2016-11-01

A copula-based approach for the estimation of wave height records through spatial correlation

Jane, R

10.1016/j.coastaleng.2016.06.008
Coastal Engineering
Elsevier

Coastal Engineering 117 (2016) 1–18

Contents lists available at ScienceDirect

Coastal Engineering

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

A copula-based approach for the estimation of wave height records through spatial correlation

R. Jane[a,*], L. Dalla Valle[b], D. Simmonds[a], A. Raby[a]

[a] University of Plymouth, School of Marine Science and Engineering, Drakes Circus, Plymouth PL4 8AA, UK
[b] University of Plymouth, School of Computing, Electronics and Mathematics, Drakes Circus, Plymouth PL4 8AA, UK

ARTICLE INFO

Article history:
Received 17 December 2015
Received in revised form 17 June 2016
Accepted 25 June 2016
Available online 29 July 2016

Keywords:
Significant wave height
Wave hindcasting
Multivariate copula
Spatial correlation
Wave model
SWAN

ABSTRACT

Information on the wave climate at a particular location is essential in many areas of coastal engineering from the design of coastal structures to flood risk analysis. It is most commonly obtained either by direct measurements or hindcast from meteorological data. The extended deployment of a wave buoy to directly measure wave conditions and the application of wave transformation models used in hindcasting, including public domain models such as Wavewatch and SWAN, are both expensive. The accuracy of the results given by the latter are also highly sensitive to the quality of the wind data used as input. In this paper a new copula-based approach for predicting the wave height at a given location by exploiting the spatial dependence of the wave height at nearby locations is proposed. By working directly with wave heights, it provides an alternative method to hindcasting from observed or predicted wind fields when limited information on the wave climate at a particular location is available. It is shown to provide predictions of a comparable accuracy to those given by existing numerical models.

1. Introduction

In a standards-based approach, structures are designed to withstand an event of a given severity. For coastal structures an event is traditionally characterised by a significant wave height and its severity defined in terms of a return period. The return period is defined as the average time elapsing between two consecutive occurrences of a prescribed event. In order to uncover the significant wave height corresponding to a given return period, information on the local wave climate is required. Depending on the available budget and richness of wave data in the locality a range of approaches exist for obtaining such data including long term deployment of a wave buoy, short term deployment and subsequent hindcasting to give a sufficiently long record of the local wave climate or the use of wave data from a nearby location. Taking advantage of the increase in readily available computational power and advances in the modelling of the structure of the dependance between non-independent random variables over the last couple of decades, this work presents an alternative method for deriving design conditions which has the potential to outperform the existing approaches.

Extensive networks of wave buoys are in place along many populated coastlines where flood risk is actively managed. To name but two, the Channel Coastal Observatory (CCO) maintains a network of around 40 wave buoys along the UK coast with the National Oceanic and Atmospheric Administration's (NOAA) National Data Buoy Center maintaining a similar network of over 100 wave buoys in the USA. Some wave buoy deployments such as these are, in principle, intended to be permanent with the requirement for regular maintenance. In others a wave buoy may be deployed for a short period of time to gather wave data in a new or critical location over a few seasons or years. The wave time series from these can be used to calibrate numerical hindcast models driven by meteorological forcing, or to validate wave transformation models driven by other wind and wave data. In both cases, the existing data sets may need to be extended, either because of missing data created by a damaged or vagrant buoy or because observations at a temporary site are not sufficient to build an accurate model for localised extremes.

It is reasonable to assume that observations of wave fields generated by the same physical processes in the same region, should be correlated. These observations made along the same natural coastline will be modified by variation of the direction and degree of exposure and variability in the wave transformation processes created by the variable bathymetry. Local generative processes may also prove a

* Corresponding author.
E-mail addresses: robert.jane@plymouth.ac.uk (R. Jane), luciana.dallavalle@plymouth.ac.uk (L. Valle), D.Simmonds@plymouth.ac.uk (D. Simmonds), alison.raby@plymouth.ac.uk (A. Raby).

http://dx.doi.org/10.1016/j.coastaleng.2016.06.008

reason for variability. By exploiting and modelling such correlations between the wave height observed at a given location and at two or more neighbouring locations, the proposed approach offers a simple and computationally efficient alternative method for generating missing or extended data at a site of interest. In the case of extremes analysis, buoys frequently malfunction during energetic events so such an approach would be useful to infer the absent extreme values.

In the case of buoy networks, once the correlations are suitably determined a virtual wave buoy model will have been created, an idea postulated in Londhe (2008). This might enable a reduction in the number of buoys, or allowing the swapping in and out of buoys during a programme of maintenance with minimal loss of measurement system accuracy. Removed wave buoys might also be redeployed at other locations. The method thus promises a means for increasing the detailed knowledge of wave climate along a stretch of coast by removing redundant measurements.

In each of the applications discussed so far our interest lies in the extreme values of these variables. Extreme values of the variables concerned are modelled through fitting extreme value distributions to the observed extremes (Section 2). With longer length of data sets more extreme values are likely to be captured leading to more accurate definition of the return periods. However, due to time and financial constraints the long term deployment of a wave buoy is often not practical and therefore other means of deriving the local wave conditions have been contrived. For coastal areas where no data is available, historic offshore wave conditions are commonly derived from records of the local wind field, a process referred to as hindcasting. Originating in the 1940s and 50s with simple empirical models such as SMB and SPM, today computationally intensive numerical wave transformation models such as WAM (WAve Modelling) (WAMDI group, 1988), Wavewatch III (Tolman, 1991) and SWAN (Simulating WAves Nearshore) (Booij et al., 1999) dominate the field. These models work on the principal that most statistical properties of wind waves are captured in their wave energy or action spectrum. Assuming the principle of linear superposition, observations of the water surface at a given location can be decomposed into a spectrum of wave energy as a function of frequency. With additional local observations, a two dimensional spectrum can be constructed to describe the directional variation of the wave energy spectrum. These are used in spectral wave models. The action spectrum is the energy spectrum divided by the intrinsic frequency of the spectral components. The models numerically integrate the wave energy or action transport equation which governs the evolution in space and time of their respective spectra to predict future wave conditions at different locations.

Numerical wave models such as WAM and Wavewatch were developed primarily for predicting deep ocean wave conditions where the waves are mostly wind driven. The source terms present in the transport equations represent the physical processes that can result in generation, dissipation, or redistribution of wave energy. For the WAM model, inputs required for the source terms include wind data, information on non-linear (Quadruplet) wave-wave interactions and white capping dissipation. On the other hand, numerical models such as SWAN were developed for predicting nearshore wave conditions. This includes not only each of the source terms present in deep water models but also terms for triad wave–wave interactions which become much more significant than the higher order wave-wave interactions in the nearshore region. It also includes a term to account for depth induced breaking which can lead to a significant reduction in the maximum significant wave height in coastal regions. The SWAN model uses the wave action spectrum rather than the wave energy spectrum since, in contrast to the energy spectrum, it is conserved in the presence of currents thus allowing the effect of a mean current on the evolution of the wave field to be included in the modelling. In practice, to take advantage of both models, it is common for the generated wave field from a WAM model to subsequently form the boundary conditions for a higher resolution wave transformation model, such as SWAN for propagating the waves inshore (Wornom et al., 2001).

The accuracy of the hindcast offshore wave conditions are highly dependent on the accuracy of the data on the forcing wind field (Holthuijsen et al., 1996; Moeini et al., 2010; Teixeira et al., 1995). These wave fields are derived from the measured wind fields and act as the primary input for the prediction of the coastal wave conditions. Thus, the quality of the predicted coastal conditions relies on accurate descriptions of the forcing wind field. Also, the exercise of assimilating the data and running numerical wave models, which may involve, long run times and multiple sensitivity analyses to validate the output, makes such an exercise non-trivial and often significantly costly. In order to reduce the computational cost meta-models, that are simplified approximations of computationally intensive models, have found favor. For example, Camus et al. (2011) developed a model that uses radial basis functions as a meta model for SWAN. Other methods for reducing the computation burden include coupling numerical and soft computing techniques as is done in Malekmohamadi et al. (2008) by way of WAM and a neural network. Again, this can require a considerable time and cost, but once set up, these models can run very efficiently.

In addition to the finite available computational power, these deterministic modelling approaches are also hampered by a lack of knowledge of the physical processes that generate waves, the interrelated parameters that drive them as well as insufficient detail concerning the modelled region. Moreover, wave energy spectrum-based models are believed to be reaching the limits of the accuracy with which they can simulate wave hydrodynamics (Liu et al., 2002). Any further improvements are likely to occur at a much slower pace than in the past and these may not lead to any practical improvement to the accuracy of their predictions (Cavaleri et al., 2007). In reality, even if substantial advancements were made in capturing the detail of physical processes and their interactions, the complexity of the situation will limit deterministic prediction.

In acknowledgment of the uncertainties associated with physics based models, over recent decades scientists have started to considere the problem probabilistically. Although not normally considered to be a Gaussian process, Gaussian models of varying complexity were the first to be investigated (Cunha and Soares, 1999; Scotto and Soares, 2000; Soares et al., 1996). More recently soft computational techniques that can be applied directly to observations have become increasingly popular. Artificial Neural Networks (ANN's) are perhaps the most widely adopted of these techniques. ANN's aim to mimic how biological neural systems such as the human brain process information. They consist of a set of interconnected processing units or nodes, and a set of weightings along the connections or synapses, that are analogous to synaptic strength. The arrangement of the nodes are prescribed by the user with the weights adjusted through a learning or training procedure. This references a cost function so that the relationship between the input stimuli and the output responses becomes optimised. They have been shown to outperform the auto regressive models when forecasting wave heights (Deo and Naidu, 1998). As such they have prevailed as the dominant approach for prediction using either previous wave height observations at a location (Deo and Naidu, 1998; Gopinath and Dwarakish, 2015; Hadadpour et al., 2014; Londhe and Panchang, 2006; Mandal and Prabaharan, 2006) or wind wave data (Deo et al., 2001; Kamranzad et al., 2011; Malekmohamadi et al., 2008). They have also been applied to improve the accuracy of physics based process models (Zhang et al., 2006). Other soft computational approaches such as genetic programming (Gaur and Deo, 2008; Nitsure et al., 2012), fuzzy inference systems (Kazeminezhad et al., 2005; Özger and Şen, 2007) and support vector machines (Mahjoobi and Mosabbeb, 2009) have also been shown to provide useful predictions (Malekmohamadi et al., 2011). For a more detailed

discussion of statistical methods used to predict wave heights see Vanem (2011).

Applications for predicting wave heights via spatial correlations have also been proposed. Ho and Yim (2006) manually fitted an autoregressive time series model to monthly wave height data, producing a transfer function between observations made at two locations. By first filling in any gaps in one of the observed time series, predictions can then be made for the other even when the wave height observations are missing at both locations; thus the approach offers an improvement on a similar work undertaken by Hidalgo et al. (1995). Predicting the wave height at a location as a fixed function of a sequence of past observations at another location is likely to be insufficient in regions with spatially or temporally variable wind/wave regimes. To overcome this issue, predictions of the significant wave height at a wave buoy given the wave height at other wave buoys at a single instance in time rather than at previous time periods have been proposed. Although at the cost of being able to directly forecast ahead, this is likely to provide more accurate hindcasting when highly changeable conditions persist. ANN's were adopted by Tsai et al. (2009) to predict the wave height at a location where a wave buoy had temporarily been located on the basis of the recorded wave heights at other sites in Taichung Harbor, Taiwan. Similarly in Jain and Deo (2007) an ANN was set up to fill in gaps in the time series of two nearshore buoys deployed off the coast of west India using wave heights from another, deployed in deep water further offshore between the two nearshore buoys as input. When these completed records were incorporated into a forecasting ANN they were shown to greatly improve the learning capability of a forecasting ANN and subsequently improve the accuracy of the predictions it gave. Londhe (2008) used an ANN to predict the significant wave height at the location of a buoy given the wave heights recorded at five other buoys deployed as part of a network in the Gulf of Mexico. They also applied a Genetic Programming (GP) approach. This uses an analogue of Darwinian evolution to determine optimised model parameters in software. At each stage, a population of candidate model parameter sets may be kept, mutated, swapped or combined. The parameter sets are then selected using a cost function based on the model output and the observed output to determine which of the population either evolves further or dies out. Both methods gave acceptable results when compared to the results given by a numerical model, with the GP providing more accurate predictions of the extremes. A likely result of an ANN's high sensitivity to the training data which are unlikely to include many extreme values and thus result in poor prediction.

The approach proposed in this paper provides a similarly cheap and computationally efficient method for deriving coastal wave conditions. By offering a fully statistical approach it will reduce the sensitivity of the quality of the predictions to the observations used in the fitting of the models, a major drawback of the soft computational techniques. Yet the adoption of a copula-based approach will ensure it retains greater flexibility in the modelling as possessed by these soft computational techniques over the traditional fully statistical models. In addition, as with most of the statistical techniques it will remove the reliance of the quality of the predictions on factors such as the accuracy of the driving wind field data or how accurately the physical processes are represented as is the case in wave modelling. If applied correctly it can be used to model the wave conditions over a much larger spatial area provided the locations are subjected to similar wind/wave regimes. The paper can be broadly divided into two parts. The first part, consisting of Section 2 provides a description of the methodology. Initially the proposed approach for modelling the wave heights at the individual case study sites is put forward before a brief introduction to copulas focusing on utilising them to capture the dependence between the wave heights at different locations is given. The sequence of steps required to simulate wave heights from the fitted model is then described. The second part of the paper focuses on the application of the method to a set of sites on the south coast of the UK. Section 3 provides a description of the case study sites as well as an outline of the approach adopted for dealing with missing data which is subsequently implemented before the proposed model is fitted. The simulated wave heights are then shown and analysed in Section 4. Finally, in Section 5 appropriate conclusions are drawn on the basis of these results.

