Some Thoughts on Quantifying Storm Surge Risk

Whitney Huang
Joint work with SAMSI Storm Surge Working Group

JSM 2020 August 2-6
VIRTUAL CONFERENCE
Everyone Counts: Data for the Public Good

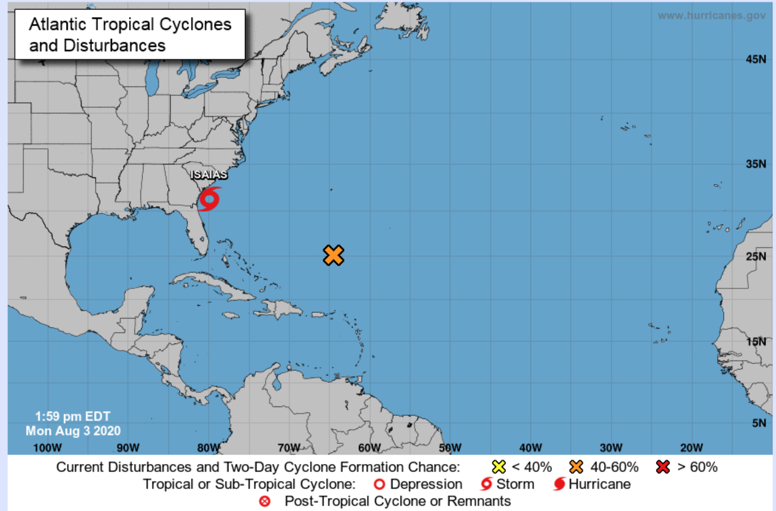
Uncertainty Quantification for Environmental Applications
August 6, 2020

CLEMSON UNIVERSITY
samsi NSF Duke NCSU UNC
CANSSI INCASS
COASTAL RESILIENCE CENTER

SAMSI Storm Surge Working Group

- Domain experts: **Taylor Asher**, Rick Luettich, Brian Blanton, Wei Mei (UNC), Michelle Bensi (UMD), Alex Taflanidis (Notre Dame), Gabriel Toro (Lettis Consultants International), Ben Timmermans, Christina Patricola (LBL)
- Emulation: Won Chang, Bledar Konomi (UCincinnati), Pulong Ma (SAMSI/Duke), Matthew Plumlee (Northwestern), William Welch (UBC), Alexi Rodrguez-Arelis (UBC)
- Input modeling: Veronica Berrocal (UMich), Qiuyi Wu (URochester), Anirban Mondal, Kai Yin (CWRU)
- Tail estimation: Richard Smith (UNC)
- SAMSI Volcanic Hazards Group: Bruce Pitman (SUNY Buffalo), Elaine Spiller (Marquette), Robert Wolpert (Duke), James Berger (Duke)

Hurricane ISAIAS



Hurricane, storm surge, and flooding

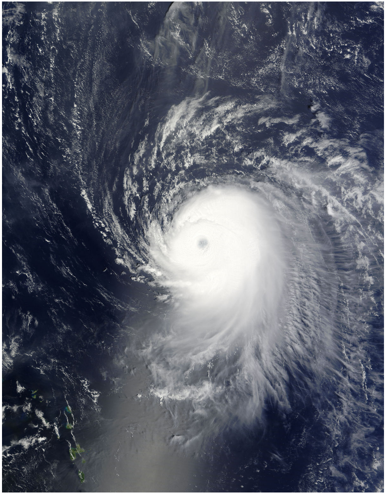


Figure: Source: NASA

Figure: Courtesy of NOAA

The problem: Storm surge height estimation

Applications

- Flood insurance: The Federal Emergency Management Agency (FEMA) requires an estimation of the magnitude of surges in terms of 10-, 50-, 100-, and 500-year return levels for coastal areas to determine insurance rates
- Flood mitigation: To assess whether coastal nuclear plants meet 1 in 10,000 year flood protection criteria.

One could use tide gauge observations to estimate, say, 100-year return level at a given location. But
1) Limited data coverage; 2) Hurricane are rare

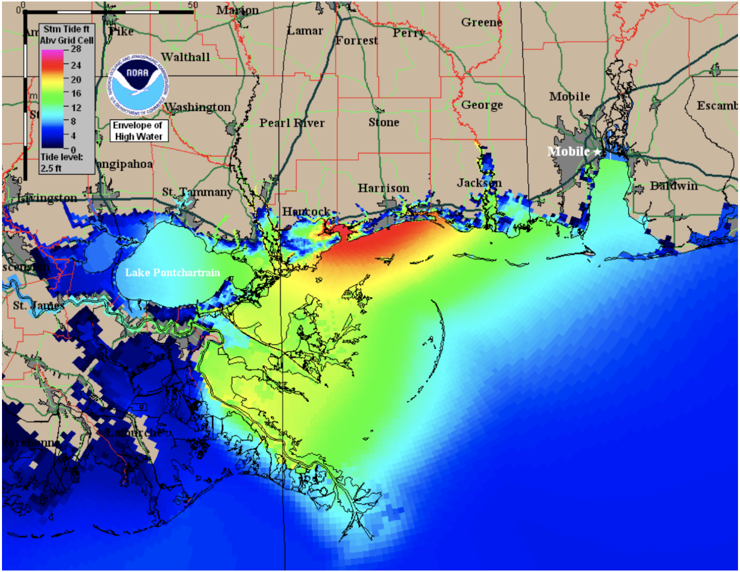
The problem: Storm surge height estimation

Applications

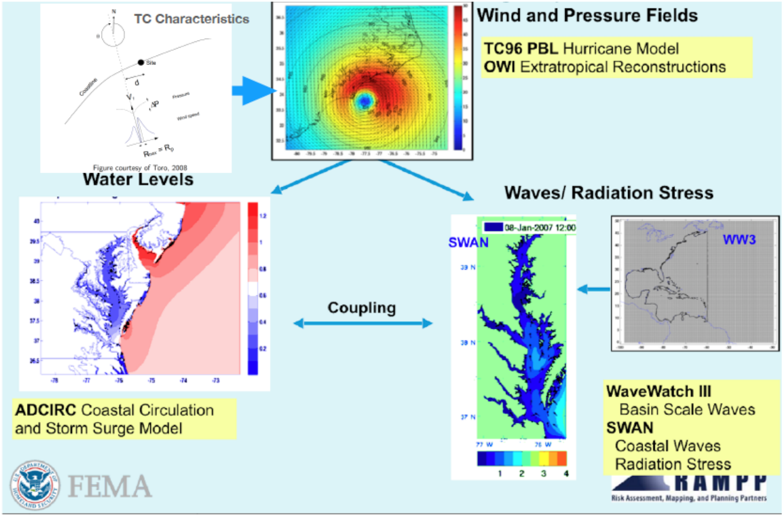
- Flood insurance: The Federal Emergency Management Agency (FEMA) requires an estimation of the magnitude of surges in terms of 10-, 50-, 100-, and 500-year return levels for coastal areas to determine insurance rates
- Flood mitigation: To assess whether coastal nuclear plants meet 1 in 10,000 year flood protection criteria.

One could use tide gauge observations to estimate, say, 100-year return level at a given location. But
1) Limited data coverage; 2) Hurricane are rare

This is what National Hurricane Center (NHC) will use



Modeling storm surge using computer simulations



Courtesy of Gangai (Dewberry) & Danforth (FEMA)

Workflow for Quantifying Storm Surge Risk

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

Input → Model → Output

TC Characteristics

- limited data
- sparse in space and time

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

Computer Model

- deterministic
- computationally extensive

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

Surge Level

- focus on extremes
- r-year return level y_r

Task: Estimating y_r

> Requires proper propagation of estimation uncertainties to obtain a sensible statistical inference for y_r

Estimating $\eta(\boldsymbol{x})$

- **Goal**: Build a statistical emulator based on (limited) compute model runs $\{\boldsymbol{x}_i, y_i = \eta(\boldsymbol{x}_i)\}_{i=1}^N$ to infer $\eta(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}$, the surge response function, especially for large surge values

- **Strategy**: **assume** $\eta(\boldsymbol{x})$ is a realization of a stochastic process, typically a Gaussian Process (GP), that interpolates $\{\boldsymbol{x}_i, y_i = \eta(\boldsymbol{x}_i)\}_{i=1}^N$.

