Multivariate extreme value modelling of sea conditions around the coast of England

Ben Gouldby[1], David Wyncoll[1], Mike Panzeri[1], Mark Franklin[2], Tim Hunt[2], Dominic Hames[1], Nigel Tozer[1], Peter Hawkes[1], Uwe Dornbusch[2] and Tim Pullen[1]

[1] HR Wallingford, Howbery Park, Wallingford, Oxfordshire OX10 8BA, UK
[2] Environment Agency
Corresponding author: Ben Gouldby, b.gouldby@hrwallingford.com

Published in the Proceedings of the Institution of Civil Engineers, Maritime Engineering, 170, Issue MA1, Pages 3-20 (March 2017).

Abstract

It is widely recognised that coastal flood events can arise from combinations of extreme waves and sea levels. For flood risk analysis and the design of coastal structures it is therefore necessary to assess the joint probability of occurrence of these variables. Traditional methods have involved the application of joint probability contours (JPC), defined in terms of extremes of the sea conditions, that can, if applied without correction factors, lead to the underestimation of flood risk and under design of coastal structures. This paper describes the application of a robust multivariate statistical model to analyse extreme offshore waves, wind and sea levels around the coast of England. The approach described here is risk-based in that it seeks to define extremes of response variables directly, rather than the joint extremes of the sea conditions. The output of the statistical model comprises a Monte-Carlo simulation of extreme events. These distributions of extreme events have been transformed from offshore to the nearshore using a statistical emulator of a wave transformation model. The resulting nearshore extreme sea condition distributions have the potential to be applied for a range of purposes. The application is demonstrated using two structures located on the South Coast of England.

Keywords

flooding; flood defences; flood risk analysis, probabilistic modelling

Notations

$\boldsymbol{X}$	Vector of sea condition variables
f_X	Joint density of X
Z	Random variable for the response variable, eg. Overtopping rate
S	Deepwater wave steepness
H_s	Significant wave height
T_m	Mean wave period
$\mathrm{T_e}$	Energy period, used in wave overtopping calculations
Y	$\boldsymbol{X}$ transformed onto Laplace Scales
$\boldsymbol{w}$	Vector of residuals from fitted regression model
a	Vector of parameters from the fitted regression model
b	Vector of parameters from the fitted regression model
v	Threshold used in the fitting of the regression model

1. Introduction

The UK has a long history of coastal flooding (Haigh *et al* 2015). The Winter of 2013/2014 further highlighted the threat to the UK posed by coastal flooding . From December through to February a series of low pressure weather systems crossing the country bought unprecedented levels of disruption, caused by flooding, to large parts of the country. Flooding from multiple sources was in evidence as mainline railway infrastructure was severely damaged and large numbers of properties were inundated. Some properties, notably on the Somerset Levels, experienced inundation lasting over a month.

In response to this flooding the government implemented a series of initiatives, including collation of information relating to the state of flood defences by the military and obtaining improved information on the national level of flood risk. As part of this process it was recognised that improved information relating to the analysis of coastal flood risks was required.

It is well-known that coastal flooding in England arises as a combination of extreme sea levels and wave conditions occurring together and consideration of extremes of their joint likelihood of occurrence is important (Bruun and Tawn, 1998, Hawkes *et al,* 2002, Defra/Environment Agency, 2005). The Environment Agency (EA) has produced a national coastal flood boundary conditions dataset that provides industry with return period estimates of extreme sea levels around the coastline of the UK, (Environment Agency, 2011a). Information relating to extreme wave conditions and their joint likelihood of occurrence with extreme sea levels is however, also required to undertake coastal flood risk analysis and to support the design of coastal structures that protect critical infrastructure, including nuclear facilities.

Traditional approaches to joint probability analysis that use joint probability contouring (JPC) methods are known to have limitations that can result in an underestimation of flood risk, or the under-design of structures (HR Wallingford/Lancaster University, 1998, Hawkes *et al.,* 2002). This study, undertaken for the EA, describes the application of a statistically robust multivariate analysis of offshore extreme waves, sea levels and wind speeds around the coastline of England. The output of the multivariate analysis is a Monte-Carlo simulation of extreme offshore sea conditions. Whilst this analysis has been undertaken offshore, the results have been translated to the nearshore using a combination of a wave transformation model and a statistical emulation method that significantly increases computational efficiency. It is envisaged the resulting outputs from the study can potentially be used for a range of purposes, including national and local-scale flood risk analysis and future climate change impact assessments.

To understand the motivation for the new approach adopted here, the method applied in current practice has been reviewed in Section 2. Section 3 goes on to describe the study area and data sets. The multivariate and wave transformation methods are described in Sections 4 and 5. An example application of the methodology to estimate overtopping rates at a location on the south coast is described in Section 6. Finally, a discussion and conclusions are provided in Sections 6 and 7, respectively.

2. Background and limitations of current practice

2.1. Background

There are two distinct joint probability approaches in widespread use in coastal engineering practice in the UK, (Defra/Environment Agency, 2005). These two approaches comprise a simplified method that involves the use of joint probability contours (JPC) and a robust risk-based statistical method. Both approaches are implemented within the widely-used JOIN-SEA software system (HR Wallingford/Lancaster University, 1998, Hawkes *et al*, 2002).

It is of note however, the practice of deriving and applying JPC's has known limitations (HR Wallingford, 1998, Hawkes *et al.,* 2002).

The JPC method is motivated by the traditional deterministic univariate design framework that specifies structures are to be designed to "withstand a 100 year (for example) design event". Where the design event is defined in terms of the loading or forcing variables. The underpinning quantity of interest in this type of analysis is the probability (or return period) of failure of the structure, defined either through serviceability failure or failure of the ultimate limit state (Melchers, 1999). Unfortunately, this traditional framework does not translate well to the coastal environment, where the hydraulic loading is defined by multiple sea condition variables. There are three main reasons for this. Firstly, there are multiple definitions of return period that can be applied to the case when the loading is multivariate and hence there is ambiguity associated with a multivariate return period. This is in contrast to the univariate case. In addition, within any specified definition of a multivariate return period, there are an infinite number of "design events", lying on a specified contour and hence no unique "design event". Moreover, in general, none of the definitions of multivariate return periods, and associated contours and design events, relate directly to the quantity of interest, the return period associated with failure of the structure or associated flood consequences. These aspects are discussed in more detail below in order to describe the motivation for the approach adopted here.

2.2. Technical description of the JPC method

The JPC approach applied in coastal engineering practice, shown in concept Figure 1, requires the creation of contours of the extreme variables (e.g. waves and sea levels) that have an equal likelihood of simultaneous exceedance. These contours are defined as a locus of (x_1, x_2) that satisfies:

$$\Pr(X_1 > x_1 \cap X_2 > x_2) = p \qquad (1)$$

where X_1 and X_2 are the random variables for sea levels and wave heights respectively and p is a fixed probability. In practice, the contours are derived for probabilities associated with specific return periods of interest. The contours are used for two specific purposes: specification of design events that form the boundary conditions for numerical and physical models for the purposes of structural design; and quantification of return period estimates of overtopping and overflow rates for use in flood mapping and risk analysis. In practice, combinations of the wave and sea level variables that lie on a specific contour are extracted and then applied to the response. This could mean a marginal extreme sea level with a return period of 10 years is combined with a marginal extreme wave height of 5 years and tested as one possible combination, for example. A series of combinations are tested in order to find the "worst case" value of the response variable. This "worst case" value of the response is then assumed to have the same return period as the return period associated with the JPC (Hawkes *et al.,* 2002, HR Wallingford/Lancaster University, 1998). This assumption can be defined explicitly.

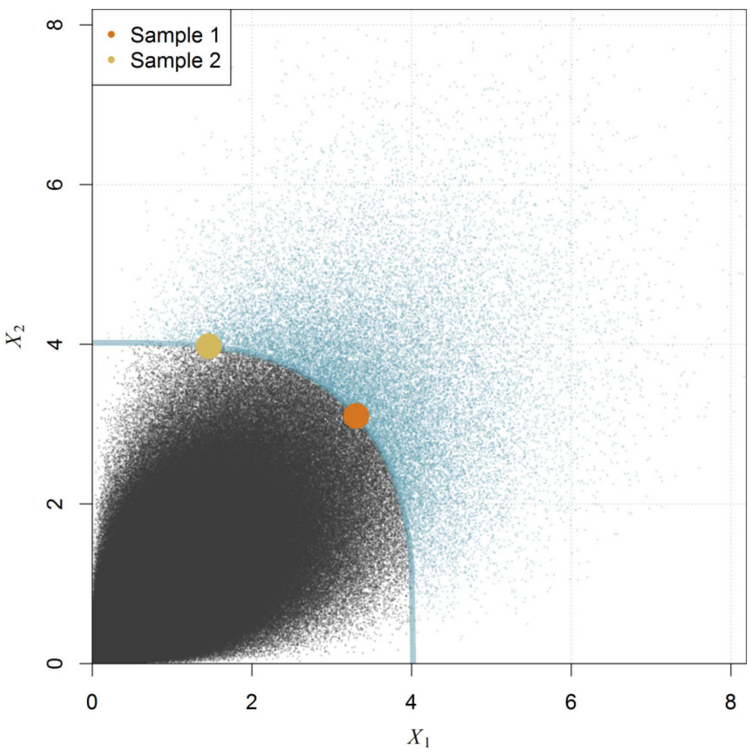


Figure 1: Simulated joint events from a BVN with an associated 100-year joint probability contour (AND definition), with two "design events" highlighted

If Z is the random variable for the response function, it follows that:

$$Z = f(X_1, X_2) \tag{2}$$

The JPC method implements the following false assumption :

$$\Pr(Z > \max\{f(x_{1,i}, x_{2,i}) | i = 1, \dots, n\}) = \Pr(X_1 > x_1 \cap X_2 > x_2) \tag{3}$$

where the $(x_{1,i}, x_{2,i})$ pairs each satisfy Eqn. 1.

