Characterizing and modeling hydroclimate extremes in the western US

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

Dissertation Defense
April 7, 2016

Dissertation Advised by Prof. Balaji Rajagopalan

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

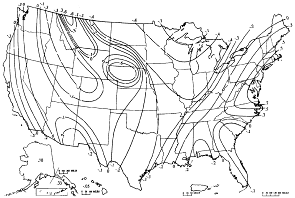
Extreme value analysis has been standard practice in civil engineering design and management dating back to [Gumbel, 1941].

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]*

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

Improvements over traditional methods

1. **Seasonal**: Sub-annual variability. (Chapter 1)
2. **Spatial**: Spatial dependence. (Chapter 2)
3. **Nonstationary**: Due to a changing climate. (Chapter 3,4)
4. **Multivariate**: Joint behavior of multiple hydrologic variables (Chapter 3,4)
5. **Uncertainty**: The uncertainty in return level estimates is rarely quantified in practice. (All chapters)

Broad research goal

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

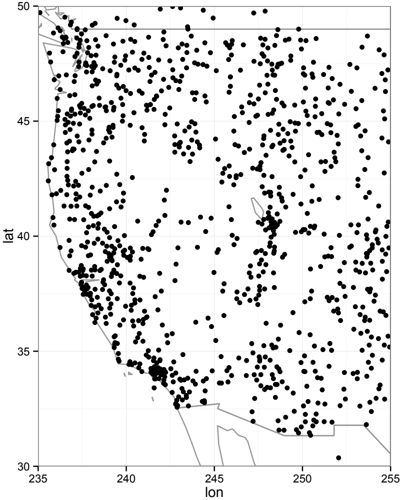
Research questions

- How does extreme precipitation vary **spatially and seasonally** in large diverse geographic regions like the western US?
- 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

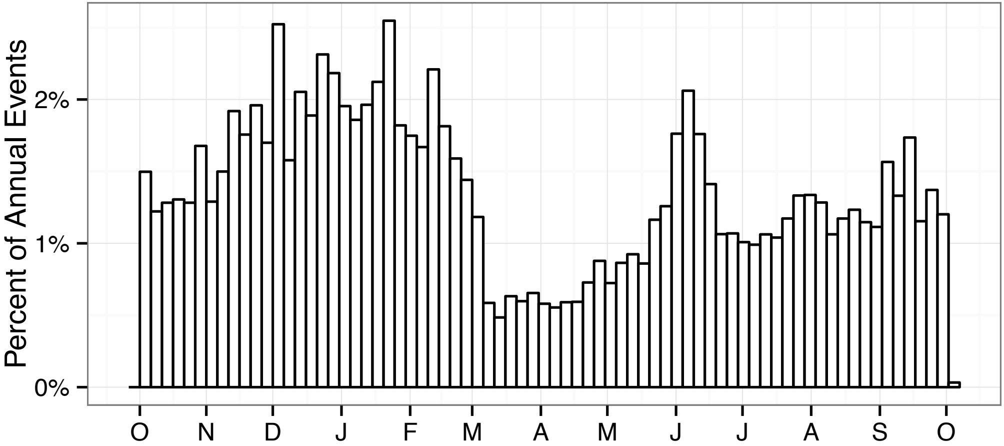
Global Historical Climatology Network (GHCN), daily data from over 90,000 stations worldwide, about 14,000 in the western US.

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



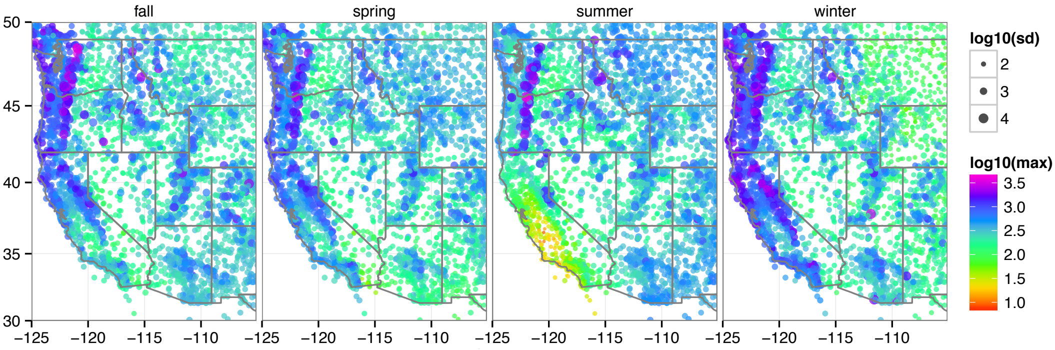
When does extreme precipitation occur in the western US?

Annual 3day max occurrence day.



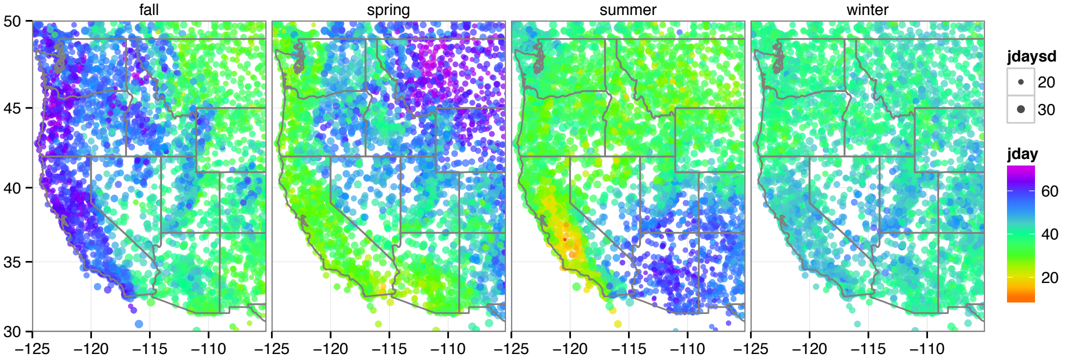
Extreme Precipitation Magnitude

GHCNd mean 3day precip.



Extreme Precipitation Timing

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



Clustering of Maxima

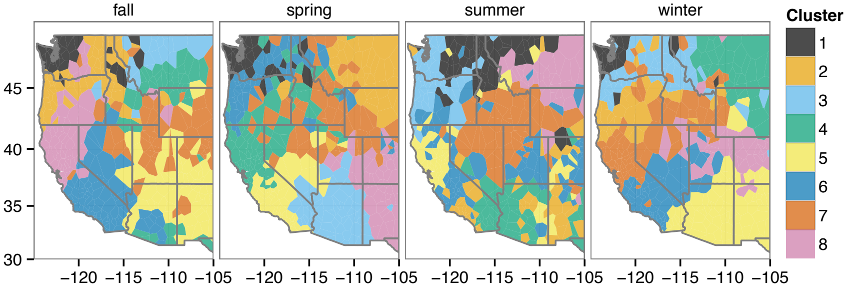
Goal: Objectively determine spatially coherent regions over which extreme values behave similarly.

Clustering of Maxima

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

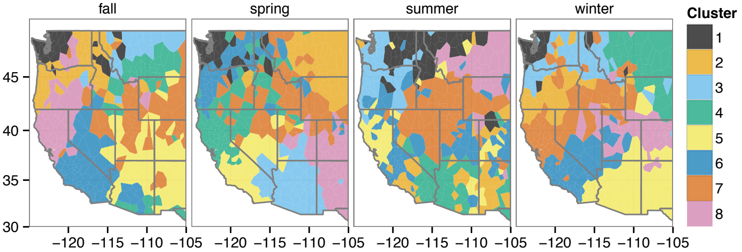
- Bernard et al. 2013 developed an extreme value oriented clustering algorithm specifically tailored to extreme values.
- Dissimilarity matrix based on the F-Madogram (non-parametic, based on shape of empirical CDF)

Method was tested on small geographic regions.

