Understanding and modeling hydroclimate extremes in the western US

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

Portland State University
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].

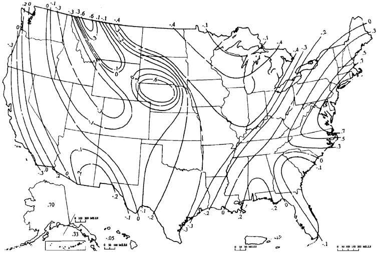
> *"In order to apply any theory we have to suppose that the data are homogeneous, i.e. that no systematical change of climate and no important change in the basin have occurred within the observation period and that no such changes will take place in the period for which extrapolations are made." [Gumbel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

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

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

Part 3 - Coupled Bayesian hierarchical modeling of streamflow and precipitation extremes

Part 4 - Proposed future work

Extras

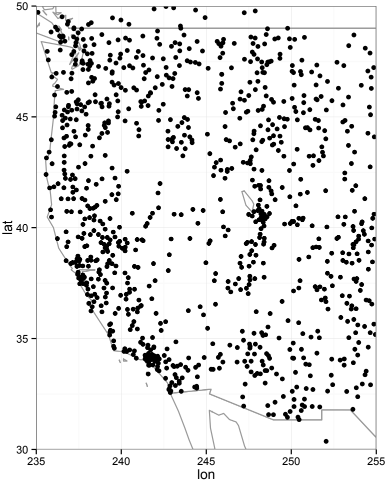
Research questions

- **How does 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

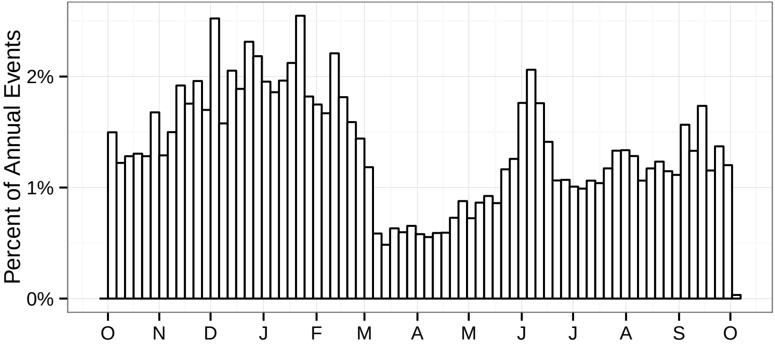
Global Historical Climatology Network (GHCN), daily data from over 90,000 stations worldwide, about 14,000 in the western US.

- Limit to ∼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?

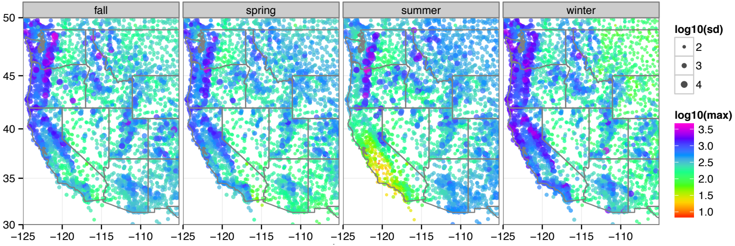
Annual 3day max occurrence day.



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

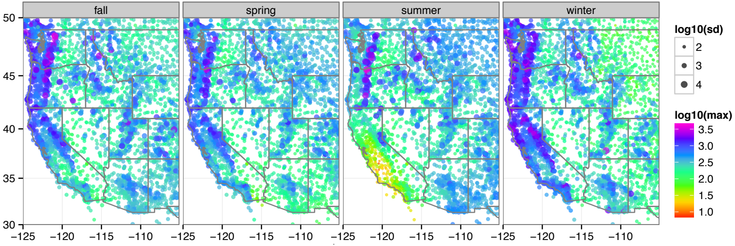
Extreme Event Timing and Magnitude

GHCNd mean 3day precip.

