Doctoral Dissertation Proposal

Understanding and modeling hydroclimate extremes in the western United States

by

Cameron W. Bracken

B.S., Humboldt State University, 2009
M.S., University of Colorado at Boulder, 2012

Committee members:
Prof. Balaji Rajagopalan
Dr. Subhrendu Gangopadhyay
Prof. Will Kleiber
Prof. Joseph Kasprzyk
Prof. Ben Livneh

Department of Civil, Environmental and Architectural Engineering
University of Colorado at Boulder
July 2015

Contents

1 Overview

Extreme hydroclimate events are by definition rare but have a disproportionate effect on infrastructure, the environment, the economy and human life. The incorporation of extreme value analysis has long been standard practice in civil engineering design and management dating back to [*Gumbel*, 1941]. The typical goal of an extreme value analysis is the estimation of a return level, or the magnitude of an extreme event associated with a given return period. While the motivation to provide better estimates of return levels is self-evident, engineering practice remains somewhat rudimentary despite substantial research improvements in recent years [*Katz et al.*, 2002]. The assumptions of traditional extreme value analysis are adequately summarized by Gumbel:

> "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]

Analysis of hydroclimate extremes began as a single site procedure and surprisingly remains the recommended procedure for determining flood frequency estimates for a basin [*USGS*, 1981]. This approach has some obvious drawbacks; namely, it ignores any spatial coherence that may be present across multiple sites and it cannot be applied to sites with no observations. Regional frequency analysis (RFA) is a popular methodology that addresses both issues [*Hosking et al.*, 1985]. RFA aims to improve frequency estimates at all sites in a homogeneous region and enables the transfer of frequency estimates to ungaged sites. The probability weighted moments (PWM) or "L-moments" approach is a common component of RFA because of its speed of computation and good performance with small samples [*Katz et al.*, 2002]. While RFA provides obvious advantages over at-site analyses, it still requires the (often subjective) delineation of homogeneous regions, and traditional implementations do not have the ability to incorporate climate information.

In the western US, there is clear evidence for strong spatial and temporal variability in hydroclimate extremes and links to large scale climate [*Arriaga Ramírez and Cavazos*, 2010; *Balling Jr and Goodrich*, 2011; *Cayan et al.*, 2010, 1999; *Dominguez et al.*, 2012; *Dulière et al.*, 2013b; *Gershunov and Barnett*, 1998; *Higgins et al.*, 2010; *Kunkel*, 2003b, a; *Palecki et al.*, 2005; *Pryor et al.*, 2009; *Schumacher and Johnson*, 2010]. Despite an increasing understanding of how extremes behave historically in space and time, tools to incorporate this information are not widespread. In particular, there is a need for tools which provide seasonal and nonstationary estimates of risk throughout the western US at a scale that is useful for decision making. It is also crucial that these tools provide robust quantifications of uncertainty in return level estimates.

Return levels are traditionally viewed as fixed quantities, though in reality return level estimates have their own probability distributions. Robust quantification of uncertainty is a fundamental component of any frequency analysis. Bayesian model fitting methods provide an avenue to represent the full uncertainty associated with return levels given the data. In recent years Bayesian approaches to modeling extremes have seen an explosion in the literature [*Cooley et al.*, 2007; *Cooley and Sain*, 2010; *Aryal et al.*, 2010; *Atyeo and Walshaw*, 2012; *Davison et al.*, 2012; *Ghosh and Mallick*, 2011; *Reich and Shaby*, 2012; *Sang and Gelfand*, 2010, 2009; *Apputhurai and Stephenson*, 2013; *Dyrrdal et al.*, 2014]. Chapters 3-5 will utilize Bayesian model fitting procedures in order to quantify the uncertainty in return level estimates.

This proposal will discuss methods to analyze and model seasonal, spatial and non-stationary hydroclimate extremes to address a number of simplifying assumptions made in traditional extreme value analysis:

1. **Annual**: Typically only annual maxima are modeled, or maxima in the single wettest season, which masks seasonal variability in extremes.
2. **Single site**: Extreme events do not occur as point processes. Considering spatial dependence among data can reduce overall uncertainty in return levels and allows for interpolation of return levels.
3. **Stationary**: Extreme event distributions are often assumed to be stationary in time, a tenuous assumption under a changing climate.
4. **Uncertainty**: While not strictly an assumption, the uncertainty in return level estimates is rarely quantified in practice.

Chapter 2 describes a comprehensive analysis of the seasonal spatial behavior of extreme precipitation in the western US. Chapter 3 describes a Bayesian hierarchical model for spatial extremes which is applied to model seasonal extreme precipitation. Chapter 4 will discuss a Bayesian hierarchical model for coupled spacetime extreme precipitation and streamflow analysis. Chapter 5 will focus on precipitation and flow frequency analysis for dam safety, highlighting the differences between traditional approaches and those developed in previous chapters.

2 Spatial variability of seasonal extreme precipitation in the western United States

This research has been published in the Journal of Geophysical Research: Atmospheres with the following citation:

Bracken, C., B. Rajagopalan, M. Alexander, and S. Gangopadhyay (2015), Spatial variability of seasonal extreme precipitation in the western United States, J. Geophys. Res. Atmos., 120, 45224533. doi:10.1002/2015JD023205.

2.1 Introduction

The nature of extreme precipitation varies widely across the Western United States. For example, the Southwest is most likely to receive extreme precipitation in the summer, while the Pacific Northwest and California typically see the largest extreme events in the winter, and the intermountain west tends to see a more even distribution of occurrence across seasons [*Kunkel et al.*, 1999]. The mean occurrence day of annual maxima for all stations in the Western US clearly shows a peak in the winter months (DJF), but in more than 50% of stations, annual maxima tend to occur between February and July (Figure 1). This motivates a careful examination of how extremes behave both seasonally and spatially. Past studies have classified the behavior of extreme events regionally throughout the Western US [*Groisman et al.*, 2001; *Arriaga Ramírez and Cavazos*, 2010; *Dos Santos et al.*, 2011; *Dulière et al.*, 2013a; *Mullens et al.*, 2013; *Janssen et al.*, 2014] and for specific seasons [*Warner et al.*, 2012; *Pal et al.*, 2013; *Jiang et al.*, 2014], but few have taken a detailed look at both of these pieces. A seasonal analysis is especially important because while the most likely season for an extreme event may be identified for a specific region, extreme events can occur at any time of the year and the characteristics of these events vary both seasonally and spatially [*Kunkel et al.*, 1999].

Past studies have grouped stations by climate divisions [*Kunkel*, 2003b], though these regions are defined for climatological mean and are not appropriate for extreme values [*Jones et al.*, 2014]. Other studies have used grid based groupings [*Kunkel et al.*, 2003], subjective [*Maraun et al.*, 2008; *Alexander et al.*, 2006] and objective [*Jones et al.*, 2014; *Wigley et al.*, 1984; *Dales and Reed*, 1989; *Neal and Phillips*, 2009] regions. *Jones et al.* [2014] defined regions in the UK using principal component analysis and clustering of a number of extreme value statistics. *DeGaetano* [1998] used a method for defining clusters based on similarity of extreme value cumulative distributions function and spatial proximity. *Bernard et al.* [2013] used a clustering method specifically tailored to the characteristics of extreme value distributions.

Given homogeneous clusters of extreme precipitation, dominant moisture pathways and sources of extreme precipitation can be examined regionally. Moisture pathway and delivery questions can be answered with back trajectory analysis. Back trajectory analysis calculates the pathway that an air parcel followed such that it arrives at the location and time of an observed extreme event. Trajectory analysis has been used previously to identify

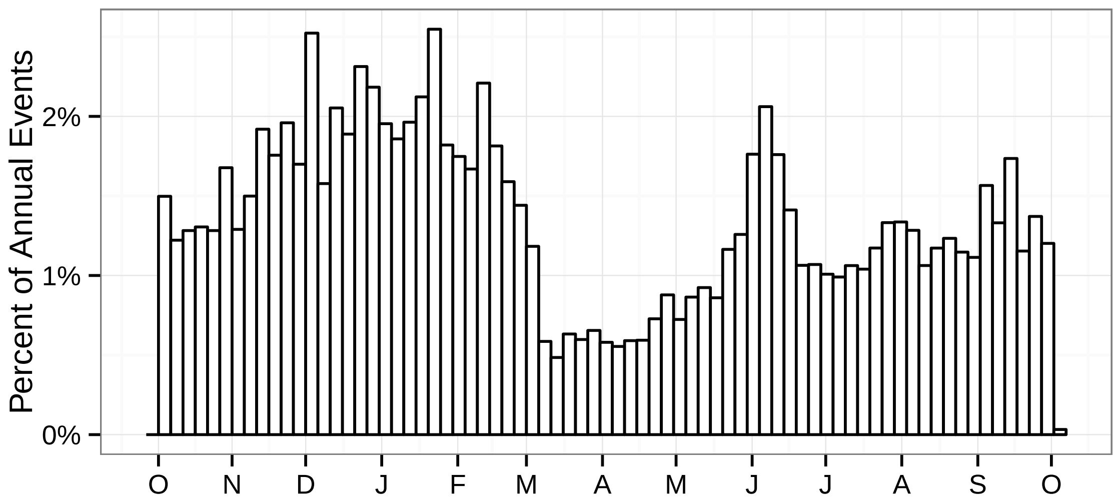

Figure 1: Occurrence day of annual 3-day precipitation maxima at ∼14,000 stations in the western US (5 day bins).

moisture sources and pathways for extreme precipitation [*Gustafsson et al.*, 2010; *Massacand et al.*, 1998; *Reale et al.*, 2001; *Alexander et al.*, 2015]. In this study we investigate the statistically likely moisture sources and pathways for extreme events and how these vary between seasons and between regions [*DeGaetano*, 1998; *Jorba et al.*, 2004]. This is of importance in modeling, simulating, and predicting extremes in space and time and, consequently, for resource management.

