Coastal Engineering 125 (2017) 42–50

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

Applying a Bayesian network based on Gaussian copulas to model the hydraulic boundary conditions for hurricane flood risk analysis in a coastal watershed

A. Sebastian[a,b,*], E.J.C. Dupuits[b], O. Morales-Nápoles[b]

[a] Department of Civil and Environmental Engineering, Rice University, Houston, TX, United States
[b] Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:
Tropical cyclone
Flood risk
Compound flooding
Storm surge
Bayesian networks
Copula

MSC:
00-01
99-00

ABSTRACT

In recent years significant emphasis has been placed on quantifying coastal flood hazards in the U.S. using high resolution 2-D hydrodynamic and nearshore wave models. However, these studies are computationally expensive and often neglect to consider the flooding that arises from the combined hazards of precipitation and storm surge in coastal watersheds. This paper describes a method to stochastically simulate a large number of combinations of peak storm surge and cumulative precipitation to determine the hydraulic boundary conditions for a low-lying coastal watershed draining into a semi-enclosed tidal bay. The method is computationally efficient and takes into consideration five tropical cyclone characteristics at landfall: wind-speed, angle of approach, landfall location, radius of maximum winds, and forward speed. A precipitation gage network and tidal gage data were used, along with observations from over 300 tropical cyclones in the Gulf of Mexico. A Non-parametric Bayesian Network was built to generate 100,000 synthetic storm events and used as input to an empirical wind set-up model to simulate storm surge within a tidal bay and at the downstream boundary of the watershed. Based on the results, probable combinations of cumulative precipitation and peak storm surge for the watershed during hurricane conditions are determined. These boundary conditions can be easily incorporated into a coastal riverine model to determine flood risk in the watershed.

1. Introduction

Integrated flood risk assessment is critical for the determination of appropriate flood mitigation strategies for heavily populated, low-lying coastal areas. While the number of deaths from tropical cyclones have decreased with the development of advanced prediction and early warning systems, economic losses have increased exponentially due to rapid population growth and urban development near the coast. It is estimated that today almost half of the global population lives within 150 km of a coastline [1]. These highly urbanized coastal areas are threatened by the combined impacts of severe storms, especially hurricane-induced storm surge and heavy precipitation. Communities along the U.S. Gulf Coast and in delta regions around the world, where storm surge often coincides with heavy precipitation, are especially vulnerable.

In the United States, the 100-year Federal Emergency Management Agency (FEMA) floodplain is used as the primary instrument for delineating and mitigating flood risk. This boundary, indicating the 1% percent chance of inundation each year from riverine *or* coastal flooding, drives federal flood insurance requirements, household protective actions, and local mitigation policies. It also determines where future development can take place and what it will look like. However, increasing evidence suggests that the FEMA 100-year floodplain is a poor predictor of actual flood damage and that in some watersheds, upwards of 50% of insured losses are occurring outside of the demarcated flood hazard areas [2]. This is especially apparent along the Gulf Coast where insured assets account for 41% of flood insurance policies in the United States, but amounted to more than 80% of claim payouts between 1978 and 2010 [3].

One of the primary contributing factors to floodplain inaccuracy is the age of existing FEMA floodplain maps; more than 60% of the maps are at least 10 years old and in some coastal areas, maps date back as far as the mid-to-late 1970s [4,5]. In addition to their age, riverine floodplains are derived using deterministic hydrologic and hydraulic models based on a single design storm. This leads to compounding uncertainties since the modeled floodplain is only as accurate as the information used to define them, including the original assumptions made about the design storm. Other sources of inaccuracy include:

* Corresponding author at: Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, the Netherlands.
E-mail address: a.g.sebastian@tudelft.nl (A. Sebastian).

http://dx.doi.org/10.1016/j.coastaleng.2017.03.008
Received 30 August 2016; Received in revised form 24 January 2017; Accepted 4 March 2017

limited historical hydrometeorological observations [6]; spatio-temporal variations in precipitation [7]; and changes in climate, topography, and land use conditions [8].

Similarly, early FEMA coastal floodplains were generated using a design storm, or Standard Project Hurricane (SPH), which was intended to represent a probable, yet infrequent hurricane along a section of the coast [9]. These design storms were often based on a single storm characteristic: intensity, derived from historical storm data prior to 1960. The relative calm in the period prior to 1960 and the oversimplification of hurricane behavior typically led to an underestimation of surge heights at a particular coastal location [10]. In subsequent studies, attempts were made to assess return period surge using historical water level records; however, the lack of extreme water level events at a single coastal location coupled with anonymously high single-storm records made these estimates difficult to resolve [10]. In response, a move toward generating a probabilistic suite of storms was made in the 1970s and 1980s and a Joint Probability Method (JPM) approach based on cluster analysis of hurricane characteristics at landfall was developed to generate parametric wind fields at landfall [11]. However, the sources of uncertainty in this approach were quite high, primarily due to lack of historical data [10].

