Spatial-Temporal Extreme Modeling through Point-to-Area Random Effects (PARE)

Carlynn Fagnant

Julia Schedler (speaker)

Katherine B. Ensor

Thursday, May 25 2023

Dr. Carlynn Fagnant

Dr. Katherine Ensor

Dr. Julia Schedler

Modeling Context

Extreme precipitation

Greater Houston Area

Hurricane Harvey

Answer two key questions:

- Have extreme precipitation trends in the Houston area shifted over the past century?
- How are these changes related spatially?

Answers are communicated through **RETURN LEVELS (quantiles incorporating event frequency)** – on average we should see this level every “25”, “100” and “500” years

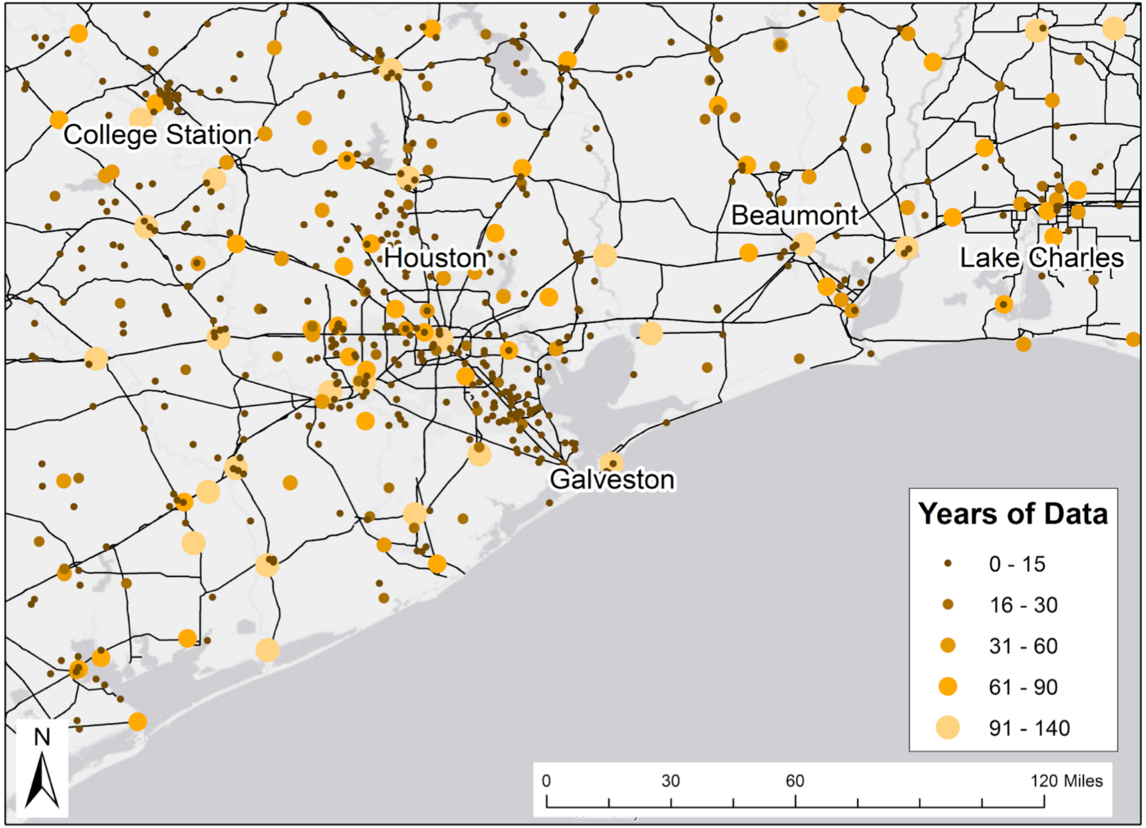
Focus on GAUGE MEASURED RAINFALL

Constraints

- on time!! With not much time and limited resources.
- modeling is accessible to a wide range of engineers (CEVE community)
- uses existing software and methods
- scientifically defensible / legally defensible?

