Ocean Engineering 54 (2012) 119–131

Contents lists available at SciVerse ScienceDirect

Ocean Engineering

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

Modelling significant wave height distributions with quantile functions for estimation of extreme wave heights

G. Muraleedharan [a], Cláudia Lucas [a], C. Guedes Soares [a,*], N. Unnikrishnan Nair [b], P.G. Kurup [c]

[a] Centre for Marine Technology and Engineering (CENTEC), Instituto Superior Tecnico, Technical University of Lisbon, Av. Rovisco Pais 1049-001, Lisboa, Portugal
[b] Department of Statistics, Cochin University of Science and Technology, Cochin-682022, India
[c] School of Arts and Sciences, Amrita Institutions, Cochin-682024, India

ARTICLE INFO

Article history:
Received 15 October 2011
Accepted 1 July 2012
Available online 3 August 2012

Keywords:
North Atlantic Ocean
Indian Ocean
Wave data
HIPOCAS database
Quantile functions
Wave height distribution
Significant wave height
Extreme wave height
Return periods

ABSTRACT

This paper starts by introducing extreme wave height analysis using quantile functions, which are an alternative to the classical approaches to model long term maxima or extreme values. The long-term distribution of significant wave heights from four locations are modelled with Davies, 3-parameter Weibull, generalized extreme value (*GEV*) and generalized Pareto (*GP*3) quantile functions. Even though the 3-parameter Weibull and *GP*3 quantile functions are adequate wave height models in this study, the performance of the Davies quantile function for extreme wave analysis seems to be consistently good both temporally and spatially.

1. Introduction

Ocean sea surface wave climatology is of vital importance in all enhancement and exploratory activities that include coastal and offshore engineering, defence, navigation, fishing and ocean mining. The intensity of the sea states is characterized by their significant wave height and this is the reference parameter used in long term distributions and in the prediction of extreme values. Typically this is represented by a probability distribution function $F(x)$ or equivalently the probability density function $f(x)$ of the random variable X that is assumed to represent the significant wave heights.

The initial approaches used visual observations (Guedes Soares, 1986), which was the data generally available in earlier stages and at later phases measured data by wave rider buoys became available. The traditional approach to determine extreme values of significant wave height has been to model the initial distribution, often by a Weibull distribution (Isaacson and Mackenzie, 1981; Muir and El Shaarawi, 1986), although other models have been used (Ferreira and Guedes Soares, 1999; Stefanakos, 1999; Lopatoukhin et al., 2001; Stefanakos and Athanassoulis, 2006), leading to a large modelling uncertainty in the predictions (Guedes Soares and Scotto, 2001) as compared with the statistical one (Guedes Soares and Henriques, 1996).

As a better alternative to using the initial distribution method, extreme value distributions have been used sometimes associated with the peak-over-threshold method (Van Vledder et al., 1993; Ferreira and Guedes Soares, 1998), which can be generalized to include more observations so as to increase its robustness (Sobey and Orloff, 1995; Guedes Soares and Scotto, 2004). When using only data larger than a high threshold, one normally deals with independent observations as the correlation between successive observations is reduced, although there may still occur some clustering as a result of a storm, in which more than one exceedance occurs.

One way to deal with this problem is to use storm models, which were initially proposed by Petruaskas and Aagaard (1971). A more recent approach is the Equivalent Triangular Storm model, proposed by Arena and Pavone (2009) and Fedele and Arena (2010).

An alternative to increase robustness is to adopt non-parametric descriptions such as the kernel of the distributions, which can be done whenever there is a sizable amount of data available. This has been proposed for univariate distributions by Ferreira and Guedes Soares (2000) and by Athanassoulis and Belibassakis (2002), and for bivariate distributions of significant wave height and mean period by Ferreira and Guedes Soares (2002).

An alternative towards having robust models is to use the quantile function $Q(p)$ which is the inverse of the cumulative distribution function that shares equivalent properties with $F(x)$,

* Corresponding author
E-mail address: guedess@mar.ist.utl.pt (C. Guedes Soares).

0029-8018/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.oceaneng.2012.07.007

where x is a particular value of the wave height random variable X. The present paper examines the usefulness of the quantile functions in modelling long term data of significant wave heights.

The advantages of the quantile function approach are several. The properties of $Q(p)$ like (i) the sum of two quantile functions (the product of two positive quantile functions) is again a quantile function (ii) the reciprocal of X is distributed as $(Q(1-p))^{-1}$, (iii) for a non-decreasing function $H(p)$ with $H(0)=0$, $H(1)=1$, $Q(H(p))$ is again a quantile function over the same range of $Q(p)$ etc. are not shared by $F(x)$ (Gilchrist, 2000). 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. In many model estimation procedures, sample moments and their functions are used as the estimates. The non-robustness of these statistics, their susceptibility to extreme observations, instability to match the corresponding population values are some of the problems in the conventional methods. This can be reduced to a marked extent by adopting quantile based methods (Gilchrist, 1997, 2000).

