Ocean Engineering 142 (2017) 268–279

Contents lists available at ScienceDirect

Ocean Engineering

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

Detection and prediction of segments containing extreme significant wave heights

A.M. Durán-Rosal*, J.C. Fernández, P.A. Gutiérrez, C. Hervás-Martínez

Department of Computer Science and Numerical Analysis, University of Córdoba, Rabanales Campus, Albert Einstein Building, 3rd Floor, 14071, Córdoba, Spain

A R T I C L E I N F O

Keywords:
Time series segmentation
Multiobjective evolutionary algorithm
Local search
Prediction
Detection
Extreme significant wave height
Minimum sensitivity

A B S T R A C T

This paper presents a methodology for the detection and prediction of Segments containing very high Significant Wave Height (SSWH) values in oceans. This kind of prediction is needed in order to account for potential changes in a long-term future operational environment of marine and coastal structures. The methodology firstly characterizes the wave height time series by approximating it using a sequence of labeled segments, and then a binary classifier is trained to predict the occurrence of SSWH periods based on past height values. A genetic algorithm (GA) combined with a likelihood-based local search is proposed for the first stage (detection), and the second stage (prediction) is tackled by an Artificial Neural Network (ANN) trained with a Multiobjective Evolutionary Algorithm (MOEA). Given the unbalanced nature of the dataset (SSWH are rarer than non SSWH), the MOEA is specifically designed to obtain a balance between global accuracy and individual sensitivities for both classes. The results obtained show that the GA is able to group SSWH in a specific cluster of segments and that the MOEA obtains ANN models able to perform an acceptable prediction of these SSWH.

1. Introduction

Large ocean waves pose significant risks to ships and offshore structures. The development of offshore installations for oil and gas extraction requires knowledge of the wave fields and any potential changes in them. Moreover, in order to accurately predict the long-term energy resource and performance of ocean wave energy converters, long-term prediction of extreme wave heights is particularly important. Additionally, high ocean waves represent significant risks in ship movements and port activity, and a reliable measurement of these extreme and critical events is crucial from the point of view of navigation and civil protection.

In recent years, different statistical and mathematical methods have been proposed for calculating and predicting Significant Wave Height (SWH) Mahjoobi et al. (2008), Mahjoobi and Mosabbeb (2009). SWH can be defined either in the temporal domain or in the frequency domain. In the former case, it is noted $H_{1/3}$ and is defined as the average height of the highest one-third of wave heights, measured from the time series of free surface by up or down-crossing. In the latter case, it is noted H_{m0} and is defined from the frequency spectrum. In deep water, $H_{1/3}$ and H_{m0} are quite close (less than 5% of difference) and they are generally confused in the generic term H_s. For this reason, even if the definitions of wave height are formally expressed, it is advisable to use the generic term H_s or simply SWH. According to the National Data Buoy Center (NDBC) and the National Oceanic and Atmospheric Administration (NOAA), SWH is the average trough to crest in meters of the highest one-third of all the wave heights during a 20-min sampling period (2016). NOAA uses hydrographic stations and ocean buoys with special sensors to collect data, and this paper uses this source of information. There are other statistical measures of the wave height, such as the Root Mean Square (RMS) wave height, which is defined as the squared root of the average of the squares of all wave heights and is approximately equal to SWH divided by 1.4 Holthuijsen (2007).

Recently, a more specific field, the determination and prediction of Extreme SWH (ESWH), has gained significant attention. In general, the previously proposed methods are based on considering the probability distributions of the Extreme Values (EV) of SWH. For example, the work of Muraleedharan et al. Muraleedharan et al. (2016) proposes the use of quantile regression to model the ESWH distribution, as an alternative to fitting EV distributions based on the tails of data samples. Another popular methodology is the Peaks Over Threshold (POT) Davison and Smith (1990) (i.e. considering only those values of the time series higher than a predefined threshold, that is, those values which are a sample of exceedances), which has been used as a standard approach for these predictions Caires and Sterl (2005), Viselli et al.

* Corresponding author.
E-mail address: i92duroa@uco.es (A.M. Durán-Rosal).

http://dx.doi.org/10.1016/j.oceaneng.2017.07.009
Received 29 August 2016; Received in revised form 20 June 2017; Accepted 2 July 2017
0029-8018/

(2015). Moreover, the IAHR Working Group on Extreme Wave Analysis suggested the use of the POT method along with a Weibull distribution estimated by maximum likelihood, for the determination of ESWH Mathiesen et al. (1994). A similar study is described by Mazas and Hamm (2011), where, in a first step of peak selection, a POT declustering above a physical threshold is considered, and the GPD-Poisson model is used for the second step (see also Bernardara et al., 2014, for the justification of the double threshold approach). This model is then extended to other statistical distributions (Weibull and Gamma).

A similar idea was used in Wimmer et al. (2006) for determining ESWH when considering a GPD for all values above a threshold. In addition, the GPD was used in combination with a Generalized EV (GEV) distribution in Petrov et al. (2013). To this purpose, a maximum entropy (MaxEnt) was used to estimate the parameters of GPDs. The MaxEnt method can be seen as a very robust and efficient tool in statistics, which serves for characterization of the probability distribution functions in the case that the information about the process is limited and it solely relies on data from the sample. MaxEnt was applied to the samples generated with the POT approach, where the lower boundary was naturally given by the selected threshold while the upper boundary was selected arbitrarily. The reason why MaxEnt was used with the GPD and the GEV model is the general belief that these models provide the best estimates of high quantiles among all available parametric models Coles et al. (2001). Recently, Galiatsatou et al. (2016) considered GEV to assess non stationarity in annual maximum wave heights, both in present and future climates. Finally, Wada et al. (2016) applied these methodologies in rarest and severest of ocean wave events (e.g. the storm peaks over threshold of significant wave heights in a tropical cyclone at a location).

