4th Annual Probabilistic Flood Hazard Assessment Workshop
Rockville, MD – April 30-May 2, 2019

ERDC
ENGINEER RESEARCH & DEVELOPMENT CENTER

Advancements in Probabilistic Storm Surge Models and Uncertainty Quantification using Gaussian Process Metamodeling

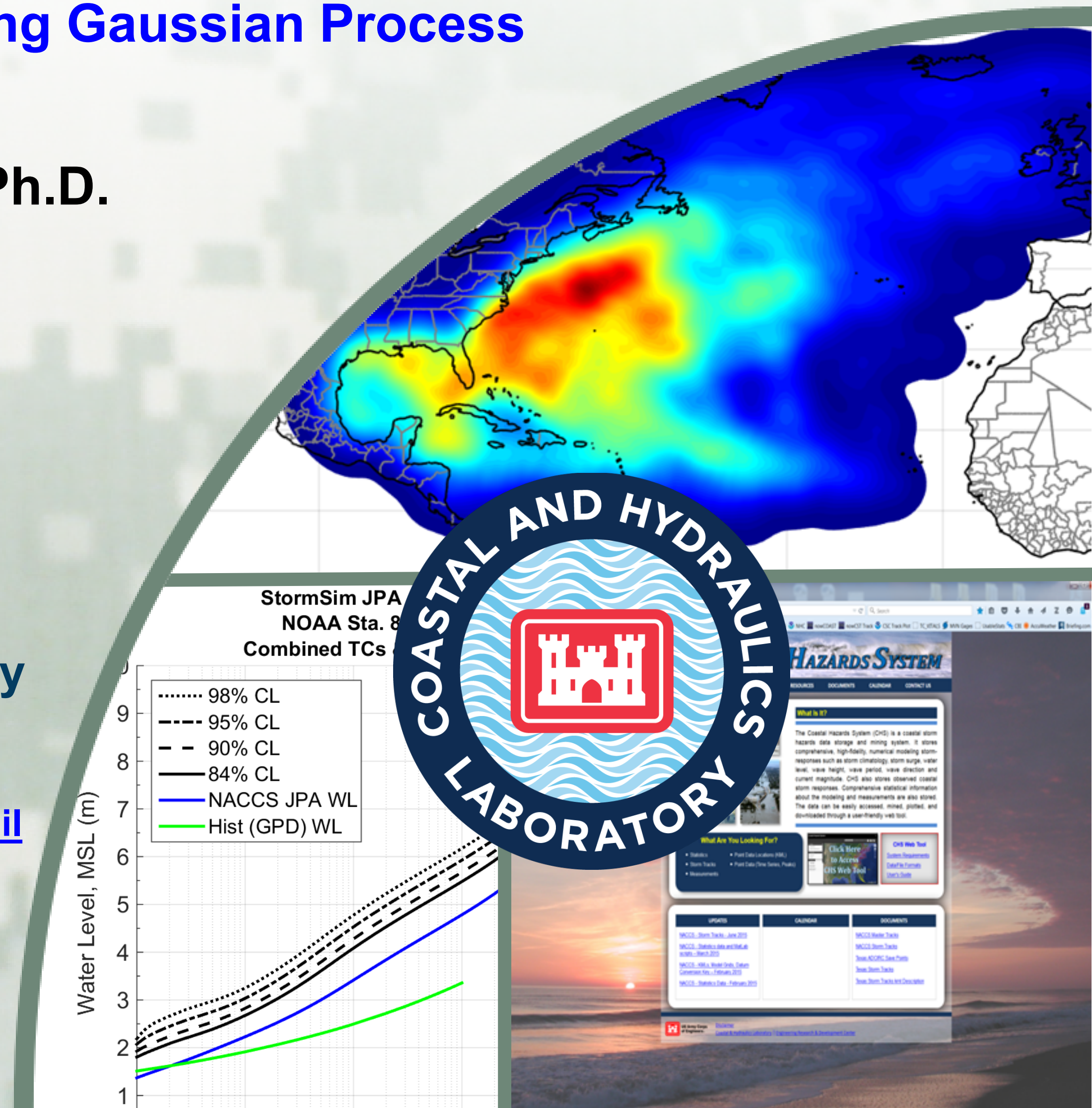
Norberto C. Nadal-Caraballo, Ph.D.
Victor M. Gonzalez, P.E.

Team: Efrain Ramos-Santiago

Alexandros Taflanidis, Ph.D.
(University of Notre Dame)

Coastal and Hydraulics Laboratory
US Army Engineer R&D Center

Norberto.C.Nadal-Caraballo@usace.army.mil



U.S. ARMY

US Army Corps of Engineers

Outline

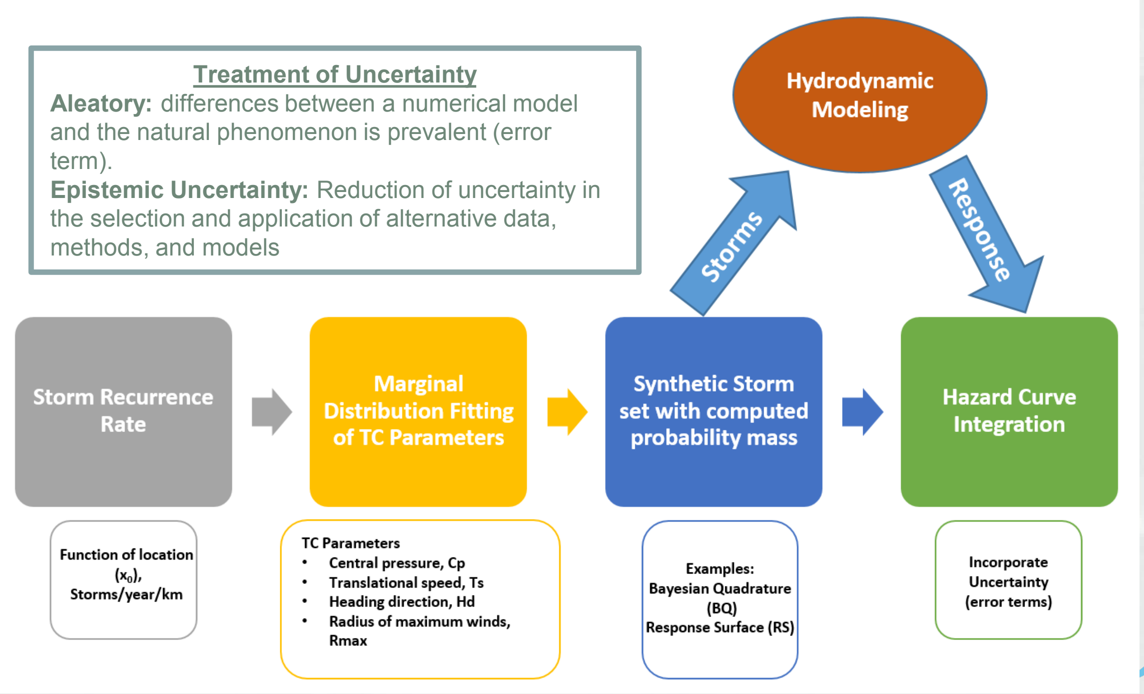
- Introduction.
- Quantification of uncertainty in probabilistic storm surge models.
- Gaussian process metamodeling (GPM): definition and error quantification.
- GPM Applications in probabilistic storm surge modeling:
  - Reference set
  - Monte Carlo simulation
  - Meta-Gaussian distribution (multivariate Gaussian copula)
- Epistemic uncertainty/Logic tree approach.

Introduction

- Study objectives:
  - Identification of technically defensible data sources, models, and methods for the computation of storm surge.
  - Assessment for carrying forward for evaluation of epistemic uncertainty.
  - Epistemic uncertainty is quantified and propagated through logic tree approach (PSHA approach).
- Gaussian process metamodeling.
  - Augmented storm suite for:
    - Reference set
    - Monte Carlo simulation
    - Meta-Gaussian distribution (multivariate Gaussian copula)
  - Error quantification.

Quantification of Uncertainty in Probabilistic Storm Surge Models

- Characterize, quantify, and propagate both aleatory and epistemic uncertainties through the probabilistic framework of storm surge assessment.