In addition to a seasonal analysis, we also seek to understand the impact of the El Niño Southern Oscillation (ENSO) on dominant moisture pathways and sources in the Western US. The link between ENSO and extreme events has been previously explored [*Gershunov and Barnett*, 1998; *Cayan et al.*, 1999; *Higgins et al.*, 2010; *Feldl and Roe*, 2011; *DeFlorio et al.*, 2013]. In El Niño conditions, days with high daily precipitation are seen to be more frequent than average over the Southwest and less frequent over the Northwest. During La Niña conditions the signal is typically reversed, but the ENSO events are not all the same and the spatial patterns are complex.

We propose the following research questions related to extreme events in the western US:

- Can we objectively define coherent extreme value regions and how many regions are appropriate?
- What are the dominant moisture sources and pathways for each season for these regions?
- How do seasonal moisture sources pathways change under ENSO regimes?

2.2 Selected results

One research contribution of this chapter was the development of an automated method for determining homogeneous extreme regions based on station observation. The method built on a previous method making an improvement that drastically improved the coherence of the regions while still maintaining the theoretical benefits of the method *Bernard et al.* [2013]. Figure 2 shows the results of the improved method compared to the results of *Bernard et al.* [2013].

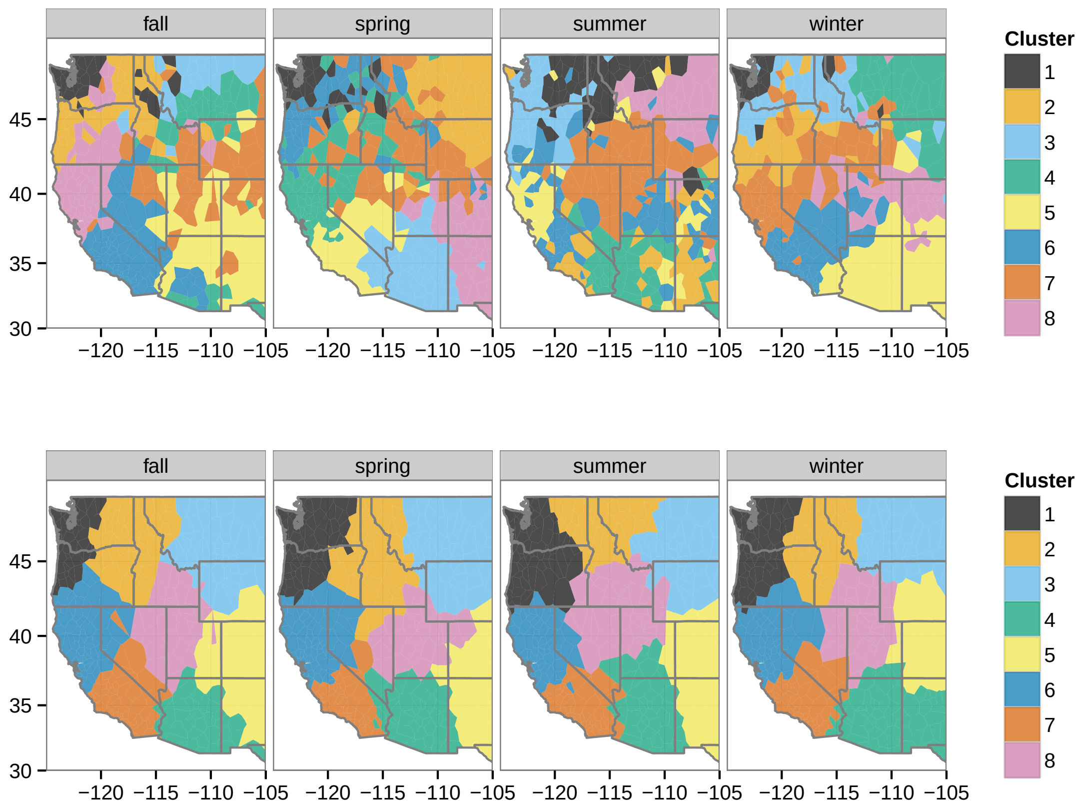

Figure 2: (top) Extreme regions defined by the clustering method of *Bernard et al.* [2013] and (bottom) regions defined with the modified extremes clustering method for large geographic regions.

Moisture sources for extremes occurring in the previously defined regions were identified using the HYSPLIT trajectory model [*Draxler et al.*, 1999; *Draxler and Hess*, 1998, 1997]. For each event, the model was run for 8 days back in time. The maximum specific humidity along this trajectory was identified as the moisture source for this trajectory. By collecting the moisture sources for all trajectories that produced rain at their teminus, and binning these points on a 1 degree grid, 3 was created. This figure shows sharp regional and seasonal differences in moisture source locations for extreme rainfall in the western US.

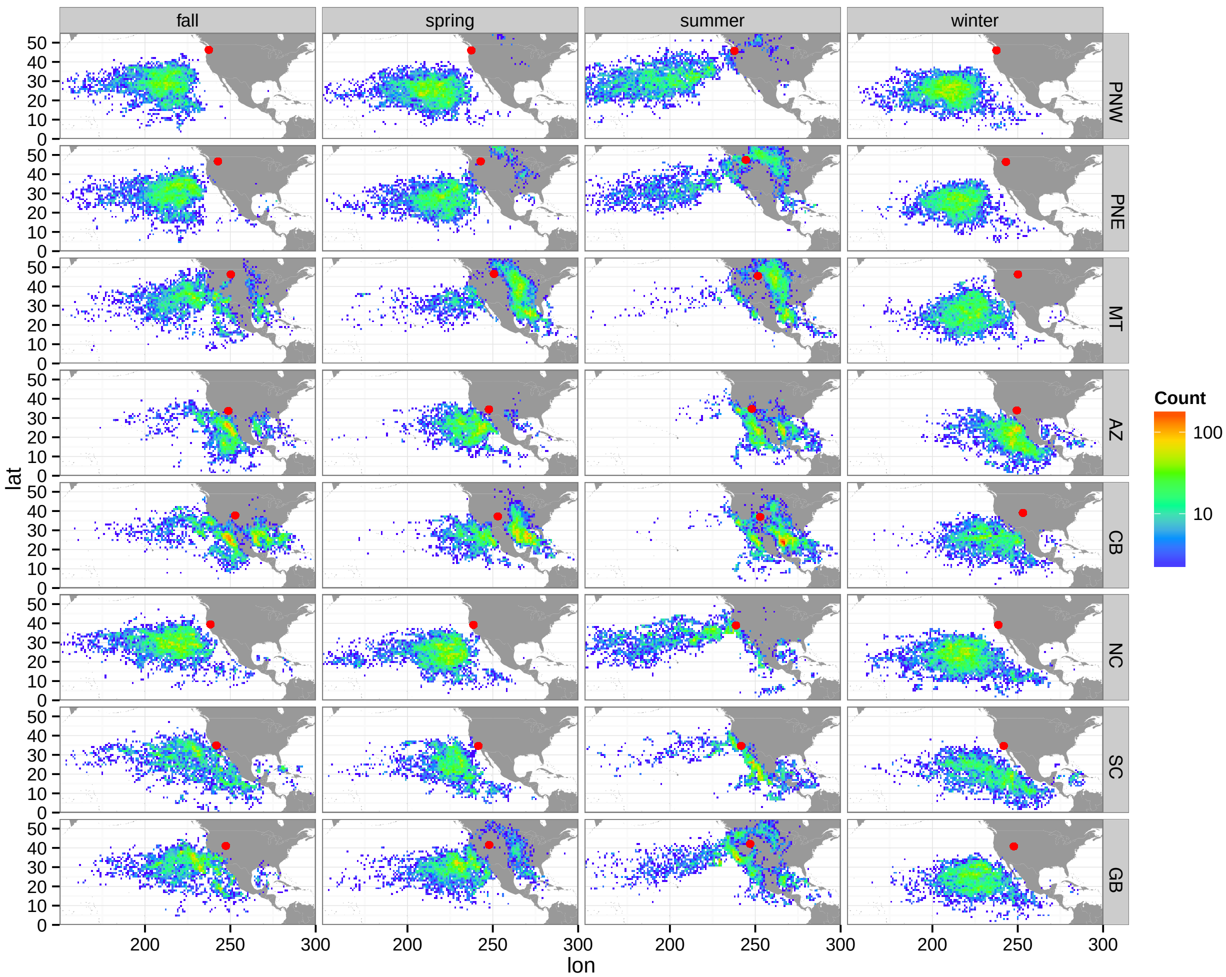

Figure 3: Moisture source location counts for each season and region, binned on a 1° grid. Red dots indicate the center of each extreme region.

