The joint probability of waves and water levels in coastal engineering design

La probabilite conjointe de la houle et des hauteurs d'eau dans la conception de defences cotieres

PETER J. HAWKES, *Section Leader, Coastal Department, HR Wallingford, Howbery Park, Wallingford, Oxfordshire, OX10 8BA, UK*

BEN P. GOULDBY, *Oceanographer, Coastal Department, HR Wallingford*

JONATHAN A. TAWN, *Professor, Department of Mathematics and Statistics, Lancaster University*

MICHAEL W. OWEN, *Independent Consultant, Powys, UK*

ABSTRACT

On coasts with high tidal ranges, or subject to high surges, both still water levels and waves can be important in assessing flood risk; their relative importance depends on location and on the type of sea defence. The simultaneous occurrence of large waves and a high still water level is therefore important in estimating their combined effect on sea defences. Wave period can also be important in assessing run-up and overtopping, and so it is useful also to have information on the joint distribution of wave height and period.

Unless the variables are either completely independent or completely dependent, multivariate extremes are difficult to predict directly from observational data, as there may be too few events of the relevant type amongst the observations. In the past, the fitting and extrapolation of the dependence functions between the variables has often involved complicated and/or subjective approaches. This paper presents a method for joint probability analysis, using a Monte Carlo simulation approach, based on distributions fitted to water level, wave height and wave steepness, and to the dependence between them.

RÉSUMÉ

Sur les côtes présentant un fort marnage, ou soumises à de fortes surcotes, les hauteurs d'eau à mer calme et les houles peuvent être importantes dans l'évaluation du risque d'inondation ; leur importance relative dépend de l'emplacement et du type de défense côtière. La présence simultanée de fortes houles et de niveaux d'eau élevés à marée haute est donc importante dans l'estimation de leurs effets combinés sur les défenses côtières. La période de la houle peut également être importante dans l'analyse du flot d'assaut et des franchissements, et il est donc utile d'avoir des informations sur la distribution conjointe des hauteurs et des périodes de houle.

A moins que les variables soient complètement indépendantes ou complètement dépendantes, les extrêmes à variables multiples sont difficiles à prévoir directement à partir de données observées, et il peut y avoir un nombre insuffisant d'événements du type qui convient parmi les observations. Par le passé, l'adaptation et l'extrapolation des fonctions de dépendance entre les variables ont souvent impliqué des approches compliquées et/ou subjectives. Ce papier présente une méthode pour l'analyse de probabilités conjointes, utilisant une approche de simulation de Monte Carlo, qui repose sur des distributions adaptées aux niveaux d'eau, aux hauteurs et à la pente de la houle, et à la dépendance entre elles.

1 Introduction

Background

Joint probability refers to two or more partially related environmental variables occurring simultaneously to produce a response of interest. Examples are large wave heights and high water levels, large river flows and high coastal water levels, and large surges and high astronomical tidal levels.

In design or assessment of a sea defence, a key step is the estimation of the probability of failure (usually expressed in terms of a return period) to protect against extreme sea conditions. This probability assessment is a four-stage process.

1. Select a design (either the existing design or one based on some preliminary analysis).
2. Identify all types of failure for that design, including structural failure and excessive overtopping.
3. For each mode of failure, identify the combinations of sea condition variables which cause failure.
4. Estimate the probability of these combinations which give failure.

For any particular mode of failure, we term the associated variable of interest the structure function (Δ), which is dependent on the sea condition variables: water level (X_1), significant wave height (X_2) and mean wave period (X_3).

The astronomical (tide) and non-astronomical (surge) components of still water level can be considered as separate variables, as can wave direction, but typically only the three primary variables above are considered.

The UK Ministry of Agriculture, Fisheries and Food (MAFF) has been funding research on joint probability for several years (HR Wallingford, 1994 and 2000; Coles and Tawn, 1994; POL, 1995 and 1997, Owen *et al*, 1997) and is increasingly expecting the results to be applied in sea defence scheme assessments.

Townend (1994) demonstrated the different results that could be achieved using alternative joint probability approaches in engineering analysis. Scheffner and Borgman (1992) developed a method for generation of a realistic wave sequence from a small

---

Revision received April 17, 2002. Open for discussion till October 31, 2002.

number of stored parameters, including approximately correct distributions of height, period and direction, and dependence and sequencing between them. However, the accuracy of the reproduction of high values was not suited to extremes analysis. Adamson *et al* (1999) used a Monte Carlo simulation approach to investigate riverine flooding, based on the use of a number of different conditional distributions to represent discretised dependence between the input variables.

Dependence between wave heights and water levels

If two or more variables were either completely independent or completely dependent, then the joint probability calculations would be relatively easy. However, this is rarely a good approximation for waves and water levels, which are usually partially dependent.

One might expect to see dependence between surges and waves, since both are related to the local weather conditions. This meteorological dependence (which may not be particularly high to begin with) is masked in consideration of the dependence between wave heights and total water levels, because the astronomical tidal component of water level is unrelated to the weather conditions. It is possible to detect and quantify dependence using simultaneous data on the two variables, but it is not easy to extrapolate from the observed dependence to the behaviour in the extreme tail of the distributions. Hence the need for statistical methods comparable with those used for marginal extreme value predictions.

Typically, there is only a modest correlation between waves and water levels for open coast locations around the UK, the strength of which varies from one area of coast to another. On the east coast of England, strong northerly winds produce both surges and high waves, but there tends to be a time lag between the occurrences of the peak values of the two variables. Conversely, on the west coast, there is little time lag, but wave conditions tend to include a higher swell component unrelated to local weather conditions. The most extreme sea conditions (usually including a high surge component) will tend to show more dependence than more commonly occurring conditions, particularly where the surge to tide variability is large.

An example of high dependence between large waves and the highest water levels occurs in Hong Kong, where the tidal range is relatively small, and extremes of both wave heights and water levels occur during typhoon conditions. However, even in this situation there will not be complete dependence: there may be a time lag between peak surge and the highest waves, and neither may occur at high tide.

There are some locations (eg Dover) where surges are associated with winds from one direction (south-west at Dover), whilst the highest waves are associated with winds from a different direction (south and south-east at Dover). At Dover there is a slight negative correlation between high surges and large wave heights.

Additional dependence, either positive or negative, may arise for hydraulic reasons, especially close to the coast. The extent to which waves are affected by currents may depend on the state and range of the tide. If wave heights are limited by breaking in shallow water, there is a close relationship with water level. These hydraulic effects are not explicitly considered in the statistical analysis methods presented, but if the effects are present in the input wave and water level data then they will be included implicitly. Joint probability analysis in shallow water is therefore site-specific, and although it may be appropriate in some applications, a deeper water analysis is more generic.