U.S. ARMY BUILDING STRONG®

COASTAL & HYDRAULICS LABORATORY

Logic Tree Approach

JPM Integral

$$\lambda_{r(\hat{x})>r} = \lambda \int P[r(\hat{x}) + \varepsilon > r|\hat{x}, \varepsilon]\, f_{\hat{x}}(\hat{x}) f_{\varepsilon}(\varepsilon) d\hat{x} d\varepsilon$$

$$\approx \sum_i^n \lambda_i\, P[r(\hat{x}) + \varepsilon > r|\hat{x}, \varepsilon]$$

where:

$\lambda_{r(\hat{x})>r}$ = AEP of TC response r due to forcing vector $\hat{x}$

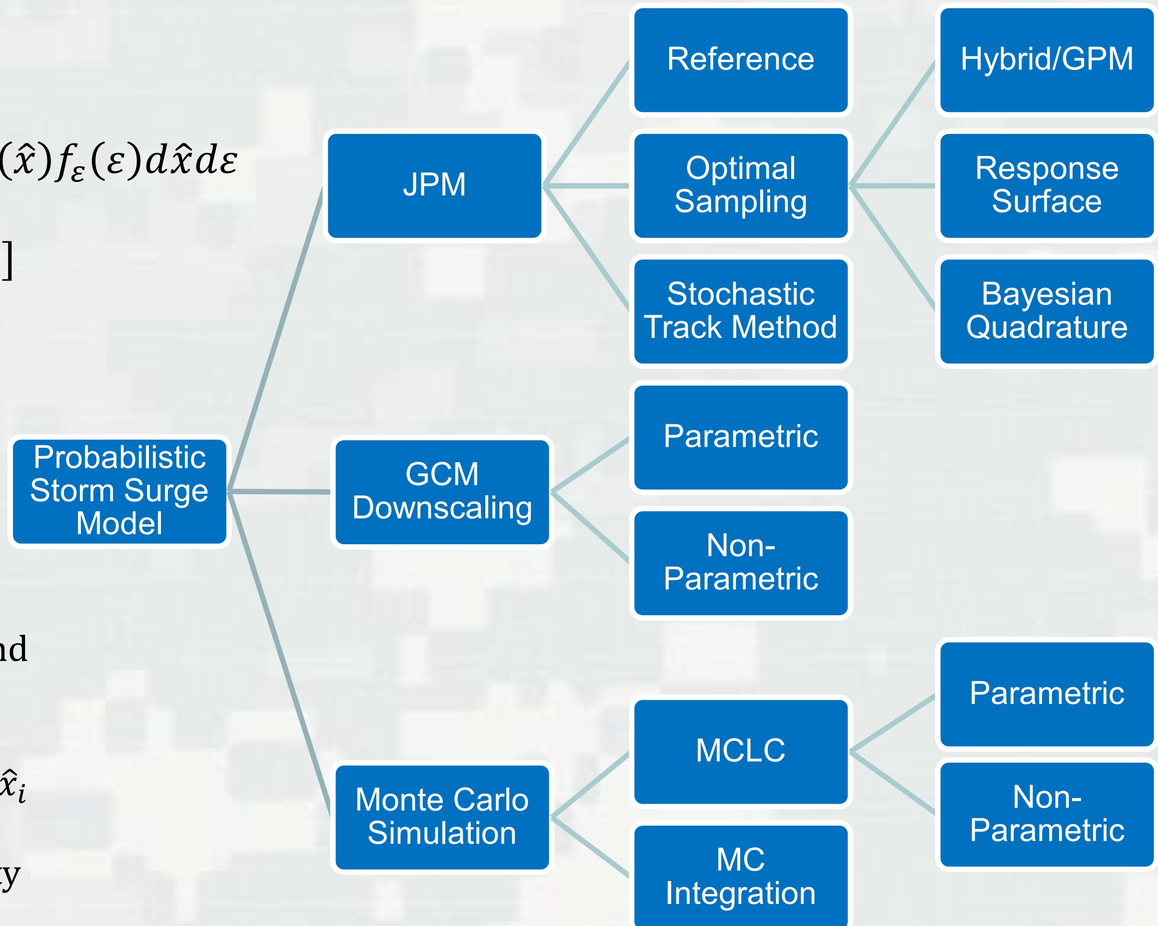
$\hat{x} = f(x_o, \theta, \Delta p, R_{max}, V_t)$

λ = SRR (storms/yr/km)

$\hat{\lambda}_i$ = probability mass (storms/yr) or $\lambda\, p_i$, with p_i=product of discrete probability and TC track spacing (km)

$P[r(\hat{x}) + \varepsilon > r|\hat{x}, \varepsilon]$ conditional probability that storm i with parameters $\hat{x}_i$ generates a response larger than r

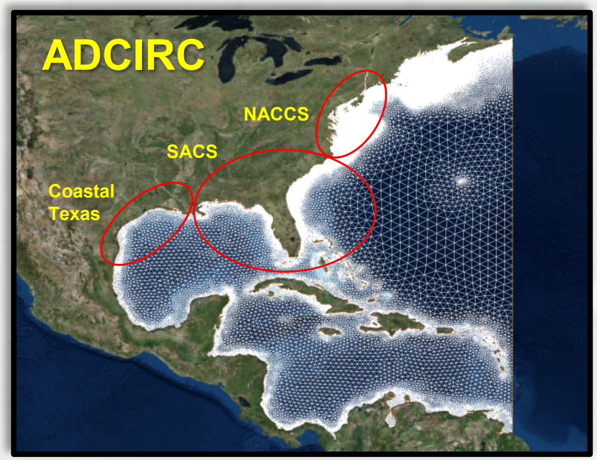
ε = unbiased error or aleatory uncertainty of r



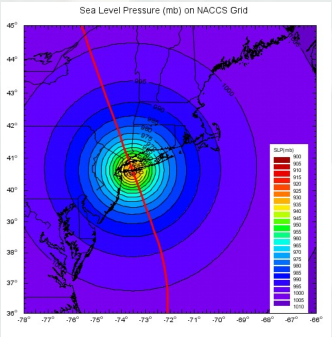
Metamodeling

- Mathematical approximation for the **input/output** (x/z) relationship of complex numerical models.
- "State-of-practice" in coastal hydrodynamic modeling for coastal hazard studies:

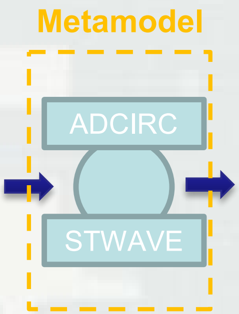
  Computationally-intensive coupled ocean circulation models and spectral wave models over large domains
- Metamodel is designed (trained) using parameterized TC inputs and hydrodynamic model outputs.

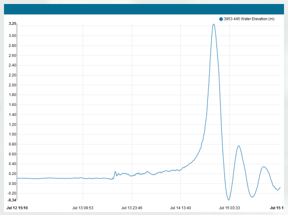


$$\text{Response} = f(\hat{x}) = f\left(x_0, \Delta p, R_{\max}, V_f, \theta\right)$$



Synthetic Storm 445, NACCS




Water Elevation

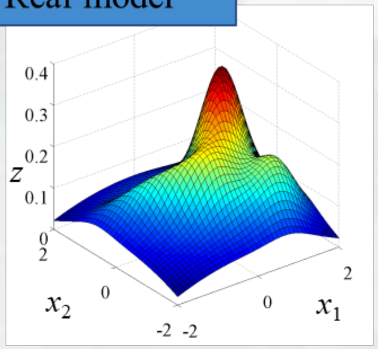
U.S. ARMY BUILDING STRONG®

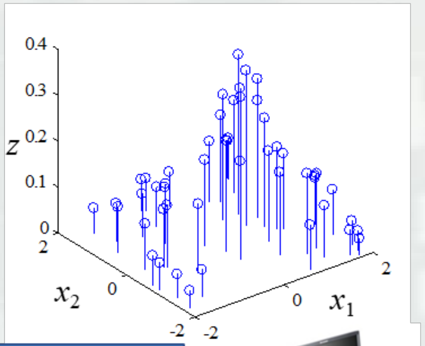
COASTAL & HYDRAULICS LABORATORY

Gaussian Process Metamodeling

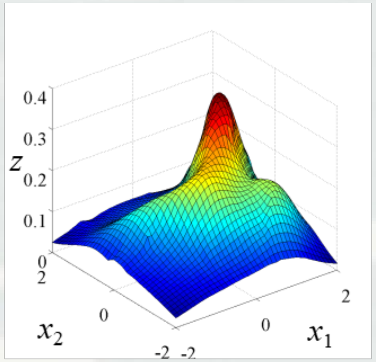
- GPM formulation has 2 parts:
  - Global regression model
  - Gaussian process or local correction
- GPM advantages:
  - Computational efficient
  - Handles complex models
  - Uncertainty of predictions