2.3. Multivariate return periods

To highlight the limitations of the current JPC method, seven example alternative multivariate contouring definitions have been applied, (Serinaldi, 2014) and include those developed and explored by a range of authors, (Hawkes *et al.,* 2002, Salvadori *et al.,* 2004, De Michele *et al.,* 2007, Salvadori *et al.,* 2011, Jonathan *et al.,* 2013, Corbella & Stretch, 2013, Volpi & Fiori, 2014). Whilst these alternatives are not routinely applied in coastal engineering practice within the UK, they have been introduced here to raise awareness of the limitations of multivariate contouring approaches.

In this convention, (Serinaldi, 2014), the JPC method is known as AND, Eqn. 1. The alternative definitions are OR, PCOND1, PCOND2, PCOND3 and Kendall AND. Their definitions are detailed in Appendix 1.

The alternative multivariate JPC methods have been compared using samples drawn from a simple bivariate normal model (BVN), Figure 1. The commonly used AND probabilities, Eqn. (1) were calculated for each sample using the known function of the data. A contour with an AND return period of 100 years was adopted. Two sample points on this contour were then identified, (i.e. candidate 100-year “design event” points) as shown in Figure 1. The multivariate return periods were then calculated for these two sample points using the alternative definitions and the known function of the data. Table 1 gives the return periods that were obtained for each definition.

The results from Table 1 show the “100-year design event” obtained from the widely applied coastal JPC (AND method), Eqn. 1, can have a return period that ranges between 1 and 5000 years, depending upon which multivariate return period definition is used. An alternative view to the problem is shown in Figure 2, where “100-year” contours have been derived and plotted for the alternative multivariate return period definitions. It is apparent there is little in common between the various different definitions.

Whilst the AND approach is adopted in coastal engineering practice, there are no specific properties that give this definition a meaningful advantage over the others. Hence adoption of the AND method in practice can be considered a somewhat arbitrary choice. This is because it is the expected frequency (or return period) of structural failure or overtopping or flood risk (ie the response variable) that is of most interest. Or, in risk parlance, it is the probability of outcomes, response or consequences that are of most relevance.

Table 1. Comparison of the return period of two 100-year events (AND definition) with alternative JPC definitions

JPC method name	Multivariate return period: years	
AND[a]	100	100
OR	20	4
COND1 (conditional on X_1)	3	25
COND1 (conditional on X_2)	3	1
COND2 (conditional on X_1)	42	1210
COND2 (conditional on X_2)	62	4
COND3 (conditional on X_1)	6	5616
COND3 (conditional on X_2)	10	1
Kendal AND	37	37

[a] *Definition currently applied in practice*

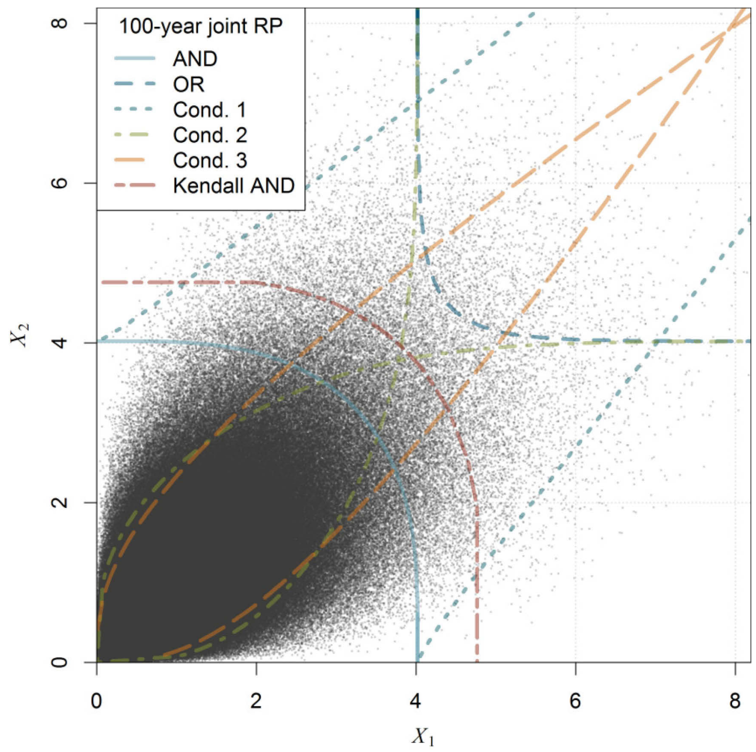


Figure 2: A range of “100-year” joint probability contours obtained from the simulated data

2.4. Practical application of the JPC method

When applying the current JPC method it is important that these limitations are accounted for. With particular regard to the wave overtopping rate response function, evidence on a limited number of tests (Hawkes *et al.,* 2002, HR Wallingford/Lancaster University, 1998), has shown that application of the JPC method can underestimate the probability (and hence overestimate the return period) associated with exceeding a specific overtopping rate leading to potential under design of coastal defences. This systematic error is illustrated in concept in two dimensions in Figure 3. A conceptual contour of a response variable (y) has been defined (labelled “y” in Figure 3) and exceedance of the contour is represented by the dashed shaded region and solid line shaded region. The JPC definition is conceptualised by the right angle region shaded with solid lines. The systematic error can be viewed as the difference between these regions (the dashed shaded region). The JPC method approach also contains limitations with regard to the treatment of other variables, wave period, for example. In practice, the application involves assuming wave period is a direct deterministic function of wave height, often implemented through an offshore steepness equation. In reality, however, there is variability between these parameters and it is desirable to account for this variability. This can be important as it is well-known overtopping rates and damage potential are sensitive to wave period, (Pullen *et al.,* 2009).

The objective of the analysis described here is to overcome the limitations of the JPC approach. This is achieved by enabling extremes of response variables to be determined directly, through the provision of a full multivariate extreme sea condition distribution in the nearshore region around the coast of England.

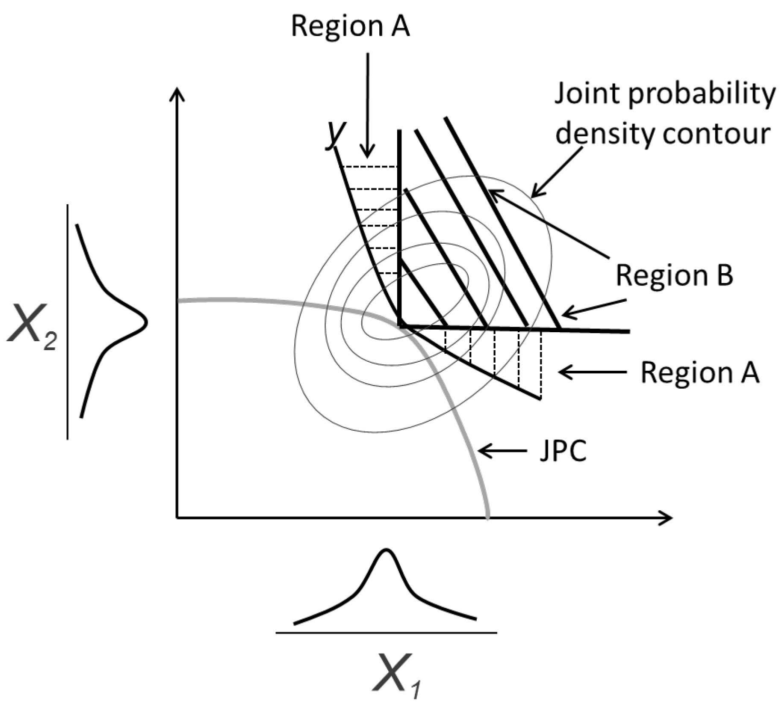


Figure 3: Conceptual diagram showing the systematic error introduced by the joint probability contour method represented by the differences between Regions A and B

3. Study area and data overview

The method adopted on this study comprises two main components:

- Multivariate (joint) probability analysis – offshore wave and wind data has been combined with sea level data and extrapolated to extreme values using a robust statistical method.
- Offshore to nearshore wave transformation –The offshore wave conditions have been translated to nearshore (approximately to the -5mODN sea bed contour), using the SWAN wave transformation model and a statistical emulation method.

To implement the method, the coastline has been sub-divided into 24 different regions, each region comprising a SWAN wave transformation model domain, Figure 4. These regions were defined through consideration of a number of factors, including the exposure and orientation of the coastline and the spatial variability of the wave conditions along the offshore boundary. The data used in the analysis is summarised in Table 2 and detailed below.

