

Spatial analysis and simulation of extreme coastal flooding scenarios for national-scale emergency planning

David Wyncoll^{1,a}, Ivan Haigh², Ben Gouldby¹, Dominic Hames¹, Stefan Laeger³, Andy Wall⁴, Peter Hawkes¹ and Anthony Hammond⁵

¹HR Wallingford, Howbery Business Park, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BA, UK

²Ocean and Earth Science, National Oceanography Centre, University of Southampton, European Way, Southampton, SO14 3ZH, UK

³Environment Agency, Manley House, Kestrel Way, Exeter, Devon, EX2 7LQ, UK

⁴Natural Resources Wales, Wales, UK

⁵Environment Agency, Thames Barrier, Eastmoor Street, London, SE7 8LX, UK

Abstract. The UK has a long history of coastal flooding, driven by large-scale low-pressure weather systems which can result in flooding over large spatial areas. Traditional coastal flood risk analysis is, however, often undertaken at local scales and hence does not consider the likelihood of simultaneous flooding over larger areas. The flooding within the UK over the Winter of 2013/2014 was notable both for its long duration, lasting over two months, and its spatial extent, affecting many different areas of England and Wales. It is thus apparent that to plan and prepare for these types of extreme event it is necessary to consider the likelihood of flood events arising at different locations simultaneously (i.e. to consider the spatial dependence of extreme flood events). This paper describes the application of a state-of-the-art multivariate extreme value methodology to extreme sea levels and wave conditions around the coast of England and Wales. The output of the analysis comprises a synthetic set of extreme but plausible events that explicitly captures the dependence between sea conditions at different spatial locations around the coast. These simulated extreme events can be used for emergency management and advanced flood risk analysis.

1 Introduction

The Winter of 2013/2014 saw widespread coastal flooding, with extreme events recorded simultaneously at a number of locations across the UK. These events brought the challenges of emergency response and management into sharp focus and served as a reminder of the vulnerability of coastal communities to flooding. There have been numerous significant historical coastal flood events that have been observed over the past century or more [1]. These events vary in terms of the spatial areas affected and there is a general trend that indicates a significant spatial correlation with increasing severity of event at any one location. This is indicated by Figure 1 which shows the number of sites recording sea levels with a return period greater than five years at one or more of the Environment Agency's "A" Class sea level recording network locations, against the maximum return period recorded at any single location. This indicates, for example, five occasions where ten or more gauges record an extreme sea level, approximately a quarter of the operational tide gauges. These events are often associated with well-known and understood phenomena, whereby particular storm tracks can induce widespread coastal flooding. This can often result in extreme sea level events along large stretches of coastline, particularly when they coincide with high water

of large spring tides. This is demonstrated in Figure 2 which shows storm tracks associated with large surges that propagate from north to south down the east coast.

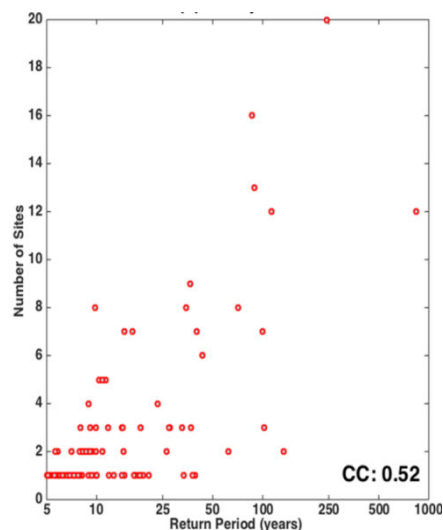


Figure 1. The number of Environment Agency "A" Class sea level sites recording events with a return period greater than five plotted against the maximum sea level return period at any one site.

^a Corresponding author: d.wyncoll@hrwallingford.com



Figure 2. Storm tracks associated with east coast surges.

As noted, however, the spatial “footprint” associated with any single event is not necessarily restricted to a particular geographical region. Figures 3 and 4 show the storm track and the associated extreme event footprint of an event that affected the south, west and north-east coasts of the UK in January 2014. This shows a large low pressure region over the northern parts of the UK with a storm track passing to the west of Ireland. Coinciding with large spring tides, this resulted in extreme tides being recorded at most tide gauges across the UK in the regions affected.