Real model





Support points



GPM approximation (global) + correction (local)

Gaussian Process Metamodeling

- Fundamental GPM formulation:
  - Regression: de-trend data.
  - GP: interpolates within the residuals of the regression.
- The predictive mean combines both the regression and GP.
- Regression is not optimized irrespective of GP.
- Validation metrics.
  - Example Types: average, per save point, per storm.

$$h(\boldsymbol{x}) = \underbrace{\boldsymbol{f}_k(\boldsymbol{x})^T \beta_k}_{\text{Global regression}} + \underbrace{z_k(\boldsymbol{x})}_{\text{Gaussian process (GP)}} \quad (1)$$

$$\bar{h}(\boldsymbol{x}|M_k) = \boldsymbol{f}_k(\boldsymbol{x})^T \beta_k^{MLE} + \boldsymbol{r}_{\theta_k^{MLE}}(\boldsymbol{x})^T \boldsymbol{R}_{\theta_k^{MLE}}^{-1}\left(\boldsymbol{H} - \boldsymbol{F}_k \beta_k^{MLE}\right) \quad (2)$$

$$\beta_k^{MLE} = \left(\boldsymbol{F}_k^T \boldsymbol{R}_{\theta_k^{MLE}}^{-1} \boldsymbol{F}_k\right)^{-1} \boldsymbol{F}_k^T \boldsymbol{R}_{\theta_k^{MLE}}^{-1} \boldsymbol{H} \quad (3)$$

Where:
β_k: regression coefficient
$\boldsymbol{f}_k$: basis functions (regression models)
$h(\boldsymbol{x})$: output
$\boldsymbol{R}_{\theta_k}$: Gaussian correlation function
θ_k: vector of parameters for correlation

e.g. R^2, Mean Absolute Error, Correlation Coefficient, RMS.

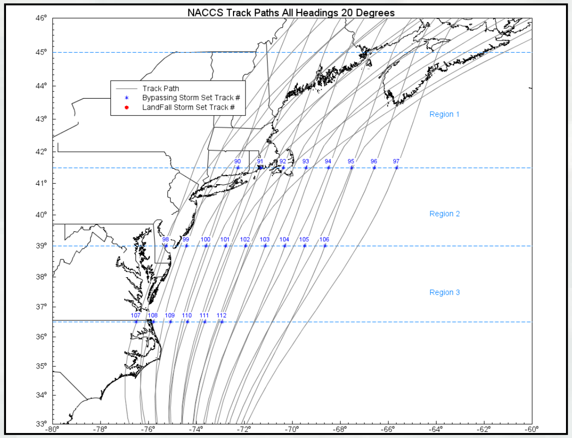
GPM Training – NACCS

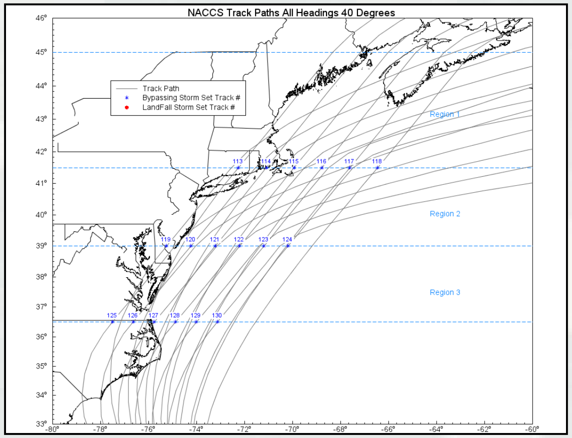
- The GPM used in this study was trained using the 1050 synthetic TCs developed as part of the NACCS (Nadal-Caraballo et al. 2015).
- Trained on WL peaks (time series training also possible).
- Validation of landfalling TCs.

Coastal Reference Location	NACCS Save Point	R^2	Mean Absolute Error	Correlation Coefficient	RMS
Virginia Beach, VA	6488	0.988	0.071	0.994	0.064
Chesapeake Bay, MD	5951	0.974	0.072	0.988	0.082
The Battery, NY	7672	0.989	0.066	0.994	0.090
Newport, RI	1082	0.985	0.077	0.993	0.068
Boston, MA	1884	0.982	0.069	0.992	0.057

GPM Training – NACCS

- Validation of bypassing TCs.



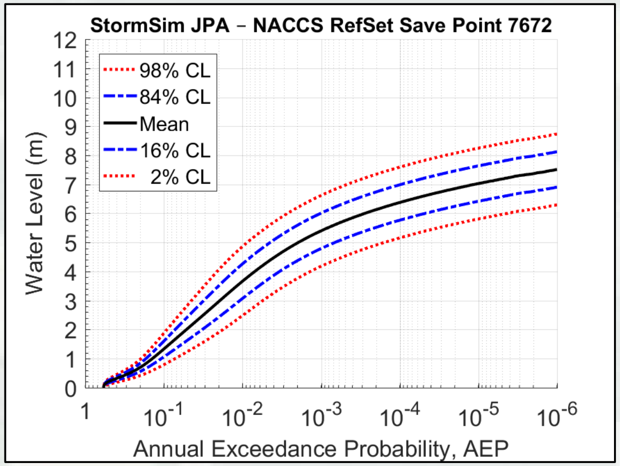


Coastal Reference Location	NACCS Save Point	R^2	Mean Absolute Error	Correlation Coefficient	RMS
Virginia Beach, VA	6488	0.970	0.091	0.986	0.089
Chesapeake Bay, MD	5951	0.963	0.075	0.983	0.060
The Battery, NY	7672	0.974	0.086	0.988	0.100
Newport, RI	1082	0.952	0.111	0.979	0.121
Boston, MA	1884	0.946	0.093	0.977	0.070

GPM Applications: Reference Set

- Developed a reference set that considers all parameter combinations.
- The reference set was developed for five coastal reference locations located within the NACCS project area.

Region	Number of tracks	Number of Δp, R_{max} and V_t unique combinations	Number of tropical cyclones
Bypassing			
1	14	495	6,930
2	15	585	8,775
3	12	675	8,100
Landfalling			
1	40	495	19,800
2	25	585	14,625
3	24	675	16,200
Total			74,430

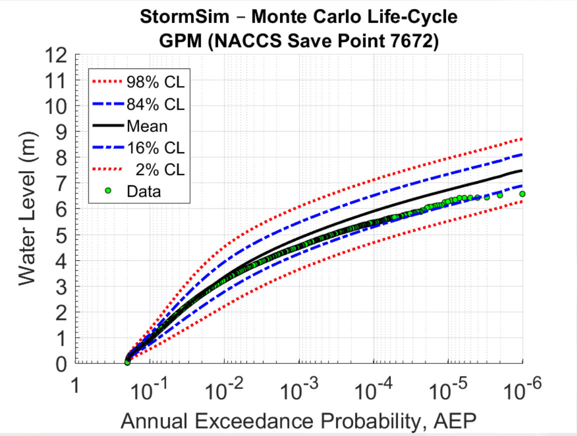


U.S.ARMY BUILDING STRONG®

COASTAL & HYDRAULICS LABORATORY

GPM Applications: Monte Carlo Simulation

- Monte Carlo Life-Cycle.
  - Univariate distributions of TC parameters were sampled for a 1,000,000-yr period, which resulted in 211,997 TCs.
  - No probability masses required.
    - TC's sampled based on their likelihood of occurrence and parameters joint probability.
    - Storm surge hazard curve from **empirical distribution** (Weibull plotting position).
  - Mean hazard curve and confidence levels calculated through bootstrap resampling using replicated storm surge values with added discretized uncertainty.