Definitions used in joint probability analysis

At any particular time (t), structure variable (Y_t, eg overtopping, run-up, force) will be related to sea condition vector ($\mathbf{X}_t$), comprising sea condition variables ($X_1, X_2, ..., X_n$), via structure function (Δ):

$$Y_t = \Delta(\mathbf{X}_t) = \Delta(X_{1,t}, X_{2,t}, ..., X_{n,t}) \qquad (1)$$

The probability of some critical value (v) being exceeded is $\Pr(Y_t > v)$, which can be expressed as:

$$\Pr(Y_t > v) = \int_{A>v} f\mathbf{x}(\mathbf{x})\, d\mathbf{x} \qquad (2)$$

ie integration of the estimated joint density (f**x**) over the failure region (A>v) of **X**, containing data for which Y would exceed v. For the two-variable case of wave heights and water levels, the curve bounding a shaded area in Figure 1 illustrates the typical shape of contours of equal structure variable Y = v, with the shaded area within the curve indicating Pr (Y > v). If v is chosen to correspond to structural failure, then Pr (Y > v) represents the probability of failure and the shaded area represents the corresponding failure region. In determining Pr (Y > v) it is necessary to establish the extreme joint probability density of the multivariate distribution, using either analytical extrapolation or Monte Carlo simulation. Low values of Pr (Y > v) are often referred to in terms of return period (RP) ie the average period of time, in years, between occasions of v being exceeded:

$$RP = 1 / \{N(\Pr(Y > v))\} \qquad (3)$$

where N is the number of observations per year.

In coastal engineering, joint probability results are often quoted in terms of joint exceedance combinations of water levels (X_1) and wave heights (X_2) for a given probability of occurrence (usually expressed as a return period). The probability corresponds to the chance of a given water level (x_1) being exceeded at the same time as a given wave height (x_2) being exceeded ($X_1 > x_1$ *and* $X_2 > x_2$). (Wave period and direction are also important, but in this analysis method, they are usually just assigned, rather than being derived statistically.)

The bold curve in Figure 1 is a contour of equal joint exceedance probability for wave heights and water levels, with the points indicating particular examples which might be tested in design. The rectangular heavily shaded area illustrates an example range of wave height and water level with the given joint exceedance probability. This area, and the probability it represents, provide an approximation to the failure region shown in Figure 1 and the probability it represents.

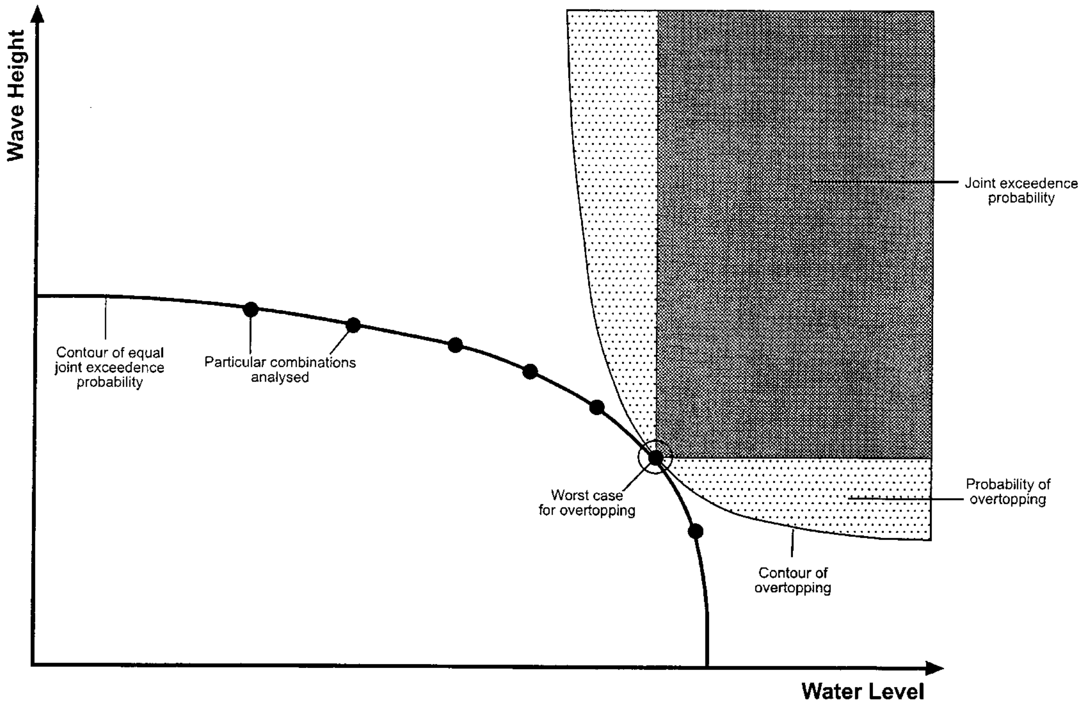

Fig. 1 Joint exceedance and structure variable probabilities

Alternative joint probability approaches

Structure variable approach

One way of estimating the probability of the failure region, Pr (Y > v), is the structure variable approach (sometimes called continuous simulation). The multivariate input data are transformed to equivalent structure variable values, the distribution of which is then extrapolated to extreme values. This can be an adequate method where there is a long period of input data and where only a small number of designs are to be tested.

Joint exceedance approach

In a typical coastal engineering study, a number of combinations of waves and water levels would be derived for each joint exceedance return period of interest. These are sometimes called 'design sea states' and can be applied to a number of different design scenarios. Only one of these sea states will be a worst case in terms of the structure variable, and it may not be the same one for each structure function.

The probability of occurrence of the structure variable calculated from the worst case combination of wave height and water level will generally be higher than the joint exceedance probability. In other words, joint exceedance return period sea conditions tend to under-predict structure variables, if the structure variables are assumed to have the same return period. This is because the same structure variable value might be obtained by other sea conditions in which only one or other of wave height and water level takes a very high value. This is illustrated in Figure 1 by the difference in extent of the two shaded areas. In practice, a margin of safety would be included in the joint exceedance probability predictions to offset this discrepancy with the return period of the structure variable.

Joint density approach

The new approach we present here works in terms of the joint probability density of the three variables (water level, wave height and wave period). As well as providing the standard joint exceedance extremes (if required) the return periods of the structure variable(s) (where known) can be calculated directly.

In this method, the multivariate input data are extrapolated to extreme values, before transformation (at least of the highest values) to equivalent structure variable values. This is potentially more accurate for extremes than the structure variable method, in that it uses much more information from the original data. It uses the distributions and extrapolations of the individual sea condition variables, and information on the dependence between them. Several different structure functions and/or sea defence types can be applied to the same extrapolated joint probability density, be they from physical models, numerical models or design curves.

Outline of the new joint density approach

The method has five main elements.

1. Preparation of input data, consisting of many independent records of wave height, wave period and water level.
2. Fitting of statistical distributions to the wave heights, the water levels and the wave steepnesses.
3. Fitting the dependence between wave heights and water levels, and between wave heights and steepnesses.
4. Simulation of a large sample of wave height, wave period and water level data, using the fitted distributions.
5. Extremes analysis based on the simulated data.

