CLEMSON UNIVERSITY

Estimating Precipitation Return Levels in the Carolinas via Spatial Extremes Methods with Focus on South Carolina's October 2015 Precipitation Event

Brook Russell, CU Department of Mathematical Sciences

Motivation

Primary Questions:

- How unusual was this event?
- What is a X-year precipitation event?
- Have SC precipitation extremes changed through time?

Secondary Questions:

- How does a spatial extremes model compare to the NOAA Atlas?
- Can we use a gridded data product instead of station data?

Outline

October 2015 Precipitation Event

- Hurricane Joaquin stalls off coast
- Systems from North and South funnel moisture to SC

Satellite image courtesy of NOAA

Radar Precipitation Estimates

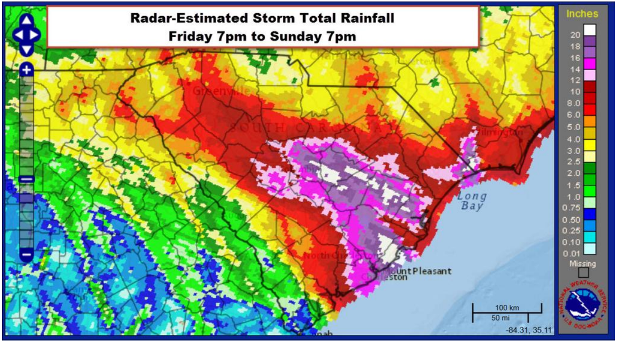


Image courtesy of NOAA

Record Setting Precipitation in Carolinas

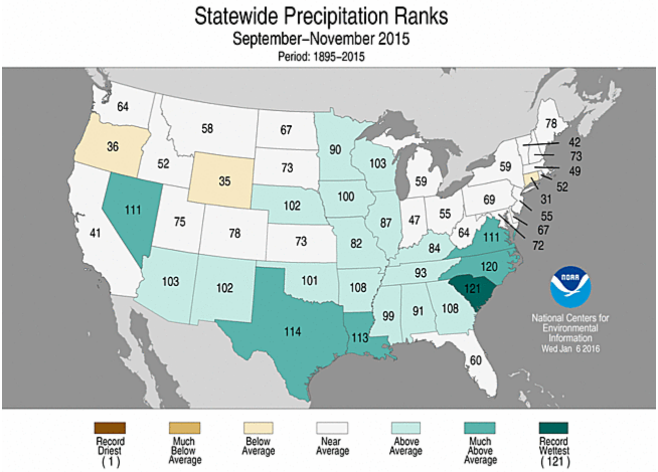


Image courtesy of NOAA

Clemson University vs. Notre Dame

Images courtesy Clemson University Athletic Department

Flood Damage in Columbia, SC

Image courtesy South Carolina National Guard

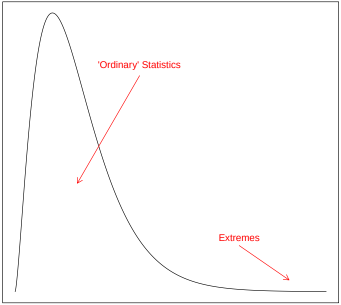
How Unusual was this Event?

- Wikipedia: "Rainfall across parts of South Carolina reached 500-year event levels, with areas near Columbia experiencing a 1-in-1000 year event"
- SC Governer: "We haven't seen this level of rain in the lowcountry in 1,000 years."

Image courtesy South Carolina National Guard

Extremes

- EVT offers methods to address this question
- In extremes we only use “extreme” observations



Return Levels

- Precipitation Return Level: The rainfall amount that is exceeded by the annual maximum in any particular year with probability p
- Return period is $1/p$
- Interpretations:
  - Average waiting time until next event exceeding this amount is $1/p$
  - Average number of events exceeding this amount occurring within return period is one

Fisher-Tippett-Gnedenko Theorem

For iid sample $X_1, \ldots, X_n$, if there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$P((\text{Max}(\{X_i\}_{i=1\ldots,n}) - b_n)/a_n \leq z) \xrightarrow{d} G(z)$$

for non-degenerate G, then G belongs to one of the following families:

- (Reverse) Weibull,
- Gumbel, or
- Fréchet

Generalized Extreme Value Distribution

- Generalized extreme value (GEV) distribution encompasses Weibull, Gumbel, and Fréchet
- GEV – three parameter family: $\mu \in \mathbb{R}, \sigma > 0, \xi \in \mathbb{R}$ (location, scale, shape)
  - $\xi < 0 \Rightarrow$ Weibull – bounded tail
  - $\xi = 0 \Rightarrow$ Gumbel – light tail
  - $\xi > 0 \Rightarrow$ Fréchet – heavy tail

