

# National scale multivariate extreme value modelling of waves, winds and sea levels

Ben Gouldby<sup>1</sup>, David Wyncoll<sup>1</sup>, Mike Panzeri<sup>1</sup>, Mark Franklin<sup>2</sup>, Tim Hunt<sup>2</sup>, Nigel Tozer<sup>1</sup>, Uwe Dornbusch<sup>2</sup>, Dominic Hames<sup>1</sup>, Tim Pullen<sup>1</sup>, Peter Hawkes<sup>1</sup>

<sup>1</sup>HR Wallingford Ltd., Benson Lane, Crowmarsh, Wallingford, UK  
<sup>2</sup>Environment Agency, Ergon House, Horseferry Rd, London, UK

**Abstract.** It has long been recognised that extreme coastal flooding can arise from the joint occurrence of extreme waves, winds and sea levels. The standard simplified joint probability approach used in England and Wales can result in an underestimation of flood risk unless correction factors are applied. This paper describes the application of a state-of-the-art multivariate extreme value model to offshore winds, waves and sea levels around the coast of England. The methodology overcomes the limitations of the traditional method. The output of the new statistical analysis is a Monte-Carlo (MC) simulation comprising many thousands of offshore extreme events and it is necessary to translate all of these events into overtopping rates for use as input to flood risk assessments. It is computationally impractical to transform all of these MC events from the offshore to the nearshore. Computationally efficient statistical emulators of the SWAN wave transformation model have therefore been constructed. The emulators translate the thousands of MC events offshore. Whilst the methodology has been applied for national flood risk assessment, it has the potential to be implemented for wider use, including climate change impact assessment, nearshore wave climates for detailed local assessments and coastal flood forecasting.

## 1 Introduction and background

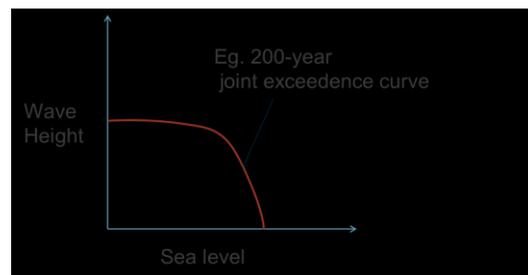
It is well-known that coastal flooding in England arises as a combination of extreme sea levels and wave conditions occurring together, and consideration of extremes of their joint likelihood of occurrence is important [1, 2]. The Environment Agency (EA) of England has produced a national coastal flood boundary conditions report that provides industry with return period estimates of extreme sea levels around the coastline of the UK. Information relating to extreme wave conditions and their joint likelihood of occurrence with extreme sea levels is, however, required to undertake coastal flood risk analysis and to support the design of coastal structures that protect critical infrastructure, including nuclear facilities.

This study, undertaken for the EA, comprises a multivariate analysis of extreme waves, sea levels and wind speeds around the coastline of England. Whilst this analysis has been undertaken offshore, the results have been translated to the nearshore using a combination of a wave transformation model and a statistical emulation method. 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.

There is a long tradition of undertaking analysis of the joint probability of waves and sea conditions around the coastline of England with two distinct approaches in widespread use. These two approaches comprise a simplified method that involves the use of environmental joint probability contours and a robust risk-based statistical method.

Joint probability contouring methods involve the development of contours that have some defined equal probability of exceedance, and include those developed

and explored by a range of authors [3,4,5,6]. An example of the approach adopted in England is shown in concept in Figure 1.



**Figure 1** Conceptual diagram showing a traditional joint probability contour

It is of note, however, that for flood risk analysis and risk-based structural design, it is the probability of exceeding a response (or consequence) of interest that is of relevance. In general terms this latter aspect requires integration of the joint probability density of the loading variables, over the response function of interest.

$$\Pr(Z > z) = \int_{Z > z} f_X(x) dx \tag{1}$$

Where  $Z$  is a response variable of interest and  $X$  is a vector comprising the forcing sea conditions. Joint probability contours are generally developed in the absence of knowledge of the response function. Their application does not lend itself to the direct evaluation of the required quantity of interest, i.e. the probability of exceeding a specified value of the response variable.

In this analysis Equation 1 is solved directly using a multivariate method that involves extrapolation of the joint probability density to extreme values. The output of

<sup>a</sup> Corresponding author: b.gouldby@hrwallingford.com

the multivariate extremes analysis is a Monte-Carlo simulation comprising many 1000's of extreme sea condition events. In principle, it is necessary to transform all of the events output from the offshore Monte Carlo simulation through to the nearshore. It is, however, of note that 2D wave transformation models can be computationally time consuming to run. Rather than attempt to run the model for all of these events a statistical model emulation method was employed. This method ensures practical runtimes for this national scale analysis.

## 2 Model set up and data

The objective of the analysis described here is to provide multivariate extreme sea conditions in the nearshore region around the coast of England for potential use in structural design and flood risk analysis. The method comprises two main components: Offshore multivariate (joint) probability analysis; offshore to nearshore wave transformation of the extreme events.

To implement the method the coastline was sub-divided into 24 different regions, each region comprising a SWAN wave transformation model domain, Figure 2. A separate offshore multivariate extreme value model was developed for each of the 24 regions.

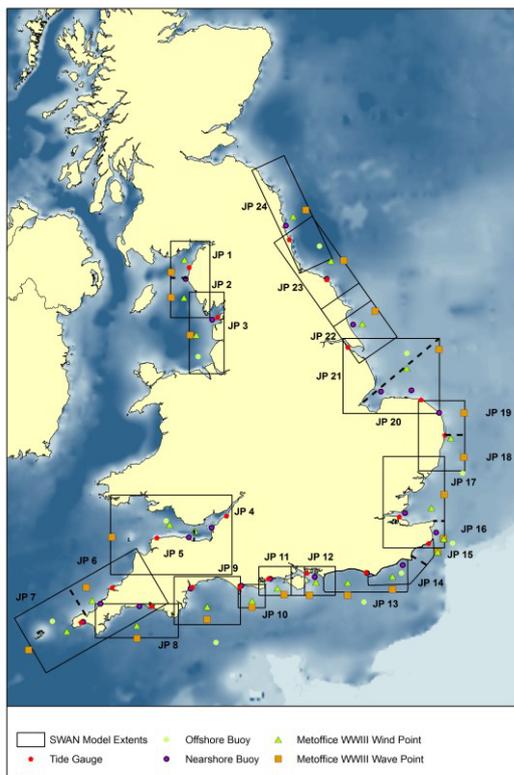


Figure 2 SWAN wave model grids and multivariate data set locations

Time series sea level data was obtained from the network comprised within the UK Coastal Monitoring and Forecasting (UKCMF) service operated by the EA. Prior to implementation within the multivariate analysis, the water level data was de-trended and updated to present. Wave and wind data was obtained from a hindcast of

wave conditions using the WaveWatch III Model (WWIII), undertaken by the Met Office. The grid resolution for this model is 8km and the timespan of the hindcast is from January 1980 to June 2014 which therefore includes the severe winter storms of 2013/2014 that caused significant flooding in England. Data from 1980 to 2000 is available at a 3-hour resolution and from 2000 onwards at a 1-hour resolution. The wave model was driven with wind data from the ECMWF ERA-interim (global) and Unified (regional) models. The hindcast study provided spectral components of waves. 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 2.

## 3 Multivariate extreme value method

The objective of the multivariate extreme value analysis was to extrapolate the joint probability density of the waves, winds and sea level information to extreme values whilst ensuring the appropriate dependence between the variables was captured. 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, wind speed and sea level required extrapolation to extremes. The approach adopted for undertaking this extrapolation was that developed by [7]. Further description on the justification for the use of this model in the context of coastal wave and water level analysis, and flood risk modelling, is provided by [8,9,10].

The method works pairwise by fitting non-linear regression models above specified thresholds. This analysis is undertaken after de-clustering of the time series data, marginal univariate extremes fitting and transformation to uniform scales of each of the variables. Once fitted, the multivariate model enables simulation of synthetic extreme sea conditions. The simulation method involves use of the residuals obtained from fitting the regression model, hence capturing some of the natural variability within the dependence structure. The properties of the synthesised events include preservation of the marginal extremes of waves, winds and sea levels and also the dependence within the extremes. A conceptual diagram depicting synthetic and empirical events from which the multivariate model was fitted, on transformed scales, is shown in Figure 3.