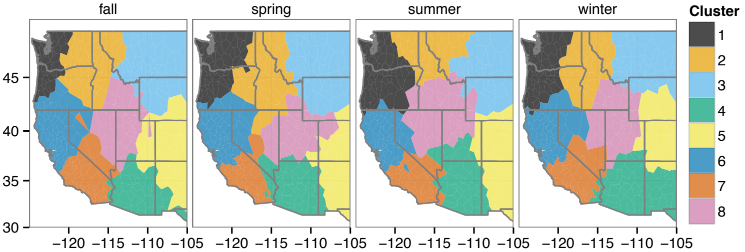


Distance weighted F-Madogram clustering

Original Method (tested on small regions):



Improved clustering method using weighted dissimilarity matrix:



Moisture Source Identification

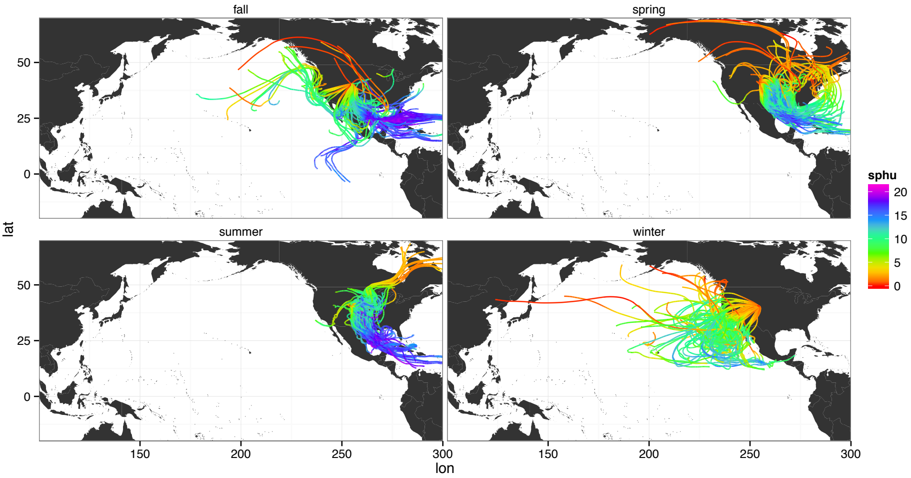
HYSPLIT - Hybrid Single Particle Lagrangian Integrated Trajectory Model

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

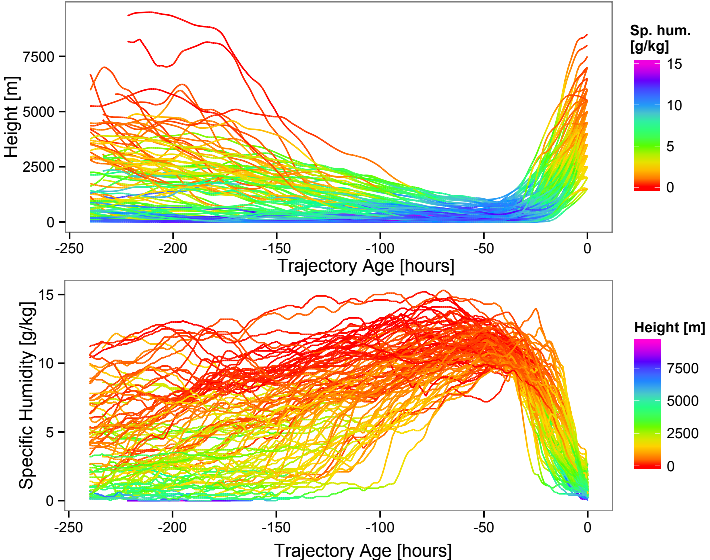
Runs:

- 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**

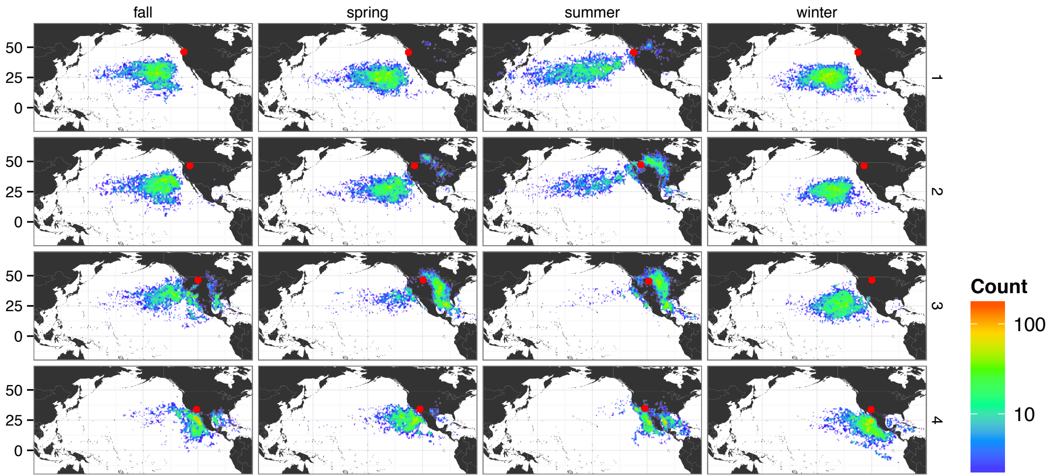
Example rain trajectories – Colorado



Moisture Source identification



Moisture Source identification

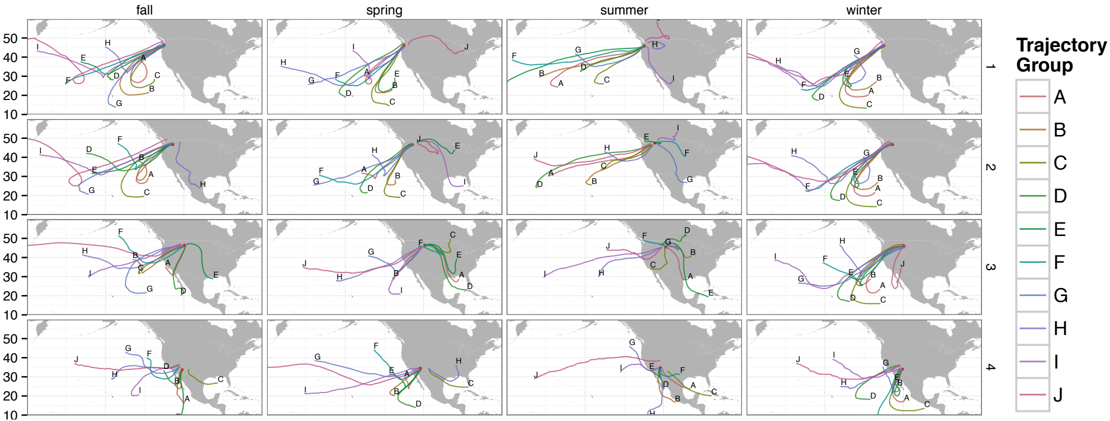


Moisture Pathways

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

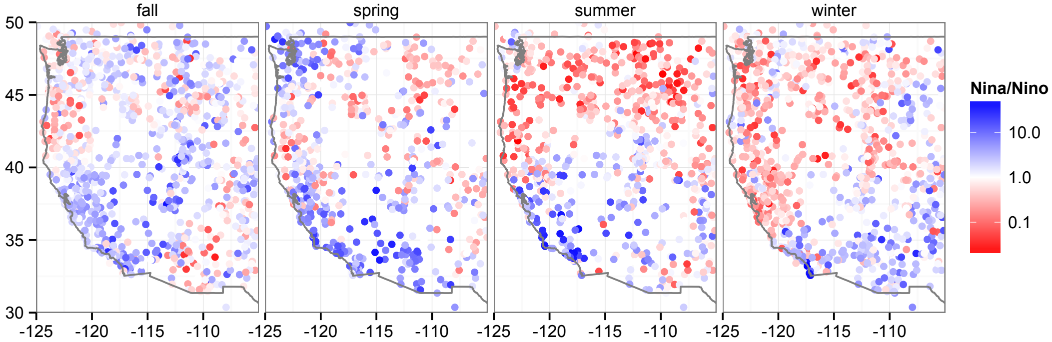
Moisture Pathways

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



ENSO influence

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



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

Contributions - Chapter 1

- Seasonal approach reveals important differences in the genesis of extreme precip between seasons
- Improved extreme clustering method for large geographic regions
- Seasonally dependent extreme regions
- Seasonally dependent moisture sources and pathways
- ENSO effect on frequency of extreme precip – agrees with previous results

