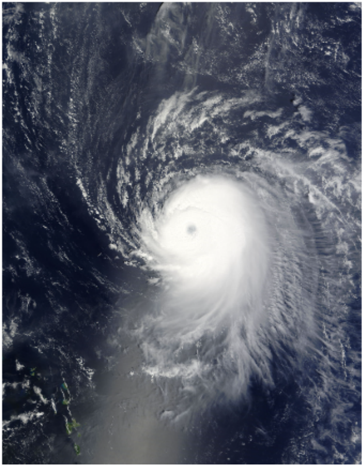
A Combined Physical-Statistical Approach for Estimating Storm Surge Risk

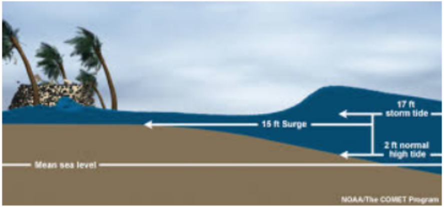
CLEMSON UNIVERSITY

wkhuang@clemson.edu

Math 4500, February 28, 2023

Hurricanes, Storm Surges, and Flooding





Source: NASA (left); National Oceanic and Atmospheric Administration (right)

Quantifying Storm Surge Risk

- Flood insurance: The Federal Emergency Management Agency (FEMA) requires estimates of the magnitude of surges in terms of 10-, 50-, 100-, and 500-year return levels for coastal areas to determine insurance rates, infrastructure design, and emergency planning

- Flood mitigation: To assess whether coastal nuclear plants meet 1 in 10,000 year flood protection criteria.

U.S. Department of Homeland Security
500 C Street, SW
Washington, DC 20472

FEMA

March 22, 2012

Operating Guidance No. 8-12
For use by FEMA staff and Flood Hazard Mapping Partners

Title:	Joint Probability – Optimal Sampling Method for Tropical Storm Surge Frequency Analysis
Effective Date:	March 22, 2012
Approval:	Luis Rodriguez Branch Chief, Engineering Management Branch Risk Analysis Division Federal Insurance and Mitigation Administration

U.S.NRC
UNITED STATES NUCLEAR REGULATORY COMMISSION
Protecting People and the Environment

JAPAN LESSONS-LEARNED PROJECT DIRECTORATE

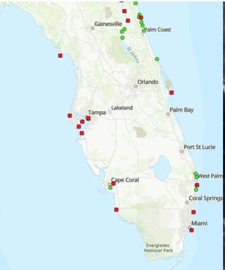
JLD-ISG-2012-06

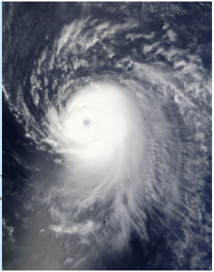
Guidance for Performing a Tsunami, Surge, or Seiche Hazard Assessment

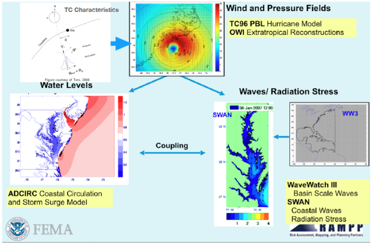
Interim Staff Guidance
Revision 0

Quantifying Storm Surge Risk: Data Sources

Data \ Variable	y ("Output")	x ("input")
Observation	Observed storm surges $\Rightarrow$ Very limited in space and time	Storm characteristics $\Rightarrow$ Limited but well observed
Simulation	Simulated storm surge responses	"Synthetic" storm characteristics







Courtesy of Gangai (Dewberry) & Danforth (FEMA)