For flood risk analysis and emergency management planning it is desirable to be able to predict and simulate coastal flood events that are more extreme than those that have occurred in the past. This is particularly important when developing adaptive flood risk management strategies in relation to sea level rise and climate change.

To date, however, extreme value analysis of sea levels and extreme coastal wave conditions has tended to have been localised in nature and not focused on the larger-scale spatial correlation of extremes between sites. This imposes limitations on the information that is available to help support emergency planning decisions, which needs to consider actions to mitigate widespread flood events. In addition, the insurance sector requires estimates of losses associated with specific widespread event scenarios that are likely to impact multiple sites.

Historically, these limitations have tended to be imposed because of restrictions within the range of statistical models that are capable of extrapolating to extremes of multiple variables (i.e. extreme sea levels and wave conditions at multiple sites). Restrictions on the flexibility of multivariate extreme value methods to capture the complex dependence structure between such variables were a dominating factor. More recently,

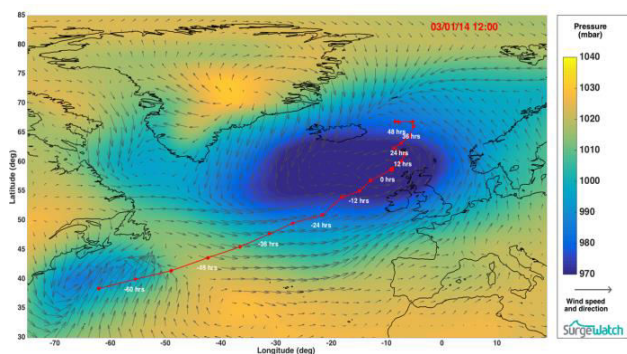


Figure 4. Storm track associated with the January 2014 extreme event.

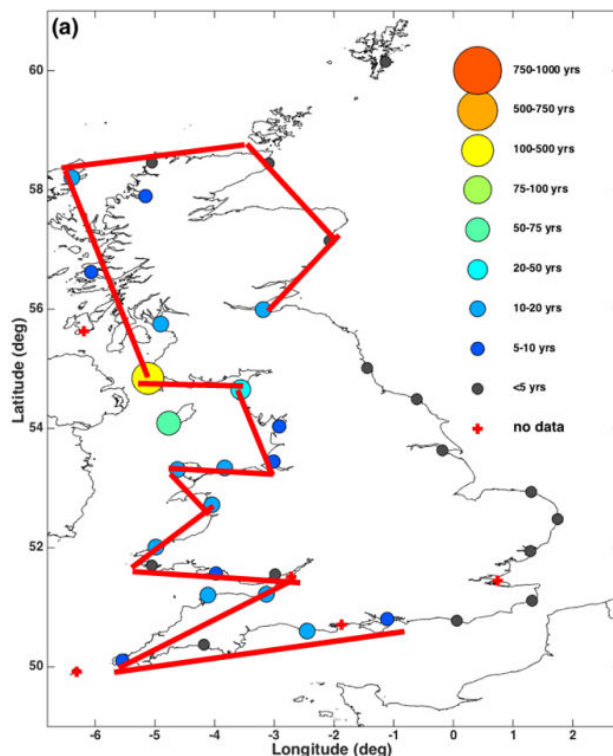


Figure 3. Extreme sea level return periods associated with the January 2014 extreme event.

advances in multivariate extreme methods [2], have been shown to be capable of capturing these complexities in the context of flood risk [3–6]. The approach described here extends these previous analyses to undertake the generation of spatially coherent extreme multivariate events of sea conditions (sea levels and wave conditions) around the coastline of England and Wales.

The following sections provide information on the study area extents and data sources, the multivariate extremes methodology, verification of the generated synthetic events and, finally, conclusions.

2 Study area and data

2.1 Study area

The overall aim of the analysis described here to is to provide multivariate extreme sea conditions offshore of the coast of England and Wales, for the purposes of supporting emergency management planning and flood risk analysis.

