Statistical Modeling of Hydroclimate Extremes

Balaji Rajagopalan
Department of Civil, Environmental and Architectural Engineering
&
Co-operative Institute for Research in Environmental Sciences (CIRES)

University of Colorado
Boulder, CO

CLIVAR PPAI Breakout

August 9, 2017

Motivation for Extremes

Flood Damages

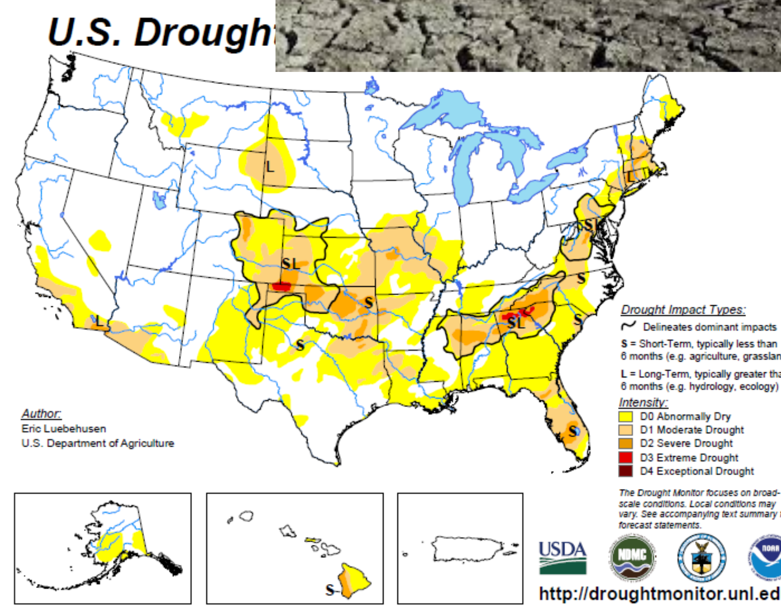
- From 1983 – 2000 Western States experienced ~$24.7 Billion in flood damages
- ~$1.5 Billion annually

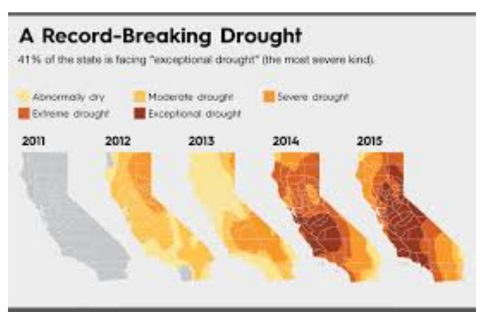
Droughts cause slow and long term Damages

- Recent prolonged drought in the West

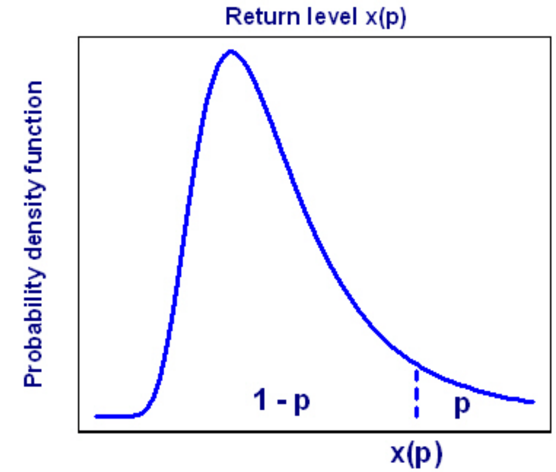
Flood drought whiplash in California

Increasing frequency/strength/damages of Climate extremes





Scientific Needs

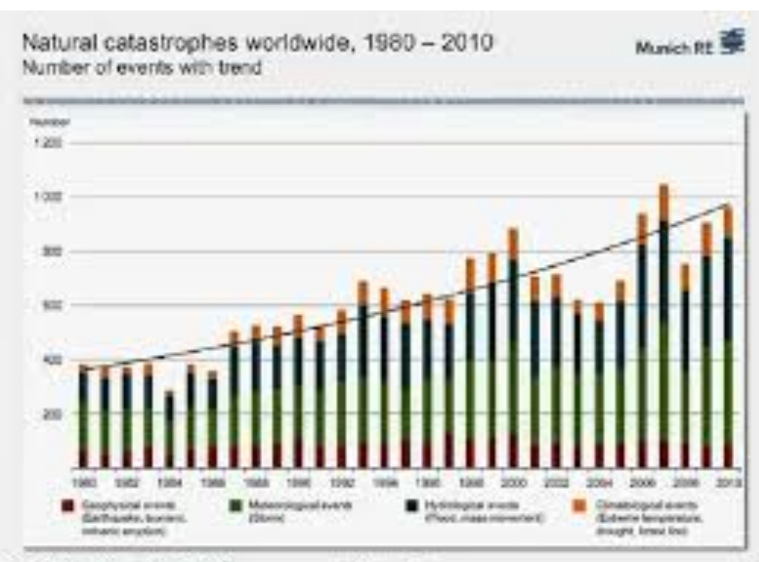


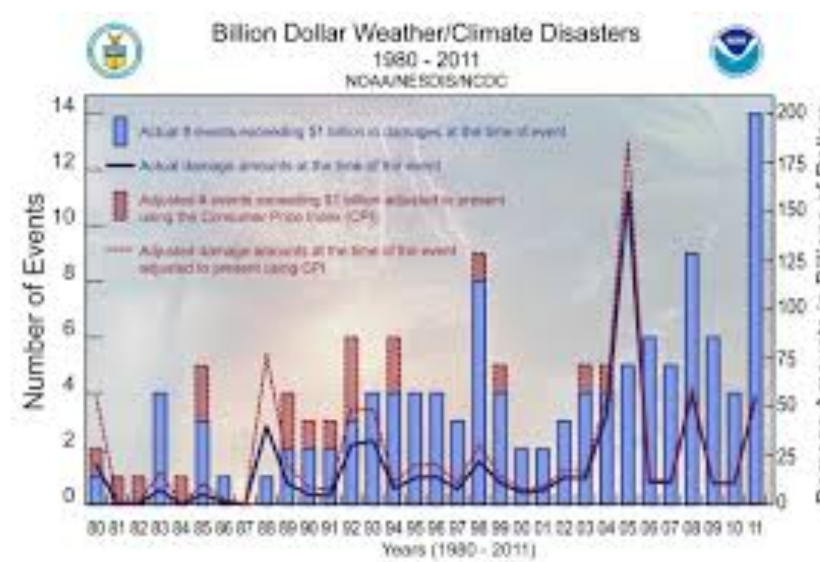
- Understanding how extremes are enabled

- Space-Time Modeling – Return Levels
  - Climate extremes
  - Extremes of decision variables (Multivariate Extremes)

- Tools for modeling extremes at multiple time scales - Downscaling
  - Sub-seasonal, Seasonal, Interannual and Multi-decadal

- Modeling Extremes in Space-time is crucial for effective planning management of natural resources



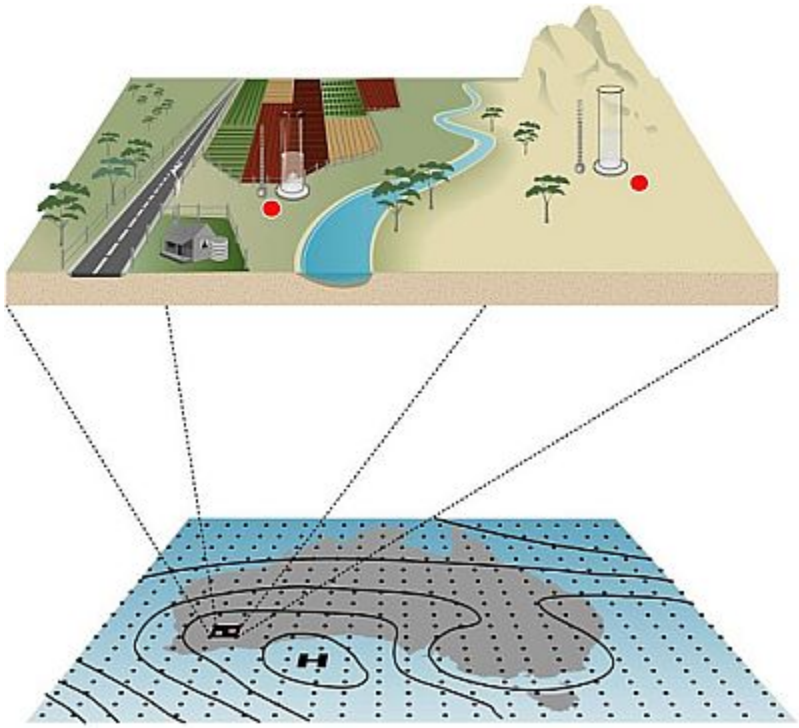


Methods/Applications Suite

Understanding how extremes are enabled

- Moisture sources and pathways
- Clustering of Extremes

Application: Western US



Tools for modeling extremes at multiple time scales – Downscaling

- Downscaling precipitation extremes –
  - QR, BMA
- Stochastic Weather Generator

Application: Upper Colorado River Basin

Space-Time Modeling of extremes and Multivariate Extremes

- Bayesian Hierarchical Modeling

Application: Upper Colorado River Basin (Taylor Park Dam)