Model Input: Hurricane Characteristics

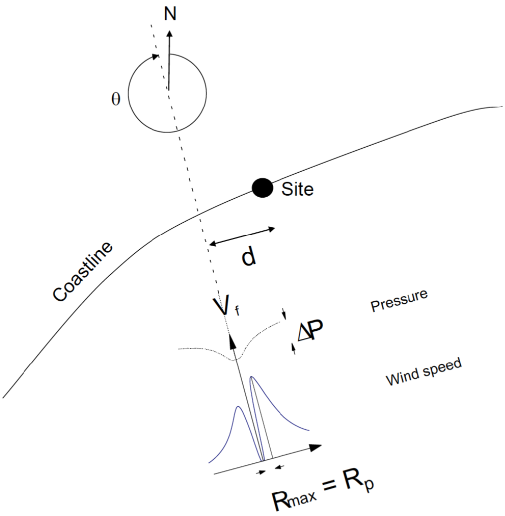
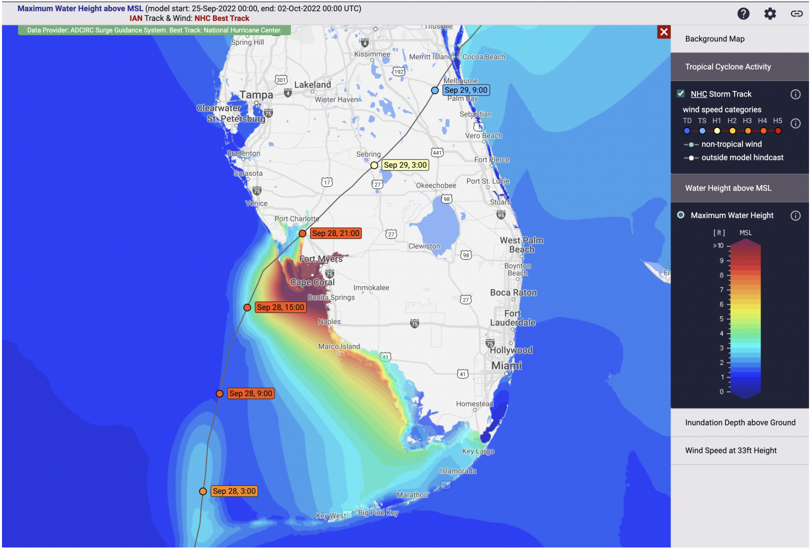


Figure courtesy of Toro, 2008

Model Output: Simulated Storm Surge Field



Source: Coastal Emergency Risks Assessment (CERA) Link: https://cera.coastalrisk.live/cerarisk/

Estimating Extreme Surges: Physical-Statistical Approach

$$\boldsymbol{x} \in \mathcal{X} \qquad \eta : \mathcal{X} \mapsto \mathcal{Y} \qquad y = \eta(\boldsymbol{x})$$

$$\text{Input} \longrightarrow \text{Model} \longrightarrow \text{Output}$$

TC Characteristics

- Records are more complete than surge levels
- Input to simulate storm surge levels

Task: Estimating $f(\boldsymbol{x})$
Density estimation

Computer Model

- Simulate high fidelity surge response
- computationally extensive

Task: Estimating $\eta(\boldsymbol{x})$
Gaussian process

Surge Level

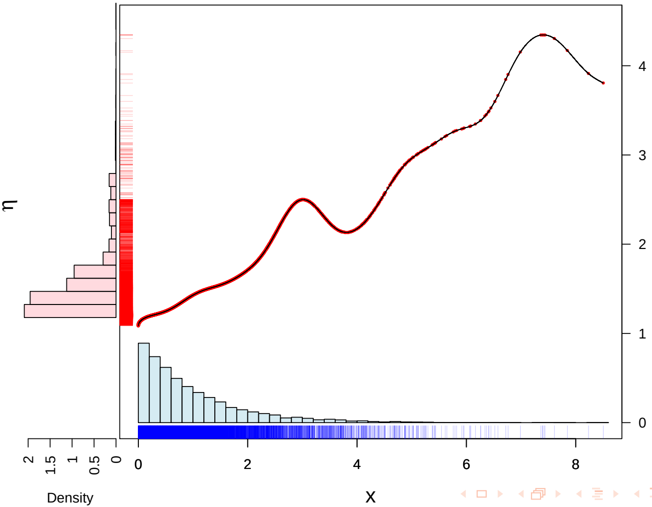
- simulate synthetic storms
- generate surge response for risk analysis