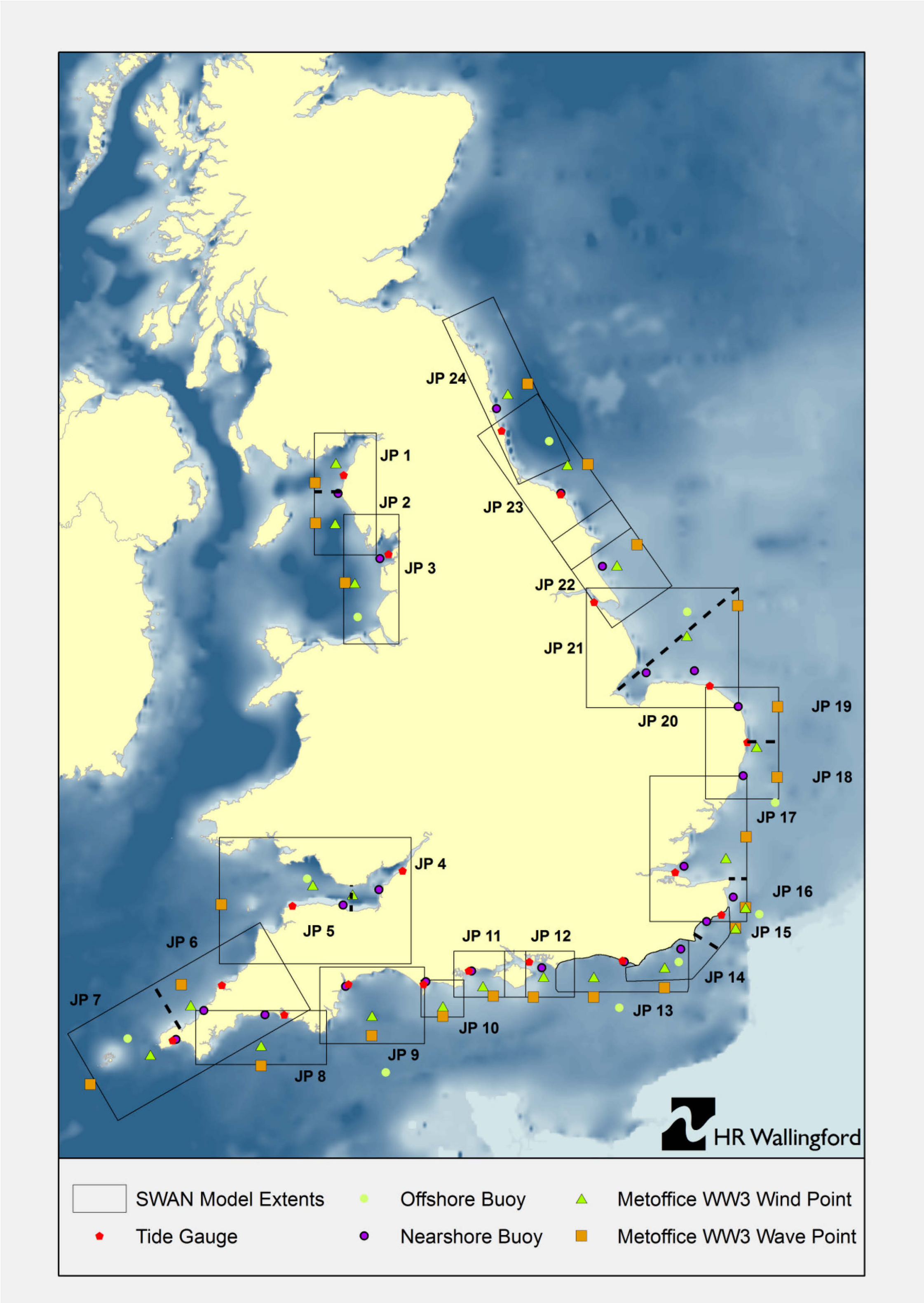


Figure 4: SWAN wave model domains and locations of the associated wave, wind and sea level data

Table 2. Summary of data sources

Variable	Source	Comment
Sea-level time series	NTSLF national class 'A' network of tide gauges supplemented with additional processed EA tide gauge data	Class A tide gauges are supplemented with records from EA tide gauges located at Padstow and Exmouth
Sea-level extremes	Coastal flood boundaries study	EA (2011a) study that provides extreme sea levels at a 2 km resolution around the coast
Wave conditions	WWIII hindcast	Hindcast run by the Met Office model grid is 8km resolution for a timespan from January 1980 to June 2014 so it includes the winter storm of 2013/2014 Some locations have been corrected for bias following comparison with measured data sets (EA, 2011a)
Wind conditions	WWIII hindcast	As wave conditions, but it is of note than no bias corrections have been applied to wind speeds
Bathymetry offshore	SeaZone TruDepth	Approximately 30 m resolution, July 2014 download
Bathymetry nearshore	2 m resolution combination of EA light imaging detection and ranging, multibeams and single beam surveys from channel coastal observatory, known as the EA surf zone composite bathymetry	Compiled by EA Geomatics Group

A separate offshore multivariate extreme value model has been developed at a single offshore WWIII grid point of each of the 24 regions. The wave conditions were imposed along the offshore boundary for the relevant wave directions. To account for the spatial variation in the water levels across the model domain, when undertaking the wave transformation modelling, an algebraic relationship between the water level at a specified control point and the distance from this control point was defined. These relationships were implemented at the nearshore points (approximately the -5mODN contour), during the emulation phase of the wave transformation modelling. The specific relationships and further details are published in an accompanying report, (HR Wallingford, 2015a).

Time series sea level data was obtained from the National Tide and Sea Level Facility (NTSLF) National Class "A" network of tide gauges (Figure 4), owned by the Environment Agency. Prior to implementation within the multivariate analysis, the water level data was de-trended and updated to present day levels, using standard approaches, (HR Wallingford, 2015). Within the statistical model this sea level data was supplemented with information on extreme sea levels at a 2km resolution around the entire coastline, (Environment Agency, 2011a), again updated to the present day.

Wave and wind data was obtained from a hindcast of wave conditions using the WaveWatch III Model (WWIII), undertaken by the UKMO, (Mitchell *et al.,* 2016) . The grid resolution for this model is 8km and the timespan of the hindcast from January 1980 to June 2014 which therefore includes the winter storms of 2013/2014. Data from 1980 to 2000 is available at a 3-hour resolution and from 2000 onwards at a 1-hour resolution. The wave model was driven with wind data from the ECMWF ERA-interim (global) and Unified (regional) models. The hindcast study provided spectral components of waves including H_s, T_m and direction. The locations of the WWIII points where wave and wind information was extracted for the multivariate extreme value analysis for each region is shown in Figure 4. An example plot showing the performance of the WWIII model when compared with measured wave data is shown in Figure 5. Measured wave data from wave buoys was obtained from CEFAS and the Regional Coastal Monitoring Programmes for these purposes. Where appropriate, bias corrections were introduced and applied to the offshore wave conditions. Further results of this analysis are detailed in an accompanying report, (HR Wallingford, 2015a).

The new SWAN models are based upon bathymetry from the SeaZone TruDepth data set which is at a resolution of approximately 30m.

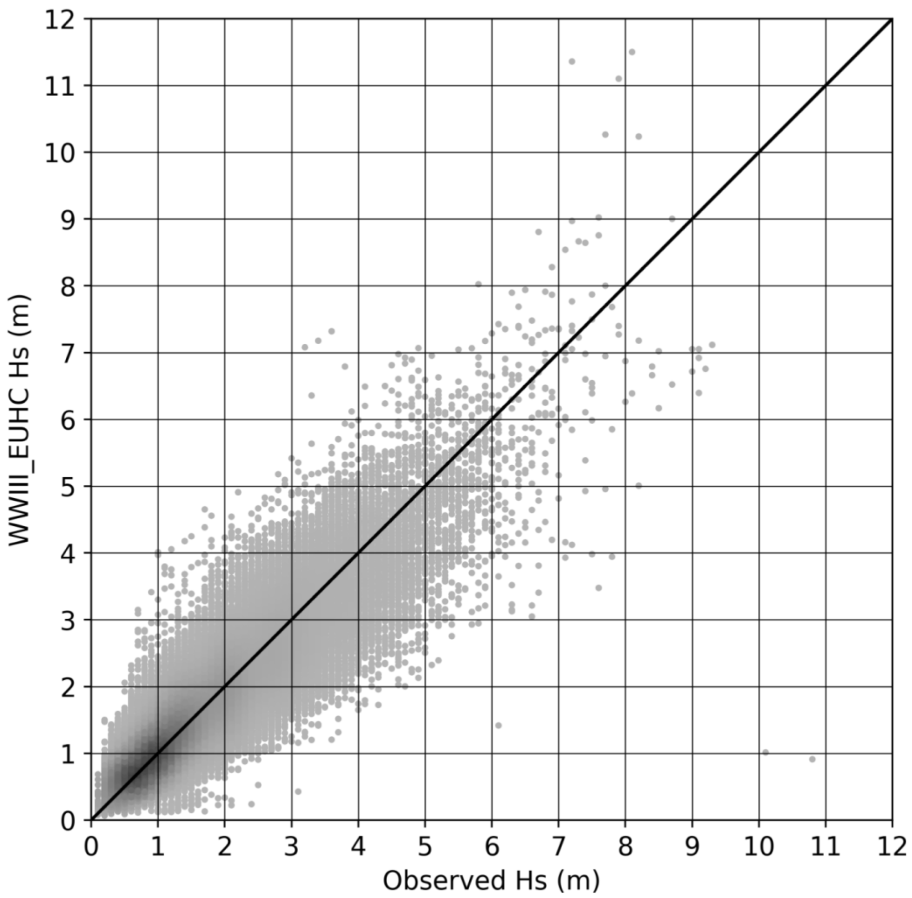


Figure 5: Comparison between the WaveWatch III Hindcast data and Channel Light Vessel measured wave conditions

4. Multivariate extreme value modelling

To quantify the probability of exceeding a particular value of the response variable, it is necessary to extrapolate the joint probability density of the sea condition variables to extremes, whilst preserving the dependence between the variables, and then integrate the joint density over the region where the response variable Z has been exceeded:

$$\Pr(Z > z) = \int_{Z>z} f_X(x)dx \tag{4}$$

This integration is illustrated in concept in Figure 3. This robust statistical approach is facilitated within the JOIN-SEA system and has been implemented in practice by specialists for many years. For this analysis, the same basic method is used (i.e. Eqn. 4 is solved directly). However, an alternative method to that employed by JOIN-SEA has been used to extrapolate the joint density of the sea condition variables to extreme values. The approach adopted for undertaking this extrapolation is more advanced, (Heffernan & Tawn, 2004). This method has greater flexibility in terms of the handling of the dependence structure with a greater number of variables, when compared to the approach adopted within JOIN-SEA and other algebraic copula approaches. Further description on the justification for the use of this model in the context of coastal wave and water level analysis and flood risk modelling is provided by others, (Wyncoll & Gouldby, 2012, Gouldby *et al.,* 2014, Jonathan *et al.*, 2013, Lamb *et al.,* 2012, Keef *et al.*, 2012, Environment Agency, 2011b). A description of its implementation here is provided below.