The above models are based on the hypothesis that there are no climatic variation patterns and the yearly samples of weather data are independent and identically distributed. This is not the case for the work of Galiatsatou et al. (2016), who specifically examine these variations. Recent researches show several evidences suggesting that the intensity of storms over the ocean is changing, due to the climate change Wang et al. (2004), Fan et al. (2013) (although some other studies did not find these evidences Feng et al., 2014). Kitano et al. (2015) found that the difference between past and future values is significant. All these studies agree in the fact that the modelling EV involves unavoidable limitations Bitner-Gregersen and Guedes Soares (2007). All previous works show the great interest in determining a maximum threshold that allow us to fit the Gamma, Beta or Weibull distributions for ESWH (and, in general, GPD and GEV distributions). For models with POT declustering, this process usually involves the assumptions of different hypothesis which can change over the years, leading to under or over-estimations of the ESWH.

Given that climate change clearly affects the determination of ESWH, long-term prediction of SWHs have been previously tackled by other researchers Vanem (2016), Woo and Park (2016), Sierra et al. (2017), Feng et al. (2014). The study in Vanem (2016) is based on various joint models for the simultaneous distribution of significant wave height and zero-crossing wave period, where each of the models is fitted to data generated from a numerical wave model for the current climate and for two future climates consistent with alternative climate scenarios. The results obtained suggest that ESWH and zero-crossing wave period tend to be more correlated in a future climate compared to the current climate. In Woo and Park (2016), the authors agree that there are substantial changes associated to climate change for the determination of ESWH in the East/Japan Sea. They conclude that the annual mean of ESWHs was dramatically increased by 3.45 m in this sea, which is significantly higher than the normal mean of about 1.44 m Woo and Park (2016). In Sierra et al. (2017), the authors analyse how changes in wave patterns, due to the effect of climate change, can affect wave energy power and yield around Menorca (Spain), where more consistent long-term predictions of ESWH are found to be possible. In Feng et al. (2014), the authors analyse 10 years of in-situ measurements of SWH and maximum wave height from the ocean weather ship Polar front in the Norwegian Sea. All these studies confirm that the long-term analysis of ESWH in the context of climate change is interesting and also very challenging.

On the other hand, ESWH predictions are helpful for operational decisions, such as vessel operations, subsea operations (diving and remotely operated vehicles), laying of submarine pipelines, deciding whether operational works in the sea should continue, crane lift operations, warning mariners about wave heights, tanker loading and drilling, deciding about leaving ports according to wave heights, etc.

In this paper, a new approach for ESWH prediction is proposed. Significant Wave Height (SSWH) values can be defined as sea states with a very high absolute height or those which can be considered to be high in relation to other waves close in time. The methodology autonomously finds a set of segments which are grouped in a clustering step to discover whether one of the clusters is representing SSWH. Moreover, the clustering step transforms the time series into a sequence of labels, and the final phase of the proposed system is the construction of a predictive model able to determine if an SSWH will be produced (or not) after a given subsequence of a time series. Therefore, the final phase is a binary classification problem, where the positive class corresponds to the prediction of a SSWH, while the negative class represent non SSWH. The problem is highly unbalanced, given that SSWHs are less frequent. It should be noted that unbalanced binary classification problems often produce models biased towards the majority class Weiss and Provost (2003), He and Garcia (2009). Traditional measures as the Correctly Classified Rate (CCR) or accuracy capture only the global precision of a classifier, not considering whether the minority class is well classified. Designing a classifier with an unbalanced dataset using only CCR as the objective function could perform poorly on the minority class. As can be seen, the proposed methodology is related to the detection and prediction of extreme values, but from a completely different and novel point of view. The method proposed in this paper is better adapted for long-term variations in the SWH regime than other methods such as standard regression (given the limitations of linear auto-regressive models), and, in contrast to previous methods performing long-term prediction of SWH Vanem (2016), Woo and Park (2016), Sierra et al. (2017), Feng et al. (2014), it does not make any assumption about the probability distribution.

The methodology developed in this work incorporates algorithms based on Evolutionary Computation (EC) Back (1996), Bäck et al. (1997), a subfield of Artificial Intelligence (AI). In EC, this algorithms are named in a general way as Evolutionary Algorithms (EA), which are based on adopting Darwinian principles (population, selection, recombination, mutation, survival). Technically these algorithms belong to the family of trial and error problem solvers and they can be considered global optimization methods with a metaheuristic or stochastic optimization character, distinguished by the use of a population of candidate solutions (rather than just iterating over one point in the search space). The application of recombination and evolutionary strategies makes them less prone to get stuck in local optima than alternative classic methods. Training a classifier with EC techniques has also the problem of use the *CCR* measure in unbalanced datasets as the objective function. However, if the classifier is trained with an additional objective function that maximizes the worst classified class, more balanced models could be obtained, as is demonstrated in Fernández et al. (2010). That second objective function is the Minimum Sensitivity (*MS*), defined as the sensitivity (or accuracy ratio) of the worst classified class. For this reason, a Multiobjective Evolutionary Algorithm (MOEA) Coello and Lamont (2004), Deb (2004), Coello et al. (2007) that simultaneously optimizes the *CCR* and *MS* objective functions is used in a second prediction phase, trying to maximize the global precision maintaining the level of accuracy for the both classes as balanced as possible.