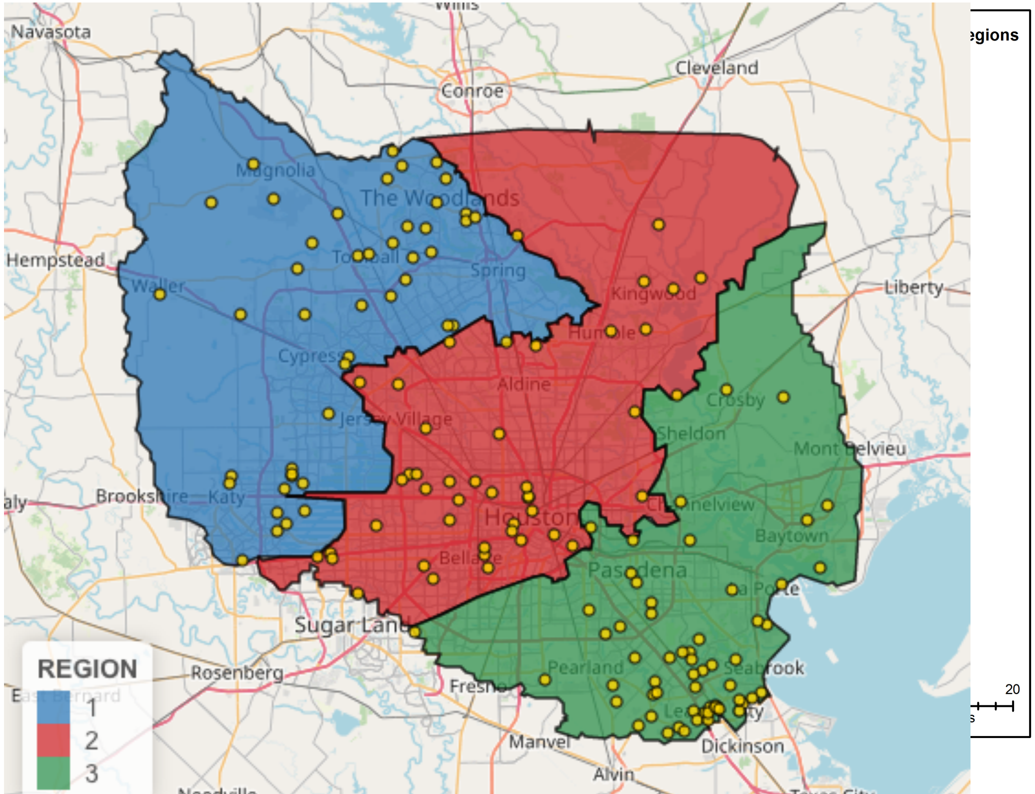
Data - Measured Rainfall

- Daily rainfall
- 600 stations
- Earliest date 1870
- From NOAA's Global Historical Climatology Network (GHCN)



Expert estimates- hydrologic regions

- Each of the 3 hydrologic regions is a collection of watersheds
- Determined by experts
- Historic return levels available at hydrologic region level



What do we do when…

- the target of estimation and inference are not central/“typical” values?
- the data is collected at the point level but estimates are desired at the regional level?