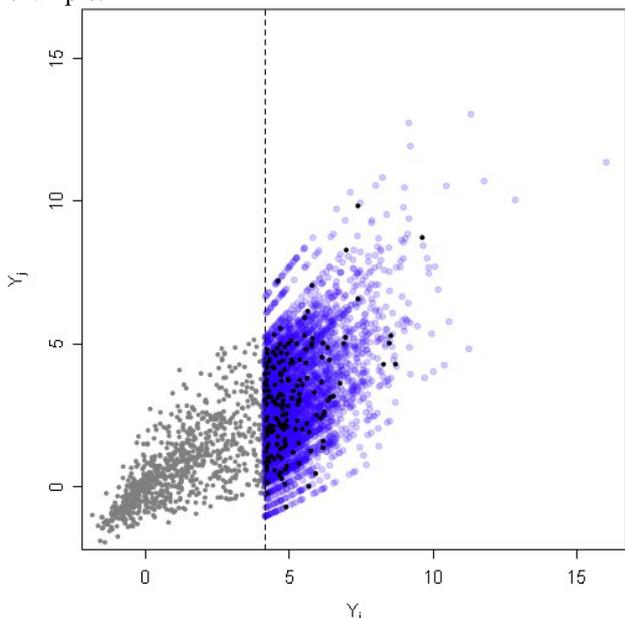
The multivariate method was used to generate synthetic events that were representative of 10,000 years offshore (typically > 0.5million events) of each of the 24 SWAN models.

## 4 Offshore to Nearshore wave transformation modelling

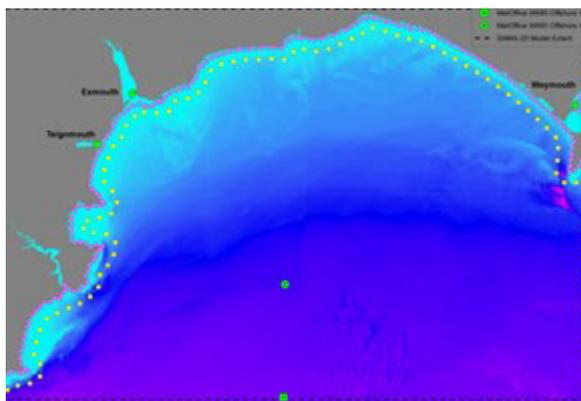
The model chosen to transform the wave conditions from offshore to inshore was the well-known SWAN model [11]. The objective of the SWAN modelling was to transform the offshore multivariate extreme sea condition data to a series of nearshore locations with a

1km spacing along the coastline, at approximately the -5mODN contour, Figure 4.

This sea bed elevation was chosen because in shallower water, wave breaking increases and the bed levels vary significantly. It is thus desirable to separate out the complex surf zone and structure related aspects to enable flexibility with regard to sensitivity analysis relating to beach levels on the foreshore and at the structure, for example.



**Figure 3** Diagram showing synthetic events simulated from the multivariate model and underlying empirical data.



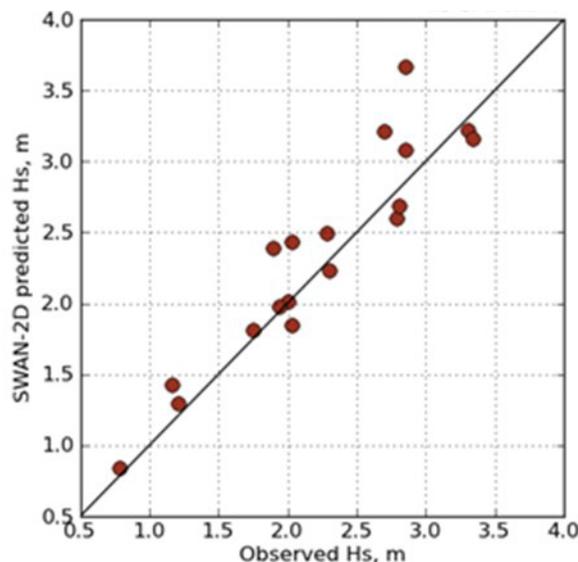
**Figure 4** Example SWAN model domain with nearshore output points highlighted (approx. -5mODN contour)

For reasons relating to numerical stability and runtime efficiency, the new SWAN models were set up using a 200m regular mesh. The SWAN models were set up in a stationary mode with a constant wind direction and speed applied. Each of the new models was calibrated using a range of different events selected, based upon analysis of historical peak events. An example of the performance of a calibrated SWAN model is shown in Figure 5.

