New Exploratory Tools for Extremal Dependence: χ and Annual Extremal Networks

Dan Cooley & Whitney Huang
Joint work with SAMSI Networks and Extremes working group

Colorado State University & Clemson University

JSM2019 Session on Climate Networks and Extremes, July, 30, 2019

Outline of the talk

- Motivation and Background
  - 2017 Atlantic hurricane season and rainfall extremes
  - To develop climate network methods for exploring extremal dependence
- χ and Annual Extremal Networks
  - An empirical and its bias-corrected estimator of χ network
  - The use of annual extremal network to explore the year-to-year variation of extremal dependence
- Gulf Coast Extreme Rainfall Application

2017 Atlantic hurricane season

Source: NOAA/NASA

Motivation

We would like to explore the spatial dependence of hurricane seasons rainfall extremes at the Gulf Coast and surrounding area.

- A standard treatment of the data from these three storms would likely treat them as independent events, due to their temporal lag and spatial distance.
- However, it is largely acknowledged that these storms are related, arising from conditions in 2017 conducive to tropical cyclone formation and intensication
- We seek to use an climate network-type approach to explore the extremal depedence of rainfall extremes

Climates network for extremes

- A climate network [Tsonis and Roebber 2004] consists of **nodes** (stations or grid cells) and **edges**. Two nodes are connected by an edge depending on the degree of statistical dependence between the corresponding pairs of time series

- Most studies use correlation to construct climate networks $\Rightarrow$ may fail to capture tail dependence structure (see next slide for an illustration)

- We use the upper tail dependence (χ) as the measure of the tail dependence between a pair of annual maxima series

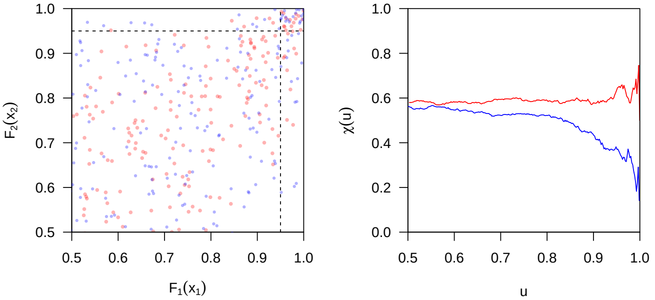
Characterizing (bivariate) tail dependence

Let $\mathbf{X} = (X_1, X_2)$ be a bivariate random vector with marginal CDF F_1, F_2

Upper tail dependence: $\chi = \lim_{u \to 1^-} \chi(u)$, where

$$\chi(u) = \mathbb{P}(F_2(X_2) > u | F_1(X_1) > u)$$

$\Rightarrow$ the probability of one variable being extreme given that the other is extreme



An empirical estimator of χ network

Input: $\{m_{i,j}\}$: annual maximum series (in year $i = 1, \cdots, n_j$) at locations $j \in \mathcal{S}$ and a threshold χ_{min}.

1. Compute the empirical distribution function of M_j

$$u_{i,j} = \frac{\mathsf{Rank}_j(m_{i,j})}{n_j + 1}$$

2. Compute the **F-madogram** (Cooley et al. 2006):

$$\hat{\nu}_{jj'} = \frac{1}{2}\frac{1}{n_{j,j'}}\sum_{i=1}^{n_{j,j'}} |u_{i,j} - u_{i,j'}|, \quad j, j' \in \mathcal{S}$$

χ network estimation cont'd

3. Compute the **extremal coefficient** (Smith, 1990):

$$\hat{\theta}_{jj'} = \frac{1 + 2\hat{\nu}_{jj'}}{1 - 2\hat{\nu}_{jj'}}, \quad j, j' \in \mathcal{S}$$

4. Compute χ:

$$\hat{\chi}_{jj'} = 2 - \hat{\theta}_{jj'}, \quad j, j' \in \mathcal{S}$$

5. Connect the pairs s.t. $\hat{\chi}_{jj'} > \chi_{\text{min}}$

Output: The χ network $G = (V, E)$ for the given threshold χ_{min}.

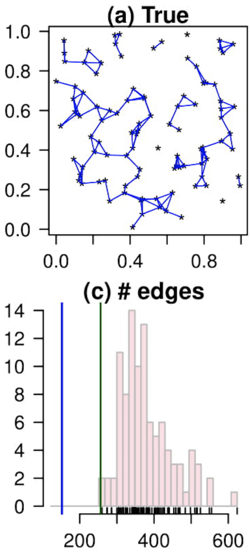
An simulation study

- We simulate 100 realizations from a Brown–Resnick max–stable process, each with 50 “annual maxima” at 100 locations