Estimating Return Levels

Block Maxima Approach:

- Assume data made of large independent 'blocks' (often years)
- If blocks are large enough, block maxima could be considered GEV realizations
- Use sample of block maxima to estimate GEV parameters
- Estimate return level via function of GEV parameters

$$RL_p = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - \{-\log(1-p)\}^{-\xi}) \text{ for } \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} \text{ for } \xi = 0 \end{cases}$$

South Carolina Estimates

GEV parameter estimates

Charleston

- $\hat{\mu} = 3.29$, $\hat{\sigma} = 1.09$, $\hat{\xi} = .22$

Columbia

- $\hat{\mu} = 2.67$, $\hat{\sigma} = .63$, $\hat{\xi} = .27$

Greenville

- $\hat{\mu} = 2.71$, $\hat{\sigma} = .63$, $\hat{\xi} = .14$

South Carolina Estimated RLs

Charleston

- 100-yr RL: 11.9

Columbia

- 100-yr RL: 8.4

Greenville

- 100-yr RL: 6.8

NOAA Atlas 14

- Estimates precip return levels
- Not available in all locations
- Pac NW and Texas current documents created in 1960s, 70s, and 80s

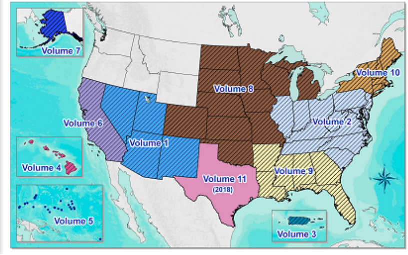


Image courtesy of NOAA

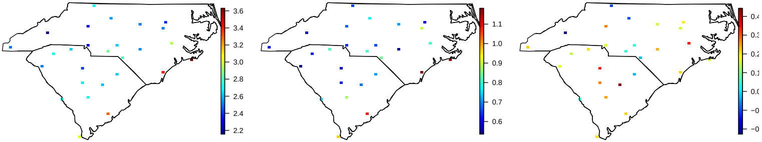
NOAA Atlas 14

100-year 24-hour precip RLs (NOAA Atlas 14 Volume 2 Version 3)

- Charleston
  - 100-yr RL: 10.1
- Columbia
  - 100-yr RL: 8.4
- Greenville
  - 100-yr RL: 9.23

Spatial Dependence

- Estimate GEV parameters at several locations
- Spatial dependence in GEV parameters
- Estimate RLs at each location



Outline

1. Introduction
   - Oct 2015 Precip Event
   - Background
   - Exploratory Analysis
2. Estimating Return Levels
   - Bayesian Hierarchical Model
   - Carolina Analysis
   - Comparison to NOAA Atlas
3. Gridded Data Products
   - NARR and PERSIANN-CDR
4. Conclusion

Bayesian Hierarchical Model

- Idea: let parameters of GEV vary through space and time, similar to Apputhurai and Stephenson (2013)
- At time t and location $\mathbf{s}$, assume

$$Y_{\mathbf{s},t} | \mu_{\mathbf{s},t}, \sigma_{\mathbf{s},t}, \xi_{\mathbf{s},t} \sim \text{GEV}(\mu_{\mathbf{s},t}, \sigma_{\mathbf{s},t}, \xi_{\mathbf{s},t})$$

- Build models for $\mu_{\mathbf{s},t}, \sigma_{\mathbf{s},t}, \xi_{\mathbf{s},t}$ based on functions of t and $\mathbf{x}_{\mathbf{s},\mu}, \mathbf{x}_{\mathbf{s},\sigma}, \mathbf{x}_{\mathbf{s},\xi}$ (vectors of spatial covariates at $\mathbf{s}$)
  - $\mu_{\mathbf{s},t} = \mathbf{x}_{\mathbf{s},\mu}^T \boldsymbol{\beta}_\mu + f_\mu(t) + W_\mu(\mathbf{s})$
  - $\log(\sigma_{\mathbf{s},t}) = \mathbf{x}_{\mathbf{s},\sigma}^T \boldsymbol{\beta}_\sigma + f_\sigma(t) + W_\sigma(\mathbf{s})$
  - $\xi_{\mathbf{s},t} = \mathbf{x}_{\mathbf{s},\xi}^T \boldsymbol{\beta}_\xi + W_\xi(\mathbf{s})$
- $\boldsymbol{\beta}_.$ – vectors of regression coefficients
- $f_.$ – temporal functions
- $W_.$ – spatially correlated, mean zero Gaussian random effects

