Multivariate Spatial Extreme Value Analysis of Reconstructed Coastal Sea-Level Time-Series

Brian White[1] Richard L. Smith[2] Brian Blanton[3] Rick Luettich[4]

May 23, 2022

[1]Ph.D. Student|Department of Statistics & Operations Research|UNC

[2]Distinguished Professor|Department of Statistics & Operations Research|UNC

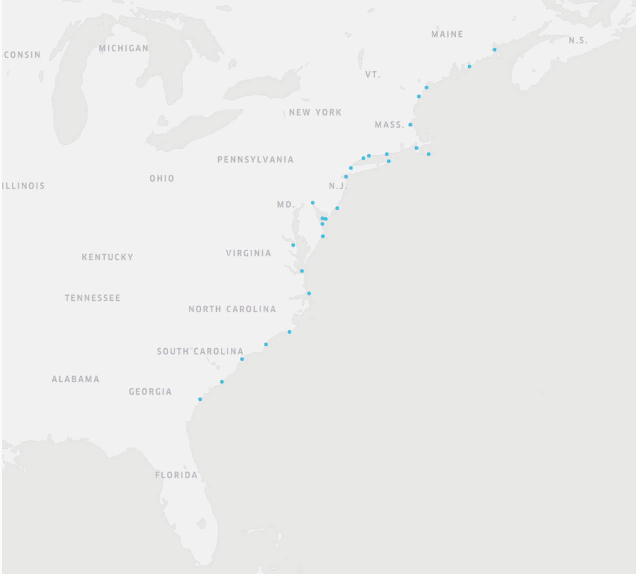
[3]Director of Environmental Initiatives|Renaissance Computing Institute

[4]Distinguished Professor|Department of Earth, Marine & Environmental Sciences|UNC

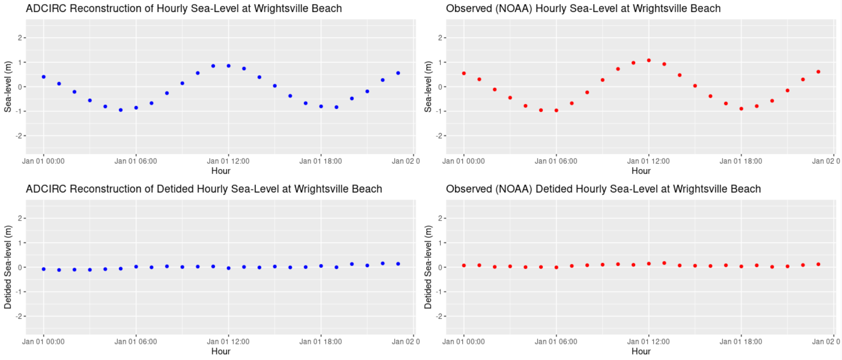
Introduction

1. Research aim: To understand how sea-level extremes along the U.S. East Coast have varied over space and time in the past and to make predictions for how these patterns will evolve in the future under different climate change projections.
2. Tools: Oceanography, climate-science, statistical methodology (i.e. extreme value theory. spatial statistics)
3. Data:
   (a) Hourly sea-level time-series taken from NOAA observation stations along the U.S. East Coast over a 40-year period.
   (b) Corresponding model-generated (ADCIRC) reconstruction of historic sea-level time-series.

