Coastal Engineering 100 (2015) 37–47

Contents lists available at ScienceDirect

Coastal Engineering

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

Comparing different extreme wave analysis models for wave climate assessment along the Italian coast

L. Sartini *, L. Mentaschi, G. Besio

DICCA — Dipartimento di Ingegneria Civile, Chimica e Ambientale, Università degli studi di Genova, Via Montallegro, 1-16145 Genova, Italy

ARTICLE INFO

Article history:
Received 23 May 2014
Received in revised form 16 January 2015
Accepted 18 March 2015
Available online 15 April 2015

Keywords:
Wave climate
GEV
GPD–Poisson
POT
Wave modeling
Wave hindcasting

ABSTRACT

By utilizing various statistical models to quantify the return levels of extreme significant wave height this study seeks to achieve two objectives: the updating of the state-of-the-art concerning extreme wave analysis along the Italian coast and the creation of a long-term predictive model. To these ends, four different methods widely used in the field of metocean engineering are employed to analyze both buoy data (Rete Ondametrica Nazionale, RON) and wave data obtained by means of dynamical hindcasting techniques such as the forecast/hindcast operational model chain in use at the University of Genoa (www.dicca.unige.it/meteocean). Return levels are estimated by the Goda method, the Generalized Extreme Value (GEV) and the Generalized Pareto Distribution–Poisson point process models and the Equivalent Triangular Storm (ETS) algorithm. All models follow the Peak-Over-Threshold (POT) approach which require an optimal threshold implementation, save for the GEV analysis, which is applied to model significant wave height maxima pertaining to time-blocks. The models exhibit different performance characteristics, presented here and treated in depth. In general, noteworthy versatility characterizes the GPD–Poisson model, which often recovers Goda results, while the GEV and ETS models exhibit limitations in assessment of a variety of wave fields, greatly diversified in a semi-closed basin such as the Mediterranean Sea. Long-term estimates have been provided by means of the most appropriate model selected, thus offering a complete overview of wave climate in the Mediterranean basin based on wave hindcast data.

1. Introduction

The application of Extreme Value (EV) theory to coastal engineering is problematic and complicated because quality data are difficult to obtain and also because many different statistical methods are used to generate return level estimates. Pioneer contributions, largely applied in the fields of hydrology and climatology, were made by Gumbel (1958) and by Fréchet (1927), who introduced asymptotic models, by Weibull (1939) as well as Fisher and Tippett (1928) whose work, later extended by Gnedenko (1943), focused on the convergence of extreme distributions. Any asymptotic model used to select extreme data reflects one of two radically different approaches: the Fisher–Tippet theorem (theorem I) (Galambos, 1987) or the Pickands theorem (theorem II) (Pickands, 1975). The former approach collects data in fixed time periods; as a result, maxima are distributed identically. The latter approach relies on exceedance beyond a given threshold (Peaks Over Threshold, POT) (Davison and Smith, 1990; Goda, 1988; Naess and Clausen, 2001; Smith, 2001). The first method decreases statistical error because it requires no threshold selection. However, this leads to data scarcity. The second method benefits from larger data sets (Hawkes et al., 2008; Méndez et al., 2006), which are basically dependent on the threshold choice. Consequently, threshold selection is critical, more so when covariate effects are present because they could influence the threshold selected. Moreover, in the specific case of multivariate extremes modelling, threshold selection is further complicated by potential interaction between variables and by inadequate criteria for joint tail estimation (Jonathan and Ewans, 2013).

It could be possible to obviate the need to determine a threshold by extending EV models into the body of distributions, for example, by integrating the threshold choice within a Bayesian framework (MacDonald et al., 2011; Tancredi et al., 2006; Wadsworth et al., 2010) or by basing the model on an average of multiple thresholds. Jonathan and Ewans (2013) note that none of these options has been widely adopted in standard statistical methodology (Dupuis, 1999). On the contrary, concerning covariate effects, graphical approaches, such as non-linear threshold (Scotto and Soares, 2000) and quantile regression (Koenker, 2005; Thompson et al., 2009) methods have routinely been preferred. Obviously, when visual analysis is involved, as with the mean residual life plot (Coles, 2001) a model's fit reliability depends on the accurate interpretation of threshold plots. Thompson et al. (2009) illustrate the principal difficulties associated with threshold plot interpretation.

