07 - NEW STANDARDS AND METHODS FOR VALIDATING COASTAL FLOOD FORECASTING SYSTEMS

Dr Mark Lawless[1] and Clive Powis[2]

[1] *JBA Consulting, Broughton Hall, Skipton, North Yorkshire, BD23 3AE, England*

[2] *Plan B (UK) Limited, 5 Raglan Villas, Fairfield Park, Bath, BA1 6HZ, England.*

Abstract

The general approach for the development of regional coastal flood forecasting systems has remained unchanged for nearly a decade. Whilst there have been improvements in the quality of the underlying models and data used in the development of these systems, many are still known to perform poorly. This is especially the case from the perspective of the prediction of wave overtopping. Unlike fluvial flood forecasting systems development, little has been done in terms of long-term performance testing for coastal flood forecasting systems. This has led to relatively low confidence in system performance. This paper discusses datasets available and recommends new methods for evaluating the performance of regional coastal flood forecasting systems using long archives of meteorological forecast data now available. By testing a forecasting system against these archives, it is possible to evaluate how it would have performed had it been in operation historically. This information can then be used to further calibrate and validate the system, providing a level of operational confidence that has not been possible in the past.

1. INTRODUCTION

The development of regional coastal flood forecasting systems normally involves the creation and use of a suite of coupled oceanographic models (i.e. wave transformation, wave overtopping, flood inundation models). Rather than running these models live for forecasting purposes, they are used to pre-compute the flooding consequences associated with a wide range of potential combinations of key storm driving variables such as sea-level, offshore wave properties and wind conditions. For instance, recent flood forecasting systems have included the simulation of upwards of 10,000 ensemble combinations of these variables. These pre-computed simulations are then used to develop look-up tables that relate offshore forecasts of sea-level, wind and wave conditions (supplied by the UK Met Office) to the expected flooding consequences on land.

Whilst this general approach for developing coastal flood forecasting systems has been in practice for some time (e.g. Dublin Coastal Forecasting System, NW England Forecasting System), these systems have normally been implemented with little or no historical performance testing, limiting confidence in the system. In particular, some systems are known to perform poorly from the perspective of wave overtopping prediction, the key source of coastal flood risk in many places. This paper discusses datasets available and recommends new methods for evaluating the performance of regional coastal flood forecasting systems using long archives of meteorological forecast data now available. By testing a forecasting system against these archives, it is possible to evaluate how it would have performed had it been in operation historically. This information can then be used to further calibrate and validate the system, providing a level of operational confidence that has not been possible in the past.

Whilst some of the methods discussed herein are standard practice with respect to the development of fluvial flood forecasting systems (Environment Agency 2007), they have not traditionally been used for coastal flood forecasting system development due to a number of barriers discussed within. At present, tailored formal best practice guidance is not available with respect to the development of regional coastal flood forecasting systems. It is recommended that a study is commissioned to prepare such guidelines, drawing on the concepts and methods discussed herein.

2. WHAT DRIVES COASTAL FLOOD RISK?

Before discussing how coastal flood forecasting systems can be developed and tested, it is worth reviewing the processes that drive coastal flood risk, thereby setting the context for the system requirements. Figure 1 illustrates the main components of sea-level variation that contribute to coastal flooding during a storm event. The still water sea-level is comprised of the underlying astronomical tide and any storm surge affects. These two components determine the average sea-level for a particular location at a particular time. Whilst this variable is very important in terms of coastal flooding, still water-induced flooding is normally limited to sheltered locations such as tidal rivers and harbours. Not surprisingly, the sea is not "still" during a storm event for more exposed locations. For these locations, most flooding occurs through wave action, rather than still water flooding.