Extremes
Clusters
Moisture Sources/Pathways

Bracken et al., 2015, JGR

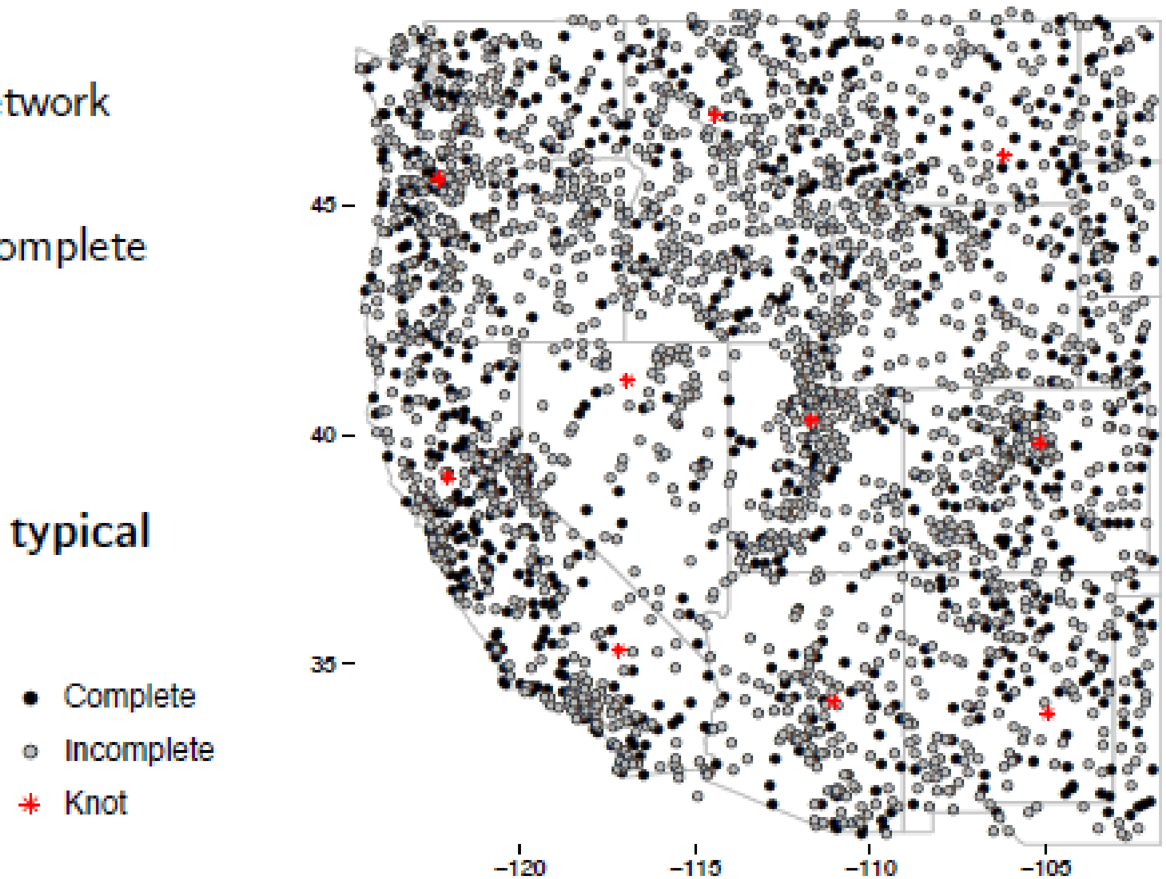
Data for Modeling

Precipitation Data

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

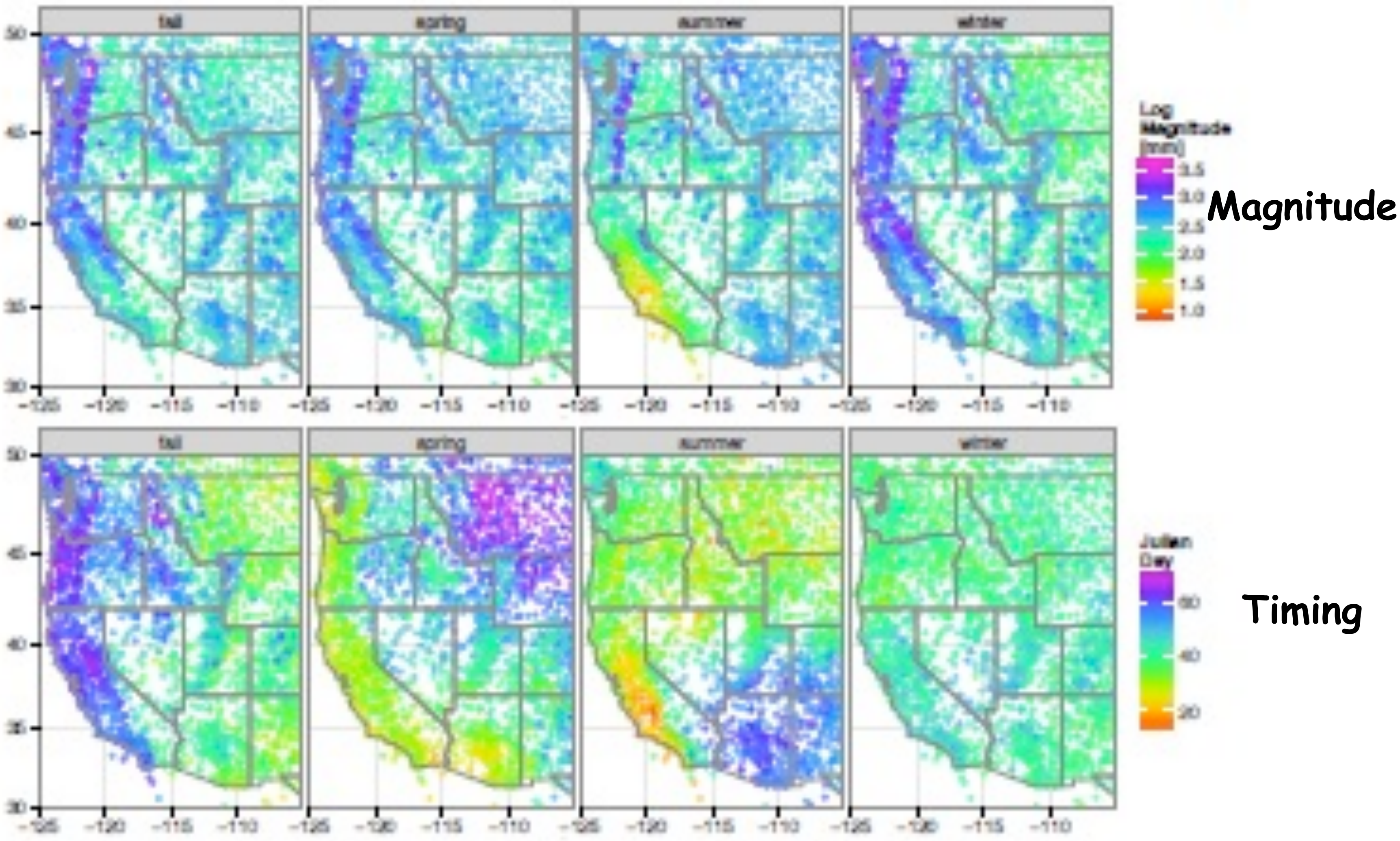
- ~2500 stations with near complete data from 1948-2013
- 3 day aggregation window
- Fall maxima

Very large region/dataset for typical Bayesian spatial model



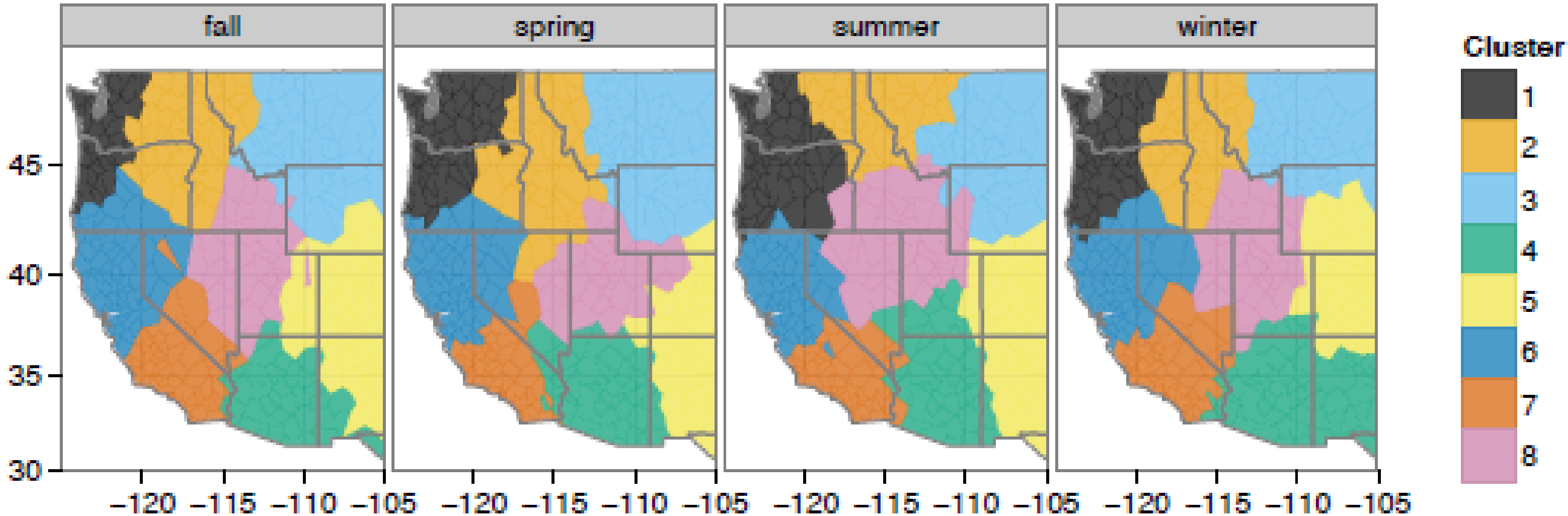
Bracken et al., 2015, JGR

Extremes – Magnitude and Timing



Bracken et al., 2015, JGR

Extremes – Magnitude and Timing Clusters

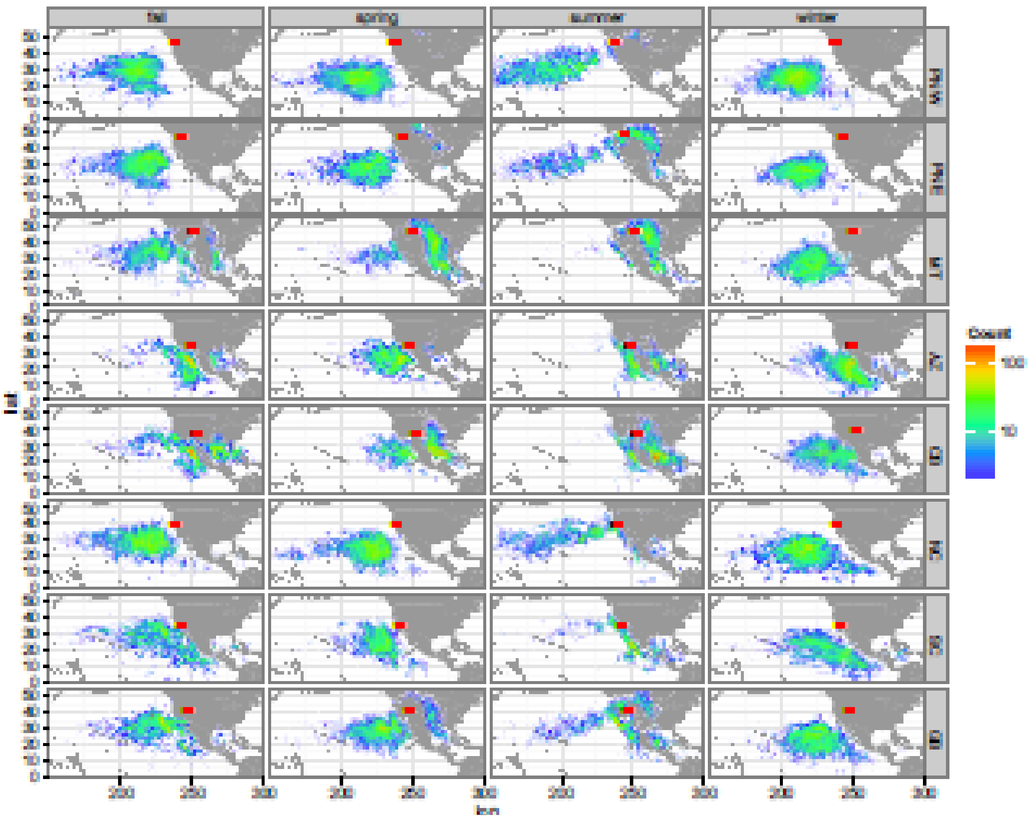


Magnitude Cluster

- **Consistent with topography and seasonal climatology**
  - **Winter precipitation/snow and Summer monsoonal Rainfall in SouthWestern US**

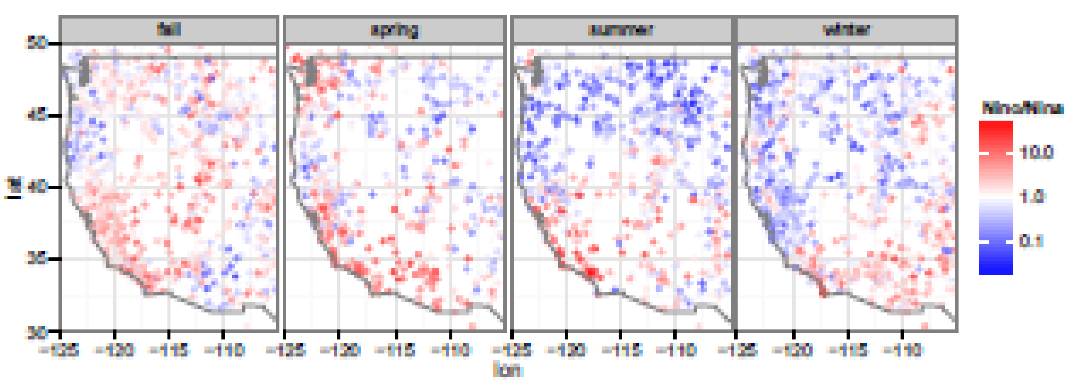
Bracken et al., 2015, JGR

Extremes - Moisture sources and Pathways



- North Central Pacific an Important source
  - Reminiscent of Atmospheric Rivers (ARs) in winter
- Land source important During summer for inland regions

- ~1000 stations with near Complete data 1948 – 2013
- 3-day rainfall maximum for each year and each season
- HYSPLIT trajectories.



Ratio of number of rain trajectories during La Nina vs l Nino

Red – More trajectories during l Nino

Bracken et al., 2015, JGR

Downscaling Extremes
(Post Processing Dynamical Model Output)

Mendoza et al., 2016, MWR

Motivation

- Dynamical model Forecasts – NWP, Seasonal and Multi-decadal Forecasts – are on Spatial grid and **far from being perfect**

- Information and Decisions are made at point or regional scale
- Need for *Downscaling/Postprocessing*

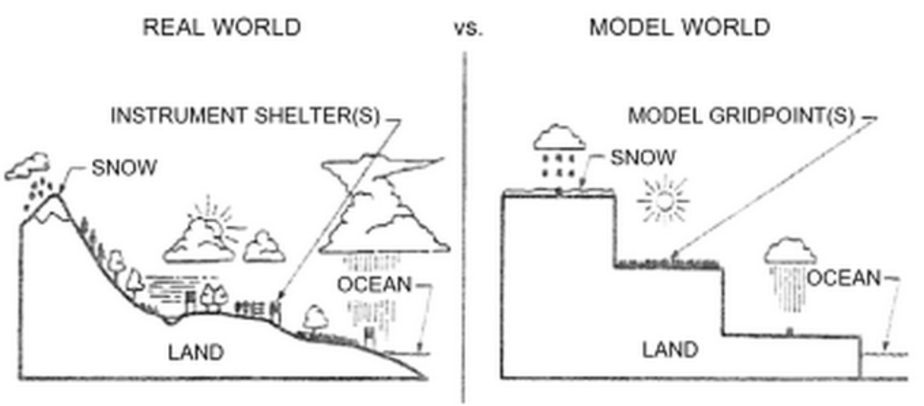
- Why would we want a **statistical reinterpretation** of **dynamical model outputs**? (Wilks, 2011)

- There are several techniques for post-processing Extremes from Dynamical Models:

i. Multinomial Logistic Regression.

ii. Quantile Regression.

iii. Bayesian Model Averaging.

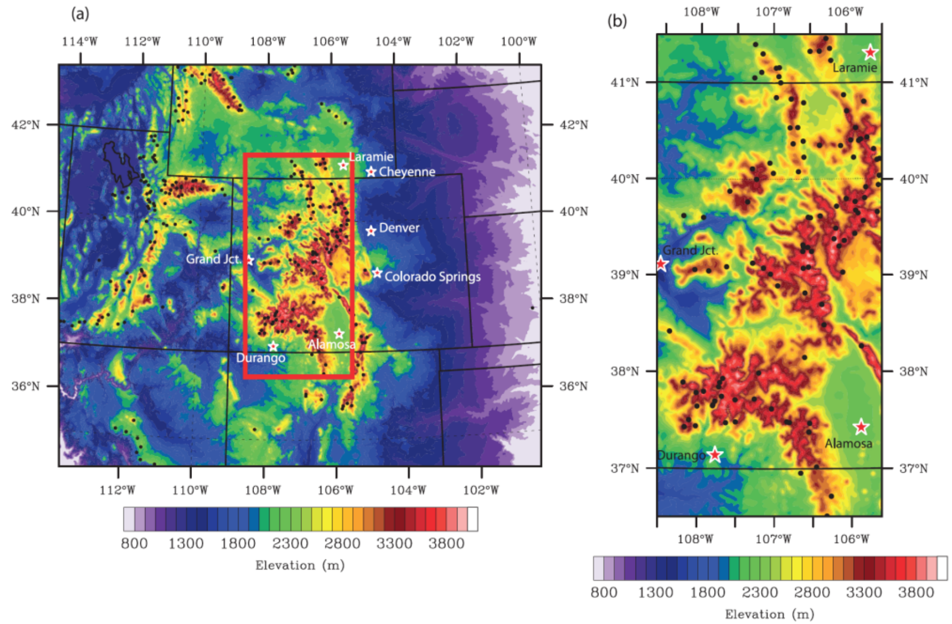


From Karl et al. (1989)

Study area & data

Site

- The Colorado Headwaters Region offers a major renewable water supply in the southwestern US, with approximately 85 % of the streamflow coming from snowmelt.



