Coastal Engineering 117 (2016) 138–156

Contents lists available at ScienceDirect

Coastal Engineering

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

A multivariate statistical model of extreme events: An application to the Catalan coast

J. Lin-Ye*, M. Garcia-Leon, V. Gracia, A. Sanchez-Arcilla

Laboratory of Maritime Engineering, Barcelona Tech, D1 Campus Nord, Jordi Girona 1-3, 08034 Barcelona, Spain

ARTICLE INFO

Article history:
Received 26 February 2015
Received in revised form 6 July 2016
Accepted 2 August 2016
Available online xxxx

Keywords:
Wave storms
Catalan coast
von Mises distribution
Multivariate logit function
Hierarchical Archimedean copula
Generalized Pareto distribution

ABSTRACT

Wave extreme events can be understood as the combination of Storm-intensity, Directionality and Intra-time distribution. However, the dependence structure among these factors is still unclear. A methodology has been developed to model wave-storms whose components are linked together. The model is composed by three parts: an intensity module, a wave directionality module, and an intra-time distribution module. In the Storm-intensity sub-model, generalized Pareto distributions and hierarchical Archimedean copulas have been used to characterize the storm energy, unitary energy, peak wave-period and duration. In the Directionality and Intra-time sub-models, the wave direction (at the peak of the storm) and the storm growth–decay rates are linked to the variables from the intensity model, respectively. The model is applied to the Catalan coast (NW Mediterranean). The outcomes denote spatial patterns that coincide with the state of knowledge. The proposed methodology is able to provide boundary conditions for wave and near-shore studies, saving computational time and establishing the dependence of the proposed variables. Such synthetic storms reproduce the inter-variable co-dependence of the original data.

1. Introduction

Wave storms strongly perturb the state of coastal environments, becoming such changes concomitant with episodic coastal hazards such as coastal flooding and erosion. These extreme phenomena drive complex hydrodynamic processes whose understanding is paramount for proper infrastructure design (Goda, 2010). The conventional approach is usually based on the probabilistic definition of a single parameter, typically the wave height. Other concurrent components as the duration of the storm, the storm total energy and the associated wave period influence the final response of a beach or the damage evolution of a structure (Martin-Soldevilla et al., 2015; Melby and Kobayashi, 2011). These variables are known to be semi-dependent (de Waal and van Gelder, 2005; Salvadori et al., 2007), but the classical methodology either a) assumes one variable to be stochastic and the other ones to be deterministic or, b) assumes all variables to be stochastic but completely independent. In the latter case, the lack of dependence structure hampers finding sets of physically plausible storm components, and requires expert guidance plus local knowledge to discern the suitable combinations.

A common modeling approach is to hindcast high energy events or to synthesize storms to a representative extreme sea-state, which is generally predisposed by the degree of knowledge of the area. For the latter case, dependency structures among the hydrodynamic variables pose a hurdle, as they tend to be unknown. Exploratory methods, such as 2D scatter plots, have been widely used as a rule-of-thumb for the most frequent problem, wave-height vs. wave-period. However, the interpretation of existing co-dependences among several variables is challenging. Recurrently, a wide scatter cloud can mislead about biased co-dependence structures, due to subjective criteria. Storm modeling requires to consider a multivariate analysis of storm parameters (Corbella and Stretch, 2012), as univariate analyses may oversimplify coastal processes, often leading to over or under-estimation of the storm induced damages.

Specialized statistical techniques such as copulas can be used for finding existing relationships among storm variables (Genest and Favre, 2007; Trivedi and Zimmer, 2007) with more objective criteria. Copulas were once described by Sklar (1959), for bivariate models. They were popularized in the 1990s in financial, insurance, econometrical, risk management and actuarial analyses (Cherubini et al., 2004). Applications can also be found in hydrology (De Michele and Salvadori, 2003; Salvadori and De Michele, 2004) and more recently, in coastal engineering (Corbella and Stretch (2012), Wahl et al. (2011); among others).

* Corresponding author.
E-mail address: jue.lin@upc.edu (J. Lin-Ye).

http://dx.doi.org/10.1016/j.coastaleng.2016.08.002
0378-3839/

Corbella and Stretch (2012) employed copula based return-periods to identify the most probable combination of wave-height, wave-period, storm-duration, and water-level for a given probability of exceeding at South Africa. The threshold in the peak-over-threshold method was defined as a critical layer of multiple dimensions that prescribe both a safe and super-critical combination of storm conditions. In the study, the extreme events were fitted to Generalized extreme value distributions (GEVD). They also noted the importance that their statistical model was constrained, to avoid unrealistic results. Hence, they proposed wave steepness as a restriction that can increase model rigidity and enhance system robustness.

Li et al. (2014) fitted maximum significant wave height, peak-wave-period and storm-duration measured in the Dutch Coast with generalized Pareto distributions (GPD). They had used the Kolmogorov–Smirnov and Chi-square tests to evaluate the goodness-of-fit. A similar approach had also been followed by Corbella and Stretch (2013). Salvadori et al. (2014), on the other hand, fitted the significant wave-height and the duration to a Generalized Weibull model (GW) distribution and used Akaike Information Criterion (AIC) to select the suitable copula.

