Coastal Engineering 80 (2013) 107–121

Contents lists available at SciVerse ScienceDirect

Coastal Engineering

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

Climate-based Monte Carlo simulation of trivariate sea states

Y. Guanche [1], R. Mínguez *, F.J. Méndez [1]

Environmental Hydraulics Institute, "IH Cantabria", Universidad de Cantabria, C/Isabel Torres n 15, Parque Científico y Tecnológico de Cantabria, Santander 39011, Spain

ARTICLE INFO

Article history:
Received 22 February 2013
Received in revised form 7 May 2013
Accepted 21 May 2013
Available online 2 July 2013

Keywords:
Stochastic processes
Time series
Autoregressive logistic regression
Univariate ARMA
Simulation

ABSTRACT

Accurate wave climate characterization, which is vital to understand wave-driven coastal processes and to design coastal and offshore structures, requires the availability of long term data series. Where existing data are sparse, synthetically generated time series offer a practical alternative. The main purpose of this paper is to propose a methodology to simulate multivariate hourly sea state time series that preserve the statistical characteristics of the existing empirical data. This methodology combines different techniques such as univariate ARMAs, autoregressive logistic regression and K-means clusterization algorithms, and is able to take into account different time and space scales. The proposed methodology can be broken down into three interrelated steps: i) simulation of sea level pressure fields, ii) simulation of daily mean sea conditions time series and iii) simulation of hourly sea state time series. Its effectiveness is demonstrated by synthetically generating multivariate hourly sea states from a specific location near the Spanish Coast. The direct comparison between simulated and empirical time series confirms the ability of the developed methodology to generate multivariate hourly time series of sea states. Finally, the potential of the proposed methodology to simulate multivariate time series at multiple locations and incorporate climate change issues is discussed.

1. Introduction

Sea condition data is required for the assessment of long-term morphological changes along coastlines, designing coastal structures and flood risk assessments. Depending on the kind of structure to be designed, the extreme value distribution, the long-term distribution or even both are required. There are several different data sources available for designers such as: buoy, satellite, reanalysis (or hindcast) data and visual observation records. To obtain an accurate characterization of the extreme conditions at a specific location, a long term data set is required, which is rarely available. The reanalysis data usually provide longer records and avoid missing data and sparse spatial resolution. In the last decade, long-term databases from numerical models have been developed and improved the knowledge of deep water wave climate (Chawla et al., 2012; Rascle and Ardhuin, 2013; Reguero et al., 2012) and its propagation to obtain nearshore conditions (Camus et al., 2011). It is of note however, that the length of these hindcast databases is limited, typically up to 60 years. The case of instrumental records is even worse, since their length is limited up to 30–40 years for the longest records and they may contain missing data. Therefore, synthetically generated time series data can be used to extend historically limited data. These extended data sets can be used within structural design and flood and erosion risk analysis, for example.

Simulating synthetic time series in order to represent sea conditions is not new. In 1952, Longuett-Higgins gave an approach for the statistical distribution of wave heights based on a short term model (Longuett-Higgins, 1952). Since then, many authors have made contributions to achieve improved models representing time series of significant wave heights. A revision and an application to the Portuguese Coast can be found in Guedes Soares and Ferreira (1996). More recently, Solari and Losada (2012) proposed a unified distribution model that mixes different fits for central, minimum and maximum regimes, objectively determining the thresholds within the different regimes. Mínguez et al. (2013) proposes an alternative model to that from Solari and Losada (2012), also reproducing the central and maximum regimes simultaneously but including the temporal autocorrelation structure of the stochastic process.

Despite being one of the most important variables to define a sea state, the wave height (H_s) alone is not sufficient to fully characterize the prevailing wave conditions. As a minimum, the mean period (or peak period) associated with the significant wave height is required. It is well known that wave period can be an influencing variable in many situations, such as overtopping or transmission through permeable breakwaters, for example. There exist in the technical literature different analyses that explore the joint distribution of $H_s - T_m$. For example, in Guedes Soares and Cunha (2000), a bivariate autoregressive model is described that reproduces time series of significant wave height and mean period. In their work, an ARMA (Autoregressive Moving Average) model is used to represent both variables, significant wave height and mean period. Recently, in Dong et al. (2013) a maximum entropy distribution of significant

* Corresponding author. Tel.: +34 942 20 16 16; fax: +34 942 26 63 61.
E-mail address: roberto.minguez@unican.com (R. Mínguez).
[1] Tel.: +34 942 20 16 16; fax: +34 942 26 63 61.

0378-3839/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.coastaleng.2013.05.005

wave heights and peak periods was proposed. One of their approaches uses a univariate maximum entropy distribution, while an alternative approach consists of adjusting the maximum entropy marginal distribution of wave height, followed by conditioning the wave period distribution on the wave height. With an extension from the ARMA models to the VARMA models (Vector Autoregressive Moving Average), Cai (2011) proposed a multivariate simulation able to deal simultaneously with more than two variables but with the inherent complexity of the multivariate ARMA parameter estimation. In Castillo et al. (2006), long-term statistics of storms are characterized by a set of three variables that represent the maximum significant wave height $H_{s_{\max}}$ during each storm, its maximum wave height $H_{\max}$, and the associated wave period $T_{z_{\max}}$ (that occur with $H_{\max}$). The joint probability distribution and dependence structure is derived from real data so that once a storm has occurred, its intensity and characteristics can be derived from this joint distribution, i.e., a set of values $(H_{s_{\max}}, H_{\max}, T_{z_{\max}})$ can be drawn at random from a population with the corresponding distribution. Their model defines i) the marginal distribution of $H_{s_{\max}}$, ii) the conditional distribution of $H_{\max}$ given $H_{s_{\max}}$ and iii) the conditional distribution of $T_{z_{\max}}$ given $H_{\max}$, $H_{s_{\max}}$. Alternatively, if more than two variables are included in the analysis, copula functions may be used instead. This way, De Michele et al. (2007) presents a multivariate model to study sea storms. In their study, up to four variables (wave height, storm duration, storm direction and time in between storms) are taken into account to develop a model capable of simulating sea storm behavior.

