Evaluating the Sensitivity of Flood Return Levels to Data Length by Maximum Likelihood Estimation

Alexandra Klufas[1], Tony Wong[2], Klaus Keller[3]

aklufas@wellesley.edu

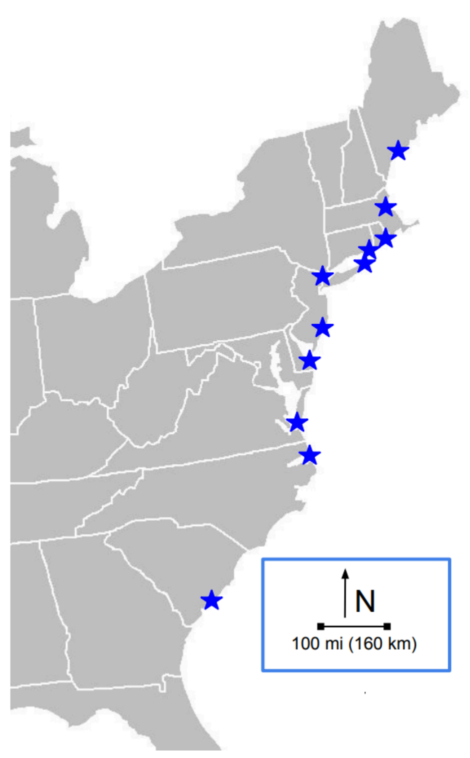
Wellesley College[1]
Earth and Environmental Systems Institute[2], Geoscience Department[3], Penn State University
Department of Engineering and Public Policy[3], Carnegie Mellon University

1. Introduction

- Across the East Coast of the United States, flooding is a persistent threat to coastal communities
- With a predicted increase in sea level over the next century, risk analysis risk of coastal flooding is important to assessing possible future economic loss
- Extreme value theory is a way to model the statistics of extreme events, such as coastal flooding
- Non-stationary modeling of storm surge levels incorporates changing temperature
    - These models are commonly dismissed because they predict larger intervals of future storm surge

2. Methods

2.1 Observational Data



★ Tide Stations Used in Study

Tide Station	Observational Record (Years)	Number of Years Used
Atlantic City, NJ	103	91
Boston	93	91
Charleston, SC	93	89
Chesapeake Bay, VA	39	39
Duck Pier, NC	36	33
Lewes, DE	57	53
Montauk, NY	55	38
Newport, RI	84	90
New London, CT	76	66
New York City	94	68
Portland, ME	104	90

Tide Stations Used in Study

2.2 Statistical Modeling of Extreme Storm Surges

- Generalized Extreme Value (GEV) Distribution
    - Limiting distribution for a series of block maxima
    - Three parameters influence density and distribution functions
        - μ – the location of the mean of the distribution curve
        - σ – the width of the distribution curve
        - ξ – the thickness of the tails of the distributions curve
- Density function

$$f(x \mid \mu, \sigma, \xi) = 1/\sigma\, z(x)^{\xi+1} e^{-z(x)} \qquad z(x) = \begin{cases} \left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ e^{-\frac{x-\mu}{\sigma}} & \text{if } \xi = 0 \end{cases}$$

- Non-Stationarity
    - Increasing global temperatures will increase the number of stronger storms which include larger storm surges
    - We focus on the upper tail of the GEV, the probability that a flood will overtop the effective height
    - We use New London, CT, as a test case
        - In this community, floods over 2.85 m cause damage

$$\mu(t) = \mu_0 + \mu_1 T(t)$$
$$\sigma(t) = e^{\sigma_0 + \sigma_1 T(t)}$$
$$\xi(t) = \xi_0 + \xi_1 T(t)$$

Model Name	Non-Stationary Parameters	Estimated Parameters
ST	None	μ_0, σ_0, ξ_0
NS1	μ	$\mu_0, \mu_1, \sigma_0, \xi_0$
NS2	μ, σ	$\mu_0, \mu_1, \sigma_0, \sigma_1, \xi_0$
NS3	μ, σ, ξ	$\mu_0, \mu_1, \sigma_0, \sigma_1, \xi_0, \xi_1$

Candidate Models. Our candidate models provide a range of free parameters for analysis. Using these models, we study how much data is needed to estimate storm surge heights.