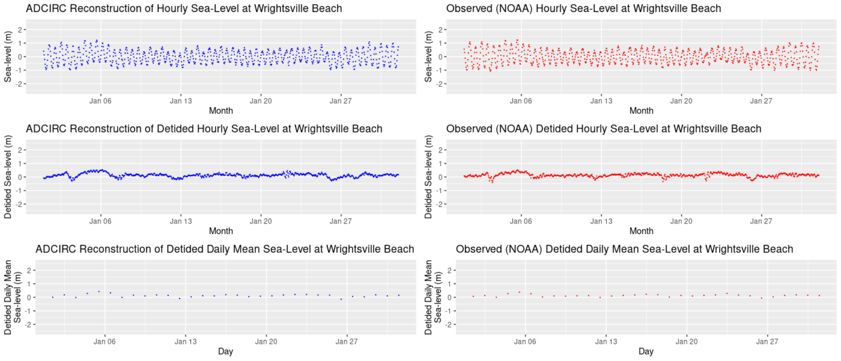
NOAA Sea-Level Observation Stations



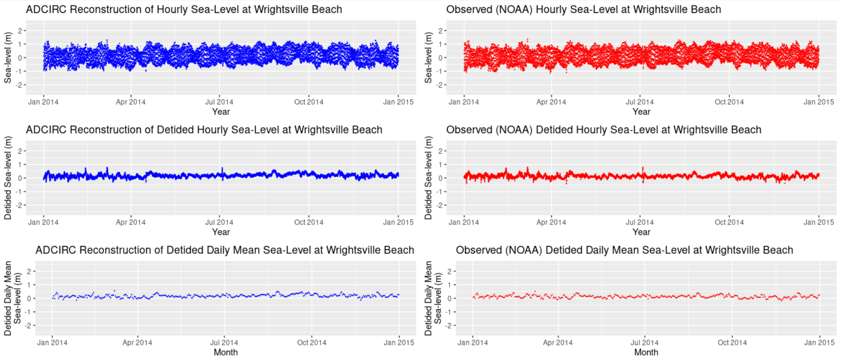
Exploratory Data Analysis



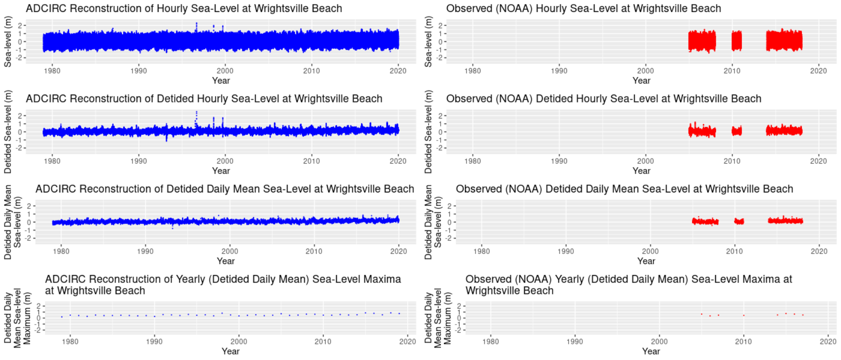
Exploratory Data Analysis



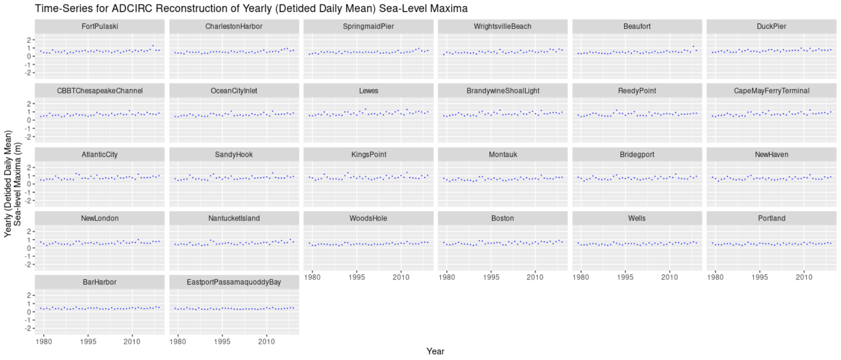
Exploratory Data Analysis



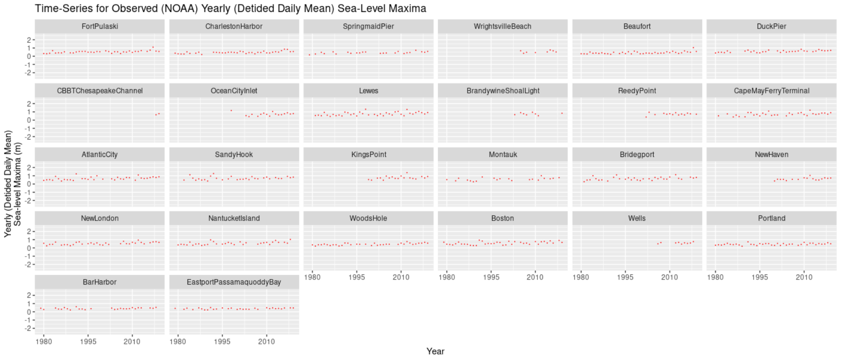
Exploratory Data Analysis



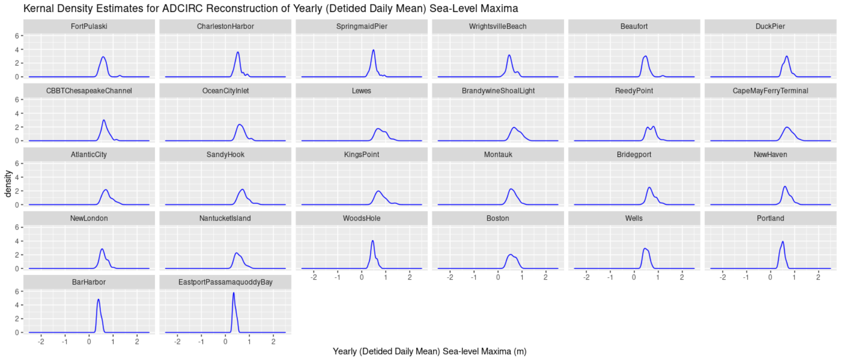
Exploratory Data Analysis