Simplifying Assumptions

- Simplifying assumptions
  - Exponential covariance functions

$$Cov(W.(\mathbf{s}), W.(\mathbf{s}')) = \alpha.\exp(-\lambda.d(\mathbf{s}, \mathbf{s}'))$$

  - No temporal component for σ

$$f_\sigma(t) = 0$$

  - Linear temporal trend for μ

$$f_\mu(t) = \rho_\mu t$$

  - No spatial covariates for ξ

$$\mathbf{x}_{\mathbf{s},\xi} = 1$$

- Analysis indicates $\rho_\mu = 0$

Bayesian Hierarchical Model

$$\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathbf{y}) \propto \underbrace{\pi(\mathbf{y}|\boldsymbol{\theta}_1)}_{\text{data}} \underbrace{\pi(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)}_{\text{process}} \underbrace{\pi(\boldsymbol{\theta}_2)}_{\text{prior}}$$

- Data Level: Likelihood characterizing distribution of data given value of parameters at process level – GEV
- Process Level: Latent processes modeling μ, σ, and ξ – GP
- Prior Level: Prior distributions for parameters
  - Regression Parameters – MV Normal
  - Sill – Inverse Gamma
  - Range – Gamma

Selecting Spatial Covariates

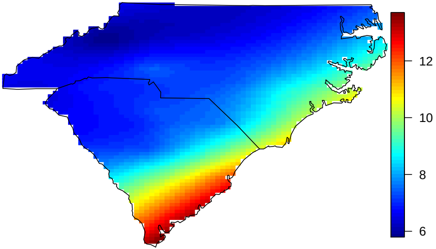
- Typical spatial covariates
  - Latitude
  - Longitude
  - Elevation
  - Annual average precipitation
  - Distance to coast
- Several covariates correlated

MCMC

- Three parallel chains
- 25,000 iterations after burn-in for each model
- Best model (highest order term)
  - Location – $dist^2$
  - Scale – $dist^2$
  - Shape – constant
- Draws used to perform pointwise spatial interpolation

Map Results

Estimated 100-yr One-day RLs



NOAA Atlas 14

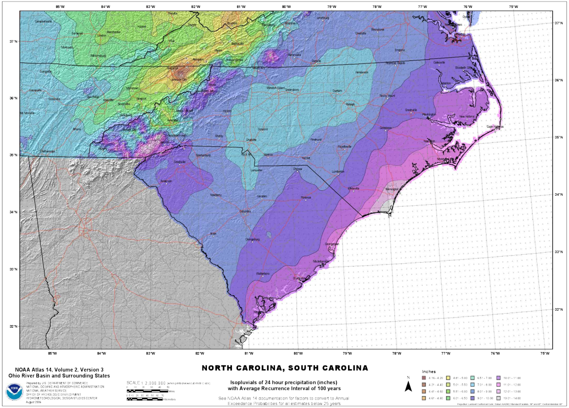
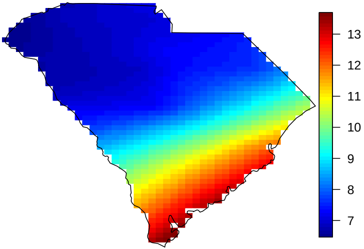


Image courtesy of NOAA

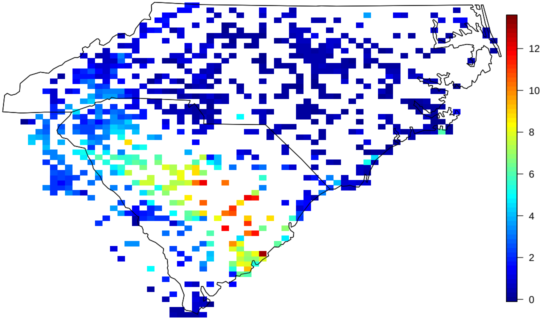
Map Results

Estimated 100-yr One-day RLs

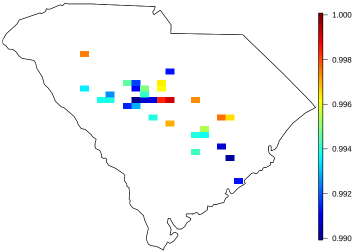


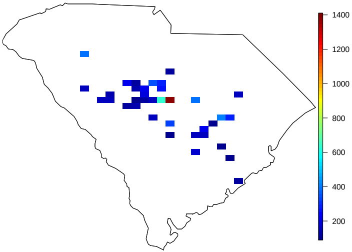
How Unusual Was 10/04/2015?