The trajectory data was used to investigate the effects of ENSO on on the frequency of heavy rainfall. 100 trajectories per station and season were kept that showed the highest loss of specific humidity during the event period. Figure 4 was developed by taking the ratio of the number of trajectories occurring in El Niño vs. La Niña years. Points in red indicate the station experiences more heavy rainfall in El Ninõ years and blue indicates more events in La Niña years. Strong spatial coherence is eveident, indicating regional responses to ENSO regimes that corroborates previous research [*Gershunov and Barnett*, 1998; *Cayan et al.*, 1999; *Higgins et al.*, 2010; *Feldl and Roe*, 2011; *DeFlorio et al.*, 2013]. The results also indicate a seasonally diverse response.

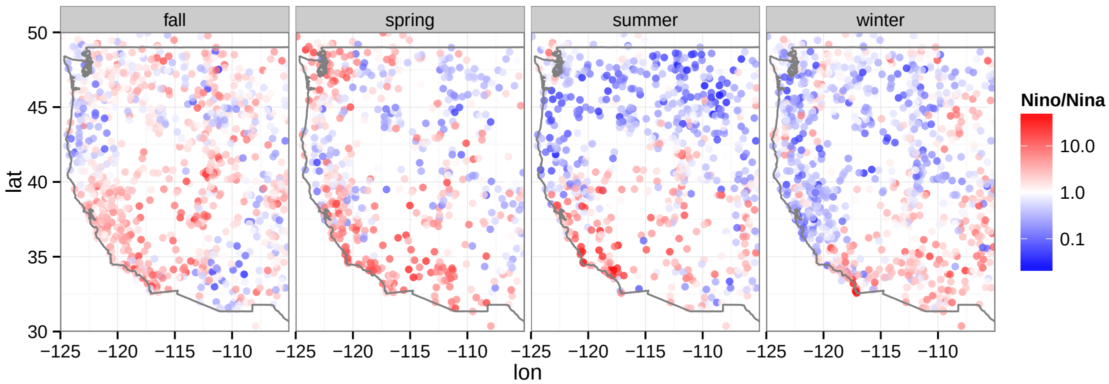

Figure 4: Ratio of the number of rain trajectories occurring in La Niña versus El Niño years. For points in red, more rain trajectories occurred in El Niño seasons and for points in blue, more rain trajectories occurred in La Niña seasons.

3 Efficient hierarchical spatial modeling of seasonal precipitation extremes

Status: Results are finalized and the papers in undergoing internal review.

3.1 Introduction

Modeling extreme precipitation is of crucial importance to engineering design. Accurate estimates of return levels and associated errors provide important information about risk for a variety of applications, including water supply management and flood control. Spatial modeling of extremes can capture spatial dependence between stations and reduce overall uncertainty in at-site return level estimates by borrowing strength across spatial locations [*Cooley et al.*, 2007]. Hierarchical Bayesian modeling of extremes precipitation was first introduced by [*Cooley et al.*, 2007] and since has been widely discussed in the literature [*Cooley and Sain*, 2010; *Aryal et al.*, 2010; *Atyeo and Walshaw*, 2012; *Davison et al.*, 2012; *Ghosh and Mallick*, 2011; *Reich and Shaby*, 2012; *Sang and Gelfand*, 2010, 2009; *Apputhurai and Stephenson*, 2013; *Dyrrdal et al.*, 2014]. Hierarchical modeling is an alternative to regional frequency analysis providing gridded or pointwise estimates of return levels within a study region [*Renard*, 2011].

Due to computational limitations, Bayesian spatial extremes models have typically been limited to small geographic regions that include on the order 100 stations. Large geographic regions with many stations present a computational challenge for heirachical Bayesian models. The computational bottleneck comes in when computing the likelihood of a Gaussian process which for n data points, requires inverting (or computing the Cholesky factor of) an $n \times n$ matrix, an $O(n^3)$ operation. *Banerjee et al.* [2008] present a method for estimating Gaussian processes for large data sets using Gaussian predictive processes (GPPs). In a GPP, knots are placed throughout the spatial domain, and the Gaussian process is estimated at only the knot locations. The gaussian field is then kriged to the station locations. This simplification allows for the incorporation of large datasets without substantially increasing computation time.

Some attempts have been made to model extremes in large regions and large datasets in a Bayesian hierarchical context. *Reich and Shaby* [2012] use a hierarchical max-stable model with climate model output in the east coast to examine spatially varying GEV parameters, with a max stable process for the data dependence level. [*Ghosh and Mallick*, 2011] model gridded precipitation data over the entire US, for annual maxima at a 5x5 degree resolution (43 grid cells) and copula for data dependence, incorporating spatial dependence directly in a spatial model on the data, not parameters. [*Cooley and Sain*, 2010] and [*Sang and Gelfand*, 2009] model over 1000 grid cells of climate model output using spatial autoregressive models which take advantage of data on a regular lattice to simplify computations.

In the western US, extremes behave differently in each season and between regions due to complex topography and climate interactions [*Kunkel et al.*, 1999, 2003; *Kunkel*, 2003b] (Bracken et al. 2015). The NOAA Atlas 14 is the foremost guide for precipitation return levels in the western US [*National Oceanic Atmospheric Administarion*, 2004]. Despite its comprehensive database of return levels at a multitude of return periods, the NOAA only provides information about annual extremes. Modeling annual maxima alone masks the seasonal variability of risk in this dynamic region. A seasonal analysis helps to pinpoint the magnitude of the expected extreme event throughout the year and quantifies the associated risk.

The research contributions of this study are as follows. We develop a Bayesian hierarchical model capable of incorporating thousands of observation locations by applying Gaussian predictive processes at the process layer. In addition the model is capable of incorporating stations with missing data with little computational overhead. This model can be used to efficiently incorporate large observational datasets for any spatially observed extremes. Finally, the model is applied to observed precipitation extremes in each season, providing estimated seasonal return levels for the western US.

3.2 Model Structure

Precipitation values at each site are modeled as conditionally independent draws from a generalized extreme value (GEV) distribution where the spatial dependence is captured through spatial processes on the location $\mu(\mathbf{s})$, scale $\sigma(\mathbf{s})$ and $\xi(\mathbf{s})$ parameters. We assume the parameters can be described through a latent spatial regression where the residual component $w_\gamma(\mathbf{s})$ follows a mean 0, stationary, isotropic Gaussian process (GP) with covariance function $C_\gamma(\mathbf{s}, \mathbf{s}')$. The corresponding covariance matrix is $C_\gamma(\boldsymbol{\theta}_\gamma) = [C(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}_\gamma)]_{i,j=1}^n$, where γ represents any GEV parameter (μ, σ, ξ) and $\boldsymbol{\theta}_\gamma$ represents the covariance parameters. The hierarchical model structure is:

$$Y(\mathbf{s})|\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s}) \sim \text{GEV}[\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})] \tag{1}$$

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

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

$$\xi(\mathbf{s}) = \mathbf{x}_\xi^T(\mathbf{s})\, \boldsymbol{\beta}_\xi + w_\xi(\mathbf{s}). \tag{4}$$

In the data layer (Equation 1), $Y(\mathbf{s})|\mu(\mathbf{s}), \sigma(\mathbf{s}), \xi(\mathbf{s})$ indicates conditional independence of the data given the GEV parameters, a common assumption in spatial extremes modeling [*Cooley and Sain*, 2010]. This assumption precludes the simulation of realistic fields of extremes but does not inhibit the computation of return levels, the main goal of this study. Alternatives to the conditionally independent assumption include multivariate extreme value copulas [*Renard*, 2011; *Ghosh and Mallick*, 2011; *Renard and Lang*, 2007] and max-stable processes [*Smith*, 1990; *Schlather*, 2002; *Cooley et al.*, 2006; *Shang et al.*, 2011; *Padoan et al.*, 2010], which proved to be computationally intractable for this application.