Of particular interest in the quantile function approach is the family of models known by the general name; lambda distributions involving several parameters that express x as a some simple function of p. Such models have been successfully employed in several problems, e.g., hydrology (Houghton, 1978), process control (Gilchrist, 1997), reliability (Gilchrist, 2000), air pollution (Okur, 1988), finance (Mc Nichols, 1987), climate studies (Abouammoh and Osturk, 1987) and inventory modeling (Nahmriss, 1994).

Lambda distributions form a highly flexible family of distributions capable of representing many of the known distributions in statistical literature either exactly or approximately (Gilchrist, 2000). Lambda distributions are particularly useful in situations where the choice of a relevant model is difficult on the basis of the physical conditions that govern the phenomenon and different forms of $F(x)$ are prescribed under different conditions. In such cases, a single form of lambda distribution can be found as a satisfactory fit by varying the parameter values. This enables subsequent analyses under a unified framework.

Muraleedharan et al. (2009) conducted an initial and exploratory study at the application of quantile functions to wave data. They simulated the long-term distribution of daily maximum shallow water wave heights, including rough sea and swell dominated southwest monsoon (active weather condition) season (May to October), spread over a period of 5 years (1980–1984) off Valiathura, southwest coast of India. They adopted a 2-parameter tuning coefficient incorporated modified Weibull (Muraleedharan et al., 2007) a general Weibull, a truncated Gumbel (up to $x=0$ for non-negative wave height data), a 3-parameter generalized Pareto, Rayleigh and Davies distribution and compared their performance. The Davies quantile function was competitive to other models and the various estimated wave height statistics using quantile function were in good agreement with computed values from the data.

This study is a continuation of the previous work and applies that type of approach to a larger number of ocean sites and is not restricted to data from the Indian Ocean.

The 44 years North Atlantic Ocean daily maximum hindcast significant wave height distributions off Azores, Figueira da Foz and Sines calculated in the European Project HIPOCAS (Guedes Soares, 2008) are modelled by three parameter quantile functions of Weibull, generalized extreme value, generalized Pareto and Davies quantile function analysis. The hindcast data produced has been validated with buoy data (Pilar et al., 2008), although no specific study has been made of their performance in modelling extremes. Therefore the use of the data in this paper aims at demonstrating the method proposed and not at determining specific extreme values at a given location.

In order to determine the applicability of the functions on oceanic regions of different wave characteristics, the 3-hour recorded significant wave height distributions off Machilipatnam in the Bay of Bengal (1997–2005) are also modelled and subjected to extreme wave analysis.

The paper is organized into four sections. In Section 2, the basic properties of the model required in the sequel are discussed. Analysis and discussion of results on the most widely used 3-parameter Weibull, generalized extreme value (*GEV*), and generalized Pareto (*GP3*) for wave height distribution and Davies *QF*, in respect of the significant wave heights; extreme wave heights and their return periods are presented in Section 3. The study is concluded in Section 4 with discussions on the models and the derived wave parameters.

2. Quantile function models

2.1. Quantile function approach

The quantile function at a probability level p is defined as

$$Q(p)=F^{-1}(p)=\inf\{x:F(x)\geq p\},\quad 0\leq p\leq 1 \tag{1}$$

One of the quantile functions or lambda distributions in common use is the symmetric form Tukey (1960):

$$Q(p)=\lambda+\frac{\eta}{\alpha}\left[p^{\alpha}-(1-p)^{\alpha}\right],\quad 0\leq p\leq 1 \tag{2}$$

where λ, η and α are the location, scale and shape parameters.

Another form of the quantile functions or lambda distributions is the Ramberg and Schmeiser (1974) model:

$$Q(p)=\lambda+\eta[p^{\alpha}-(1-p)^{\beta}] \tag{3}$$

where λ and η, are the location and scale parameters and α and β are shape parameters.

This latter one contains four parameters and its modification by Freimer et al. (1988) is

$$Q(p)=\lambda+\eta\left[\frac{p^{\alpha}-1}{\alpha}-\frac{(1-p)^{\beta}-1}{\beta}\right] \tag{4}$$

In the present paper, a simple model called Davies distribution is adopted:

$$x=Q(p)=\frac{Cp^{\lambda_1}}{(1-p)^{\lambda_2}};\quad C,\ \lambda_1,\ \lambda_2>0,\ 0\leq p<1 \tag{5}$$

where C is a scale parameter and λ_1 and λ_2 are shape parameters. This distribution, introduced by Hankin and Lee (2006) is also proposed along with the extreme value distributions: 3-parameter Weibull, the generalized extreme value (*GEV*) and the 3-parameter generalized Pareto to represent the distribution of wave heights and is applied to real data for further analyses.

2.2. The model and its properties

The derivative of $Q(p)$ is called the quantile density function denoted by $q(p)$ which is directly related to the density function $f(x)$ as

$$q(p)=\frac{dQ(p)}{dp}=\frac{dx}{dp}=\frac{1}{f(x)}$$

$$\text{since } F(x)=p,\ \text{then } \frac{dx}{dp}=\frac{1}{f(x)} \tag{6}$$

Thus the moments of the distribution can be found from

$$\mu_r'=\int_0^{\infty}x^r f(x)dx=\int_0^1[Q(p)]^r dp$$