JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD033557

Probabilistic Characterization of Extreme Storm Surges Induced by Tropical Cyclones

B. Zhang[1] and S. Wang[1,2]

[1]Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China, [2]The Hong Kong Polytechnic University Shenzhen Research Institute, Shenzhen, China

Key Points:

- A robust data-driven framework was proposed to characterize extreme storm surges induced by tropical cyclones
- The dependence structure of multiple mechanisms causing extreme storm surges was uncovered through an ensemble multi-structure vine copula simulation
- The framework demonstrated higher skill than existing vine- and regression-based models in extreme storm surge characterization

Supporting Information:

- Supporting Information S1

Correspondence to:
S. Wang,
shuo.s.wang@polyu.edu.hk

Citation:
Zhang, B., & Wang, S. (2021). Probabilistic characterization of extreme storm surges induced by tropical cyclones. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD033557. https://doi.org/10.1029/2020JD033557

Received 20 JUL 2020
Accepted 21 DEC 2020

Author Contributions:

Conceptualization: S. Wang
Data curation: B. Zhang
Formal analysis: B. Zhang
Funding acquisition: S. Wang
Investigation: S. Wang
Methodology: B. Zhang, S. Wang
Project Administration: S. Wang
Resources: S. Wang
Software: B. Zhang
Supervision: S. Wang
Validation: B. Zhang
Writing – original draft: B. Zhang
Writing – review & editing: S. Wang

Abstract The overtopping of flood defenses by extreme storm surges during tropical cyclones poses a significant threat to life and property in coastal and estuarine regions. Since little effort has been devoted to investigating the complex interaction of multiple mechanisms causing extreme storm surges, for the first time, we propose a robust data-driven framework to explicitly uncover the complex interaction and thus to improve the characterization of extreme storm surges induced by tropical cyclones. The framework constructs a probabilistic ensemble of dependence structures of multiple climatological forcing factors that potentially cause extreme storm surges based on a multi-structure regular vine copula approach. The uncertainty in the vine-based model structure is explicitly addressed in a Bayesian framework. The climatological forcing factors sensitive to extreme storm surge levels are selected using partial correlations, including 10-m wind, 850-mb temperature, 700-mb geopotential height, precipitation, sea level pressure, and its spatial gradient extracted from the ERA5 reanalysis data. We demonstrate the framework by an in-depth analysis of the extreme storm surge levels observed over two tidal gauging stations in Hong Kong during 1979–2018. Our findings show that the proposed framework substantially improves the characterization of extreme storm surges by taking into account the joint evolution of multiple mechanisms causing extreme storm surges and underlying uncertainties. Furthermore, the framework not only demonstrates higher skill than previous single-structure vine-based models but also can outperform the principal components regression and the random forest regression in terms of characterizing extreme storm surges.

1. Introduction

Storm surge is a type of coastal hazards characterized by an abnormal rise of seawater level over the estimated high tide that is caused by low-pressure systems such as tropical cyclones (Bevacqua et al., 2019; Cid et al., 2016; Howard et al., 2010; Zheng et al., 2014). Extreme storm surge has been considered more of a threat to coastal communities and infrastructure than strong winds from a hurricane, as it often leads to coastal inundations and can cause substantial loss of life and economic damage (Grinsted et al., 2013; MacPherson et al., 2019; Marcos et al., 2019; Morrow et al., 2015; Wahl et al., 2015). For example, Typhoon Mangkhut in 2018 caused record-breaking storm surges of 2–3 m and widespread inundation in low-lying coastal areas of Hong Kong. Thus, a reliable characterization of extreme storm surges induced by tropical cyclones is vital for coastal flood risk management and hazard mitigation, which has attracted increasing attention in hydroclimate and ocean science communities in recent years (Calafat & Marcos, 2020; Couasnon et al., 2020; Dube et al., 2009; Flowerdew et al., 2013; Haigh et al., 2014; Horsburgh & Wilson, 2007; Needham et al., 2015; Qing et al., 2020; Reed et al., 2015).

Storm surges during the passage of tropical cyclones have been commonly characterized by the process-driven hydrodynamic models, such as ADCIRC (Bilskie et al., 2016; Luettich et al., 1992), Delft3D (Hu et al., 2009), ROMS (Colberg & McInnes, 2012), FVCOM (Rego & Li, 2010), and HYCOM (Shriver et al., 2014). These process-driven models use the meteorological fields (e.g., wind and pressure) as forcing to simulate surge/wave conditions during the passage of tropical cyclones (Bloemendaal et al., 2019; Brown et al., 2007; Fernández-Montblanc et al., 2019; Gräwe & Burchard, 2012; Howard et al., 2010; Muis et al., 2016). For example, Vousdoukas et al. (2016) used the hydrodynamic model Delft3D-Flow driven by ERA-Interim atmospheric forcing to evaluate dynamics in storm surge level (SSL) along the European coastline. Wu et al. (2018) used ROMS forced with meteorological forcing obtained from the U.S. National Centers for Environmental Prediction Climate Forecast System Reanalysis to simulate SSL along the

Australian coastline. The process-driven hydrodynamic models contribute to the understanding of physical mechanisms of the storm surge process and can predict storm surges accurately (Dietrich et al., 2011; Ramos-Valle et al., 2020); however, they are often numerically expensive and may fail to fully resolve bathymetric and geometric features due to limited details in the available data sets or the resolution of computational grids, thereby affecting the applicability and accuracy of the SSL simulation (Arns et al., 2020; Beisiegel et al., 2020; Cyriac et al., 2018; Fernández-Montblanc et al., 2019; Zou et al., 2013).

Since the process-driven models have their own limitations in SSL simulations, the data-driven models have emerged as an important alternative in recent years since they provide a simple and efficient method to characterize extreme storm surges with a low data requirement and a reasonable level of accuracy (Cid et al., 2017, 2018; Dangendorf et al., 2014; Kim et al., 2015; T. L. Lee, 2006; Muis et al., 2018; Sahoo & Bhaskaran, 2019). For example, Cid et al. (2018) used a multivariate linear regression model to link the daily maximum SSL and the principal components of the local atmospheric conditions in Southeast Asia. Tadesse et al. (2020) used a random forest regression model to simulate storm surges globally. These parametric data-driven approaches can yield reasonable SSL simulations, but they commonly require the assumption of normality and linearity. Such an assumption, however, does not often hold for the extreme SSL that obeys the skewed distribution, thereby diminishing the reliability of coastal flood risk assessments. In addition to the parametric data-driven approaches, the nonparametric ones have also been extensively used to simulate SSL, such as the linguistic decision tree approach (Royston et al., 2013) and the NARX (nonlinear autoregressive models with exogenous input) model (Sheridan et al., 2019). However, the nonparametric approaches often require a large number of samples to ensure the performance of extreme SSL simulations, which can be barely satisfied for the extreme storm surge characterization (Lall et al., 2016).

Copulas have been proven to be a powerful tool to assess the joint dependence between hydroclimatic variables regardless of their marginal distributions (Chen et al., 2020b, 2020a; Masina et al., 2015; Moftakhari et al., 2017; Trepanier et al., 2017; Wahl et al., 2015; Wang & Wang, 2019; Wang & Zhu, 2020; Zhang et al., 2019). In order to extend the parametric copulas to higher dimensions, vine copula has been proposed to decompose an arbitrary multivariate probability density into a cascade of bivariate copulas, thereby allowing for flexible simulations of the complex interactions among hydroclimate variables (Aas et al., 2009). In recent year, vine copula has been successfully used to improve multivariate simulations and to understand compound extremes (Bevacqua et al., 2017; Liu et al., 2018, 2015; Manning et al., 2018; Pereira & Veiga, 2018; Tosunoglu & Singh, 2018; Wang et al., 2019). However, little effort has been devoted to using the flexible structures of vine copula for storm surge simulations. Since the meteorological factors (e.g., wind and pressure) contributing to extreme storm surges are interdependent, assessing such multidimensional interdependence based on vine copula can improve the accuracy and reliability of SSL simulations. It is thus desired to use vine copula for improving the characterization of extreme storm surges induced by tropical cyclones. In addition, the vine copula simulations are subject to the uncertainties in both model parameters and model structures. Although the uncertainty in vine copula parameters has been explored in previous studies (Gruber & Czado, 2015, 2018; Min & Czado, 2010), few studies have explicitly addressed the uncertainty in vine copula model structures (Chang et al., 2020). Previous studies commonly used the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) to select an optimal vine structure. Such a procedure not only fails to guarantee the optimal model performance but also discards the vine structures that are equally good at characterizing hydroclimate variables, thereby leading to biased simulation results. For example, existing vine copula models can underestimate high values and overestimate low values even though they often perform well for normal values (Wang et al., 2019). Thus, it is necessary to develop a generalized framework to address the vine structure uncertainty for improving the reliability and robustness of SSL characterization.

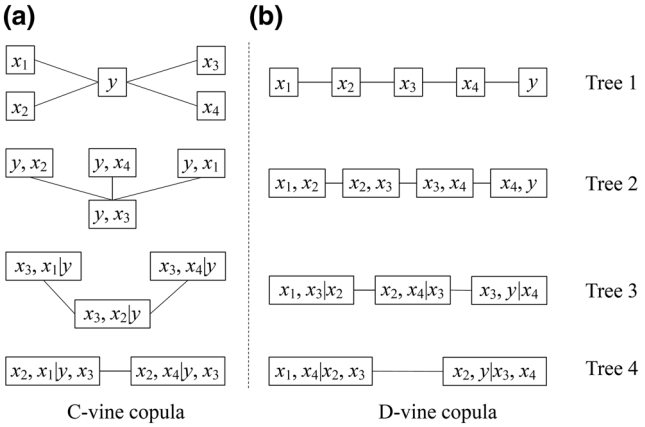

Figure 1. The structures of the (a) canonical vine (C-vine) and (b) drawable vine (D-vine) copula for five random variables (x_1, x_2, x_3, x_4, and y) with a given order. The edges represent the bivariate copulas.

In this study, we propose a robust data-driven framework for improving the characterization of extreme storm surges induced by tropical cyclones. Specifically, a multi-structure approach is developed to address

Table 1
Summary of seven Bivariate Copula Families for Vine Copula Simulation

Name	Mathematical description for $C(u, v)$	Parameter range
Gaussian	$\int_{-\infty}^{\phi^{-1}(u)}\int_{-\infty}^{\phi^{-1}(v)}\frac{1}{2\pi\sqrt{1-\theta^2}}\exp\left(\frac{2\theta xy-x^2-y^2}{2\left(1-\theta^2\right)}\right)dxdy$	$\theta\in[-1,1]$
Student t	$\int_{-\infty}^{t_{\theta_2}^{-1}(u)}\int_{-\infty}^{t_{\theta_2}^{-1}(v)}\frac{\Gamma\left(\left(\theta_2+2\right)/2\right)}{\Gamma\left(\theta_2/2\right)\pi\theta_2\sqrt{1-\theta_1^2}}\left(1+\frac{x^2-2\theta_1xy+y^2}{\theta_2}\right)^{\left(\theta_2+2\right)/2}dxdy$	$\theta_1\in[-1,1],\theta_2\in(0,\infty)$
Clayton	$\max\left(u^{-\theta}+v^{-\theta}-1,0\right)^{-1/\theta}$	$\theta\in[-1,\infty)\setminus 0$
Frank	$-\frac{1}{\theta}\ln\left[1+\frac{\left(\exp\left(-\theta u\right)-1\right)\left(\exp\left(-\theta v\right)-1\right)}{\exp\left(-\theta\right)-1}\right]$	$\theta\in\mathbb{R}\setminus 0$
Gumbel	$\exp\left\{-\left[\left(-\ln\left(u\right)\right)^{\theta}+\left(-\ln\left(v\right)\right)^{\theta}\right]^{1/\theta}\right\}$	$\theta\in[1,\infty)$
Joe	$1-\left[\left(1-u\right)^{\theta}+\left(1-v\right)^{\theta}-\left(1-u\right)^{\theta}\left(1-v\right)^{\theta}\right]^{1/\theta}$	$\theta\in[1,\infty)$
BB1	$\left\{1+\left[\left(u^{-\theta_1}-1\right)^{\theta_2}+\left(v^{-\theta_1}-1\right)^{\theta_2}\right]^{1/\theta_2}\right\}^{-1/\theta_1}$	$\theta_1\in(0,\infty),\theta_2\in(1,\infty)$

the uncertainty in the vine copula model structure. The Bayesian model averaging (BMA) technique is used to construct a probabilistic ensemble of dependence structures of multiple mechanisms causing extreme storm surges. In addition, the Markov chain Monte Carlo (MCMC) simulations are performed to address the uncertainty in the relative performance of vine structures, enhancing the reliability of SSL characterizations. We demonstrate the proposed data-driven framework by an in-depth analysis of the extreme SSL observed over two tidal gauging stations in Hong Kong which is prone to tropical cyclones and has been seriously damaged by storm surges over the past decades. The extreme SSL observations are collected from the Hong Kong Observatory, and the climatological forcing factors are collected from the ERA5 reanalysis data set.

