Ocean Engineering 165 (2018) 164–183

Contents lists available at ScienceDirect

Ocean Engineering

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

Long-term distributions of individual wave and crest heights

Ed Mackay*, Lars Johanning

Renewable Energy Research Group, University of Exeter, Penryn, Cornwall, UK

ARTICLE INFO

Keywords:
Wave height
Crest height
Long-term distribution
Return period
Serial correlation

ABSTRACT

This paper considers three types of method for calculating return periods of individual wave and crest heights. The methods considered differ in the assumptions made about serial correlation in wave conditions. The long-term distribution of individual waves is formed under the assumption that either (1) individual waves, (2) the maximum wave height in each sea state or (3) the maximum wave height in each storm are independent events. The three types of method are compared using long time series of synthesised storms, where the return periods of individual wave heights are known. The methods which neglect serial correlation in sea states are shown to produce a positive bias in predicted return values of wave heights. The size of the bias is dependent on the shape of the tail of the distribution of storm peak significant wave height, with longer-tailed distributions resulting in larger biases. It is shown that storm-based methods give accurate predictions of return periods of individual wave heights. In particular, a Monte Carlo storm-based method is recommend for calculating return periods of individual wave and crest heights. Of all the models considered, the Monte Carlo method requires the fewest assumptions about the data, the fewest subjective judgements from the user and is simplest to implement.

1. Introduction

Estimating the long-term statistics of individual wave or crest heights is an important problem in the design of offshore and coastal structures. The long-term statistics of individual waves are dependent on both the long-term distribution of sea states and the short-term distribution of wave heights or crests heights, conditional on sea state. To produce an accurate estimate of the heights of extreme individual waves, information from the long-term and short-term distributions must be combined in an appropriate manner. The approaches that have been proposed for combining these distributions are equally applicable to both wave heights and crest heights, so to avoid referring to both throughout the text, the following discussion is presented in terms of wave heights.

The simplest approach for estimating extreme individual wave heights at a given exceedance probability, e.g. once in 100 years, is to estimate the significant wave height, H_s, at a return period of 100-years then calculate the most probable maximum wave height in that sea state, assuming a duration of somewhere between 3 and 6 h (see e.g. Hogben, 1990; Tucker and Pitt, 2001). There are several problems with this approach. Firstly, the appropriate duration of sea state to use for calculating the most probable maximum wave height is not clear. Secondly, this approach neglects the probability that the largest wave could occur in a sea state other than the 1 in 100-year H_s. This can lead to significant underestimates in predictions of extremes, since there will be several sea states with H_s close to the most severe value, either within the same storm or in separate storms (Carter and Challenor, 1990).

To overcome the problems of the simple approach, various methods have been proposed to combine the long-term and short-term distributions that account for the probability of large waves occurring in any sea state. Battjes (1970) calculated the total number of waves exceeding a level in a given interval and divided this by the total number of waves in the interval to derive an estimate of the long-term distribution of all individual wave heights. Krogstad (1985) proposed a method for calculating the long-term distribution of the maximum wave height in an interval, derived in terms of the distribution of the maximum wave height in each sea state during the interval (see also Prevosto et al., 2000; Krogstad and Barstow, 2004). Various methods have also been proposed for calculating return periods of individual wave heights from the distribution of the maximum wave height in each storm (Jahns and Wheeler, 1973; Ward et al., 1979; Haring and Heideman, 1980; Boccotti, 1986, 2000; Forristall et al., 1991; Tromans and Vanderschuren, 1995; Arena and Pavone, 2006; Fedele and Arena, 2010; Laface and Arena, 2016; Mackay and Johanning, 2018a; b).

Although the assumption is not always explicit in the derivations of each approach, the various methods proposed all have the common feature of calculating the distribution of the maximum wave height in a

* Corresponding author.
E-mail address: e.mackay@exeter.ac.uk (E. Mackay).

https://doi.org/10.1016/j.oceaneng.2018.07.047
Received 22 May 2018; Received in revised form 5 July 2018; Accepted 15 July 2018
Available online 21 July 2018
0029-8018/

random “event”, where the event is either a storm, a sea state, or a single wave. In the last case the maximum wave height in the event is just the individual wave height. Long-term statistics of individual waves are then calculated from the random event distribution, assuming that occurrences of events are independent.

The assumption of independence of these events is not true in general, with records of wave measurements exhibiting serial correlation at multiple scales, with correlation between successive individual wave heights, sea states and storms. In the statistical literature, the effect of serial correlation on estimates of extreme values is often quantified using the extremal index, which can be introduced as follows. Suppose that $X_1, \ldots, X_n$ are a sequence of n independent variables with common distribution function F. In this case, the distribution of the maximum observation is given by $\Pr(\max\{X_1,\ldots,X_n\} < x) = F^n(x)$. If instead $Y_1, \ldots, Y_n$ are a stationary process, also with common distribution function F, but with some level of serial correlation, then subject to certain regularity conditions, it can be shown that $\Pr(\max\{Y_1,\ldots,Y_n\} < x) = F^{\theta n}(x)$, where $\theta \in [0,1]$ is known as the extremal index (see e.g. Coles, 2001). Serial correlation effectively reduces the probability of large observations in a sequence of a given length. Therefore, for processes with $\theta < 1$, assuming that observations are independent will lead to a positive bias in the estimates of extreme values. Some studies have proposed estimating θ explicitly and using this in the estimation of extremes (e.g. Fawcett and Walshaw, 2012). However, obtaining a reliable estimate of θ is difficult in practice and subject to considerable uncertainties (Davis et al., 2013). For oceanographic data it is more common to adopt the peaks-over-threshold scheme which selects events that are approximately independent, ensuring that $\theta \approx 1$ (Jonathan and Ewans, 2013).

