [Gumbel 1941]*

Predicated on stationarity.
Elephant in the room: Climate change



Generalized skew coefficients of annual maximum streamflow
Source: Guidelines for determining flood flow frequency, Bulletin 17B, Hydrology Subcommittee, Interagency Advisory Committee on Water Data, March 1982

Drawbacks of traditional methods

Drawbacks of traditional methods

1. **Annual**: Typically only annual maxima are modeled, or maxima in the single wettest season, which masks seasonal variability in extremes.

Drawbacks of traditional methods

1. **Annual**: Typically only annual maxima are modeled, or maxima in the single wettest season, which masks seasonal variability in extremes.
2. **Single site**: Considering spatial dependence among data can reduce overall uncertainty in and allows for interpolation to ungaged location.

Drawbacks of traditional methods

1. **Annual**: Typically only annual maxima are modeled, or maxima in the single wettest season, which masks seasonal variability in extremes.
2. **Single site**: Considering spatial dependence among data can reduce overall uncertainty in and allows for interpolation to ungaged location.
3. **Stationary**: Extreme event distributions are often assumed to be stationary in time, a tenuous assumption under a changing climate.

Drawbacks of traditional methods

1. **Annual**: Typically only annual maxima are modeled, or maxima in the single wettest season, which masks seasonal variability in extremes.
2. **Single site**: Considering spatial dependence among data can reduce overall uncertainty in and allows for interpolation to ungaged location.
3. **Stationary**: Extreme event distributions are often assumed to be stationary in time, a tenuous assumption under a changing climate.
4. **Climate**: Hydroclimate extremes have been increasingly linked to large scale atmospheric ocean oscillations (eg. ENSO)

Drawbacks of traditional methods

1. **Annual**: Typically only annual maxima are modeled, or maxima in the single wettest season, which masks seasonal variability in extremes.
2. **Single site**: Considering spatial dependence among data can reduce overall uncertainty in and allows for interpolation to ungaged location.
3. **Stationary**: Extreme event distributions are often assumed to be stationary in time, a tenuous assumption under a changing climate.
4. **Climate**: Hydroclimate extremes have been increasingly linked to large scale atmospheric ocean oscillations (eg. ENSO)
5. **Uncertainty**: The uncertainty in return level estimates is rarely quantified in practice.

Broad research goal

Develop tools and methodologies to better understand and model hydroclimate extremes in the western US to aid in flood mitigation and water resources planning and management.

Outline

Introduction

Chapter 1 - Spatial variability of seasonal extreme precipitation in the western United States

Crash Courses

Chapter 2 - Efficient hierarchical spatial modeling of seasonal precipitation extremes

Chapter 3 - Coupled hierarchical modeling of streamflow and precipitation extremes

Chapter 4 - Hydroclimate frequency analyses for dam safety: case studies using traditional and modern methodologies

Progress

- Chapter 1 - Published online, May 19 2015, Journal of Geophisical Research: Atmospheres
- Chapter 2 - Under internal review, expected submission this Summer 2015
- Chapter 3 - Fall 2015
- Chapter 4 - Spring 2016

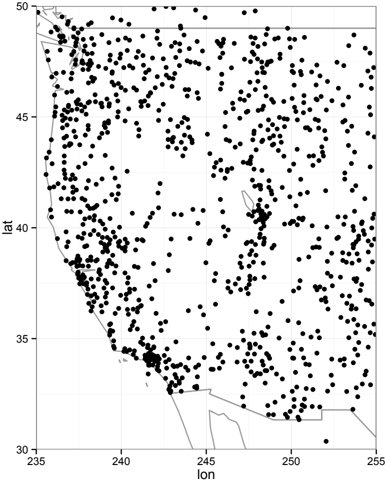
Research questions

- **How are extreme precipitation vary spatially** in large diverse geographic regions like the western US?
- How does extreme precipitation **vary in each season**?
- What are dominant **moisture sources** and **delivery pathways** for extreme events in the west?
- Are extreme events **spatially coherent** and how is this **linked to large scale climate**?

Precipitation Data

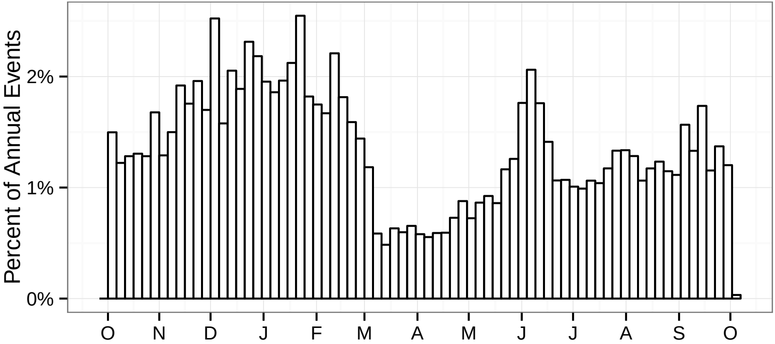
Global Historical Climatology Network (GHCN), daily data from over 90,000 stations worldwide, about 14,000 in the western US.

- Limit to $\sim$1000 stations with near complete data from 1948-2013
- 3 day aggregation window
- Seasonal (Winter, Spring, Summer, Fall) block maxima approach



When does extreme precipitation occur in the western US?

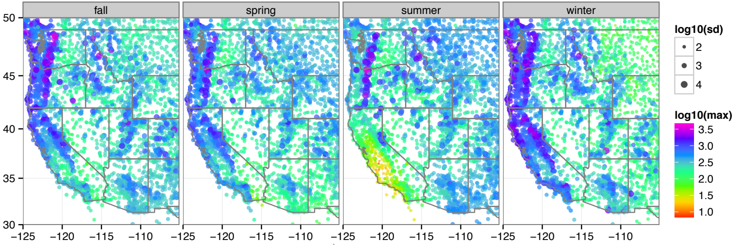
Annual 3day max occurrence day.



Motivates a closer look at how extreme events behave in all seasons.

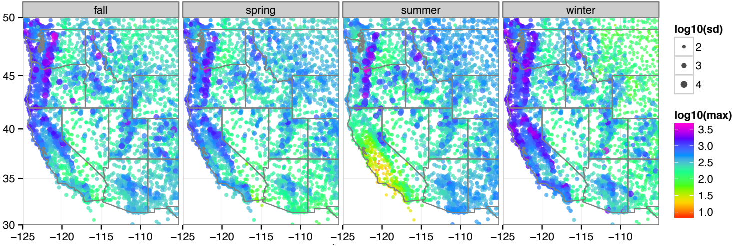
Extreme Event Timing and Magnitude

GHCNd mean 3day precip.

