ABSTRACT

Title of Dissertation:	EXTENSION OF EXISTING PROBABILISTIC COASTAL HAZARD ANALYSIS FOR BAYESIAN NETWORK COASTAL COMPOUND FLOOD ANALYSIS
	Ziyue Liu, Doctor of Philosophy, 2023
Dissertation directed by:	Assistant Professor, Michelle T Bensi, Civil and Environment Engineering

In the past decades, coastal compound floods (CCF) caused significant losses to coastal communities. To develop an accurate and complete probabilistic framework for assessing coastal hazards, the U.S. Army Corps of Engineers established the Coastal Hazard System's Probabilistic Coastal Hazard Analysis (PCHA) framework. The current PCHA framework has focused primarily on a subset of CCF drivers, particularly storm surges.

This dissertation documents four studies that contribute to efforts to advance and extend the PCHA for CCF analysis.

In the first study, machine learning-based data imputation models are developed to fill in missing records in the historical storm dataset. The performance of the machine learning-based data imputation models under different parameterizations are assessed considering statistical and physical factors and also compared against existing prediction models.

In the second study, a series of Joint Probability Method (JPM) assumptions to model the dependence among tropical cyclone (TC) atmospheric parameters are comparatively investigated. JPM assumptions considered in the analysis include parameter independence, partial dependence, and full dependence. Candidate full-dependence models include meta-Gaussian copula and vine copulas combining linear-circular copulas. Emphases are put on modeling the circular behavior of storm heading and its dependencies with other linear parameters. Full dependence models are compared based on the predicted probability of large CPD and RMW combinations.

In the third study, a Bayesian Network (BN) of multiple coastal hazards is constructed, where the conditional probability tables (CPT) of TC atmospheric parameters are computed using copula. A deaggregation methodology is developed to identify the dominant TC for significant coastal hazard events to support risk-informing and decision-making processes and refined analyses.

In the fourth study, leveraging the aforementioned study results, the extended PCHA is leveraged to develop a multi-tiered BN CCF analysis framework. In this framework, multiple tiers of BN models with different complexities are designed for study cases with varying levels of resource availability. A case study is conducted in New Orleans, LA and a series of joint distribution, numerical, machine learning, and experimental models are used to compute CPTs needed for BNs.

EXTENSION OF EXISTING PROBABILISTIC COASTAL HAZARD ANALYSIS FOR BAYESIAN NETWORK COASTAL COMPOUND FLOOD ANALYSIS

by

Ziyue Liu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

2023

Advisory Committee:

Professor Michelle Bensi, Chair

Professor Nii O. Attoh-Okine

Professor Gregory Baecher

Dr. Meredith Carr, Special Member

Professor Mohammad Modarres, Dean's Representative

© Copyright by

Ziyue Liu

2023

Acknowledgments

I am deeply thankful to have spent my Ph.D. studies at the University of Maryland, where I have had the opportunity to meet exceptional individuals, acquire invaluable research expertise, and undergo personal growth. First and foremost, I have my heartfelt gratitude to my advisor, Dr. Michelle Bensi, for her strong mentorship and guidance throughout my Ph.D. studies and research. Dr. Bensi's compassionate and sensitive mentoring style not only enriched my academic experience but also fueled my passion and motivation for research in my life.

I also want to express my sincere appreciation to Dr. Meredith Carr, Dr. Norberto Nadal-Caraballo, and other members of the United States Army Corps of Engineers (USACE) Research and Development Center Coastal and Hydraulics Laboratory (ERDC-CHL), who provided guidance and support during my research. I am immensely grateful to my dissertation committee members, Dr. Nii O. Attoh-Okine, Dr. Gregory Baecher, Dr. Mohammad Modarre, Dr. Meredith Carr, and Dr. Allison Reilly for their willingness to invest their time in reviewing my dissertation and participating in my defense.

My research was supported by the USACE's ERDC-CHL under contract award number ID W912HZ20C0050. All opinions expressed in this dissertation are the authors' and do not necessarily reflect the policies and views of the research sponsor (USACE) or any other organization.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AEF	Annual exceedance frequency
AMC	Antecedent moisture condition
ANN	Artificial neural network
BN	Bayesian network
CCF	Coastal compound flood
CHS	Coastal hazard system
CPT	Conditional probability table
ΔP	Storm central pressure deficit
EBTRK	Tropical Cyclone Extended Best Track Dataset
EC	Evidence case
GKF	Gaussian kernel function
GPM	Gaussian process met-model
GPR	Gaussian process regression
HI	High intensity
HURDAT2	Hurricane Database 2
JPM	Joint probability method
LCGV	Linear-circular Gaussian vine copula
LCFV	Linear-circular Frank vine copula
LI	Low intensity
MCS	Monte Carlo simulation
MGC	Meta-Gaussian copula
MI	Moderate intensity

NACCS	North Atlantic Coast Comprehensive Study
NOAA	National Oceanic and Atmospheric Administration
P_{c}	Storm central pressure
PCHA	Probabilistic coastal hazard analysis
PM	Probability mass
R_{34}	34 kt wind radii maximum extent
R_{50}	50 kt wind radii maximum extent
R_{64}	64 kt wind radii maximum extent
R_{max}	Radius of maximum wind
SLR	Sear level rise
SSR	Storm recurrence rate
SST	Sea surface temperture
SWAN	Simulating WAves Nearshore model
TC	Tropical cyclone
TCPC	Tropical cyclone parameter combination
TCR	Tropical cyclone rainfall model
Θ	Storm heading direction
USACE	United States Army Corps of Engineers
V_{f}	Storm forward velocity
V_{w}	Storm wind speed
VKF	Von-misses kernel function
X_0	Storm landfall location

Chapter 1: Introduction

1.1 Background

Coastal hazards bring significant damage to coastal communities and have caused over one trillion dollars of losses in the United States since 1980 (NOAA 2022a). Among the top 20 most costly tropical cyclone (TC) disasters in U.S. history, eighteen have happened since 1998, resulting in over 800 billion dollars in losses (NOAA 2022b). As a result, there is an increasing demand for more accurate quantification of coastal hazards arising from tropical cyclone events. Comprehensive assessments of coastal hazards from tropical cyclones must account for individual hazard mechanisms (e.g., storm surge, storm precipitation, and river flow) as well as compound hazard events in which mechanisms interact and lead to more severe impacts.

In the past decades, the joint probability method (JPM) has been used as the primary method for probabilistic assessment of coastal hazards (Myers 1954; Ho and Myers 1975; Russell 1969; Myers 1970), with a particular focus on storm surge hazards. Since its initial development, enhancements have been made to the JPM to increase efficiency and accuracy (e.g., Resio 2007; Toro 2008; Toro et al. 2010). Recently, a series of studies of the United States Army Corps of Engineers (USACE) (Cialone et al. 2015; Nadal-Caraballo et al. 2015, 2019, 2020, 2022a; b) leveraged a series of technological advancements including machine learning and copula models to make further advancements and formalize a procedure for the USACE's Coastal Hazard System's (CHS) probabilistic coastal hazard analysis (PCHA) framework. Figure 1 is a flowchart of the USACE's CHS-PCHA procedure, as adapted from (Nadal-Caraballo et al. 2020).

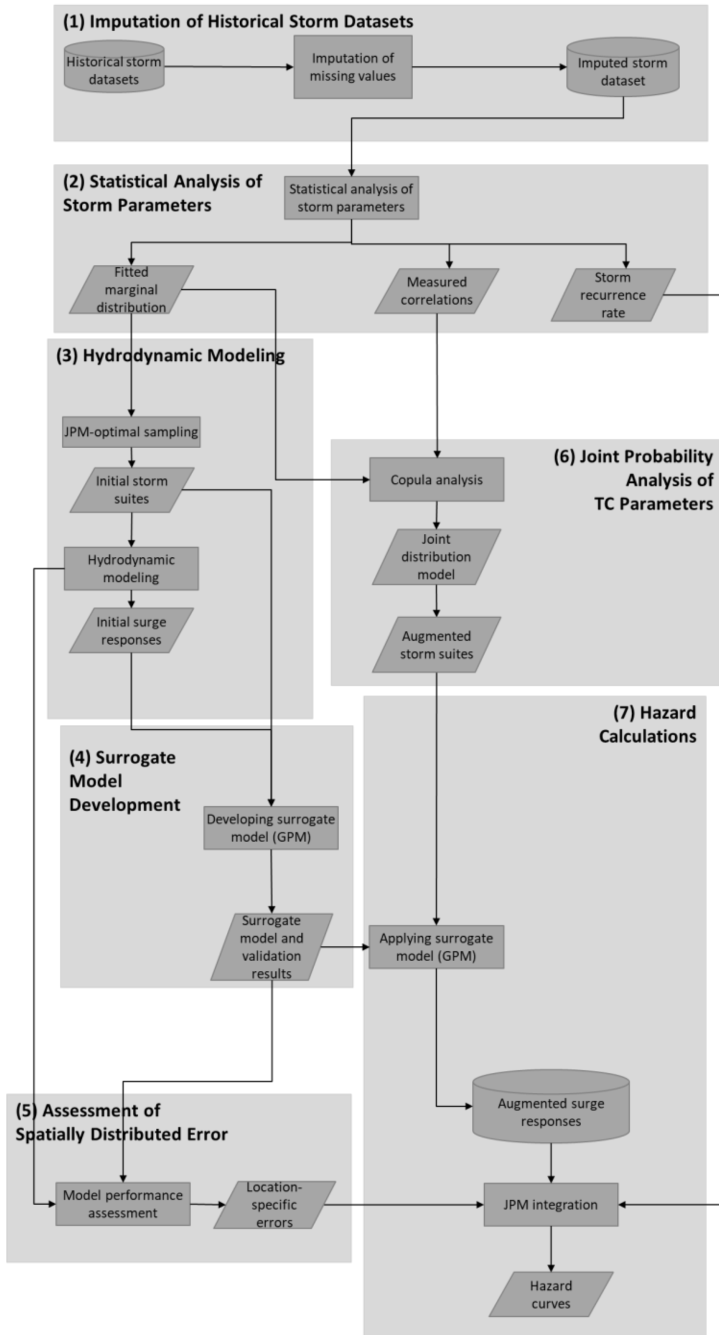


Figure 1: Flowchart of the PCHA procedure (adapted from Nadal-Caraballo et al. 2020)

In this USACE's procedure, historical data for TC atmospheric parameters (i.e., central pressure deficient (ΔP), forward velocity (V_{f}), radius to maximum winds ($R_{\max}$) and storm heading (Θ)) is first collected or computed from existing databases. Because existing data are not complete, a data imputation process is applied to fill in missing values in historical storm datasets, as illustrated in box (1) of Figure 1. In box (2), imputed historical storm datasets are used for statistical analysis of storms affecting the study region. The marginal distribution of each TC atmospheric parameter is fitted, the correlation coefficients among TC atmospheric parameters are estimated, and the storm recurrence rate (SRR) is computed. In box (3), based on the marginal distributions, a JPM optimal sampling process (Resio 2007; Toro 2008; Toro et al. 2010) is applied to discretize the TC atmospheric parameter space and generate a relatively small number (usually a few hundred) of synthetic storms. The initial storm suite is input into a high-fidelity (computationally expensive) hydrodynamic model to generate storm surge responses across the study region. In box (4), the initial storm suite and associated storm surge responses are used as input and output to train a more efficient storm surge surrogate model. In box (5), the errors of the storm surge surrogate model are quantified. In box (6), copulas are fitted and used to build joint probability distribution models for TC atmospheric parameters. The TC atmospheric parameter space is then finely discretized to create a set of tens of thousands of synthetic storms (referred to as the augmented storm suite). The copula-based joint distribution is then similarly discretized to assign a probability mass to each synthetic storm in the augmented storm suite. In box (7), storm surge responses of the augmented storm suite are computed with the storm surge surrogate models, and the JPM integral is used to estimate hazard curves, which provides a measure of storm response (e.g., peak storm surge) on one axis and the corresponding annual frequency (or annual probability) of exceedance on the other axis.

1.2 *Dissertation objectives and outline*

To date, the USACE's CHS-PCHA procedure shown in Figure 1 has been primarily used to assess coastal hazards resulted from storm surge. In this dissertation, research efforts are devoted to extending the USACE's CHS-PCHA for coastal compound flood (CCF) hazards. To account for differences in resources that are available to support CCF hazard assessment, a multi-tiered framework is established. Four studies presenting these research efforts are documented separately from Chapter 2 to Chapter 5.

As shown in box (1) of Figure 1, historical storm datasets are used as the sources of statistical analysis. Existing historical storm datasets of the Atlantic basin date back to 1851. However, these datasets are not complete throughout the period of record. Specifically, a substantial part of the record for certain parameters (central pressure (p_{c}) and $R_{\max}$) is missing because of the measurement and sensing technology limitations in the early years. A data imputation process that can fill in these missing values is needed to make the storm datasets complete. In this study, machine learning technologies are leveraged for this imputation process. The details and outcomes of this effort are documented in Chapter 2.

As shown in box (6) of Figure 1, copulas are used to capture the dependence among TC atmospheric parameters in defining their joint probabilities. The meta-Gaussian copula[1] (MGC) is straightforward to implement due to the ease of parameter estimation and model construction. However, the MGC has a drawback in dealing with circular variables such as storm heading, which may be more important when assessing multiple coastal hazards (e.g., storm rainfall) than

[1] In this dissertation, a multivariate Gaussian copula that links a set of marginal distributions is refered as a meta-Gaussian copula

storm surge alone. Chapter 3 documents the efforts to (1) assess the robustness of the MGC for characterizing the joint distribution of TC atmospheric parameters in PCHA and (2) comparatively investigate different joint distribution models.

Performing an analysis that captures the whole breadth of coastal hazards is important but can also be expensive. As a result, there is a demand to focus resources on identifying and understanding TC scenarios (i.e., TC parameter combinations; TCPCs) that are dominant contributors to significant coastal hazards that equal or exceed a specified threshold. The deaggregation (also referred to as disaggregation) method has been commonly employed for probabilistic seismic hazards analysis to identify earthquake event rupture scenarios that contribute to seismic hazards with specified return periods (Baker 2013). In this work, a Bayesian network (BN) based deaggregation method has been developed to identify the dominant storm scenarios associated with individual and multiple coastal hazards that equal or exceed a specified threshold. The details of this method are documented in Chapter 4.

Ultimately, the studies described above provide input information that enables a multi-tiered BN framework for CCF analysis. Specifically, this work develops multiple tiers of BN CCF analysis framework that can be adapted to provide results with varying degrees of sophistication. The framework leverages and extends the CCF BN model established by Mohammadi et al. (2023) and USACE's CHS-PCHA. In this framework, multiple tiers of BN models with different complexities are designed for study cases with varying levels of data availability and computational resources. The details of this framework are documented in Chapter 5.

Chapter 2: Machine Learning Motivated Data Imputation of Storm Data Used in Coastal Hazard Assessments[2]

2.1 *Introduction*

Probabilistic hazard analysis assesses the frequency and severity of hazards. The Joint Probability Method (JPM) has become the foundational tool for probabilistic analysis of coastal hazards, particularly those associated with storm surges (Myers 1954; Ho and Myers 1975; Russell 1969; Myers 1970). JPM-based methods are increasingly being explored and used as part of assessments focused on compound hazards. Computation-enabling extensions such as response surface methods (Lee et al. 2021; Resio 2007), optimal sampling approaches (Toro et al. 2010), and the use of machine learning (Jia et al. 2016; Taflanidis et al. 2014; Zhang et al. 2018) have been fundamental in enabling the JPM for probabilistic coastal hazard assessment. The United States Army Corps of Engineers (USACE) implements the Coastal Hazard System's (CHS) (https://chs.erdc.dren.mil) Probabilistic Coastal Hazard Analysis (PCHA) framework (Nadal-Caraballo et al. 2020), which builds off and extends the JPM. This implementation of JPM uses a hybrid set of sample approaches in the development of the joint probability distribution of storm atmospheric-forcing parameters; central pressure deficit (Δp), forward velocity (V_{f}), radius of maximum winds ($R_{\max}$) and storm heading (θ). The approaches used by USACE to develop these joint distributions have evolved from assuming independence between storm atmospheric-forcing parameters (Nadal-Caraballo et al. 2015) to use of models that capture the full

[2] This chapter was submitted to the journal Coastal Engineering on September 1, 2023 (Manuscript number: CENG-D-23-00364) and as of October 31, 2023, it is "Under Review."

dependence between quantities using copula-based methods (Nadal-Caraballo et al. 2020, 2022a; b).

To provide the climatological history for JPM analyses, statistical analysis to develop the joint distribution of storm parameters typically use historical storm databases (CIRA 2021; Demuth et al. 2006; Landsea and Franklin 2013; NOAA 2022c). However, these records are incomplete due to short record lengths, low spatial resolution in observations, and missing observations for some historical storms. Most notably, central pressure (p_c) and R_max are often not available. Missing observations can adversely affect the results of statistical assessments of atmospheric storm parameters, potentially leading to bias in the resulting joint distributions. This challenge is exacerbated when seeking to build joint probability models that capture the full dependence between random variables (Nadal-Caraballo et al. 2020, 2022a; b).

To augment historical records, data imputation can be used to fill in missing gaps in storm parameter observations. Data imputation is the process of substituting missing values with values generated with a model built using available observations. The data imputation "model" can be as simple as replacing missing values with the mean of the available dataset or a randomly selected observation. More sophisticated approaches use regression or machine learning methods to build an imputation model from available data.

In this work, we explore and assess the performance of multiple machine learning strategies to impute missing p_c and R_max observations in the historical storm datasets. Specifically, we consider multiple imputation models in conjunction with multiple candidate model parameterizations (i.e., alternate sets of predictor variables). Section 2 provides background information and describes available data. Section 3 introduces the candidate model parameterizations and candidate machine learning models used in this study to impute p_c and

$R_{\max}$. Section 3 describes an assessment of the performance of candidate imputation models. Section 4 systematically compares the performance of candidate imputation models with existing storm parameter prediction models. Section 4 also investigates the performance of different imputation models and the effect of the data imputation process on the statistical analysis in PCHA. Section 5 provides discussions and conclusions.

2.2 Background and Data

The primary historical database used to develop the joint distribution of storm parameters (e.g., (Nadal-Caraballo et al. 2015, 2020, 2022a; b) is the Hurricane Database 2 (HURDAT2) (Landsea and Franklin 2013; NOAA 2022c). A National Oceanic and Atmospheric Administration's (NOAA) National Hurricane Center Re-Analysis Project product, the HURDAT2 database, contains validated, reanalyzed data of all historical storms recorded in the Atlantic basin from 1851 to 2021. The primary sources for HURDAT2 data include satellite observations, Dvorak analysis, radar data, aircraft reconnaissance, ship records, and buoy and meteorological station data (NOAA 2022c). The database includes track data (typically at a 6-hour time step) related to date, time, latitude (lat), longitude (lon), maximum sustained wind speed (V_{w}), p_{c}, 34 kt wind radii maximum extent (R_{34}), 50 kt wind radii maximum extent (R_{50}), and 64 kt wind radii maximum extent (R_{64}). Data related to storm heading and speed can be derived from temporal and geospatial track information. The record for p_{c} in HURDAT2 is limited prior to 1960, when satellite technology was introduced. Specifically, p_{c} values in HURDAT2 are reported consistently after 1979 (Landsea and Franklin 2013).

In addition to the limited p_{c} observations, another limitation of HURDAT2 is the lack of $R_{\max}$ observations prior to the 2021 update. The $R_{\max}$ parameter, which describes the distance to the strongest wind from the storm's center, provides information on the extent of the wind field for ocean response forcing. Chavas and Knaff (Chavas and Knaff 2022) provided a comprehensive literature review of meteorologists' efforts to understand the properties of R_{max}. The Tropical Cyclone Extended Best Track Dataset (EBTRK) database (CIRA 2021; Demuth et al. 2006) aggregates information from HURDAT2 with other data sources, providing extensive information but is somewhat less consistent. It also includes storm structure information, such as R_{max} and wind radii. In EBTRK, radii data (R_{34}, R_{50}, R_{64}) from 2004 onward are based on the HURDAT2 re-analysis (CIRA 2021). Radii data prior to 2004 are estimated from operational data sources, which include ship and other surface reports, aircraft reconnaissance data, and satellite imagery. The R_{max} observations in EBTRK are all forecaster estimated and only available after 1988 (CIRA 2021).

Figure 2 shows the historical storm data availability in HURDAT2 and EBTRK (using data downloaded on the date of 08/30/2022). Specifically, it shows the fraction of observations associated with a storm that includes data for each parameter. It shows that R_{max} is only available after 1988. p_{c} is only well recorded after 1979, while other parameters have longer coverage and a higher fraction of storms with available data. For example, in the historical databases, over 50,000 V_{w} data points are recorded while only around 23,000 p_{c} data points and around 13,000 R_{max} data points are recorded. Few observations exist for which R_{max} is available but p_{c} is not.

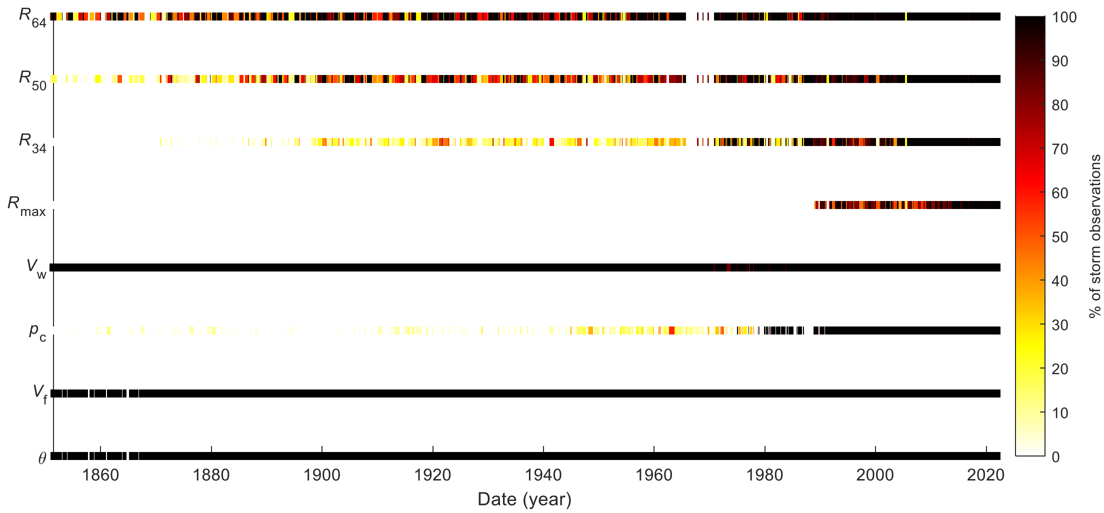


Figure 2: Historical storm data availability, expressed as ratio of track points that include a specified variable for each storm in HURDAT2 and EBTRK

In this study, we develop a series of p_c and R_max machine learning imputation models to impute missing values in historical datasets using available data of other storm parameters. The performance of these models is compared against existing models that relate p_c and R_max to other storm parameters. For example, a strong correlation exists between V_w and p_c and Nadal-Caraballo et al. (Nadal-Caraballo et al. 2015) developed a regression model estimating V_w as a function of p_c. We use the inverse of the regression model developed in Nadal-Caraballo et al. (Nadal-Caraballo et al. 2015) to predict p_c as a function of V_w, which we refer to in this study as "NACCS15." Similarly, Vickery and Wadhera (Vickery and Wadhera 2008) introduced regression models for R_max using Δp and latitude information as predictors. We refer to the "all hurricane" regression model expressed by Equation 11 in Vickery and Wadhera (Vickery and Wadhera 2008) hereafter as "VW08." More recently, Chavas and Knaff (Chavas and Knaff 2022) developed an R_max prediction model based on a physical storm structure model (Chavas et al. 2015), referred to hereafter as "CK22." We assess the performance of our R_max imputation models by comparing to the VW08 and CK22 models (with their regression/model parameters re-estimated for consistency across the model validations, as described in Section 4.2).

2.3 *Imputation Methodology*

The combined database containing the HURDAT2 and EBTRK data is referred to herein as the "original dataset." As noted above, the original dataset contains gaps, typically missing values of p_{c} and $R_{\max}$. A data imputation process, which is used to fill in missing p_{c} and $R_{\max}$ values in the historical record, is proposed based on a series of imputation models and consists of two stages. In the first stage, available storm parameters in the original dataset are used to predict (impute) missing p_{c} values. In the second stage, the original dataset, augmented with imputed p_{c} values, are used to predict (impute) missing $R_{\max}$ values. When missing p_{c} and $R_{\max}$ historical records are filled in using this two-stage process, the data matrix is described herein as the "imputed dataset."

Figure 3 presents a flowchart of the data imputation process. In the "Data Sources Integration" step (Figure 3; left shaded box), identical storms in HURDAT2 and EBTRK are paired. Then data from the two sources are integrated to create the original storm data matrix. In most cases, the timesteps for storm tracks included in HURDAT2 and EBTRK coincide; however, in some cases, values may be available in HURDAT2 at intermediate timesteps that are not available in EBTRK. In these cases, linear interpolation is used to temporally interpolate EBTRK observations onto the same temporal grid as HURDAT2. For each storm track and each timestep included in the HURDAT2 data, the original storm data matrix contains:

- Storm track data point locations (latitude and longitude of the eye of the storm).
- Central pressure (p_{c}) in hPa.

- Maximum sustained wind speed (V_{w}) in km/hr.
- Storm heading (θ) in degrees; the heading assigned to point i is computed as the angle the storm is heading between track point $i-1$ and track point i; θ is measured clockwise in [-180,180] degrees using due North as zero-degree direction.
- Storm forward velocity V_{f} in km/hr; the V_{f} assigned to point i is computed using the distance traveled and the time elapsed between the track point $i-1$ and track point i.
- The radius of maximum winds ($R_{\max}$).
- The four quadrants (northwest, northeast, southeast, and southwest) of R_{34}, R_{50} and R_{64} (in km).

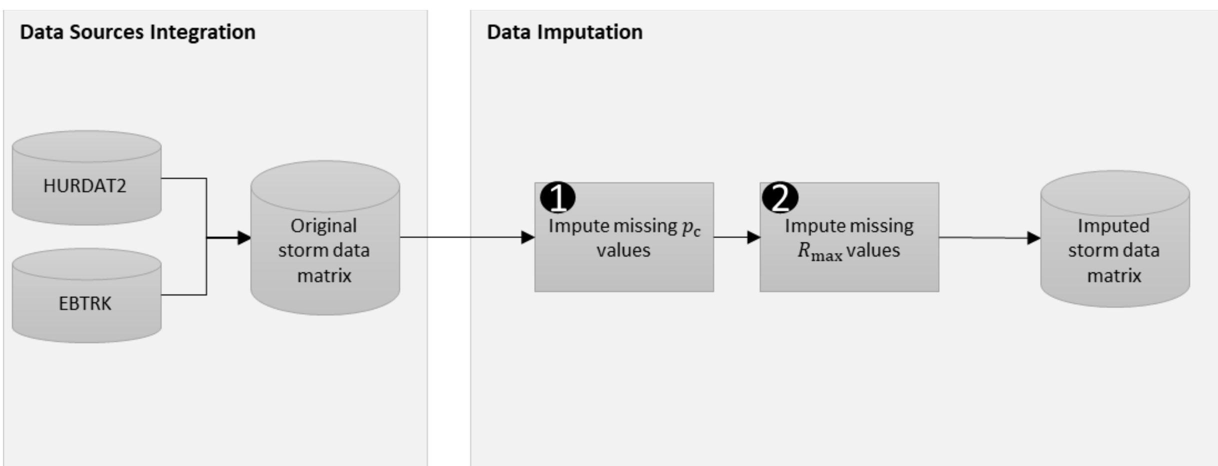


Figure 3: Flowchart of the data imputation process

Next, as illustrated in the right shaded box (labeled Data Imputation) of Figure 3, the original dataset is used to build candidate imputation models for p_{c}, as shown in Box 1. The model takes the form:

$$\hat{p}_{\mathrm{c}}{}_i^{[m,v]} = g_{p_{\mathrm{c}}}^{[m]}(\mathbf{x}_i^{[v]}) \tag{1}$$

Where: ${\hat{p}_{\mathrm{c}}}_i^{[m,v]}$ indicates the imputed value of p_{c} at data point i using model m and input variable parameterization (predictor variables) option v. In this notation, the superscripts in brackets indicate models and variable parameterizations, while subscripts indicate data points $g_{p_{\mathrm{c}}}^{[m]}(\square)$ is imputation model m and $\mathbf{x}_i^{[v]}$ is the vector of storm parameters at data point i using parameterization option v. Similarly, the imputation model for $R_{\max}$ (Box 2 of Figure 3) takes the following form:

$$
{\hat{R}_{\max}}_i^{[m,v]} = g_{R_{\max}}^{[m]}(\mathbf{x}_i^{[v]}) \tag{2}
$$

Where terms are defined analogously to those in Equation (1).

2.3.1 Candidate variable parameterizations

Based on the tropical cyclone parameter coverage, as shown in Figure 2, p_{c} and $R_{\max}$ are generally only available after 1988, while some of the other storm parameters (e.g., lat, lon, V_{f}, V_{w} and θ) have almost full historical record coverage (from 1851 to 2020). Therefore, to impute missing p_{c} and $R_{\max}$ for full historical record coverage, these "full coverage" storm parameters represent the set of potential imputation model predictor parameters.

In Table 1, candidate parameterizations for the p_{c} imputation model are listed, where $\mathbf{x}_i^{[1]}$ is treated as the baseline parameterization. As shown, parameterizations $\mathbf{x}_i^{[2]}$ to $\mathbf{x}_i^{[6]}$ are developed by removing a single individual parameter from the baseline parameterization set ($\mathbf{x}_i^{[1]}$) for each. This provides insights regarding the importance of each of the individual parameters to the performance of the p_{c} imputation model.

Table 1: Tested parameterizations for p_c imputation model

ν	**Parameterization name**	$\mathbf{x}_i^{[\nu]}$ **(included variables)**	**Excluded variables**
1	Baseline	lat, lon, V_f, V_w, $\sin\theta$ and $\cos\theta$[1]	
2	Baseline ***without lat***	lon, V_f, V_w, $\sin\theta$ and $cos\theta$	lat
3	Baseline ***without lon***	lat, V_f, V_w, $\sin\theta$ and $cos\theta$	lon
4	Baseline ***without*** $\boldsymbol{V_f}$	lat, lon, V_w, $\sin\theta$ and $cos\theta$	V_f
5	Baseline ***without*** $\boldsymbol{V_w}$	lat, lon, V_f, $\sin\theta$ and $cos\theta$	V_w
6	Baseline ***without*** $\boldsymbol{\theta}$	lat, lon, V_f, V_w	θ

[1] $\sin\theta$ and $\cos\theta$ are used to account for the circular nature of the variable θ.

For the $R_{\max}$ imputation model, the selection of candidate parameterizations is informed by existing studies. In the VW08 model (Vickery and Wadhera 2008), Δp and lat are used as model predictors, and separate subregional models are developed to improve model performance. In the CK22 model (Chavas and Knaff 2022), a physical model with parameters estimated from regression analysis is used to predict $R_{\max}$. The predictors in the CK22 model include V_w, R_{34} and lat. Values of R_{34} are available for four quadrants. In this paper, and consistent with CK22, the mean value of R_{34} is computed across non- zero quadrants. Hereafter, the symbol R_{34} indicates the mean value of non-zero quadrants. Table 2 shows the candidate parameterizations for $R_{\max}$ imputation model. The record for R_{34} is limited in the historical database (see Figure 2). This adversely affects its value in helping to expand the amount of available data. Therefore, candidate parameterizations are considered both with and without R_{34}.

Table 2: Tested parameterizations for R_{max} imputation model

ν	Parameterization name	$\mathbf{x}_i^{[\nu]}$ (included variables)[1]	Excluded variables
1	Baseline	lat, lon, V_{f}, p_c, V_{w}, sin θ and cos θ, R_{34}	
2	Baseline ***without lat***		lat
3	Baseline ***without lon***	lat, V_{f}, p_c, V_{w}, sin θ and cos θ, R_{34}	lon
4	Baseline ***without*** $\boldsymbol{V}_{\mathbf{f}}$	lat, lon, p_c, V_{w}, sin θ and cos θ, R_{34}	V_{f}
5	Baseline ***without*** $\boldsymbol{p}_{\mathbf{c}}$	lat, lon, V_{f}, V_{w}, sin θ and cos θ, R_{34}	p_c
5	Baseline ***without*** $\boldsymbol{V}_{\mathbf{w}}$	lat, lon, V_{f}, p_c, sin θ and cos θ, R_{34}	V_{w}
6	Baseline ***without*** $\boldsymbol{\theta}$	lat, lon, V_{f}, p_c, V_{w}, R_{34}	θ

[1] For all rows, candidate parameterizations for R_{max} are considered that include and exclude R_{34}.

2.3.2 Candidate machine learning models

The implementation of machine learning techniques in coastal hazard analysis has increased substantially in the past decade. For example, Jia and Taflanidis (Jia and Taflanidis 2013) implemented a Gaussian process regression (GPR) model for storm surge prediction with follow-up studies expanding and augmenting the use of GPR (e.g., (Jia et al. 2016; Taflanidis et al. 2014; Zhang et al. 2018)) or leveraging or comparing alternate approaches (Al Kajbaf and Bensi 2020; Lee et al. 2021). Nadal-Caraballo et al. (Nadal-Caraballo et al. 2020) added GPR-based metamodels to the USACE's CHS-PCHA framework for water level prediction.

In this study, GPR models are investigated as candidate machine learning models for p_{c} and R_{max} imputation. For comparison, the performance of two-hidden-layers forward-propagate artificial neural network (ANN) models for p_{c} and R_{max} imputation are also investigated. To support reproducibility, calculations were performed using standard functions available in the MATLAB Statistics and Machine Learning Toolbox and Deep Learning Toolbox (MathWorks 2022).

For the GPR models, a series of basis functions and kernel functions were tested. The linear basis function and the exponential kernel function (with exponent coefficient fixed to 1) provided by the MATLAB Statistics and Machine Learning Toolbox were found to have robust performance as well as good efficiency. Therefore, this study uses the linear basis function and the exponential kernel function in the training process of all GPR models. To address noise in the observation database, a nugget can be included in the GPR kernel, and its value can be explicitly optimized during the GPR calibration. However, for performance comparison with other candidate models, this was not adopted in this study.

The ANN models implemented in this study use the sigmoid function as the hidden layers transfer function, linear function as the output layer transfer function, Bayesian regularization back-propagation as the regularization strategy, 50 as the training epochs number, and ten as neurons number in each hidden layer. The epoch number and neuron number in each hidden layer were set to optimize the performance of ANN models.

2.3.3 Model validation

K-fold cross-validation is applied to assess the performance of the GPR and ANN imputation models with different parameterizations. The validation uses the full dataset, filtered to include only observations with all necessary parameters available.

For the cross-validation, the data is randomly split into $k = 10$ equal-size sets (folds). For each fold, one of the ten partitions is reserved as a testing set, while the remaining nine sets are treated as training data. All options in the candidate suite (i.e., combinations of candidate parameterizations and candidate machine learning models) are tested on the same set of cross-validation partitions. This ensures that all models are trained and tested on the same data and assures a "head-to-head" comparison. Within each fold, the trained model is applied to the testing set, and the predictions are compared to the held-out testing data to calculate the root mean square error (RMSE) and the correlation coefficient (R). RMSE is calculated using the difference between the predicted and holdout values of p_c (or R_{max}). R is computed as the correlation coefficient between the predicted and holdout values of p_c (or R_{max}).

Figure 4 shows the cross-validation results for the GPR and ANN p_c imputation models under the candidate parameterizations. The bars in Figure 4 (top) show the average value of RMSE (ARMSE) computed across the $k = 10$ cross-validation folds for GPR and ANN models. Figure 4 (bottom) similarly shows average values of R ($R_{averaged}$) of each model. The vertical red line sections indicate the range of RMSE and R computed across the the k=10 cross-validation folds. When imputing p_c, removing V_w has a substantial impact on model performance, which is expected due to the comparatively high correlation between V_w and p_c (Nadal-Caraballo et al.

2015). The impact of removing lat, lon, V_{f} or θ is not as notable. The performance of the baseline parameterization models is later compared with the NACCS15 model for predicting p_{c}.

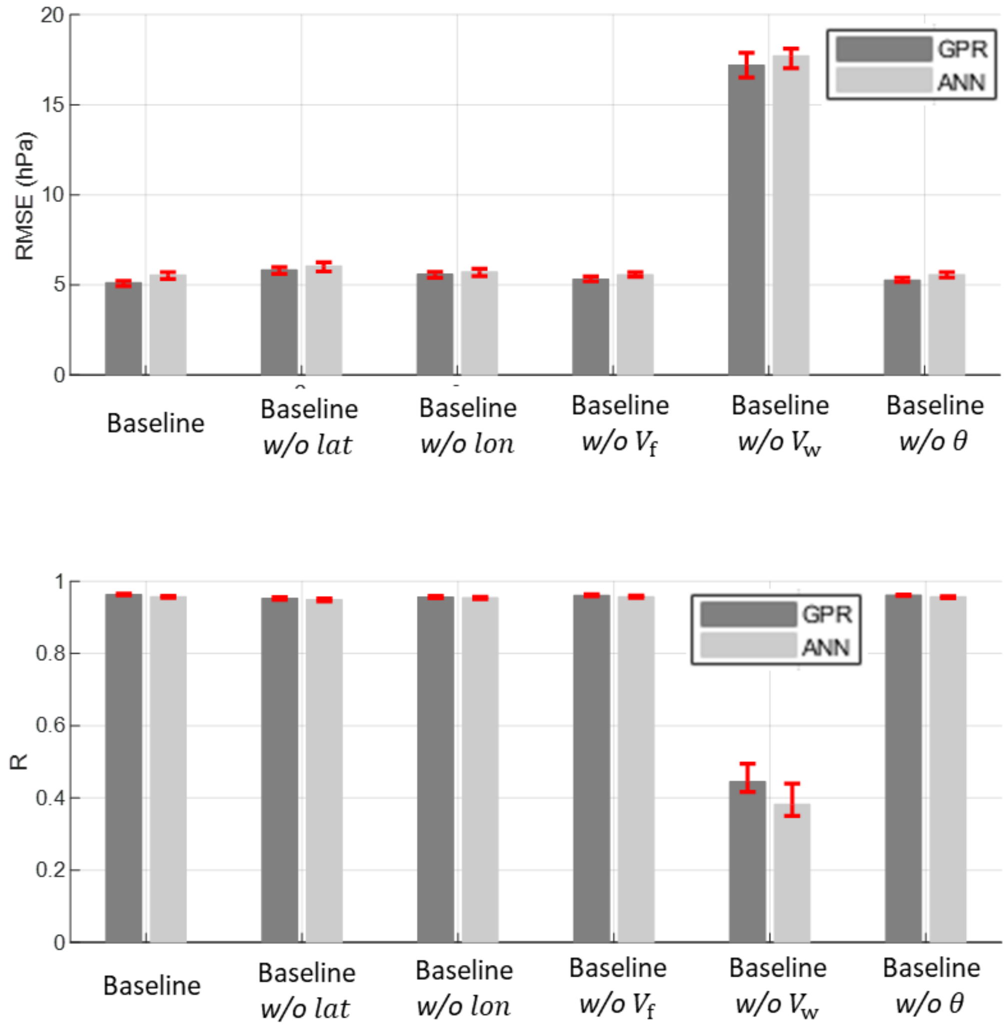
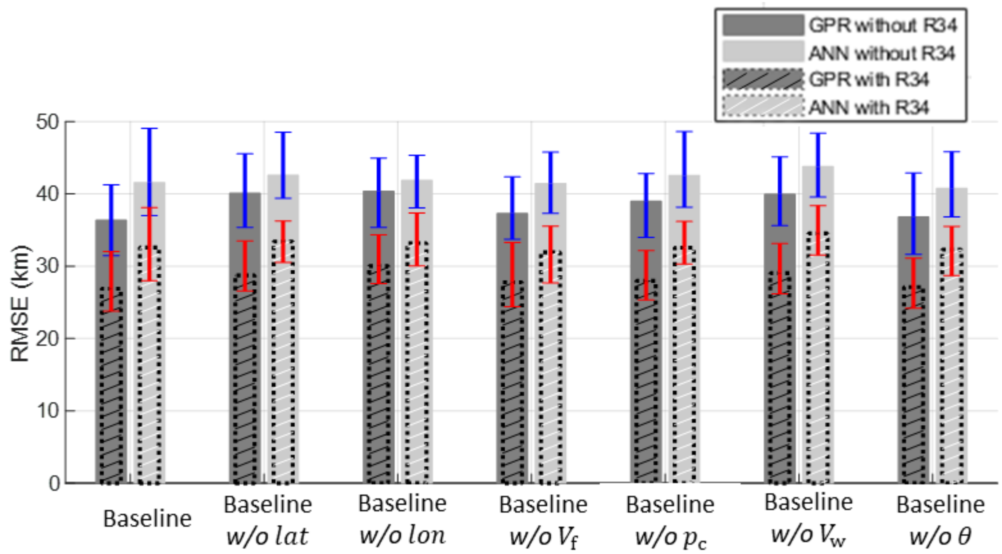


Figure 4: Cross-validation results (top: RMSE; bottom: R) for p_{c} imputation models with different parameterizations. Bar heights indicate average values of metrics, and red line sections show the ranges of computed metrics computed across validation folds

Figure 5 shows the cross-validation results for the $R_{\max}$ imputation models, with and without R_{34}.The performance of $R_{\max}$ imputation models is not as strong as seen for p_{c} imputations, which is seen from the lower R values and reflects the experience that $R_{\max}$ is harder to estimate. Regardless of the parameterization selected, the inclusion of R_{34} has a notable positive impact on the performance imputation models (i.e., the bars without dashes are systematically higher than those with dashed edges). When R_{34} is included in the model, the impact of removing other parameters is relatively minor, as seen by the similarity in heights of the bars with dashed edges. However, when R_{34} is not included, the removal of spatial information (lat/lon) and windspeed have moderate bad effect on performance.



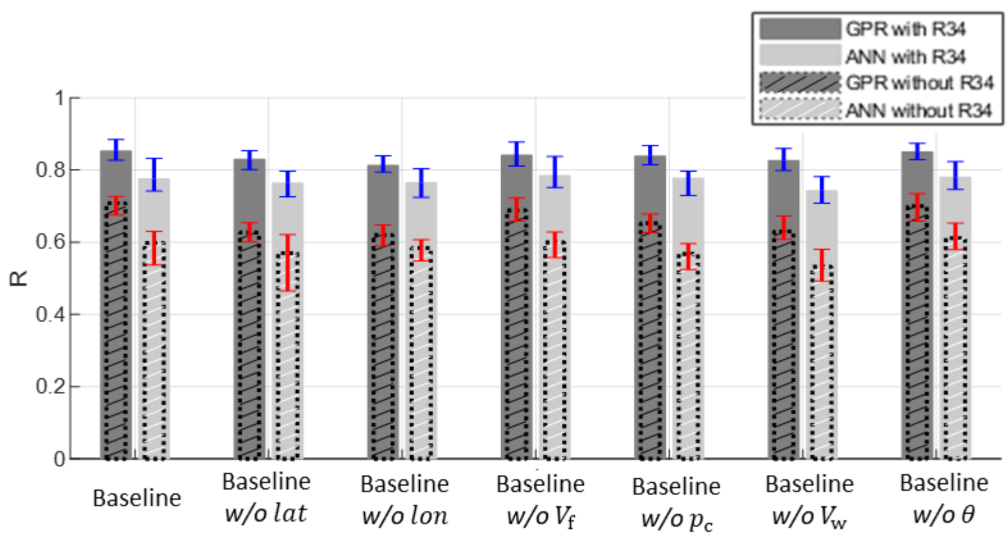


Figure 5: Cross-validation results (top: RMSE; bottom: R) for $R_{\max}$ imputation models of different parameterizations with and without R_{34}

Note: For graphical clarity, the colors/legends differ between the top (RMSE) and bottom (R) figures.

While the inclusion of R_{34} as a predictor results in the best performance, it is important to recognize that R_{34} is only available for a limited portion of the historical record (see Figure 2). Therefore, in the analysis that follows, the performance of both the baseline and "baseline without R_{34}" parameterization models are compared with the VW08 and CK22 models for predicting $R_{\max}$.

2.4 Performance and Effect of Data Imputation

In this section, the performance of machine learning-based imputation models is assessed. In Section 4.1, p_c machine learning-based imputation models were compared with NACCS15. In Section 4.2, $R_{\max}$ machine learning-based imputation models were compared with VW08 and CK22. In Section 4.3, statistical analysis results generated from the original dataset and imputed dataset are visualized and quantified to assess the effect of the data imputation process.

2.4.1 p_c model performance assessment

To assess the performance of p_c imputation models, the same set of 10-fold cross-validation were applied to the two candidate machine learning models in the North Atlantic region (GPR and ANN with baseline parameterizations) as well as NACCS15 (i.e., the inverse of the Nadal-Caraballo et al. (Nadal-Caraballo et al. 2015) regression model). Figure 6 plots the RMSE and R

of the p_c models for the North Atlantic region. Among the three p_c models, the GPR baseline p_c imputation model has the highest accuracy (i.e., the lowest ARMSE and highest R_{averaged}).

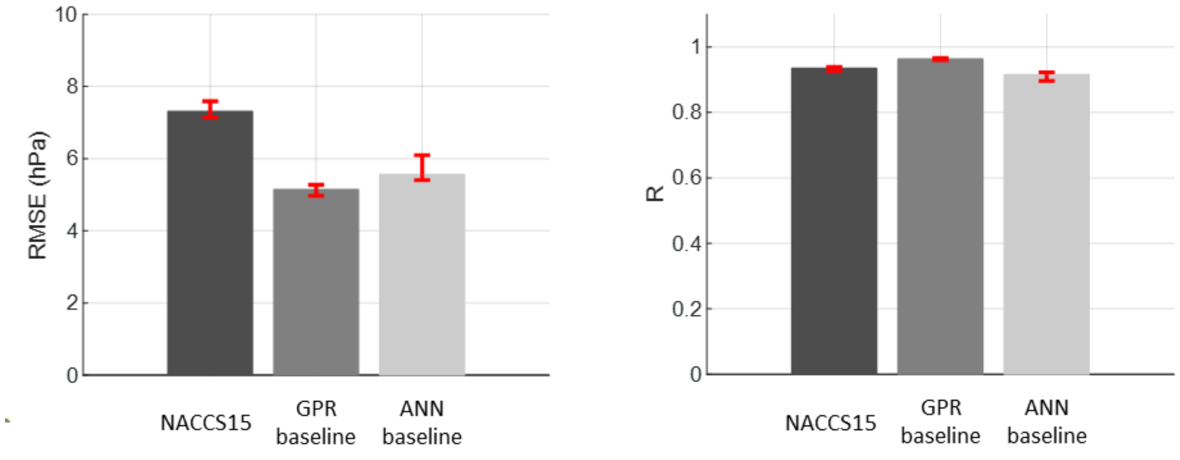


Figure 6: RMSE and R for candidate p_c models. The bars show ARMSE and R_{averaged} of each model, while the red line sections show the range of RMSE and R of each model

Figure 7 presents density plots (2D histograms) of the predicted (imputed) and observed values for the testing data across the 10-fold cross-validation results for the three p_c models. In the density plots, the x-axis shows the observation value (e.g., as reflected in the original dataset), the y-axis shows the prediction value from the imputation model, and the color bar shows the number of observations within the specified bin. The color bar is presented in log-scale to aid visualization. The NACCS15 model generates gridded prediction values that are expected because NACCS15 model's single input, V_w, is recorded in increments of five knots. It also can be observed that in all three models, the predicted p_c are predominantly symmetrically distributed along the 45-degree line, which means the p_c model has a generally consistent performance across the range of p_c values.

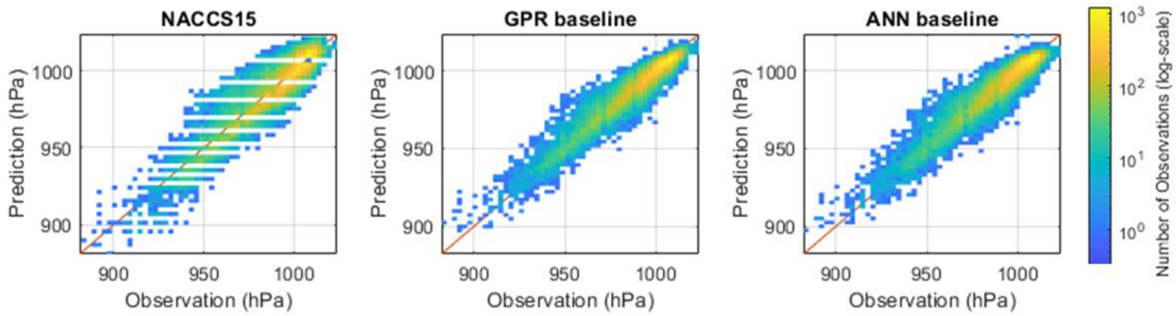


Figure 7: Density plot of cross-validation results for candidate p_c models

2.4.2 R_{max} model performance assessment

To assess the performance of the R_{max} imputation models, six models are compared using the cross-validation approach described in Section 3.3: (1) VW08, (2) CK22, (3) GPR baseline, (4) GPR baseline without R_{34}, (5) ANN baseline, and (6) ANN baseline without R_{34}. The models are applied to three study cases. The first case includes all data points within the North Atlantic region. The second case refines the geographic region to focus on the Gulf of Mexico. The third case considers only high-intensity storm data points within the North Atlantic region, consistent with data used in VW08. Table 3 lists the three cases, their geographic boundaries, and the associated number of data observations available for analysis.

Table 3: Study cases used for R_{max} model performance assessment

Study case	Bounding box	# points
North Atlantic region (Case 1)	Lat: 69° N to 0° N; Lon: 12° E to 98° W	9833
Gulf of Mexico region (Case 2)	Lat: 30° N to 18° N; Lon: 82° W to 98° W	1326

North Atlantic region with p_c < 980 hPa (Case 3)	Lat: 69° N to 0° N; Lon: 12° E to 98° W	2880

We compare the six models mentioned above using the K-fold cross-validation approach as described in Section 3.3. The VW08 and CK22 model functions are used with model parameters (e.g., regression coefficients) re-estimated for each study case using the training data contained in each validation fold (note these parameters may therefore differ from the parameters of a specific published model).

The performance of the six $R_{\max}$ models are compared using three criteria:

(1) Model accuracy quantified by model ARMSE and R_{averaged}

(2) Model capability to predict large $R_{\max}$ values

(3) Effect of models on the correlation between Δp and $R_{\max}$

2.4.2.1. Model accuracy

Figure 8 plots the RMSE and R of candidate $R_{\max}$ models for the three study cases. In all three study cases, it is found that the GPR baseline imputation model has the minimum ARMSE and the maximum R_{averaged} among all tested $R_{\max}$ models. Among different study cases, the size of the study region, comparing the larger study case 1 with the smaller study case 2, has a moderate effect on the accuracy of different $R_{\max}$ models, with the smaller study region resulting in a slight reduction in performance when measured using ARMSE. This is likely due to the disadvantage of the smaller quantity of training data exceeding the advantage of the expected geographic homogeneity in the data. Applying models to a subset of high-intensity storms (i.e., comparing study case 1 with study case 3) improves the model performance. This is consistent with the field's understanding that different intensities of tropical cyclones are connected to different environmental conditions (Emanuel et al. 2004) and are sometimes treated as different populations. However, focusing on high-intensity storms limits the number of data points for which $R_{\max}$ can be imputed.

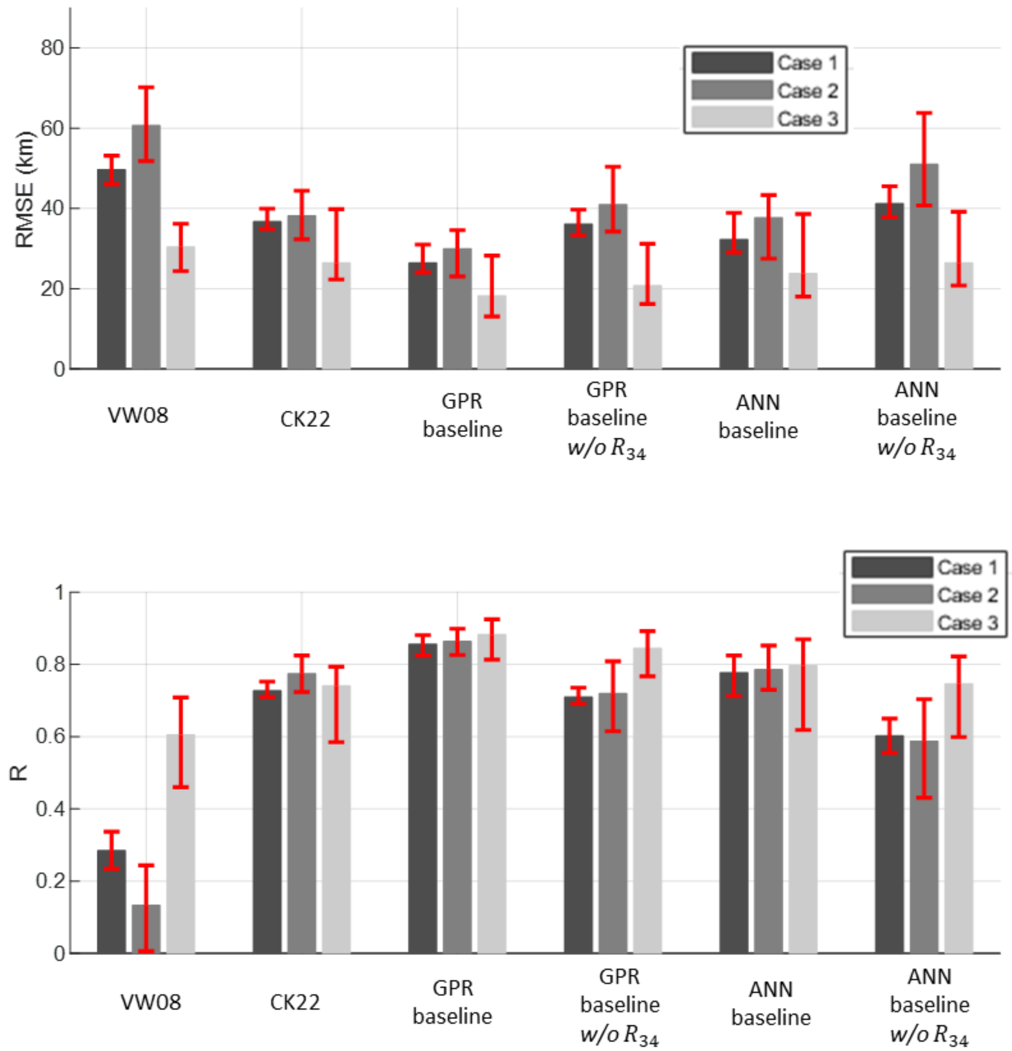


Figure 8: RMSE and R of candidate $R_{\max}$ models. The bars show ARMSE and R_{averaged} of each model, while the red line sections show the range of RMSE and R of each model

2.4.2.2. Model capability to predict large $R_{\max}$ values

Figure 9 through Figure 11 present density plots (2D histograms) of the 10-fold cross-validation results of the candidate $R_{\max}$ models applied to the three study cases. In the density plots, the x-axis shows the observation bin center values (e.g., as reflected in the original dataset), the y-axis shows the prediction bin center values from the imputation model, and the color bar shows the number of observations in the bin. The VW08 model shows a notable upper bound on predicted $R_{\max}$ values for the first two study cases. This upper bound is not observed for the third study case, where the value of p_c is restricted. These results are consistent with the conditions under which the VW08 model was developed, which only include storms with Δp> 33 hPa. For all study cases, the GPR baseline without R_{34} model likewise shows challenges in predicting large $R_{\max}$ values, though the effect is not as pronounced as for the VW08 model. The CK22 and ANN models do not show similar issues related to large $R_{\max}$ values.