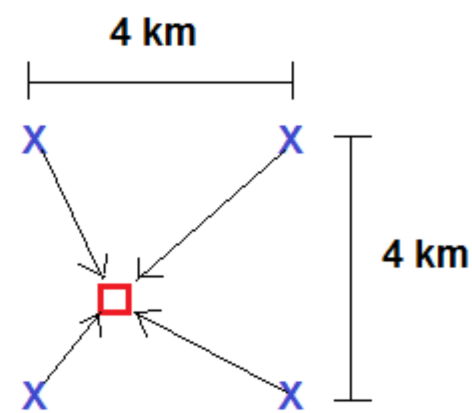
Colorado Headwaters Region

Data

•Daily outputs from the WRF-4km reanalysis (**predictors**): precipitation, air temperature, air pressure, specific humidity and wind speed

•Verification data (**predictand**): precipitation at 93 SNOTEL sites for eight water years (October 1st, 2000 – September 30th, 2008).

Predictors (WRF)
Predictand (SNOTEL)



Approach

QR:

$$q_\theta(y \mid \mathbf{x}) = \mathbf{x}^{\mathrm{T}} \boldsymbol{\beta}_\theta ,$$

$$\hat{\boldsymbol{\beta}}_\theta = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \rho_\theta (y_i - \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}).$$

MnLR:

for $j = 1, \ldots, J$, where

$$\eta_j = \log\left(\frac{\pi_j}{\pi_J}\right) = \alpha_j + \mathbf{x}^{\mathrm{T}} \boldsymbol{\beta}_j ,$$

Step 1

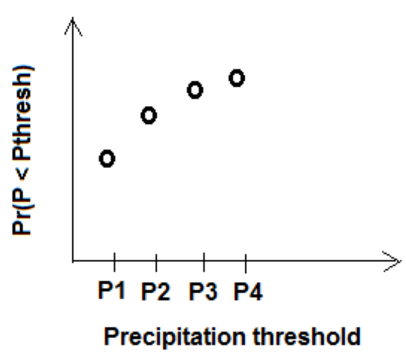
Select the four nearest neighbors from the reanalysis grid as potential predictors (**MnLR** and **QR**) or as ensemble members (**BMA**).

Step 2

Perform experiments at specific SNOTEL sites to find suitable power transformation of the form $x_p = x^{\eta}$

Step 3

Fit models at all sites (93), and estimate the cumulative distribution function (cdf) at each station and time step.



Step 4

Generate ensembles of precipitation (Nens = 100) by sampling randomly from the cdfs generated in the previous step.

Step 5

Perform probabilistic verification over the period Oct/2000 – Sep/2008:

- -Brier Skill Score.
- -Reliability diagrams.
- -Discrimination diagrams.
- -Rank histograms.

Bayesian Model Averaging, BMA (Sloughter et al., 2007)

- The predictive pdf is a mixture of a discrete component at zero precipitation and a Gamma distribution

- There are 2 steps:
  1. Estimate PoP as a function of the forecasts *fk*
  2. Specify the PDF of the amount of precipitation given that it is not zero.

Estimation of pdf

$$p(y \mid f_1, \ldots, f_K) = \sum_{i=1}^{K} w_i h_i(y \mid f_i) \qquad \sum_{i=1}^{K} w_i = 1$$

Estimation of hi

$$h_i(y \mid f_i) = P(y=0 \mid f_i) I[y=0] + P(y>0 \mid f_i) g_i(y \mid f_i) I[y>0]$$

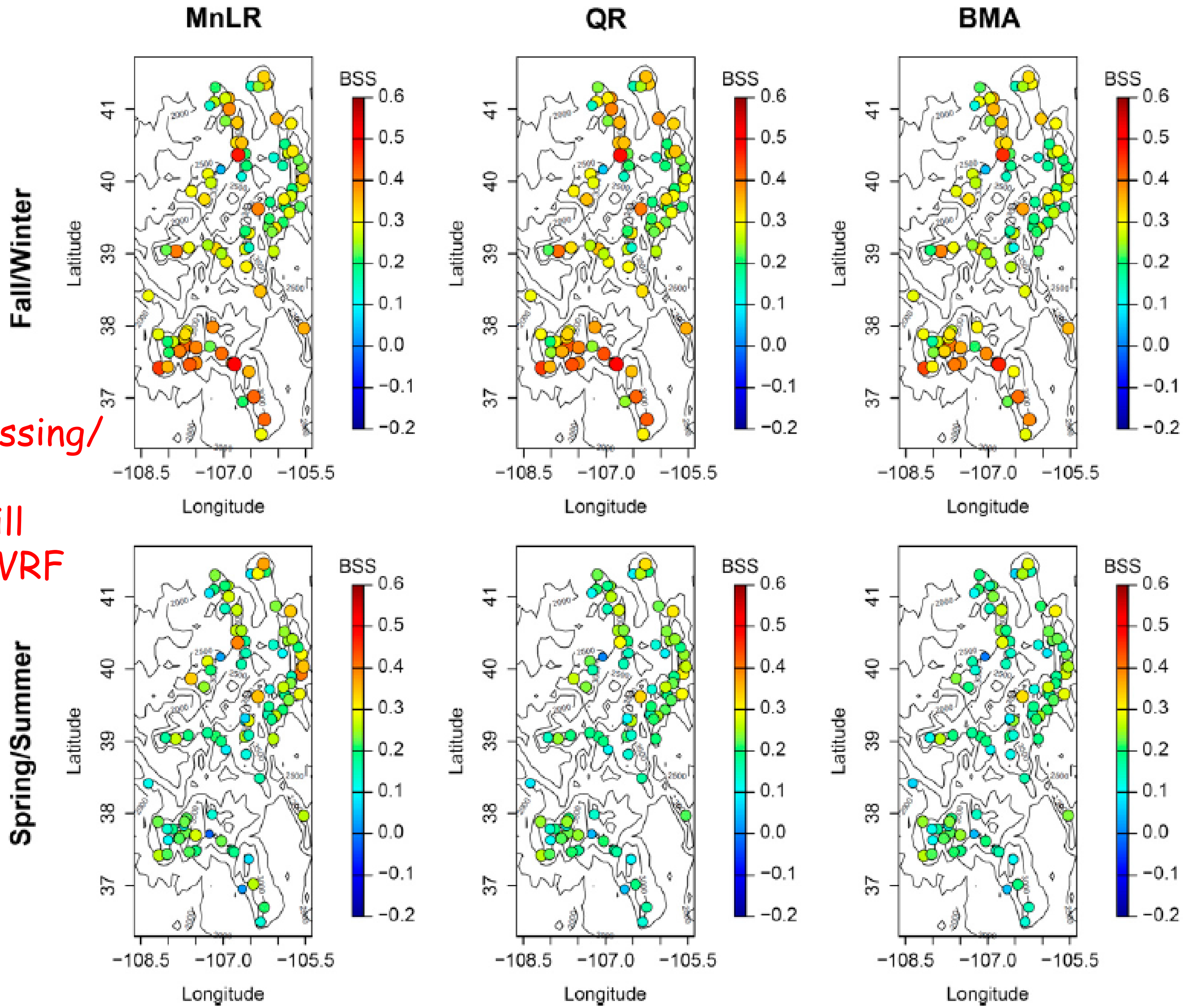
gi has the shape of a Gamma function

$$g_i(y \mid f_i) = \frac{1}{\beta_i^{\alpha_k} \Gamma(\alpha_i)} y^{\alpha_i - 1} \exp(-y / \beta_i)$$