To implement the method, the coastline of England was sub-divided into 24 different regions with each region comprising a SWAN wave transformation model domain obtained from a recent national flood risk assessment undertaken for the Environment Agency. In further work carried out for Natural Resources Wales (NRW), these model grids were extended to cover Wales, so the total number of regions considered was 32, as shown in Figure 5.

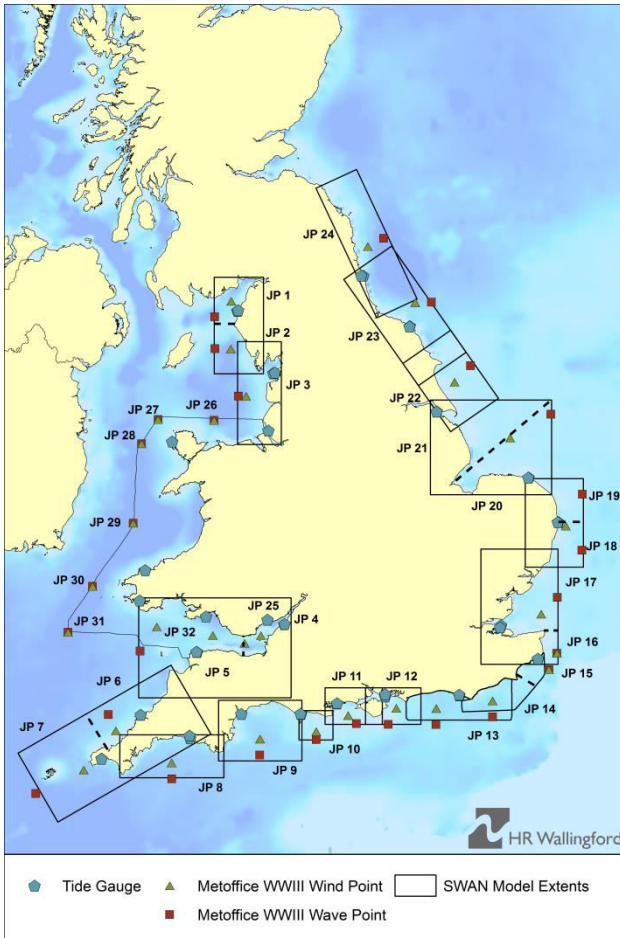


Figure 5. SWAN model domains with wave, wind and sea level data points covering England and Wales.

The seven variables considered in this analysis are: significant wave height, wave period, wave direction, directional spreading, wind speed, wind direction and sea level. Wave and wind data was obtained from a hindcast of wave conditions using the WaveWatch III Model (WWIII), undertaken by the UK 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 and Wales. 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) reanalyses. 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 5. Prior to implementation within the multivariate analysis, the water level data was de-trended and updated to present day levels. Gaps in the surge data were filled using multiple linear regression analysis.

2.2 Event definition

Of the variables considered in the analysis, only significant wave height, wave period, wind speed and sea level require extrapolation to extreme values.

Following the approach of [7], the wave period T may be described as a function of the wave height H_s and wave steepness s given by:

$$s = \frac{2\pi H_s}{g T^2}. \quad (1)$$

Since the steepness is expected to lie within a narrow range for large wave heights, it can be treated as a dependent but non-extreme variable in the analysis in place of the wave period to reduce the dimension of the extreme value analysis. The sea level is similarly replaced by the skew surge in the extreme value analysis to remove the astronomical tidal component.

The extreme value analysis, therefore, aims to extrapolate significant wave height, wind speed and skew surge at each of the 32 offshore locations while maintaining the dependencies between the remaining non-extreme variables, including wave steepness. Before this analysis can be performed, datasets of independent joint peak events must be extracted from the raw time series.

Since extreme wave, wind and surge events do not necessarily peak concurrently, the peaks of each of these variables are identified separately to form distinct sets of peak events for each extreme variable. This allows each type of extreme event to be extrapolated separately to avoid defining a single type of joint event under a false assumption that the variables peak concurrently. Peak wave, wind and surge events are each identified at each location as local maxima separated by at least 2.5 days (the average storm length). In each case, the peak in the primary variable is paired to the concurrent values for the remaining variables at the same location after accounting for the approximate travel time of the offshore waves to the nearshore water level gauge. Since skew surges are defined only at every high tide, the nearest value is taken when matching the peak of another variable.