In the process layer (Equations 2-4), $\mathbf{x}_\gamma^T(\mathbf{s}_i)$ is a vector of p_γ spatially varying covariates (prepended with a 1) and $\boldsymbol{\beta}_\gamma$ is a vector of $p_\gamma + 1$ regression coefficients (including an intercept term). Choice of covariates will be discussed in Section 3.3.2.

The shape parameter ξ is notoriously difficult to estimate, its value determing the support of the GEV distribution. Positive values of ξ indicate a lower bound to the distribution, negative values indicate an upper bound and zero indicates no bounds. In many studies, ξ is modeled as a single value per study area or per region within the study area [*Cooley et al.*, 2007; *Renard*, 2011; *Atyeo and Walshaw*, 2012; *Apputhurai and Stephenson*, 2013]. As in [*Cooley and Sain*, 2010], we cannot assume that this parameter is constant over the large study region and so it is modeled spatially along with the other GEV parameters.

3.3 Application to the Western US

3.3.1 Precipitation Data

Daily precipitation data is obtained from the Global Historical Climatology Network (GHCN). We use all available stations in the western US that contain more than 30 years of data from 1950-2013. 3-day maxima were computed for each season, winter (DJF), spring (MAM), summer (JJA) and fall (SON). For a data point to be included from a season, we require no more than 25% of the days be missing. The number of stations in each season (with the number of complete stations in parentheses) is, winter: 2326 (932), spring: 2286 (774), summer: 2288 (794) and fall: 2286 (791) stations. Figure 5 shows the staion locations for fall, with solid black points indicating stations with complete data and filled grey points indicating stations with incomplete data. Each other season has nearly identical distribution of stations.

3.3.2 Covariates

For all GEV parameters the same covariates are used, i.e., $\mathbf{x}_\mu(\mathbf{s}) = \mathbf{x}_\sigma(\mathbf{s}) = \mathbf{x}_\xi(\mathbf{s}) = \mathbf{x}(\mathbf{s})$. The covariates are elevation, mean seasonal precipitation, latitude and longitude. Covariates were obtained at knot locations, station locations and at a 1/8th degree grid throughout the study area. Elevation data was obtained from the NASA Land Data Assimilation Systems (NLDAS) website [1] [*Xia et al.*, 2012a, b]. Mean seasonal precipitation was computed from the Maurer dataset [*Maurer et al.*, 2002].

3.4 Selected results

Figures 6 and 7 show the median return level and the associated width of the 95% credible interval for fall, winter, spring and summer, respectively. Note the logarithmic color scale. Seasonal differences in the return levels are immediately apparent. The highest return levels are present in in coastal mountain ranges, these areas also tend to have the largest credible intervals due to the strong positive correlation between μ and σ. The dots in the credible interval plots are due to the knot placement and conditional simulation. At the knots, the credible interval is small, and it increases further from the the knot locations.

[1] http://ldas.gsfc.nasa.gov/nldas/NLDASelevation.php

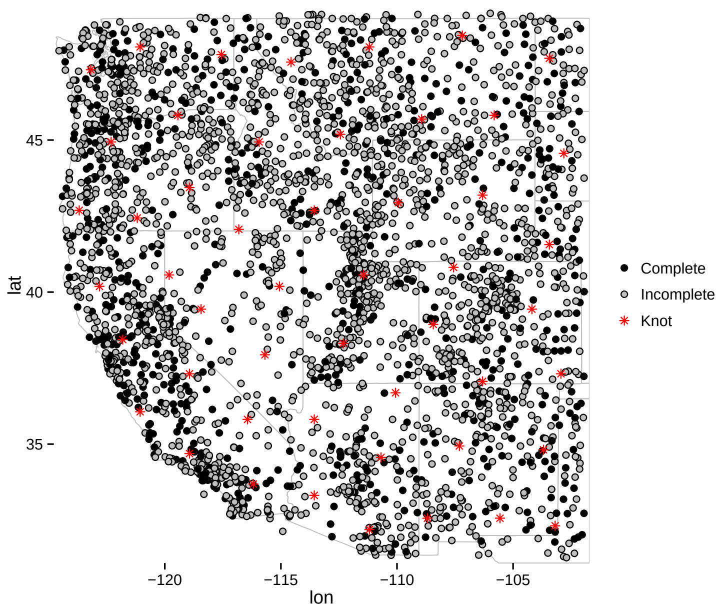

Figure 5: Station locations with complete data (black solid dots) and station locations with incomplete data (grey filled dots) for Fall. Note that station locations change slightly for each season.

To illustrate the benefit of spatially modeling ξ, we show the median of the shape parameter field after interpolation by conditional simulation. Figure 8 shows the fields for each season. The shape parameter fields reveal intricate spatial patterns. Values are nearly always positive, indicating a heavy upper tail. A region of the southwest US has consistently high values in each season.

A crux of this model is the use of appropriate spatial covariates. Mean seasonal precipitation (MSP) had a correlation of 95% with the MLE estimates of μ and 75% with the MLE estimates of σ. This covariate went a long way in generating realistic spatial variability, leaving small residuals. The covariates also help to reveal a complex spatial pattern for ξ. The strongest covariate for ξ was elevation. The seasonally dependent spatial variability in ξ shows that it is inappropriate to model it as a single value for anything but the smallest regions.

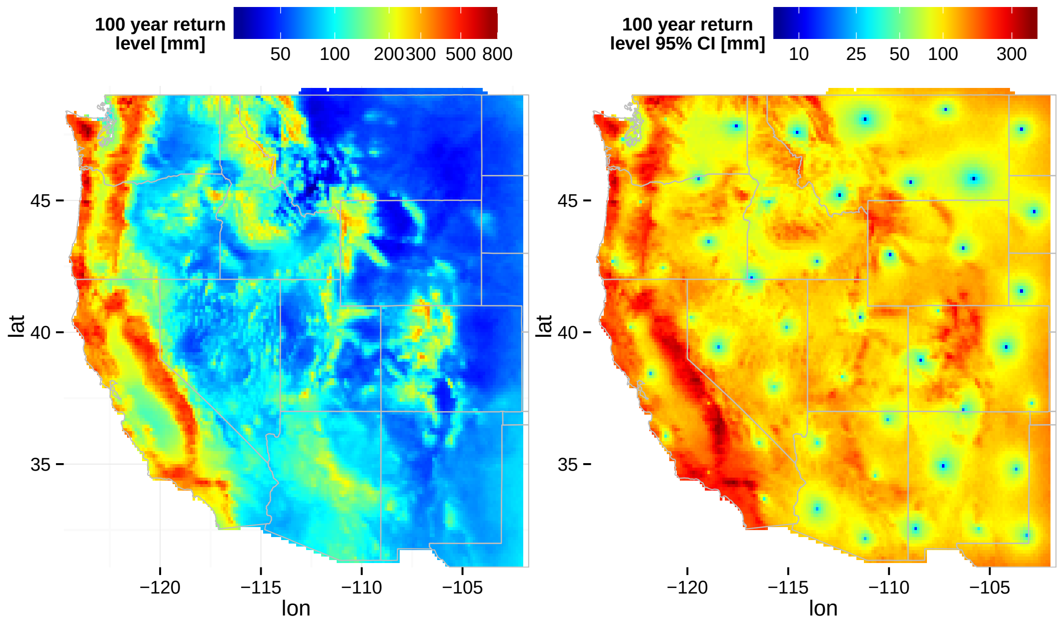

Figure 6: Median 100-year return levels for winter (left) and width of corresponding 95% credible interval (right). Note the logarithmic color scale.

