Ocean Engineering 168 (2018) 173–184

Contents lists available at ScienceDirect

Ocean Engineering

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

Study of sampling methods for assessment of extreme significant wave heights in the South China Sea

Zhuxiao Shao[a], Bingchen Liang[a,b,*], Huajun Li[a,b], Dongyoung Lee[a,c]

[a] College of Engineering, Ocean University of China, 238 Songling Road, Qingdao, 266100, China
[b] Shandong Province Key Laboratory of Ocean Engineering, Ocean University of China, 238 Songling Road, Qingdao, 266100, China
[c] Korea Institute of Ocean Science & Technology, Ansan, South Korea

ARTICLE INFO

Keywords:
Extreme significant wave heights
Tropical cyclone waves
Annual maxima
Peak over threshold
Threshold selection
South China Sea

ABSTRACT

Adequate knowledge of extreme wave events is crucial for the coastal and offshore engineering community. In this study, assessment of extreme significant wave heights is performed in the South China Sea (SCS). Significant wave heights from a 40-year (1975–2014) hindcast of tropical cyclone waves are adopted as the initial database. The annual maxima (AM) method with the Gumbel model, a conventional method, is used to assess extreme significant wave heights, which fits the annual maximal significant wave height for extrapolation. The peak over threshold (POT) method with the generalized Pareto distribution (GPD) model is applied to determine return significant wave heights with certain return periods by fitting the peak excesses over a sufficiently large value (the threshold). To select the suitable threshold for the POT method, the sensitivity of the return significant wave height to the threshold is analysed. Based on the sampling theory and the characteristic of hindcasted significant wave heights of tropical cyclone waves, applicability of two sampling methods in the SCS is investigated.

1. Introduction

Reasonable estimation of extreme significant wave heights is highly important for coastal and offshore engineering practices. It affects the design, construction, installation and operation activities of coastal and offshore structures (Karunarathna et al., 2014, 2016; Vanem, 2016; Thompson et al., 2017). To adequately determine the extreme significant wave height, a reliable sampling method and an appropriate probability distribution model are needed.

The annual maxima (AM) method (DNV, 2014) employs the annual maximal significant wave height as the sample for extrapolation (Ruggiero et al., 2010; Arns et al., 2013; Chen et al., 2013; Vanem and Walker, 2013; Li et al., 2016). Compared with the POT method, the AM method is an easy sampling method without additional work. The Gumbel (1958) model is extensively used as a probabilistic model for working with a sample under the AM method (Castillo, 1988; Soares and Scotto, 2001; Calderon-Vega et al., 2013; Hong et al., 2013; Niemann and Diburg, 2013), which may reduce the uncertainty caused by an insufficient number of samples (Suh et al., 2013; Hong et al., 2013). This method (the AM/Gumbel method) is a conventional method for the extreme wave analysis (Jonathan and Ewans, 2013; Xu et al., 2016), especially in the field of regional study of extreme values (Chini et al., 2010; Perez et al., 2017; Vanem, 2017; Polnikov et al., 2017; Li et al., 2018).

The peak over threshold (POT) method (Goda et al., 2001) has usually been employed to identify and select independent peak significant wave heights from initial data (Mathiesen et al., 1994; Ferreira and Soares, 1998; Caires and Sterl, 2005; Ambühl et al., 2014; Liu et al., 2018). Compared with the AM method, the POT method is a natural sampling method without additional limitation. The generalized Pareto distribution (GPD) model (Coles, 2001) is widely applied to extrapolate return significant wave heights based on a sample of the POT method (Alves and Young, 2003; Martucci et al., 2010; Os et al., 2011; Mackay et al., 2011; You, 2011). This method (the POT/GPD method) takes all higher peak significant wave heights above a certain threshold as the sample to adjust the parametric distribution (Caires, 2007; Vanem, 2015; Hawkes et al., 2008; Wang, 2017; Durán-Rosal et al., 2017). Therefore, it might generate reasonable estimates of extreme significant wave heights due to its merits of a reasonable sample and natural distribution, based on a suitable threshold.

In this study, extreme significant wave heights are estimated in the South China Sea (SCS). Considering that the SCS is a tropical cyclone wave-dominated area and these tropical cyclones are major weather systems that drive storm waves (Liu et al., 2008; Doong et al., 2011, 2015), 40-year wave hindcast data obtained during tropical cyclones are used as initial data for extrapolation. These tropical cyclone waves

* Corresponding author. College of Engineering, Ocean University of China, 238 Songling Road, Qingdao, 266100, China.
E-mail address: bingchen@ouc.edu.cn (B. Liang).

https://doi.org/10.1016/j.oceaneng.2018.09.015
Received 4 April 2018; Received in revised form 2 September 2018; Accepted 5 September 2018
Available online 15 September 2018
0029-8018/

are only simulated during every tropical cyclone from 1975 to 2014 in the SCS. The POT method and the AM method are used to extract the sample from initial data, and the GPD model and the Gumbel model are used to extrapolate significant wave heights for different return periods, respectively. To assess the feasibility of candidate thresholds and select the suitable threshold, the analysis of the sensitivity of the return significant wave height to the threshold is presented. Through studying this method, the influence of the excluded samples on the return significant wave height can be found; thus, a stable threshold range covering the suitable threshold can be found. The asymptotic tail approximation and the estimation uncertainty of the POT/GPD method reveal that the selected threshold is reliable. Considering that the sampling method is a basic for assessment of extreme significant wave heights, applicability of two sampling methods in the SCS is investigated based on the sampling theory and the characteristic of hindcasted significant wave heights of tropical cyclone waves. In previous studies (You and Callaghan, 2012; Shao et al., 2017), the deficiency of the AM method for an insufficient number of samples has commonly been accepted. However, as shown by the presented study that the sample of the AM method may be unreasonable when the return period is close to the size of the database, due to the characteristic of hindcasted significant wave heights of tropical cyclone waves (the distribution and representative of the sample).