Previous approaches focus their interest in the simulation of univariate or multivariate time series at one particular location. Alternatively, Morales et al. (2010) proposed a methodology to generate statistically dependent wind speed scenarios at different locations, decoupling the process into univariate ARMA models and their cross-correlations. This allows the reproduction of more than two time series associated with the same variable at multiple locations, and avoiding the complexity of multivariate ARMA parameter estimation or the use of copula functions. This methodology provides good results not only in the joint distribution but also in the marginals.

Alternatively, to deal efficiently with long time-series of multivariate data, several clustering methods have been developed in the field of data mining. These techniques enable the general features of the data to be well represented within a small subset of selected points. The K-means algorithm (KMA) and self-organizing maps (SOM) are two of the most popular clustering techniques in this field. These allow, for instance, the definition of a number of synoptic patterns over a specific region. These algorithms have been widely used in both: atmospheric and marine climate data. For example, Izaguirre et al. (2012) classify SLP (sea level pressure) fields over the Atlantic Ocean and use them to explain the wave climate and its variability; while Camus et al. (2011) use a clusterization of the met-ocean parameters to propose a wave propagation methodology.

Moreover, plausible time series of circulation patterns can be simulated using an autoregressive logistic model as described by Guanche et al. (2013). In that approach, a model that takes into account seasonality, interannual variability in terms of sea level pressure anomalies, long-term trends and Markov Chains is developed to accurately reproduce stochastically similar time series of circulation patterns.

However, although considerable advances on the proper characterization of multivariate stochastic processes and Monte Carlo simulation techniques have been achieved in recent years. There are still several methodological gaps which in our opinion should be fulfilled to get safer and more economical structures from a stochastic viewpoint:

1. Life-cycle of coastal and ocean engineering structures encompasses the consideration of different phases: pre-design, design, construction, service life and dismantling. Traditionally, each of these phases has been faced by engineers under the assumption of stationarity, behind the idea that historical natural patterns are reliable predictors for the future. However, it is recognized and accepted in the technical literature that this hypothesis is no longer valid (Milly et al., 2008), and we need to find ways to characterize and model the non-stationarity of relevant environmental variables, and use those models to optimize engineering designs including adaptation strategies during lifetime. However, there are significant knowledge gaps about how climate change issues will affect the practice in coastal and ocean engineering.
2. Studies, such as flooding risk or long-term beach morphodynamics, may require not only the multivariate characterization of marine climate variables at one location but at multiple locations as well.

The main objective of this paper is to provide a general non-stationary stochastic framework to incorporate the uncertainty associated with available climate information into the day-to-day coastal and ocean engineering decisions, filling the knowledge gaps about how climate change issues will affect the practice in coastal and ocean engineering. In particular, we combine different space and time scales to generate plausible long-term hourly time-series of trivariate (significant wave height, H_s, mean period, T_m and mean direction θ_m) sea state parameters by using some of the aforementioned techniques. To achieve this objective, a methodology has been developed that comprises three interrelated steps. In the first step, synthetic daily sea level pressure fields (DSLP) in the wave generation area, decomposed into principal components, are simulated by using the multivariate simulation technique proposed by Morales et al. (2010). During the second step, daily mean sea conditions (DMSC), clustered by K-means as proposed by Camus et al. (2011), are simulated by applying an autoregressive logistic model and taking into account the previously simulated DSLP as covariates. The third step consists of a modified version of the methodology proposed by Morales et al. (2010), to be used with hourly sea state (HSS) parameters and conditioned to the catalog of synoptic DMSC patterns simulated in the previous step.

Note that the aim of the paper is quite ambitious, and thus it requires the combination of several sophisticated statistical techniques. We provide concise descriptions of the complex steps used within the proposed simulation approach, including a summary of the physical meaning of the associated suite of statistical steps. However, and since the paper is still quite dense, we have decided to only apply the methodology to the study of hourly trivariate sea states at one location along the northwest Spanish coast. Nevertheless, and in order to show readers the potential of the presented approach, we also provide some hints about how this general framework could be used to incorporate more variables at one location, at multiple locations, or climate change issues. This fact and the accuracy shown by the method to reproduce both joint and marginal distributions prevent us from comparing with existing approaches, which may provide analogous results for this particular application, but do not have the potential to be extended for including additional variables, multiple locations and/or climate change issues straightforwardly.

The rest of the paper is organized as follows. Section 2 provides an in-depth exposition of the methodology, explaining the possible extension at multiple locations and including climate change effects. In contrast, in Section 3, the methodology is applied to a specific location on the north-western coast of the Iberian Peninsula. Here the results at each step of the process are verified, validating this way the performance of the method. Finally, in Section 4 relevant conclusions are duly drawn, including further applications and possible limitations.

2. Methodology

In this section, the procedure to generate plausible synthetic multivariate sea state time series is described step-by-step. As mentioned above, the entire methodology can be divided into three interrelated