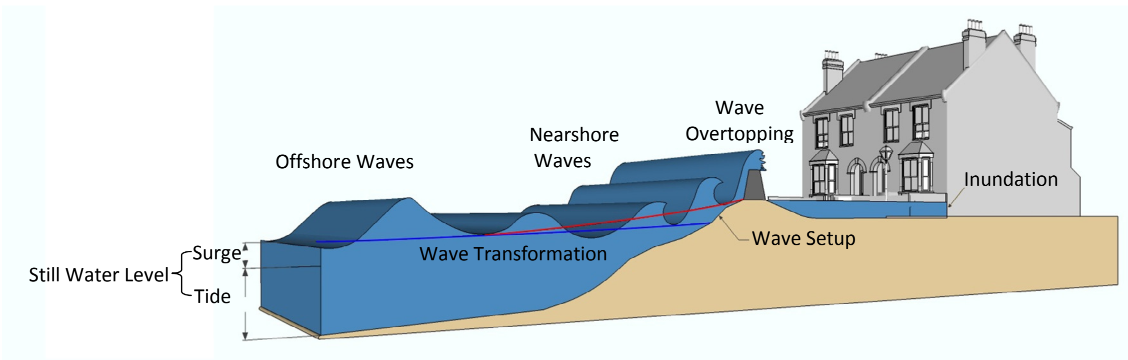

Figure 1: *Components of Sea-level Variation that lead to Coastal Flooding*

Wave action is a complex process controlled by many contributing factors. The manner in which these factors combine determines the magnitude of any wave-induced flood impacts. Waves are generated in deep water by strong winds and then propagate towards land. As they do so, they enter shallower bathymetry where wave transformation processes occur, including shoaling, diffraction, refraction, depth limitation and breaking. The consequence of these processes is that the properties of the waves, when they reach the base of coastal flood defences, are quite different to the waves in deep water. It is these nearshore waves that are of most importance in terms of flood forecasting because they interact with beaches and defences and lead to wave overtopping.

The magnitude of wave overtopping is controlled by the state of the sea (i.e. depth, nearshore wave properties), the geometry of the beach and local flood defences, and the directional characteristics of the storm (i.e. wave and wind direction). Once wave water overtops a defence, flood inundation commences. The nature of this inundation is related to the magnitude of the overtopping, any additional still water flooding and the characteristics of the land and its drainage. Coastal flooding can also be exacerbated by fluvial or surface water flooding.

3. COASTAL FLOOD FORECASTING SYSTEM DEVELOPMENT

No one numerical model is capable of simulating all of the processes that drive coastal flood risk simultaneously. The development of regional coastal flood forecasting systems therefore normally involves the development and use of a range of models and data inputs. The cumulative results from these components are then drawn together to forecast the expected consequences on land. Figure 2 provides a schematic representation of the general approach used.

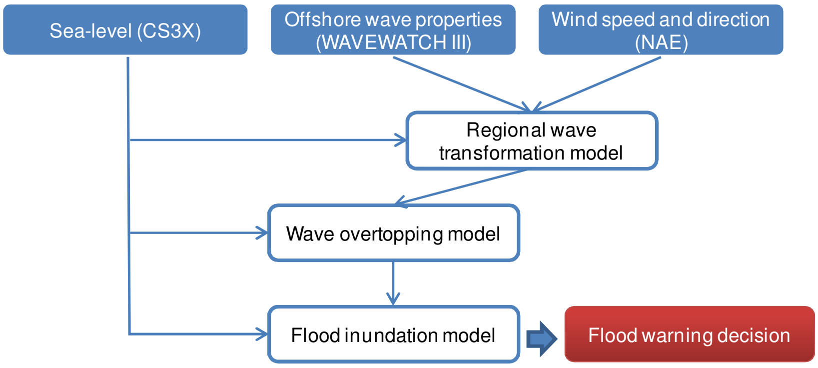

***Figure 2**: Coastal Flood Modelling System Architecture*

The starting point for most coastal flood forecasting systems in Ireland and the UK is 36 hour forecasts of sea-level, offshore wave conditions and wind conditions provided by the Storm Tide Forecasting Service (STFS), run by the UK Met Office. These forecasts comprise of the following components:

- **Sea-level**. Storm surge magnitude is currently forecasted 36 hours into the future 4 times a day using the 12km resolution CS3X model developed by the National Oceanographic Centre (NOC) in the UK. This model is forced with meteorological data provided by the Met Office's 12km grid North Atlantic European (NAE) weather forecast model. Total sea level at standard ports is determined by adding the storm surge values from the CS3X model to predicted tide levels derived from tidal harmonics.
- **Offshore wave height, period and direction** are also currently forecasted 36 hours into the future 4 times a day using the 12km resolution WAVEWATCH III model run by the UK Met Office. This model is also forced using weather data from the NAE model.
- **Wind speed and direction** are currently forecasted 36 hours into the future 4 times a day using the NAE model.

Whilst the information provided through the STFS is a critical component of any coastal regional coastal flood forecasting system constructed for Ireland and the UK, this information is only a component part. In particular, the wave information provided is for offshore waves only; no specific information is provided in terms of nearshore wave conditions, wave overtopping or flood inundation. To develop a regional system that accounts for these processes, it is normally the case that three additional model components are required to be constructed, calibrated and validated. These include:

- A spectral **Wave Transformation Model**, used to transform offshore waves from the WAVEWATCH III model to the toe of beaches and flood defences.

- **Wave Overtopping Models**, used to compute wave overtopping as a function of sea-level (from CSX3), nearshore wave height, period and direction (from the WM model) and beach and defence geometry. Wave overtopping is usually computed using the methods contained in the European Wave Overtopping Manual (EurOtop, 2010).