The methods proposed by Battjes (1970) and Krogstad (1985) (referred to as the ‘all-wave’ (AW) and ‘sea state maxima’ (SSM) methods respectively from here onwards) and the various storm-based methods make implicit assumptions about independence between events. The AW method assumes that all wave heights are independent, the SSM method assumes that sea state maxima are independent, and storm-based methods assumes that the maximum wave heights in separate storms are independent. Given the serial correlation in wave height time series, the three methods would be expected to give different results, with the AW method producing the highest estimates and storm-based methods producing the lowest estimates.

Forristall (2008) compared estimates of the long-term distribution of individual wave heights from the AW, SSM and storm-based methods. Forristall conducted Monte Carlo simulations of 100,000 years of individual wave heights from synthetic storms, assuming that the time series of H_s in a storm follows a triangular shape with a fixed relationship between the peak H_s and duration of the storm. The duration, D, of the storm was defined to be the time for which $H_s > 0$ and a value of $D = 8H_{s,peak}$ was used, where D is in hours. The zero-crossing period, T_z, was assumed to be constant at 10s and wave heights were assumed to follow a Rayleigh distribution. It was shown that AW and SSM methods produce a positive bias in estimates of the 100-year wave height, consistent with both models neglecting serial-correlation effects. Forristall showed that the storm-based method of Tromans and Vanderschuren (1995) gave the correct return values of individual wave heights when applied to the synthetic triangular storm data with Rayleigh distributed wave heights.

Forristall's study provides a useful insight into the differences between various methods for calculating return periods of individual waves. However, due to the assumptions about the fixed shape of the storms and constant wave period, it is difficult to draw conclusions about the accuracy of storm-based methods when applied to real data. The purpose of this paper is to compare methods for estimating return periods of individual wave heights based on more realistic simulations of synthetic storms, where the wave period varies throughout the storm and the temporal evolution of sea state parameters is based on measured data. In the current work, a block-resampling method is used generate random time series of realistic storm histories, which are used to compare various methods of estimating return periods of individual wave heights. The block-resampling method divides a time series of measured wave data into discrete blocks consisting of storms where the peak wave heights can be considered approximately independent. The problem of determining time scales over which storm peak wave heights can be considered independent is also discussed in some detail.

The results presented in this paper also have implications for the estimation of extreme load values on marine structures. The “full long-term response analysis” method advocated by some authors (e.g. Sagrilo et al., 2011; Naess and Moan, 2012; Giske et al., 2017) is essentially a method for forming the long-term distribution of the maximum load in each sea state, analogous to the SSM method for wave and crest heights. This method is therefore likely to be subject to the same problems associated with neglecting serial correlation in sea states. Methods for calculating extreme loads which account for serial correlation in sea states have been proposed by other authors (e.g. Forristall et al., 1991; Tromans and Vanderschuren, 1995). However, the focus of this work is on wave and crest heights, and the effect of serial correlation on extreme load values is beyond the scope of the paper.

The paper starts in Section 2 with a brief review of how return periods and return values are defined in the context of the various types of long-term distributions considered. Section 3 presents a short discussion of models for the short-term distribution of wave and crest heights conditional on sea state. Section 4 presents the mathematical derivation of the methods for combining the long-term and short-term distributions, and highlights where various assumptions about independence are made – either implicitly or explicitly. The methods are compared in a simplified example in Section 5, which isolates the effects of serial correlation in sea states. The methods are then applied to measured data in Section 6, providing a quantitative comparison of the effect the various assumptions made in each method in a real situation. The accuracy of the methods is compared in Section 7 using Monte Carlo simulations of synthetic storms. Finally, conclusions are presented in Section 8.

2. Return periods & return values

The methods for estimating the long-term distribution of individual wave heights considered in this paper are used to define return periods and return values in slightly different ways. It is therefore useful to review how return periods and return values are defined in each context. Return values are defined in terms of the distribution of the maximum wave height in a year. We denote the probability that the maximum individual wave height, H_{max}, does not exceed h in any year selected at random as $\Pr(H_{max} \le h|\, 1\text{ year})$. The T-year return value of individual wave height, H_T, is then defined as the value which has an exceedance probability $1/T$ in any year:

$$\Pr(H_{max} > H_T\,|\,1\text{ year}) = \frac{1}{T}, \quad T > 1 \tag{1}$$

The duration T is referred to as the return period and is the average period between exceedances of H_T. Over the last few decades, the peaks-over-threshold (POT) method has gained popularity over the annual-maxima method (see Jonathan and Ewans, 2013, for a review of the use of POT in an oceanographic context). In this method, the distribution of the annual maximum is not estimated explicitly. Instead, the distribution of independent threshold exceedances is estimated. If each independent threshold exceedance is described as an ‘event’ then return values of wave heights can be defined in terms of the distribution of the maximum wave height in a random event, as the solution of:

$$\Pr(H_{max} > H_T\,|\text{event}) = \frac{1}{Tm}, \quad T > \frac{1}{m} \tag{2}$$

where m is the expected number of events per year (see e.g. Coles, 2001). In the present context, the event is either a storm, sea state or