Individual extreme events cannot be expected to affect all parts of the coastline, so it is difficult to define a single set of peak events that peak at all locations. Instead, a separate set of joint events is identified for each location and extreme variable, each representing peaks of a single variable at a single location with corresponding values at all other locations and for all other variables. To do this, the identified peaks for the primary variable at the primary location are matched where possible to a previously identified event at a neighbouring location that peaks within a single-day window and this is continued in both directions along the coastline. When no peak is found within the search window, the concurrent values are taken instead and, in either case, the remaining variable values are taken as concurrent with the primary variable peak.

This approach produces a separate set of joint events for waves, winds and surges at each location. Each represents independent peak events for a single variable at a single location paired appropriately with values for all other variables and at all other locations. Each dataset is used only to extrapolate the peaks of the same primary variable and location while modelling the dependence between the remaining variables and locations.

3 Multivariate extreme value analysis

The aim of the multivariate extreme value analysis is to extrapolate the identified joint events to extreme values while preserving the dependencies both spatially around the coastline and between variables. The peak wave height events are used to extrapolate wave heights to extremes and similarly for wind speed and skew surge. The Heffernan and Tawn approach [2] is used to model the dependencies between the extreme variables as each primary variable is extrapolated. The extension of [7] is used to maintain dependencies between the remaining non-extreme variables.

Since variables such as wave height are known to have strong dependencies on season as well as direction, it is important to account for these too. The output of this analysis is a large Monte Carlo sample of potential peak events that each provide a value for all variables at all offshore locations.

3.1 Pre-processing

To account for the known seasonal and directional dependencies in the three extreme variables, the data for these variables is de-trended with respect to season and direction prior to use in the extreme value analysis. By tracking the seasons and directions through the analysis with the other non-extreme variables, the de-trending can be reversed to return the Monte Carlo simulation to the original scale.

The significant wave height is de-trended first by season and then by wave direction. Wind speed is similarly transformed using the wind direction and skew surge is de-trended only by season. In each case, the de-trending is carried out by subtracting a smooth mean and dividing by a smooth standard deviation estimate that are each created from Gaussian kernel smoothers fitted to the peaks of each variable. For this, a continuous season variable is defined from the date and time of the primary peak as a fraction within a year. Care is taken to ensure the smooth estimates are continuous over the directional 360°–0° and seasonal 1–0 boundaries by appropriately taking account of the periodicity in the de-trending variables.

3.2 Marginal extreme value analysis

As with any copula-based extreme value analysis, the marginals of the transformed extreme variables are first analysed separately before modelling the dependencies between variables. For this, the standard peaks-over-threshold (POT) approach of [8] is applied whereby peaks of each variable above a high threshold are fitted to the Generalised Pareto distribution (GPD).

The GPD is fitted separately to the transformed peaks of each of the extreme variables at each offshore location using penalised maximum likelihood. A penalisation term mirroring a $N(0,0.5)$ prior is used to ensure the shape parameter is close to zero to provide robust estimates. For skew surge, the GPD is fitted only to data

calculated from observed surges and not those estimated using multiple linear regression.

Standardised parameter estimate and mean residual life plots are used to select a robust threshold for each fit that is high enough for the GPD limit to hold but low enough to provide enough data for a stable fit. Probability-probability and quantile-quantile plots are used to assess the goodness-of-fit.

The GPD fit provides a statistical model only for values above the threshold. This is supplemented with the empirical distribution of the transformed data below the threshold to define a complete distribution for each margin, see [9].

3.3 Extremal dependence model

In order to separate the marginal characteristics from the dependence analysis, the three extreme variables at each location are transformed further to common marginal distributions by applying the probability integral transform. Following the approach of [10], common Laplace margins are used.

