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

Ben Gouldby BSc (Hons)

Chief Technical Director, Flood Management Group, HR Wallingford Ltd, Wallingford, UK (corresponding author: b.gouldby@hrwallingford.com)

David Wyncoll PhD

Senior Scientist, Flood Management Group, HR Wallingford Ltd, Wallingford, UK

Mike Panzeri BSc (Hons)

Technical Director, Flood Management Group, HR Wallingford Ltd, Wallingford, UK

Mark Franklin BSc (Hons), FRMets, MCIWEM, C.WEM

Project Executive, Environment Agency, Bristol, UK

Tim Hunt BEng, CEng, MICE

Modelling and Forecasting Advisor, Environment Agency, Bristol, UK

Dominic Hames PhD

Principal Scientist, Flood Management Group, HR Wallingford Ltd, Wallingford, UK

Nigel Tozer BSc (Hons), MSc

Principal Scientist, Hydrodynamics and Metocean Group, HR Wallingford Ltd, Wallingford, UK

Peter Hawkes PhD

Principal Engineer, Hydrodynamics and Metocean Group, HR Wallingford Ltd, Wallingford, UK

Uwe Dornbusch PhD, CGeog (geomorph), FRGS

Senior Specialist – Coast, Environment Agency, Bristol, UK

Tim Pullen PhD

Principal Engineer, Coastal Structures Group, HR Wallingford Ltd, Wallingford, UK

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 the occurrence of these variables. Traditional methods have involved the application of joint probability contours, defined in terms of extremes of 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 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 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.

Notation

$\mathbf{a}$	vector of parameters from the fitted regression model
$\mathbf{b}$	vector of parameters from the fitted regression model
f_X	joint density of X
g	gravitational acceleration constant
H_s	significant wave height
S	deepwater wave steepness
T_e	energy period, used in wave overtopping calculations
T_m	mean wave period
v	threshold used in fitting of the regression model
$\mathbf{w}$	vector of residuals from fitted regression model
X	vector of sea condition variables
Y	X transformed onto Laplace scale
Z	random variable for the response variable – for example, overtopping rate

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 brought unprecedented levels of disruption, caused by flooding, to large parts of the country. Flooding from multiple sources was in evidence as the 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;

Defra/EA, 2005; Hawkes *et al.*, 2002). The Environment Agency (EA) has produced a national coastal flood boundary conditions data set that provides the industry with return period estimates of extreme sea levels around the coastline of the UK (EA, 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 contour (JPC) methods are known to have limitations that can result in an underestimation of flood risk or the under-design of structures (Hawkes *et al.*, 2002; HR Wallingford/Lancaster University, 1998). 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. While this analysis has been undertaken offshore, the results have been translated to nearshore using a combination of a wave transformation model and a statistical emulation method that significantly increases computational efficiency. It is envisaged that 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 is 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 of 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 7 and 8, 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/EA, 2005). These two approaches comprise a simplified method that involves the use of JPCs and a robust risk-based statistical method. Both approaches are implemented within the widely used JOIN-SEA software system (Hawkes *et al.*, 2002; HR Wallingford/Lancaster University, 1998).

It is of note, however, that the practice of deriving and applying JPCs has known limitations (Hawkes *et al.*, 2002; HR Wallingford/Lancaster University, 1998).

The JPC method is motivated by the traditional deterministic univariate design framework that specifies that 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 hydraulic loading is defined by multiple sea condition variables. There are three main reasons for this. First, 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 the associated contours and design events, relate directly to the quantity of interest, the return period associated with failure of the structure or the 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 in Figure 1, requires the creation of the 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

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

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 that 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 is 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;

Figure 1. Simulated joint events from a BVN with an associated 100-year JPC (AND definition), with two ‘design events’ highlighted

HR Wallingford/Lancaster University, 1998). This assumption can be defined explicitly.

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

2. $Z = f(X_1, X_2)$

The JPC method implements the following false assumption

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

where the $(x_{1,i}, x_{2,i})$ pairs each satisfy Equation 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 (Corbella and Stretch, 2013; De Michele et al., 2007; Hawkes et al., 2002; Jonathan et al., 2013; Salvadori and De Michele, 2004; Salvadori et al., 2011; Volpi and Fiori, 2014). While 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 (Equation 1). The alternative definitions are OR, PCOND1, PCOND2, PCOND3 and Kendall AND. Their definitions are detailed in Appendix 1.