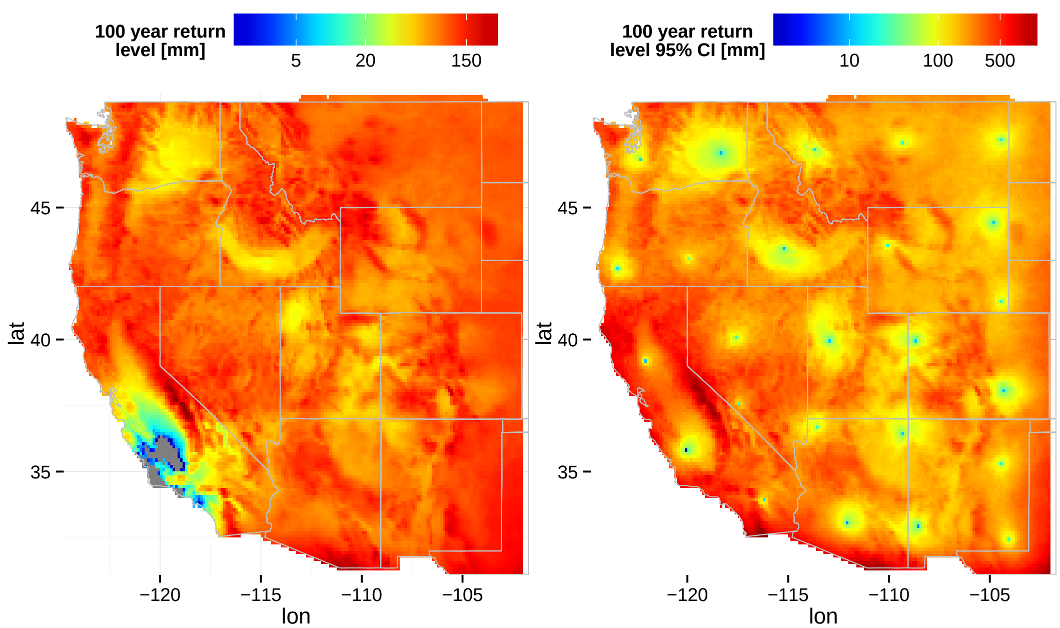

Figure 7: Median 100-year return levels for winter (left) and width of corresponding 95% credible interval (right). Note the logarithmic color scale.

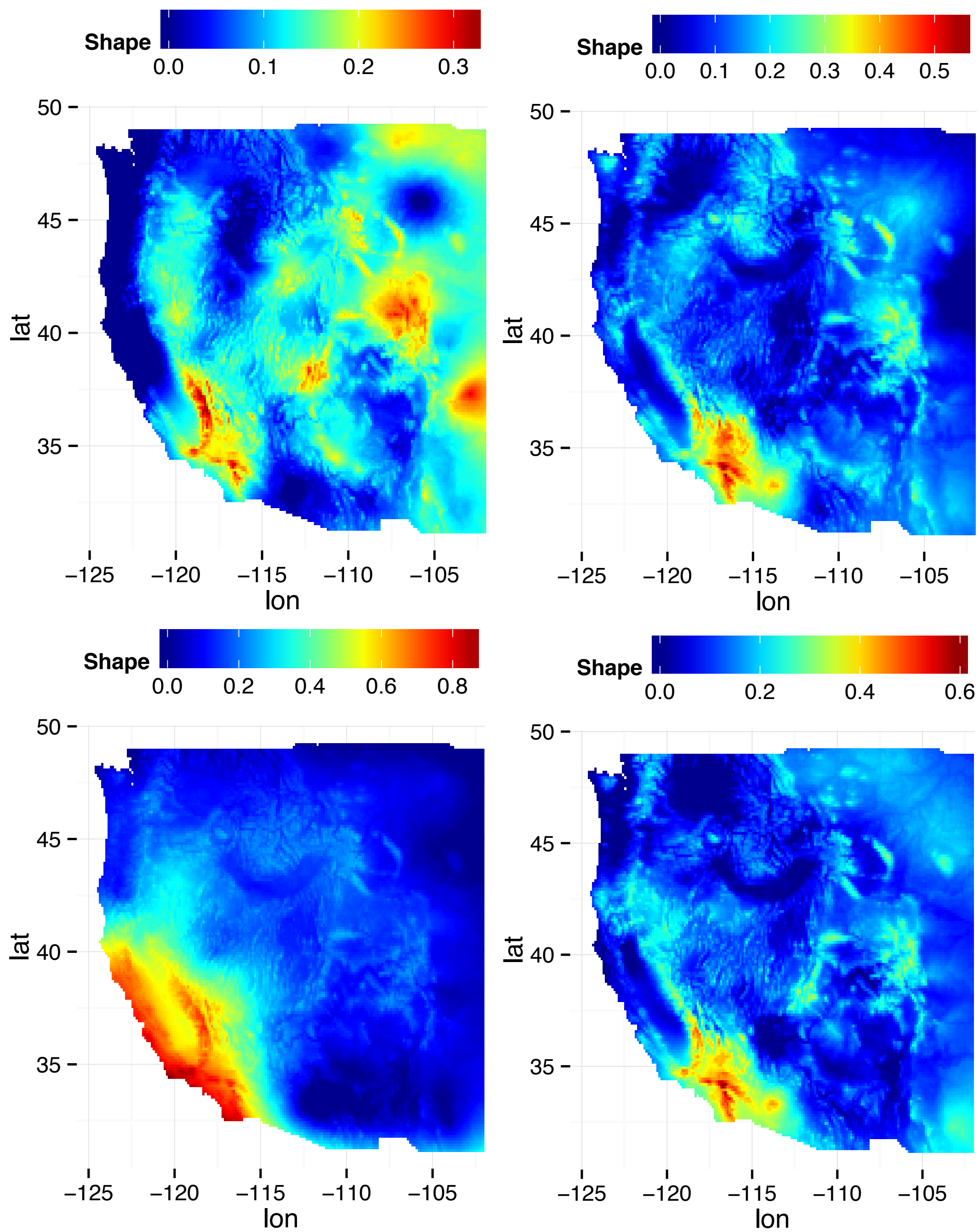

Figure 8: Median of underlying shape parameter field for winter (top left), spring (top right), summer (bottom right) and (fall) bottom right.

4 Coupled hierarchical modeling of streamflow and precipitation extremes

Status: Data is ready and preliminary model development is underway

4.1 Introduction

There is strong evidence for nonstationarity of streamflow and precipitation extremes in the Western US [*Bracken et al.*, 2015; *Kunkel et al.*, 2003; *Groisman et al.*, 2001; *Kunkel et al.*, 1999; *Cayan et al.*, 1999], both due to linear trends in time and large scale atmospheric oscillations (AMO, PDO, ENSO). In the literature, there are many examples of nonstationary spatial precipitation [*Apputhurai and Stephenson*, 2013; *Ghosh and Mallick*, 2011; *Sang and Gelfand*, 2009] and streamflow [*Najafi and Moradkhani*, 2014; *Renard and Lang*, 2007; *Reza Najafi and Moradkhani*, 2013; *Yan and Moradkhani*, 2014] models, yet these analyses are typically done separately. It is reasonable to assume, especially in rainfall-runoff dominated regions, that annual or seasonal maximum precipitation and streamflow are closely related.

This chapter will focus on the development and application of a Bayesian nonstationary spatial extremes model for precipitation in which streamflow is linked into the hierarchy at the parameter level. The goal of building such a model is to estimate spacetime return levels for streamflow and precipitation simultaneously.

4.2 Model Structure

The joint distribution of the m data in each year is modeled as a realization from a Gaussian elliptical copula with generalized extreme value (GEV) distribution marginals. The copula is characterized by pairwise dependence matrix Σ. Spatial dependence is further captured through spatial processes on the location $\mu(\mathbf{s})$, scale $\sigma(\mathbf{s})$ and $\xi(\mathbf{s})$ parameters. We assume the parameters can be described through a latent spatial regression where the residual component $w_\gamma(\mathbf{s})$ follows a mean 0, stationary, isotropic Gaussian process (GP) with covariance function $C(\mathbf{s}, \mathbf{s}')$. The corresponding covariance matrix is $C(\boldsymbol{\theta}_\gamma) = [C_\gamma(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}_\gamma)]_{i,j=1}^m$, where γ represents the location or scale paramter (μ, σ) and $\boldsymbol{\theta}_\gamma$ represents the covariance parameters. The precipitation model structure is:

$$y(\mathbf{s}, t) \sim Gcop_m[\Sigma; \mu_y(\mathbf{s}, t), \sigma_y(\mathbf{s}, t), \xi_y] \tag{5}$$

$$\mu_y(\mathbf{s}, t) = \mathbf{x}_\mathbf{s}^T(\mathbf{s})\, \boldsymbol{\beta}_{\mu,\mathbf{s}} + \mathbf{x}_t^T(t)\, \boldsymbol{\beta}_{\mu,t} \tag{6}$$

$$\sigma_y(\mathbf{s}, t) = \mathbf{x}_\mathbf{s}^T(\mathbf{s})\, \boldsymbol{\beta}_{\sigma,\mathbf{s}} + \mathbf{x}_t^T(t)\, \boldsymbol{\beta}_{\sigma,t} \tag{7}$$

where $y(\mathbf{s}, t)$ is the precipitation response at generic site s and time t and $Gcop_m$ stands for "m-dimensional Gaussian elliptical copula" with dependence matrix Σ [*Renard*, 2011; *Renard and Lang*, 2007]. The data layer processes in each year are assumed independent and identically distributed. Alternatives to using a copula to construct the joint distribution are an assumption of conditional independence [*Cooley et al.*, 2007; *Sang and Gelfand*, 2009] and max-stability [*Smith*, 1990; *Schlather*, 2002; *Cooley et al.*, 2006; *Shang et al.*, 2011; *Padoan et al.*, 2010]. In contrast to the conditional independence assumption, the spatial dependence is captured at the data layer and there is no need to model μ_y and σ_y with additional spatial processes.