2. Model structure

2.1. Marginal distributions

For all of the applications discussed above our interest lies in the occurrences of extreme significant wave height measurements at a location. In order to ensure that distribution of the wave heights lying in the upper tails are modelled accurately an extreme value distribution will be applied to the tails. Two approaches exist for extreme value modelling: block maxima and peaks over threshold (POT). The block maxima approach consists of dividing the observation period into non-overlapping, equal length time intervals before selecting the observation with the maximum value within each of these intervals or 'blocks'. The Fisher–Tippett–Gnedenko Theorem (Fisher and Tippett, 1928; Gnedenko, 1943) states that these maxima converge to the generalised extreme value distribution (GEV) as the number of such values $\to\infty$. If X_i is the significant wave height at a particular site, i, then GEV has the cumulative distribution function:

$$G(x_i) = \exp\left(-\left[1+\xi_i\frac{(x_i-\mu_i)}{\sigma_i}\right]^{-\frac{1}{\xi_i}}\right) \tag{1}$$

for $1+\xi_i\frac{(x_i-\mu_i)}{\sigma_i} \geq 0$, where $\mu_i \in \mathbb{R}$, $\xi_i \in \mathbb{R}$ and $\sigma_i > 0$ are the location, shape and scale parameters of the GEV respectively. By dividing the data into these intervals and only using the maximum values this approach can provide, in the instances where there are several "large" values in a single block, an inefficient use of the data. This is a particular issue when dealing with a small data set where extreme values will be scarce. The POT combats this issue by modelling all of the independent observations above some carefully selected sufficiently high threshold. Picklands (1975), showed that all independent exceedances above a suitably selected threshold converge to a Generalised Pareto Distribution (GPD):

$$Pr\,(X_i < x_i | X_i > u_i) = 1-\left[1+\xi_i\frac{(x_i-u_i)}{\sigma_i}\right]^{-\frac{1}{\xi_i}} \tag{2}$$

where subscript i denotes the site, $\xi_i \in \mathbb{R}, \sigma_i > 0$ are the shape and scale parameters of the GPD and u_i is the threshold chosen to ensure stable estimates of the GPD parameters. To ensure that each of the exceedances are independent and stable parameter estimates are obtained a double threshold approach is applied (Bernardara et al., 2014). The first threshold v_i ensures that the events are independent from each other (a requirement when fitting the GPD). In this study once a significant wave height exceeding this first threshold is recorded a storm event is said to have initiated and is said to have stopped once the wave height has been below v_i for a period of 12 h. This is analogous to the criterion for independence adopted when fitting a GPD to records of significant wave height recorded off the UK coast in Hawkes et al. (2002). The peak significant wave height within each storm event or cluster is then picked out giving the sample of independent wave heights to which the GPD can be applied. The second so-called statistical threshold u_i where $u_i > v_i$ is selected so as to ensure unbiased and stable estimates of the GPD parameters. To determine a suitable value of u_i, the GPD is fitted for a range of thresholds. The resulting parameter estimates can then be plotted

against the threshold values in, so called, parameter stability plots. If the excesses above a given threshold follow a GPD then it follows that the excesses above a greater threshold will also follow a GPD. Consequently, the mean of the excesses above a suitable threshold will be a linear function of the threshold. For a thorough derivation of this result and more information on extreme value theory consult Coles (2001). In order to check for linearity in the mean excesses they are plotted against the threshold in what is known as a mean residual life plot. The lowest threshold value above which the mean residual life plot is linear, and for which the parameter stability plots demonstrate that the estimates of both the shape and scale parameters are stable, is selected as u_i.

A Bayesian approach for estimating the parameters of the Pareto distribution when modelling the significant wave height at a particular location was proposed by Sánchez-Arcilla et al. (2008). It provided a means of incorporating *a priori* information which for example could take the form of wave records from buoys located in a similar climate but for a longer time period or known physical constraints relating to the maximum possible significant wave height into the estimation procedure. To retain the simplicity of the modelling, however, the more standard maximum likelihood approach was adopted in this paper. Below the threshold the significant wave heights are modelled by their empirical distribution function. This gives rise to a semi-parametric cumulative distribution function as seen in Coles and Tawn (1991)

$$f(x) = \begin{cases} \hat{F}(x) & : x < u_i \\ 1 - (1 - F(u_i))\left[1 + \xi_i \frac{(x_i - u_i)}{\sigma_i}\right]^{\frac{1}{\xi_i}} & : x > u_i \end{cases} \quad (3)$$

where for site i, ξ_i, σ_i are the parameters of the fitted GPD and u_i is a suitably selected threshold.

2.2. Dependence

2.2.1. Introduction

A copula is a multivariate distribution function used to capture the dependence between a set of variables. In the past, flexibility in the modelling of the dependence structure between a set of random variables has been restricted by the distributions selected to model the marginal distributions. Sklar's Theorem (Sklar, 1959) states that any multivariate distribution can be fully characterised by its marginal (i.e. univariate) distribution functions and a copula. As a consequence, by adopting a copula approach the dependence structure between a set of random variables may be specified independently to that of their marginal distributions, overcoming these previous drawbacks. A comprehensive review of copulas are given in Joe (1997). Copulas have been used to derive multivariate return periods in fully parametric approaches (Corbella and Stretch, 2012; deWaal and van Gelder, 2005; Michele et al., 2007; Salvadori et al., 2014) as well as to model the dependence between the extreme values of the hydraulic loading conditions at a given location (Gouldby et al., 2014; Hawkes et al., 2002; Jonathan et al., 2010) using the semi-empirical method put forward in Heffernan and Tawn (2004). For a set of locations where the significant wave height is produced as a result of the same broad-scale wave field, although each will be subject to the different nearshore processes to differing extents, it seems reasonable to expect the wave heights at these locations to be correlated. A copula approach is therefore considered for its suitability in this paper for modelling the dependence between the wave heights at the different locations.

There are two main classes of copula: Archimedean and elliptical. Although both are capable of modelling the dependence between any number of variables, Archimedean copulas are restricted in their ability to capture the dependence between a set of more than two variables as they rely on a single parameter to capture dependence between each pair of variables. Therefore, in practice these copulas are only used as building blocks for more complex copulas such as nested (hierarchical) Archimedean or vine copulas. Nested copulas consist of Archimedean copulas connected by other Archimedean copulas of a dimension of two or higher (Savu and Trede, 2010). Vine copulas (Bedford and Cooke, 2001, 2002; Joe, 1996; Kurowicka and Cooke, 2006) capture the dependence structure between a set of variables by arranging a series of bivariate copulas in a tree structure. In both cases more flexibility in the modelling of the dependence is permitted than using standard multivariate copulas. In the trivariate setting, however, elliptical copulas, in particular the Student t-copula, have proven themselves more than capable of capturing the dependance structure between a set of random variables (Ma et al., 2013; Poulomi and Reddy, 2013; Wang et al., 2010; Wong et al., 2010) and thus will be considered before more complex arrangements of Archimedean copulas are explored.

The elliptical copulas are constituted by the normal and Student t-copula (see Appendix A). Just as the Student's t-distribution is a generalisation of the Gaussian distribution the Student t-copula is a generalisation of the Gaussian copula. In fact, the Gaussian copula is a limiting case of the t-copula as the degrees of freedom of the t-copula υ goes to infinity. The extra parameter, υ, in the Student's t-copula allows an increase in the probability density located in the tail regions compared to the Gaussian copula giving the Student's t-copula the ability to better capture the dependence between the variables in data sets with joint extreme events. In this paper both of the copulas will be fitted via maximum likelihood estimation with a goodness of fit criterion subsequently used to select between them.

2.2.2. Copula construction

Maximum likelihood estimators have a range of desirable limiting properties (Hogg and Craig, 1995) and as such have become the most common means of estimating a copula's parameters. When estimating a copula's parameters several maximum likelihood estimation methods exist. They include standard maximum likelihood estimation (MLE) where all of the (marginal and dependence) parameters

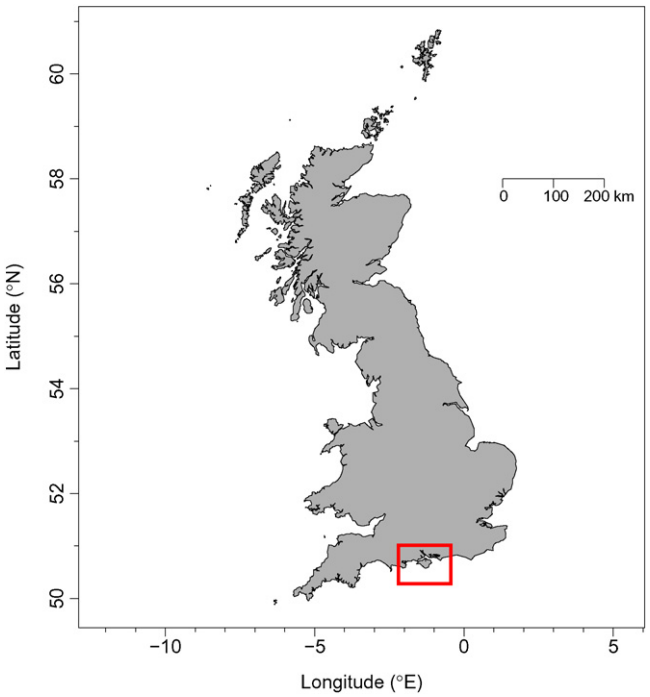
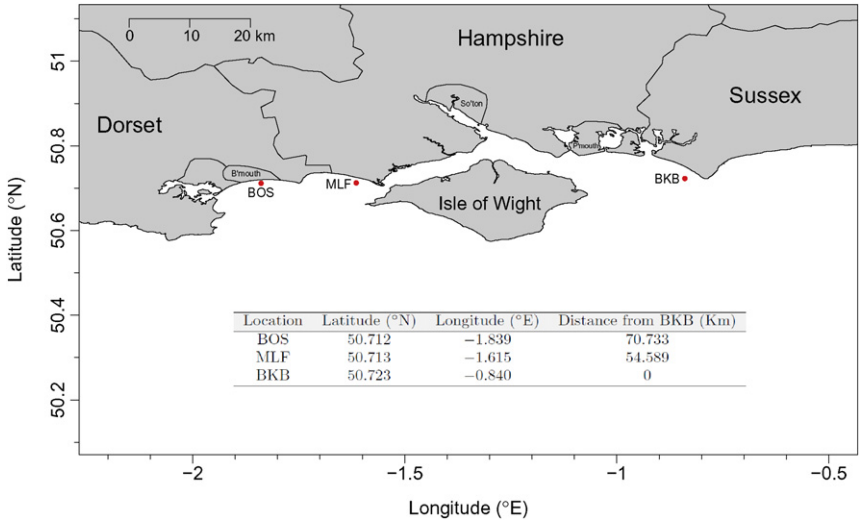


Fig. 1. Location of case study sites.



Location	Latitude (°N)	Longitude (°E)	Distance from BKB (Km)
BOS	50.712	−1.839	70.733
MLF	50.713	−1.615	54.589
BKB	50.723	−0.840	0

Fig. 2. Case study area where ● denotes the position of a wave buoy.

are estimated simultaneously. Alternatively, canonical or pseudo-maximum likelihood (PMLE) estimation (Genest et al., 1995) can be employed, in which the non-parametrically derived ranked or pseudo-observations are used to estimate the copula's parameters. The calculation of the pseudo-observations ensures that the marginal characteristic of the variables are removed and only information on