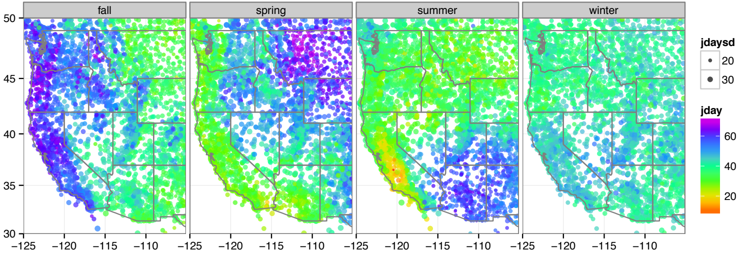


Extreme Event Timing and Magnitude

GHCNd mean 3day precip.



GHCNd mean 3day precip, days since start of season.



Clustering of Maxima

Goal: Objectively determine spatially coherent regions over which extreme values behave similarly.

Clustering of Maxima

Goal: Objectively determine spatially coherent regions over which extreme values behave similarly.

- Bernard et al. 2013 developed an extreme value oriented clustering algorithm specifically tailored to extreme values.
- Dissimilarity matrix based on the F-Madogram (non-parametic, based on shape of empirical CDF)

$$\hat{d}_{ij} = \frac{1}{2T} \sum_{t=1}^{T} \left| \hat{F}_i(M_i^{(t)}) - \hat{F}_j(M_j^{(t)}) \right|, \tag{1}$$

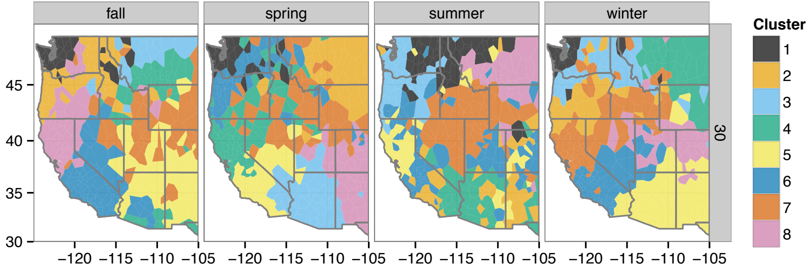
where

$$\hat{F}_i(u) = \frac{1}{T} \sum_{t=1}^{T} \mathbf{1}_{\{M_i^{(t)} \leq u\}} \tag{2}$$

where $\mathbf{1}_{\{M_i^{(t)} \leq u\}}$ is the indicator function for the event $\{M_i^{(t)} \leq u\}$ which returns 1 if the statement is true or 0 otherwise.

Clustering of Maxima

Method was tested on small geographic regions.



Weighted F-Madogram clustering

For large regions, weighting $\hat{d}_{ij}$ on physical distance

$$\hat{\hat{d}}_{ij} = \hat{d}_{ij} + p_{ij}, \tag{3}$$

where

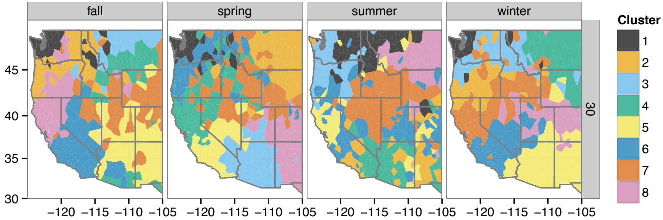
$$p_{ij} = \frac{q_{ij}}{\sum_{i=1}^{N} q_{ij}} \max_{ij} \hat{d}_{ij} \tag{4}$$

and

$$q_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \tag{5}$$

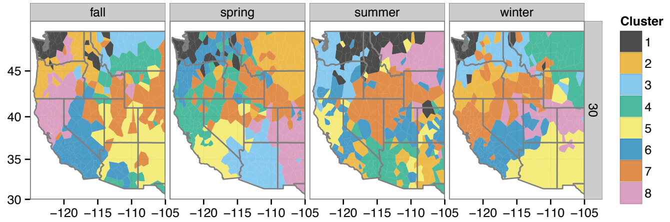
Clustering of Maxima

Original Method (tested on small regions):

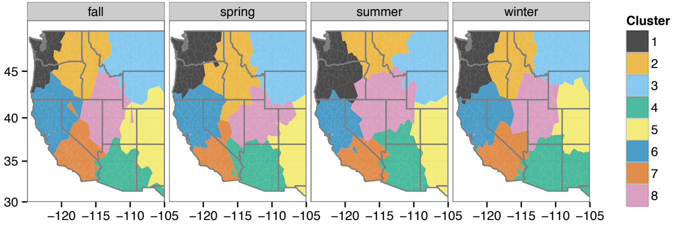


Clustering of Maxima

Original Method (tested on small regions):



Improved clustering method using weighted dissimilarity matrix:



Moisture Source Identification

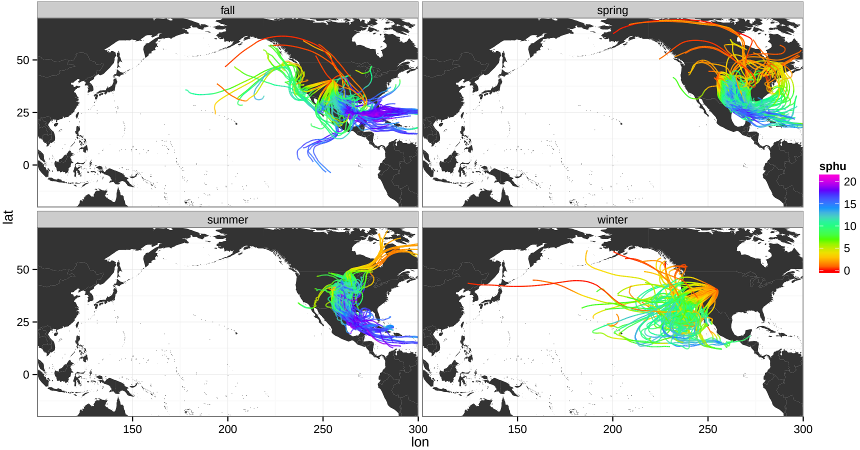
HYSPLIT - Hybrid Single Particle Lagrangian Integrated Trajectory Model

- Inputs: Reanalysis data, starting location, time and height
- Output: Trajectory of air parcel

Runs:

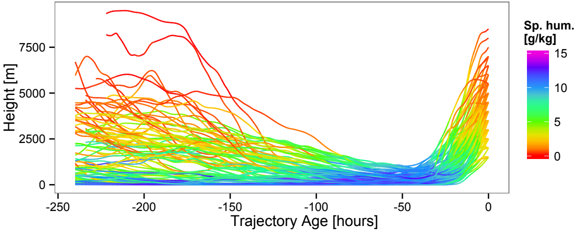
- A single trajectory has a lot of uncertainty, necessitating a statistical approach
- Number of trajectories = 12 time trajectories/event * 21 height trajectories/event * 65 years * 1000 stations * 4 seasons = 65,520,000 trajectories = 15,724,800,000 points
- For each station and season find the top 100 trajectories that lost the most specific humidity during the event period = **400,000 rain trajectories**