Previous work

Article

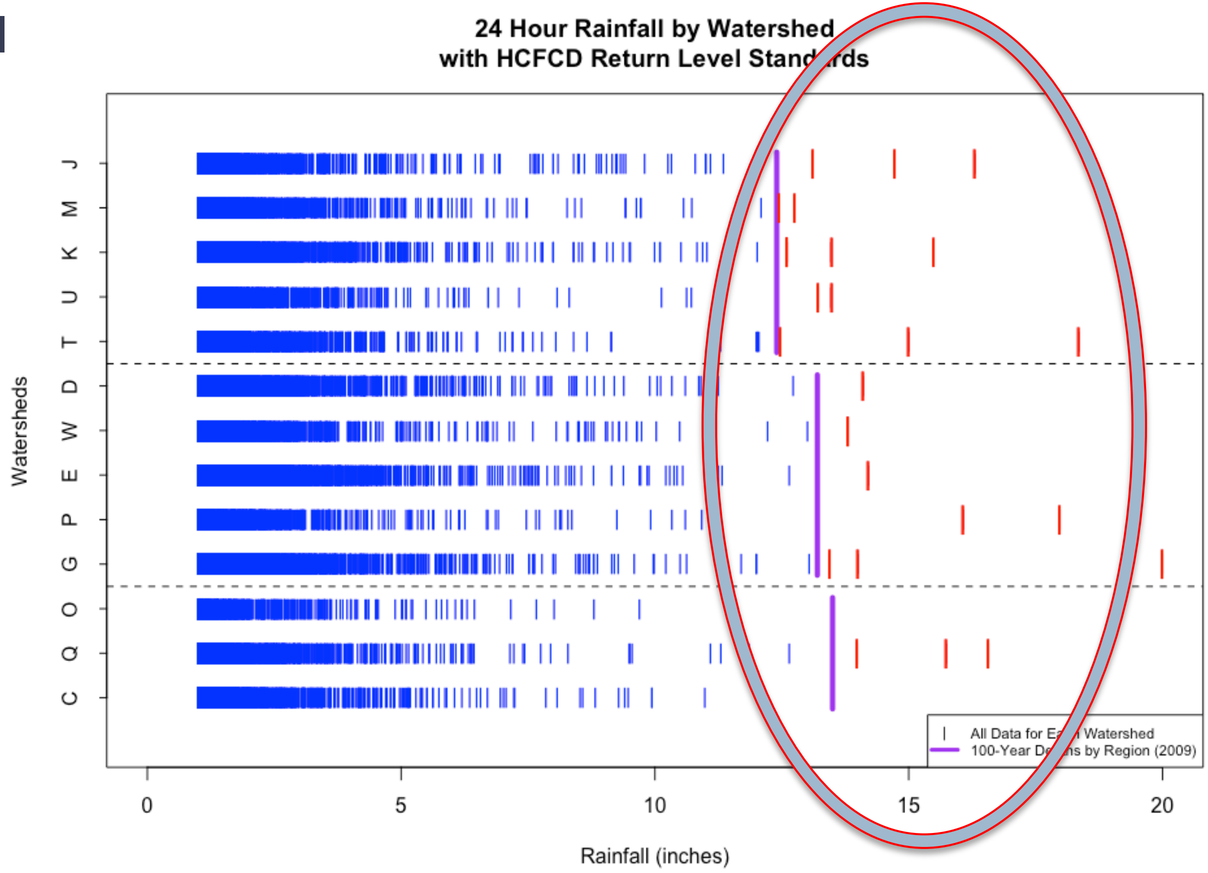
Characterizing spatiotemporal trends in extreme precipitation in Southeast Texas

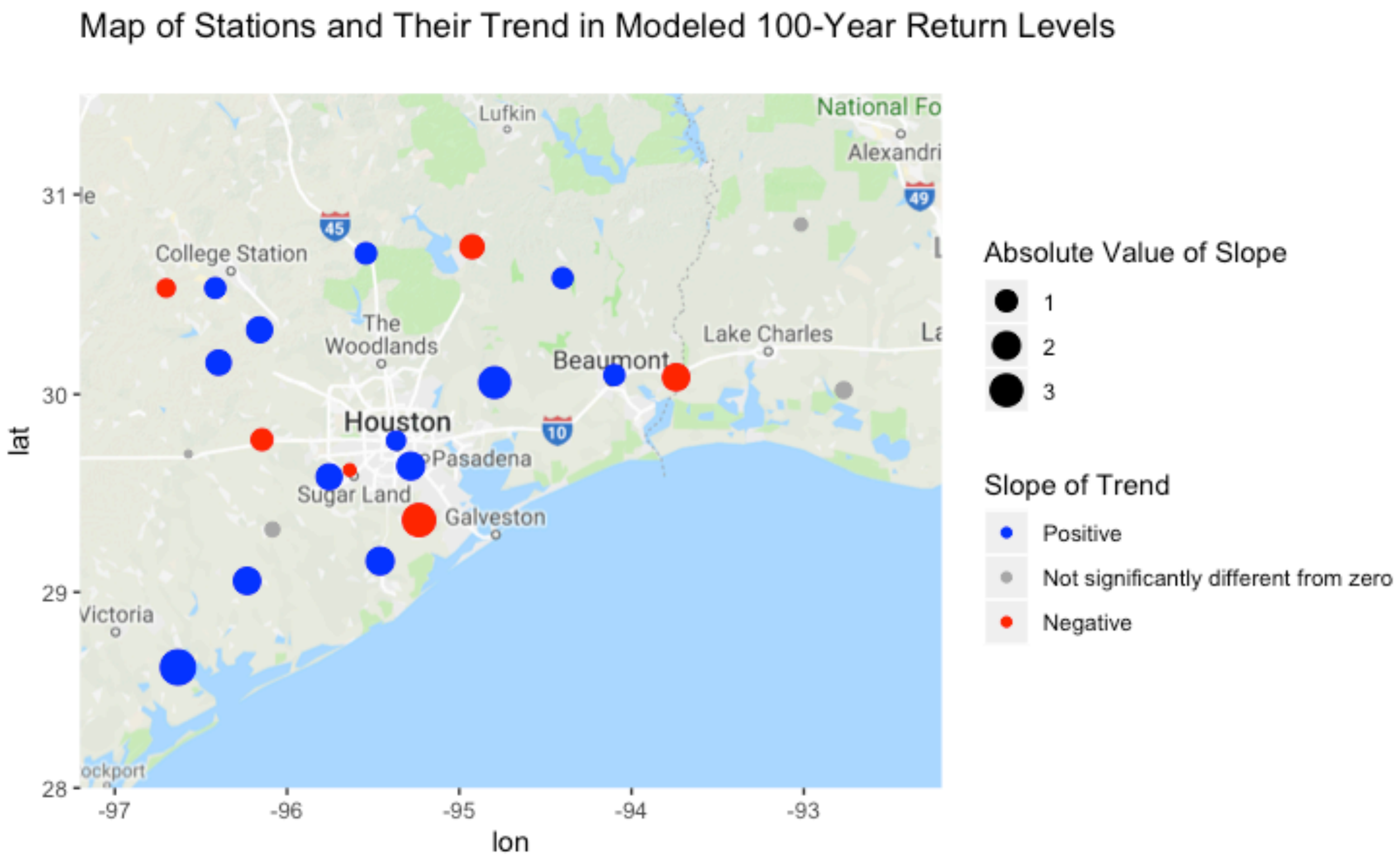
Fagnant, C.[1], Gori, A.[2,3], Ensor, K.B.[1*], Sebastian, A.[2,4] and Bedient, P.B.[2]