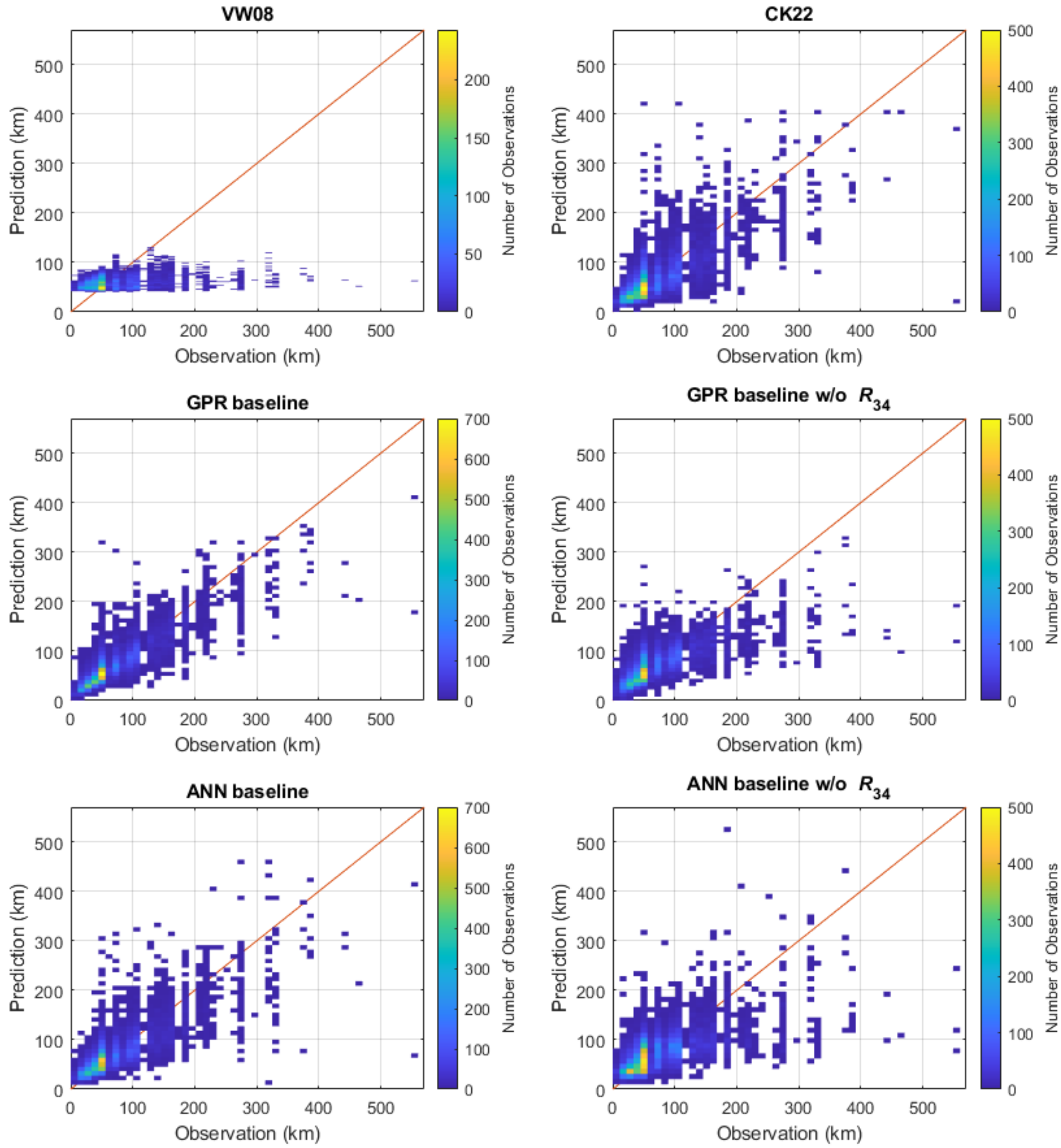


Figure 9: Study Case 1, North Atlantic region; Density plot of cross-validation results for candidate $R_{\max}$ models (note that color scales differ across plots)

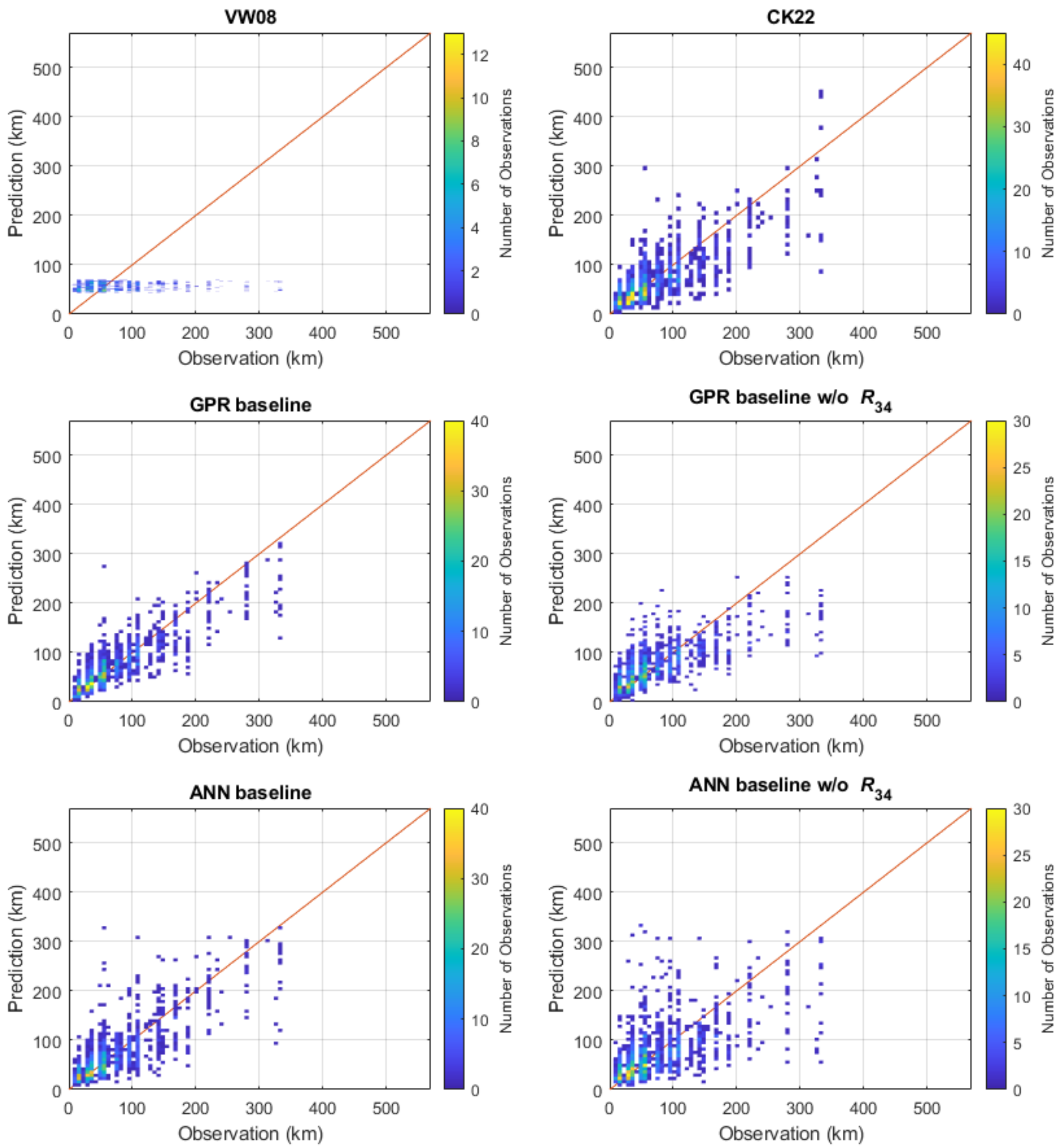


Figure 10: Study Case 2, Gulf of Mexico region; Density plot of cross-validation results for candidate $R_{\max}$ models (note that color scales differ across plots)

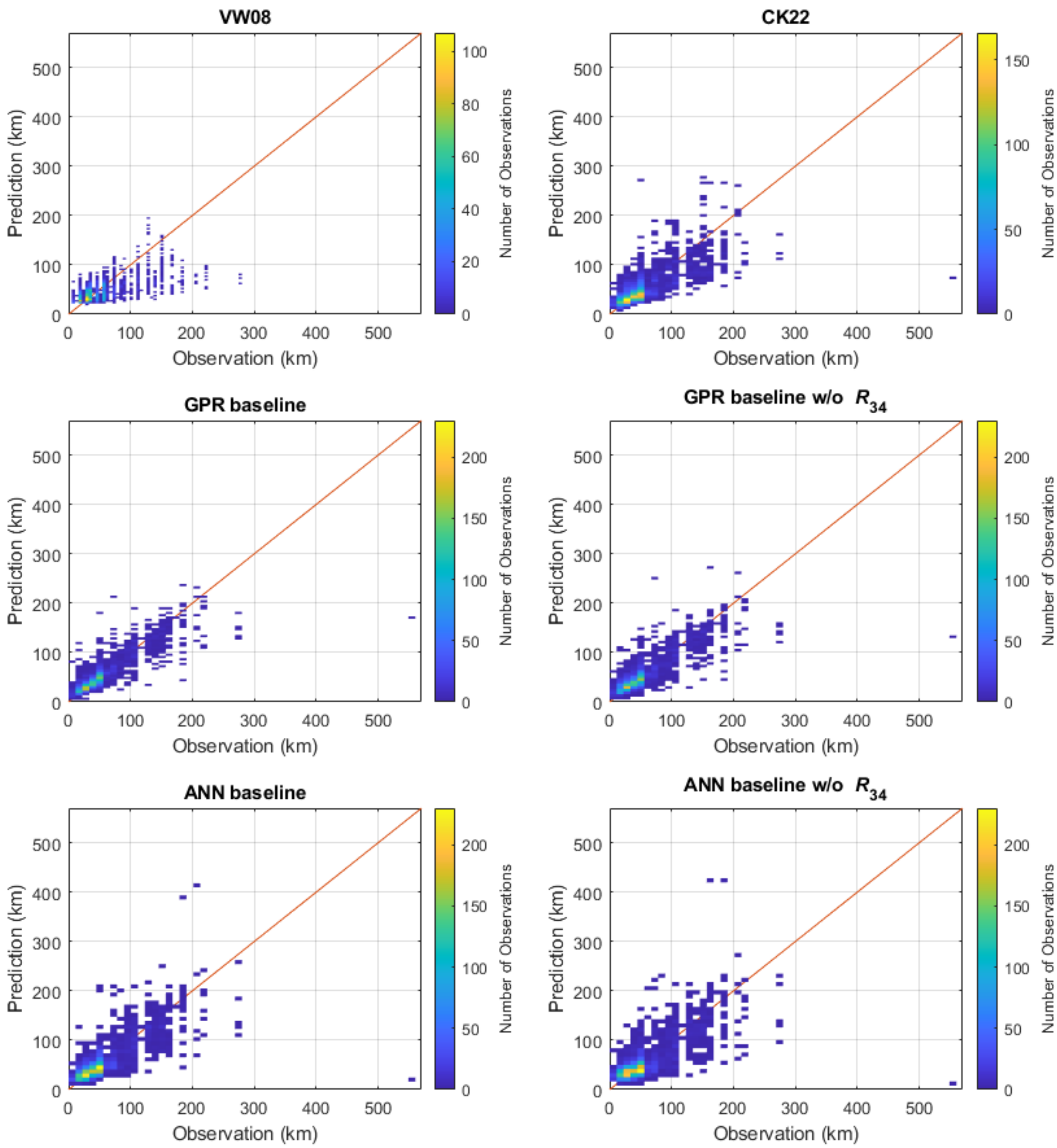
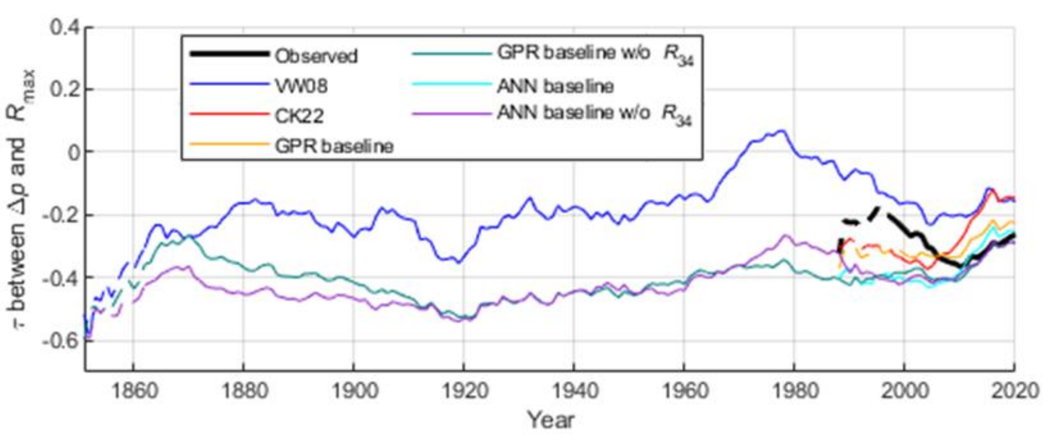


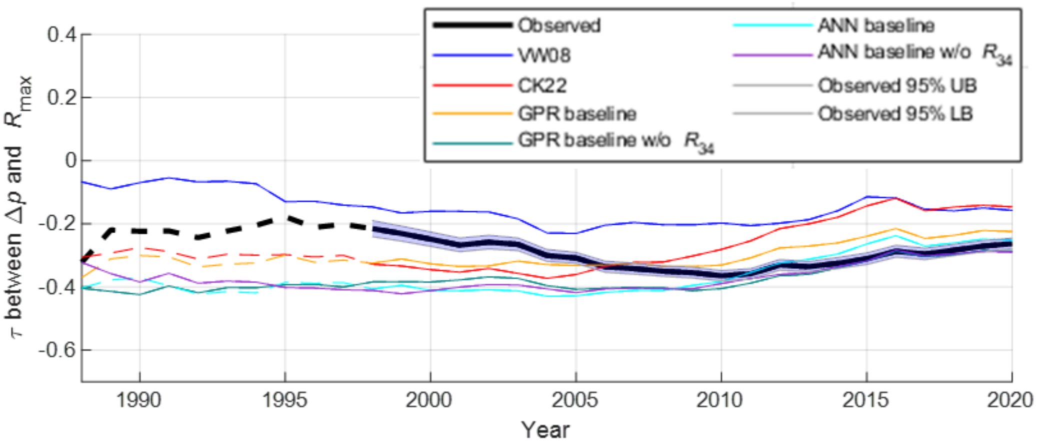
Figure 11: Study Case 3, North Atlantic region with p_c smaller than 980 hPa; Density plot of cross-validation results for candidate $R_{\max}$ models (note that color scales differ across plots)

2.4.2.3. Effect on the correlation between Δp and $R_{\max}$

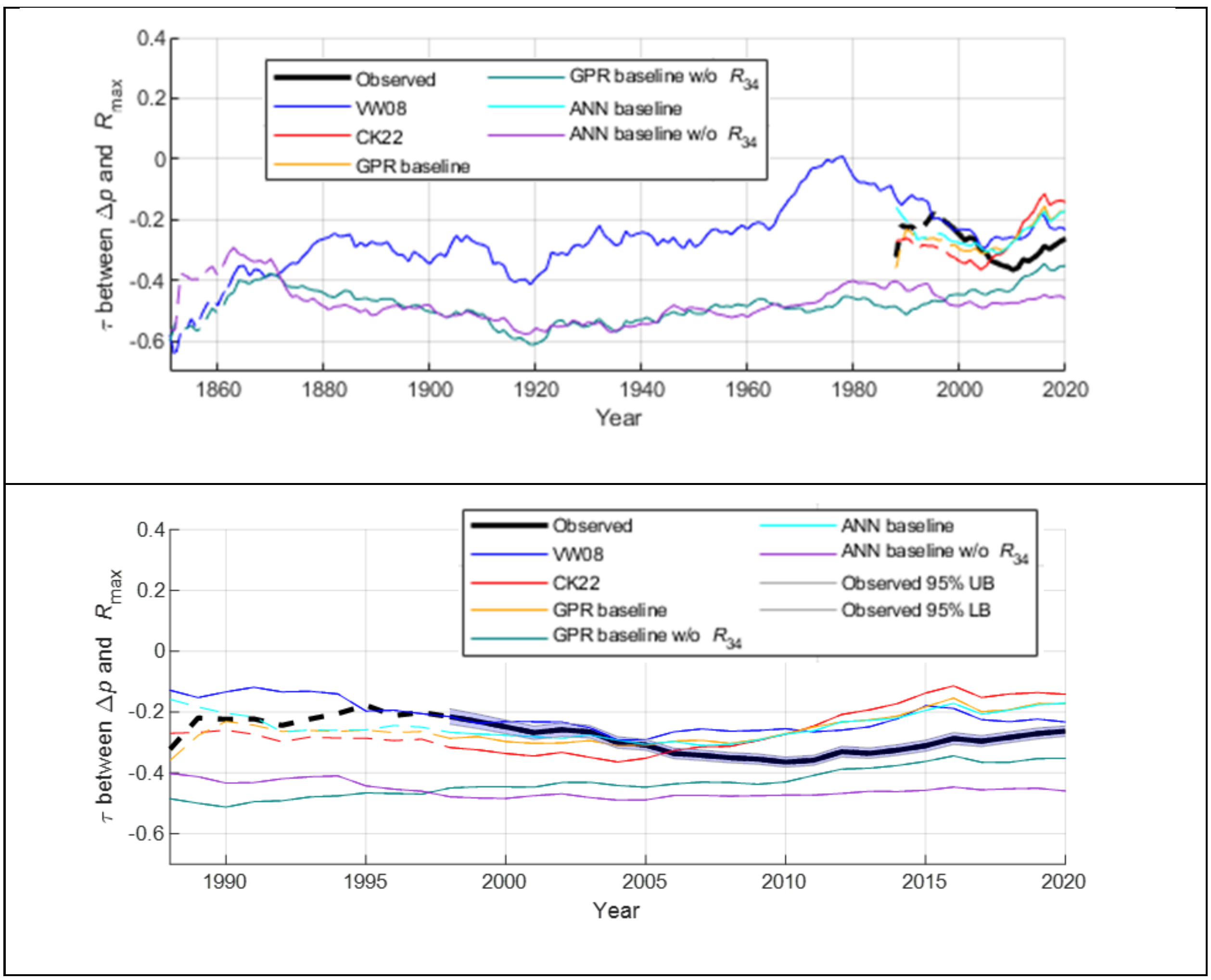
For PCHA studies, particularly those focused on storm surge hazards, Δp and $R_{\max}$ are key parameters. Tropical cyclone parameter combinations involving larger values of Δp and $R_{\max}$ typically cause higher surges. However, negative correlation between them makes those combinations less likely to occur. For this reason, it is important to understand the impact of the data imputation process on the correlation between these two quantities.

Figure 12 plots the values of Kendall's τ between Δp and $R_{\max}$ (computed using a ten-year moving window) vs. time for each candidate $R_{\max}$ model as well as when using the original dataset. This study calculates Δp as $\Delta p = p_{\mathrm{p}} - p_{\mathrm{c}}$, where p_p is peripheral pressure of the storm (taken as 1013 hPa). Correlations for the observed data are only available after 1988, when $R_{\max}$ data became available. Before 1988, only the VW08, GPR w/o R34 and ANN w/o R34 models were assessed and plotted, as they do not require R34, which likewise became available after 1988. Therefore, time-series plots are provided for each study case for time horizons going back to 1851 (top plot) and beginning in 1988 (bottom plot). In the bottom plots, the 95% confidence interval of observed data is generated using the MATLAB bootstrap confidence interval function (MathWorks 2022).

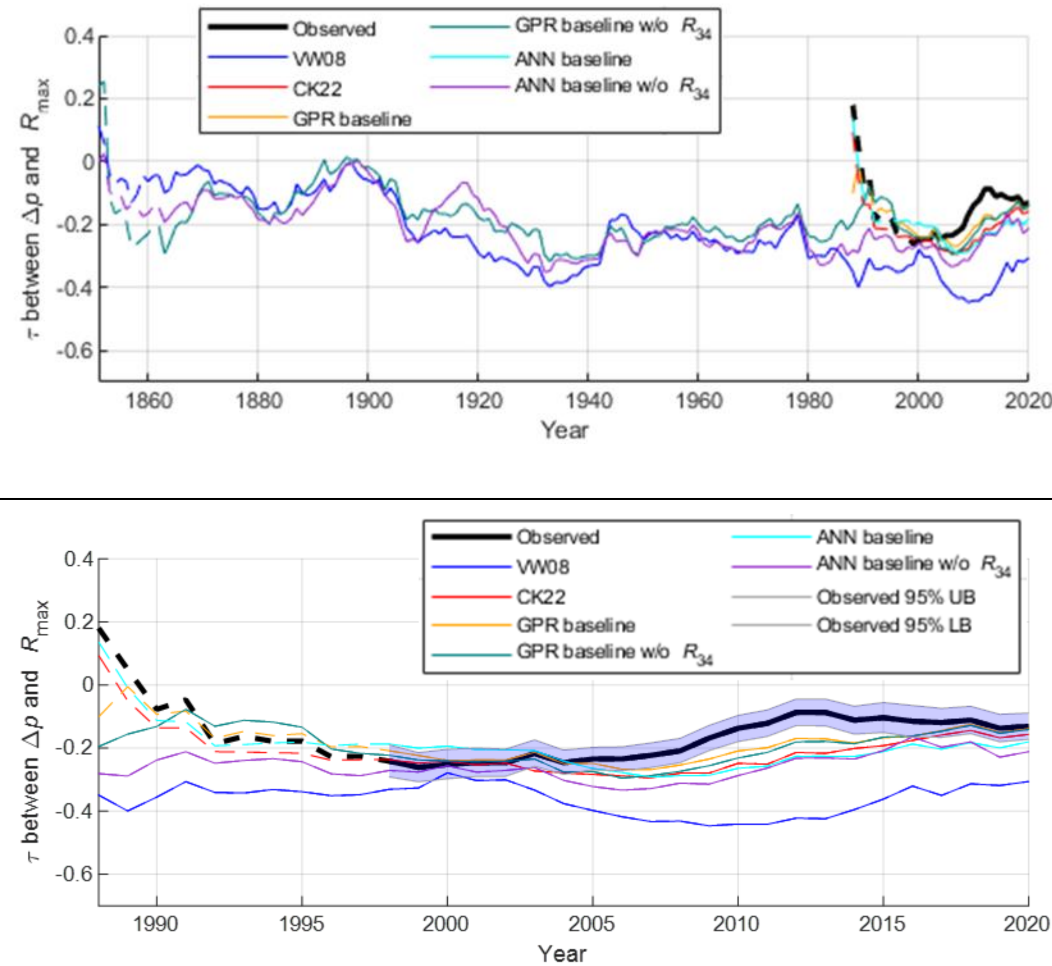




(a) Study Case 1, North Atlantic region



(b) Study Case 2, Gulf of Mexico region



(c) Study Case 3, North Atlantic region with p_c smaller than 980 hPa

Figure 12: Time-series of Kendall's τ between Δp and R_{max} for candidate R_{max} models for three study cases. For each study case, the top figure plots curves for 1851 to 2020, and the bottom figure plots the curves for 1988 to 2020. For each point, data from the previous ten years are used to compute Kendall's τ. When the interval of data used for computation is less than ten years, a dashed line is used. In the process of computing Kendall τ, the imputed dataset includes imputed p_c values. In the bottom plots, the 95% confidence interval of observed data after 1998 are plotted in the light blue shaded region

To quantify the difference between modeled values and observed values of Kendall's τ between Δp and $R_{\max}$, the RMSE between each modeled value and observed value are calculated using all data from 1988 to 2020. The calculated RMSEs are presented in Table 4.

Table 4: RMSE on Kendall's τ (between Δp and $R_{\max}$) computed using imputed and observed datasets

	VW08	CK22	GPR baseline	GPR baseline w/o R_{34}	ANN baseline	ANN baseline w/o R_{34}
Study Case 1	0.14	0.099	0.072	0.12	0.13	0.12
Study Case 2	0.078	0.098	0.072	0.18	0.072	0.19
Study Case 3	0.23	0.068	0.065	0.090	0.073	0.14

Based on the plots and RMSEs for study case 3 (except the VW08 model), Kendall's τ values for the imputed dataset are more consistent with the observed dataset's correlation values than in study case 1 and study case 2. Assuming the observed dataset correlation is more representative of the 'true' correlation, these results indicate the performance of models will be improved if only high-intensity storms are included in the training set (with the disadvantage that such a model cannot be used to impute data for lower-intensity storms). It is also noticed that in all three study cases, Kendall's τ modeled by the GPR baseline without R_{34} and ANN baseline without R_{34} models (which can be used prior to 1988) are generally consistent with each other through time. Based on the RMSE value in Table 4, the GPR baseline $R_{\max}$ imputation model can best capture the value and trend of the observed Kendall's τ quantifying the correlation between Δp and $R_{\max}$.

2.4.2.4. Assessment results

Overall, the GPR baseline imputation model is judged as the best performing $R_{\max}$ imputation model based on its ARMSE and the ability to capture the observed correlation between Δp and $R_{\max}$. However, the limited number of R_{34} records in the historical database restricts the use of the baseline GPR model to impute records prior to 1988. Among the models not requiring R_{34} data, the GPR baseline without R_{34} imputation model offers the best performance, based on ARMSE and the ability to capture the correlation between Δp and $R_{\max}$. However, it is associated with a notable challenge in predicting large $R_{\max}$ values. The ANN baseline without R_{34} imputation model offers reduced performance compared to the GPR baseline without R_{34} imputation model from the perspective of ARMSE but is better able to capture large $R_{\max}$ values. It also performs well in capturing correlations.

2.4.3 Imputed dataset assessment

In this section, the data imputation process using a subset of candidate models is applied to the historical storm dataset described in Section 2. The GPR baseline parameterization p_{c} imputation model is used to impute around 30,000 missing p_{c} values in the dataset. The GPR baseline parameterization $R_{\max}$ imputation model is used to impute over 700 missing $R_{\max}$ at data points with R_{34} records available. The GPR baseline without R_{34} parameterization $R_{\max}$ imputation model is used to impute around 40,000 missing $R_{\max}$ at data points for which R_{34} records are not available. In the process of imputing $R_{\max}$, the imputed p_{c} values are used as input data for the $R_{\max}$ imputation models.

In the USACE's CHS-PCHA framework (Nadal-Caraballo et al. 2020), a series of coastline nodes are selected as coastal reference locations (CRLs), and statistical analysis is performed for each CRL individually. The statistical sampling data of each CRL is generated with a distance-weight-adjustment process using representative storm track points with the maximum intensity indices within a 600 km radius capture zone (Nadal-Caraballo et al. 2015, 2022b). Information about the statistical sampling process is provided in Appendix A. Data resulting from the application of this process will be referred to as "statistical sampling data" in this paper.

Three representative CRLs along the Atlantic and Gulf coastline are selected to assess the effect of the data imputation on the resulting statistical sampling data. Figure 13 shows the location of the three CRLs and their 600 km radius capture zone. Consistent with recent CHS studies (Nadal-Caraballo et al. 2022a; b), storm data from 1938 to 2020 are used to generate statistical sampling data.

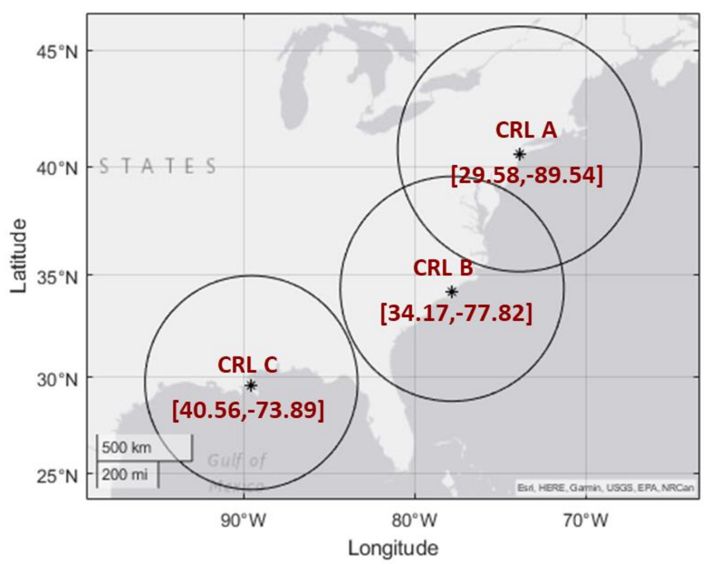


Figure 13: Location of example CRLs with their 600 km statistical analysis capture zones

Figure 14 to 15 show the comparison of statistical sampling data generated from the original and imputed datasets (with adjustments made according to the distance-weight adjustment process). The data imputation process increased the quantity of both Δp and $R_{\max}$ statistical sampling data and consequently smoothed the empirical cumulative density functions (eCDF) curves for Δp and $R_{\max}$. This effect is especially significant for $R_{\max}$ because, prior to 1988, there is no $R_{\max}$ record in original dataset, the increase of statistical sampling data can improve the robustness of fitted marginal distribution, regardless of the choice of distribution function form. A slight reduction of variance is observed, as evidenced by the steeper slope of the eCDF for the imputed dataset, reflecting the reduced variability that results from the use of model output in place of a missing observation.

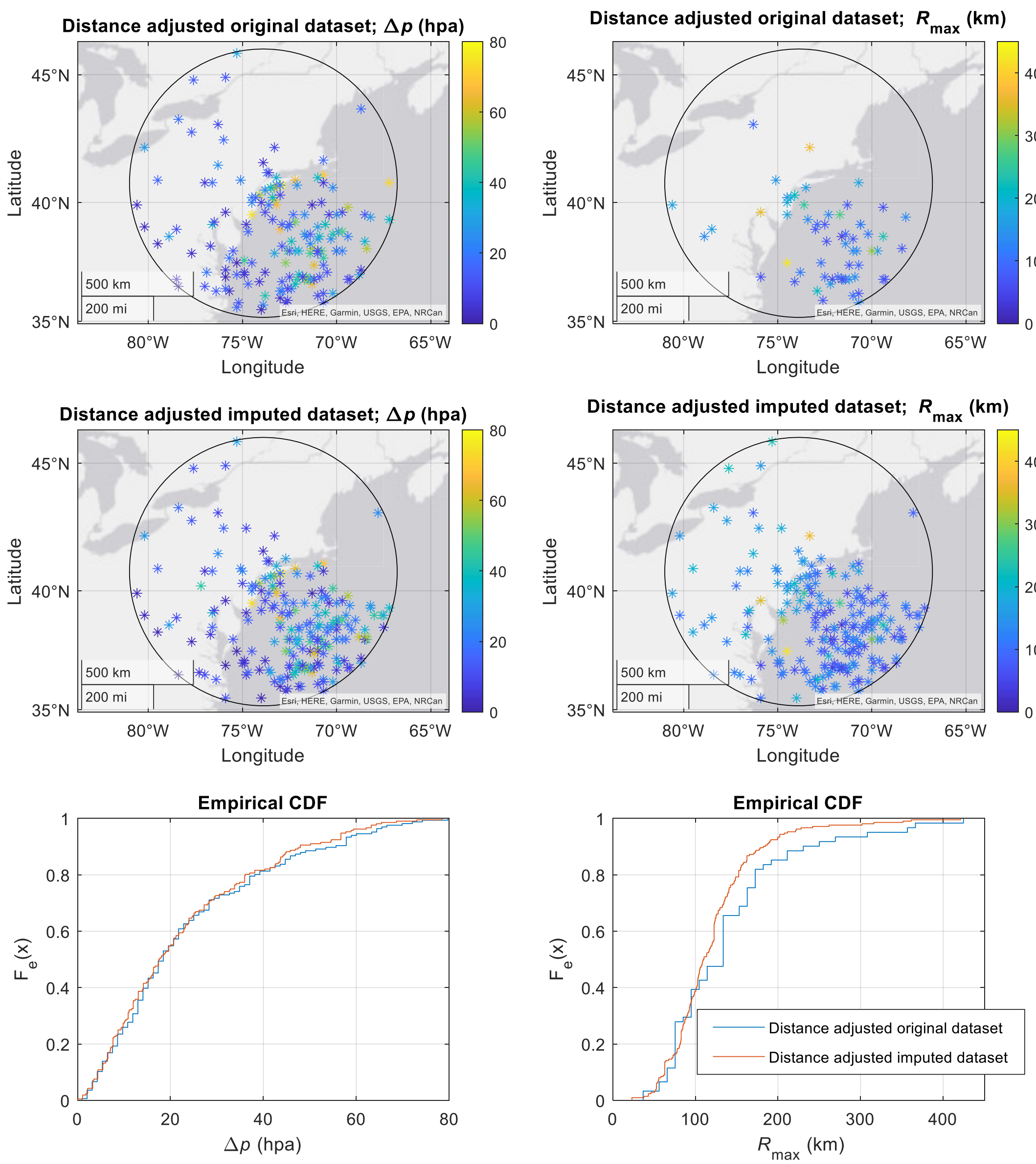


Figure 14: Statistical sampling data for CRL A; The scatter plot maps show the statistical sampling data locations (lat/lon) and values (color bar) for Δp and $R_{\max}$ using the original dataset and imputed dataset. The bottom row plots the eCDF of Δp and $R_{\max}$ from statistical sampling data generated from the original dataset and imputed dataset

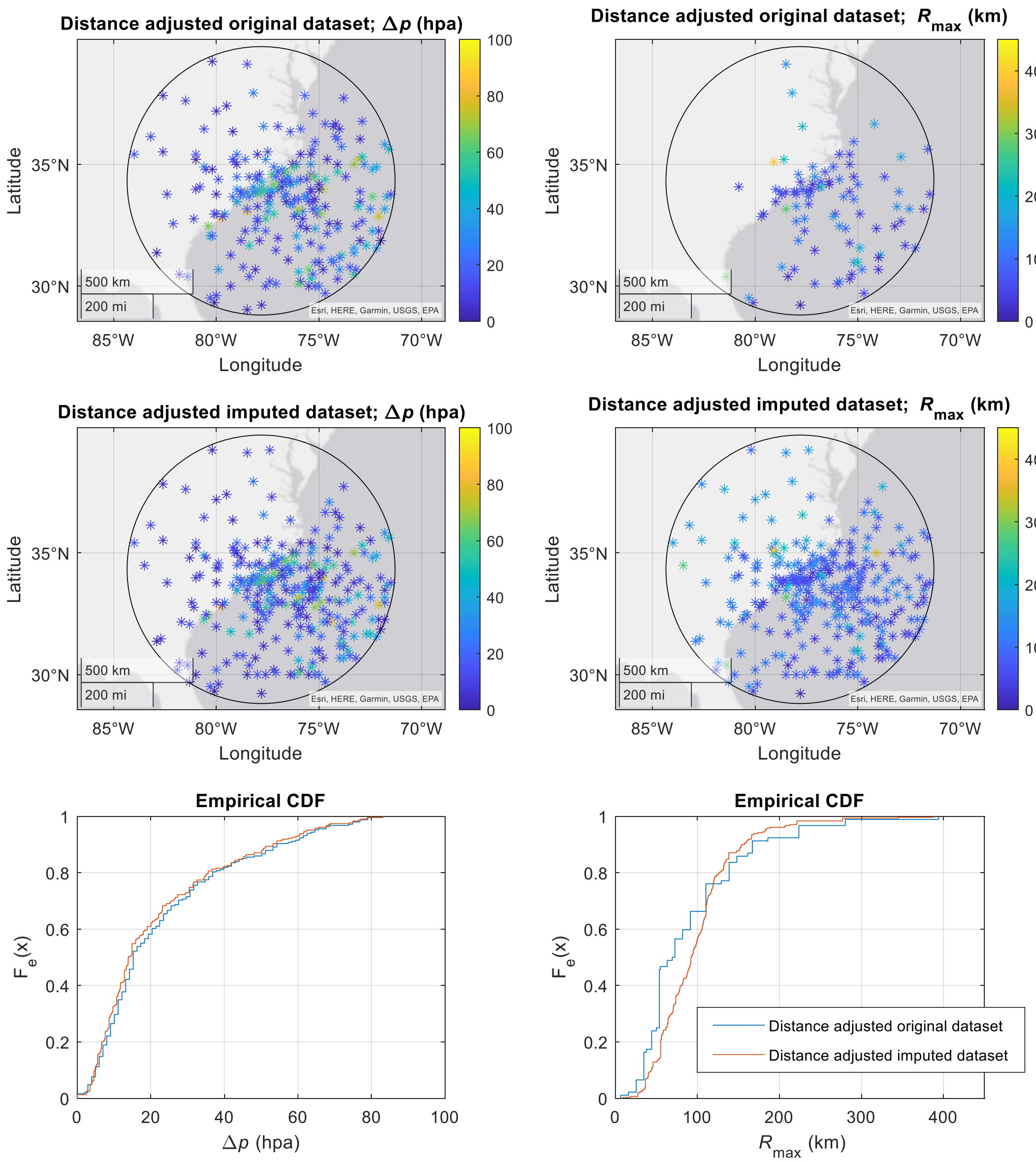


Figure 15: Statistical sampling data for CRL B; The scatter plot maps show the statistical sampling data locations (lat/lon) and values (color bar) for Δp and $R_{\max}$ using the original dataset and imputed dataset. The bottom row plots the eCDF of Δp and $R_{\max}$ from statistical sampling data generated from the original dataset and imputed dataset

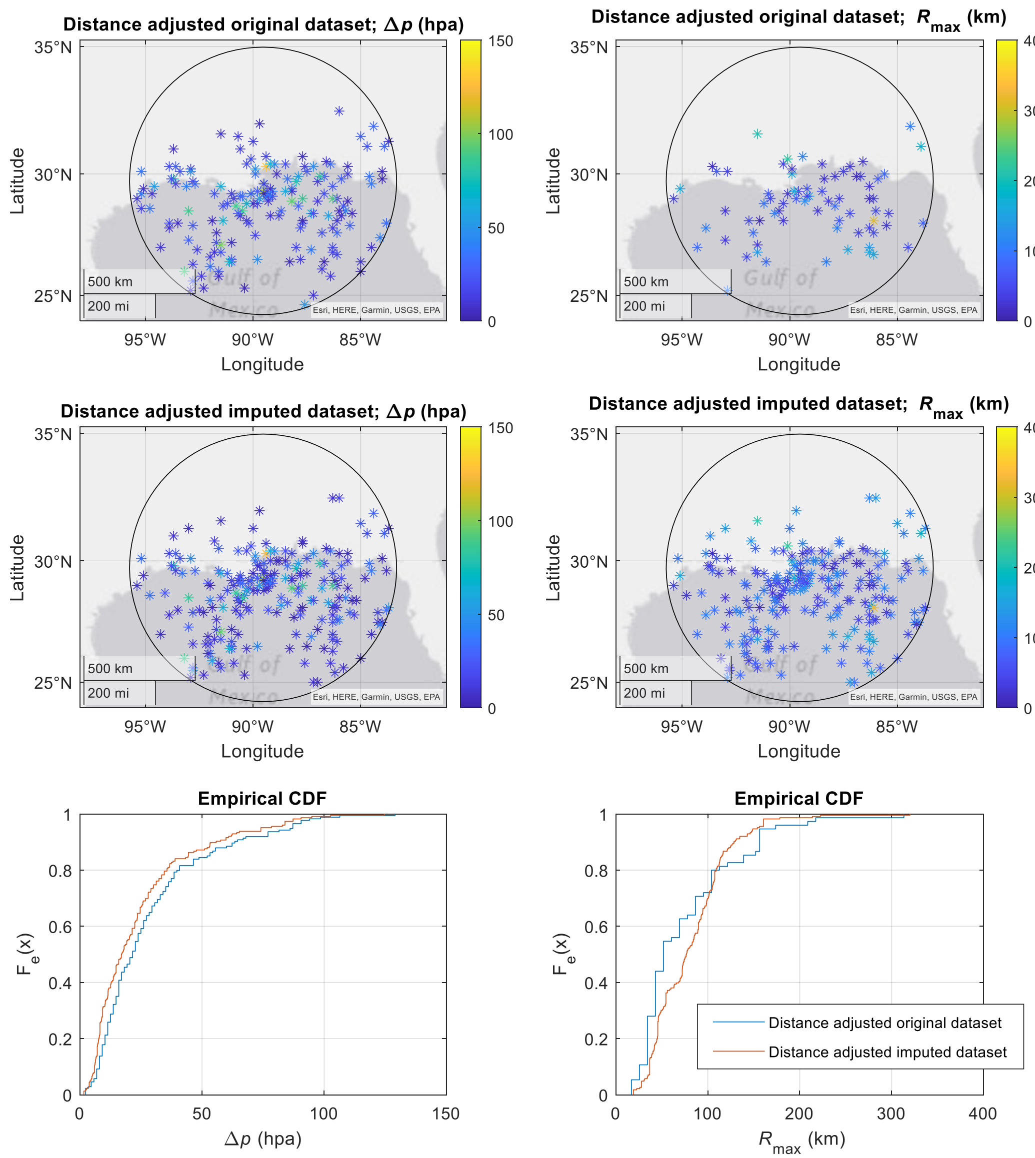


Figure 16: Statistical sampling data for CRL C; The scatter plot maps show the statistical sampling data locations (lat/lon) and values (color bar) for Δp and $R_{\max}$ using the original dataset and imputed dataset. The bottom row plots the eCDF of Δp and $R_{\max}$ from statistical sampling data generated from the original dataset and imputed dataset

In the USACE's CHS-PCHA framework (Nadal-Caraballo et al. 2020), copulas are used for constructing joint distribution models. A copula's parameters are often estimated from measured correlation (Kendall's τ) between storm atmospheric parameters. Here we compute the measured Kendall's τ between storm atmospheric parameters from the statistical sampling data generated from the original dataset and imputed dataset for the three CRLs. The calculated Kendall's τ are shown in .

Table 5. Here, we focus on Kendall's τ between Δp and R_{max} due to the impact of these parameters on surge generation. The 95% bootstrap confidence intervals (MathWorks 2022) are calculated and added for the Kendall's τ between Δp and R_{max}. It can be observed that, even when only using data within the time period of 1988~2020 (during which only a small part of Δp and R_{max} are imputed), the data imputation process results in a more negative correlation between Δp and R_{max} than the original dataset (noting that the imputation model for R_{max} includes p_{c} as an input parameter (predictor)). The enhanced negative correlation can likely be attributed to the lack of observational noise in data imputation. The data imputation process also reduces the width of the confidence interval of Kendall's τ between Δp and R_{max}. When using the imputed dataset for storm surge frequency analysis, a stronger negative correlation between Δp and R_{max} may reduce the likelihood of large Δp and R_{max} combinations. The potential tradeoffs between the larger sample size and increased correlation strength in coastal hazard analysis is an area for future work.

Table 5: Comparison of Kendall's τ of Δp and R_{max} computed from the original dataset and imputed dataset

CRL A

	τ of Δp	Δp	v_f	$R_{\max}$	θ
original dataset	Δp	1.0	0.20	-0.038	-0.22
imputed dataset	Δp	1.0	0.19	-0.27	-0.20
	τ of $R_{\max}$	Δp	v_f	$R_{\max}$	θ
original dataset	$R_{\max}$	-0.038 [-0.23,0.18]	-0.041	1.0	-0.070
imputed dataset	$R_{\max}$	-0.27 [-0.36,-0.19]	0.00045	1.0	0.042
imputed dataset (1988~2020)[1]	$R_{\max}$	-0.060 [-0.21,0.08]	-	-	-

CRL B

	τ of Δp	Δp	v_f	$R_{\max}$	θ
original dataset	Δp	1.0	0.066	-0.11	-0.25
imputed dataset	Δp	1.0	0.10	-0.33	-0.20
	τ of $R_{\max}$	Δp	v_f	$R_{\max}$	θ
original dataset	$R_{\max}$	-0.11 [-0.23,0.04]	0.14	1.0	0.088
imputed dataset	$R_{\max}$	-0.33 [-0.39,-0.26]	0.12	1.0	0.19
imputed dataset (1988~2020)[1]	$R_{\max}$	-0.17 [-0.27,-0.05]	-	-	-

CRL C

	τ of Δp	Δp	v_f	$R_{\max}$	θ
original dataset	Δp	1.0	0.093	-0.30	0.021
imputed dataset	Δp	1.0	0.14	-0.39	0.073
	τ of $R_{\max}$	**Δp**	**v_f**	**$R_{\max}$**	**θ**
original dataset	$R_{\max}$	-0.30 [-0.43,-0.17]	-0.051	1.0	0.064
imputed dataset	$R_{\max}$	-0.39 [-0.45,-0.30]	-0.081	1.0	0.043
imputed dataset (1988~2020)[1]	$R_{\max}$	-0.31 [-0.41,-0.20]	-	-	-

[1] In the original dataset, the Δp and $R_{\max}$ are only available after 1988, while the imputed dataset extends this availability to 1938. To make a comparison for temporal consistency between the original dataset and the imputed dataset, Kendall's τ of Δp and $R_{\max}$ computed using imputed dataset data within 1988~2020 are also listed.

2.5 Conclusion of Chapter 1

In this study, a set of machine learning-derived data imputation models has been developed to impute the missing p_c and $R_{\max}$ in the North Atlantic tropical cyclone historical database, i.e., a combination of HURDAT2 and EBTRK databases. GPR and ANN models are investigated as candidate imputation models and compared against existing statistical relationships. The observed performance of the GPR and ANN models demonstrates the capability and potential for using machine learning-derived methods to address data limitations.

For p_{c} imputation, results indicated that V_w is the most important parameter. Both GPR and ANN can achieve good model performance with R greater than 0.9 when using lat, lon, V_{f}, V_{w}, storm heading (specifically $\sin\theta$ and $\cos\theta$) as model predictors. The performance of these machine learning-based p_{c} imputation models are also compared with the p_{c} model adapted from NACCS15. As expected, both GPR and ANN p_{c} imputation models have better model accuracy than the NACCS15-derived model.

For $R_{\max}$ imputation, the model selection is more complicated. $R_{\max}$ is a highly idealized quantity intended to capture the radial extent of wind. Defining this quantity for historical storms requires modeling and judgment, leading to uncertainty in reported data, and researchers are still making efforts to understand and more accurately define $R_{\max}$. This context appears reflected by the relatively weaker performance of the $R_{\max}$ imputation models compared to p_{c} imputation models. For the $R_{\max}$ imputation parameterization, we leverage insights from the recent $R_{\max}$ prediction model CK22 (Chavas and Knaff 2022) and use R_{34} as one parameter in the $R_{\max}$ imputation baseline parameterization (i.e., using lat, lon, V_{f}, p_{c}, V_{w}, R_{34}, $\sin\theta$ and $\cos\theta$ as model predictors). This results in the best overall performance when combined with a GPR model. However, given the fact that R_{34} data has a limited availability, we consider an alternate $R_{\max}$ imputation parameterization (i.e., using lat, lon, V_{f}, p_{c}, V_{w}, $\sin\theta$ and $\cos\theta$ as model predictors) for data points without R_{34}. It is found that the GPR models have the highest accuracy (i.e., the minimum ARMSE) among all tested models for imputing $R_{\max}$. Additionally, compared with using all storm data, the $R_{\max}$ imputation model accuracy is improved when developed using storm data with high-intensity and only applied for high-intensity storms. This can be explained by the fact that tropical cyclones with different intensities are connected to different environmental conditions (Emanuel et al. 2004). We also assessed the $R_{\max}$ imputation models' capability in predicting large $R_{\max}$ values (e.g., values greater than 250 km) that appear in the EBTRK database. While physical arguments may lead to the eventual treatment of large $R_{\max}$ values, which treat them as outliers in a statistical aspect in some contexts and applications, we opted to keep the feature of original $R_{\max}$ data. It is found that for model parameterizations where R_{34} is not used, the GPR is challenged in predicting large $R_{\max}$ values though the performance was strong in other aspects. The ANN baseline without R_{34} parameterization model has a better capability for predicting large $R_{\max}$ values. Additionally, we compared the effect of different $R_{\max}$ imputation models on the correlation between Δp and $R_{\max}$, which is a critical component of joint distribution analysis. The rank correlation coefficient Kendall's τ was measured considering multiple candidate $R_{\max}$ models using a ten-year moving time window. It turns out that the GPR baseline $R_{\max}$ imputation model can best capture the value and trend of the observed Kendall's τ quantifying the correlation between Δp and $R_{\max}$.

Finally, we assessed the effect of data imputation on statistical analysis using USACE's CHS-PCHA procedure. The data imputation process significantly increases the number of statistical sampling data of Δp and $R_{\max}$, and therefore improves the robustness of fitted marginal distribution. However, the imputation process can also reduce the variability in the resulting distribution and may amplify the negative correlation between Δp and $R_{\max}$, which might induce a potential underestimation of the frequency of severe storm surges. One alternative option is to use correlation estimated from recently observed data. The tradeoffs between the benefits of increasing sample size and the statistical implications of using imputed data (with less variability) should be considered. Possible imputation model alternatives that can preserve expected variability and retain the observed correlation between storm parameters should be explored in the future.

Chapter 3: Comparative Analysis of Joint Distribution Models for Tropical Cyclone Atmospheric Parameters in Probabilistic Coastal Hazard Analysis[3]

3.1 Introduction

A tropical cyclone (TC) is characterized by its atmospheric-forcing parameters, which can be used to construct synthetic storms comprised of wind and pressure fields, allowing for the prediction of coastal hazards such as storm surges and storm rainfall. These coastal hazards play significant roles in the mechanisms of coastal compound floods. Over the past few decades, the Joint Probability Method (JPM) has become the foundational tool for probabilistic characterizations of hurricane hazards, particularly those associated with storm surges (Myers 1954; Ho and Myers 1975; Russell 1969; Myers 1970; IPET 2009).

A key input to the JPM is the joint probability distribution of TC atmospheric-forcing parameters, which are typically represented by central pressure deficit (Δp), forward velocity, radius of maximum wind ($R_{\max}$), and landfall heading direction (θ). A range of joint distribution models of TC atmospheric parameters have been constructed and used for probabilistic coastal hazard studies. Specifically, previous studies have made assumptions regarding the relationship between parameters that range from independence to partial-dependence (i.e., assuming a subset

[3] This chapter was submitted to the journal Stochastic Environmental Research and Risk Assessment on June 20, 2023 (Submission ID: 6175b70d-c951-4533-9740-ab95d525e8b5) and as of October 31, 2023, it is "Under Review."

of parameters are dependent) to full dependence (i.e., all parameters are dependent) (e.g., FEMA 2009; Nadal-Caraballo et al. 2015, 2022b).

For the past decade, the United States Army Corps of Engineers (USACE) has been developing the Coastal Hazard System (CHS) (https://chs.erdc.dren.mil) and its Probabilistic Coastal Hazard Analysis (PCHA) framework in collaboration with other U.S. Federal agencies and academia (Nadal-Caraballo et al. 2020). The main goal of the CHS is the systematic quantification of coastal hazards caused by TCs and other storm events. The techniques used in the CHS for TCs build off and extend the previously established JPM, resulting in the development of the Joint Probability Method Aided by Metamodel Prediction (JPM-AMP). The CHS' JPM-AMP encompasses climatological data imputation via metamodeling to resolve data discontinuities within historical TC records, metamodel prediction of hydrodynamic storm responses, implementation of multivariate meta-Gaussian copula (MGC) to estimate TC probabilities, and estimation of geospatial bias and associated uncertainties, among other advancements. The primary results of the CHS are expressed as hazard curves, which provide measures of storm response (e.g., peak storm surge elevation) on one axis and the corresponding annual exceedance frequency (AEF) on the other axis (Nadal-Caraballo et al. 2015; 2022a; b).

In the CHS' JPM-AMP, the MGC approach has been implemented to characterize the joint distribution of TC atmospheric parameters (Nadal-Caraballo et al. 2020, 2022a; b). Copulas have also been used in other coastal and ocean applications, such as modeling the dependence between coastal response quantities (e.g., storm surge and precipitation) (Sebastian et al. 2017), wave heights at multiple locations (Jane et al. 2016), and wave height and wave period and wave direction (Lin and Dong 2019). While existing research has explored a range of candidate copulas (e.g., elliptical copulas and Archimedean copulas), the MGC used in the CHS' JPM-

AMP has the advantage of being straightforward in capturing the full dependence between variables in the joint distribution. However, the linear correlation parameter used in the MGC may become unstable because the storm heading direction θ is a circular variable (e.g., angles of 0 and 360 degrees have the same physical meaning but are not equivalent in linear space). Alternate copulas can explicitly account for both linear and circular variables (e.g., García-Portugués et al. 2013; Johnson and Wehrly 1978) and have been used for applications in winds (García-Portugués et al. 2013; Wang et al. 2021) and waves (Lin and Dong 2019). However, circular-linear copulas are less straightforward to implement. To date, no research has explored the relative trade-offs of using the MGC and circular variable compatible copulas in the context of the joint distribution of TC atmospheric parameters.