The Heffernan and Tawn approach [2] consists of modelling pairwise dependencies on common scales between a single variable above a high threshold and all remaining transformed variables. If \mathbf{Y}_{-i} denotes the vector of all transformed extreme variables excluding a particular peak variable Y_i , the following multivariate non-linear regression model is applied to \mathbf{Y}_{-i} where $Y_i > v$:

$$\mathbf{Y}_{-i} = \mathbf{a} Y_i + Y_i^b \mathbf{Z} \quad \text{for } Y_i > v \quad (2)$$

where \mathbf{a} and \mathbf{b} are vectors of parameters, v is a specified threshold, \mathbf{Z} is a vector of residuals and the vector algebra is applied component-wise. The regression parameters are estimated pair-wise using maximum likelihood under the temporary assumption that the residuals follow a Gaussian distribution with unknown mean and standard deviation. As with the marginal GPD fits, parameter stability plots are used to select the thresholds to ensure robust estimates of each parameter.

Once fitted to all peak observations of $Y_i > v$, an empirical distribution of residuals \mathbf{Z} is constructed. Each residual is augmented with the non-extreme variables associated with the observation Y_i following the approach of [7]. This is repeated to condition upon each of the three extreme variables at each of the 32 offshore locations.

3.4 Monte Carlo simulation

The fitted models each describe the dependence between all remaining variables at all locations when a single peak variable at a single location is extreme. Together they describe the full distribution of spatial and between-variable dependencies when at least one peak variable is extreme. When accompanied by the empirical distribution of all non-extreme peak events, the models can be applied to simulate a large dataset of synthetic events with appropriate extrapolations to extreme values.

For a selected peak variable Y_i , a new extreme value is sampled from the Laplace distribution above a fixed

sampling threshold. This is paired with a residual resampled from the empirical distribution which is used to calculate the remaining extreme variables by inverting Equation 2. The non-extreme variables associated with the residual are each copied to the new sample as in [7].

Rejection sampling is used to ensure the peak variable is most extreme to avoid double-counting events where more than one variable is extreme. The observed proportion of identified events for which each variable is extreme is used to determine the proportion of samples to produce from each conditioning model.

The probability integral transform is used to convert the marginal distribution of each extreme variable from the Laplace scale to the de-trended scale with the fitted GPD tail. The associated season and directional variables are then used to apply the seasonal and directional trends by reversing the de-trending process using the same smoothed mean and standard deviation. The wave height and steepness are used to calculate the wave period by inverting Equation 1.

The empirical distribution of high tides that are seasonally close to each simulated event is resampled and added to the skew surges to assign peak sea level values to each event. The resulting simulated dataset then consists of peak wave height, wind speed and sea level events with appropriate dependencies between all variables at all offshore locations. This may be combined with a resampled subset of observed events below the sampling threshold in all variables to produce a complete simulated dataset of all offshore variables.

Figure 7 provides pair-wise scatter plots of simulated significant wave height values at a subset of the 32 locations which are indicated on the diagonal maps. It is evident that the correlation between waves at nearby locations is high, unlike the dependence with other coastlines. Figure 6 similarly compares the values of a selection of fitted variables at a single location. This shows a strong correlation between the wave height and wind speed at this location as well as a strong dependence on direction.

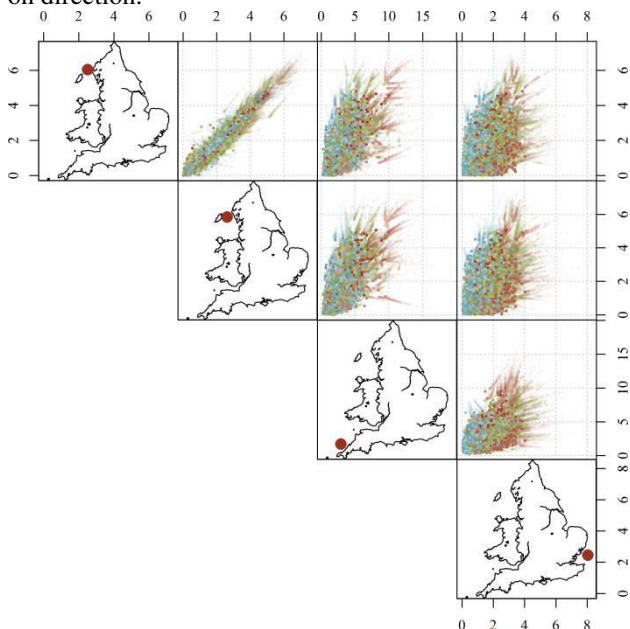


Figure 7. Pair-wise scatter plots of simulated significant wave height values at a subset of the 32 locations.