The variables considered in this analysis were significant wave height, wave period, wave direction, directional spreading, wind speed, wind direction and sea level. Of these, only wave height, wave period, wind speed and sea level required extrapolation to extremes.

The offshore analysis for a single region started with a concurrent time series of the seven variables of interest, created by combining a representative wave and wind series with a de-trended water level series. In order to ensure a good representation of the known strong dependence between wave height H_s and period T_e, the deep water wave steepness was calculated via:

$$S = \frac{2\pi H_s}{g T_e^2} \tag{5}$$

and used in place of the wave period in the multivariate analysis as a non-extreme variable. In contrast to the traditional contour methods, this enables some random variability between wave height and period, rather than assuming wave period is a deterministic function of wave height. This left three variables to be extrapolated to extremes: wave height, wind speed and sea level; with four additional variables whose dependence with these and with each other requiring preservation within the modelling procedure.

Potential peak wave overtopping events were then identified from the joint time series. Since overtopping can potentially be caused by extreme offshore waves, local winds generating waves or sea levels which do not necessarily peak concurrently, the peaks of each of these variables were identified separately to form a separate set of peak events for each extreme variable. Peak waves and winds were identified as local maxima separated by at least a day whereas for water level each high tide was identified. In each case, the peak in the primary variable was paired to the concurrent values for the remaining variables (Figure 6). Each peak dataset was used only to extrapolate extremes of the same primary variable and sampled only when this was most extreme to avoid double-counting the events.

In order to account for known seasonal and directional dependencies, the extreme variables were de-trended with respect to season and direction before their extremes were analysed (Figure 7). The EA has already published industry standard extreme sea levels (Environment Agency, 2011a) and there was a requirement to ensure compatibility with these published results. Since these published results have no dependence on season, this could only be done to the two remaining variables. Wave heights and wind speeds were first de-trended with respect to season followed by wave and wind direction, respectively. For this, a continuous season variable was constructed from the date and time of the peak representing the fraction of the year. In each case, the de-trending was carried out by subtracting a smoothed mean and then dividing by a smoothed standard deviation that were each created from Gaussian Kernel smoothers fitted to the peaks of each variable that account for the periodicity in the smoothing variables. The seasonality was added to the list of non-extreme variables associated with each joint observation so that its dependencies could be modelled to allow variables on the original scale to be reconstructed after simulation.

Prior to analysis of the dependencies between each variable, the marginals of the extreme variables were first analysed. As it was a requirement for the sea levels to follow a predefined distribution based upon the industry standard levels, (Environment Agency, 2011a), these were imposed within the modelling process during specification of the marginal distribution of extreme sea levels. But for wave height and wind speed, the standard peaks-over-threshold (POT) approach of Davison and Smith (1990) was applied, whereby the peaks of each variables that fall above a suitably high threshold were fitted to the Generalised Pareto distribution (GPD). This defined a probability model for large values of X_i. To provide a full specification of the marginal distributions of the extreme variables, the empirical distribution of the X_i values below the threshold were combined with the GPD above the threshold to provide a semi-parametric function for the cumulative marginal distribution (Coles & Tawn, 1991).

The multivariate method (Heffernan & Tawn, 2004) was applied in the standard way. The extreme variables in each peak dataset were therefore transformed from their original scales $\boldsymbol{X}$ to Laplace scales $\boldsymbol{Y}$ using the standard probability integral transformation applied to the marginal distributions of the peaks. The analysis of the dependence between the variables on the transformed scales $\boldsymbol{Y}$ was then undertaken.

If $\boldsymbol{Y}_{-i}$ denotes the vector of all extreme variables excluding a particular variable Y_i, the following multivariate non-linear regression model was applied to $\boldsymbol{Y}_{-i} \mid Y_i > v$:

$$\boldsymbol{Y}_{-i} = \boldsymbol{a}\, Y_i + Y_i^{\boldsymbol{b}}\, \boldsymbol{w} \quad \text{for } Y_i > v \qquad (6)$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors of parameters, v is a specified threshold, $\boldsymbol{w}$ is a vector of residuals. The fitted models each describe the dependence between all remaining variables when a primary peak variable was extreme and together they describe the full distribution of potential peak overtopping events when at least one source variable is extreme. The fitted model was then applied using a Monte-Carlo simulation procedure whereby a large synthetic dataset representing 10,000 years' worth (0.2 million) of potential peak overtopping events was randomly sampled. Rejection sampling was used to ensure Y_i was the most extreme to avoid double-counting events where more than one source was extreme. The simulated dataset was then transformed back to the original scales. The resulting output was a large multivariate sample of extreme offshore sea condition data that captures the characteristics of dependencies between the variables, as well as preserving the marginal extremes, see Figure 8.

To quantify the probability of exceeding a value of the response variable, Z (and hence solve Eqn. 4), it is, in principle, necessary to evaluate the response function for each Monte-Carlo realisation output from the multivariate analysis. Given the chain of models involved, this can be computationally demanding and hence an efficient statistical model emulation procedure was implemented to facilitate this process. This is described in more detail in Section 5.

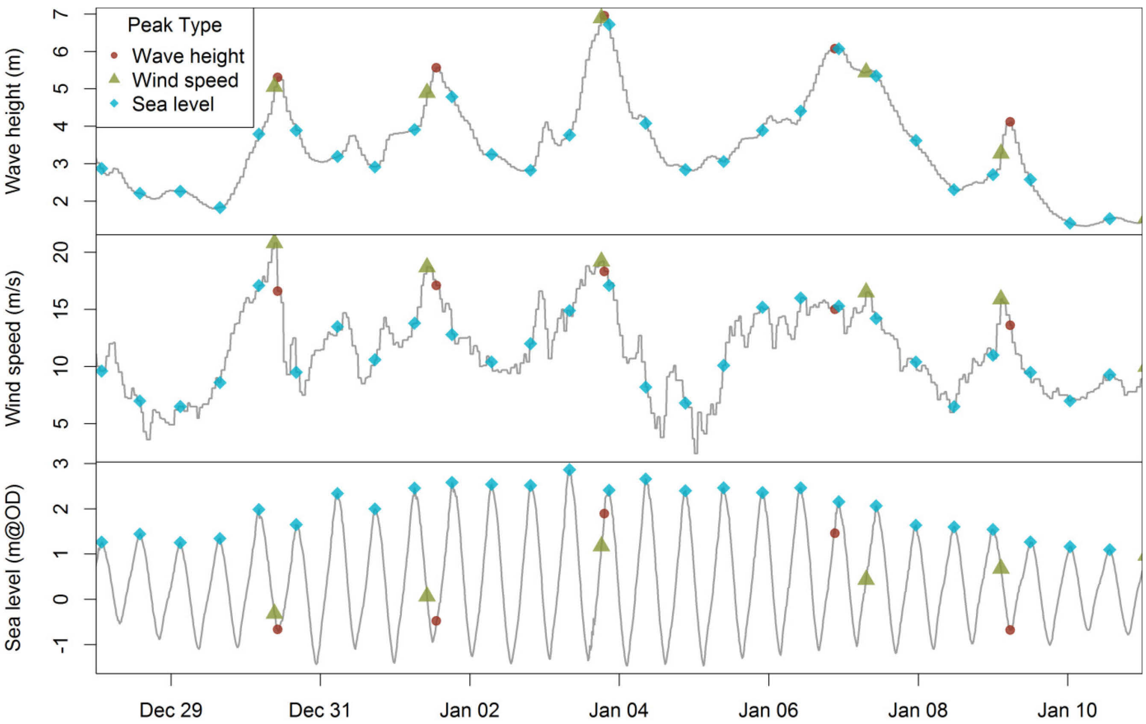


Figure 6: Illustration of the method used to select events

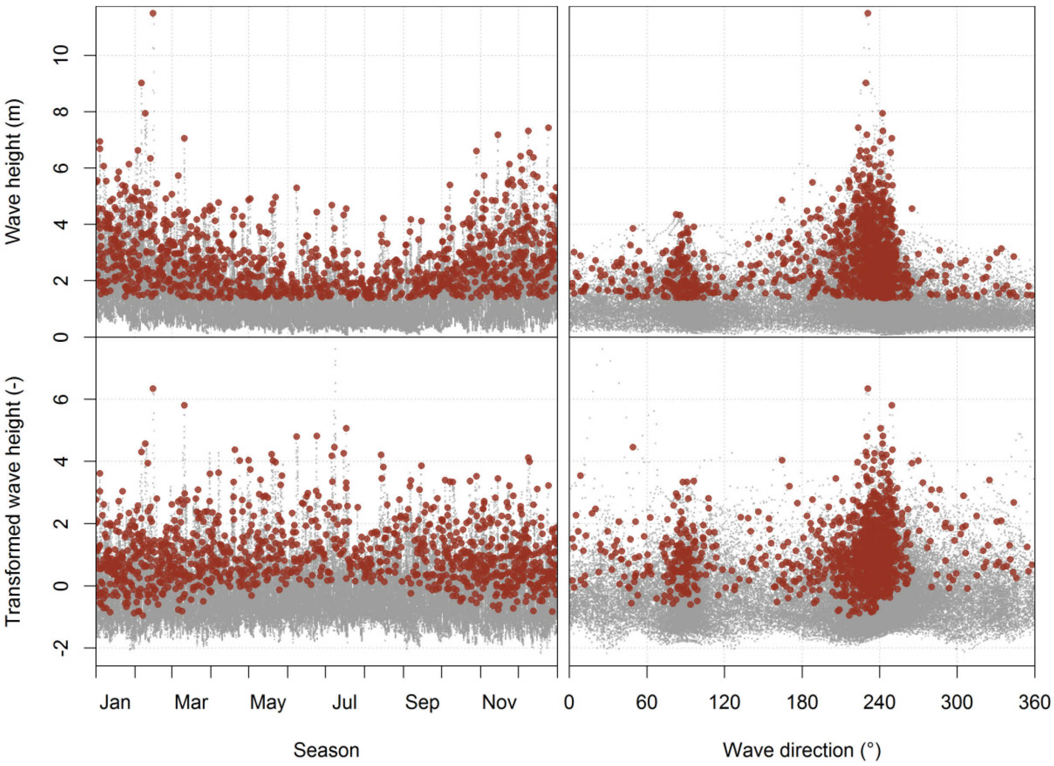


Figure 7: Illustration of the method used to de-trend peak wave height events with respect to wave direction and season.