Considering the trade-offs between model stability and implementation practicality, this study aims to investigate and comparatively assess the performance of several candidate JPM assumptions for TC atmospheric parameters. JPM assumptions evaluated herein include parameter independence, partial dependence (i.e., only considering dependence between Δp and $R_{\max}$), and full dependence copula models. The details of each tested assumption are described in Section 2.3.

The performance of candidate models is assessed from several perspectives. We compare the surge hazard curves generated under multiple joint distribution models and quantify their differences at a series of case study locations around the region of New Orleans, LA. We take the MGC assumption as the base case. We also explore the sensitivity of estimated tail dependence to JPM assumptions. Given the practical advantages of the MGC, we investigate the sensitivity of the hazard curve developed using the MGC assumption to the shift of heading direction θ's angular coordinate (i.e., the convention used to set the zero-degree direction) to understand the stability of the MGC assumption in the case study region.

3.2 Method

Figure 17 illustrates the procedure used to compare JPM assumptions as well as references to the sections of this paper where additional information is provided. This figure will be referenced in the subsections that follow.

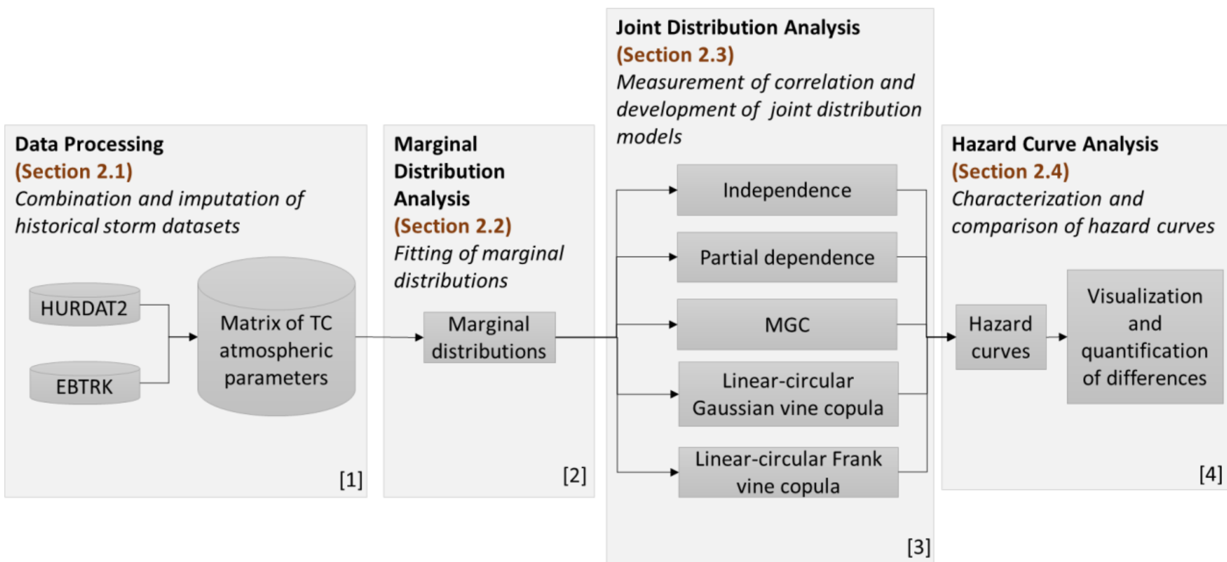


Figure 17: Procedure used to compare JPM assumptions

3.2.1 Data Processing

In the Data Processing step (Box [1] in Figure 17), historical data from multiple databases are combined and imputed to develop a matrix of TC atmospheric parameters for use in subsequent statistical analysis. Historical storm-related climatology records were collected from the National Oceanographic and Atmospheric Administration's National Hurricane Center (NHC) HURDAT2 (HURricane DATa 2nd generation) (Landsea and Franklin 2013; NOAA 2022c) for the years

1938 to 2020. Specifically, values of central pressure, wind speed, location (latitude/longitude), and date/time at discrete points along historical storm tracks are available in the database. Central pressure deficit is computed as the difference between recorded central pressure and an assumed peripheral pressure (taken as 1013 hPa in this study, consistent with recent studies (e.g., Nadal-Caraballo et al. 2022b). Spatial and temporal information is then used to compute storm heading (θ) and forward velocity (V_{f}). Heading and forward velocity values assigned to a track point i are computed based on the angle and distance traveled from point $[i-1]$ to point $[i]$. In addition, $R_{\max}$ data was collected from the extended best track (EBTRK) database (CIRA 2021; Demuth et al. 2006) because $R_{\max}$ values are not included in HURDAT2. The HURDAT2 and EBTRK data were temporally interpolated onto the same temporal grid. A data imputation process based on Gaussian process regression (GPR; also known as Kriging) models is applied to the combination of HURDAT2 and EBTRK datasets to "fill in" missing central pressure and $R_{\max}$ values (Liu et al. 2023b; Nadal-Caraballo et al. 2022a; b).

3.2.2 Marginal Distribution Analysis

Developing joint distributions using any of the candidate models used in this study first requires the development of marginal distributions. In the Marginal Distribution Analysis step (Box [2] in Figure 17), marginal distributions for (Δp, V_{f}, $R_{\max}$ and θ) are estimated based on the sampling method developed and implemented by Nadal-Caraballo et al. (2015). Specifically, the Gaussian kernel function (GKF) distance-based weight method developed by Chouinard and Liu (1997) is used to compute the distance-weight adjusted values of Δp, $R_{\max}$, and V_{f}. Weighted Δp data is then used to estimate the parameters of a truncated Weibull distribution with maximum likelihood estimation (MLE). Similarly, the probability models of V_{f} and $R_{\max}$ are both represented by lognormal distributions with MLE.

In this study, a modification to the CHS' JPM-AMP (Nadal-Caraballo et al. 2020, 2022a; b) is investigated for building the marginal distribution of θ. In JPM-AMP, the probability model of θ is estimated by using scaled values of the directional storm recurrence rate (DSRR) to represent its evaluated marginal PDF. Specifically, the DSRR (λ_θ) is derived from the directional intensity function in Chouinard and Liu (1997), which is expressed as:

$$\lambda_\theta = \frac{1}{T}\sum_{i}^{n} w(d_i)\, w(\theta_i - \theta) \tag{3}$$

In the above expression, T is the record length (years), and the sum is taken over $i = 1, .., n$ statistical sampling data points. θ_i is the heading direction of the storm track point associated with the statistical sampling data location i (computed as the direction from point $i-1$ to point i). $w(d_i)$ is the distance weight of ith statistical sampling data point calculated using a GKF. $w(\theta_i - \theta)$ is also conventionally calculated using a GKF:

$$w(\theta_i - \theta) = \frac{1}{\sqrt{2\pi}h_\theta} exp\left[-\frac{1}{2}\left(\frac{\theta_i - \theta}{h_\theta}\right)^2\right] \tag{4}$$

Where the kernel size h_θ here is set to be $30°$ as recommended by FEMA (2012). The value of marginal PDF is then evaluated at θ as proportional to λ_θ in Equation (3). A normalizing factor is set to make the integral of the PDF over the domain [-180,180] equal to 1.

In this study, we use a von Mises kernel function (VKF) (Taylor 2008) to replace the GKF in Equation (4). The VKF is given by:

$$w(\theta_i - \theta) = \frac{1}{2\pi I_0(\kappa)} \exp\left[\kappa \cos(\theta_i - \theta)\right] \tag{5}$$

The parameter κ can be regarded as analogous to the bandwidth parameter of the VKF and $I_0(\kappa)$ is the zeroth-order modified Bessel function of the first kind and is expressed as:

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \exp[\kappa \cos(\theta)]\, d\theta \tag{6}$$

The VKF is derived from the von Mises distribution, which is also called the wrapped normal distribution. The VKF is proposed herein because it can be used to build a stable marginal distribution for a circular variable (Zar 1999).

Three sets of marginal distribution are fitted for Δp, V_{f}, $R_{\max}$ and θ for low intensity (LI), moderate intensity (MI), and high intensity (HI) TCs. The three intensity categories are classified based on Δp as LI (28 hPa $> \Delta p \geq 8$ hPa), MI (48 hPa $> \Delta p \geq 28$ hPa), and HI ($\Delta p \geq 48$ hPa).

The probability model of the landfall location parameter x_0 is treated as independent from Δp, V_{f}, $R_{\max}$ and θ.

3.2.3 Joint Distribution Analysis

In the Joint Distribution Analysis step (Box [3] in Figure 17), a series of assumptions are considered to construct the joint distribution of TC atmospheric parameters Δp, V_{f}, $R_{\max}$ and θ. The assumptions reflect the range of approaches that have been used throughout the evolution of the JPM, leading to JPM-AMP. Candidate assumptions include:

- **Independence assumption**: In this assumption, the joint PDF is taken as the product of the marginal PDFs.

- **Partial-dependence assumption**: In this assumption, there is dependence between Δp and $R_{\max}$, using the statistical model for $R_{\max}|\Delta p$ developed by Vickery and Wadhera (2008).
- **MGC assumption**: In this assumption, there is full dependence among TC atmospheric parameters, and an MGC is used to construct the joint distribution model.
- **Linear-circular Gaussian vine (LCGV) assumption**: In this assumption, there is full dependence among TC atmospheric parameters. A 2-tier canonical vine copula based on linear-circular copulas and MGCs is used to construct the joint distribution model.
- **Linear-circular Frank vine (LCFV) assumption**: In this assumption, there is full dependence among TC atmospheric parameters. A 3-tier canonical vine copula based on linear-circular copulas and Frank copulas is used to construct the joint distribution model.

Each of these assumptions is further described in the sections that follow. In the mathematical expressions that follow, notation is adopted such that capital letters (i.e., $\Delta P, R_{\max}, V_{\mathrm{f}}, \Theta$) represent random variables and lowercase letters (i.e., $\delta p, r_{\max}, v_{\mathrm{f}}, \theta$) represent realizations (values) of the random variable.

3.2.3.1. Independence assumption

Under the independence assumption, the joint PDF of the TC atmospheric parameters is calculated as the product of TC atmospheric parameters' marginal PDFs as:

$$f(\delta p, v_{\mathrm{f}}, r_{\max}, \theta) = f_{\Delta P}(\delta p) * f_{V_{\mathrm{f}}}(v_f) * f_{R_{\max}}(r_{\max}) * f_{\Theta}(\theta) \tag{7}$$

Where $f(\square)$ is the PDF of corresponding variables.

Partial-dependence assumption

For the partial-dependence assumption, the joint PDF of TC atmospheric parameters considering the dependence between Δp and $R_{\max}$ is expressed as:

$$f(\delta p, v_{\mathrm{f}}, r_{\max}, \theta) = f_{R_{\max}|\Delta P}(r_{\max}\,|\delta p) * f_{\Delta P}(\delta p) * f_{V_{\mathrm{f}}}(v_{\mathrm{f}}) * f_{\Theta}(\theta) \tag{8}$$

To define the conditional PDF $f_{R_{\max}|\Delta P}(r_{\max}\,|\delta p)$, we leverage the work of Vickery and Wadhera (2008), which developed a statistical model for $R_{\max}$ using data for almost all hurricanes making landfall along the U.S. coastline between 1900 and 1985. General and region-specific models are presented in their work. In this study, we leverage their regression model for the Gulf of Mexico. Equation 13 of Vickery and Wadhera (2008) is used to calculate $r_{\max}$. In this study, we assume $f_{R_{\max}|\Delta P}(r_{\max}\,|\delta p)$ is a normal distribution PDF with standard deviation and mean value derived from Vickery and Wadhera (2008).

3.2.3.2. MGC assumption

A copula can be regarded as a function that couples multiple marginal distribution functions to create a multivariate distribution function. According to Sklar's theorem (Sklar 1959), the joint cumulative distribution function (CDF) of multiple marginal distributions can be expressed as:

$$H(x_1,\ldots,x_{\mathrm{n}}) = C(F_1(x_1),\ldots,F_{\mathrm{n}}(x_{\mathrm{n}})) \tag{9}$$

Where: $x_i, i = 1, \ldots, n$ are realizations of random variables X_i. In this case, the x_1, x_2, x_3 and x_4 represent the $n = 4$ TC atmospheric parameters δp, v_{f}, $r_{\max}$ and θ, respectively. $F_i(x_i)$ is the CDF of X_i.

For $n = 4$, the joint PDF $f(x_1, x_2, x_3, x_4)$ can be calculated as:

$$f(x_1,x_2,x_3,x_4) = c\big(F_1(x_1), F_2(x_2), F_3(x_3), F_4(x_4)\big)\, f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \tag{10}$$

where $c(\square)$ is the copula PDF and $f_i(x_i), i = 1, \ldots, 4$ is the marginal PDF of X_i.

The Gaussian copula has been adopted in the JPM-AMP as part of USACE's CHS-PCHA framework (Nadal-Caraballo et al. 2020). In general, a meta-Gaussian joint distribution (resulting from the use of a Gaussian copula) relaxes the multivariate Gaussian distribution requirement to have Gaussian marginals by allowing arbitrary marginals. Herein, we refer to a multivariate Gaussian copula derived from a meta-Gaussian joint distribution as MGC (Zhang and Singh 2019). In the MGC assumption, four-dimensional MGCs are used to characterize the full dependence among TC atmospheric parameters.

The MGC CDF is expressed as:

$$C\big(F_1(x_1), \dots, F_4(x_4)\big) = \Phi_{\mathrm{R}}\big(\Phi^{-1}(F_1(x_1)), \Phi^{-1}(F_2(x_2)), \Phi^{-1}(F_3(x_3)), \Phi^{-1}(F_4(x_4))\big) \tag{11}$$

Where: $\Phi^{-1}(\square)$ is the inverse standard normal (marginal) CDF and $\Phi_{\mathrm{R}}^{-1}(\square)$ is the joint normal multivariate CDF with correlation matrix R. The MGC uses a matrix of Pearson's ρ as a parameter. In this study, Pearson's ρ values were computed from estimated Kendall's τ (i.e., τ values estimated from the distance-adjusted imputed data series) using the relationship $\rho = \sin\frac{\tau\pi}{2}$ (Fang et al. 2002).

The MGC assumption has the advantage of being straightforward in constructing a high-dimensional joint distribution model. However, its correlation parameter R matrix is calculated using Pearson's ρ, which is a linear correlation coefficient. Because Θ is a circular variable, Pearson's ρ can differ with a change of the zero-degree direction/convention defined for θ. In addition, the Gaussian copula is known to have potential limitations for modeling dependence in the tails of distributions (Donnelly and Embrechts 2010).

3.2.3.3. Linear-circular and Gaussian vine copula assumption

Because Θ is a circular variable, there is no "true zero point," and its designation of high and low values is arbitrary (Zar 1999). Pearson's ρ, which is used to capture dependence in the MGC assumption, is unstable when estimating correlation related to Θ; that is, the correlation between Θ and other TC atmospheric parameters will change when the zero-degree convention for Θ changes.To address this issue, a linear-circular copula (García-Portugués et al. 2013; Johnson and Wehrly 1978; Lin and Dong 2019; Wang et al. 2021) is applied herein.

Two candidate linear-circular copulas are considered: (1) the linear-circular copula introduced by Johnson and Wehrly (1978) (JW linear-circular copula) and (2) the quadratic section (QS) linear-circular copula (García-Portugués et al. 2013; Nelsen 2007; Wang et al. 2021). The likelihood values of these two copulas are calculated (see Appendix B.1), and it is found that the JW linear-circular copula exhibits a higher likelihood value than the QS linear-circular copula and is thus selected for hazard curve construction in this study.

The challenge of implementing this linear-circular copula is that, unlike the MGC, this linear-circular copula can only be used to depict the dependence between two variables (i.e., to define a pair copula). Vine theory (Bedford and Cooke 2002; Nelsen 2007) provides a method that can be used to combine multiple pair copulas into a multiple-variable joint probability model for more than two variables. In vine theory, it is assumed that the conditional dependence between variables can be characterized using a copula with fixed parameters regardless of the variance of a given conditional variable. In the rest of this section and Section 2.3.5, two vine copula models (i.e., an LCGV and an LCFV) are constructed.

Leveraging the advantages of the MGC assumption, an MGC can be combined with linear-circular copulas to construct a joint distribution model that contains circular variables. In this study, an LCGV, which is a 2-tier canonical vine (Jaworski et al. 2010), was built to develop the joint distribution for TC atmospheric parameters. A vine copula can usually be represented with a tree graph. The tree graph of this 2-tier canonical vine is presented in Figure 18, where the gray ovals represent the random variables (Tier 1) or conditional copulas (Tier 2) used to estimate the parameters of the copulas in the tier, while the red boxes represent the fitted copulas. In Tier 1, three joint linear-circular copulas (C_{14}, C_{24}, C_{34}) are used to depict the dependence between Θ (Θ is represented by X_4 in Tier 1 of Figure 18) and the other three TC atmospheric parameters ΔP, V_{f} and $R_{\max}$ (represented by X_1, X_2 and X_3, respectively). Next, the joint linear-circular copulas (C_{12}, C_{24}, C_{34}) are used to calculate the conditional copula values ($C_{1|4}, C_{2|4}, C_{3|4}$) at data points (see Equation A-6) as shown in the middle box of Figure 18. In Tier 2, a three-dimensional MGC (the red box labeled $C_{123|4}$) is used to depict the dependence among the three linear-circular copulas (gray ovals). The expressions for deriving the joint PDF using this two-tier vine copula are provided in Appendix B.2.

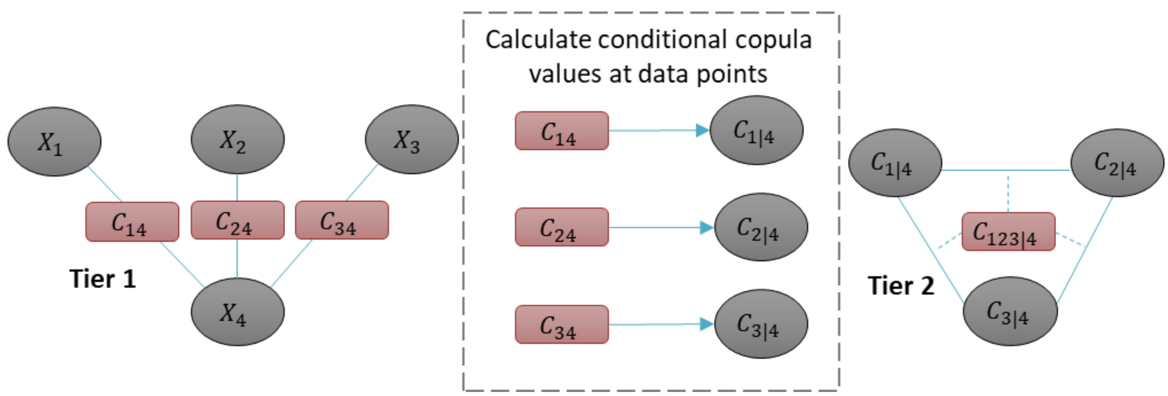


Figure 18: Tree graph of the 2-tier canonical vine. In Tier 1, the gray ovals represent the random variables used to estimate the parameters of copulas, while the red boxes represent the fitted copulas. In Tier 2, the gray ovals represent the conditional copulas (estimated from Tier 1 copulas), while the red box represents the fitted copula

Based on the derivation in Appendix B.2, we can calculate the joint PDF as:

$$f_{1234}(x_1,x_2,x_3,x_4) = c_{123|4}\left(F_{1|4}(x_1|x_4), F_{2|4}(x_2|x_4), F_{3|4}(x_3|x_4)\right) * c_{14}\left(F_1(x_1),F_4(x_4)\right) c_{24}\left(F_2(x_2),F_4(x_4)\right) c_{34}\left(F_3(x_3),F_4(x_4)\right) * f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_4) \quad (12)$$

3.2.3.4. Linear-circular and Frank vine copula assumption

The LCGV introduced above allows for the inclusion of the circular variable but inherits the MGC assumption, which is based on Gaussian copulas. The Gaussian copula has a potential drawback in terms of the ability to model tail dependence (Donnelly and Embrechts 2010), which might lead to the underestimation of the probability of an extreme event.

The Frank copula is selected as an alternative choice to replace the MGC in the vine copula model. The Frank copula is selected here because it belongs to the Archimedean copula family, which is the most commonly implemented copula family in coastal and ocean engineering (Joe, 1997; Lin and Dong, 2019; Nelsen, 2007). Moreover, the Frank copula has flexibility in modeling both positive dependence and negative dependence (i.e., the Frank copula accepts a wide range of Kendall's τ values, which is not possible with all copulas in the Archimedean copula family).

The CDF of the Frank copula is presented here for convenience:

$$C\left(F_i(x_i),F_j(x_j)\right) = -\frac{1}{\alpha}\ln\left[1+\frac{(e^{-\alpha F_i(x_i)}-1)(e^{-\alpha F_j(x_j)}-1)}{(e^{-\alpha}-1)}\right], \alpha \neq 0 \quad (13)$$

Where: $F_i(x_i)$ and $F_j(x_j)$ are the marginal CDFs of random variables x_i and x_j, respectively, and α is the dependence parameter. α has a relationship with Kendall's τ as described by:

$$\tau = 1+\frac{4}{\alpha}\left[\frac{1}{\alpha}\int_0^{\alpha}\frac{t}{e^t-1}dt-1\right] \quad (14)$$

An LCFV, which is a 3-tier canonical vine based on a Frank copula, is used to build the joint distribution model of TC atmospheric parameters. The graphical model of this LCFV is presented in Figure 19. In Tier 1, a linear-circular copula is used to depict the dependence between θ (θ is represented by X_4 in Tier 1) and the other three TC atmospheric parameters Δp, V_f and $R_{\max}$ (corresponding to X_1,X_2 and X_3) similar to the Tier 1 in LCGV (see Figure 18). In Tier 2, two Frank copulas are used to capture the dependence among the three linear-circular copulas fitted in Tier 1. In Tier 3, a Frank copula is used to capture the dependence between the two Frank copulas fitted in Tier 2.

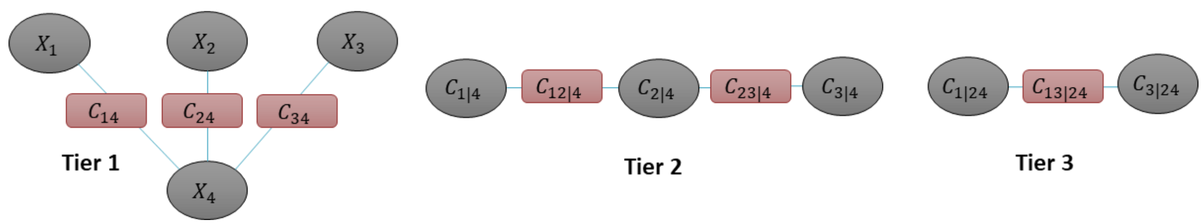


Figure 19: Tree graph of the 3-tier canonical vine. The gray ovals represent the random variables used to estimate the parameters of copulas, while the red boxes represent the fitted copulas

The joint PDF $f_{1234}(x_1,x_2,x_3,x_4)$ can be expressed as:

$$f_{1234}(x_1,x_2,x_3,x_4) = f_{1|234}(x_1|x_2,x_3,x_4)f_{3|24}(x_3|x_2,x_4)f_{2|4}(x_2|x_4)f_4(x_4) \tag{15}$$

where $f_{1|234}(x_1|x_2,x_3,x_4)$ is the conditional PDF of X_1 given X_2, X_3 and X_4; $f_{3|24}(x_3|x_2,x_4)$ is the conditional PDF of X_3 given X_2 and X_4.

Using Skalar's theorem, the joint PDF under LCFV assumption can be calculated as:

$$\begin{aligned} f_{1234}(x_1,x_2,x_3,x_4) = {} & c_{13|24}(F_{1|24},F_{3|24})\ c_{12|4}(F_{1|4},F_{2|4})c_{23|4}(F_{2|4},F_{3|4}) \\ & * c_{14}(F_1,F_4)\ c_{24}(F_2,F_4)c_{34}(F_3,F_4) \\ & * f_1(x_1)f_2(x_2)f_3(x_3)f_4(x_4) \end{aligned} \tag{16}$$

3.2.3.5. Hazard curve construction

In the Hazard Curve Analysis step (Box [4] in Figure 17), hazard curves are constructed using the five JPM assumptions outlined above, and comparisons are made among the resulting hazard curves. As a supplemental analysis, the hazard curve construction using the MGC assumption is repeated with different conventions for setting the "zero-degree direction." The sensitivity of constructed hazard curves to different zero-degree conventions is then assessed.

In this study, the quantification of hazard curves follows the general approach outlined in Nadal-Caraballo et al. (2022b). The joint PDF values of discretized TC atmospheric parameters combination are calculated using Equations (7), (8), (10), (12), and (16) for different joint distribution models considered in this study. Joint PDF values corresponding to each discretized parameter combination are then normalized and assigned as the corresponding probability mass. The JPM integral (in discretized form) is processed to compute AEF, and results are presented as estimated hazard curves. A confidence interval is later added to account for random error (uncertainty) by assuming it follows a Gaussian distribution, with its mean equal to zero and standard deviation accounting for both absolute and relative uncertainties (Nadal-Caraballo et al. 2022b).

3.3 Case Study

The method described in Section 2 was applied to a case study in the New Orleans region. Historical storm datasets HURDAT2 and EBTRK are processed following the procedure described in Section 2.1. The marginal distribution analysis is processed following the procedure described in Section 2.2 and centered on a coastal reference location (CRL) near the City of New Orleans (latitude 29.576 and longitude -89.538). It considers data within 600 km of the CRL, as shown by the black circle in Figure 20 (a). Figure 20 (a) also shows the statistical sampling locations (red circles) used for statistical analysis. The results of the marginal distribution analysis are presented in Appendix C. Sixteen reference storm landfall locations as presented in Figure 20 (c) are used in the case study and assumed equally-likelihood.

The storm surge hazard curve is highly location-dependent because it depends on the physical processes governing surge amplification and attenuation. To cover the location variability in the study region, a set of representative study nodes were selected for the hazard curve sensitivity analysis. Figure 20b shows the location of selected study nodes. All of these selected nodes are permanently inundated. In this study, NAVD88 2009.65 is used as the datum of all elevation values. The labels and elevations of study nodes are:

- Node 1: A river node at Plaquemine [30.31, -91.21], elevation: 1.67m.
- Node 2: A lake node at New Canal Station [30.03, -90.11], elevation: -3.49m.
- Node 3: A river node near the Center of New Orleans City [29.93, -90.14], elevation: -19.75m.
- Node 4: A coastal node at Grand Isle [29.25, -89.95], elevation: -3.30m.

This case study uses a GPR metamodel developed in the USACE's CHS-LA study (Nadal-Caraballo et al. 2022b) as a storm surge prediction model. GPR-based storm surge prediction models have been well studied in a series of studies (Jia et al. 2016; Jia and Taflanidis 2013; Taflanidis et al. 2014; Zhang et al. 2018). In the GPR metamodel used in this study, ΔP, V_{f}, $R_{\max}$, Θ and X_0 are treated as input (predictors), and peak surge elevation at 1.2 million nodes around the case study region are treated as output (target). The training data includes 645 synthetic storms and peaks surge response for 1.2 million nodes generated by the ADCIRC model and planetary boundary layer model. Principal component analysis was applied in this GPR metamodel to reduce overall dimensionality. Given the purpose of this study, the probability models used for TC parameters differ from those used in the CHS-LA study (Nadal-Caraballo et al. 2022b), and model refinements and judgements applied in the CHS-LA study are not applied here.

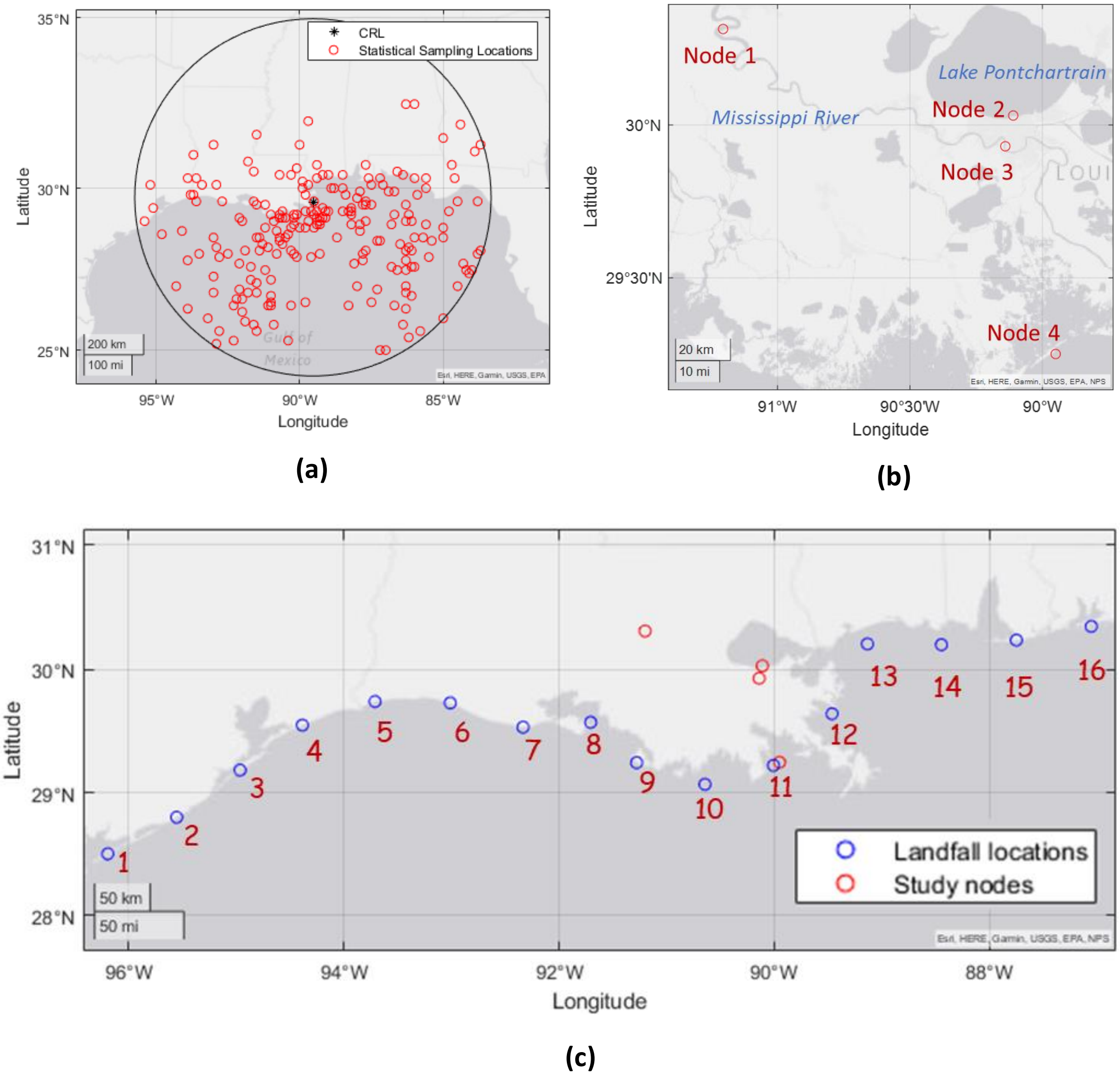


Figure 20: (a) Center of statistical analysis of study region (i.e., the CRL), locations of sample data, and capture zone (the black circle) for the case study; (b) Study node locations; (c) Storm landfall locations (red circles show study nodes)

3.4 *Results*

For the case study region, joint distribution models are built based on the JPM assumptions introduced in Section 2.3. In Section 3.4.1 , we explore the sensitivity of the marginal PDF of Θ and the correlation coefficient between Θ and the other TC atmospheric parameters (i.e., ΔP,

$R_{\max}$, V_{f}) to the zero-degree convention used to define Θ. Then, in Section 3.4.2 , we graphically explore the pair-wise joint PDFs for TC atmospheric parameters and quantify tail dependence between ΔP and $R_{\max}$ under different joint distribution models. The sensitivity of storm surge hazard curves in the case study region to different joint distribution models is described in Section 3.4.3 .

3.4.1 Sensitivity of statistical analysis to the Θ zero-degree convention

In this case study, the evaluated marginal PDF value of Θ is represented by the scaled VKF DSRR as described in Section 2.2. The VKF's bandwidth parameter κ is subjectively selected to make the overall "shape" of the marginal PDF close to the marginal obtained with the GKF DSRR. To visualize the advantage of using the VKF, we test the sensitivity of the marginal distribution of Θ to the zero-degree convention. Figure 21 plots Θ's marginal PDFs fitted using the GKF DSRR and VKF DSRR with differing zero-degree conventions for LI, MI, and HI TCs. As can be seen, the VKF produces results that are invariant to the zero-degree convention and thus cannot be distinguished in the small inset plots in Figure 21. It is worth noting that the similarity between the invariant VKF results and the "North-centered" zero-degree convention for the GKF in Figure 21 is, in part, an artifact of the geographic configuration of the case study region. In the case study region, landfalling storms tend to head in a northerly direction. For example, different results may be found along the Atlantic coast, where landfalling storms tend to head in a more westerly direction.

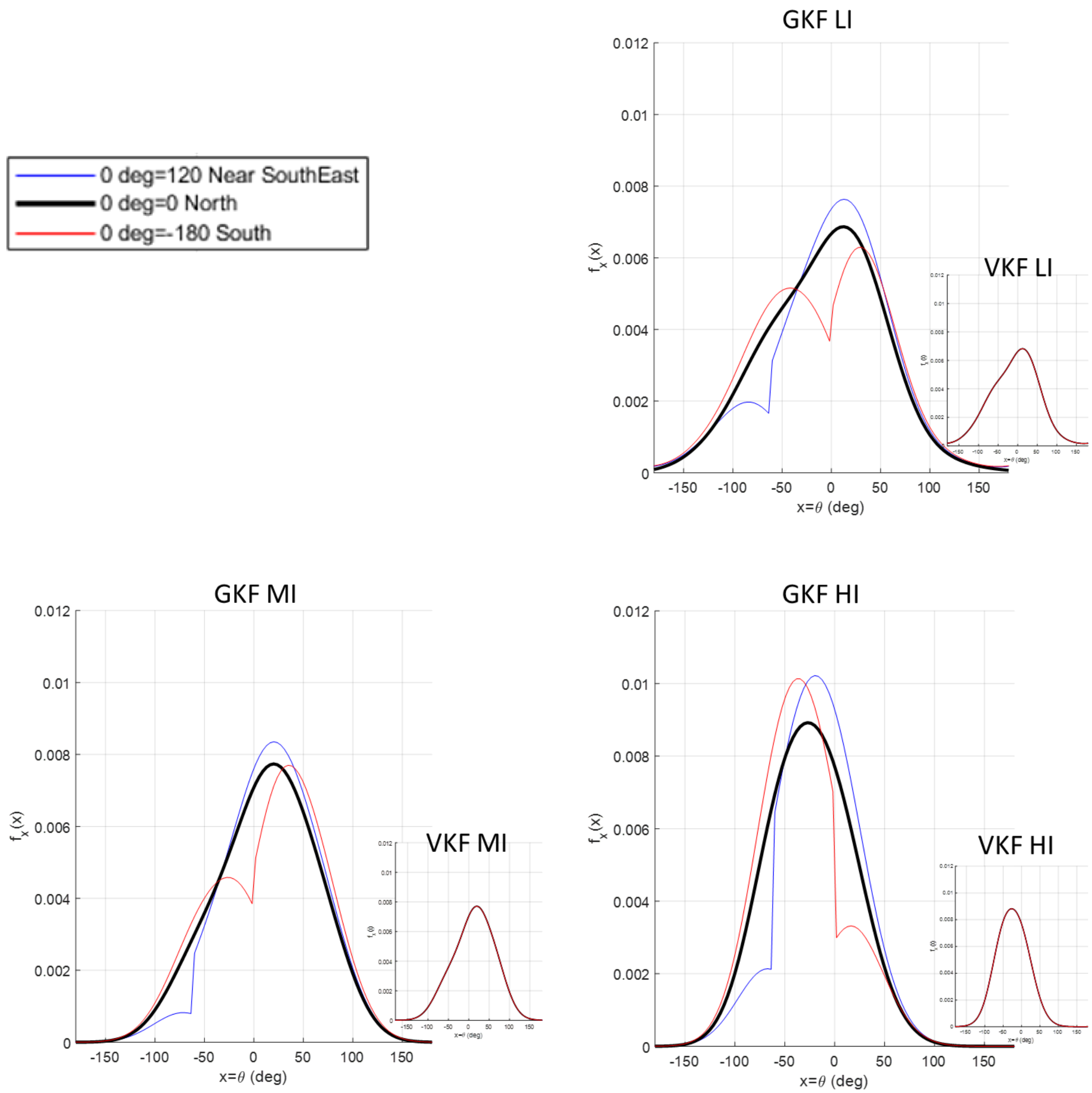
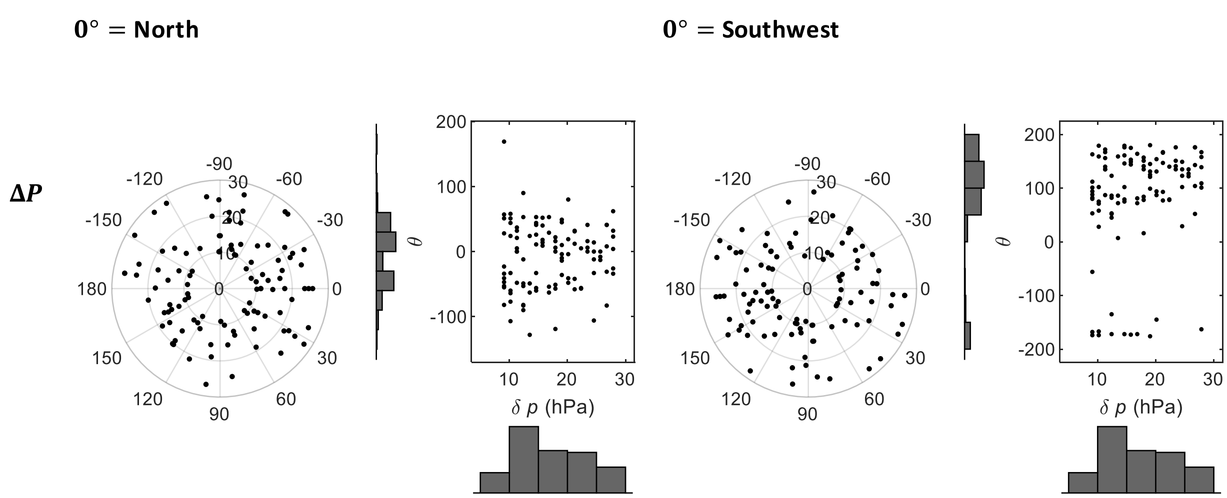


Figure 21: Marginal PDF for Θ, fitted using GKF DSRR (large plots) and VKF DSRR (inset small plots) considering different zero-degree conventions

In addition to the effects from the marginal distributions, the joint distribution built using the MGC assumption could also be affected by other instabilities related to the circular variable Θ. As described in Section 2.3.3, the correlation matrix R of the MGC assumption is derived from the estimated rank correlation Kendall's τ and is used to characterize the dependence among TC atmospheric parameters. However, the heading direction Θ is a circular variable, and its rank correlation will change depending on the selection of the zero-degree direction in the angular coordinate. Conventionally, North is assigned as the zero-degree direction, and Θ is measured from -180 degrees to 180 degrees (where positive angles are associated with the clockwise direction). A “conventional” selection of zero-degree direction might lead to the underestimation of the dependencies between the Θ and other variables.

To visualize how a shifted angular coordinate can affect the estimated dependence between Θ and other parameters, Figure 22 shows scatter plots that depict the difference between the data measured in a conventional angular coordinate (using North as the zero-degree direction) and data measured using a shifted angular coordinate (using Southwest, i.e., -135° relative to due North, as the zero-degree direction). Results are shown for the LI category, which is associated with the largest sample size.



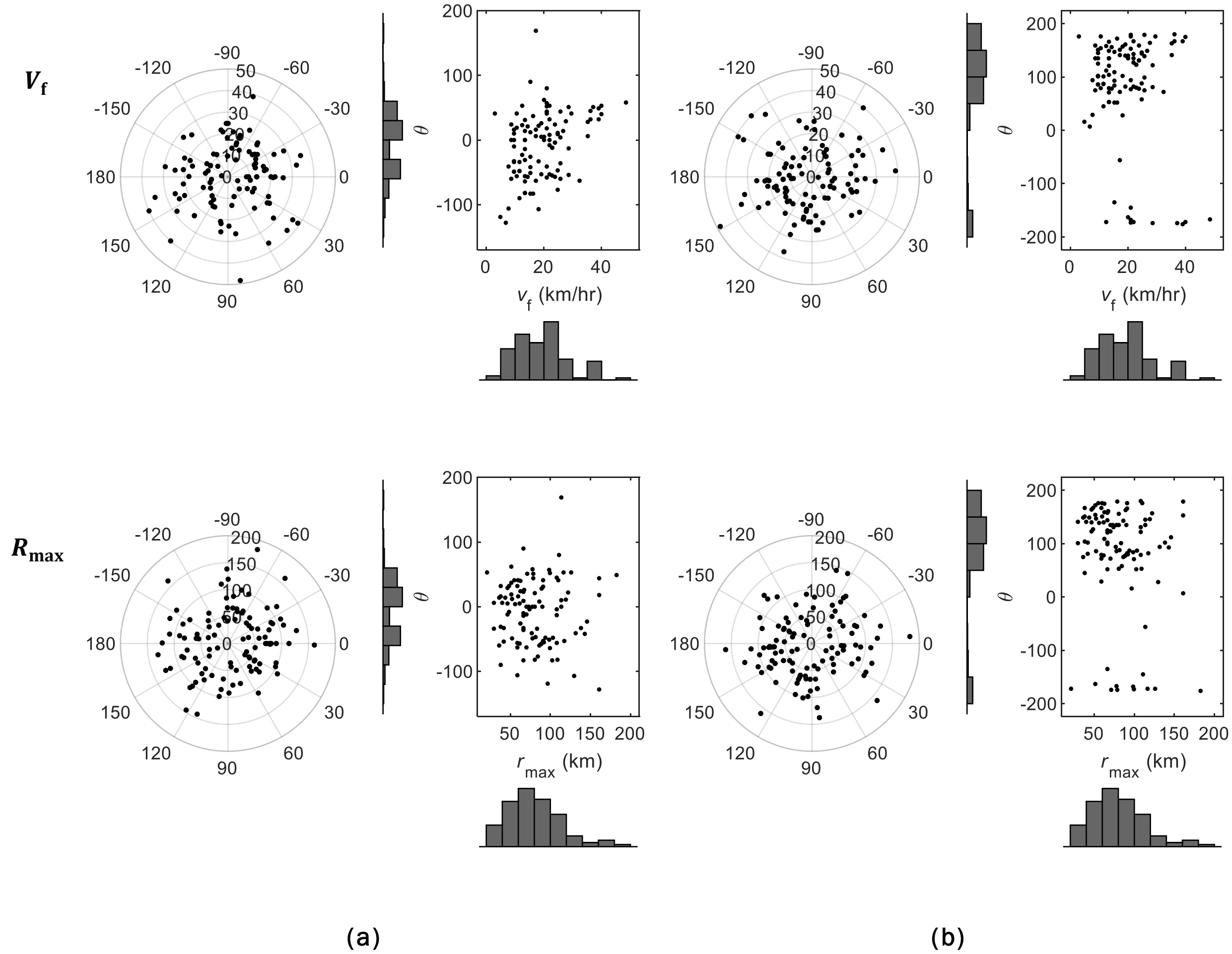


Figure 22: Scatterplots of θ and other TC parameters using two options for zero-degree conventions with LI data in polar and cartesian coordinates. (a) Scatterplot using North as the zero-degree direction; (b) Scatterplot using Southwest as the zero-degree direction. In polar coordinates, the r axis represents the magnitude of linear variable data, and the angle represents θ

It can be observed from Figure 22 that the shifted zero-degree direction did not change the "shape" of data distribution (it only rotates data distribution, as seen in the polar coordinate plots) but significantly affects the visual correlation between Θ and the other three linear variable parameters (as seen in the cartesian plots).

Figure 23 presents the variations of estimated Kendall's τ (a linear correlation coefficient) between Θ and the three linear TC atmospheric parameters (ΔP, V_f and R_{max}) when shifting the zero-degree direction in Θ's angular coordinate. Results are shown for Kendall's τ when the angle is measured in [-180,180] and [0,360]. For comparison, a linear-circular correlation coefficient (Zar 1999) was also plotted, which is calculated as:

$$\rho_{\mathrm{cl}} = \sqrt{\frac{r_{cx}^2 + r_{sx}^2 - 2r_{cx}r_{sx}r_{cs}}{1 - r_{cs}^2}} \tag{17}$$

where r_{sx} is the Pearson's ρ between $\sin\theta$ and x; r_{cx} is the Pearson's ρ between $\sin\theta$ and x; r_{cs} is the Pearson's ρ between $\sin\theta$ and $\cos\theta$; x is a linear variable (ΔP, V_{f} and R_{max}).

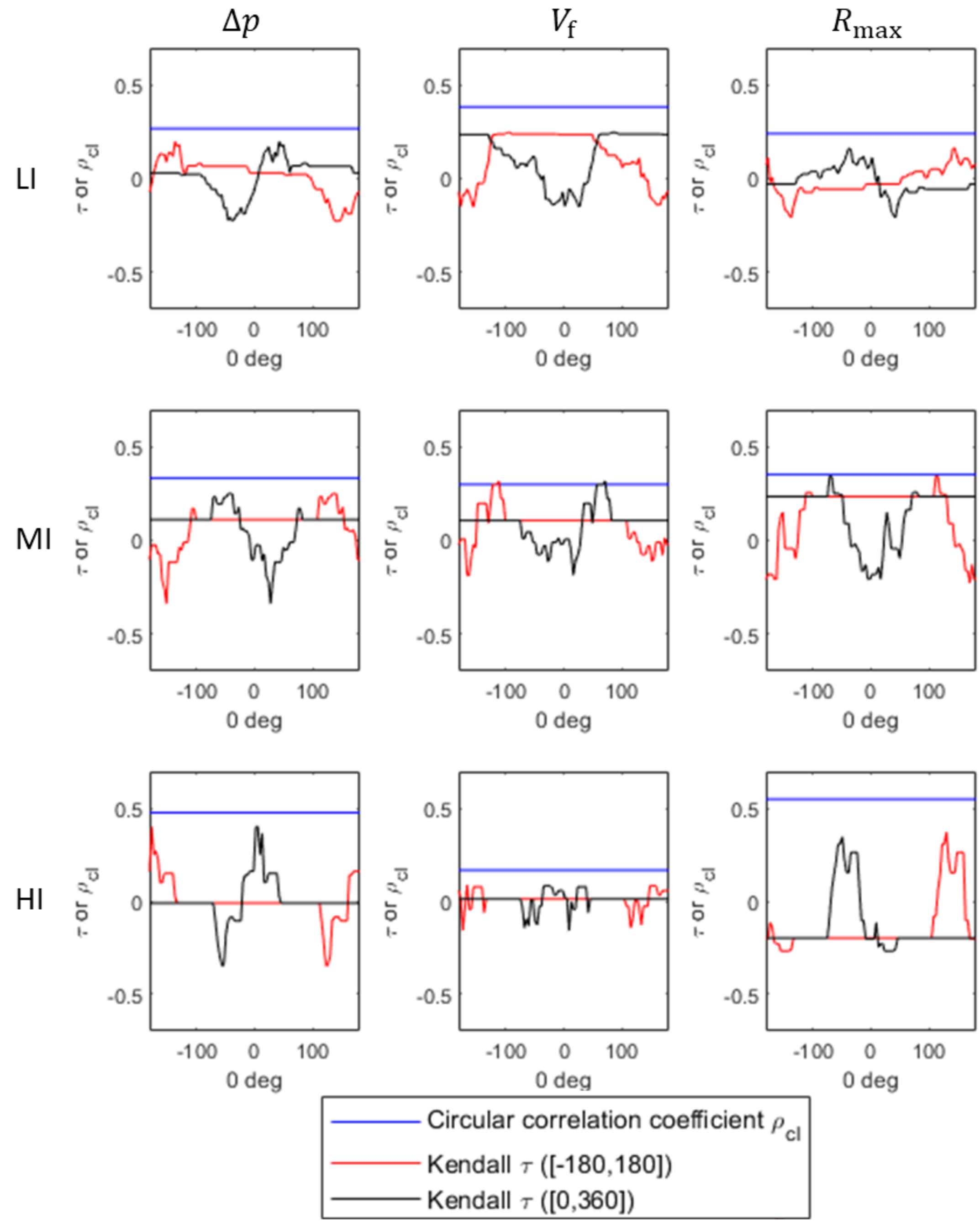


Figure 23: Variations of estimated Kendall's τ between Θ and the three linear TC atmospheric parameters. The zero-degree direction (with $0°$ indicating due North) is shown on the x-axis, and the correlation coefficients are on the y-axis

It can be observed from Figure 23 that Kendall's τ is highly variant with the shift of the zero-degree direction. Moreover, Kendall's τ is often lower than the invariant linear-circular correlation coefficient ρ_{cl}. The change of Θ's scale range (i.e., [-180,180] vs. [0,360]) does not affect the above property of variance. Using [0,360] as Θ's scale range is just equivalent to using -180 as the zero-degree direction in the [-180,180] case. There are some "flat" parts in Kendall's τ curves, this is induced by the scarcity of data.

3.4.2 Dependence analysis

In this section, the characterized full dependence in the MGC, LCGV, and LCFV assumptions are investigated. These full-dependence joint distribution models are visually presented by plotting their pair-wise joint PDF. The pair-wise joint PDF of two variables shown in the graphics is generated by numerically integrating the four-variable joint PDF over the other two variables that are not shown. In Figure 24 through Figure 26, the pair-wise joint PDFs are shown for the HI TC category (here, HI was chosen to present because HI TCs are most relevant to severe surge events). Results are shown for the MGC model, LCGV model, and LCFV model in Figures 8, 9, and 10, respectively. The off-diagonals are the plan view of the pair-wise joint PDFs; the diagonals are the marginal PDFs.