This paper is organized as follows. Section 2 describes the regular vine copula, and Section 3 introduces the proposed data-driven framework as well as the relevant models and algorithms. Section 4 introduces the extreme storm surge observations and the climatological forcing factors used in this study. Section 5 introduces two benchmark approaches for the SSL simulation. Section 6 presents the characterization of extreme SSL based on the robust data-driven framework and a comparison with the benchmark approaches. Finally, Section 7 summarizes the major remarks of the study.

2. Regular Vine Copula

Copulas are cumulative joint distributions of multiple random variables regardless of their marginal distributions. Assume that $X=(X_1,...,X_n)$ signifies an n-dimensional random variable with marginal cumulative distribution functions (CDFs), $F_1(x_1),...,F_n(x_n)$, and marginal probability density functions (pdfs), $f_1(x_1),...,f_n(x_n)$. According to Sklar's theorem (Sklar, 1959), the n-dimensional joint CDF $F(x_1,...,x_n)$ can be expressed by

$$F\left(x_1,...,x_n\right)=C\left(F_1\left(x_1\right),...,F_n\left(x_n\right)\right)=C\left(u_1,...,u_n\right) \quad (1)$$

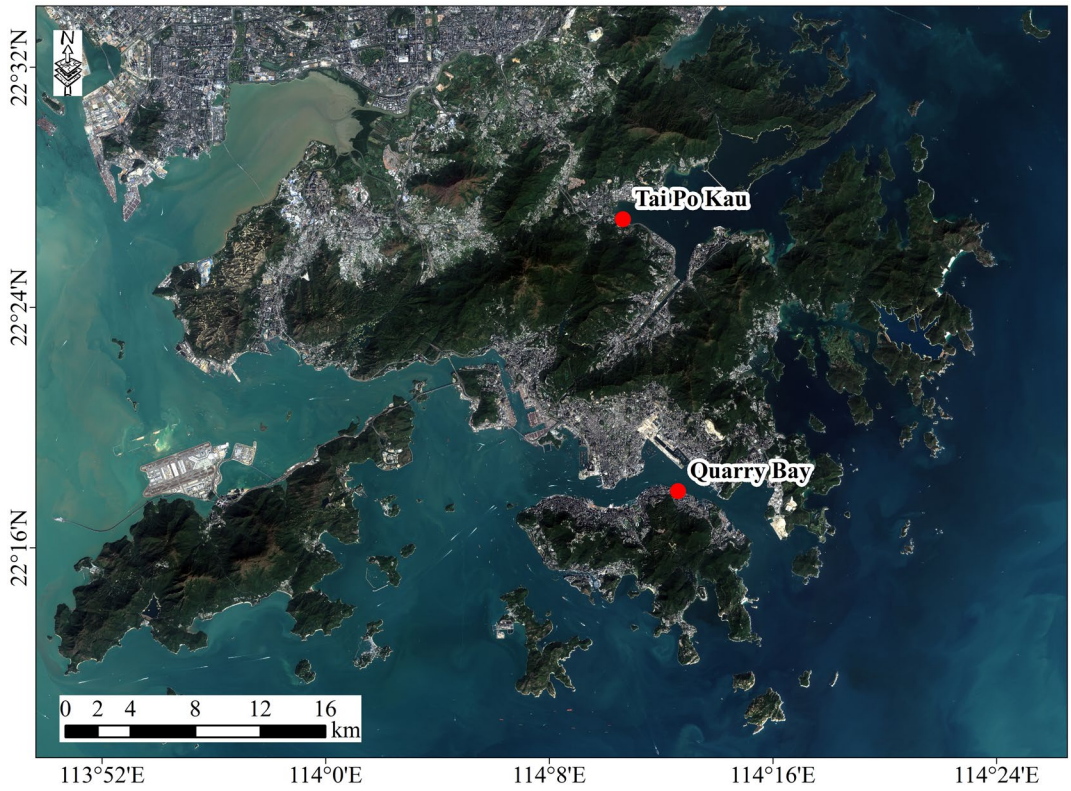

Figure 2. Locations of two tidal gauging stations in Hong Kong.

where F is the joint CDF; C is an n-dimensional copula. Table 1 presents formulas of seven commonly used copula families. $F_i(x_i)$ is the marginal CDF of random variable x_i and $u_i = F_i(x_i)$, $i = 1,..., n$. The corresponding joint pdf can be calculated by

$$f(x_1,...,x_n) = f_1(x_1) \cdot ... \cdot f_n(x_n) \cdot c(u_1,...,u_n) \tag{2}$$

where c is the copula density that can be expressed by

$$c(u_1,...,u_n) = \frac{\partial^n C(u_1,u_2,...,u_n)}{\partial u_1 \partial u_2 \cdots \partial u_n} \tag{3}$$

Since copulas are inflexible in high dimensions and high-dimensional copula families are limited, vine copula, also known as pair-copula construction (PCC), has been proposed to graphically represent the high-dimensional dependence structure as vines comprising a nested set of trees with nodes which are joined by edges. For an n-dimensional vine, the nodes in the first tree represent marginal densities of each variable, while the nodes in tree i represent conditional bivariate densities (edges) in tree $i-1$, $i = 2, ..., n-1$, which are calculated by bivariate copulas. A vine in which two edges in tree i are joined by an edge in tree $i+1$ only if these edges share a common node, $i = 1, ..., n-2$, is called regular vine. There are two special regular vines: the canonical vine (C-vine) and the drawable vine (D-vine), which are also the most widely used decompositions. Each tree in a C-vine has a unique node connected to all other nodes, while a D-vine has no node connected to more than two edges. There are a large number of possible vine structures resulting from the different order of variables and different bivariate copula families in a high-dimensional PCC. Figure 1 presents the structures of C-vine and D-vine copulas for five random variables (x_1, x_2, x_3, x_4, and y) with a given order. The joint densities for an n-dimensional C-vine and D-vine copula can be expressed by Equations 4 and 5, respectively.

$$f(x_1,\ldots,x_n)=\prod_{k=1}^{n}f(x_k)\prod_{j=1}^{n-1}\prod_{i=1}^{n-j}c_{j,j+i|1,\ldots,j-1}\left\{F\left(x_j\mid x_1,\ldots,x_{j-1}\right),F\left(x_{j+i}\mid x_1,\ldots,x_{j-1}\right)\right\} \tag{4}$$

$$f(x_1,\ldots,x_n)=\prod_{k=1}^{n}f(x_k)\prod_{j=1}^{n-1}\prod_{i=1}^{n-j}c_{i,i+j|i+1,\ldots,i+j-1}\left\{F\left(x_i\mid x_{i+1},\ldots,x_{i+j-1}\right),F\left(x_{i+j}\mid x_{i+1},\ldots,x_{i+j-1}\right)\right\} \tag{5}$$

where $F(\cdot|\cdot)$ denotes marginal conditional distributions; $c_{\cdot|\cdot}$ represents bivariate copula densities. The five-dimensional C-vine and D-vine copulas can be expressed by Equations 6 and 7, respectively.

$$f_{12345}=f_1\cdot f_2\cdot f_3\cdot f_4\cdot f_5\cdot c_{12}\cdot c_{13}\cdot c_{14}\cdot c_{15}\cdot c_{23|1}\cdot c_{24|1}\cdot c_{25|1}\cdot c_{34|12}\cdot c_{35|12}\cdot c_{45|123} \tag{6}$$

$$f_{12345}=f_1\cdot f_2\cdot f_3\cdot f_4\cdot f_5\cdot c_{12}\cdot c_{23}\cdot c_{34}\cdot c_{45}\cdot c_{13|2}\cdot c_{24|3}\cdot c_{35|4}\cdot c_{14|23}\cdot c_{25|34}\cdot c_{15|234} \tag{7}$$

Such a decomposition improves upon the corresponding five-dimensional copulas (e.g., Gaussian copula) since it permits a heterogeneous assignment of the 10 bivariate copula families in Equations 6 and 7. In addition, the PCC involves marginal conditional distributions $F(\cdot|\cdot)$, which can be expressed using the general formula (Joe, 1996):

$$F\left(x\mid\mathbf{v}\right)=\frac{\partial C_{x,\upsilon_j|\mathbf{v}_{-j}}\left\{F\left(x\mid\mathbf{v}_{-j}\right),F\left(\upsilon_j\mid\mathbf{v}_{-j}\right)\right\}}{\partial F\left(\upsilon_j\mid\mathbf{v}_{-j}\right)} \tag{8}$$

where $\boldsymbol{v}$ is a d-dimensional vector; υ_j is one arbitrarily chosen component of $\boldsymbol{v}$ and $\boldsymbol{v}_{-j}$ denotes the remaining vector of $\boldsymbol{v}$ after excluding υ_j; $C_{x,\upsilon_j|\mathbf{v}_{-j}}$ denotes the bivariate copula. For the special case where v is univariate, Equation 8 can be expressed by Equation 9, which is also commonly defined as h-functions:

$$F\left(x\mid v\right)=h\left(x,v,\theta\right)=\frac{\partial C_{x,v}\left\{F\left(x\right),F\left(v\right),\theta\right\}}{\partial F\left(v\right)} \tag{9}$$

where θ refers to the parameter set of bivariate parametric copulas (see Table 1). More details of vine copula can be seen in Aas et al. (2009).

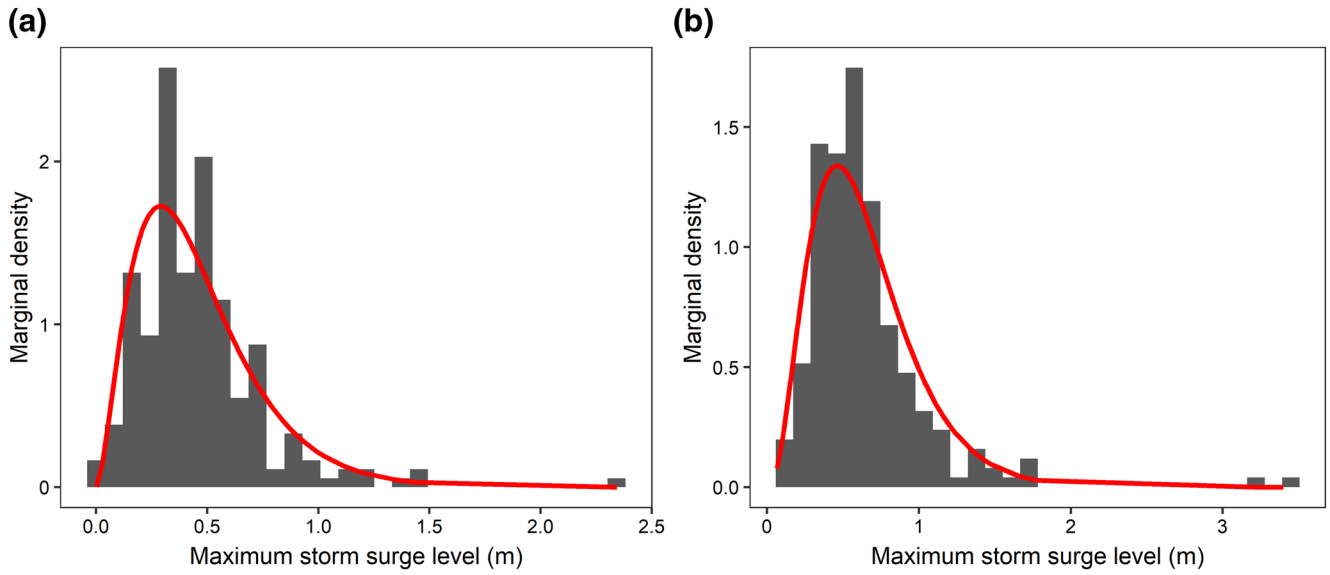

Figure 3. Histograms for the extreme storm surge level during the passage of tropical cyclones for the (a) Quarry Bay and (b) Tai Po Kau tidal gauging stations.