The alternative multivariate JPC methods were compared using samples drawn from a simple bivariate normal model (BVN) (Figure 1). The commonly used AND probabilities (Equation 1) were calculated for each sample using the known function of the data. A contour with an AND return period of 100 years was then identified (Figure 1). Two sample points on this contour were then selected (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 that a potential ‘100-year design event’ obtained from the widely applied coastal JPC (AND method) (Equation 1) can have a return period that ranges between 1 and 5000 years, depending on which multivariate return period definition is used. An alternative view of 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 that there is little in common between the various different definitions.

While 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 (i.e. the response variable) that is of most

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	
	Sample 1	Sample 2
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

^aDefinition currently applied in practice

Figure 2. A range of '100-year' JPCs obtained from the simulated data

interest. Or, in risk parlance, it is the probability of outcomes, response or consequences that are of most relevance.

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 combination of Regions A and B. 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 (Region B). The JPC method approach also contains limitations with regard to the treatment of other variables, the wave period, for example. In practice, the application involves assuming the 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

Figure 3. Conceptual diagram showing the systematic error introduced by the JPC method, represented by the differences between the two shaded regions (dashed and solid lines respectively)

that overtopping rates and damage potential are sensitive to the wave period (Pullen *et al.*, 2007).

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.

3. Study area and data overview

The method adopted in this study comprises two main components.

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

To implement the method, the coastline was 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 are summarised in Table 2 and detailed below.

A separate offshore multivariate extreme value model was developed at a single offshore WaveWatch III (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 -5 mODN contour) during the emulation phase of the wave

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 8 km resolution for a timespan from January 1980 to June 2014 so it includes the winter storms 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 that 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, multibeam and single beam surveys from channel coastal observatory, known as the EA surfzone composite bathymetry	Compiled by EA Geomatics Group

transformation modelling. The specific relationships and further details are published in an accompanying report (HR Wallingford, 2015a).

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

Wave and wind data were obtained from a hindcast of wave conditions using the WWIII model, undertaken by the UK Met Office (Mitchell *et al.*, 2017). The grid resolution for this model is 8 km and the timespan of the hindcast ranges from January 1980 to June 2014, which therefore includes the winter storms of 2013/2014. Data from 1980 to 2000 are available at a 3 h resolution, and from 2000 onwards, at a 1 h resolution. The wave model was driven with wind data from the European centre for medium range weather forecasting, European re-analysis-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 are 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 were obtained from the Centre for Environment,

Figure 5. Comparison between the WWIII hindcast data and channel light vessel measured wave conditions

Fisheries and Aquaculture Science (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, 2015b).

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

4. Multivariate extreme value modelling

To quantify appropriately 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, while appropriately preserving the dependence between the variables, and then integrate the joint density over the region where the response variable Z has been exceeded

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

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. Equation 4 is solved directly). An alternative method to that employed by JOIN-SEA has however, 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 and Tawn, 2004). This method has greater flexibility in terms of the handling of the dependence structure with a greater number of variables, when compared with the approach adopted within JOIN-SEA and other algebraic copula approaches. Further explanation of 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 (EA, 2011b; Gouldby *et al.*, 2014; Jonathan *et al.*, 2013; Keef *et al.*, 2012; Lamb *et al.*, 2010; Wyncoll and Gouldby, 2015). A description of its implementation here is provided in the following.

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 , deepwater wave steepness was calculated by way of

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

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 that 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 and on each other required 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 that 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 data set was used only to extrapolate extremes of the same primary variable and sampled only when this was the 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). It is of note, however, that the EA (2011a) has already published industry standard extreme sea levels 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 implemented on 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 was each created from Gaussian kernel smoothers fitted to the peaks of each variable that accounts 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 on the industry standard levels (EA, 2011a), these were imposed within the modelling process during specification of the marginal distribution of extreme sea levels. However, for wave height and wind speed, the standard peaks-over-threshold 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

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

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 and Tawn, 1991).

The multivariate method (Heffernan and Tawn, 2004) was applied in the standard way. The extreme variables in each peak data set were therefore transformed from their original scales X to Laplace scales 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 Y was then undertaken.