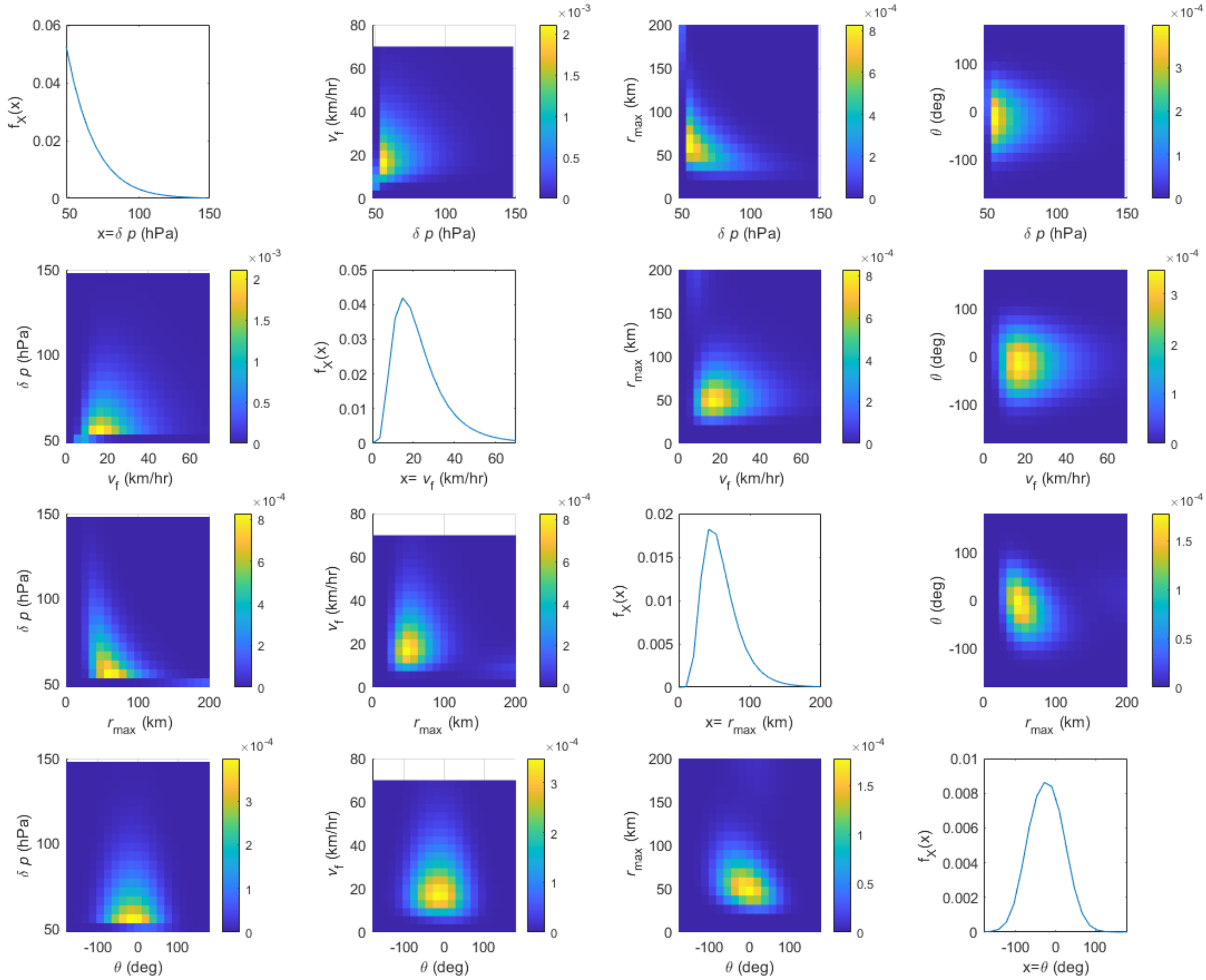


Figure 24: MGC model pair-wise joint PDF for HI TC category

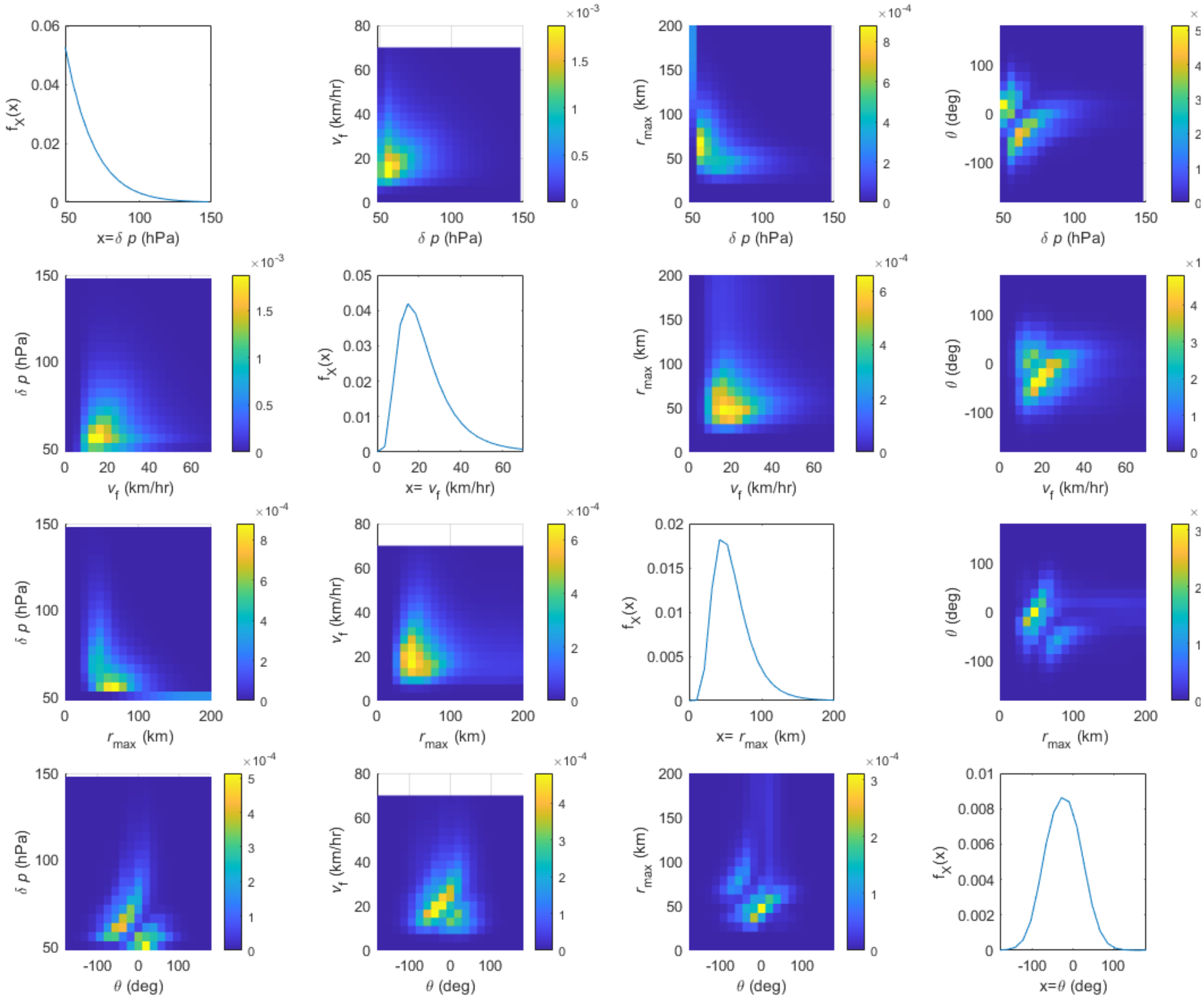


Figure 25: LCGV model pair-wise joint PDF for HI TC category

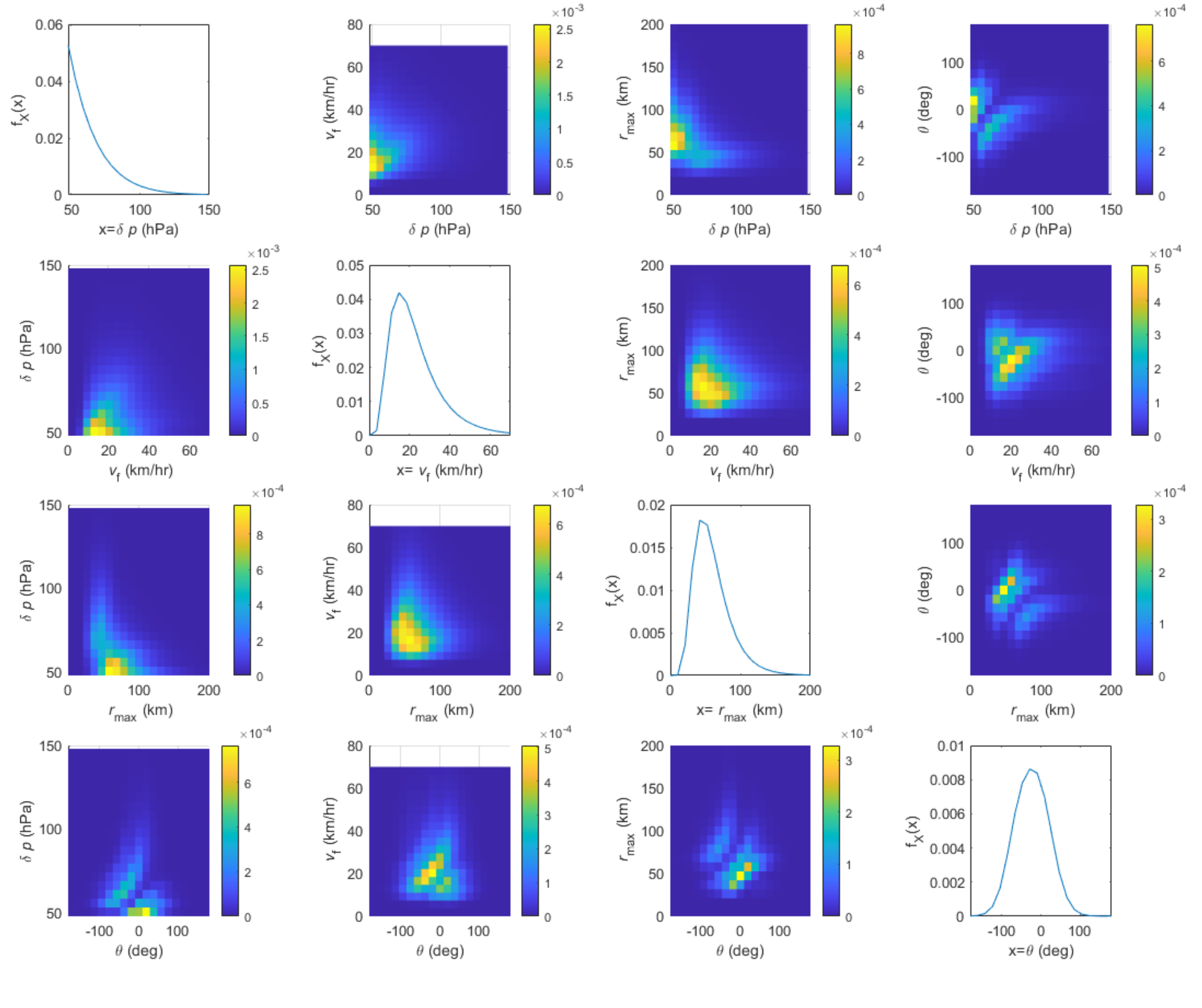


Figure 26: LCFV model pair-wise joint PDF for HI TC category

It can be observed that the MGC, LCGV, and LCFV models generated notably different pair-wise joint PDF surface plots. However, there is almost no difference for the pair-wise joint PDF related to Θ between Figure 25 and Figure 26 (i.e., the bottom row and the rightmost column in both figures) because, in both vine copula models, their dependence between Θ and other parameters are characterized by the same set of linear-circular copulas, the minimum difference can be explained by numerical errors in integration. The estimated copula parameters are presented in Appendix B.3.

While the pair-wise joint PDF surfaces can be used to visualize the differences among the candidate full dependence models, quantitative metrics can provide further insights. Tail dependence is a concept that can be used to describe how large (or small) values of one random variable couple with large (or small) values of another random variable in a joint distribution model (Nelsen 2007). In surge generation, the mechanics indicate that large surge heights are likely to occur with combinations of large ΔP and $R_{\max}$. This means the tail dependence between large ΔP and large $R_{\max}$ is strongly relevant to the frequency of severe surge events. To quantify the difference of the tail dependence between large ΔP and large $R_{\max}$ in full dependence models, a tail dependence estimator λ_t is defined here as the conditional probability of a large $R_{\max}$ value given a large ΔP value, and is expressed by:

$$\lambda_t = P(R_{\max} \geq r_{\max}^{+} \,|\, \Delta P \geq \delta p^{+}) \tag{18}$$

Where $r_{\max}^{+}$ represents a large $R_{\max}$ value and δp^{+} represents a large ΔP value.

In this study, we used tens of thousands of discretized TC atmospheric parameters combinations to numerically estimate λ_t (denoted as $\hat{\lambda}_t$) for a series of combinations of large values of ΔP and $R_{\max}$ in the MGC, LCGV, and LCFV models. The $\hat{\lambda}_t$ is thus calculated as:

$$\hat{\lambda}_t = \frac{\sum_{\delta p_i \geq \delta p^{+}} \sum_{r_{\max_i} \geq r_{\max}^{+}} \sum_{v_{\mathrm{f}_i}} \sum_{\theta_i} p\left(\delta p_i, v_{\mathrm{f}_i}, r_{\max_i}, \theta_i\right)}{\sum_{\delta p_i \geq \delta p^{+}} \sum_{r_{\max_i}} \sum_{v_{\mathrm{f}_i}} \sum_{\theta_i} p\left(\delta p_i, v_{\mathrm{f}_i}, r_{\max_i}, \theta_i\right)} \tag{19}$$

Where $\delta p_i, v_{\mathrm{f}_i}, r_{\max_i}, \theta_i$ are the TC atmospheric parameter in the i th combination. The calculated $\hat{\lambda}_t$ values are listed in Table 6.

Table 6: Calculated tail dependence estimator in the MGC, LCGV, and LCFV models

	$\delta p^{+} = 132.21$ hPa	$\delta p^{+} = 137.47$ hPa	$\delta p^{+} = 142.74$ hPa

	$r^{+}_{\max} = 151.58$ km	$r^{+}_{\max} = 161.05$ km	$r^{+}_{\max} = 170.53$ km
MGC	$\hat{\lambda}_t$ =8.0e-6	$\hat{\lambda}_t$ =2.2e-6	$\hat{\lambda}_t$ =5.3e-7
LCGV	$\hat{\lambda}_t$ =1.4e-4	$\hat{\lambda}_t$ =4.9e-5	$\hat{\lambda}_t$ =1.4e-5
LCFV	$\hat{\lambda}_t$ =2.1e-3	$\hat{\lambda}_t$ =1.2e-3	$\hat{\lambda}_t$ =6.2e-4

Table 6 shows that the MGC model always has a lower tail dependence between large ΔP and large $R_{\max}$ than vine models (LCGV and LCFV). Thus, the conditional probability of large $R_{\max}$ values given large ΔP values is smaller for the MGC. The LCFV model always has a higher tail dependence between large ΔP and large $R_{\max}$ than the LCGV model. It is challenging to ascertain a "truth" regarding tail dependence from limited historical observations. Higher values of tail dependence between large ΔP and large $R_{\max}$ can be "conservative" because they assign a higher probability to combinations involving concurrently high values of ΔP and $R_{\max}$, which will induce higher frequency of high storm surges at locations where high storm surges are connected with combinations of large ΔP and large $R_{\max}$. However, this effect might not be notable to the hazard curve if combinations of high values ΔP and $R_{\max}$ are associated with relatively low probability regardless of the dependence between them.

3.4.3 Hazard curves sensitivity assessment

To compare storm surge hazard curves, the estimated surge curves are plotted in Figure 27. Confidence intervals are also shown for the surge curve generated using the MGC with zero-degree convention set as due North. In some regions, alternate strategies may be used to define the zero-degree convention for heading (e.g., through strategic selection of the zero-degree convention to correspond to the mean heading relative to due North), which may help to mitigate the effect of the circular variable. However, in this region, the mean of the heading distribution relative to due north is approximately zero, so this strategy was not explored herein. The confidence intervals are computed by assuming the uncertainty follows a Gaussian distribution with zero mean and standard deviation equal to an uncertainty term combining both absolute and relative uncertainties (Nadal-Caraballo et al. 2022b).

Figure 27 plots the surge curves generated using the five JPM assumptions for the four study nodes. Note that the plot range of η for each node is different.

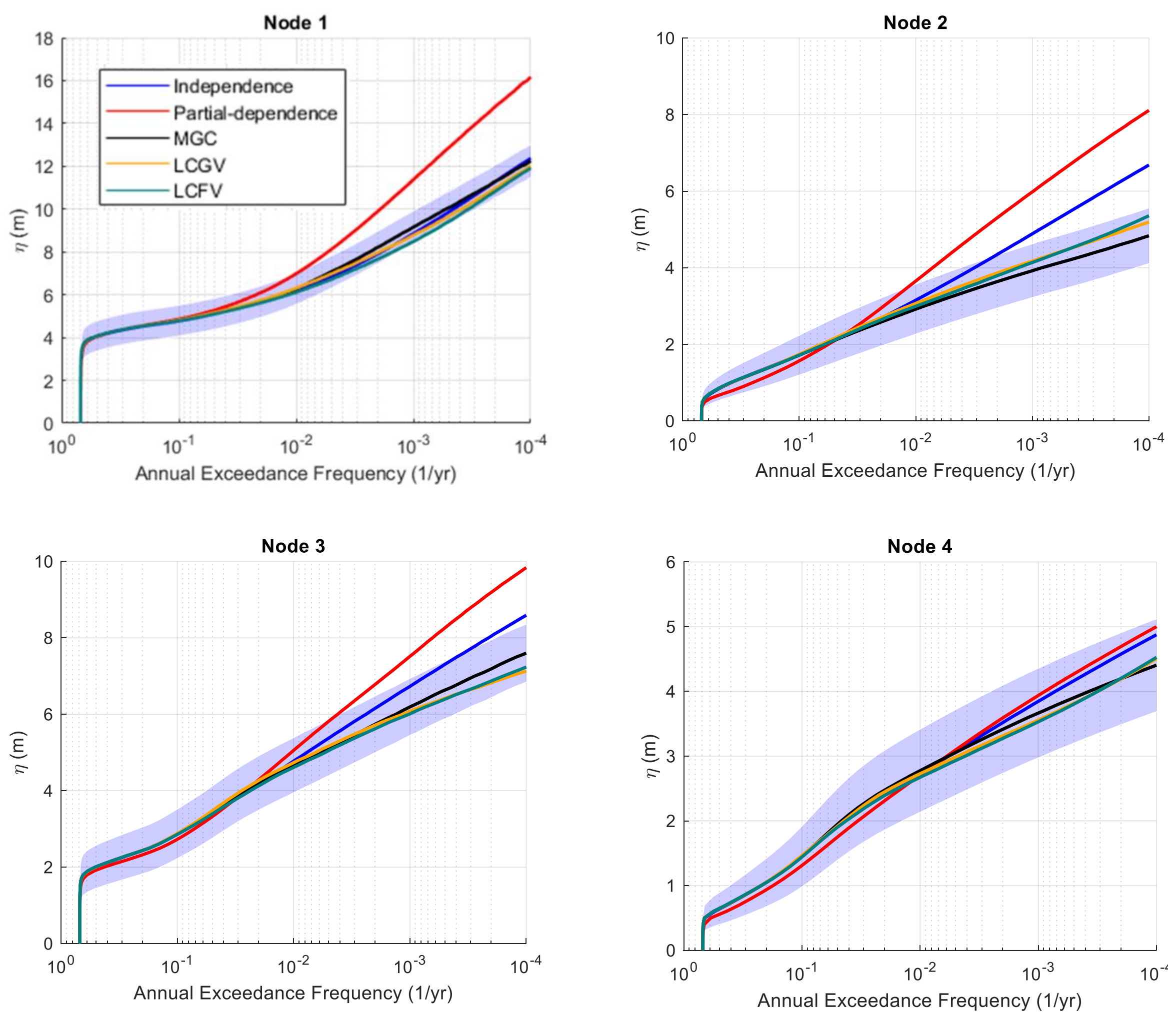


Figure 27: Comparison of storm surge hazard curves using the five JPM assumptions for the four study nodes (shaded area shows the 90% confidence interval of MGC)

To quantify the surge curves' sensitivity to JPM assumptions, Table 7 lists the 10^{-3} AEF surge elevation estimated (with a 90% confidence interval) across the JPM assumptions for the four study nodes. All confidence intervals are computed using the combined uncertainty term (Nadal-Caraballo et al. 2022b).

Table 7: Estimated 10^{-3} AEF surge elevation

JPM assumption	Independence	Partial-dependence	MGC	LCGV	LCFV
Node 1	8.9±0.8 m	11.4±0.8 m	9.2±0.8 m	8.8±0.8 m	8.5±0.8 m
Node 2	4.9±0.7 m	6.0±0.7 m	3.9±0.7 m	4.2±0.7 m	4.1±0.7 m
Node 3	6.7±0.8 m	7.5±0.8 m	6.2±0.7 m	6.1±0.7 m	6.0±0.7 m
Node 4	3.8±0.7 m	3.9±0.7 m	3.7±0.7 m	3.6±0.7 m	3.5±0.7 m

From the above sensitivity analysis visualization and quantification, we observe that those surge curves (and 10^{-3} AEF storm surge elevations) generated from full dependence joint probability models (i.e., MGC, LCGV, and LCFV) can be notably different from those generated with the independence and partial-dependence joint probability models (for Node 1~3). Among those full dependence models, differences are generally less notable and vary by location.

While we know that the linear-circular copula-based vine copulas (LCGV and LCFV) are able to model the linear-circular dependence between Θ and other TC atmospheric parameters, they are more challenging to implement compared to the MGC assumption. Therefore, we investigate the sensitivity of the surge curve generated from the MGC assumption to the zero-degree convention.

Figure 28 plots MGC surge curves with different zero-degree conventions. For visualization, only three zero-degree conventions are included in the plots. It can be observed that the surge curves generated with MGC vary with the change of zero-degree convention. This sensitivity can be caused by the use of linear correlation coefficients in the MGC as described in Section 4.1.

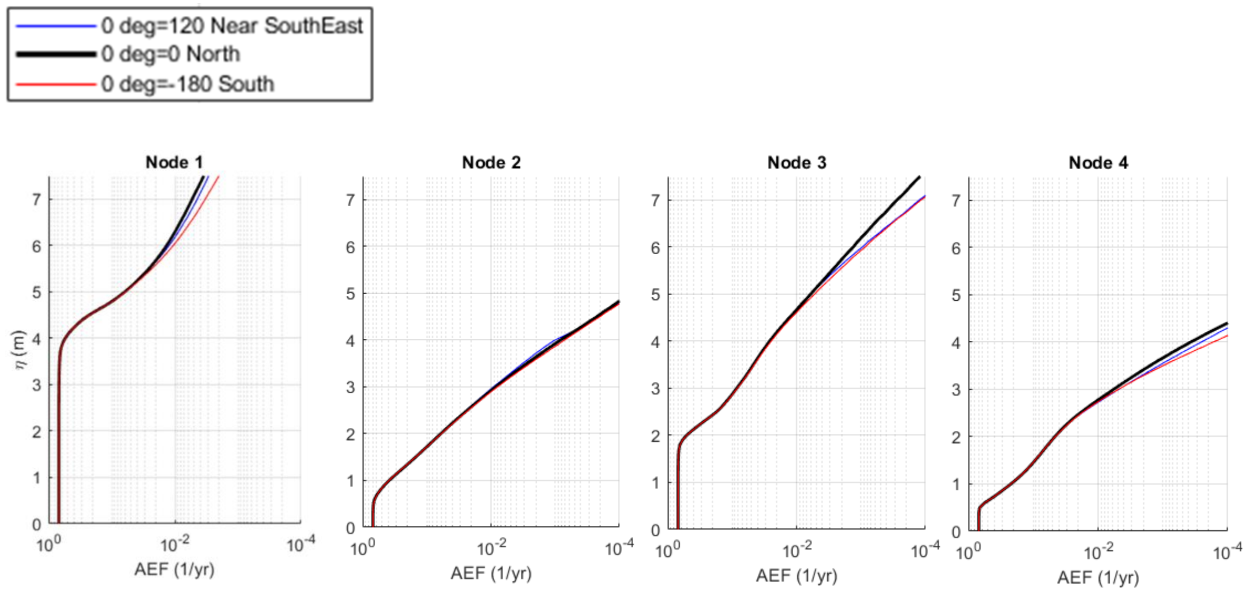


Figure 28: Surge curve with multiple zero-degree conventions using MGC assumption

To visualize the advantage of stability in linear-circular copula-based vine copulas, two cases that can show this advantage (Node 1 MI conditional surge curve and Node 2 MI conditional surge curve) are plotted in Figure 29. The conditional surge curve refers to the curve with the conditional probability of exceedance (i.e., probability of exceedance given the occurrence of a TC) on the x-axis and the associated surge height on the y-axis. The conditional probability of exceedance curve does not include the storm recurrence rate. The left panel of Figure 29 shows how the surge curve generated using the MGC varies when the zero-degree convention for Θ is changed. The impact on the 10^{-2} conditional probability of exceedance surge value is approximately 0.4 m for Node 1 MI and 0.2 m for Node 2 MI. Under LCGV and LCFV (shown in the center and right panels of Figure 29, respectively), the shown surge curves have better stability with the varying of zero-degree convention. In other cases (not shown), either the instability of surge curves under the MGC is trivial, or the LCGV and LCFV are notably unstable because of the sensitivity of the procedure used to fit the linear-circular copula to the limited available statistical sampling data.

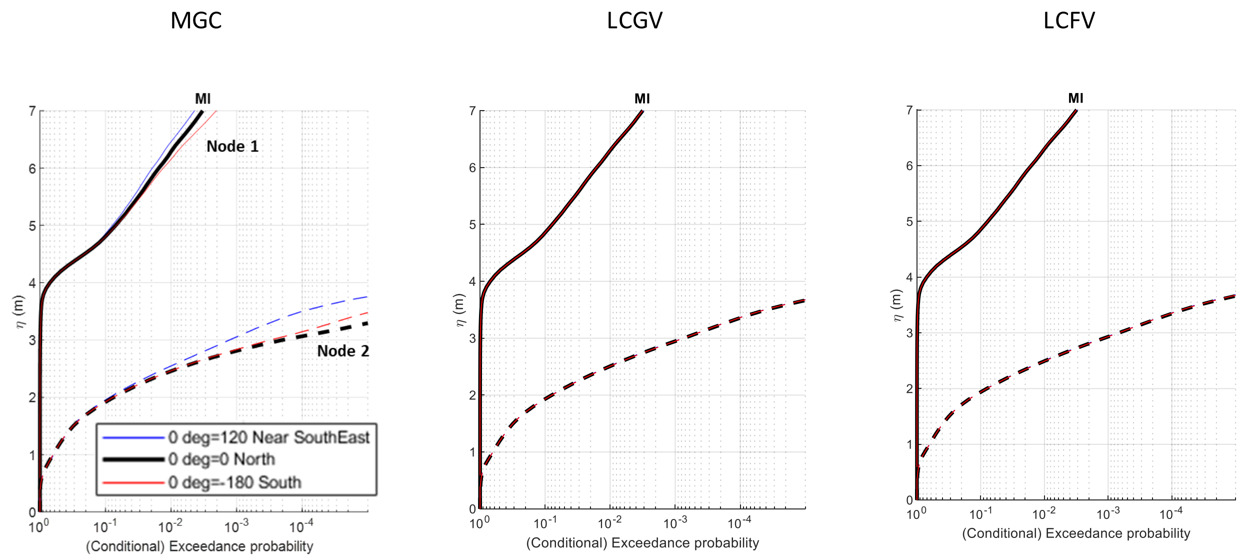


Figure 29: Comparison of storm surge hazard curve's sensitivity to different zero-degree conventions between MGC assumption (left panel) and vine copula assumptions (right and center panel). The x axis is the conditional exceedance probability for MI TC

From the above sensitivity visualization plots, we observed that the surge curves generated from the MGC assumption might be affected by the selection of the zero-degree convention. Implementing Linear-circular copula-based vine copula models can mitigate the instability issues caused by circular variables in MGC assumption in some cases. However, it may result in increased instability due to the sensitivity of the linear-circular vine copula procedure to limited amounts of sampling data.

3.5 Conclusion of Chapter 3

In this paper, a series of joint distribution models for TC atmospheric parameters were constructed with different JPM assumptions. It is found that full dependence models generate notably different hazard curves relative to independence and partial-dependence assumptions. Full dependence models considered in this study include the MGC model, LCGV model, and LCFV model. The MGC has the advantage of ease of implementation. However, it cannot accommodate circular variables such as storm heading (Θ). It is observed that the estimated linear dependence between the circular Θ and other (linear) TC atmospheric parameters varies with the shift of zero-degree convention. The constructed hazard curves also showed sensitivity to the shift of zero-degree convention in some cases.

The linear-circular vine copulas discussed in this paper can be implemented to capture the full dependence among TC atmospheric parameters and mitigate the potential hazard curve instability induced by the circular variable when using the MGC. The MGC assumption's sensitivity to the shift of zero-degree convention was investigated and compared against vine models that include linear-circular copulas. This paper assessed the use of a VKF (in lieu of the GKF) for characterizing the marginal distribution of Θ, which offers stability to the marginal distribution of Θ regardless of the selection of zero-degree convention.

It is also found that the MGC might result in a lower tail dependence between large ΔP and large $R_{\max}$ (compared to LCGV and LCFV), which may induce a lower estimation of the AEF of extremely severe hazards at locations where high storm surges are connected with a combination of large ΔP and large $R_{\max}$. However, this difference might not have a substantial effect on the hazard curves if the likelihood of high values ΔP and large $R_{\max}$ remains relatively small, regardless of the strength of dependence.

In the case study presented in this paper, the result of the investigation of the MGC assumption's sensitivity to the shift of zero-degree convention shows that its instability issue related to circular variables is moderate. However, this does not imply that the MGC assumption has robust performance in all cases. TC's heading direction in the New Orleans region, which is located on the southern U.S. coast on the Gulf of Mexico, has a distribution concentrated around the North direction. This fact might mitigate the instability issue in MGC. Additional research is needed to assess the robustness of hazard curves to this convention in other regions.

Chapter 4: A Deaggregation Method of Coastal Hazard Combinations Based on Bayesian Network

4.1 Introduction

Over the past few decades, the joint probability method (JPM) has been implemented for coastal hazard frequency analysis, particularly focusing on storm surge hazards. In the JPM, the joint distribution models of storm parameters are constructed, and the joint probability of storm parameters combinations are computed and assigned to their associated storm surge values (Nadal-Caraballo et al. 2015, 2020; Toro 2008; Toro et al. 2010). A numerical integration method is then typically applied to construct a hazard curve. Constructing a surge hazard curve that captures the possible breadth of tropical cyclone (TC) storm parameter combinations and their associated surge values requires running many synthetic storms in high-fidelity hydrodynamic numerical models. These simulations can be highly computationally expensive. In response, several strategies have been developed to reduce the computational burden of implementing the JPM. Quadrature-based (Toro 2008; Toro et al. 2010) and response-surface approaches (Irish et al. 2009; Resio et al. 2009), typically referred to as optimal-sampling approaches, were developed to improve the efficiency of storm surge hazard analysis by reducing the number of required synthetic storms. However, these techniques may still require substantial effort and may have other potential disadvantages, such as the dependence of certain quadrature approaches on the probability assumptions made regarding the joint distribution of storm parameters. Recently, machine-learning methods have been used to generate surrogate models that augment high-fidelity numerical models (Jia and Taflanidis 2013; Jia et al. 2016;

Taflanidis et al. 2014; Nadal-Caraballo et al. 2020; Lee et al. 2021; Nadal-Caraballo et al. 2022a; b). However, surrogate models are not as accurate as high-fidelity models and are not inherently constrained by physical considerations or limits. Therefore, their use to predict surges for parameter combinations outside the training data may lead to inaccurate results. These challenges may be exacerbated when predicting and assessing hazards associated with compound events.

Alternatively, a hybrid approach can be used in which a lower fidelity model is used to develop an initial probabilistic model. Then, strategic probabilistic inference may be used to identify the portions of the TC parameter domain for which high-fidelity simulations should be performed to more accurately model coastal hazard effects. A lower-fidelity model may, for example, be a computationally efficient numerical model or a machine-learning-derived surrogate model trained using a limited number of high-fidelity simulations for a refined grid. Such a hybrid approach may also be used to process the results of a regional study and make selections regarding high-fidelity numerical model runs to be performed for a local-scale study.

We posit that the limited computing resources for high-fidelity simulation be focused on storm parameter combinations associated with low exceedance probability coastal hazards that are most likely to bring consequential impacts to a study region and might be most relevant for risk-informed decision-making. This study proposes a method to identify the dominant TC parameter combinations (TCPCs) that account for low exceedance probability coastal hazard events in specific locations. Specifically, a Bayesian modeling approach is implemented to compute the under-evidence distribution of TC parameters and deaggregate TCPCs contributing to low exceedance probability coastal hazards associated with two types: storm surge and storm total rainfall (STR). This study leverages the Coastal Hazard System's (CHS) (ERDC 2023)

Probabilistic Coastal Hazard Analysis (PCHA) framework developed by the U.S Army Corp of Engineers (USACE) (Nadal-Caraballo et al. 2020) and adapts it within a Bayesian network (BN)-based modeling approach that leverages the work of Mohammadi et al. (2023). Leveraging the BN's capabilities, analyses are performed under the occurrence of low exceedance probability storm surge and storm rainfall to identify storm characteristics associated with multiple hazards. In addition, the dependence between different coastal hazards is investigated. The distributions of one type of coastal hazard under evidence cases of low exceedance probability event of another type of coastal hazard are investigated. These analyses can provide insights to inform the TC parameter domain for which refined analyses are particularly relevant. A case study is applied to the New Orleans region to demonstrate the proposed approach.

4.2 Method

Figure 30 presents a flowchart of the analysis procedure proposed for deaggregation of low exceedance probability coastal hazards. First, a BN is developed for modeling one or more coastal hazards as a function of TC parameters, as shown in Box [1] of Figure 1. The general BN structure proposed in this study is described in Section 2.1. Next, statistical analyses and modeling activities are performed to define the relationship between random variables modeled in the BN. This includes the collection and processing of historical TC data, as shown in Box [2] and described further in Section 2.2. Using historical TC data, marginal and joint distribution models of TC parameters are developed (Box [3], Section 2.3). Then, functions are developed to map TC parameters to coastal hazard response parameters (Box [4], Section 2.4). The results of statistical and numerical modeling activities are used to develop conditional probability tables

(CPT) required by the BN (Box [5], Section 2.5). Finally, the deaggregation is applied and used to identify dominant TCPCs (Box [6], Section 2.6).

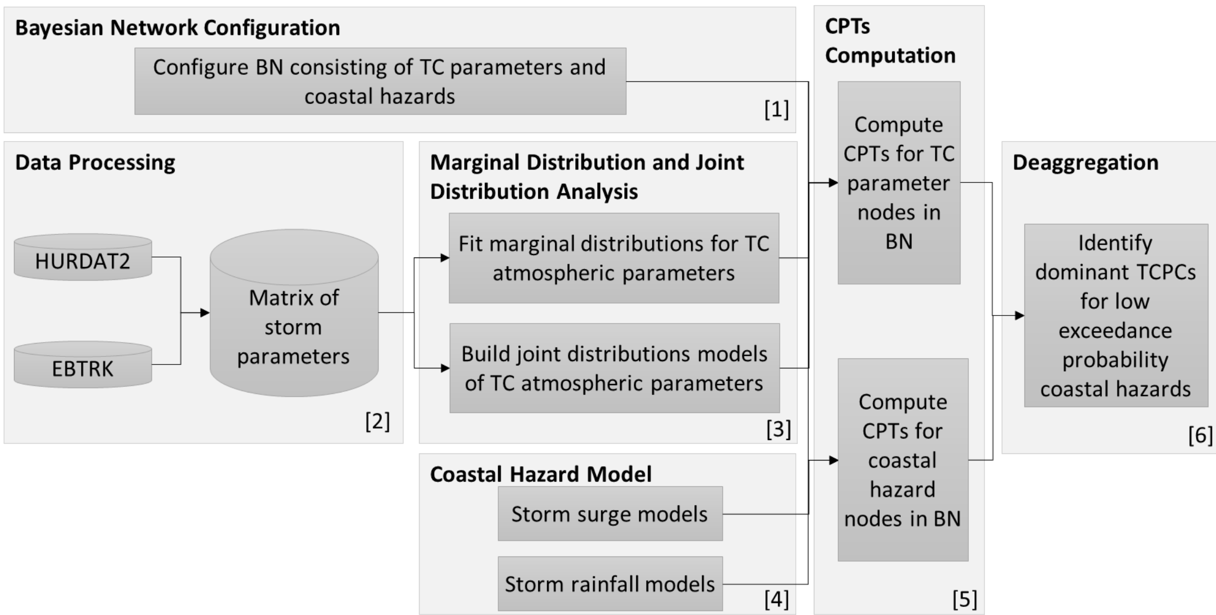


Figure 30: Analysis procedure for the deaggregation of low exceedance probability coastal hazards

4.2.1 Bayesian network configuration

A BN is a Bayesian rule-motivated graphical model that represents the probabilistic relationships among multiple dependent variables. It excels in visualizing the dependence relationship in complex probabilistic models and supports efficient calculations of models involving complex dependencies. BNs have been selectively implemented in coastal and flood hazard assessments. For example, Sebastian et al. (2017) described a non-parametric BN-based storm events simulation method to determine the hydraulic boundary conditions for a low-lying coastal watershed in the Houston-Galveston region. Couasnon et al. (2018) built a BN based on a Gaussian copula to generate coastal watershed boundary conditions in Southeast Texas. Flores et al. (2019) created a hybrid dynamic object-oriented BN to model the risk of flooding in a

Mediterranean catchment in the south of Spain. Wu et al. (2020) coupled ontology and BN for urban pluvial flood risk evaluation in Zhengzhou City. Sanuy et al. (2020) built a BN to characterize the relationship between wave conditions, coastal erosion, and coastal inundation in the Catalan coast region. Mohammadi et al. (2023) implemented a BN framework for coastal compound flood analysis. The foundational concept in Mohammadi et al. (2023) is leveraged in the approach proposed in this study.

To illustrate key concepts associated with BNs, Figure 31 presents a BN with four fully dependent variables. In a BN, each node represents a variable, and each arrow represents the (typically causal) relationship from a "parent" node to a "child" node. For example, in the BN shown in Figure 31, X_1 is a parent node of X_2, X_3 and X_4. X_4 is a child node of X_1, X_2 and X_3. To build a BN model that can be implemented for probabilistic analysis, every variable typically needs to be discretized (with the exception of selected special cases), and a CPT needs to be assigned for each variable. Additional information regarding BNs can be found in a range of references, e.g., Kjaerulff and Madsen (2008).

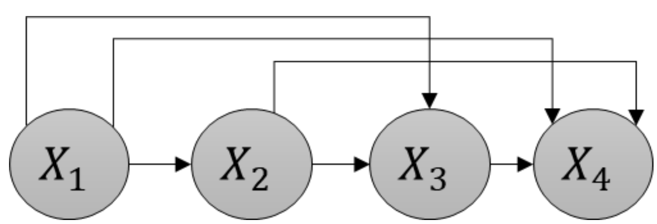


Figure 31: Example BN with four fully dependent variables

The BN enables multiple forms of probabilistic inference. Forward (predictive) inference calculates the probability distribution of nodes by propagating the information in the direction of the arrows. In this study, this forward inference of BN is leveraged to generate the probability distribution of coastal hazards given the probability distribution of TC parameters (see Section 3.5.1), which is consistent with the direction of inference used in probabilistic coastal hazard analysis. In addition, evidence cases (i.e., cases in which the values of nodes are assigned a particular known or assumed state) can be set for TC parameters to generate under-evidence probability distributions of coastal hazards. Backward inference calculates the probability distribution of nodes by propagating the information in the opposite direction of the arrows. In this study, this backward inference of BN is leveraged to generate the under-evidence probability distribution of TC parameters given evidence cases in which coastal hazards equal or exceed a particular value (see Section 3.5.2). It is also possible to perform analyses involving combinations of inference in which information is first propagated in the direction opposite of the arrows. In turn, the information propagates back down the network in the direction of the arrows. This type of inference is leveraged to generate the under-evidence probability distribution of one coastal hazard, given the evidence case in which another coastal hazard equals or exceeds a particular value (see Section 3.5.2).

A general BN structure depicting the probabilistic model of a TC-induced coastal hazard is configured and presented in Figure 32. The physical definition of each node is listed in Table 8. The joint distribution over all involved variables is expressed using the BN factorization as:

$$\begin{aligned} p(\delta p, v_{\mathrm{f}}, r_{\max}, \theta, x_0, \hat{r}, r, \varepsilon_R) \\ = p(r|\hat{r}, \varepsilon_R) p(\hat{r}|\delta p, v_{\mathrm{f}}, r_{\max}, \theta, x_0) p(\varepsilon_R) p(\delta p, v_{\mathrm{f}}, r_{\max}, \theta, x_0) \end{aligned} \tag{20}$$

$$p(\delta p, v_{\mathrm{f}}, r_{\max}, \theta, x_0) = p(\theta|\delta p, v_{\mathrm{f}}, r_{\max}) p(r_{\max}|\delta p, v_{\mathrm{f}}) p(v_{\mathrm{f}}|\delta p) p(\delta p) f(x_0) \tag{21}$$

Where $p(\square)$ is a discretized probability mass function, and lowercase notation denotes a particular outcome of the random variables presented in Table 8 (e.g., δp denotes an outcome of the continuous random variable ΔP) . Implementing the BN requires the specification of CPTs corresponding to the conditional and marginal distributions specified above. BN inference algorithms are then used to efficiently perform computation (inference) involving the joint distribution over all random variables.

This BN configuration leverages the USACE's CHS-PCHA framework (Nadal-Caraballo et al. 2020) as well as the work of Mohammadi et al. (2023). However, modifications of the BN topology can be made to reflect new or different information regarding TC parameters and their relationship to coastal hazard response quantities. This general structure is extended to multiple hazards in conjunction with the case study (see Section 3.1).

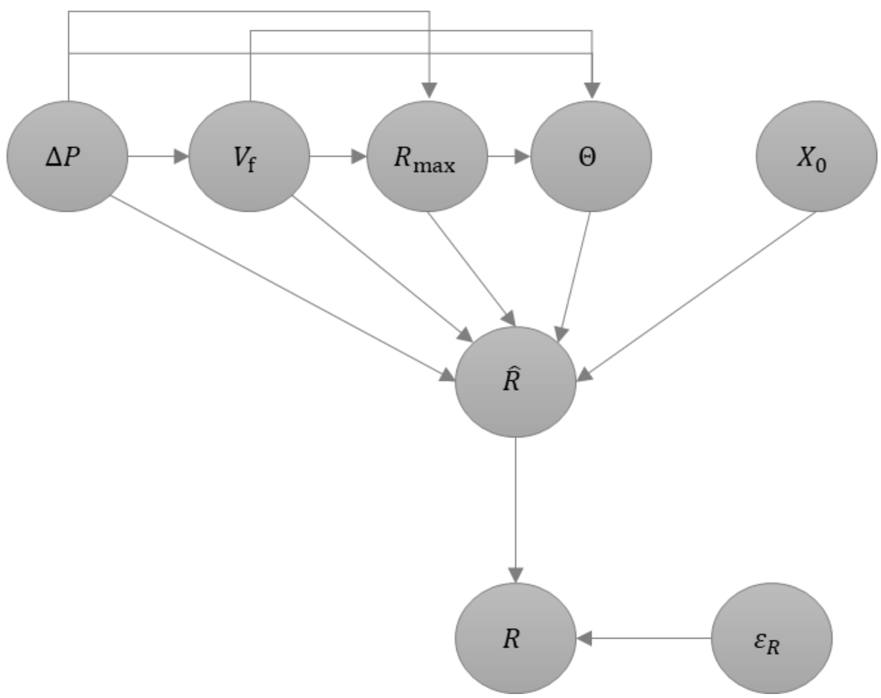


Figure 32: BN that depicts the probabilistic model of TC-induced coastal hazard

Table 8: Physical definition of each node in BN

Notation	Physical definition	Notation	Physical definition
ΔP:	Central pressure deficit	X_0:	Storm landfall location

V_{f}:	Forward velocity	$\hat{R}$:	Coastal hazard severity measure, as predicted by a model
$R_{\max}$:	Radius of maximum winds	R:	Coastal hazard severity measure
Θ:	Storm heading direction	ε_R:	Error node accounting for prediction model error

4.2.2 Data processing

Following the development of the BN model topology, data are collected and processed to support the development of probabilistic relationships between random variables included in the BN. In most probabilistic coastal hazard assessments, the relationship among TC parameters is based on observational or reanalysis data associated with historical TCs, e.g., Hurricane Database 2 (HURDAT2) (Landsea and Franklin 2013; NOAA 2022c) and Tropical Cyclone Extended Best Track Dataset (EBTRK) database (CIRA 2021; Demuth et al. 2006).

4.2.3 Marginal distribution and joint distribution analysis

Next, statistical analysis is performed to characterize the relationships among TC atmospheric parameter nodes in Figure 32 based on the joint distribution models of TC atmospheric parameters. For example, in the USACE's CHS-PCHA framework (Nadal-Caraballo et al. 2020, 2022a; b), a series of coastline nodes are selected as coastal reference locations (CRL), and statistical analysis is applied to each CRL individually. The statistical sampling data of each CRL is generated with a distance-weight-adjustment process (Chouinard and Liu 1997; Nadal-Caraballo et al. 2015) using storm track points with a maximum intensity index within a specified capture zone (Nadal-Caraballo et al. 2015). Then, the statistical sampling data is used for marginal distribution fitting, dependence estimation, and development of the joint distributions model using copula analysis.

4.2.4 Coastal hazard model

Next, coastal hazard models are implemented to capture the relationship between TC parameter nodes and the coastal hazard node in Figure 32. In the past decades, different models have been applied to coastal hazard analysis with varying degrees of fidelity. For example, for storm surge, Dietrich et al. (2012) examined the performance of the high-fidelity Advanced CIRCulation model (ADCIRC) and Simulating WAves Nearshore model (SWAN) in computing hurricane waves and surges in the Gulf of Mexico and Louisiana regions. Gonçalves and Soares (2012) compared the performance of SWAN and steady-state spectral wave models in shallow water. Recently, a series of USACE studies leveraged ADCIRC and wave models for storm response simulations (Gonzalez et al. 2019; Nadal-Caraballo et al. 2015, 2020, 2022b).

For storm rainfall, a series of parametric models including R-CLIPER (Lonfat et al. 2004; Marks and DeMaria 2003), IPET (Chen et al. 2006; Lonfat et al. 2004), PHRaM (Grieser and Jewson 2011; Lonfat et al. 2004) and P-CLIPER (Lonfat et al. 2004; Marks and DeMaria 2003) has been developed. Brackins and Kalyanapu (2020) provide a comparison of these storm rainfall parametric models. While these parametric storm rainfall models are efficient, a physics-based numerical rainfall model can generally provide more accurate storm rainfall simulations (Lu et al. 2018; Xi et al. 2020; Zhu et al. 2013) but with higher computational cost.

Although high-fidelity numerical models can generate accurate coastal hazard simulations, they are usually time-consuming and computationally expensive to develop and use. Surrogate models and lower fidelity models (e.g., Jelesnianski 1992) can be taken as alternatives to numerical models and can make efficient simulations with varying degrees of accuracy. We propose that deaggregation can be implemented to extract dominant TCPCs for coastal hazards with a return period of interest to reduce computational burdens, and then a high-accuracy model can be applied to the portion of dominant TCPCs.

4.2.5 Conditional probability table computation

The CPT for a node defines the conditional probability of a node given each of the mutually exclusive combinations of the states of its parent nodes. Quantification of the BN requires specification of CPTs that provide the probability that a node X_i is in a specified state ($x_i^- < X_i < x_i^+$) given the state of the parent node $pa(X_i)$: $P\left(x_i^- < X_i < x_i^+ \middle| pa(X_i)\right)$. Given a joint probability density function of a set of m random variables $f_{\mathbf{X}}(x_1, x_2, \dots x_m)$ (e.g., derived from copula analysis), the joint probability mass (PM) of a discretized variable combination $p_{\mathbf{X}}(x_1, x_2, \dots x_m)$ can be calculated as:

$$p_{\mathbf{X}}(x_1, x_2, \dots x_m) = P(x_1^- < X_1 < x_1^+ \cap \dots \cap x_m^- < X_m < x_m^+) = \int_{x_1^-}^{x_1^+} \int_{x_2^-}^{x_2^+} \dots \int_{x_m^-}^{x_m^+} f_{\mathbf{X}}(x_1, x_2, \dots x_m) dx_1 dx_2 \dots dx_m \quad (22)$$

where x_i^- is the lower limit of discretized bin corresponding to point $X_i = x_i$ and x_i^+ is the corresponding upper limit. Therefore, the conditional probability of any subset of variables can be derived as:

$$p_{\mathbf{X}}(x_1, x_2, \dots x_i | x_{i+1}, \dots x_m) = \frac{p_{\mathbf{X}}(x_1, x_2, \dots x_m)}{p_{\mathbf{X}}(x_{i+1}, \dots x_m)} \quad (23)$$

Given the implementation described later in Section 3, the joint distribution of TC atmospheric parameter ΔP, V_{f}, $R_{\max}$ and Θ will be defined via copula analysis, and their CPTs required by the

BN will be derived from the joint distribution. In this study, the TC landfall location X_0 is assumed to be uniformly distributed along TC track spacing.

The CPT of the model-estimated coastal hazard node $\hat{R}$ is computed using coastal hazard models as described in Section 2.4. The predicted value of the coastal hazard severity measure ($\hat{R}$) is a function of TC parameters (ΔP, V_{f}, R_{max}, Θ, X_0), as indicated by the arrows from the TC parameters to $\hat{R}$. Then, the value of the coastal hazard R is computed as:

$$R(\mathbf{x}) = \hat{R}(\mathbf{x}) + \varepsilon_R \sigma_c(\hat{R}(\mathbf{x})) \tag{24}$$

Where: $\mathbf{x}$ represents the vector of TC parameters, ε_R is the normalized residuals of R, which follows a standard Gaussian distribution, σ_c is the standard deviation of a combined uncertainty term, which is a function of $\hat{R}(\mathbf{x})$ (Nadal-Caraballo et al. 2022b).

Monte Carlo simulation (MCS) is implemented for the computation of the CPT of nodes $\hat{R}$ and R for the purpose of reducing the error induced by the discretization of TC parameters Mohammadi et al. (2023). In this MCS process, a number of TCPCs are generated within each combination of TC parameters' discretized bins. For the CPT of node R, tens of thousands of normalized residuals ε_R are generated using a standard Gaussian distribution.

4.2.6 Deaggregation

Deaggregation (also known as disaggregation) is a method leveraged in probabilistic seismic hazard analysis for the purpose of identifying dominant scenarios for a given seismic hazard (Baker et al. 2021). In this study, we adapted the deaggregation method from probabilistic seismic hazard analysis (PSHA) described in (Baker et al. 2021) to PCHA to identify dominant TCPCs (scenarios of coastal hazards) under evidence cases of coastal hazard events.

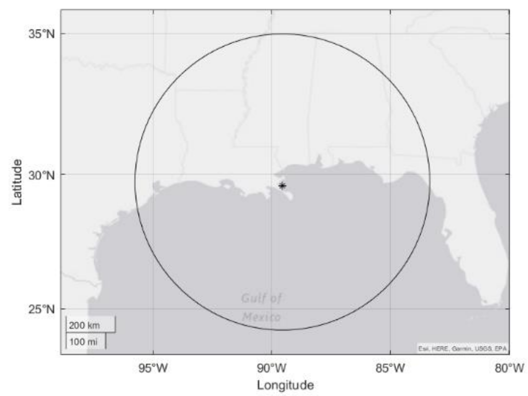
The probability of TCPC vector $\mathbf{X}$ taking on states $\mathbf{x}$, given coastal hazard $R > r$ can be expressed as:

$$P(\mathbf{X} = \mathbf{x}|R > r) = \frac{P(\mathbf{X}=\mathbf{x}, R>r)}{P(R>r)} \tag{25}$$

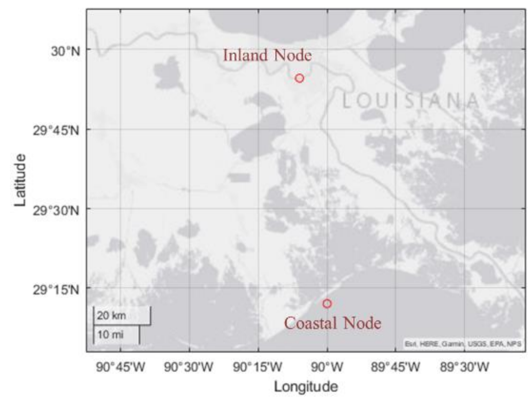
Where $\mathbf{x} = [\delta p, v_{\mathrm{f}}, r_{\max}, \theta, x_0]$ is a combination of TC parameters (each discrete value corresponding to a bin, as described in Section 2.5). We leverage the BN to perform the deaggregation (as described in Bensi 2010) by entering evidence in the BN corresponding to $R > r$ and using backward inference to obtain the joint PM of TC parameters $\mathbf{X}$ given the specified evidence state. In this study, the GeNIe software (BayesFusion 2023) is used for BN computation. The clustering algorithm (Dawid 1992; Jensen 1990; Lauritzen and Spiegelhalter 1988) is used to efficiently calculate the joint PMs of TCPCs. A dummy child node under the clustering algorithm allows the preservation of clique potentials, from which the joint PMs of its parent nodes can be computed (BayesFusion 2023). Then, the range of TCPCs with the total joint PMs that dominantly contribute to coastal hazards under corresponding evidence can be identified as the dominant TCPCs (see Section 3.5.2).

4.3 Case Study

A case study was performed in the region of New Orleans, LA. A coastal node (latitude = 29.58 and longitude = -89.54) was selected as the CRL for TC atmospheric parameter statistical analysis (see Section 3.3). Figure 33(a) plots the location of the CRL and its 600 km radius capture zone. Historical storm data from 1938 to 2020 are included in the statistical analysis, consistent with Nadal-Caraballo et al. (2022b). One inland river node (Node 1) and one coastal node (Node 2) are selected as study nodes for computing coastal hazards. The locations of these two study nodes are identified in Figure 33(b).



(a)



(b)

Figure 33: (a) CRL [29.58, -89.54]and capture zone; (b)Locations of study nodes. Inland Node (Node 1) [29.91, -90.10]; Coastal Node (Node 2) [29.20, -90.00]

4.3.1 Bayesian network construction

In this case study, a BN consisting of TC parameters and coastal hazards (storm surge and STR) is constructed using the GeNIe software (BayesFusion 2023). Figure 34 shows this BN. Table 9 lists the physical meaning of each node in the BN. The discretization information of each node can be found in Appendix D.

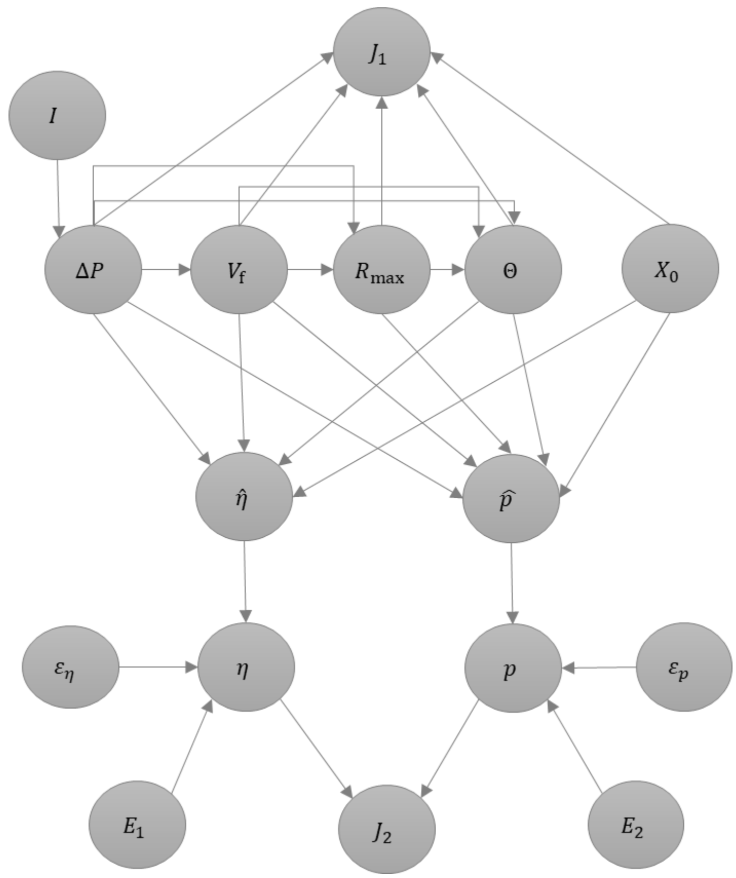


Figure 34: BN consisting of TC atmospheric parameters and coastal hazards

Table 9: Physical definition of each node in BN

Notation	Physical definition	Notation	Physical meaning
I:	Intensity category of TC (LI, MI, HI)	η:	Storm surge (m)
J_1:	Node used to compute the joint PMs of TCPCs	ε_η:	Normalized residuals of storm surge
ΔP:	Central pressure deficit (hPa)	$\hat{p}$:	STR predicted by the surrogate model (mm)
V_f:	Forward velocity (km/hr)	p:	STR (mm)
R_{max}:	Radius of maximum winds (km/hr)	ε_p:	Normalized residuals of STR
Θ:	Storm heading direction (deg)	J_2:	Node used to compute the joint PMs of storm surge and STR
X_0:	Storm landfall location index	E_1:	Node used to enter evidence to η
$\hat{\eta}$:	Storm surge predicted by the surrogate model (m)	E_2:	Node used to enter evidence to p

In this study, node J_1 is assigned as a dummy child node of TC parameter nodes and the clique potentials are preserved (BayesFusion 2023) for the purpose of computing the joint PMs of TCPCs. For computing the joint PMs of node η and node p, the memory demand restricts the application of the preserving clique potential method (BayesFusion 2023). Instead, Node J_2 is set as a child node of node η and node p, with CPT defined as an identity matrix and states representing combinations of discretized bins of node η and node p to compute their joint PMs.