Wahl et al. (2012) applied fully nested Archimedean copulas to consider both storm surge parameters (defined with the highest turning point and the intensity) and the wave height, at the German coast. Nested copulas can characterize multivariate random variables by determining a priori nesting architecture that composes simpler copulas structures into larger and more complex ones. Wahl et al. (2012) firstly characterized the highest turning point and intensity; and then incorporated the significant wave height.

The main objective of this paper is to propose a methodology for inferring multivariate wave storm parameters that shares a common structure. To this aim, one of the main points of the paper has been to propose a dependence structure that links the parameters that explain wave storms. The paper is divided into two steps: Model building and Applicability. The proposed wave storm model has been split into three modules: intensity, wave directionality and intra-time storm distributions. This methodology has been tested on the Catalan coast, a fetch limited environment.

The structure of the paper is as follows: Section 2 deals with the methods for building the proposed statistical model. Section 3 presents the study area and, Section 4, the database used. Results are summarized in Section 5 and discussed in Section 6. Finally, Section 7 sets out the conclusions.

2. Methods

2.1. Storm definition and variables

The determination of storms has three criteria: 1) intensity definition and associated threshold, 2) minimum time-lapse between storms (D^*_{min}), and 3) minimum duration of the storm (D_{min}). Wave storms are extreme phenomena that can be dealt with the peak-over-threshold description (Embrechts et al., 1997). The threshold separates storm conditions from non-storm conditions. The D^*_{min} helps satisfy independence of the samples. The independence is one part of the “independent and equidistributed” assumption for data in many statistical techniques. D_{min} discards the storms of insufficient duration and which are, therefore, of lesser significance.

The usual procedure associates the threshold with the percentile 90 of the wave height (Bernardara et al., 2014; Eastoe et al., 2013). Here, other approaches are proposed. For instance, the occurrence in time of extreme events, for any given geographical location, follows a Poisson distribution. Therefore, it can be deduced that the time lapse between storms must be approximately an exponential distribution; if not, these events are not extreme. Apart from this, the threshold should belong to the linear segment of a mean-excess wave-height function (Ortego et al., 2012). At the same time, the events must be statistically significant in number. The wave-height threshold has been varied ranging from 1.5 m to 3 m, whose minimum doubles the mean wave heights (CIIRC, 2010). The finally selected value of the wave-height threshold is exposed in Section 5 and discussed in Section 6.

Turning to the independence and equal distribution of storm samples, neighboring storms are clustered if the D^* that separates them is below D^*_{min}, which means that both episodes belong to the same storm event. After clustering, each storm can be considered to be independent from the others. On the other hand, it is assumed that the marine extreme events are generated by a limited subset of synoptic conditions (Lionello, 2012), which is true in Western Europe (Mazas et al., 2014). Therefore, the storms are regarded as identically distributed.

Three candidates for D^*_{min} are proposed: 72 h, 48 h, and 12 h. $D^*_{min} = 72$ h is because the two sub-storms in a twin storm tend to be less than 72 h apart. Approximately 20–30% of the total storm events on the Catalan coast are twin, depending on the location (Wojtanowicz, 2010). The consideration of $D^*_{min} = 48$ h is conceptually similar to Tolosana-Delgado et al. (2011), whereas $D^*_{min} = 12$ h is based on direct observations of Catalan sea-storms. A sensitivity test is performed to select the most correct D^*_{min} value. The test consists of representing storms for different values of D^*_{min}. The D^*_{min} selected and the reasons leading to this choice are stated in Section 5 and discussed in Sub-section 6.1.

D is the duration of the event between the first and last threshold crossing (Fig. 1a). It is not to be confounded with D^*. The value of D_{min} is given in Section 5.

From each independent storm, the total storm-energy (E), the maximum storm-unitary-energy ($E_{u,p}$), the peak wave period (T_p), the duration D, the direction of the peak-wave (θ^*_p), the growth-rate and the decay-rate are obtained.

The Storm-intensity sub-model includes E, $E_{u,p}$, T_p, and D.

The E is defined as

$$E = \int_{iniT}^{endT} H^2_{m0} dt, \tag{1}$$

where H_{m0} is the spectral significant wave-height, and t is time. In case that the wave-height returns below the threshold, during the event, the duration and the energy of these low intensity periods are included in the sums of D and E.

It has been highlighted in Sánchez-Arcilla et al. (2014) that the capture with numerical models of the peak-wave-height lacks of exactitude, whereas a better skill is found for the existing temporal trend. Therefore, a new definition of the m_0-wave-height during the storm peak ($H_{m0,peak}$) is proposed through the definition of $E_{u,p}$:

$$E_{u,p} = \max_i \left(mean \left(E_{u,(i-1)} + E_{u,i} + E_{u,(i+1)} \right) \right), \tag{2}$$

where E_u is the unitary storm-energy at each hour. The square root of $E_{u,p}$ is proposed, here, as an improved definition of $H_{m0,peak}$, and is herein called H^*_p.

The H^*_p synthesizes the energy shortly before and after the peak. The subset (see Fig. 1b) presents a) point ($t - 1$): growing to reach the peak, b) point (t): Storm peak and c) point ($t + 1$): decreasing or maintaining. The differential energy at ($t + 1$) in decreasing or maintaining the energy is a crucial assumption for point t. The reason is that Mediterranean storms usually present a sharp gradient during wave height growth and a milder one during decay. The variables E and H^*_p provide more complete metrics for the storm hazard rather than a representative wave height, as they describe the behavior of the entire storm, rather than a snapshot.