If 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 $Y_{-i} | Y_i > v$

$$6. \quad Y_{-i} = aY_i + Y_i^b w \quad \text{for } Y_i > v$$

where a and b are vectors of parameters, v is a specified threshold and 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 data set representing 10 000 years' worth (0.2 million) of potential peak overtopping events was randomly sampled. Rejection sampling was used to ensure that Y_i was the most extreme to avoid double counting events where more than one source was extreme. The simulated data set was then transformed back to the original scales. The resulting output was a large multivariate sample of extreme offshore sea condition data that capture the characteristics of dependencies between the variables, as well as preserving the marginal extremes (Figure 8).

To quantify the probability of exceeding a value of the response variable Z (and hence solve Equation 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.

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

et al., 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 1 km spacing along the coastline, at approximately the -5 mODN contour. This depth was chosen as, in shallower water (surfzone), wave breaking increases and the bed levels vary significantly. It was thus desirable to separate out the complex, highly dynamic surfzone and structure-related aspects from the more stable regime in deeper water. This de-coupling enables flexibility with regard to the surfzone models. The surfzone 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 setup using a 200 m regular mesh.

All the SWAN models were setup 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 on analysis of historical peak events. Details of the calibration methodology and results are provided in an accompanying report (HR Wallingford, 2015c).

5.2 SWAN emulation

In principle, it is necessary to transform all of the event outputs from the offshore Monte Carlo simulation (Figure 8) through to the nearshore. It is, however, of note that 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 two-dimensional (2D) models 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 rectangular 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 (GPEs) (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

Figure 8. Example of Monte Carlo simulation output from the multivariate extreme value model

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 with 17 000 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

Figure 9. Comparison of the performance of a GPE fitted to varying numbers of SWAN simulations (solid line) with a traditional 'look-up' table approach based on 17 000 simulations (dashed line)

using a previously established methodology (Camus *et al.*, 2011a, 2011b). The use of the MDA ensures that the multivariate parameter space is captured efficiently. Figure 10 shows an example of the design point 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 to the nearshore wave points located on (approximately) the -5 mODN contour. A series of separate emulators was created for each nearshore wave point (1 km resolution). A more detailed description of the emulator approach is provided in Appendix 2. An 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 hind-cast WWIII data (Figure 5), model structural uncertainties associated with SWAN and model structural uncertainties associated with the emulator. A 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.

6. Example application

6.1 Case study application

To demonstrate the application of the method to estimate wave overtopping rates, two particular structures were selected from the EA's Asset Information Management System. Both sites are located in the southwest of England. 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 was 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 (Battjes and Janssen, 1978; Goda, 2000). The latter method is used for the calculation of surfzone breaking in both SWAN one-dimensional (1D) and 2D models and SWAN 1D was applied in this study. The profile information was extracted using an automated routine applied to the Surfzone digital elevation model 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 Bayesian overtopping neural network (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 were transformed through SWAN 1D and Bayonet into peak overtopping rates. The overtopping rate samples were then ranked and empirical return periods assigned. These results were then 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 JPCs

To demonstrate the difference between the routinely applied JPC approach and the robust statistical method applied here, a series of comparisons were 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. It is thus evident, in

Figure 10. The emulator design points (larger points) output from the MDA overlaid on the extrapolated Monte Carlo samples

these particular examples, that 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 were generated around the coast of England at a 1 km 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 that probabilities of failure are expressed directly in terms of the response variable/s of interest. This is

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

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 (~ 5 m

contour) extreme event data through the surfzone 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 that the SWAN emulation approach, due to its computational efficiency, has the potential to be incorporated within coastal flood forecasting systems.

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

Figure 13. Exmouth

Figure 14. Lyme Regis

Ensemble and probabilistic forecasting are readily achievable with this type of approach.

Should specific combinations of wave height, period and water level conditions be required for deterministic ‘design event’ purposes? This 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 the 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 JPC 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 off-shore wave and wind condition data were derived from a model hindcast of the WWII model. An insight into the uncertainties

Figure 15. SWAN 1D profile example

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

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 the record being analysed. For example, the widely applied guidance on extreme sea levels (EA, 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 offshore waves to nearshore. Where nearshore wave measurements exist, these models have been calibrated to reduce the uncertainties, however. Additional uncertainty is introduced by the use of the emulator. Combined uncertainties from the offshore boundary condition waves, the SWAN model and the 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 the uncertainties associated with the various model components and appropriately communicate these uncertainties, and further research into these aspects is recommended. Given the many different sources of uncertainty and the complexity involved in defining and propagating them through the modelling chain, it was not possible to undertake this analysis in 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 into where future priorities lie for model and data improvements.