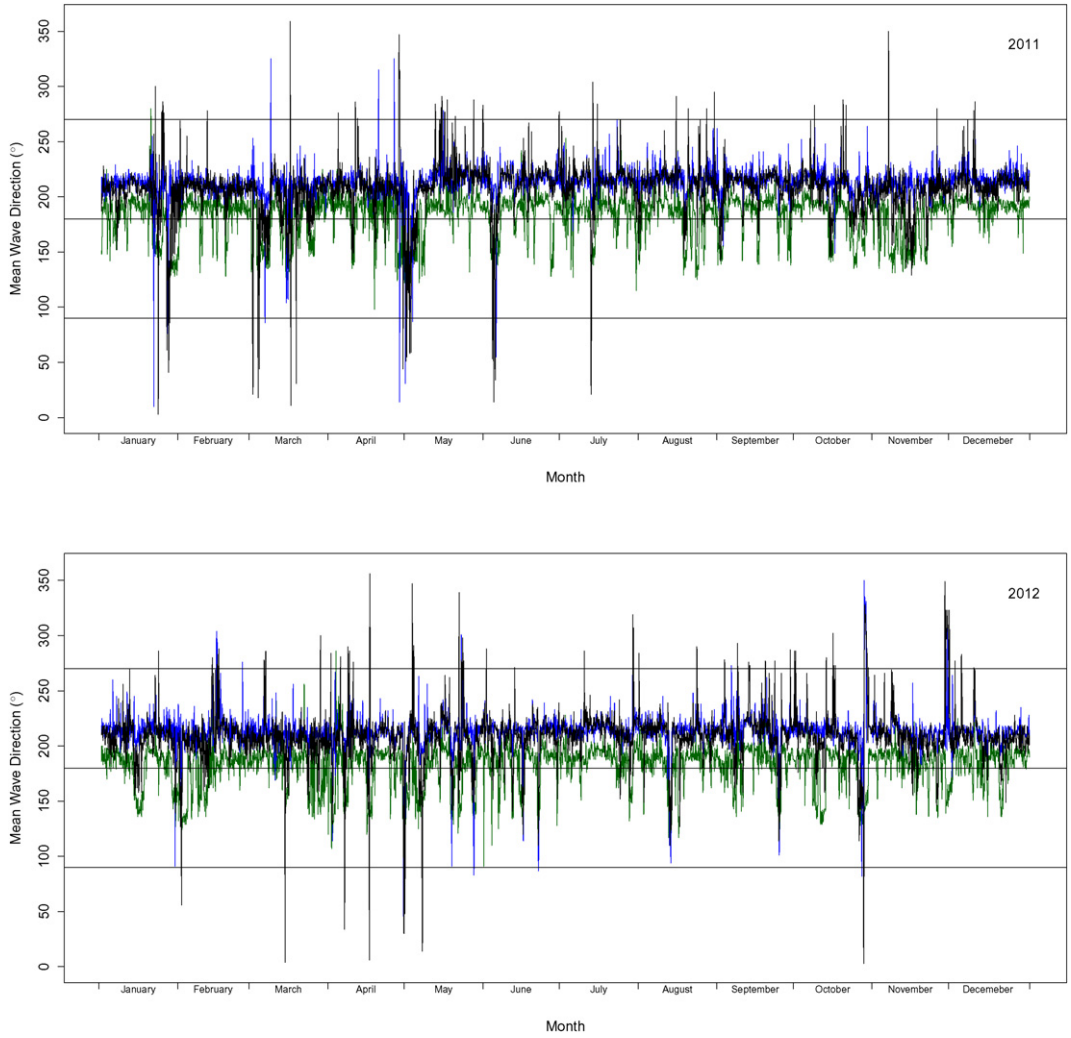


Fig. 3. Mean wave approach direction at: – BOS, – MLF & – BKB in 2011 and 2012. A similar range of conditions can be seen in the 2011 and 2012 data sets, indicating either year sufficiently captures the dominant wind wave regimes at this site.

the dependence structure remains (Genest and Favre, 2007). The inference function for margins (IMLE) method (Joe, 1997) is conceptually similar to maximum likelihood and designed to be used in situations where MLE is computationally too difficult or unfeasible. It consists of a two step procedure where the parameters of each marginal distribution are estimated separately via MLE and subsequently substituted into the likelihood function which is then maximised to obtain estimates of the dependence parameters. It has been shown by Kim et al. (2007) that PMLE is better suited for copula parameter estimation than the other two methods since, using the ranks of the observations rather than parametrically fitted marginals, its performance is unaffected by any misspecification of marginal distributions. Any copula fitted in this paper will therefore be fitted via PMLE using the *copula* package (Hofert et al., 2015; Hofert and Mächler, 2011; Kojadinovic and Yan, 2010; Yan, 2007) in R, a free software environment for statistical computing. To estimate the parameters of a copula by this approach the following likelihood function, L_U, is constructed:

$$l_U(\boldsymbol{\theta}) = log(L_U(\boldsymbol{\theta})) = \sum_{i=1}^{n} \log\left[c\left(u_{i,1}, u_{i,2}, .., u_{i,D}; \boldsymbol{\theta}\right)\right] \tag{4}$$

where the pseudo-observation $u_{i,j}$ is given by $\frac{Rank(x_{i,j})}{n+1}$ for $i \in \{1,\ldots,n\}$, $j \in \{1,\ldots,D\}$, where $Rank(x_{i,j})$ is the rank of $x_{i,j}$ among $x_{i,j}$, $i \in \{1,\ldots,n\}$, $\boldsymbol{\theta}$ is the (vector of) copula parameters and c is the density of the D-dimensional copula C given by,

$$c\left(u_1, u_2, \ldots, u_D; \boldsymbol{\theta}\right) = \frac{\partial D \;\; C\left(u_1, u_2, \ldots, u_D; \boldsymbol{\theta}\right)}{\partial u_1 \partial u_2, .., \partial u_D}. \tag{5}$$

The parameter estimates are then obtained by maximising this function

$$\boldsymbol{\theta}_{mle} = \arg\max l_U(\boldsymbol{\theta}) \tag{6}$$

using the BFGS (Broyden, 1969; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) algorithm, a quasi-Newton method for solving unconstrained non-linear optimisation problems.

2.2.3. Copula selection

The Akaike information criterion (AIC) (Akaike, 1973; Burnham and Anderson, 2002), perhaps the most widely employed of these measures, is an elegantly derived measure which assesses the goodness of fit of a given model whilst penalising complexity. It is given by

$$AIC = -2l_U(\theta) + 2p \tag{7}$$

where $l_U(\theta)$ is, again, the log-likelihood and p is the number of model parameters.

2.3. Construction of synthetic records

The first step of the simulation procedure is to transform the long term wave records to lie between 0 and 1, by invoking the probability integral transform (see Appendix B) and Eq. (3). For a given set of simultaneously observed wave heights, the copula is then conditionally sampled to obtain a set of pseudo-observations with the correct dependence structure. The pseudo-observation corresponding to the prediction site, generated from the copula, is then transformed back to the original scale, this time by invoking the inverse probability integral transform and the inverse of Eq. (3) to hence obtain an estimate of the significant wave height at the prediction site. The process is then repeated for each set of simultaneously observed wave heights in the long term wave records to produce a simulated long term record of wave heights at the prediction site. These synthetic wave height records can subsequently be used to determine the wave heights corresponding to a given return period by the conventional methods as in Ferreira and Soares (1998, 2000), Goda (1988), Soares and Scotto (2004).

3. Application

3.1. Study location & data

To demonstrate the application of the method, it will now be applied to a case study area on the south coast of England, Fig. 1. The set of sites comprises of three sites as shown in Fig. 2. Each has an associated nearshore wave buoy, maintained by the CCO and situated in approximately 10–12 m water depth (CD). The most westerly wave buoy is located off the coast of Boscombe (BOS), an eastern suburb of Bournemouth. The central wave buoy is situated off the coast of Milford-On-Sea (MLF) a small coastal village which lies approximately a mile east of Hurst Spit. The third wave buoy is located to the east of the Isle of Wight off the coast of Brackelsham Bay (BKB) a mixed sand and gravel beach located next to a seaside village of the same name.

The wave climate along the south coast of the UK is dominated by Atlantic swell and locally generated wind waves which primarily approach from a south westerly direction, the direction of the prevailing wind; the latter being responsible for most of the larger waves hitting the sites. Locally generated wind waves from the south east are also observed but are rare and generally much smaller due to the significantly reduced fetch length compared with the waves that approach from a south westerly direction. The total distance between the furthest apart wave buoys is 70.37 km, see Fig. 2. Over such a spatial extent the waves can be considered manifestations of the same system as the waves reaching each of the sites are likely to be generated under the same conditions and subsequently subjected to similar processes as they propagate shoreward. Evidence of the correlation between conditions over the range of sites is given in Fig. 3. The high correlation between the wave approach direction at the three locations is clearly visible indicating the likelihood that waves at these regionally connected sites are closely related in terms of their origin, as might be expected. Thus the strong dependence of the wave conditions at the different sites should be expected, subject to local variations in exposure and transformation over the particular bathymetries. This suggests the application of the approach put forward here is worth testing further. In the case of spatially extensive locally generated wind and wave fields, such as those that occurred along the south UK coast in early 2014 (Sibley et al., 2015), it might be argued that these correlations be extended over greater distances. This could also be true for greatly separated coasts with very similarly facing exposures dominated by swells from the same origin.

Table 1
Structure of the missing observations in the data set; 1 present and 0 not present.

BOS	MLF	BKB	Total
1	1	1	15,821
0	1	1	456
1	0	1	718
1	1	0	464
0	0	1	25
0	1	0	15
1	0	0	18
0	0	0	2
498	763	499	17,600

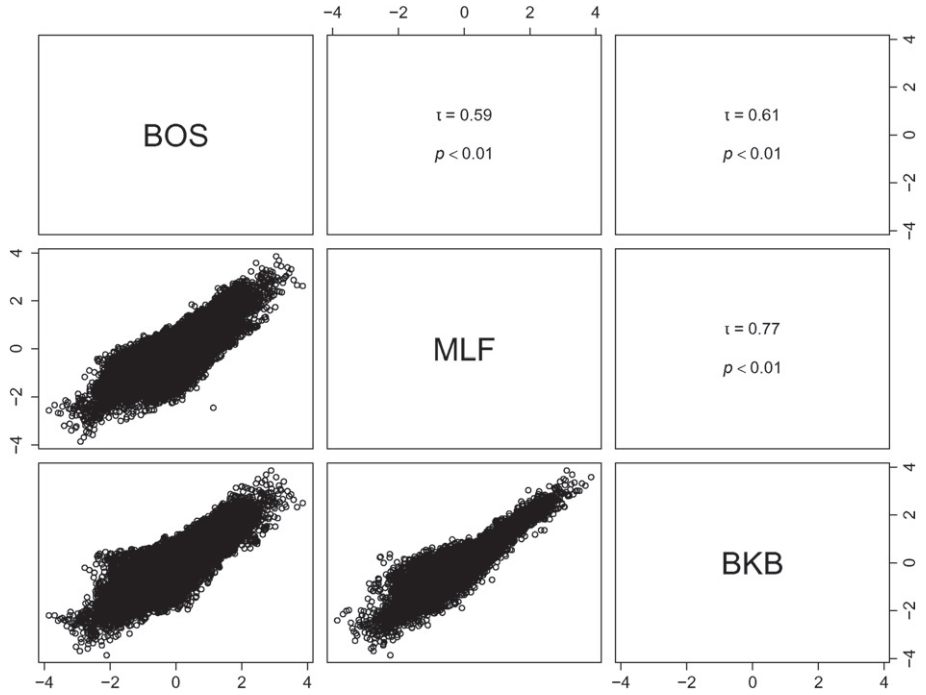


Fig. 4. Scatter plots of the normalised wave heights provide evidence of the positive correlation between the wave heights at the different locations. Each of the Kendall's τ, p-values are significant at the 0.01 level, so there is strong evidence against the null hypothesis that the wave heights at each of the sites are uncorrelated with one another.

However, where the direction of exposure of a swell dominated coastline changes abruptly it is likely that two buoys may be only weakly correlated. It may, however, be possible to predict the behavior of a single buoy exposed to a range of waves conditions if sufficient information can be encapsulated from observations using two or more buoys exposed to the full range of conditions jointly.

One of the main applications of the approach is to predict the wave height once a wave buoy that forms part of a network of buoys has been removed by exploiting the correlations between the wave heights at sites when the wave buoys were simultaneously deployed. In general the stronger the correlation between a set of wave records the more accurate the predictions that can be expected when such an approach is adopted. In terms of optimising the deployment of wave buoys, if a selection is possible it therefore appears logical to remove the wave buoy with the strongest correlation with each of the other buoys and continue deployment of the set of wave buoys that are least correlated with each other.

3.2. Missing observations

In order to use the method, concurrent records of the wave height at three locations over a given period of time are required. Records of the significant wave height and a host of other sea state characteristics including the direction of wave approach, each recorded at half hourly intervals are available from wave buoys maintained by the CCO at each of the locations in 2012. However, the records of the significant wave heights contain missing observations see Table 1. Reasons for missing observations typically include the wave buoy sustaining damage which prevents it from measuring and/or recording the significant wave height or it being removed from its normal location for a period of maintenance.

A test for whether the observations are Missing Completely At Random (MCAR) (Little, 1988) gave a significant result and so the possibility that there is a discernible pattern to the missing observations cannot be ruled out. Simply leaving out missing observations of this nature has the potential to introduce bias into the analysis, so the missing values will therefore have to be estimated. Imputation is the process of replacing missing data with plausible values. Multiple imputation is an imputation technique first formally put forward in Rubin (1987) for estimating each of the missing values whilst accounting for the uncertainty associated with each of the estimates. Fully conditional specification (FCS) (van Buuren, 2007) is a form of multiple imputation whereby initially the missing observations of a variable are replaced by the respective variable's mean value. An iterative procedure is then implemented where the missing values are predicted by regressing each variable on the observed and most recently imputated values of the remaining variables. The process is terminated once the parameters in each of the regression models are stable. The whole procedure is then repeated to obtain a set of imputed data sets. By adopting a conditional approach to the modelling it removes the assumptions regarding the form of the multivariate distribution of the variables that are required for other multiple imputation approaches such as joint modelling. In this study each of the imputed data sets were found to be very similar and so one was selected at random for use in the analysis.

Kendalls τ coefficient $[-1, 1]$ is a non-parametric measure of the degree of association between two variables, where 1 indicates a perfect positive linear association between the ranks of the two variables and -1 a perfect negative linear association (see Appendix C). The values of the Kendalls τ statistic between each pair of completed and subsequently ranked wave records are given in the upper panels of the scatter plots, Fig. 4. The plots show the positive correlations between the wave heights at the three locations. It is clear that there is particularly strong positive correlation between the wave heights

Table 2
Parameter estimates and goodness of fit statistics of the fitted copulas.