BUILDING STRONG®

COASTAL & HYDRAULICS LABORATORY

GPM Applications: Meta-Gaussian Distribution

Joint Probability: TC Parameter Dependence

Typical approaches in previous studies:

- Assumed independence.

$$P(x_0, \Delta p, R_{\max}, V_f, \theta) = P(x_0) \cdot P(\Delta p) \cdot P(R_{\max}) \cdot P(V_f) \cdot P(\theta)$$

- Correlation tree (1:1 dependence).

$$P(x_0, \Delta p, R_{\max}, V_f, \theta) =$$

$$\begin{tikzcd}
 & P(x_0) \arrow[dl] \arrow[dr] & \\
P(\Delta p \mid x_0) \arrow[d] & & P(\theta \mid x_0) \arrow[d] \\
P(R_{\max} \mid \Delta p) & & P(V_f \mid \theta)
\end{tikzcd}$$

GPM Applications: Meta-Gaussian Distribution

Previous approaches for TC parameter dependence:

Study	Δp	R_{max}	θ	V_f
LA/TX	$P(\Delta p \mid x_0)$	$P(R_{max} \mid \Delta p)$	$P(\theta \mid x_0)$	$P(V_f \mid \theta)$
Mississippi	$f(x_0)$	$f(\Delta p)$	Δp *slices*	Δp *slices*
FEMA R2	$f(x_0)$	$f(\Delta p)$	$f(x_0)$	$f(\Delta p)$
SFL	$f(x_0)$	$f(\Delta p)$	$f(x_0)$	*independent*
SWFL	$f(x_0)$	$f(\Delta p)$	$f(x_0)$	*independent*
WFL	$f(x_0)$	$f(\Delta p)$	$f(x_0)$	*independent*
Big Bend	$f(x_0)$	$f(\Delta p)$	$f(x_0)$	*independent*

Meta-Gaussian Distribution (MGD)

- **MGD** refers to a set of marginal (univariate) probability distributions with a multivariate Gaussian copula as dependence structure.

- Sklar's theorem (1959):

  - Any joint (multivariate) distribution, H, can be deconstructed into marginal distributions, $F_1, \ldots, F_n$, and a copula, C.

$$H(x_1, \ldots, x_n) = C\left(F_1(x_1), \ldots, F_n(x_n)\right)$$

$$F_1(x_1), \ldots, F_n(x_n) \rightarrow \Delta p, R_{max}, V_f, \theta$$

Meta-Gaussian Distribution (MGD)

- **Copula** – dependence function that "links" a set of marginal distributions to form a joint distribution.
  - Must be expressed in terms of

    u_n = uniform margins defined on [0,1]

    $$C(u_1, \ldots, u_n) = H\left(F_1^{-1}(x_1), \ldots, F_d^{-1}(x_d)\right)$$

- **Gaussian Copula** – "elliptical" copula.

  $$C_R^{Gauss}(u) = \Phi_R\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)$$

  Φ_R = CDF of a multivariate standard normal distribution, $\mathcal{N}(0,1)$

Meta-Gaussian Distribution (MGD)

Limitations of Gaussian Copula

- "A recipe for disaster"
  - Tail dependence is 0, regardless of the correlation matrix

$$\lambda_l = \lambda_u = 2 \lim_{x \to \infty} \Phi\left(x \frac{\sqrt{1-\rho}}{\sqrt{1+\rho}} \right) = 0$$

- Correlation ρ (X, Y) is static and symmetrical along [0,1] range
  - Slices (e.g., Δp)
- Pearson's ρ – measures linear correlation; affected by outliers
  - Spearman's ρ (rank correlation coefficient)
  - Kendall's τ

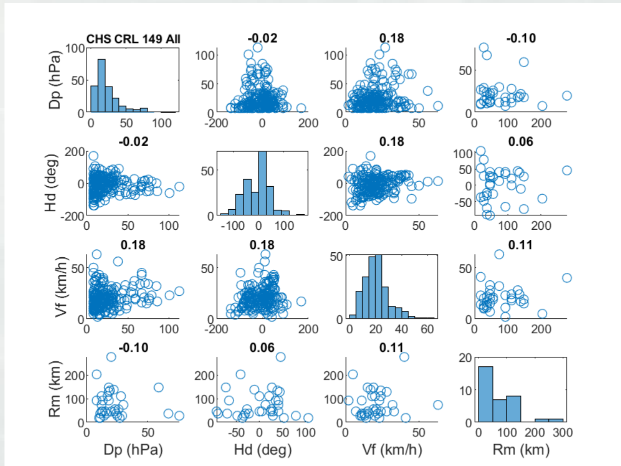
Gaussian Copula

- Correlation matrix, R

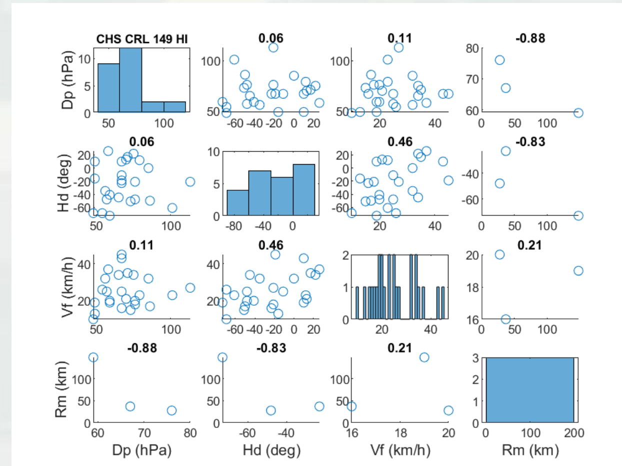
$$R = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,d} \\ \rho_{2,1} & 1 & \cdots & \rho_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d,1} & \rho_{d,2} & \cdots & 1 \end{pmatrix}$$

Meta-Gaussian Distribution (MGD)
(Mississippi Example)