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 were demonstrated with reference to alternative contouring methods using a conceptual data set. The results 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 was then undertaken around the coast of England. A robust statistical method was applied at 24 different locations. The output of the multivariate extremes analysis comprises a Monte Carlo sample of ~0.2 million events. To estimate robustly the return period of response variables such as overtopping rate

or economic damage for flood risk analysis, it is, in principle, necessary to transform all of these events to the nearshore and then to overtopping and subsequent flood inundation.

To undertake the offshore to nearshore wave transformation, 24 separate SWAN wave models were set up. To minimise the computational effort involved in transforming all of the Monte Carlo events through the SWAN models, a series of emulators was developed that provides nearshore outputs at a 1 km resolution approximately along the -5 mODN 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 were calculated for two structures located on the south coast of England. These results were compared with the results obtained by applying the widely used JPC method where the known limitations were observed.

It is suggested that 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 were 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)

$$\begin{aligned} p_{\text{AND}} &= \Pr(X_1 > x_1 \cap X_2 > x_2) \\ p_{\text{OR}} &= \Pr(X_1 > x_1 \cup X_2 > x_2) \\ p_{\text{COND1}} &= \Pr(X_1 > x_1 | X_2 > x_2) \\ p_{\text{COND2}} &= \Pr(X_1 > x_1 | X_2 \leq x_2) \\ p_{\text{COND3}} &= \Pr(X_1 > x_1 | X_2 = x_2) \\ p_{\text{Kendal AND}} &= \Pr[p_{\text{AND}}(X_1, X_2) > p_{\text{AND}}(x_1, x_2)] \end{aligned}$$

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

Appendix 2: GPE description

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

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

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

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

where β 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 was used, which is defined by

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

for smoothing parameters θ_i .

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

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

where

$$\begin{aligned} \mathbf{t}(\mathbf{x})^T &= [c(\mathbf{x}, \mathbf{x}_1), \dots, c(\mathbf{x}, \mathbf{x}_n)] \\ A &= [c(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n \\ \mathbf{z}^T &= [f(\mathbf{x}_1), \dots, f(\mathbf{x}_n)] \\ H^T &= [\mathbf{h}(\mathbf{x}_1), \dots, \mathbf{h}(\mathbf{x}_n)] \end{aligned}$$

If applied to one of the design events $\mathbf{x}_i$, this formula returns the known SWAN output with zero error. The prediction equation was applied for every offshore event to produce a large simulated data set of nearshore events.