The main developments of the new approach are in elements 2, 3 and 4. HR Wallingford (2000) provides a full description of the developments and their validation.

2 The analysis method

Input data requirements

The first of the five stages involves preparation of the input data. Each input record consists of a simultaneous wave height, wave period and water level (or alternatively a surge) preferably using identical measurement or hindcast locations for both waves and water levels. Ideally, at least four years of data should be used in the dependence analysis.

Usually, in the context of sea defences, only conditions at high water are of interest, and typically peak surges and wave conditions persist for less than half a day. Therefore, each high water (about 706 per year) can conveniently be taken as an independent 'record', which is assumed to persist over the duration of high water. Therefore, 1-year and 100-year return period 'events', for example, have probabilities of occurrence of 1/706 and 1/(100x706) and are assumed to persist over the duration of high water.

Fitting of statistical distributions to wave heights, water levels and wave steepnesses

Although commonly occurring conditions are well represented within a data sample, there will be considerable uncertainty and irregularity within the upper tail of any particular sample, which can be smoothed out by the use of fitted probability distributions. The second stage, therefore, involves the fitting of statistical models to wave heights, water levels and wave steepnesses, above user-defined thresholds. Below the thresholds, the population distributions are represented by the sample distributions.

The Generalised Pareto Distribution (GPD) is appropriate for high and extreme values above a threshold and is used to fit the top few percent of wave heights and water levels.

$$\mathrm{GPD}(\sigma,\xi): \quad P(X \le x \mid X > u) = 1 - \{1 + \xi(x-u)/\sigma\}^{-1/\xi} \tag{4}$$

This defines the cumulative probability distribution of variable X, given that X is greater than the threshold value u, where σ ($\sigma > 0$) is a scale parameter and ξ is a shape parameter. Features of the GPD are that it is invariant to the threshold level, and that it need not increase approximately in proportion to log return period, but rather it can increase either more or less rapidly at the highest levels.

In some situations wave period can be as important as wave height in calculating structure variables, especially when wave heights are depth-limited at a sea wall. Wave steepness represents the ratio of height to length, and from linear wave theory is given by $2\pi H_s/gT_m^2$, where H_s is the significant wave height and T_m is the mean wave period. In the joint exceedance approach, the wave periods associated with calculated extreme wave heights are usually based on the assumption that they have the same wave steepness as the average for the highest few percent of the waves in the sample, neglecting the potentially important wave steepness variability. The new joint density approach models the variability of wave steepness (a more robust variable than T_m for the statistical calculations, although input and output are in terms of T_m). The empirical distribution of wave steepnesses is modelled by a Normal regression on wave height.

$$\text{Normal}\,(\mu,\sigma^2): \quad p(x) = \{1/[\sigma(2\pi)^{0.5}]\}\exp\{-(x-\mu^2/2\sigma^2\} \tag{5}$$

This defines the probability density function of variable X where μ is the mean value and σ^2 is the variance. Features of this distribution are that it is symmetrical and that it has a short tail.

Statistical models for dependence

To model dependence, and its variability with threshold, it is convenient to work with a bivariate distribution whose dependence characteristics are well understood. The third stage, therefore, involves conversion to Normal scales, and the fitting of a dependence function to the wave height and water level data. Simple diagnostic tests have been developed (HR Wallingford, 2000; Bortot *et al*, 2000) to assess whether full dependence or independence models are adequate approximations. For situations when these simplifications cannot be made, two alternative partial dependence statistical models have been developed to represent the dependence between wave heights and water levels. These consist of a single Bivariate Normal (BVN) Distribution and a mixture of two BVN's (HR Wallingford, 2000). These dependence models are just two flexible possibilities from a wide range of potential dependence models for extreme values, see Coles, Heffernan and Tawn (1999) and Heffernan (2000).

For the single BVN, the distributions of wave height and water level are transformed to normalised Normal distributions (Equation 5, with $\mu = 0$ and $\sigma = 1$). A correlation coefficient (ρ) is calculated from the resulting joint distribution, above each of a series of thresholds of exceedance, one of which is later selected to represent the distribution.

The mixture model has seven parameters. One represents the proportions of data between the two distributions, two are correlation coefficients for each of the two BVN's, and four represent the position and spread of the second BVN, relative to the first normalised BVN.

The choice between one and two BVN's is determined by the relative goodness of fit to the data, and by the stability of ρ with threshold. Typically, the single BVN is adequate for a location at which all the wave conditions belong to a single population. However, where the wave conditions belong to more than one population, for example wind-sea and swell, the mixture model is more likely to capture the different dependences in the two populations.

Monte Carlo simulation

The usual purpose of joint probability analysis is to estimate the probability of exceeding some critical value (Equation 2). It would have been possible to develop an analytical approach to extrapolate the fitted distributions to extreme values, from which that probability could be estimated. However, this would be unwieldy and computationally difficult in view of the number of distributions involved, and in view of the fact that the different individual variables contribute in different (and possibly as yet unknown) ways to the magnitudes of the subsequent structure variables. A Monte Carlo simulation approach offers greater flexibility.

The fourth stage, therefore, involves simulation of a large sample of synthetic records of H_s, T_m and water level, based on the fitted distributions, and with the same statistical characteristics as the input data. This permits thousands of years of sea conditions to be simulated with the fitted distributions, extremes and dependences for wave height, water level and wave period, thus providing a sample containing much larger sea conditions than occurring in the observed sample.

Joint probability analysis is based on simultaneous information on the variables of interest. It is quite likely that there will be additional non-simultaneous data on at least one of the variables, with which to refine the extremes predictions for that one variable. For example, there may be 20 years of water level data from which to derive extreme values but only 10 years of simultaneous wave data for the dependence analysis. More accurate extreme values for one or other variable may have already been established (eg from spatial analysis of water levels (POL, 1997) or from application of the Weibull or truncated Weibull distribution to wave heights (Mathiesen et al, 1994)). There may also be anecdotal evidence of severe sea conditions outside the period of the measurements.

Joint probability analysis should use this additional information. This might involve modification of the parameters of the fitted distribution(s) or scaling of the predicted extremes to achieve better agreement with the refined marginal predictions. The present method incorporates such refinements during the Monte Carlo simulation of data, by re-scaling any simulated values (based on the fitted distributions) above a specified return period threshold to new values with equal probabilities taken from the refined predictions. Thus the refinements are permanently built into the synthesised sea state data to be used in subsequent structure variable analysis.

Analysis of joint exceedance extremes and structure variables

The fifth stage involves analysis of the Monte Carlo simulated records to produce extreme values for use in design and assessment of sea defences. These can take the form of extreme wave heights (and associated periods), extreme water levels, or extreme combinations of the two. In addition, any structure function which can be defined in terms of constants (eg wall slope, toe depth, crest elevation etc) and variables H_s, T_m and water level, can be synthesised directly for every record. Direct analysis of the distribution and extremes of the structure variable is then relatively easy: extreme values can be estimated from the appropriate exceedance probability in the synthesised data.