2. Methods (cont.)

2.3 Bayesian Calibration

- We use the likelihood function,

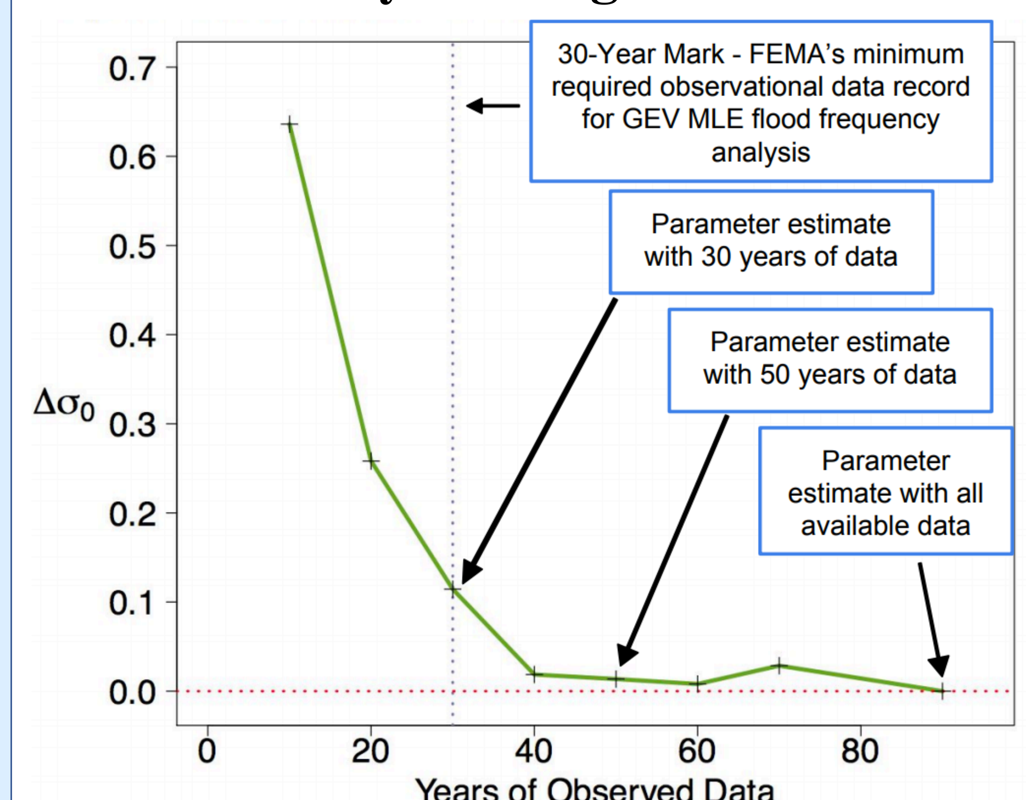
$$f(x^i \mid \mu^i, \sigma^i, \xi^i) = \prod_{t=1}^{n} 1/\sigma^i\, z(x^i)^{\xi^i+1} e^{-z(x^i)}$$

to quantify the goodness of fit between the i-th annual block maximum, x^i, and the GEV distribution given by the parameters, μ^i, σ^i, and ξ^i

- We use a differential optimization algorithm to estimate the maximum likelihood parameters (MLEs) and examine how these estimates change with length of data
- We use an adaptive Markov chain Monte Carlo (MCMC) Bayesian calibration to obtain posterior distributions of each parameter

