Coastal Engineering 117 (2016) 1–18

Contents lists available at ScienceDirect

Coastal Engineering

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

A copula-based approach for the estimation of wave height records through spatial correlation

R. Jane[a,*], L. Dalla Valle[b], D. Simmonds[a], A. Raby[a]

[a]University of Plymouth, School of Marine Science and Engineering, Drakes Circus, Plymouth PL4 8AA, UK
[b]University of Plymouth, School of Computing, Electronics and Mathematics, Drakes Circus, Plymouth PL4 8AA, UK

ARTICLE INFO

Article history:
Received 17 December 2015
Received in revised form 17 June 2016
Accepted 25 June 2016
Available online xxxx

Keywords:
Significant wave height
Wave hindcasting
Multivariate copula
Spatial correlation
Wave model
SWAN

ABSTRACT

Information on the wave climate at a particular location is essential in many areas of coastal engineering from the design of coastal structures to flood risk analysis. It is most commonly obtained either by direct measurements or hindcast from meteorological data. The extended deployment of a wave buoy to directly measure wave conditions and the application of wave transformation models used in hindcasting, including public domain models such as Wavewatch and SWAN, are both expensive. The accuracy of the results given by the latter are also highly sensitive to the quality of the wind data used as input. In this paper a new copula-based approach for predicting the wave height at a given location by exploiting the spatial dependence of the wave height at nearby locations is proposed. By working directly with wave heights, it provides an alternative method to hindcasting from observed or predicted wind fields when limited information on the wave climate at a particular location is available. It is shown to provide predictions of a comparable accuracy to those given by existing numerical models.

1. Introduction

In a standards-based approach, structures are designed to withstand an event of a given severity. For coastal structures an event is traditionally characterised by a significant wave height and its severity defined in terms of a return period. The return period is defined as the average time elapsing between two consecutive occurrences of a prescribed event. In order to uncover the significant wave height corresponding to a given return period, information on the local wave climate is required. Depending on the available budget and richness of wave data in the locality a range of approaches exist for obtaining such data including long term deployment of a wave buoy, short term deployment and subsequent hindcasting to give a sufficiently long record of the local wave climate or the use of wave data from a nearby location. Taking advantage of the increase in readily available computational power and advances in the modelling of the structure of the dependance between non-independent random variables over the last couple of decades, this work presents an alternative method for deriving design conditions which has the potential to outperform the existing approaches.

Extensive networks of wave buoys are in place along many populated coastlines where flood risk is actively managed. To name but two, the Channel Coastal Observatory (CCO) maintains a network of around 40 wave buoys along the UK coast with the National Oceanic and Atmospheric Administration's (NOAA) National Data Buoy Center maintaining a similar network of over 100 wave buoys in the USA. Some wave buoy deployments such as these are, in principle, intended to be permanent with the requirement for regular maintenance. In others a wave buoy may be deployed for a short period of time to gather wave data in a new or critical location over a few seasons or years. The wave time series from these can be used to calibrate numerical hindcast models driven by meteorological forcing, or to validate wave transformation models driven by other wind and wave data. In both cases, the existing data sets may need to be extended, either because of missing data created by a damaged or vagrant buoy or because observations at a temporary site are not sufficient to build an accurate model for localised extremes.

It is reasonable to assume that observations of wave fields generated by the same physical processes in the same region, should be correlated. These observations made along the same natural coastline will be modified by variation of the direction and degree of exposure and variability in the wave transformation processes created by the variable bathymetry. Local generative processes may also prove a

* Corresponding author.
E-mail addresses: robert.jane@plymouth.ac.uk (R. Jane), luciana.dallavalle@plymouth.ac.uk (L. Valle), D.Simmonds@plymouth.ac.uk (D. Simmonds), alison.raby@plymouth.ac.uk (A. Raby).

http://dx.doi.org/10.1016/j.coastaleng.2016.06.008
0378-3839/

reason for variability. By exploiting and modelling such correlations between the wave height observed at a given location and at two or more neighbouring locations, the proposed approach offers a simple and computationally efficient alternative method for generating missing or extended data at a site of interest. In the case of extremes analysis, buoys frequently malfunction during energetic events so such an approach would be useful to infer the absent extreme values.

In the case of buoy networks, once the correlations are suitably determined a virtual wave buoy model will have been created, an idea postulated in Londhe (2008). This might enable a reduction in the number of buoys, or allowing the swapping in and out of buoys during a programme of maintenance with minimal loss of measurement system accuracy. Removed wave buoys might also be redeployed at other locations. The method thus promises a means for increasing the detailed knowledge of wave climate along a stretch of coast by removing redundant measurements.