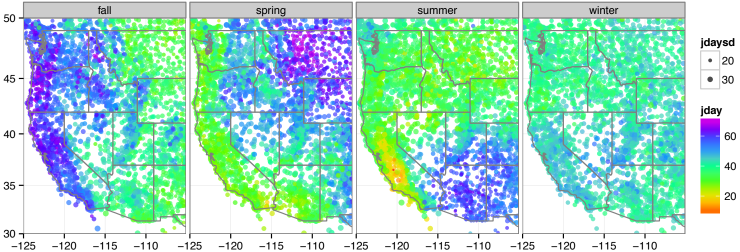


Extreme Event Timing and Magnitude

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)

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

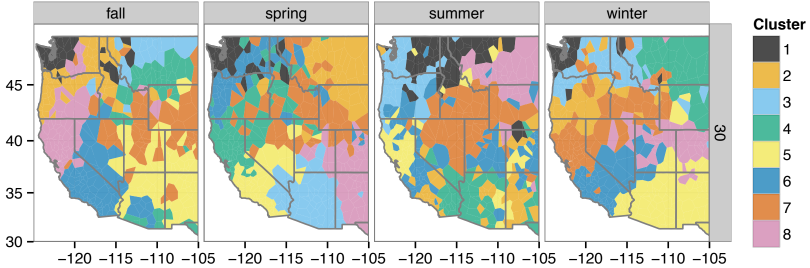
where

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

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.

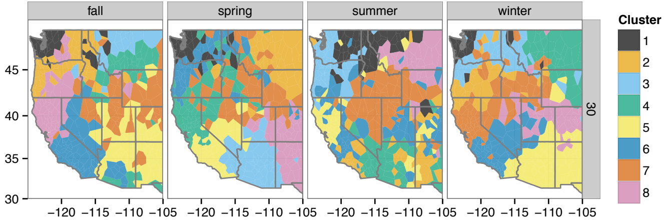
Clustering of Maxima

Method was tested on small geographic regions.



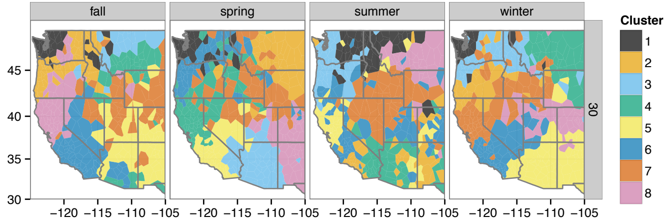
Clustering of Maxima

Original Method (tested on small regions):

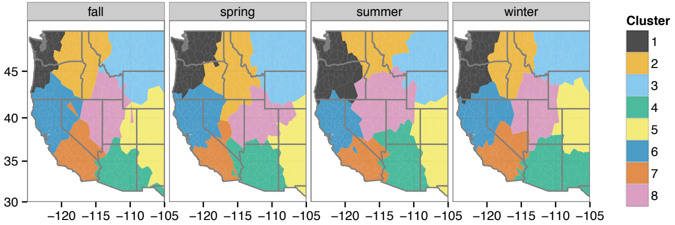


Clustering of Maxima

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

Moisture Source Identification

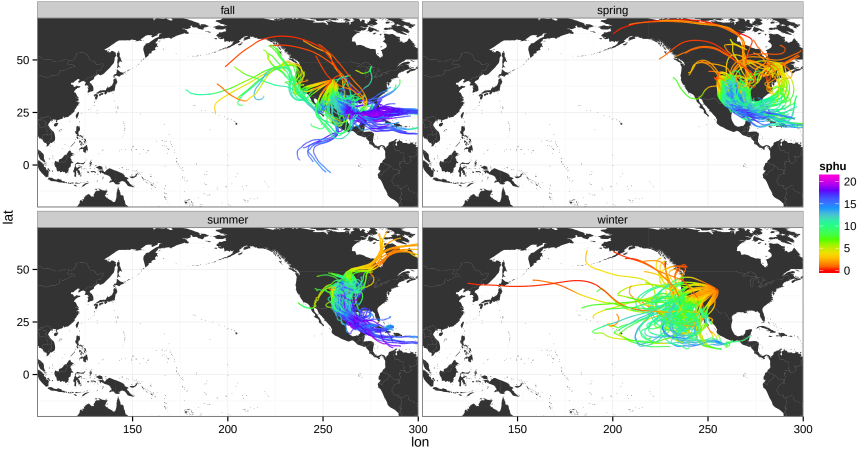
HYSPLIT - Hybrid Single Particle Lagrangian Integrated Trajectory Model

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

Runs:

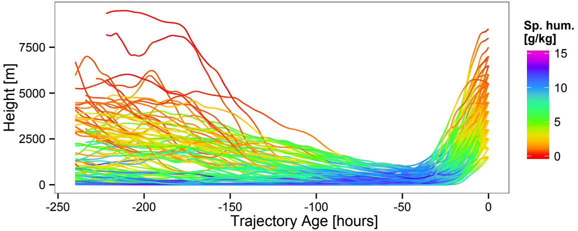
- A single trajectory has a lot of uncertainty, necessitating a statistical approach
- Number of trajectories = 12 time trajectories/event * 21 height trajectories/event * 65 years * 1000 stations * 4 seasons = 65,520,000 trajectories = 15,724,800,000 points
- For each station and season find the top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

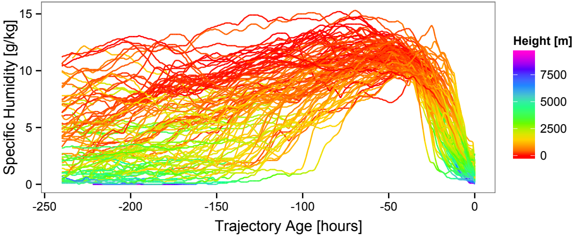
Example rain trajectories – Colorado



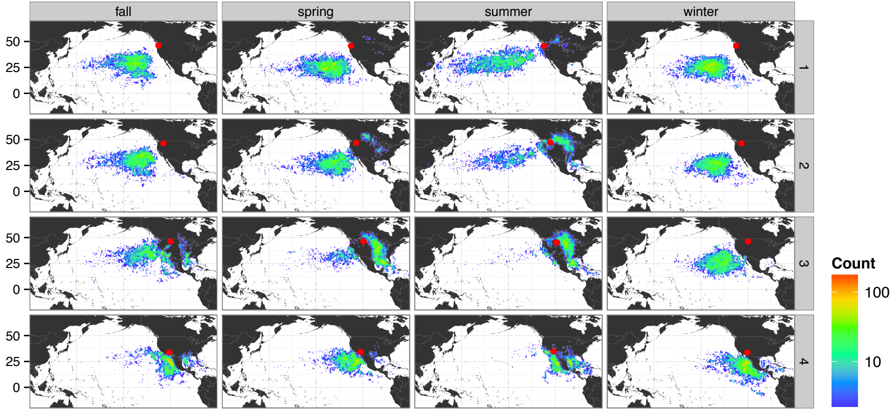
Moisture Source identification

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

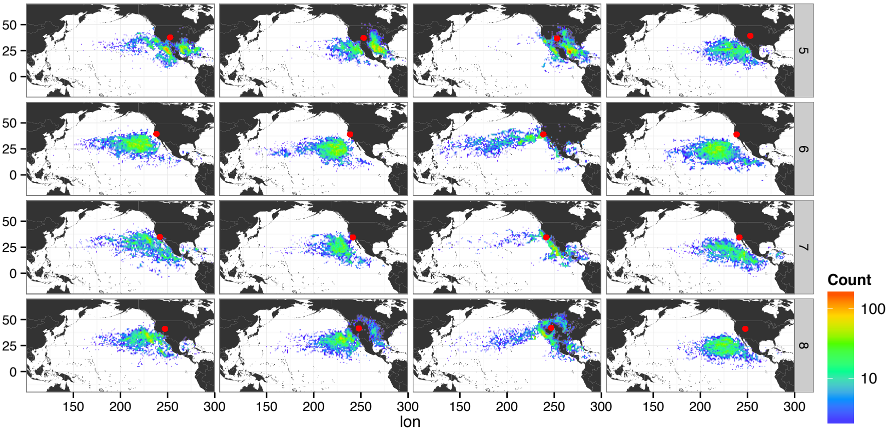




Moisture Source identification



Moisture Source identification

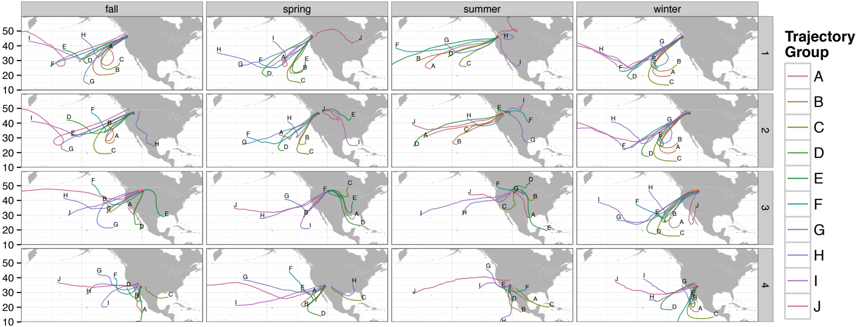


Moisture Pathways

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

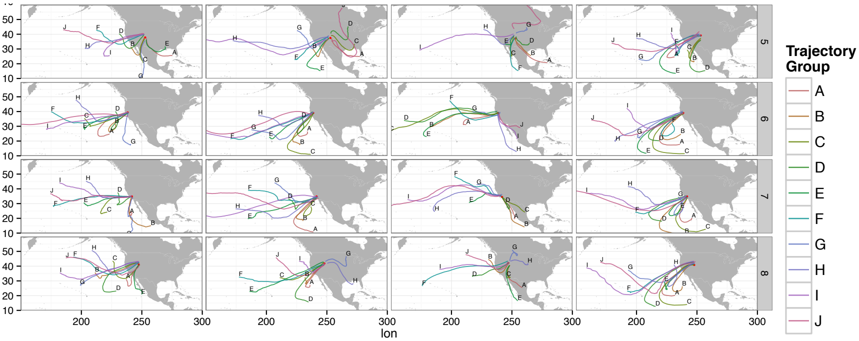
Moisture Pathways

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



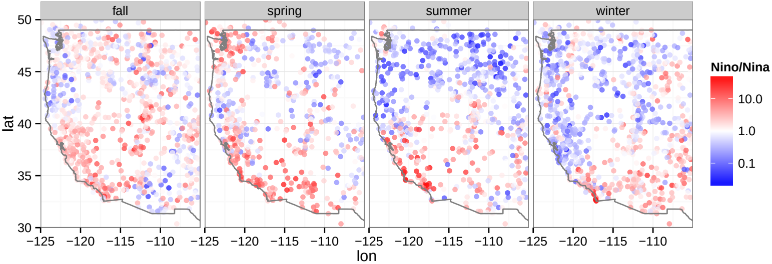
Moisture Pathways

- 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.

Research Goals

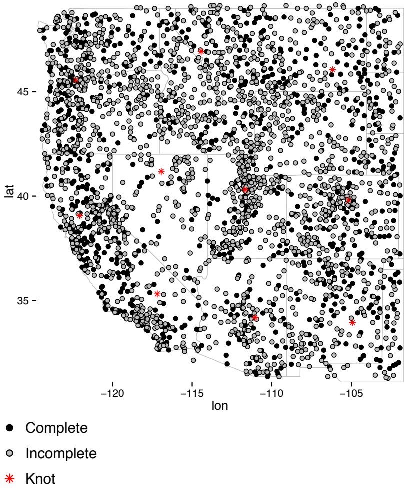
- How do extreme precipitation return levels vary spatially?
- What covariates affect precipitation return levels?