For computational reasons, rather than attempt to run SWAN 2D for the many 1000's of events within each region, a statistical model emulation method was employed [12]. A statistical emulator is similar in

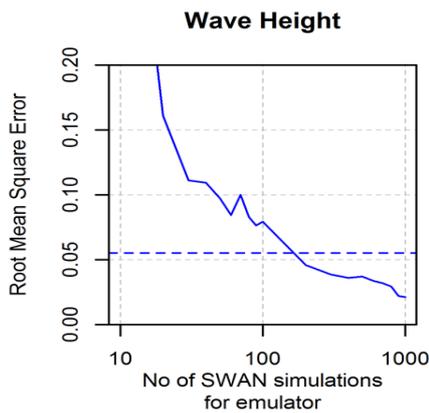
concept to a traditional “look-up table” approach used in coastal flood forecasting systems, for example. The process involves running the SWAN 2D model for a subset of events (known as the design points). Interpolation techniques are then applied to predict the results for other events (not run in SWAN 2D).

Traditional look-up table approaches are typically applied using regular or recti-linear grids and linear interpolation techniques. As the output from SWAN is generally not a linear function of the inputs, these traditional look-up tables can be inefficient and require a large number of design point simulations.



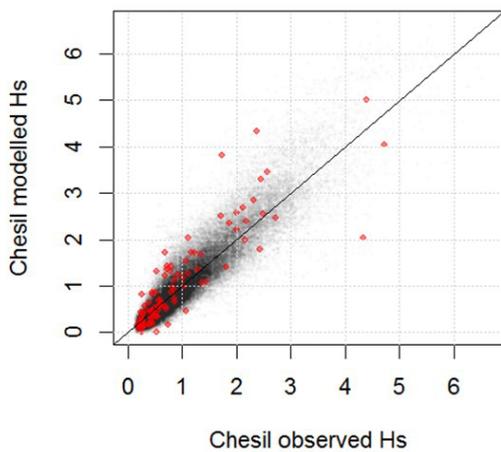
**Figure 5** Example SWAN wave calibration performance – Central South Coast.

Gaussian Process Emulators (GPE's) [12], are a more sophisticated interpolation technique and have been shown to be efficient when used in the context of wave transformation modelling [13]. To select the design points used to fit the emulator and hence used to define the boundary conditions for the SWAN model, the Maximum Dissimilarity Algorithm (MDA) was applied using a previously established methodology [14]. The use of the MDA ensures the multivariate parameter space is captured efficiently. Once fitted, the GPE can then be used in place of SWAN to transform all of the extreme events from offshore to nearshore with significant computational savings when compared to traditional look-up table approaches. Figure 6 shows the performance of the GPE, when compared to a traditional look-up table approach, in terms of RMSE of  $H_s$ . The analysis, undertaken using a benchmark dataset involving simulation using the SWAN model for all events, shows the performance of a traditional look-up table created using 17000 SWAN model simulations (dashed line). The GPE shows significant efficiencies and reaches the same RMSE with around 200 SWAN simulations, a significant computational saving, particularly in the context of this analysis comprising 24 regions.



**Figure 6** GPE comparison against a traditional look-up table (dashed line). Y axis units are metres.

To verify the performance of the GPE, outputs at selected nearshore points were compared with measured data from nearshore wave buoys operated by the Channel Coastal Observatory (CCO). An example comparison is shown in Figure 7. The larger highlighted points are “design points” from the SWAN model itself. The other data points are GPE predictions. The comparison shows good agreement and the emulator outputs fall within the general scatter arising as a result of uncertainties within the input boundary conditions and those associated with the SWAN model, as well as the emulator itself.



**Figure 7** GPE comparison against measured nearshore wave data (highlighted points are actual SWAN model runs). Axis units are metres.

