Ocean Engineering 118 (2016) 204–215

Contents lists available at ScienceDirect

Ocean Engineering

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

Regression quantile models for estimating trends in extreme significant wave heights

G. Muraleedharan, C. Lucas, C. Guedes Soares *

Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Portugal

ARTICLE INFO

Article history:
Received 27 July 2015
Received in revised form
8 March 2016
Accepted 10 April 2016
Available online 20 April 2016

Keywords:
Long term data
Extreme significant wave height
Climate trends
Weibull quantiles
Pareto quantiles
GEV quantiles

ABSTRACT

Regression quantile models are proposed to model extreme significant wave height distributions in two Portuguese locations of Figueira da Foz and Azores in the North Atlantic Ocean. A data set from a 44 years hindcast produced in the HIPOCAS project was used in this study. In order to identify trends with time, the data set was divided in 11 samples of 4 year data. Three-parameter *Weibull,* generalised extreme value and generalised *Pareto* quantile functions are fitted to data and extrapolated to 50 and 100 year significant wave height return periods. The algorithm of distributional least absolutes is used to estimate the model parameters. The regression quantile models showed the ability to model the historical trends that may subsist in long term data sets, a feature that the traditional fitting of extreme distributions does not account for.

1. Introduction

The maximum significant wave height expected in a long period is a reference value used for the design of ocean and coastal structures and hence, several probabilistic models have been developed for its assessment and applied in various problems.

Typical approaches are to construct long-term distributions based on all observations accumulated during a period, which are afterwards extrapolated to appropriate return periods (e.g. Isaacson and Mackenzie, 1981; Haver, 1985; Muir and El Shaarawi, 1986; Ferreira and Guedes Soares, 2000; Guedes Soares and Scotto, 2001). While this approach is appropriate when one wants to model the whole history of the loading as required for fatigue assessment (Guedes Soares and Moan 1991), it is better to concentrate on the higher wave heights when one is only interested in extreme values, in which case extreme value theory (Coles, 2001) is then relevant. The Peak Over Threshold (POT) method (Davison and Smith, 1990) has become a standard approach for these predictions (e.g. Mathiesen et al., 1994; Ferreira and Guedes Soares, 1998; Elsinghorst et al., 1998; Caires and Sterl, 2005). Various improvements over the basic approach have been proposed by various authors since then (e.g. Sobey and Orloff, 1995; Guedes Soares and Scotto, 2004; Stefanakos and Athanassoulis, 2006; Scotto and Guedes Soares, 2007; Huerta and Sansó, 2007; Petrov et al., 2013; Jonathan and Ewans, 2013).

An alternative to the use of the tail of data samples to fit extreme value distributions is to operate with the quantile function $Q(p)$, which is the inverse of the cumulative distribution function $F(x)$ that shares equivalent properties with $F(x)$.

In many model estimation procedures, sample moments and their functions are used as estimates. The non-robustness of these statistics, their susceptibility to extreme observations and instability to match the corresponding population values are some of the problems in the conventional methods, which can be reduced to a certain extent by adopting quantile based methods. For example various characteristics of the distribution like location, dispersion, skewness and kurtosis can be directly derived from $Q(p)$, whereas the use of $f(x)$ requires integration of functions to derive such quantities. The use of quantile functions owe much to Gilchrist (1997, 2000) where the main properties are described. Quantile functions have been used to derive return values of sea state parameters by Muraleedharan et al. (2009, 2012), demonstrating the usefulness of such approach.

All of the referred models predict return values for long time periods on the postulate that there are no climatic variation patterns and thus, that the yearly samples of weather data are statistically independent and identically distributed, despite the large inter year variability of parameters that is often observed (Guedes Soares and Henriques, 1996; Ferreira and Guedes Soares, 1999). Due to climate change, several evidences (Bouws et al., 1996; Kushnir et al., 1997; Sergy and Lutz, 1999; Grevemeyer et al., 2000; Xiaolan et al., 2004; Fan et al., 2013) suggest that the storm

* Corresponding author.
E-mail address: c.guedes.soares@centec.tecnico.ulisboa.pt (C. Guedes Soares).

http://dx.doi.org/10.1016/j.oceaneng.2016.04.009
0029-8018/

intensity over the ocean is changing, which would imply the existence of a trend on the maximum significant wave height experienced for long periods of the order of 50–100 years. However, because until recently there were no data sets available with that type of duration, it was not possible to study those trends based on data.

The relatively recent availability of consistent data sets of hindcast data, of the order of 45 years (Uppala et al., 2005; Guedes Soares, 2008; Dee et al., 2011), permits the possibility of having attempts of identifying trends in the data and to incorporate them in the extrapolations to the long-term time frame.

Regression quantile models are adopted in the present paper to model the long-term trends of significant wave height, and will be fitted to data from the hindcast study produced in the HIPOCAS project for the North Atlantic Ocean (Pilar et al., 2008).

As the *Weibull* model has been the one most commonly adopted in the investigations of the long-term distribution of sea states (Battjes, 1972; Isaacson and Mackenzie, 1981; Haver, 1985; Guedes Soares and Scotto, 2001; Muraleedharan et al., 2007) this study will adopt the *Weibull* regression quantile model along with generalised extreme value model. The generalised *Pareto* (*GP3*) distribution is also efficiently used along with extreme value distributions (Coles, 2001) and thus, the *GP*3 regression quantile model is also considered in this work.