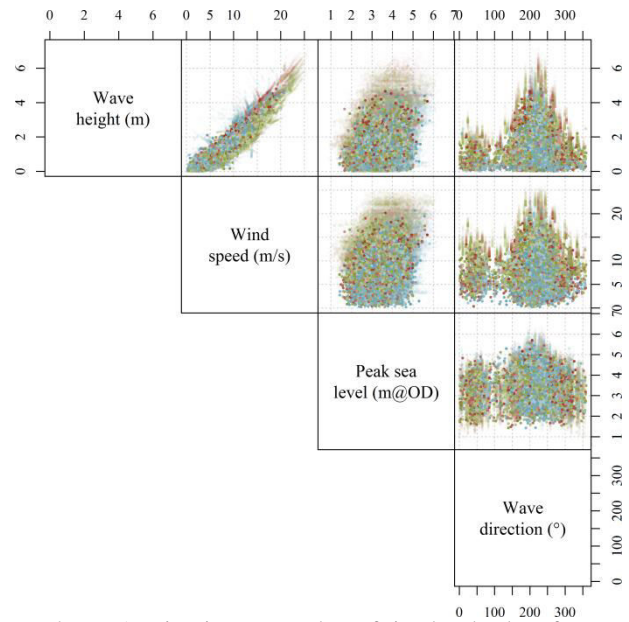


Figure 6. Pair-wise scatter plots of simulated values for a selection of variables at a single location.

4 Validation

Verifying statistical models that extrapolate historical observations to extreme values is always challenging. Having a robust theoretical motivation for applying the method is essential. The statistical model applied here has a strong theoretical justification and has been applied in the context of flood risk analysis, on a number of previous studies [3–6]. This helps lend confidence to the application of the method here.

One of the main characteristics of the new model is to capture the spatial dependence within the extremes of different variables at different locations. The historical database of extreme events can aid the verification process by comparison with the synthetically generated events. Figure 8 shows a comparison between a synthetic event that has been generated by the statistical model and an historical observed event. Visual inspection shows the spatial “footprint” of the events to be similar but the synthetic event to be more extreme. This gives an indication the statistical model is capturing the spatial dependence in the extremes as intended.

5 Conclusions

There is a long history of coastal flooding within the UK. It is well-known that coastal flooding can affect large areas, which can pose challenges for the development of emergency response and management plans. To date, coastal flood risk analysis has focused on local scales, or made simplifying assumptions regarding the spatial correlation associated with extreme events at larger spatial scales. This has posed limitations regarding the use of this type of information to help inform high level emergency management plans and risk analysis for the insurance sector, for example. The analysis undertaken here has sought to apply an advanced multivariate extreme value model to sea conditions