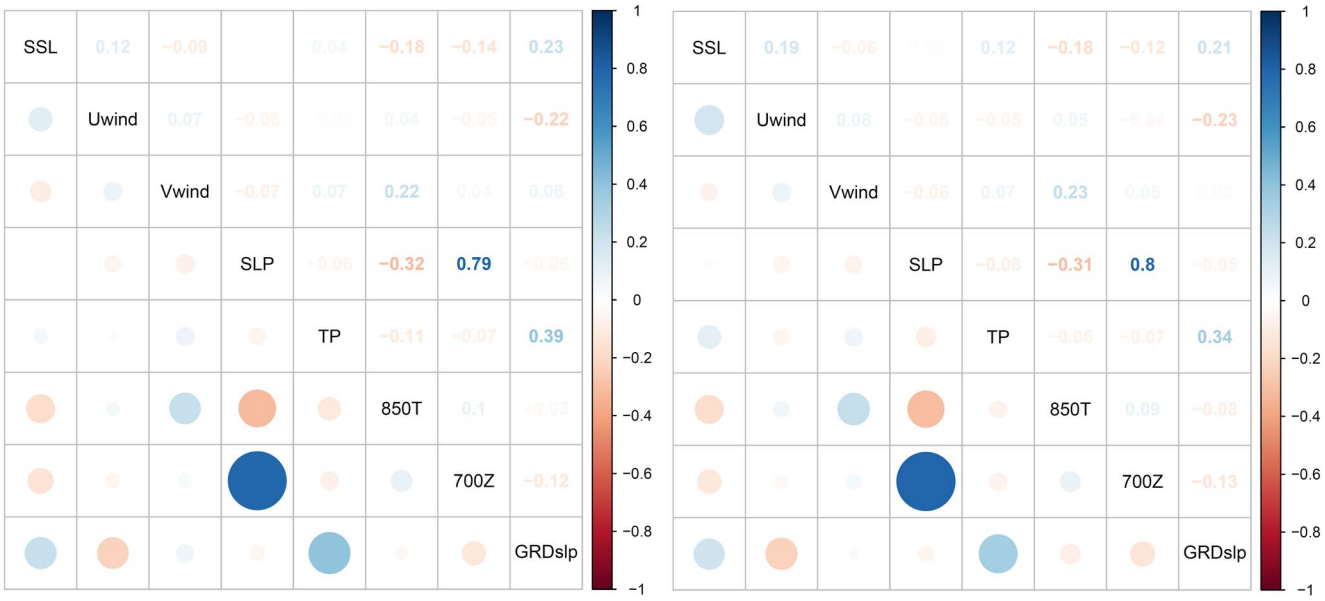

Figure 4. Partial Rank Kendall's Tau correlogram between storm surge level (SSL) and its forcing factors, including the geopotential heights in 700 hPa (700Z), the air pressure at mean sea level (SLP) and its gradient (GRDslp), the meridional and zonal wind components at 10 m (Uwind and Vwind, respectively), the temperature in 850 hPa (850T), as well as the total precipitation (TP) over the Quarry Bay (left panel) and Tai Po Kau (right panel) tidal gauging stations. The size of circles represents the relative magnitude of partial correlation. The colored numbers represent the detailed correlation coefficients.

An appropriate regular vine copula construction requires the structure specification and parameter estimation. There are $n!/2$ C-vine or D-vine structures for an n-dimensional vine copula resulting from different node orders in the vine, and each structure also varies with the pair-copula family selection, leading to a challenging selection of an optimal vine structure. Most bivariate copulas have at least one parameter, and an n-dimensional vine copula typically has at least $n(n-1)/2$ pair-copula parameters to be estimated. Thus, the global optimum of vine copula inference is challenging due to the numerous possible vine structures and the enormous parameter space. To conduct a feasible optimization and reduce the computation cost, the sequential maximal spanning tree algorithm has been widely used and proven to yield reasonable constructions of vine copula (Dißmann et al., 2013). The sequential algorithm refers to that the pair-copula parameters are estimated from top tree to down tree sequentially by maximizing the log-likelihood function, while the optimal pair-copula family is commonly selected based on the AIC or the BIC. The log-likelihood function of the bivariate copula with parameters θ given the uniform margins u_1 and u_2 can be calculated as Equation 10, where N is the sample size.

$$\log\left(L\left(u_1,u_2,\theta\right)\right)=\sum_{i=1}^{N}\log\left(c\left(u_{1,i},u_{2,i},\theta\right)\right) \tag{10}$$

3. The Data-Driven Storm Surge Characterization Framework

To improve the characterization of extreme storm surges, we propose a robust data-driven framework by intelligently integrating a multi-structure ensemble of conditional quantile functions of SSL based on regular vine copulas and by explicitly addressing the uncertainty in vine structures in a Bayesian framework. The marginal distribution of the SSL observation and its climatological forcing factors were selected from a total of 13 types of probability distributions, including Gaussian, gamma, exponential, Weibull, logistic, log-normal, log-logistic, Cauchy, Gumbel, generalized extreme value, generalized Pareto, Pearson Type III, and inverse Gaussian. The optimal marginal distribution was identified using the Kolmogorov-Smirnov (K-S) test, while the corresponding distribution parameters were estimated by the maximum likelihood method. Vine copula was used to link the marginal distributions of the SSL and its climatological forcing factors together for constructing the conditional quantile functions of SSL. In the vine copula construction,

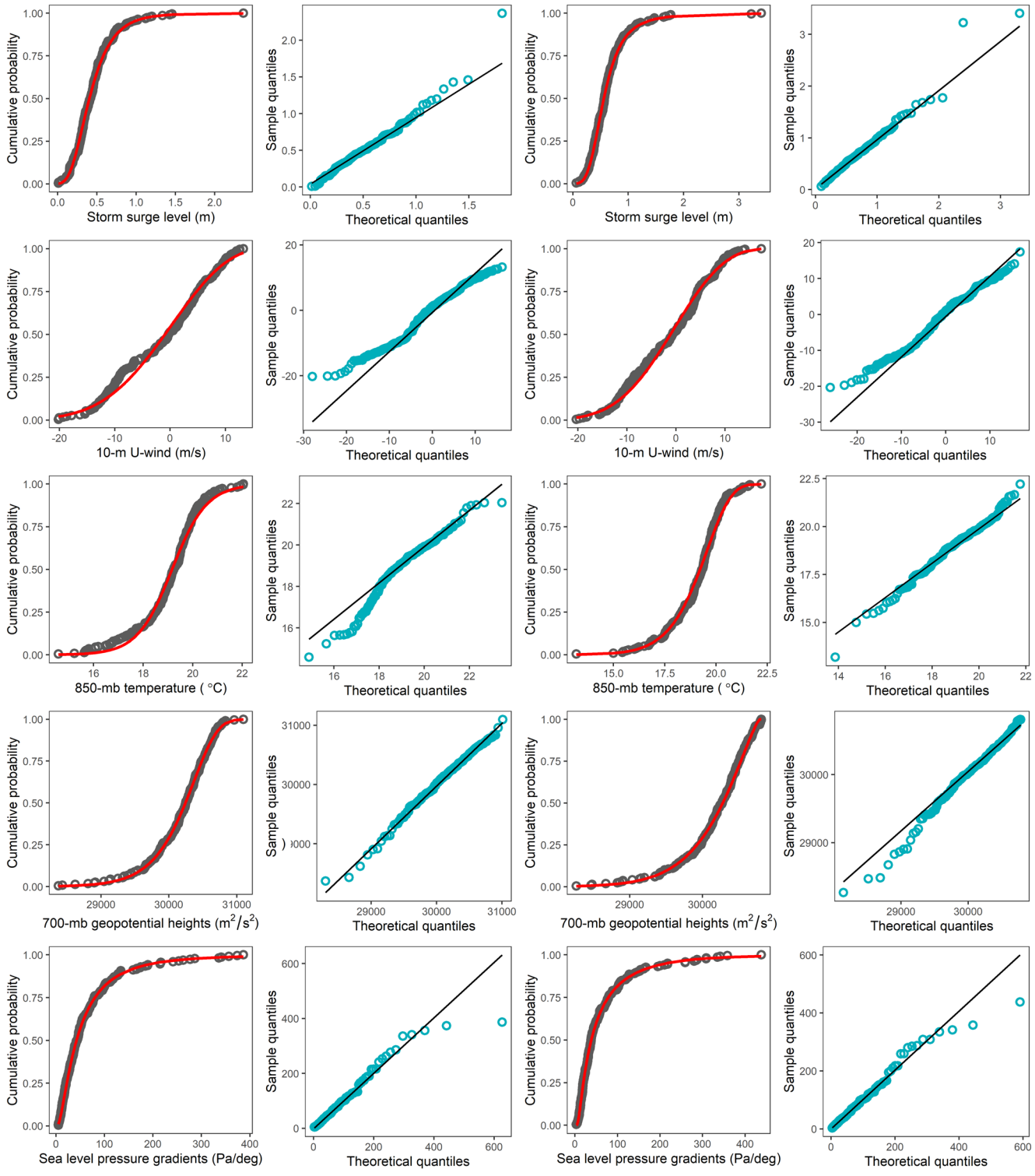

Figure 5. Cumulative distribution functions (CDF) of storm surge level and its forcing factors, as well as their Q-Q plots over the Quarry Bay and Tai Po Kau tidal gauging stations.

seven commonly used bivariate copulas were included as the pair-copula family candidates, including the elliptical (Gaussian and Student *t*) and Archimedean copulas (Clayton, Gumbel, Frank, and Joe), as well as a widely used two-parameters hybrid Archimedean copula, BB1. Formulas of these copulas are provided in Table 1. Suppose there are $n-1$ forcing factors of the SSL denoted by $x_1,\ldots, x_{n-1}$ and the SSL is denoted by y,

the conditional distribution of the SSL based on the conditional C-vine and D-vine copulas can be expressed by Equations 11 and 12, respectively.

$$F\left(y \mid x_1,\ldots,x_{n-1}\right)=\frac{\partial C_{n,n-1|1,\ldots,n-2}\left\{F\left(y \mid x_1,\ldots,x_{n-2}\right),F\left(x_{n-1} \mid x_1,\ldots,x_{n-2}\right)\right\}}{\partial F\left(x_{n-1} \mid x_1,\ldots,x_{n-2}\right)} \tag{11}$$

$$F\left(y \mid x_1,\ldots,x_{n-1}\right)=\frac{\partial C_{n,1|2,\ldots,n-1}\left\{F\left(y \mid x_2,\ldots,x_{n-1}\right),F\left(x_1 \mid x_2,\ldots,x_{n-1}\right)\right\}}{\partial F\left(x_1 \mid x_2,\ldots,x_{n-1}\right)} \tag{12}$$

To better explain the construction of the conditional distribution, an example of three forcing factors (x_1, x_2, x_3) for C-vine and D-vine copulas can be expressed by Equations 13 and 14, respectively, where u_1, u_2, u_3, and u_y denote the corresponding marginal cumulative probability.

$$\begin{aligned}F\left(y \mid x_1,x_2,x_3\right)&=\frac{\partial C_{y,3|1,2}\left(F\left(y \mid x_1,x_2\right),F\left(x_3 \mid x_1,x_2\right)\right)}{\partial F\left(x_3 \mid x_1,x_2\right)}\\&=\frac{\partial C_{y,3|1,2}\left(\frac{\partial C_{y,2|1}\left(F\left(y \mid x_1\right),F\left(x_2 \mid x_1\right)\right)}{\partial F\left(x_2 \mid x_1\right)},\frac{\partial C_{3,2|1}\left(F\left(x_3 \mid x_1\right),F\left(x_2 \mid x_1\right)\right)}{\partial F\left(x_2 \mid x_1\right)}\right)}{\partial\left(\frac{\partial C_{3,2|1}\left(F\left(x_3 \mid x_1\right),F\left(x_2 \mid x_1\right)\right)}{\partial F\left(x_2 \mid x_1\right)}\right)}\\&=h\left\{h\left[h\left(u_y,u_1,\theta_{1y}\right),h\left(u_2,u_1,\theta_{12}\right),\theta_{2y|1}\right],h\left[h\left(u_3,u_1,\theta_{13}\right),h\left(u_2,u_1,\theta_{12}\right),\theta_{23|1}\right],\theta_{3y|12}\right\}\end{aligned} \tag{13}$$

$$\begin{aligned}F\left(y \mid x_1,x_2,x_3\right)&=\frac{\partial C_{y,1|2,3}\left(F\left(y \mid x_2,x_3\right),F\left(x_1 \mid x_2,x_3\right)\right)}{\partial F\left(x_1 \mid x_2,x_3\right)}\\&=\frac{\partial C_{y,1|2,3}\left(\frac{\partial C_{y,2|3}\left(F\left(y \mid x_3\right),F\left(x_2 \mid x_3\right)\right)}{\partial F\left(x_2 \mid x_3\right)},\frac{\partial C_{1,3|2}\left(F\left(x_1 \mid x_2\right),F\left(x_3 \mid x_2\right)\right)}{\partial F\left(x_3 \mid x_2\right)}\right)}{\partial\left(\frac{\partial C_{1,3|2}\left(F\left(x_1 \mid x_2\right),F\left(x_3 \mid x_2\right)\right)}{\partial F\left(x_3 \mid x_2\right)}\right)}\\&=h\left\{h\left[h\left(u_y,u_3,\theta_{3y}\right),h\left(u_2,u_3,\theta_{23}\right),\theta_{2y|3}\right],h\left[h\left(u_1,u_2,\theta_{12}\right),h\left(u_3,u_2,\theta_{23}\right),\theta_{13|2}\right],\theta_{1y|23}\right\}\end{aligned} \tag{14}$$

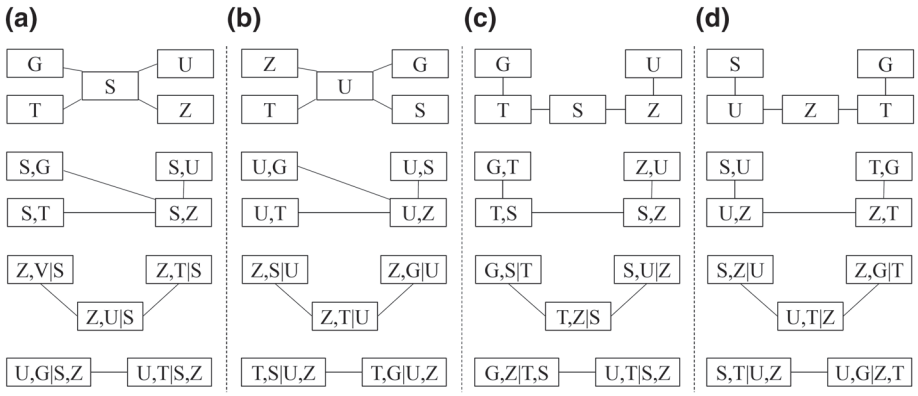

Figure 6. Four representative structures for five-dimensional vine copulas. (a–b) denote two canonical vine structures and (c–d) denote two drawable vine structures. U, T, Z, G, and S represent the meridional wind components at 10 m, air temperature, geopotential heights, the spatial gradient of sea level pressure, and storm surge level, respectively. The edges represent the bivariate copulas.