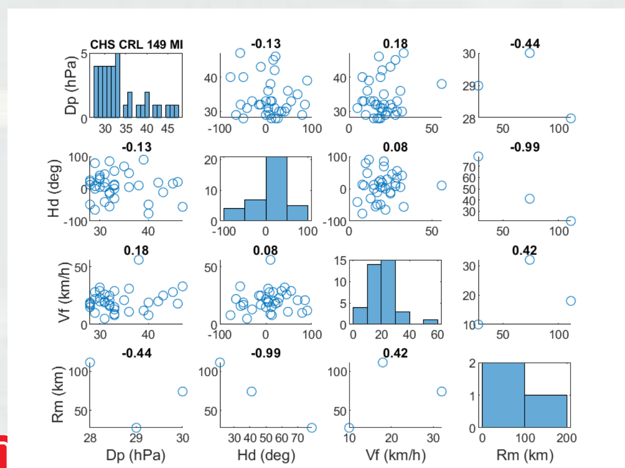
Correlation Plots



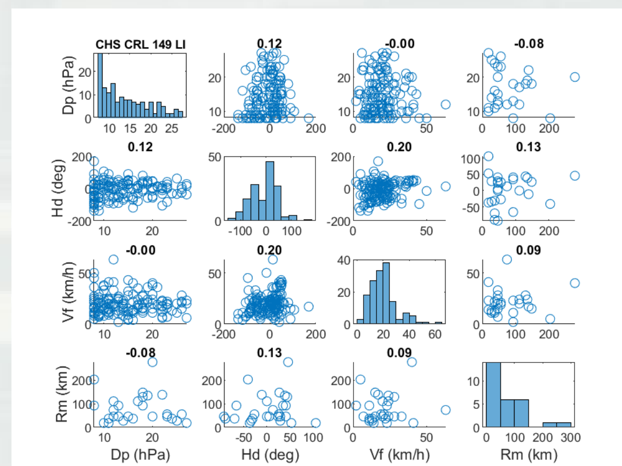
All TCs



High Intensity TCs



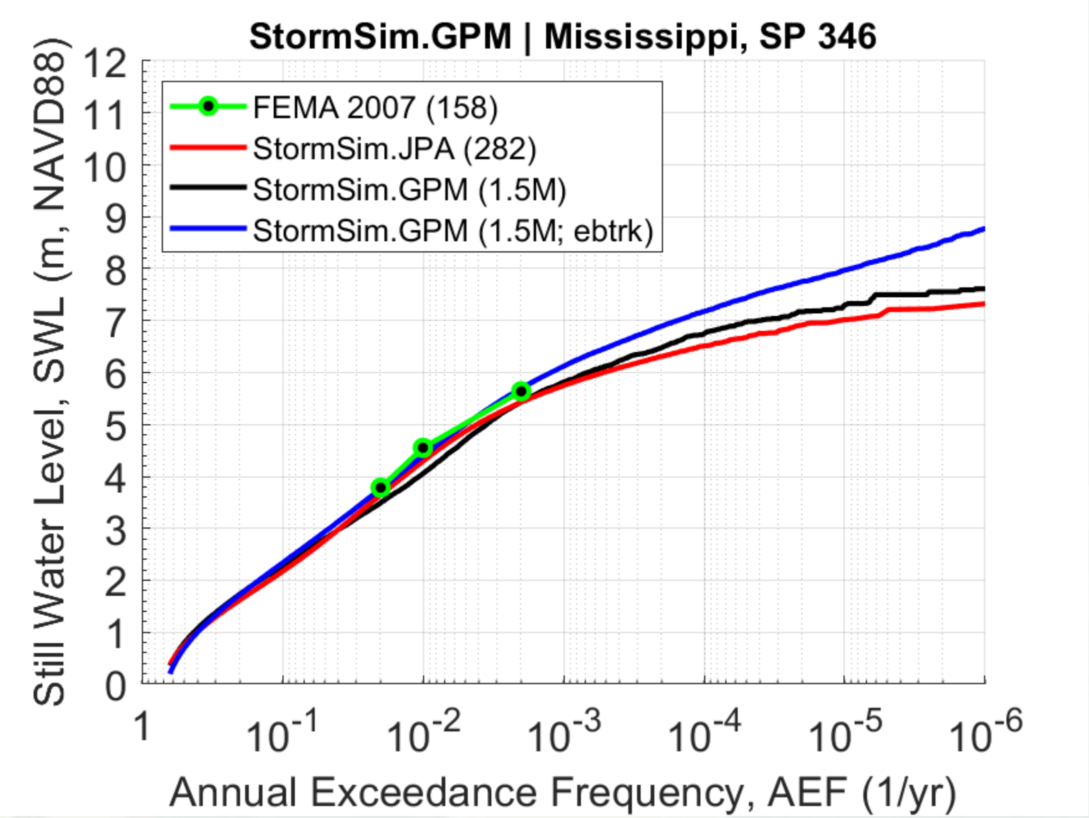
Medium Intensity TCs



Low Intensity TCs

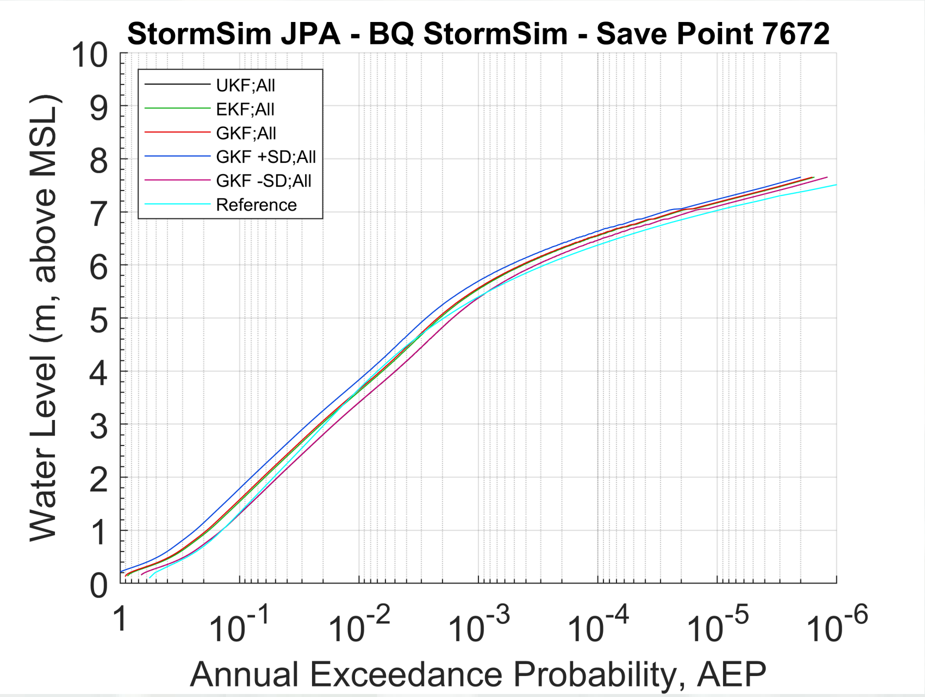
Meta-Gaussian Distribution (MGD)
(Mississippi Example)

- **MGD**
  - Marginal distributions are defined for each JPM parameter.
  - A unique copula can be fitted per Δp "slice" to maintain parameter dependencies for different TC intensity ranges.
  - The correlation between Δp and R_{max} can be updated.



Epistemic Uncertainty in SRR Models

- Models for Calculating SRR.
  - Uniform kernel function (UKF) or capture zone.
  - Gaussian kernel function (GKF).
  - Epanechnikov kernel function (EKF).
- Incorporated +/- 1 standard deviation (SD).
  - SRR uncertainty contribution ($\Delta p \geq$ 28 hPa):
    - Sampling uncertainty – 65%
    - Selected period of record – 19%
    - Gaussian kernel size – 15%
    - Observational data – 1%



$$\lambda = \frac{1}{T}\sum_{i}^{n} w(d_i)$$

$$w(d_i) = \frac{1}{h_d}\begin{cases} 0.5, & if \left|\frac{d_i}{h_d}\right| < 1 \\ 0, & otherwise \end{cases}$$

UKF

$$w(d_i) = \frac{1}{\sqrt{2\pi}h_d} exp\left[-\frac{1}{2}\left(\frac{d_i}{h_d}\right)^2\right]$$

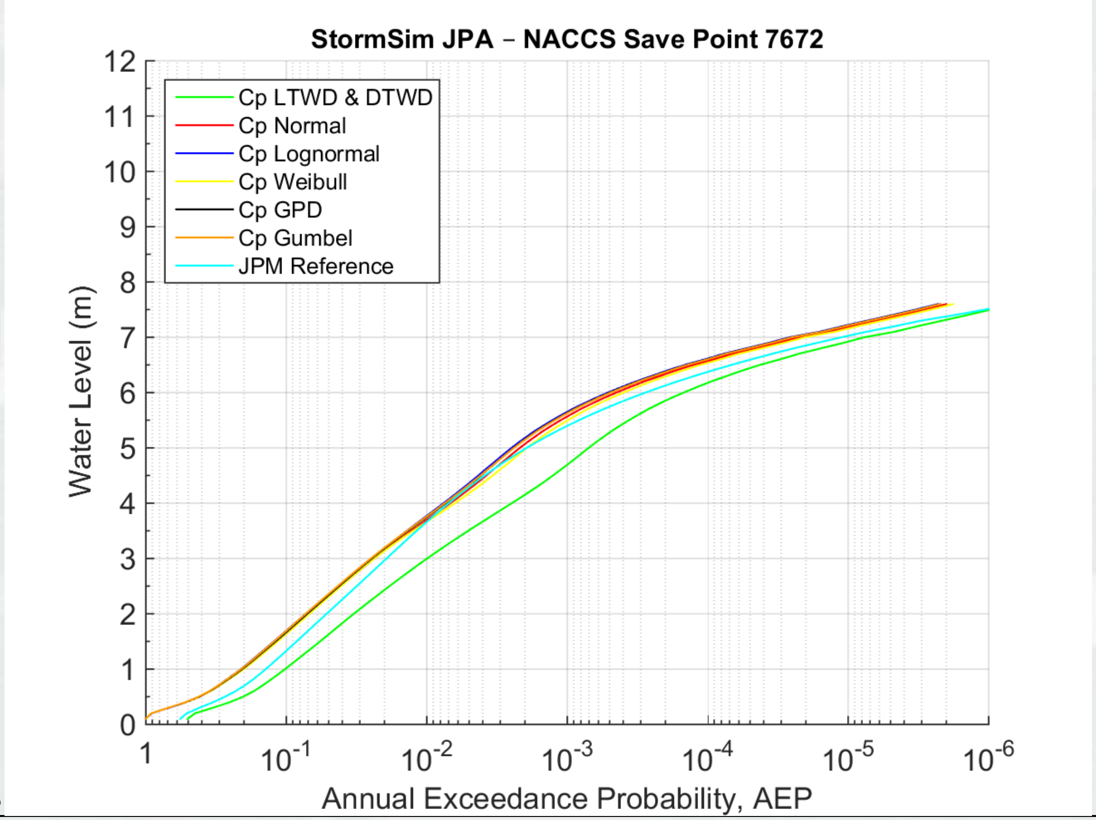
GKF

$$w(d_i) = \frac{1}{h_d}\begin{cases} \frac{3}{4}\left[1 - \left(\frac{d_i}{h_d}\right)^2\right], & if \left|\frac{d_i}{h_d}\right| < 1 \\ 0, & otherwise \end{cases}$$

EKF

Defining Joint Probability of Storm Parameters

- Effect of selection of Δp distribution on hazard curve.