$f(x)$	Parameters	AIC
Normal	$\rho_{1,2} = 0.801, \rho_{1,3} = 0.822, \rho_{2,3} = 0.919$	−53,070
Student's t	$\rho_{1,2} = 0.798, \rho_{1,3} = 0.815, \rho_{2,3} = 0.927, \upsilon = 7.266$	−54,308

Table 3
Key statistics for assessing a model's ability to accurately predict wave heights are presented for the WAM, WAM/SWAN & copula models. Max. resolution refers to the highest resolution grid used within a given model and n is the total number of wave heights predicted by the model. The statistics are defined in the text.

Location	Max. resolution	n	Bmed	Smed	Bias	RMSE	SI	r
Weather Station Boyle (UK)	0.5° × 0.5°	7158	2.149	2.569	0.421	0.815	0.379	0.884
Cabo de Peas (Spain)	0.25° × 0.25°	4464	1.911	1.914	0.003	0.499	0.261	0.899
Sines (Portugal)	0.25° × 0.25°	31,646	1.688	1.783	0.095	0.456	0.270	0.905
Sines (Portugal)	0.005° × 0.005°	1515	2.397	2.591	0.194	0.502	0.209	0.934
Peniche (Portugal)	0.01° × 0.01°	135	2.120	2.403	0.283	0.608	0.287	0.908
Brackelsham Bay (UK)	–	17,519	0.769	0.766	0.003	0.282	0.366	0.857

at Milford and Bracklesham Bay. Given the relative proximity of Boscombe and Milford compared with Brakelsham Bay and the other two sites this is perhaps a surprising result. It is likely to be due to the fact that the Boscombe site is afforded protection from waves that approach from the south west by the headland located to the west of the site. The headland reduces the fetch lengths of waves approaching from this direction leading to less extreme wave heights being experienced at Boscombe relative to the other two locations. Finally,

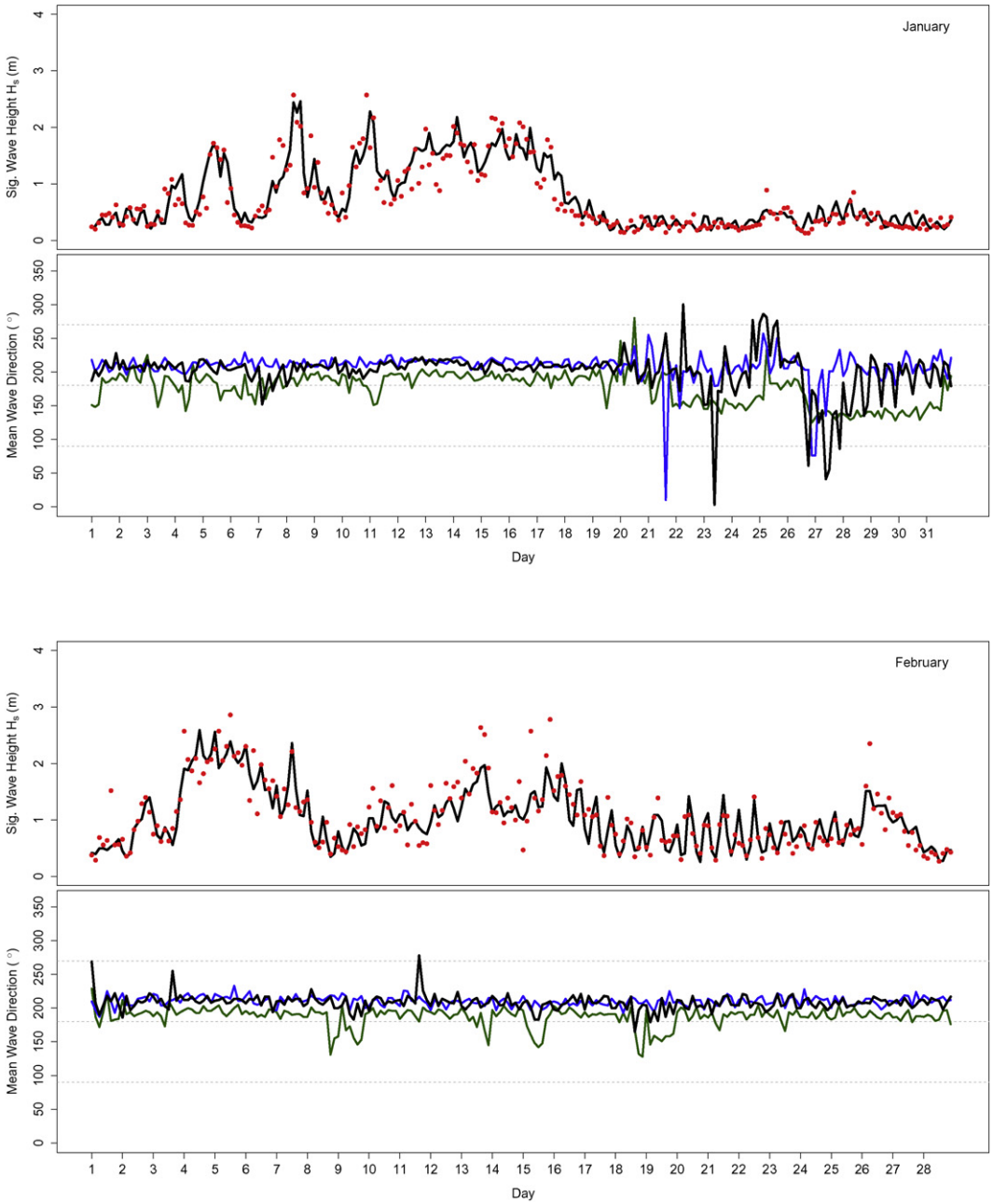


Fig. 5. Significant wave height at 3 hourly intervals (January–February 2011): – Observed vs ● Predicted. With associated mean wave direction at: – BOS, – MLF & – BKB.

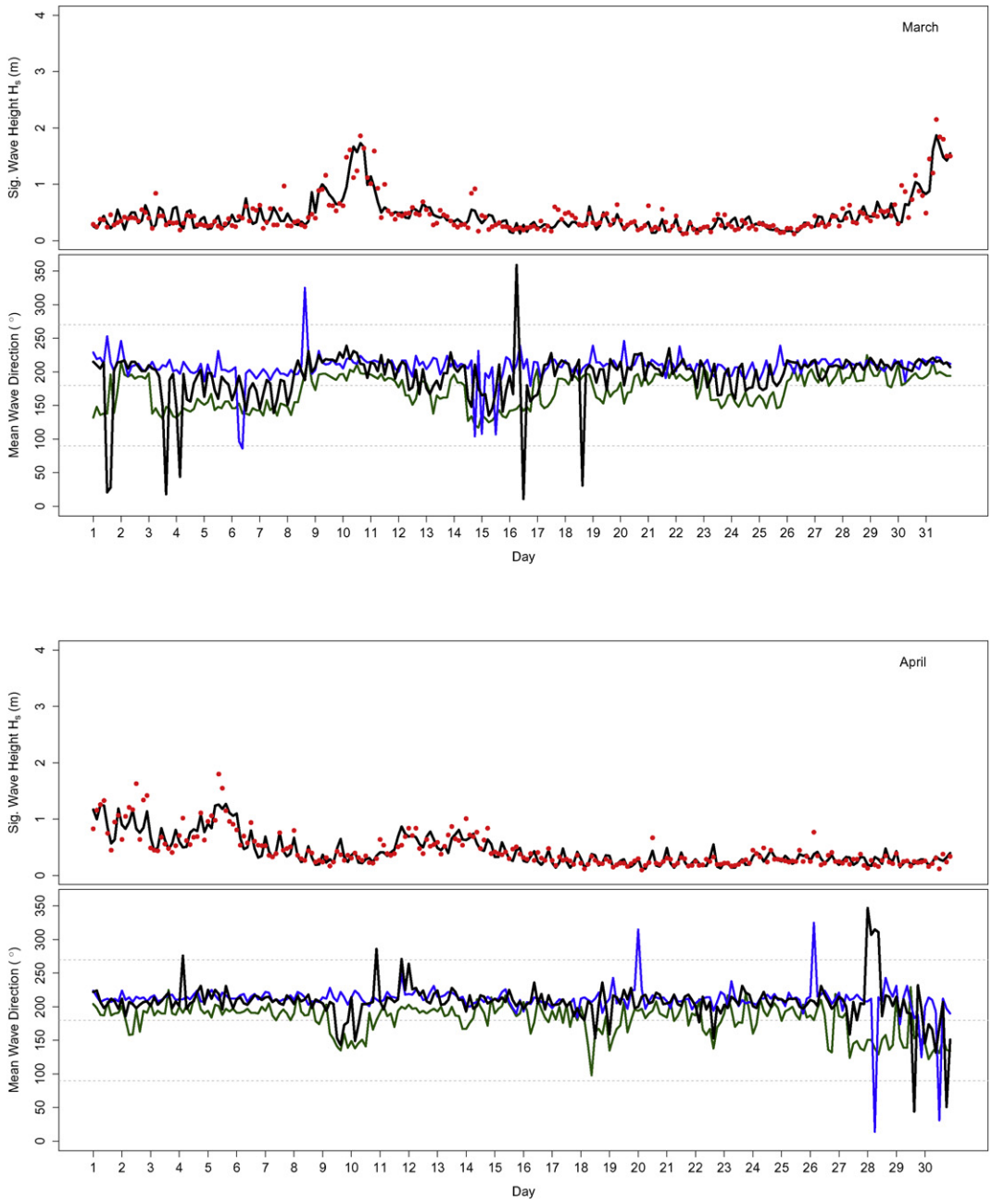


Fig. 6. Significant wave height at 3 hourly intervals (March–April 2011): – Observed vs ● Predicted. With associated mean wave direction at: – BOS, – MLF & – BKB.

the statistically significant p-values indicate definite concordance and thus dependence between the wave heights.

The parameters and goodness-of-fit criterion for each of the fitted copulas are given in Table 2. The lower AIC suggests that the Student's t-copula offers the overall better fit. This is a likely result of the differences in the tail dependencies between the copulas (Demarta and McNeil, 2005). The Student's t-copula with its non-zero tail dependence assumes asymptotic dependence between the wave heights which in doing so gives non-negligible probabilities of joint extreme events. This is in contrast to the normal copula which assumes that in the tails the wave heights are asymptotically independent of one another, meaning that the most extreme wave heights occur more or less independently which from the figure is clearly not the case. The limitations of the normal copula for modelling data with tail dependence and the superiority of using the t-copula in such situations are shown in Breymann et al. (2003), Mashal and Zeevi (2002).

The fitted model will be used to predict the significant wave height at Braklesham Bay for the previous year, 2011, given the significant wave heights at Boscombe and Milford. The wave height is predicted at the site most correlated with the other two sites. The approach is therefore applied in line with how it would be adopted in practice if it were to be applied to a network of wave buoys, as discussed in Section 3.1. The wave heights at all three sites are known throughout 2011 thus the application will provide an opportunity to assess the quality of the model. The annual wave climate observed

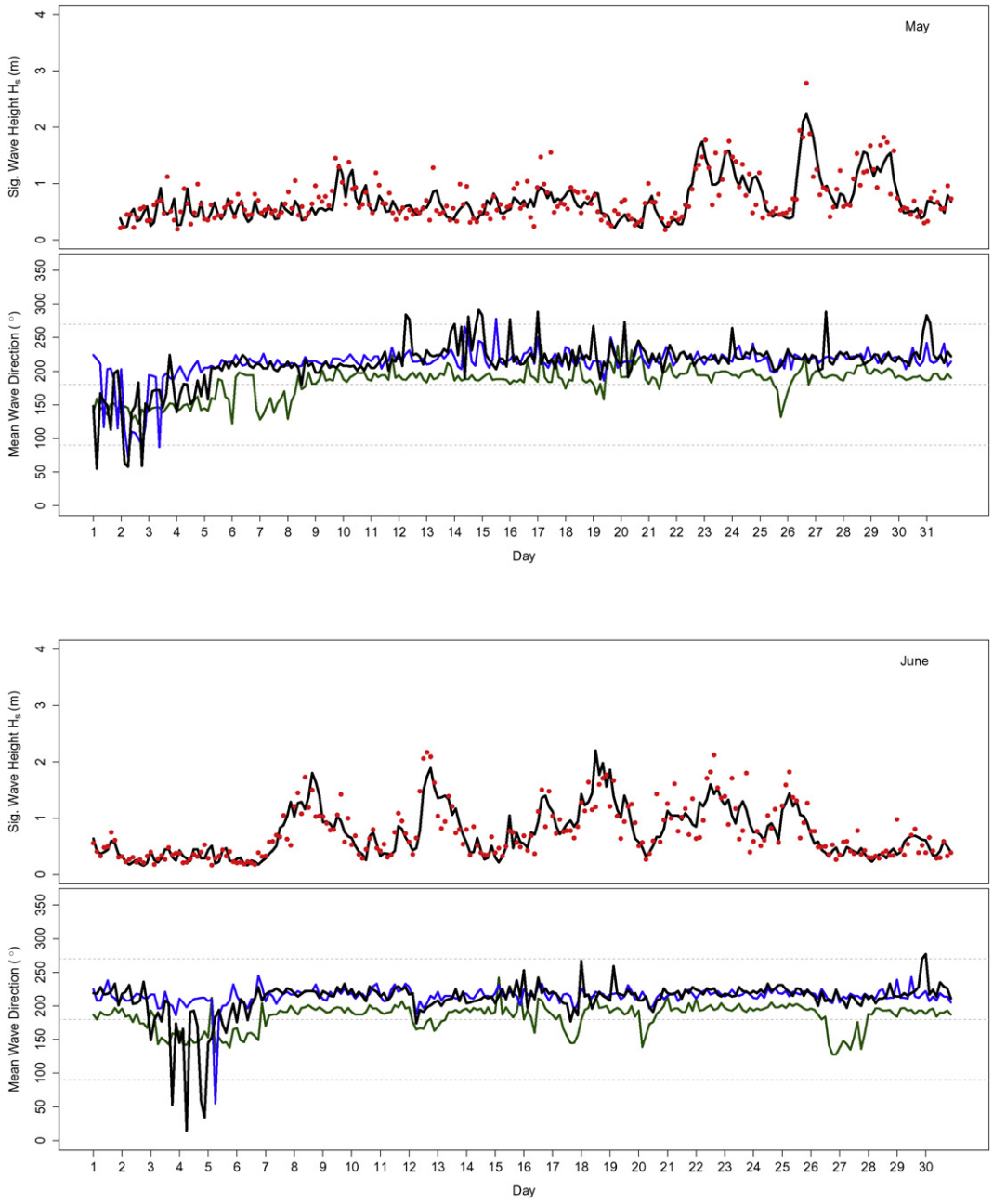


Fig. 7. Significant wave height at 3 hourly intervals (May–June 2011): – Observed vs ● Predicted. With associated mean wave direction at: – BOS,– MLF & – BKB.

during 2011 was found to be consistent with that observed during other years in the case study area. Consequently it was deemed suitable as a test data set for assessing the model.