$$x_N = \begin{cases} u + \frac{\sigma}{\xi}\left[(\lambda N)^{\xi} - 1\right] & \text{if } \xi \neq 0 \\ u + \sigma\, log(\lambda N) & \text{if } \xi = 0 \end{cases}$$

Return level based on GPD

- The N-year return level is the value that has a 1/N probability of being reached or exceeded in a given year.





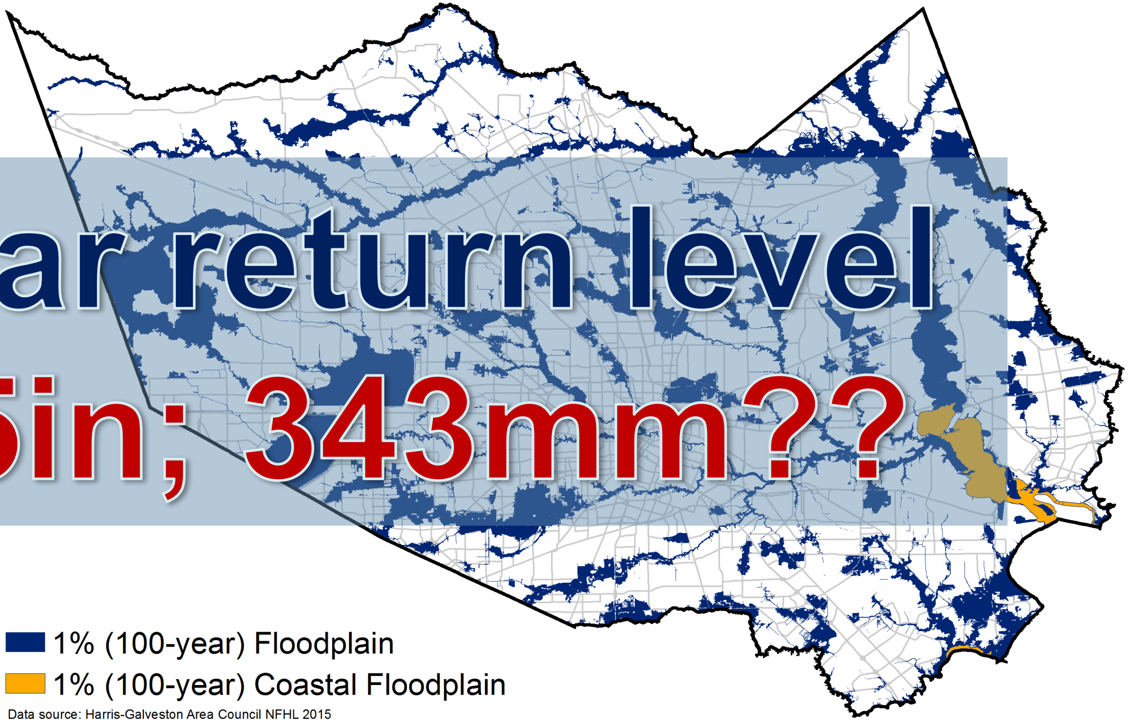
Map of Stations and Their Trend in Modeled 100-Year Return Levels
Absolute Value of Slope
1
2
3
Slope of Trend
Positive
Not significantly different from zero
Negative
lat
lon
31
30
29
28
-97
-96
-95
-94
-93
Lufkin
National Fo
Alexandri
College Station
The Woodlands
Lake Charles
Beaumont
Houston
Pasadena
Sugar Land
Galveston
Victoria
ockport
45
49
10