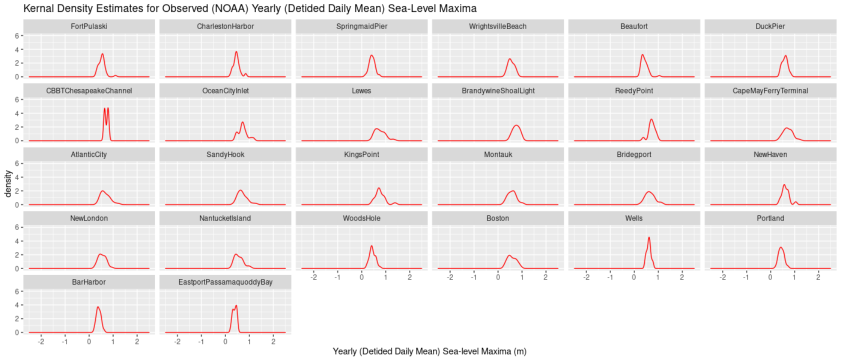
Exploratory Data Analysis



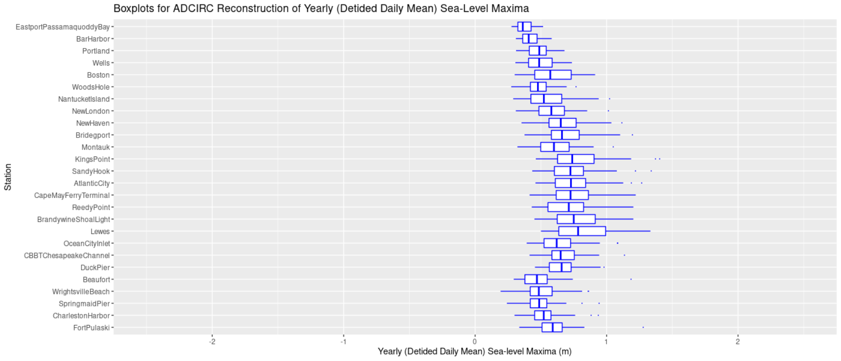
Exploratory Data Analysis



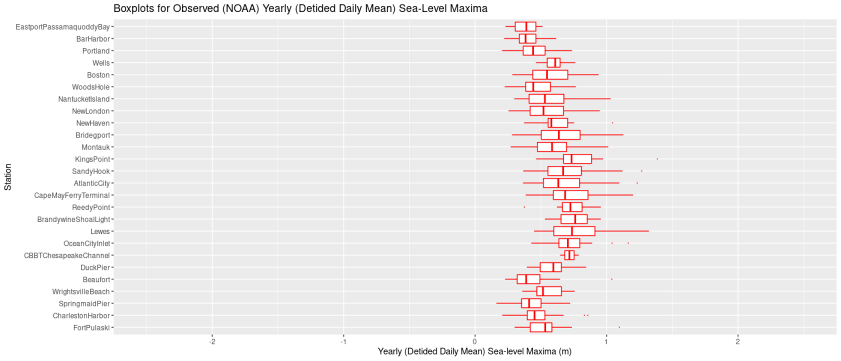
Exploratory Data Analysis



Exploratory Data Analysis



Exploratory Data Analysis



Generalized Extreme Value (GEV) Distribution

Let $X_1, X_2, ..., X_n$ be i.i.d. random variables with common CDF F and let $M_n := \max\{X_1, ..., X_n\}$.