In the process layer (Equations 6 and 7), $\mathbf{x}_{\mathbf{s}}^T(\mathbf{s})$ and $\mathbf{x}_t^T(t)$ are vectors of spatial and temporal covariates (including an intercept), respectively, and $\boldsymbol{\beta}$'s are corresponding regression coefficients. Spatial covariates will include latitude, longitude, elevation and mean seasonal precipitation. Temporal covariates will include the year (linear trend), ENSO, PDO and AMO. Given the small region, the shape parameter ξ_z is modeled as a single unknown parameter for the entire region constant in time.

The streamflow model structure is

$$z(t) \sim \mathrm{GEV}[\mu_z(t), \sigma_z(t), \xi_z] \tag{8}$$

$$\mu_z(t) = b_\mu \mu_y(\mathbf{s}^*, t) + c_\mu \tag{9}$$

$$\sigma_z(t) = b_\sigma \sigma_y(\mathbf{s}^*, t) + c_\sigma \tag{10}$$

where $z(t)$ is the flow in year t at location s^*. We assume the the location and scale parameters of the precipitation field at the basin outlet are linearly related, b and c are regression coefficients. This simple hierarchical formulation allows the streamflow distribution in each year to be linked to the precipitation field. The shape parameter ξ_z is modeled as a single unknown paramter constant in time.

4.3 Elliptical copula for data dependence

Elliptical copulas are a flexible tool for modeling multivariate data [*Renard*, 2011; *Sang and Gelfand*, 2010; *Ghosh and Mallick*, 2011; *Renard and Lang*, 2007]. This class of copulas can represent spatial data with any marginal distribution, a particularly attractive feature for extremal data. The Gaussian copula constructs the joint pdf of a random vector $(Y_1, ...Y_m)$ as

$$F_{Gaussian}(y_1, \ldots, y_m) = \Phi_\Sigma(u_1, ...u_m) \tag{11}$$

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

$$f_{Gaussian}(y_1, \ldots, y_m) = \frac{\displaystyle\prod_{i=1}^{m} f_i[y_i]}{\displaystyle\prod_{i=1}^{m} \psi[u_i]} \Psi_\Sigma(u_1, ...u_m) \tag{12}$$

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

The dependence between sites is assumed to be a function of distance [*Renard*, 2011]. The dependence matrix is constructed with a simple exponential model

$$\Sigma(i,j) = \exp(-\|\,\mathbf{s}_i - \mathbf{s}_j\,\|/a_0) \tag{13}$$

where a_0 is the copula range parameter. Note that the values in this dependence matrix are not covariances. So by analogy with the variogram, the dependence model is termed the dependogram [*Renard*, 2011].

4.4 Application

The model will be applied to a small basin in southern California which is rainfall-runoff dominated (Figure 9). This basin was chosen for having a relatively long corresponding unimpaired flow and precipitation records (∼30 years). The flow data is from the Large-Sample Hydrometeorological Dataset [*Newman et al.*, 2014, 2015].

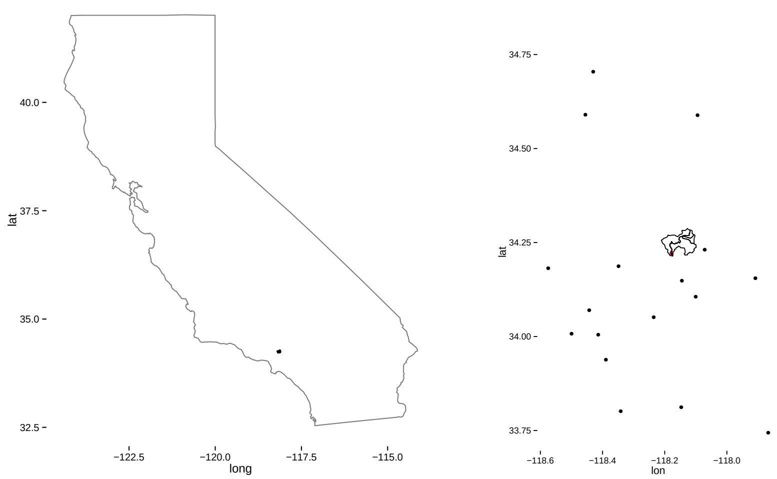

Figure 9: Study area for the linked streamflow-precipitation model.

4.5 Expected Results

The results are expected to show spatial and temporal variability in return levels even in such a small basin. Comparisons will be made with frequency estimates conducted separately for both precipitation and streamflow. Looking at the relative effects of spatial and temporal covariates will be a focus of the analysis.

5 Hydroclimate frequency analyses for dam safety: case studies using traditional and modern methodologies

Status: In collaboration with Reclamation, a potential case studies has been identified

5.1 Introduction

The truth of the matter is, we will never know how accurate any return level estimates are, particularly for long return periods (eg. 10,000-year). The best that can be done is to provide an honest representation of the uncertainty in our estimates given the data. Bayesian methods excel at this. The question then becomes, how much do the estimates of return levels differ when using spatial and nonstationary Bayesian models versus traditional methods.

The goal of this chapter is to (1) provide an overview of the state of the practice for frequency analyses for dam safety in the western US and (2) provide comparisons of return level estimates with a number of dam safety studies. One study has already been identified and is listed here:

- Friant Dam Hydrologic Hazard for Issue Evaluation, Reclamation, 2013

This report uses a variety of methods to estimate return levels for precipitation, streamflow and reservoir elevation. When possible, the methods described in the previous two chapters will be compared directly to the results in this chapter and other dam safety reports.

5.2 Current methods

Reclamation dams in the western US regularly undergo hydrologic hazard assessments such as in the Friant Dam study listed above. The main steps in such an analysis are:

1. Collect spatial and temporal profiles of extreme storms and historic and paleo peak flow data for a basin.
2. Develop precipitation and flood frequency curves for the basin (by assuming some probability distribution for the data).
3. Develop reservoir elevation frequency curve by routing peak flow event hydrographs (Figure 10, for example).
4. Quantify uncertainty with resampling approach (via Latin Hypercube Sampling for example).

5.3 Proposed methodology

There are two main drawbacks in the steps above: (1) frequency curves for precipitation, flow and reservoir elevation are estimated independently, making uncertainty propaga-