Task: Estimating y_r
Extreme value analysis

A 1-D toy example: $Y = \eta(X)$

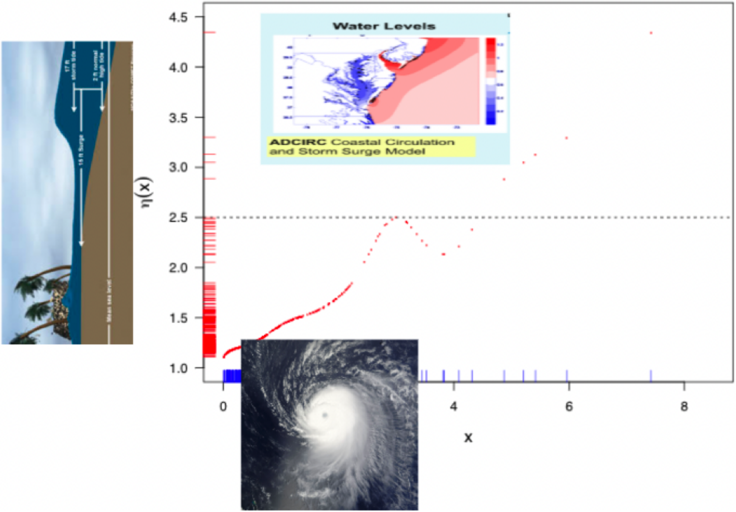
Ideal situation:

1) $f(x)$ is given and 2) can run the model many times



Storm Surge Application

Real situation: 1) limited observations to infer $f(x)$ and 2) computationally expensive to run the model

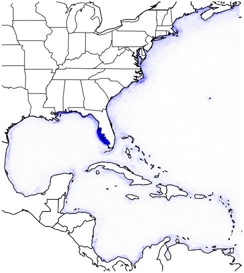


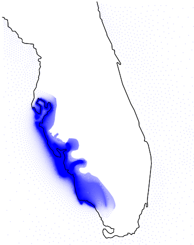
Next, two studies will be presented:

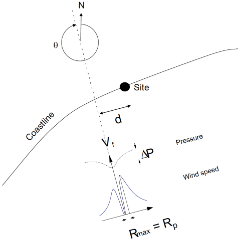
1. emulating storm-wise maximum response $\eta_{max}(\boldsymbol{x})$

2. estimating r-year return level y_r at a give location

Southwest Florida Storm Surge Case Study

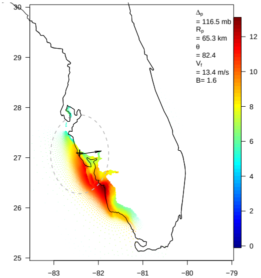


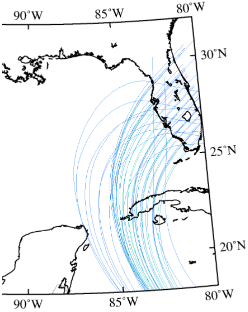


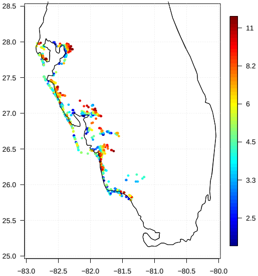


- Peak surge of 3,156 "synthetic storms" via ADvanced CIRCulation model (ADCRIC, Luettich & Westerink, 1992)
- These synthetic storms are parameterized by some storm characteristics $\boldsymbol{x}$
- 148,055 spatial nodes ($\sim 120,000$ in/near study region)

Surge Response: y_{sim}



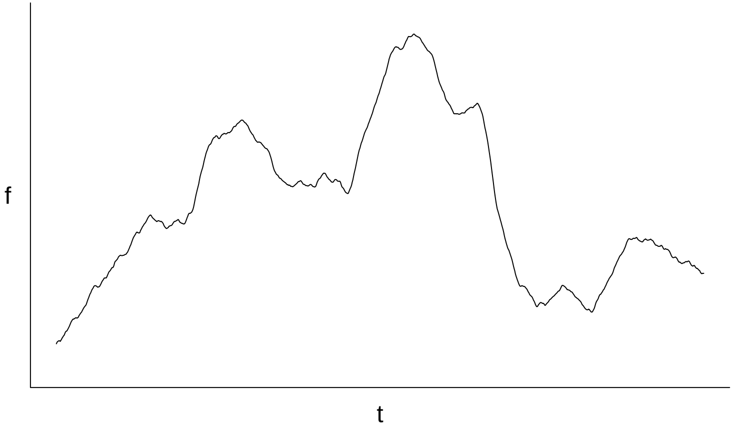




Goal: to emulate the **storm-wise maximum surge** (Note: one could also emulate the spatial location where the storm-wise maximum occur)