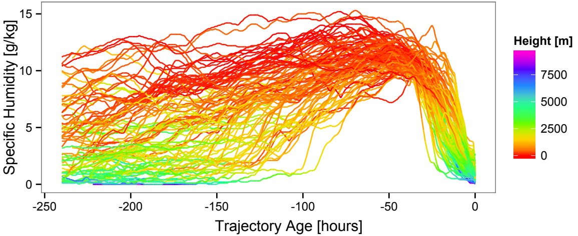
Example rain trajectories – Colorado



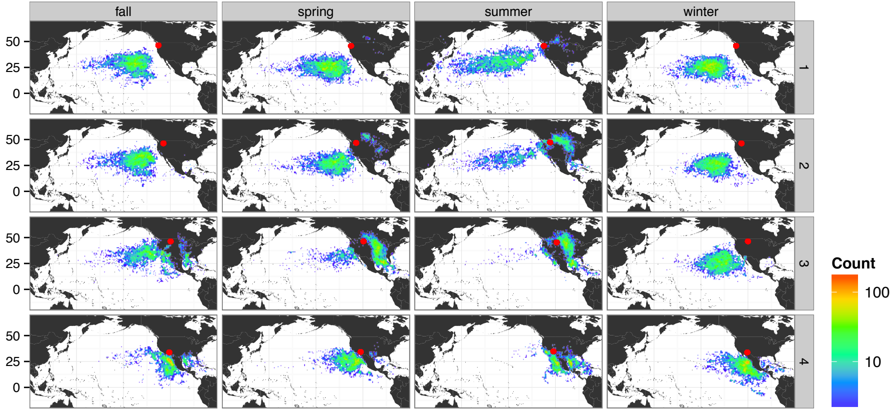
Moisture Source identification

- For each trajectory find the location where peak specific humidity occurred, bin on 1° grid

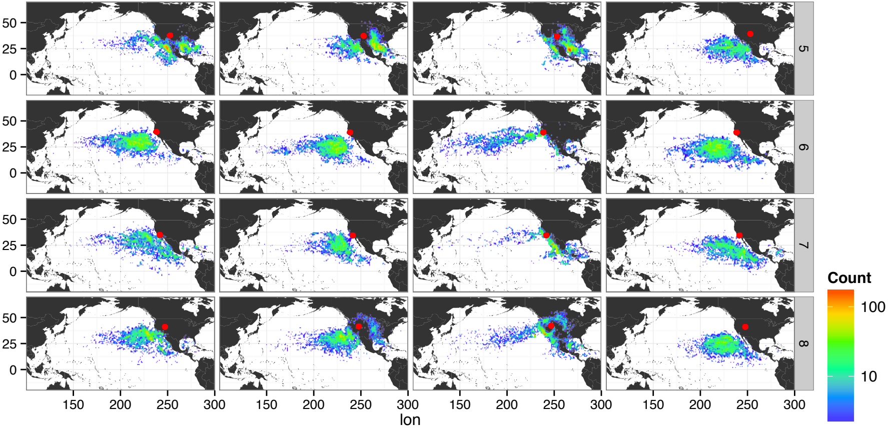




Moisture Source identification



Moisture Source identification

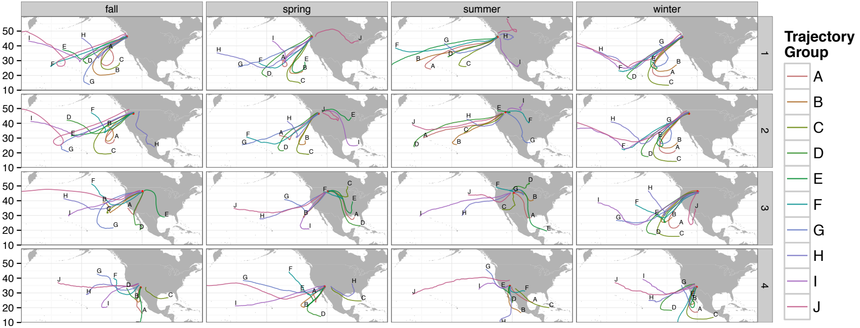


Moisture Pathways

- Longitudinal clustering on trajectories
- Compute mean location at each time point to get "dominant pathways"

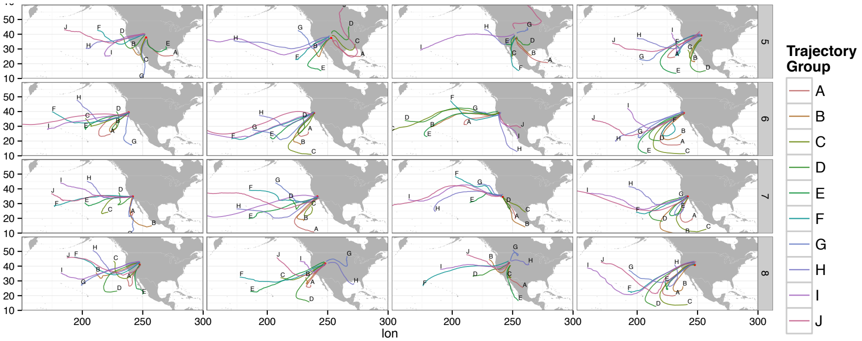
Moisture Pathways

- Longitudinal clustering on trajectories
- Compute mean location at each time point to get "dominant pathways"



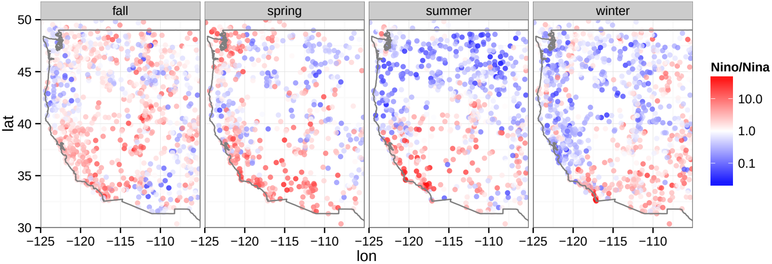
Moisture Pathways

- Longitudinal clustering on trajectories
- Compute mean location at each time point to get "dominant pathways"



ENSO influence

Ratio of rain trajectories occurring in La Nina Vs. El Nino



ENSO tends to affect frequency of extremes, not a strong effect on moisture sources or pathways.