3. Results

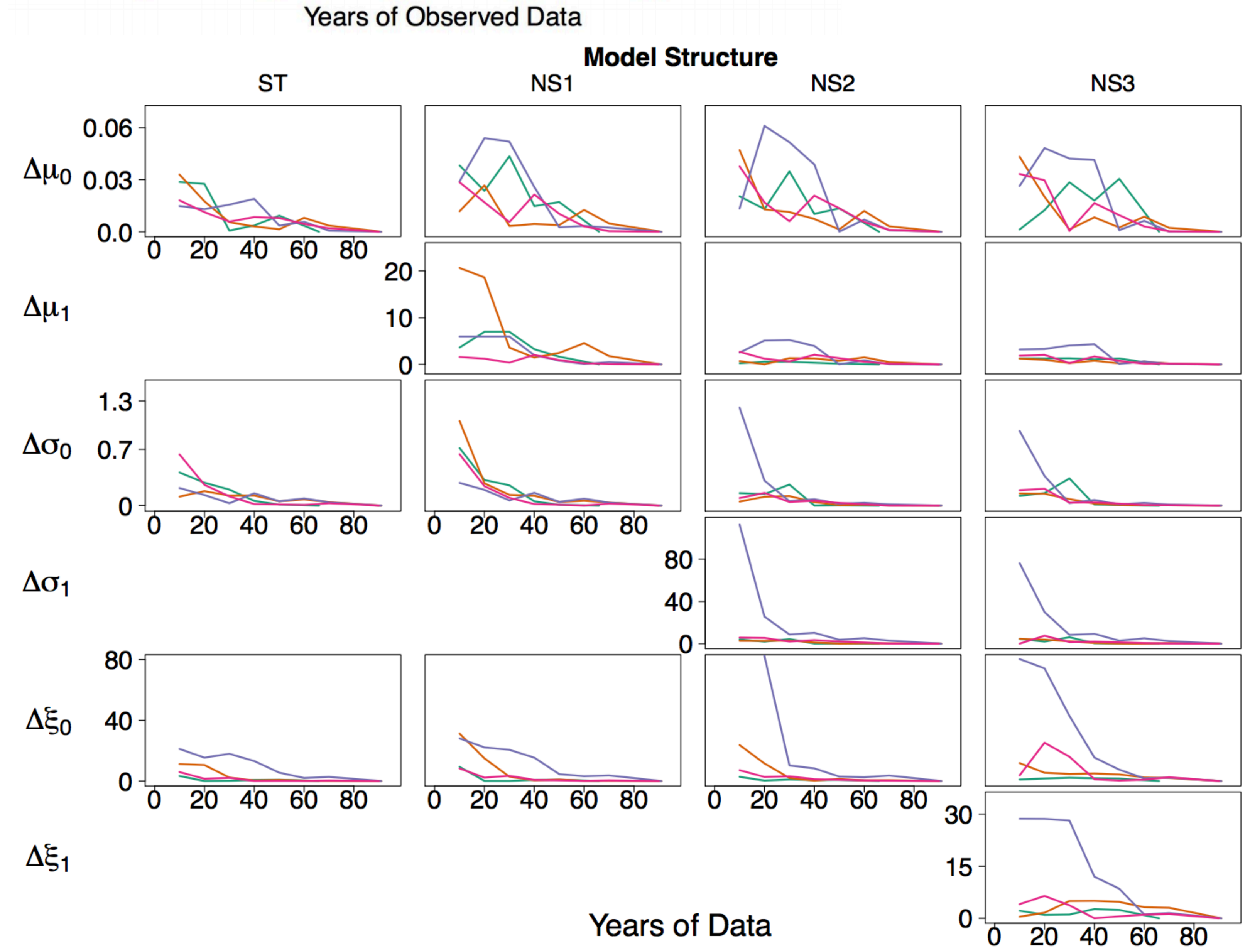
3.1 Sensitivity to Length of Observational Record

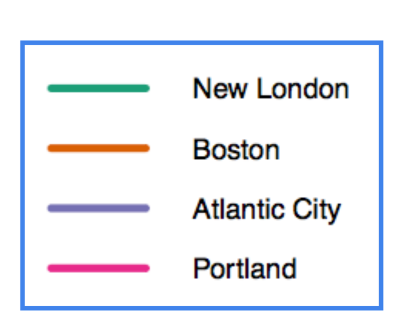


- In order to quantify how well shorter data sets represent the full data set, we use:

$$\Delta\sigma_0 = \left| \frac{\sigma_0^i - \sigma_0}{\sigma_0} \right|$$

- The closer to zero $\Delta\sigma_0$ is, the more closely the parameter estimate represents the parameter estimate of the entire data set





Stabilization of Parameter Estimates with Increase in Data Length. Indicates minimum length of tide record needed to reliably use MLEs for flood risk analysis for each candidate model. ST stabilizes with less data than the other candidate models. ξ is the most difficult parameter to estimate with shorter data records.