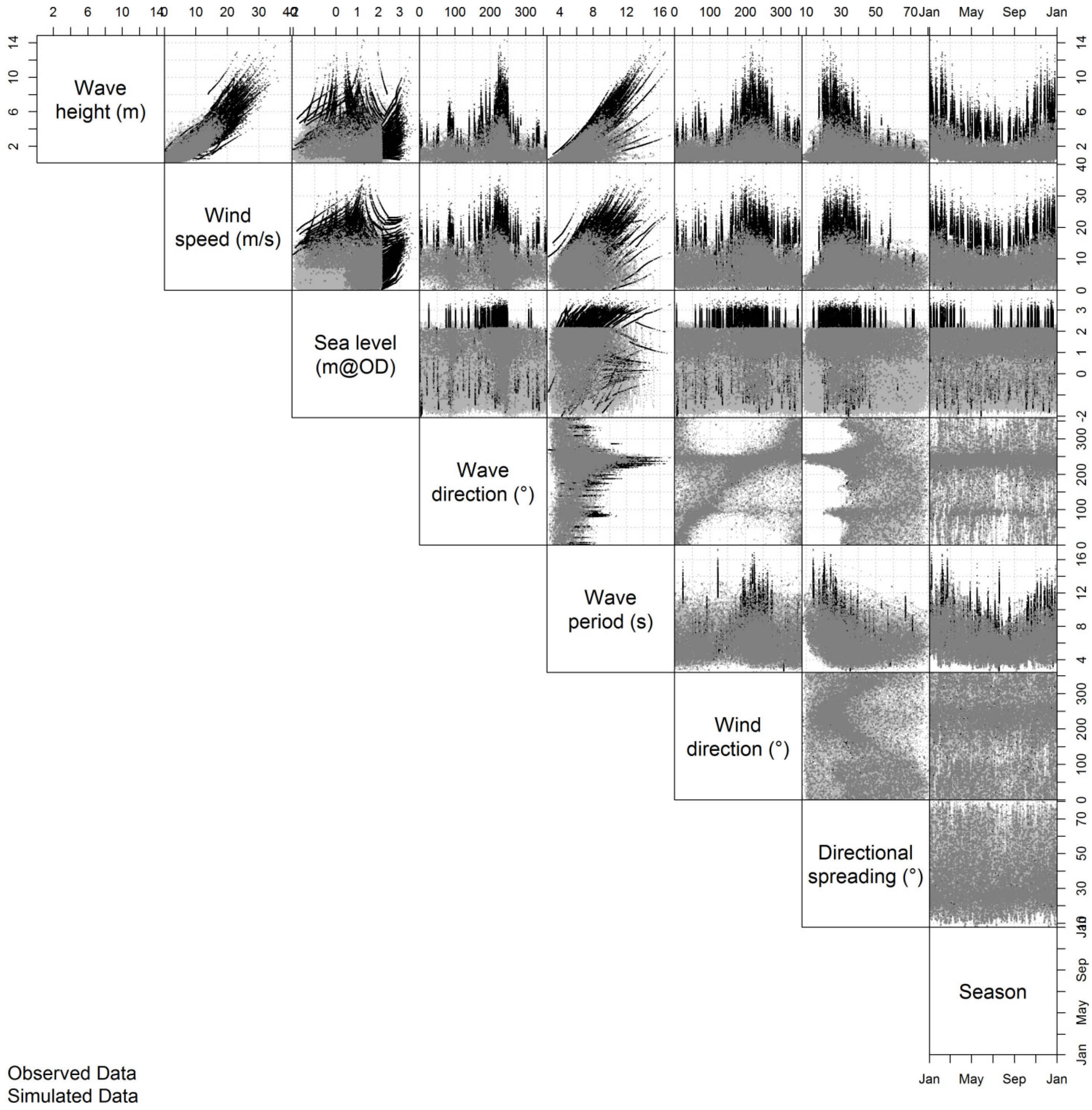


Figure 8: Example Monte-Carlo simulation output from the multivariate extreme value model

5. Offshore to nearshore wave transformation modelling

5.1. Wave transformation model set up and calibration

The model chosen to transform the wave conditions from offshore to inshore was the well-known SWAN model, (Booij, 1999). The objective of the SWAN modelling was to transform the offshore multivariate extreme sea condition data to a series of nearshore locations with a 1km spacing along the coastline, at approximately the -5mODN contour. This depth was chosen, as in shallower water (the surf zone), wave breaking increases and the bed levels vary significantly. It was thus desirable to separate out the complex, highly dynamic surf zone and structure related aspects from the more stable regime in deeper water. This de-coupling enables flexibility with regard to the surf zone models. The surf zone modelling can easily be repeated when new beach level information becomes available, for example.

Figure 4 shows the SWAN model extents used for the study. For reasons relating to numerical stability and runtime efficiency, the SWAN models were set up using a 200m regular mesh.

All the SWAN models were set up in a stationary mode with a constant wind direction and speed applied. Each of the new models was calibrated using a range of different events selected based upon analysis of historical peak events Details of the calibration methodology and results are provided in an accompanying report, (HR Wallingford, 2015).

5.2. SWAN Emulation

In principle it is necessary to transform all of the events output from the offshore Monte-Carlo simulation (Figure 8) through to the nearshore. However, SWAN can be computationally time consuming to run, particularly given the number of SWAN models and the number of events that required simulation in each model (of the order of 200,000 or more per region). Rather than attempt to run SWAN 2D for all of these events a statistical model emulation method was therefore employed.

A statistical emulator is similar in concept to a traditional “look-up table” approach used in coastal flood forecasting systems, for example. The process involves running the SWAN 2D model for a subset of events (known as the design points). Interpolation techniques are then applied to predict the results for other events (not run in SWAN 2D). Traditional look-up table approaches are typically applied using regular or recti-linear grids and linear interpolation techniques. As the output from SWAN is generally not a linear function of the inputs, these traditional look-up tables can be inefficient and require a large number of design point simulations. There has, however, been extensive research into more sophisticated interpolation techniques, in particular Gaussian Process Emulators (GPE’s), Kennedy *et al.,* (2006), for example. These more sophisticated approaches have been shown to be efficient when used in the context of wave transformation modelling, Camus *et al.* (2011a). Figure 9 shows the computational efficiency gains that are possible when comparing an emulator to a traditional “look-up table” of the SWAN wave model. Within Figure 9, the same root mean squared error (RMSE) was achievable with 200 separate event simulations of the SWAN model using a GPE when compared to 17000 simulations using a traditional look-up table approach. To obtain these RMSE statistics the SWAN model was run for the full data set to establish the benchmark.

To select the design points used to fit the emulator and hence used to define the boundary conditions for the SWAN model, the Maximum Dissimilarity Algorithm (MDA) was applied using a previously established methodology, (Camus *et al.*, 2011a, 2011b). The use of the MDA ensures the multivariate parameter space is captured efficiently. Figure 10 shows an example of the design points output from the MDA (larger dots) in relation to the full space covered by the Monte-Carlo realisations.

The emulator approach was used to translate the large sample of offshore Monte-Carlo events through to the nearshore wave points located on (approximately) the -5mODN contour. A series of separate emulators was created for each nearshore wave point (1km resolution). A more detailed description of the emulator approach is provided in Appendix 2. Example output from the emulator at a nearshore point is shown in Figure 11.

Verification of the wave transformation modelling process was undertaken by comparing nearshore wave conditions with measurements at selected locations, using the Regional Coastal Monitoring Programme wave buoys. An example of this process is shown in Figure 12 It is of note that within Figure 12, departures from the one-to-one prediction line can arise as a result of uncertainties relating to the offshore hindcast WaveWatch III data (Figure 5), model structural uncertainties associated with SWAN and model structural uncertainties associated with the emulator. Detailed analysis that considers the multiple sources of uncertainty and how these propagate through the modelling chain has, however, not been undertaken. It is recommended that this type of analysis is conducted in future research.

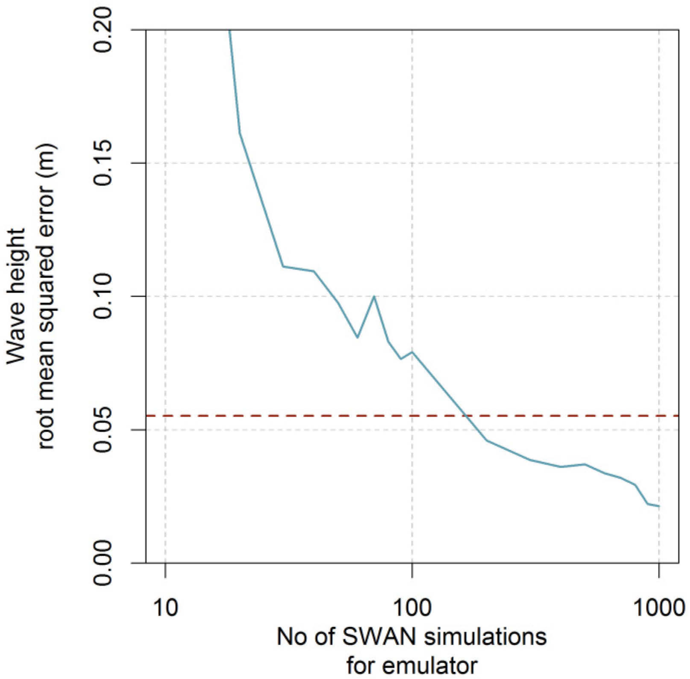


Figure 9: Comparison of the performance of a Gaussian process emulator fitted to varying numbers of SWAN simulations (solid line) with a traditional “look-up” table approach based on 17,000 simulations (dashed line)