Statistics of Extremes

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

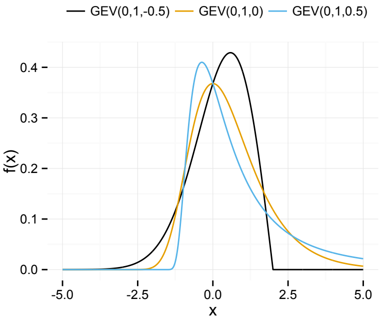
$$\mathrm{GEV}(x;\mu,\sigma,\xi) = \frac{1}{\sigma} b^{(-1/\xi)-1} \exp\left\{-b^{-1/\xi}\right\}$$

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

Return Level (quantile function):

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

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



Spatial Statistics

Typical spatial model

$$\underbrace{Y(\mathrm{s})}_{\text{A}} = \underbrace{\mu(\mathrm{s})}_{\text{B}} + \underbrace{Z(\mathrm{s})}_{\text{C}} + \underbrace{\varepsilon(\mathrm{s})}_{\text{D}}$$

- s – Space

A **Data** – Observations

B **Mean** – $\mu(s) = \mathrm{x}^T \boldsymbol{\beta}$; Covariates and regression parameters

C **Spatial residual** – Random function incorporating spatial correlation

D **Nonspatial residual** – Random variation (nugget effect)

Sometimes $Z(s)$ and $\varepsilon(s)$ are lumped together.

$$Y(s) = \mathrm{x}^T \boldsymbol{\beta} + w(s)$$

Gaussian Processes

$Z(s)$ is a Gaussian process if $[Z(S_1), ..., Z(S_n)] \sim MVN(\boldsymbol{\mu}, \Sigma)$.

$$E[Z(s)] = \mu(s)$$
$$\mathrm{Cov}(Z(s_1), Z(s_2)) = C(s_1, s_2)$$

Common Assumptions:

- Stationary: Covariance is only a function of separation between sites
- Isotropic: Covariance is only a function of distance $||s_i - s_j||$ (also implies $E[Z(s)] = \mu$ is constant)

Example parametric isotropic covariance model (with no nugget):

$$C(r) = \sigma^2 \exp(-r/a) \quad a > 0$$
$$\Sigma(i,j) = \sigma^2 \exp(-||s_i - s_j||/a) \quad a > 0$$

Hierarchical Modeling

Hierarchical statistical models represent relationships between data and parameters in several layers, allowing for flexible model specifications. For example, recall simple linear regression:

$$y_i = \alpha + \beta x_i + \varepsilon_i \text{ where } \varepsilon_i \sim \mathrm{N}(0, \sigma^2)$$

This can be viewed as a hierarchical model with 2 layers:

$$\begin{aligned} y_i | \mu_i, \sigma^2, x_i &\overset{ind.}{\sim} \mathrm{N}(\mu_i, \sigma^2) \\ \mu_i &= \alpha + \beta x_i \end{aligned}$$

Viewing linear regression this way, it is easy to make tweaks to model many kinds of behavior, for example different data distributions such as $y_i \sim \mathrm{GEV}$.

Background on Bayesian Statistics

From Bayes' rule:

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

- θ Parameters
- y Dependent data (response)
- x Independent data (covariates/predictors/constants)

Posterior: The answer, probability distributions of parameters

Liklihood: A computable function of the parameters, model specific

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

The key to building Bayesian models is specifying the likelihood function, same as frequentist.

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

Marginal parameter distributions

In practice, we want marginal distributions of parameters. Consider the simple linear regression case, say we want to marginal posterior distribution for σ^2

$$
\begin{aligned}
p(\sigma^2|\,\mathrm{y}) &= \int\int p(\alpha,\beta,\sigma^2|\,\mathrm{y})p(\alpha)p(\theta)d\alpha d\beta \\
&\propto p(\sigma^2)\int\int\left(\prod_{i=1}^{n} N(\alpha+\beta x_i,\sigma^2)\right)p(\alpha)p(\theta)d\alpha d\beta
\end{aligned}
$$

Even for simple models we get complex integrals to solve, this is where MCMC comes in.

Monte Carlo Markov Chain (MCMC) in a nutshell

- We want to generate random draws from a target distribution (the posterior). We then identify a way to construct a 'nice' Markov chain such that its equilibrium probability distribution is our target distribution.
- Eventually, the draws from the chain will appear as if they are coming from our target distribution.
- There are several ways to construct 'nice' Markov chains (e.g., gibbs sampler, Metropolis-Hastings algorithm).

(explanation from Cross Validated)

MCMC is really a way to approximate integrals (and get marginal distributions).

Research Goals

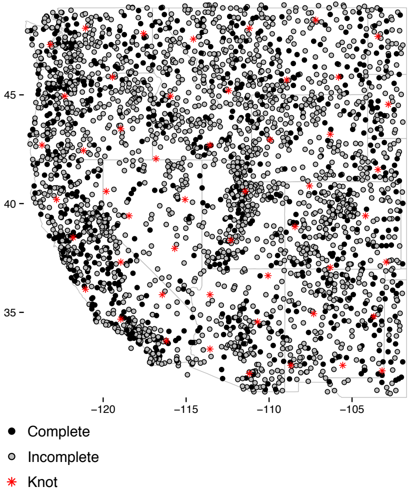
- How do extreme precipitation return levels vary spatially?
- What covariates affect precipitation return levels?
- How can gridded return levels be produced?
- What is the associated uncertainty in these return level estimates?
- How can we model return levels in the entire western US simultaneously?
- How can we incorporate stations with missing data?

Precipitation Data

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

- ∼2500 stations with near complete data from 1948-2013
- 3 day aggregation window
- Seasonal (Winter, Spring, Summer, Fall) block maxima approach

Very large region/dataset for typical Bayesian spatial model



Hierarchical spatial model

Data layer (weather), process layer (climate), priors

$$Y(\mathrm{s})|\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s}) \overset{ind.}{\sim} \mathrm{GEV}[\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s})] \tag{6}$$

Hierarchical spatial model

Data layer (weather), process layer (climate), priors

$$Y(\mathrm{s})|\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s}) \overset{ind.}{\sim} \mathrm{GEV}[\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s})] \quad (6)$$

$$\mu(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\mu + w_\mu(\mathrm{s}) \quad (7)$$

$$\sigma(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\sigma + w_\sigma(\mathrm{s}) \quad (8)$$

$$\boldsymbol{\xi}(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\xi + w_\xi(\mathrm{s}) \quad (9)$$

Hierarchical spatial model

Data layer (weather), process layer (climate), priors

$$Y(\mathrm{s})|\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s}) \overset{ind.}{\sim} \mathrm{GEV}[\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s})] \tag{6}$$

$$\mu(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\mu + w_\mu(\mathrm{s}) \tag{7}$$

$$\sigma(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\sigma + w_\sigma(\mathrm{s}) \tag{8}$$

$$\boldsymbol{\xi}(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\xi + w_\xi(\mathrm{s}) \tag{9}$$

$$w_\mu(\mathrm{s}) \sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\mu)) \tag{10}$$

$$w_\sigma(\mathrm{s}) \sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\sigma)) \tag{11}$$

$$w_\xi(\mathrm{s}) \sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\xi)) \tag{12}$$

