Coastal Engineering 126 (2017) 81–95

Contents lists available at ScienceDirect

Coastal Engineering

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

Regional frequency analysis of extreme waves in a coastal area

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

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

ARTICLE INFO

Keywords:
Regional frequency analysis
Cluster analysis
Regional extreme wave height quantile
L-moments
Discordant site
Sub-region

ABSTRACT

This study analysed the wave data from several locations in a coastal region to identify the areas with comparable wave height statistics and to estimate regional extreme significant wave heights of designated return periods. The *regional frequency analysis* algorithm executed in this work utilised the significant wave height data from 35 sites (35°–45°N, ̄6.5°–̄11.0°W) in a coastal grid (0.25° × 0.25°) off Portugal in the North Atlantic Ocean, extracted from 44 years HIPOCAS hindcast wave database. *Regional frequency analysis* algorithms based on *L-moments* identified the discordant sites (discordant site statistics are substantially diffused from the group of sites statistics), assisted in the formation of homogeneous regions by *cluster analyses* (data vectors include site characteristics and site statistics) and consequently selected appropriate regional frequency distributions by *Z* goodness-of-fit test statistic at 90% level of significance to estimate regional extreme significant wave height quantiles of designated return periods. The *regional frequency analysis* algorithm recognised in certain cases, neighbouring sites as members of diverse regions, revealing that geographical proximity of sites are not ascertaining factors to form homogeneous regions. *At-site analyses* along with *regional frequency analysis* enabled to apprehend the precision of the regional extreme quantile as genuine feature of its at-site extreme quantiles.

1. Introduction

The knowledge of the frequency of the expected extreme significant wave heights is important to design marine structures and to plan maritime operations. However, extreme event data are scarce and thus they have to be predicted by fitting distributions to relatively small samples of data. In general, this is done by adjusting different types of distributions to buoy data sets, identifying distributions and return periods of extremes that are representative of a certain region (e.g. Ferreira and Guedes Soares, 2000; Guedes Soares and Scotto, 2004).

Extreme wave height modelling has been addressed by many researchers, as for example, Ferreira and Guedes Soares (1998); Coles (2001); Ewans and Jonathan (2008); Menendez et al. (2009); Thomson et al. (2009); Ruggiero et al. (2010); Jonathan and Ewans (2013); Randell et al. (2015). However, these methods have been applied to data of a specific location, without any information of the extent of the geographical area that can be described by the same probabilistic law. This resulted from the absence of available wave data spatially distributed in wide areas.

The increased availability of hindcast data such as for instance ERA-40 (European Centre for Medium-Range Weather Forecasts (ECMWF) (Caires and Sterl, 2005) and HIPOCAS (Hindcast of Dynamic Processes of the Ocean and Coastal Areas of Europe) (Guedes Soares, 2008) improved the situation. These data sets have wave and wind data on a spatial grid that allow performing extreme wave analysis at a large number of neighbouring sites, and also allow studying the spatial variability of the wave climate. *Regional frequency analysis* (*RFA*) provides a suitable method to deal with this problem, as it uses data from several sites with similar wave statistics to estimate the frequency distribution of a homogeneous region. By using this procedure, the data sets are increased by “trading space for time”, i.e., it pools the data from the neighbouring sites evaluated as having similar frequency distributions to the site of interest. In the *regional frequency analysis* approach, a single frequency distribution is fitted to data from several sites of similar wave statistics.

The *regional frequency analysis* approach has been applied by several authors, such has Bernardara et al. (2011) who applied the *regional frequency analysis* to extreme storm surges. The study was conducted with surge data of 18 French harbours located in the Atlantic coast. Weiss et al. (2014a) derived a theoretical model of inter-site dependence for *regional frequency analysis* of extreme marine events, such as storms, using wave data from shallow and deep water points located in the North-East part of the Atlantic Ocean. Weiss et al. (2014b) also conducted a detailed study using extreme significant wave height data of coastal and deep waters, and proposed a method to form homogeneous regions by the

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

http://dx.doi.org/10.1016/j.coastaleng.2017.06.002
Received 2 March 2016; Received in revised form 1 April 2017; Accepted 2 June 2017
Available online 10 June 2017
0378-3839/

identification of typical storms footprints.

In *regional frequency analysis* more information is used than in an *at-site analysis* which uses only a single site's data, and hence there is potential for greater accuracy in the final quantile estimates (Hosking and Wallis, 1997).

The mean of the at-site quantiles is its regional quantile. In this study, in order to increase the robustness of regional extreme quantile estimations as the true representatives of its at-site quantiles, *at-site analyses* are also performed along with *regional frequency analysis*. The basic concept in *regional frequency analysis* is that the at-site distributions of a region are identical apart from a scaling factor, which is taken to be the mean of the at-site frequency distributions.

L-moments are utilized in the estimation process of *regional frequency analysis* (Hosking and Wallis, 1997). These quantities are similar to the conventional moments and can be estimated by linear combinations of probability weighted moments (*PWMs*), i.e., *L-statistics,* where the *L* in the word *L-moments* is for linear functions of the data. *L-moments* are a relatively recent development within statistics and they facilitate the estimation process in *regional frequency analysis*. The advantage of the *L-moments* over the conventional moments is the ability of being able to characterize a large number of distributions and when estimated from a sample of being more robust to the presence of outliers in the data. From the different methods of *regional frequency analysis* that exist, the regional probability weighted moments algorithm is considered to be more efficient (Cunnane, 1988) and therefore is the basis of the *RFA* method.