In each of the applications discussed so far our interest lies in the extreme values of these variables. Extreme values of the variables concerned are modelled through fitting extreme value distributions to the observed extremes (Section 2). With longer length of data sets more extreme values are likely to be captured leading to more accurate definition of the return periods. However, due to time and financial constraints the long term deployment of a wave buoy is often not practical and therefore other means of deriving the local wave conditions have been contrived. For coastal areas where no data is available, historic offshore wave conditions are commonly derived from records of the local wind field, a process referred to as hindcasting. Originating in the 1940s and 50s with simple empirical models such as SMB and SPM, today computationally intensive numerical wave transformation models such as WAM (WAve Modelling) (WAMDI group, 1988), Wavewatch III (Tolman, 1991) and SWAN (Simulating WAves Nearshore) (Booij et al., 1999) dominate the field. These models work on the principal that most statistical properties of wind waves are captured in their wave energy or action spectrum. Assuming the principle of linear superposition, observations of the water surface at a given location can be decomposed into a spectrum of wave energy as a function of frequency. With additional local observations, a two dimensional spectrum can be constructed to describe the directional variation of the wave energy spectrum. These are used in spectral wave models. The action spectrum is the energy spectrum divided by the intrinsic frequency of the spectral components. The models numerically integrate the wave energy or action transport equation which governs the evolution in space and time of their respective spectra to predict future wave conditions at different locations.

Numerical wave models such as WAM and Wavewatch were developed primarily for predicting deep ocean wave conditions where the waves are mostly wind driven. The source terms present in the transport equations represent the physical processes that can result in generation, dissipation, or redistribution of wave energy. For the WAM model, inputs required for the source terms include wind data, information on non-linear (Quadruplet) wave-wave interactions and white capping dissipation. On the other hand, numerical models such as SWAN were developed for predicting nearshore wave conditions. This includes not only each of the source terms present in deep water models but also terms for triad wave–wave interactions which become much more significant than the higher order wave-wave interactions in the nearshore region. It also includes a term to account for depth induced breaking which can lead to a significant reduction in the maximum significant wave height in coastal regions. The SWAN model uses the wave action spectrum rather than the wave energy spectrum since, in contrast to the energy spectrum, it is conserved in the presence of currents thus allowing the effect of a mean current on the evolution of the wave field to be included in the modelling. In practice, to take advantage of both models, it is common for the generated wave field from a WAM model to subsequently form the boundary conditions for a higher resolution wave transformation model, such as SWAN for propagating the waves inshore (Wornom et al., 2001).

The accuracy of the hindcast offshore wave conditions are highly dependent on the accuracy of the data on the forcing wind field (Holthuijsen et al., 1996; Moeini et al., 2010; Teixeira et al., 1995). These wave fields are derived from the measured wind fields and act as the primary input for the prediction of the coastal wave conditions. Thus, the quality of the predicted coastal conditions relies on accurate descriptions of the forcing wind field. Also, the exercise of assimilating the data and running numerical wave models, which may involve, long run times and multiple sensitivity analyses to validate the output, makes such an exercise non-trivial and often significantly costly. In order to reduce the computational cost meta-models, that are simplified approximations of computationally intensive models, have found favor. For example, Camus et al. (2011) developed a model that uses radial basis functions as a meta model for SWAN. Other methods for reducing the computation burden include coupling numerical and soft computing techniques as is done in Malekmohamadi et al. (2008) by way of WAM and a neural network. Again, this can require a considerable time and cost, but once set up, these models can run very efficiently.

In addition to the finite available computational power, these deterministic modelling approaches are also hampered by a lack of knowledge of the physical processes that generate waves, the interrelated parameters that drive them as well as insufficient detail concerning the modelled region. Moreover, wave energy spectrum-based models are believed to be reaching the limits of the accuracy with which they can simulate wave hydrodynamics (Liu et al., 2002). Any further improvements are likely to occur at a much slower pace than in the past and these may not lead to any practical improvement to the accuracy of their predictions (Cavaleri et al., 2007). In reality, even if substantial advancements were made in capturing the detail of physical processes and their interactions, the complexity of the situation will limit deterministic prediction.

In acknowledgment of the uncertainties associated with physics based models, over recent decades scientists have started to considere the problem probabilistically. Although not normally considered to be a Gaussian process, Gaussian models of varying complexity were the first to be investigated (Cunha and Soares, 1999; Scotto and Soares, 2000; Soares et al., 1996). More recently soft computational techniques that can be applied directly to observations have become increasingly popular. Artificial Neural Networks (ANN's) are perhaps the most widely adopted of these techniques. ANN's aim to mimic how biological neural systems such as the human brain process information. They consist of a set of interconnected processing units or nodes, and a set of weightings along the connections or synapses, that are analogous to synaptic strength. The arrangement of the nodes are prescribed by the user with the weights adjusted through a learning or training procedure. This references a cost function so that the relationship between the input stimuli and the output responses becomes optimised. They have been shown to outperform the auto regressive models when forecasting wave heights (Deo and Naidu, 1998). As such they have prevailed as the dominant approach for prediction using either previous wave height observations at a location (Deo and Naidu, 1998; Gopinath and Dwarakish, 2015; Hadadpour et al., 2014; Londhe and Panchang, 2006; Mandal and Prabaharan, 2006) or wind wave data (Deo et al., 2001; Kamranzad et al., 2011; Malekmohamadi et al., 2008). They have also been applied to improve the accuracy of physics based process models (Zhang et al., 2006). Other soft computational approaches such as genetic programming (Gaur and Deo, 2008; Nitsure et al., 2012), fuzzy inference systems (Kazeminezhad et al., 2005; Özger and Şen, 2007) and support vector machines (Mahjoobi and Mosabbeb, 2009) have also been shown to provide useful predictions (Malekmohamadi et al., 2011). For a more detailed