In addition to marginal and joint exceedance extremes, for demonstration purposes four simplified structure functions were tested during development of the methods (HR Wallingford, 2000). These are overtopping rate on a smooth slope (Owen, 1980), 2% run-up on a smooth slope (CIRIA/CUR, 1991; Section 5.1.2.1), force on a high vertical wall (Allsop *et al*, 1996) and armour size for a sea wall (CIRIA/CUR, 1991; Section 5.1.3.2).

More complex variables, for example the simultaneous occurrence of a high force and a high overtopping rate, which would have been difficult to assess previously, can now be routinely studied. Sensitivity and alternative designs can also be assessed relatively easily, by making appropriate changes to the structure function(s) analysed.

3 Case studies

Site selection and data preparation

A number of joint probability data sets were created for locations around England and Wales. Each 'field' data set comprised ten years or so of records of still water level at high water (from measurements) and the simultaneous significant wave height (H_s) and mean wave period (T_m) (from wave hindcasts based on long-term wind data). Corresponding 'synthetic' data sets, again consisting of about ten years of high water records of water level, H_s and T_m, were generated from known statistical distributions based on the field data sets. These had the advantage that the 'true' extremes and joint probabilities could be calculated from extremely long simulations (hereafter called 'truth simulations') based on the known distributions.

All of the data sets were analysed during development and testing of the joint density approach, using several alternative analysis methods (HR Wallingford, 2000). The most interesting results and conclusions for the field data sets at Dowsing (off the Wash, east coast of England), North Wales and Somerset are summarised here. Results are also shown for one of the synthetic data sets, NSIM, based approximately on the North Wales field data set.

Dowsing and North Wales were chosen as typical open coast locations with high waves and high tidal ranges affected by surges: one location in the North Sea and the other in the Irish Sea. (Note: although the wave heights are representative of conditions off the North Wales coast, the water levels quoted for North Wales and for NSIM are based on conditions at Liverpool, where the original source data were recorded.)

Somerset was chosen, again because of its high tidal range, but also to demonstrate the importance of nearshore wave transformation effects and the sensitivity to selection and use of appropriate wave data. The location is exposed to a mixture of locally generated waves and waves arriving from the Atlantic. Wave refraction effects cause a significant increase in T_m as waves propagate from offshore to nearshore, and there is increasing dependence between wave heights and water levels as wave height increases. 28 years of simultaneous measured water level data and hindcast wave data were available to produce two 'field' data sets, one based on deep water wave conditions and the other based on nearshore wave conditions within a bay after application of a wave transformation model.

Predictions are presented and compared in three different forms: firstly, as joint exceedance curves for wave heights and water levels; secondly, as a structure variable derived from the curves (the necessary wave period being calculated based on the average wave steepness of the highest few percent of wave conditions); and thirdly, as the same structure variable, but this time derived directly from conversion of the original data and the Monte Carlo simulation to equivalent structure variable values.

The first and second forms correspond to typical practice in coastal engineering, whilst the third illustrates a refined option available with the new joint density approach. For the synthetic data set NSIM, where the underlying statistical distributions are known, results are also compared with target extremes based on a 'truth simulation'.

A relatively simple structure function is illustrated, namely mean overtopping rate on a smooth sea wall, in one instance with wave heights limited by water depth at the toe of the wall. The equation used for mean overtopping rate (Q) is based on Owen (1980).

$$Q = A\, g\, H_s\, T_m \exp\{-B\, R / [T_m\, (g\, H_s)^{0.5}]\} \qquad (6)$$

where A and B are parameters dependent upon wall slope, R is freeboard (wall crest level minus water level), g is acceleration due to gravity, H_s is significant wave height and T_m is mean wave period.

The hypothetical sea wall parameters for the three sites used in this paper are as follows: A = 0.0192; B = 46.96; NWales/NSIM wall crest level = 14.5mCD; Dowsing wall crest level = 9.0mCD; Somerset wall crest level = 2.8m above MHWS; Somerset wall toe level = 2.7m below MHWS.

Results for case studies using deep water coastal wave data

Diagnostic information from the fitted distributions was used to examine the variability with threshold of the dependence between wave heights and water levels in the data sets. The mixture model (of two Bivariate Normals) was chosen to represent the dependence characteristics of the North Wales and Dowsing data sets, and the single BVN for NSIM. The Monte Carlo simulation was then used to produce the equivalent of 1,000 years of data for each location. Figures 2 and 3 show the original data and the Monte Carlo simulation for North Wales, in the form of scatter diagrams of H_s against water level. Figures 4 and 5 show the original data and the Monte Carlo simulation for Dowsing in the form of scatter diagrams of H_s against T_m. (The 'striping' effect seen in Figures 3 and 5 occurs only below the thresholds used for GPD fitting, where the Monte Carlo simulation can select only from amongst the discrete values existing within the original data sets.)

Figures 2-5 show that the fitting and simulation approach reproduces the original data, that it smoothly extrapolates the probabilities in the extreme tail for wave height and water level, and that it maintains a realistic distribution for wave period in the tail.

Figures 6-9 contrast the likely relative accuracies of the structure variable and joint density approaches, using histograms of the structure variable derived from the North Wales and Dowsing data sets, and from the corresponding Monte Carlo simulations. Apart from the highest few values, the field data sets (Figures 6 and 8) provide fairly smooth distributions of mean overtopping rate, and it would not be unreasonable to estimate up to about 3-year return period values directly from the original data sets. However, Dowsing has only one value (9.9m^3/s/m) above 5m^3/s/m in 9 years of data, and so it would be difficult to estimate 5-year or higher values directly from the data, and extrapolation would be impractical. Similarly, North Wales has three values distinctly higher than the others in 14 years of data, giving difficulty in estimating 10-year or higher values. In both cases, the Monte Carlo simulation (Figures 7 and 9) provides a smooth distribution up to much higher values, without the need for extrapolation of the structure variable itself.