Next, I will present a quick (and incomplete) case study

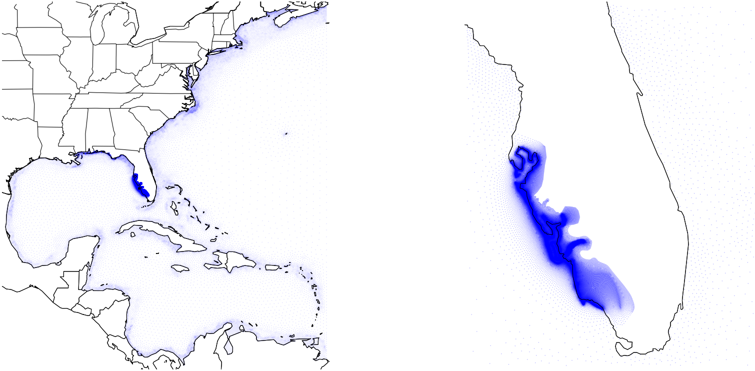
Estimating $\eta(\boldsymbol{x})$

- **Goal**: Build a statistical emulator based on (limited) compute model runs $\{\boldsymbol{x}_i, y_i = \eta(\boldsymbol{x}_i)\}_{i=1}^N$ to infer $\eta(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{X}$, the surge response function, especially for large surge values

- **Strategy**: **assume** $\eta(\boldsymbol{x})$ is a realization of a stochastic process, typically a Gaussian Process (GP), that interpolates $\{\boldsymbol{x}_i, y_i = \eta(\boldsymbol{x}_i)\}_{i=1}^N$.

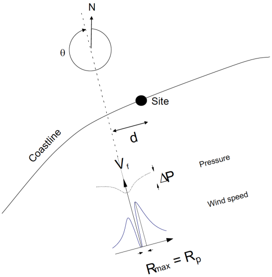
Next, I will present a quick (and incomplete) case study

Southwest Florida storm surge case study



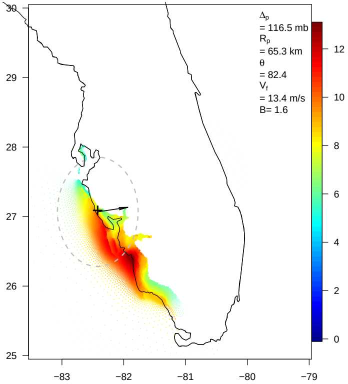
- 148,055 nodes ($\sim 120,000$ in/near study region)
- Peak surge of 3,156 storms (200 input configurations)

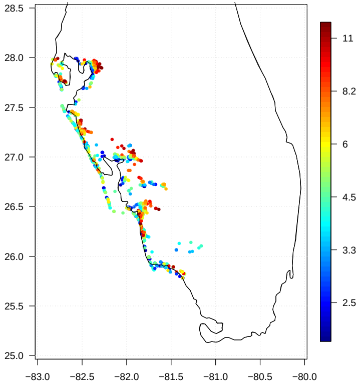
Model input




```
> summary(input_par[, 3:9])
  dp.mb.coast.     rp.nmi.offshore.    Vf.m.s.          theta.deg.             B
 Min.   : 13.71   Min.   : 5.985   Min.   : 1.049   Min.   :  5.125   Min.   :0.515
 1st Qu.: 41.70   1st Qu.:14.589   1st Qu.: 4.490   1st Qu.: 35.208   1st Qu.:0.880
 Median : 67.93   Median :30.190   Median : 8.521   Median : 69.846   Median :1.270
 Mean   : 69.15   Mean   :34.759   Mean   : 8.419   Mean   : 64.430   Mean   :1.256
 3rd Qu.: 98.00   3rd Qu.:54.136   3rd Qu.:12.325   3rd Qu.: 92.700   3rd Qu.:1.613
 Max.   :122.71   Max.   :80.362   Max.   :14.970   Max.   :119.930   Max.   :1.984
  X_Land.deg.      Y_Land.deg.
 Min.   :-83.87   Min.   :25.00
 1st Qu.:-82.76   1st Qu.:26.17
 Median :-82.43   Median :27.08
 Mean   :-82.31   Mean   :27.25
 3rd Qu.:-81.83   3rd Qu.:28.15
 Max.   :-81.06   Max.   :30.00
```