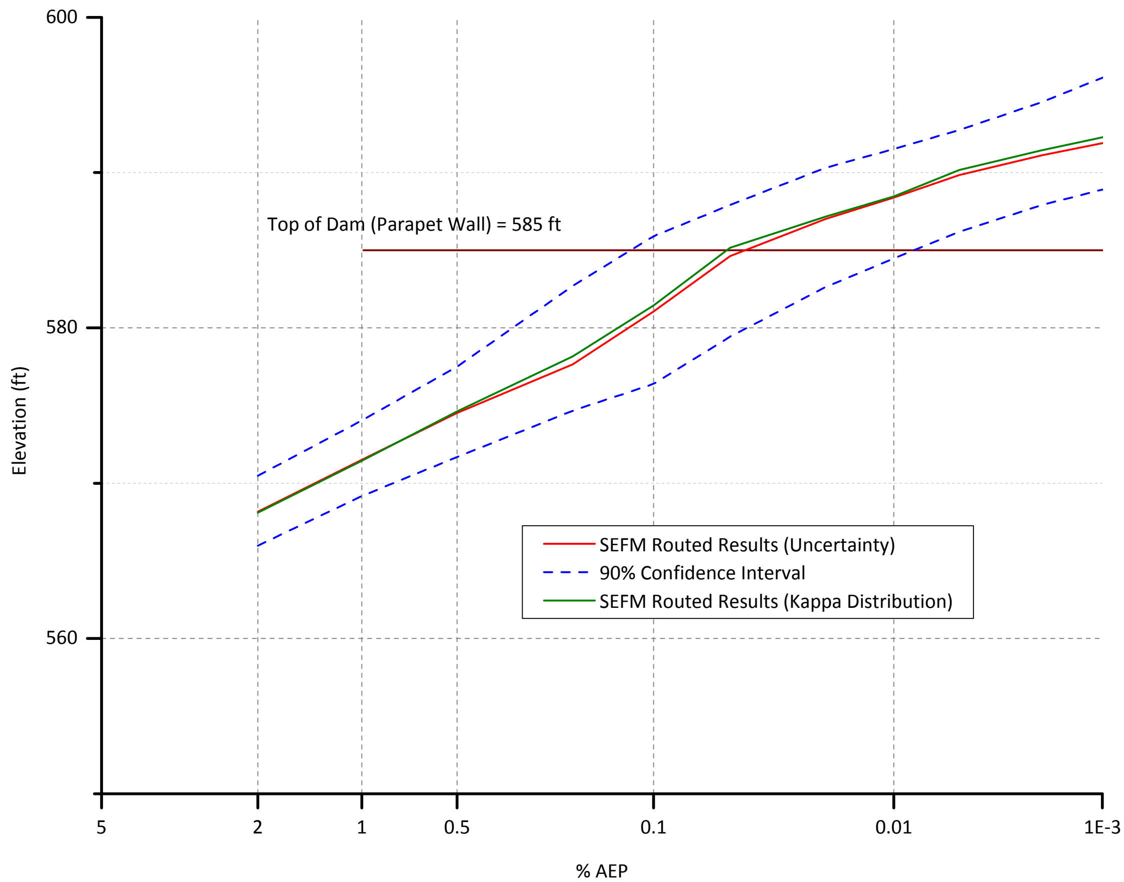

Figure 10: Maximum reservoir elevation frequency curve for Friant Dam (Reclamation, 2010).

tion difficult, and (2) return levels are developed under an assumption of stationarity. The methodology proposed for this chapter is to develop a spatial nonstationary hierarchical extremes model that can be directly compared to the results in the dam safety studies. Precipitation, flow and elevation will be linked in a nonstationary Bayesian hierarchical model similar to that in Chapter 4. A main focus of these comparisons will be the uncertainty associated with these frequency curves.

5.4 Expected results

The results of this chapter are expected to look somewhat similar to the results of previous dam safety studies but with potentially very different uncertainty profiles. For example, in Figure 10 the uncertainty bounds might greatly increase with return period, reflecting the true uncertainty in this quantity. One challenge will be comparing the nonstationary estimates provided by the Bayesian model to the stationary estimates in the dam safety studies.

6 Contributions

The novel contributions of this research are as follows:

Chapter 2

- Development of an improved extreme value clustering method that incorporates spatial proximity of stations.
- Identification of dominant moisture sources and pathways for seasonal extreme precipitation throughout the western US.
- Independent confirmation of ENSO effects on extreme precipitation frequency in the western US.

Chapter 3

- Development of a Bayesian hierarchical spatial extremes model capable of incorporating thousands of observation locations by applying Gaussian predictive processes at the process layer.
- Development of a procedure for incorporating stations with missing data in the above model with little computational overhead.
- Application of the above model to seasonal precipitation extremes, providing maps of return level with associated uncertainty for the western US.

Chapter 4

- Development and application of a nonstationary spatial Bayesian hierarchical model linking streamflow and precipitation extremes for simultaneous frequency analysis.

Chapter 5

- Development of a nonstationary spatial hierarchical model for linking precipitation, flow and reservoir elevation extremes.
- Direct comparison of the frequency estimates from the above model with state of the practice dam hydrologic hazard studies, particularly focusing on uncertainty estimates.

7 Status and timeline

Chapter 2

Status: Completed and published in JGR: Atmospheres

Chapter 3

Status: Draft manuscript completed, under internal review

Timeline: Expect to submit Summer 2015

Chapter 4

Status: Data is ready and preliminary model development is underway

Timeline: Expect to complete Fall 2015

Chapter 5

Status: In collaboration with Reclamation, one potential case studies has been identified, more are expected

Timeline: Expect to complete Spring 2016

Thesis defense

Status: In collaboration with Reclamation, one potential case studies has been identified, more are expected

Timeline: Expected Spring/Summer 2016

References

Alexander, L. V., et al., Global observed changes in daily climate extremes of temperature and precipitation, *Journal of Geophysical Research: Atmospheres*, *111*(D5), D05,109, 2006.

Alexander, M. A., J. D. Scott, D. Swales, M. Hughes, K. Mahoney, and C. A. Smith, Moisture Pathways into the US Intermountain West Associated with Heavy Winter Precipitation Events, *Journal of Hydrometeorology*, 2015.

Apputhurai, P., and A. G. Stephenson, Spatiotemporal hierarchical modelling of extreme precipitation in Western Australia using anisotropic Gaussian random fields, *Environmental and Ecological Statistics*, *20*(4), 667–677, 2013.

Arriaga Ramírez, S., and T. Cavazos, Regional trends of daily precipitation indices in northwest Mexico and southwest United States, *Journal of Geophysical Research: Atmospheres*, *115*(D14), D14,111, 2010.

Aryal, S. K., B. C. Bates, E. P. Campbell, Y. Li, M. J. Palmer, and N. R. Viney, Characterizing and Modeling Temporal and Spatial Trends in Rainfall Extremes, *dx.doi.org*, *10*(1), 241–253, 2010.

Atyeo, J., and D. Walshaw, A region-based hierarchical model for extreme rainfall over the UK, incorporating spatial dependence and temporal trend, *Environmetrics*, *23*(6), 509–521, 2012.

Balling Jr, R. C., and G. B. Goodrich, Spatial analysis of variations in precipitation intensity in the USA, *Theoretical and Applied Climatology*, *104*(3-4), 415–421, 2011.

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

Bernard, E., P. Naveau, M. Vrac, and O. Mestre, Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France, *Journal of Climate*, 2013.

Bracken, C., B. Rajagopalan, M. Alexander, and S. Gangopadhyay, Spatial variability of seasonal extreme precipitation in the western United States, *Journal of Geophysical Research: Atmospheres*, *120*(10), 4522–4533, 2015.

Cayan, D. R., K. T. Redmond, and L. G. Riddle, ENSO and Hydrologic Extremes in the Western United States*, *Journal of Climate*, *12*(9), 2881–2893, 1999.

Cayan, D. R., M. D. Dettinger, H. F. Diaz, and N. E. Graham, Decadal Variability of Precipitation over Western North America, *http://dx.doi.org/10.1175/1520-0442(1998)011¡3148:DVOPOW¿2.0.CO;2*, *11*(12), 3148–3166, 2010.

Cooley, D., and S. R. Sain, Spatial Hierarchical Modeling of Precipitation Extremes From a Regional Climate Model, *Journal of Agricultural, Biological, and Environmental Statistics*, *15*(3), 381–402, 2010.

Cooley, D., P. Naveau, and P. Poncet, Variograms for spatial max-stable random fields, *Dependence in probability and statistics*, 2006.

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

Dales, M. Y., and D. W. Reed, Regional flood and storm hazard assessment - NERC Open Research Archive, 1989.

Davison, A. C., S. A. Padoan, and M. Ribatet, Statistical Modeling of Spatial Extremes, *Statistical Science*, *27*(2), 161–186, 2012.

DeFlorio, M. J., D. W. Pierce, D. R. Cayan, and A. J. Miller, Western U.S. Extreme Precipitation Events and Their Relation to ENSO and PDO in CCSM4, *Journal of Climate*, *26*(12), 4231–4243, 2013.

DeGaetano, A. T., A Smirnov test–based clustering algorithm with application to extreme precipitation data, *Water Resources Research*, 1998.

Dominguez, F., E. Rivera, and D. P. Lettenmaier, Changes in winter precipitation extremes for the western United States under a warmer climate as simulated by regional climate models - Dominguez - 2012 - Geophysical Research Letters - Wiley Online Library, *Geophysical Research . . .* , 2012.

Dos Santos, C., C. Neale, and T. Rao, Trends in indices for extremes in daily temperature and precipitation over Utah, USA, *International Journal . . .* , 2011.

Draxler, R. R., and G. D. Hess, Description of the HYSPLIT4 modeling system, *Techreport*, p. 24pp, 1997.

Draxler, R. R., and G. D. Hess, An overview of the HYSPLIT4 modelling system for trajectories, *Australian Meteorological Magazine*, 1998.

Draxler, R. R., B. Stunder, and G. Rolph, HYSPLIT4 user's guide, *Techreport*, 1999.

Dulière, V., Y. Zhang, and E. P. Salathé Jr., Changes in Twentieth-Century Extreme Temperature and Precipitation over the Western United States Based on Observations and Regional Climate Model Simulations*, *Journal of Climate*, *26*(21), 8556–8575, 2013a.