Among the methods requiring threshold implementation, in recent years the Generalized Pareto Distribution (from now on GPD) has been largely used in order to study maxima distribution tails (Castillo

* Corresponding author. Tel.: +39 010 3536576.
E-mail address: ludovica.sartini@dicca.unige.it (L. Sartini).

http://dx.doi.org/10.1016/j.coastaleng.2015.03.006

et al., 2004; Coles, 2001; Holthuijsen, 2007; Mazas and Hamm, 2011; Méndez et al., 2006; Thompson et al., 2009) as an improvement to data wastefulness typical of the Generalized Extreme Value (GEV) models, based on identical-distributed maxima (Cañellas et al., 2007; Martucci et al., 2009).

Beginning with Borgman (1970, 1973) and later studies (Krogstad, 1985) focused on the observation of the properties shown by waves exceeding a threshold during storms, Boccotti (2000) developed the Equivalent Triangular Storm (ETS) algorithm later reviewed and applied by Arena and Pavone (2006, 2009).

Recent advances representing valid perspectives in metocean engineering allowed innovative methods for extreme values analysis, such as the works of Vrac and Naveau (2007) and Michelangeli et al. (2009), which combine reliably extreme characteristics of local short-term hindcast data with large-scale long-term climate simulations to provide estimates of return levels. The studies of Bernier et al. (2007) and Kyselỳ et al. (2010) incorporate climate change effects within return level estimation, while the development of generalized linear models able to simulate spatio-temporal systems with climatology purposes has been proposed by Chandler (2005). Further analysis focused on the suitability of single locations compared with the regional ones to provide return levels, similarly to the regional frequency models, can be found in Kyselỳ et al. (2011), with the purpose of defining best single site estimates.

Despite the fact that extreme wave analysis has greatly improved on the basis of field measurements, the growing need for reliable data sets which cover longer periods of time has shifted attention toward statistical models applied to hindcast data (Breivik et al., 2009; Golshani et al., 2007; Silva and Mendes, 2013; Stephens and Gorman, 2006). The present study, undertaken within the framework of the EV analysis, attempts to perform a detailed analysis of extreme wave height estimates along the Italian coast by making use of four different models (Goda POT, GEV, GPD–Poisson point process and a revised ETS method). The primary objective is to identify the best model for the processing of return level estimates for maritime design. To this end, the above-cited models were applied to 32 years (1979–2010) of wave simulation to create a wave climate atlas of the Mediterranean Basin. This period of simulation brought to light the problems and the performance characteristics of the various models. Further, the entire analysis rests upon the employment of an up-to-date wave hindcast model in order to allow for the development of long-term models for extreme sea waves. The manuscript is organized as follows. In Section 2, first of all the four statistical methods are presented; then, both the hindcast and field data will be described. Section 3 shows obtained results and some relevant aspects will be discussed. The final section gives some suggestions for future developments and perspectives.

2. Methods

2.1. Statistical models

This section provides a short overview of models applied on both wave hindcast and buoy data.