Introduction

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

Chapter 2 - Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain

Chapter 3 - Joint Bayesian hierarchical nonstationary regional frequency analysis of precipitation and streamflow extremes

Chapter 4 - A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis

Research Goals

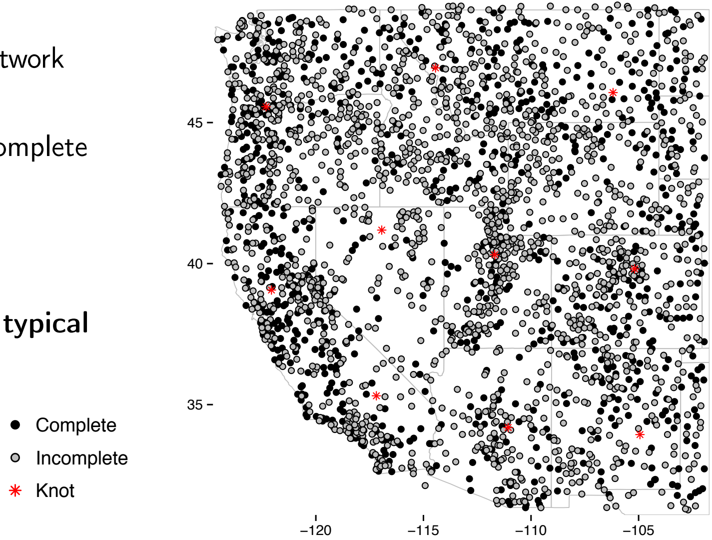
- How do extreme precipitation return levels vary spatially?
- How can gridded return levels be produced?
- What is the associated uncertainty in these return level estimates?
- How can we model return levels in the entire western US simultaneously?

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



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

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

Statistics of Extremes

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

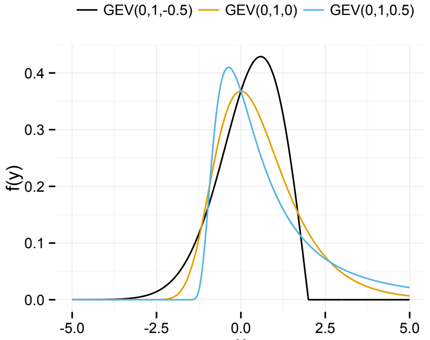
$$\mathrm{GEV}(y;\mu,\sigma,\xi) = \frac{1}{\sigma} b^{(-1/\xi)-1} \exp\left\{-b^{-1/\xi}\right\}$$

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

Return Level (quantile function):

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

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



Hierarchical spatial model

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

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

$$
\begin{aligned}
(Y(\mathbf{s}_1,t),\ldots,Y(\mathbf{s}_n,t)) &\sim C_g[\Sigma;\{\mu(\mathbf{s}),\sigma(\mathbf{s}),\xi(\mathbf{s})\}]\\
Y(\mathbf{s},t) &\sim GEV[\mu(\mathbf{s}),\sigma(\mathbf{s}),\xi(\mathbf{s})]\\
\mu(\mathbf{s}) &= \beta_{\mu,0}+\mathbf{x}_\mu^T(\mathbf{s})\boldsymbol{\beta}_\mu(\mathbf{s})+w_\mu(\mathbf{s})\\
\sigma(\mathbf{s}) &= \beta_{\sigma,0}+\mathbf{x}_\sigma^T(\mathbf{s})\boldsymbol{\beta}_\sigma(\mathbf{s})+w_\sigma(\mathbf{s})\\
\xi(\mathbf{s}) &= \beta_{\xi,0}+\mathbf{x}_\xi^T(\mathbf{s})\boldsymbol{\beta}_\xi(\mathbf{s})+w_\xi(\mathbf{s})\\
\boldsymbol{\beta}_\mu(\mathbf{s}) &= \sum_{i=1}^{k} c_i^\mu \eta_i^\mu(\mathbf{s}) \ \ \text{(Similarly for } \sigma,\xi)
\end{aligned}
$$

Hierarchical spatial model

$$
\begin{aligned}
(Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t)) &\sim C_g[\Sigma; \{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}] \\
Y(\mathbf{s}, t) &\sim GEV[\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})] \\
\mu(\mathbf{s}) &= \beta_{\mu,0} + \mathbf{x}_\mu^T(\mathbf{s})\boldsymbol{\beta}_\mu(\mathbf{s}) + w_\mu(\mathbf{s}) \\
\sigma(\mathbf{s}) &= \beta_{\sigma,0} + \mathbf{x}_\sigma^T(\mathbf{s})\boldsymbol{\beta}_\sigma(\mathbf{s}) + w_\sigma(\mathbf{s}) \\
\xi(\mathbf{s}) &= \beta_{\xi,0} + \mathbf{x}_\xi^T(\mathbf{s})\boldsymbol{\beta}_\xi(\mathbf{s}) + w_\xi(\mathbf{s}) \\
\boldsymbol{\beta}_\mu(\mathbf{s}) &= \sum_{i=1}^{k} c_i^\mu \eta_i^\mu(\mathbf{s}) \;\; \text{(Similarly for } \sigma, \xi\text{)} \\
w_\mu(\mathbf{s}) &\sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\mu)) \;\; \text{(Similarly for } \sigma, \xi\text{)}
\end{aligned}
$$

Hierarchical spatial model

$$
\begin{aligned}
(Y(\mathbf{s}_1, t), \ldots, Y(\mathbf{s}_n, t)) &\sim C_g[\Sigma; \{\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})\}] \\
Y(\mathbf{s}, t) &\sim GEV[\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})] \\
\mu(\mathbf{s}) &= \beta_{\mu,0} + \mathbf{x}_\mu^T(\mathbf{s})\boldsymbol{\beta}_\mu(\mathbf{s}) + w_\mu(\mathbf{s}) \\
\sigma(\mathbf{s}) &= \beta_{\sigma,0} + \mathbf{x}_\sigma^T(\mathbf{s})\boldsymbol{\beta}_\sigma(\mathbf{s}) + w_\sigma(\mathbf{s}) \\
\xi(\mathbf{s}) &= \beta_{\xi,0} + \mathbf{x}_\xi^T(\mathbf{s})\boldsymbol{\beta}_\xi(\mathbf{s}) + w_\xi(\mathbf{s}) \\
\boldsymbol{\beta}_\mu(\mathbf{s}) &= \sum_{i=1}^{k} c_i^\mu \eta_i^\mu(\mathbf{s}) \ \text{(Similarly for } \sigma, \xi) \\
w_\mu(\mathbf{s}) &\sim GP(\mathbf{0}, C(\boldsymbol{\theta}_\mu)) \ \text{(Similarly for } \sigma, \xi)
\end{aligned}
$$