Flood plain maps are inaccurate

- 47% off the mark in Harris county
- Costs are staggering and rising
- Global coastal flooding expected to reach $1 trillion per year by 2050 (Hellegate et al 2013)
- Brody, Sebastian, Blessing (2014)

100 year return level

~13.5in; 343mm??



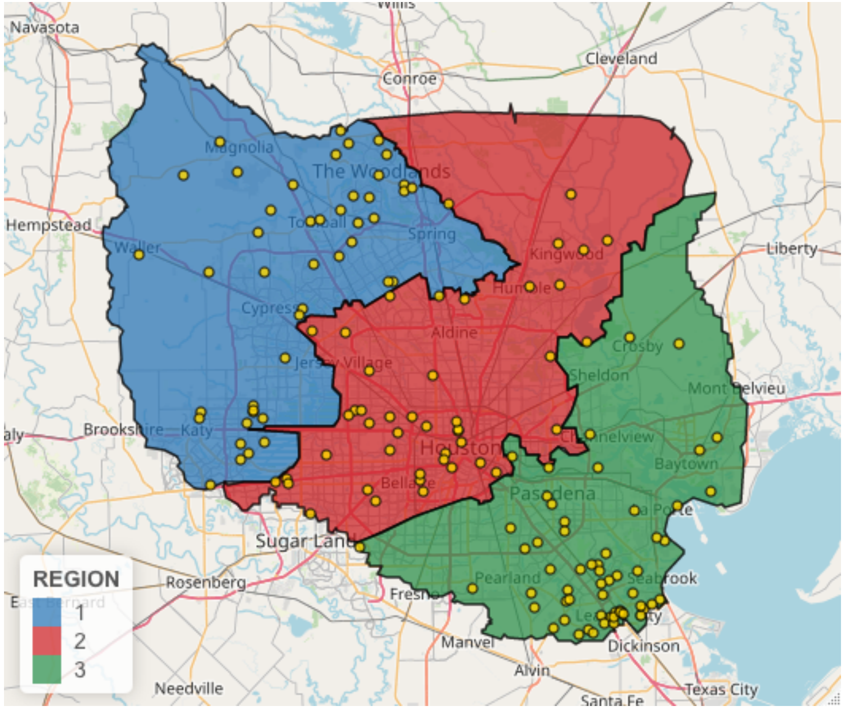
Recap

- Using GPD and 40-year rolling window works well for modeling the extreme precipitation
- Spatial variation in the temporal trends

Modeling the spatial structure

Change-of-support problem

- Data at point level, estimates/inference desired at regional level
- Gotway and Young (2002)



Modeling plan

- Obtain **regional** estimates of shape, scale, and rate using 3 spatial change-of-support structures
  1. Regional max (“data” are observed rainfall for each station)
  2. Block Kriging (“data” are GPD MLEs for each station)
  3. PARE (“data” are GPD MLEs for each station)
- Calculate estimated regional return levels and compare results