REFERENCES

- Battjes JA and Janssen JPFM (1978) Energy loss and set-up due to breaking of random waves. *Proceedings of 16th International Conference on Coastal Engineering*. ASCE, New York, NY, USA, pp. 569–587.
- Booij N, Ris RC and Holthuijsen LH (1999) A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research: Oceans* **104(C4)**: 7649–7666.
- Bruun JT and Tawn JA (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 FJ and Medina R (2011a) A hybrid efficient method to downscale wave climate to coastal areas. *Coastal Engineering* **58(9)**: 851–862.
- Camus P, Mendez FJ, Medina R and Cofiño AS (2011b) Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering* **58(6)**: 453–462.
- Coles SG and Tawn JA (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* **76**: 68–78, <https://doi.org/10.1016/j.coastaleng.2013.01.011>.
- Davison AC and Smith RL (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **52(3)**: 393–442.
- Defra/EA (Department for Environment, Food and Rural Affairs/Environment Agency) (2005) *Use of Joint Probability Methods in Flood Management A Guide to Best Practice*. Defra/Environment Agency, London, UK, 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. *Coastal Engineering* **54(10)**: 734–751.
- EA (Environment Agency) (2011a) *Coastal Flood Boundary Conditions for UK Mainland and Islands Project: SC060064/TR2: Design Sea Levels*. EA, Bristol, UK.
- EA (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*. EA, Bristol, UK.
- Goda Y (2000) *Random Seas and the Design of Maritime Structures*. World Scientific, Singapore.
- 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. *Coastal Engineering* **88**: 15–26, <https://doi.org/10.1016/j.coastaleng.2014.01.012>.
- Haigh ID, Wadley MP, Gallop SL et al. (2015) A user-friendly database of 100 years (1915–2014) of coastal flooding in the UK. *Scientific Data* **2**: 150021, <http://dx.doi.org/10.1038/sdata.2015.21>.
- Hawkes PJ, Gouldby BP, Tawn JA and Owen MW (2002) The joint probability of waves and water levels in coastal engineering design. *Journal of Hydraulic Research* **40(3)**: 241–251.
- Heffernan JE and Tawn JA (2004) A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66(3)**: 497–546.
- HR Wallingford (2015a) *State of the Nation: Assessment of the Met Office WaveWatch III Hindcast Wave Data*. HR Wallingford, Wallingford, UK, report number RT013, MCR5289-13-R004.
- HR Wallingford (2015b) *State of the Nation: SWAN 2D Wave Calibration Report*. HR Wallingford, Wallingford, UK, report number RT02, MCR5289-R-02-00.
- HR Wallingford (2015c) *State of the Nation: Coastal Boundary Conditions Report for the Environment Agency*. HR Wallingford, Wallingford, UK, report number 30, MCR5389-30-R00-01.
- HR Wallingford/Lancaster University (1998) *The Joint Probability of Waves and Water Levels: JOIN-SEA, A Rigorous But Practical New Approach*. HR Wallingford, Wallingford, UK, report SR 537.
- Jonathan P, Ewans K and Randell D (2013) Joint modelling of extreme ocean environments incorporating covariate effects. *Coastal Engineering* **79**: 22–31, <https://doi.org/10.1016/j.coastaleng.2013.04.005>.
- Keef C, Tawn J and Lamb R (2012) Estimating the probability of widespread flood events. *Environmetrics* **24(1)**: 13–21.
- Kennedy MC, Anderson CW, Conti S and O'Hagan A (2006) Case studies in Gaussian process modelling of computer codes. *Reliability Engineering and 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. *Proceedings of FLOODrisk 2008, Oxford, UK*, pp. 561–565.
- Lamb R, Keef C, Tawn J *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, NY, USA.
- Mitchell JA, Bett PE, Hanlon HM and Saulter A (2017) Investigating the impact of climate change on the UK wave power climate. *Meteorologische Zeitschrift* **26**(3): 291–306.
- Pullen T, Allsop NWH, Bruce T *et al.* (2007) *Eurotop: Wave Overtopping of Sea Defences and Related Structures – Assessment Manual*. HR Wallingford, Wallingford, UK. See <http://www.overtopping-manual.com> (accessed 12/06/2017).
- Saltelli A, Tarantola S, Campolongo F and Ratto M (2004) *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Wiley, Chichester, UK.
- Salvadori G and De Michele C (2004) Frequency analysis via copulas: theoretical aspects and applications to hydrological events. *Water Resources Research* **40**(12): W12511.
- Salvadori G, De Michele C and Durante F (2011) On the return period and design in a multivariate framework. *Hydrology and Earth System Sciences* **15**: 3293–3305, <http://dx.doi.org/10.5194/hess-15-3293-2011>.
- Serinaldi F (2014) Dismissing return periods. *Stochastic Environmental Research and Risk Assessment* **29**(4): 1179–1189.
- Van Gent MRA, Van den Boogaard HFP, Pozueta B and Medina JR (2007) Neural network modelling of wave overtopping at coastal structures. *Coastal Engineering* **54**(8): 586–593.
- Volpi E and Fiori A (2014) Hydraulic structures subject to bivariate hydrological loads: return period, design, and risk assessment. *Water Resources Research* **50**(3): 885–897.
- Wyncoll D and Gouldby B (2015) Integrating a multivariate extreme value method within a system flood risk analysis model. *Journal of Flood Risk Management* **8**(2): 145–160.

How can you contribute?

To discuss this paper, please email up to 500 words to the editor at journals@ice.org.uk. Your contribution will be forwarded to the author(s) for a reply and, if considered appropriate by the editorial board, it will be published as discussion in a future issue of the journal.

Proceedings journals rely entirely on contributions from the civil engineering profession (and allied disciplines). Information about how to submit your paper online is available at www.icevirtuallibrary.com/page/authors, where you will also find detailed author guidelines.