The solid lines in Figures 10-12 show the wave height and water level joint exceedance curves predicted from the Monte Carlo simulations for North Wales, Dowsing and NSIM (the 1,000-year predictions coming from a separate 10,000 year simulation). For NSIM, there are two additional curves for each joint return period. The predictions indicated by the dashed curves are adjusted to have marginal extremes in agreement with target values, so as

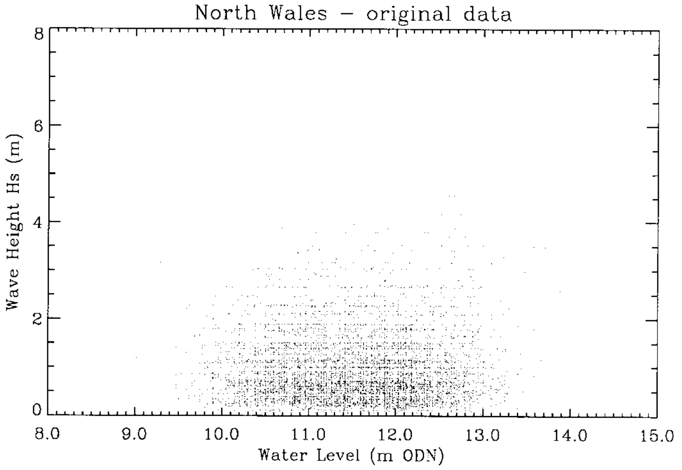

Fig. 2 Original data for North Wales: H_s and water level

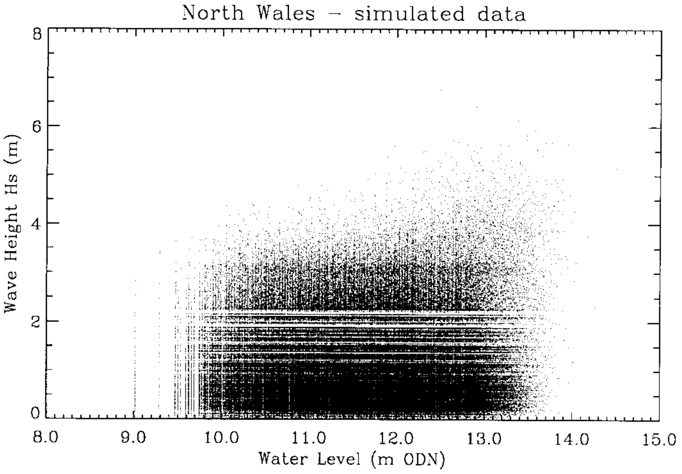

Fig. 3 Monte Carlo simulation data for North Wales: H_s and water level

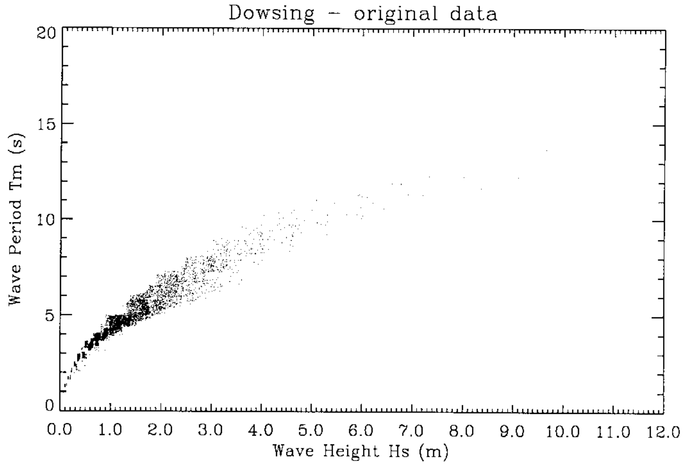

Fig. 4 Original data for Dowsing: H_s and T_m

to highlight the modelling and extrapolation of dependence. The dotted curves show the target values from the 'truth simulations'. The joint exceedance contours for each site are roughly parallel to each other, suggesting no sudden change in dependence at very extreme values. The three alternative predictions in Figure 12 show that when the marginal extremes are known, results from the Monte Carlo simulation are close to the truth, and (in this case) that the majority of the error would come from uncertainties in the marginal extremes predictions. It is not unusual to be able to refine the marginal extremes predictions in this way using additional (non-simultaneous) data, particular on water levels.

Tables 1-3 show the predicted structure variable values for the three data sets. To illustrate the sensitivity to the assumed wave period in the joint exceedance approach (Figures 10-12), two alternative wave steepnesses are used, one being a typical value of 0.06 and the other being taken directly from the data. The tables also include direct predictions of the same structure variables from the Monte Carlo simulations (Figures 7 and 9).

Table 3 includes target extreme values from the 'truth simulation', against which other predictions can be judged. Although there is no corresponding 'true' value available for North Wales or Dowsing, Tables 1 and 2 include 1-year return period estimates of the structure variable directly from the highest few values in the source data sets (Figures 6 and 8). These can be regarded as target values for the field data sets.

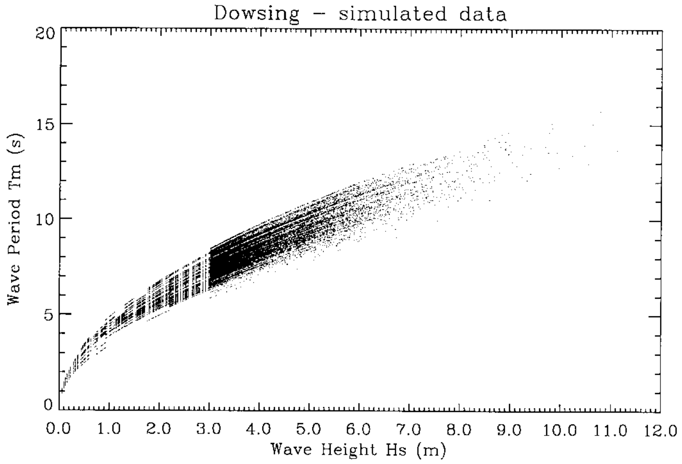

Fig. 5 Monte Carlo simulation data for Dowsing: H_s and T_m

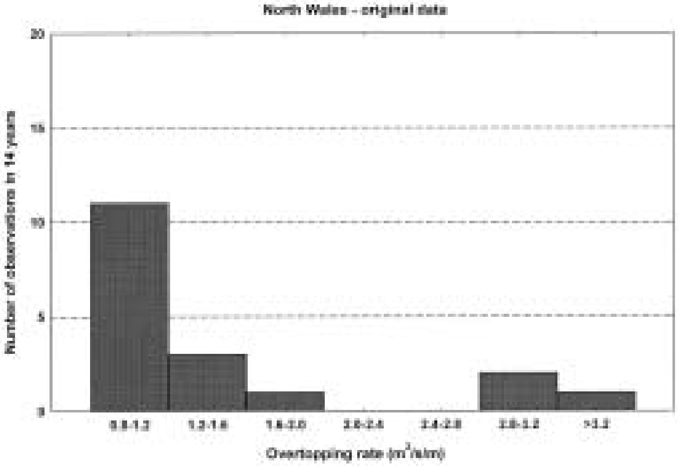

Fig. 6 Histogram of overtopping rate ($m^3/s/m$) for North Wales original data

Table 1 Joint probability overtopping rate results ($m^3/s/m$) for North Wales

Return period (years)	Analysis of Monte Carlo simulation			Structure variable simulation direct from data set
	Joint exceedance		Structure variable simulation	
	Typical steepness of 0.06	Data set steepness of 0.047		
1	0.38	0.57	1.04	0.82
10	1.43	1.86	2.59	
20	1.89	2.44	3.22	
100	3.01	3.85	4.42	
200	3.59	4.56	5.28	