4. Validation

4.1. Comparison of results with existing wave prediction models

For the approach to offer a viable alternative to the existing methods for predicting wave heights, confidence in the accuracy of the results is essential. The accuracy of the model's output can be assessed for a given location by simulating wave heights for a period of time where the wave heights have been observed, and comparing the two respective time series. The times series can be compared first by considering key summary statistics such as the sample mean as well as by statistical measures such as the bias, root mean squared error (RMSE), scatter index (SI) and Pearson's Correlation Coefficient (r). This combination of statistics provides a comprehensive assessment of the quality of the model and as such have become the standard criterion for assessing the quality of wave prediction models. If X_i is an observed wave height, Y_i is a predicted wave height and n is the number of observations, then the aforementioned statistics can be defined as follows:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \quad \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \quad Bias = \frac{\sum_{i=1}^{n} (X_i - Y_i)}{n}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_i - Y_i)^2}{n}} \quad SI = \frac{RMSE}{\bar{X}} \quad r = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n} \left(Xi - \bar{X}\right)^2 \sum_{i=1}^{n} \left(Yi - \bar{Y}\right)^2}}$$

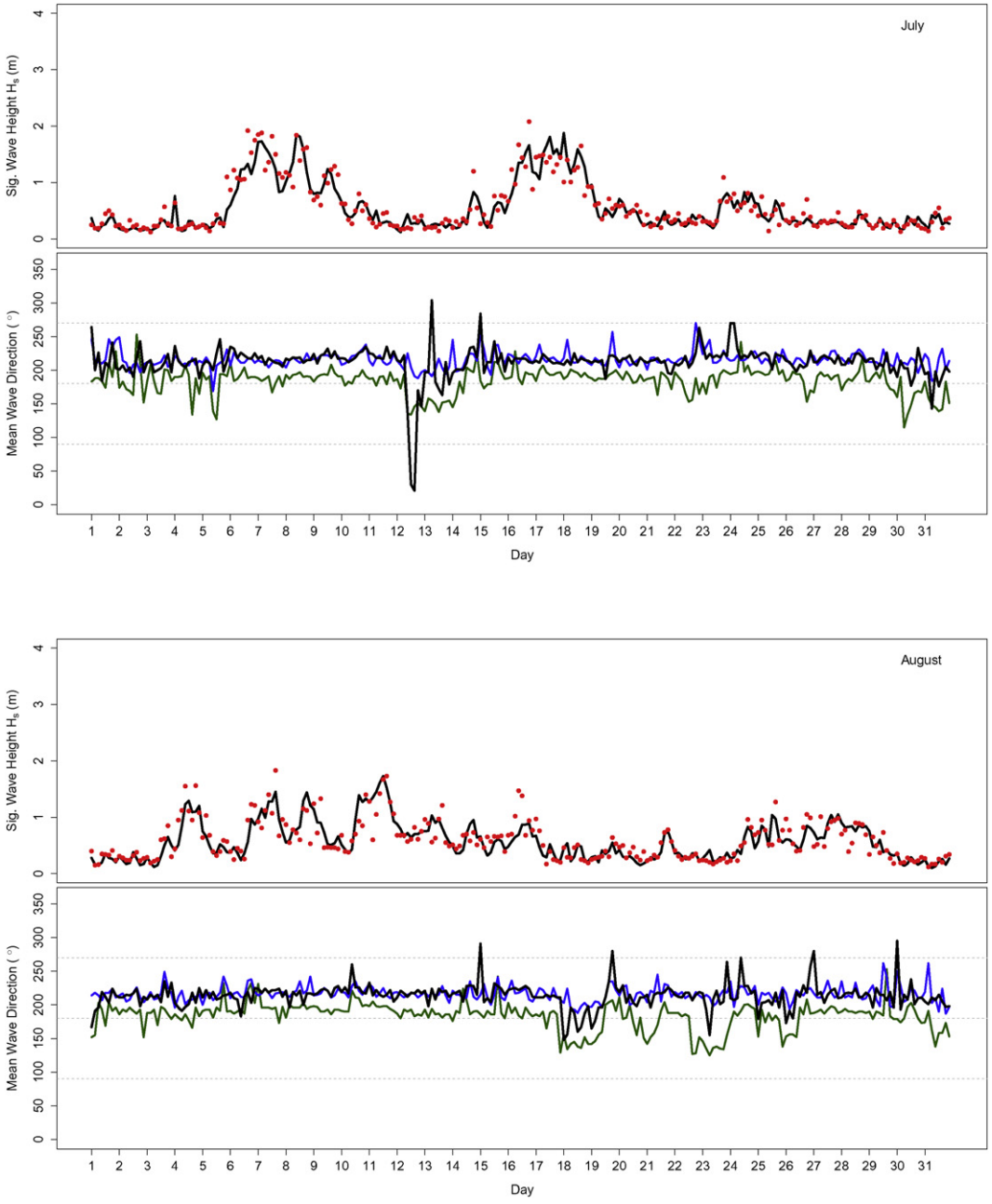


Fig. 8. Significant wave height at 3 hourly intervals (July–August 2011): – Observed vs ● Predicted. With associated mean wave direction at: – BOS, – MLF & – BKB.

The values of these statistical parameters when using the approach proposed in this paper to predict the significant wave height at BKB for the whole of 2011 are given in the final row of Table 3. The table also contains the key statistics from a selection of wave prediction models. Each of these models use initial wind input conditions with a spatial resolution of 0.5° and a time step of 1 h. The first three results were achieved by Rusu et al. (2008) using a WAM model consisting of six nested grids. The latter two were produced by Pilar et al. (2008) with a model comprising of a nested SWAN model coupled with a two way nested WAM model. The selected simulations represent a range of resolutions. Simulations with a minimum grid spacing of 0.5° × 0.5° or coarser are considered low resolution, those with a grid spacings of 0.01° × 0.01° or finer are considered high resolution whilst those with grid spaces between these values considered medium resolution. Since each of the sites where the SWAN modelling has been carried out will have their own unique characteristics and intricacies, the difficulty in predicting the wave heights to a specified degree of accuracy will vary between the sites. The accuracy of the results from each of these simulations are therefore not directly comparable, however, these statistics will make it possible to gauge whether the new approach provides an acceptable level of accuracy.

After a short perusal of Table 3 it is apparent that the statistics for the model described in this paper are comparable to those obtained from the existing models. Most striking is that there is almost no bias in the results given by the newly fitted model, indicating that there is very little systematic error in the prediction of the significant wave height. This is likely due to the fact that the model is derived directly from wave data rather than from a physical process based model comprising representations of a number of interacting sub-processes which, due to their very nature, can introduce systematic error. For instance, in SWAN, factors such as a steep

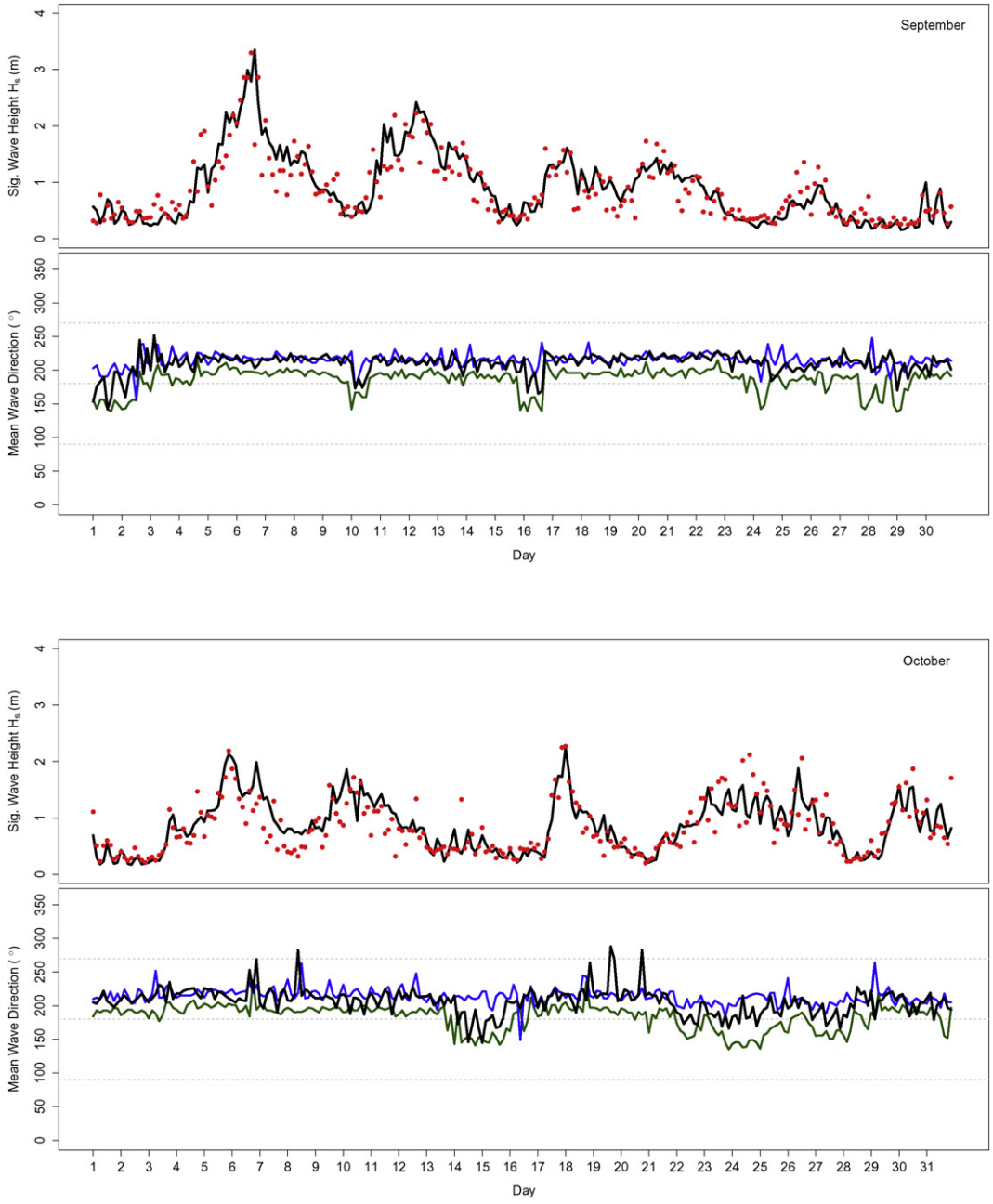


Fig. 9. Significant wave height at 3 hourly intervals (September–October 2011): – Observed vs ● Predicted. With associated mean wave direction at: – BOS, – MLF & – BKB.

wind gradient, poorly described orography and complex bathymetry have been shown to induce systematic error into significant wave height predictions (Sánchez-Arcilla et al., 2008) particularly in semi enclosed basins where they have been shown to contribute to over prediction of the smallest wave heights and under prediction of the peak wave heights (Bolaños et al., 2007; Rusu et al., 2008). This indicates an advantage of working directly with observed wave heights rather than a model which takes a component form such as SWAN where there are many possible routes of introducing systematic error into the predictions.

The scatter index is arguably the most valid measure when comparing predictions made at different locations. It is the RMSE normalised by the mean observed value. It is therefore a dimensionless measure which gives the percentage of RMS difference with respect to mean observation. The normalisation means that the results from the different simulations are compared on the same terms, independent of scale which is important in this situation since locations with smaller average wave heights will produce relatively smaller dimensional goodness of fit measures compared with those locations where the wave heights are generally larger. The scatter index from the fitted model is 34.4% which is between those values observed from the medium and high resolution process based models. The correlation between the observed and predicted wave heights is close to 1 indicating a high degree of linear correlation between the values which is comparable with those from the higher resolution models.