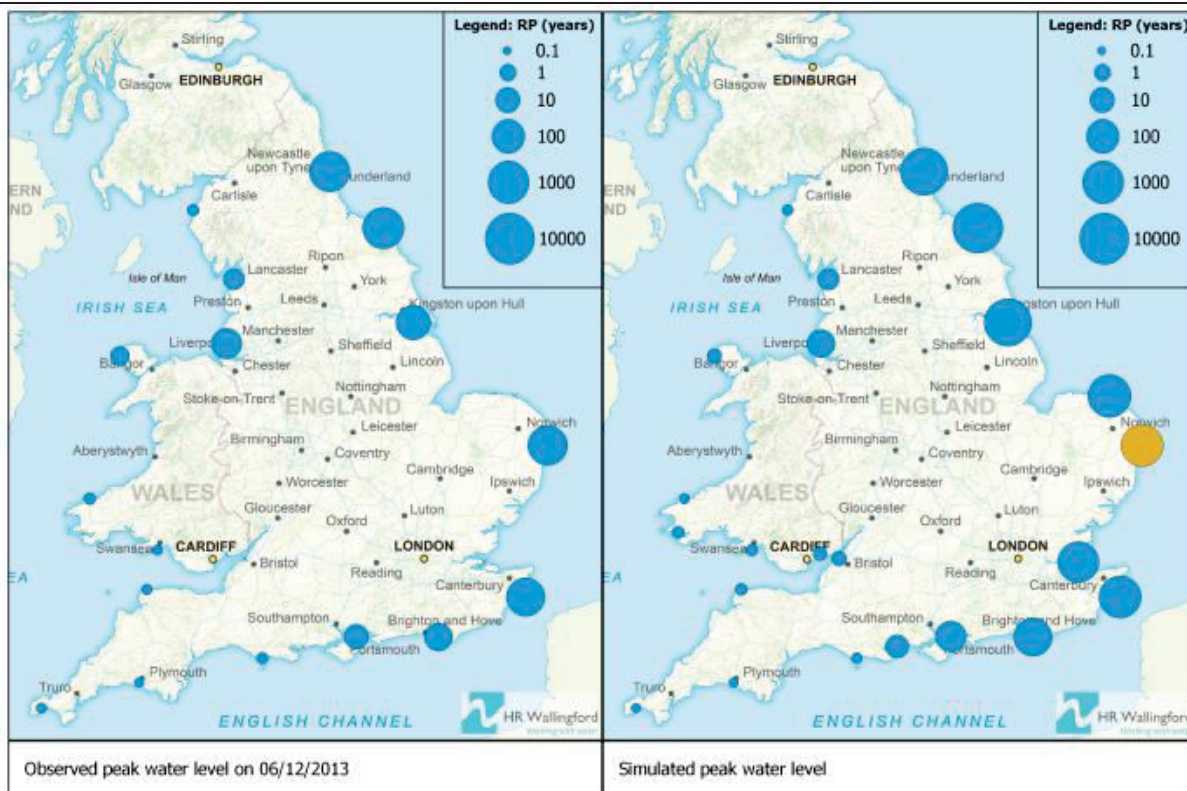


Figure 8. Spatial comparison between an observed event (left) and a more extreme simulated event with similar characteristics.

around the coast of England and Wales. The statistical model enables historical observations of sea conditions to be extrapolated to extreme values ensuring the spatial dependence between extremes is captured. The methodology results in the generation of a synthetic set of extreme sea condition events. An historical database of coastal flood events has been used to aid the verification of the output of the statistical model. These synthetically generated events can be used to form the boundary conditions for coastal flood simulations. The flood simulations are more realistic in terms of the spatial extreme characteristics than other similar simulations that have been undertaken to date. These simulations can be used to provide more advanced information to aid the development of emergency action plans and risk analysis studies.

6 References

1. Haigh I. D., Wadey M. P., Gallop S. L., Leohr H., Nicholls R. J., Horsburgh K., Brown J. M. and Bradshaw E. (2015). A user-friendly database of coastal flooding in the United Kingdom from 1915–2014. *Scientific Data*, **2**
2. 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
3. Lamb R., Keef C., Tawn J., Laeger S., Meadowcroft I., Surendran S., Dunning P. and Batstone C. (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
4. Environment Agency (2011). 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*
5. Wyncoll D. and Gouldby B. (2015). Integrating a multivariate extreme value method within a system flood risk model, *Journal of Flood Risk Management*, **8(2)**, 145–160
6. Gouldby B., Mendez F., Guanache 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
7. Wyncoll D., Gouldby B. and Hames D. (2013). Application of a conditional approach for multivariate extreme values to flood risk. *Proceedings of the 1st International Short Conference on Advances in Extreme Value Analysis and Application to Natural Hazards, Germany, Sept 2013*, 87–96
8. Davison A. C. and Smith R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **52(3)**, 393–442
9. Coles S. G. and Tawn J. A. (1991). Modelling extreme multivariate events. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, **53(2)**, 377–392
10. Keef C., Papastathopoulos I. and Tawn J. A. (2013). Estimation of the conditional distribution of a multivariate variable given that one of its components is large: Additional constraints for the Heffernan and Tawn model, *Journal of Multivariate Analysis*, **115**, 396–404