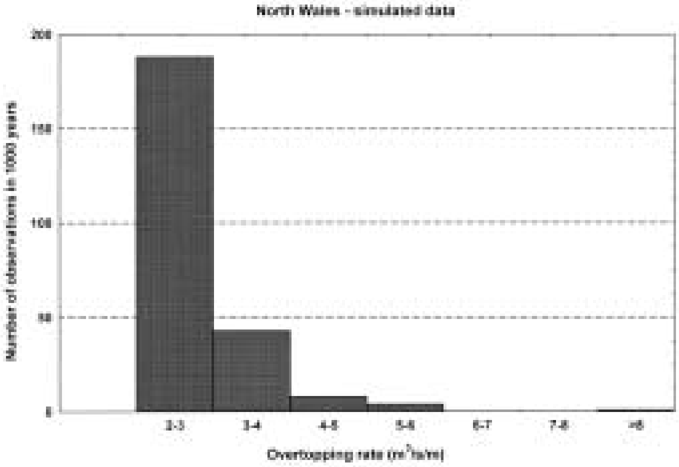

Fig. 7 Histogram of overtopping rate ($m^3/s/m$) for North Wales Monte Carlo simulation

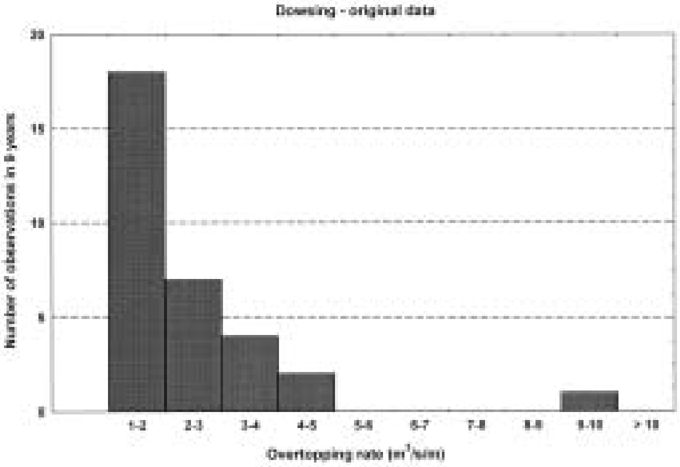

Fig. 8 Histogram of overtopping rate (m^3/s/m) for Dowsing original data

Table 2 Joint probability overtopping rate results (m^3/s/m) for Dowsing

Return period (years)	Analysis of Monte Carlo simulation			Structure variable simulation direct from data set
	Joint exceedance		Structure variable simulation	
	Typical steepness of 0.06	Data set steepness of 0.053		
1	0.99	1.19	2.20	2.6
10	3.35	3.83	5.76	
20	4.26	4.84	7.19	
100	7.39	8.31	10.9	
200	8.77	9.84	12.5	

Comparison of the first two columns of results in each of the tables indicates that use of too high a wave steepness (therefore too low a wave period) could under-estimate overtopping rates by about a factor of two on return period. Comparison between the middle two columns of results in each of the tables indicates that the joint exceedance approach could under-estimate the return period of the structure variable by up to about a factor of three on return period. Comparison with the 'truth simulation' results, and the 1-year results direct from the field data sets, shows that the structure variable simulation based on the Monte Carlo simulation gives predictions closest to target values.

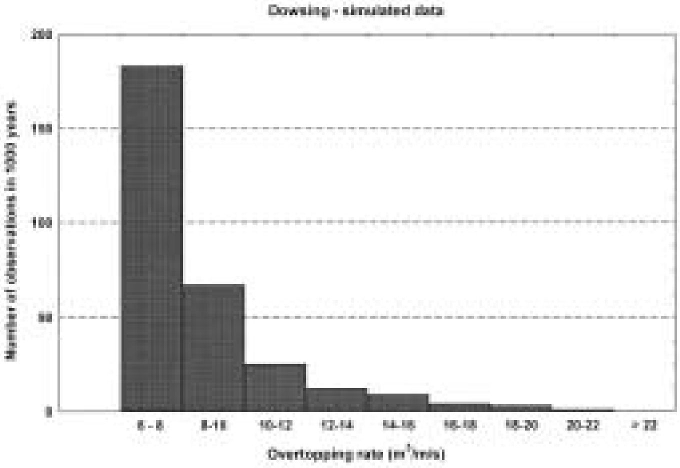

Fig. 9 Histogram of overtopping rate (m^3/s/m) for Dowsing Monte Carlo simulation

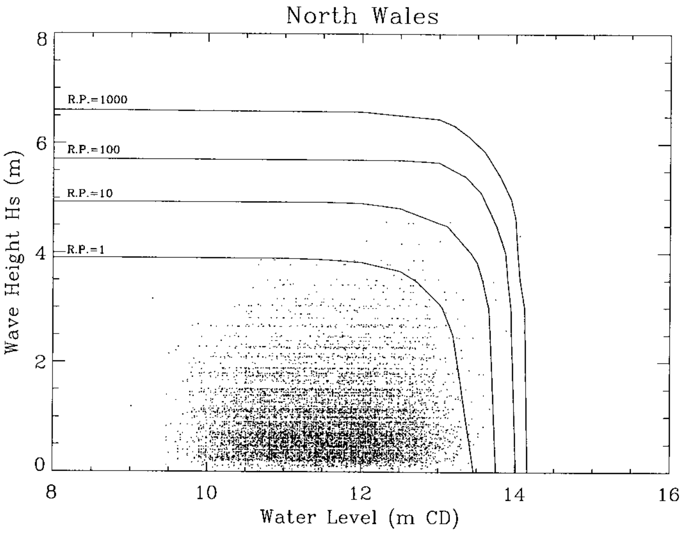

Fig. 10 Joint exceedance curves for North Wales

Table 3 Joint probability overtopping rate results (m^3/s/m) for NSIM

Return period (years)	Analysis of Monte Carlo simulation			Target results from 'truth simulation'
	Joint exceedance analysis		Structure variable simulation	
	Typical steepness of 0.06	Data set steepness of 0.048		
10	0.90	1.22	1.49	1.41
100	1.86	2.30	2.50	2.58