Then, when appropriately centered and scaled, M_n converges in distribution to a member of the GEV family:

$$G(z) := \exp\left\{ -\left[1+\xi(\frac{z-\mu}{\sigma})\right]_+^{-\frac{1}{\xi}} \right\} \tag{1}$$

- $\mu \in \mathbb{R}$ is the location parameter
- $\sigma \in \mathbb{R}^+$ is the scale parameter
- $\xi \in \mathbb{R}$ is the shape parameter

r-Year Return Level

The GEV parameters can be used to characterize extremes via the quantile function of the GEV.

For $Z \sim GEV(\mu, \sigma, \xi)$, the 1-p quantile of Z (i.e. the value that is exceeded with probability p) is given by:

$$Z_p(\mu, \sigma, \xi) := \begin{cases} \mu - \frac{\sigma}{\xi}[1 - \{-\log(1-p)\}^{-\xi}] & \xi \neq 0 \\ \mu - \sigma \log\{-\log(1-p)\} & \xi = 0 \end{cases} \qquad (2)$$

When each year contains exactly one block the r-year return level is given by $Z_{r^{-1}}(\mu, \sigma, \xi)$

Inference

The GEV parameters are usually estimated by a likelihood based method.

In particular, for $Z_1, ..., Z_m \overset{iid}{\sim} GEV(\mu, \sigma, \xi)$ the maximum likelihood estimate (MLE) of $\theta := (\mu, \sigma, \xi)^t$ is:

$$\hat{\theta} := (\hat{\mu}, \hat{\sigma}, \hat{\xi})^t := \underset{\mu,\sigma,\xi}{argmax} \left\{ \ell(\mu, \sigma, \xi | Z_1, ..., Z_m) \right\} \tag{3}$$

where $\ell(\mu, \sigma, \xi | Z_1, ..., Z_m)$ is the log-likelihood of $Z_1, ..., Z_m$.

The r-year return level estimate is simply $Z_{r^{-1}}(\hat{\mu}, \hat{\sigma}, \hat{\xi})$.

Inference

From likelihood theory it follows that the distribution of $\hat{\theta}$ is approximately $\mathcal{N}_3(\theta, V)$ where V is the inverse of the observed information matrix evaluated at $\hat{\theta}$.

Therefore, the delta-method implies that the distribution of $Z_p(\hat{\theta})$ is approximately $\mathcal{N}_1\big(Z_p(\theta), \nabla Z_p(\hat{\theta})^t V \nabla Z_p(\hat{\theta})\big)$.

Modeling Questions

1. How do the GEV parameters and r-year return levels depend upon their spatial location?
2. How should the NOAA data be used to validate the ADCIRC reconstruction?
3. How should one incorporate global climatic information?

Methods: Overview

Idea (Russell et al. 2019): Multivariate spatial extreme value model fit by a 2-stage inference procedure.

1. a. Independently model the yearly (detided daily mean) sea-level maxima at each station using the GEV distribution.
   b. Perform inference via MLE.
2. a. Model the MLE output from stage 1 as a multi-dimensional Gaussian process with measurement error.
   b. Perform inference via MLE.

The output of stage 2 can then be used to spatially interpolate the GEV parameters and return-levels along the coastline via Kriging.

Let $Y(\mathsf{s})$ be the yearly (detided daily mean) maximum sea-level at location $\mathsf{s} \in \mathcal{D} \subset \mathbb{R}^2$ and assume

$$Y(\mathsf{s}) \sim GEV\big(\mu(\mathsf{s}), \sigma(\mathsf{s}), \xi(\mathsf{s})\big)$$

To characterize how the sea-level extremes vary spatially define, at both observed and unobserved locations, the latent Gaussian process

$$\boldsymbol{\theta}(\mathsf{s}) = \boldsymbol{\beta} + \boldsymbol{\eta}(\mathsf{s}) \tag{4}$$

for $\boldsymbol{\theta}(\mathsf{s}) := \big(\mu(\mathsf{s}), \log(\sigma(\mathsf{s})), \xi(\mathsf{s})\big)^t$.