4.3.2 Data processing for case study

In this study, a historical storm dataset with data imputation (Liu et al. 2023b) is used for the statistical analysis of TC atmospheric parameters. The historical dataset is a combination of HURDAT2 (Landsea and Franklin 2013; NOAA 2022c) and EBTRK databases (CIRA 2021; Demuth et al. 2006) with missing central pressure and $R_{\max}$ record imputed using Gaussian process regression (GPR) data imputation models (Liu et al. 2023b) consistent with (Nadal-Caraballo et al. 2022b).

Based on the historical storm dataset, statistical sampling data were generated by extracting the maximum intensity index track point of each storm pass through the capture zone and applying the distance-weight-adjustment (Nadal-Caraballo et al. 2015). These statistical sampling data are divided into three categories based on the value of ΔP, i.e., low intensity (LI) storms (8 hPa $\leq \Delta P <$28 hPa), moderate intensity (MI) storms (28 hPa $\leq \Delta P <$48 hPa), and high intensity (HI) storms (48 hPa $\leq \Delta P$).

Storm surge data generated from the ADCIRC and SWAN is used for storm surge surrogate model training (Nadal-Caraballo et al. 2022b), where peak surge values for each synthetic storm at each location are extracted and taken as responses in training data of storm surge surrogate model. The rainfall data used in this study for surrogate model training is generated by a physics-based TC rainfall model (TCR) (Lu et al. 2018; Xi et al. 2020; Zhu et al. 2013), which is able to simulate the time series of rainfall under TC. The rainfall time series data is accumulated to compute the STR, and then, the STR data is taken as responses in the training data of the STR surrogate model. Additional details regarding surrogate models are provided in Section 3.4.

4.3.3 Statistical analysis

In the case study, the joint distribution of storm parameters is constructed using a meta Gaussian copula (MGC), consistent with recent implementations of the USACE's CHS-PCHA framework (Nadal-Caraballo et al. 2020, 2022b). The performance of MGC is assessed in Liu et al. (2023a). Specification of the MGC requires estimations of marginal distributions as well as the correlation coefficient between the involved random variables.

In this study, the truncated Weibull distribution is used for ΔP because of its flexibility in modeling the heavy tail of ΔP. The distribution is truncated based on the three categories of storms (LI, MI, HI). The lognormal distribution is used for V_{f} and $R_{\max}$. For heading direction Θ, a directional storm recurrence rate (DSRR) (Chouinard and Liu 1997) is used to represent its probability model as:

$$\lambda_{\theta} = \frac{1}{T}\sum_{i}^{n} w(d_i)\, w(\theta_i - \theta) \tag{26}$$

where λ_{θ} is the DSRR of heading direction θ, T is the record length (yr) of historical data, $w(d_i)$ is a Gaussian kernel function (GKF) distance-based weight, and d_i is the distance between the location associated with data point i and the CRL (Nadal-Caraballo et al. 2019). Leveraging the work of Liu et al. (2023a), $w(\theta_i - \theta)$ is a von Mises Kernel function (VKF) (Taylor 2008).

Three sets of marginal distribution are fitted separately to the statistical sampling data of LI, MI and HI. For each intensity category, an individual MGC is constructed, and statistical sampling data are used for fitting marginals and correlation estimation. The fitted marginals and correlation information among TC atmospheric parameters are presented in Appendix D. Given the marginal distributions and correlation coefficients, the MGC is specified using the method of Nadal-Caraballo et al. (2022b).

In this study, landfall location X_0 is not included in the MGC model. A horizontal reference line with latitude = 29.5 is used to build a uniform probability distribution model of X_0. Fourteen reference points with index 1~14 are set on the reference line from longitude = -93 to longitude = -86 with equal spacing and equal likelihood assigned. Figure 35 plots the locations of the 14 reference points.

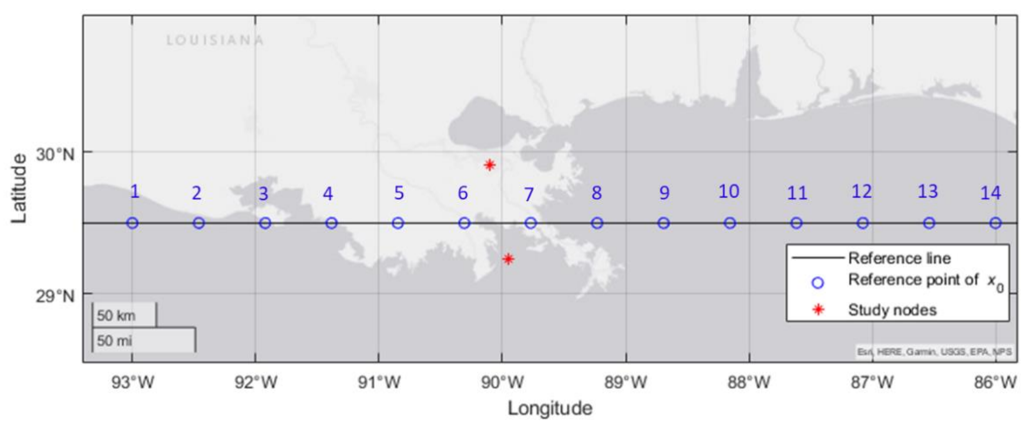


Figure 35: Representatives of x_0

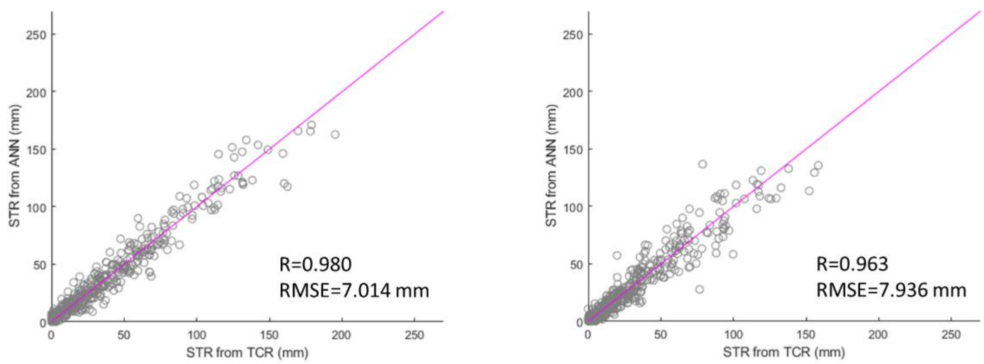
Straight line TC tracks are used to determine projected landfall locations for each combination of X_0 and discretized Θ value. The projected landfall location is used as input in coastal hazard surrogate models. The projected landfall locations are plotted in Appendix D.

4.3.4 Coastal hazard model of case study

In the case study, machine-learning-derived surrogate models are used to predict coastal hazards as a function of TC parameters. Machine-learning-derived surrogate models have been used increasingly in the estimation of storm surge hazards. For example, Jia and Taflanidis (Jia and Taflanidis 2013) implemented a GPR model for storm surge prediction with a series of follow-up studies expanding and augmenting the use of GPR (Jia et al. 2016; Taflanidis et al. 2014; Zhang et al. 2018) or leveraging or comparing alternate approaches (Al Kajbaf and Bensi 2020; Lee et al. 2021). Nadal-Caraballo et al. (Nadal-Caraballo et al. 2020) added GPR-based metamodels to the USACE's CHS-PCHA framework for water level prediction. In this case study, a GPR-based metamodel developed for USACE's CHS-LA study (Nadal-Caraballo et al. 2022b) is used as the storm surge surrogate model.

There are a number of studies that develop storm surge surrogate models, but few storm rainfall surrogate models have been developed. In this study, an artificial neural network (ANN) model is used as a storm rainfall surrogate model. The MATLAB Deep Learning Toolbox (MathWorks 2022) is used to train a multilayer feedforward back-propagation ANN model for the prediction of STR in study nodes. This ANN model contains one input layer, two hidden layers, and one output layer. Each hidden layer contains ten neurons. The training algorithm used is Levenberg-Marquardt (Levenberg 1944; Marquardt 1963). The transfer function used in hidden layers is linear. To prevent the ANN model from producing negative outcomes, which is invalid for STR values, the Tanh Sigmoid function is used as the transfer function in the output layer. The regularization strategy used is Bayesian regularization back-propagation.

In this study, TC parameters of synthetic TCs and corresponding STR generated from a numerical model are used as training data for the ANN storm rainfall surrogate model. The 645 synthetic TCs developed in the USACE's CHS-LA study (Nadal-Caraballo et al. 2022b) are input into the TCR model (Lu et al. 2018; Xi et al. 2020; Zhu et al. 2013) for TC rainfall time series simulations. The time series rainfall responses are accumulated to get the corresponding STR. Leave-one-out cross-validation (LOOCV) is applied to assess the performance of this STR surrogate model. Figure 36 presents the results of LOOCV of this STR surrogate model. It can be observed that the trained STR surrogate model has reliable performance.



(a) (b)

Figure 36: Scatter plots of LOOCV with error information of STR surrogate model at Node 1 (a) and Node 2 (b)

4.3.5 Results of Bayesian network analyses

In Section 3.5.1, the forward propagation capabilities of the BN are used to calculate the marginal and joint distributions of coastal hazard nodes under the no-evidence case (i.e., considering only the initial assumptions used to build the model prior to entering any evidence in the BN). In Section 3.5.2, forward and backward propagations are used to calculate the distributions of nodes in the BN conditioned on evidence indicating low exceedance probability coastal hazards.

4.3.5.1. Analyses under no evidence

The distribution information of coastal response quantities for the no-evidence case is represented by the PM of discretized bins and plotted in Figure 37 for nodes 1 and 2. The stem plot represents the PM value of each discretized bin, and dashed curves are created by interpolating the PM of discretized bins to approximate the distribution shapes.

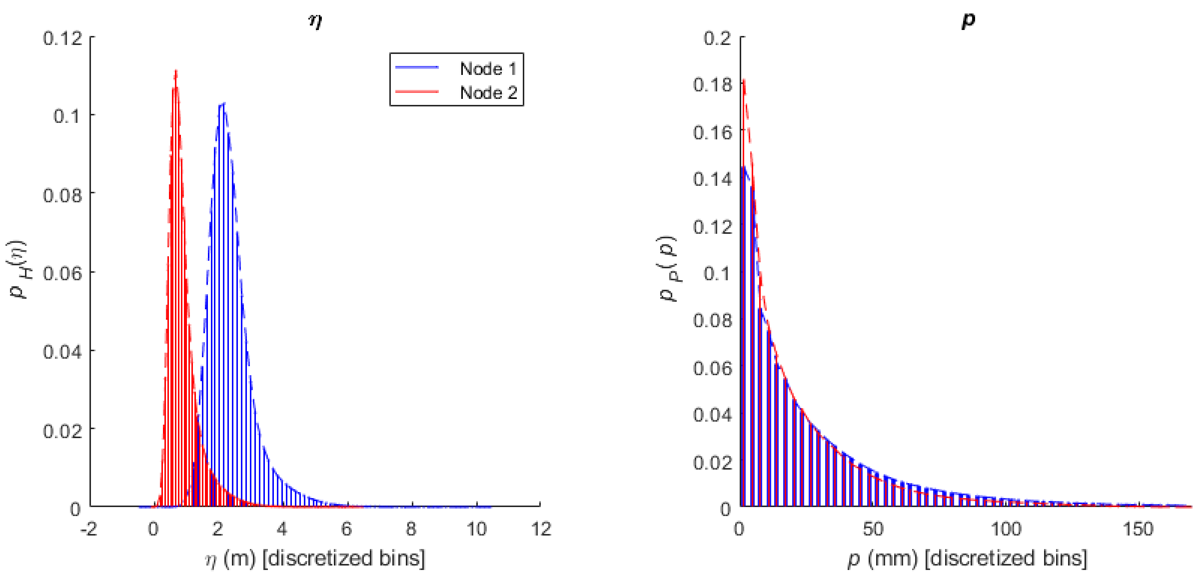


Figure 37: Coastal hazard distribution generated from BN

Using the approach described in Section 3.1, the joint PMs of storm surge and STR are determined with node J_2. Then, the PDF surfaces and joint exceedance probability surfaces are numerically approximated by scaling the computed joint PMs. Figure 38 plots the approximated joint PDF and joint exceedance probability surfaces.

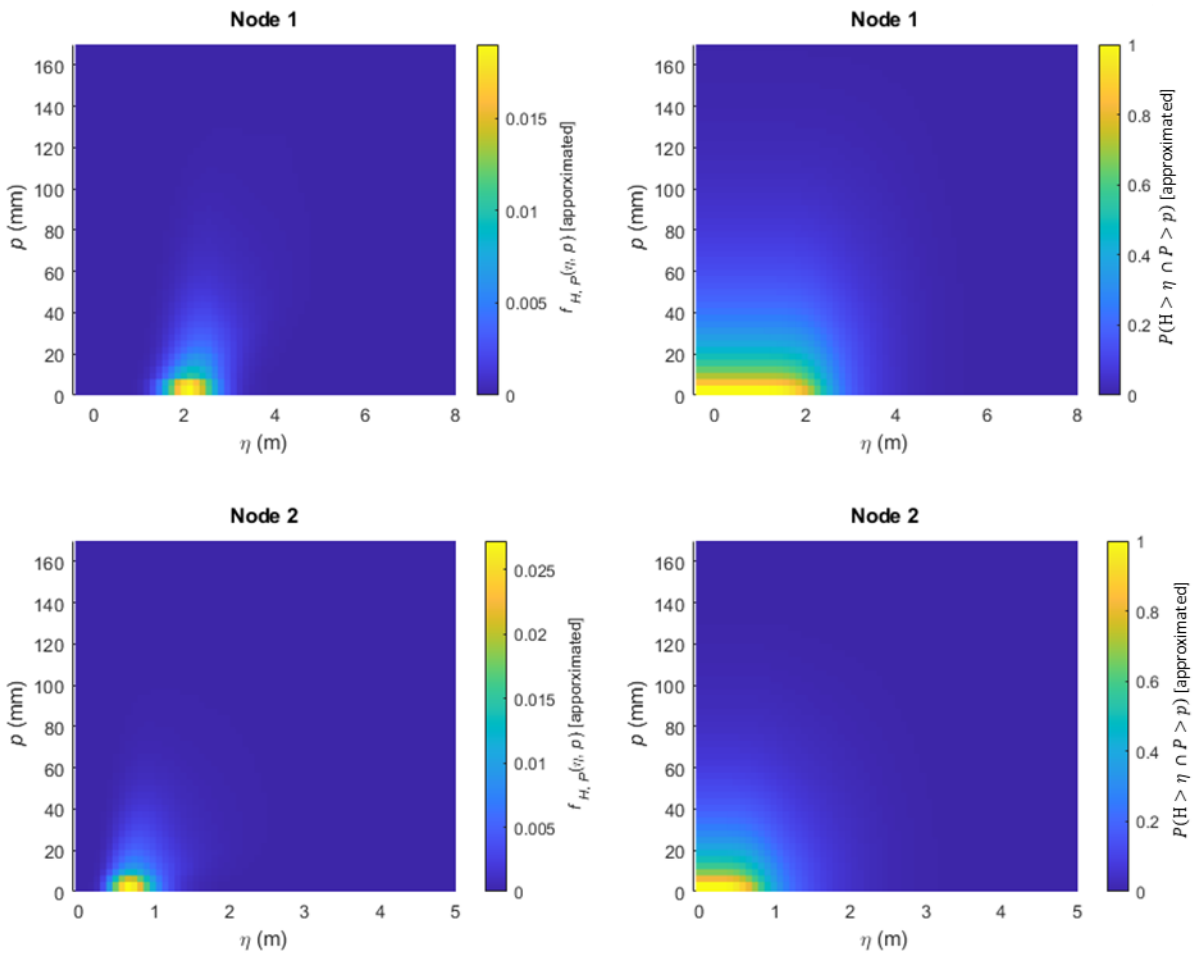


Figure 38: Approximated joint PDF and joint exceedance probability of storm surge and STR. Note that color scales are set differently for visualization

As expected, the peaks of the joint PDF and joint exceedance probability are concentrated at the left bottom area of their surface plot, reflecting the higher likelihood of smaller combinations of response quantities. Moderate correlations between storm surge and STR values are observed from the approximated joint PDF plots. The joint PMs are leveraged to approximately compute the correlation coefficients $\rho_{\eta,p}$, which is evaluated as $\rho_{\eta,p}$ = 0.5 at Node 1 and $\rho_{\eta,p}$ = 0.48 at Node 2.

4.3.5.2. Analyses under evidence

The forward and backward inference capabilities of BNs are leveraged to generate the probability distributions of TC parameters and coastal hazards for each study node under three evidence cases. The evidence cases correspond to events involving the occurrence of low exceedance probability coastal hazards. The three evidence cases are:

EC1: restricting storm surge to be greater than 4.1 m at Node 1 or 2.1 m at Node 2.

EC2: restricting STR to be greater than 107.8 mm at Node 1 or 85.9 mm at Node 2.

EC3: a multiple hazards evidence case that combines EC1 and EC2.

The EC1 and EC2 thresholds are selected based on the value of the severity measure that has a 0.03 exceedance probability under the no-evidence case.

Figure 39 shows the PMs of ΔP, V_{f} and $R_{\max}$. The dashed curves are created by interpolating the discretized bins' PM to approximate the distributions' shapes.

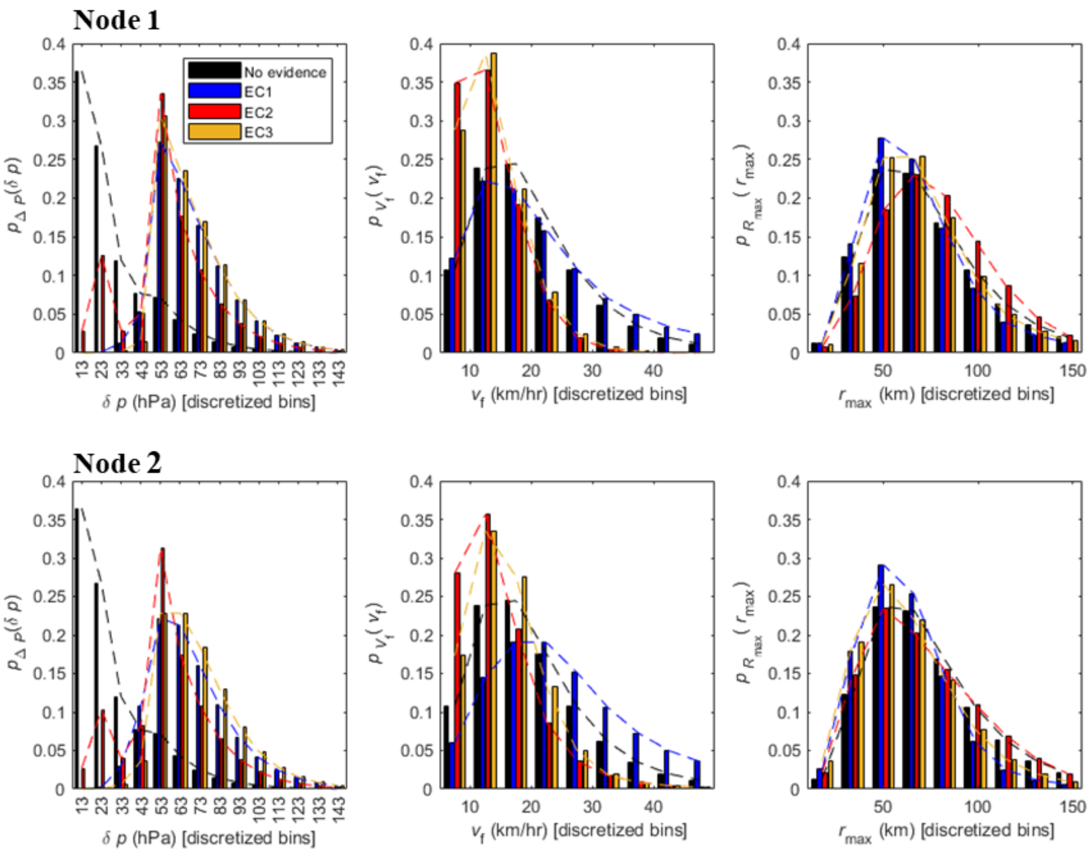


Figure 39: Distribution information of ΔP, V_{f} and $R_{\max}$ under EC1 to EC3 at Node 1 (row 1) and Node 2 (row 2)

EC1 (high storm surge) shifts the distribution of ΔP towards higher values, reflecting the higher likelihood of larger ΔP values to have produced high storm surge events. EC2 (high STR) is associated with a moderately bimodal distribution with lower likelihood of large ΔP values than EC1. However, EC3 (high storm surge and STR) results in a more substantial shift towards higher ΔP than observed under EC1 and EC2 individually.

EC1 (high storm surge) shifts the distribution of V_f towards higher values reflecting the higher likelihood of faster moving storms given high storm surge events. On the other hand, EC2 (high STR) shifts the distribution of V_f towards lower values, reflecting the higher likelihood of slower moving storms given high STR events. EC3 (high storm surge and STR) results in a moderated shift between the two evidence cases. The impact of evidence on the distribution of R_{max} is relatively small when compared with ΔP and V_{f}.

The effects of evidence on the distribution of X_0 and Θ are considered together because they reflect the TC track information. In each case, the J_1 node in the BN, as shown in Figure 34, is used to compute the joint PMs of all discretized TCPCs. The joint distribution of TC track parameter combinations (pair-wise joint PM of X_0 and Θ) for each case is computed by integrating the joint PMs computed by the J_1 node in the BN over ΔP, V_{f} and R_{max} . Figure 40 shows the pair-wise joint PM plot of each case. The no-evidence case is plotted for comparison and shows the equal likelihood assumption for X_0 and the marginal distribution shape for Θ.

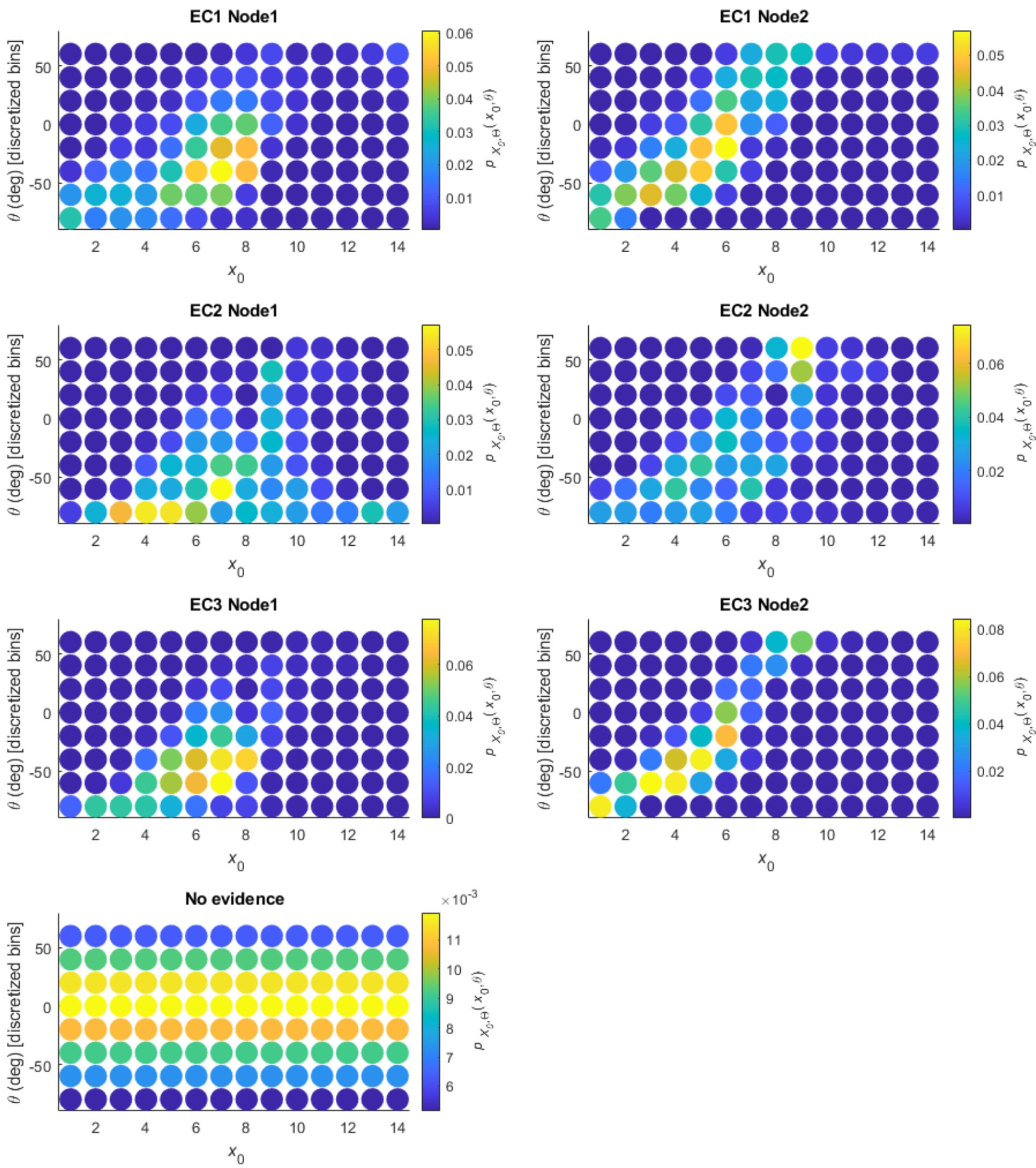


Figure 40: Pair-wise joint PM of X_0 and Θ under each case

From Figure 40, it can be observed that the input of evidence case notably affects the pair-wise joint PM of X_0 and Θ (compared with no-evidence case). Under EC1, high storm surge values at Node 1 lead to a distribution with mode associated with northwesterly tracks crossing nearly the midpoint of the X_0 reference line, while the distribution is a bit more dispersed when a high storm surge is observed as Node 2. This pattern is somewhat moderately reflected under EC2.

However, the TC tracks in EC3 show a more concentrated dominant area than EC1 and EC2, reflecting the smaller range of tracks given large values of storm surge and STR.

The contributions of ΔP,V_{f} and $R_{\max}$ along with the TC track variables are visualized in Figure 41 to Figure 43 as 3D stacked bar plots with partitions of ΔP, V_{f} and $R_{\max}$ identified with different colors.

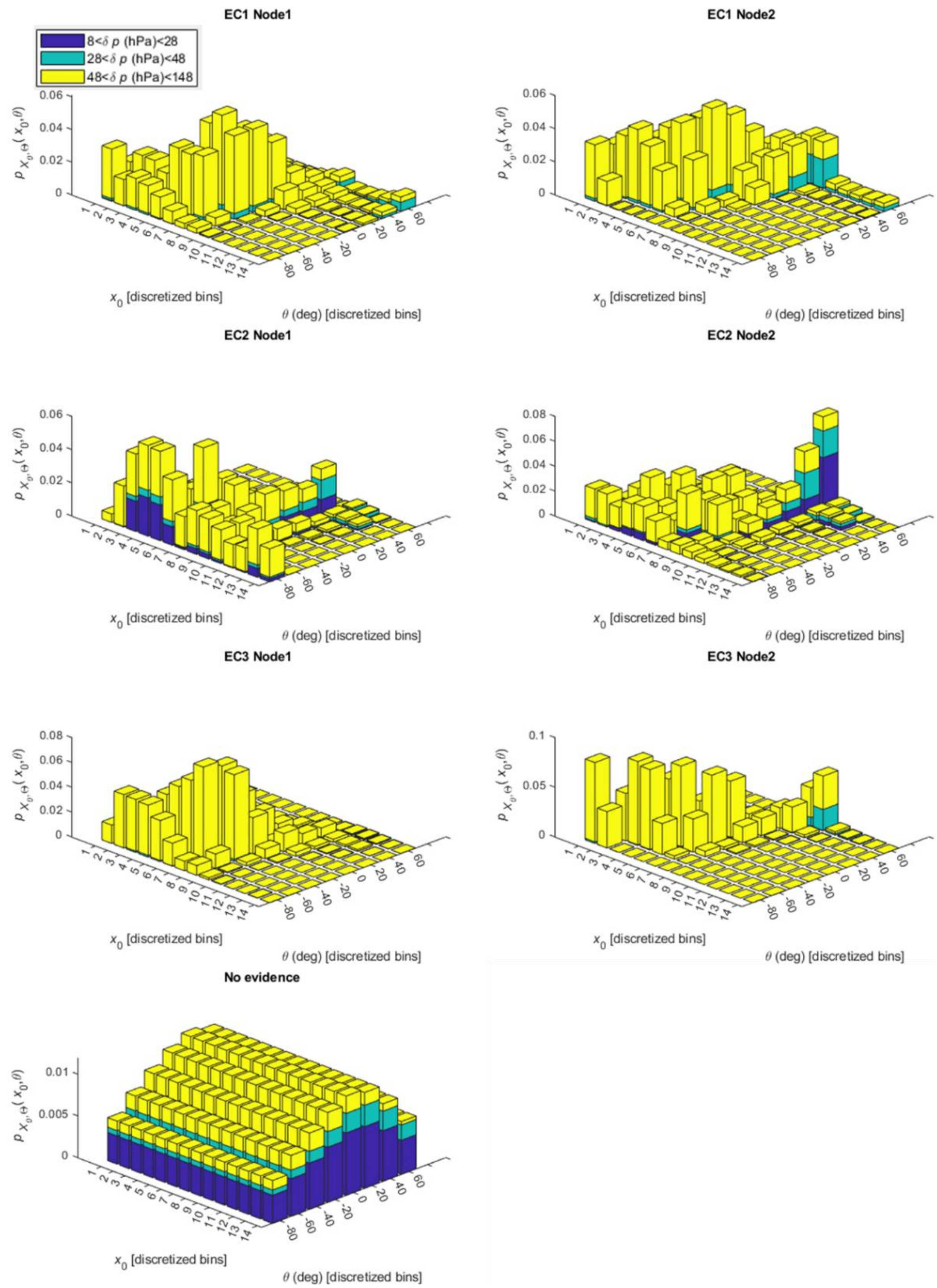


Figure 41: 3D stacked bar plot of TC track information with ΔP identified with colors

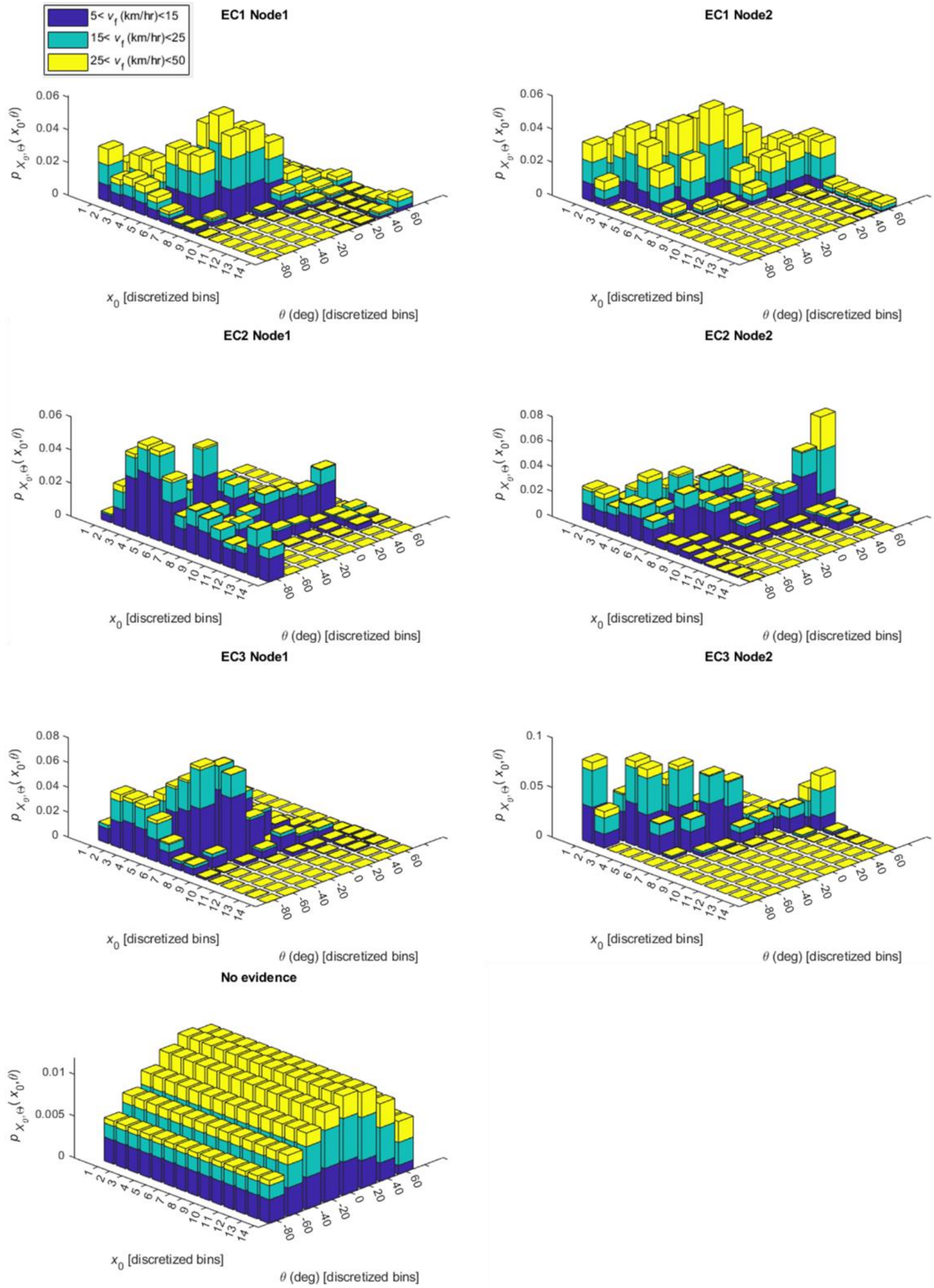


Figure 42: 3D stacked bar plot of TC track information with V_{f} identified with colors

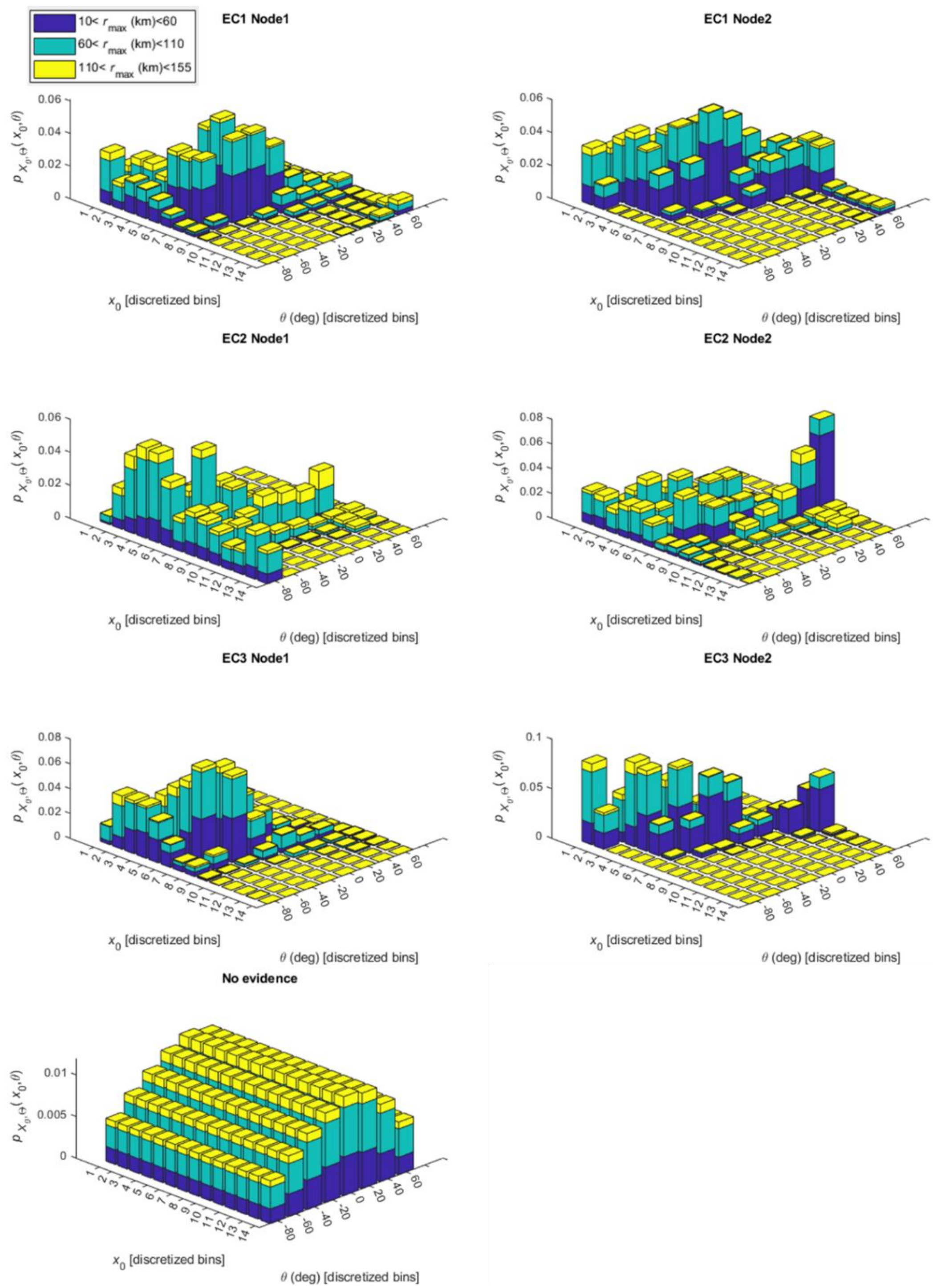


Figure 43: 3D stacked bar plot of TC track information with $R_{\max}$ identified with colors

From Figure 41, it can be observed that TCs with high ΔP (ΔP >48 hPa) dominant under most evidence cases, which is very different from the no-evidence case, which is dominated by lower Δp values. Depiste the general dominance of high Δp values across evidence cases, EC2 is associated with a higher likelihood of TCs with low ΔP (ΔP <48 hPa). Figure 42 shows that the dominant range of V_{f} can be notably different under evidence cases. The likelihood of faster storms (V_{f} >25 km/hr) under EC1 is greater than under EC2. Figure 43 shows that, for the inland node (Node 1), the contribution of smaller R_{max} (R_{max} <60 km) to dominant TC tracks is greater under EC1 than under EC2, however, this difference is not oberved in the coastal node (Node 2) for the two most dominant TC tracks with $\Theta \geq 40°$ (the two most right bars of EC2 Node2). Leveraging backward and forward propagation in the BN, information updating is performed to calculate the distribution of one coastal hazard given evidence associated with the occurrence of severe values of another coastal hazard. That is, the distributions of STR are computed under EC1 (high storm surge), and the distributions of storm surge are computed under EC2 (high STR). Figure 44 shows the resulting distribution. Note that the plotted range of variables in each figure is different.

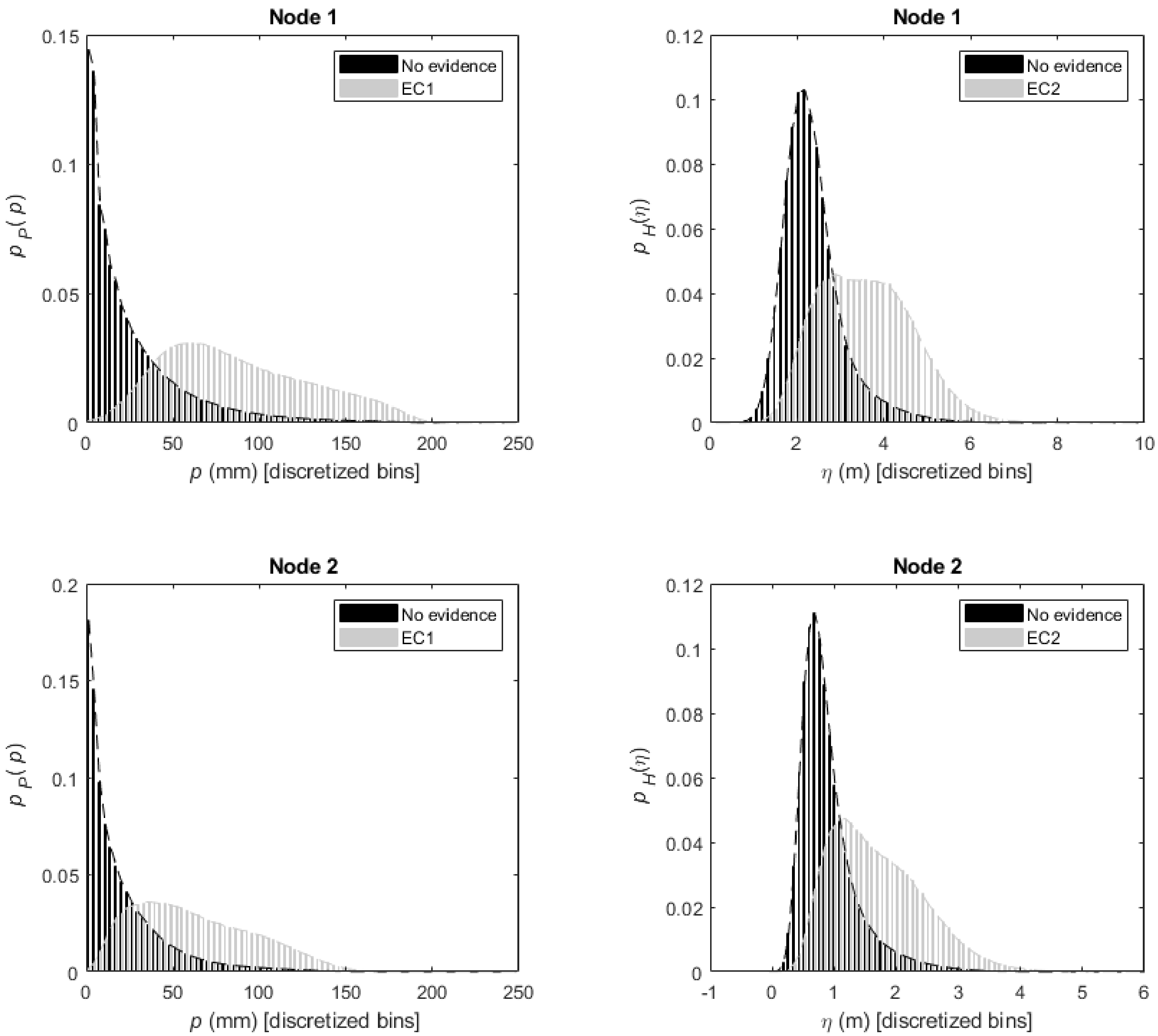


Figure 44: Distribution of coastal hazard with no evidence and evidence case of low exceedance probability event of another coastal hazard

It is observed that the peak PMs of storm surge and STR shift to the right under the evidence case of a low exceedance probability event of the other coastal hazard, which can be explained by the connection between intense storms and intense coastal hazards. Namely, EC1 (high storm surge) leads to a higher likelihood of high STR, while EC2 (high STR) leads to a higher likelihood of high storm surge at both nodes.

4.4 *Conclusion of Chapter 4*

This study has developed a framework for the deaggregation of low exceedance probability coastal hazards, which is based on a BN approach. To generate the BN, copulas are used to capture the dependence among TC atmospheric parameters, and surrogate models are used to capture the relationship between TC atmospheric parameters and coastal hazards. The joint PMs of TCPCs under different evidence cases are computed with the BN. The ranges of dominant TCPCs can be identified based on the joint PMs of TCPCs.

In the case study, an inland location and a coastal location near New Orleans, LA are selected as study nodes. A BN was constructed for this case study using the GeNIe software (BayesFusion 2023). The statistical analysis and JPM based on the USACE's CHS-PCHA framework are applied for the BN construction, where MGCs are used to characterize the dependence among TC atmospheric parameters. A GPR-based metamodel developed in the CHS-LA study (Nadal-Caraballo et al. 2022b) is used as a storm surge surrogate model. An ANN model is trained as an STR surrogate model. The statistical analysis results, MGCs, and surrogate models are applied to compute the CPTs of TC parameters nodes and coastal hazard nodes in the BN.

Leveraging the BN inference, the distributions of TC parameters are computed under evidence cases of low exceedance probability coastal hazards, providing insights regarding dominant TC parameter combinations, particularly relative to track information (landfall location and heading). Additionally, the dependence between coastal hazards was investigated to explore the impact of observations related to one hazard on the distribution of the other hazard. The distributions under evidence cases can be used to inform model refinement and augmentation activities, by focusing efforts on the events most likely to have consequential effects.

Chapter 5: A Multi-Tiered Bayesian Network Coastal Compound Flood Analysis Framework

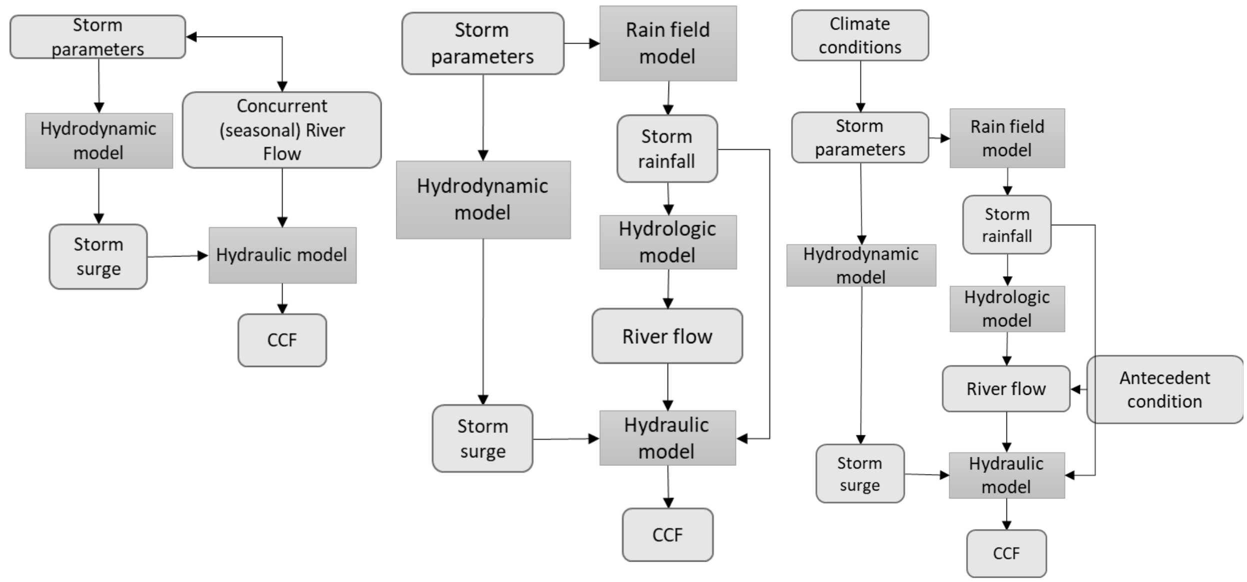
5.1 *Introduction*

Coastal areas are vulnerable to coastal compound floods (CCFs) that result from the combined effects of storm surges, tides, intense rainfall, and (in some cases) river flooding. CCFs can bring substantial losses to coastal communities, and these impacts are expected to increase as a result of climate change (Kirshen et al. 2008; NOAA 2022a; b). As a result, there is an increasing demand for accurate and efficient probabilistic analyses of CCFs to support risk assessments and decision-making processes.

Xu et al. (2022) summarized a literature review of the development of CCF studies over the past two decades. The literature review covered CCF mechanism analysis, research methodology, and probabilistic analysis. They found there is limited literature that integrates statistical and numerical models for CCF hazard analysis. Zhang and Najafi (2020) provided a literature review of probabilistic CCF analysis in conjunction with a probabilistic study on the southern coast of Saint Lucia in the Caribbean Sea. Their work demonstrates the potential of using simplified hydrodynamic models for CCF simulations. Bensi et al. (2020) provided a review of a range of studies addressing compound flooding hazards and identified several main classes of studies, including those that seek to characterize the joint distribution of multiple severity measures associated with CCFs (e.g., using parametric, non-parametric, and copula-based joint distribution) and apply Bayesian approaches.

In the past decade, the US Army Corps of Engineers (USACE) has undertaken a series of efforts to probabilistically analyze coastal hazards along coastal regions in the U.S. and has developed a Coastal Hazard Systems (CHS) Probabilistic Coastal Hazard Analysis (PCHA) framework (Melby et al. 2017; Nadal-Caraballo et al. 2015, 2020, 2022b). The USACE's CHS-PCHA framework has been augmented with a range of meta-modeling strategies to increase efficiency and enable large-scale studies. This study explores the use of Bayesian networks (BNs) to enable the extension of the CHS-PCHA framework to model CCFs. Three tiers of BN models are conceptualized and constructed. These three tiers of BN models have different complexity and are designed for cases with varying levels of data and computational resource availability.

The BN-based CCF modeling approach builds off the proof-of-concept BN presented in Mohammadi et al. (2023) and subsequent BN-based applications in Chapter 4. Mohammadi et al. (2023) built a BN for a coastal compound flood cases study in a site located on the Delaware River near Trenton, New Jersey, where storm surge, storm precipitation, river flow, and tide have been included in their model. In Chapter 4, a BN built based on USACE's CHS-PCHA has been implemented to develop a deaggregation method for multiple coastal hazards without addressing the interaction of coastal hazards.



(a) (b) (c)

Figure 45: Conceptions of three tiers of BN models: (a) Tier-1 BN; (b) Tier-2 BN; (c) Tier-3 BN

Figure 30 presents a conceptual illustration of the three tiers of BN models. In the Tier-1 BN model (Figure 30 (a)), which reflects the approach used in Nadal-Caraballo et al. (2022), storm surge and river flow are treated as the dominant CCF drivers. The dependence between storm events and river flow is modeled through seasonality during hurricane seasons. In the Tier-2 BN model conception (Figure 30 (b)), storm rainfall is added as a contributor to CCF. In the Tier-3 BN conception (Figure 30 (c)), climate conditions and antecedent river conditions are added to build a more comprehensive CCF model.

To illustrate the proposed framework, these three tiers of BN are implemented using a case study in the region of New Orleans, LA, and illustrative results are presented for two study nodes illustrated in Figure 33 (a). Mississippi River (MR) is treated as the source of river flow contributor of CCF in the study region. While the case study is used to illustrate the framework, the multi-tiered BN approach is intended to be more generally valid.

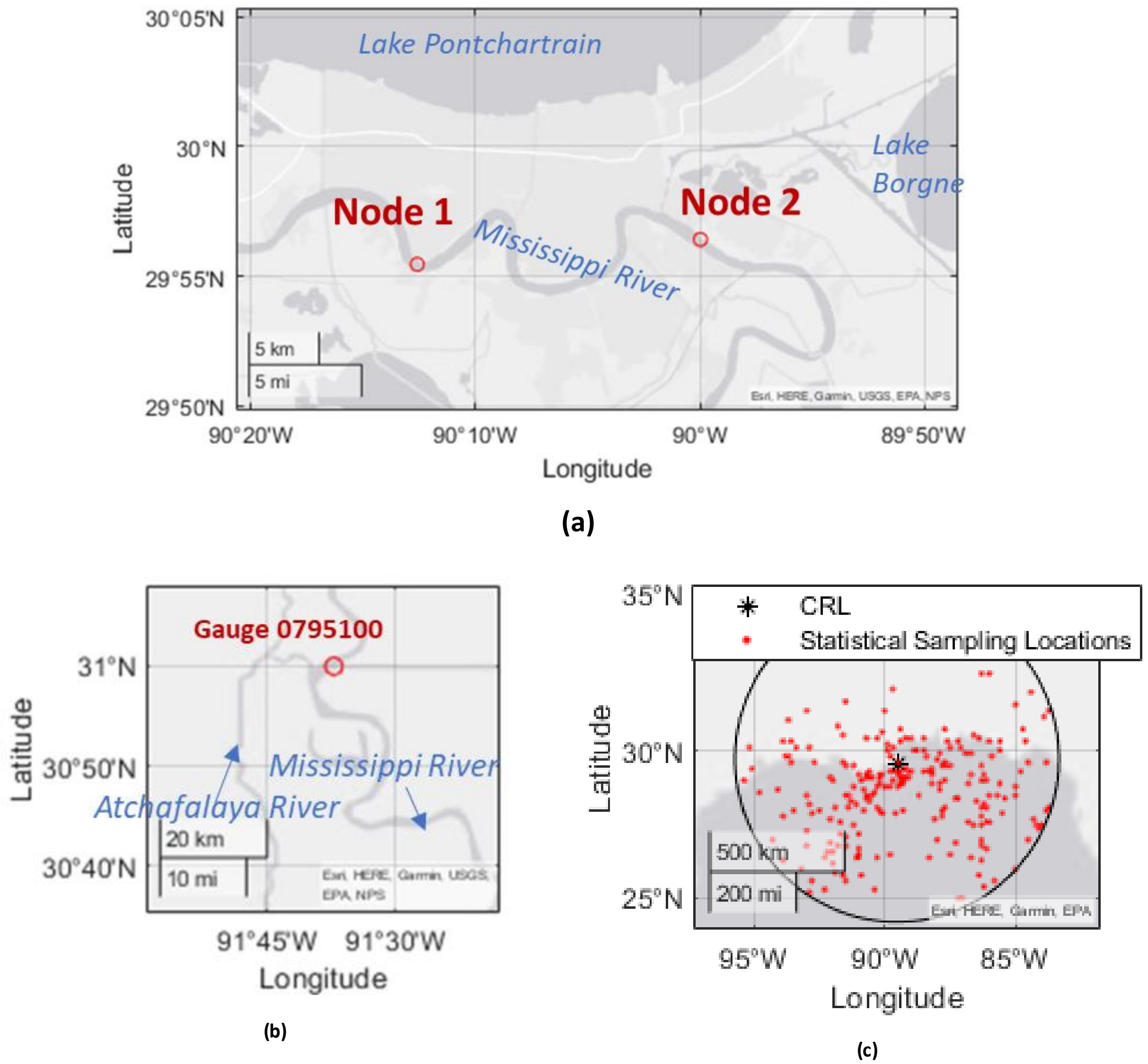


Figure 46: (a)Locations of study nodes; (b) Location of USGS gauge 07295100; (c) Center of statistical analysis study region, locations of sample data, and capture zone (the black circle)

5.2 Tier-1 Bayesian Network Construction

A Tier-1 BN model is constructed for the case study region leveraging the modeling approach first employed in the USACE's CHS-LA (Coastal Hazards System–Louisiana) study (Nadal-Caraballo et al. 2022b). Figure 47 presents this Tier-1 BN, and Table 9 lists the physical definition of each node. The discretization information of each node is provided in Appendix E.