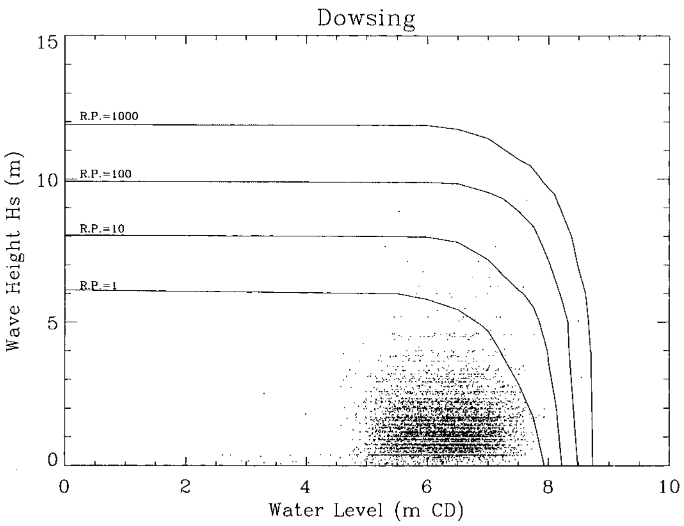

Fig. 11 Joint exceedance curves for Dowsing

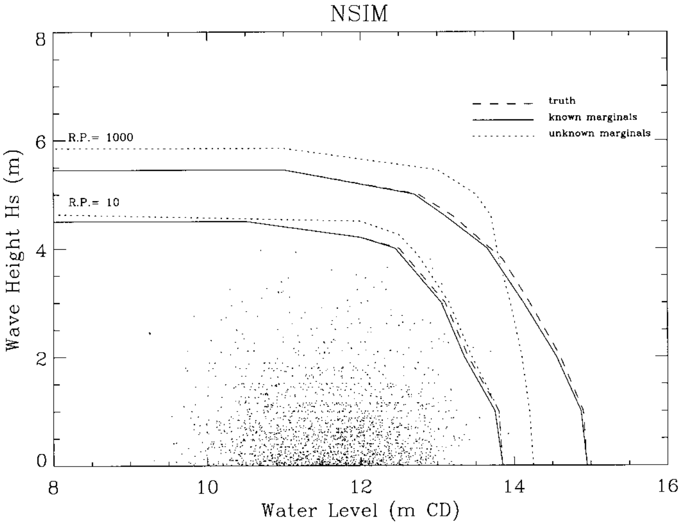

Fig. 12 Joint exceedance curves for NSIM

Results for the case study using nearshore wave data

The analysis for Somerset was undertaken firstly using offshore wave data and then using nearshore wave data, calculating overtopping for a hypothetical sea wall, using the joint density approach. Both depth-limited and non-depth-limited conditions at the sea wall were tested.

The mixture of locally generated and swell waves at the site suggested that the mixture of BVN's would be a more appropriate probability model than the single BVN. Therefore the MIX model was run to produce direct predictions of the structure variable via Monte Carlo simulation. The BVN model was also run for comparison purposes, firstly using a best estimate of correlation between wave heights and water levels directly from the data, and then assuming independence ('$\rho = 0$') and then complete dependence ('$\rho = 1$'). The alternative extremes predictions are listed in Table 4.

Predictions from MIX are quite similar to those from BVN (with a best estimate of ρ) for return period of 1-year, and it is reassuring to note that both methods produce plausible results for a frequency of occurrence well within the data set. However, predictions diverge at higher return periods, those from MIX being consistently a little higher.

The BVN best estimates are only a little higher than those based on the assumption of independence. The independent ($\rho = 0$) predictions for 10-year and 200-year return periods correspond roughly to best estimate predictions for 5-year and 50-year return periods, respectively. A factor of four difference in return period at this site corresponds to a difference of about a quarter of a metre in still water level, and perhaps a half a metre difference in sea defence level (allowing for the larger waves which could reach the wall at a higher water level). Conversely, the dependent predictions for 1-year and 20-year return periods correspond roughly to best estimate predictions for 10-50-year and 800-3000-year return periods, respectively. This degree of over-prediction at this site would correspond to a difference of about two thirds of a metre of water level and perhaps a metre or so in sea defence level.

Table 4 Joint probability overtopping rate results ($m^3/s/m$) for Somerset

Wave data	Sea wall	Return period (years)	MIX model	BVN model		
				ρ=0	Best	P=1
Offshore	Deep at toe	1	0.10	0.06	0.09	0.41
		10	0.37	0.21	0.31	0.99
		20	0.50	0.25	0.38	1.21
		100	0.85	0.47	0.69	1.81
		200	1.04	0.58	0.85	2.00
	Depth limited at toe	1	0.05	0.03	0.05	0.27
		10	0.19	0.10	0.16	0.57
		20	0.25	0.13	0.21	0.65
		100	0.42	0.22	0.31	0.87
		200	0.48	0.27	0.37	0.98
Nearshore	Deep at toe	1	0.15	0.11	0.14	0.68
		10	0.73	0.54	0.67	2.07
		20	1.01	0.77	0.90	2.45
		100	1.72	1.44	1.61	3.50
		200	2.25	1.56	1.78	4.31
	Depth limited at toe	1	0.11	0.07	0.09	0.67
		10	0.53	0.34	0.48	1.96
		20	0.73	0.46	0.65	2.22
		100	1.26	0.76	0.99	2.78
		200	1.78	0.84	1.32	2.94

The predictions are quite dependent on which source of wave data was used. Higher overtopping rates are predicted using the nearshore wave data, because of the increase in T_m in the approaches to the coast, and slightly lower rates are calculated in the cases of waves being depth-limited at the wall, compared to those with unlimited depth.

Comments on the case studies

The larger number of comparisons presented in HR Wallingford (2000) suggest that the new joint density approach, working directly in terms of the structure function, is consistent and reliable, giving robust predictions even for high return periods. The return periods of structure variables calculated (as is usually done at present) from joint exceedance extremes tend to be about half the size of the return periods of the joint extremes themselves. The new Monte Carlo simulation approach allows a more direct and accurate assessment of the structure function value with a given return period.

The importance of wave period is shown in several ways: the use of wave periods which are too low (perhaps due to using a constant wave steepness which is too high) does not affect the joint exceedance extremes directly but it does lead to an under prediction of overtopping rate; the use of a variable wave steepness, rather than a constant wave steepness, tends to increase overtopping predictions.

4 Discussion and conclusions

Advantages of the new joint density approach

The new approach is potentially more objective and flexible than alternative methods previously used in consultancy studies. A single statistical analysis can provide self-consistent estimates of the marginal distributions and extremes, joint exceedance extremes, joint probability density, direct calculation of the distribution and extremes of structure variables, and parameters of all fitted distributions. It can thus provide input to both 'design sea state' and 'risk-based' analyses of sea defences.

The variability of wave period (or wave steepness), which is important for some structure functions, can be included in the analysis. Additional (non-simultaneous) data on any of the variables can also be incorporated.

The new approach permits a more direct analysis of structure variable(s) such as run-up or overtopping, thereby addressing the probability of failure or damage more directly than the more commonly used joint exceedance extremes approach which provides only an approximation to the probability of the structure variable(s). Once the Monte Carlo simulation has been carried out, several different structure functions and/or different designs can be tested consistently and relatively easily. Sensitivity to uncertainty in the input variables or the dependence between them can be tested in a similar way, by multiple simulations of the structure function(s).