Figure 7. Akaike information criterion (AIC) for all C-vine and D-vine copulas as well as root mean squared error (RMSE) for the simulated storm surge level over the Quarry Bay station during the calibration and validation periods. The number on the *y*-axis indicates the order of the vine copula model based on AIC.

To generate SSL simulations from the conditional vine copula, the inverse forms of *h*-functions are applied. For instance, assume that only one forcing factor, x_1, is included in the SSL simulation, the conditional distribution of the SSL and x_1 is $h(u_y|u_1, \theta)$. The SSL can be simulated by generating random probability levels τ (e.g., $\tau = 0.01, 0.1, ..., 0.99$) as follows:

$$y = F^{-1}\left[h^{-1}\left(\tau \mid u_1, \theta_{1y}\right)\right] \tag{15}$$

where θ_{1y} represents the copula parameter for the joint distribution of (x_1, y). Thus, the inverse forms of h-functions can be recursively applied to transform Equations 13 and 14 into the SSL simulations conditioned by multiple forcing factors for C-vine and D-vine copulas as Equations 16 and 17, respectively. The multi-structure conditional vine copula simulation was performed in R version 3.4.0 (Aas et al., 2009; Bevacqua et al., 2017; Nagler et al., 2019; R Core Team, 2018).

$$y_{\text{C-vine}} = F^{-1}\left\{h^{-1}\left[h^{-1}\left(h^{-1}\left(\tau \mid h\left(h\left(u_3 \mid u_1\right) \mid h\left(u_2 \mid u_1\right)\right)\right) \mid h\left(u_2 \mid u_1\right)\right) \mid u_1\right]\right\} \tag{16}$$

$$y_{\text{D-vine}} = F^{-1}\left\{h^{-1}\left[h^{-1}\left(h^{-1}\left(\tau \mid h\left(h\left(u_1 \mid u_2\right) \mid h\left(u_3 \mid u_2\right)\right)\right) \mid h\left(u_2 \mid u_3\right)\right) \mid u_3\right]\right\} \tag{17}$$

Since the large number of possible vine structures can lead to considerable uncertainty in the SSL simulation, the proposed data-driven framework uses BMA to generate an "intelligent" consensus forecast, weighted by the model performance for each vine structure. Assume that $\boldsymbol{x} = x_1,..., x_K$ signify an ensemble of SSL simulations obtained from K different vine structures and y denotes the SSL observations. Each ensemble member forecast, x_k, in BMA is associated with a conditional pdf, $p_k(y|x_k)$, which can be interpreted as the conditional pdf of y on x_k, given that x_k is the optimal forecast in the ensemble. The probabilistic prediction of y based on the multi-structure model ensemble can be expressed as a finite mixture model:

$$p\left(y \mid x_1 \ldots x_k\right) = \sum_{k=1}^{K} w_k p_k\left(y \mid x_k\right) \tag{18}$$

where w_k is the posterior probability distribution of forecast k being the optimal one. All the w_k values are nonnegative and add up to one, and they can be viewed as weights reflecting an individual model's relative contribution to predictive skill in the training period (Madadgar & Moradkhani, 2014; Raftery et al., 2005, 1997; Vrugt, 2016). Since the SSL observations commonly obey the skew distribution, the conditional pdfs, $p_k(y|x_k)$, of the different ensemble members are approximated by a gamma distribution with PDF

$$p\left(y \mid x_k\right) = \frac{1}{\beta_k^{\alpha_k}\Gamma\left(\alpha_k\right)} y^{\alpha_k - 1} \exp\left(-y / \beta_k\right) \tag{19}$$

The parameters of the gamma distribution depend on the ensemble member forecast, x_k, through the relationships

$$u_k = b_{0k} + b_{1k} x_k \tag{20}$$

and

$$\sigma_k = c_0 + c_{1k} x_k \tag{21}$$

where $\mu_k = \alpha_k \beta_k$ is the mean of the distribution, and $\sigma_k^2 = \alpha_k \beta_k^2$ is its variance. The parameters b_{0k} and b_{1k} are determined by linear regression of the observed SSL on the simulated SSL for each of the K ensemble members.

The BMA weights, w_k, $k = 1,..., K$; c_0; and c_{1k}, $k = 1,..., K$, were estimated using the MCMC simulation in this study. The MCMC simulation has been demonstrated in a number of studies to outperform the Expectation-Maximization algorithm, since it not only guarantees the global convergence of the BMA weights and other parameters but also explicitly uncovers the underlying uncertainty (Duan & Phillips, 2010; Vrugt et al., 2008; Wong et al., 2018). The MCMC simulation is implemented using the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt, 2016). According to the Bayes' theorem, the posterior distribution $p(\boldsymbol{w}, c_0, \boldsymbol{c}_1|y)$ of the BMA weights, $\boldsymbol{w} = (w_1,..., w_K)$; c_0; and $\boldsymbol{c}_1 = (c_{11},..., c_{1K})$ can be expressed as:

Table 2
Kolmogorov-Smirnov Test for the Optimal Distributions Fitted to Different Variables

Gauging station	Variable	Distribution	p-value
Quarry Bay	Storm surge level	Logistic	0.99
Quarry Bay	10-m Uwind	GEV	0.95
Quarry Bay	Sea level pressure gradient	Log-normal	0.99
Quarry Bay	700-mb geopotential height	Weibull	0.99
Quarry Bay	850-mb temperature	Logistic	0.85
Tai Po Kau	Storm surge level	Log-logistic	0.98
Tai Po Kau	10-m Uwind	GEV	0.84
Tai Po Kau	Sea level pressure gradient	Inverse Gaussian	0.99
Tai Po Kau	700-mb geopotential height	GEV	0.99
Tai Po Kau	850-mb temperature	Weibull	0.90

Figure 8. Akaike information criterion (AIC) for all C-vine and D-vine copulas as well as root mean squared error (RMSE) for the simulated storm surge level over the Tai Po Kau station during the calibration and validation periods. The number on the *y*-axis indicates the order of the vine copula model based on AIC.

$$p\left(\mathrm{w}, c_0, \boldsymbol{c}_1 \mid \mathbf{x}, y\right) = \frac{p\left(\mathrm{w}, c_0, \boldsymbol{c}_1\right) \times p\left(\mathbf{x}, y \mid \mathrm{w}, c_0, \boldsymbol{c}_1\right)}{p\left(\mathbf{x}, y\right)} \tag{22}$$

where $p(\boldsymbol{x}, y \mid \boldsymbol{w}, c_0, \boldsymbol{c}_1) \cong L(\boldsymbol{w}, c_0, \boldsymbol{c}_1 \mid \boldsymbol{x}, y)$ denotes the likelihood function; $p(\boldsymbol{w}, c_0, \boldsymbol{c}_1)$ is the prior distribution of the BMA weights and other parameters. $p(\boldsymbol{x}, y)$ denotes the evidence that acts as a normalization constant, which is excluded from the Bayesian analysis in practice. Thus, the posterior distribution can be simplified as:

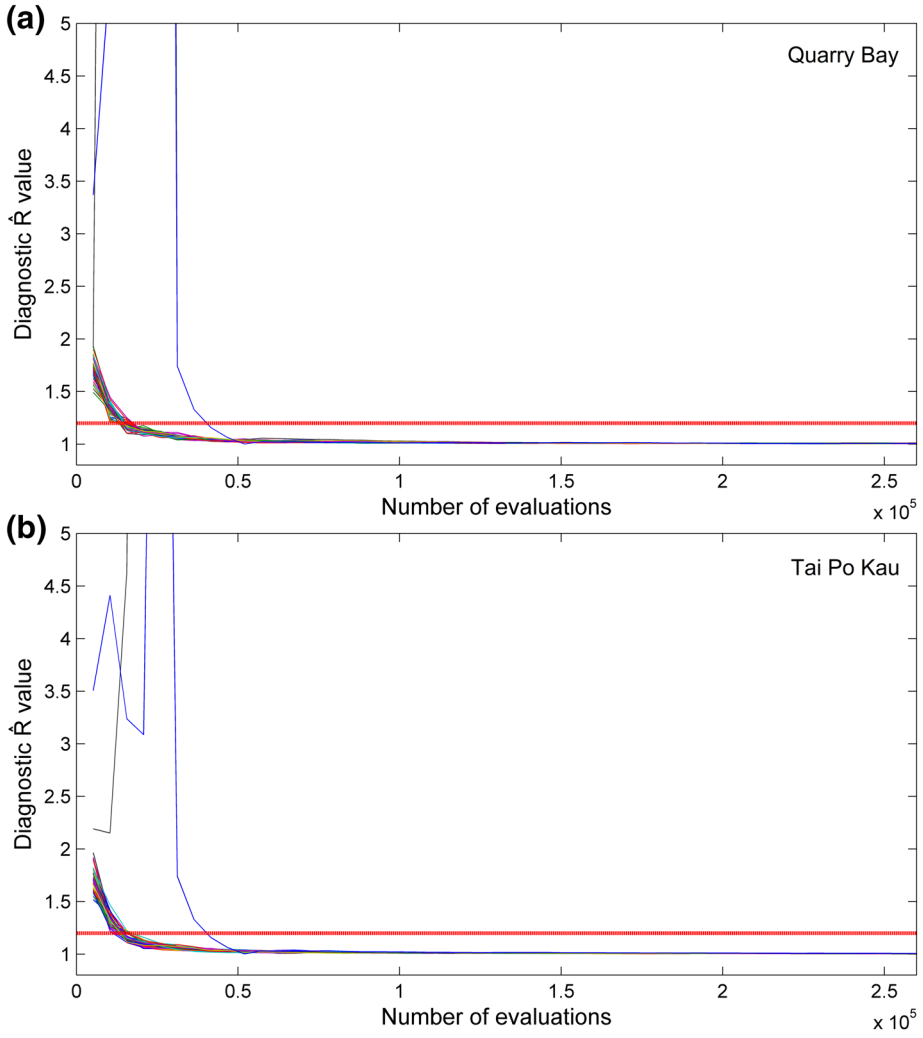

Figure 9. Evolution of the convergence diagnostic $\hat{R}$ values for different parameters in the Markov chain Monte Carlo (MCMC)-based Bayesian model averaging (BMA) simulation.

$$p\left(\mathrm{w},c_0,\boldsymbol{c}_1 \mid \mathbf{x},y\right) \propto p\left(\mathrm{w},c_0,\boldsymbol{c}_1\right) \times L\left(\mathrm{w},c_0,\boldsymbol{c}_1 \mid \mathbf{x},y\right) \tag{23}$$

For numerical stability and simplicity, the likelihood function $L(\cdot|\cdot)$ is commonly logarithmically transformed to the log-likelihood function as:

$$\ell\left(w_1,\ldots,w_K,c_0,c_1,\ldots c_K \mid x_1,\ldots x_K,y\right) = \sum_{t=1}^{n} \log\left(\sum_{k=1}^{K} w_k p_k\left(y \mid x_k\right)\right) \tag{24}$$

where n denotes the total number of SSL observations in the training data set. When the prior distributions are specified, the global convergence of BMA weights and other parameters to the stationary distribution can be achieved by repeated Monte Carlo sampling in the prior parameter space, together with an acceptance/rejection rule. Specifically, the MCMC simulation proceeds by running multiple Markov chains simultaneously and proposing a candidate point z_p at each step (S. Wang et al., 2018). The candidate can be either accepted or rejected by calculating the Metropolis acceptance probability:

$$p_{\text{accept}}\left(z_C \rightarrow z_p\right) = \min\left[1, \frac{p\left(z_p\right)}{p\left(z_C\right)}\right] \tag{25}$$