Hierarchical spatial model

Data layer (weather), process layer (climate), priors

$$Y(\mathrm{s})|\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s}) \overset{ind.}{\sim} \mathrm{GEV}[\mu(\mathrm{s}),\sigma(\mathrm{s}),\xi(\mathrm{s})] \quad (6)$$

$$\mu(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\mu + w_\mu(\mathrm{s}) \quad (7)$$

$$\sigma(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\sigma + w_\sigma(\mathrm{s}) \quad (8)$$

$$\boldsymbol{\xi}(\mathrm{s}) = \mathrm{x}^T(\mathrm{s})\,\boldsymbol{\beta}_\xi + w_\xi(\mathrm{s}) \quad (9)$$

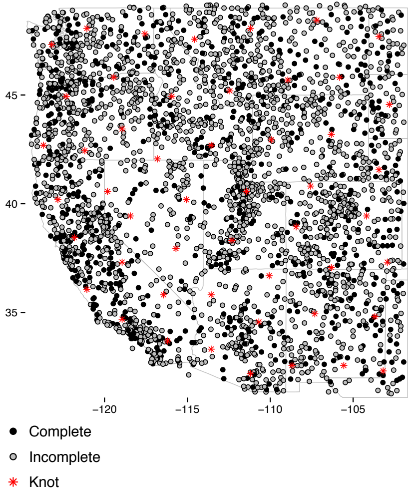
$$w_\mu(\mathrm{s}) \sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\mu)) \quad (10)$$

$$w_\sigma(\mathrm{s}) \sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\sigma)) \quad (11)$$

$$w_\xi(\mathrm{s}) \sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\xi)) \quad (12)$$

Covariates: $x^T(\mathrm{s})$ = (Lat, Lon, Elevation, Mean Seasonal Precip)^T
C: Stationary, isotropic, exponential covariance model,
$\boldsymbol{\theta}_\mu = (\rho_\mu, a_\mu)$: $\Sigma_\mu(i,j;\boldsymbol{\theta}_\mu) = \rho_\mu \exp(-1/a_\mu||\,\mathrm{s}_i - \mathrm{s}_j\,||))$

Gaussian Predictive Process



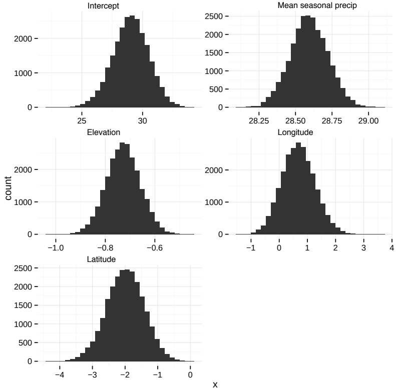
- Computation of GP likelihood requires inversion of large covariance matrix, bottleneck for computation
- Gaussian predictive process estimates w as $\widetilde{w}$ at knot locations $\mathcal{S}^* = \{s_1^*, \ldots, s_n^*\}$.
- Actual process at n sites is approximated at k knots, where $k << n$
- GP value at stations is computed with $\widetilde{w}(s_0) = \mathbf{c}^T C^{*-1}(\boldsymbol{\theta})\, w^*$ where $\mathbf{c}(s_0; \boldsymbol{\theta}) = [C(s_0, s_j^*; \boldsymbol{\theta})]_{j=1}^k$
- First application to spatial extremes

Model fit – GEV location (μ) regression parameters, Winter

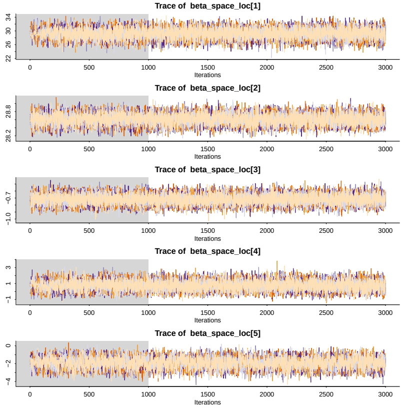
Priors: vague normal priors centered on MLE estimates (N($\cdot$,100))

Fit using MCMC (Stan, NUTS)