1. Regional Max (procedure)

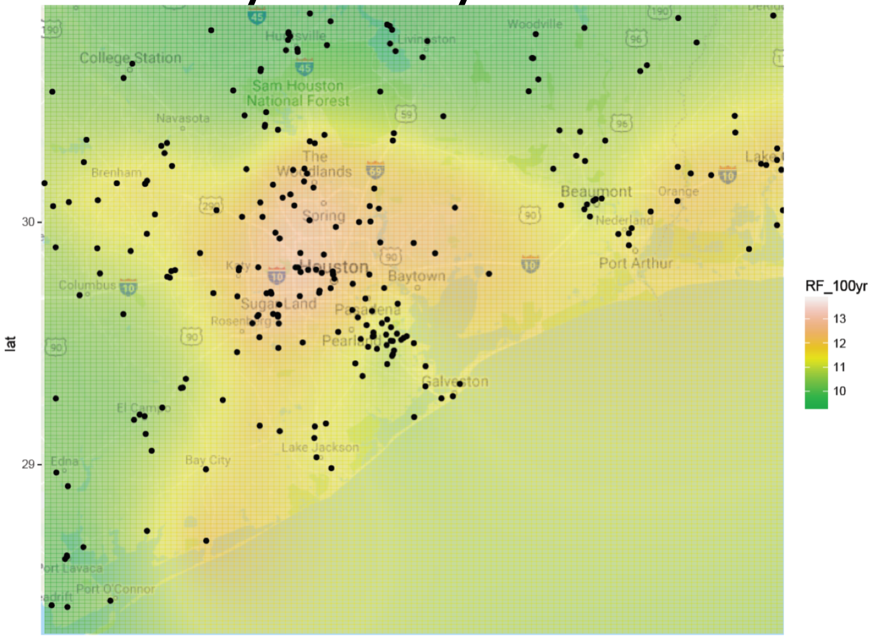
1. Create a single rainfall data series for each region by taking the maximum of the daily values
2. Decluster each region's series using a run of 1 day
3. Apply univariate extreme value modeling to obtain GPD parameter estimates for each region
4. Calculate regional return levels using the parameter estimates

2. Block Kriging (procedure)

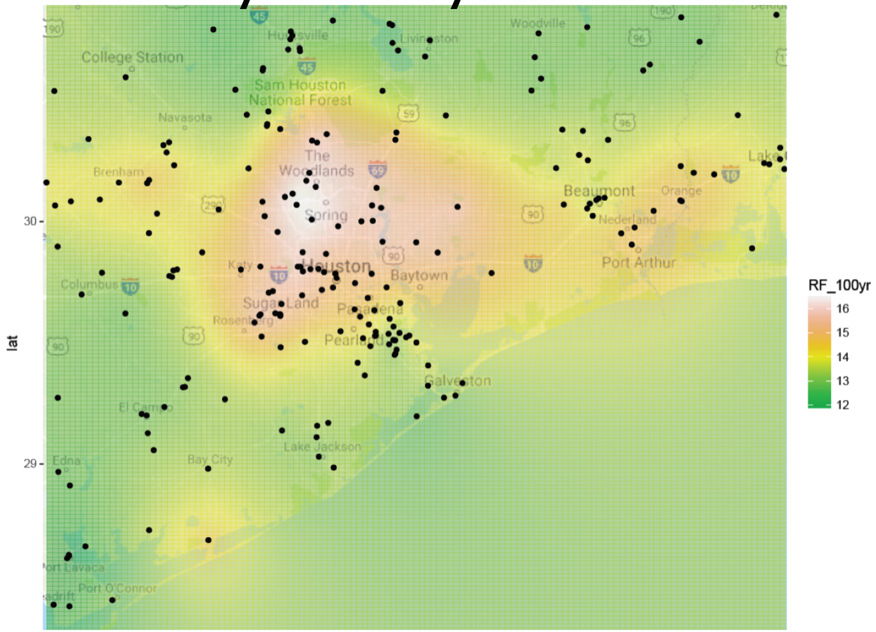
1. Apply univariate extreme value approach to obtain MLE estimates of GPD parameters for each *station*
2. (cross) variogram modeling for each parameter over grid
3. Apply kriging (rate) /cokriging (shape, scale) to the MLEs
4. Average the (co)kriged estimates over each region to obtain regional parameter estimates
5. Calculate regional return levels

Kriging-based Return Level

1-day 100-year



2-day 100-year



Hierarchical model

Data Model: $Z(s_i)|Y(s_i), \sigma_\varepsilon^2 \sim N\left(Y(s_i),\ \sigma_\varepsilon^2\right)$

Process Model: $Y(\cdot)|\mu, C_Y \sim N(\mu,\ C_Y)$

Predictive Distribution: $Y(s_0)|\mathbf{Z}, \mu, c_Y, \sigma_\varepsilon^2 \sim N(E[Y(s_0)|\mathbf{Z}],\ var(Y(s_0)|\mathbf{Z}))$

$$= N\left(Y^*(s_0),\ \sigma_Y^2(s_0)\right)$$

3. PARE model (procedure)

1. Extreme value fit to each station
2. Create spatial weight matrix (median Hausdorff) and regional indicators
3. Fit CAR model to obtain regional estimates of shape, scale, and rate
4. Calculate regional return levels

PARE (continued)

$$Z_i | z_{(-i)} \sim N\left(\mathbf{x}_i{}'\beta + \sum_{j \neq i} \rho w_{ij}(z_j - \mathbf{x}_i{}'\beta),\ m_{ii}\right)$$