- A **Flood Inundation Model**, used to evaluate how flood water originating from still water flooding and/or wave overtopping will spread on the land.

Due to the lengthy run-times of these regional models, it is not normally possible nor sensible to run them live within a forecasting system. Rather, the standard practice is to run these models offline in order to pre-compute the consequences associated with a wide range of potential combinations of the key flood risk drivers (i.e., sea-level; offshore wave height, period and direction; wind speed and direction). In recently developed systems and a number currently under development for the Environment Agency for England and Wales, up to 10,000 ensemble combinations of these driving variables and their associated flooding consequences have been simulated. Whilst it may not be necessary to simulate this high density of simulations for all locations, it is important to ensure that the simulations undertaken cover the full range of potential storm driving variables combinations possible. This should cover simulations ranging from those that represent the initiation of flooding through to the maximum possible flooding.

Once all of the ensemble simulations have been computed, the information is used to produce 'look-up tables' that relate the forecasts provided through the STFS (i.e. sea-level, offshore wave conditions) to the closest match in the ensemble database. The look-up tables then returns the expected consequences in terms of nearshore sea-level, nearshore wave conditions and wave overtopping. This information is then compared to pre-determined thresholds for operational measures, and the issuing of Flood Alerts, Flood Warnings. With this information, a Flood Warning Duty Officer can then decide what action should be taken.
The above approaches have been in operation largely unchanged for nearly a decade; although there have been significant advances in terms of model quality and resolution and the density of the ensemble simulations. A common criticism of previous systems has been that despite the sophistication of the modelling, the results do not tend to reflect reality (i.e. incorrectly forecasting events or missing real events). The key issue associated with the development of previous regional coastal forecasting systems has been that they have not been properly tested in terms of their long-term performance; testing that is the key to refining forecasting systems and improving their reliability. This omission has been related to a perception that relevant data do not exist for long term testing combined with a lack of appropriate tools. In the next section, new methods for undertaking long-term performance testing currently under development are outlined, including a discussion of the available data.

4. PERFORMANCE TESTING

The key to performance testing for coastal forecasting systems is data. Whilst it is important to calibrate and validate the individual model components of the system as best possible using available recorded data (e.g. sea-level and wave buoy data), calibrating and validating the overall forecasting system requires a different approach. Each of the components in the system is associated with high levels of uncertainty, and by coupling these components together, there is an escalating potential for error in the forecasts. It is therefore essential to ensure that the final outputs of the system (i.e. the alerts, alarms and flood outlines), do indeed reflect reality when tested against the right kind of historical data. In the case of coastal flood forecasting systems, the only inputs are the offshore forecasts from the STFS (sea-level, offshore wave properties). Performance testing is therefore ideally done by driving the system with long archives of these forecasts. By testing a forecasting system against these archives, it is possible to determine how it would have performed had it been in operation historically. This information can then be used to calibrate and validate the overall system, improving its overall performance.

4.1 Available Data

Data that are now available for long-term performance testing in Ireland and the UK include:

- Sea-level: Archives of surge forecasts from NOC dating from 1990 onwards for standard ports. These data are provided as a seamless time-series. They are stitched together by extracting the data from Time T+6hr to T+12hr (where T=0 is the time of issue of the forecast) from each of the four forecasts simulated each day.
- Wave and wind: Archives of wave and wind data are also available from the UK Met Office, dating back as far at 1990. Whilst no one dataset is available that covers this time period, a time-series of data can be assimilated using the following sources:
    - *WAVEWATCH III.* Archives of offshore wave and wind forecasts are available from this model for the period 2008 to 2012. These data are stitched together by extracting data from time T-6hr to T=0 from each of the four forecasts issued each day [where T=0 is the time of issue and time T-6 is the six hours leading up the forecast (discussed further below)]. The wind data provided as part of this dataset are from on the NAE model.
    - UK Waters Wave Model. Archives of offshore wave and wind forecast data are available from this model for the period 2000 to 2008. These data are also stitched together by extracting data from time T-6hr to T=0 from each of the four forecasts issued each day.
    - European 2G Model. Archives of offshore wave and wind forecast data are available from this model for the period 1990 to 2008. These data are stitched together by extracting data from time T-12hr to T=0 from each of the two forecasts issued each day.

By assimilating the above data, a performance analysis can be undertaken using more than 20 years of data. Whilst this is a major improvement in terms of coastal flood forecasting performance testing, it is not without limitation. Some of the key limitations are as follows:

- For all of the above wave models, the archive data are from the forecast period T-6hr to T=0 (or T-12 for the European 2G model). This means that these data are not, in fact, true forecast data. When a forecast is issued, the time period associated with these data has already passed. Furthermore, this is the period termed the data assimilation period, meaning that the model results are influenced by observed data. This means that the data are certain to be more accurate than the data contained in the time T=0 to T+36 hours (i.e. the true forecast data). Nevertheless, these data provide the best available source of data that can be used to drive the performance testing and provide a very good approximation of how the overall system would have performed (based on methods described further below). Of most importance is that any systematic biases in the models should be able to be identified and incorporated into the forecasting system (discussed further below).