Here $\boldsymbol{\beta}$ is a vector of mean parameter values over $\mathcal{D}$ and $\boldsymbol{\eta}(\mathsf{s})$ a vector of spatially correlated random effects.

The spatially correlated random effects are defined by the relation

$$\boldsymbol{\eta}(\mathsf{s}) := A\boldsymbol{\delta}(\mathsf{s}) \tag{5}$$

where A is a lower-triangular matrix and $\boldsymbol{\delta}(\mathsf{s})$ is a vector of independent second-order stationary Gaussian processes with mean 0 and covariance function

$$Cov\big(\delta_i(\mathsf{s}), \delta_i(\mathsf{s'})\big) = \exp\left(\frac{-||\mathsf{s} - \mathsf{s'}||}{\rho_i}\right) \tag{6}$$

for $\mathsf{s}, \mathsf{s'} \in \mathcal{D}$ where $\rho_i > 0$ is the range parameter.

Methods: Latent Process with Measurement Error

For NOAA station $l \in \{1, \ldots, 26\}$, let $\hat{\boldsymbol{\theta}}(\mathsf{s}_l)$ be the point-wise MLE for the GEV distribution associated with $Y(\mathsf{s}_l)$. We assume that

$$\hat{\boldsymbol{\theta}}(\mathsf{s}_l) = \boldsymbol{\theta}(\mathsf{s}_l) + \epsilon(\mathsf{s}_l) \tag{7}$$

where $\epsilon(\mathsf{s}_l)$ is estimation error that is independent of $\boldsymbol{\eta}$.

Thus, the latent process with measurement error at station l is

$$\begin{aligned} \hat{\boldsymbol{\theta}}(\mathsf{s}_l) &= \boldsymbol{\theta}(\mathsf{s}_l) + \epsilon(\mathsf{s}_l) \\ &= \boldsymbol{\beta} + \boldsymbol{\eta}(\mathsf{s}_l) + \epsilon(\mathsf{s}_l) \\ &= \boldsymbol{\beta} + A\boldsymbol{\delta}(\mathsf{s}_l) + \epsilon(\mathsf{s}_l) \end{aligned} \tag{8}$$

Now, let

$$\boldsymbol{\Theta} := \left(\boldsymbol{\theta}(\mathsf{s}_1), ..., \boldsymbol{\theta}(\mathsf{s}_{26})\right)^t \tag{9}$$

and

$$\hat{\boldsymbol{\Theta}} := \left(\hat{\boldsymbol{\theta}}(\mathsf{s}_1), ..., \hat{\boldsymbol{\theta}}(\mathsf{s}_{26})\right)^t \tag{10}$$

where

$$Cov\left(\hat{\boldsymbol{\Theta}}\right) := \boldsymbol{\Sigma}_{\boldsymbol{\rho},A}.$$

Methods: Latent Process with Measurement Error

Further, assume that

$$\boldsymbol{\epsilon} := \left(\epsilon(\mathbf{s}_1), \ldots, \epsilon(\mathbf{s}_{26})\right)^t \sim \mathcal{N}_{78}(\mathbf{0}, W) \tag{11}$$

where W is unknown and estimated via a regularized non-parametric bootstrap procedure:

$$W_{tap} := W_{bs} \circ T_{tap}(\lambda) \tag{12}$$

where W_{bs} is the non-parametric bootstrap estimate of W and $T_{tap}(\lambda)$ is a taper matrix with range parameter $\lambda > 0$

Methods: Latent Process with Measurement Error

The taper matrix is defined by the relation

$$W_{tap} := C_{W_2}(\lambda) \otimes (\mathbf{1}_3 \mathbf{1}_3^t) \tag{13}$$

where $C_{W_2}(\lambda)$ is a matrix whose entries are computed using the Wendland 2 covariance function:

$$[C_{W_2}(\lambda)]_{ij} := \begin{cases} (1 - \frac{||\mathbf{s}_i - \mathbf{s}_j||}{\lambda})^6 (\frac{35}{3} (\frac{||\mathbf{s}_i - \mathbf{s}_j||}{\lambda})^2 + 6(\frac{||\mathbf{s}_i - \mathbf{s}_j||}{\lambda}) + 1 & ||\mathbf{s}_i - \mathbf{s}_j|| \leq 0 \\ 0 & ||\mathbf{s}_i - \mathbf{s}_j|| > 0 \end{cases}$$