Marginal posteriors

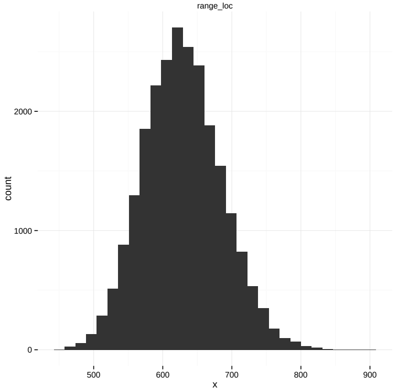


Traceplots

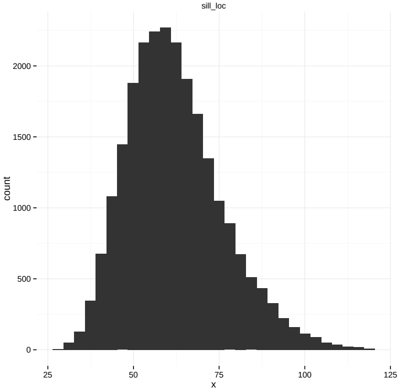


Model fit – Covariance parameters, location, Winter

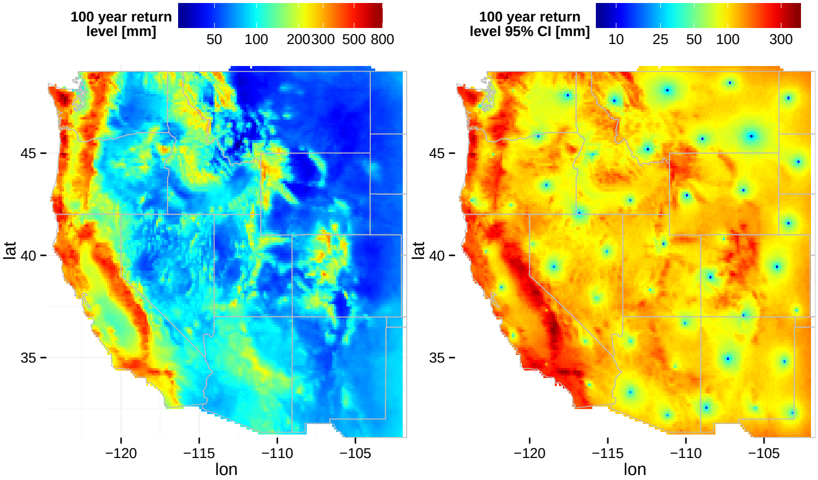
Marginal posterior, Range, a_μ



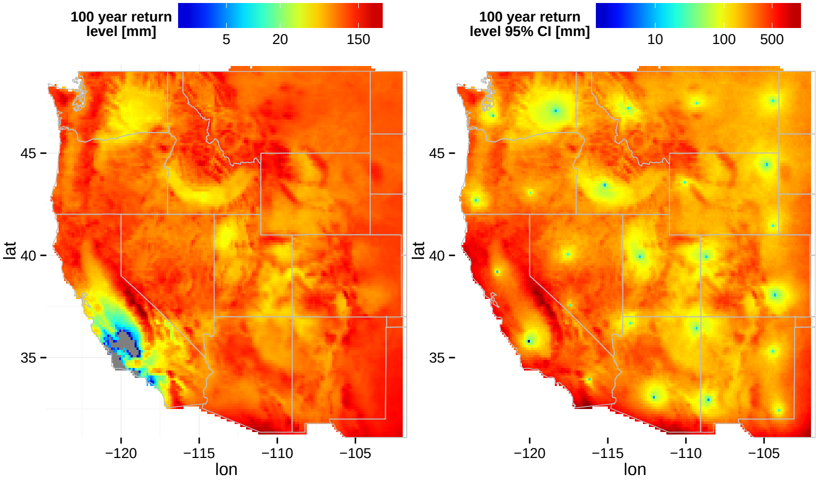
Marginal posterior, Partial sill, ρ_μ



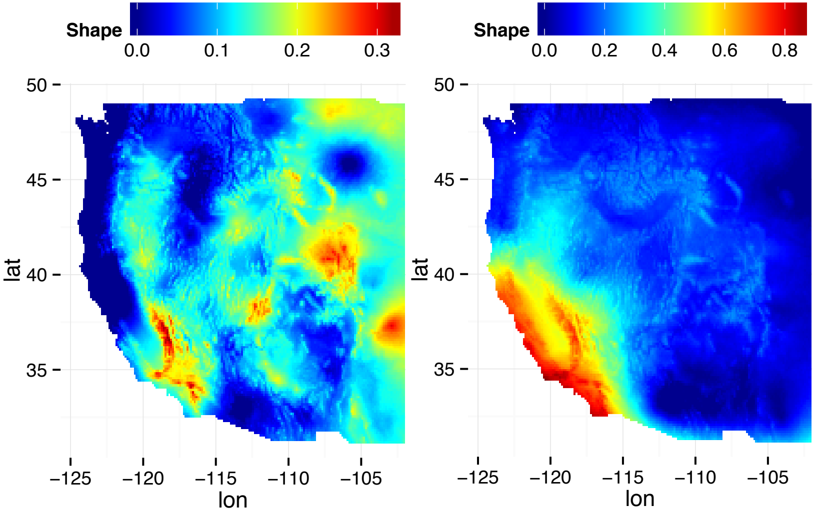
Results – Median return levels, winter



Results – Median return levels, summer



Results – Median shape (ξ), winter, summer



Discussion

- Models converge quickly, 100-200 iterations
- Computation takes 5-10 days!
- Seasonal total precip is the strongest covariate for μ and σ
  - μ and σ are highly correlated
- Elevation is the strongest covariate for ξ
- Credible interval is low near knots

Introduction

Chapter 1 - Spatial variability of seasonal extreme precipitation in the western United States

Crash Courses

Chapter 2 - Efficient hierarchical spatial modeling of seasonal precipitation extremes

Chapter 3 - Coupled hierarchical modeling of streamflow and precipitation extremes

Chapter 4 - Hydroclimate frequency analyses for dam safety: case studies using traditional and modern methodologies

Introduction and Motivation

- There is strong evidence for nonstationarity of streamflow and precipitation extremes in the Western US [Kunkel et al. 2003; Groisman et al. 2001; Kunkel et al. 1999; Cayan et al. 1999]
- Nonstationarity is due to linear trends in time, large scale atmospheric oscillations (AMO, PDO, ENSO), and climate change
- Climate change signal for extremes is not definitive
- There are many examples of nonstationary spatial precipitation and streamflow models but none explicitly consider both
- It is reasonable to assume, especially in rainfall-runoff dominated regions, that annual or seasonal maximum precipitation and streamflow are closely related

Research questions

- How can we simultaneously model precipitation and streamflow return levels in space and time?
- What temporal covariates are appropriate?
- How can we make future projections of precipitation and streamflow return levels?

Multivariate extremes

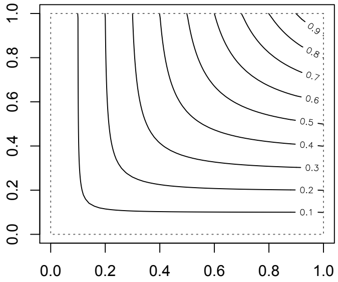
- Previously, conditional independence of the data was assumed given the GEV parameters.
- The data layer should be modeled as coming from a multivariate distribution, but the data is highly non-normal.
- Likelihood computations for multivariate extreme value distributions extremely computationally expensive $O(n^n)$, and prohibitive for $n > 15$ sites [Castruccio et al. 2014].
- **Need a practical tool to model multivariate extremes**

Elliptical copula for data dependence

The Gaussian copula constructs the joint pdf of a random vector $(Y_1, ...Y_m)$ as

$$F_{Gaussian}(y_1, \ldots, y_m) = \Phi_\Sigma(u_1, ...u_m) \tag{13}$$

where $\Phi_\Sigma(u_1, ...u_m)$ is the joint cdf of an m-dimensional multivariate normal distribution with covariance matrix Σ, $u_i = \phi^{-1}(F_i[y_i])$, ϕ is the cdf of the standard normal distribution and F_i is the marginal GEV cdf at site i.



The corresponding joint pdf is

$$f_{Gaussian}(y_1, \ldots, y_m) = \frac{\prod_{i=1}^{m} f_i[y_i]}{\prod_{i=1}^{m} \psi[u_i]} \Psi_\Sigma(u_1, ...u_m) \tag{14}$$

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:

$$\Sigma(i,j) = \exp(-\|\, \mathrm{s}_i - \mathrm{s}_j \,\|/a_0) \tag{15}$$

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

Model Structure

Precipitation model:

$$y(\mathrm{s},t) \sim Gcop_m[\Sigma; \mu_y(\mathrm{s},t), \sigma_y(\mathrm{s},t), \xi_y] \tag{16}$$

Model Structure

Precipitation model:

$$y(\mathrm{s},t) \sim Gcop_m[\Sigma; \mu_y(\mathrm{s},t), \sigma_y(\mathrm{s},t), \xi_y] \tag{16}$$

$$\mu_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\mu,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\mu,t} + \mathrm{w}_\mu(\mathrm{s},t) \tag{17}$$

Model Structure

Precipitation model:

$$y(\mathrm{s},t) \sim Gcop_m[\Sigma; \mu_y(\mathrm{s},t), \sigma_y(\mathrm{s},t), \xi_y] \tag{16}$$

$$\mu_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\mu,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\mu,t} + \mathrm{w}_\mu(\mathrm{s},t) \tag{17}$$

$$\sigma_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\sigma,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\sigma,t} + \mathrm{w}_\sigma(\mathrm{s},t) \tag{18}$$