Surge response





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

Emulation setup

- Constant mean, (product) exponential (`Exp`), Matérn $\nu = 1.5$ (`Mat32`), Matérn $\nu = 2.5$ (`Mat52`), or squared-exponential (`SqExp`) covariance functions. Include (and estimate) nugget effect.

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

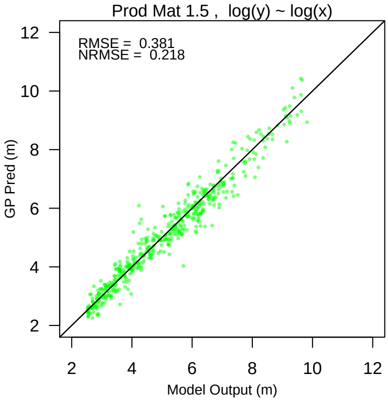
- Use normalized root mean squared prediction error (NRMSPE) to assess out-of-sample prediction performance:

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

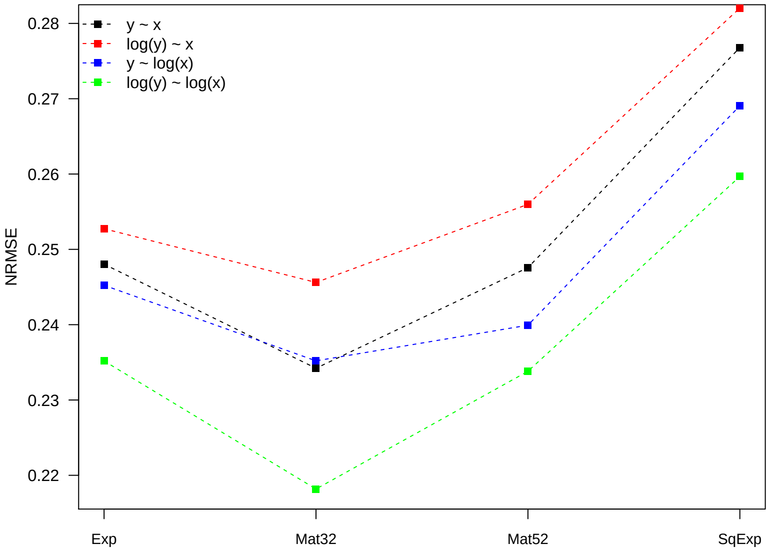
where $\mathrm{RMSPE}_0 = \sqrt{\frac{1}{N_{test}} \sum_{i=1}^{N} (\bar{y} - y^{(i)})^2}$