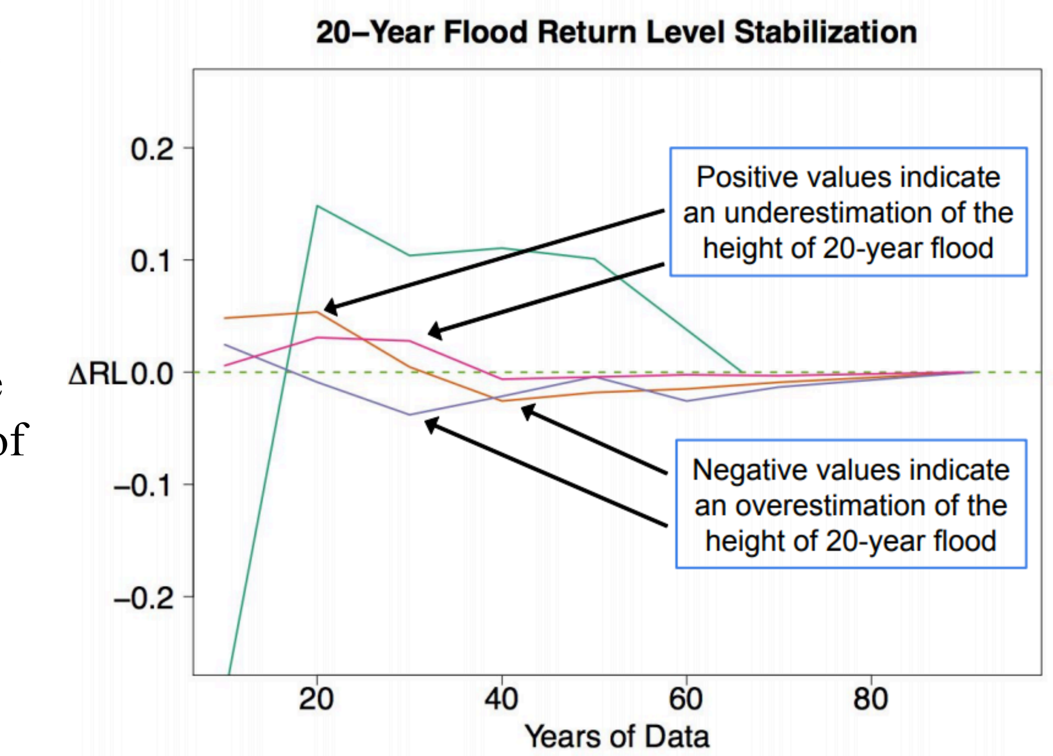
3. Results (cont.)