- How can gridded return levels be produced?
- What is the associated uncertainty in these return level estimates?
- How can we model return levels in the entire western US simultaneously?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

Precipitation Data

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

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

Very large region/dataset for typical Bayesian spatial model



Hierarchical spatial model

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

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

Hierarchical spatial model

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

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

Hierarchical spatial model

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

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

Hierarchical spatial model

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

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

Hierarchical spatial model

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

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

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

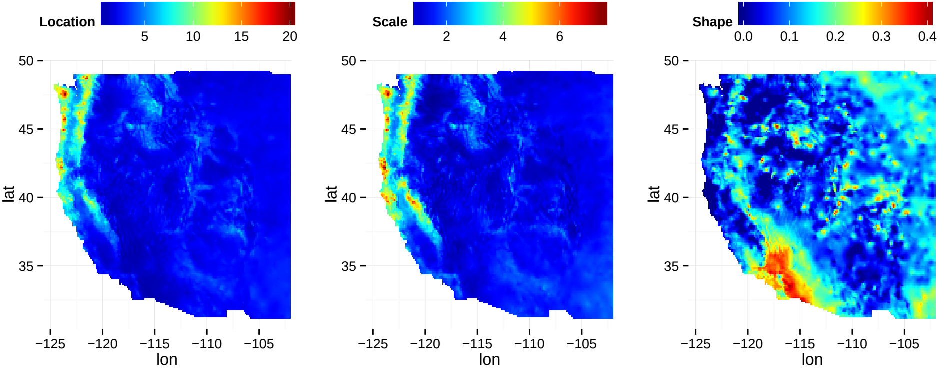
Composite Likelihood (CL)