Out-of-sample prediction

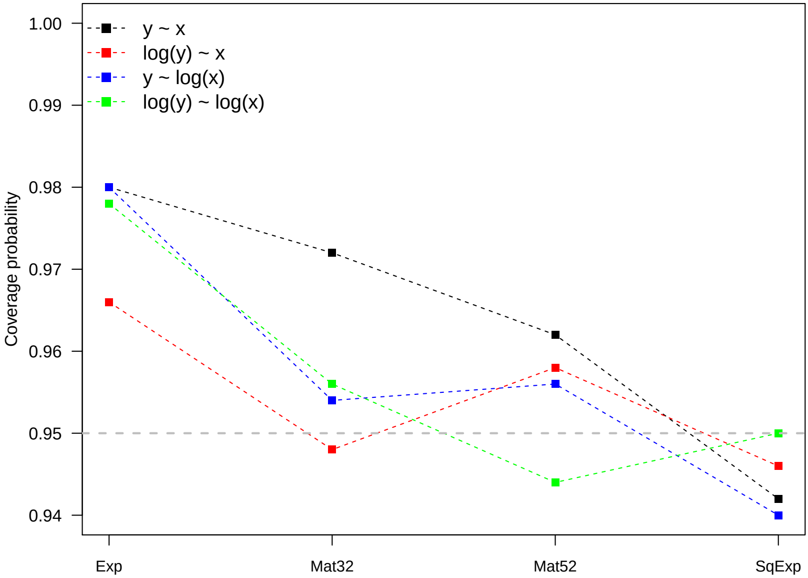
- Focus on $2\text{m} \leq y \leq 12\text{m}$
- $N = 1458,\ N_{test} = 500$



Normalized Root Mean Squared Prediction Error (NRMSE)



Empirical coverage probabilities



Summary and discussion

- Combining Physical and Statistical modeling in the context of storm surges risk modeling
- Emulating storm-wise maximum
  - Nugget
  - Covaraince function matters, taking logs of both sides help
- Things not Addressed
  - Estimating $f(\boldsymbol{x})$, the joint distribution of input (i.e., TC characterise)
  - Make use of Dimensional Analysis to build a better emulator; emulating the whole surge field; nonstationarity; non-GP emulator
  - Other sources of uncertainty: e.g. wind field parameterization.

SAMSI

Our storm surge working group would not have been possible without the support from SAMSI. In particular the following programs:

- Model Uncertainty: Mathematical and Statistical (MUMS) 2018-2019
- Mathematical and Statistical Methods for Climate and the Earth System (CLIM) 2017-2018
- Development, Assessment and Utilization of Complex Computer Models 2006-2007

Thank you for your attention! Thanks you SAMSI!

SAMSI

Our storm surge working group would not have been possible without the support from SAMSI. In particular the following programs:

- Model Uncertainty: Mathematical and Statistical (MUMS) 2018-2019
- Mathematical and Statistical Methods for Climate and the Earth System (CLIM) 2017-2018
- Development, Assessment and Utilization of Complex Computer Models 2006-2007

Thank you for your attention! Thanks you SAMSI!

samsi
NSF Duke NCSU UNC

Backup Slides

Input Modeling: Estimating $f(\boldsymbol{x})$ and $\lambda(\boldsymbol{s})$

Main effort: Utilizing all sources of information to better estimate TC characteristics and its estimation uncertainty

- **Data fusion**: Combining historical storms and synthetic storms (high resolution GCM, even higher resolution RCM, Kerry Emanuel's simulation) to better estimate $f(\boldsymbol{x})$ and $\lambda(\boldsymbol{s})$

- **Borrow information across space:** Exploit and explicitly model the spatial pattern of $\{f_{\boldsymbol{s}}(\boldsymbol{x})\}_{\boldsymbol{s}\in\mathcal{S}}$ (e.g., Regional Frequency Analysis, Model-based Geostatistics).

- **Estimation Uncertainty:** Using posterior predictive distribution: $[\boldsymbol{X}_{new}|\boldsymbol{X}_{obs}] = \int_{\Theta}[\boldsymbol{X}_{new}|\theta][\theta|\boldsymbol{X}_{obs}]\,d\theta$ instead of the plug-in estimate $f_{\hat{\theta}}(\boldsymbol{x})$ ($[\boldsymbol{X}_{new}|\hat{\theta}]$)

Estimating y_r

- Should be relatively straightforward given that one can estimate $f(\boldsymbol{x})$, $\lambda(\boldsymbol{s})$, and $\eta(\boldsymbol{x})$ reasonably well and properly account for their estimation uncertainties
- Still worthwhile to exploit extreme value theory-based methods for estimating return levels with longer return periods e.g. NRC probabilistic risk assessment
- It would be of interest to compare the estimates of y_r based on EVA at locations with long/reliable enough observations to assess the "gain" by using the JPM-type of approach