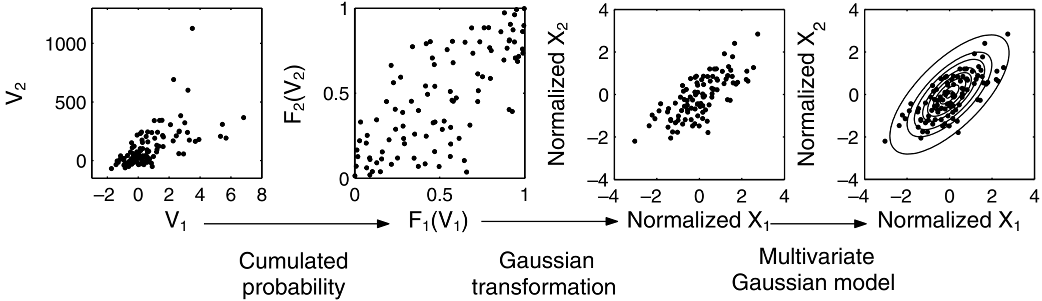
Covariates: $x^T(\mathrm{s})$ = (Elevation, Mean Seasonal Precip)^T
C: Stationary, isotropic, exponential covariance model

Elliptical copula for data dependence

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

$$F_C(v_1, \ldots, v_n) = \Phi_\Sigma(u_1, \ldots, u_n) \quad (1)$$

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

Exponential dependence matrix for spatial data:

$$\Sigma(i,j) = \exp(-||\,s_i - s_j\,||/a_0) \quad (2)$$

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

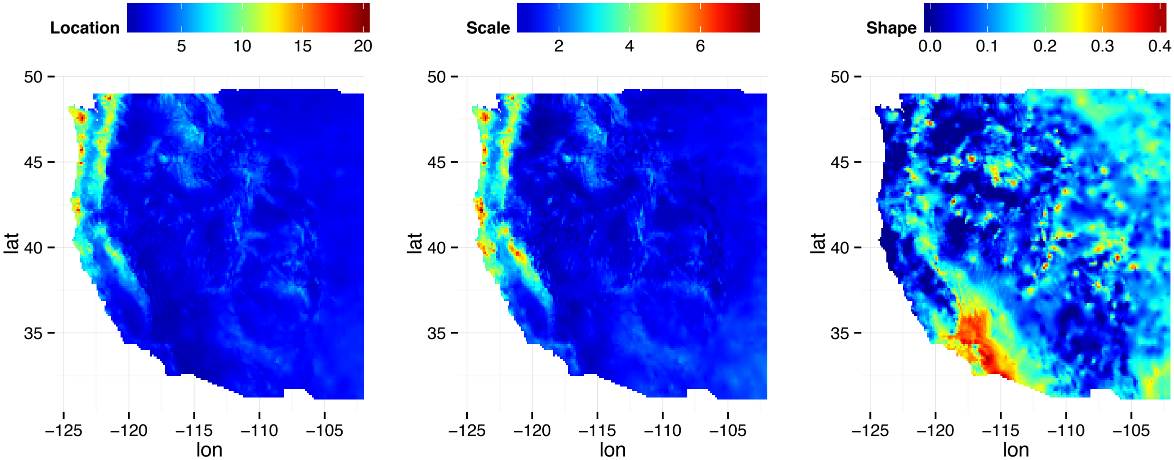
Composite Likelihood (CL)

- 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

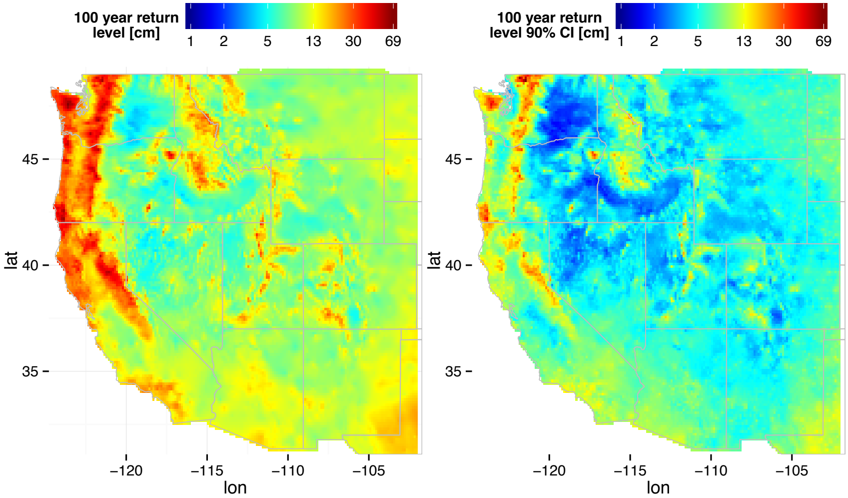
$$L_c(\boldsymbol{\theta}|\mathbf{y}_1, \ldots, \mathbf{y}_G) = \prod_{g=1}^{G} L_g(\boldsymbol{\theta}|\mathbf{y}_g) \tag{3}$$

- 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.

Results – Median GEV parameters, fall



Results – Median 100 year return levels, fall



Discussion and Contributions

Discussion:

- Computation takes 3-5 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

Contributions:

- Bayesian spatial extremes model for large regions with thousands of stations
- Composite likelihood
- Spatially varying regression coefficients
- Gridded return levels for any return period at any resolution

Introduction

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

Chapter 2 - Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain

Chapter 3 - Joint Bayesian hierarchical nonstationary regional frequency analysis of precipitation and streamflow extremes

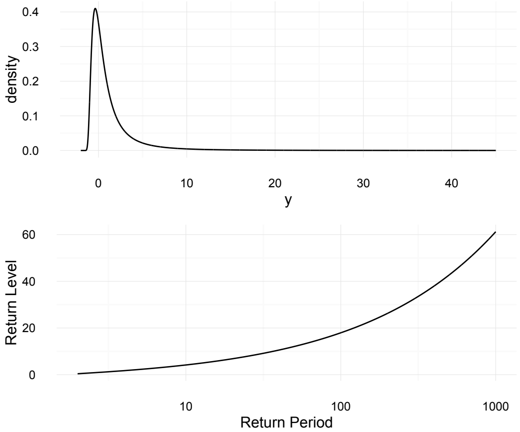
Chapter 4 - A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis

What is hydrologic frequency analysis?

Process of estimating recurrence probabilities of rare hydrologic events (floods, heavy rainfall, etc.).

General procedure:

1. Generate extreme data. For example take the maximum daily flow value from each year from a daily flow dataset.
2. Fit a probability distribution. For example generalized extreme value.
3. Compute return levels (quantiles). A 100-year return level will be the (1-1/100)th quantile.



Motivation for nonstationary frequency analysis

- 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

Lay of the land - Bayesian hierarchical frequency analysis

- Bayesian hierarchical modeling of precipitation and streamflow extremes
  - Active area of research in the last 10-15 years
  - Alternative to regional frequency analysis
  - End goal is to estimate distributions of return levels
- Hydrologic frequency models come in many flavors
  - Single site and spatial
  - Stationary and nonstationary
- Typically these analyses are conducted independently

Lay of the land - Bayesian hierarchical frequency analysis

- Bayesian hierarchical modeling of precipitation and streamflow extremes
  - Active area of research in the last 10-15 years
  - Alternative to regional frequency analysis
  - End goal is to estimate distributions of return levels
- Hydrologic frequency models come in many flavors
  - Single site and spatial
  - Stationary and nonstationary
- Typically these analyses are conducted independently
- **How should a joint frequency analysis be conducted?**

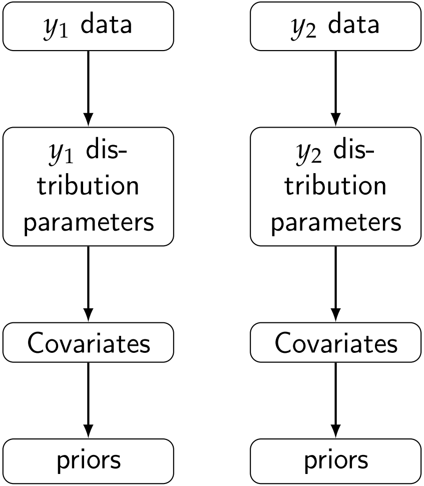
Lay of the land - Bayesian hierarchical frequency analysis

- Bayesian hierarchical modeling of precipitation and streamflow extremes
  - Active area of research in the last 10-15 years
  - Alternative to regional frequency analysis
  - End goal is to estimate distributions of return levels
- Hydrologic frequency models come in many flavors
  - Single site and spatial
  - Stationary and nonstationary
- Typically these analyses are conducted independently
- **How should a joint frequency analysis be conducted?**
- **What joint frequency models are appropriate?**

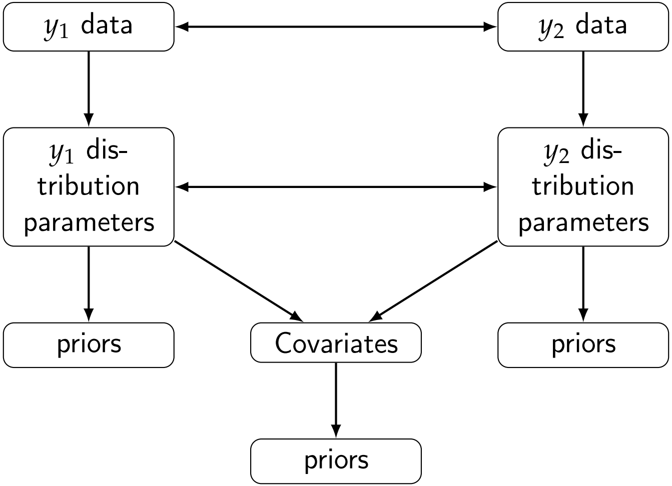
Lay of the land - Bayesian hierarchical frequency analysis

- Bayesian hierarchical modeling of precipitation and streamflow extremes
  - Active area of research in the last 10-15 years
  - Alternative to regional frequency analysis
  - End goal is to estimate distributions of return levels
- Hydrologic frequency models come in many flavors
  - Single site and spatial
  - Stationary and nonstationary
- Typically these analyses are conducted independently
- **How should a joint frequency analysis be conducted?**
- **What joint frequency models are appropriate?**
- **What is gained by a joint analysis?**

Typical model framework



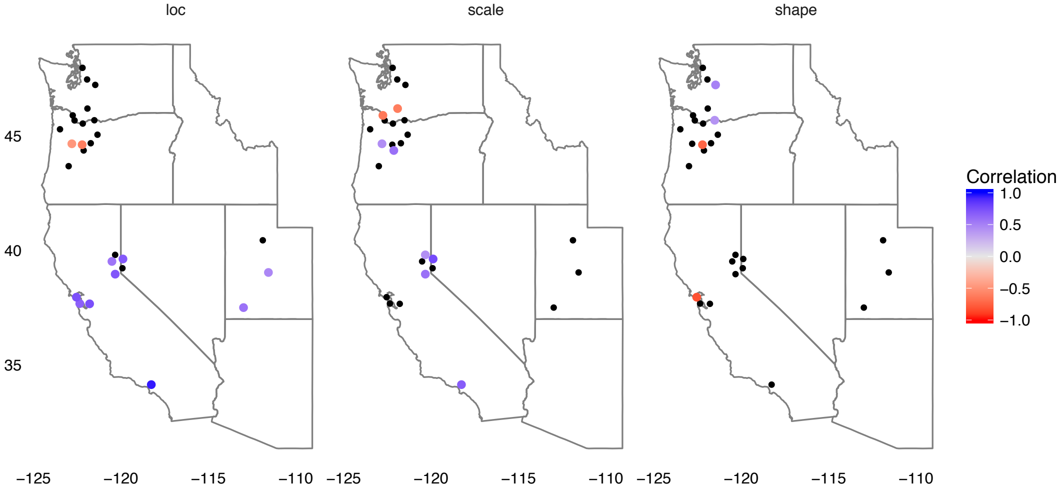
Model framework



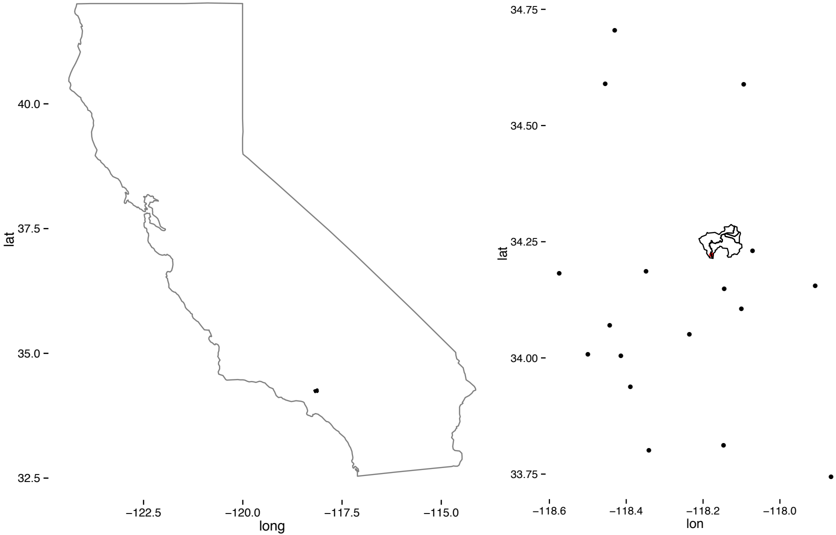
Exploratory Analysis - Are extremes linearly linked at the parameter level?

1. Fit nonstationary GEV models to flow gage and surrounding precip gages using maximum likelihood
2. Correlate the nonstationary parameter estimates
3. High correlation implies GEV parameters are related linearly

Exploratory Analysis - Are extremes linearly linked at the parameter level?



Study area


lat
40.0
37.5
35.0
32.5
−122.5
−120.0
−117.5
−115.0
long
lat
34.75
34.50
34.25
34.00
33.75
−118.6
−118.4
−118.2
−118.0
lon

Streamflow and precipitation

- 32 years of winter (DJF) 3-day flow maxima (Neuman et. al 2015):

$$z(t) \sim GEV(\mu_z(t), \sigma_z(t), \xi_z(t))$$

Streamflow and precipitation

- 32 years of winter (DJF) 3-day flow maxima (Neuman et. al 2015):

$$z(t) \sim GEV(\mu_z(t), \sigma_z(t), \xi_z(t))$$

- 32 years of winter (DJF) 3-day precipitation maxima (GHCNd):

$$y(s_i, t) \sim GEV(\mu_y(t), \sigma_y(t), \xi_y(t)),\ i = 1, ..., n$$

Streamflow and precipitation

- 32 years of winter (DJF) 3-day flow maxima (Neuman et. al 2015):

$$z(t) \sim GEV(\mu_z(t), \sigma_z(t), \xi_z(t))$$

- 32 years of winter (DJF) 3-day precipitation maxima (GHCNd):

$$y(s_i, t) \sim GEV(\mu_y(t), \sigma_y(t), \xi_y(t)),\ i = 1, ..., n$$

- Covariates:

$$x(t) = (\text{seasonal total precip, enso, pdo, amo})$$

Model Structure

$$(y(s_1,t),\ldots,y(s_n,t),z(t)) \sim C_g(\Sigma,\{\boldsymbol{\mu}(t),\boldsymbol{\sigma}(t),\boldsymbol{\xi}(t)\})$$

Regional nonstationary precip model:

$$y(s_i,t) \sim GEV(\mu_y(t),\sigma_y(t),\xi_y)$$

Model Structure

$$(y(s_1,t),\ldots,y(s_n,t),z(t)) \sim C_g(\Sigma,\{\boldsymbol{\mu}(t),\boldsymbol{\sigma}(t),\boldsymbol{\xi}(t)\})$$

Regional nonstationary precip model:

$$y(s_i,t) \sim GEV(\mu_y(t),\sigma_y(t),\xi_y)$$

$$\boxed{\mu_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\mu}$$

$$\sigma_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\sigma$$

$$\xi_y(t) = \xi_y$$

Model Structure

$$(y(s_1,t),\ldots,y(s_n,t),z(t)) \sim C_g(\Sigma,\{\boldsymbol{\mu}(t),\boldsymbol{\sigma}(t),\boldsymbol{\xi}(t)\})$$

Regional nonstationary precip model:

$$y(s_i,t) \sim GEV(\mu_y(t),\sigma_y(t),\xi_y)$$

$$\boxed{\mu_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\mu}$$

$$\sigma_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\sigma$$

$$\xi_y(t) = \xi_y$$

Nonstationary flow model:

$$z(t) \sim GEV(\mu_z(t),\sigma_z(t),\xi_z)$$

Model Structure

$$(y(s_1,t),\ldots,y(s_n,t),z(t)) \sim C_g(\Sigma,\{\boldsymbol{\mu}(t),\boldsymbol{\sigma}(t),\boldsymbol{\xi}(t)\})$$

Regional nonstationary precip model:

$$y(s_i,t) \sim GEV(\mu_y(t),\sigma_y(t),\xi_y)$$

$$\boxed{\mu_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\mu}$$

$$\sigma_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\sigma$$

$$\xi_y(t) = \xi_y$$

Nonstationary flow model:

$$z(t) \sim GEV(\mu_z(t),\sigma_z(t),\xi_z)$$

$$\boxed{\mu_z(t) = a + \mu_y(t)b}$$

$$\sigma_z(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_{\sigma_y}$$

$$\xi_z(t) = \xi_z$$

Model Structure

$$(y(s_1,t),\ldots,y(s_n,t),z(t)) \sim C_g(\Sigma,\{\boldsymbol{\mu}(t),\boldsymbol{\sigma}(t),\boldsymbol{\xi}(t)\})$$

Regional nonstationary precip model:

$$y(s_i,t) \sim GEV(\mu_y(t),\sigma_y(t),\xi_y)$$

$$\boxed{\mu_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\mu}$$

$$\sigma_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\sigma$$

$$\xi_y(t) = \xi_y$$

Nonstationary flow model:

$$z(t) \sim GEV(\mu_z(t),\sigma_z(t),\xi_z)$$

$$\boxed{\mu_z(t) = a + \mu_y(t)b}$$

$$\sigma_z(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_{\sigma_y}$$

$$\xi_z(t) = \xi_z$$

a,b,c,d are latent regression coefficients.

From nonstationary GEV parameter estimates we can compute nonstationary return levels.

Copula dependence - Hybrid spatial model

The copula dependence matrix Σ is a positive definite symetric matrix with diagonal elements equal to 1 and all other elements are between -1 and 1.

$$\Sigma = \begin{bmatrix} D & \boldsymbol{\nu} \\ \boldsymbol{\nu} & 1 \end{bmatrix} \tag{4}$$

We assume the correlation between observation locations decreases exponentially with distance, in which case

$$\nu_i = \exp(||\mathbf{p}' - \mathbf{p}_j||/\phi_1)$$

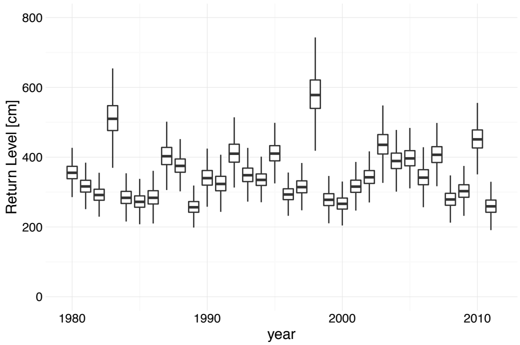
and

$$D_{ij} = \exp(||\mathbf{p}_i - \mathbf{p}_j||/\phi_2)$$

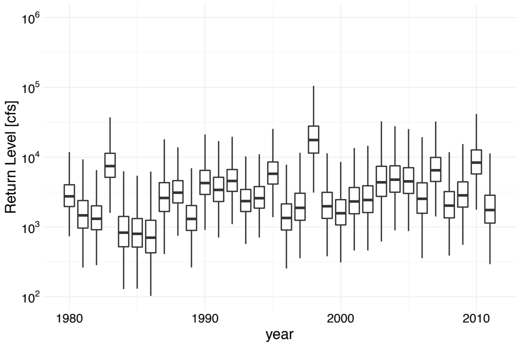
$\mathbf{p}'$: Flow location, $\mathbf{p}_i$: Precip locations, ϕ_1, ϕ_2: range parameters

Results: Nonstationary 100 year return levels

Precip:

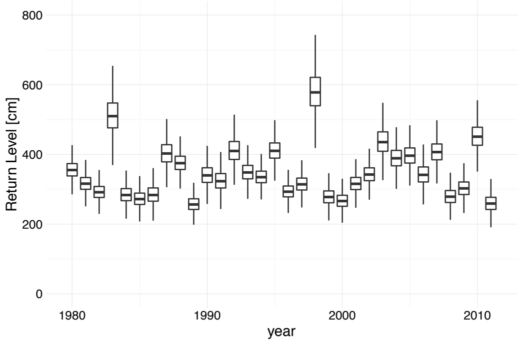


Flow:

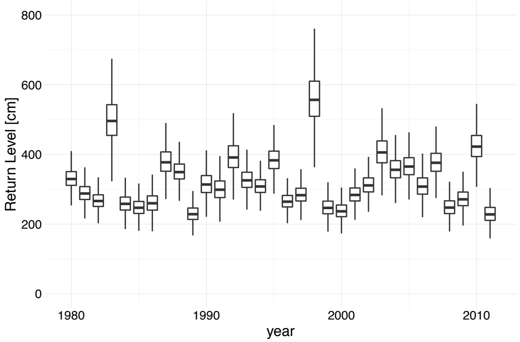


Results: Nonstationary Precipitation return levels (100 year)

Joint Precip:

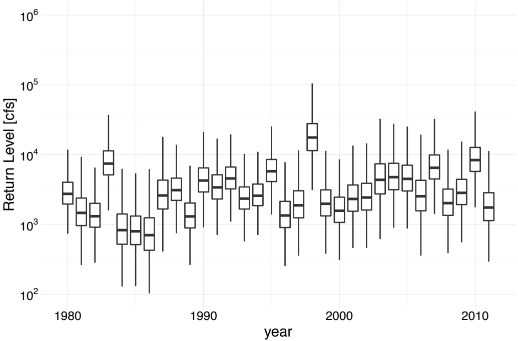


Independent Precip:

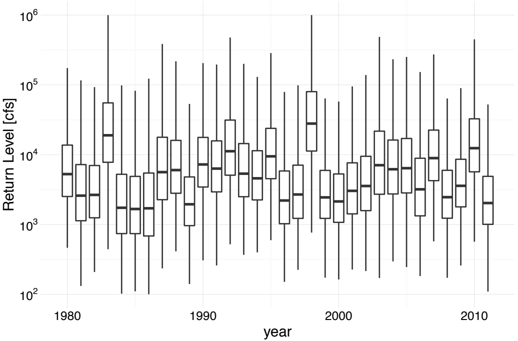


Results: Nonstationary Flow return levels (100 year)

Joint flow:

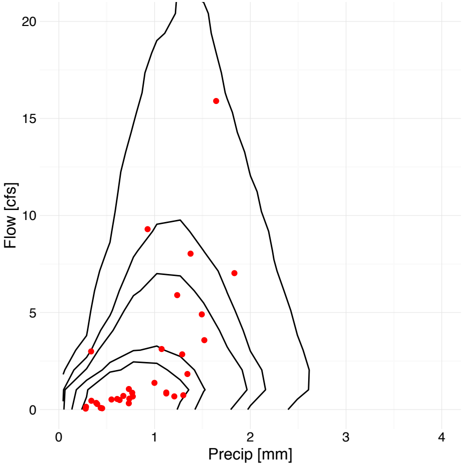


Independent flow:

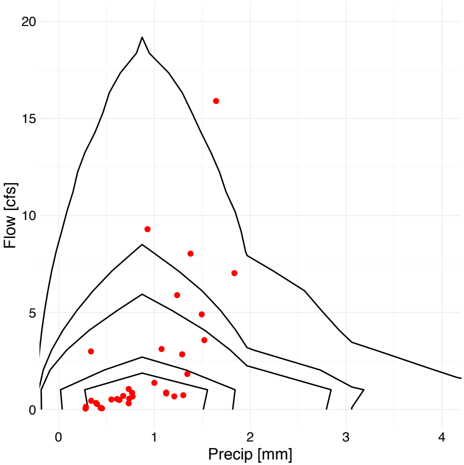


Results: Multivariate dependence

Joint:



Independent:



Nonlinear data dependence captured with linear parameter link!

Contributions

- Joint nonstationary risk estimation for streamflow and precip – applicable to rainfall-runoff dominated basins
- Incorporates climate covariates
- Simulation captures multivariate nonlinear dependence
- Reduction of uncertainty in return levels for flow
- Potential for seasonal forecasting and future projections of risk

Introduction

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

Chapter 2 - Spatial Bayesian hierarchical modeling of precipitation extremes over a large domain

Chapter 3 - Joint Bayesian hierarchical nonstationary regional frequency analysis of precipitation and streamflow extremes

Chapter 4 - A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis

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

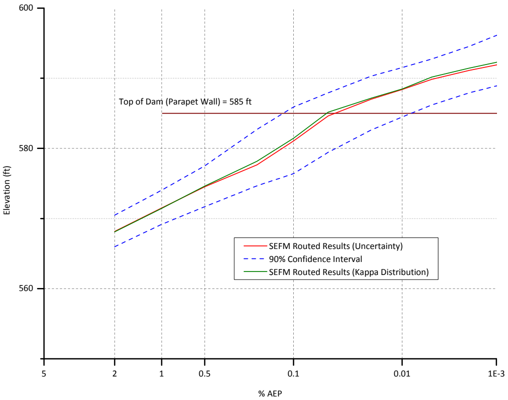
The current approach:

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



General Multivariate Model Structure

Let $y_1, \ldots, y_n$ be n block maxima variables we wish to conduct frequency analysis on.

$$
\begin{aligned}
(y_1(t), \ldots, y_n(t)) &\sim C_g(\Sigma; \{\boldsymbol{\mu}(t), \boldsymbol{\sigma}(t), \boldsymbol{\xi}(t)\}) && && (5)\\
y_i(t) &\sim GEV(\mu_i(t), \sigma_i(t), \xi_i(t)), && i = 1 \ldots n && (6)\\
\mu_i(t) &= g_{\mu i}(\mathbf{x}_i(t)^T, \boldsymbol{\mu}(t), \boldsymbol{\sigma}(t), \boldsymbol{\xi}(t)), && i = 1 \ldots n && (7)\\
\sigma_i(t) &= g_{\sigma i}(\mathbf{x}_i(t)^T, \boldsymbol{\mu}(t), \boldsymbol{\sigma}(t), \boldsymbol{\xi}(t)), && i = 1 \ldots n && (8)\\
\xi_i(t) &= g_{\xi i}(\mathbf{x}_i(t)^T, \boldsymbol{\mu}(t), \boldsymbol{\sigma}(t), \boldsymbol{\xi}(t)), && i = 1 \ldots n && (9)
\end{aligned}
$$

where C_g is a Gaussian elliptical copula joint distribution and the functions $g_{\mu i}(\cdot)$, $g_{\sigma i}(\cdot)$ and $g_{\xi i}(\cdot)$ are (potentially nonlinear) functions of covariates $\mathbf{x}_i(t)^T$ and parameters of other variables.

$$\boldsymbol{\mu}(t) = [\mu_i(t)]_{i=1}^n, \;\; \boldsymbol{\sigma}(t) = [\sigma_i(t)]_{i=1}^n, \;\; \boldsymbol{\xi}(t) = [\xi_i(t)]_{i=1}^n$$

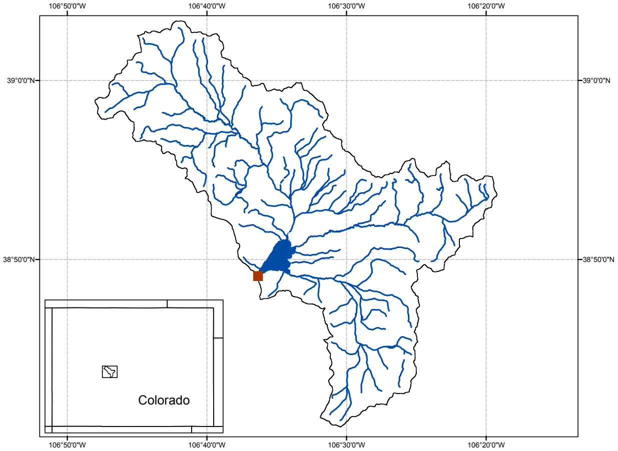
Copula Dependence

The copula dependence matrix, Σ is a symmetric positive definite matrix capturing the strength of dependence between each pairwise variable. The i, jth element of Σ measures the dependence between variables i and j and can take values between -1 and 1. By definition the dependence between a variable and itself is unity so the diagonal elements of Σ are 1's

$$\Sigma = \begin{bmatrix} 1 & \nu_{12} & \cdots & \nu_{1,n-1} & \nu_{1n} \\ \nu_{12} & 1 & \ddots & & \nu_{2n} \\ \nu_{13} & \nu_{23} & \ddots & & \vdots \\ \vdots & \vdots & \ddots & 1 & \nu_{n-1,n} \\ \nu_{1n} & \nu_{2n} & \cdots & \nu_{n-1,n} & 1 \end{bmatrix} \tag{10}$$

Note that since Σ is symmetric, there are $n(n-1)/2$ dependence parameters to fit (values in the lower or upper triangle of Σ).

Application - Taylor Park Dam



Reservoir frequency analysis

Taylor Park Reservoir, Colorado, USA.

- 35 years of annual 1-day flow maxima:

$$z(t) \sim GEV(\mu_z(t), \sigma_z, \xi_z)$$

Reservoir frequency analysis

Taylor Park Reservoir, Colorado, USA.

- 35 years of annual 1-day flow maxima:

$$z(t) \sim GEV(\mu_z(t), \sigma_z, \xi_z)$$

- 35 years of annual 1-day peak SWE (GHCNd):

$$y(t) \sim GEV(\mu_y(t), \sigma_y, \xi_y)$$

Reservoir frequency analysis

Taylor Park Reservoir, Colorado, USA.

- 35 years of annual 1-day flow maxima:

$$z(t) \sim GEV(\mu_z(t), \sigma_z, \xi_z)$$

- 35 years of annual 1-day peak SWE (GHCNd):

$$y(t) \sim GEV(\mu_y(t), \sigma_y, \xi_y)$$

Reservoir frequency analysis

Taylor Park Reservoir, Colorado, USA.