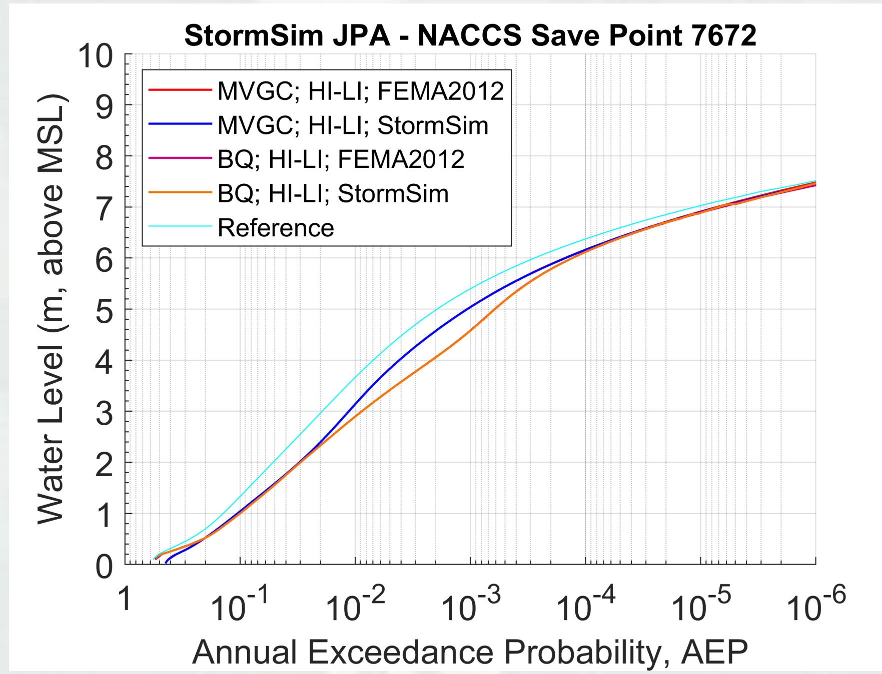
LTWD & DTWD curve considers the discretization of TCs into high and low intensity.

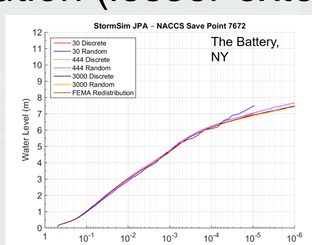
The effect is to lower the hazard curve.

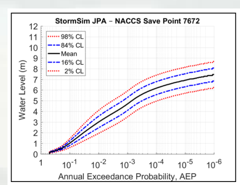
Choice of Δp distribution showed limited impact

Computation probability masses/integration

- MGD vs BQ.
- Hazard curve integration method did not have an effect.
- Elements of integration process that affect curve:
  - Characterization method:
    - Constant
    - Proportional
    - Constrained [min(20%, 0.61m)]
  - Discretization of normal distribution (lesser extent).

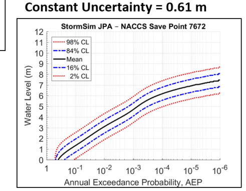


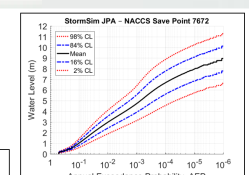




Uncertainty = min(20%, 0.61m)

Constant Uncertainty = 0.61 m

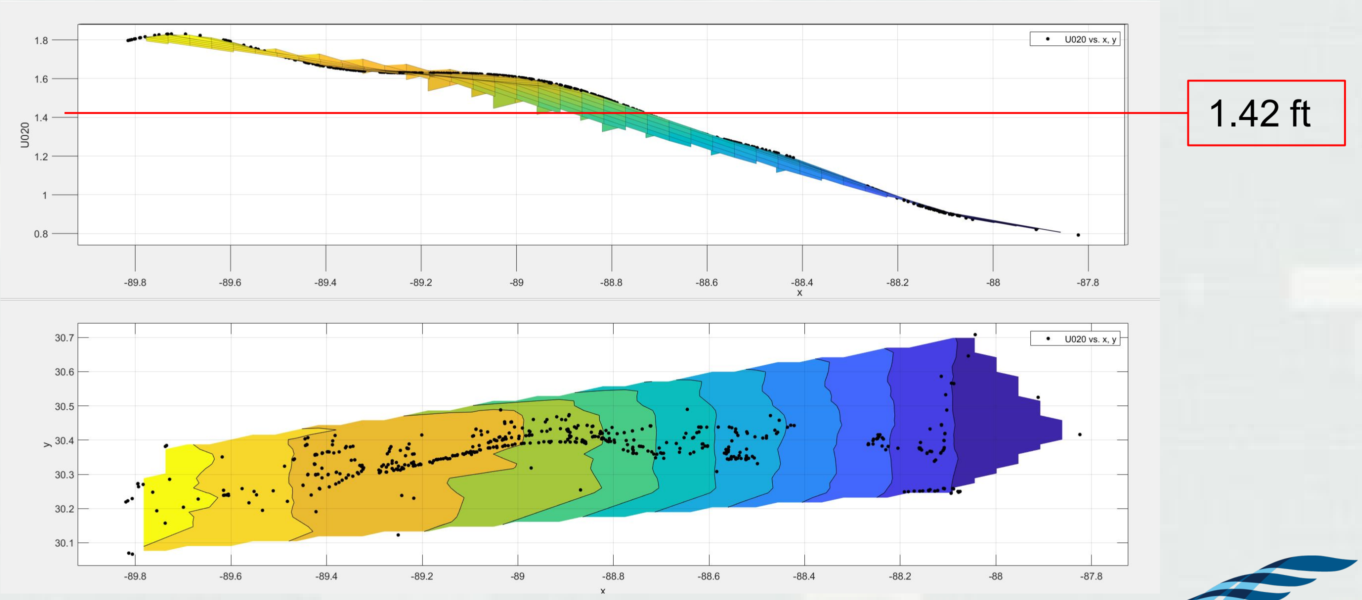




Proportional Uncertainty = 20%

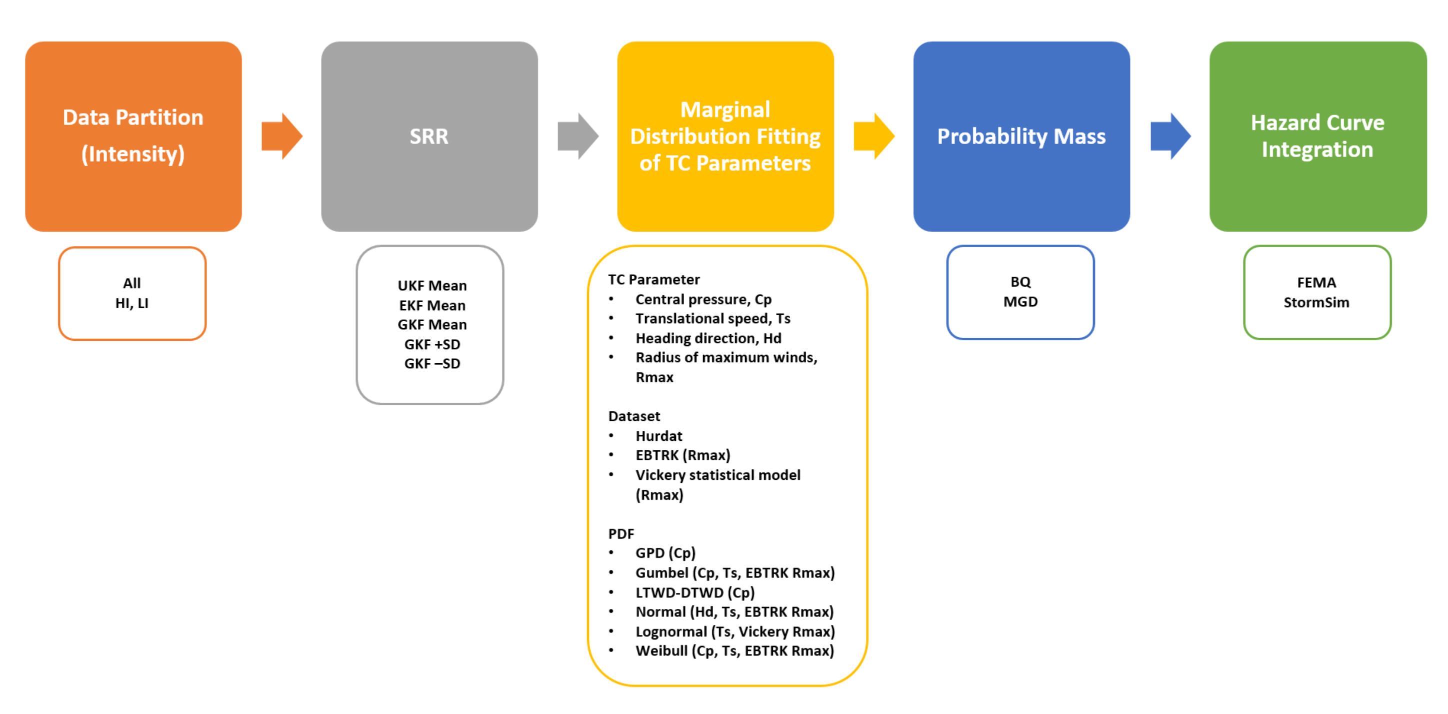
Spatially Varying Modeling Error
(Mississippi Example)