Observed Precip Values:



How Unusual Was 10/04/2015?





Outline

Station Data vs. Gridded Data

Station Data:

- + No averaging, should pick up extreme observations
- – Can be "messy" (missing observations, etc.)
- – Limited data in some regions

Gridded Data:

- + No missing observations
- + Good spatial coverage
- – Shorter data record
- – Average over space/time, could miss extreme events

Gridded Data Products

NARR

- North American Regional Reanalysis
- Extension of the NCEP Global Reanalysis run over NA
- 3-hour data 1979-present, approximately .3° (32km) resolution
- Incorporates data from many sources

PERSIANN-CDR

- Precipitation Estimation from Remote Sensing Information using Artificial Neural Networks
- Daily data 1983-present, .25° global resolution (60°S – 60°N)
- Basic process
  - Uses artificial neural networks to convert GridSat-B1 IR satellite data to rain rate
  - Bias corrected with monthly GPCP (Global Precipitation Climatology Project) data

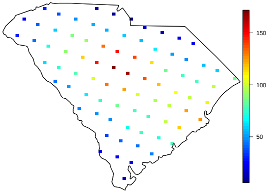
Comparison of Gridded Products

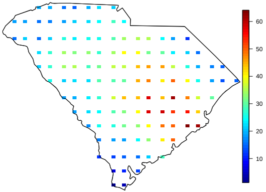
How well do gridded products capture extremes?

- NARR – known to underestimate extreme precipitation
- PERSIANN-CDR – Performance varies by location and season, “slightly underestimates the values of extreme heavy precipitation” Miao et al. (2015)

Comparison of Gridded Products

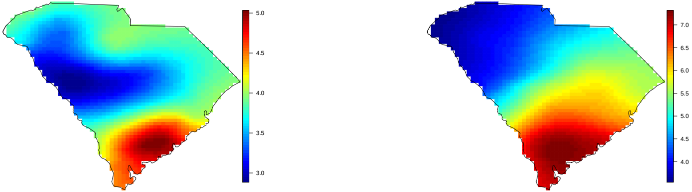
Precipitation on 10/04/2015: NARR (L) and PERSIANN-CDR (R)





Estimated RLs based on Gridded Data

100-year RLs: NARR (L) and PERSIANN-CDR (R)



Potential Solutions:

- Downscaling
- Mannshardt-Shamseldin et al. (2010) – regression relationship

Conclusions

Primary Questions:

- How unusual was this event?

Very unusual in much of SC, extremely unusual in Columbia area

- What is a X-year precipitation event?

Average waiting time until next event exceeding this amount is X

- Have SC precipitation extremes changed through time?

No evidence of change – need more data to address this question...

Secondary Questions:

- How does a spatial extremes model compare to the NOAA Atlas?

They give similar answers for the state of SC

- Can we use a gridded data product instead of station data?

The two we looked at underestimate precip extremes

Interesting Work

- Cooley et al. (2007) – Threshold exceedance approach
- Apputhurai and Stephenson (2013) – Spatiotemporal model for GEV parameters
- Dyrrdal et al. (2015) – Spatially model GEV parameters incorporating BMA
- Schliep et al. (2010) – Look at extreme precip from RCMs
- Davison et al. (2012) – Good spatial extremes overview

References I

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

Cooley, D., Nychka, D., and Naveau, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.

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

Dyrrdal, A. V., Lenkoski, A., Thorarinsdottir, T. L., and Stordal, F. (2015). Bayesian hierarchical modeling of extreme hourly precipitation in Norway. *Environmetrics*, 26(2):89–106.

References II

Mannshardt-Shamseldin, E. C., Smith, R. L., Sain, S. R., Mearns, L. O., and Cooley, D. (2010). Downscaling extremes: A comparison of extreme value distributions in point-source and gridded precipitation data. *The Annals of Applied Statistics*, pages 484–502.

Miao, C., Ashouri, H., Hsu, K.-L., Sorooshian, S., and Duan, Q. (2015). Evaluation of the PERSIANN-CDR daily rainfall estimates in capturing the behavior of extreme precipitation events over China. *Journal of Hydrometeorology*, 16(3):1387–1396.

Schliep, E. M., Cooley, D., Sain, S. R., and Hoeting, J. A. (2010). A comparison study of extreme precipitation from six different regional climate models via spatial hierarchical modeling. *Extremes*, 13(2):219–239.