In this Tier-1 BN, seasonality is used to capture the dependence between TC atmospheric parameters and concurrent river flow. Meteorological seasonality can have a significant impact on river discharge. Typically, seasonal events like snowmelt and rainfall will increase river discharge. Meanwhile, varying temperatures and moisture affect the formation of storms (Gray 1968; Palmen 1948). Michele and Salvadori (2002) use river seasonality to capture flood antecedent conditions. Walega and Młyński (2017) evaluated the seasonality of median monthly discharge in selected Carpathian rivers in Poland.

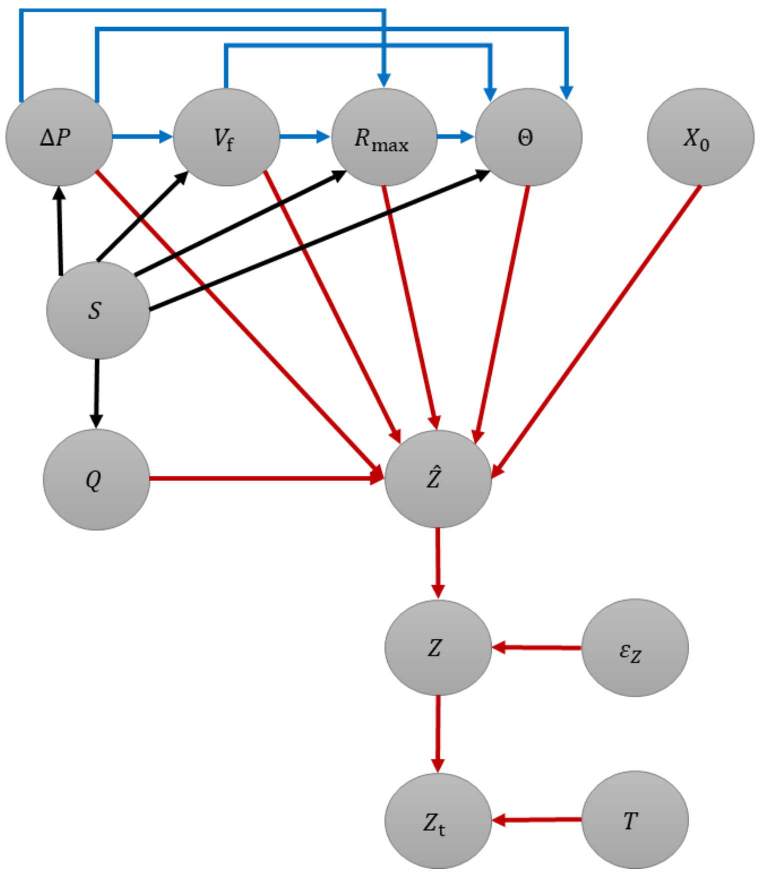


Figure 47: Tier-1 BN

Table 10 physical definition of each node in Tier-1 BN

Notation	Physical definition	Notation	Physical meaning

S:	Time period within hurricane season	Q:	MR river flow (kcfs)
ΔP:	Central pressure deficit (hPa)	$\hat{Z}$:	Water surface level (WSL) resulting from storm surge and river flow predicted by model (m)
V_{f}:	Forward velocity (km/hr)	Z:	WSL resulting from storm surge and river flow with model error considered (m)
$R_{\max}$:	Radius to maximum winds (km)	ε_Z:	Normalized residuals of WSL
Θ:	Storm heading direction (deg)	T:	Normalized residuals of astronomical tide
X_0:	Storm landfall location index	Z_{t}:	WSL with astronomical tide considered (m)

This Tier-1 BN consists of three primary models:

- **Seasonality-storm-river model** (as depicted by black links in Figure 47): The node S, representing the time period within the hurricane season, is used to capture the dependence between TC atmospheric parameters (nodes ΔP, V_{f}, $R_{\max}$, and Θ) and concurrent river flow (node Q), as reflected by the links from the node S to its children.
- **Tier-1 joint probability model for TC atmospheric parameters** (as depicted by blue links in Figure 47): The dependence between TC parameters (for a given season) is reflected by the series of links connecting them.
- **ADCIRC-derived water surface level model** (as depicted by red links in Figure 47): The estimated CCF water surface level (WSL) resulting from storm surge and river flow is represented by node $\hat{Z}$, which is a child of nodes ΔP, V_{f}, $R_{\max}$, Θ and X_0. The associated actual CCF WSL is a function of the estimated value and an error term, as

reflected by node Z with parent nodes $\hat{Z}$ and ε_Z. Finally, the effects of astronomical tide are included through the addition of a tidal residual, resulting in a final WSL, as reflected by node Z_{t} with parent nodes Z and T.

Additional details regarding the generation of CPTs associated with each of the above models are provided in the subsections that follow.

5.2.1 Seasonality-storm-river model

The seasonality node S is discretized into time periods within the hurricane season. While multiple discretization strategies can be used, the illustrative case study application uses two time periods: Season 1 (May 1 to August 31) and Season 2 (September 1 to November 30). The dividing point between meteorological summer and meteorological fall (i.e., Sept 1st) is subjectively chosen as the dividing point between the two seasons. A weighting strategy is used to mitigate the effect of the hard breakpoint when using time-series data. For each season, its probability is assigned based on the portion of observed storms that initiate inside that season. CPTs of TC atmospheric parameters nodes (ΔP, V_{f}, R_{max}, and Θ) are generated based on statistical analysis of observational data. A historical storm dataset combining Hurricane Database 2 (HURDAT2) (Landsea and Franklin 2013; NOAA 2022c) and Tropical Cyclone Extended Best Track Dataset (EBTRK) database (CIRA 2021; Demuth et al. 2006) with data imputation (Liu et al. 2023b) is used to generate statistical sampling data using the procedure of (Chouinard and Liu 1997; Nadal-Caraballo et al. 2015). A coastal reference location (CRL) with latitude = 29.576 and longitude = -89.538 is selected for statistical sampling. Figure 30 (c) plots the location of the CRL and its 600 km radius capture zone and statistical sampling locations.

Statistical sampling data is partitioned for each season and used to perform a seasonally-dependent joint distribution analysis using copulas, as described further in Section 2.2.

The daily discharge data of the United States Geological Survey (USGS) gauge 07295100 is used to compute the CPT of the river (MR) node Q. Consistent with Nadal-Caraballo et al. (2022), it is assumed that there is a 70% and 30% split in discharge directed through MR and Atchafalaya Rivers, respectively, at the USGS gauge 07295100 at Red River Landing (Tarbert Landing). The location of the gauge is shown in Figure 30 (b). Gauge records from 1976 to 2019 and inside the window of hurricane season (May 1-November 30) are used, and the maximum MR daily discharge is regulated to 35396 m^3/s (1250 kcfs). Figure 48 plots the temporal distribution of MR daily discharge data with the red line depicting the subjectively chosen meteorological breakpoint between seasons.

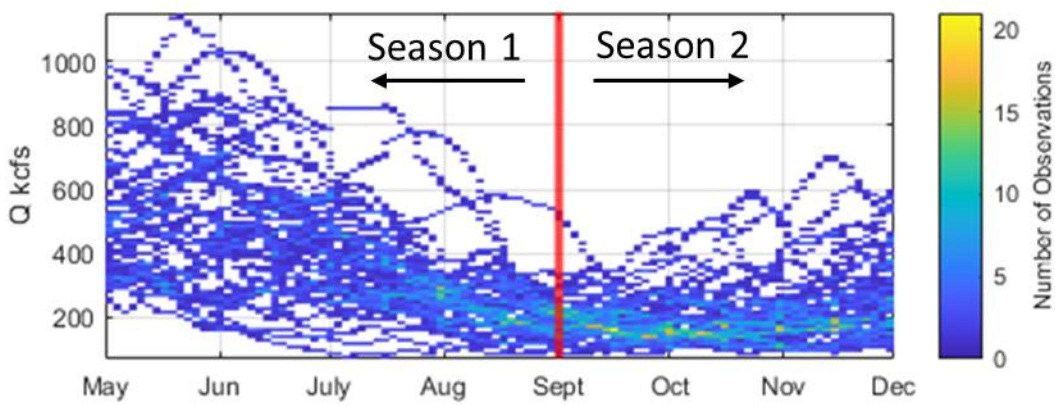


Figure 48: Temporal distribution of sample data of MR daily discharge with red line depicting the dividing point between Season 1 and Season 2

A Gaussian kernel function (GKF) based temporal distance weighting method is used here to smooth out the effect of the season division in the time series. To generate the statistical sample data for a season, each record inside hurricane season (i.e., recordes within the window of May

1to November 30) is assigned a temporal distance d_i to the boundaries of the corresponding season S_k . The distance is calculated as:

$$d_i = \begin{cases} S_k^- - t_i, & t_i < S_k^- \\ 0, & S_k^- \leq t_i \leq S_k^+ \\ t_i - S_k^+, & S_k^+ \leq t_i \end{cases} \quad (27)$$

Where S_k^- is the date of the first day of the season S_k, S_k^+ is the last day of season S_k, t_i is the date of the ith record. The calculation of Equation (27) uses the integer of day.

For each record, a temporal distance weight is calculated using a GKF:

$$w_t(d_i) = \frac{1}{\sqrt{2\pi}h_d} \exp\left[-\frac{1}{2}\left(\frac{d_i}{h_d}\right)^2\right] \quad (28)$$

Where $w_t(d_i)$ is the temporal distance weight of the ith record, h_d is the kernel bandwidth, and an $h_d = 15$ days is selected in this study.

Using this method, an identical highest weight is assigned for each data point within the seasonal time window, and the assigned weight decreases as the distance between the data and the time window increases. This is demonstrated in Figure 49.

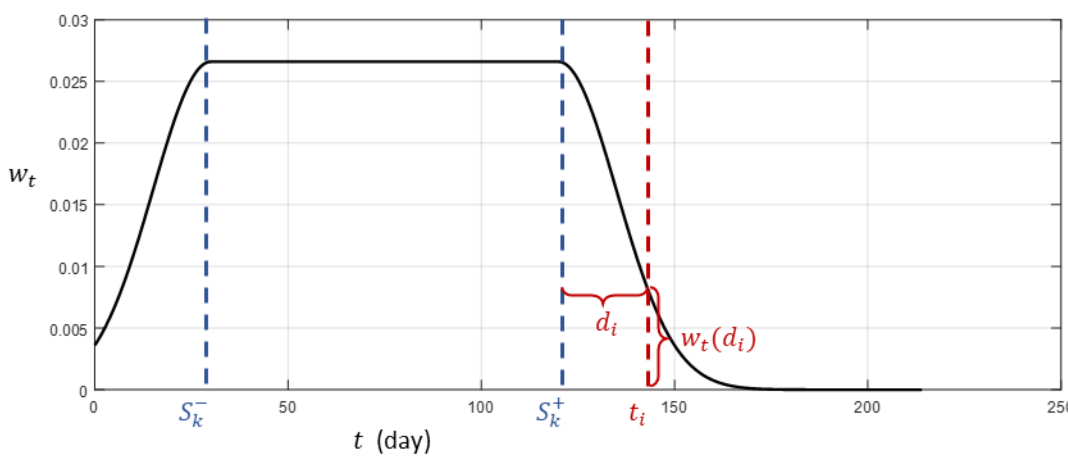


Figure 49: Assigned weight along the time axis

After calculating the temporal distance weight using the method described above, a distance weight adjustment method (Chouinard and Liu 1997; Nadal-Caraballo et al. 2015) is applied to all records within hurricane season. To mitigate the effect from the autocorrelation in the time series data, processed data are randomly selected and used as statistical samples for MR daily discharge statistical analysis in each season.

Figure 50 plots the fitted marginal PDF curves with the corresponding sampling data of MR daily discharge in each month. The generalized extreme value distributions (MathWorks 2022) are used to fit marginal distributions of MR daily discharge in each season.

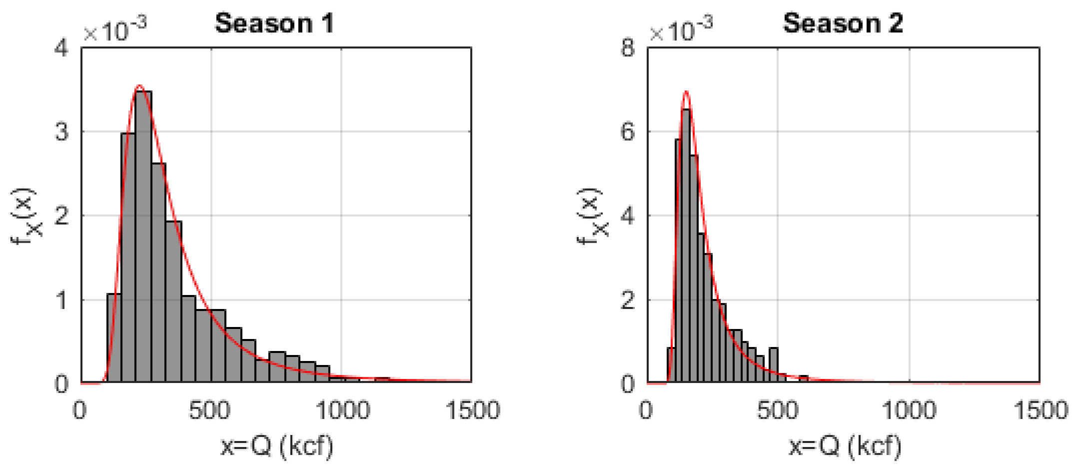


Figure 50: Fitted marginal PDF curves with the corresponding statistical sampling data of MR daily discharge; Note scales of y-axises are different

5.2.2 Tier-1 Joint probability model for TC parameters

Leveraging the work of Nadal-Caraballo et al. (2020, 2022) and Liu et al. (2023a), an MGC is used to construct the joint probability model of TC atmospheric parameters for each season. In this Tier-1 BN, the dependence among TC atmospheric parameters is characterized by one MGC in each season, with the marginal distributions and parameters (i.e., the correlation matrix) estimated using each season's TC atmospheric parameter statistical sampling data. Figure 51 plots the fitted marginal PDF curves with the corresponding statistical sampling data of TC atmospheric parameters in each season. The marginal distribution model (distribution functional form) for the TC atmospheric parameters used in this study is consistent with Chapter 4. The conditional probability table of each TC parameter node is computed using the same method as described by Chapter 4.

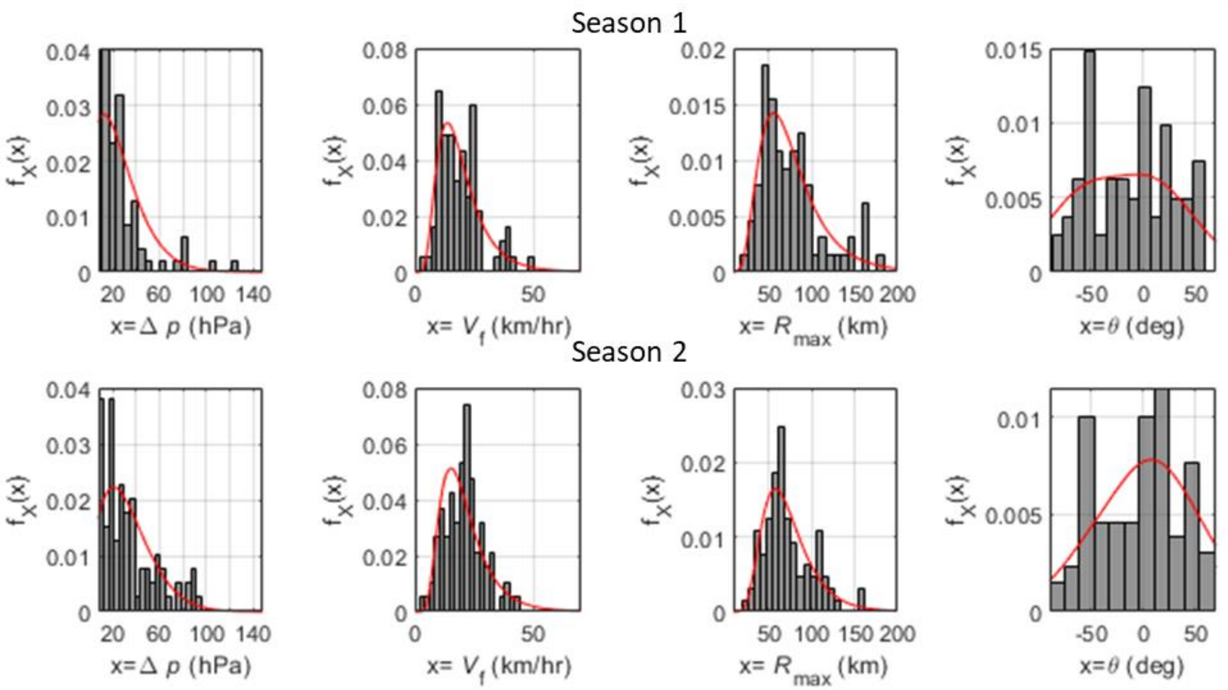


Figure 51: Fitted marginal PDF curves with the corresponding statistical sampling data of TC atmospheric parameters; Note scales of y-axises are different

5.2.3 ADCIRC-derived water surface level model

To compute the conditional probability table (CPT) of node $\hat{Z}$ in the Tier-1 BN, a CCF WSL model with storm surge and river flow involved is needed. In the past decades, different models have been applied to coastal hazard analysis. A series of USACE studies leveraged ADCIRC and wave models for storm surge simulations (Gonzalez et al. 2019; Nadal-Caraballo et al. 2015, 2020, 2022b). Meanwhile, surrogate models based on machine learning have been well-studied for efficiently predicting storm surges based on ADCIRC-derived training data. For example, Jia and Taflanidis (Jia and Taflanidis 2013) implemented a Gaussian process regression (GPR) model for storm surge prediction with a series of follow-up studies expanding and augmenting the use of GPR (Jia et al. 2016; Taflanidis et al. 2014; Zhang et al. 2018) or leveraging or comparing alternate approaches (Al Kajbaf and Bensi 2020; Lee et al. 2021). Nadal-Caraballo et al. (Nadal-Caraballo et al. 2020) added GPR-based metamodels (GPM) to the USACE's CHS-PCHA framework for WSL prediction.

A GPR-based metamodels (GPM) WSL surrogate model developed in USACE's CHS-LA study (Nadal-Caraballo et al. 2022b) is used here to predict WSL (i.e., to develop the CPT for node $\hat{Z}$ given each combination of TC parameters) considering storm surge and MR river flow. The surrogate model training data is derived from ADCIRC simulations of 1145 synthetic storm events that cover multiple MR river flow scenarios. The resulting surrogate model can be used to predict peak WSL using TC parameters and MR river flow as input. The uncertainty information of this WSL model is represented by the normalized residual node ε_Z, which carries combined uncertainty accounting for both the relative and absolute uncertainty of modeled value (Nadal-Caraballo et al. 2022b). In this study, all normalized residual nodes carry the combined uncertainties. The actual WSL is represented by the node Z, which is defined via the superposition of estimated WSL and an error term defined as a function of the normalized residual: i.e., $Z = \hat{Z} + f(\varepsilon_Z)$.

Monte Carlo simulation (MCS) is implemented for the computation of the CPTs of nodes $\hat{Z}$ and Z for the purpose of reducing the error caused by discretization Mohammadi et al. (2023).

Applying this MCS process to generate the CPT for the node $\hat{Z}$, a large number of TC parameter values are generated for each combination of TC parameters' discretized bins and used to define the corresponding conditional distribution of Z. For the CPT of node Z, a large number of normalized residual values corresponding to each state of ε_Z are generated to computing error term and superimposed with simulated values from the discretized states of the node $\hat{Z}$. Similar MCS strategy is applied to other parts of BNs in this study to reduce the error induced by discretization. A graphical explanation of this MCS in BN can be found in Mohammadi et al. (2023).

In the node Z_{t}, the tides component is accounted for by superimposing WSL and random generalized astronomical tides. The standard deviation of the astronomical tide computed in the CHS-LA study (Nadal-Caraballo et al. 2022b) is used. The CPT of tide residual node T is set based on a discretized standard normal distribution.

5.3 Tier-2 Bayesian Network Construction

In this section, a Tier-2 BN model is constructed for the case study region. In this Tier-2 BN model, the river flow seasonality is preserved to capture base flow, while a TC intensity is used to categorize TCs into different intensities. Storm rainfall is involved in the relationship between TC parameters and WSL. A surrogate model derived from a regional HEC-RAS model is used to simulate the resultant WSL (Z) under storm surge, storm rainfall, and MR river flow. Figure 52 presents this Tier-2 BN, and Table 11 lists the physical definition of each node of this Tier-2 BN.

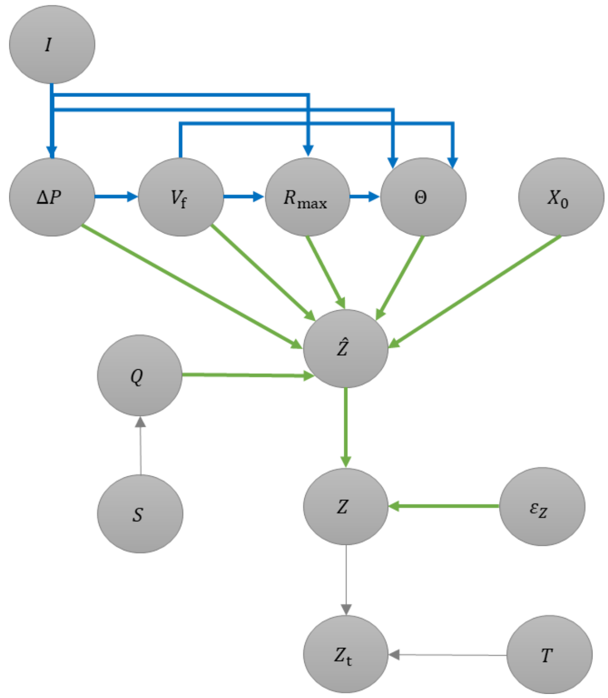


Figure 52: Tier-2 BN

Table 11 physical definition of each node in Tier-2 BN

Notation	Physical definition	Notation	Physical meaning
I:	Intensity category of TC	Z:	WSL resultant from storm surge, storm rainfall and river flow with model error considered (m)
$\hat{Z}$:	WSL resultant from storm surge, storm rainfall and river flow predicted by model (m)		

Note: Nodes not included have the same physical definition as Tier-1 BN (see Table 9).

The Tier-2 BN contains three models:

- Tier-2 joint probability model for TC parameters (as depicted by blue links in Figure 52): The dependence between TC parameters nodes ΔP, V_f, R_{max}, and Θ (for a given TC intensity category) is reflected by the links connecting them.

- HEC-RAS-derived water surface level model (as depicted by green links in Figure 52): The estimated CCF WSL resulting from storm surge, storm rainfall, and river flow is represented by node $\hat{Z}$, which is a child of nodes ΔP, V_{f}, $R_{\max}$, Θ,X_0, and Q.
- Storm rainfall model: A storm rainfall model is used to simulate TC-induced rainfall. The rainfall model is not directly reflected in the BN. Instead, it is used to incorporate storm rainfall of the HEC-RAS-derived WSL model as a function of the TC parameters.

Additional details regarding the generation of CPTs associated with each of the above models are provided in the subsections that follow.

5.3.1 Tier-2 Joint probability model for TC atmospheric parameters

In this Tier-2 BN, the statistical sampling data of TC atmospheric parameters are partitioned based on the value of ΔP into three categories: low intensity (LI, $28\ \mathrm{hPa} > \Delta P \geq 8\ \mathrm{hPa}$), moderate intensity (MI, $48\ \mathrm{hPa} > \Delta P \geq 28\ \mathrm{hPa}$), and high intensity (HI, $148\ \mathrm{hPa} > \Delta P \geq 48\ \mathrm{hPa}$). For each TC intensity category, a separate set of marginal distributions is fitted to the intensity-partitioned statistical sampling data of ΔP, V_{f}, $R_{\max}$ and Θ, and a separate MGC is fitted using the general procedure described in Chapter 4.

5.3.2 Storm rainfall model

To incorporate storm rainfall in this Tier-2 BN, a storm rainfall model is used to generate storm rainfall under synthetic storm events (i.e., to estimate TC rainfall for each combination of TC parameters). In this study, the storm rainfall for synthetic storms is simulated using a physical-based tropical cyclone rainfall model (TCR) (Lu et al. 2018; Xi et al. 2020; Zhu et al. 2013). The computed rainfall time series is later used as rainfall input to HEC-RAS for CCF modeling, as described below.

5.3.3 HEC-RAS-derived water surface level model

To compute the CPT of the node $\hat{Z}$ in the Tier-2 BN, a CCF WSL model with TC rainfall involved is needed. HEC-RAS (the USACE Hydrologic Engineering Center's River Analysis System software (USACE 2023)) is a widely used software application that can perform one and two-dimensional hydraulic calculations with rainfall involved. In this Tier-2 BN model construction, a limited-scope HEC-RAS 2D regional model is used to simulate CCFs under storm surge (modeled as a time-series boundary), storm rainfall (time-series gridded precipitation), and MR river flow (baseflow at the upstream boundary and storm rainfall-induced runoff). Figure 53 presents this limited-scope HEC-RAS 2D regional model. Then, a CCF surrogate model is trained using the HEC-RAS-derived data. The resulting model combination provides a "mapping" between TC parameters, base flow, and WSL, which allows the construction of the CPT for the node $\hat{Z}$.

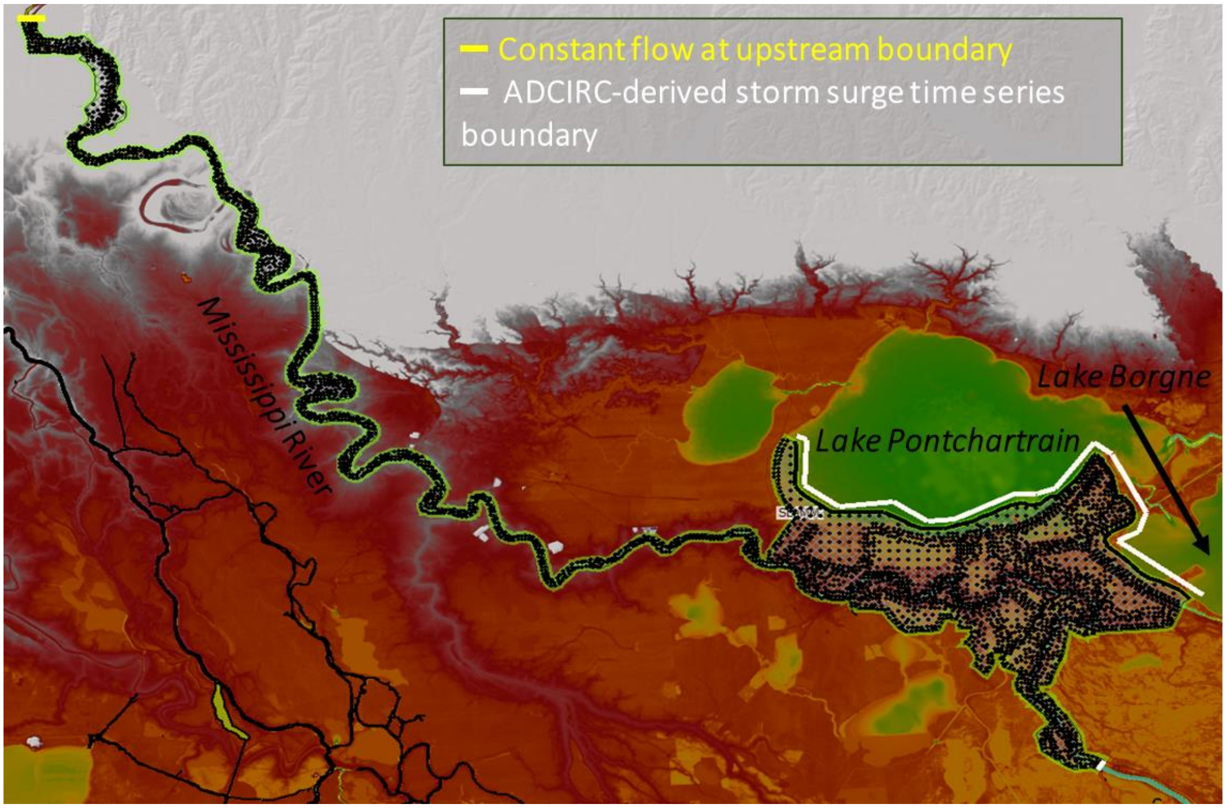


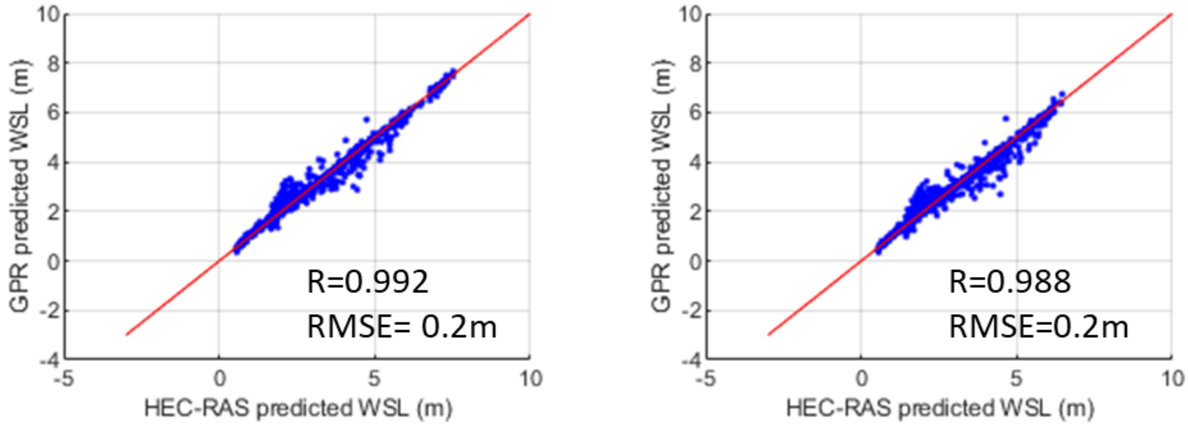
Figure 53: HEC-RAS 2D regional model; White curves depict boundary condition lines

This limited-scope HEC-RAS 2D regional model covers the MR reach starting from around [30.96, -91.66] and ending at a downstream site around [29.63, -89.93]. It includes the region of Hurricane & Storm Damage Risk Reduction System (HSDRRS) (IPET 2009). The perimeter of this model is set based on the contour of the watershed with USGS Hydrologic Unit Code 080701 ("Lower Mississippi-Baton Rouge") and levees of the HSDRRS.

The HEC-RAS software allows users to add hydraulic structure models into a 2D flow area to model the hydraulic mechanics of different types of hydraulic structures, which will affect the WSL distribution inside a 2D flow area, especially for an urban area with comprehensive flood risk reduction system like the HSDRRS region. In this HEC-RAS 2D regional model, a series of pump stations, flood gates, and barriers are involved inside the area of HSDRRS.

As shown in Figure 53, constant flow is used as the upstream boundary condition of MR, and ADCIRC-derived storm surge time series are used as a downstream boundary condition for the MR and as boundaries along the coastal line of Lake Pontchartrain and Lake Borgne. TCR model-derived grided rainfall time series, as mentioned in Section 3.2, is used to cover the whole model area to facilitate storm rainfall input.

A synthetic TC suite adapted/interpolated from the CHS-LA study (Nadal-Caraballo et al. 2022b) is used to generate the aforementioned ADCIRC-derived storm surge time series and TCR model-derived grided rainfall time series. A MATLAB program adapted from Leon and Goodell (2016) is developed to automatically input boundary conditions and collect the result (e.g., WSL) of each synthetic TC event, considering the combined inputs from storm surge and rainfall. This provided a set of synthetic results that "maps" TC parameters predictors (ΔP, V_{f}, R_{max}, Θ, lat, and lon) to peak CCF WSL response. Then, given the the demonstrated strong performance of GPR prediction models (e.g., Jia and Taflanidis 2013; Jia et al. 2016; Taflanidis et al. 2014; Zhang et al. 2018), a GPR surrogate model was trained using the simulation results and used to efficiently compute the CPT of node $\hat{Z}$ in conjunction with the MCS approach described above. Figure 54 shows the 10-fold cross-validation performance of this surrogate model, which demonstrates sufficient accuracy for the purposes of this illustrative study.



(a) (b)

Figure 54: 10-fold cross-validation performance of the surrogate model. (a) Node 1; (b) Node 2

5.4 *Tier-3 Bayesian Network Construction*

In the Tier-3 BN, climate conditions and flood antecedent conditions are added to build a more comprehensive CCF BN. In this study, the theory of Kang and Elsner (2015) is used to guide assumptions regarding the impact of climate change on TC activity. As a result, it is assumed that, under sea surface temperature (SST) rise, the ratio of HI TC to all intensity TCs increased, but the SRR of all intensity TCs decreased. However, the BN is sufficiently flexible to accommodate alternate approaches. In addition, the Tier-3 BN incorporates expanded models for antecedent conditions, namely seasonality of the MR and antecedent precipitation in the MR basin. Figure 55 presents the Tier-3 BN constructed in this study, and Table 12 lists the physical definition of each node inside the Tier-3 BN.

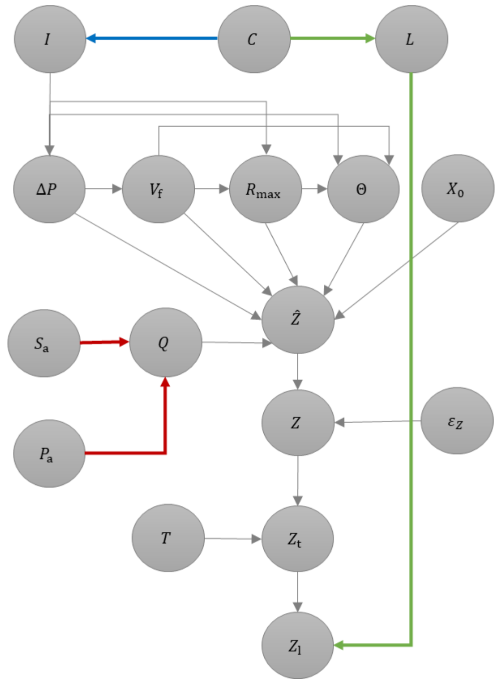


Figure 55: Tier-3 BN

Table 12 physical meaning of each node in Tier-3 BN

Notation	Physical definition	Notation	Physical meaning
C:	Global averaged SST rise relative to the year of 2020 (°C)	$\hat{Z}$:	WSL resultant from storm surge, storm rainfall and river flow predicted by model (m)
L:	Sea level rise (SLR) (m)	Z:	WSL resultant from storm surge, storm rainfall and river flow with model error considered (m)
S_a:	Antecedent river season	Z_l:	WSL with SLR involved (m)
P_a	Antecedent river basin precipitation		

Note: Nodes not included have the same physical definition as Tier-1 BN and Tier-2 BN (See Table 9 and Table 11).

The Tier-3 BN contains three models:

- **SST-tropical cyclone model** (as depicted by the blue link in Figure 55): Potential SST rise scenarios are captured by node C, which is set as the a parent node of TC intensity (node I) to account for the impact of climate change on TC activity.
- **Antecedent moisture conditions-river models** (as depicted by red links in Figure 55): Antecedent condition nodes (S_{a} and P_{a}) are set as parents of river flow (node Q) to account for the impact of antecedent conditions on river flow.
- **SLR model** (as depicted by green links in Figure 55): Potential SST rise scenarios are captured by node C, which is set as a parent node of SLR node (L) reflecting the impact of climate on sea-level. The final CCF WSL node (Z_{l}) is a child of nodel L to capture impact of SLR on WSL.

In addition, the Tier-3 BN leverages models described in the Tier-2 BN. These models are reflected by the grey links in Figure 55. Additional details regarding the generation of CPTs associated with each of the above models are provided in the subsections that follow.

5.4.1 Sea surface temperature-tropical cyclone model

A series of studies have investigated the impact of climate condition change on TC activity. Webster et al. (2005) found that, during 1970~2004, an increase in the number of HI TCs was observed with a decrease in the number of all TCs, except the case of the North Atlantic Ocean during 1995~2004. Kang and Elsner (2015) concluded that the TC intensity increased along with a decrease in TC frequency during 1985~2014. Wu et al. (2022) performed a review of current understanding of the effect of climate change on basin-wide TC intensity and concluded that the degree of climate change to TC is unclear, and it is also unclear when this impact will be detectable. Knutson et al. (2019) assessed whether the detectable change in TC activity is identifiable and focused on the possible detection error. Roberts et al. (2020) simulated the impact of climate change on TC properties with multiple models and made comparisons. Gutmann et al. (2018) performed a high-resolution regional climate modeling experiment with the Weather Research and Forecasting model and found that storm rainfall increased in the study region under climate change. While there is currently a lack of well-established models that can predict the characteristics of TCs under future climate change, the goal of the Tier-3 BN is to facilitate the inclusion of climate change considerations.

Based on the findings of Kang and Elsner (2015), this case study implements the assumption that climate change will cause a decrease in the storm recurrence rate (SRR) of TC in the North Atlantic but an increase in their intensity. Considering the lack of an elaborated SST-TC model and uncertainty regarding what degree climate change will impact TC (Wu et al. 2022), only a small range of SST rise is considered. In the Tier-3 BN, a climate condition node (C) is used to represent the rise in SST. A simplified SST-TC intensity relation is used in this case study. It assumed that as the SST increases, the SRR decreases linearly. Furthermore, a trade-off is assumed to exist between high-intensity (HI) and low-intensity (LI) cyclones, i.e., with the increase of SST, the proportion of HI TC increases while the proportion of LI TC decreases. The simplified SST-TC relation is used to build the CPT of the intensity node I in the Tier-3 BN.

5.4.2 Antecedent moisture conditions-river models

Previous studies have investigated the impact of flood antecedent conditions on flood events. For example, Michele and Salvadori (2002) present a peak flood analytical formulation that involves antecedent moisture conditions based on river season and antecedent rainfall. Bennett et al.

(2018) investigate the effects of antecedent precipitation on flood volume. Grillakis et al. (2016) investigate the importance of the initial soil moisture state for flash flood magnitudes. In this case study, the river seasonality and antecedent precipitation are combined to build a flood antecedent moisture (AMC) condition model to enable increased refinement in the modeling of antecedent conditions.

Michele and Salvadori (2002) identified antecedent precipitation and river season as the two factors that contribute to river AMC. In this case study, two nodes are used to represent AMC of the MR, i.e., antecedent river season (S_{a}) and antecedent river basin precipitation (P_{a}). The node S_{a} is defined in the same way that S node is defined in Tier-1 BN. For node P_{a}, the antecedent river basin precipitation value is calculated as the basin-averaged accumulated rainfall for the prior five days. The “PRISM Climate Data” (PRISM 2023) is used as a database to inform the development of the marginal distribution of this node. Specifically, five-day accumulated basin-averaged rainfall from 1980~2015 are randomly selected to generate independent statistical samples for P_{a} and a GEV distribution (MathWorks 2022) is used for fitting the marginal distribution. In this case study, a linear regression rainfall-runoff model derived from HEC-RAS rainfall-runoff simulations is used to characterize the relationship between P_{a} and Q. This linear regression rainfall-runoff model is plotted in Figure 56.

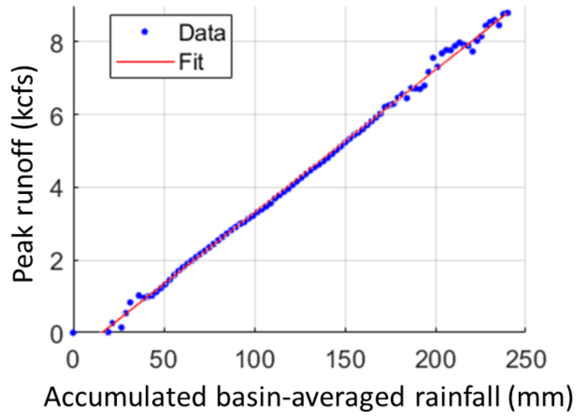


Figure 56: Linear regression rainfall-runoff model

5.4.3 Sea level rise model

In this case study, a node representing the SLR (node L) is added as a child on node C representing climate conditions. A cubic SLR function from Roper (2016) is used to characterize the relationship between SST and SLR and define the CPT between node C and node L. Kyprioti et al. (2021) investigated the performance of incorporating SLR scenarios into storm surge surrogate models and concluded a series of strategies to develop a high-accuracy storm surge surrogate model incorporating SLR. In the North Atlantic Coast Comprehensive Study (NACCS) (Nadal-Caraballo et al. 2015), the linear superposition method is used to incorporate SLR into the water level results due to the complexity and extent of the hydrodynamic model. It is noted that the NACCS discusses the potential error induced by linear superposition. In this study, for simplicity, SLRs are superimposed on WSL (i.e., $Z_l = Z_t + L$) as a means of generating the CPT for node Z_l.

5.5 *Results*

As illustrated in Section 2 ~ Section 4, three tiers of BN were constructed for a case study in New Orleans, LA. Leveraging these three BNs, probabilistic analyses of CCF are performed. In Section 5.1, the CCF WSL distributions generated from different BNs are assessed. It is emphasized that the BNs are built using illustrative, limited-scope models and statistical analysis results that differ from other studies. Thus, the results do not reflect actual hazards at the study nodes. Nonetheless, the results presented in Section 5.1 demonstrate the benefits and challenges associated with using the proposed framework. As a supplemental benefit of the use of BNs, Section 5.2 shows how a BN-based deaggregation method scheme Chapter 4 can be applied to better understand CCFs.

5.5.1 Coastal compound flood distribution

Figure 57 plots the CCF WSL distribution under two seasonality categories generated by Tier-1 BN, assuming the occurrence of a tropical cyclone of unknown characteristics. It can be observed that the CCF WSL distribution under Season 1 has a slightly higher modal WSL value than Season 2, which captures the seasonal differences in MR river flow and storm parameter probabilities.

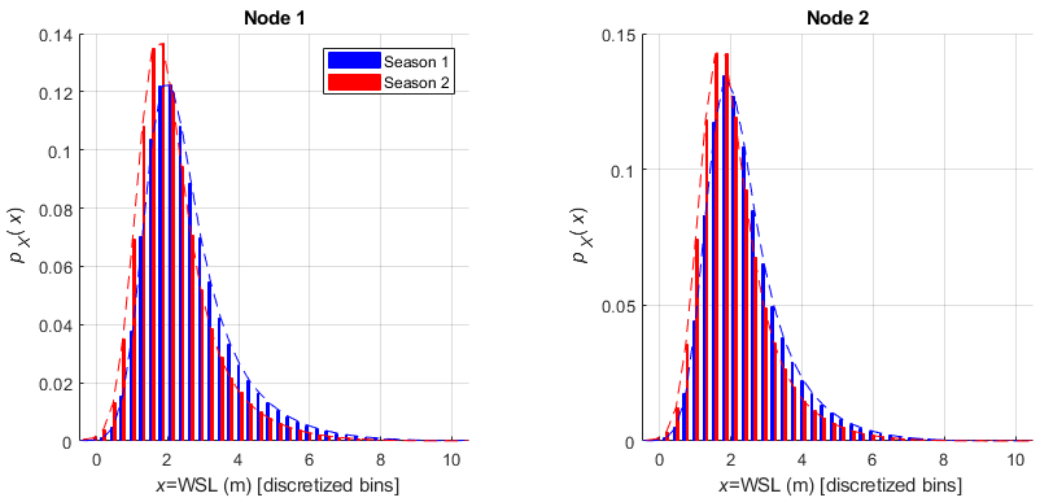


Figure 57: CCF WSL distribution generated by Tier-1 BN

Figure 58 plots the CCF WSL distribution under three TC intensity categories generated by Tier-2 BN. It can be observed that CCF WSL increases with the TC intensity (i.e., the probability distributions shift rightward), which is primarily caused by the connection between higher ΔP and higher CCF WSL. However, it is noted that the joint distribution model for TC parameters captures the complex dependencies in storm parameters and their resulting influence on CCF WSL.

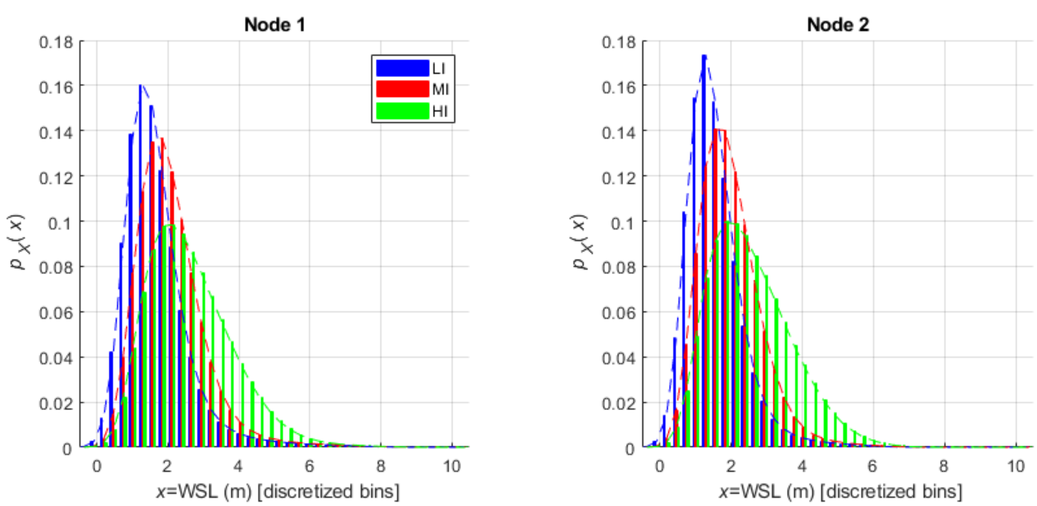


Figure 58: CCF WSL distribution generated by Tier-2 BN

Figure 59 plots the CCF WSL distributions and hazard curves under illustrative SST rise scenarios generated by Tier-3 BN. A clear trend can be observed that with the rise of SST, the CCF WSL distributions shift rightward, which represents a more critical CCF prediction. Meanwhile, with the rise of SST, the predicted hazard curve (hazard curve reflected reduced SRR under the rise of SST) rises.

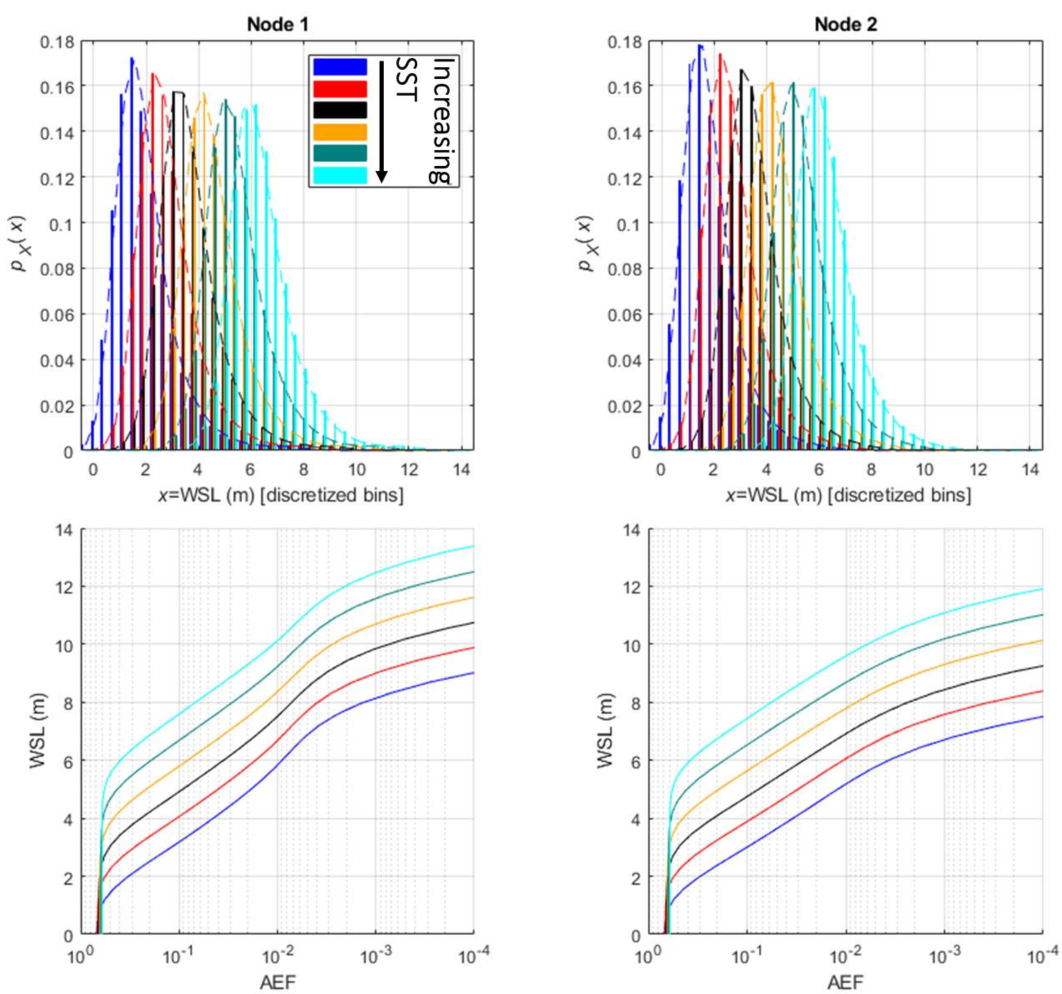


Figure 59: CCF WSL distribution under Tier-3 BN (top row); Hazard curve (CCF WSL) under Tier-3 BN (bottom row)

To quantitatively compare CCF predicted by multiple models, Table 13 lists examples of 100-, 200-, and 500-year CCF WSL predicted by Tier-1 BN, Tier-2 BN, and the Tier-3 BN (under a moderate SST rise scenario) considering the illustrative modeling choices made in this study.

Table 13 Tabulated CCF predicted under different models and scenarios

		100-year	200-year	500-year
Node 1	Tier-1 BN	6.2 m	6.9 m	7.8 m
	Tier-2 BN	5.4 m	6.1 m	7.0 m
	Tier-3 BN (moderate SST rise)	8.4 m	9.3 m	10.3 m
Node 2	Tier-1 BN	5.7 m	6.3 m	7.0 m
	Tier-2 BN	4.8 m	5.3 m	5.9 m
	Tier-3 BN (moderate SST rise)	7.8 m	8.4 m	9.0 m

5.5.2 Coastal compound flood deaggregation

Chapter 4 developed a BN-based coastal hazard deaggregation method. This deaggregation method can provide insights to inform the TC parameter domain of potentially severe coastal hazards, which is helpful for risk-informing and decision-making processes and refined analyses. Here, this deaggregation method is applied for CCF. An evidence case (EC) of CCF WSL > 5.13 m is utilized here to evaluate the performance of this tiered BN framework in identifying the dominant TC parameter combinations of significant CCF events.

First, to visualize the track information of dominant TCs, Figure 60 shows the pair-wise joint probability mass (PM) plot of X_0 (landfall location) and Θ (track angle) under EC generated from Tier-1 BN, Tier-2 BN and Tier-3 BN (moderate SST rise scenario). The no-evidence case is plotted for comparison. It can be observed that the input of EC notably affects the pair-wise joint PM of X_0 and Θ (compared with no-evidence case). The dominate track parameters are concentrated around the same dominate values across all tiers. However, Tier-2 BN generates a more coherent range of dominant TC tracks compared to Tier-1 BN. Under Tier-3 BN (with a moderate SST rise scenario), a notable expansion of the range of dominant TCs tracks compared to Tier-2 BN is observed.