$$\mu_i = b_{0i} + b_{1i} f_i^{1/3}$$

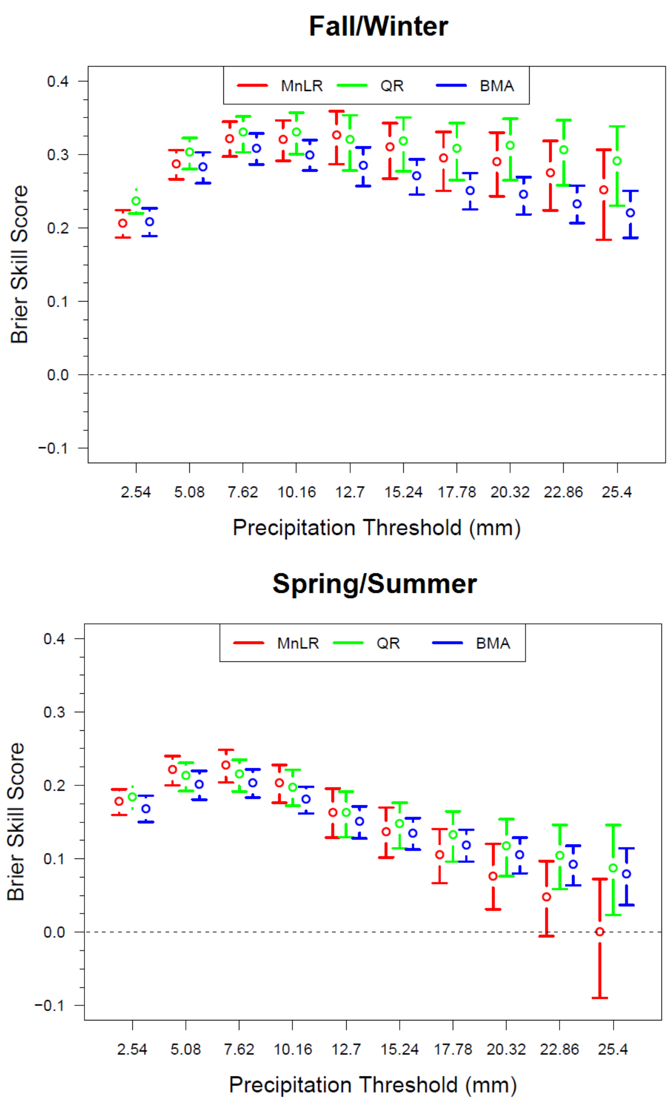
$$\sigma_i^2 = c_{0i} + c_{1i} f_i$$

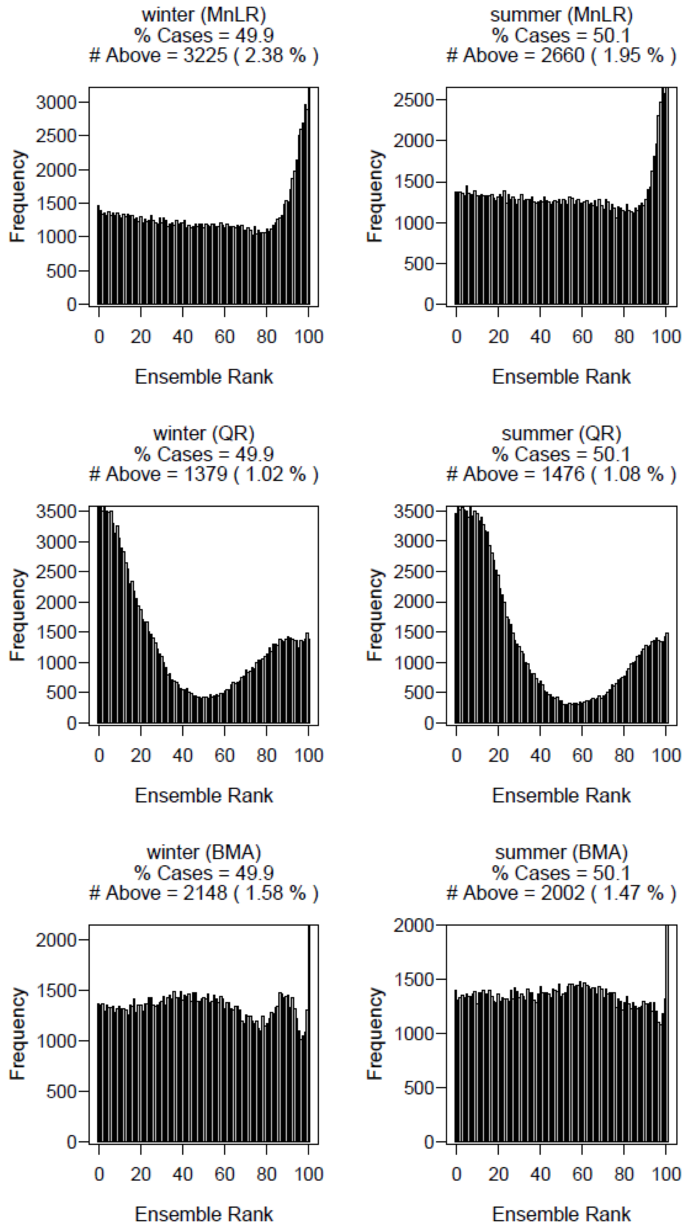
Results



- Postprocessing/ Downscaling Improves skill Relative to WRF

Results





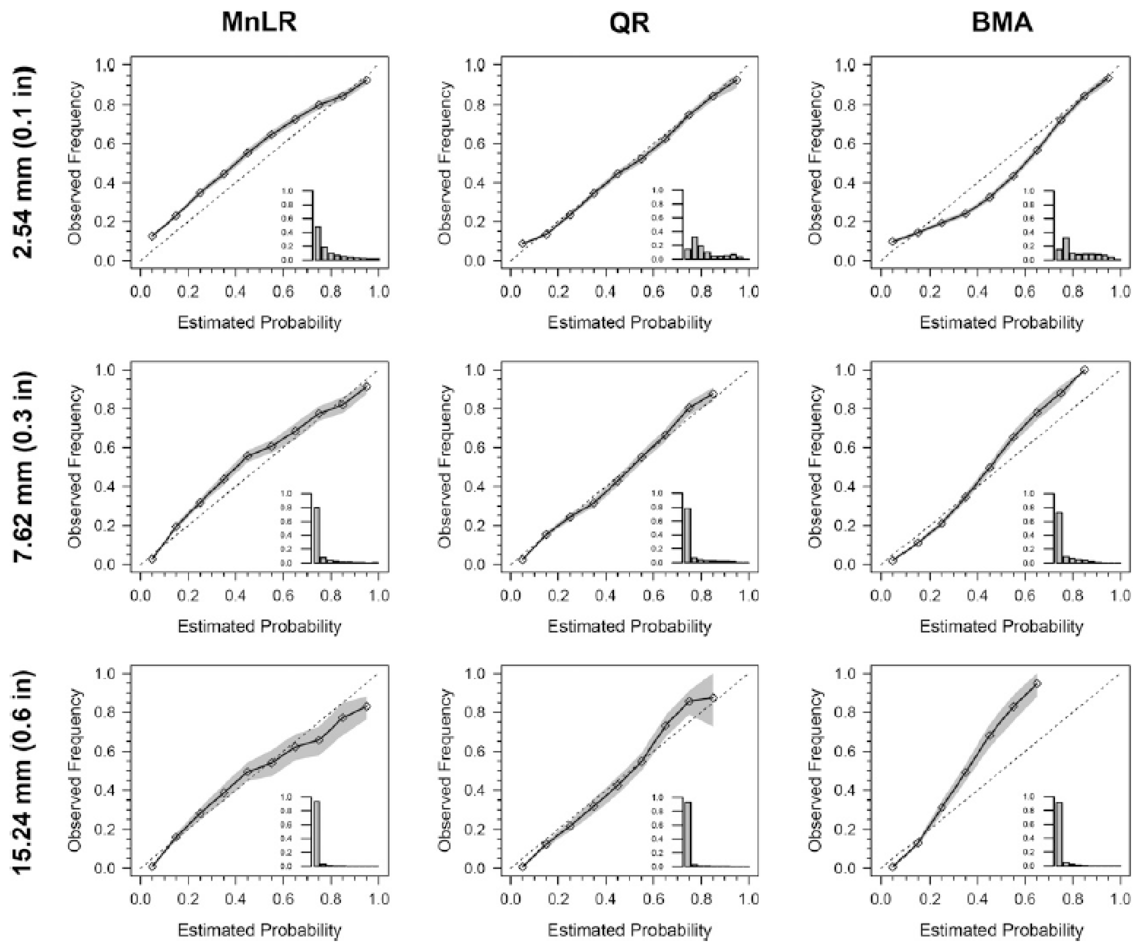
Brier skill scores (BSS)

Rank histograms

Results

- WRF has skill in Fall/Winter
  - Spatial skill of post-processing consistent with spatial skill of WRF

- MnLR shows poor performance
- QR best for skill and reliability
- BMA is best for discrimination, statistical consistency and robust estimates of uncertainty



Downscaling Using Stochastic Weather Generators
(Post Processing Dynamical Model Output)

Verdin et al., 2016, J. Hydrology

Generalized Linear Model (GLM) Based Weather Generator

GLMs can model a variety of distributions of the response variable Y

- Skewed distribution (e.g., Gamma, Weibull)
- Discrete/Binary (e.g., Binomial, Poisson)
- Mean of the distribution – i.e., E(Y | X1, X2 , ..., Xp) linearly related to Covariates

Precipitation occurrence

$$O(\mathbf{s},t) = \mathbb{1}_{[W_O(\mathbf{s},t)\geq 0]}$$

$$W_O(\mathbf{s},t) \sim \mathrm{GP}(\mathbf{X}_O(\mathbf{s},t)^{\mathrm{T}}\beta_O(\mathbf{s}), C_O)$$

$$\mathbf{X}_O(s,t) = (1, O(s,t-1), \cos(2\pi t/365), \sin(2\pi t/365$$

Precipitation amount

$$A(\mathbf{s},t) \sim \mathrm{Gamma}(\alpha_A(\mathbf{s}), \alpha_A(\mathbf{s})/\mu_A(\mathbf{s},t))$$

$$\mu_A(\mathbf{s},t) = \exp(\mathbf{X}_A(\mathbf{s},t)^{\mathrm{T}}\beta_A)$$

Temperature

$$Z_N(\mathbf{s},t) = \mathbf{X}_N(\mathbf{s},t)^{\mathrm{T}}\beta_N(\mathbf{s}) + W_N(\mathbf{s},t)$$

$$Z_X(\mathbf{s},t) = \mathbf{X}_X(\mathbf{s},t)^{\mathrm{T}}\beta_X(\mathbf{s}) + W_X(\mathbf{s},t)$$

$$\mathbf{X}_i(s,t) = (1, Z_N(s,t-1), Z_X(s,t-1), \cos(2\pi t/365), \sin(2\pi t/365), r(t), O(s,t)) \qquad \text{for } i = N, X$$

- Fit GLMs at each location
- Maximum Likelihood Estimation of Parameters
- Spatial models on βs to enable simulation

At any location

Verdin et al. (2015); Kleiber et al. (2012,2013)

Conditional Generation
Seasonal and Multidecadal

- Additional Covariates

Precipitation occurrence

$$O(\mathbf{s},t) = \mathbb{1}_{[W_O(\mathbf{s},t)\geq 0]}$$
$$W_O(\mathbf{s},t) \sim \mathrm{GP}(\mathbf{X}_O(\mathbf{s},t)^{\mathrm{T}}\boldsymbol{\beta}_O(\mathbf{s}), C_O)$$

Precipitation amount

$$A(\mathbf{s},t) \sim \mathrm{Gamma}(\alpha_A(\mathbf{s}), \alpha_A(\mathbf{s})/\mu_A(\mathbf{s}, t$$
$$\mu_A(\mathbf{s},t) = \exp(\mathbf{X}_A(\mathbf{s},t)^{\mathrm{T}}\boldsymbol{\beta}_A)$$

Temperature

$$Z_N(\mathbf{s},t) = \mathbf{X}_N(\mathbf{s},t)^{\mathrm{T}}\boldsymbol{\beta}_N(\mathbf{s}) + W_N(\mathbf{s},t)$$
$$Z_X(\mathbf{s},t) = \mathbf{X}_X(\mathbf{s},t)^{\mathrm{T}}\boldsymbol{\beta}_X(\mathbf{s}) + W_X(\mathbf{s},t)$$

$$\mathbf{X}_O(s,t) = (1,\ldots,ST1(t),ST2(t),ST3(t),ST4(t))$$

$$\mathbf{X}_A(s,t) = (1,\ldots,ST1(t),ST2(t),ST3(t),ST4(t))$$

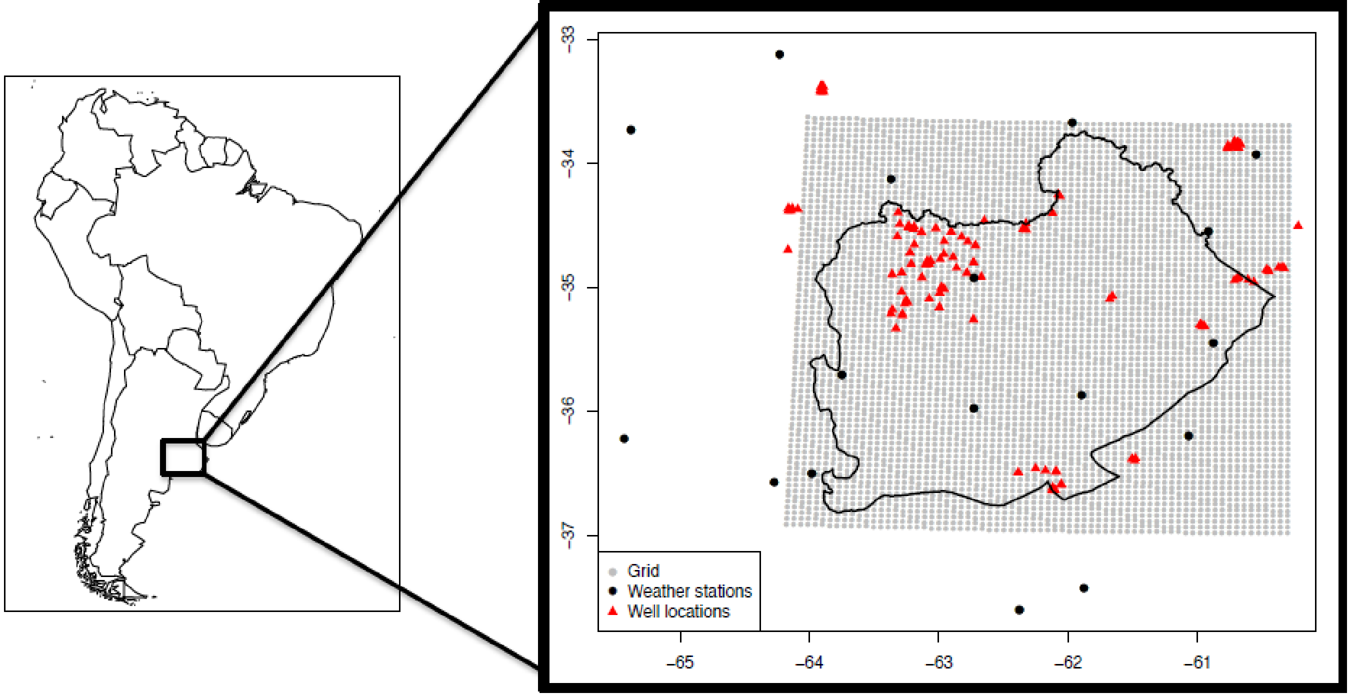
$$\mathbf{X}_i(s,t) = (1,\ldots,SMN1(t),SMN2(t),SMN3(t),SMN4(t), SMX1(t),SMX2(t),SMX3(t),SMX4(t)) \quad \text{for } i = N, X$$

- Domain averaged seasonal total precipitation, ST1; ST2; ST3; ST4 – for the four seasons
- Domain averaged seasonal max and min temperatures – SMN and SMX..
- Other covariates can be added as needed – e.g., climate indices – ENSO, PDO etc.

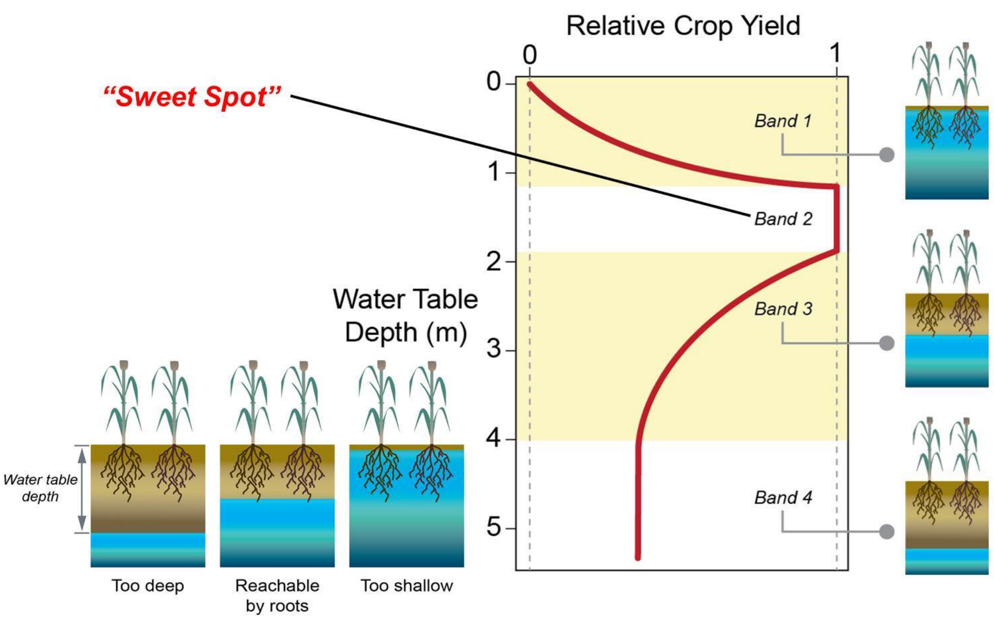
Verdin et al. (2016)

Application
Agriculture Management
Crop Modeling
Seasonal and Multidecadal
La Pampa ~ Argentina

Verdin et al., 2014 and 2016



- Crop Simulation model (DSSAT)
- Crop yields with Water Table Depth (MIKE-SHE)
- Stochastic daily weather on a 5km x 5km grid
- Ensemble of WTD and crop yields
- Agriculture planning and management