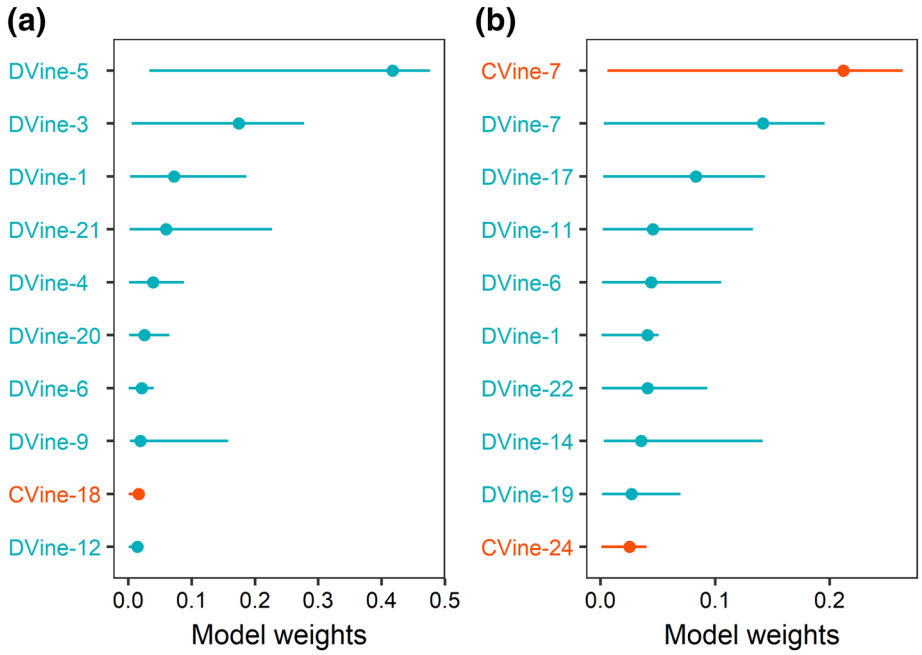

Figure 10. Posterior estimates of the 10 largest Markov chain Monte Carlo (MCMC)-derived Bayesian model averaging (BMA) weights and their uncertainty ranges of the individual single-structure storm surge level (SSL) simulation for the (a) Quarry Bay and (b) Tai Po Kau tidal gauging stations. The number on the *y*-axis indicates the order of the vine copula model based on Akaike information criterion (AIC).

where $p(z_c)$ and $p(z_p)$ denote the probability density of current point z_c and candidate point z_p, respectively. The Markov chain moves to z_p if the candidate point is accepted, or it remains at its current location. The MCMC evolution proceeds until the convergence of the Markov chain to a limiting distribution, which can be monitored with the multi-chain $\hat{R}$ diagnostic of Gelman and Rubin (1992). Typically, the convergence of the posterior distribution is achieved when the $\hat{R}$-statistic drops below a value of 1.2. More details about the MCMC simulation using the DREAM algorithm can be found in Vrugt (2016) and Vrugt et al. (2008).

4. Extreme Storm Surge Observations and Climatological Forcing Factors

The non-tidal residual component is considered as the SSL in this study, which represents the difference between the recorded sea level and the predicted astronomical tide (Cid et al., 2018; Pawlowicz et al., 2002; Tadesse et al., 2020). Since tropical cyclones do not necessarily lead to significant storm surge events, we focus on the maximum SSL during the passage of each of tropical cyclones (and tropical depression) that necessitate the issuance of tropical cyclone warning signals. Thus, we collected 226 and 219 such SSL records over the Quarry Bay and Tai Po Kau (hereafter referred to as QB and TPK, respectively, see Figure 2) tidal gauging stations during 1979–2018 since QB and TPK have the longest SSL records in Hong Kong. The maximum SSL records are separated from the recorded sea level time series by the Hong Kong Observatory through harmonic analysis (Devlin et al., 2019; T. C. Lee & Wong, 2007). It should be noted that the maximum SSL records are temporally discrete with a sub-hourly time accuracy. Figure 3 presents the histograms of the maximum SSL records for the (a) QB and (b) TPK tidal gauging stations, which significantly obey the gamma distribution at the significance level of 0.05.

The ERA5 gridded reanalysis data set was used to select the climatological forcing factors, including the pressure at mean sea level (SLP) and the corresponding gradients (GRDslp), the meridional and zonal wind components at 10 m (Uwind and Vwind, respectively), the temperature in 850 hPa (850T), the geopotential

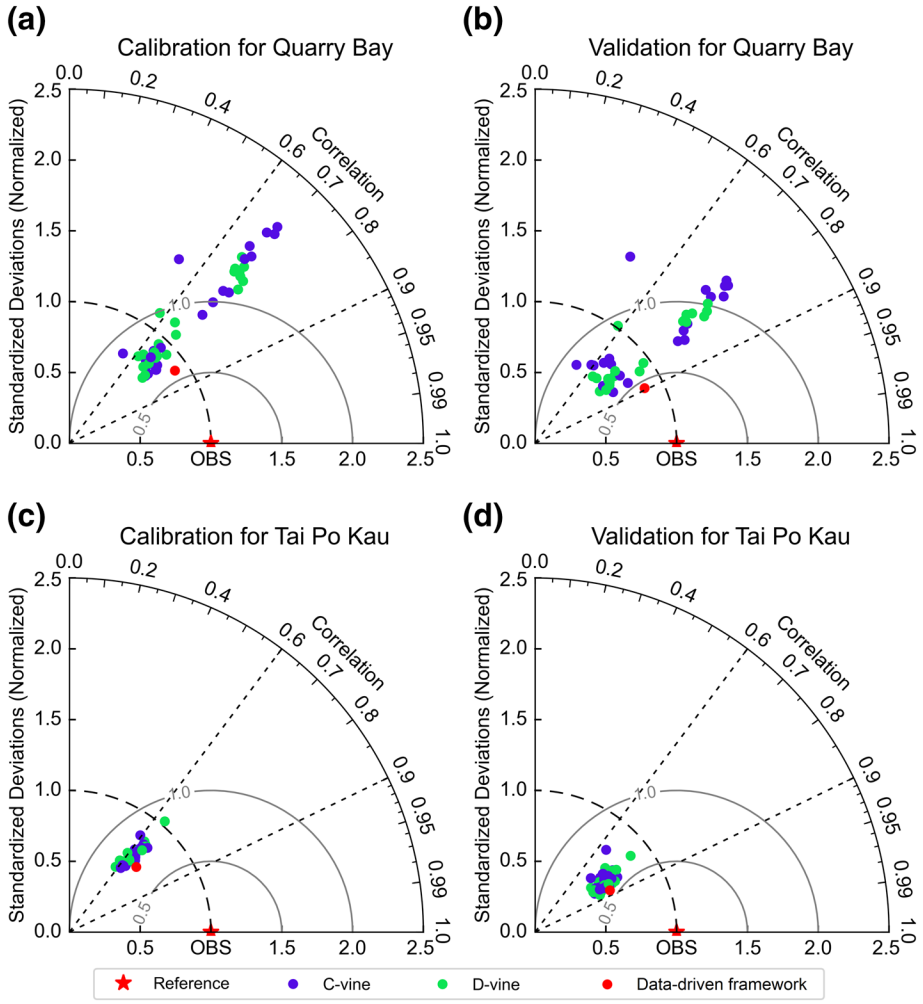

Figure 11. Comparison of SSL derived from the proposed data-driven framework and the single-structure vine copula simulations over the Quarry Bay and Tai Po Kau stations during a, c the calibration and b, d validation periods.

heights in 700 hPa (700Z), as well as the total precipitation (TP). The ERA5 reanalysis data set has a spatial resolution of 0.25 × 0.25° and a one-hour temporal resolution, which is available from 1 January 1979 to present (Dullaart et al., 2020). To link the ERA5 reanalysis data set with the SSL records, the meteorological variables were averaged over the neighboring window of size 3 × 3 centered at each tidal gauging station to eliminate the bias caused by the spatially heterogeneous surface (see Text S1 of the supplementary material for sensitivity analysis of grid sizes). In addition, a 10-h lag was used to consider the delay effects of meteorological variables on extreme storm surges (i.e., the temporal average between the time of extreme storm surge occurrence and 10 h before for each variable). For sensitivity analysis of time lags, please see Text S2 of the supplementary material.

5. Benchmark Approaches

To justify the robustness of the proposed data-driven framework, it is necessary to quantitatively compare with previous approaches. The principal components regression (PCR) and the random forest (RF) regression were selected as comparative benchmarks since they have been successfully used in the SSL characterization by linking with the surrounding atmospheric conditions (Cid et al., 2018; Tadesse et al., 2020). The PCR is a multivariate linear regression model that relates the SSL with the principal components of the climatological forcing factors, while the RF is a supervised machine learning algo-

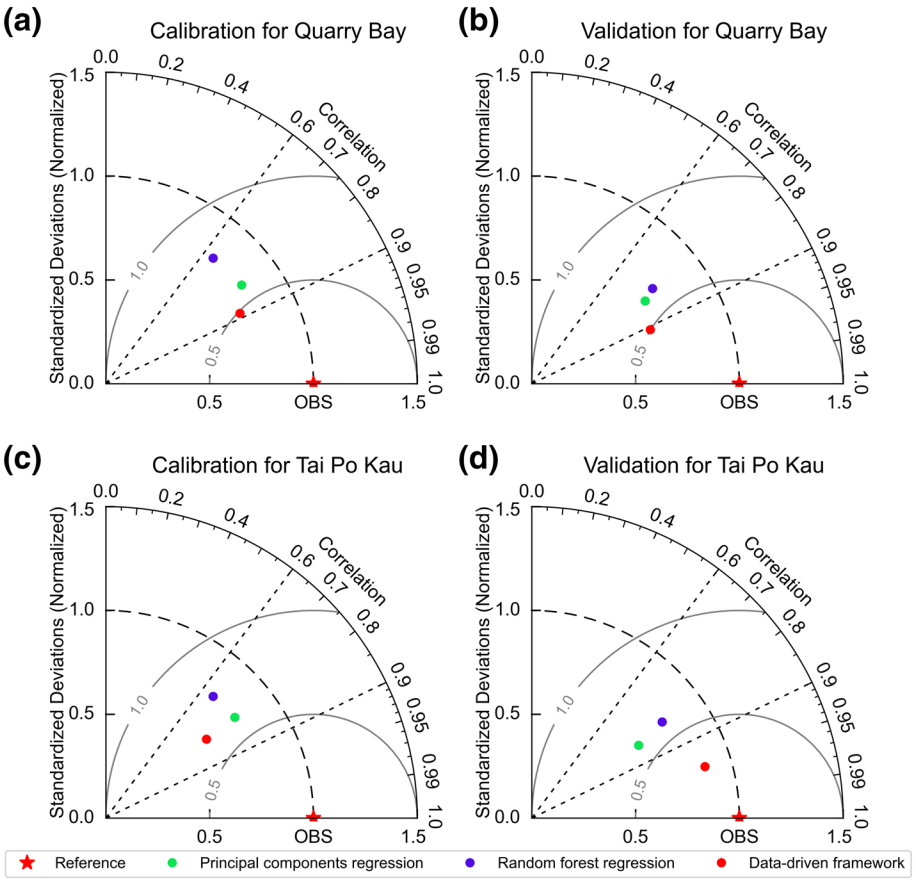

Figure 12. Comparison of SSL derived from the proposed data-driven framework, principal components regression, and random forest regression over the Quarry Bay and Tai Po Kau stations during a, c the calibration and b, d validation periods.

rithm that combines predictions from multiple machine learning algorithms based on the concepts of classification and regression trees, thereby leading to a more reliable prediction than a single model. Since the two benchmark approaches are both capable of handling predictors with a large dimension, the aforementioned climatological forcing factors within a 3×3 grid centered at each tidal gauging stations were used as the predictors. The predictors were also temporally averaged within a 10-h lag period to consider the delay effects. Hence, a total of 63 predictors (i.e., 9 grid cells $\times$ 7 variables) were used for the PCR and RF simulations.

To implement the PCR-based SSL simulation, a principal component analysis was performed on the predictor, thereby generating 63 principal components (PCs). Each PC was used as the independent variable to fit a simple linear regression model where the SSL was the dependent variable, and the PC with the best fit (smallest sum of squared errors) was selected as PC_1 in Equation 26.

$$\mathrm{SSL} = a + \sum_{i=1}^{N} b_i \times PC_i \tag{26}$$

where a and b_i are the regression coefficients; N represents the number of PCs. Then, a forward procedure was used to iteratively add PCs with the best fit into Equation 26 until the model was not improved at the significance level of 0.05 based on the F-test. On the other hand, the number of regression trees should be chosen properly to better implement the RF-based SSL simulation. We chose 25 as the optimal number of regression trees based on sensitivity analysis that can be seen in Text S3 of the supplementary material.