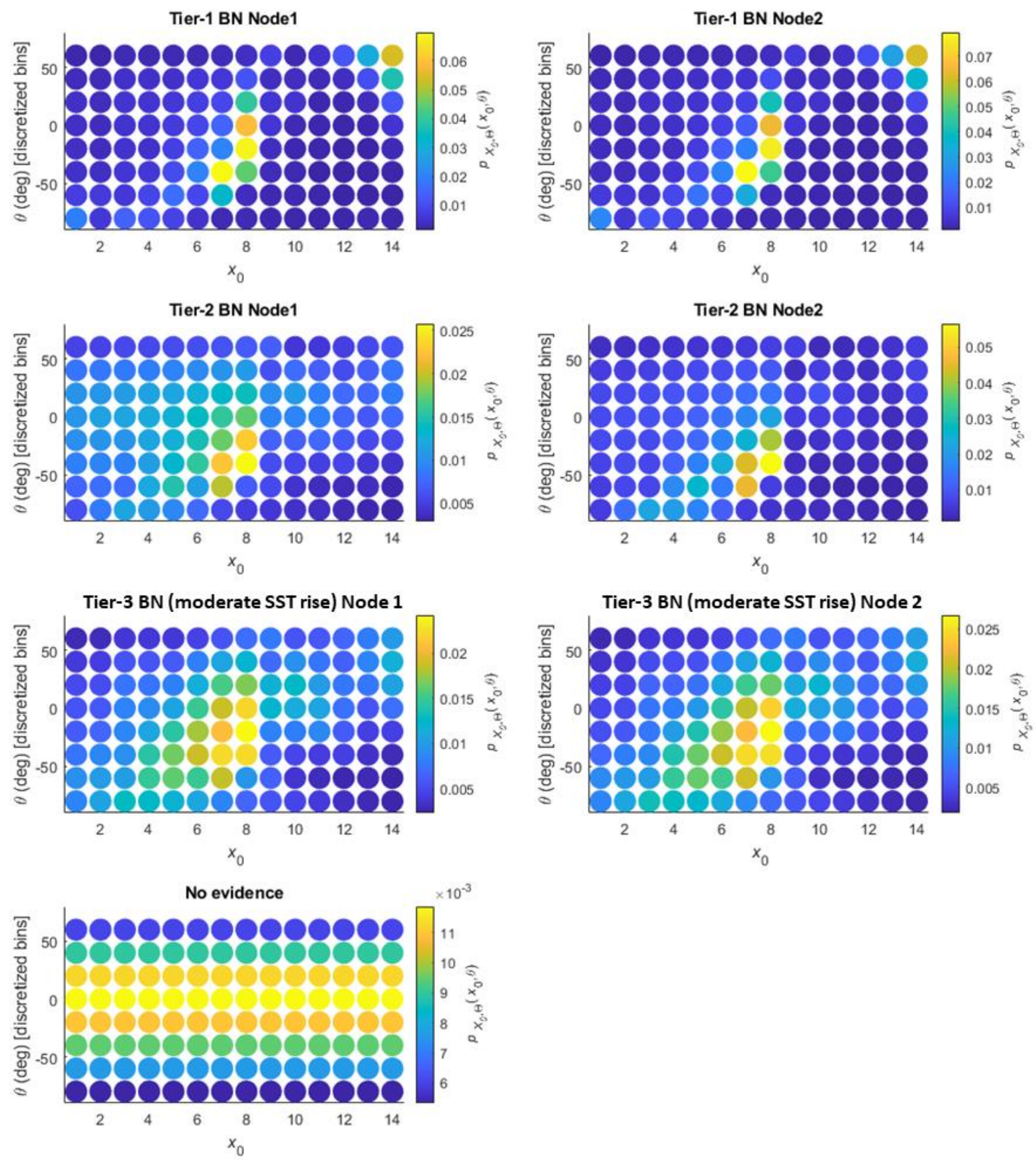


Figure 60: Pair-wise joint PM of X_0 and Θ under EC (WSL > 5.13 m)

The contributions of ΔP,V_f and R_max in conjunction with the TC track variables are visualized in Figure 61 to Figure 63 as 3D stacked bar plots with partitions of $\Delta P, V_\mathrm{f}$ and R_max identified with different colors. From Figure 61, it can be observed that TCs with high ΔP (ΔP >48 hPa) dominant most dominant TC tracks for the high WSL EC. Under the Tier-3 BN (with a moderate SST rise scenario), the portion of relatively low ΔP slightly increased relative to the inissights from other tiers. Figure 62 shows that moderate V_f (25 km/hr >V_f >15 km/hr) dominant most dominant TC tracks under EC, which is similar to the no-evidence case. Figure 63 shows that the contribution of small $R_{\max}$ ($R_{\max}$ <60 km) dominant most dominant TC tracks under EC, which is also similar to the no-evidence case.

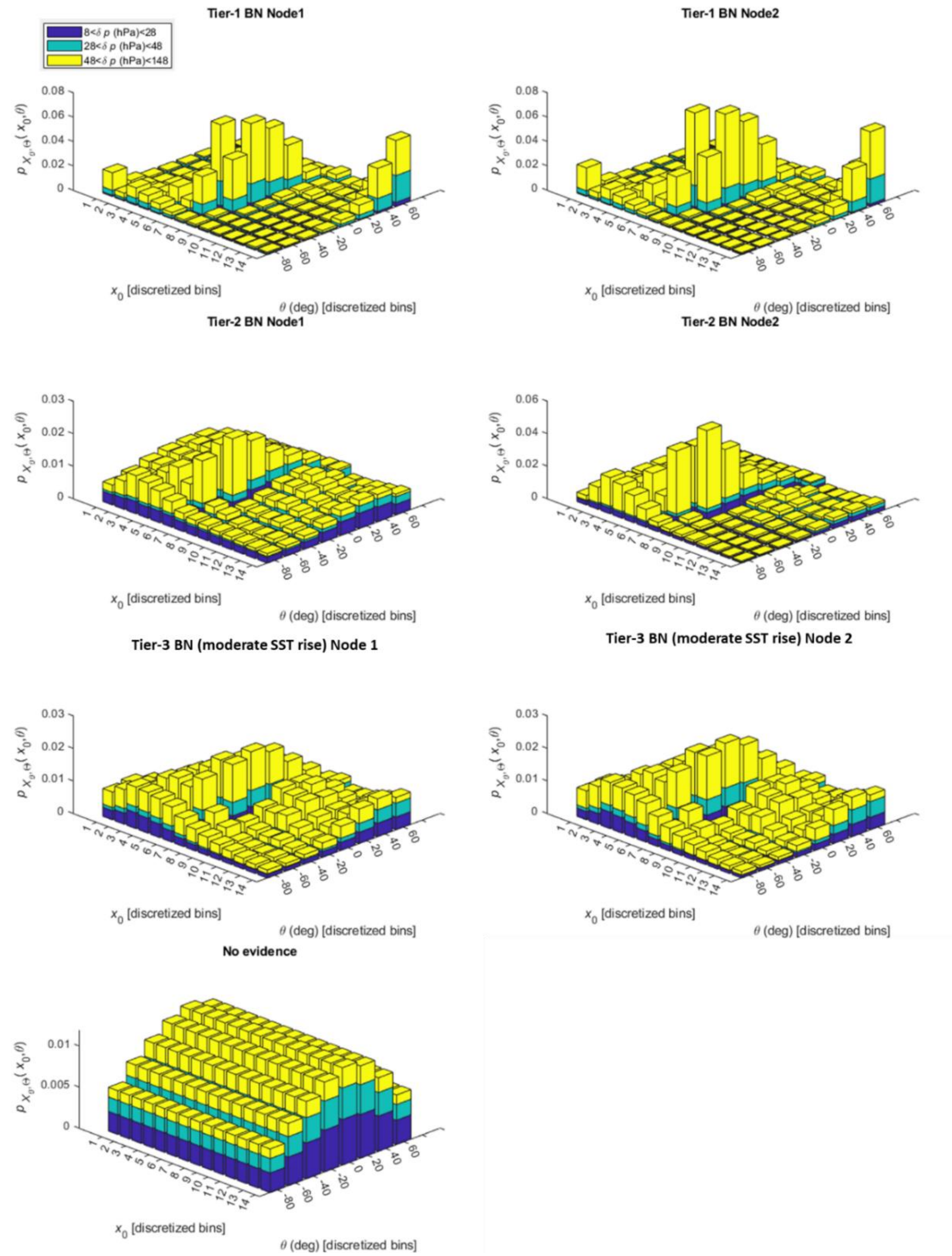


Figure 61: 3D stacked bar plot of TC track information with ΔP identified with colors

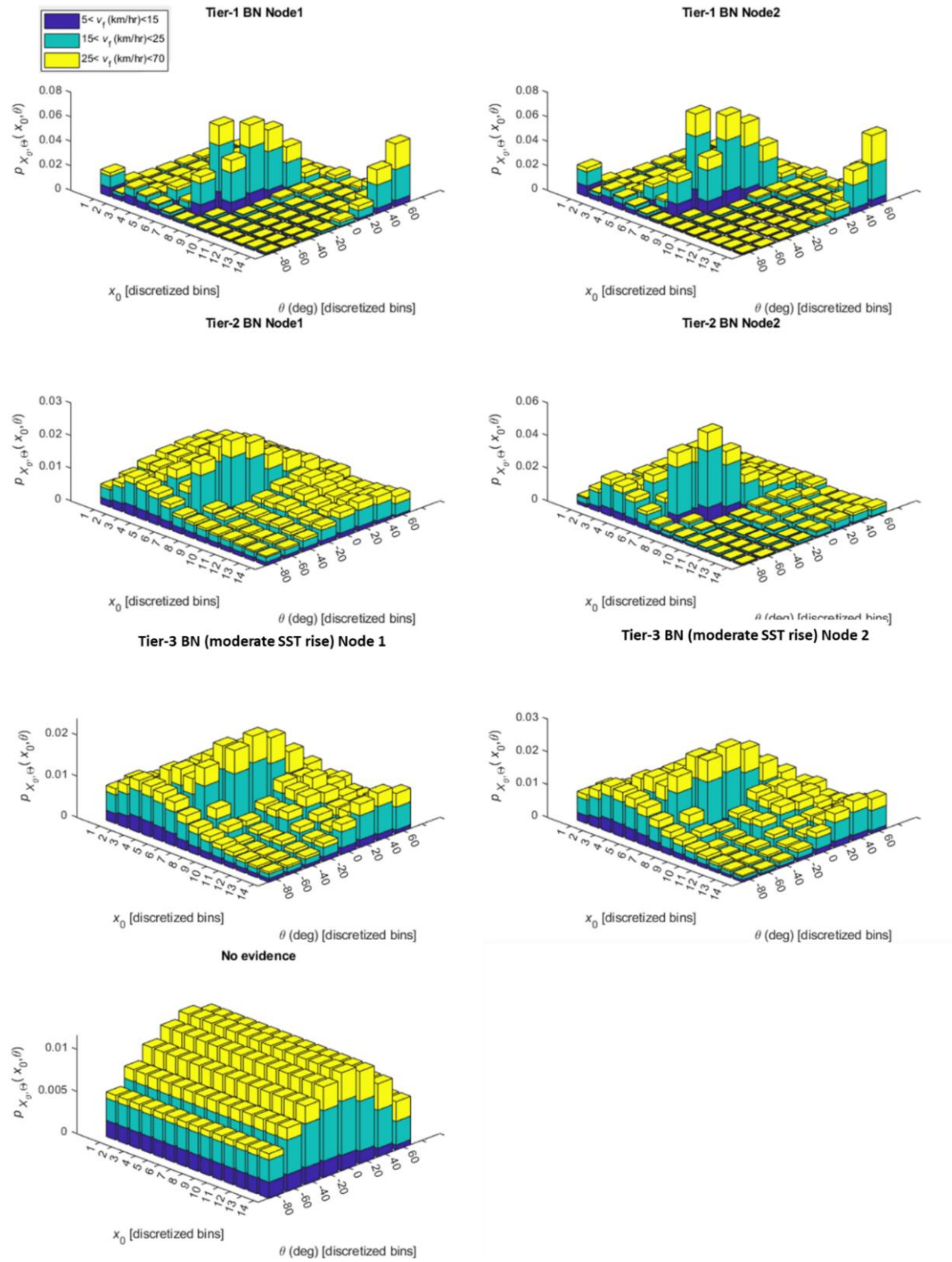


Figure 62: 3D stacked bar plot of TC track information with V_{f} identified with colors

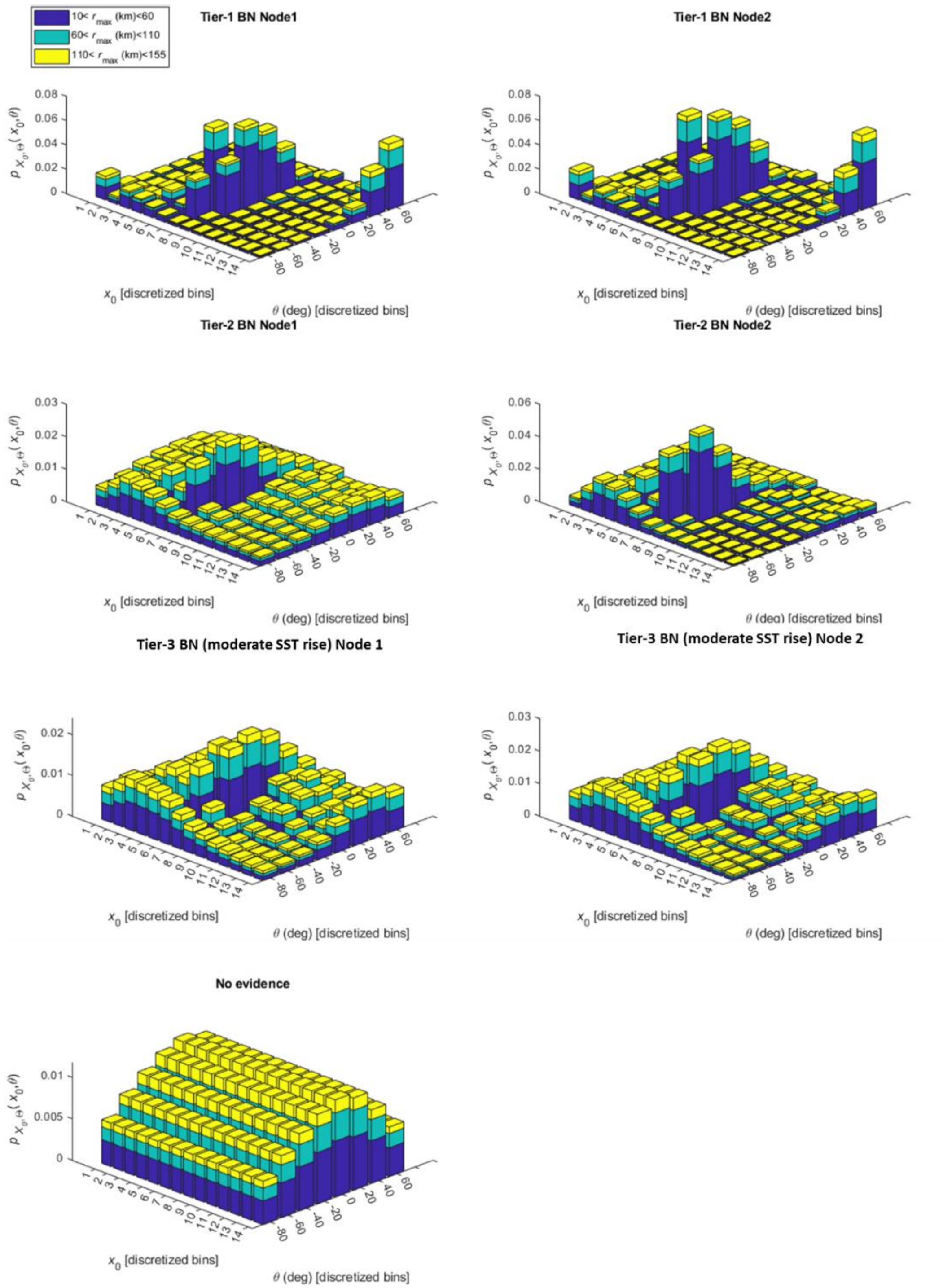


Figure 63: 3D stacked bar plot of TC track information with R_{max} identified with colors

5.6 *Conclusion of Chapter 5*

In this work, a framework for constructing multiple tiers of CCF BN model is established. The motivation behind this framework lies in the idea of creating tiered BN models that become progressively more complex as more data and computing resources become available. The simplest BN only considers storm surge and river processes as mechanisms contributing to CCF, while the most sophisticated BN incorporates both rainfall and potential climate conditions as additional CCF mechanisms.

The framework is implemented for a case study conducted in New Orleans, LA. A series of joint distribution, numerical, machine learning, and experimental models are used to build specific multiple tiers of BN for this study region. CCF Probabilistic analysis and deaggregation are performed based on established multiple tiers of BN. Although a series of simplifications and coarse assumptions are used herein, the results demonstrate the promise of these BN models in generating CCF distribution under various resource availability levels. Furthermore, the deaggreagtion results highlight the potential of this series of BN models for the identification of dominant TC of CCF events that might bring substantial damage to coastal communities, which is helpful for both risk-informing decision-making related to CCFs and identifying the potential need for high-fidelity numerical CCF events modeling.

It is important to note that data and computing resources needed to implement this tiered BN framework may differ depending on the study region. In particular, additional nodes, dependence models, and more sophisticated process models may be necessary to account for the temporal process of CCF mechanisms in a given region (Joyce et al. 2018). In the Tier-3 BN, alternative strategies like linking climate change and moisture and rainfall can be explored. As climate models continue to evolve, there is also room for integrating more sophisticated regional climate change and TC activity models into this framework.

In summary, the adaptability and complexity of the BN framework should be tailored to the unique characteristics and data availability of the study region, allowing for a more precise representation of CCF events over time.

Chapter 6: Summary and Future Work

6.1 Summary

This dissertation documents a series of efforts to extend the current PCHA (Nadal-Caraballo et al. 2020, 2022a; b) and enables its usage for coastal compound flood analysis.

In Chapter 2, machine learning techniques are used to develop a data imputation method to “fill in” missing storm parameter records in historical datasets used for JPM-based coastal hazard analysis. The study extends upon previous imputation work performed by the USACE. Specifically, ANN and GPR models are investigated as candidate machine learning-derived data imputation models, and the performance of different model parameterizations is assessed. Candidate imputation models are compared against existing statistical relationships. The effect of the data imputation process on statistical analyses (marginal distributions and correlation measures) is also evaluated for a series of example coastal reference locations. It is found that a GPR model (with baseline parameterization) outperforms other models assessed in this work. Overall, the potential of using efficient machine learning models for imputing missing historical storm data is demonstrated.

In Chapter 3, a series of TC parameters joint distribution models are comparatively assessed. Those models are based on assumptions of parameter independence, partial dependence (i.e., dependence between central pressure deficit and radius of maximum wind), and full dependence. Full dependence models consider a range of copula models, such as the Gaussian copula and vine copulas that combine linear-circular copulas with Gaussian or Frank copulas. The Gaussian copula represents the current state of practice used by the USACE. However, the consideration of linear-circular copulas allows for the characterization of correlation between linear variables (e.g., ΔP, V_{f} and $R_{\max}$) and circular variables (e.g., Θ). The sensitivity of the results to different joint distribution models is assessed by comparing hazard curves at representative locations in New Orleans, LA (USA). The stability of hazard curves generated using an MGC considering variation in the selection of the zero-degree convention is also assessed. The tail dependence between large central pressure deficit and large radius of maximum wind associated with various copula models are also compared. It is found that the linear-circular copula and Frank copula-based vine copula model improve the stability of hazard curves and maximize tail dependence between large ΔP and large $R_{\max}$. However, the MGC models exhibit good performance within this study region (due to the limited range of heading values of landfalling storms in the region) and have the advantage of being easier to implement than vine copulas.

In Chapter 4, a BN-based framework for coastal hazard deaggregation is developed. Copula-based joint distribution models are applied to characterize the dependence among TC atmospheric parameters and compute CPTs for TC atmospheric parameters nodes in the BN. Machine-learning-based surrogate models are applied to characterize the relationship between TC atmospheric parameters and coastal hazards and compute CPTs for coastal hazard nodes in BN. Case studies are applied to New Orleans, LA. The primary accomplishment of this work is the development of a deaggregation method for multiple coastal hazards. Leveraging this deaggregation method, the dominant TCs that have the potential to generate large storm surges and storm rainfall (and their combinations) in the study region are identified.

In Chapter 5, a multiple tiers BN CCF analysis framework is established. In this framework, multiple tiers of BN models with different complexities are designed for study cases with varying levels of data availability and computational resources. A case study is conducted in New Orleans, LA to implement this framework. In the Tier-1 BN model, storm surge and river flow are involved based on ADCIRC simulations. A seasonality node is used to capture the dependence between concurrent river flow and TC parameters. Historical river discharge records and historical storm datasets are leveraged to build probability models of concurrent river flow and joint distribution models of TC parameters under separate meteorological seasons. In the Tier-2 BN model, joint distribution models of TC parameters are built for separate TC intensity categories. Storm rainfall is incorporated based on HEC-RAS simulations. In the Tier-3 BN model, potential climate change scenarios are incorporated and directly impact TC activity and SLR. Flood antecedent conditions are also involved to enhance the overall model completeness. In the case study, a series of joint distribution, numerical, machine learning, and experimental models are used to compute conditional probability tables needed for BNs. Probabilistic analyses are performed based on established BN models. The results demonstrate the promise of this framework in performing CCF hazard analysis under various resource availability constraints.

6.2 Key Contribution

This dissertation offers the following major contributions that support the extension of the USACE's PCHA framework for analysis of CCFs:

- Explored and comprehensively assessed the performance of multiple machine learning strategies to impute missing observations in historical storm datasets using a holistic statistical and physically-informed evaluation.
- Implemented a series of circular-variable statistical techniques (e.g., linear-circular vine copulas, VKF) in the JPM analysis of TC parameters and improved the robustness of hazard curves under modeling conventions associated with storm heading.
- Comparatively assessed the performance of different joint distribution models for TC parameters ranging from independence to full dependence models (including MGC and different vine copulas) and provided insights regarding the tradeoffs between statistical accuracy and implementation practicality.
- Developed a BN-motivated deaggregation method for multiple coastal hazards that enables the identification of dominant contributors to significant coastal hazards (i.e., coastal hazards that equal or exceed a specified threshold) to inform decision-making processes and model refinements.
- Leveraged BN inference analysis to investigate the correlation between storm surge and storm rainfall in inland and coastal locations of the case study region (i.e., New Orleans, LA) and visualized this correlation.
- Conceptualized and developed CCF analysis framework comprised of multiple tiers of BN models that can be scaled based on case needs and resources for probabilistic coastal compound flood analysis.

- Applied the multi-tiered BN CCF analysis framework to an illustrative case study and demonstrated its promise for predicting CCF hazards under potential climate change scenarios.

6.3 *Future Work*

Following a series of efforts and accomplishments presented in this dissertation, a discussion of future work is presented here. The majority of this dissertation focused on using the region of New Orleans, LA, as a case study area. Situated in the Gulf of Mexico, this region is exposed to a notably high risk of coastal hazards (IPET 2009). However, it’s important to note that factors such as TC parameter statistical property (explored in Chapter 3), the dominant TC range of potentially severe coastal hazard (as investigated in Chapter 4), and CCF mechanisms (as established in Chapter 5) may vary in other regions, such as the eastern coast of the U.S, which is also under high risk of coastal hazards (Cialone et al. 2015; Nadal-Caraballo et al. 2015). Therefore, future work will extend and apply the methodologies and frameworks developed in this dissertation to other study regions. It is anticipated that statistical conclusions (e.g., regarding the relative tradeoffs between the use of circular-variable enabling vine copulas and the MGC) may differ for other study regions.

In addition to the extension of the effort devoted to this dissertation, there is a pressing need to develop more advanced tools and models for coastal hazard analysis. Surrogate models play a pivotal role in efficiently generating storm response in coastal hazard analysis. For CCF analysis, a CCF WSL surrogate model can significantly reduce the computational expenses associated with numerical modeling. In Chapter 5, a simplified CCF WSL surrogate model trained based on synthetic storm suites is employed. However, there is a need for more sophisticated surrogate models capable of predicting CCF under realistic storm scenarios. One promising approach can be combining HEC-RAS simulations with deep learning techniques, using storm time-series data as predictors (Lee et al. 2021).

Under the context of climate change, the extent to which TC activity will be impacted is uncertain (Wu et al. 2022). To enhance our ability to predict coastal hazards under potential climate change, it is of high interest to invest effort in developing models that can capture the relationship between regional climate conditions and TC activity.

Meanwhile, it is important to emphasize that, under coastal hazards, coastal urban areas might suffer more losses than remote areas. However, conducting coastal hazard analysis within urban areas becomes notably complex due to the presence of urban infrastructure and human factors (Gallien et al. 2018). More sophisticated BN models with human factors involved can be a potential strategy to enhance the capability of PCHA inside urban areas.

Lastly, considering the significance of the temporal effect during a CCF (Joyce et al. 2018), it is of high interest to explore the potential of implementing dynamic BN(Ghahramani 1998; Sen et al. 2022) in CCF analysis, which is an extension of standard BN that can account for the varying dependence between variables under temporal process.

Appendix A: Supplemental information for Chapter 2

To facilitate a consistent characterization of storm climatology, the USACE's CHS-PCHA framework develops marginal distributions and computes storm recurrence rates and parameter correlations along an idealized coastline discretized into a series of CRLs. The application of the CHS-PCHA framework is illustrated in **Error! Reference source not found.**.

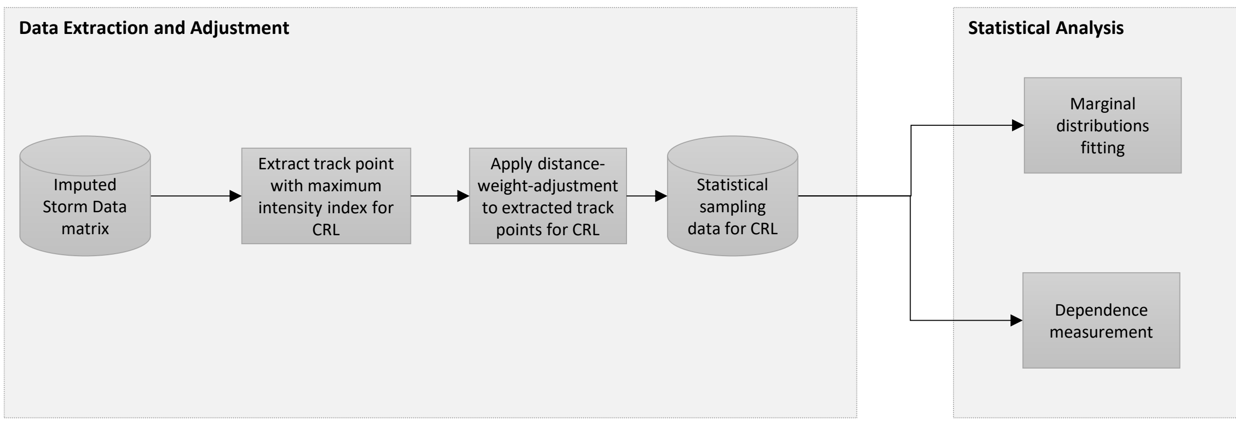


Figure A-1: Flowchart of the USACE's CHS-PCHA statistical analysis

The process uses the imputed storm data matrix as the input data. An assessment of storm climate is performed by first extracting storms within 600 km of each CRL. To estimate the marginal distributions of the storm parameters, the framework develops the distributions in a manner that accounts for the spatial distance of storm points from the CRLs. Therefore, a distance-weight adjustment is conducted for all data sampled at a given CRL. The steps for performing the statistical data sampling in the USACE's CHS-PCHA are as follows:

For each storm passing through the 600 km radius capture zone of a CRL, calculate a distance-based weight $w(d_i)$ for each track point of the storm.

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

where d_i is the distance between the location associated with data point i and CRL. h_d is the optimal kernel size, which is taken as 200 km (Nadal-Caraballo et al. 2019).

For each storm, extract the track point with the largest intensity index $I_{\Delta p}$ as the sampling location.

$$I_{\Delta p_i} = w(d_i)\Delta p_i \tag{A-2}$$

where Δp_i is the central pressure deficit recorded (or imputed) at track point i.

Compute the mean μ and standard deviation σ of the storm parameters at sampling locations

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \tag{A-3}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N-1}} \tag{A-4}$$

Where x_i is the observed value of storm parameter at the ith sampling location, N is the number of sampling locations.

Use the values computed in Step 2 to compute a z-score for normalizing the parameters

$$z_i = \frac{x_i-\mu}{\sigma} \tag{A-5}$$

The value z_i is the normalized index of optimal sampling location point i's data

Use the distance-based weight to compute a distance-weighted mean and standard deviation for each parameter

$$\mu_{DW} = \frac{\sum_{i=1}^{N} w(d_i)x_i}{\sum_{i=1}^{N} w(d_i)} \tag{A-6}$$

$$\sigma_{DW} = \sqrt{\frac{\sum_{i=1}^{N} w(d_i)(x_i - \mu_{DW})^2}{\left(\frac{N-1}{N}\right)\sum_{i=1}^{N} w(d_i)}} \tag{A-7}$$

where μ_{DW} is the distance-adjusted mean and σ_{DW} is the distance-weight adjusted standard deviation.

Apply the distance-adjusted mean and standard deviation and the z-score to compute a new, distance-adjusted storm parameter

$$x'_i = z_i \sigma_{DW} + \mu_{DW} \tag{A-8}$$

As a result, this process is a transformation of the storm parameters relative to a CRL based on the parameter's distance from that location. The development of the marginal distribution and the correlation analysis is then performed by using the distance-adjusted values.

Appendix B: Supplemental information for Chapter 3, Part 1

B.1 Linear-circular copula comparison

The JW linear-circular copula is derived from a linear-circular joint distribution provided by Johnson and Wehrly (1978). The JW linear-circular joint distribution contains a circular random variable θ and a linear random variable x, and the joint PDF $f(\theta, x)$ can be expressed as (Lin and Dong 2019):

$$f(\theta, x) = 2\pi g(\xi) f_\Theta(\theta) f_X(x) \tag{B-1}$$

where $f_\Theta(\theta)$ and $f_X(x)$ are the marginal PDFs of circular variable Θ and linear variable X. $g(\xi)$ is the PDF of the "coupling variable" ξ, which is defined as:

$$\xi = \begin{cases} 2\pi(u - v), u \geq v \\ 2\pi(u - v + 1), u < v \end{cases} \tag{B-2}$$

where $u = F_\Theta(\theta)$ and $v = F_X(x)$. The distribution of ξ is usually represented by a mixture of two von Mises distributions (MvM2) (Carta et al. 2008). According to Skalar's theorem, we can write the JW linear-circular copula PDF as:

$$\begin{aligned} c_{JW}\left(F_\Theta(\theta), F_X(x)\right) &= \frac{f(\theta, x)}{f_\Theta(\theta) f_X(x)} = \frac{2\pi g(\xi) f_\Theta(\theta) f_X(x)}{f_\Theta(\theta) f_X(x)} \\ &= \begin{cases} 2\pi g\left(F_\Theta(\theta) - F_X(x)\right), F_\Theta(\theta) \geq F_X(x) \\ 2\pi g\left(F_\Theta(\theta) - F_X(x) + 1\right), F_\Theta(\theta) < F_X(x) \end{cases} \end{aligned} \tag{B-3}$$

The QS linear-circular copula was introduced by (García-Portugués et al. 2013), which is derived from corollary 3.2.5 of the QS copula in Nelsen (2007). The QS copula PDF is defined as follow:

$$c_{QS}\left(F_\Theta(\theta), F_X(x)\right) = 1 + 2\pi\alpha(1 - 2F_X(x))\cos\left(2\pi F_\Theta(\theta)\right) \tag{B-4}$$

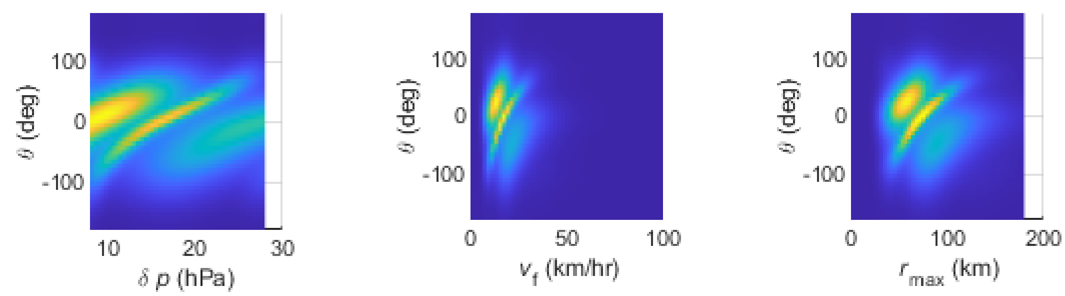
Where α is the copula parameter, $|\alpha| \leq (2\pi)^{-1}$. Here we used our statistical sampling data to fit both two linear-circular copulas. Figure B-1 plots the pair-wise PDF surface of each fitted linear-circular copula. We use the copula log-likelihood $\mathcal{L}$ to compare the fit of copulas to the data. The copula log-likelihood $\mathcal{L}$ is calculated as:

$$\mathcal{L} = \sum \ln\left(c(F_\Theta(\theta_i), F_X(x_i))\right) \tag{B-5}$$

Where θ_i is the θ value of the i th data point, x_i is the corresponding linear variable value of the i th data point. Larger values of $\mathcal{L}$ indicate a better fitted copula of the data. The computed values of $\mathcal{L}$ for each pair copula are provided in Figure B-1. For each pair copula, except the case of pair copula between ΔP and Θ of LI, the JW linear-circular copula has a higher log-likelihood value than the QS linear-circular copula. Additionally, for each case, the QS linear-circular copula generates a symmetrical pair-wise PDF surface while the JW linear-circular copula does not, which suggests the JW linear-circular copula has higher flexibility than the QS linear-circular copula, though perhaps also exhibiting a higher potential for overfitting.

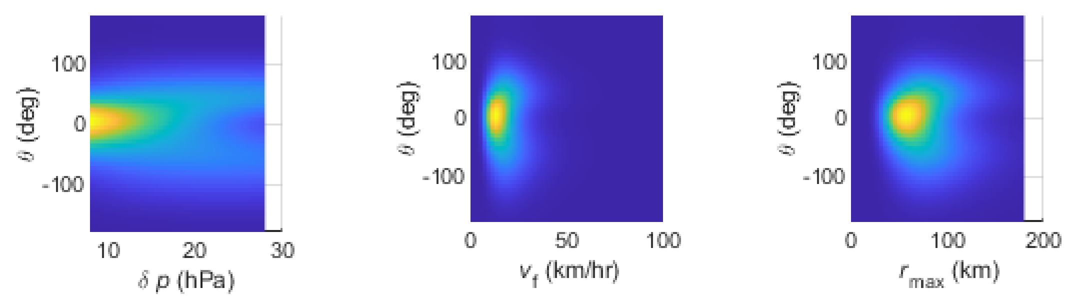
LI

JW



$\mathcal{L} = 1.22$ $\mathcal{L} = 2.55$ $\mathcal{L} = 3.72$

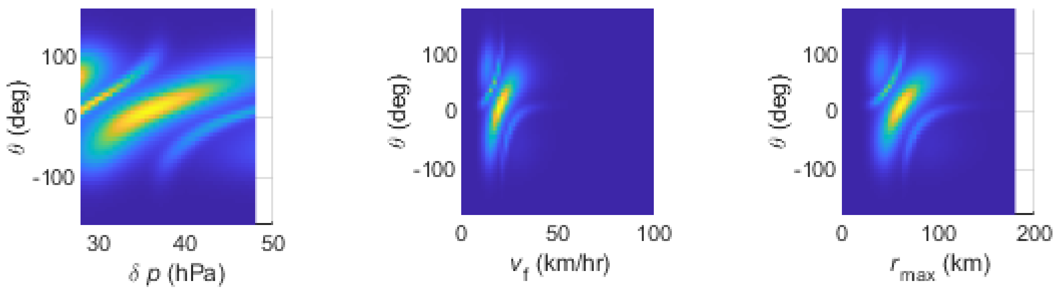
QS



$\mathcal{L} = 5.52$ $\mathcal{L} = 0.05$ $\mathcal{L} = 2.7$

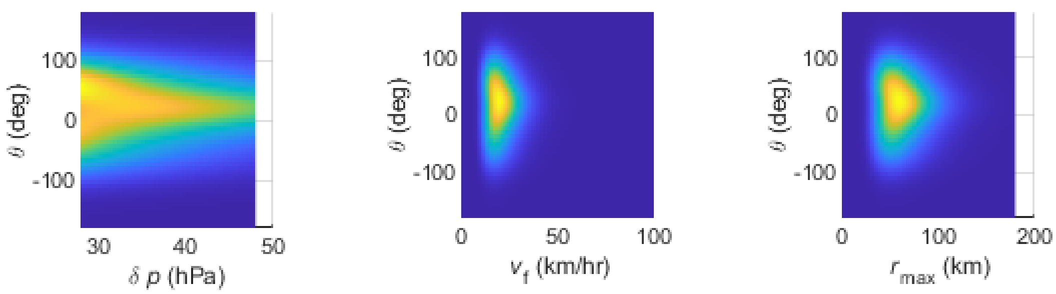
MI

JW



$\mathcal{L} = 3.22$ $\mathcal{L} = 2.92$ $\mathcal{L} = 3.94$

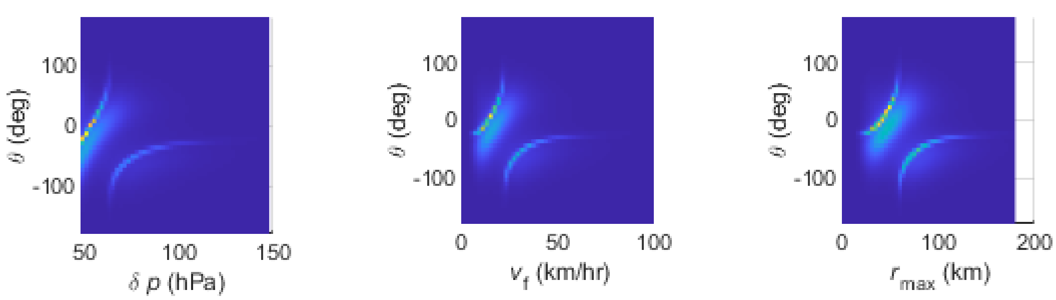
QS



$\mathcal{L} = 0.74$ $\mathcal{L} = 0.99$ $\mathcal{L} = 0.19$

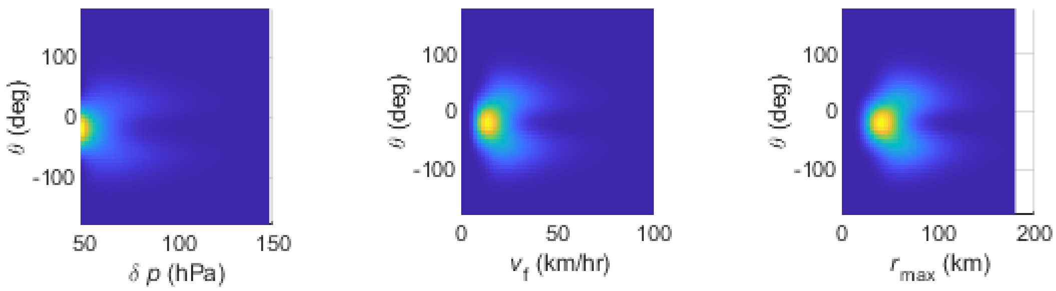
HI

JW



$\mathcal{L} = 5.43$ $\mathcal{L} = 1.53$ $\mathcal{L} = 7.83$

QS



$\mathcal{L} = 4.19$ $\mathcal{L} = 0.37$ $\mathcal{L} = 3.19$

Figure B-1: The pair-wise PDF surface of JW linear-circular copula and QS linear-circular copula. The corresponding copula log-likelihood values are presented under each plot

B.2 Vine copula joint PDF derivation

In the LCGV, TC atmospheric parameter data (i.e., the distance-adjusted data) are used to fit the three pair-wise joint linear copulas (C_{14}, C_{24}, C_{34}) in Tier 1, as illustrated in the left graphic of Figure 18. Next, the joint linear-circular copulas are used to calculate the conditional copula values ($C_{1|4}, C_{2|4}, C_{3|4}$) at data points as shown in the middle box of Figure 18. First, we derive the expressions used to calculate $C_{i|j}$. Letting $u_i = F_i(x_i)$, $C_{i|j}$ is calculated as:

$$C_{i|j}(u_i|u_j) = F_{i|j}(x_i|x_j) = P(X_i \leq x_i|X_j = x_j) = \int_{-\infty}^{x_i} f_{i|j}(x_i|x_j)dx_i \tag{B-6}$$
$$= \int_{-\infty}^{x_i} \frac{f_{ij}(x_i,x_j)dx_i}{f_j(x_j)} = \int_{-\infty}^{x_i} c_{ij}(u_i,u_j) f_i(x_i)dx_i = \int_0^{u_i} c_{ij}(u_i,u_j)du_i$$

Where it is noted that the copula density $c_{ij}(u_i,u_j)$ may be expressed as the ratio of the joint PDF of X_i and X_j to the product of the marginal PDFs, i.e., $c_{ij}(u_i,u_j) = \frac{f_{ij}(x_i,x_j)}{f_i(x_i)f_j(x_j)}$.

The computed conditional copula CDF values $C_{i|j}$ are used to fit the Tier 2 copula $C_{123|4}$, as shown in the rightmost graphic of Figure 18. Given the $C_{123|4}$, which is a three-dimensional MGC the copula density function $c_{123|4}$ can be derived. We have:

$$c_{123|4}(u_1,u_2,u_3|u_4) = \frac{\partial^3 C(u_1,u_2,u_3|u_4)}{\partial u_1 \partial u_2 \partial u_3} \tag{B-7}$$

Given the copula density $c_{123|4}(u_1,u_2,u_3|u_4)$, the joint density can be derived as follows. First, it is noted that the joint PDF $f_{1234}(x_1,x_2,x_3,x_4)$ can be expressed as:

$$f_{1234}(x_1,x_2,x_3,x_4) = f_{123|4}(x_1,x_2,x_3|x_4)f_4(x_4) \tag{B-8}$$

where the $f_{123|4}(x_1, x_2, x_3|x_4)$ is the conditional joint PDF of X_1, X_2 and X_3 given X_4.

According to Skalar's theorem for the 3-dimensional copula and 2-dimensional copula, we have

$$f_{123|4}(x_1, x_2, x_3|x_4) = c_{123|4}\left(F_{1|4}(x_1|x_4), F_{2|4}(x_2|x_4), F_{3|4}(x_3|x_4)\right) * f_{1|4}(x_1|x_4) f_{2|4}(x_2|x_4) f_{3|4}(x_3|x_4) \quad \text{(B-9)}$$

Where:

$$f_{1|4}(x_1|x_4) = c_{14}\left(F_1(x_1), F_4(x_4)\right) f_1(x_1) \quad \text{(B-10)}$$

$$f_{2|4}(x_2|x_4) = c_{24}\left(F_2(x_2), F_4(x_4)\right) f_2(x_2)$$

$$f_{3|4}(x_3|x_4) = c_{34}\left(F_3(x_3), F_4(x_4)\right) f_3(x_3)$$

Combing the above expressions, we can calculate the joint PDF as:

$$f_{1234}(x_1, x_2, x_3, x_4) = c_{123|4}\left(F_{1|4}(x_1|x_4), F_{2|4}(x_2|x_4), F_{3|4}(x_3|x_4)\right) * c_{14}\left(F_1(x_1), F_4(x_4)\right) c_{24}\left(F_2(x_2), F_4(x_4)\right) c_{34}\left(F_3(x_3), F_4(x_4)\right) * f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4) \quad \text{(B-11)}$$

B.3 Estimated copula parameters

Table B-1: Estimated copula parameters

	LI	MI	HI
MGC	$R = \begin{bmatrix} 1 & -0.0599 & -0.3165 & 0.0678 \\ 0.0599 & 1 & -0.2092 & 0.3130 \\ -0.3165 & -0.2092 & 1 & -0.0032 \\ 0.0678 & 0.3130 & -0.0032 & 1 \end{bmatrix}$	$R = \begin{bmatrix} 1 & -0.0292 & -0.1246 & 0.1156 \\ -0.0292 & 1 & 0.1000 & 0.2152 \\ -0.1246 & 0.1000 & 1 & 0.3017 \\ 0.1156 & 0.2152 & 0.3017 & 1 \end{bmatrix}$	$R = \begin{bmatrix} 1 & 0.1948 & -0.4334 & 0.0514 \\ 0.1948 & 1 & 0.0991 & -0.0237 \\ -0.4334 & 0.0991 & 1 & -0.1135 \\ 0.0514 & -0.0237 & -0.1135 & 1 \end{bmatrix}$
JW Linear-circular copula (MvM2)	C_{14}: $\mu_1 = 1.30; \kappa_1 = 0.461;$ $P_1 = 0.504$ $\mu_2 = -2.099; \kappa_2 = 0.864;$ $P_2 = 0.496$ C_{24}:	C_{14}: $\mu_1 = -1.396; \kappa_1 = 1.261;$ $P_1 = 0.419$ $\mu_2 = 2.036; \kappa_2 = 1.271;$ $P_2 = 0.581$ C_{24}:	C_{14}: $\mu_1 = 0.375; \kappa_1 = 4.920;$ $P_1 = 0.200$ $\mu_2 = -1.870; \kappa_2 = 1.071;$ $P_2 = 0.800$ C_{24}:

	$\mu_1 = 0.064; \kappa_1 = 8.876;$ $P_1 = 0.091$ $\mu_2 = -2.886; \kappa_2 = 0.597;$ $P_2 = 0.909$ C_{34}: $\mu_1 = 2.114; \kappa_1 = 0.340;$ $P_1 = 0.506$ $\mu_2 = -0.491; \kappa_2 = 0.506;$ $P_2 = 0.494$	$\mu_1 = -1.569; \kappa_1 = 1.066;$ $P_1 = 0.842$ $\mu_2 = 2.165; \kappa_2 = 1.939;$ $P_2 = 0.158$ C_{34}: $\mu_1 = -0.296; \kappa_1 = 5.174;$ $P_1 = 0.154$ $\mu_2 = -2.751; \kappa_2 = 1.126;$ $P_2 = 0.846$	$\mu_1 = -2.622; \kappa_1 = 3.228;$ $P_1 = 0.299$ $\mu_2 = -0.461; \kappa_2 = 0.183;$ $P_2 = 0.702$ C_{34}: $\mu_1 = -2.860; \kappa_1 = 6.148;$ $P_1 = 0.207$ $\mu_2 = 1.091; \kappa_2 = 1.211;$ $P_2 = 0.793$
QS Linear-circular copula	$C_{14}: \alpha = 0.129$ $C_{24}: \alpha = -0.013$ $C_{34}: \alpha = -0.096$	$C_{14}: \alpha = -0.095$ $C_{24}: \alpha = 0.092$ $C_{34}: \alpha = 0.049$	$C_{14}: \alpha = 0.159$ $C_{24}: \alpha = 0.081$ $C_{34}: \alpha = -0.159$
LCGV Tier 2	$R = \begin{bmatrix} 1 & -0.1184 & -0.2713 \\ -0.1184 & 1 & -0.1503 \\ -0.2713 & -0.1503 & 1 \end{bmatrix}$	$R = \begin{bmatrix} 1 & -0.0119 & -0.0654 \\ -0.0119 & 1 & 0.0178 \\ -0.0654 & 0.0178 & 1 \end{bmatrix}$	$R = \begin{bmatrix} 1 & 0.0232 & -0.4408 \\ 0.0232 & 1 & 0.1158 \\ -0.4408 & 0.1158 & 1 \end{bmatrix}$
LCFV Tier 2	$C_{12\|4}$: α =-0.579 $C_{23\|4}$: α =-0.740	$C_{12\|4}$: α =0.057 $C_{23\|4}$: α =0.159	$C_{12\|4}$: α =1.317 $C_{23\|4}$: α =-0.205
LCFV Tier 3	$\alpha = -1.599$	$\alpha = -0.839$	$\alpha = -2.933$

B.4 Log-likelihood of different JPM assumptions

The log-likelihood of the j th JPM assumption is calculated as:

$$\mathcal{L}^j = \sum \ln (f_i^j) \tag{B-12}$$

Where f_i^j is the PDF value of the j th JPM assumptions at the i th statistical sampling points. The calculated log-likelihood values are presented in Table B-2, where the "-JW" and "-QS" specified the linear-circular copula used in corresponding vine models.

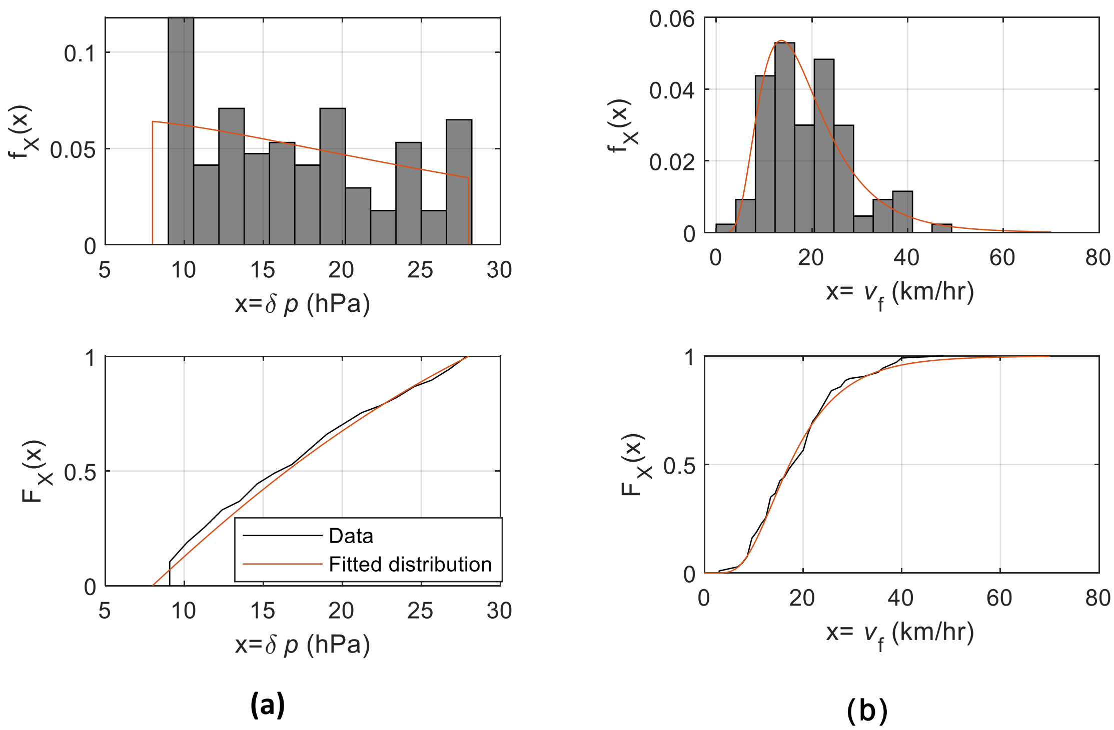
Table B-2: Log-likelihood of different JPM assumptions

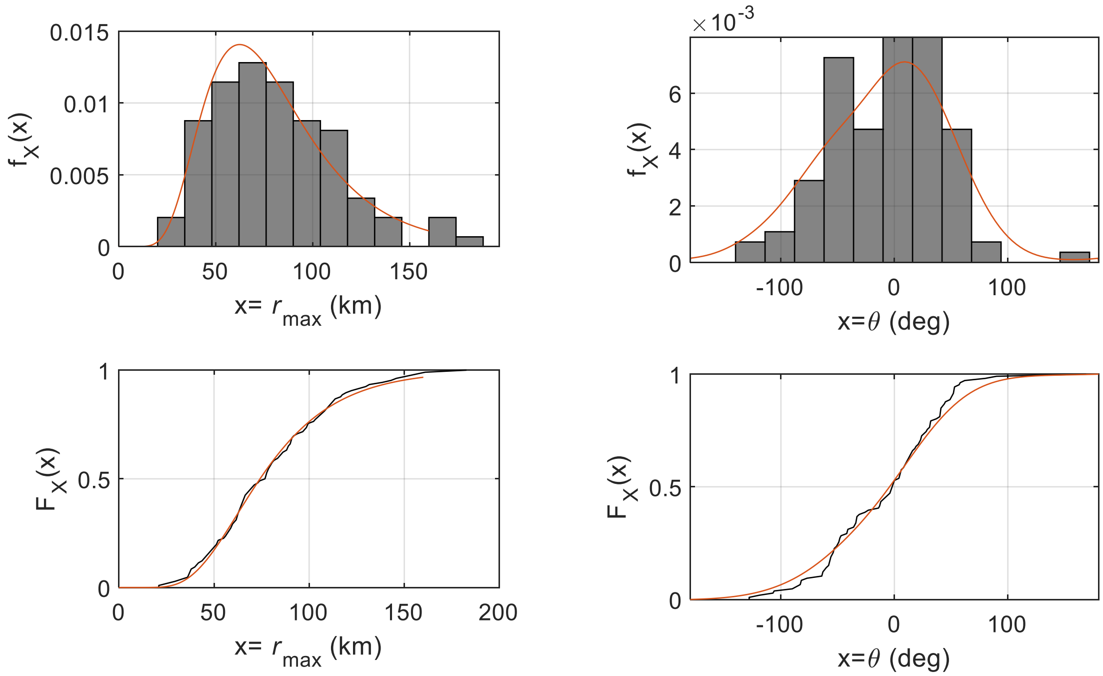
	Independence	Partial-dependence	MGC	LCGV-JW	LCFV-JW	LCGV-QS	LCFV-QS
LI	-1769.879	-2391.408	-1756.027	-1756.929	-1756.957	-1756.064	-1756.460
MI	-525.264	-721.070	-523.434	-515.167	-514.952	-523.270	-523.264
HI	-509.011	-666.949	-503.310	-492.677	-490.893	-498.820	-497.193

It can be observed that in both the MI category and HI category, the LCFV model with JW linear-circular copula has the highest log-likelihood value among all JPM assumptions. However, in the LI category, the MGC has the highest log-likelihood value.