L-moment ratios i.e., the *L-CV* (τ), the *L-skewness* (τ_3), the *L-kurtosis* (τ_4) and τ_5 characterise the shape of a distribution independently of its scale of measurement. *L-CV* is a dimensionless measure of variability and for a probability distribution that has only positive values, which is the case in this work, the *L-CV* (τ) is in the range $0 \leq \tau < 1$. Negative values are only conceivable if the at-site mean has a negative value. The *L-skewness* (τ_3) is a dimensionless measure of asymmetry, which can have positive or negative values. For a distribution, the *L-skewness* is in the range $0 < |\tau_r| \leq 1; r \geq 3$.

The first phase in *regional frequency analysis* is to find discordant sites, which are those that are grossly incompatible with the group and need to be discarded. The second stage is to perform *cluster analysis* based on *Ward*'s method (Hosking and Wallis, 1997) to compose sets of sites that more or less satisfy the homogeneity condition, with frequency distributions that are identical apart from a site categorical scale factor. This is to be achieved by defining for each site of interest a region containing those sites whose information are advantageously used to estimate the frequency distribution at the site of concern. The importance of defining sub-regions relies in the fact of being possible to characterize the area in terms of its member sites *L-moment ratios* and suggest relevant regional distributions to estimate regional extreme significant wave height quantiles of assigned return periods.

Hosking and Wallis (1997) considered that when the data are generated from a physical process that can generate outliers, then a distribution that is a close fit to the current observed data will not guarantee the same with the future data. So, it is suitable to use a robust approach that will yield reasonably accurate quantile estimates even when the true at-site frequency distribution deviates from the fitted frequency distribution. By *regional frequency analysis* method there will be no single genuine distribution that applies to every site. Therefore, the aim is not to identify the “true” distribution but rather to discover a distribution that will yield accurate quantile estimates for every site.

It is essential to fit a probability model to the observed significant wave height frequency distribution to evaluate the extreme quantiles of given return periods. Albeit the region is moderately heterogeneous, *regional analysis* will still yield much more accurate quantile estimates than *at-site analysis* (Hosking and Wallis, 1988; Potter and Lettenmaier, 1990).

This study aims to apply the *regional frequency analysis* approach based on *L-moments* to identify discordant sites, perform *cluster analysis* to form homogeneous regions, find out appropriate regional frequency distribution and derive reasonably precise estimates of extreme wave height quantiles. Lucas et al. (2011, 2012a,b) has explored the application of this approach to limited areas or to limited data sets and now a systematic study is presented covering the whole area around Portugal and a large period of collected data.

The 44 years of HIPOCAS wave data (Pilar et al., 2008) are used in the analysis. Recently this data set has been compared with others (Campos and Guedes Soares, 2016a, b), showing its consistency, although with some tendency to over predict the extremes, which should be taken into consideration in interpreting the present results.

2. Regional frequency analysis

The *regional frequency analysis*, *RFA* (Hosking and Wallis, 1997), is briefly described here. The various steps in *regional frequency analysis* are:

2.1. Screening of the data

The data should be screened for homogeneity (stationary) over time, i.e., the sampled data should be representative of the population, which keeps its random characteristics constant along the time. The *NCEP* (US National Centre for Environmental Protection) global atmospheric re-analysis wind data set were used to produce 44 years homogeneous and high resolution atmospheric data sets (Weisse and Feser, 2003), which were used to force the numerical wave model *WAM* (third generation wave model) to generate the wave data of the ocean and the coastal areas of Europe. Therefore, the wave data can be assumed to be homogeneous over time. In *RFA*, vital information can be obtained by comparing the sample *L-moment ratios* for different sites. Guedes Soares and Ferreira (1995) removed the seasonal components of the three hourly non-stationary time series making the resulting series stationary and enabling the modelling with ARMA (autoregressive-moving-average) models. In *RFA*, incorrect data values, anomalies, trends and shifts in the mean of a sample are all reflected in the *L-moments* of the sample.

A measure of the discordance between the *L-moment ratios* of a site and the average *L-moment ratios* of a group of similar sites is applied to the data to identify discordant sites, i.e. those sites that are grossly incompatible with the group.

2.2. Identification of homogeneous regions

The next step in *regional frequency analysis* is the assignment of sites to regions. A region consists of sites whose frequency distributions are considered to be approximately similar. Approximate homogeneity is sufficient to ensure that *regional frequency analysis* is much more accurate than *at-site analysis*. Sites with similar site statistics (*L-moment ratios*) constitute a region. *Cluster analysis* based on *Ward's* method assists to form regions and takes into account site characteristics; latitude and longitude and site statistics; *L-CV* and *L-kurtosis* as in this study. The motivation to consider *L-CV* and *L-kurtosis* in *cluster analysis* algorithm to form homogeneous regions is that the heterogeneity measure is defined in terms of *L-CV* and the goodness-of-fit measure, the *Z*-test statistic of the fitted regional frequency distribution is defined on *L-kurtosis*.

2.3. A heterogeneity measure

The heterogeneity measure compares the between-site variations in sample *L-moments* for the group of sites with what would be expected for a homogeneous region. The between-site variation in *L-CV* has a much larger effect than variation in *L-skewness* or *L-kurtosis*.

If the proposed region has *N* sites, with site *i* having sample size n_i and sample *L-moment ratios L-CV, L-skewness, L-kurtosis* denoted by $t^{(i)}$, $t_3^{(i)}$,