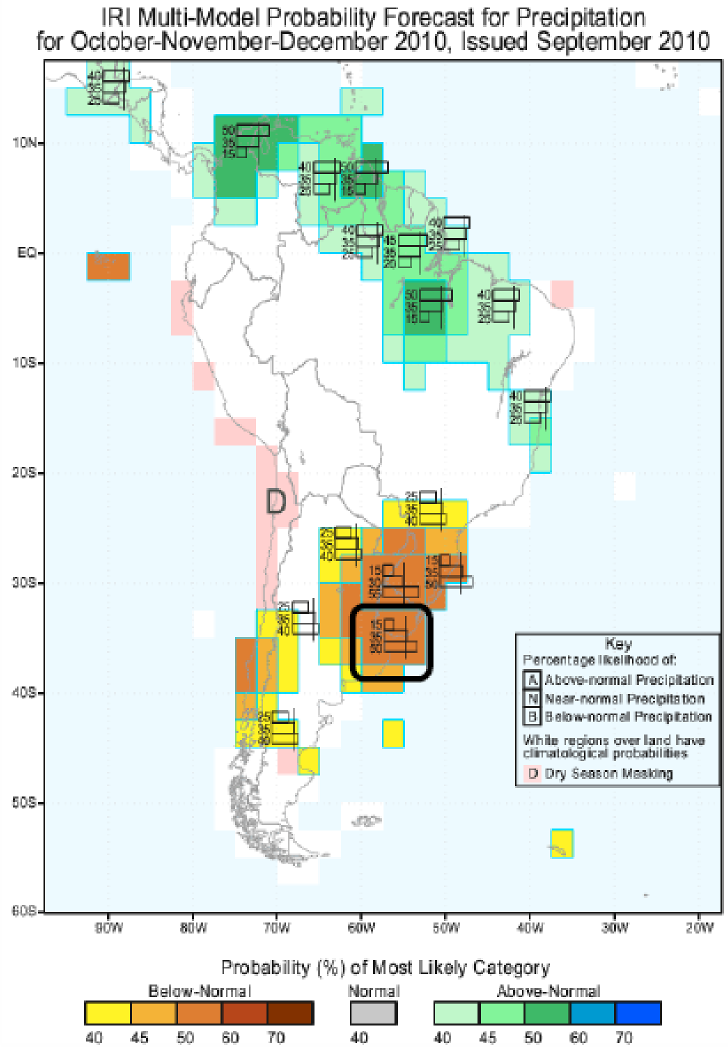


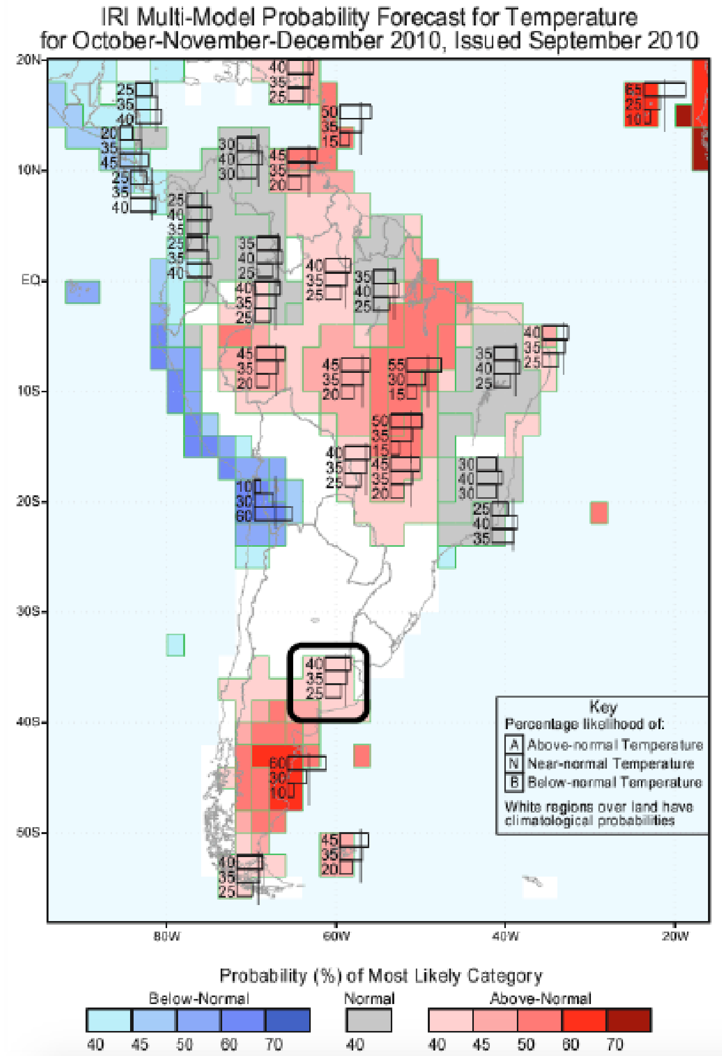
- 1961-2013 daily weather data

Verdin et al. (2016)

courtesy: Guillermo Podestá

Seasonal Simulation – OND 2010 – Dry year





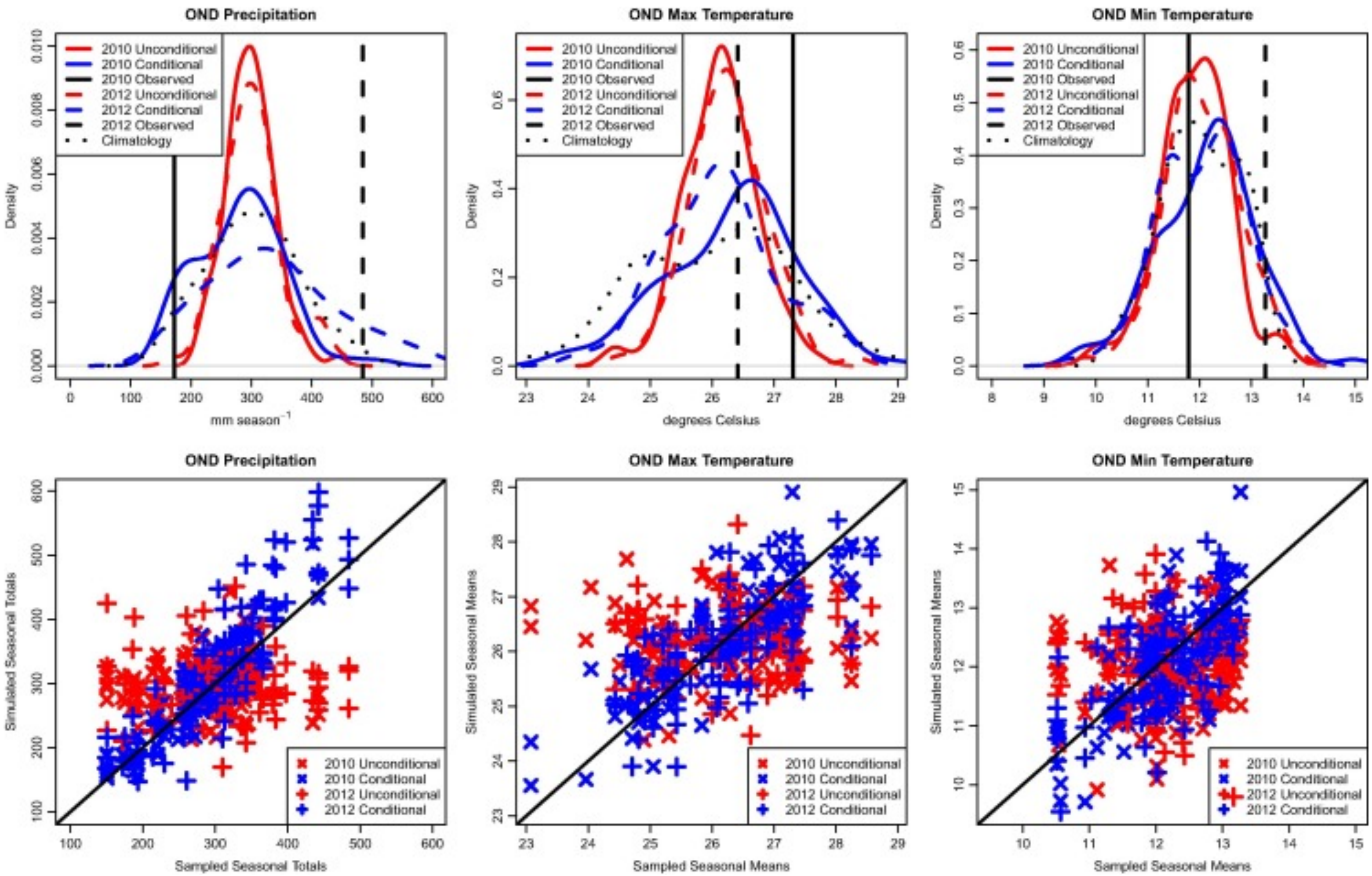
OND 2010 Precip.
15:35:50 – A:N:B

OND 2010 Temp.
40:35:25 – A:N:B

- 2012 Wet Year

- Re-sample ensemble of climatology OND season Precip/temp with A:N:B as weights
- Generate daily weather Ensemble with the above Covariates
- Ensemble of weather Reflects uncertainty

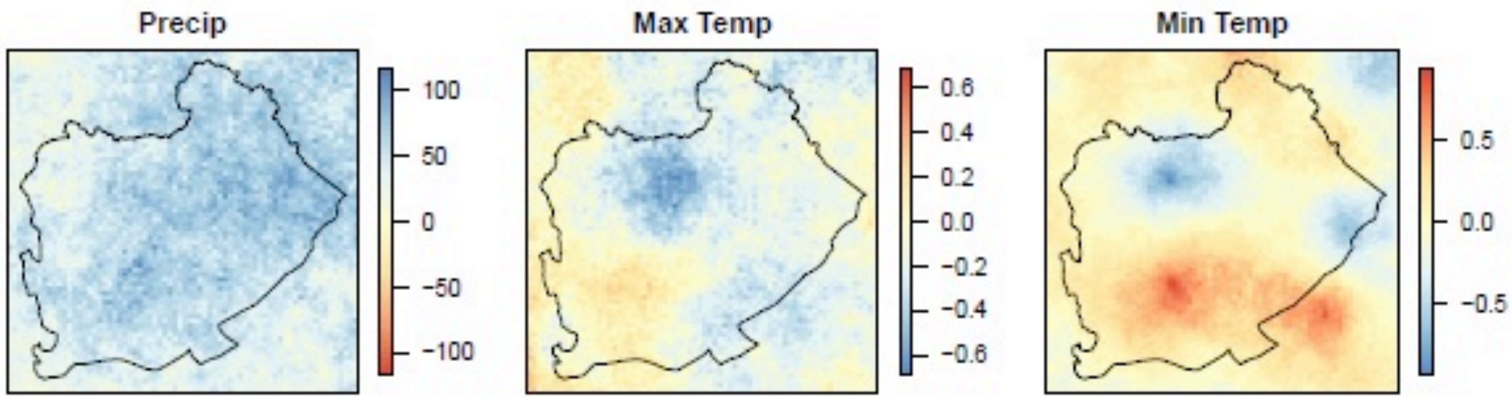
Seasonal Simulation – Conditioned on Climate Forecast



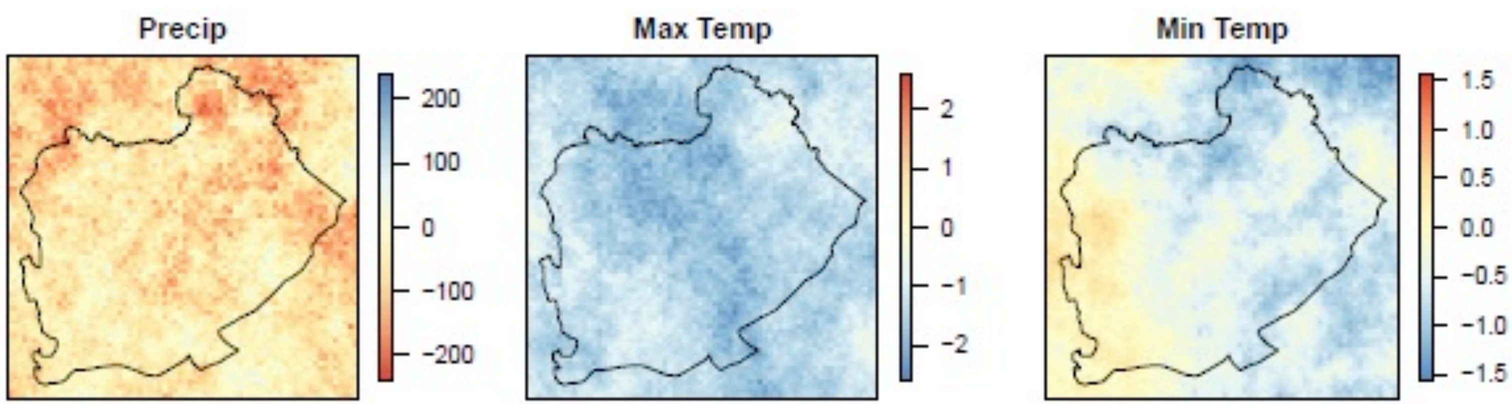
- **Conditional simulation captures the observed variability quite well**
- **Unconditional reproduces climatology**

Seasonal Simulation - Conditioned on Climate Forecast

Differences in ensemble mean (unconditional minus conditional):

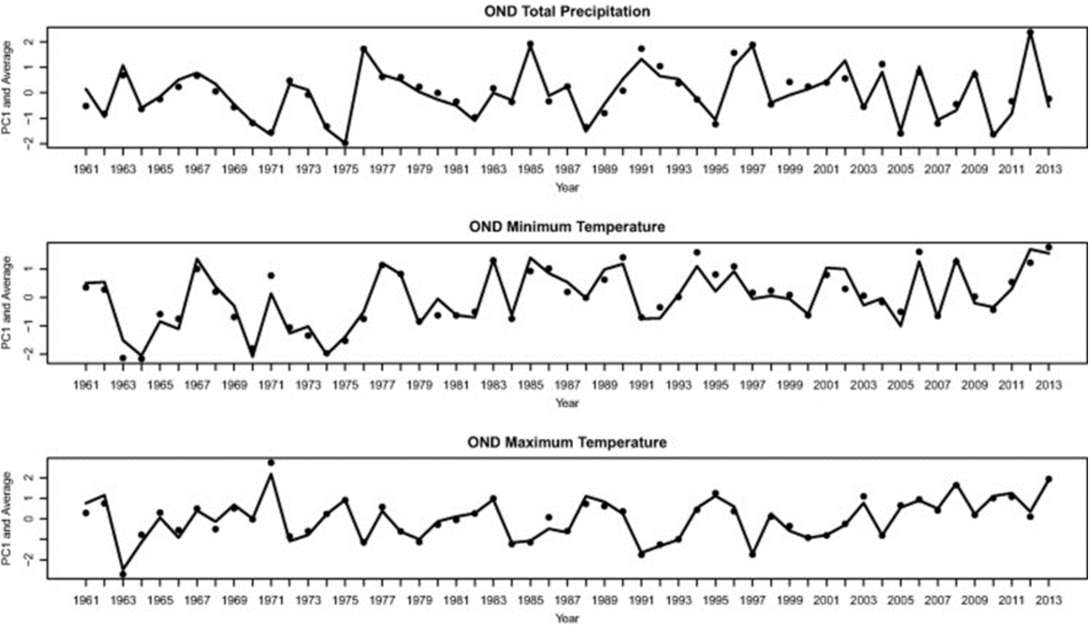


Differences in 95% ensemble spread (unconditional minus conditional):