Thus,

$$\hat{\mathbf{\Theta}} = \mathbf{\Theta} + \boldsymbol{\epsilon} \tag{14}$$

and hence

$$\hat{\mathbf{\Theta}} \sim \mathcal{N}_{78}\left(\mathbf{1}_{26} \otimes \boldsymbol{\beta}, \mathbf{\Sigma}_{A,\boldsymbol{\rho}} + W_{tap}\right) \tag{15}$$

Therefore, given $\hat{\mathbf{\Theta}}$ (i.e. the output from the 1st stage of inference) and W_{tap}, we can obtain $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\rho}}$ and $\hat{A}$ via MLE.

Methods: Kriging (Gaussian Process Regression)

Given $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\rho}}$ and $\hat{A}$ (i.e. the output from the 2nd stage of inference) we then interpolate $\hat{\boldsymbol{\theta}}(\mathsf{s})$ over $\mathsf{s} \in \mathcal{D}$ via Kriging:

$$\hat{\boldsymbol{\theta}}(\mathsf{s}) = \hat{\boldsymbol{\beta}} + Cov\big(\hat{\boldsymbol{\theta}}(\mathsf{s}), \hat{\boldsymbol{\Theta}}\big)(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\rho}},\hat{A}} + W_{tap})^{-1}(\hat{\boldsymbol{\Theta}} - \mathbf{1}_{26} \otimes \hat{\boldsymbol{\beta}})$$

$$Var\big(\hat{\boldsymbol{\theta}}(\mathsf{s})\big) = Var\big(\hat{\boldsymbol{\theta}}(\mathsf{s})\big) - Cov\big(\hat{\boldsymbol{\theta}}(\mathsf{s}), \hat{\boldsymbol{\Theta}}\big)(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\rho}},\hat{A}} + W_{tap})^{-1}Cov\big(\hat{\boldsymbol{\theta}}(\mathsf{s}), \hat{\boldsymbol{\Theta}}\big)^{t}$$

and compute the 100-Year Return Level estimates for the Yearly (Detided Daily Mean) Sea Level Maxima:

$$Z_{100^{-1}}(\mathsf{s}) = Z_{100^{-1}}(\hat{\boldsymbol{\theta}}(\mathsf{s}))$$

$$Var\big(Z_{100^{-1}}(\mathsf{s})\big) = \nabla Z_p(\hat{\boldsymbol{\theta}}(\mathsf{s}))^t Var\big(\hat{\boldsymbol{\theta}}(\mathsf{s})\big) \nabla Z_p(\hat{\boldsymbol{\theta}}(\mathsf{s}))$$

Preliminary Results: Stage 1 Output (ADCIRC)

GEV Parameter Estimate at Each Station (ADCIRC)

station	*location*	*log.scale*	*shape*
FortPulaski	0.53	-2.14	0.00
CharlestonHarbor	0.47	-2.20	-0.05
SpringmaidPier	0.44	-2.18	-0.08
WrightsvilleBeach	0.45	-2.00	-0.14
Beaufort	0.42	-2.34	0.13
DuckPier	0.61	-2.15	-0.15
CBBTChesapeakeChannel	0.60	-2.10	-0.05
OceanCityInlet	0.58	-2.02	-0.03
Lewes	0.72	-1.79	-0.03
BrandywineShoalLight	0.68	-1.81	-0.11
ReedyPoint	0.63	-1.93	-0.08
CapeMayFerryTerminal	0.67	-1.80	-0.13
AtlanticCity	0.67	-1.89	0.02

Preliminary Results: Stage 1 Output (ADCIRC)