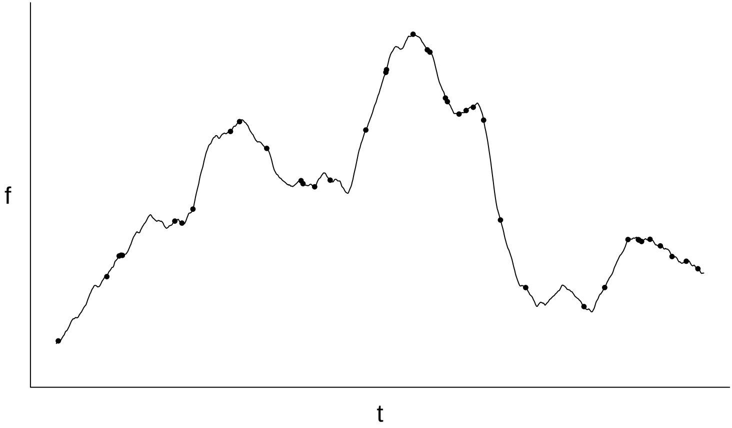
Gaussian Processes

Function Estimation



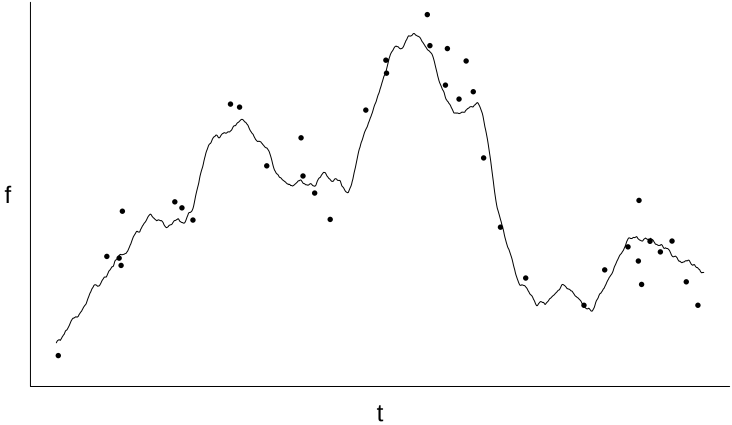
Consider a function $f(t), t \in \mathcal{T}$

Function Estimation



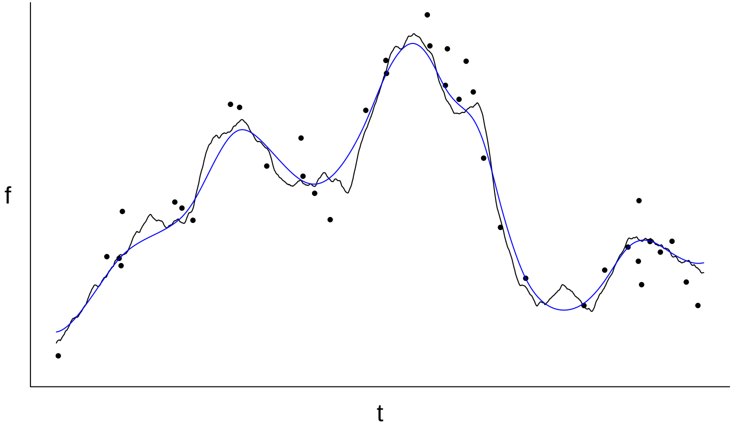
Consider $f(t)$, observed incompletely $\{f(t_i)\}_{i=1}^{n}$

Function Estimation



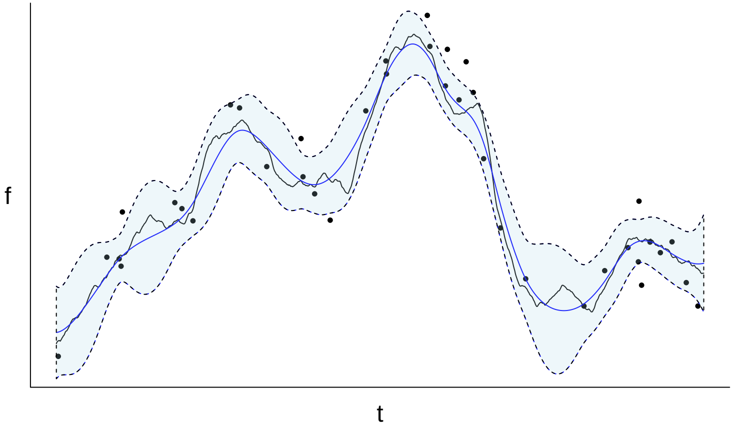
Consider $f(t)$, observed incompletely $\{f(t_i)\}_{i=1}^{n}$, and with noise $\{\varepsilon_i\}_{i=1}^{n}$