- Modeling error: has a direct effect on hazard curve shape and confidence limits.
  - Global uncertainty: **1.42 ft.**
  - Spatially varying uncertainty:



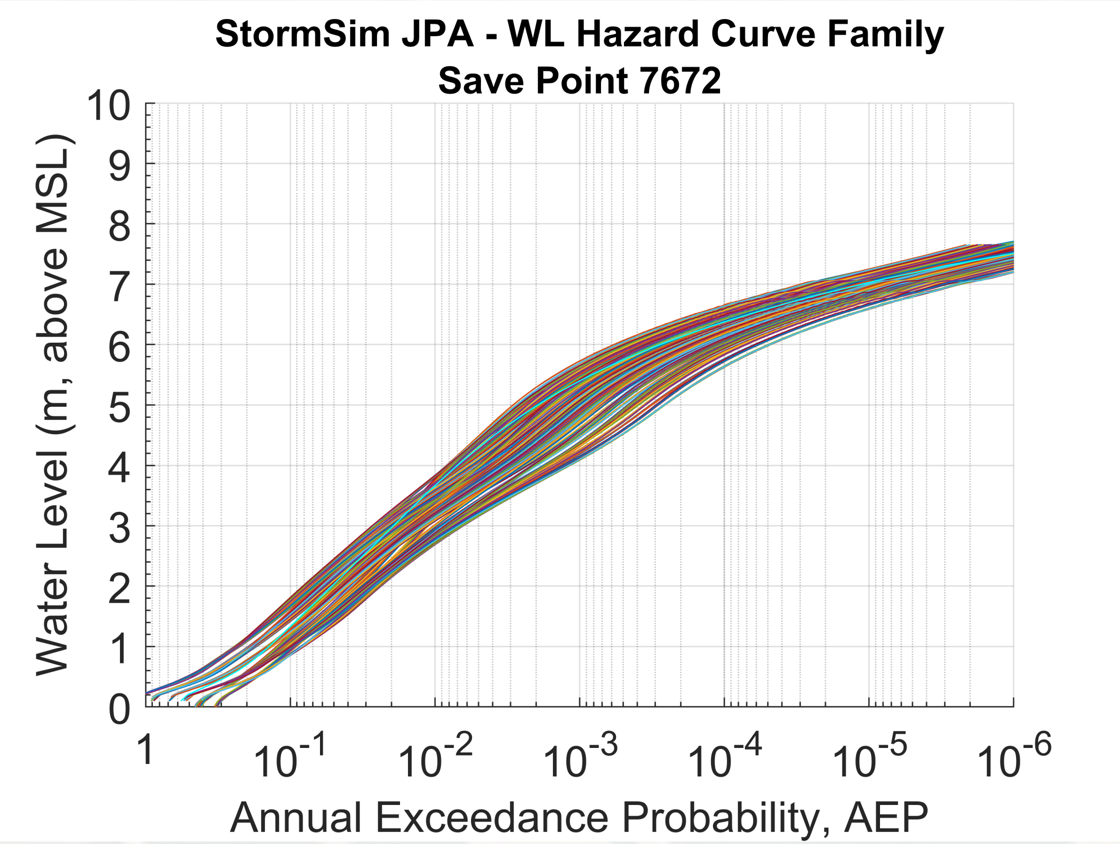
Quantification and Propagation of Epistemic Uncertainty

- Evaluated alternate data, models, and methods (logic tree branches).



Quantification of Epistemic Uncertainty

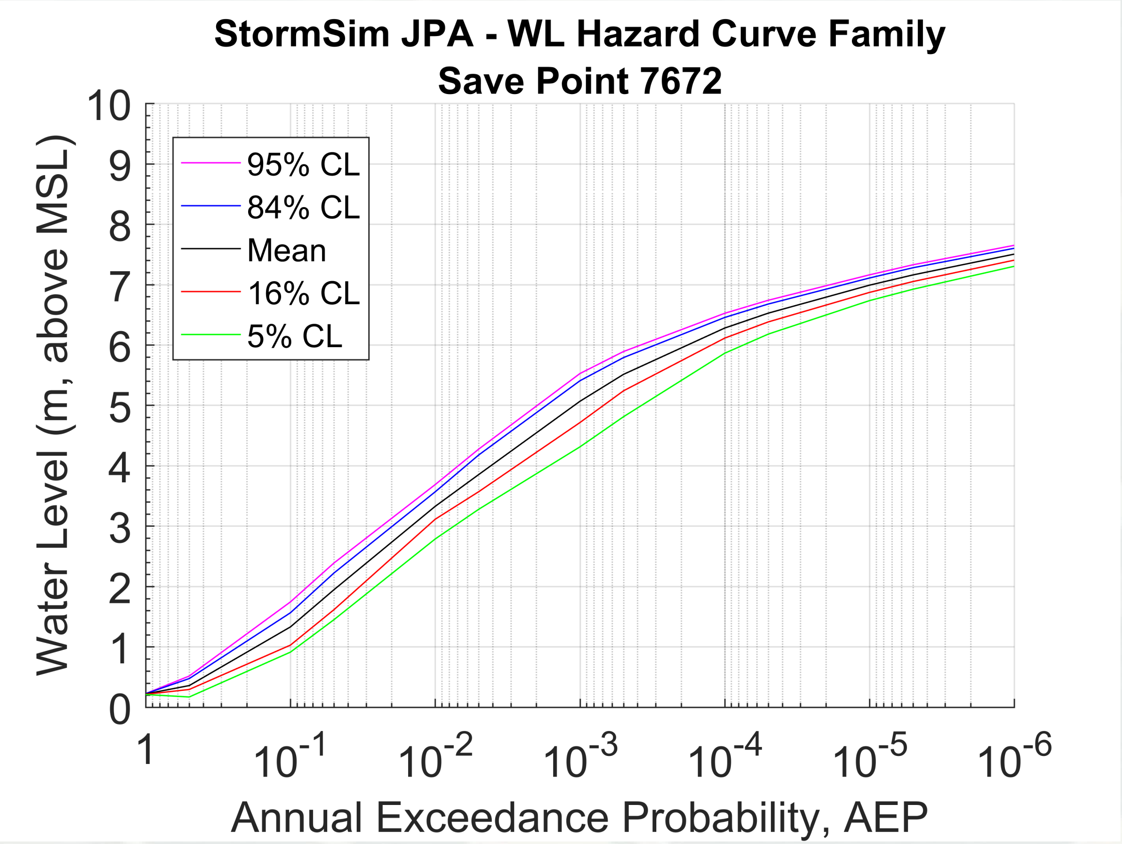
- Family of hazard curves representing alternate data, model and methods.
- Number of curves: 1,261.
- About 5 ft spread at 100 years.