The article is structured as follows. In the next section, details of the extreme value theory, including the sampling theory and the distribution equation, are provided. Information on initial data and study sites is presented in Section 3. In Section 4, the POT/GPD method is performed in the SCS, and the analysis of the sensitivity of the return significant wave height to the threshold is presented for threshold selection. Applicability of the POT method and the AM method for hindcasted significant wave heights of tropical cyclone waves is investigated in Section 5. Finally, the conclusions are presented in Section 6.

2. Background

2.1. POT/GPD method

The POT method extracts a series of independent peak significant wave heights above a threshold as the sample.

Extract N peak significant wave heights (X_1, X_2, …, X_N with common distribution function $F(x)$) as the initial sample from the initial database under an independent and identically distributed assumption. For a threshold, u, that is sufficiently large, the distribution of the peak excess over the threshold, i.e., $Y = X - u$, is given by

$$G(y) = P[Y \le y | X > u] = \frac{F(u+y) - F(u)}{1 - F(u)} \tag{1}$$

Thus, the unconditional distribution $F(x)$ can be expressed, in terms of $G(y)$:

$$F(x) = (1 - p_u) + p_u G(y) \tag{2}$$

Where $p_u = 1 - F(u)$ is the probability of threshold exceedance.

According to the study of Pickands (1975), the distribution $G(y)$ for peak excesses over the threshold can be approximated by a member of the GPD:

$$G_u(y) = \begin{cases} 1 - \left(1 + k\frac{y}{\sigma}\right)^{-\frac{1}{k}} & \text{k} \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right) & \text{k} = 0 \end{cases} \tag{3}$$

Where σ is the scale parameter and k is the shape parameter. These GPD parameters (σ and k) are estimated using the maximum likelihood estimation method, which is recommended by Mazas and Hamm (2011):

$$\ln L(k, \sigma; x) = \begin{cases} -N_u \ln\sigma + \left(\frac{1}{k} - 1\right)\sum_{j=1}^{N_u} \ln\left(1 - \frac{kX_j}{\sigma}\right) & \text{k} \neq 0 \\ -N_u \ln\sigma - \frac{1}{\sigma}\sum_{j=1}^{N_u} X_j, & k = 0 \end{cases} \tag{4}$$

where N_u is the number of events exceeding the threshold (the number of samples).

The probability of threshold exceedance p_u is estimated by means of the empirical distribution function (Embrechts et al., 1997):

$$\widehat{p_u} = \frac{N_u}{N} \tag{5}$$

Through combing Eq. (2) with Eqs. (3) and (5), the approximation of the distribution function can be obtained,

$$F(x) = \begin{cases} 1 - \frac{N_u}{N}\left(1 + k\frac{y}{\sigma}\right)^{-\frac{1}{k}} & k \neq 0 \\ 1 - \frac{N_u}{N}\exp\left(-\frac{y}{\sigma}\right) & k = 0 \end{cases} \tag{6}$$

The return period, i, associated with a certain return level, X_i, is defined as follows:

$$i = \frac{1}{1 - F(X_i)} \tag{7}$$

Accordingly, X_i is calculated by means of the inverse function F^{-1} of distribution F:

$$X_i = F^{-1}\left(1 - \frac{1}{i}\right) = \begin{cases} u + \left[\left(\frac{N_u}{N}i\right)^k - 1\right]\sigma/k & k \neq 0 \\ u + \sigma\ln\left(\frac{N_u}{N}i\right) & k = 0 \end{cases} \tag{8}$$

2.2. AM/Gumbel method

The AM method directly extracts the annual maximal significant wave heights as the sample. The widely used probabilistic model for working with a sample under the AM method is the Gumbel model (Gumbel, 1958). The distribution function of the Gumbel model can be defined as follows:

$$F(x) = \exp(-exp(-(x - \beta)/\alpha)) \tag{9}$$

where β represents the location parameter, and α represents the scale parameter. These parameters (β and α) are estimated using the maximum likelihood estimation method (Embrechts et al., 1997).

The i -year return level is defined as follows:

$$X_i = \beta + \alpha y_i \tag{10}$$

where $y_i = -\ln(-\ln(1 - 1/i))$.

3. Initial data and study sites

3.1. Initial data

The dataset analysed in this study consists of a wave hindcast dataset of tropical cyclones in the SCS and is evaluated using the third-generation spectral wind-wave model SWAN (an acronym for Simulating WAves Nearshore). This model can simulate the growth, decay and transformation of wind-generated waves and swells during tropical cyclones.

To avoid the influence of boundary conditions, a wider area than the SCS is selected as the computational area (at 0°N–40°N and 100°E–150°E, shown in Fig. 1). The space resolution is 0.0625° for both longitude and latitude. The directional space is resolved in 48 equal directions. In the frequency space, the number of frequencies is 36, with a minimal frequency of 0.03 Hz and a maximal frequency of 1 Hz. Both linear and exponential growths of waves by winds are included in the model (Cavaleri and Rizzoli, 1981; Snyder et al., 1981). Dissipation due to depth-induced wave breaking is treated by the spectral formulation