- 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

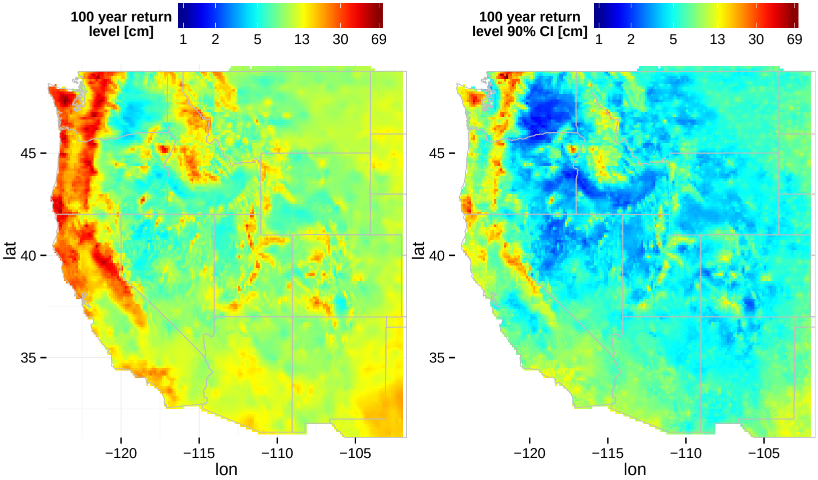
$$L_{CL} = \prod_{g=1}^{G} \mathcal{N}(\mathbf{0}, \Sigma_g(\boldsymbol{\theta}))$$

- 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 return levels, fall



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

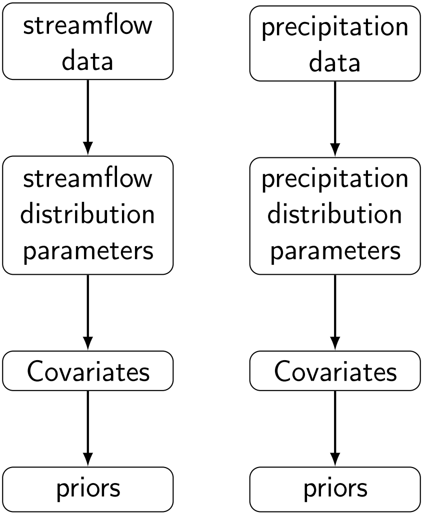
Motivation

- 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
- 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

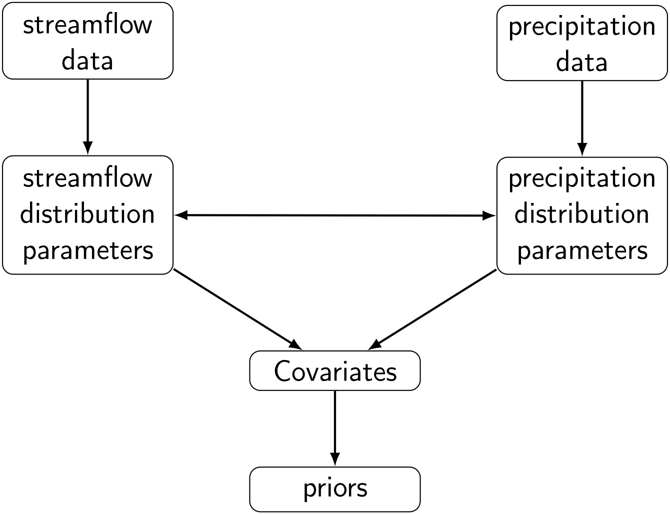
Research Questions

- **How can precip and streamflow frequency analysis be linked?**
- **Where geographically is this linkage appropriate?**
- **What is gained by linking?**

Typical hierarchical model framework



Proposed hierarchical model framework



Data

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

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

Data

- 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):