Dulière, V., Y. Zhang, and E. P. Salathé Jr., Changes in Twentieth-Century Extreme Temperature and Precipitation over the Western United States Based on Observations and Regional Climate Model Simulations*, *Journal of Climate*, *26*(21), 8556–8575, 2013b.

Dyrrdal, A. V., A. Lenkoski, T. L. Thorarinsdottir, and F. Stordal, Bayesian hierarchical modeling of extreme hourly precipitation in Norway, *Environmetrics*, pp. n/a–n/a, 2014.

Feldl, N., and G. H. Roe, Climate Variability and the Shape of Daily Precipitation: A Case Study of ENSO and the American West, *dx.doi.org*, *24*(10), 2483–2499, 2011.

Gershunov, A., and T. P. Barnett, ENSO Influence on Intraseasonal Extreme Rainfall and Temperature Frequencies in the Contiguous United States: Observations and Model Results, *Journal of Climate*, *11*(7), 1575–1586, 1998.

Ghosh, S., and B. K. Mallick, A hierarchical Bayesian spatio-temporal model for extreme precipitation events, *Environmetrics*, *22*(2), 192–204, 2011.

Groisman, P. Y., R. W. Knight, and T. 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, 2001.

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

Gustafsson, M., D. Rayner, and D. Chen, Extreme rainfall events in southern Sweden: where does the moisture come from?, *Tellus A*, *62*(5), 605–616, 2010.

Higgins, R. W., J.-K. E. Schemm, W. Shi, and A. Leetmaa, Extreme Precipitation Events in the Western United States Related to Tropical Forcing, *Journal of Climate*, *13*(4), 793–820, 2010.

Hosking, J. R. M., J. R. WALLIS, and E. F. WOOD, An appraisal of the regional flood frequency procedure in the UK Flood Studies Report, *Hydrological Sciences Journal*, *30*(1), 85–109, 1985.

Janssen, E., D. J. Wuebbles, K. E. Kunkel, S. C. Olsen, and A. Goodman, Observational- and model-based trends and projections of extreme precipitation over the contiguous United States, *Earth's Future*, *2*(2), 99–113, 2014.

Jiang, T., K. J. Evans, Y. Deng, and X. Dong, Intermediate frequency atmospheric disturbances: A dynamical bridge connecting western U.S. extreme precipitation with East Asian cold surges, *Journal of Geophysical Research: Atmospheres*, *119*(7), 3723–3735, 2014.

Jones, M. R., S. Blenkinsop, and H. J. Fowler, Objective classification of extreme rainfall regions for the UK and updated estimates of trends in regional extreme rainfall, *International Journal of Climatology*, 2014.

Jorba, O., C. Pérez, F. Rocadenbosch, and J. M. Baldasano, Cluster Analysis of 4-Day Back Trajectories Arriving in the Barcelona Area, Spain, from 1997 to 2002, *Journal of Applied Meteorology*, *43*(6), 887–901, 2004.

Katz, R. W., M. B. Parlange, and P. Naveau, Statistics of extremes in hydrology, *Advances in Water Resources*, *25*(8-12), 1287–1304, 2002.

Kunkel, K. E., Sea surface temperature forcing of the upward trend in U.S. extreme precipitation, *Journal of Geophysical Research: Atmospheres*, *108*(D1), ACL 6–1–ACL 6–10, 2003a.

Kunkel, K. E., North American Trends in Extreme Precipitation, *Natural Hazards*, *29*(2), 291–305, 2003b.

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

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

Maraun, D., T. J. Osborn, and N. P. Gillett, United Kingdom daily precipitation intensity:

improved early data, error estimates and an update from 2000 to 2006, *International Journal of Climatology*, 2008.

Massacand, A. C., H. Wernli, and H. C. Davies, Heavy precipitation on the alpine southside: An upper-level precursor, *Geophysical Research Letters*, *25*, 1438, 1998.

Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen, A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States, *Journal of Climate*, *15(22)*(22), 3237–3251, 2002.

Mullens, E. D., M. Shafer, and J. Hocker, Trends in heavy precipitation inthe southern USA, *Weather*, *68*(12), 311–316, 2013.

Najafi, M. R., and H. Moradkhani, A hierarchical Bayesian approach for the analysis of climate change impact on runoff extremes, *Hydrological Processes*, *28*(26), 6292–6308, 2014.

National Oceanic Atmospheric Administarion, Precipitation-Frequency Atlas of the United States, *Tech. rep.*, NOAA, 2004.

Neal, R. A., and I. D. Phillips, Summer daily precipitation variability over the East Anglian region of Great Britain, *International Journal of Climatology*, 2009.

Newman, A., K. sampson, M. P. Clark, A. Bock, R. J. Viger, and D Blodgett, A large-sample watershed-scale hydrometeorological dataset for the contiguous USA, 2014.

Newman, A. J., et al., Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, *Hydrology and Earth System Sciences*, *19*(1), 209–223, 2015.

Padoan, S. A., M. Ribatet, and S. A. Sisson, Likelihood-Based Inference for Max-Stable Processes, *dx.doi.org*, *105*(489), 263–277, 2010.

Pal, I., B. T. Anderson, G. D. Salvucci, and D. J. Gianotti, Shifting seasonality and increasing frequency of precipitation in wet and dry seasons across the U.S., *Geophysical Research Letters*, *40*(15), 4030–4035, 2013.

Palecki, M. A., J. R. Angel, and S. E. Hollinger, Storm Precipitation in the United States. Part I: Meteorological Characteristics, *Journal of Applied Meteorology*, *44*(6), 933–946, 2005.

Pryor, S. C., J. A. Howe, and K. E. Kunkel, How spatially coherent and statistically robust are temporal changes in extreme precipitation in the contiguous USA?, *International Journal of Climatology*, *29*(1), 31–45, 2009.

Reale, O., L. Feudale, and B. Turato, Evaporative moisture sources during a sequence of floods in the Mediterranean region, *Geophysical Research Letters*, *28*(10), 2085–2088, 2001.

Reich, B. J., and B. Shaby, A HIERARCHICAL MAX-STABLE SPATIAL MODEL FOR EXTREME PRECIPITATION, *The annals of applied statistics*, *6*(4), 1430–1451, 2012.

Renard, B., A Bayesian hierarchical approach to regional frequency analysis, *Water Resources Research*, 2011.

Renard, B., and M. Lang, Use of a Gaussian copula for multivariate extreme value analysis: Some case studies in hydrology, *Advances in Water Resources*, *30*(4), 897–912, 2007.

Reza Najafi, M., and H. Moradkhani, Analysis of runoff extremes using spatial hierarchical Bayesian modeling, *Water Resources Research*, *49*(10), 6656–6670, 2013.

Sang, H., and A. E. Gelfand, Hierarchical modeling for extreme values observed over space and time, *Environmental and Ecological Statistics*, *16*(3), 407–426, 2009.

Sang, H., and A. E. Gelfand, Continuous Spatial Process Models for Spatial Extreme Values, *Journal of Agricultural, Biological, and Environmental Statistics*, *15*(1), 49–65, 2010.

Schlather, M., Models for Stationary Max-Stable Random Fields, *Extremes*, *5*(1), 33–44, 2002.

Schumacher, R. S., and R. H. Johnson, Characteristics of U.S. Extreme Rain Events during 1999–2003, *dx.doi.org*, *21*(1), 69–85, 2010.

Shang, H., J. Yan, and X. Zhang, El Niño–Southern Oscillation influence on winter maximum daily precipitation in California in a spatial model, *Water Resources Research*, *47*(11), n/a–n/a, 2011.

Smith, R. L., Max-stable processes and spatial extremes, *Unpublished manuscript*, 1990.

USGS, Guidelines for determining flood flow frequency, *Tech. rep.*, United States Geological Survey, 1981.

Warner, M. D., C. F. Mass, and E. P. Salathé Jr., Wintertime Extreme Precipitation Events along the Pacific Northwest Coast: Climatology and Synoptic Evolution, *Monthly Weather Review*, *140*(7), 2021–2043, 2012.

Wigley, T., J. M. Lough, and P. D. Jones, Spatial patterns of precipitation in England and Wales and a revised, homogeneous England and Wales precipitation series, *Journal of Climatology*, 1984.

Xia, Y., et al., Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, *J. Geophys. Res.*, *117*(D3), D03,110, 2012a.

Xia, Y., et al., Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products, *J. Geophys. Res.*, *117*(D3), D03,109, 2012b.

Yan, H., and H. Moradkhani, A regional Bayesian hierarchical model for flood frequency analysis, *Stochastic Environmental Research and Risk Assessment*, *29*(3), 1019–1036, 2014.