The time series of observed and predicted wave heights along with the wave approach direction are given in Figs. 5– 10 whilst the observed and predicted waves are also shown in the form of a scatter plot in Fig. 11. The plots demonstrate that in general the copula model predicts the wave heights to a good degree of accuracy. The lack of bias in the predictions is also clearly evident from

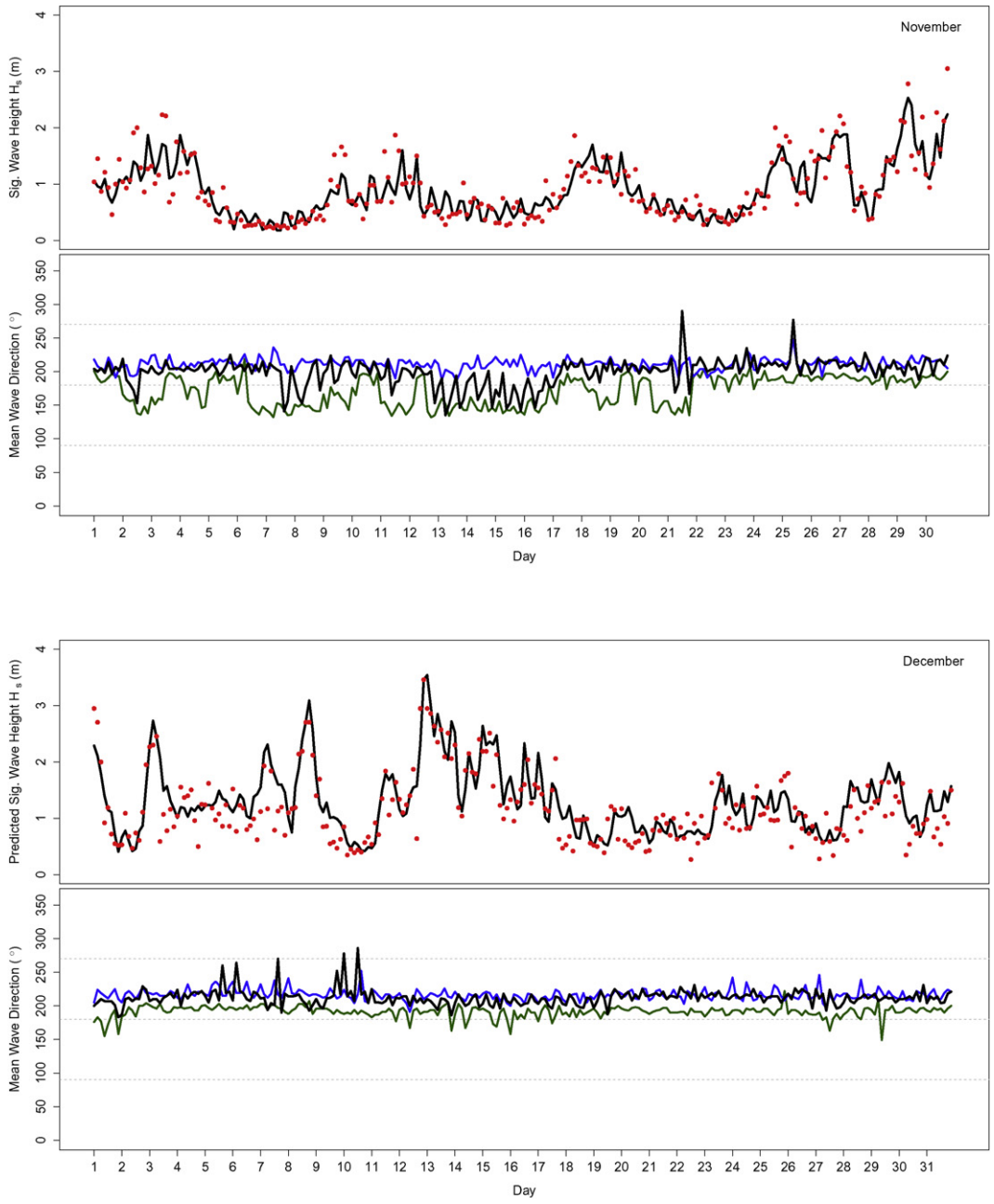


Fig. 10. Significant wave height at 3 hourly intervals (November–December 2011): – Observed vs ● Predicted. With associated mean wave direction at: – BOS,– MLF & – BKB.

the scatter plot where approximately an equal number of points lie above and below the diagonal. By comparing the time series plots containing the observations and predictions with the wave direction plots beneath them it is possible to gauge the sensitivity of the predictions to the wave approach direction. The wave direction plots demonstrate that for the majority of the prediction period the waves approach MLF and BKB from the south west whilst they approach BOS from a more southerly direction. In such instances it is likely that the waves initially approached all of the sites from the south west but due to the headland to the west of the BOS site the waves undergo a degree of diffraction and consequently approach the BOS site from a more southerly direction. The largest error was recorded on the 6th of September, it is circled on the scatter plot, Fig. 11. On this occasion the waves approached each of the locations from the south west, the dominant wave direction, which would imply that the quality of the predictions given by the model are insensitive to wave direction. Another of the largest errors occurred on the first day of February. On this occasion the waves approach the BOS site from the south east and the other two from the south west — a rare combination of directions. As opposed to the previous error, this error indicates the possibility that a masking effect has occurred as a result of grouping all the wave heights together when fitting the copula which can lead to the clouding out of the relationship between the wave heights at the different sites when these conditions prevail. It would therefore seem a logical next step to incorporate wave direction into the model.

In order to carry out a more detailed analysis of the influence of the wave direction on the quality of the predictions, the wave approach direction was discretised into sectors of 60°, comparable with the approach used when analysing the wave climate around the

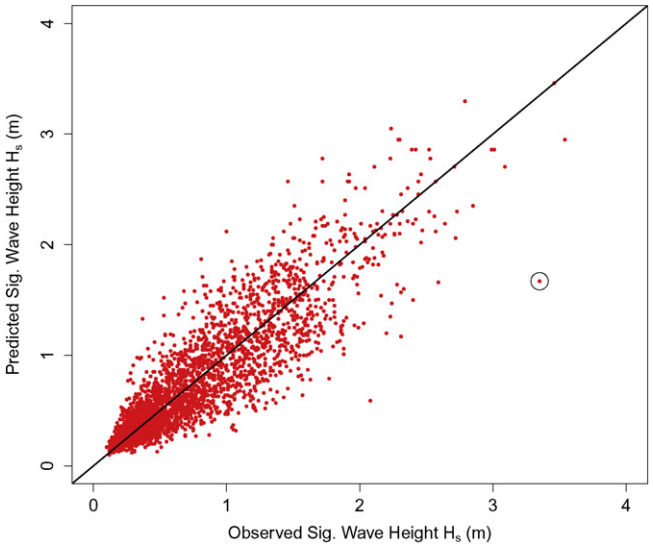


Fig. 11. Scatter plot of significant wave heights.

Table 4
Wave direction combinations observed at the three sites during the prediction period and the average, denoted by the overbar, RMSE of the prediction when they persisted.

BOS	MLF	BKB	Number of observations	$\overline{RMSE}$
↑	↗	↗	1196	0.200
↑	↗	↑	593	0.191
↑	↑	↑	388	0.180
↑	↑	↗	227	0.153
↖	↑	↑	118	0.135
↖	↗	↑	100	0.158
↗	↗	↗	46	0.194
↑	↗	↙	39	0.197
↖	↗	↗	38	0.120
↗	↗	↑	26	0.190
↖	↑	↖	25	0.166
	Other		124	0.118

UK coast in Environment Agency (2011). The average RMSE for each combination of typical wave approach directions are given in Table 4. The largest RMSE (0.20) is associated with the most persistent combination of directions. In general, the errors are of a very similar magnitude regardless of their relative frequency of occurrence, indicating that the quality of the predictions are insensitive to wind direction. These results dampen the argument for incorporation of a directional covariate into the model and remind us of the remarkable agreement between the predicted and observed wave heights particularly considering the limited amount of the information and effort required to fit the model.

In the majority of the applications of the proposed model, most interest lies in the extreme values. They are also the values that are most scarcely observed and consequently the most challenging to predict. It is thus essential when assessing the merits of the technique that its ability to predict these values is examined. A statistical analysis of the recorded wave heights from CCO wave buoys deployed along the south coast of the UK was undertaken by Bradbury and Mason (2014) as part of their investigations into the responses of beaches during the storm events of the winter of 2013–2014. For the buoy at BKB they found that there had been seven events with a return period >1 in 1 year since its deployment in 2008. For each of these events, multiple predictions rather than a single estimate are generated from the proposed model in order to gauge an improved understanding of its ability to predict these values. To assess the bias and error associated with the predictions slightly redefined bias and RMSE statistics are required. If X is an observed wave height, Y_i is a predicted wave height and n is the number of predictions of the observation simulated from the model, then the bias and RMSE for a given extreme observation can be re-defined as follows:

$$Bias = \frac{\sum_{i=1}^{n}(X - Y_i)}{n} \qquad RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X - Y_i)^2}{n}}$$

The values of these statistical parameters when the approach proposed was used to predict the significant wave height for the observed extreme wave heights at BKB are given in the Table 5. Firstly, it is interesting to note that in each extreme event the waves approach all three sites from very similar directions. This indicates the existence of a corridor for storms that lead to the most extreme conditions on this coast. For the majority of the time period under analysis it is most probable that waves initially approach the case study area from a south westerly direction. The waves on final approach to BOS will have strongly refracted and diffracted, whilst the south westerly direction of approach is maintained at MLF and BKB. This is consistent with the historical observations that the storms impinging the south coast of the UK are a product of the low pressure systems that develop over the Atlantic Ocean (Zong and Tooley, 2003). These track eastwards in the process generating long period swell before they reach the UK where the high wind speeds also result in large locally generated wind waves on their final approach. Depending on the velocity of the low pressure center, the swell and locally generated wind waves do not necessarily reach the coast simultaneously. Although there appears to be a small bias in the predictions, the RMSE is of a similar magnitude to that when predicting the wave heights in energetic conditions using SWAN, see for example Table 5 in Rusu et al. (2008). This result is remarkable given that each of these values are greater than any value to which the copulas were fitted. In this area the extreme conditions approach from the same direction as the dominant wave conditions.

Although the dependence of the wave fields is likely to increase during strong, extreme events, the application of the copula over the entire range observations at a site apparently helps to strengthen the representation of the extremes also. Caution would therefore be required when applying the approach in areas where the extreme conditions have a direction of approach which is different from the normal conditions, although from Table 4 the importance of the

Table 5
Events with a return period $>$ 1 in 1 year observed at BKB since 2008. The wave heights at each of the sites are given with the wave directions (in °'s) provided in the brackets along with the average, denoted by the overbar, predicted wave height at BKB and the relevant model fit assessment statistics.

Date	Time	Return period	BOS	MLF	BKB	$\overline{BKB}$	Bias	RMSE
23rd November 2009	13 : 00	$>$ 1 in 1 year	1.590(190)	2.970(215)	3.83(203)	3.215	0.615	0.651
28th October 2013	04 : 30	$>$ 1 in 2 years	2.620(188)	3.480(219)	4.03(207)	3.693	0.337	0.453
24th December 2013	02 : 00	$>$ 1 in 3 years	2.830(176)	3.370(218)	4.13(208)	3.673	0.457	0.522
3rd January 2014	23 : 30	$>$ 1 in 1 year	2.030(186)	3.610(211)	3.89(207)	3.740	0.150	0.369
5th February 2014	13 : 30	$>$ 1 in 3 years	2.990(177)	3.670(208)	4.07(198)	3.747	0.323	0.463
8th February 2014	16 : 30	$>$ 1 in 1 year	2.520(186)	3.550(205)	3.80(218)	3.671	0.129	0.445
15th February 2014	00 : 00	$>$ 1 in 20 years	3.360(193)	4.250(217)	4.47(208)	4.024	0.446	0.476

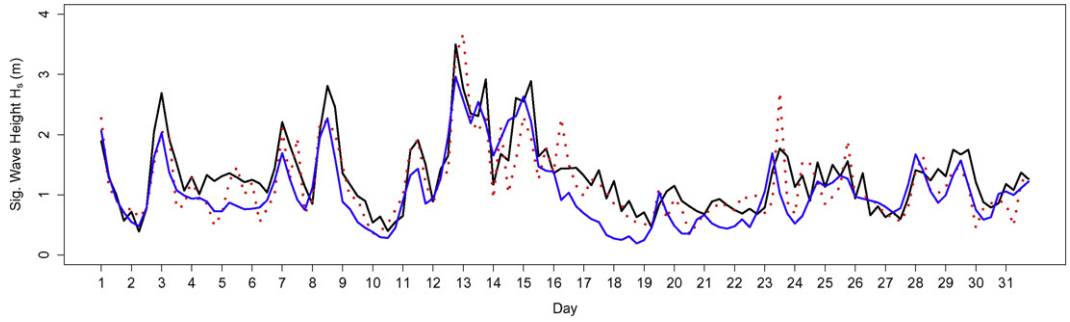


Fig. 12. Significant wave height at 6h intervals (December 2011): – Observed vs Predicted by – SWAN and – – the copula model.

wave direction on the quality of the predictions is questionable. Even though the model has been shown to predict wave heights outside the bounds of the input data set to a good degree of accuracy, caution would also have to be exercised when predicting waves heights more extreme than those recorded at the site to date.

4.2. Direct comparison with a SWAN model

The final part of the validation will compare the accuracy of predictions achieved by the proposed model with those obtained from a high resolution SWAN model. The SWAN model was run in third generation two dimensional stationary mode, the spectral space was resolved in 36 directions and 24 logarithmically spaced frequencies between 0.05 Hz and 1 Hz. The Komen et al. (1984) formulation was used for the wind growth parameterisation, a key parameter in determining the prevalence of whitecapping which in SWAN is characterised by the pulse-based model of Hasselmann (1974). For depth induced wave breaking, the Battjes and Janssen (1978) model is applied, the lumped triad approximation (Eldeberky and Battjes, 1996) is adopted to parameterise the triad–triad non-linear interactions and diffraction accounted for using the phase-decoupled refraction–diffraction approximation as described in Holthuijsen et al. (2003). The latter is a significant consideration in the determination of wave heights at this particular location due to the presence of the Isle of Wight.