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

Data

- 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):

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

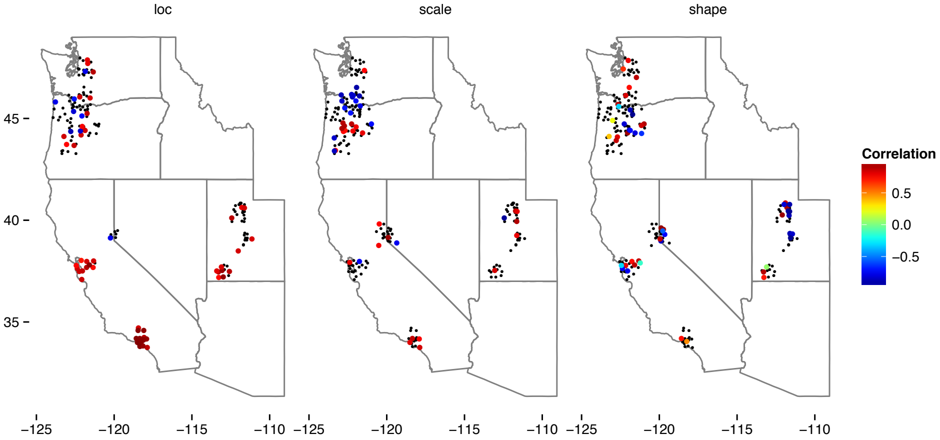
- Covariates:

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

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

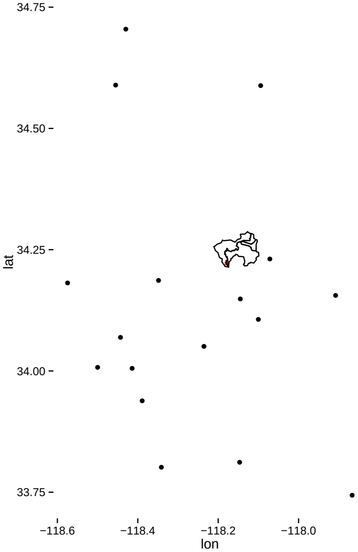
1. Fit nonstationary GEV models to flow gage and surrounding precip gages using maximum likelihood using the same covariates
2. Correlate the nonstationary parameter estimates
3. High correlation implies GEV parameters are related linearly

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



Study area


40.0
37.5
35.0
32.5
lat
−122.5
−120.0
−117.5
−115.0
long


34.75
34.50
34.25
34.00
33.75
lat
−118.6
−118.4
−118.2
−118.0
lon

Model Structure

Nonstationary precip model:

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

Model Structure

Nonstationary precip model:

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

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

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

$$\xi_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\xi + \varepsilon_\xi$$

Model Structure

Nonstationary precip model:

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

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

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

$$\xi_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\xi + \varepsilon_\xi$$

Nonstationary flow model:

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

Model Structure

Nonstationary precip model:

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

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

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

$$\xi_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\xi + \varepsilon_\xi$$

Nonstationary flow model:

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

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

$$\boxed{\sigma_z(t) = c + \sigma_y(s)d + \nu_\sigma}$$

$$\xi_z(t) = \mathrm{x}^T(t)\boldsymbol{\beta}'_\xi + \nu_\xi$$

Model Structure

Nonstationary precip model:

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

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

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

$$\xi_y(t) = \mathrm{x}^T(t)\boldsymbol{\beta}_\xi + \varepsilon_\xi$$

Nonstationary flow model:

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

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

$$\boxed{\sigma_z(t) = c + \sigma_y(s)d + \nu_\sigma}$$

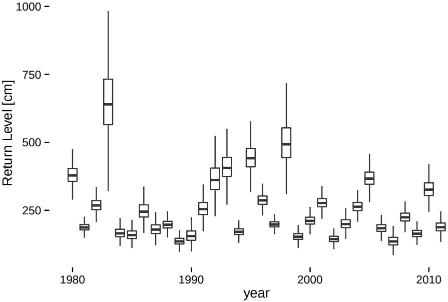
$$\xi_z(t) = \mathrm{x}^T(t)\boldsymbol{\beta}'_\xi + \nu_\xi$$

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

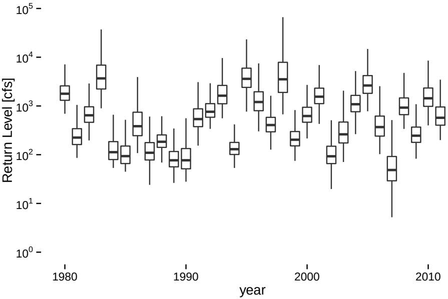
From nonstationary GEV parameter estimates we can compute nonstationary return levels.