- **Unconditional simulation shows**
  - ***A wet bias relative to conditional***
  - ***A cool bias in Max temperature***

Multi-decadal Simulation

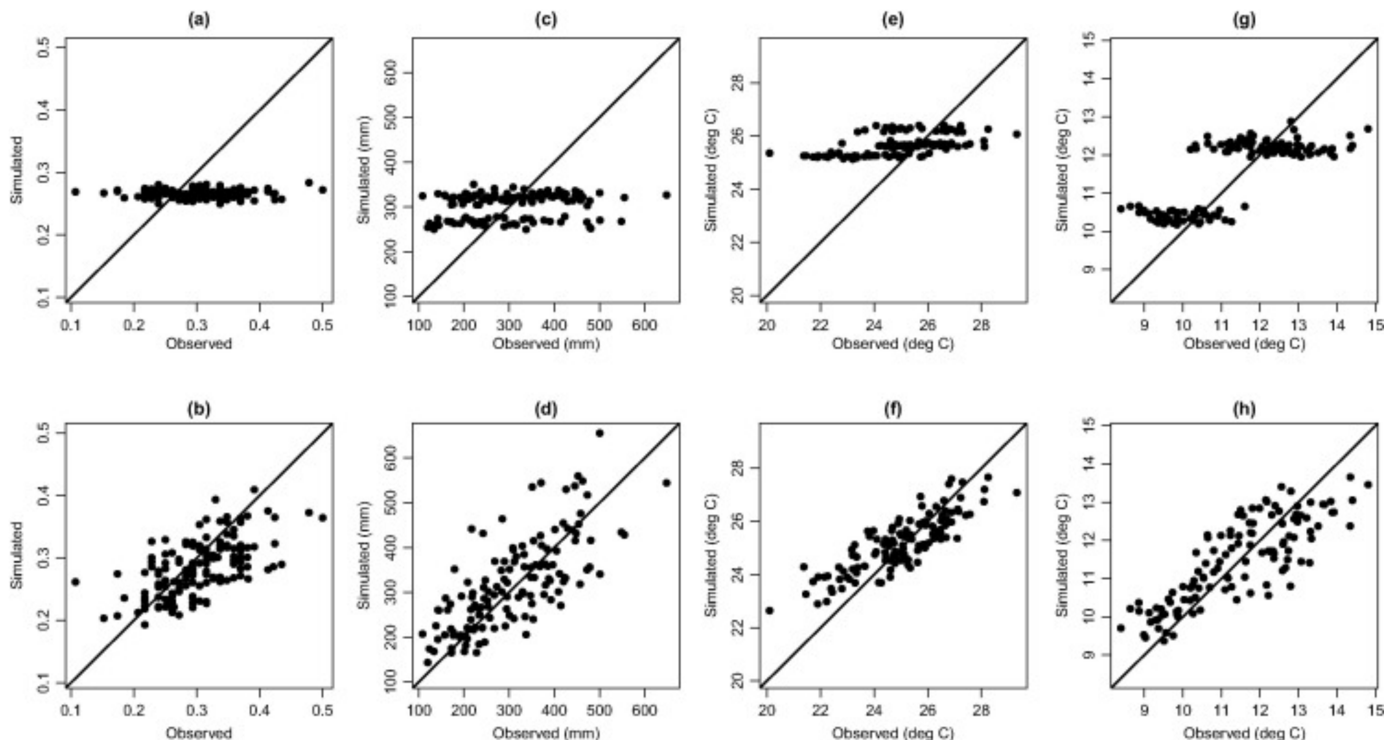


- Seasonal Precipitation, Max Temp And Min Temp, 1961 - 2013

- Unconditional Simulations

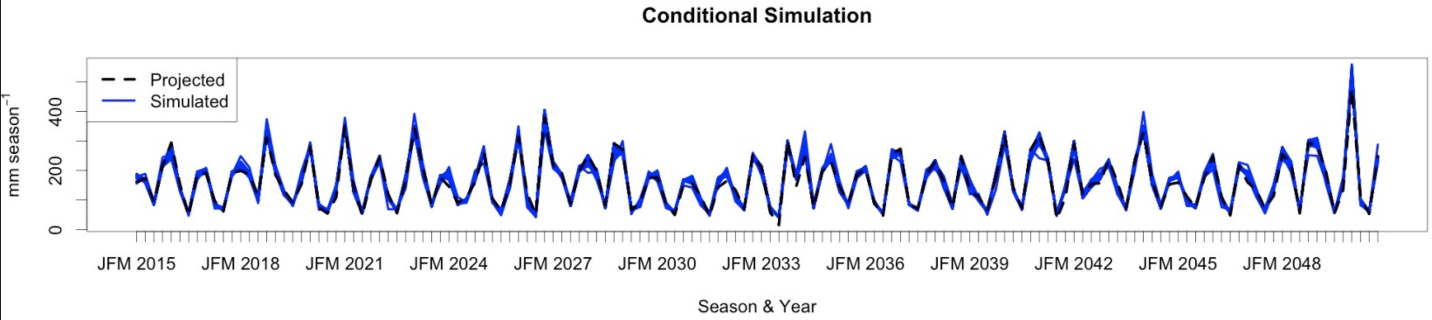
(a) Precip. Occurr.
(b) Precip. Amounts
(c) Max Temp.
(d) Min Temp

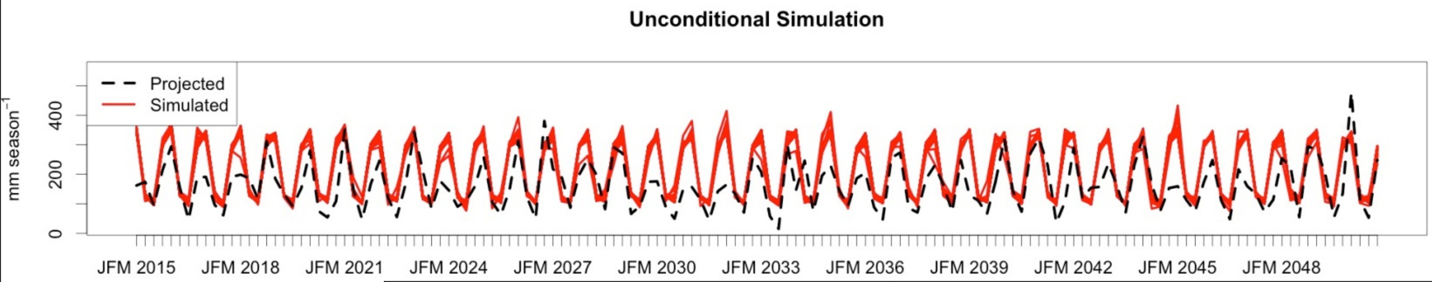
- Conditional Simulations



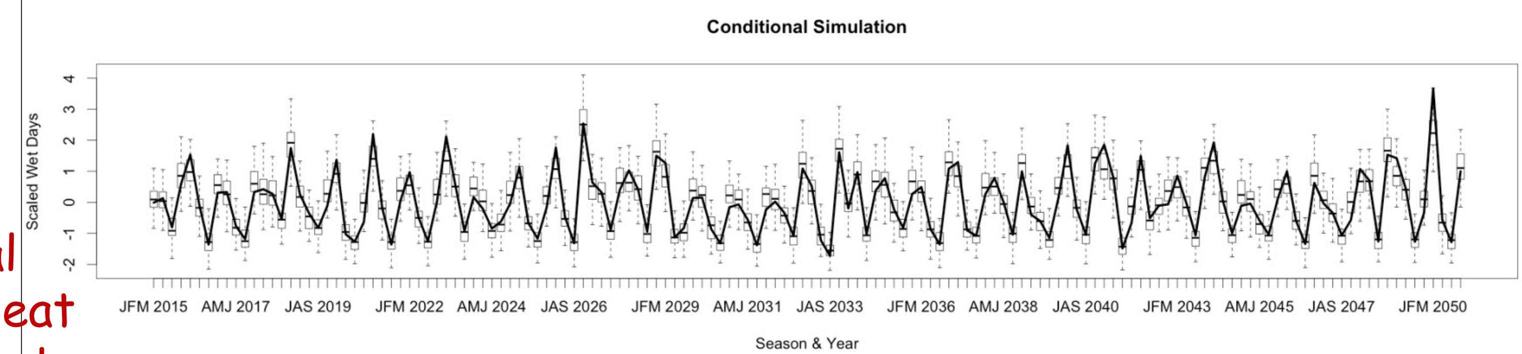
Simulations for the Period 2015 – 2050

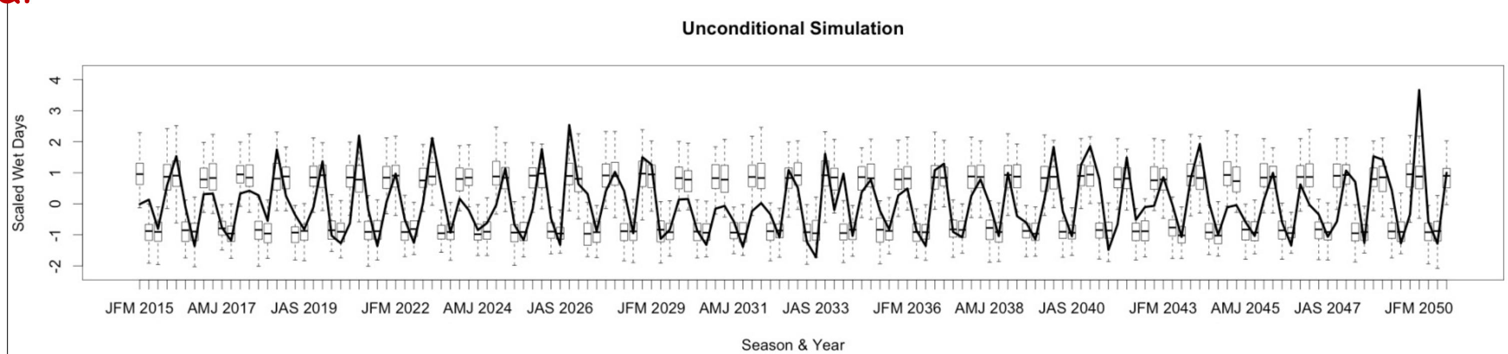
- Precipitation





- Wet Days





- Unconditional Simulations repeat the climatological Cycle

- Conditional Simulations are consistent with the projections

Space-Time Modeling of Extremes
Bayesian Hierarchical Model

Bracken et al., 2016, WRR

Bayesian and Extreme Values

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:

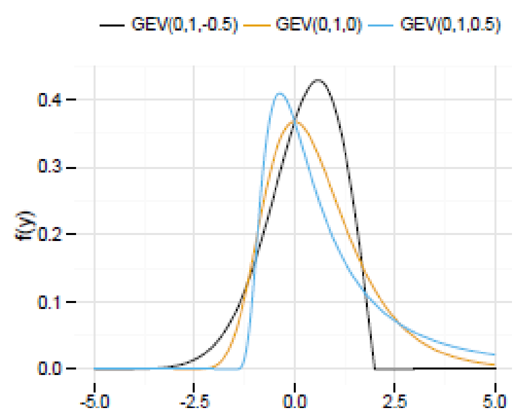
$$\text{GEV}(y;\mu,\sigma,\xi) = \frac{1}{\sigma} b^{(-1/\xi)-1} \exp\left\{-b^{-1/\xi}\right\}$$

$b = 1 + \xi\left(\frac{x-\mu}{\sigma}\right)$, 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).