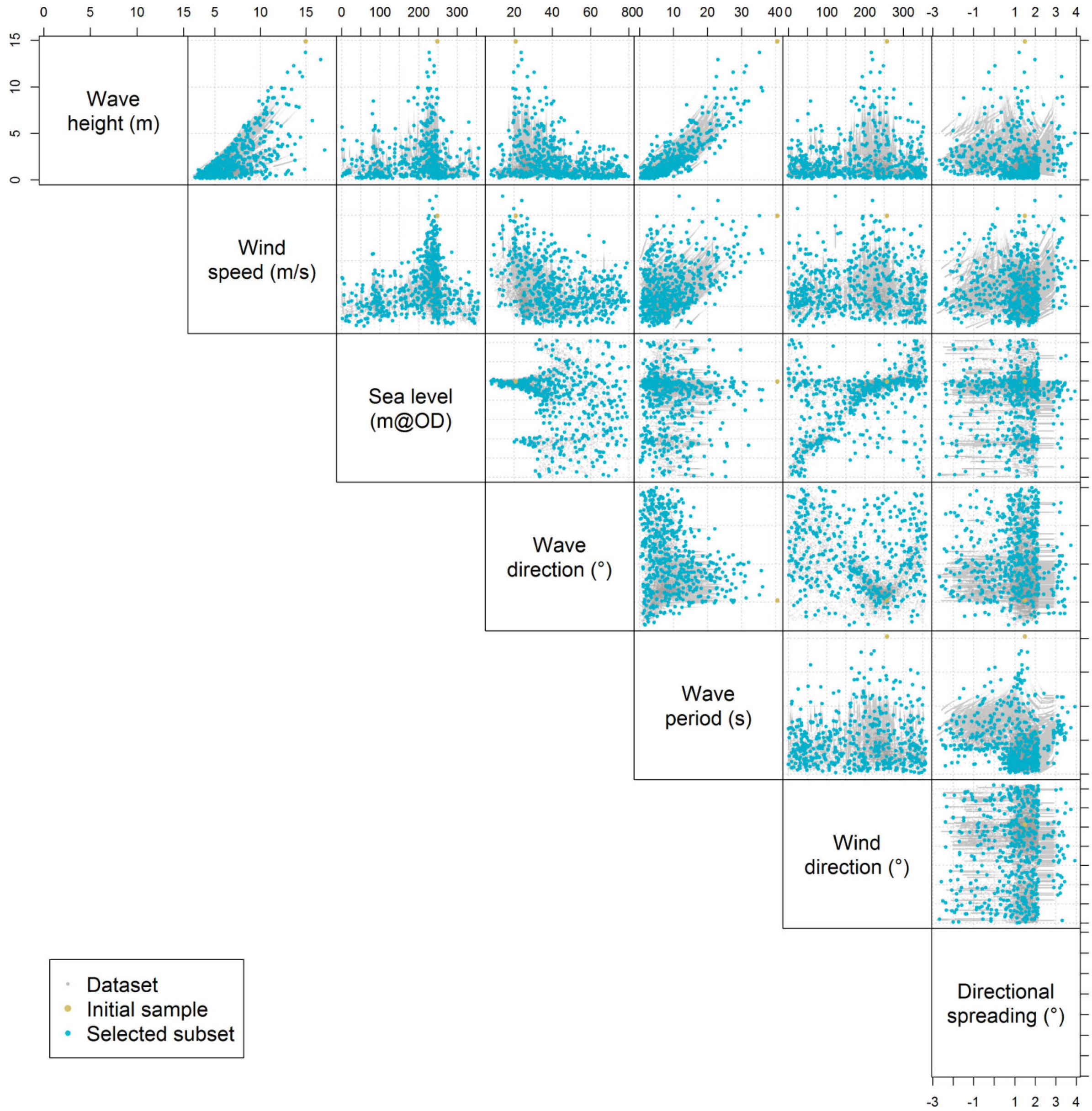


Figure 10: The emulator design points (larger dots), output from the MDA overlaid on the extrapolated MC samples

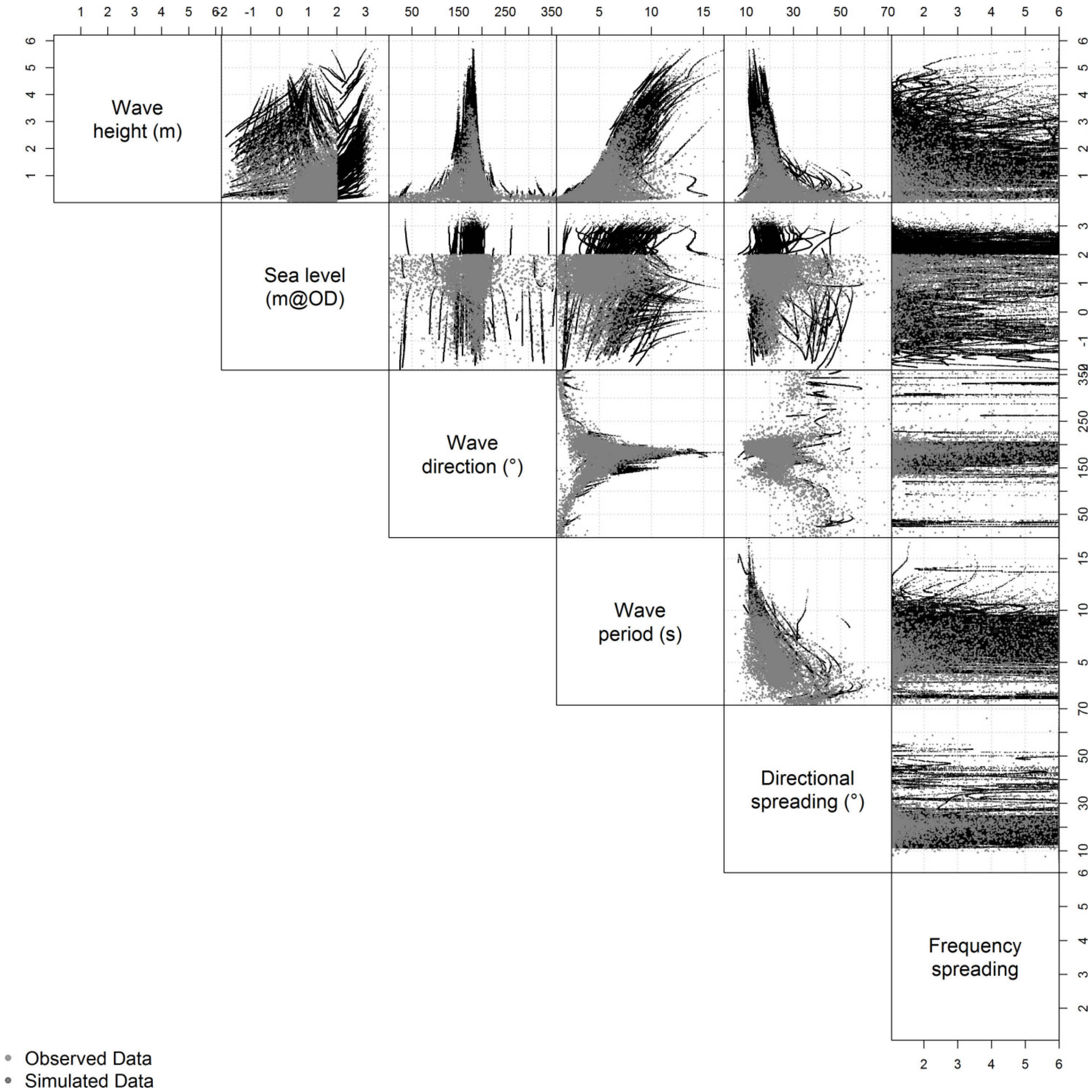


Figure 11: Example output from the SWAN emulator at a specific nearshore point

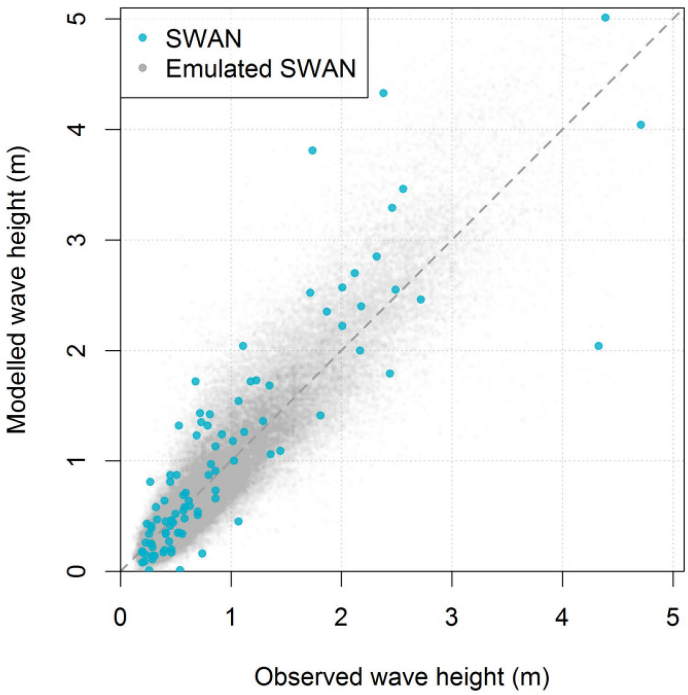


Figure 12: Comparison of the output from the wave transformation process and nearshore measured wave data

6. Example application

6.1. Case study application

To demonstrate the application of the method to estimate wave overtopping rates, two particular structures have been selected from the EA's Asset Information Management System (AIMS). Both sites are located in the South west. The first comprises a smoothly sloping stone wall with vertical upstand, located in Exmouth (Figure 13). The second is a shingle beach fronting a promenade in Lyme Regis (Figure 14).