Results: 100 year return levels

Precip:

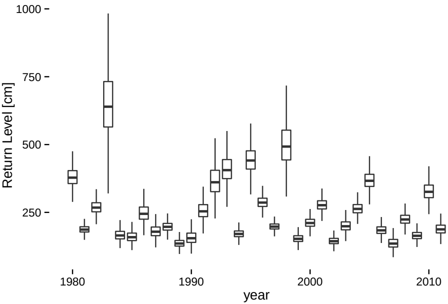


Flow:

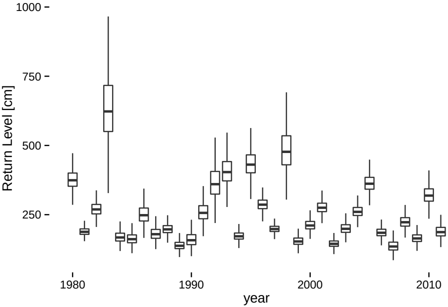


Results: Precipitation return levels (100 year)

Coupled Precip:

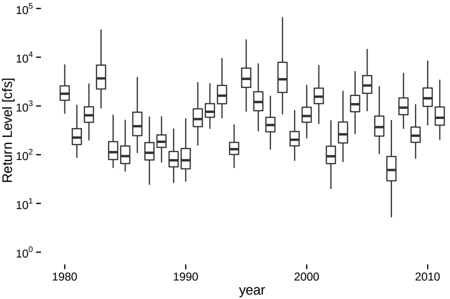


Uncoupled precip:

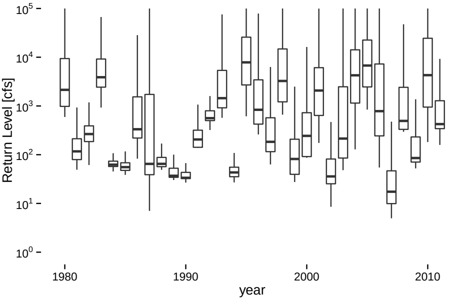


Results: Flow return levels (100 year)

Coupled flow:



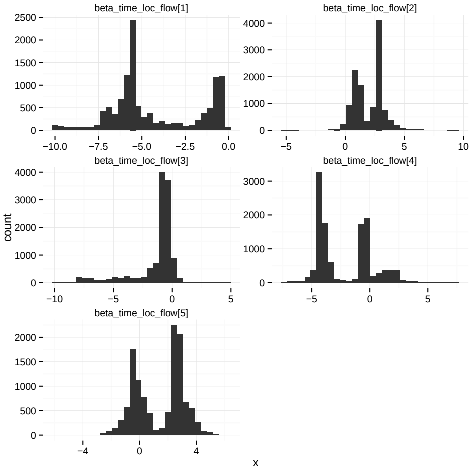
Uncoupled flow:



Why is this useful?

Lurking problems:

- Convergence issues and lack of identifiability —>
- Coupled: 24 parameters, 544 data points
- Uncoupled Flow: 30 parameters, 32 datapoints



Conclusions

Pros:

- Flow return levels strongly affected by coupling but precip is not
- Coupling helps convergence for flow

Cons:

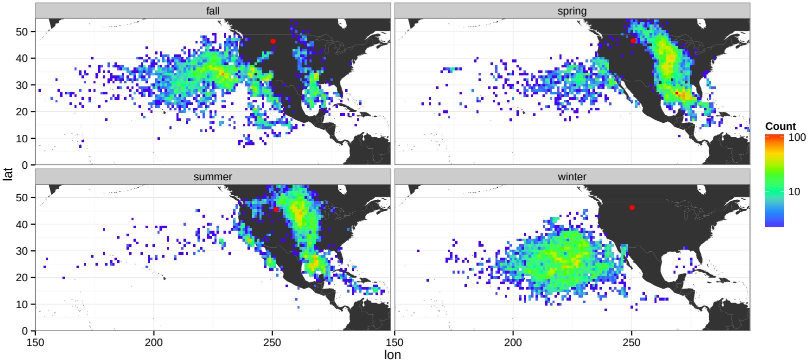
- Data availability
- Appears to only be appropriate for rainfall runoff dominated basins (I am working on this)