- Can be run in R using spatialreg::spautolm with family="CAR"

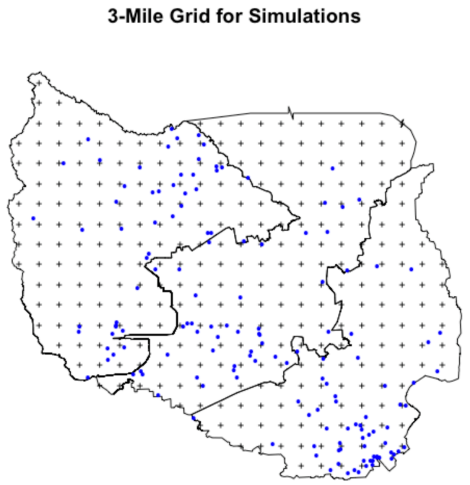
PARE (continued)

- Conditional Autoregressive model to move from point-level to regional-level relationship.
- W: Describe distance as in an areal CAR model (region to region (3x3)), but bring structure to the individual stations (nxn)
  - For example, distance between every point in R1 and every point in R2 is based on the median Hausdorff distance between R1 and R2.
  - Distance between points in a single region is a constant, eg. 1 mile
  - Inverting W: Add small amount of noise to each value eg, N(0, .1) to “jitter” our weights
- Region parameter: Create covariate of indicator variables for region. The estimate of this parameter is what we are seeking.

Simulations

Simulation procedure

- Using GPD first for 1981-2020, compute regional means for each parameter as “true” values
- Simulate the rainfall data on a grid
  - Simulate the **independent excendances**
  - **Impose a pseudo-time ordering for regional max**
  - **Impose daily-level correlations via ranking**
  - **Rate parameter held constant**
- **Fit the 3 models for 50 replicates**



Simulation results

		Truth	Model 1 - PARE Model			Model 2 - Block Kriging			Model 3 - Regional Max		
			Mean	RMSE	MAE	Mean	RMSE	MAE	Mean	RMSE	MAE
Scale	Region 1	233.64	233.96	1.3466	1.0536	235.85	2.5194	2.2470	250.84	17.5993	17.1950
	Region 2	246.78	247.07	1.3125	1.0537	245.11	2.1282	1.8337	264.27	17.8354	17.4853
	Region 3	229.38	229.70	1.1994	0.9160	229.61	1.1697	0.9225	247.01	17.9624	17.6309
Shape	Region 1	0.2044	0.2013	0.0056	0.0044	0.1984	0.0072	0.0065	0.3326	0.1291	0.1282
	Region 2	0.2319	0.2295	0.0050	0.0041	0.2261	0.0073	0.0061	0.3676	0.1363	0.1357
	Region 3	0.1641	0.1621	0.0044	0.0035	0.1700	0.0071	0.0060	0.2775	0.1143	0.1134

- With a few exceptions, PARE performs best.

Application to precipitation

Estimated region parameters

Method	Parameter	Region 1	Region 2	Region 3
1. PARE Model	scale	214.39 (4.077)	217.08 (4.337)	221.40 (3.322)
	shape	0.2039 (0.0163)	0.2320 (0.0172)	0.1643 (0.0129)
	rate	0.0570 (0.0029)	0.0514 (0.0031)	0.0545 (0.0023)
2. Block Kriging	scale	208.83 (16.521)	228.30 (18.557)	229.41 (17.530)
	shape	0.2205 (0.0696)	0.1771 (0.0715)	0.1121 (0.0672)
	rate	0.0559 (0.0058)	0.0554 (0.0058)	0.0557 (0.0057)
3. Regional Max	scale	291.21 (14.897)	343.46 (16.778)	331.38 (16.747)
	shape	0.1523 (0.0377)	0.1397 (0.0355)	0.1351 (0.0366)
	rate	0.0565 (0.0019)	0.0601 (0.0020)	0.0557 (0.0019)

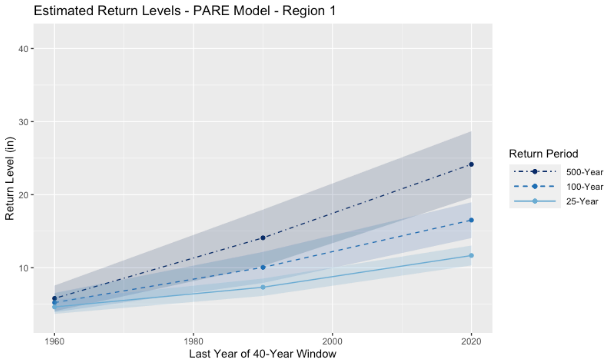
Table 5.2 : Comparison of extreme value parameter estimates (standard errors in parentheses) across our proposed models for the last 40 years of data, 1981-2020. As we predicted, the regional max results differ more than the other two models with its scale parameters being much higher, which may lead to larger return level estimates.