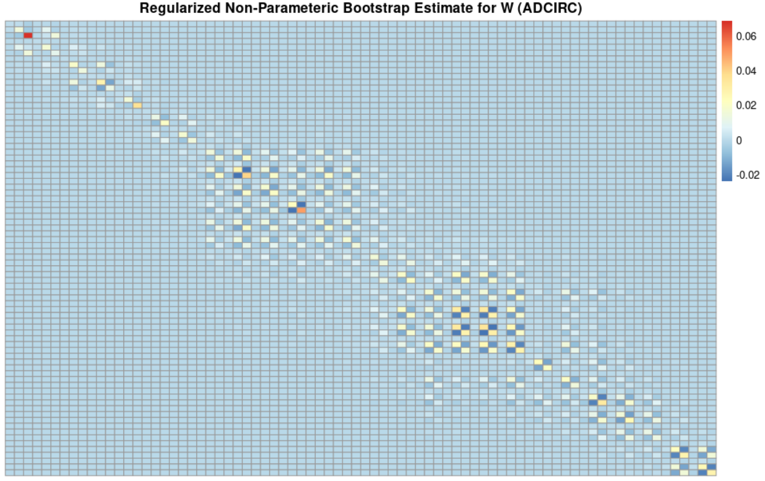
GEV Parameter Estimates at Each Station (ADCIRC)

station	*location*	*log.scale*	*shape*
SandyHook	0.66	-1.88	0.00
KingsPoint	0.69	-1.84	0.06
Montauk	0.54	-1.94	-0.13
Bridegport	0.63	-1.94	-0.06
NewHaven	0.61	-1.99	-0.09
NewLondon	0.54	-2.07	-0.10
NantucketIsland	0.49	-2.02	0.00
WoodsHole	0.44	-2.33	-0.14
Boston	0.53	-1.89	-0.25
Wells	0.46	-2.29	-0.21
Portland	0.45	-2.43	-0.26
BarHarbor	0.39	-2.78	-0.07
EastportPassamaquoddyBay	0.35	-2.92	-0.04

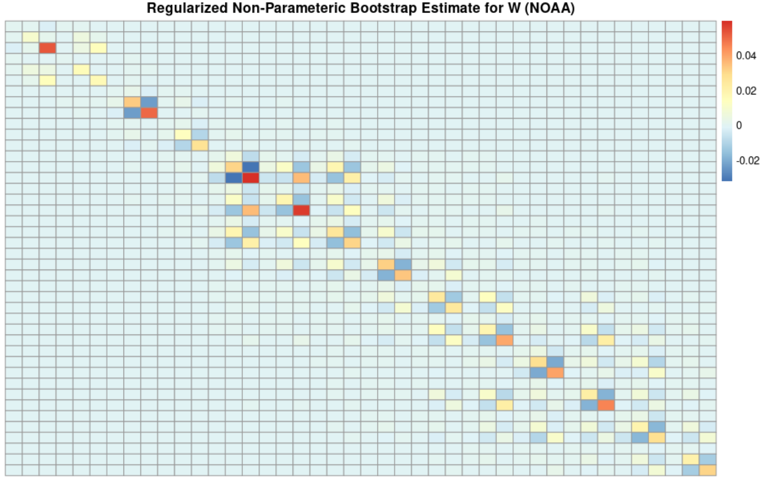
Preliminary Results: Stage 1 Output (NOAA)

GEV Parameter Estimates at Each Station (NOAA)			
station	*location*	*log.scale*	*shape*
FortPulaski	0.467	-2.159	-0.031
CharlestonHarbor	0.405	-2.140	-0.087
Beaufort	0.356	-2.406	0.156
DuckPier	0.549	-2.197	-0.248
Lewes	0.660	-1.820	0.055
CapeMayFerryTerminal	0.642	-1.744	-0.167
AtlanticCity	0.582	-1.892	0.032
SandyHook	0.611	-1.857	-0.010
Bridegport	0.576	-1.738	-0.126
NewLondon	0.481	-1.910	-0.182
NantucketIsland	0.473	-1.987	0.075
WoodsHole	0.400	-2.181	-0.166
Boston	0.508	-1.783	-0.216
Portland	0.403	-2.238	-0.190

Preliminary Resutls: Stage 1 Output (ADCIRC)



Preliminary Resutls: Stage 1 Output (NOAA)