Nonstationarity hydroclimate extremes in the United States

- A **nonstationary** physical process is one whose statistical properties such as mean, variance and autocorrelation change in time.
- Nonstationarity in hydroclimate extremes (precipitation, temperature and streamflow) arises due to:
  - Natural variability
    - Seasonal
    - Decadal to multidecadal
  - Anthropogenic influences
    - Land use changes (agriculture, urbanization)
    - Climate change

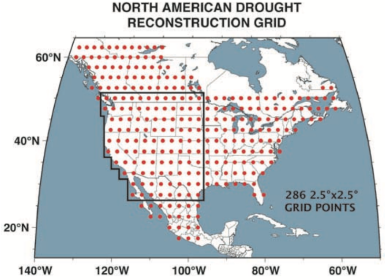
My proposed research addresses a number of aspects of **nonstationary hydroclimate extremes** in the United States across a **variety of timescales** through a lens of **appropriate quantification of uncertainty**.

Nonstationarity in moisture sources and pathways of precipitation extremes



- How have moisture sources and pathways for extremes changed over the last 100 years?
- Are there identifiable anthropogenic impacts?

Uncertainty quantification for paleo estimates of PDSI



- Tree-ring chronologies + Regression = Paleo drought dataset [Cook et al., 2004]
- Tree-ring chronologies + PMIP simulations + spatial model = New Paleo PDSI dataset
  - Drought maps at finer resolution with quantified uncertainty
  - Capture spatial and temporal correlation structure
  - Input to planning models

Nonstationary signals in hydroclimate extremes

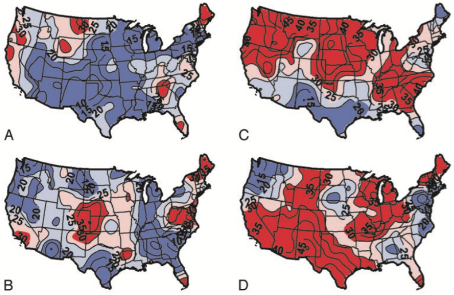
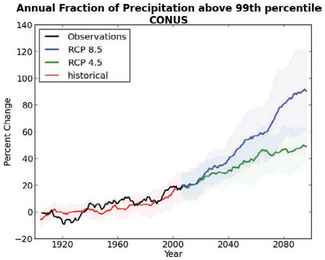


Fig. 5. Drought frequency (in percent of years) for positive and negative regimes of the PDO and AMO. (*A*) Positive PDO, negative AMO. (*B*) Negative PDO, negative AMO. (*C*) Positive PDO, positive AMO. (*D*) Negative PDO, positive AMO.

- How do climate indices combine to produce extremes?
- Dynamic models allow climate covariates to have varying influence over space and time
- We can answer specific questions about the sensitivity of extremes in certain regions to changing atmospheric oscillations

Blending climate models and observations for nonstationary extremes analysis



- Climate model provide information about future extremes but with limited uncertainty quantification
- Climate models can be blended with observations using Bayesian hierarchical models to make future projections of extremes that correctly quantify uncertainty

The end. Thanks for listening!

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 1/8th$^\circ$ 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:

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

Parameters/Likelihood/Priors

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

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

Statistics of Extremes

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

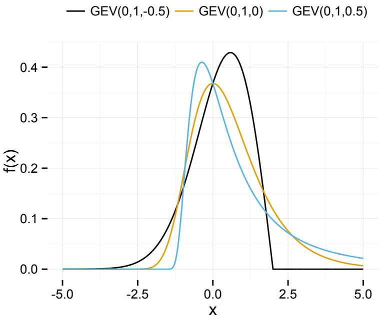
$$\mathrm{GEV}(x;\mu,\sigma,\xi) = \frac{1}{\sigma} b^{(-1/\xi)-1} \exp\left\{-b^{-1/\xi}\right\}$$

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

Return Level (quantile function):

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

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:

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

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).

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.