Model Structure

Precipitation model:

$$y(\mathrm{s},t) \sim Gcop_m[\Sigma; \mu_y(\mathrm{s},t), \sigma_y(\mathrm{s},t), \xi_y] \tag{16}$$

$$\mu_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\mu,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\mu,t} + \mathrm{w}_\mu(\mathrm{s},t) \tag{17}$$

$$\sigma_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\sigma,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\sigma,t} + \mathrm{w}_\sigma(\mathrm{s},t) \tag{18}$$

Streamflow model:

$$z(t) \sim \mathrm{GEV}[\mu_z(t), \sigma_z(t), \xi_z] \tag{19}$$

Model Structure

Precipitation model:

$$y(\mathrm{s}, t) \sim Gcop_m[\Sigma; \mu_y(\mathrm{s}, t), \sigma_y(\mathrm{s}, t), \xi_y] \tag{16}$$

$$\mu_y(\mathrm{s}, t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\, \boldsymbol{\beta}_{\mu,\mathrm{s}} + \mathrm{x}_t^T(t)\, \boldsymbol{\beta}_{\mu,t} + \mathrm{w}_\mu(\mathrm{s}, t) \tag{17}$$

$$\sigma_y(\mathrm{s}, t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\, \boldsymbol{\beta}_{\sigma,\mathrm{s}} + \mathrm{x}_t^T(t)\, \boldsymbol{\beta}_{\sigma,t} + \mathrm{w}_\sigma(\mathrm{s}, t) \tag{18}$$

Streamflow model:

$$z(t) \sim \mathrm{GEV}[\mu_z(t), \sigma_z(t), \xi_z] \tag{19}$$

$$\mu_z(t) = b_\mu \mu_y(\mathrm{s}^*, t) + c_\mu + \varepsilon_\mu \tag{20}$$

Model Structure

Precipitation model:

$$y(\mathrm{s},t) \sim Gcop_m[\Sigma; \mu_y(\mathrm{s},t), \sigma_y(\mathrm{s},t), \xi_y] \tag{16}$$

$$\mu_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\mu,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\mu,t} + \mathrm{w}_\mu(\mathrm{s},t) \tag{17}$$

$$\sigma_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\sigma,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\sigma,t} + \mathrm{w}_\sigma(\mathrm{s},t) \tag{18}$$

Streamflow model:

$$z(t) \sim \mathrm{GEV}[\mu_z(t), \sigma_z(t), \xi_z] \tag{19}$$

$$\mu_z(t) = b_\mu \mu_y(\mathrm{s}^*, t) + c_\mu + \varepsilon_\mu \tag{20}$$

$$\sigma_z(t) = b_\sigma \sigma_y(\mathrm{s}^*, t) + c_\sigma + \varepsilon_\sigma \tag{21}$$

Model Structure

Precipitation model:

$$y(\mathrm{s},t) \sim Gcop_m[\Sigma; \mu_y(\mathrm{s},t), \sigma_y(\mathrm{s},t), \xi_y] \tag{16}$$

$$\mu_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\mu,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\mu,t} + \mathrm{w}_\mu(\mathrm{s},t) \tag{17}$$

$$\sigma_y(\mathrm{s},t) = \mathrm{x}_\mathrm{s}^T(\mathrm{s})\,\boldsymbol{\beta}_{\sigma,\mathrm{s}} + \mathrm{x}_t^T(t)\,\boldsymbol{\beta}_{\sigma,t} + \mathrm{w}_\sigma(\mathrm{s},t) \tag{18}$$

Streamflow model:

$$z(t) \sim \mathrm{GEV}[\mu_z(t), \sigma_z(t), \xi_z] \tag{19}$$

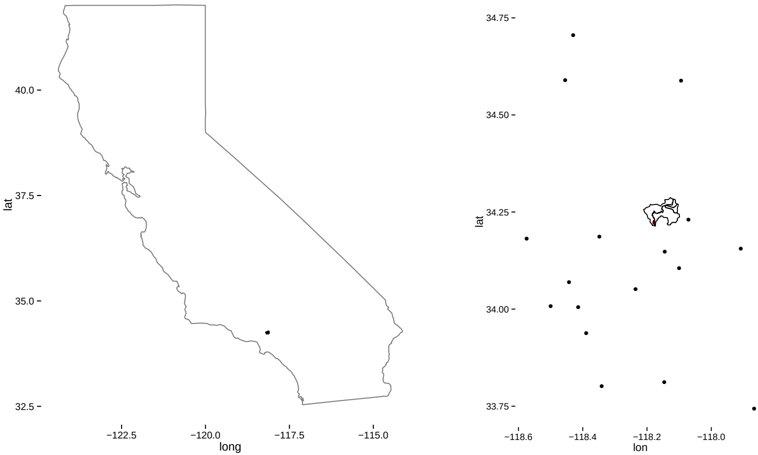
$$\mu_z(t) = b_\mu \mu_y(\mathrm{s}^*,t) + c_\mu + \varepsilon_\mu \tag{20}$$

$$\sigma_z(t) = b_\sigma \sigma_y(\mathrm{s}^*,t) + c_\sigma + \varepsilon_\sigma \tag{21}$$

$\mathrm{w}_\mu \sim \ \mathbf{GP}(0, C(\boldsymbol{\theta}_\mu))$, $\mathrm{w}_\sigma \sim \ \mathbf{GP}(0, C(\boldsymbol{\theta}_\sigma))$
$\varepsilon_\mu \sim \mathrm{N}(0, \sigma_\mu)$, $\varepsilon_\sigma \sim \mathrm{N}(0, \sigma_\sigma)$
s^* - Streamflow region $b_\mu, c_\mu, b_\sigma, c_\sigma$ - Regression coefficients

Study Region

Small HCDN basin in Southern California, rainfall runoff dominated



Expected Results

- Nonstationary gridded return levels
- Future projections of precipitation and streamflow return levels and uncertainty quantification

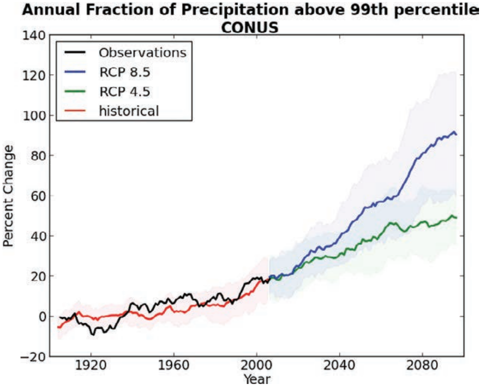


Figure from [Wuebbles et al. 2014].