A wave overtopping response function has been applied. The overtopping method requires sea conditions at the toe of the structure. It is thus necessary to transform the sea conditions from the nearshore to the structure toe. There are a range of methods that can be applied to undertake this transformation, Goda, (2000) and Battjes & Jansen (1978). This latter method is used for the calculation of surf zone breaking in both SWAN 1D and 2D and SWAN 1D and was applied for this study. The profile information was extracted using an automated routine applied to the Surf zone DEM composite compiled by the EA. This formed the bathymetric input to the SWAN 1D Model (Figure 15) and enabled the translation of the nearshore wave conditions through to the structure toe.

To translate the data at the toe of the flood defence structures to overtopping discharges for use in flood inundation analysis, the BAYONET wave overtopping model was applied. BAYONET Kingston *et al* (2008) is a neural network overtopping tool. It is based on the widely used CLASH overtopping database and follows the general model of the CLASH neural network, van Gent *et al.* (2007) but incorporates additional information relating to uncertainty. The Monte-Carlo realisations at the nearshore point have been transformed through SWAN 1D and BAYONET into peak overtopping rates. The overtopping rate samples were then ranked and empirical return periods assigned. These results have then been analysed to determine an empirical distribution of overtopping rates as shown in Figure 16. Given site-specific structural geometry, it is straightforward to undertake this analysis for any structure along the coastline of England.

6.2. Comparison of the robust statistical approach with the joint probability contours

To demonstrate the difference between the routinely applied JPC approach and the robust statistical method applied here, a series of comparisons have been made. A set of '100-year' JPC conditions was extracted that satisfy the relevant criteria (Equation 1) (large points on Figure 16). These were applied to the same wave transformation and overtopping process, and the resulting outputs, in terms of wave overtopping rate, are shown for two defences in Figure 16. In these particular examples, the 'worst case 100-year design event', defined in terms of the JPC approach, without any correction factors having been applied, provides estimates of overtopping rates that actually have return periods of 40 and 16 years, respectively. This demonstrates the limitations of the routinely applied AND JPC approach and highlights the importance of applying correction factors should this approach be adopted.

7. Discussion of results

Extreme multivariate data sets have been generated around the coast of England at a 1km resolution. The methodology applied has significant advantages over the JPC method that is routinely applied in practice for detailed site-specific flood risk analysis and the design of coastal structures. It is particularly important to note probabilities of failure are expressed directly in terms of the response variable/s of interest. This is in contrast to the traditional approach that expresses probabilities of extreme sea conditions, which suffers from ambiguity and does not directly relate to the response of interest.

The outputs from this study can be applied within a wide range of studies including, structural design, flood risk analysis and related climate change impact assessment, for example. It is relatively straightforward to translate the nearshore (-5m contour) extreme event data through the surf zone and into overtopping rates for a range of nearshore structures and hence to determine statistically robust return period overtopping estimates around the entire coastline.

In addition, it is apparent the SWAN emulation approach, due to its computational efficiency, has the potential to be incorporated within coastal flood forecasting systems. Ensemble and probabilistic forecasting is readily achievable with this type of approach.

Specific combinations of wave height, period and water level conditions may be required for deterministic "design event" purposes. It is a trivial process, having determined the empirical return period estimates of the response function of interest using the robust statistical approach, to extract specific combinations of the sea conditions that do actually yield 100-year estimates of the required response function, be this overtopping rate or any other relevant structural response function. It is suggested that it is preferable to adopt design events whose return period is defined in terms of the response function of interest, rather than applying the joint probability contour approach with associated error correction factors. This is because there is no ambiguity associated with the univariate response function return period and it is defined in terms of the variable of interest. The analysis undertaken here can potentially be used to support this type of approach.

As with any coastal modelling analysis of this type, there are substantial uncertainties associated with the multivariate extreme value analysis undertaken here. For example, the offshore wave and wind condition data has been derived from a model hindcast of the WWIII model. An insight to uncertainties associated with this data set can be gained through comparison with offshore measurement data (Figure 5).

For flood risk analysis, it is necessary to extrapolate historical data to extreme values. It is well-known that significant uncertainties can be introduced at this stage. Standard methods for assessing confidence in the extremes of each (marginal) variable are available that take account of the variability of the data and the length of record being analysed. For example, the widely applied guidance on extreme sea levels (Environment Agency, 2011a) gives confidence intervals for the extrapolated extreme sea levels at each location around the coast. For multivariate extreme value analysis it is necessary to capture the uncertainties associated with the marginal extremes and the dependence model.

The SWAN model has been used to transfer the offshore waves to the nearshore. Where nearshore wave measurements exist, these models have been calibrated to reduce the uncertainties. Additional uncertainty is introduced by the use of the emulator. Combined uncertainties from the offshore boundary condition waves, SWAN model and emulator are illustrated in Figure 12. Overtopping formulae are known to contain substantial uncertainty (Kingston *et al.,* 2008) and are sensitive to input data relating to toe levels and crest levels, for example.

In principle it is desirable to quantify uncertainties associated with the various model components and appropriately communicate this uncertainty and further research into these aspects is recommended. Given the many different sources of uncertainty and complexity involved in defining them and propagating them through the modelling chain, it has not been possible to undertake this analysis on the study described here. Further work to quantify uncertainties is, however, desirable. In addition, extending this to formal sensitivity analysis using well used methods (Saltelli *et al.,* 2004), can aid insights into the dominant sources of uncertainty and where future priorities lie for model and data improvements.

Figure 13: Exmouth

Figure 14: Lyme Regis

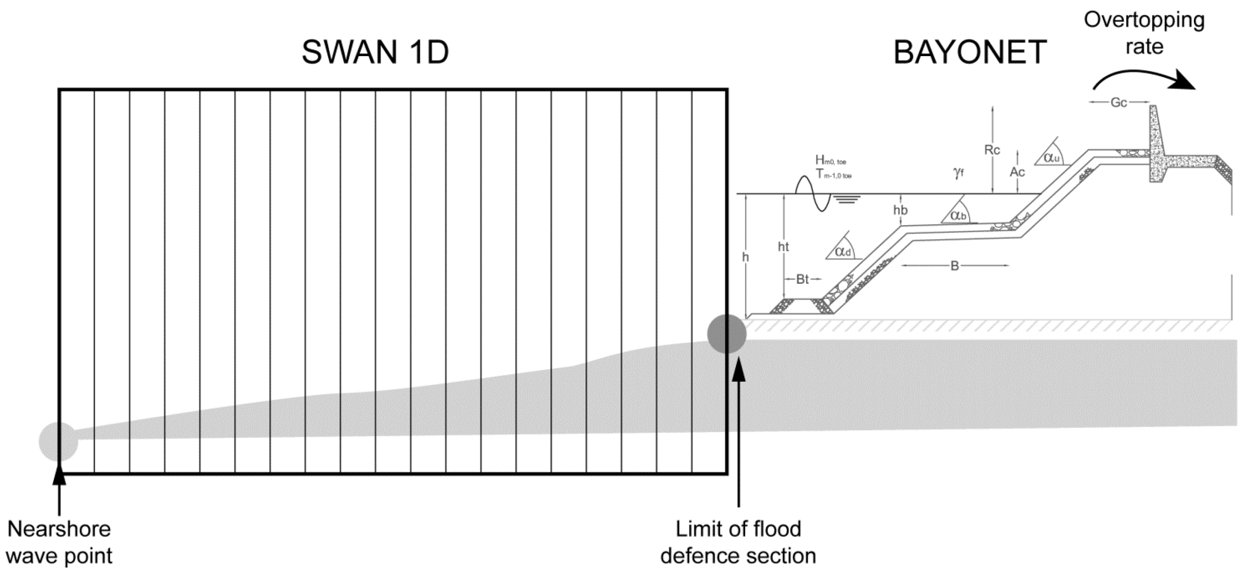
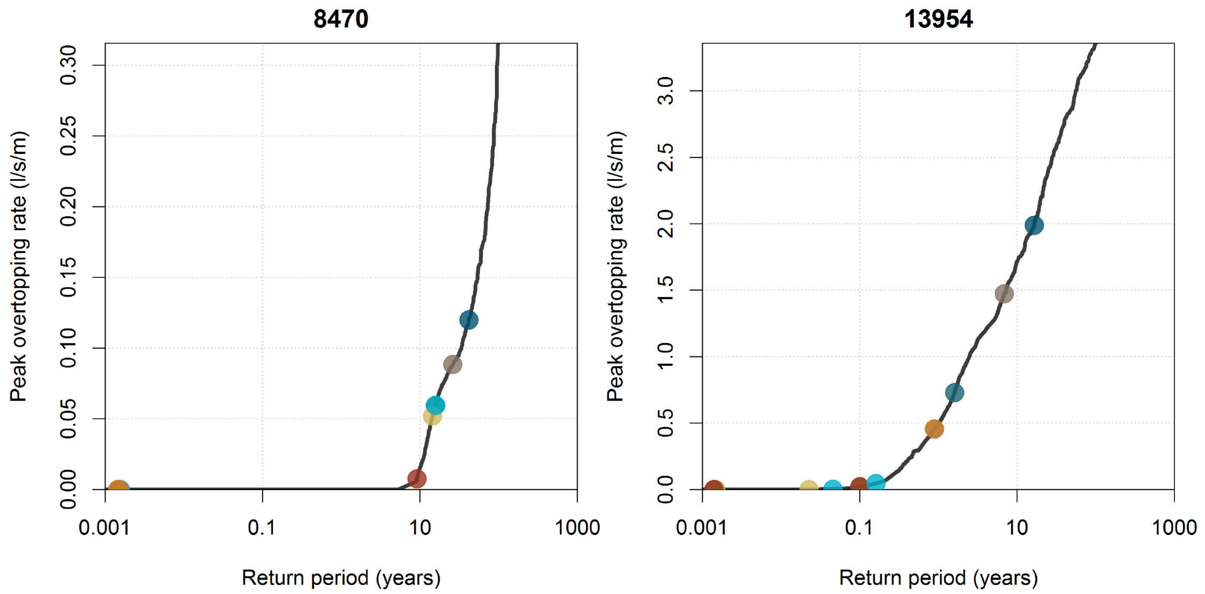


Figure 15: SWAN 1D profile example



	Defence number 8470		Defence number 13954	
	Return period: years	Overtopping rate: (l/s)/m	Return period: years	Overtopping rate: (l/s)/m
Uncorrected JPC (100-years JPC)	40	0.12	16	2
Robust risk based	100	0.315	100	3.4

Figure 16: Results from the traditional Joint exceedance contour method (100-year joint probability) overlaid on the empirical distribution of wave overtopping rates from the multivariate simulation, highlighting the known error

8. Conclusions

The limitations of the JPC approach, widely used in practice for coastal flood risk analysis and the design of coastal structures, have been described. These limitations have been demonstrated with reference to alternative contouring methods using a conceptual data set. The results have highlighted the ambiguity and potential confusion that can arise using the traditional JPC approach.