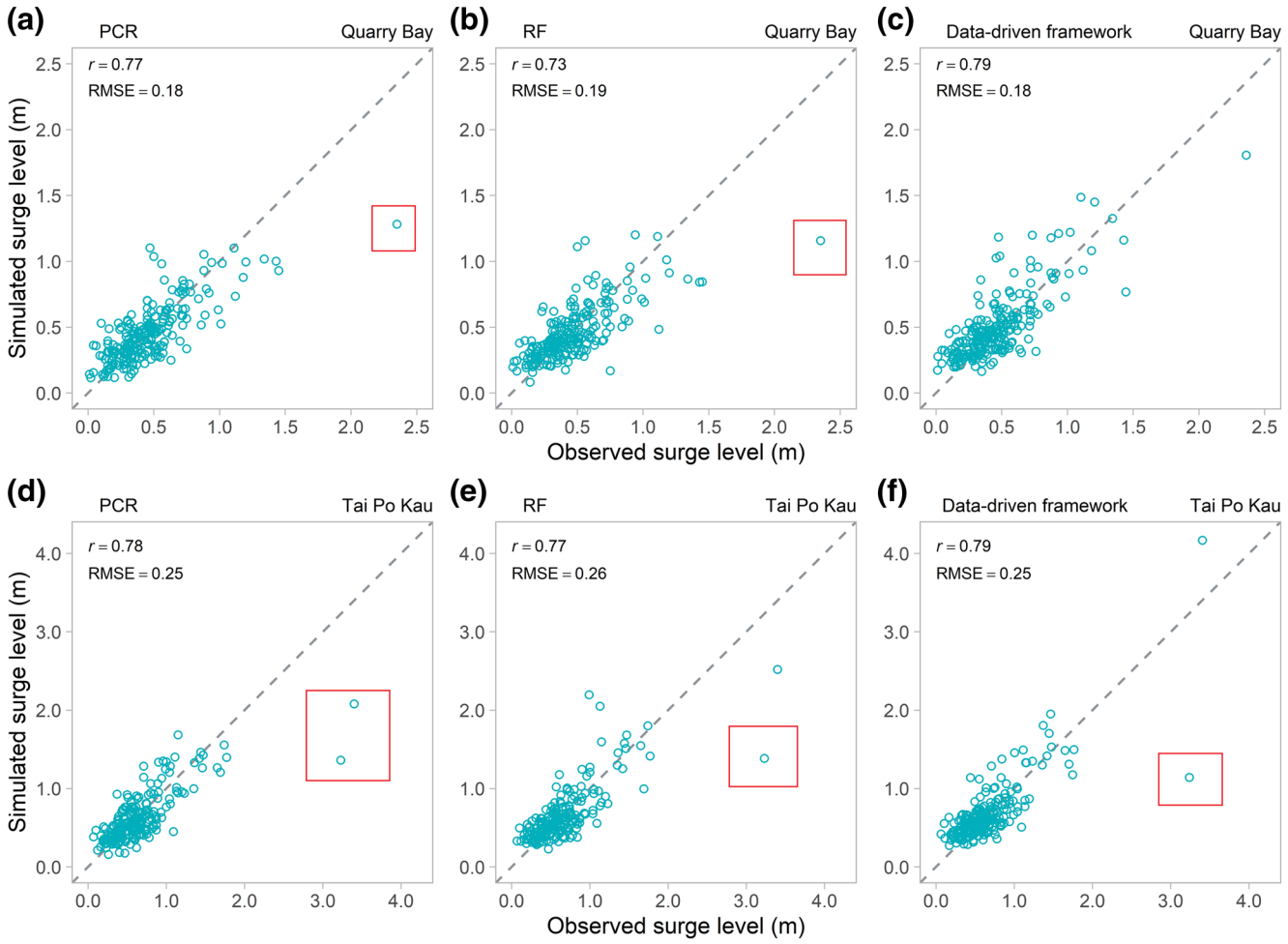

Figure 13. Scatter plots of the 10-fold cross validation for SSL simulations generated from a, d the principal components regression (PCR), b, e random forest regression (RF), and c, f the proposed data-driven framework.

6. Results and Discussion

The proposed data-driven framework was used to characterize the extreme SSL during the passage of tropical cyclones over two tidal gauging stations in Hong Kong during 1979–2018. To evaluate robustness of the framework, the first two-thirds of SSL samples and the corresponding climatological forcing factors were used to select the predictors and to calibrate the framework. The remaining one-third of SSL samples were used to validate the framework. Figure 4 presents Kendall's τ partial rank correlation coefficients between the extreme SSL and the climatological forcing factors. The extreme SSL is highly correlated with the geopotential heights in 700 hPa (700Z), the horizontal wind (Uwind), the temperature in 850 hPa (850T), and the spatial gradient of sea level pressure (GRDslp) over the two stations. Precipitation does not show a significant correlation with the SSL as previous studies (Bevacqua et al., 2019; Wahl et al., 2015) since the correlation is measured with the effect of other forcing factors removed, and precipitation is not as important as other forcing factors (e.g., wind and pressure gradient). The four forcing factors and the SSL were transformed to uniform margins on [0, 1] based on their optimal parametric probability distributions, which were selected based on the K-S test and estimated by the maximum likelihood method. Table 2 presents the K-S test results for the optimal distributions of the SSL and the four forcing factors. Figure 5 presents the empirical cumulative probability of observations and the theoretical cumulative probability calculated by the optimal probability distribution. The theoretical cumulative probability is overall consistent with the empirical one for all variables. The five uniform margins were used to construct the multivariate vine copula models, including 24 C-vine and 24 D-vine models, thereby leading to the model structural uncer-

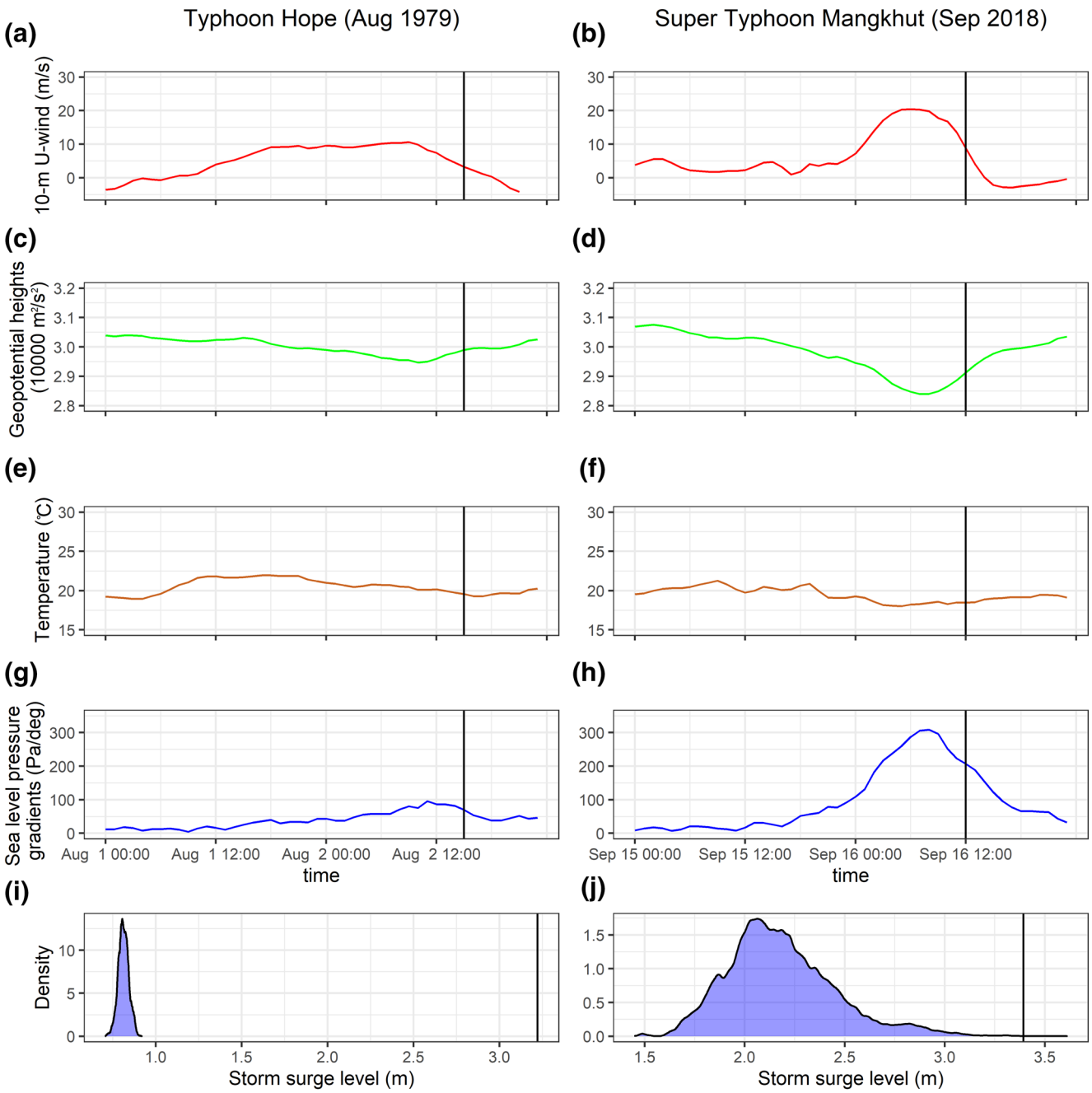

Figure 14. The meteorological factors over the Tai Po Kau station during the passage of Typhoon Hope (1979 August) and Super Typhoon Mangkhut (2018 September). (i) and (j) present the probability density of the storm surge levels predicted by the proposed data-driven framework. The vertical lines in (a–h) and (i–j) correspond to the time and magnitude of observed extreme storm surges, respectively.

tainty in extreme SSL simulations. Figure 6 presents four representative structures for the five-dimensional vine copulas, including two C-vine and two D-vine copulas. The multi-structure vine copula models were calibrated based on the first two-thirds samples (i.e., 151 and 146 samples over the QB and TPK stations, respectively, during 1979–2006), and then validated for the last one-third samples (i.e., 75 and 73 samples over the QB and TPK stations, respectively, during 2007–2018) by comparing against extreme SSL observations. Figures 7 and 8 present the AIC values of the 48 conditional vine copula models constructed based on the sequential maximal spanning tree algorithm proposed by Dißmann et al. (2013) and the corresponding root mean squared error (RMSE) values over the QB and TPK stations, respectively, for the calibration and

validation periods. The C-vine and D-vine models are distinguished by different colors. The number on the y-axis indicates the order of the vine copula model based on AIC. For example, d-vine-24 represents a D-vine model with the highest AIC among the 24 D-vine copulas. Results show that a lower AIC value does not necessarily indicate a better SSL characterization for the conditional vine copula models. For example, d-vine-24 is commonly not selected as the optimal model structure since it has the highest AIC, but its model bias is considerably lower than c-vine-1 that has the lowest AIC over the QB station for the calibration and validation periods.

Since the model selection based on AIC fails to guarantee the optimal SSL characterization, it is necessary to use an advanced model selection approach to improve the multi-structure SSL simulation. The proposed data-driven framework uses BMA to intelligently assign weights to each single-vine forecast based on their model performance, enhancing the reliability of extreme SSL characterization. Moreover, the MCMC simulations using the DREAM algorithm were performed to estimate the posterior distributions of the BMA parameters. Figure 9 presents the evolution of the convergence diagnostic $\hat{R}$ values for the MCMC simulations. The $\hat{R}$ values of different BMA parameters are depicted with colored lines. All the BMA parameters converge to the posterior distribution since the $\hat{R}$ values drop below the critical threshold of 1.2 within the total number of 250,000 evaluations.

Figure 10 presents the posterior estimates of the MCMC-derived BMA weights and their uncertainty ranges of the individual single-structure SSL simulation. Only the 10 largest weights are shown for better visualization. Results show that the D-vine structure makes a dominant contribution in the multi-structure simulation to reproducing the extreme SSL. To evaluate the performance of the BMA-based data-driven framework in characterizing the extreme SSL, Figure 11 presents the Taylor Diagram to visualize the consistency between the predicted and observed SSL for the calibration and validation periods. The single-structure simulations, including C-vine and D-vine copulas, were also presented and distinguished by different colors in Figure 11. Each point in Figure 11 represents an SSL forecast, and the forecast agrees better with the observation if it has a higher correlation and a more consistent standard deviation with the observation, as well as it lies nearer the "OBS." Results indicate that the proposed data-driven framework substantially improves the SSL simulation upon each single-structure forecast by calculating the weighted average forecasts. The D-vine structures lead to lower model biases compared to the C-vine structures for the QB station (see Figures 11a and 11b).