Gaussian Processes (GPs): Function Estimators



Main idea: exploit the inter-point correlation to estimate $f(t)$

GPs: Probabilistic Function Estimators



GP provides an "optimal" estimate $\hat{f}(t)$ along with "localized" uncertainty quantification ("error bars")

GP Emulation Setup

$$y = \eta(\boldsymbol{x}) + \epsilon, \quad \eta(\cdot) \sim \mathsf{GP}(m(\cdot), K_\phi(\cdot,\cdot)), \quad \epsilon \sim N(0, \tau^2)$$

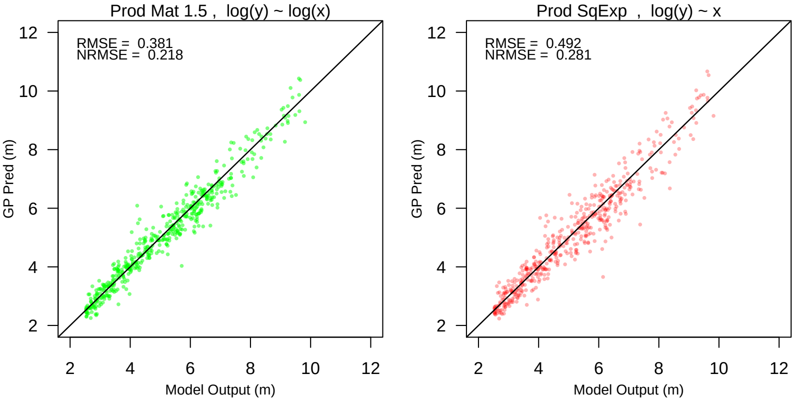
- Constant mean, (product) exponential (`Exp`), Matérn $\nu = 1.5$ (`Mat32`), Matérn $\nu = 2.5$ (`Mat52`), or squared-exponential (`SqExp`) covariance functions.
- Explore the use of logarithm transform: $(y \sim x)$, $(\log(y) \sim x)$, $(y \sim \log(x))$, $(\log(y) \sim \log(x))$
- Use normalized root mean squared prediction error (NRMSPE) to assess out-of-sample prediction performance:

$$\text{NRMSPE} = \frac{\text{RMSPE}}{\text{RMSPE}_0},$$

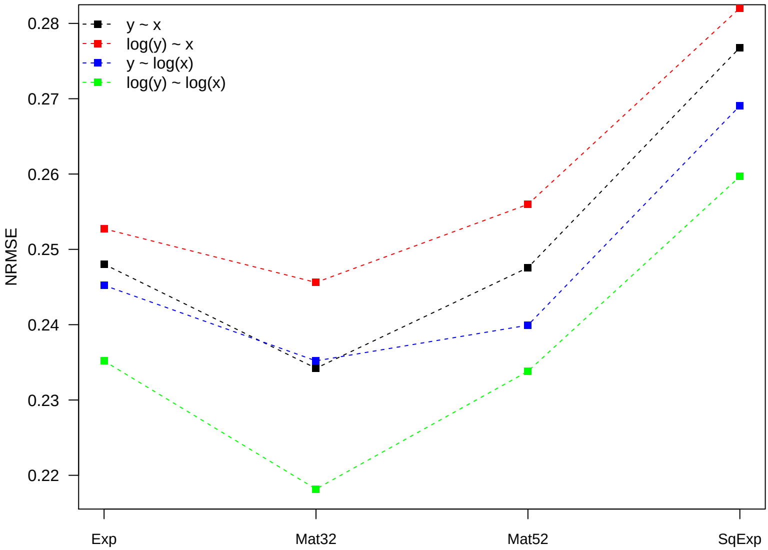
where $\text{RMSPE}_0 = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N} (\bar{y} - y^{(i)})^2}$