A multivariate extreme value analysis of offshore waves, winds and sea levels has then been undertaken around the coast of England**.** A robust statistical method has been applied at 24 different locations. The output of the multivariate extremes analysis comprises a Monte-Carlo sample of approximately 0.2 million events. To robustly estimate the return period of response variables like overtopping rate, or economic damage for flood risk analysis, for example, it is, in principle, necessary to transform all of these events through to the nearshore and then to overtopping and subsequent flood inundation.

To undertake the offshore to nearshore wave transformation 24 separate SWAN wave models have been set up. To minimise the computational effort involved in transforming all of the Monte-Carlo events through the SWAN models, a series of emulators have been developed that provide nearshore outputs at a 1km resolution approximately along the -5mODN contour. A nearshore data set comprising many thousands of extreme wave and water level events has therefore been created. To demonstrate the application of this data set, wave overtopping rates have been calculated for two structures located on the South Coast. These results have been compared with the results obtained by applying the widely applied JPC method where the known limitations have been observed.

It is suggested the methodology adopted here can be applied for the purposes of traditional deterministic design and optimised risk based design.

Appendix 1. Alternative definitions of multivariate events

A number of methods for defining joint events have been summarised (Serinaldi, 2014) . These methods have been used to provide a range of multivariate return period estimates for the same event, Table 1. The definition of each method is provided below, adopting an existing naming convention (Serinaldi, 2014).

$p_{\text{AND}} = \Pr(X_1 > x_1 \cap X_2 > x_2)$

$p_{\text{OR}} = \Pr(X_1 > x_1 \cup X_2 > x_2)$

$p_{\text{COND1}} = \Pr(X_1 > x_1 | X_2 > x_2)$

$p_{\text{COND2}} = \Pr(X_1 > x_1 | X_2 \leq x_2)$

$p_{\text{COND3}} = \Pr(X_1 > x_1 | X_2 = x_2)$

$p_{\text{Kendal AND}} = \Pr\,[p_{\text{AND}}(X_1, X_2) > p_{\text{AND}}(x_1, x_2)]$

where $p_{\text{AND}}(x_1, x_2) = \Pr(X_1 > x_1 \cap X_2 > x_2)$.

Appendix 2. Gaussian process emulator description

Gaussian process emulators have been used to predict the SWAN model output at each nearshore point for every event in the large offshore simulated dataset. A separate emulator has been applied for each nearshore point and for each output variable of interest. Let $z = f(\boldsymbol{x})$ represent the SWAN-2D output for a single nearshore variable as a function of the offshore variables $\boldsymbol{x} = (x_1, \dots, x_p)$. The Gaussian process emulator approximates the unknown function $f(\boldsymbol{x})$ by treating it as a random Gaussian process. The statistical model is defined by a mean function $\boldsymbol{h}(\boldsymbol{x})$ satisfying

$$\mathrm{E}[f(\boldsymbol{x})|\boldsymbol{\beta}] = \boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{\beta}$$

and a covariance function $c(\boldsymbol{x}, \boldsymbol{x}')$ satisfying

$$\mathrm{Cov}[f(\boldsymbol{x}), f(\boldsymbol{x}')|\sigma^2] = \sigma^2 c(\boldsymbol{x}, \boldsymbol{x}')$$

where $\boldsymbol{\beta}$ and σ^2 are parameters. The mean function is typically taken to be a linear function of the input variables. For this analysis, the Gaussian covariance function is used which is defined by

$$\mathrm{c}(\boldsymbol{x}, \boldsymbol{x}') = \exp\left[-\textstyle\sum_{i=1}^{p} \theta_i (x_i - x_i')^2\right]$$

for smoothing parameters θ_i.

A Bayesian formulation has been used to estimate the function output probabilistically given the n known outputs at the design points. For a simulated offshore event $\boldsymbol{x}$, the best estimate of $f(\boldsymbol{x})$ in light of the known model outputs is given by:

$$\boldsymbol{h}(\boldsymbol{x})^T\boldsymbol{\beta} + \boldsymbol{t}(\boldsymbol{x})^T A^{-1}(\boldsymbol{z} - H\boldsymbol{\beta})$$

where

$$\boldsymbol{t}(\boldsymbol{x})^T = [\mathrm{c}(\boldsymbol{x}, \boldsymbol{x}_1), \dots, \mathrm{c}(\boldsymbol{x}, \boldsymbol{x}_n)]$$

$$A = \left[\mathrm{c}\left(\boldsymbol{x}_i, \boldsymbol{x}_j\right)\right]_{i,j=1}^{n},$$

$$\boldsymbol{z}^T = [f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n)],$$

$$H^T = [\boldsymbol{h}(\boldsymbol{x}_1), \dots, \boldsymbol{h}(\boldsymbol{x}_n)].$$

If applied to one of the design events $\boldsymbol{x_i}$, this formula returns the known SWAN output with zero error. The prediction equation is applied for every offshore event to produce a large simulated dataset of nearshore events.

References

Battjes,J . A., and J.P. F. M. Janssen, 1978. Energy loss and set-up due to breaking of random waves, Proc. 16th International Conference on Coastal Engineering, p, p. 569-587, Am. Soc.of Civ. Eng.,New York.

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.

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): 4.

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.

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., 2013. Simulating a multivariate sea storm using Archimedean copulas, Coastal Engineering, Vol. 76, Pg. 68–78, June.

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.

Defra/Environment Agency, 2005. Use of Joint Probability Methods in Flood Management A Guide to Best Practice R&D Technical Report FD2308/TR2.

De Michele, C., Salvadori, G., Passoni, G. and Vezzoli, R., 2007. A multivariate model of sea storm using copulas, Coast. Eng. Vol. 54, Issue 10, Pgs 734–751, Oct.

Environment Agency, 2011a. Coastal flood boundary conditions for UK mainland and islands Project: SC060064/TR2: Design sea levels.

Environment Agency, 2011b. The risk of widespread flooding – Capturing spatial patterns in flood risk from rivers and coasts, SC060088/R1 Spatial Coherence of Flood Risk – Technical Methodology Report.

Goda Y, 2000. Random seas and the Design of Maritime Structures, World Scientific Publishing, ISBN 981-02-3256-X.

Gouldby B, Mendez F, Guanche Y, Rueda A and Minguez R, 2014. A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis, Coast. Eng, vol. 88, June.

Haigh I.D., Wadey M.P., Gallop S.L., Loehr H., Nicholls R.J., Horsburgh K., Brown J. M. and Bradshaw E., 2015. A user-friendly database of 100 years (1915-2014) of coastal flooding in the UK, Scientific Data (2), Article number: 1500 doi:10.1038.

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.

HR Wallingford/Lancaster University, 1998. The joint probability of waves and water levels: JOIN-SEA, A rigorous but practical new approach, HR Wallingford Report SR 537.

HR Wallingford. 2015. State of the Nation: Coastal Boundary Conditions Report for the Environment Agency. Report number 30, reference MCR5289-30-R00-01.

HR Wallingford (2015a) State of the Nation: Assessment of the Met Office WaveWatch III hindcast wave data, Report Number RT013 MCR5289-13-R004.

HR Wallingford (2015b) State of the Nation: SWAN 2D wave calibration report, Report Number RT02, MCR5289-R-02-00.

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.storms using copulas. Coastal Engineering, 54(10): 734-751.

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

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.

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

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.

Melchers R (1999). *Structural Reliability Analysis and Prediction*, 2nd edn. Wiley, New York.

Mitchell, J.A., Bett, P.E, Hanlon, H.M. and A. Saulter, 2016. Investigating the impact of climate change on the UK wave power climate. Meteorologische Zeitschrift, http://dx.doi.org/10.1127/metz/2016/0757

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

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. 2004. Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. John Wiley & Sons, Ltd., Chichester, England

Salvadori, G., and C. De Michele (2004), Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, Water Resour. Res., 40, W12511, doi:10.1029/2004WR003133.

Salvadori, G., De Michele, C., and Durante, F. 2011 On the return period and design in a multivariate framework, Hydrol. Earth Syst. Sci., 15, 3293-3305.

Serinaldi, 2014 Dismissing Return Periods, Stochastic Environmental Research and Risk Assessment, Volume 29, Issue 4, pp 1179-1189, May.

Van Gent, M.R.A., Van den Boogaard, H.F.P., Pozueta, B. & Medina, J.R. 2007. Neural network modelling of wave overtopping at coastal structures. Coastal Engineering 54(8): 586-593.

Volpi, E., and A. Fiori (2014), Hydraulic structures subject to bivariate hydrological loads: Return period, design, and risk assessment, Water Resour. Res., 50, 885–897, doi:10.1002/2013WR014214.

Wyncoll and Gouldby (2014) Application of a multivariate extreme value method to flood risk analysis, J. Flood Risk Man. DOI: 10.1111/jfr3.12069.