GPEs have been fitted between the offshore and each nearshore output point. The resulting output of the GPEs is therefore a full multivariate distribution of extreme events at a 1km resolution. This data set can be used for a wide variety of uses as discussed below.

## 5 Case Study application

The nearshore multivariate data set has the potential to be used for a number of applications including flood risk analysis, engineering design and climate change impact assessment. To demonstrate the application of the data here, a wave overtopping response function has been used for a structure located on the South Coast of England.

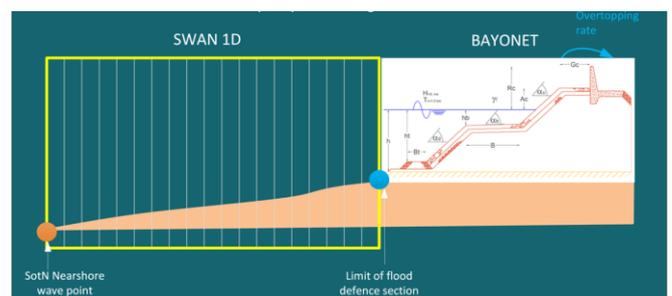
Wave overtopping methods generally require sea conditions at the toe of the structure to be estimated. It is thus necessary to transform the sea conditions from the nearshore to the structure toe. There are a range of methods that can be applied to undertake this transformation [15,16]. The method of Battjes and Janssen method is used for the calculation of surf zone breaking in both SWAN 1D and 2D and the former was applied for this study.

To translate the data at the toe of the flood defence structures to overtopping discharges for use in flood inundation analysis, the BAYONET wave overtopping model was applied [17]. BAYONET 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 [18] but incorporates additional information relating to uncertainty.

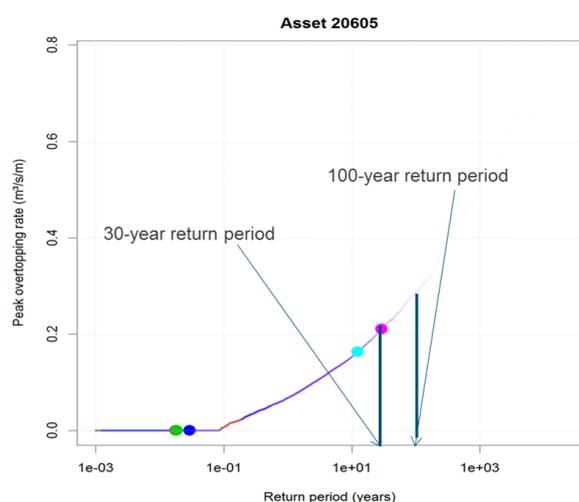
The Monte Carlo realisations at the nearshore point have been transformed through SWAN 1D and BAYONET into peak overtopping rates, Figure 8. The overtopping rate samples were then ranked and empirical return periods assigned. These results have then been analysed to determine an empirical distribution of overtopping rates as shown in Figure 9 .

Highlighted in Figure 9 (larger points) are events that have been extracted from the traditional joint probability 100 year contour (Figure 1) and used to calculate overtopping rates. In practice, points on the 100 year contour are analysed and the highest overtopping rate obtained would be assigned a return period of 100 years. It is, however, evident from this analysis the overtopping rate obtained using this method is closer to a 30 year return period. This highlights how flood risk can be underestimated unless correction factors are applied. These findings are consistent with earlier work in this area [2].

The new method estimates return period overtopping rates directly and hence avoids the use of these correction factors.



**Figure 8** 1D profile representation used for overtopping rate estimation on the case study site.



**Figure 9** Distribution of extreme overtopping rates obtained from the new multivariate method, compared with those from a traditional contour (large points)

Given site-specific bathymetry and coastal defence structural geometry, it is straightforward to undertake this analysis for any/all structures along the coastline of England. This can provide a consistent source of information for coastal flood risk analysis.

## 6 Discussion and conclusions

The limitations of the joint probability contour approach, widely used in practice for coastal flood risk analysis and the design of coastal structures in England have been described. A multivariate extreme value analysis of offshore waves, winds and sea levels has then been undertaken around the coast of England. A robust statistical method has been applied in 24 different areas. The output of the multivariate extremes analysis comprises a Monte Carlo sample of 1000's of events. To robustly estimate the return period of response variables it is, in principle, necessary to transform all of these events through to the nearshore and then to relevant response functions, including wave overtopping rates.