Appendix C: Supplemental Information for Chapter 3, Part 2

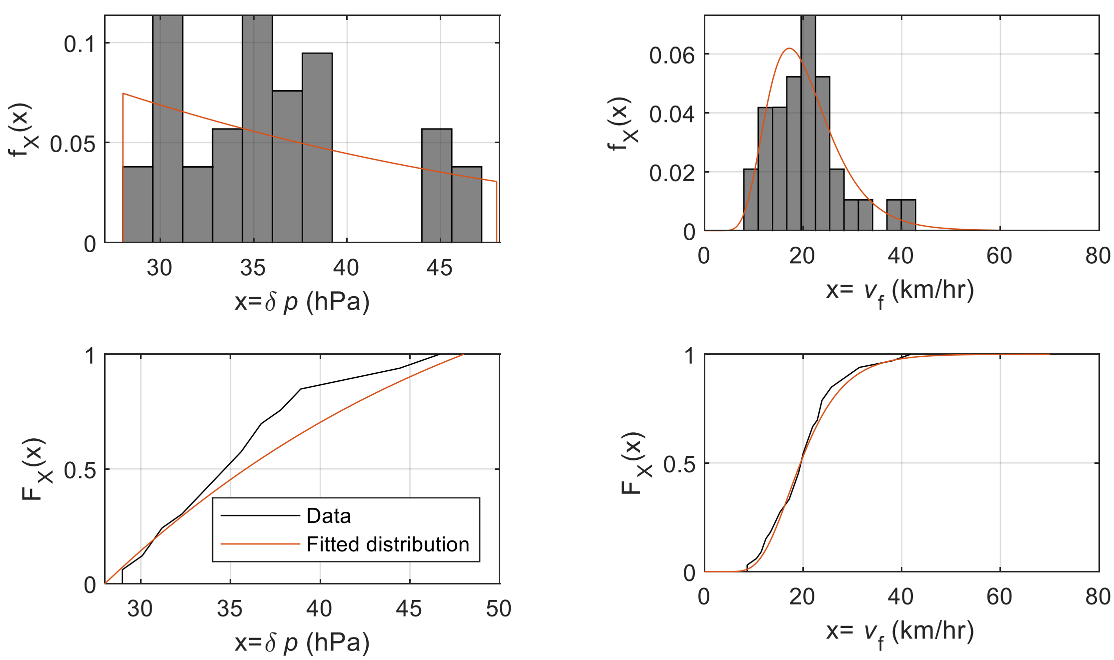
Figure C-1 through Figure C-3 shows the fitted marginal PDF and CDF curves with the corresponding statistical sampling data. In these figures, Θ is measured in [-180,180] clockwise with the zero-degree direction set to due North (consistent with typical practice). For each TC atmospheric parameter, the top figure shows the estimated marginal PDF (red curve) plotted over the normalized frequency histogram (black bar), and the bottom figure shows the estimated marginal CDF (red curve) plotted over the empirical CDF (black curve). The parameters of all fitted marginal distributions can be found in Table C-1. For ΔP, a single Weibull distribution is fitted to statistical sampling data of all storms and then truncated on the bounds (8,28], (28,48], and (48,148], for LI, MI, and HI intensity storms, respectively. For V_{f}, $R_{\max}$ and Θ, separate distributions are estimated for storms of LI, MI, and HI.



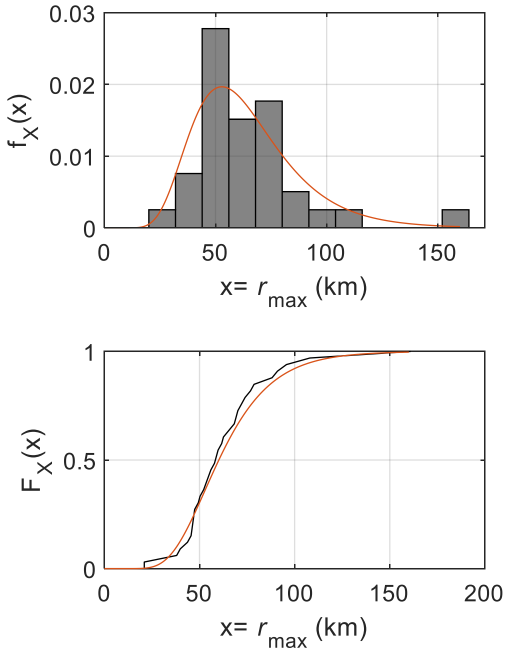


(c) **(d)**

Figure C-1: Marginal distributions fitted for LI TC atmospheric parameters

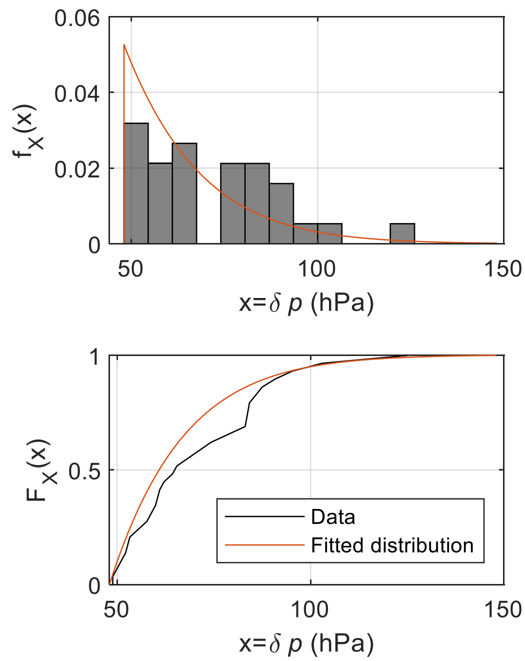


(a) (b)

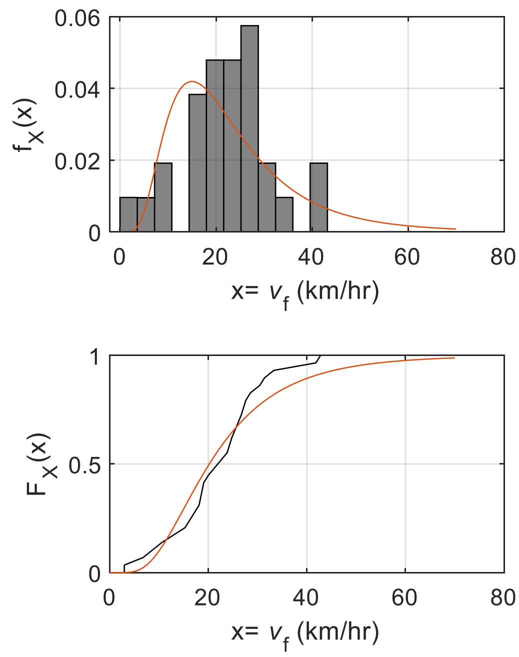


(c) (d)

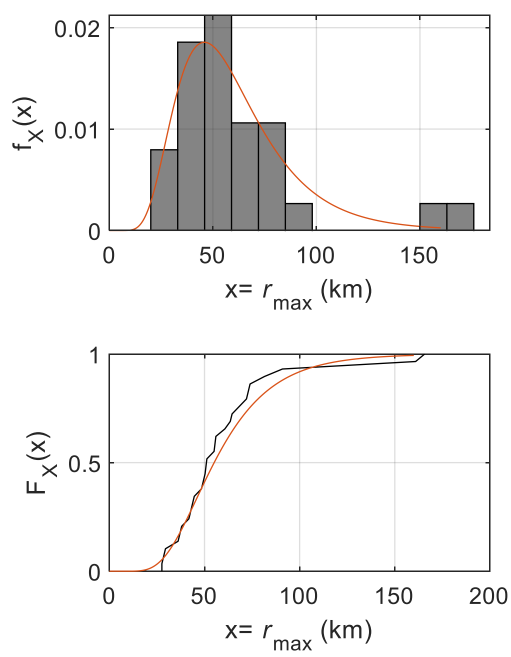
Figure C-2: Marginal distributions fitted for MI TC atmospheric parameters



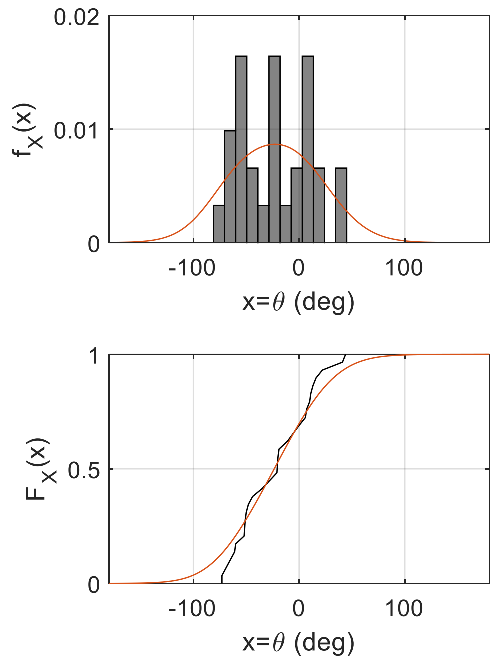
(a)



(b)



(c)



(d)

Figure C-3:Marginal distributions fitted for HI TC atmospheric parameters

Table C-1: Marginal distribution parameters

Variable	Name of distribution	Truncate	Distribution parameters	
ΔP	Weibull	LI：[8,28]		$a = 25.79$
		MI：[28,48]		$b = 1.197$
		HI：[48,148]		
V_{f}	Lognormal		LI：	$\lambda = 2.848$
				$\zeta = 0.4857$
			MI：	$\lambda = 2.970$
				$\zeta = 0.3518$
			HI：	$\lambda = 3.006$
				$\zeta = 0.5465$
$R_{\max}$	Lognormal		LI：	$\lambda = 4.307$
				$\zeta = 0.4170$
			MI：	$\lambda = 4.097$
				$\zeta = 0.3597$
			HI：	$\lambda = 4.009$
				$\zeta = 0.4276$
Θ	Scaled GKF DSRR	[-180,180]	$h_d = 200$ km	
			$h_\theta = 30$ deg	
	Scaled VKF DSRR	[-180,180]	$h_d = 200$ km	
			$\kappa = 4$	

Appendix D: Supplemental information for Chapter 4

Figure D-1 to Figure D-3 presents the fitted marginals and correlation information among TC atmospheric parameters. In each figure, the diagonals are histogram plots of statistical sampling data with fitted PDF curves (in blue) and cumulative density function (CDF) curves (in magenta). The off-diagonals are scatter plots of two TC atmospheric parameters. On top of each off-diagonal scatter plot, the Pearson's correlation coefficient ρ are reported, which are derived from the measured Kendall's τ shown in parentheses (Fang et al. 2002). The derived Pearson's ρ are used as the parameter of corresponding MGC.

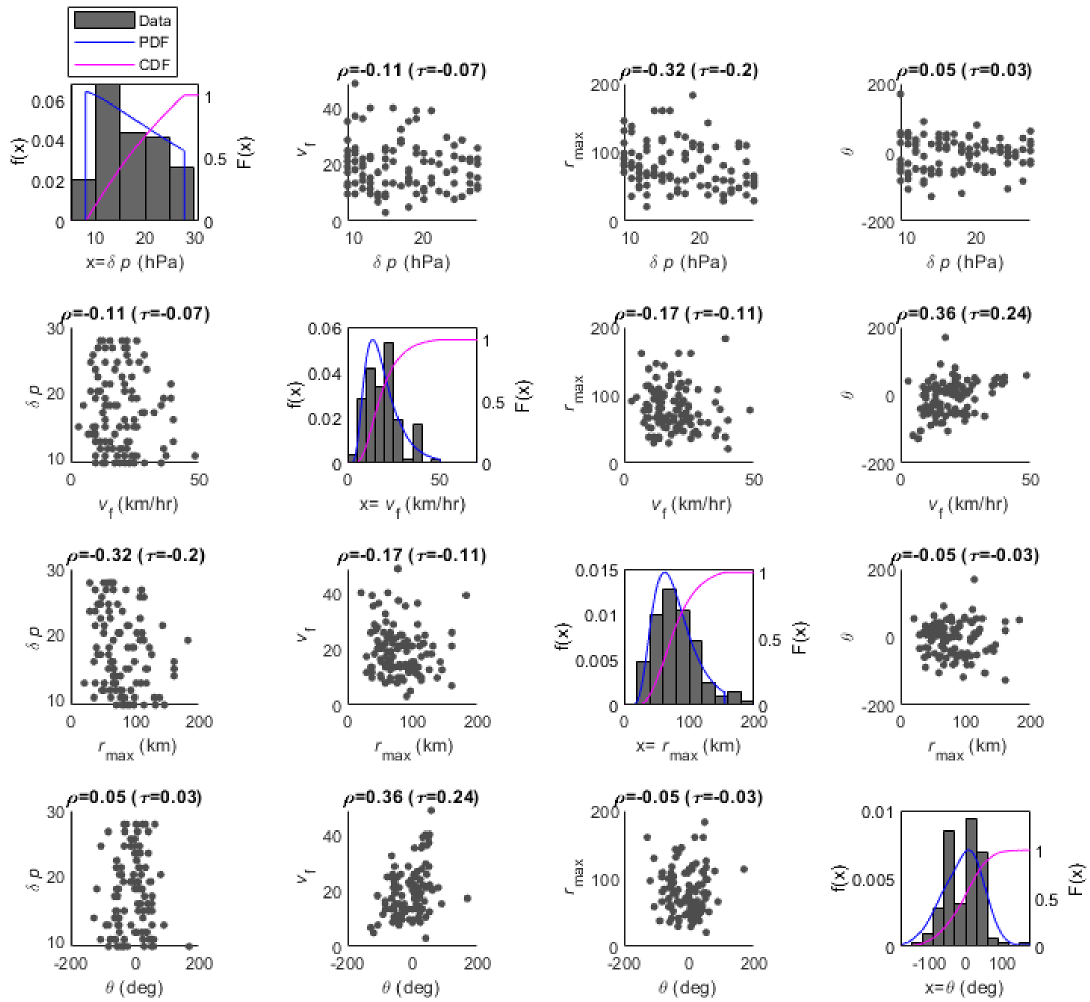


Figure D-1: Correlation information of TC atmospheric parameters for LI TCs

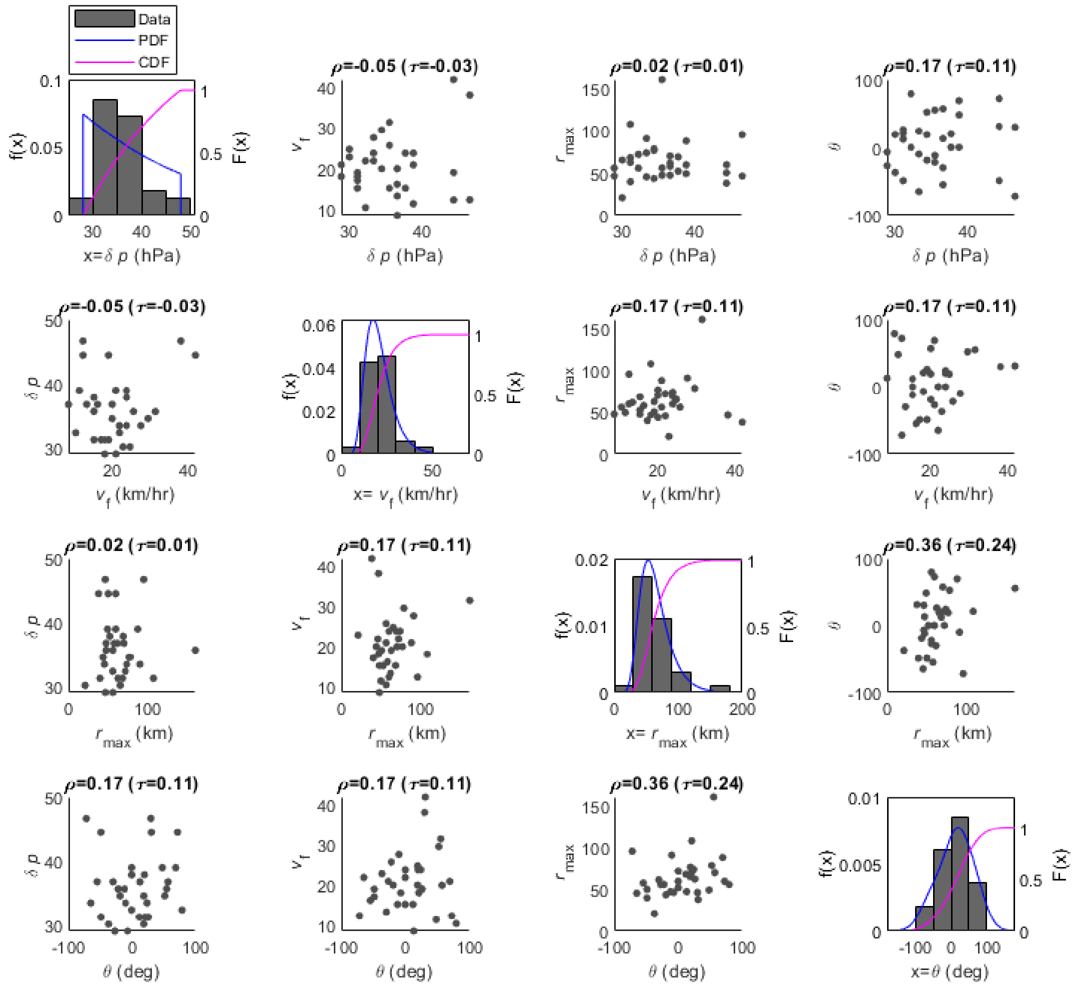


Figure D-2: Correlation information of TC atmospheric parameters for MI TCs

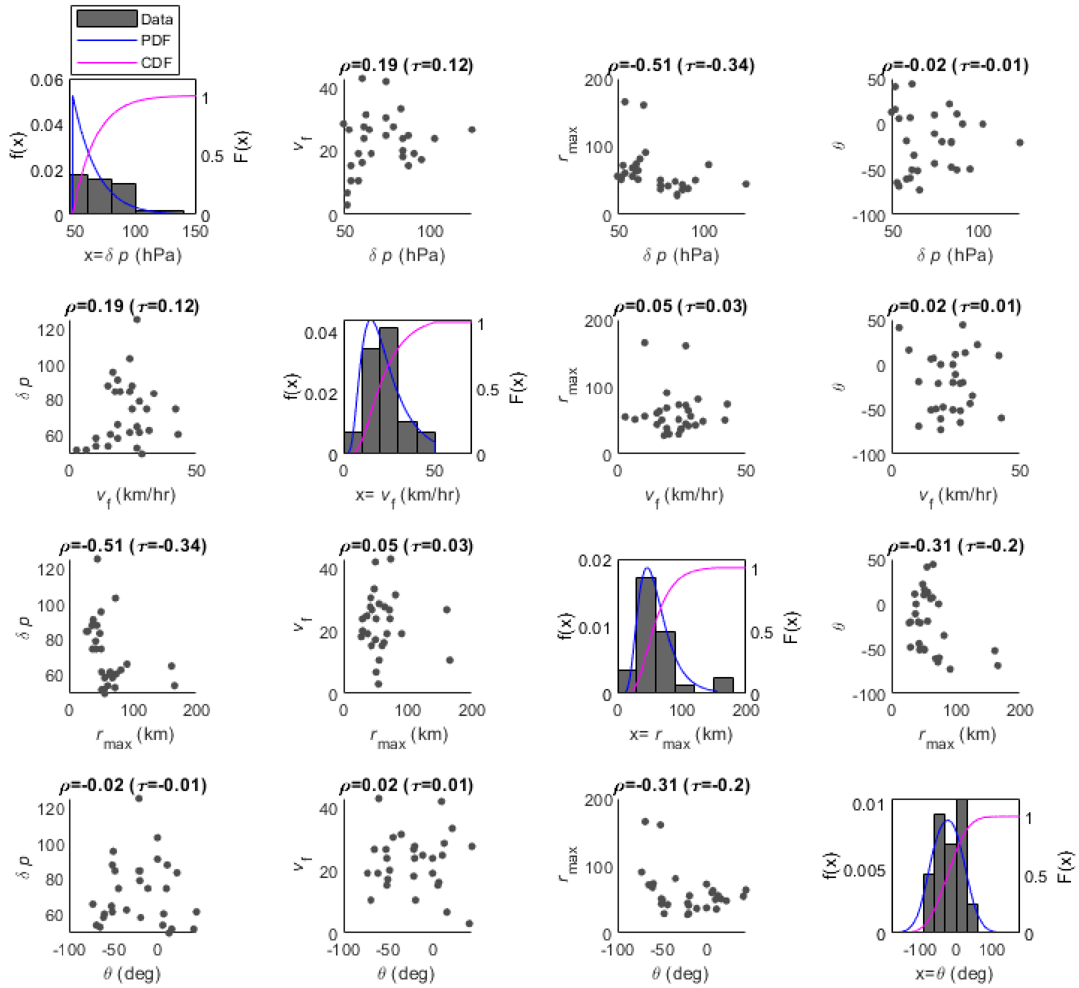


Figure D-3: Correlation information of TC atmospheric parameters for HI TCs

Table D-1: The Discretization information of BN nodes

NO.	Variable	Discretized values
1	ΔP (hPa)	8, 18, 28, 38, 48, 58, 68, 78, 88, 98, 108, 118, 128, 138
2	V_f (km/hr)	5, 10, 15, 20, 25, 30, 35, 40, 45

3	$R_{\max}$ (km)	10, 26.11, 42.22, 58.33, 74.44, 90.56, 106.67, 122.78, 138.89
4	$\hat{p}$ (mm)	Node 1 and Node 2: 0, 6.25, 12.5, …, 243.75 (40 bins with equivalent interval)
5	p (mm)	Node 1 and Node 2: 0,3.125, 6.25, …, 246.875 (80 bins with equivalent intervals)
6	$\hat{\eta}$ (m)	Node 1: -0.5, -0.225, 0.05, …, 10.225 (40 bins with equivalent intervals) Node 2: -0.1, 0.065, 0.23, …, 6.335 (40 bins with equivalent intervals)
7	η (m)	Node 1: -0.5, -0.3625, 0.225, …, 10.3625 (80 bins with equivalent intervals) Node 2: -0.1,-0.0175, 0.065…, 6.4175 (80 bins with equivalent intervals)
8	ε_η	-∞,-3,-2,-1,0,1,2,3
9	ε_p	-∞,-3,-2,-1,0,1,2,3
10	Θ (deg)	-80, -60, -40, -20, 0, 20, 40, 60
11	X_0	Projected landfall locations as shown in **Figure D-4**

Note: The discretized values presented in rows 1 to 9 provide the lower edge (value) of each bin. The values in row 1 provide the landfall location. In row 10, eight values of Θ are used for TC tracks.

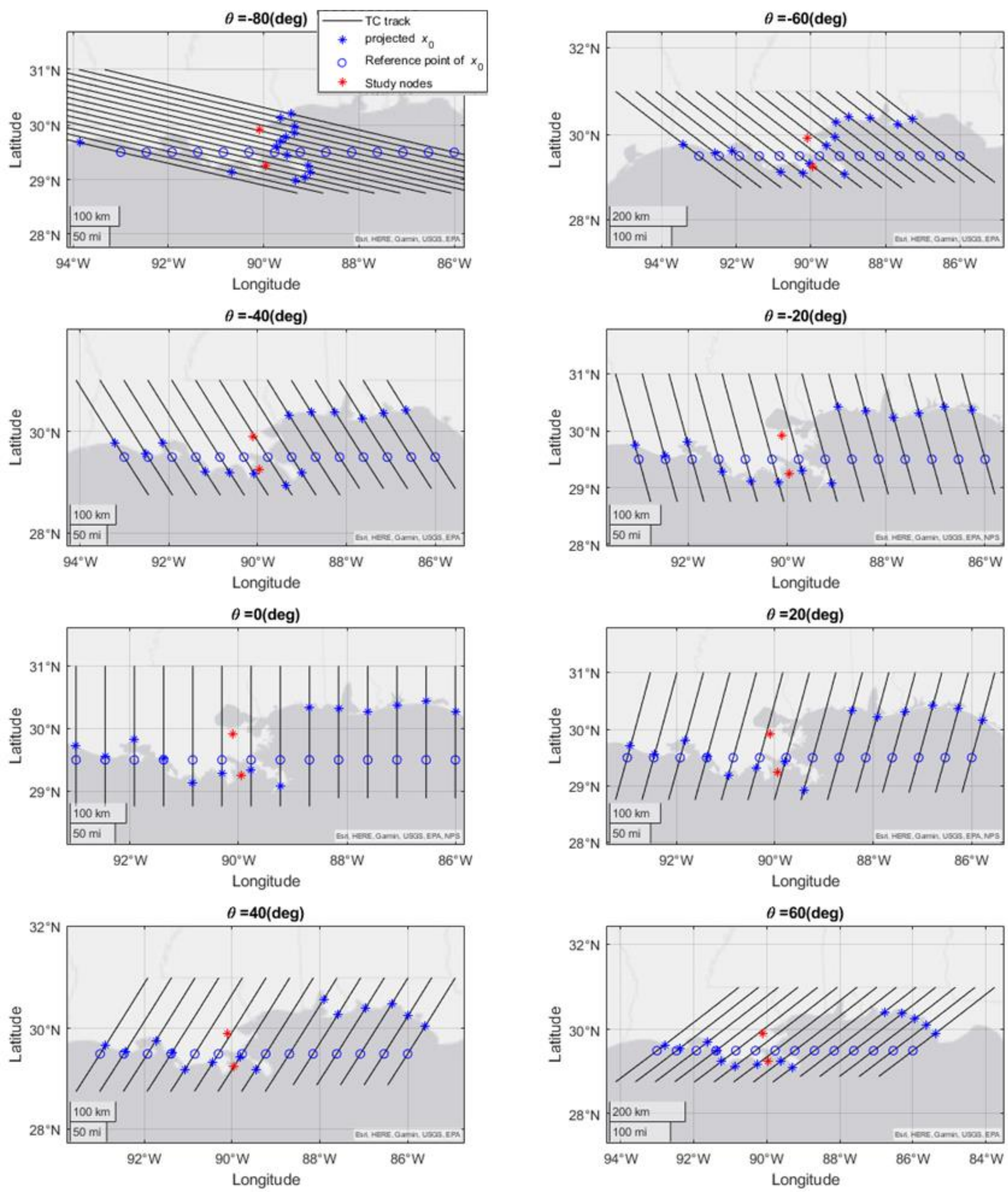


Figure D-4: Projected landfall locations

Appendix E: Supplemental information for Chapter 5

Table E-1 The Discretization information of Tier-1 BN nodes

NO.	Node	Discretized values
1	ΔP (hPa)	8, 28, 48, 68, 88, 108, 128
2	V_{f} (km/hr)	0,11.67, 23.33, 35, 46.67, 58.33
3	$R_{\max}$ (km)	8, 40, 72, 104, 136, 168
4	Q (kcfs)	0, 150, 300,… 1350 (equal interval)
5	$\hat{Z}$ (m)	-0.5, -0.23, 0.05, …, 10.23 (equal interval)
6	Z (m)	-0.5, -0.23, 0.05, …, 10.23 (equal interval)
7	ε_z	-∞,-3,-2,-1,0,1,2,3
8	Z_{t} (m)	-0.5, -0.23, 0.05, …, 10.23 (equal interval)
9	T	-∞,-3,-2,-1,0,1,2,3
10	X_0	See Figure D-4
11	Θ (deg)	-80, -60, -40, -20, 0, 20, 40, 60
12	S	Season 1, Season 2

Note: The discretized values presented in rows 1 to 10 provide the lower edge (value) of each bin.

Table E-2 The Discretization information of Tier-2 BN nodes

NO.	Node	Discretized values
1	I	LI, MI, HI

Note: Nodes not mentioned use the same discretized information as Tier-1 BN.

Table E-3 The Discretization information of Tier-3 BN nodes

NO.	Node	Discretized values
1	C ($°C$)	-0.05, 0.05, 0.15, 0.25, 0.35, 0.45
2	L (m)	-0.44, 0.44, 1.32, 2.2, 3.08, 3.96
3	P_{a} (mm)	0, 41.67, 83.33, 125, 166.67, 208.33
4	Z_{l} (m)	-0.5, -0.13, 0.25,… 14.13 (equal interval)
5	S_{a}	Season 1, Season 2

Note: The discretized values presented in rows 1 to 4 provide the lower edge (value) of each bin.

Nodes not mentioned use the same discretized information as Tier-1 BN and Tier-2 BN.

Bibliography

Al Kajbaf, A., and M. Bensi. 2020. "Application of surrogate models in estimation of storm surge:A comparative assessment." *Applied Soft Computing*, 91: 106184. https://doi.org/10.1016/j.asoc.2020.106184.

Baker, J., B. Bradley, and P. Stafford. 2021. *Seismic Hazard and Risk Analysis*. Cambridge: Cambridge University Press.

Baker, J. W. 2013. *An introduction to probabilistic seismic hazard analysis*.

BayesFusion. 2023. "GeNIe Modeler USER MANUAL."

Bedford, T., and R. M. Cooke. 2002. "Vines--a new graphical model for dependent random variables." *The Annals of Statistics*, 30 (4): 1031–1068. Institute of Mathematical Statistics. https://doi.org/10.1214/aos/1031689016.

Bennett, B., M. Leonard, Y. Deng, and S. Westra. 2018. "An empirical investigation into the effect of antecedent precipitation on flood volume." *Journal of Hydrology*, 567: 435–445. https://doi.org/10.1016/j.jhydrol.2018.10.025.

Bensi, M., S. Mohammadi, S.-C. Kao, and S. T. DeNeale. 2020. *Multi-Mechanism Flood Hazard Assessment: Critical Review of Current Practice and Approaches*. Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States).

Bensi, M. T. 2010. "A Bayesian Network Methodology for Infrastructure Seismic Risk Assessment and Decision Support." UC Berkeley.

Brackins, J. T., and A. J. Kalyanapu. 2020. "Evaluation of parametric precipitation models in reproducing tropical cyclone rainfall patterns." *Journal of Hydrology*, 580: 124255. https://doi.org/10.1016/j.jhydrol.2019.124255.

Carta, J. A., P. Ramírez, and C. Bueno. 2008. "A joint probability density function of wind speed and direction for wind energy analysis." *Energy Conversion and Management*, 49 (6): 1309–1320. https://doi.org/10.1016/j.enconman.2008.01.010.

Chavas, D. R., and J. A. Knaff. 2022. "A simple model for predicting the hurricane radius of maximum wind from outer size." *arXiv:2106.11809 [physics]*. https://doi.org/10.1175/WAF-D-21-0103.1.

Chavas, D. R., N. Lin, and K. Emanuel. 2015. "A Model for the Complete Radial Structure of the Tropical Cyclone Wind Field. Part I: Comparison with Observed Structure." *Journal of the Atmospheric Sciences*, 72 (9): 3647–3662. American Meteorological Society. https://doi.org/10.1175/JAS-D-15-0014.1.

Chen, S. S., J. A. Knaff, and F. D. Marks. 2006. "Effects of Vertical Wind Shear and Storm Motion on Tropical Cyclone Rainfall Asymmetries Deduced from TRMM." *Monthly Weather Review*, 134 (11): 3190–3208. American Meteorological Society. https://doi.org/10.1175/MWR3245.1.

Chouinard, L. E., and C. Liu. 1997. “Model for Recurrence Rate of Hurricanes in Gulf of Mexico.” *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 123 (3): 113–119. https://doi.org/10.1061/(ASCE)0733-950X(1997)123:3(113).

Cialone, M. A., L. M. Bocamazo, N. C. Nadal-Caraballo, T. C. Massey, and J. A. Melby. 2015. “North atlantic coast comprehensive study (naccs) storm simulation and statistical analysis: part i ? overview.” *Coastal Sediments 2015*. WORLD SCIENTIFIC.

CIRA. 2021. “The Tropical Cyclone Extended Best Track Dataset (EBTRK).” *Regional and Mesoscale Meteorology Branch*. Accessed March 28, 2022.

Couasnon, A., A. Sebastian, and O. Morales-Nápoles. 2018. “A Copula-Based Bayesian Network for Modeling Compound Flood Hazard from Riverine and Coastal Interactions at the Catchment Scale: An Application to the Houston Ship Channel, Texas.” *Water*, 10 (9): 1190. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/w10091190.

Dawid, A. P. 1992. “Applications of a general propagation algorithm for probabilistic expert systems.” *Stat Comput*, 2 (1): 25–36. https://doi.org/10.1007/BF01890546.

Demuth, J. L., M. DeMaria, and J. A. Knaff. 2006a. “Improvement of Advanced Microwave Sounding Unit Tropical Cyclone Intensity and Size Estimation Algorithms.” *Journal of Applied Meteorology and Climatology*, 45 (11): 1573–1581. American Meteorological Society. https://doi.org/10.1175/JAM2429.1.

Demuth, J. L., M. DeMaria, and J. A. Knaff. 2006b. “Improvement of Advanced Microwave Sounding Unit Tropical Cyclone Intensity and Size Estimation Algorithms.” *Journal of Applied Meteorology and Climatology*, 45 (11): 1573–1581. American Meteorological Society. https://doi.org/10.1175/JAM2429.1.

Dietrich, J. C., S. Tanaka, J. J. Westerink, C. N. Dawson, R. A. Luettich, M. Zijlema, L. H. Holthuijsen, J. M. Smith, L. G. Westerink, and H. J. Westerink. 2012. “Performance of the Unstructured-Mesh, SWAN+ADCIRC Model in Computing Hurricane Waves and Surge.” *J Sci Comput*, 52 (2): 468–497. https://doi.org/10.1007/s10915-011-9555-6.

Donnelly, C., and P. Embrechts. 2010. “The Devil is in the Tails: Actuarial Mathematics and the Subprime Mortgage Crisis.” *ASTIN Bulletin: The Journal of the IAA*, 40 (1): 1–33. Cambridge University Press. https://doi.org/10.2143/AST.40.1.2049222.

Emanuel, K., C. DesAutels, C. Holloway, and R. Korty. 2004. “Environmental Control of Tropical Cyclone Intensity.” *Journal of the Atmospheric Sciences*, 61 (7): 843–858. American Meteorological Society. https://doi.org/10.1175/1520-0469(2004)061<0843:ECOTCI>2.0.CO;2.

ERDC. 2023. “Coastal Hazards System.” Accessed April 20, 2023. https://chs.erdc.dren.mil/.

Fang, H.-B., K.-T. Fang, and S. Kotz. 2002. "The Meta-elliptical Distributions with Given Marginals." *Journal of Multivariate Analysis*, 82 (1): 1–16. https://doi.org/10.1006/jmva.2001.2017.

FEMA. 2009. *Flood Insurance Study: Commonwealth of Puerto Rico and Municipalities*.

FEMA. 2012. *Joint Probability – Optimal Sampling Method for Tropical Storm Surge Frequency Analysis*. Federal Emergency Management Agency.

Flores, M. J., R. F. Ropero, and R. Rumí. 2019. "Assessment of flood risk in Mediterranean catchments: an approach based on Bayesian networks." *Stoch Environ Res Risk Assess*, 33 (11): 1991–2005. https://doi.org/10.1007/s00477-019-01746-3.

Gallien, T. W., N. Kalligeris, M.-P. C. Delisle, B.-X. Tang, J. T. D. Lucey, and M. A. Winters. 2018. "Coastal Flood Modeling Challenges in Defended Urban Backshores." *Geosciences*, 8 (12): 450. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/geosciences8120450.

García-Portugués, E., R. M. Crujeiras, and W. González-Manteiga. 2013. "Exploring wind direction and SO2 concentration by circular–linear density estimation." *Stoch Environ Res Risk Assess*, 27 (5): 1055–1067. https://doi.org/10.1007/s00477-012-0642-5.

Ghahramani, Z. 1998. "Learning dynamic Bayesian networks." *Adaptive Processing of Sequences and Data Structures: International Summer School on Neural Networks "E.R. Caianiello" Vietri sul Mare, Salerno, Italy September 6–13, 1997 Tutorial Lectures*, Lecture Notes in Computer Science, C. L. Giles and M. Gori, eds., 168–197. Berlin, Heidelberg: Springer.

Gonçalves, M., and E. R. & C. G. Soares. 2012. "Evaluation of the wave models SWAN and STWAVE in shallow water using nested schemes." *Maritime Engineering and Technology*. CRC Press.

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*.

Goodbell, C. 2014. *Breaking the Hec-Ras Code: A User's Guide to Automating Hec-Ras*.

Gray, W. M. 1968. "GLOBAL VIEW OF THE ORIGIN OF TROPICAL DISTURBANCES AND STORMS." *Monthly Weather Review*, 96 (10): 669–700. American Meteorological Society. https://doi.org/10.1175/1520-0493(1968)096<0669:GVOTOO>2.0.CO;2.

Grieser, J., and S. Jewson. 2011. "The RMS TC-Rain Model."

Grillakis, M. G., A. G. Koutroulis, J. Komma, I. K. Tsanis, W. Wagner, and G. Blöschl. 2016. "Initial soil moisture effects on flash flood generation – A comparison between basins of contrasting hydro-climatic conditions." *Journal of Hydrology*, Flash floods, hydro-geomorphic response and risk management, 541: 206–217. https://doi.org/10.1016/j.jhydrol.2016.03.007.

Gutmann, E. D., R. M. Rasmussen, C. Liu, K. Ikeda, C. L. Bruyere, J. M. Done, L. Garrè, P. Friis-Hansen, and V. Veldore. 2018. "Changes in Hurricanes from a 13-Yr Convection-Permitting Pseudo–Global Warming Simulation." *Journal of Climate*, 31 (9): 3643–3657. American Meteorological Society. https://doi.org/10.1175/JCLI-D-17-0391.1.

Ho, F. P., and V. A. Myers. 1975. *Joint Probability Method of Tide Frequency Analysis Applied to Apalachicola Bay and St. George Sound, Florida*.

IPET. 2009. *Performance evaluation of the New Orleans and southeast Louisiana hurricane protection system final report of the Interagency Performance Evaluation Task Force*. US Army Corps of Engineers.

Irish, J. L., D. T. Resio, and M. A. Cialone. 2009. "A surge response function approach to coastal hazard assessment. Part 2: Quantification of spatial attributes of response functions." *Nat Hazards*, 51 (1): 183–205. https://doi.org/10.1007/s11069-009-9381-4.

Jane, R., L. Dalla Valle, D. Simmonds, and A. Raby. 2016. "A copula-based approach for the estimation of wave height records through spatial correlation." *Coastal Engineering*, 117: 1–18. https://doi.org/10.1016/j.coastaleng.2016.06.008.

Jaworski, P., F. Durante, W. K. Hardle, and T. Rychlik. 2010. *Copula theory and its applications*. Springer.

Jelesnianski, C. P. 1992. *SLOSH: Sea, Lake, and Overland Surges from Hurricanes*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Weather Service.

Jensen, F. 1990. "Bayesian updating in recursive graphical models by local comutations." *Computational Statistics and Data Analysis*, 4: 269–282.

Jia, G., and A. A. Taflanidis. 2013. "Kriging metamodeling for approximation of high-dimensional wave and surge responses in real-time storm/hurricane risk assessment." *Computer Methods in Applied Mechanics and Engineering*, 261: 24–38. Elsevier.

Jia, G., 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. Springer.

Joe, H. 1997. *Multivariate Models and Multivariate Dependence Concepts*. CRC Press.

Johnson, R. A., and T. E. Wehrly. 1978. "Some Angular-Linear Distributions and Related Regression Models." *Journal of the American Statistical Association*, 73 (363): 602–606. Taylor & Francis. https://doi.org/10.1080/01621459.1978.10480062.

Joyce, J., N.-B. Chang, R. Harji, T. Ruppert, and P. Singhofen. 2018. "Cascade impact of hurricane movement, storm tidal surge, sea level rise and precipitation variability on flood assessment in a coastal urban watershed." *Clim Dyn*, 51 (1): 383–409. https://doi.org/10.1007/s00382-017-3930-4.

Kang, N.-Y., and J. B. Elsner. 2015. "Trade-off between intensity and frequency of global tropical cyclones." *Nature Clim Change*, 5 (7): 661–664. https://doi.org/10.1038/nclimate2646.

Kirshen, P., C. Watson, E. Douglas, A. Gontz, J. Lee, and Y. Tian. 2008. "Coastal flooding in the Northeastern United States due to climate change." *Mitig Adapt Strateg Glob Change*, 13 (5): 437–451. https://doi.org/10.1007/s11027-007-9130-5.

Kjaerulff, U. B., and A. L. Madsen. 2008. *Bayesian networks and influence diagrams*. Springer.

Knutson, T., S. J. Camargo, J. C. L. Chan, K. Emanuel, C.-H. Ho, J. Kossin, M. Mohapatra, M. Satoh, M. Sugi, K. Walsh, and L. Wu. 2019. "Tropical Cyclones and Climate Change Assessment: Part I: Detection and Attribution." *Bulletin of the American Meteorological Society*, 100 (10): 1987–2007. American Meteorological Society. https://doi.org/10.1175/BAMS-D-18-0189.1.

Kyprioti, A. P., A. A. Taflanidis, N. C. Nadal-Caraballo, and M. O. Campbell. 2021. "Incorporation of sea level rise in storm surge surrogate modeling." *Nat Hazards*, 105 (1): 531–563. https://doi.org/10.1007/s11069-020-04322-z.

Landsea, C. W., and J. L. Franklin. 2013. "Atlantic Hurricane Database Uncertainty and Presentation of a New Database Format." *Monthly Weather Review*, 141 (10): 3576–3592. American Meteorological Society. https://doi.org/10.1175/MWR-D-12-00254.1.

Lauritzen, S. L., and D. J. Spiegelhalter. 1988. "Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems." *Journal of the Royal Statistical Society: Series B (Methodological)*, 50 (2): 157–194. https://doi.org/10.1111/j.2517-6161.1988.tb01721.x.

Lee, J.-W., J. L. Irish, M. T. Bensi, and D. C. Marcy. 2021. "Rapid prediction of peak storm surge from tropical cyclone track time series using machine learning." *Coastal Engineering*, 170: 104024. https://doi.org/10.1016/j.coastaleng.2021.104024.

Leon, A. S., and C. Goodell. 2016. "Controlling HEC-RAS using MATLAB." *Environmental Modelling & Software*, 84: 339–348. https://doi.org/10.1016/j.envsoft.2016.06.026.

Levenberg, K. 1944. "A method for the solution of certain non-linear problems in least squares." *Quart. Appl. Math.*, 2 (2): 164–168. https://doi.org/10.1090/qam/10666.

Lin, Y., and S. Dong. 2019. “Wave energy assessment based on trivariate distribution of significant wave height, mean period and direction.” *Applied Ocean Research*, 87: 47–63. https://doi.org/10.1016/j.apor.2019.03.017.

Liu, Z., M. L. Carr, N. C. Nadal-Caraballo, L. A. Aucoin, M. C. Yawn, and M. Bensi. 2023a. “Comparative Analysis of Joint Distribution Models for Tropical Cyclone Atmospheric Parameters in Probabilistic Coastal Hazard Analysis.” *Stochastic Environmental Research and Risk Assessment (under review)*.

Liu, Z., M. L. Carr, N. C. Nadal-Caraballo, M. C. Yawn, A. A. Taflanidis, and M. Bensi. 2023b. “Machine Learning Motivated Data Imputation of Storm Data Used in Coastal Hazard Assessments.” *Coastal Engineering (under review)*.

Lonfat, M., F. D. Marks, and S. S. Chen. 2004. “Precipitation Distribution in Tropical Cyclones Using the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager: A Global Perspective.” *Monthly Weather Review*, 132 (7): 1645–1660. American Meteorological Society. https://doi.org/10.1175/1520-0493(2004)132<1645:PDITCU>2.0.CO;2.

Lu, P., N. Lin, K. Emanuel, D. Chavas, and J. Smith. 2018. “Assessing Hurricane Rainfall Mechanisms Using a Physics-Based Model: Hurricanes Isabel (2003) and Irene (2011).” *Journal of the Atmospheric Sciences*, 75 (7): 2337–2358. American Meteorological Society. https://doi.org/10.1175/JAS-D-17-0264.1.

Marks, F. D., and M. DeMaria. 2003. *Development of a Tropical Cyclone Rainfall Climatology and Persistence (R-CLIPER) Model*.

Marquardt, D. W. 1963. “An Algorithm for Least-Squares Estimation of Nonlinear Parameters.” *Journal of the Society for Industrial and Applied Mathematics*, 11 (2): 431–441. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/0111030.

MathWorks. 2022. “MATLAB, Release 2022a.”

Melby, J. A., F. Diop, N. C. Nadal-Caraballo, D. Green, and V. Gonzalez. 2017. “Coastal Hazards System.” 219–229. American Society of Civil Engineers. https://doi.org/10.1061/9780784480304.023.

Michele, C. D., and G. Salvadori. 2002. “On the derived flood frequency distribution: analytical formulation and the influence of antecedent soil moisture condition.” *Journal of Hydrology*, 262 (1): 245–258. https://doi.org/10.1016/S0022-1694(02)00025-2.

Mohammadi, S., M. T. Bensi, S.-C. Kao, and S. T. DeNeale. 2021. *Multi-Mechanism Flood Hazard Assessment: Example Use Case Studies*. Oak Ridge National Lab.

Mohammadi, S., M. T. Bensi, S.-C. Kao, S. T. DeNeale, J. Kanney, E. Yegorova, and M. L. Carr. 2023. “Bayesian-Motivated Probabilistic Model of Hurricane-Induced Multimechanism Flood Hazards.” *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 149 (4): 04023007. American Society of Civil Engineers. https://doi.org/10.1061/JWPED5.WWENG-1921.

Myers, V. A. 1954. *Characteristics of United States Hurricanes Pertinent to Levee Design for Lake Okeechobee, Florida*. U.S. Government Printing Office.

Myers, V. A. 1970. “Joint Probability Method of Tide Frequency Analysis Applied to Atlantic City and Long Beach Island.”

Nadal-Caraballo, N. C., M. O. Campbell, V. M. Gonzalez, M. J. Torres, J. A. Melby, and A. A. Taflanidis. 2020a. “Coastal Hazards System: A Probabilistic Coastal Hazard Analysis Framework.” *Journal of Coastal Research*, 95 (SI): 1211–1216. https://doi.org/10.2112/SI95-235.1.

Nadal-Caraballo, N. C., M. O. Campbell, V. M. Gonzalez, M. J. Torres, J. A. Melby, and A. A. Taflanidis. 2020b. “Coastal Hazards System: A Probabilistic Coastal Hazard Analysis Framework.” *Journal of Coastal Research*, 95 (SI): 1211–1216. https://doi.org/10.2112/SI95-235.1.

Nadal-Caraballo, N. C., V. M. Gonzalez, and L. Chouinard. 2019. *Storm Recurrence Rate Models for Tropical Cyclones: Report 1of a series on the quantification of uncertainties in probabilistic storm surge models*. Vicksburg, MS: US Army Engineer Research and Development Center.

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*. Vicksburg, MS: US Army Engineer Research and Development Center.

Nadal-Caraballo, N. C., M. C. Yawn, L. A. Aucoin, M. L. Carr, J. A. Melby, E. Ramos-Santiago, F. A. Garcia-Moreno, V. M. Gonzalez, T. C. Massey, M. B. Owensby, A. A. Taflanidis, A. P. Kyprioti, A. T. Cox, and J. Gonzalez-Lopez. 2022a. *Coastal Hazards System – Puerto Rico (CHS-PR)*. Vicksburg, MS: US Army Engineer Research and Development Center.

Nadal-Caraballo, N. C., M. C. Yawn, L. A. Aucoin, M. L. Carr, J. A. Melby, E. Ramos-Santiago, V. M. Gonzalez, A. A. Taflanidis, A. A. Kyprioti, Z. Cobell, and A. T. Cox. 2022b. *Coastal Hazards System–Louisiana (CHS-LA)*. Vicksburg, MS: US Army Engineer Research and Development Center.

Nelsen, R. B. 2007. *An introduction to copulas*. Springer Science & Business Media.

NOAA. 2022a. “Hurricane Costs.” https://coast.noaa.gov/states/fast-facts/hurricane-costs.html.

NOAA. 2022b. “National Centers for Environmental Information (NCEI) U.S. Billion-Dollar Weather and Climate Disasters.” Accessed July 25, 2022. https://www.ncdc.noaa.gov/billions/.

NOAA. 2022c. “Re-Analysis Project Hurricane Database.” https://www.aoml.noaa.gov/hrd/hurdat/Data_Storm.html.

Palmen, E. 1948. “On the formation and structure of tropical hurricanes.” *Geophysica*, 3(1): 26–38.

PRISM. 2023. “PRISM Climate Data.” *PRISM Climate Data*. https://prism.oregonstate.edu.

Resio, D. T. 2007. “White paper on estimating hurricane inundation probabilities.” *This Digital Resource was created from scans of the Print Resource*. Coastal and Hydraulics Laboratory (U.S.).

Resio, D. T., J. Irish, and M. Cialone. 2009. “A surge response function approach to coastal hazard assessment – part 1: basic concepts.” *Nat Hazards*, 51 (1): 163–182. https://doi.org/10.1007/s11069-009-9379-y.

Roberts, M. J., J. Camp, J. Seddon, P. L. Vidale, K. Hodges, B. Vannière, J. Mecking, R. Haarsma, A. Bellucci, E. Scoccimarro, L.-P. Caron, F. Chauvin, L. Terray, S. Valcke, M.-P. Moine, D. Putrasahan, C. D. Roberts, R. Senan, C. Zarzycki, P. Ullrich, Y. Yamada, R. Mizuta, C. Kodama, D. Fu, Q. Zhang, G. Danabasoglu, N. Rosenbloom, H. Wang, and L. Wu. 2020. “Projected Future Changes in Tropical Cyclones Using the CMIP6 HighResMIP Multimodel Ensemble.” *Geophysical Research Letters*, 47 (14): e2020GL088662. https://doi.org/10.1029/2020GL088662.

Roper, D. 2016. “Sea Level Versus Temperature.” http://roperld.com/personal/RoperLDavid.htm.

Russell, L. R. 1969. *Probability distributions for Texas Gulf coast hurricane effects of engineering interest*. Stanford University.

Sanuy, M., J. A. Jiménez, and N. Plant. 2020. “A Bayesian Network methodology for coastal hazard assessments on a regional scale: The BN-CRAF.” *Coastal Engineering*, 157: 103627. https://doi.org/10.1016/j.coastaleng.2019.103627.

Sebastian, A., E. J. C. Dupuits, and O. Morales-Nápoles. 2017. “Applying a Bayesian network based on Gaussian copulas to model the hydraulic boundary conditions for hurricane flood risk analysis in a coastal watershed.” *Coastal Engineering*, 125: 42–50. https://doi.org/10.1016/j.coastaleng.2017.03.008.

Sen, M. K., S. Dutta, and G. Kabir. 2022. “Modelling and quantification of time-varying flood resilience for housing infrastructure using dynamic Bayesian Network.” *Journal of Cleaner Production*, 361: 132266. https://doi.org/10.1016/j.jclepro.2022.132266.

Sklar, M. 1959. “Fonctions de repartition an dimensions et leurs marges.” *Publ. inst. statist. univ. Paris*, 8: 229–231.

Taflanidis, A. A., G. Jia, N. C. Nadal-Caraballo, A. B. Kennedy, J. A. Melby, and J. M. Smith. 2014. “Development of Real-Time Tools for Hurricane Risk Assessment.” *Vulnerability, Uncertainty, and Risk: Quantification, Mitigation, and Management*, 1341–1350. American Society of Civil Engineers. https://doi.org/10.1061/9780784413609.135.

Taylor, C. C. 2008. “Automatic bandwidth selection for circular density estimation.” *Computational Statistics & Data Analysis*, 52 (7): 3493–3500. Elsevier.

Toro, G. R. 2008. *Joint Probability Analysis of Hurricane Flood Hazards for Mississippi, Final report in support of the FEMA-HMTAP flood study of the State of Mississippi (Rev. 1).*

Toro, G. R., A. W. Niedoroda, C. W. Reed, and D. Divoky. 2010. "Quadrature-based approach for the efficient evaluation of surge hazard." *Ocean Engineering*, A Forensic Analysis of Hurricane Katrina's Impact: Methods and Findings, 37 (1): 114–124. https://doi.org/10.1016/j.oceaneng.2009.09.005.

USACE. 2023. "HEC-RAS." Accessed February 2, 2023. https://www.hec.usace.army.mil/software/hec-ras/.

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. American Meteorological Society. https://doi.org/10.1175/2008JAMC1837.1.

Walega, A., and D. Młyński. 2017. "SEASONALITY OF MEDIAN MONTHLY DISCHARGE IN SELECTED CARPATHIAN RIVERS OF THE UPPER VISTULA BASIN." *Carpathian journal of earth and environmental sciences*, 12: 617–628.

Wang, Z., W. Zhang, Y. Zhang, and Z. Liu. 2021. "Circular-linear-linear probabilistic model based on vine copulas: An application to the joint distribution of wind direction, wind speed, and air temperature." *Journal of Wind Engineering and Industrial Aerodynamics*, 215: 104704. https://doi.org/10.1016/j.jweia.2021.104704.

Webster, P. J., G. J. Holland, J. A. Curry, and H.-R. Chang. 2005. "Changes in Tropical Cyclone Number, Duration, and Intensity in a Warming Environment." *Science*, 309 (5742): 1844–1846. American Association for the Advancement of Science. https://doi.org/10.1126/science.1116448.

Wu, L., H. Zhao, C. Wang, J. Cao, and J. Liang. 2022. "Understanding of the Effect of Climate Change on Tropical Cyclone Intensity: A Review." *Adv. Atmos. Sci.*, 39 (2): 205–221. https://doi.org/10.1007/s00376-021-1026-x.

Wu, Z., Y. Shen, H. Wang, and M. Wu. 2020. "Urban flood disaster risk evaluation based on ontology and Bayesian Network." *Journal of Hydrology*, 583: 124596. https://doi.org/10.1016/j.jhydrol.2020.124596.

Xi, D., N. Lin, and J. Smith. 2020. "Evaluation of a Physics-Based Tropical Cyclone Rainfall Model for Risk Assessment." *Journal of Hydrometeorology*, 21 (9): 2197–2218. American Meteorological Society. https://doi.org/10.1175/JHM-D-20-0035.1.

Xu, K., C. Wang, and L. Bin. 2022. "Compound flood models in coastal areas: a review of methods and uncertainty analysis." *Nat Hazards*. https://doi.org/10.1007/s11069-022-05683-3.

Zar, J. H. 1999. *Biostatistical analysis*. Pearson Education India.

Zhang, J., A. A. Taflanidis, N. C. Nadal-Caraballo, J. A. Melby, and F. Diop. 2018. "Advances in surrogate modeling for storm surge prediction: storm selection and addressing characteristics related to climate change." *Nat Hazards*, 94 (3): 1225–1253. https://doi.org/10.1007/s11069-018-3470-1.

Zhang, L., and V. P. Singh. 2019. *Copulas and Their Applications in Water Resources Engineering*. Cambridge University Press.

Zhang, Y., and M. R. Najafi. 2020. "Probabilistic Numerical Modeling of Compound Flooding Caused by Tropical Storm Matthew Over a Data-Scarce Coastal Environment." *Water Resources Research*, 56 (10): e2020WR028565. https://doi.org/10.1029/2020WR028565.

Zhu, L., S. M. Quiring, and K. A. Emanuel. 2013. "Estimating tropical cyclone precipitation risk in Texas." *Geophysical Research Letters*, 40 (23): 6225–6230. https://doi.org/10.1002/2013GL058284.