- We apply our empirical estimator to obtain estimated χ networks

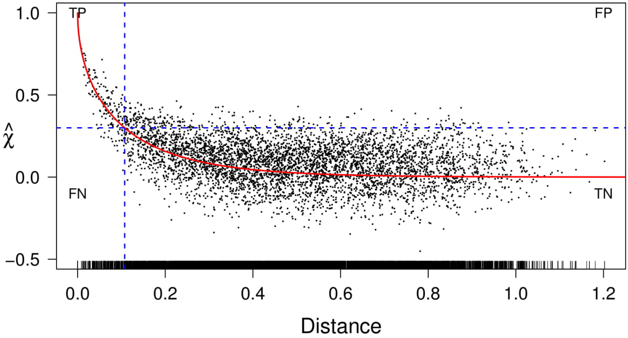
- We evaluate the estimator’s performance using some network statistics:
  - $\#$ of edges
  - True positive rate: TPR = $\frac{\#\{(j,j'):\hat{\chi}_{jj'}>\chi_{\text{min}} \text{ and } \chi_{jj'}>\chi_{\text{min}}\}}{\#\{(j,j'):\chi_{jj'}>\chi_{\text{min}}\}}$
  - Positive predictive value: PPV = $\frac{\#\{(j,j'):\hat{\chi}_{jj'}>\chi_{\text{min}} \text{ and } \chi_{jj'}>\chi_{\text{min}}\}}{\#\{(j,j'):\hat{\chi}_{jj'}>\chi_{\text{min}}\}}$

The number of edges is overestimated ☹



Understanding the network bias

Although $\{\hat{\chi}_{jj'}\}$ appear unbiased, it is the act of thresholding which introduces the bias

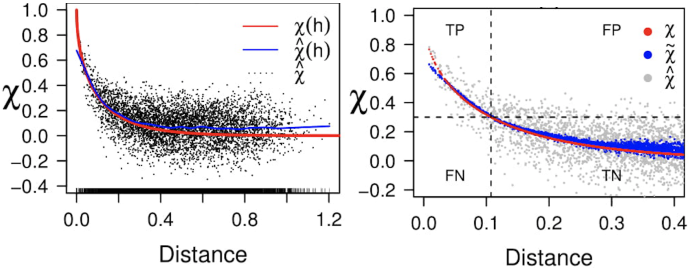


We could exploit the spatial structure of $\{\hat{\chi}_{jj'}\}$ to spatially regularize the network estimation

Network bias correction

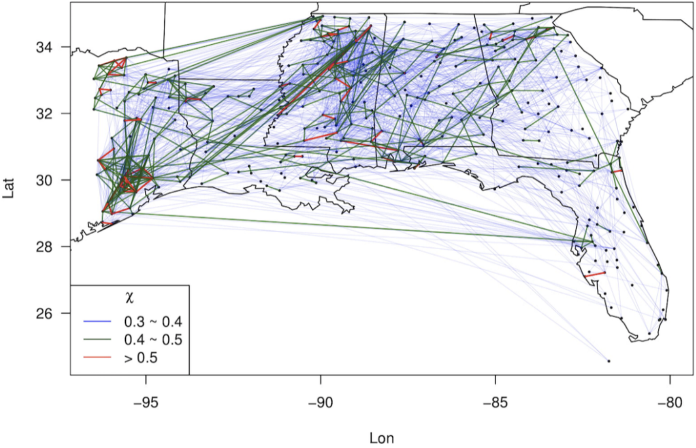
$$\tilde{\chi}_{jj'} = \lambda_{jj'}\hat{\chi}_{jj'} + (1 - \lambda_{jj'})\hat{\chi}(h_{jj'}), \quad j, j' = 1, \cdots, d.$$

- $\hat{\chi}_{jj'}$: empirical estimate; $\hat{\chi}(h_{jj'}) \sim \mathrm{N}(\hat{\chi}(h_{jj'}), \tau^2(h_{jj'}))$: spatial "prior"
- $\lambda_{jj'} = \tau^2_{jj'}/(\tau^2_{jj'} + \sigma^{2*}_{jj'})$, where $\sigma^{2*}_{jj'} = \mathrm{se}(\hat{\chi}_{jj'})$