2.1.1. GEV model

Defining z the maxima value sample of a Z independent maxima dataset, the GEV non-exceedance probability, evaluated in a positive neighborhood $[]_+$, is expressed by:

$$F(z) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{z-\mu}{\psi}\right)\right]_+^{-1/\xi}\right\} & \xi \neq 0 \\ \exp\left\{-\exp\left[-\left(\frac{z-\mu}{\psi}\right)\right]\right\} & \xi = 0, \end{cases} \tag{1}$$

where the location ($\mu > 0$), scale ($\psi > 0$) and shape (ξ) parameters define the center of distribution, the size of deviations related to the location parameter and the upper tail decay, respectively. Model parameters can be estimated by the likelihood method, where, given m observation $\{(t_1, z_1), \ldots, (t_m, z_m)\}$ of a period t_i at which the maximum z_i is attained,

$$l(\theta|t_i, z_i) = -\sum_{i=1}^{m}\left\{\log\psi_i + \left(1+\frac{1}{\xi_i}\right)\log\left[1+\xi_i\left(\frac{z_i-\mu_i}{\psi_i}\right)\right]_+ + \left[1+\xi_i\left(\frac{z_i-\mu_i}{\psi_i}\right)\right]_+^{-1/\xi_i}\right\} \tag{2}$$

is the log-likelihood function, becoming:

$$l(\theta|t_i, z_i) = -\log\psi_i - \frac{z_i-\mu_i}{\psi_i} - \exp\left[-\frac{z_j-\mu_i}{\psi_i}\right] \tag{3}$$

in case of a Gumbel distribution. Maximization of Eqs. (2) and (3) leads to θ, the vector containing model parameters. Corresponding quantiles are given by:

$$x_q = \begin{cases} \mu - \frac{\psi}{\xi}\left[1-(-\log q)^{-\xi}\right], & \xi \neq 0, \\ \mu - \psi\log(-\log q), & \xi = 0, \end{cases} \tag{4}$$

where $1 - q$ is the probability associated with the return period $1/q$. Delta method (Rice, 1994) is used to define confidence interval connected to extreme quantiles. GEV distribution is applied to model extreme events, referring them to cadenced time blocks; thus, maxima are identically distributed and stochastically independent, assuring that their marginal probability is equivalent. These statistical features imply that GEV models are not widely used in extreme wave analysis (Chini et al., 2010; Soares and Scotto, 2004; Sobey and Orloff, 1995) but largely applied to other environmental variables such as sea level, daily temperature and rainfall (Church et al., 2006; Haigh et al., 2010; Hunter, 2010). On the contrary, in the specific case of sea waves, the adoption of identically-distributed significant wave height maxima involves losing sight of the whole wave dataset in its entirety; as a consequence, return level estimates are arranged in trends characterized by high slopes. This implies that return levels related to higher levels are often overestimated, while results achieved for lower return periods could be unlikely. This behavior is strongly influenced by maxima type chosen in function of to the time span considered (e.g. annual or monthly maxima). Very good results could be achieved modeling with GEV maxima exceeding a given threshold; since the last approach is not particularly sound statistically, the GPD model and its derivates were introduced to compensate for its shortfalls.

2.1.2. GPD–Poisson, point process model

As already explained, GPD models, applied to maxima samples selected by means of a defined threshold, overcome GEV weakness induced by data scattering. According to Pickands (1975) and Embrechts et al. (1997), if the threshold is high enough, the Generalized Pareto Distribution exceeding probability is defined as:

$$G(y;\sigma,\xi) = \begin{cases} 1-\left(1+\xi\frac{y}{\sigma}\right)^{-\frac{1}{\xi}} & \xi \neq 0, \\ 1-\exp\left(-\frac{1}{\sigma}\right) & \xi = 0, \end{cases} \tag{5}$$

where y is the exceedances sample and σ and ξ are the scale and shape parameters. An advanced version of basic model is obtained by combining the GPD and Poisson models in accordance with the assumption that the number of exceedance over a threshold u follows a Poisson distribution with mean event rate ν, while the excesses are modeled using Eq. (5). Following Katz et al. (2002), Pickands (1975) and Smith (2001), the combination of the Poisson model for frequency and GPD model for intensity leads to a combined form compatible with a GEV distribution as follows:

$$\sigma = \psi + \xi(u-\mu), \quad \nu = [1+\xi(u-\mu)/\psi]^{-1/\xi} \tag{6}$$