Figure 12 presents the relative performance of the SSL simulations generated from the proposed data-driven framework, the PCR, and the RF. Results indicate that the proposed data-driven framework can lead to better performance in characterizing the extreme SSL than the PCR and RF. For example, the PCR-based and MLR-based SSL simulations have correlation coefficients of approximately 0.81 and 0.65, respectively, with the observations over the QB station for the calibration period (see Figure 12a), whereas the proposed data-driven framework leads to a higher correlation coefficient of 0.89. The model bias was also reduced by the data-driven framework compared with the PCR and RF approaches. For example, the RMSE value of the SSL simulations generated by both the PCR and RF is 0.22 m, which is reduced to 0.17 m by the proposed data-driven framework over the QB station for the validation period. Although the data-driven framework leads to a similar performance compared to the PCR approach over the TPK station for the calibration period (Figure 12c), the correlation is largely increased and the RMSE is largely reduced over the validation period, indicating a higher model performance. This indicates that the proposed data-driven framework can improve the characterization of extreme storm surges upon previous approaches by constructing a probabilistic ensemble of dependence structures of multiple driving forcings, thereby enhancing the reliability and robustness of extreme SSL simulations and coastal flood risk assessments.

To further evaluate the robustness of the data-driven framework, a 10-fold cross validation has been performed to compare the predicted and observed SSL. The Pearson's r and the RMSE have also been used to quantitatively evaluate model performance. Figure 13 presents scatter plots of the 10-fold cross validation results based on the PCR, RF, and the proposed data-driven framework. Results show that the proposed data-driven framework leads to slightly higher performance than the PCR and RF in terms of Pearson's r, while the three models reach almost similar results in terms of RMSE. For example, the Pearson's r generated from the PCR, RF, and the proposed data-driven framework is 0.77, 0.73, and 0.79, respectively, over the QB station (Figures 13a–13c), while the corresponding RMSE is 0.18, 0.19, and 0.18 m.

Figure 13 also indicates that the proposed data-driven framework and the benchmark approaches may underpredict the extreme storm surges, as highlighted by the red rectangles. To understand the underlying reason and further illustrate the data-driven framework, the extreme storm surges during the passage of two tropical cyclones, namely Typhoon Hope (1979 August) and Super Typhoon Mangkhut (2018 September) were explored. Figures 14i and 14j depict the probability density of the maximum SSL predicted by the proposed data-driven framework over the Tai Po Kau station during the passage of the two tropical cyclones, while the vertical lines represent the observed maximum SSL. The probability density results from the uncertainty in the vine copula structures and the marginal distribution of SSL. The observed maximum SSLs are 3.23 and 3.4 m during the passage of the two tropical cyclones. The extreme SSL during the passage of Super Typhoon Mangkhut is captured by the prediction intervals of the data-driven framework, although an underprediction exists (Figure 14j). In comparison, the framework shows a significant underprediction during the passage of Typhoon Hope (Figure 14i). Figures 14a–14h present the temporal evolution of the meteorological forcing factors during the passage of the two tropical cyclones. There are significant peaks of wind and pressure gradients during the passage of Super Typhoon Mangkhut (Figures 14b and 14h). Such a pattern, however, is not observed during the passage of Typhoon Hope (Figures 14a and 14g). A possible reason for such a difference is the relatively high translation speed of Typhoon Hope (Hong Kong Observatory, 1979), which renders the hourly ERA5 data set unable to capture the instant extremes of meteorological factors. This can also be the reason for the underpredicted SSL over the TBK station.

7. Summary and Conclusions

In this study, we propose a robust data-driven framework to improve the characterization of extreme storm surges induced by tropical cyclones. This framework links the extreme SSL with the climatological forcing factors by developing a multi-structure ensemble of conditional vine copula simulations. The model structural uncertainty in vine copula simulations was explicitly addressed in a Bayesian framework. The proposed framework was applied to two tidal gauging stations in Hong Kong during 1979–2018, and was also compared with the existing well-known approaches used to characterize storm surges.

The proposed data-driven framework improves upon the characterization of extreme storm surges not only by taking into account the joint evolution of multiple mechanisms causing the extreme SSL through the multi-structure vine copula simulations but also by addressing the underlying uncertainties in model structures and weights. Our findings show that the model structure selection based on AIC fails to guarantee the optimal SSL simulation for the conditional vine copula model. D-vine copula makes a dominant contribution in the multi-structure ensemble SSL simulation compared with C-vine copula. The proposed framework also shows better performance in comparison with previous approaches in characterizing extreme storm surges. This framework can be directly applicable to simulating a variety of extreme events driven by multiple interdependent factors, which is an effective and efficient approach for enhancing the credibility of the risk assessment of extreme events and natural hazards.

Data Availability Statement

The extreme storm surge observations used in this study are freely available from Hong Kong Observatory and at https://www.hko.gov.hk/en/wservice/tsheet/pms/stormsurgedb.htm#. The ERA5 gridded reanalysis products are provided by the ECMWF, which can be obtained from their website (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview).

Acknowledgments
This research was supported by the National Natural Science Foundation of China (grant no. 51809223), the Hong Kong Research Grants Council Early Career Scheme (grant no. PP5Z), and the Hong Kong Polytechnic University Research Grant (grant no. ZE8S, ZVN6).

References

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, *44*(2), 182–198. https://doi.org/10.1016/j.insmatheco.2007.02.001

Arns, A., Wahl, T., Wolff, C., Vafeidis, A. T., Haigh, I. D., Woodworth, P., et al. (2020). Non-linear interaction modulates global extreme sea levels, coastal flood exposure, and impacts. *Nature Communications*, *11*(1), 1918. https://doi.org/10.1038/s41467-020-15752-5

Beisiegel, N., Vater, S., Behrens, J., & Dias, F. (2020). An adaptive discontinuous Galerkin method for the simulation of hurricane storm surge. *Ocean Dynamics*, *70*(5), 641–666. https://doi.org/10.1007/s10236-020-01352-w

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., & Vrac, M. (2017). Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in Ravenna (Italy). *Hydrology and Earth System Sciences*, *21*(6), 2701–2723. https://doi.org/10.5194/hess-21-2701-2017

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., & Widmann, M. (2019). Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change. *Science Advances*, *5*(9), eaaw5531. https://doi.org/10.1126/sciadv.aaw5531

Bilskie, M. V., Hagen, S. C., Medeiros, S. C., Cox, A. T., Salisbury, M., & Coggin, D. (2016). Data and numerical analysis of astronomic tides, wind-waves, and hurricane storm surge along the northern Gulf of Mexico. *Journal of Geophysical Research: Oceans*, *121*, 3625–3658. https://doi.org/10.1002/2015JC011400

Bloemendaal, N., Muis, S., Haarsma, R. J., Verlaan, M., Irazoqui Apecechea, M., de Moel, H., et al. (2019). Global modeling of tropical cyclone storm surges using high-resolution forecasts. *Climate Dynamics*, *52*(7–8), 5031–5044. https://doi.org/10.1007/s00382-018-4430-x

Brown, J. D., Spencer, T., & Moeller, I. (2007). Modeling storm surge flooding of an urban area with particular reference to modeling uncertainties: A case study of Canvey Island, United Kingdom. *Water Resources Research*, *43*, W06402. https://doi.org/10.1029/2005WR004597

Calafat, F. M., & Marcos, M. (2020). Probabilistic reanalysis of storm surge extremes in Europe. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(4), 1877–1883. https://doi.org/10.1073/pnas.1913049117

Chang, B., Pan, S., & Joe, H. (2020). Vine copula structure learning via Monte Carlo tree search. In *AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics*. Okinawa, Japan.

Chen, H., Wang, S., & Wang, Y. (2020). Exploring abrupt alternations between wet and dry conditions on the basis of historical observations and convection-permitting climate model simulations. *Journal of Geophysical Research: Atmosphere*, *125*, e2019JD031982. https://doi.org/10.1029/2019JD031982

Chen, H., Wang, S., Zhu, J., & Zhang, B. (2020). Projected changes in abrupt shifts between dry and wet extremes over China through an ensemble of regional climate model simulations. *Journal of Geophysical Research: Atmosphere*, *125*, e2020JD033894. https://doi.org/10.1029/2020JD033894

Cid, A., Camus, P., Castanedo, S., Méndez, F. J., & Medina, R. (2017). Global reconstructed daily surge levels from the 20th century reanalysis (1871–2010). *Global and Planetary Change*, *148*, 9–21. https://doi.org/10.1016/j.gloplacha.2016.11.006

Cid, A., Menéndez, M., Castanedo, S., Abascal, A. J., Méndez, F. J., & Medina, R. (2016). Long-term changes in the frequency, intensity and duration of extreme storm surge events in southern Europe. *Climate Dynamics*, *46*(5–6), 1503–1516. https://doi.org/10.1007/s00382-015-2659-1

Cid, A., Wahl, T., Chambers, D. P., & Muis, S. (2018). Storm surge reconstruction and return water level estimation in southeast Asia for the 20th century. *Journal of Geophysical Research: Oceans*, *123*, 437–451. https://doi.org/10.1002/2017JC013143

Colberg, F., & McInnes, K. L. (2012). The impact of future changes in weather patterns on extreme sea levels over southern Australia. *Journal of Geophysical Research*, *117*(8). https://doi.org/10.1029/2012JC007919

Couasnon, A., Eilander, D., Muis, S., Veldkamp, T. I. E., Haigh, I. D., Wahl, T., et al. (2020). Measuring compound flood potential from river discharge and storm surge extremes at the global scale. *Natural Hazards and Earth System Sciences*, *20*(2), 489–504. https://doi.org/10.5194/nhess-20-489-2020

Cyriac, R., Dietrich, J. C., Fleming, J. G., Blanton, B. O., Kaiser, C., Dawson, C. N., & Luettich, R. A. (2018). Variability in coastal flooding predictions due to forecast errors during Hurricane Arthur. *Coastal Engineering*, *137*, 59–78. https://doi.org/10.1016/j.coastaleng.2018.02.008

Dangendorf, S., Müller-Navarra, S., Jensen, J., Schenk, F., Wahl, T., & Weisse, R. (2014). North sea storminess from a novel storm surge record since AD 1843. *Journal of Climate*, *27*(10), 3582–3595. https://doi.org/10.1175/JCLI-D-13-00427.1

Devlin, A. T., Pan, J., & Lin, H. (2019). Tidal variability in the Hong Kong region. *Ocean Science*, *15*(4), 853–864. https://doi.org/10.5194/os-15-853-2019

Dietrich, J. C., Westerink, J. J., Kennedy, A. B., Smith, J. M., Jensen, R. E., Zijlema, M., et al. (2011). Hurricane Gustav (2008) waves and storm surge: Hindcast, synoptic analysis, and validation in southern Louisiana. *Monthly Weather Review*, *139*(8), 2488–2522. https://doi.org/10.1175/2011MWR3611.1

Dißmann, J., Brechmann, E. C., Czado, C., & Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, *59*(1), 52–69. https://doi.org/10.1016/j.csda.2012.08.010

Duan, Q., & Phillips, T. J. (2010). Bayesian estimation of local signal and noise in multimodel simulations of climate change. *Journal of Geophysical Research*, *115*(18), 1–15. https://doi.org/10.1029/2009JD013654

Dube, S. K., Jain, I., Rao, A. D., & Murty, T. S. (2009). Storm surge modelling for the Bay of Bengal and Arabian Sea. *Natural Hazards*, *51*, 3–27. https://doi.org/10.1007/s11069-009-9397-9

Dullaart, J. C. M., Muis, S., Bloemendaal, N., & Aerts, J. C. J. H. (2020). Advancing global storm surge modelling using the new ERA5 climate reanalysis. *Climate Dynamics*, *54*(1–2), 1007–1021. https://doi.org/10.1007/s00382-019-05044-0

Fernández-Montblanc, T., Vousdoukas, M. I., Ciavola, P., Voukouvalas, E., Mentaschi, L., Breyiannis, G., et al. (2019). Towards robust pan-European storm surge forecasting. *Ocean Modelling*, *133*, 129–144. https://doi.org/10.1016/j.ocemod.2018.12.001

Flowerdew, J., Mylne, K., Jones, C., & Titley, H. (2013). Extending the forecast range of the UK storm surge ensemble. *Quarterly Journal of the Royal Meteorological Society*, *139*(670), 184–197. https://doi.org/10.1002/qj.1950

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gräwe, U., & Burchard, H. (2012). Storm surges in the Western Baltic Sea: The present and a possible future. *Climate Dynamics*, *39*(1–2), 165–183. https://doi.org/10.1007/s00382-011-1185-z

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2013). Projected Atlantic hurricane surge threat from rising temperatures. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(14), 5369–5373. https://doi.org/10.1073/pnas.1209980110

Gruber, L., & Czado, C. (2015). Sequential Bayesian model selection of regular vine copulas. *Bayesian Analysis*, *10*(4), 937–963.

Gruber, L., & Czado, C. (2018). Bayesian model selection of regular vine copulas. *Bayesian Analysis*, *13*(4), 1107–1131. https://doi.org/10.1214/17-BA1089

Haigh, I. D., MacPherson, L. R., Mason, M. S., Wijeratne, E. M. S., Pattiaratchi, C. B., Crompton, R. P., & George, S. (2014). Estimating present day extreme water level exceedance probabilities around the coastline of Australia: Tropical cyclone-induced storm surges. *Climate Dynamics*, *42*(1–2), 139–157. https://doi.org/10.1007/s00382-012-1653-0

Hong Kong Observatory (1979). Typhoon Hope 28 July – 3 August 1979. *Meteorological Results 1979, Part III - Tropical Cyclone Summaries. Hong Kong*.