- 35 years of annual 1-day flow maxima:

$$z(t) \sim GEV(\mu_z(t), \sigma_z, \xi_z)$$

- 35 years of annual 1-day peak SWE (GHCNd):

$$y(t) \sim GEV(\mu_y(t), \sigma_y, \xi_y)$$

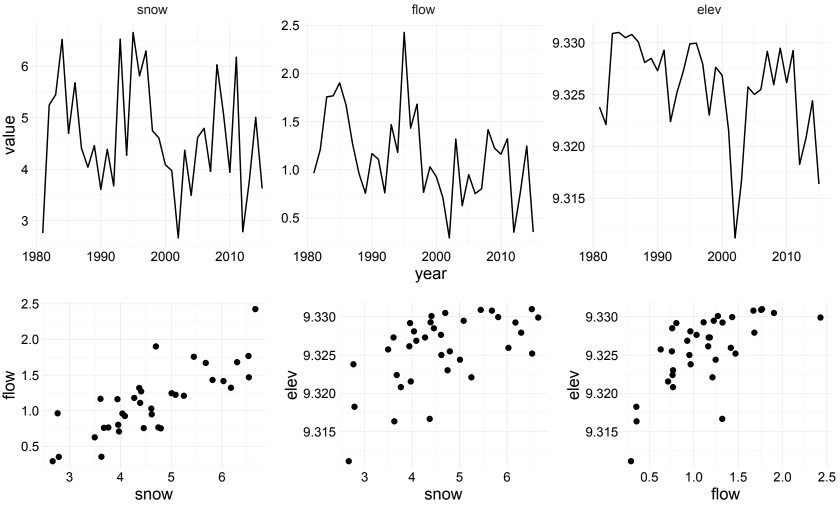
- 35 years of annual 1-day peak reservoir elevation:

$$h(t) \sim GEV(\mu_h(t), \sigma_h, \xi_h)$$

- Covariates:

$$x(t) = (\text{linear trend, enso, pdo, amo})$$

Data



Taylor Park model structure

$$(y(t), z(t), h(t)) \sim C_g(\Sigma; \{\mu_y(t), \sigma_y, \xi_y, \mu_z(t), \sigma_z, \xi_z, \mu_h(t), \sigma_h, \xi_h\}) \tag{11}$$

$$y(t) \sim GEV(\mu_y(t), \sigma_y, \xi_y) \tag{12}$$

$$z(t) \sim GEV(\mu_z(t), \sigma_z, \xi_z) \tag{13}$$

$$h(t) \sim GEV(\mu_h(t), \sigma_h, \xi_h) \tag{14}$$

$$\mu_y(t) = \mu_{y0} + x(t)^T \beta_y \tag{15}$$

$$\mu_z(t) = \mu_{z0} + x(t)^T \beta_z \tag{16}$$

$$\mu_h(t) = a - \exp(-b\mu_z(t)) \tag{17}$$

where $x(t)^T$ is a vector of climate covariates.

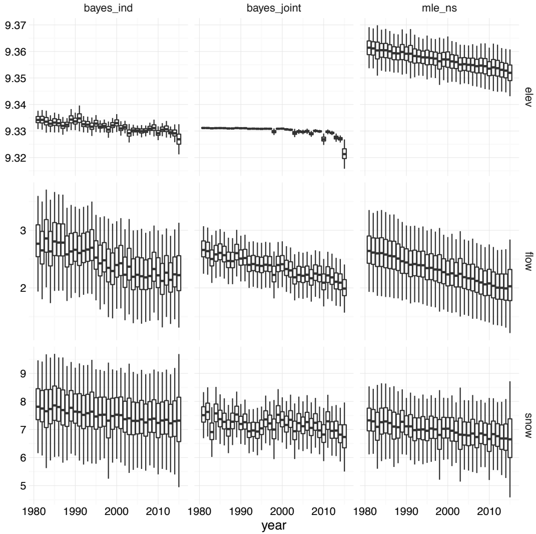
Taylor Park Copula dependence

The copula dependence matrix is

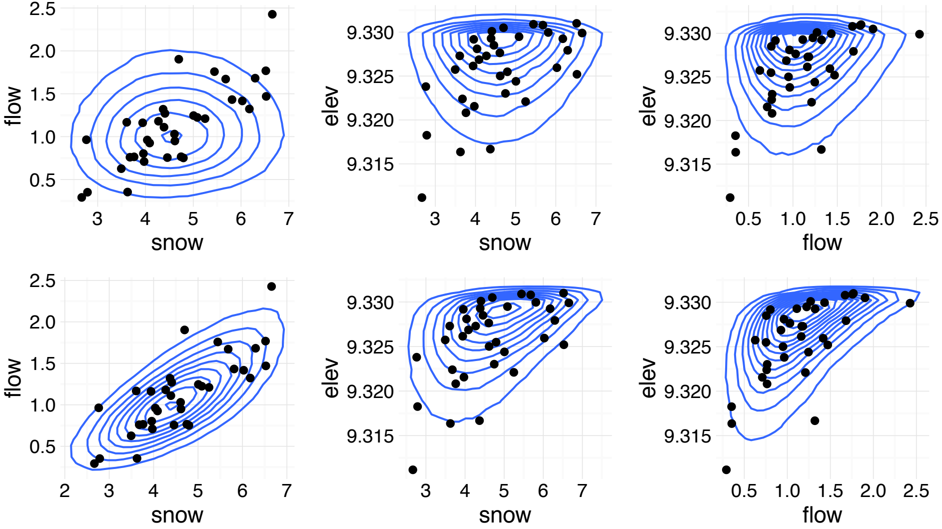
$$\Sigma = \begin{bmatrix} 1 & \nu_{yz} & \nu_{yh} \\ \nu_{yz} & 1 & \nu_{zh} \\ \nu_{yh} & \nu_{zh} & 1 \end{bmatrix} \tag{18}$$

where ν_{ij} represents the dependence (correlation) between variable i and j.

Results - Nonstationary return levels



Results - Joint distributions



Contributions

- General framework for multivariate nonstationary frequency analysis – application to dam safety
- Nonlinear hierarchical covariates
- Multivariate simulation capturing nonlinear joint dependence
- Reduction of uncertainty in return levels compared to independent models

Overview of research contributions

Chapter 1:

- Seasonal approach reveals important differences in the genesis of extreme precip between seasons
- Improved extreme clustering method for large geographic regions
- Seasonally dependent extreme regions, moisture sources and pathways
- ENSO effect on frequency of extreme precip – agrees with previous results

Chapter 2:

- Bayesian spatial extremes model for large regions with thousands of stations
- Composite likelihood
- Spatially varying regression coefficients
- Gridded return levels for any return period at any resolution

Overview of research contributions

Chapter 3:

- Joint nonstationary risk estimation for streamflow and precip – applicable to rainfall-runoff dominated basins
- Incorporates climate covariates
- Simulation captures multivariate nonlinear dependence
- Reduction of uncertainty in return levels for flow
- Potential for seasonal forecasting and future projections of risk

Chapter 4:

- General framework for multivariate nonstationary frequency analysis – application to dam safety
- Nonlinear hierarchical covariates and climate covariates
- Multivariate simulation capturing nonlinear joint dependence
- Reduction of uncertainty in return levels compared to independent models

Acknowledgments

Advisor: Balaji Rajagopalan
Reclamation Project PI: Subhrendu Gangopadhyay
Committee: Will Kleiber, Ben Livneh, Joseph Kaspryzk
Co-authors: Linyin Cheng, Kathleen Holman, Connie Woodhouse, Will Kleiber, Mike Alexander

Julie
All my friends and family

Thanks!!!

References I

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*, 47(W11513):1–21, 2011.