To undertake the offshore to nearshore wave transformation 24 separate SWAN wave models have been set up. To minimise the computational effort involved in transforming all of the Monte Carlo events through the SWAN models, a series of emulators have been developed that provide nearshore outputs at a 1km resolution approximately along the -5mODN contour of England. A nearshore data set comprising many thousands of extreme wave and water level events has therefore been created. To demonstrate the application of this data set, wave overtopping rates have been calculated using SWAN 1D and Bayonet linked to the nearshore multivariate data set. These results have been compared with the results obtained from the widely applied joint

probability contour method. The results of this analysis show the traditional method to underestimate flood risk. This confirms the findings of previous analyses.

It is suggested the dataset and methodology described here can be widely used for a range of purposes including; flood risk analysis, sea level rise impact analysis, coastal flood forecasting and the design of coastal structures.

## 7 References

1. Bruun, J.T. and Tawn, J.A., 1998. Comparison of approaches for estimating the probability of coastal flooding. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **47(3)**: 4
2. Hawkes, P.J., Gouldby, B.P., Tawn, J.A. and Owen, M.W., 2002. The joint probability of waves and water levels in coastal engineering design. *Journal of Hydraulic Research*, **40(3)**: 241-251
3. De Michele, C., Salvadori, G., Passoni, G. and Vezzoli, R., 2007. A multivariate model of sea storm using copulas, *Coast. Eng.* **54 (10)**, Pgs 734–751.
4. Salvadori G., Michele C.D., Durante F., 2011 On the return period and design in a multivariate framework *Hydrology and Earth System Sciences*
5. Corbella S. and Stretch D., 2013. Simulating a multivariate sea storm using Archimedean copulas, *Coastal Engineering*, **76**, Pg. 68–78, June
6. Serinaldi, 2015 Dismissing Return Periods, *Stochastic Environmental Research and Risk Assessment*, **29(4)**, pp 1179-1189..
7. Heffernan J. E. and Tawn J. A., 2004. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **66(3)**, 497–546
8. Lamb, R. *et al.*, 2010. A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, **3(4)**: 323-336.
9. Wyncoll D. and Gouldby B., 2014 Application of a multivariate extreme value method to flood risk analysis, *J. Flood Risk Man.* **8(2)**, 142
10. Gouldby B, Mendez F, Guancho Y, Rueda A and Minguez R.m 2014 A methodology for deriving extreme nearshore sea conditions for structural design and flood risk analysis, *Coast. Eng.* **88**
11. Booij, N., Ris, R.C. and Holthuijsen, L.H., 1999. A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research: Oceans*, **104(C4)**: 7649-7666.
12. O'Hagan, A., 2006. Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, **91(11)**: 1290-1300.
13. Camus, P., Mendez, F.J. and Medina, R., 2011a. A hybrid efficient method to downscale wave climate to coastal areas. *Coastal Engineering*, **58(9)**: 851-862.

14. Camus, P., Mendez, F.J., Medina, R. and Cofiño, A.S., 2011b. Analysis of clustering and selection algorithms for the study of multivariate wave climate. *Coastal Engineering*, **58(6)**: 453-462.
15. Goda Y 2000, Random seas and the Design of Maritime Structures, World Scientific Publishing, ISBN 981-02-3256-X
16. Battjes, J . A., and J.P. F. M. Janssen 1978 Energy loss and set-up due to breaking of random waves, Proc. 16th International Conference on Coastal Engineering, p, 569-587, *Am. Soc .of Civ. Eng.*,New York
17. Kingston, G., Robinson, D., Gouldby, B. and Pullen, T., 2008. Reliable prediction of wave overtopping volumes using bayesian neural networks. . FLOODrisk 2008, Keble College, Oxford, UK.
18. Van Gent, M.R.A., Van den Boogaard, H.F.P., Pozueta, B. & Medina, J.R. 2007. Neural network modelling of wave overtopping at coastal structures. *Coastal Engineering* **54(8)**: 586-593.