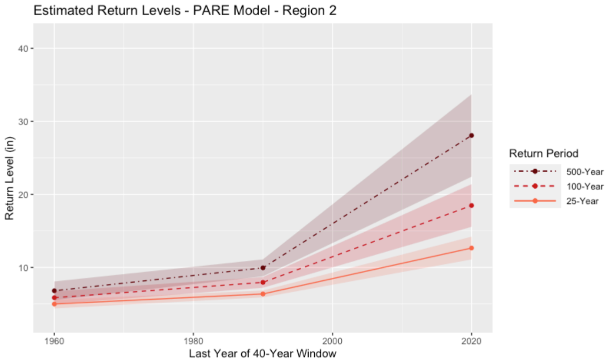
Return level estimates

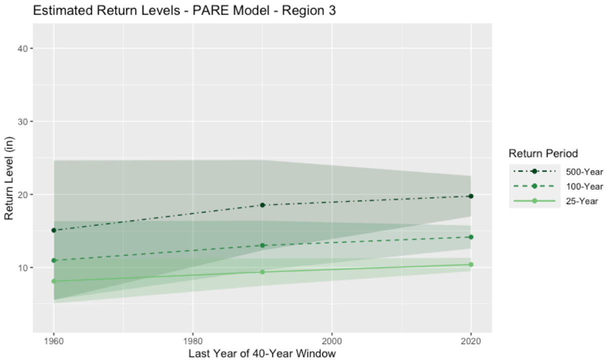
Method	Return Period	Region 1	Region 2	Region 3
1. PARE Model	25-Year	11.675 (0.689)	12.663 (0.793)	10.405 (0.462)
	100-Year	16.516 (1.246)	18.486 (1.487)	14.169 (0.800)
	500-Year	24.153 (2.312)	28.070 (2.878)	19.760 (1.412)
2. Block Kriging	25-Year	12.008 (2.266)	11.210 (2.058)	9.141 (1.443)
	100-Year	17.277 (4.485)	15.465 (3.891)	11.865 (2.545)
	500-Year	25.800 (8.969)	21.911 (7.364)	15.607 (4.437)
3. Regional Max	25-Year	12.959 (1.332)	14.682 (1.435)	13.753 (1.354)
	100-Year	17.541 (2.391)	19.676 (2.531)	18.370 (2.387)
	500-Year	24.227 (4.318)	26.828 (4.484)	24.937 (4.220)

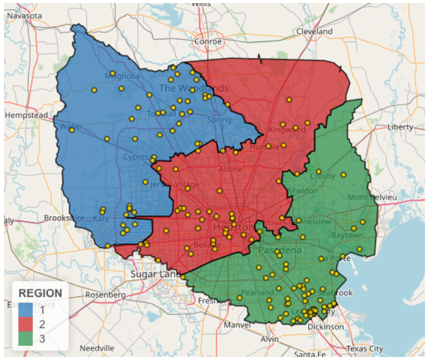
Table 5.3 : Comparison of return level estimates (in inches) across our proposed models for the last 40 years of data, 1981-2020. Standard error estimates are displayed in parentheses.

INCREASING RETURN LEVELS
(QUANTILES INCORPORATING EVENT FREQUENCY)









Technical summary

- The PARE model provides a solution for the spatial change of support problem when moving from points to expert-defined regions
- The PARE model performs best in simulations and on real data (according to error criteria studied)

Houston Flood takeaway

- Trends in return levels for some gauges
- Important that rainfall is modeled accurately within watersheds
- Inputs for flood models – output are predictions for Houston's future

Statistical Engineering in action

- Houston is being “rebuilt” and “re-designed”
  - **New RULES – build 3 ft above 500 year floodplain (return level)**
- Collaboration with key decision influencers
- *Answered the questions asked*
- *Answered on time*
- Improved estimates of rainfall return levels lead to improved floodplain modeling

Thank you!

Contact: Julia.schedler@rice.edu

ensor@rice.edu