With the advent of complex computing and high-resolution storm surge and wave models, renewed interest in improving joint probability methods (JPM) for coastal modeling has taken place. Recently, JPM-Optimum Sampling (JPM-OS) was introduced as a method to reduce the number of representative synthetic storms needed for simulating storm surge at any particular section of the coast. This method is being applied in many of the current FEMA Flood Insurance Studies (to be completed in 2020). However, even with the application of the JPM-OS approach, probabilistic modeling using high-resolution storm surge models is computationally demanding, requiring large computer clusters to run [10,12,13]. Furthermore, despite the marked improvement in coastal modeling and storm surge mapping, the new FEMA floodplain maps still neglect to consider the flood risk resulting from the interaction between rainfall-runoff and storm surge at the coast. In small, low-lying coastal watersheds, where hydrologic response to precipitation is nearly instantaneous, determining the joint exceedance of precipitation and storm surge is critical for assessing flood risk. Here, small variations in downstream elevation extend far upstream, impeding the propagation and exit of the precipitation-induced flood wave from the watershed.

This paper introduces a method for quantifying hurricane boundary conditions for small, urbanized coastal watersheds draining into a tidally influenced, semi-enclosed bay systems using a non-parametric Bayesian network (NPBN) based on Gaussian copulas. The NPBN is used to generate a suite of synthetic tropical cyclones and the events are input into an empirical wind setup model to stochastically simulate a large number of storms in the bay system. The modeled combinations of storm surge and precipitation provide an initial estimate of joint exceedance probabilities for a coastal watershed.

As a case study, the method is applied to the Clear Creek Watershed located 32 km southeast of Houston, Texas on the west side of Galveston Bay (see Section 3). Galveston Bay creates a complex environment where surge is often higher on the north and west side of the bay than on the east due to local wind-setup caused by counter-clockwise hurricane winds over the Bay [14]. Furthermore, the combined impacts of little topographic relief, slow or limited infiltration, rapid urban development, regional subsidence and sea level rise, and intense storms have led to frequent and severe flooding in the watershed. In the next section, we provide an overview of non-parametric Bayesian networks and introduce key terms. This is followed by a description of the study area and an overview of the method, model results, discussion and conclusion.

2. Bayesian networks

Bayesian Networks (BN) are probabilistic graphical models that can be used to represent a large number of interdependent variables [15–17]. The variables in the network can be either discrete or continuous, and their dependency on one another is quantified by conditional probability functions. One of the primary advantages of BNs is that the probability distribution functions in the graph can be easily updated to reflect changes in the joint distribution (i.e., inference). Given their characteristics, BNs can be efficiently sampled to generate large synthetic data sets [18].

BNs are composed of a number of children (successor) and parent (predecessor) ”nodes”, which represent a set of random variables $(X_1, X_2, \ldots, X_n)$. In this paper, the nodes are labeled on the set of positive integers. An ordered pair of elements of this set is called an ”arc” which represents the dependence between each parent-child pair in the graphical network and has a defined direction such that the graph remains acyclic. Together, the nodes and arcs represent the dependence structure between the variables in the model, where the joint distribution of the child nodes can be computed as a product of conditional probability functions:

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\mathbf{x}_{Pa(X_i)}) \tag{1}$$

where $Pa(X_i)$ is the set of parent nodes of x_i, with i=1,...,n. For nodes without parents, $Pa(X_i) = \varnothing$ and the marginal density is used in Eq. (1).

In this study, a non-parametric continuous Bayesian network (NPBN) based on copulas is applied as presented in [17,19,18]. In this model, each continuous random variable is represented by its empirical distribution (hence the non-parametric part of the name in this class of BNs) and the dependence structures in the network are built using bivariate copula of the one-parameter class. Although the term semi-parametric BNs may be more appropriate to describe this type of BN (since the procedure relies on one-parameter copulas), to remain consistent with the previous literature, we use the term NPBN in this paper. An example NPBN with three nodes is shown in Fig. 1.

NPBNs have several advantages over traditional regression methods for application to environmental data. For example, environmental data often exhibits non-linear behavior making it difficult to represent using traditional regression models. NPBNs can capture and model the interdependencies between complex environmental variables. Moreover, the graphical nature of a NPBN makes the dependence configuration between environmental variables explicit [20].

In an NPBN, the arcs between each parent-child pair (i.e., $Pa(X_i) \to X_i$) are represented by one-parameter (conditional) copulas. In its most general form, the bivariate cumulative distribution function $F_{X_i,X_j}(x_i, x_j)$ of the random variables X_i and X_j becomes

$$F_{X_i,X_j}(x_i, x_j) = C[F_{X_i}(x_i), F_{X_j}(x_j)] \tag{2}$$

where $F_{X_i}(x_i)$ and $F_{X_j}(x_j)$ are the marginal distributions of X_i and X_j and there exists a copula, C, that describes their dependence structure [21,46,47]. The advantage of using copulas is that the dependence structure of $F_{X_i,X_j}(x_i, x_j)$ is captured by the copula and the selection of the copula is independent of the marginal distributions of X_i and X_j [22–24].

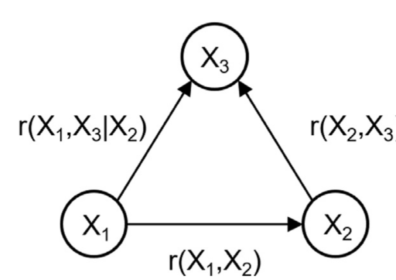

Fig. 1. Example non-parametric Bayesian network (NPBN) with three nodes ($X_1, \ldots X_3$) connected by three arcs. In this paper, each arc represents a bivariate (conditional) copula of the one-parameter class.