Can model

- **Sea Level**
- **Precipitation**
- **Streamflow**

Bracken et al., 2016, WRR

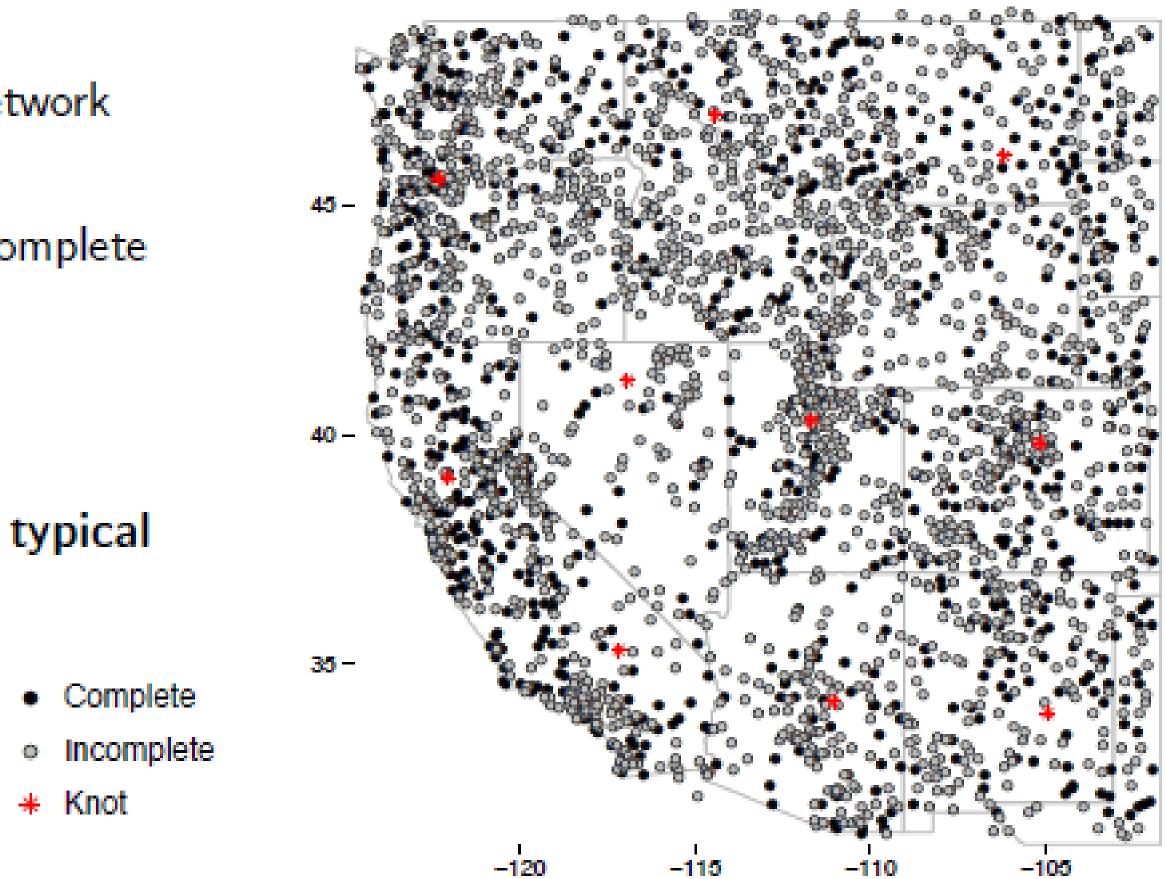
Data for Modeling

Precipitation Data

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

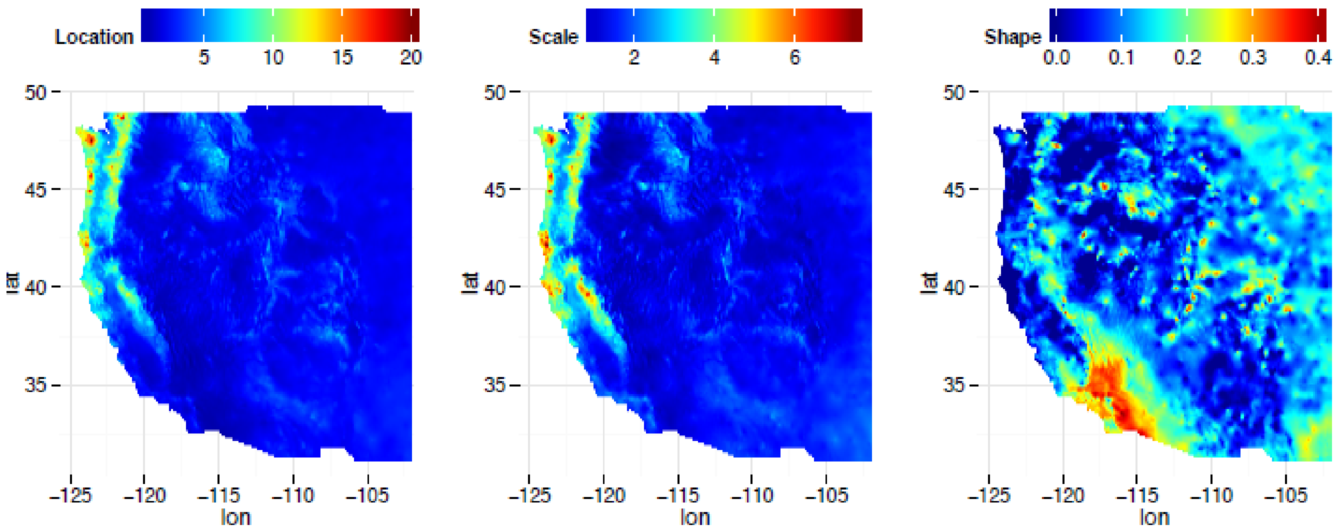
- ∼2500 stations with near complete data from 1948-2013
- 3 day aggregation window
- Fall maxima

Very large region/dataset for typical Bayesian spatial model



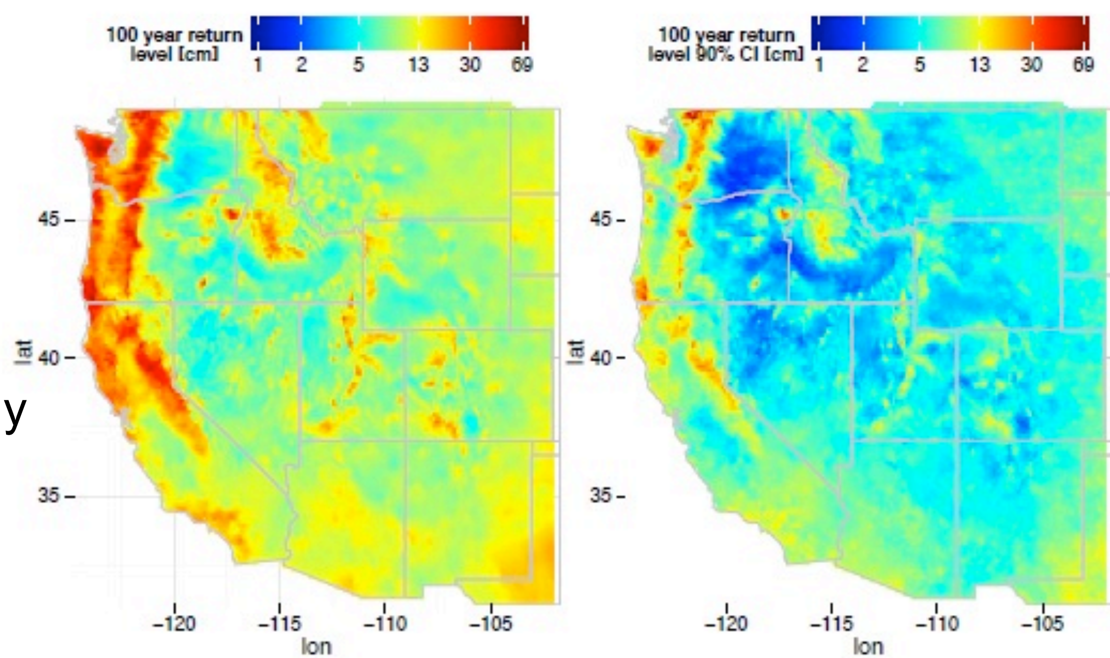
Bracken et al., 2016, WRR

Results – Median GEV parameters, fall



- **Model (GEV) parameters obtained at any location**

Results – Median 100 year return levels, fall



- **Return levels with confidence intervals**
- **Consistent with topography**

Bracken et al., 2016, WRR

Multivariate Nonstationary Extremes

- Precipitation
- Flow
- Reservoir Level

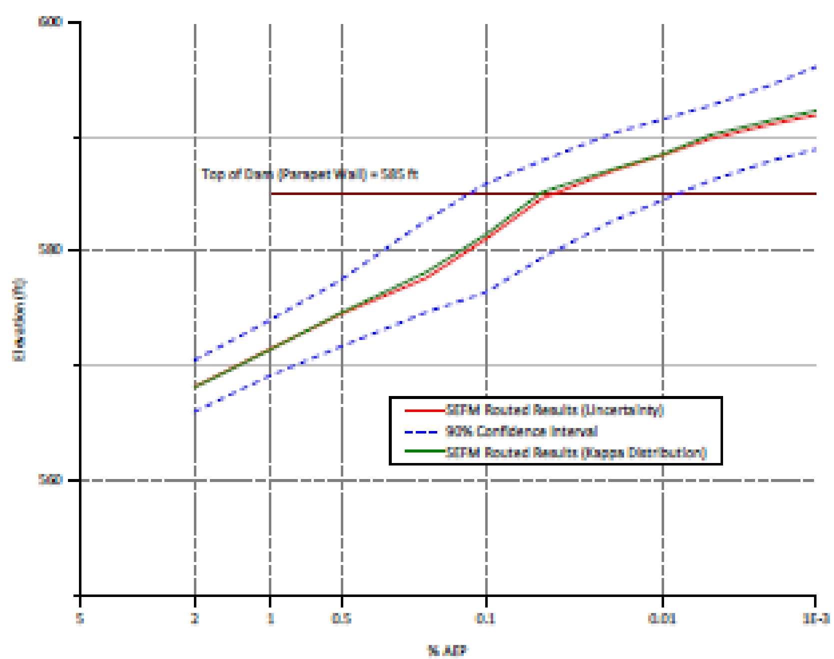
Bracken et al., 2017, in review WRR

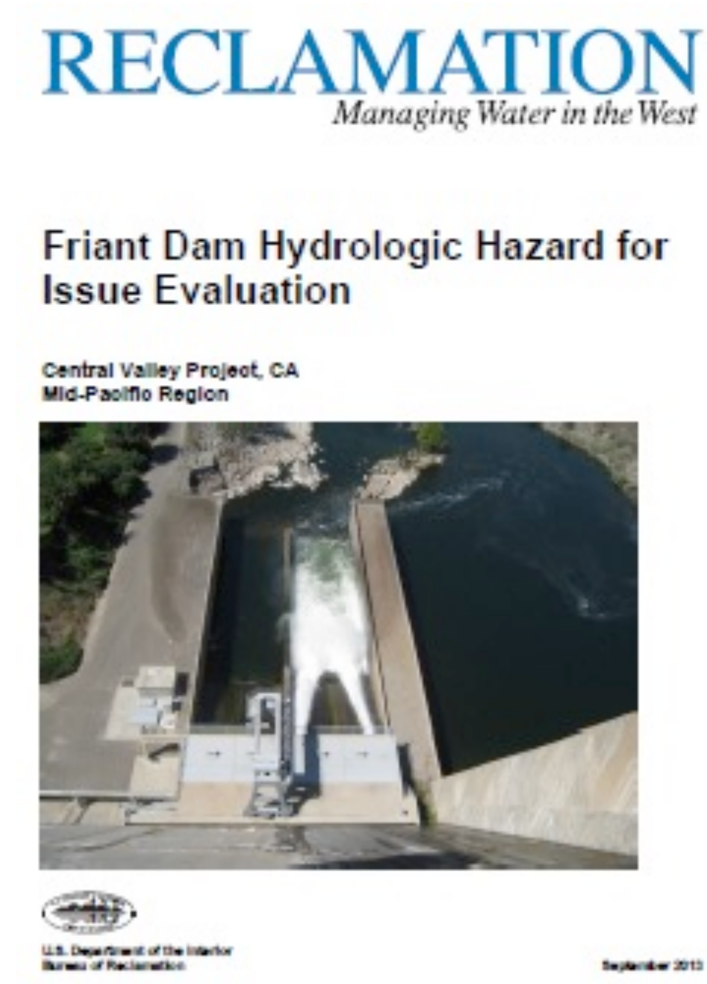
Motivation

- Frequency curves for precipitation, flow and reservoir elevation are Estimated independently, making uncertainty propagation difficult

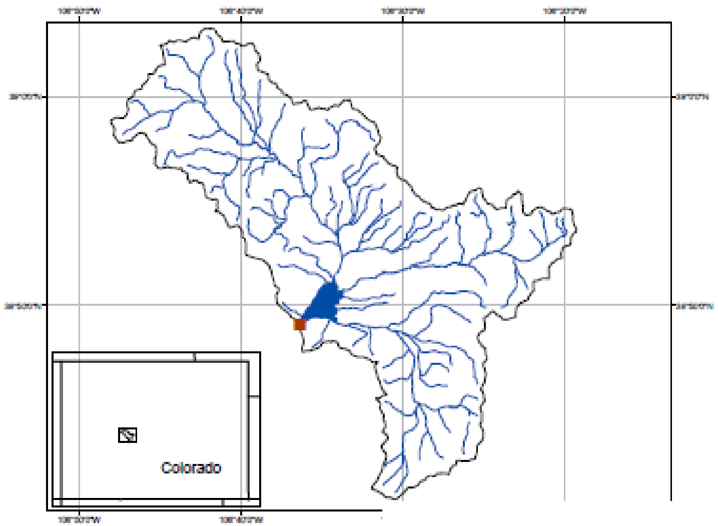
- Return levels are developed under assumption of temporal stationarity