The inputs consist of wind and offshore wave data from the European Center for Medium Range Weather Forecasts (ECMWF) and information on the water level from the British Oceanographic Data Center (BODC). The ECMWF is a meteorological data assimilation project, where observed data and forecasts are combined to give some 'best'prediction of the atmospheric condition at a specific time. Numerical weather data from the ECWMF's ERA-interim model (Dee et al., 2011), a global atmospheric reanalysis coupled with a wave model, at an offshore point (50.25°N, −0.25°E) provide the input for the SWAN model. Information on the sea level (astronomical tide + surge) is taken from the tide gauge located at Portsmouth (50.802°N, −1.112°E). The model comprised of local bathymetric data with a longitudinal spacing of approx. 300 m and a latitudinal spacing of approx. 450 m. The computation grid set-up in the model consisted of (i) an offshore grid covering an area of approx. 75 km × 60 km with a grid size of 150 m × 150 m, comprising from the offshore point where the input data are located and (ii) a high-resolution nearshore (nested) grid covering the location of the wave buoy, with dimensions of approx. 12 km × 12 km and a grid size of 25 m × 25 m, comprising from 50.620°N, −0.917°E. Each of the grids was placed at an angle of 20° to the horizontal to give the best possible coverage of the area of interest. The model was run with open boundaries to the north, south and west with a land boundary covering the east of the computational domain. It was used to predict the significant wave height at the Bracklesham Bay site at 6 hourly intervals. It is first calibrated using wave heights recorded between 1/1/2011 and 31/1/2011, a period containing a wide range of wave heights. The wave heights are then predicted by both models for December 2011, the month of the year that contained the highest average wave height and was also the month where the significant wave height most frequently exceeded 3 m.

The wave heights simulated from the two models are plotted along with the observed wave heights in Fig. 12. The quality of the models is summarised in terms of key goodness-of-fit statistics in Table 6. Although the general pattern of the observed wave heights is captured by the SWAN model it is clear from the graph that there is a consistent underestimation of the wave height throughout the prediction period, which is also alluded to by the large bias associated with the model. The systematic under prediction of wave heights in SWAN and WAM models due to under prediction of wind speed by the ECMWF is well documented (Brenner et al., 2007; Cavaleri and Sclavo, 2006; Vicinanza et al., 2013) and a likely cause of the systematic under prediction experienced by the SWAN model in this paper. In addition, when run on a quad core laptop the copula-based model took an average of just 89.64 s to estimate a significant wave height compared with the 249 s taken by the SWAN model run on the same device. Considering all of the statistics it can be reasonably argued that the newly proposed approach offers an improvement on the SWAN model for predicting significant wave heights at a reduced computational cost. Once further improvements are made to the proposed model such as incorporating covariates such as wave period into the model it is envisaged that significant further improvements in the predictions can be obtained.

5. Conclusion

Knowledge of the significant wave height is essential in the design of coastal structures. Physically collecting such data through long term deployment of wave buoys is an expensive process. To overcome this issue numerical hindcasting and wave transformation models have been developed for predicting the wave climate at a

Table 6
The key statistics for assessing a model's ability to accurately predict wave heights at BKB are presented for the high resolution SWAN & copula model. Max. resolution refers to the highest resolution grid used within a given model and n is the total number of wave heights predicted by the model. The statistics are defined in the text.

Model	Max. resolution	n	Bmed	Smed	Bias	RMSE	SI	r
SWAN	$< 0.01° \times 0.01°$	124	1.305	1.050	0.255	0.167	0.128	0.851
Copula	–	124	1.305	1.172	0.133	0.134	0.102	0.836

coastal location. However, these numerical approaches also have their limitations. These include: computational expense; sensitivity to input conditions that are often difficult to measure accurately in coastal locations; and neglect of key physical processes. As a result alternative methods for predicting significant wave height have been sought. In this paper, a method for estimating the significant wave height at a coastal location based upon spatial correlations has been put forward. In order to apply the model a limited record of the significant wave heights at the site of interest as well as simultaneously observed wave records at nearby locations are required. This requires less information and input data than that required to run and calibrate a numerical model. Furthermore, since wave buoys are often deployed for short periods at locations where the wave climate is of interest and networks of wave buoys are now commonplace along coasts under threat of flooding, the proposed method is likely to find many applications. It was shown to be able to predict the wave conditions at a particular location to a good approximation giving a similar level of performance to a medium or high resolution SWAN simulation.

By virtue of its accuracy and low computation burden it is envisaged that the method could be used, in combination with existing buoy networks, to hindcast longer term wave records at locations, after the short term deployment of a wave buoy, in place of numerical modelling. These synthetic records might then provide a means of deriving design wave conditions with greater confidence. Although a very useful application in itself only once the extensive networks of wave buoys in place along the world's coastlines are considered does its even greater potential value/significance become evident. By combining strategic placement of the buoys with the methodology proposed in the paper, information on the wave climate along an entire stretch of coastline could be derived at much greater resolution without the need to deploy further buoys. Consequently, this would alleviate the need for more extensive buoy networks, and in some instances provide an alternative to physics based numerical modelling.

Although the model provides sufficiently accurate predictions there are a number of ways in which the approach might be improved. One such improvement would be to go beyond the standard Gaussian and Student's t-copula and look at more advanced ways of modelling the dependence structure such as nested Archimedean or vine copulas. These approaches are well suited to capturing dependence structures between many variables which in turn will lead to more accurate predictions. The wave buoy measurements also contained information on sea state characteristics other than the wave height such as the mean wave direction and so an interesting improvement would be to incorporate these covariates into the model, possibly through a conditional copula (Acar et al., 2011; Patton, 2006), and to discover the effect this has on the accuracy of the predictions. The model might also be improved to include prediction of the wave period and hence provide a means of deriving multivariate design conditions.

Acknowledgments

The authors would like to thank the Channel Coastal Observatory (CCO) for their freely available wave buoy data as well as the European Centre for Medium-Range Weather Forecasts (ECMWF) and British Oceanographic Data Center (BODC) for providing the wind and water level data respectively. The authors would also like to thank the SWAN group at Delft University of Technology for providing the freely available wave propagation model. Lastly, they wish to express their gratitude to HR Wallingford and the Engineering and Physical Sciences Research Council (EPSRC) for their funding contributions that have made the work possible.

Appendix A. Copulas

A copula (C) is a multivariate probability distribution function whose marginals are all uniform (0, 1). If $U = (u_1, u_2, \ldots, u_D)$ is uniformly distributed in (0, 1) then the D-dimensional copula is given by:

$$C(u_1, \ldots, u_D) = C(U_1 < u_1, \ldots, U_n < u_D)$$

Using the result that any continuous random variable X with probability distribution function F_i can be transformed by its probability integral transform to lie between (0, 1). For a set of continuous random variables $X_1, X_2, \ldots, X_D$ with marginal probability distribution functions $F_1, F_2, \ldots, F_D$ and joint distribution function $F_{1,2,\ldots,D}$, then

$$F(x_1, \ldots, x_D) = C\left(F_1(x_1), \ldots, F_D(x_D)\right)$$

where C is unique as long as $F_1, F_2, \ldots, F_D$ are continuous, a result first shown in (Sklar, 1959). So any joint probability distribution function can be expressed as a combination of a copula and each of the variables univariate marginal distributions. For more information regarding copulas see Joe (1997) .

Two elliptical copula types can be discussed here:

A.1. Gaussian copula

If $X_1,\ldots,X_n$ is a set of correlated variables with correlation matrix Γ, then for $U = (u_1, \ldots, u_D)$ in $(0,1)^n$ the Gaussian copula is defined as

$$C_\Gamma(u) = \Phi_\Gamma\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_D)\right)$$

where Φ^{-1} is the inverse cumulative distribution function of a standard normal and Φ_Γ is the joint standard normal cumulative distribution function with mean vector 0 and correlation matrix Γ.

A.2. Student's t-copula

If $X_1,\ldots,X_n$ is a set of correlated variables with correlation matrix Γ, then for $U = (u_1, \ldots, u_D)$ in $(0,1)^n$ the Student's t-copula is defined as

$$C_{\upsilon\Gamma}(u) = t_{\upsilon\Gamma}\left(t_\upsilon^{-1}(u_1), \ldots, t_\upsilon^{-1}(u_D)\right)$$

where t^{-1} is the inverse cumulative distribution function of the one dimensional Student's t-distribution with υ degrees of freedom and $t_{\upsilon\Gamma}$ is the joint cumulative distribution function for the Student's t-distribution with υ degrees of freedom.

Appendix B. Probability integral transform

The probability integral transform states that if X has the continuous cdf $F_X(x)$ and the random variable U is defined as $U = F_X(x)$. Then U is uniformly distributed on (0, 1), that is, $P(U \le u) = u, 0 < u < 1$. Conversely, the inverse probability integral transform states $X = F_X^{-1}(U)$ has distribution function F_X. For a proof see Casella and Berger (1990). So, to transform a realisation of X to (0, 1) apply F_X and to transform a realisation of U to X with distribution function F_X, apply F_X^{-1}.

Appendix C. Kendalls τ

For a random vector (X, Y), Kendalls τ, is defined as

$$\tau = Pr\left\{(X-\tilde{X})(Y-\tilde{Y})>0\right\} - Pr\left\{(X-\tilde{X})(Y-\tilde{Y})<0\right\}$$

where $(\tilde{X}, \tilde{Y})$ is an independent copy of (X, Y). It is therefore the probability of concordance minus the probability of discordance.

References

Acar, E.F., Craiu, R.V., Yao, F., 2011. Dependence calibration in conditional copulas: a nonparametric approach. Biometrics 67 (2), 445–453.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory. Tsahkadsor, Armenia, USSR, pp. 267–281.

Battjes, J.A., Janssen, J.P.F.M., 1978. Energy loss and set-up due to breaking of random waves. Proceedings of 16th Conference on Coastal Engineering. ASCE., pp. 569–587.

Bedford, T., Cooke, R.M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. Ann. Math. Artif. Intell. 32 (1-4), 245–268.

Bedford, T., Cooke, R.M., 2002. Vines — a new graphical model for dependent random variables. Ann. Stat. 30 (4), 1031–1068.

Bernardara, P., Mazas, F., Kergadallan, X., Hamm, L., 2014. A two-step framework for over-threshold modelling of environmental extremes. Nat. Hazards Earth Syst. Sci. 14 (3), 635–647.

Bolaños, R., Sánchez-Arcilla, A., Cateura, J., 2007. Evaluation of two atmospheric models for wind-wave modelling in the NW Mediterranean. J. Marine Sys. 65 (1–4), 336–353.

Booij, N., Ris, R.C., Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions, model description and validation. J. Geophys. Res. 104 (C4), 7649–7666.

Bradbury, A., Mason, T., 2014. Review of south coast beach response to wave conditions in the winter of 2013–2014. Southeast regional coastal monitoring programme. Report SR01. Channel Coastal Observatory, National Oceanography Centre, Southampton.

Brenner, S., Gertman, I., Murashkovsky, A., 2007. Preoperational ocean forecasting in the southeastern Mediterranean Sea: implementation and evaluation of the models and selection of the atmospheric forcing. J. Marine Sys. 65 (1), 268–287.

Breymann, W., Dias, A., Embrechts, P., 2003. Dependence structures for multivariate high-frequency data in finance. Quant. Finan. 3 (1), 1–14.

Broyden, C.G., 1969. A new double-rank minimization algorithm. Not. Am. Math. Soc. 16 (4), 670.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd ed., Springer-Verlag, New York.

Camus, P., Mendez, F.J., Medina, R., 2011. A hybrid efficient method to downscale wave climate to coastal areas. Coast. Eng. 58 (9), 851–862.

Casella, G., Berger, R.L., 1990. Statistical Inference vol. 70. Duxbury Press, Belmont, California.

Cavaleri, L., Alves, J.-H., Ardhuin, F., Babanin, A., Banner, M., Belibassakis, K., Benoit, M., Donelan, M., Groeneweg, J., Herbers, T., Hwang, P., Janssen, P., Janssen, T., Lavrenov, I., Magne, R., Monbaliu, J., Onorato, M., Polnikov, V., Resio, D., Rogers, W., Sheremet, A., Smith, J.M., Tolman, H., van Vledder, G., Wolf, J., Young, I., 2007. Wave modelling The state of the art. Prog. Oceanogr. 75 (4), 603–674.

Cavaleri, L., Sclavo, M., 2006. The calibration of wind and wave model data in the Mediterranean Sea. Coast. Eng. 53 (7), 613–627.

Coles, S.G., 2001. An Introduction to Statistical Modeling of Extreme Values. 1st ed., Springer-Verlag, London.

Coles, S.G., Tawn, J.A., 1991. Modelling extreme multivariate events. J. R. Stat. Soc. Ser. B (Methodol.) 53 (2), 377–392.

Corbella, S., Stretch, D.D., 2012. Multivariate return periods of sea storms for coastal erosion risk assessment. Nat. Hazards Earth Syst. Sci. 12, 2699–2708.

Cunha, C., Soares, C.G., 1999. On the choice of data transformation for modelling time series of significant wave height. Ocean Eng. 26 (6), 489–506.