New possibilities based on Monte Carlo simulation

A small number of design sea states for a particular location are often used in design and assessment of sea defences. These sea states might correspond to wave conditions with specified return periods, or to combinations of wave conditions and water levels with given joint return periods. These design sea states may be applied to calculation of different structure variables and/or different types of sea defence. However, the same (or even higher) values of the structure variables might occur, for example, during swell wave conditions with a lower wave height but with a longer wave period. The return periods of the calculated structure variables will therefore not necessarily be the same as those of the corresponding design sea states.

A potentially better approach is possible based on Monte Carlo simulation of a wide range of sea states, from which the distribution and extremes of the structure variable(s) can be determined directly. Once the long-term sea state data has been simulated, several structure functions and/or defences can be assessed quite quickly, without needing to know the corresponding design sea states: the probability of simultaneous failure mechanisms occurring could also be evaluated. Design and assessment could thus be done directly in terms of any structure function(s) of interest, for example the overall probability of structural damage and/or unacceptable overtopping. This would imply a major change in design practice, but it is consistent with a gradual move in the UK (MAFF, 2000) towards more risk-based design.

Additional applications

The single BVN approach has been developed further to an equivalent Trivariate Normal (TVN) method. This permits analysis of three partially dependent variables, plus wave period as a fourth variable dependent on wave height. The procedure involves separate BVN fits to each of the three pairings of variables, and selection of an appropriate correlation coefficient for each, followed by a Monte Carlo simulation using the TVN. This has proved quite robust, the only requirements being that the threshold for each of the dependence fits should be the same, and that the three correlation coefficients (ρ_{12}, ρ_{23}, and ρ_{13}) should meet a consistency criterion:

$$1-\rho^2_{12}-\rho^2_{23}-\rho^2_{13}+2\rho_{12}\rho_{23}\rho_{13}\ge 0 \qquad (7)$$

The authors have also used the methods in joint extrapolation of the one, two or three variables concerned in prediction of extreme water levels, flows and waves in different parts of estuaries. An interesting difficulty to be overcome in this application arises from the way in which 'events' are defined. Whilst still water levels and wave conditions can be represented by values occurring at a specified time (for example, at successive high waters) flows are usually defined in terms of hydrographs occurring over periods of several hours or even several days.

The methods are amenable to future incorporation of sequencing, ie temporal dependence between successive records, and use of a time-varying structure function, which may be necessary to represent progressive deterioration of sea defences.

Acknowledgements

The UK Ministry of Agriculture, Fisheries and Food (now part of the Department for Environment, Food and Rural Affairs) provided the financial support for this work, and the enthusiasm of Tony Polson and John Goudie in instigating and monitoring the project. The assistance of former colleagues Dr Paola Bortot (now of the University of Bolonga), Andy Yarde (now of Bullen Consultants) and Ian Meadowcroft (now of the Environment Agency) is gratefully acknowledged.

Notation

A, B	parameters of overtopping formula
A>v	failure region of joint density
BVN	Bivariate Normal Distribution
CD	UK Admiralty Chart Datum
f**x**	estimated joint density of vector **X**
GPD	Generalised Pareto Distribution
H_s	significant wave height
MIX	mixture of two BVN's model
NSIM	simulated data set based on North Wales
Q	mean overtopping rate
R	sea wall freeboard
T_m	mean wave period
TVN	Trivariate Normal Distribution

g	acceleration due to gravity
t	time
u	threshold in GPD
v	threshold in structure variable
X, X_1, X_2, X_3	variables
$\mathbf{X}_t$	realisation of vector $\mathbf{X}$ at time t
Y	structure variable
Y_t	realisation of Y at time t
x_1, x_2	specific thresholds of X_1 and X_2
Δ	structure function
μ	mean value for Normal Distribution
ρ	correlation coefficient
σ	variance for Normal Distribution; scale parameter for GPD
ξ	shape parameter for GPD

References

Adamson, P. T., Metcalfe, A. V. and Parmentier, B. (1999). Bivariate extreme value distributions: An application of the Gibbs sampler to the analysis of floods. Water Resources Research, Vol. 35, No. 9, 2825-2832.

Allsop, N. W. H., Vicinanza, D., Calabrese, M. and Centurioni, L. (1996). Breaking wave impact loads on vertical faces. ISOPE Conference, Los Angeles.

Bortot, P., Coles, S. G. and Tawn, J. A. (2000). The multivariate Gaussian tail model: An application to oceanographic data. Applied Statistics, 49, 31-49.

CIRIA/CUR (1991). Manual on the use of rock in coastal and shoreline engineering. Construction Industry Research and Information Association Special Publication 83 / Centre for Civil Engineering Research and Codes Report 153.

Coles, S. G., Heffernan, J. E. and Tawn, J. A. (1999). Dependence measures for extreme value analysis. Extremes, 2, 339-365.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design. Applied Statistics, 43, 1-48.

Heffernan, J. E. (2000). A directory of coefficients of tail dependence. Extremes, 3, 279-290

HR Wallingford (1994). Validation of joint probability methods for large waves and high water levels. HR Wallingford Report SR 347.

HR Wallingford and Lancaster University (2000). The joint probability of waves and water levels: JOIN-SEA: A rigorous but practical new approach. HR Wallingford Report SR 537.

MAFF (2000). Flood and coastal defence project appraisal guidance: Approaches to risk. Ministry of Agriculture, Fisheries and Food Publication PB 4907, Crown copyright.

Mathiesen, M., Goda, Y., Hawkes, P. J., Mansard, E., Martin, M. J., Peltier, E., Thompson, E. F. and van Vledder, G. (1994). Recommended practice for extreme wave analysis. Journal of Hydraulic Research, Vol. 32, No. 6.

Owen, M. W. (1980). Design of seawalls allowing for wave overtopping. HR Wallingford Report EX 924.

Owen, M. W., Hawkes, P. J., Tawn, J. A. and Bortot, P. (1997). The joint probability of waves and water levels: A rigorous but practical new approach. MAFF Conference of River and Coastal Engineers, Keele, UK.

POL (1995). Estimates of extreme sea conditions: Extreme sea-levels at the UK A class sites: Optimal site-by-site analyses and spatial analyses for the east coast. Proudman Oceanographic Laboratory Internal Document No. 72.

POL (1997). Estimates of extreme sea conditions: Spatial analyses for the UK coast. Proudman Oceanographic Laboratory Internal Document No. 112.

Scheffner, N. W. and Borgman, L. E. (1992). Stochastic time-series representation of wave data. Journal of Waterway, Port, Coastal and Ocean Engineering, ASCE, Vol. 118, No. 4, 337-351.

Townend, I. H. (1994). Breakwaters and coast defences: Risk assessment of coastal defences. Proceedings of the Institution of Civil Engineers, Water, Maritime and Energy, Vol. 106, 381-384.