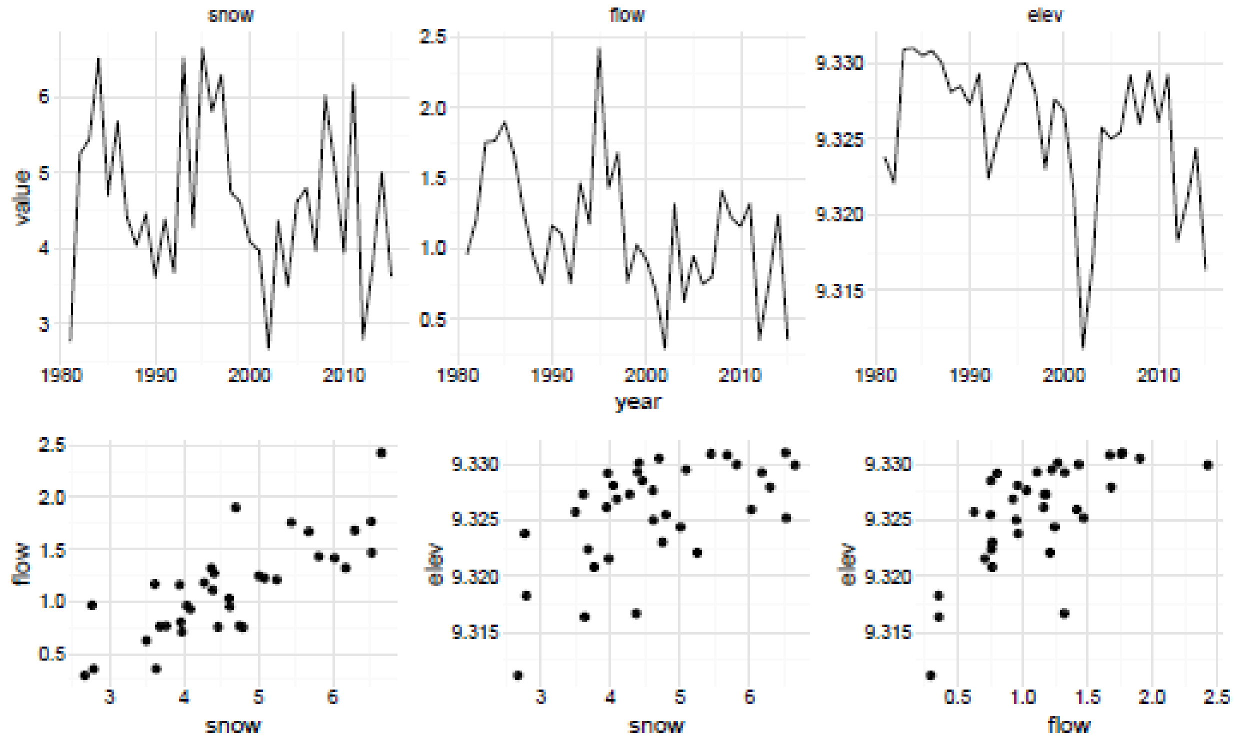
- Need for modeling extremes with temporal nonstationarity





Application – Taylor Park Dam, Colorado





Bracken et al., 2017, in review WRR

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:

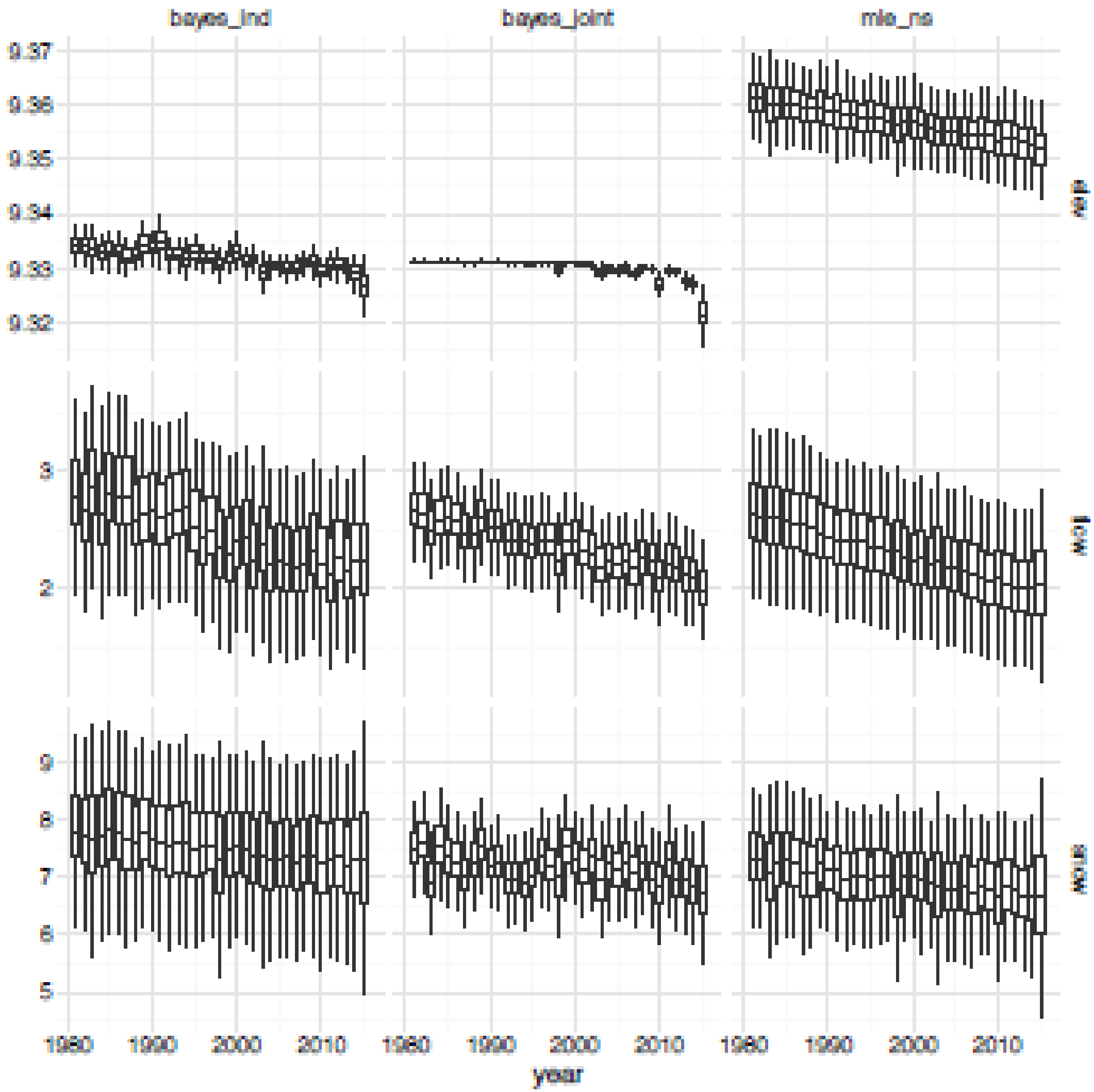
Model
- Incorporate temporal nonstationarity

$$
\begin{aligned}
(y(t), z(t), h(t)) &\sim C_g(\Sigma; \{\mu_y(t), \sigma_y, \xi_y, \mu_z(t), \sigma_z, \xi_z, \mu_h(t), \sigma_h, \xi_h\}) \\
y(t) &\sim GEV(\mu_y(t), \sigma_y, \xi_y) \\
z(t) &\sim GEV(\mu_z(t), \sigma_z, \xi_z) \\
h(t) &\sim GEV(\mu_h(t), \sigma_h, \xi_h) \\
\mu_y(t) &= \mu_{y0} + x(t)^T \beta_y \\
\mu_z(t) &= \mu_{z0} + x(t)^T \beta_z \\
\mu_h(t) &= a - \exp(-b\mu_z(t))
\end{aligned}
$$

Bracken et al., 2017, in review WRR

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

Results – Nonstationary 100-year Return Level



Reservoir Level

Flow

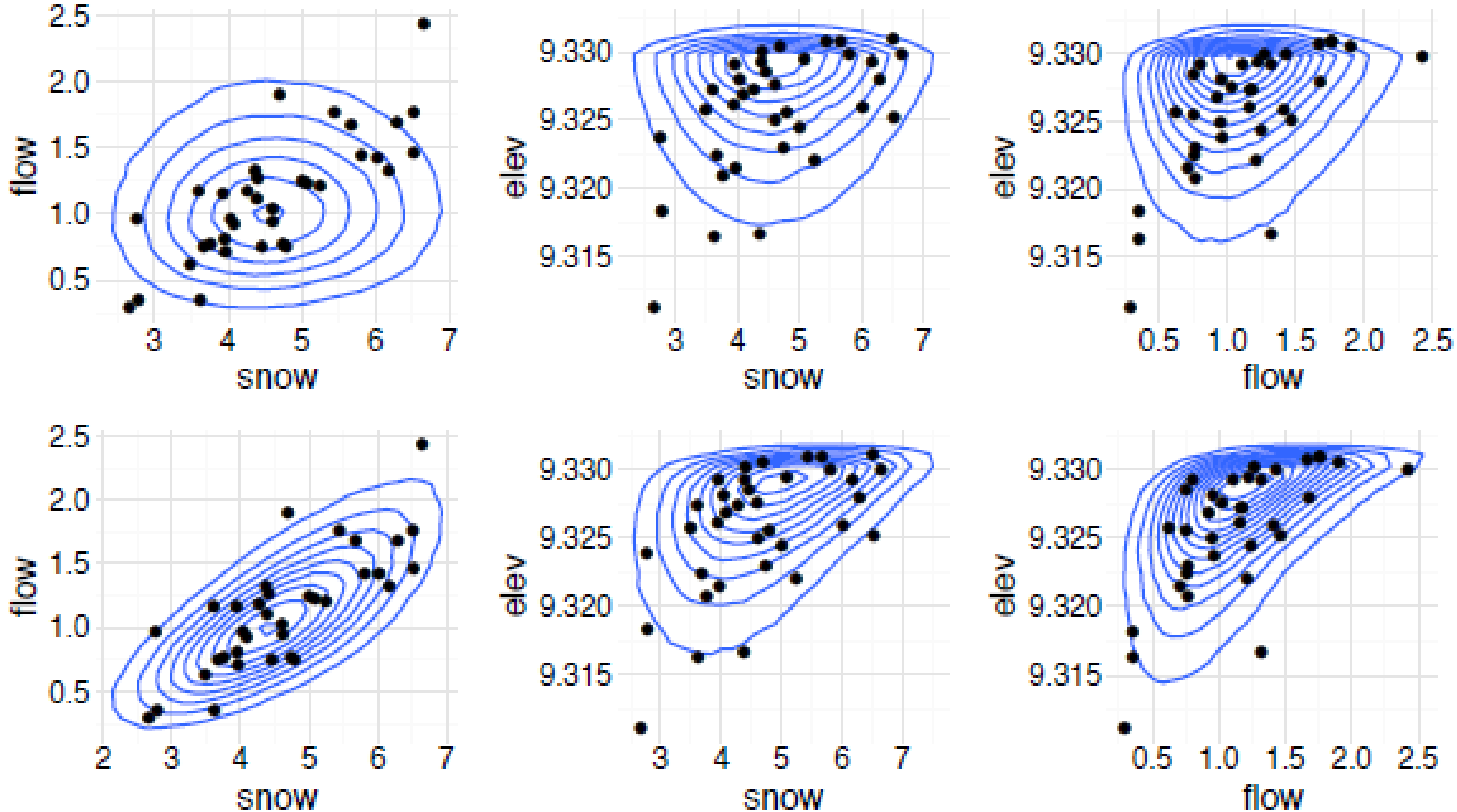
Snow

- Joint modeling Reduces uncertainty

Bracken et al., 2017, in review WRR

Results – Joint Relationships

- Modeling variables separately



- Joint Modeling

- Variable correlations are very well captured with joint modeling, compared to independent modeling

Bracken et al., 2017, in review WRR

Summary and Parting Thoughts

- Large scale climate features modulate moisture availability and transport to produce climate extremes in the Western U.S
  - Significant seasonal signatures in sources
  - ENSO effect on frequency of extremes

- Bayesian Hierarchical Modeling offers powerful and general framework for modeling extremes **with robust quantification of uncertainty**
  - In Space
  - Incorporate Temporal nonstationarity
  - Of several variables jointly (Multivariate Extremes)
  - And provide various return levels

- Climate Change projections can be incorporated as covariates

- Effective infrastructure management and societal responses for mitigating impacts of extremes are enabled

Summary and Parting Thoughts

- Weather generators offer attractive way to simulate space-time ensembles of daily weather
- Covariates are easily incorporated
  - Seasonal average precipitation, temperature etc.
    - Other covariates can also be used – weather types, NWS forecasts etc.
  - Enabling to simulate weather sequences conditioned on *Seasonal* and *Multidecadal Climate Projections*

- Can be used as an effective 'Downscaling' technique

- Easily coupled with hydrologic models to provide ensemble streamflow forecasts; capture forcing uncertainties
  - Can significantly improve upon ESP

Acknowledgements

•Bracken, C., B. Rajagopalan, W. Kleiber and L. Cheng and S. Gangopadhyay, Efficient Bayesian hierarchical modeling of spatial precipitation extremes over a large Domain, Water Resources Research, 52(8), 6643-6655, 2016

•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, 12(10), 4522-4533, 2015

•Bracken, C., K. Holman, B. Rajagopalan and H. Moradkhani, A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis, Water Resources Research, (in review, WRR), 2017

•Verdin, A., B. Rajagopalan, W. Kleiber, and R.W. Katz, 2015: "Coupled stochastic weather generation using spatial and generalized linear models." Stochastic Environmental Research and Risk Assessment, 29, 347-356

•Verdin, A., B. Rajagopalan, W. Kleiber , G. Podesta and F. Bert, 2017, "A conditional stochastic weather generator for seasonal to multi-decadal simulations." Journal of Hydrology, https://doi.org/10.1016/j.jhydrol.2015.12.036

•Mendoza, P., B. Rajagopalan, M. P. Clark, K. Ikeda and R. Rasmussen, 2015, Statistical postprocessing of high-resolution regional climate model output, Monthly Weather Review, doi: 10.1175/MWR-D-14-00159.1

Co-authors / Acknowledgements

- Pablo Mendoza - Univ. of Chile
- Cameron Bracken - Bonneville Power Administration

- Andrew Verdin - USBR, Denver

- Will Kleiber University of Colorado, Boulder, CO

- Andy Wood - NCAR
- Rick Katz - NCAR
- Guillermo Podesta - U. of Miami
- Federico Bert - U. of Buenos Aires

- Contact: balajir@Colorado.edu