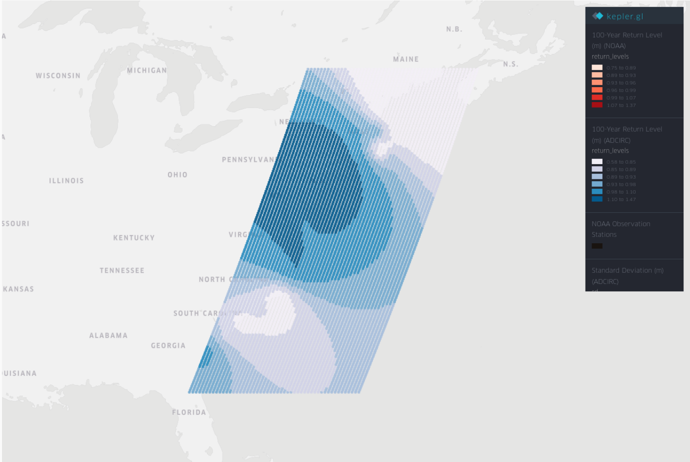
Preliminary Results: Stage 2 Output (ADCIRC)

Gaussian Process Parameter Estimates (ADCIRC)				
beta	rho	A		
0.50	396.91	0.10	0.00	0
-2.27	545.96	0.19	-0.10	0
-0.08	945.83	0.06	-0.02	0

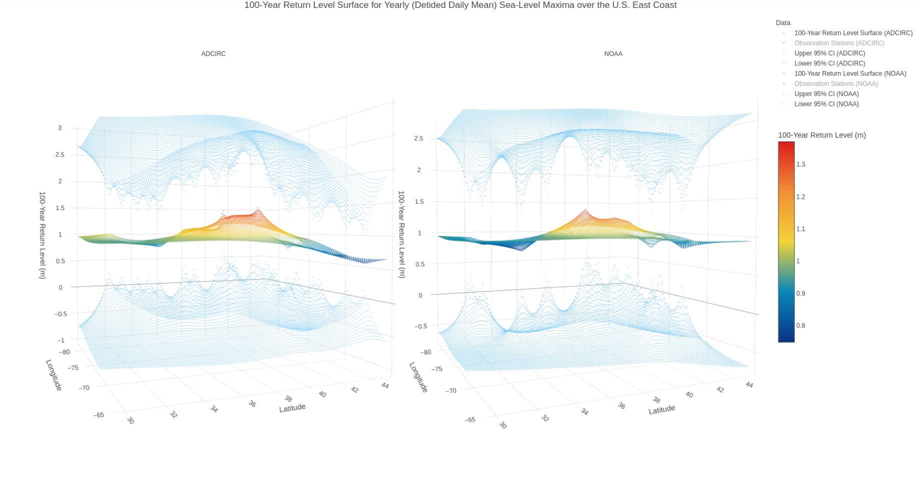
Preliminary Results: Stage 2 Output (NOAA)

Gaussian Process Parameter Estimates (NOAA)				
beta	rho		A	
0.48	228.23	0.09	0.00	0
-2.09	490.30	0.18	-0.07	0
-0.08	809.65	0.01	0.00	0

Preliminary Results: 100-Year Return Level Heat Map for Yearly (Detided Daily Mean) Sea-Level Maxima



Preliminary Results: 100-Year Return Level Surface for Yearly (Detided Daily Mean) Sea-Level Maxima



Preliminary Results: 100-Year Return Level Surface for Yearly (Detided Daily Mean) Sea-Level Maxima

Next Steps & Future Work

Next Steps

1. Use the 100-year return level surface based on the NOAA data to improve the performance of the corresponding surface based on the ADCIRC reconstruction.

Future Work

1. Introduce global climatic covariate(s) in the 1st stage of inference.
2. Examine how the r-year return-level surface along the coastline changes as as a function of these covariates.

References

1. An Introduction to Statistical Modeling of Extreme Values. Coles. Springer Series in Statistics Springer-Verlag, London, 2001.
2. Investigating the association between late spring Gulf of Mexico sea surface temperatures and US Gulf Coast precipitation extremes with focus on Hurricane Harvey. Russel, Riser, Smith and Kunkel. Environmetrics 31(1), 2019.