The change with time of the mean and spread of the significant wave height due to wave climate change will be estimated empirically from the data set.

The 44 years North Atlantic Ocean daily maximum hindcast significant wave height distributions off Azores and Figueira da Foz computed within the European project HIPOCAS (Guedes Soares, 2008) are analysed in this paper. The first location is well in the North Atlantic, in an area known for large storms and the second one is close to the Portuguese coast.

The hindcast data was validated with buoy data (Pilar et al., 2008), showing good agreement, but the ability in modelling extremes has limitations (Bitner-Gregersen and Guedes Soares, 2007). The hindcast has been compared with other hindcasts such as the NOAA data (Campos and Guedes Soares, 2012) and with ERA40 and ERA Interim (Campos and Guedes Soares, 2016). It was demonstrated that there is a very good agreement among all data sets for low and moderate sea states but for extreme sea states the data sets have some bias, with ERA40 underestimating the Hs and HIPOCAS overestimating it when compared with satellite data.

Therefore, the estimation of specific extreme values in this work does not have the purpose to indicate the best estimate for the locations considered but instead it aims to demonstrate the application of the regression quantile models. This study will allow the use of parameters of the distributions extrapolated with respect to the duration of the given return periods, rather than the constant values used in Muraleedharan et al. (2012). Section 2 will introduce the details of the methods adopted and Section 3 presents the case study analysed.

2. Methodology

2.1. Method of estimation

The model parameters of the quantile functions are estimated by the method of distributional least absolutes (*DLA*) (Gilchrist, 2000) from the sample data sets. *DLA* provides a unique, robust and universal approach for model parameter estimations based on quantile functions. The algorithm is described as:

If $x_{(1)}, x_{(2),\ldots} x_{(n)}$ are n observations of the random variable X arranged in ascending order, then the probability p_r $[P(X \le x_{(r)}) = p_r]$ associated with the rth order statistic $x_{(r)}$ is given by the rankit rule as:

$$p_r = BETAINV\,(0.5,\, r,\, n - r + 1) \tag{1}$$

$$\text{where } BETAINV\,(x,\, \alpha,\, \beta) \text{ is the inverse of } \frac{B_x(\alpha,\beta)}{B_1(\alpha,\beta)} \text{ and} \tag{2}$$

$$B_x(\alpha,\beta) = \int_0^x y^{\alpha-1}(1-y)^{\beta-1}dy,\; 0 \le x \le 1 \tag{3}$$

is the incomplete Beta function. If $\hat{\theta}$ represent the possible values of the parameters of the quantile function $Q\,(p_r, \hat{\theta})$, then the sum of absolute distributional residuals is given by

$$D = \sum_{r=1}^{n} \left| x_{(r)} - Q\left(p_r, \hat{\theta}\right) \right| \tag{4}$$

The method of distributional least absolutes criterion estimates $\hat{\theta}$, by minimising D. A standard numerical minimisation program is used here for that purpose.

2.2. Regression quantile models

To model the variability of Hs the following three parameter quantile functions $Q\,(p)$ are considered in this study:

i) The *Weibull* quantile function

$$Q\,(p) = x = \lambda + \eta\,[-\ln(1-p)]^{\alpha},\; x \ge \lambda \tag{5}$$

ii) The generalised extreme value (*GEV*) quantile function

$$Q\,(p) = x = \lambda + \frac{\eta}{\alpha}\left\{1 - [-\ln(p)]^{\alpha}\right\} \tag{6}$$

where $-\infty < x \le \lambda + \frac{\eta}{\alpha}$ if $\alpha > 0$; $\lambda + \frac{\eta}{\alpha} \le x < \infty$ if $\alpha < 0$

iii) The generalised *Pareto* (*GP*3) quantile function

$$Q\,(p) = x = \lambda + \frac{\eta}{\alpha}\left[(1-p)^{-\alpha} - 1\right] \tag{7}$$

where $\lambda \le x < \infty$ if $\alpha < 0$; $\lambda \le x \le \lambda + \frac{\eta}{\alpha}$ if $\alpha > 0$

where λ, η and α are respectively the location, scale and shape parameters of the quantile functions.

Regression quantile functional forms pave way for incorporating random elements (here *Hs*) that are functionally related to deterministic elements (here time *t*). The parameters that are estimated from the distribution of *Hs*, are regressed on time *t*. For the parameters that do not show a significant functional relationship with respect to time, the median of the parameters are used in the regression quantile models. The average (mean or median) of the parameters estimated from different data sets for the same ocean region will still be a better choice than a single valued parameter from one data set.

The appropriate functional behaviour (wherever it holds well) for the trend lines estimates $\hat{y}$ of the parameters of the models that are regressed on *t* is provided by an appropriate software. In order to minimise the sum of the errors *e* between the parameter estimate $\hat{y}$ of the trend line of the model and the computed parameter *y* by *DLA* from data, the least absolute criteria (Gilchrist, 2000) is adopted: