Title: A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis

Author 1: Gouldby B [a,b]   b.gouldby@hrwallingford.com

Author 2: Mendez F. J. [a]   fernando.mendez@unican.es

Author 3: Guanche Y [a]   yanira.guanche@unican.es

Author 4: Rueda A [a]   ana.rueda@unican.es

Author 5: Minguez R [a]   roberto.minguez@unican.es

a   Environmental Hydraulics Institute "IH Cantabria", Universidad de Cantabria, C/
         Isabel Torres, 15 Parque Científico y Tecnológico de Cantabria 39011 Santander,
         Spain Tel. +34 942 201616, Spain
b   HR Wallingford Ltd., Howbery Park, Wallingford, Oxfordshire, OX10 8BA, UK.
         Tel. +44 1491 822899

Corresponding Author: Gouldby B

Abstract

Extreme sea conditions in the nearshore zone are required for coastal flood risk analysis and structural design. Many multivariate extreme value methods that have been applied in the past have been limited by assumptions relating to the dependence structure in the extremes. A conditional extremes statistical model overcomes a number of these previous limitations. To apply the method in practice, a Monte Carlo sampling procedure is required whereby large samples of synthetically generated events are simulated. The use of Monte Carlo approaches, in combination with computationally intensive physical process models, can raise significant practical challenges in terms of computation. To overcome these challenges there has been extensive research into the use of meta-models. Meta-models are approximations of computationally intensive physical process models (simulators). They are derived by fitting functions to the outputs from simulators. Due to their simplified representation they are computationally more efficient than the simulators they approximate.

Here, a methodology for deriving a large Monte Carlo sample of extreme nearshore sea states is described. The methodology comprises the generation of a large sample of offshore sea conditions using the conditional extremes model. A meta-model of the wave transformation process is then constructed. A clustering algorithm is used to aid the development of the meta-model. The large sample of offshore data is then transformed through to the nearshore using the meta-model. The resulting nearshore sea states can be used for the probabilistic design of structures or flood risk analysis. The application of the methodology to a case study site on the North Coast of Spain is described.

Keywords: Joint probability, multivariate extremes, flood risk, meta-model, probabilistic design

1 Introduction

Extreme sea conditions in the nearshore zone are required for a variety of purposes. These include; design of coastal structures (breakwaters, port and harbour infrastructure), boundary conditions for flood risk analysis models and design of beach management schemes. Sea conditions are typically defined by a range of variables, significant wave height (H_s), mean wave period (T_m) and sea level, for example. These variables can be further disaggregated into wind-sea and swell components of wave conditions, and astronomical tide and surge levels for the sea level aspects. More often than not, formulae used for the design of coastal structures are a function of more than one of these variables. Moreover, coastal flooding can arise over large spatial scales. As the sea conditions can vary significantly in space, and are often not fully dependent, (Lewis et al., 2011), there is a need to develop nearshore sea states that are extreme, yet remain spatially coherent and plausible.

The dependence structure between the different sea condition variables (eg. H_s, T_m), and between a single variable (eg. H_s) at different points in space, can be complex. Assumptions of independence or full dependence are rarely good approximations when analysing extreme sea conditions. It is therefore often necessary to explicitly consider the dependence between the variables. Multivariate extreme value models that capture these dependencies are therefore required. These joint probability approaches have been applied in the context of coastal analyses for many years, see for example, (Coles and Tawn, 1990), (Tawn, 1992), (Bruun and Tawn, 1998), (Guedes Soares and Cunha, 2000) (Hawkes et al., 2002), (Monbet et al., 2007), (De Michele et al., 2007) (Callaghan et al., 2008), (Wahl et al., 2012) and (Corbella and Stretch, 2013).

One of the common limitations of these methods however, often relates to restrictive assumptions relating to the dependence structure of the extremes. More specifically, restrictions to joint regions in which all variables are simultaneously extreme are often imposed. This can place practical limitations on the number of variables considered and also the spatial extents of the analyses. An advance in the underlying multivariate extreme value methods, (Heffernan and Tawn, 2004) (hereinafter HT04), has however, overcome a number of these limitations. The method of HT04 imposes no specific dependence structure within the extremes, enabling greater flexibility and a wider range of application. The HT04 method has been applied in the context of fluvial flood risk analysis by (Keef et al., 2009), (Lamb et al., 2010), (Keef et al., 2012) and (Wyncoll and Gouldby, 2013). It has also been applied to offshore wave conditions, (Jonathan et al., 2013a), (Ewans and Jonathan, 2013), (Jonathan et al., 2013b) and the nearshore environment, (Environment Agency, 2013). For practical implementation, the method comprises the fitting of statistical models and then Monte Carlo approximation, involving the stochastic simulation of large (1000´s) samples of extreme conditions.

The natural boundary condition for undertaking multivariate extreme value analysis for coastal engineering and flooding applications is offshore in deep water, prior to the complex nearshore wave transformation processes of refraction, shoaling, breaking etc, (Bruun and Tawn, 1998). There are now extensive data sources for offshore wave conditions from in-situ and remote measurements as well as re-analysis (hindcast) from coupled atmospheric and ocean numerical models, (Reguero et al., 2012) and (Camus et al., 2013), for example. There are also datasets that combine these two different sources, (Mínguez et al., 2011), (Mínguez et al., 2012)

There are numerous numerical models used for simulating the propagation of waves from offshore to nearshore. Perhaps the most widely used model is SWAN, (Booij et al., 1999). Even with increases in computational resources, these numerical models can be impractical to run for large data sets. In recognition of these practical limitations, methods comprising the application of meta-models to the wave transformation process, (Camus et al., 2011a), have been developed. Meta-models are, in essence, simplified (and hence computationally efficient) representations of computationally intensive models, (Sacks et al., 1989). These methods have been applied in a wide range of fields and are closely related to the response surface method, (Box and Draper, 1959) and (Hill and Hunter, 1966), that is extensively applied in the field of structural reliability analysis . There are now a wide variety of meta-modelling techniques that have been applied. (Kingston et al., 2011) have used a neural network to approximate a finite element geotechnical model, (Camus et al., 2011a) used radial basis functions (RBF´s) to approximate the SWAN wave model and Gaussian Process emulators are increasingly being applied to a wide range of models, including global vegetation models and soil/plant atmosphere models, (Kennedy et al., 2006), for example. A general framework for dynamic emulation modelling is described by (Castelletti et al., 2012).

Meta-models are developed by undertaking simulations of the computationally intensive physical process model (so-called simulator), at a range of so-called design points (nb: the use of the expression "design point" here is consistent with its use in the meta-modelling field, (O´Hagan, 2006), for example and should not be confused with alternative uses that relate to the reliability index) that cover the input boundary space. The inputs are then mapped to the outputs using the fitted model. In the development of the meta-model, one of the challenges is to ensure the design points are selected to appropriately cover the input boundary conditions, whilst capturing the complexities in the output response surface. This has to be achieved whilst seeking to minimise the number of design points, and hence computational cost, associated with the simulator.

Data clustering methods can be used to assist in the appropriate selection of the design points. The clustering methods are capable of reducing the data to a subset of types with similar

primary properties. This information can be used in the selection of the design points to create the meta-model, thereby increasing efficiency and performance of the meta-model. (Camus et al., 2011b) have explored the performance of a range of clustering techniques (K-means algorithm (KMA), Self Organising Maps (SOM) and the Maximum Dissimilarity Algorithm (MDA)) within the context of offshore wave data.

This paper describes a practical methodology for deriving a synthetic set (stochastically generated large sample) of extreme nearshore sea conditions, without using excessive computational resources. The methodology can be applied over a broad range of spatial scales. The properties of the derived dataset are that the marginal extremes are well preserved and the dependence structure between the variables reflects the empirical data. This large sample of nearshore sea conditions can then be applied for probabilistic structural design and reliability analysis or flood risk analysis. A case study example on the North Coast of Spain is used to demonstrate the methodology. In this example, the nearshore data have been applied to a commonly used structural response function of wave overtopping.

2   Methodology

The methodology compromises a number of steps:
1. De-cluster offshore time series sea condition data
2. Fit statistical models to the marginals and dependence of the de-clustered data.
3. Simulate a large sample of offshore sea condition data from the fitted models.
4. Apply the MDA clustering approach to the large sample to define a subset of m offshore sea condition design points.
5. Run a wave transformation model for the subset of m design points to obtain a subset of m transformed nearshore sea conditions.
6. Construct a meta-model of the wave transformation process by fitting functions to the set of m offshore design points and corresponding nearshore sea conditions.
7. Transform the large sample (Step 3) to nearshore using the meta-model.
8. Translate the large sample to the response of interest.

The methodology is not prescriptive in terms of the wave transformation model or the meta-modelling approach. In the case study example below however, the SWAN wave model has been used for wave transformation and RBF´s have been used as a choice of meta-model. The following methodological description is therefore defined using these specific approaches.

A summary of the methodology is illustrated in Figure 1. Each step is described in more detail below.

2.1 Multivariate extreme value method

Let X^o be a vector of offshore sea condition variables, wave height, period, direction, sea level, wind speed and direction ($X^o_1, X^o_2....X^o_n$), for example. The problem then is to estimate the probability of exceeding some specified value of a response of interest, flood depth, wave overtopping discharge or wave impact force, for example. The response (or structure) variable is denoted as Z. X^o is related to Z through the function Δ:

$$\Pr(Z > z) = \Pr(\Delta(X^o) > z)$$

(Bruun and Tawn, 1998) termed Δ the structure function and identified two alternative approaches for solving this problem, the structure variable method (SVM) and joint probability method (JPM). The SVM involves reduction of the multivariate observed data, via the structure function, to a univariate series of the structure variable and then extrapolation using standard univariate extreme value techniques. The JPM method requires extrapolation of the joint density of X^o to extremes and then integration over the region $\Delta(X^o)>z$:

$$\Pr(Z > z) = \int_{Z>z} f_{X^o}(X^o)\,dX^o \tag{1}$$

Where $f_{X^o}(X^o)$ is the joint probability density of the offshore sea condition variables. A conceptual diagram providing an overview of the problem and the two alternative approaches is shown in Figure 2. (Bruun and Tawn, 1998) identified the JPM to offer significant advantages over the SVM, as the performance of the structure function in the extremes is fully explored. This is not the case with the SVM that relies on extrapolation in this region. For this reason, the JPM approach is adopted within the methodology described here and the extrapolation of the joint probability density is undertaken using the HT04 approach. The SVM is however, applied here for comparative purposes within the case study example below.

The objective of the HT04 method is to extrapolate the joint probability density of the offshore sea condition variables to extreme values with appropriate consideration of the dependence structure. Prior to analysis of the dependencies between each variable, the marginals are first analysed. For this, the standard peaks-over-threshold (POT) approach of (Davison and Smith, 1990) is used, whereby cluster maxima are identified and the excesses above a suitably high threshold are fitted to the Generalised Pareto distribution (GPD). This defines a probability model for large values of the variable X_i.

To provide a full specification of the marginal distributions, the empirical distribution $\widetilde{F}_i$ of the X_i values, below the threshold, is combined with the GPD above the threshold to provide the following semi-parametric function for the cumulative marginal distribution ((Coles and Tawn, 1991)):

$$\hat{F}_i(x) = \begin{cases} \widetilde{F}_i(x) & x \leq u_i, \\ 1 - (1 - \widetilde{F}_i(u_i))\left[1 + \xi_i \dfrac{(x - u_i)}{\beta_i}\right]_+^{-1/\xi_i} & x > u_i. \end{cases} \tag{3}$$

Where, β_i and ξ_i are the GPD parameters and u_i is a high threshold. The GPD is a well-established model for analysing extremes for POT sea condition variables (Hamm et al., 2010)

<div style="float:right; border:1px solid #c00; background:#fdd; padding:4px;">
Comentario [HRW1]: Do you have alternative we could use here?
</div>

In common with other copula approaches that separate the marginal characteristics from the dependence analysis, it is usual to standardise the data to common margins. Within HT04 the standard Gumbel marginal scales are used. The sea condition data are therefore transformed (transformed variables denoted as Y), from their original scales to Gumbel Scales using the standard probability integral transformation. Whilst Gumbel Scales are used within HT04 and in the analysis described below, the method is not restricted to use on these scales. Recently, (Keef et al., 2012) have undertaken analysis of fluvial flooding using Laplace Scales, for example.

The method proceeds by analysis of the dependence between the variables on the transformed scales. If Y_{-i} denotes the vector of all variables excluding Y_i, the method is typically applied using the multivariate non-linear regression model

$$Y_{-i} \mid Y_i = a\,Y_i + Y_i^b\,W \qquad \text{for } Y_i > v, \tag{4}$$

Where a and b are vectors of the parameters from the fitted pair-wise regression model, v is a specified threshold and W is a vector of the residuals. The model is fitted using maximum

likelihood assuming the residuals follow a normal distribution with a mean and standard deviation to be found. In practice the model is fitted separately with each variable treated as the conditioning variable respectively. It can therefore be extended over large spatial scales, (Lamb et al., 2010).

Once fitted, a Monte Carlo simulation procedure is used whereby samples from the residuals are combined with the parameter estimates to obtain realisations of Y. The steps involved in the Monte Carlo sampling procedure are summarised below, with further information provided by HT04, Keef et al (2009), (Lamb et al., 2010) and (Wyncoll and Gouldby, 2013):

1) Sample a value of Y_i (ie from the variable on the transformed scales) conditioned to exceed threshold v.
2) Independently sample a joint residual, W, for site i.
3) Calculate Y_{-i}, from Eqn. 4, using the sampled W and the fitted regression parameters.
4) Reject if Y_i is not a maximum.

These steps are repeated until the relative proportion of events where Y_i is a maximum, conditional on being above the threshold, is consistent with the empirical distribution. This process is then repeated conditioning on each variable in turn, to ensure the appropriate proportion of events is simulated. The output of this process is a large sample of simulated data on the transformed scales.

These data are then transformed back to the original scales by reversing the previously applied transformations. The resulting output is a large multivariate sample of extreme (in at least one variable) offshore sea condition data that captures the characteristics of dependencies between the variables, as well as preserving the marginal extremes.

2.2 Clustering method

To develop the meta-model of the nearshore wave transformation process, it is necessary to run the simulation model (SWAN) for a set of m design points. The design points are required to cover the input boundary condition space. Clustering algorithms can be used to aid the selection of the subset of input boundary conditions. The MDA, (Kennard and Stone, 1969), with further refinements described by (Willett, 1999), is particularly well-suited for this task as it enables the outer limits of the input boundary space to be appropriately represented. This is not the case for other clustering methods, SOM and K-means, for example, (Camus et al., 2011b).

The MDA algorithm is applied to the Monte Carlo realisations output from the multivariate extreme value analysis of the offshore wave conditions. Prior to implementation, it is however, first necessary to transform the data onto standard scales, including the directional component. The standardisation involves making a linear transformation to scale the variables between 0 and 1, using the maximum and minimum value for each variable. For the directional variables, the maximum distance is π (in radians), hence the directions have been divided by π to scale between 0 and 1. Further information on these transformations are described in detail by (Camus et al., 2011a).

The output from the application of the MDA defines a subset of m points that are uniformly distributed in the transformed space across the offshore boundary (i.e. the SWAN model input boundary space). This subset of design points is constructed sequentially. Firstly, the point with the maximum significant wave height is identified. The next stage is to calculate the point in the data set that is furthest, in terms of Euclidean Distance, from this point. Then, the algorithm determines the point in the data set that is furthest from these two points, and so on, until a subset of size m points is defined. The choice of the size of m requires consideration, as it represents the trade-off between the computational demand of running the simulator and the accuracy of the subsequently fitted meta-model. The required accuracy of the meta-model

will vary depending on the specific application and the uncertainties associated with subsequent modelling components. (Camus et al., 2011a) have explored the accuracy of the RBF meta-modelling approach as a function of m and this information can be used to help inform this decision.

2.3 Meta-modelling method

There are a wide range of function approximation techniques that can be used to develop meta-models of complex physical process models. These methods include: Piecewise Polynomials, Neural Networks and RBF´s. The methodology described here is not constrained to the use of a specific technique. The approach used on the case study site below, and hence described here, is that of RBF´S. This approach has been chosen because RBF´s are widely applied and have proven to work well in the context here, with SWAN being used as the simulator, (Camus et al., 2011a).

The RBF has the following general form, (Rippa, 1999), for example:

$$X^N = p(\boldsymbol{X}^o) + \sum_{i=1}^{m} a_i \Phi(\|\boldsymbol{X}^o - \boldsymbol{D}\|) \tag{5}$$

Here, $\boldsymbol{X^N}$ is the vector of the near-shore sea conditions (ie the output of the meta-model) and:

$$p(\boldsymbol{X^o}) = b_0 + b_1 X_1^o + b_2 X_2^o \ldots\ldots + b_n X_n^o \tag{6}$$

$b_{0,1,2..n}$ are coefficients to be found by fitting the RBF to the known points and Φ is a Gaussian function defined as:

$$\Phi(\|\boldsymbol{X^o} - \boldsymbol{D}\|) = \exp(-\frac{\|\boldsymbol{X^o} - \boldsymbol{D}\|^2}{2c^2}) \tag{7}$$

where c is a shape parameter and $\boldsymbol{D}$ is a vector comprising the m "known" near-shore wave conditions derived from the design point simulations of the SWAN Model.

When the coefficients have been found, Equation 5 is then used in place of the SWAN model to transform the large sample of offshore sea conditions to the nearshore with minimal computational effort. The resulting large sample of nearshore data can then be utilised for a subsequent structural risk analysis or flooding assessment for example.

3 Case study implementation

3.1 Study location and data sources

The location chosen for the case study site was Santander Bay, one of the largest inlets on the Cantabrian Coast in the North of Spain (Figure 3). The city is bordered by a number of beaches that are popular recreational amenities, these include; Sardinero, Loredo and Somo. Since the 18th century, the bay has seen many changes which have determined the evolution of the current system. The land required for the expansion of the city at the beginning of the 19th Century was obtained from the bay by reclaiming its western part, resulting in a subsequent decrease of the tidal prism. Intensive dredging activity, mainly on the offshore shoal and the spit end, has been carried out in order to maintain a navigable channel.

The global wave hindcast GOW, (Reguero et al., 2012), has been used as the primary source of wave data for this study. This reanalysis data set uses the Wave Watch III numerical model forced by 6-hourly wind fields from the atmosphere model NCEP/NCAR. The reanalysis GOW spans from 1948-2008 with hourly resolution. These data have been further downscaled to regional scale to obtain a Downscaled Ocean Waves (DOW) database, (Camus

et al., 2013). The DOW data comprise hourly data for the period 1948-2008 with spatial resolution of ~200 m along the Spanish coast. These data have been calibrated using instrumental records (Espejo et al. (2011), and (Mínguez et al., 2011).

Sea level data (astronomical and surge residuals) in the form of hourly time series from two different tide gauges, the Spanish Institution of Oceanography (1940-2005) (IEO) and from Puertos del Estado (1995-present), were used in the analysis. Where there were gaps in the time series, these were filled using a regional storm surge reanalysis of southern Europe, (Abascal et al., 2012).

3.2 Method implementation

The offshore variables considered on the case study site comprised waves (height (H_s), mean period (T_m) and direction (θ_{Hs})), winds (speed (U) and direction (θ_U)), sea level (surge (S) and astronomical component (A)). An additional variable of wave steepness (st), derived from H_S and T_m has also been used in the context of the statistical simulation and this is described further below.

The first step applied in the analysis was to de-cluster the offshore time series data into separate independent events. A notional flooding level (l), that is a function of the primary variables of interest, was used for this purpose. The notional flooding level was defined as the addition of sea level to the 2% runup level of (Stockdon et al., 2006) for a dissipative beach:

$$l = S + 0.043 \left[\frac{H_s g T^2}{2\pi} \right]^{1/2} \tag{8}$$

A time series of flood levels was constructed with events then extracted using a 3 day separation criteria and a threshold that retained the highest 20% of events. The resulting data, together with the de-clustered event set are shown in Figure 4. The event set comprised a total of 1918 records, approximately 31 per year.

Following the de-clustering process, it was then necessary to define the statistical approach for each variable. The HT04 method is appropriate for the extrapolation of the joint density of a series of variables to extreme values. In this study, Hs, U and S were selected for use with this approach. Wave period was also considered, however, this is strongly dependent on, and physically constrained by, wave height, (Toffoli et al., 2010), for example. Hence, adopting a similar approach to that of (Hawkes et al., 2002), wave period was implied through the use of the derived variable of wave steepness:

$$st = \frac{2\pi H_s}{g T_m^2} \tag{9}$$

A regression model of steepness, conditional on H_s was developed. Rather than using the (Hawkes et al., 2002) assumption of a normally distributed error model with a constant mean and variance however, the heteroscedastic nature of the relationship (steepness tends to a constant value with an increase in H_s) was captured using polynomials fitted to the mean and variance of the error, conditional on H_s. These fitted models were introduced within the simulation procedure. During the simulation procedure, rejection sampling was used to maintain the steepness variable within physically plausible ranges.

For wave and wind direction, within the simulation, the empirical distributions were utilised, conditional on the magnitude of H_s and U respectively. The total sea level was obtained by combining the stochastically simulated surge data with sampled astronomical tides. The astronomical tides were sampled by discretising the 18.6 conditional on the surge value, taking account of the surge level and month. The process involved the following steps:

- Discretisation of the empirical surge data into a 10 (discrete surge bands) by 12 (month) matrix
- Sampling a surge value using the HT04 model
- Sampling a month using the empirical surge/month joint probability density comprised within the 10 by 12 matrix
- Sampling a year from the 18.6 year tidal cycle assuming all years are equally likely.
- Sampling an astronomical tidal level, given the previously sampled month and year, assuming all tidal levels within the month were equally likely.

The statistical treatment of each variable considered in the simulation procedure is summarised in Table 1

The marginal distributions of H_S, U, S were defined using Equation 3. The threshold used for all three for the GPD fit was 97.5%, This choice was informed using the mean residual life plots,(Coles, 2001). To fit the dependence model a threshold of 85% was used, based on analysis of the stability of the dependence parameters over a range of thresholds and with the independence of the residuals (W) on the conditioning variables confirmed. Further information on these diagnostics is provided within HT04.

The resulting simulated data set, comprising approximately 314,000 realisations, representative of 10,000 years, is shown in terms of H_s in Figure 5. It is evident there is a strong dependence between surge and wave heights but combining the surge component with the astronomical tide significantly dilutes this dependence. Wave heights are highly correlated with wind speeds and it is noticeable that the highest waves always occur when the wind is from a highly specific WNW direction. The simulated distribution of H_S and T_m conforms well to the empirical data. Plots comparing the marginal distributions obtained from the multivariate simulation with those obtained from fitting the GPD to the empirical data are shown for Hs, S and U in Figures 6, 7 and 8 respectively.

The number of design points (m) used to construct the meta-model was 500. The MDA was run to define these design points. These points are shown, together with the simulated data, in Figure 9. The design points provided boundary conditions for a series of simulations from the SWAN model. SWAN was run in stationary mode. The meta-model was constructed by pairing the nearshore sea conditions output from the SWAN model with the respective offshore input conditions and fitting the RBF´s. The full set of simulated offshore data was then passed through the meta-model to obtain a corresponding nearshore set of simulated conditions. These simulated nearshore conditions are shown in Figure 10, together with the transformed design points. The computational simulation time for the 500 separate SWAN simulations was 35.3 hours on a standard desktop computer. Had the SWAN model been run for each Monte Carlo realisation, as an alternative to using the meta-model, the computational time would have been greater than 35 days. The use of the meta-model has thus resulted in a substantial reduction in terms of the computational time, hence aiding practical implementation.

To demonstrate the results of the new methodology, a comparison has been made with the SVM as defined by (Bruun and Tawn, 1998). Wave overtopping rate has been used as a structural response variable to show an application of the simulated nearshore data. The structure used was a hypothetical simply sloping embankment with a 1:4 slope. The structure was located to the north of Sardinero Beach (Figure 2). Empirical formulae for the positive, and negative freeboard situations from the EuroTop Manual, (Pullen et al., 2007), were applied to estimate the overtopping rates, Equations 10 and 11.

$$\frac{q}{gT_mH_s} = Q \ \exp\left(-b\frac{R_c}{T_m\sqrt{gH_s}}\right) \tag{10}$$

$$q = 1.71ch^{1.5} \tag{11}$$

Where q is the overtopping/overflow discharge rate $(m^3/s.m)$, Q_o and b are coefficients, R_c is the freeboard (m), c is the crest width (m) and h is the head of water (m).

It is important to note that there is a distinct change in the physical processes causing the discharge as the mechanism switches from wave overtopping (water level below the crest level) to overflow (water level above the crest level) and a different equation is used to reflect this. With the SVM approach, the majority (often all), of the observed data may relate to the situation when the water level is below the crest level (wave overtopping). Direct extrapolation of these observed overtopping rates to extremes (ie the SVM approach) will therefore not be able to reflect this change in physical process (difference between Equation 10 and 11) from wave overtopping to overflow. This can therefore lead to erroneous extrapolation. The JPM is however, capable of reflecting this changing behaviour. The extrapolation to extremes and stochastic simulation of the offshore variables, and subsequent wave transformation, ensures that both negative and positive freeboard situations are appropriately accounted for.

Figure 11 shows the comparison of wave overtopping rates calculated through the SVM and JPM respectively. The differences between the two methods in the upper tail are consistent with the findings of (Bruun and Tawn, 1998) and emphasise advantages afforded by applying the JPM.

4   Discussion

JPM methods offer advantages over SVM approaches for coastal engineering and flooding applications. A JPM that uses the method of HT04 for extrapolation of the joint probability density has been described. This method overcomes a number of the constraints relating to the dependence structure that are found in previously applied approaches. As no specific dependence structure is imposed, this enables a greater number of variables to be included within the analysis. In some locations it may be necessary to consider populations of swell and wind waves separately, perhaps with different predominant directions, for example. It is feasible to separate the data and treat these as individual variables within the context of the multivariate extremes analysis. Or, alternatively, in areas sheltered from the predominant wave and wind direction, waves generated locally from divergent winds may require separate consideration. Again, separate populations of wind speeds, conditional on direction, can feature within the multivariate analysis. Coastal flooding can occur over large spatial scales. The method of HT04 can readily be applied to large spatial scales and this aspect is well demonstrated in the context of fluvial flooding by (Lamb et al., 2010).

The output of the HT04 analysis is a large, synthetically generated, sample of extreme events. Meta-models offer a computationally efficient method of effectively using this data in practice. Once configured, the meta-model can be used to undertake extensive subsequent analysis with minimal computational effort. This could include uncertainty and sensitivity analysis or climate change impact analysis for different emission scenarios, for example.

Given the flexibility of the methodology, it is envisaged that extreme nearshore data sets can be generated over large spatial scales (nationally), stored and then made available for a multitude of different purposes.

The primary limitations within the implementation of the methodology relate to the uncertainties inherent within the simulated data obtained from the fitted statistical models and those introduced by the use of the meta-model. More specifically, in the analysis described here, point estimates of the HT04 regression parameters (including thresholds) are used to derive the simulated data. Bootstrapping approaches can be used to explore parameter uncertainties and further details are provided by HT04. These uncertainties have been quantified in the context of fluvial flooding by (Neal et al., 2012). In addition, threshold selection requires careful consideration and approaches to quantify uncertainties associated with thresholds have been developed for univariate extremes, (Tancredi et al., 2006), for example.

The use of the meta-model inevitably introduces a model structural error within the process of wave transformation. The magnitude of any error introduced is a function of the number of design points that are selected and hence is controllable. Exploration into the nature of this error has been undertaken by (Camus et al., 2011a).

The magnitude of uncertainties associated with the parameter estimation within the multivariate extreme analysis, and those introduced by the meta-model, require consideration with those associated with any subsequent analysis. It is well known that there are significant (order of magnitude) model structural uncertainties associated with empirical wave overtopping estimation, (Pullen et al., 2007) and (Kingston et al., 2008), for example. Moreover, coastal flood risk analysis often involves consideration of structural reliability and breach growth estimation, which are also subject to large uncertainties, (Morris et al., 2008). So whilst uncertainties are inevitably present in the approach for the derivation of the extreme nearshore sea conditions, it may be the case that these are not dominant when uncertainties in subsequent analyses are also considered.

The quantification of these uncertainties has not been undertaken here. It is however, envisaged that comprehensive analysis to quantify and explore the relative importance of these uncertainties will be undertaken in future work.

5   Conclusions

A methodology for simulating extreme nearshore sea conditions for use in the probabilistic design of structures and coastal flood risk analysis has been described. The methodology is computationally efficient and hence practical to apply. The methodology comprises the application of a multivariate extreme value method to offshore sea condition data. The multivariate extreme value method overcomes limitations relating to the dependence structure that were present in many previous methods. The output of the extremes analysis is a large sample of simulated offshore sea condition data. To transfer the offshore data to the nearshore, a meta-model of the SWAN wave model has been developed. A clustering algorithm has been applied to select the design points used in the construction of the meta-model. The use of the meta-model reduced the SWAN computational time from over 35 days to 35 hours (time taken to undertake the design point simulations).

The derived nearshore data preserves the marginal extremes for each variable and the dependencies between the variables. These data can be used in the probabilistic design of structures or as boundary conditions for coastal flood risk analysis. The methodology can be readily extended to more variables and larger spatial scales. The primary limitations of the method relate to the uncertainties associated with the different components. These include parameter and threshold uncertainties associated with the multivariate extremes extrapolation, model structural error introduced by the use of the emulator and model structural errors associated with the response functions. The quantification and exploration of these uncertainties should be considered in future work in this area.

It is envisaged the methodology could be applied at small or large spatial scales. Given the prevalence of offshore data, particularly from re-analyses, it is envisaged the methodology can be applied at national scales with the resulting data available for a variety of purposes.

References

Abascal, A.J., Castanedo, S., Cid, A. and Medina, R., 2012. A high resolution storm surge hindcast in southern europe: A useful database for coastal applications. 33rd International Conference on Coastal Engineering (ICCE-2012), COPRI-ASCE, Santander, Spain. , p.^pp.

Booij, N., Ris, R.C. and Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions: 1. Model description and validation. Journal of Geophysical Research: Oceans, 104(C4): 7649-7666.

Box, G.E.P. and Draper, R., 1959. A basis for the selection of a response surface design. Journal of the American Statistical Association, 54(287): 622-654.

Bruun, J.T. and Tawn, J.A., 1998. Comparison of approaches for estimating the probability of coastal flooding. Journal of the Royal Statistical Society: Series C (Applied Statistics), 47(3): 405-423.

Callaghan, D.P., Nielsen, P., Short, A. and Ranasinghe, R., 2008. Statistical simulation of wave climate and extreme beach erosion. Coastal Engineering, 55(5): 375-390.

Camus, P., Mendez, F.J. and Medina, R., 2011a. A hybrid efficient method to downscale wave climate to coastal areas. Coastal Engineering, 58(9): 851-862.

Camus, P., Mendez, F.J., Medina, R. and Cofiño, A.S., 2011b. Analysis of clustering and selection algorithms for the study of multivariate wave climate. Coastal Engineering, 58(6): 453-462.

Camus, P., Mendez, F.J., Medina, R., Tomas, A. and Izaguirre, C., 2013. High resolution downscaled ocean waves (dow) reanalysis in coastal areas. Coastal Engineering, 72(0): 56-68.

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R. and Young, P.C., 2012. A general framework for dynamic emulation modelling in environmental problems. Environmental Modelling & Software, 34(0): 5-18.

Coles, S., 2001. An introduction to statistical modelling of extreme values. Springer series in statistics.

Coles, S.G. and Tawn, J.A., 1990. Statistics of coastal flood prevention. Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences, 332(1627): 457-476.

Coles, S.G. and Tawn, J.A., 1991. Modelling extreme multivariate events. Journal of the Royal Statistical Society. Series B (Methodological), 53(2): 377-392.

Corbella, S. and Stretch, D.D., 2013. Simulating a multivariate sea storm using archimedean copulas. Coastal Engineering, 76: 68-78.

Davison, A.C. and Smith, R.L., 1990. Models for exceedances over high thresholds. Journal of the Royal Statistical Society, Series B (Statistical Methodology), 52(3): 393-442.

De Michele, C., Salvadori, G., Passoni, G. and Vezzoli, R., 2007. A multivariate model of sea storms using copulas. Coastal Engineering, 54(10): 734-751.

Environment Agency, 2013. Prototype wave overtopping tool for mdsf2, Bristol.

Ewans, K. and Jonathan, P., 2013. Evaluating environmental joint extremes for the offshore industry using the conditional extremes model. Journal of Marine Systems, In Press.

Guedes Soares, C. and Cunha, C., 2000. Bivariate autoregressive models for the time series of significant wave height and mean period. Coastal Engineering, 40(4): 297-311.

Hamm, L., Mazas, F., Garcia, N. and Bailly, B., 2010. Assessment of extreme wave heights: An attempt to reconcile theory and practice. European Journal of Environmental and Civil Engineering, 14(2): 127-148.

Hawkes, P.J., Gouldby, B.P., Tawn, J.A. and Owen, M.W., 2002. The joint probability of waves and water levels in coastal engineering design. Journal of Hydraulic Research, 40(3): 241-251.

Heffernan, J.E. and Tawn, J.A., 2004. A conditional approach for multivariate extreme values (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(3): 497-546.

Hill, W.J. and Hunter, W.G., 1966. A review of response surface methodology: A literature survey. Technometrics, 8(4): 571-590.

Jonathan, P., Ewans, K. and Randell, D., 2013a. Joint modelling of extreme ocean environments incorporating covariate effects. Coastal Engineering, 79(0): 22-31.

Jonathan, P., Flynn, J. and Ewans, K., 2013b. Joint modelling of wave spectral parameters for extreme sea states. Ocean Engineering, 37((11)12): 1070-1080.

Keef, C., Tawn J and Svensson C, 2009. Spatial risk assessment for extreme river flows. Appl Stat, 58,(5): 601–618.

Keef, C., Tawn, J. and Lamb, R., 2012. Estimating the probability of widespread flood events. Environmetrics, 24(1): 13-21.

Kennard, R.W. and Stone, L.A., 1969. Computer aided design of experiments. Technometrics, 11(1): 137-148.

Kennedy, M.C., Anderson, C.W., Conti, S. and O´Hagan, A., 2006. Case studies in gaussian process modelling of computer codes. Reliability Engineering & System Safety, 91(10): 1301-1309.

Kingston, G., Robinson, D., Gouldby, B. and Pullen, T., 2008. Reliable prediction of wave overtopping volumes using bayesian neural networks. . FLOODrisk 2008, Keble College, Oxford, UK., p.^pp.

Kingston, G.B., Rajabalinejad, M., Gouldby, B.P. and Van Gelder, P.H.A.J.M., 2011. Computational intelligence methods for the efficient reliability analysis of complex flood defence structures. Structural Safety, 33(1): 64-73.

Lamb, R. et al., 2010. A new method to assess the risk of local and widespread flooding on rivers and coasts. Journal of Flood Risk Management, 3(4): 323-336.

Lewis, M., Horsburgh, K., Bates, P. and Smith, R., 2011. Quantifying the uncertainty in future coastal flood risk estimates for the u.K. Journal of Coastal Research: 870-881.

Monbet, V., Ailliot, P. and Prevosto, M., 2007. Survey of stochastic models for wind and sea state time series. Probabilistic Engineering Mechanics, 22(2): 113-126.

Morris, M., Hanson, G. and Hassan, M., 2008. Improving the accuracy of breach modelling: Why are we not progressing faster? Journal of Flood Risk Management, 1(3): 150-161.

Mínguez, R., Espejo, A., Tomás, A., Méndez, F.J. and Losada, I.J., 2011. Directional calibration of wave reanalysis databases using instrumental data. J. Atmos. Oceanic Technol., 28: pp 1466-1485.

Mínguez, R., Tomás, A., Méndez, F.J. and Medina, R., 2012. Mixed extreme wave climate model for reanalysis databases. Stochastic Environmental Research and Risk Assessment,, 27(4): pp 757-768

Neal, J., Keef, C., Bates, P., Beven, K. and Leedal, D., 2012. Probabilistic flood risk mapping including spatial dependence. Hydrological Processes, 27(9): 1349-1363.

O´Hagan, A., 2006. Bayesian analysis of computer code outputs: A tutorial. Reliability Engineering & System Safety, 91(10(11)): 1290-1300.

Pullen, T. et al., 2007. Eurotop: Wave overtopping of sea defences and related structures - assessment manual.

Reguero, B.G., Menendez, M., Mendez, F.J., Minguez, R. and Losada, I.J., 2012. A global ocean wave (gow) calibrated reanalysis from 1948 onwards. Coastal Engineering, 65(0): 38-55.

Rippa, S., 1999. An algorithm for selecting a good value for the parameter c in radial basis function interpolation. Advances in Computational Mathematics, 11(2-3): 193-210.

Sacks, J., Welch, W.J., Toby, J.M. and Wynn, H.P., 1989. Design and analysis of computer experiments. Statistical Science, 4(4): 409-423.

Stockdon, H.F., Holman, R.A., Howd, P.A. and Sallenger Jr, A.H., 2006. Empirical parameterization of setup, swash, and runup. Coastal Engineering, 53(7): 573-588.

Tancredi, A., Anderson, C. and O´Hagan, A., 2006. Accounting for threshold uncertainty in extreme value estimation. Extremes, 9(2): 87-106.

Tawn, J.A., 1992. Estimating probabilities of extreme sea-levels. Journal of the Royal Statistical Society. Series C (Applied Statistics), 41(1): 77-93.

Toffoli, A., Babanin, A., Onorato, M. and Waseda, T., 2010. Maximum steepness of oceanic waves: Field and laboratory experiments. Geophysical Research Letters, 37(5).

Wahl, T., Mudersbach, C. and Jensen, J., 2012. Assessing the hydrodynamic boundary conditions for risk analyses in coastal areas: A multivariate statistical approach based on copula functions. Nat. Hazards Earth Syst. Sci.
, 12(2): 495–510.

Willett, P., 1999. Dissimilarity-based algorithms for selecting structurally diverse sets of compounds Journal of Computational Biology, 6(3-4): 447-457.

Wyncoll, D. and Gouldby, B., 2013. Application of a multivariate extreme value approach to system flood risk analysis. Journal of flood risk management, accepted.

Figures list

Table 1 Statistical treatment of each variable

Variable	Statistical treatment for simulation
Significant Wave Height (Hs)	HT04
Wave Period (Tm)	$f(H_s, st)$
Wave Steepness (st)	Regression, conditional on H_s
Surge level (S)	HT04
Wind Speed (U)	HT04
Wind Direction (θ_U)	Empirical distribution
Wave Direction (θ_{Hs})	Empirical Distribution
Tide level (AT)	"Empirical" distribution, conditional on S and month