Out-Of-Sample Prediction

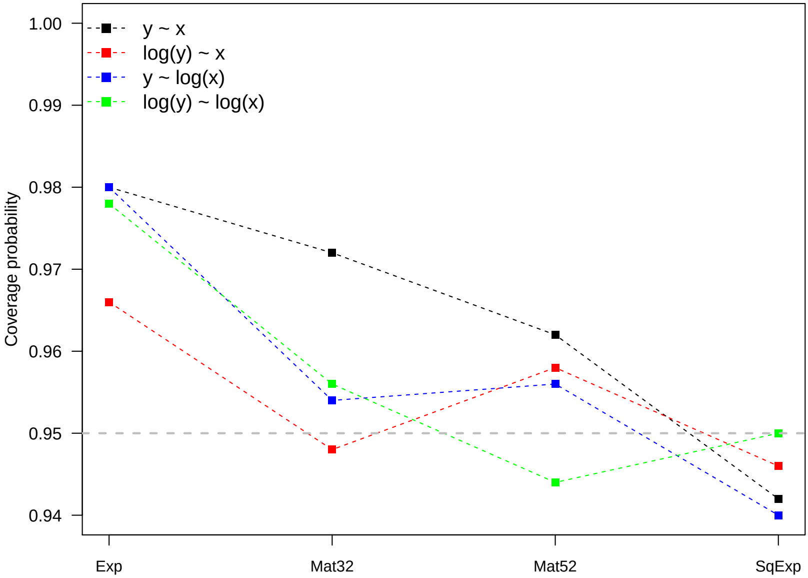
- Focus on $2\mathrm{m} \leq y \leq 12\mathrm{m}$
- $N = 1458,\ N_{test} = 500$



Normalized Root Mean Squared Prediction Error (NRMSE)



Empirical Coverage Probabilities



Part I Summary & Discussion

Emulating storm-wise maximum: covaraince function matters, taking logs of both sides help

Next, I will present a beta version of storm surge return level estimation

$x \in \mathcal{X}$ $\qquad$ $\eta : \mathcal{X} \mapsto \mathcal{Y}$ $\qquad$ $y = \eta(x)$

Input → Model → Output

where we need to estimate $f(x)$, the joint distribution of input (i.e., storm characterise)

Part I Summary & Discussion

Emulating storm-wise maximum: covaraince function matters, taking logs of both sides help

Next, I will present a beta version of storm surge return level estimation

$$\boldsymbol{x} \in \mathcal{X} \qquad \eta : \mathcal{X} \mapsto \mathcal{Y} \qquad y = \eta(\boldsymbol{x})$$

Input $\longrightarrow$ Model $\longrightarrow$ Output

where we need to estimate $f(\boldsymbol{x})$, the joint distribution of input (i.e., storm characterise)

Estimating Input Distribution $f(x)$: Data

- IBTrACS contains global tropical cyclone best-track data from 1851 - 2018 [Knapp et al., 2010]
- Only included "strong" storms (those with pressure deficit $\geq$ 13mb) from 1950[1] that made landfall within 100 km of the coastline of the study region[2] $\Rightarrow$ 28 storms were included in out Southwest Florida case study

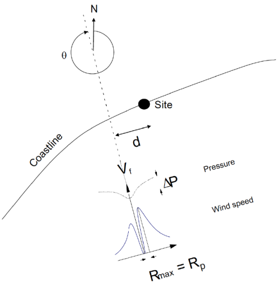
Pressure deficit Δp

Radius to max wind speed R_{max}

Forward speed V_f

Heading angle θ

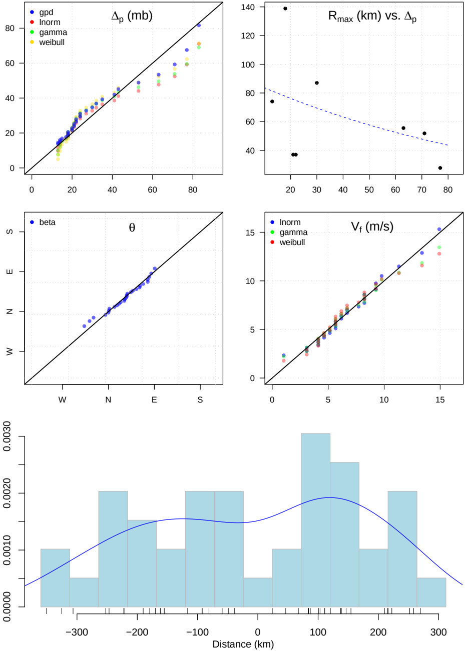
Distance from landfall d



[1] due to data quality issue

[2] surge levels generated by the excluded storms are negligible

Estimating Input Distribution: Model & Results



- Estimating $f(\boldsymbol{x}) = f(\Delta p, R_{max}, V_f, \theta, d)$ based on $\{\boldsymbol{x}_{o,i}\}_{i=1}^{28}$
  - Δp: generalized Pareto
  - $R_{max}|\Delta p$: log-normal
  - V_f: log-normal
  - θ: beta (shifted and re-scaled to $[0, 1]$)
- Distance d was not described well by a parametric distribution. We preform a kernel density estimation