The Battery, NY

Quantification of Epistemic Uncertainty

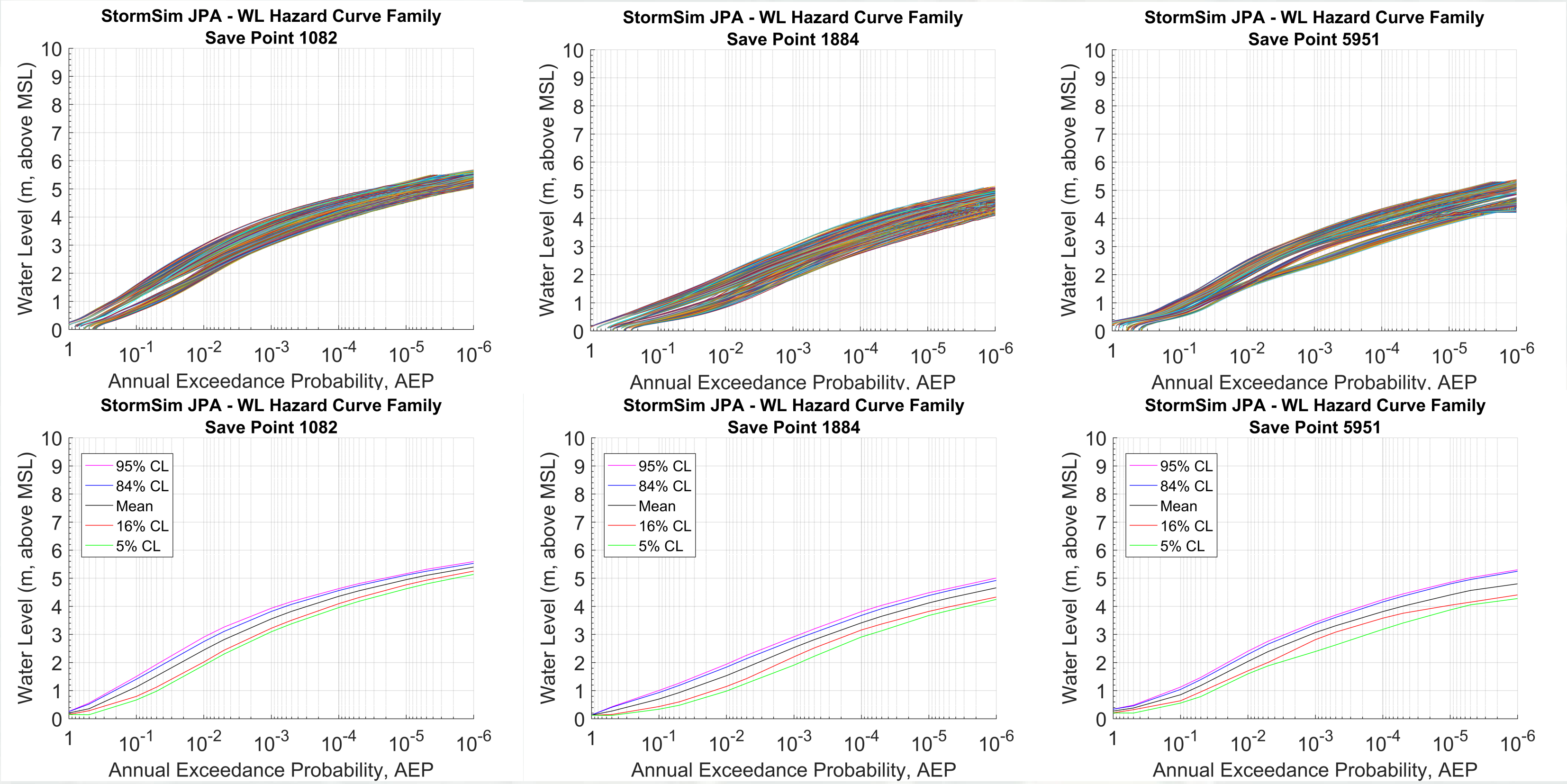
- Demonstration: all curves assigned same weight.
- Logic tree branches have been trimmed.
- Mean hazard curve with confidence limits.



The Battery, NY

US Army Corps of Engineers
BUILDING STRONG®
Coastal & Hydraulics Laboratory

Quantification of Epistemic Uncertainty



Newport, RI

Boston, MA

Chesapeake Bay, MD

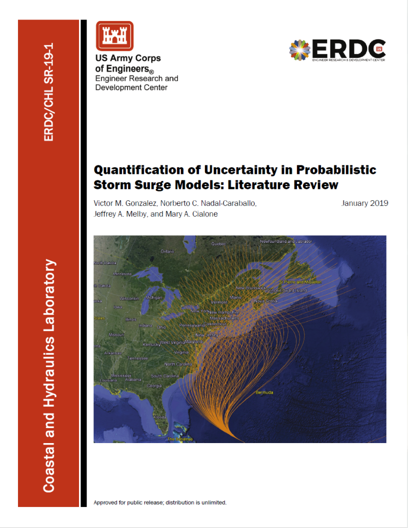
U.S.ARMY

BUILDING STRONG®

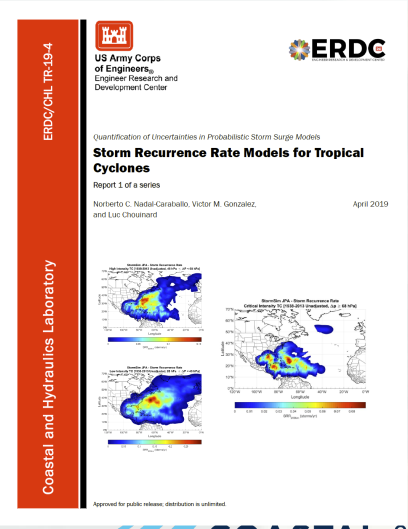
COASTAL & HYDRAULICS LABORATORY

Reports

- Gonzalez V.M., N.C. Nadal-Caraballo, J.A. Melby, and M.A. Cialone. 2019. *Quantification of Uncertainty in Probabilistic Storm Surge Models: Literature Review*. ERDC/CHL SR-19-1.


ERDC/CHL SR-19-1
Coastal and Hydraulics Laboratory
US Army Corps of Engineers® Engineer Research and Development Center
ERDC
Quantification of Uncertainty in Probabilistic Storm Surge Models: Literature Review
Victor M. Gonzalez, Norberto C. Nadal-Caraballo, Jeffrey A. Melby, and Mary A. Cialone
January 2019
Approved for public release; distribution is unlimited.

- Nadal-Caraballo, N.C., V.M. Gonzalez, and Chouinard, L. 2019. *Quantification of Uncertainties in Probabilistic Storm Surge Models: Storm Recurrence Rate Models for Tropical Cyclones*, ERDC-CHL TR-19-4. Vicksburg, MS: U.S. Army Engineer Research and Development Center.


ERDC/CHL TR-19-4
Coastal and Hydraulics Laboratory
US Army Corps of Engineers® Engineer Research and Development Center
ERDC
Quantification of Uncertainties in Probabilistic Storm Surge Models
Storm Recurrence Rate Models for Tropical Cyclones
Report 1 of a series
Norberto C. Nadal-Caraballo, Victor M. Gonzalez, and Luc Chouinard
April 2019
Approved for public release; distribution is unlimited.

References

- Chouinard, L., C. Liu, and C. Cooper. 1997. Model for Severity of Hurricanes in Gulf of Mexico. *Journal of Waterway, Port, Coastal, and Ocean Engineering* 123 (3): 120–129.
- Demuth, J., M. DeMaria, and J.A. Knaff, 2006: Improvement of advanced microwave sounder unit tropical cyclone intensity and size estimation algorithms. *Journal of Applied Meteorology and Climatology*, 45: 1573-1581.
- Jia, Gaofeng, A. A. Taflanidis, N. C. Nadal-Caraballo, J. A. Melby, A. B. Kennedy, and J. M. Smith. 2016. Surrogate Modeling for Peak or Time-Dependent Storm Surge Prediction over an Extended Coastal Region Using an Existing Database of Synthetic Storms. *Natural Hazards* 81 (2): 909–938.
- Nadal-Caraballo, N.C., J.A. Melby, V.M. Gonzalez, and A.T. Cox. 2015. North Atlantic Coast Comprehensive Study – Coastal Storm Hazards from Virginia to Maine. ERDC/CHL TR-15-5. Vicksburg, MS: U.S. Army Engineer Research and Development Center.
- Vickery, P.J., and D. Wadhera. 2008. Statistical Models of Holland Pressure Profile Parameter and Radius to Maximum Winds of Hurricanes from Flight-Level Pressure and H*Wind Data. *Journal of Applied Meteorology and Climatology* 47(10): 2497-2517.

Contact Information

U.S. Army Engineer R&D Center

Coastal and Hydraulics Laboratory

Norberto C. Nadal-Caraballo, Ph.D.

Phone: (601) 634-2008

Email: Norberto.C.Nadal-Caraballo@usace.army.mil

U.S. Nuclear Regulatory Commission

Joseph F. Kanney, Ph.D.

Phone: (301) 980-8039

Email: Joseph.Kanney@nrc.gov