B.-K., P., Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kllberg, P., Khler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J., Peubey, C., de Rosnay, P., Tavolato, C., Thpaut, J.-N., Vitart, F., 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597.

Demarta, S., McNeil, A.J., 2005. The t copula and related copulas. Int. Stat. Rev. 73 (1), 111–129.

Deo, M., Naidu, C.S., 1998. Real time wave forecasting using neural networks. Ocean Eng. 26 (3), 191–203.

Deo, M.C., Jha, A., Chaphekar, A., Ravikant, K., 2001. Neural networks for wave forecasting. Ocean Eng. 28 (7), 889–898.

deWaal, D.J., van Gelder, P.H.A.J.M., 2005. Modelling of extreme wave heights and periods through copulas. Extremes 8 (4), 345–356.

Eldeberky, Y., Battjes, J.A., 1996. Spectral modeling of wave breaking: application to boussinesq equations. J. Geophys. Res. Oceans 101 (C1), 1253–1264.

Environment Agency, 2011. Coastal flood boundary conditions for UK mainland and islands, Environment Agency Report SC060064/TR3: Design Swell Waves, Bristol.

Ferreira, J.A., Soares, C.G., 1998. An application of the peaks over threshold method to predict extremes of significant wave height. J. Offshore Mech. Arct. Eng. 120 (3), 165–176.

Ferreira, J.A., Soares, C.G., 2000. Modelling distributions of significant wave height. J. Coast. Eng. 40 (4), 361–374.

Fisher, R., Tippett, L., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society 24 (2), 180–190.

Fletcher, R., 1970. A new approach to variable metric methods. Comput. J. 13 (3), 317–322.

Gaur, S., Deo, M., 2008. Real-time wave forecasting using genetic programming. Ocean Eng. 35 (11), 1166–1172.

Genest, C., Favre, A.C., 2007. Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Eng. 12 (4), 347–368.

Genest, C., Ghoudi, K., Rivest, L.P., 1995. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82 (3), 543–552.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum dune serie aleatoire. Annals of Mathematics

Goda, Y., 1988. On the methodology of selecting design wave height. Proceedings of the 21st International Conference on Coastal Engineering. ASCE, Malaga, pp. 899–913.

Goldfarb, D., 1970. A family of variable metric methods derived by variational means. Math. Comput. 24 (109), 23–26.

Gopinath, D.I., Dwarakish, G.S., 2015. Wave prediction using neural networks at New Mangalore Port along west coast of India. Aquatic Procedia 4, 143–150.

Gouldby, B., Méndez, F.J., Guanche, Y., Rueda, A., Mínguez, R., 2014. A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis. Coast. Eng. 88, 15–26.

Hadadpour, S., Etemad-Shahidi, A., Kamranzad, B., 2014. Wave energy forecasting using artificial neural networks in the Caspian Sea. Proceedings of the Institution of Civil Engineers-Maritime Engineering 167 (1), 42–52.

Hasselmann, K., 1974. On the spectral dissipation of ocean waves due to white-capping. Bound.-Layer Meteorol. 6 (1-2), 107–127.

Hawkes, P.J., Gouldby, B.P., Tawn, J.A., Owen, M.W., 2002. The joint probability of waves and water levels in coastal engineering design. J. Hydraul. Res. 40 (3), 241–251.

Heffernan, J.E., Tawn, J.A., 2004. A conditional approach for multivariate extreme values. J. R. Stat. Soc. Ser. B (Stat Methodol.) 66 (3), 497–546.

Hidalgo, O.S., Nieto Borge, J., Cunha, C.C., Guedes Soares, C., 1995. Filling missing observations in time series of significant wave height. American Society of Mechanical Engineers, New York, NY (United States).

Ho, P.C., Yim, J.Z., 2006. Wave height forecasting by the transfer function model. Ocean Eng. 33 (8), 1230–1248.

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2015. Copula: Multivariate Dependence with Copulas. R package version 0.999-13

Hofert, M., Mächler, M., 2011. Nested Archimedean copulas meet R: the nacopula package. J. Stat. Softw. 39 (9), 1–20.

Hogg, R.V., Craig, A.T., 1995. Introduction to Mathematical Statistics. 5th, Prentice-Hall Inc.

Holthuijsen, L.H., Booij, N., Bertotti, L., 1996. The propagation of wind errors through ocean wave hindcasts. J. Offshore Mech. Arct. Eng. 118 (3), 184–189.

Holthuijsen, L.H., Herman, A., Booij, N., 2003. Phase-decoupled refraction-diffraction for spectral wave models. Coast. Eng. 49 (4), 291–305.

Jain, P., Deo, M., 2007. Real-time wave forecasts off the western Indian coast. Appl. Ocean Res. 29 (1), 72–79.

Joe, H., 1996. Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. Lecture Notes Monograph Series. vol. 28. Institute of Mathematical Statistics, Hayward, CA, pp. 120–141.

Joe, H., 1997. Multivariate Models and Multivariate Dependence Concepts (Monographs on Statistics and Applied Probability). CRC Press.

Jonathan, P., Flynn, J., Ewans, K., 2010. Joint modelling of wave spectral parameters for extreme sea states. Ocean Eng. 37 ((11)12), 1070–1080.

Kamranzad, B., Etemad-Shahidi, A., Kazeminezhad, M., 2011. Wave height forecasting in Dayyer, the Persian Gulf. Ocean Eng. 38 (1), 248–255.

Kazeminezhad, M., Etemad-Shahidi, A., Mousavi, S., 2005. Application of fuzzy inference system in the prediction of wave parameters. Ocean Eng. 32 (14), 1709–1725.

Kim, G., Silvapulle, M.J., Silvapulle, P., 2007. Comparison of semiparametric and parametric methods for estimating copulas. Comput. Stat. Data Anal. 51 (6), 2836–2850.

Kojadinovic, I., Yan, J., 2010. Modeling multivariate distributions with continuous margins using the copula r package. J. Stat. Softw. 34 (9), 1–20.

Komen, G.J., Hasselmann, S., Hasselmann, K., 1984. On the existence of a fully developed windsea spectrum. J. Phys. Oceanogr. 14 (8), 1285–1371.

Kurowicka, D., Cooke, R.M., 2006. Uncertainty Analysis with High Dimensional Dependence Modelling. Chichester: John Wiley & Sons.

Little, R.J.A., 1988. A test of missing completely at random for multivariate data with missing values. J. Am. Stat. Assoc. 83 (404), 1198–1202.
Liu, P.C., Schwab, D.J., Jensen, R.E., 2002. Has wind-wave modeling reached its limit? Ocean Eng. 29 (1), 81–98.
Londhe, S., 2008. Soft computing approach for real-time estimation of missing wave heights. Ocean Eng. 35 (11–12), 1080–1089.
Londhe, S., Panchang, V., 2006. One-day wave forecasts based on artificial neural networks. J. Atmos. Oceanic Technol. 23 (11), 1593–1603.
Ma, M., Song, S., Ren, L., Jiang, S., Song, J., 2013. Multivariate drought characteristics using trivariate Gaussian and Student t copulas. Hydrol. Process. 27 (8), 1175–1190.
Mahjoobi, J., Mosabbeb, E.A., 2009. Prediction of significant wave height using regressive support vector machines. Ocean Eng. 36 (5), 339–347.
Malekmohamadi, I., Bazargan-Lari, M.R., Kerachian, R., Nikoo, M.R., Fallahnia, M., 2011. Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. Ocean Eng. 38 (23), 487–497.
Malekmohamadi, I., Ghiassi, R., Yazdanpanah, M., 2008. Wave hindcasting by coupling numerical model and artificial neural networks. Ocean Eng. 35 (3), 417–425.
Malekmohamadi, I., Ghiassia, R., Yazdanpanah, M.J., 2008. Wave hindcasting by coupling numerical model and artificial neural networks. Ocean Eng. 35 (3), 417–425.
Mandal, S., Prabaharan, N., 2006. Ocean wave forecasting using recurrent neural networks. Ocean Eng. 33 (10), 1401–1410.
Mashal, R., Zeevi, A., 2002. Beyond Correlation: Extreme Co-movements Between Financial Assets. Unpublished. Columbia University.
Michele, C.D., Salvadori, G., Passoni, G., Vezzoli, R., 2007. A multivariate model of sea storms using copulas. Coast. Eng. 54 (10), 734–751.
Moeini, M.H., Etemad-Shahidi, A., Chegini, V., 2010. Wave modeling and extreme value analysis off the northern coast of the Persian Gulf. Appl. Ocean Res. 32 (2), 209–218.
Nitsure, S., Londhe, S., Khare, K., 2012. Wave forecasts using wind information and genetic programming. Ocean Eng. 54, 61–69.
Özger, M., Şen, Z., 2007. Prediction of wave parameters by using fuzzy logic approach. Ocean Eng. 34 (3), 460–469.
Patton, A.J., 2006. Modelling asymmetric exchange rate dependence. Int. Econ. Rev. 47 (2), 527–556.
Picklands, J., 1975. Statistical inference using extreme order statistics. Ann. Stat. 3, 119–131.
Pilar, P., Soares, C.G., Carretero, J.C., 2008. 44-year wave hindcast for the North East Atlantic European coast. Coast. Eng. 55 (11), 861–871.
Poulomi, G., Reddy, M., 2013. Probabilistic assessment of flood risks using trivariate copulas. Theor. Appl. Climatol. 111 (1), 341–360.
Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, New York.
Rusu, L., Pilar, P., Soares, C.G., 2008. Hindcast of the wave conditions along the west Iberian coast. Coast. Eng. 55 (11), 906–919.
Salvadori, G., Tomasicchio, G.R., D'Alessandro, F., 2014. Practical guidelines for multivariate analysis and design in coastal and off-shore engineering. Coast. Eng. 88, 1–14.
Sánchez-Arcilla, A., González-Marco, D., Bolaños, R., 2008. A review of wave climate and prediction along the Spanish Mediterranean coast. Nat. Hazards Earth Syst. Sci. 8 (6), 1217–1228.
Savu, C., Trede, M., 2010. Hierarchies of Archimedean copulas. Quant. Financ. 10 (3), 295–304.
Scotto, M., Soares, C.G., 2000. Modelling the long-term time series of significant wave height with non-linear threshold models. *Coast. Eng.* 40 (4), 313–327.
Shanno, D.F., 1970. Conditioning of quasi-Newton methods for function minimization. Math. Comput. 24 (111), 145–160.
Sibley, A., Cox, D., Titley, H., 2015. Coastal flooding in England and Wales from Atlantic and North Sea storms during the 2013/2014 winter. Weather 70 (2), 62–70.
Sklar, A.W., 1959. Fonctions de repartition ‘a n dimension et leurs marges. Publications de lInstitut de Statistique de lUniversite de Paris 8, pp. 229–231
Soares, C., Ferreira, A., Cunha, C., 1996. Linear models of the time series of significant wave height on the Southwest Coast of Portugal. *Coast. Eng.* 29 (1-2), 149–167.
Soares, C.G., Scotto, M.G., 2004. Application of the r largest-order statistics for long term predictions of significant wave height. J. Coast. Eng. 51 (5), 387–394.
Teixeira, J.C., Abreu, M.P., Soares, C.G., 1995. Uncertainty of ocean wave hindcasts due to wind modeling. J. Offshore Mech. Arct. Eng. 117 (4), 294–297.
Tolman, H.L., 1991. A third-generation model for wind waves on slowly varying, unsteady and inhomogeneous depths and currents. J. Phys. Oceanogr. 21 (6), 782–797.
Tsai, C., Chen, H., You, S., 2009. Toe scour of seawall on a steep seabed by breaking waves. J. Waterway Port Coast. Ocean Eng. 135 (2), 61–68.
van Buuren, S., 2007. Multiple imputation of discrete and continuous data by fully conditional specification. Stat. Methods Med. Res. 2 16 (3), 219–242.
Vanem, E., 2011. Long-term time-dependent stochastic modelling of extreme waves. *Stoch. Environ. Res. Risk Assess.* 25 (2), 185–209.
Vicinanza, D., Contestabile, P., Ferrante, V., 2013. Wave energy potential in the north-west of Sardinia (Italy). Renew. Energy 50, 506–521.
WAMDI group, 1988. The WAM model? A third generation ocean wave prediction model. J. Phys. Oceanogr. 18 (12), 1775–1810.
Wang, X., Gebremichael, M., Yan, J., 2010. Weighted likelihood copula modeling of extreme rainfall events in connecticut. J. Hydrol. 390 (1–2), 108–115.
Wong, G., Lambert, M.F., Leonard, M., Metcalfe, A.V., 2010. Drought analysis using trivariate copulas conditional on climatic states. J. Hydrol. Eng. 15 (2), 129–141.
Wornom, S.F., Welsh, D.J.S., Bedford, K.W., 2001. On coupling the SWAN and WAM wave models for accurate nearshore wave predictions. Coast. Eng. J. 43 (3), 161–201.
Yan, J., 2007. Enjoy the joy of copulas: with a package copula. J. Stat. Softw. 21 (4), 1–21.
Zhang, Z., Li, C.-W., Li, Y.-S., Qi, Y., 2006. Incorporation of artificial neural networks and data assimilation techniques into a third-generation wind-wave model for wave forecasting. J. Hydroinform. 8 (1), 65–76.
Zong, Y., Tooley, M.J., 2003. A historical record of coastal floods in Britain: frequencies and associated storm tracks. Nat. Hazards 29 (1), 13–36.