Motivation

- Reclamation regularly performs hydroclimate frequency analyses related to dam safety
- These studies are aimed at providing return levels for precipitation, flow and water elevation in particular reservoirs

RECLAMATION
Managing Water in the West

Friant Dam Hydrologic Hazard for Issue Evaluation

Central Valley Project, CA
Mid-Pacific Region

U.S. Department of the Interior
Bureau of Reclamation

September 2013

The current approach

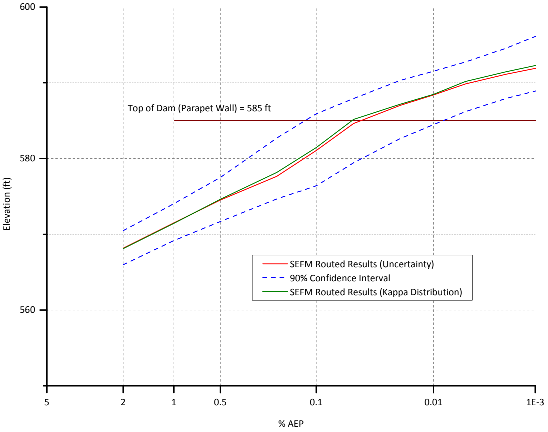
The current approach:

1. Develop precipitation and flood frequency curves for the basin (by assuming some probability distribution for the data).
2. Develop reservoir elevation frequency curve by routing peak flow event hydrographs.
3. Quantify uncertainty with resampling approach (via Latin Hypercube Sampling for example).

Drawbacks of the current approach

Drawbacks:

- Frequency curves for precipitation, flow and reservoir elevation are estimated independently, making uncertainty propagation difficult
- Return levels are developed under an assumption of temporal stationarity



Proposed approach and Expected results

Proposed approach:

- Hierarchical nonstationary spatial model for precipitation, flow and reservoir elevation, similar to Chapter 3
- Quantification of uncertainty, compare to previous dam safety studies

Expected results:

- Extreme flow, precipitation and reservoir elevation frequency curves for several dams in the western US
- Direct comparison to existing dam safety frequency analyses
- Suggested methodology for conducting dam safety frequency analyses

Progress

- Chapter 1 - Published online, May 19 2015, Journal of Geophysical Research: Atmospheres
- Chapter 2 - Under internal review, expected submission this Summer 2015
- Chapter 3 - Fall 2015
- Chapter 4 - Spring 2016
- Defense - Spring 2016

Stefano Castruccio, Raphaël Huser, and Marc Genton. High-order Composite Likelihood Inference for Max-Stable Distributions and Processes. November 2014.

Daniel R Cayan, Kelly T Redmond, and Laurence G Riddle. ENSO and Hydrologic Extremes in the Western United States*. *Journal of Climate*, 12(9):2881–2893, September 1999.

Pavel Ya Groisman, Richard W Knight, and Thomas R Karl. Heavy Precipitation and High Streamflow in the Contiguous United States: Trends in the Twentieth Century. *Bull. Amer. Meteor. Soc.*, 82(2):219–246, February 2001.

E J Gumbel. The Return Period of Flood Flows. *The annals of mathematical statistics*, 1941.

Kenneth E Kunkel, Karen Andsager, and David R Easterling. Long-Term Trends in Extreme Precipitation Events over the Conterminous United States and Canada. *Journal of Climate*, 12 (8):2515–2527, August 1999.

Kenneth E Kunkel, David R Easterling, Kelly Redmond, and Kenneth Hubbard. Temporal variations of extreme precipitation events in the United States: 1895–2000. *Geophysical Research Letters*, 30(17):1900, September 2003.

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*, 2011.

Donald Wuebbles, Gerald Meehl, Katharine Hayhoe, Thomas R Karl, Kenneth Kunkel, Benjamin Santer, Michael Wehner, Brian

Colle, Erich M Fischer, Rong Fu, Alex Goodman, Emily Janssen, Viatcheslav Kharin, Huikyo Lee, Wenhong Li, Lindsey N Long, Seth C Olsen, Zaitao Pan, Anji Seth, Justin Sheffield, and Liqiang Sun. *CMIP5 Climate Model Analyses: Climate Extremes in the United States*, volume 95. American Meteorological Society, June 2014.

Results – Return Levels

To compute posteriors of return levels on a grid:

1. For each GEV paramter, take a posterior set of GP values for at knot locations and covariance parameters
2. Conditionally simulate on $1/8\text{th}^{\circ}$ grid
3. Compute r-year return level (100 year for example) at each grid cell
4. Repeat for each posterior parameter set for posteriors on return level

Parameters/Likelihood/Priors

$$\Omega = (\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_\sigma, \boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\sigma, \boldsymbol{\theta}_\xi, \mathrm{w}^*_\mu, \mathrm{w}^*_\sigma, \mathrm{w}^*_\xi)$$

- $\boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_\xi$ – Regression parameters
- $\boldsymbol{\theta}_\mu, \boldsymbol{\theta}_\sigma, \boldsymbol{\theta}_\xi$ – Cov parameters: marginal variance, range, small fixed nugget for numerical stability
- $\mathrm{w}^*_\mu, \mathrm{w}^*_\sigma, \mathrm{w}^*_\xi$ – GP at knot locations $\mathrm{w}^*_\mu \sim \mathrm{MVN}(\mathbf{0}, \Sigma_\mu)$

Posterior probability:

$$\begin{aligned} p(\Omega|Y) \propto\, & p(Y(\mathrm{s}_i)|\, \mathrm{w}^*_\mu, \mathrm{w}^*_\sigma, \mathrm{w}^*_\xi, \boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_\xi, \mathrm{x}^T(\mathrm{s}))\cdot \\ & p(\mathrm{w}^*_\mu \mid \boldsymbol{\theta}_\mu)\, p(\mathrm{w}^*_\sigma \mid \boldsymbol{\theta}_\sigma)\, p(\mathrm{w}^*_\xi \mid \boldsymbol{\theta}_\xi)\cdot \\ & p(\boldsymbol{\beta}_\mu)\, p(\boldsymbol{\beta}_\sigma)\, p(\boldsymbol{\beta}_\xi)\cdot \\ & p(\boldsymbol{\theta}_\mu)\, p(\boldsymbol{\theta}_\sigma)\, p(\boldsymbol{\theta}_\xi) \end{aligned}$$

Parameters/Likelihood/Priors

$$log(p(Y(\mathrm{s}_i)|\,\mathrm{w}_\mu^*, \mathrm{w}_\sigma^*, \mathrm{w}_\xi^*, \boldsymbol{\beta}_\mu, \boldsymbol{\beta}_\sigma, \boldsymbol{\beta}_\xi, \mathrm{x}^T(\mathrm{s}))) = \sum_{s=1}^{n} \sum_{t} = 1^t log(GEV(y(s,$$

$$p(\mathrm{w}_\mu^* \,|\, \boldsymbol{\theta}_\mu) = MVN(\mathbf{0}, \Sigma_\mu)$$