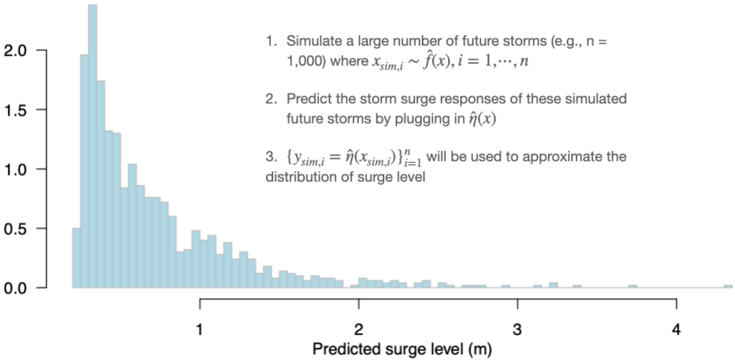
Estimating $\eta(\cdot)$: GP Emulator

- Constant mean function $m(\boldsymbol{x}) = \mu$ and product covariance function
$$K_\phi(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \prod_{i=1}^{p} C_i(d(x_i, x_i'); \phi_i),$$
where Matérn covariance with $\nu = 0.5$ (i.e., exponential) $\nu = 1.5$, $\nu = 2.5$, and $\nu = \infty$ (i.e., squared-exponential) were used

- Explore the use of logarithm transformation: $(y \sim x)$, $(\log(y) \sim x)$, $(y \sim \log(x))$, $(\log(y) \sim \log(x))$

Generating Synthetic Storms and their Surge Levels

Combining $\hat{f}(\boldsymbol{x})$ and $\hat{\eta}(\boldsymbol{x})$ to generate more storm events

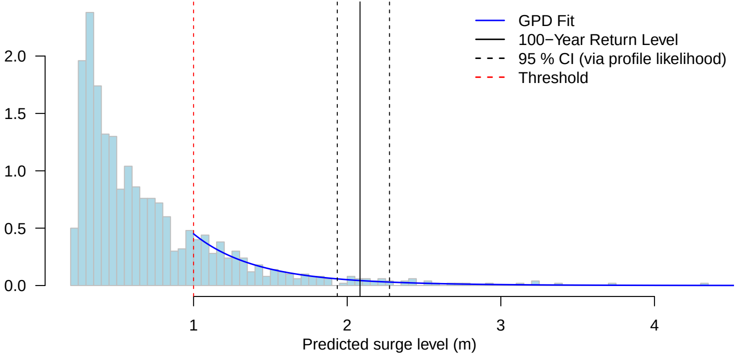


We are going to use these synthetic surge responses to estimate extreme surges

Estimating extreme surges: extreme value analysis

We employed the peaks-over-threshold method [Davison and Smith, 1990] to estimate the r-year return levels

- Assuming upper tail follow a generalized Pareto distribution (GPD)
- Using profile likelihood method to construct confidence interval (CI), which gives asymmetric interval



Part II Summary & Discussion

Recap:

- Historical storm observations were used to estimate $f(\boldsymbol{x}) \Rightarrow$ estimating the observed "input" rather than "output"
- Applied GP to emulate the $\eta(\boldsymbol{x})$ implied by the computer model $\Rightarrow$ covaraince matters, taking logs of both sides help
- Used $\hat{f}(\boldsymbol{x})$ to simulate more storms, and GP to predict surge levels, which were be used to estimate r-year storm surge levels

Limitations:

- Limited modeling for $f(\boldsymbol{x}) \Rightarrow$ "borrow strenght" across space/data sources?
- Did not account for estimation uncertainty for $\hat{\eta}(\boldsymbol{x})$
- **A comprehensive sensitivity analysis is needed to assess the effects of various modeling choices**