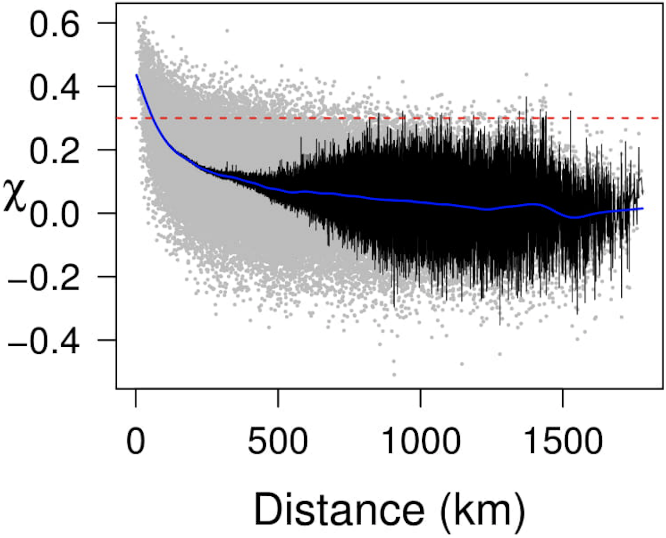


χ network for Gulf Coast rainfall extremes

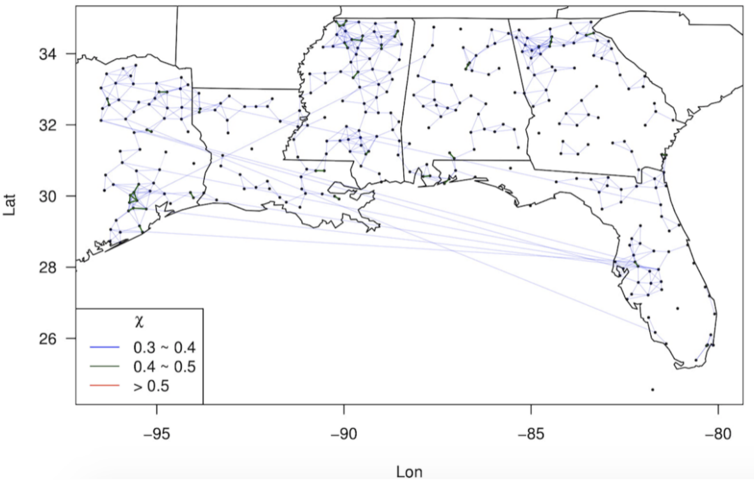
- 339 GHCN weather stations in TX, LA, MS, AL, FL, GA
- Hurricane season maxima (June - Oct.) from 1949 to 2017



Bias-corrected network estimate



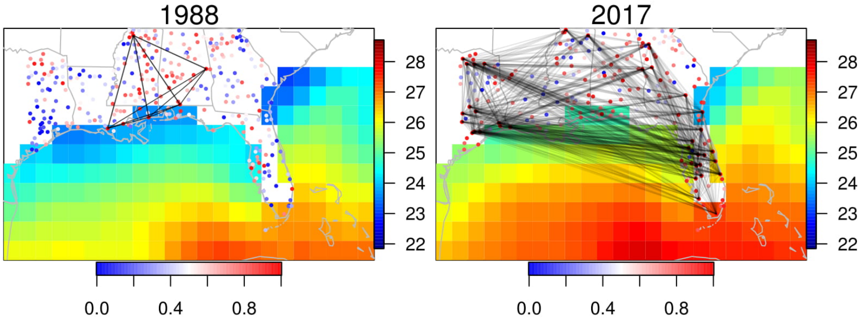
Bias-corrected network estimate cont'd



Annual extremal Network

- For each hurricane season, we connect the pairs where their EDFs of the season maximum exceed 0.95 (i.e., 20-year event)

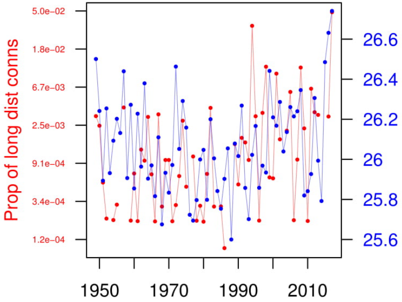
- We study how the inter-annual variability of the numbers of the “long distance” (e.g., 1000km apart) extremal pairs might be explained by some meteorological variable (e.g., SST)

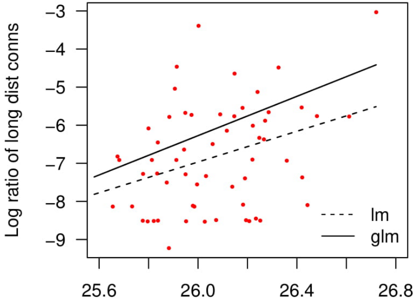


Relating the number of long distance extremal connections to SST

log ratio: $\log \frac{\text{\# long distance extremal connections}}{\text{\# all long distance pairs}}$

- `lm`: log ratio = $\alpha_0 + \alpha_1$SST + ε.
- `glm`: $\mathbb{E}[\log(\mu)|\text{SST}] = \beta_0 + \beta_1$SST.





Summary

- We develop the χ and annual extremal networks for exploring dependence structure of extreme values
- We identify the issue of network bias and we propose a bias correction method by exploiting the spatial dependence structure
- These tools allow us to quickly explore the extremal dependence structure and could provide some guidance on how to proceed the following confirmatory analysis

More details can be found in:

Huang, W. K., Cooley, D. S., Ebert-Uphoff, I., Chen, C., Chatterjee, S.B.
New Exploratory Tools for Extremal Dependence: χ Networks and Annual Extremal Networks.
Journal of Agricultural, Biological, and Environmental Statistics, 1–18, 2019