Horsburgh, K. J., & Wilson, C. (2007). Tide-surge interaction and its role in the distribution of surge residuals in the North Sea. *Journal of Geophysical Research*, *112*(C8). https://doi.org/10.1029/2006JC004033

Howard, T., Lowe, J., & Horsburgh, K. (2010). Interpreting century-scale changes in southern north sea storm surge climate derived from coupled model simulations. *Journal of Climate*, *23*(23), 6234–6247. https://doi.org/10.1175/2010JCLI3520.1

Hu, K., Ding, P., Wang, Z., & Yang, S. (2009). A 2D/3D hydrodynamic and sediment transport model for the Yangtze Estuary, China. *Journal of Marine Systems*, *77*(1–2), 114–136. https://doi.org/10.1016/j.jmarsys.2008.11.014

Joe, H. (1996). *Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters (pp. 120–141)*. https://doi.org/10.1214/lnms/1215452614

Kim, S. W., Melby, J. A., Nadal-Caraballo, N. C., & Ratcliff, J. (2015). A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling. *Natural Hazards*, *76*(1), 565–585. https://doi.org/10.1007/s11069-014-1508-6

Lall, U., Devineni, N., & Kaheil, Y. (2016). An empirical, nonparametric simulator for multivariate random variables with differing marginal densities and nonlinear dependence with hydroclimatic applications. *Risk Analysis*, *36*(1), 57–73. https://doi.org/10.1111/risa.12432

Lee, T. L. (2006). Neural network prediction of a storm surge. *Ocean Engineering*, *33*(3–4), 483–494. https://doi.org/10.1016/j.oceaneng.2005.04.012

Lee, T. C., & Wong, C. F. (2007). Historical storm surges and storm surge forecasting in Hong Kong. Hong Kong observatory, paper for the JCOMM Scientific and Technical Symposium on Storm Surges (SSS) in Seoul.

Liu, Z., Cheng, L., Hao, Z., Li, J., Thorstensen, A., & Gao, H. (2018). A framework for exploring joint effects of conditional factors on compound floods. *Water Resources Research*, *54*(4), 2681–2696. https://doi.org/10.1002/2017WR021662

Liu, Z., Zhou, P., Chen, X., & Guan, Y. (2015). A multivariate conditional model for streamflow prediction and spatial precipitation refinement. *Journal of Geophysical Research*, *120*(19), 10116–10129. https://doi.org/10.1002/2015JD023787

Luettich, R. A., Westerink, J. J., & Scheffner, N. W. (1992). ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries. Report 1, Theory and methodology of ADCIRC-2DD1 and ADCIRC-3DL. Dredging Research Program Technical Report DRP-92-6. *Vicksburg, MS: U.S. Army Engineers Waterways Experiment Station*.

MacPherson, L. R., Arns, A., Dangendorf, S., Vafeidis, A. T., & Jensen, J. (2019). A stochastic extreme sea level model for the German Baltic Sea coast. *Journal of Geophysical Research: Oceans*, *124*, 2054–2071. https://doi.org/10.1029/2018JC014718

Madadgar, S., & Moradkhani, H. (2014). Improved Bayesian multimodeling: Integration of copulas and Bayesian model averaging. *Water Resources Research*, *50*(12), 9586–9603. https://doi.org/10.1002/2014WR015965

Manning, C., Widmann, M., Bevacqua, E., Van Loon, A. F., Maraun, D., & Vrac, M. (2018). Soil moisture drought in Europe: A compound event of precipitation and potential evapotranspiration on multiple time scales. *Journal of Hydrometeorology*, *19*(8), 1255–1271. https://doi.org/10.1175/jhm-d-18-0017.1

Marcos, M., Rohmer, J., Vousdoukas, M. I., Mentaschi, L., Le Cozannet, G., & Amores, A. (2019). Increased extreme coastal water levels due to the combined action of storm surges and wind waves. *Geophysical Research Letters*, *46*(8), 4356–4364. https://doi.org/10.1029/2019GL082599

Masina, M., Lamberti, A., & Archetti, R. (2015). Coastal flooding: A copula based approach for estimating the joint probability of water levels and waves. *Coastal Engineering*, *97*, 37–52. https://doi.org/10.1016/j.coastaleng.2014.12.010

Min, A., & Czado, C. (2010). Bayesian inference for multivariate copulas using pair-copula constructions. *Journal of Financial Econometrics*, *8*(4), 511–546. https://doi.org/10.1093/jjfinec/nbp031

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Compounding effects of sea level rise and fluvial flooding. *Proceedings of the National Academy of Sciences*, *114*(37), 9785–9790. https://doi.org/10.1073/pnas.1620325114

Morrow, B. H., Lazo, J. K., Rhome, J., & Feyen, J. (2015). Improving storm surge risk communication: Stakeholder perspectives. *Bulletin of the American Meteorological Society*, *96*(1), 35–48. https://doi.org/10.1175/BAMS-D-13-00197.1

Muis, S., Haigh, I. D., Guimarães Nobre, G., Aerts, J. C. J. H., & Ward, P. J. (2018). Influence of El Niño-Southern Oscillation on global coastal flooding. *Earth's Future*, *6*(9), 1311–1322. https://doi.org/10.1029/2018EF000909

Muis, S., Verlaan, M., Winsemius, H. C., Aerts, J. C. J. H., & Ward, P. J. (2016). A global reanalysis of storm surges and extreme sea levels. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11969

Nagler, T., Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., & Erhardt, T. (2019). *VineCopula: Statistical inference of vine copulas. R package version 2.3.0*. Retrieved from https://cran.r-project.org/package=VineCopula

Needham, H. F., Keim, B. D., & Sathiaraj, D. (2015). A review of tropical cyclone-generated storm surges: Global data sources, observations, and impacts. *Reviews of Geophysics*, *53*, 545–591. https://doi.org/10.1002/2014RG000477

Pawlowicz, R., Beardsley, B., & Lentz, S. (2002). Classical tidal harmonic analysis including error estimates in MATLAB using T_TIDE. *Computers & Geosciences*, *28*(8), 929–937. https://doi.org/10.1016/S0098-3004(02)00013-4

Pereira, G., & Veiga, Á. (2018). PAR(p)-vine copula based model for stochastic streamflow scenario generation. *Stochastic Environmental Research and Risk Assessment*, *32*(3), 833–842. https://doi.org/10.1007/s00477-017-1411-2

Qing, Y., Wang, S., Zhang, B., & Wang, Y. (2020). Ultra-high resolution regional climate projections for assessing changes in hydrological extremes and underlying uncertainties. *Climate Dynamics*, *55*(7–8), 2031–2051. https://doi.org/10.1007/s00382-020-05372-6

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Raftery, A. E., Gneiting, T., Balabdaoui, F., & Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, *133*(5), 1155–1174. https://doi.org/10.1175/MWR2906.1

Raftery, A. E., Madigan, D., & Hoeting, J. A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, *92*(437), 179–191. https://doi.org/10.1080/01621459.1997.10473615

Ramos-Valle, A. N., Curchitser, E. N., & Bruyère, C. L. (2020). Impact of Tropical Cyclone Landfall Angle on Storm Surge Along the Mid-Atlantic Bight. *Journal of Geophysical Research: Atmosphere*, *125*(4). https://doi.org/10.1029/2019JD031796

Reed, A. J., Mann, M. E., Emanuel, K. A., Lin, N., Horton, B. P., Kemp, A. C., & Donnelly, J. P. (2015). Increased threat of tropical cyclones and coastal flooding to New York City during the anthropogenic era. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), 12610–12615. https://doi.org/10.1073/pnas.1513127112

Rego, J. L., & Li, C. (2010). Nonlinear terms in storm surge predictions: Effect of tide and shelf geometry with case study from Hurricane Rita. *Journal of Geophysical Research*, *115*(6). https://doi.org/10.1029/2009JC005285

Royston, S., Lawry, J., & Horsburgh, K. (2013). A linguistic decision tree approach to predicting storm surge. *Fuzzy Sets and Systems*, *215*, 90–111. https://doi.org/10.1016/j.fss.2012.10.001

Sahoo, B., & Bhaskaran, P. K. (2019). Prediction of storm surge and coastal inundation using Artificial Neural Network – A case study for 1999 Odisha Super Cyclone. *Weather and Climate Extremes*, *23*, 100196. https://doi.org/10.1016/j.wace.2019.100196

Sheridan, S. C., Lee, C. C., Adams, R. E., Smith, E. T., Pirhalla, D. E., & Ransibrahmanakul, V. (2019). Temporal modeling of anomalous coastal sea level values using synoptic climatological patterns. *Journal of Geophysical Research: Oceans*, *124*, 6531–6544. https://doi.org/10.1029/2019JC015421

Shriver, J. F., Richman, J. G., & Arbic, B. K. (2014). How stationary are the internal tides in a high-resolution global ocean circulation model? *Journal of Geophysical Research: Oceans*, *119*, 2769–2787. https://doi.org/10.1002/2013JC009423

Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris*, *8*, 229–231.

Tadesse, M., Wahl, T., & Cid, A. (2020). Data-driven modeling of global storm surges. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.00260

Tosunoglu, F., & Singh, V. P. (2018). Multivariate modeling of annual instantaneous maximum flows using copulas. *Journal of Hydrologic Engineering*, *23*(3), 04018003. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001644

Trepanier, J. C., Yuan, J., & Jagger, T. H. (2017). The combined risk of extreme tropical cyclone winds and storm surges along the U.S. Gulf of Mexico Coast. *Journal of Geophysical Research: Atmosphere*, *122*, 3299–3316. https://doi.org/10.1002/2016JD026180

Vousdoukas, M. I., Voukouvalas, E., Annunziato, A., Giardino, A., & Feyen, L. (2016). Projections of extreme storm surge levels along Europe. *Climate Dynamics*, *47*(9–10), 3171–3190. https://doi.org/10.1007/s00382-016-3019-5

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2007wr006720

Wahl, T., Jain, S., Bender, J., Meyers, S. D., & Luther, M. E. (2015). Increasing risk of compound flooding from storm surge and rainfall for major US cities. *Nature Climate Change*, *5*(12), 1093–1097. https://doi.org/10.1038/nclimate2736

Wang, S., Ancell, B. C., Huang, G. H., & Baetz, B. W. (2018). Improving robustness of hydrologic ensemble predictions through probabilistic pre- and post-processing in sequential data assimilation. *Water Resources Research*, *54*(3), 2129–2151. https://doi.org/10.1002/2018WR022546

Wang, W., Dong, Z., Lall, U., Dong, N., & Yang, M. (2019). Monthly streamflow simulation for the headwater catchment of the Yellow River basin With a hybrid statistical-dynamical model. *Water Resources Research*, *55*, 7606–7621. https://doi.org/10.1029/2019WR025103

Wang, S., & Wang, Y. (2019). Improving probabilistic hydroclimatic projections through high-resolution convection-permitting climate modeling and Markov chain Monte Carlo simulations. *Climate Dynamics*, *53*(3–4), 1613–1636. https://doi.org/10.1007/s00382-019-04702-7

Wang, S., & Zhu, J. (2020). Amplified or exaggerated changes in perceived temperature extremes under global warming. *Climate Dynamics*, *54*(1–2), 117–127. https://doi.org/10.1007/s00382-019-04994-9

Wong, T. E., Klufas, A., Srikrishnan, V., & Keller, K. (2018). Neglecting model structural uncertainty underestimates upper tails of flood hazard. *Environmental Research Letters*, *13*(7). https://doi.org/10.1088/1748-9326/aacb3d

Wu, W., McInnes, K., O'Grady, J., Hoeke, R., Leonard, M., & Westra, S. (2018). Mapping dependence between extreme rainfall and storm surge. *Journal of Geophysical Research: Oceans*, *123*, 2461–2474. https://doi.org/10.1002/2017JC013472

Zhang, B., Wang, S., & Wang, Y. (2019). Copula-based convection-permitting projections of future changes in multivariate drought characteristics. *Journal of Geophysical Research: Atmosphere*, *124*, 7460–7483. https://doi.org/10.1029/2019JD030686

Zheng, F., Westra, S., Leonard, M., & Sisson, S. A. (2014). Modeling dependence between extreme rainfall and storm surge to estimate coastal flooding risk. *Water Resources Research*, *50*(3), 2050–2071. https://doi.org/10.1002/2013WR014616

Zou, Q. P., Chen, Y., Cluckie, I., Hewston, R., Pan, S., Peng, Z., & Reeve, D. (2013). Ensemble prediction of coastal flood risk arising from overtopping by linking meteorological, ocean, coastal and surf zone models. *Quarterly Journal of the Royal Meteorological Society*, *139*(671), 298–313. https://doi.org/10.1002/qj.2078