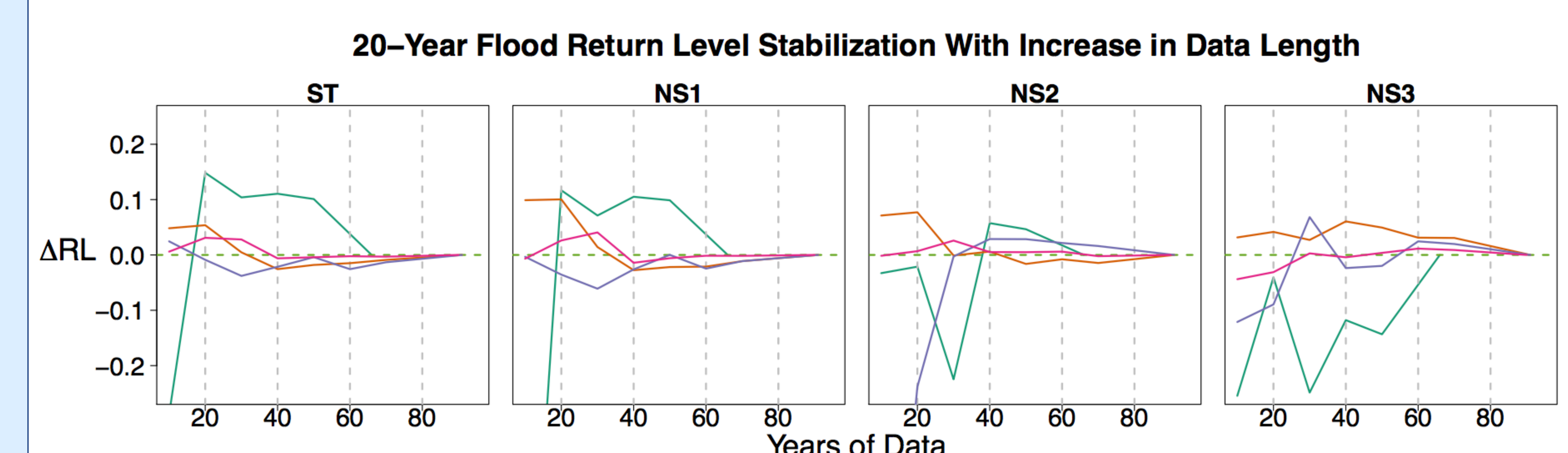
3.2 Storm Surge Projections

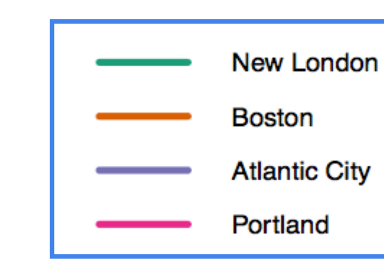
- We quantify the impacts of data length on return level estimates (ΔRL):

$$\Delta RL = \frac{RL - RL^i}{RL^i}$$

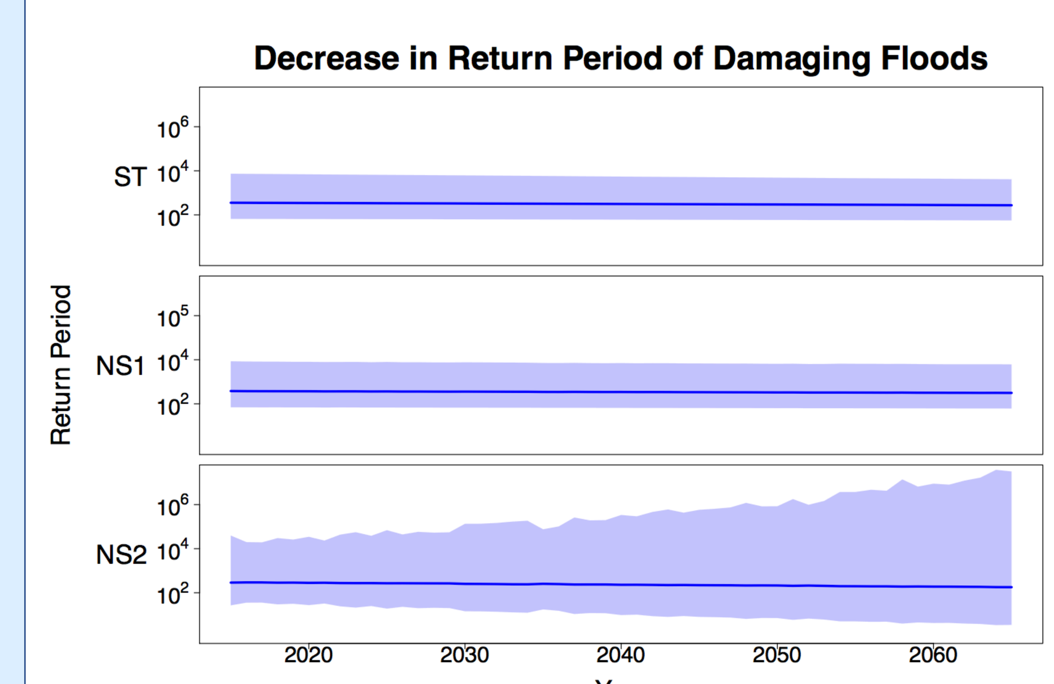
- Offers practical guidance regarding the reliability of shorter lengths of data at different stations







Comparison of the 20-Year Flood Return Level with Increase in Data Length. Each model shows that about 60 years of data are needed for reliable estimation of 20-year return levels.



Expected number of years between damaging floods in New London, CT. The blue line represents the median of the return period. The shaded region represents the 90% credible interval of the return period.

4. Discussion

- With more free parameters in a model, a record of more than 60 years is recommended for flood frequency analysis
- Damaging floods may increase with increase in global temperature in New London, CT and uncertainty of the frequency of these floods may also increase
    - Mitigating the increase of global temperatures can reduce the predicted increase of future severity of storm surge

Acknowledgements

This project was supported by the National Science Foundation Research Experience for Undergraduates (REU) Program (AGS-1560339) and through the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507.