- It is also important to highlight that the assimilated data will not all be from the latest operational models, i.e. WAVEWATCH III and the latest version of the CS3X surge model. Some of the data are from wave models now retired and earlier versions of CS3X. Use of these data, therefore, do not give an exact representation of how the system will perform when put into operation. However, they are deemed to provide the best available source of data that can be used to drive the performance testing.

4.2 Performance Measures

Once the archive forecast data and look-up tables have been assimilated and developed, the performance testing is a relatively simple operation. At JBA a software package called ForeCoast has been developed in cooperation with Plan B (UK) Limited to undertake this analysis. ForeCoast computes historical flood forecasts using available archive forecast data and look-up tables. It then analyses the performance of the system in a number of key ways, as discussed below.

4.2.1 Quantitative Measures

Probability of Detection (POD): POD can be computed to determine how well the system forecasts flood events. It can be determined using both recorded sea-level data and/or nearshore wave buoy data. Associated with POD analysis is a pre-determined threshold at which a Flood Alert or Alarm is issued. With respect to the Alarm threshold, this is generally the ground or defence level at which flooding is known to commence for a community. POD is computed as follows:

$$POD = HIT / (HIT+MISS)$$

where,

HIT is the number of occurrences where the observed sea-level exceeded the Alert or Alarm threshold level and the forecasting system also predicted a sea-level that exceeded the threshold level (i.e. the forecast was correct).

MISS is the number of occurrences where the observed sea-level exceeded the threshold level but the forecasted sea-level did not (i.e. a forecast was not issued but should have been).

Whilst POD analysis is regularly undertaken for fluvial flood forecasting systems, it has rarely if ever been used for coastal flood forecasting systems in Ireland or the UK. Although not published, there is general consensus that the target POD rate for fluvial analysis is generally 0.8. However, anything above 0.7 is considered good.

False Alarm Rate (FAR): FAR is another useful performance indicator, which is computed as follows:

$$FAR = FALSE\ ALARM / (HIT + FALSE\ ALARM)$$

where,

FALSE ALARM is the number of occurrences where the forecasted sea-level exceeded the Flood Alert or Alarm threshold but the observed sea-level did not (i.e. a forecast was issued but should not have been). The target FAR for fluvial flood forecasting systems is 0.2.

Critical Success Index (CSI): CSI is a composite function comprising of both POD and FAR. It is computed as:

$$CSI = POD\text{-}FAR$$

CSI gives a better indication of the overall system performance, including both probability of detection and false alarms. Typical targets for CSI in fluvial forecast systems are of the order of 0.6.

In the absence of any clear guidance with respect to coastal flood forecasting systems, it is recommended that the above targets for POD, FAR and CSI are also used in the development of coastal flood forecasting systems. This would provide an objective measure of system performance.

4.2.2 Qualitative Measures

Quantitative validation data do not exist for wave overtopping forecasts in the way they do for sea-level (tide gauges) or nearshore wave properties (wave buoys). This stems from the fact that observed measures of wave overtopping discharge during storm events are not normally available and are extremely difficult to collect. However, this lack of quantitative data only emphasises the importance of the validation process for wave overtopping, rather than providing an excuse to ignore it. Whilst it is not possible to compute POD, FAR and CSI for wave overtopping forecasts, forecasting systems that include wave overtopping predictions should be validated from a number of additional qualitative perspectives. This process starts at the build stage for the wave overtopping models. Where historic flood event data exist, inundation through wave overtopping should be simulated (i.e. by inputting the overtopping discharges into an inundation model) for the conditions that led to these events and the resultant flood outlines and depths should be compared. If the models re-creates the consequences o the events well, this provides good evidence that they are performing well. If they do not, an iterative process of model improvement is required before commencing the construction of the broader forecasting system. The simulation of flooding for more frequent return periods can also provide insight into whether the wave overtopping modelling is consistent with the historical evidence. It is also imperative that overtopping results are discussed in detail with client and council representatives, drawing on local experience with respect to the frequency and magnitude of historic flood events.

Once there is confidence in the individual wave overtopping models, their results can be incorporated into the overall flood forecasting system. The performance of the overall system should then be further evaluated in the following ways:

- **Event Analysis**: The most important analysis is a simple event analysis evaluating whether key historic flooding events would have been correctly forecast had the system been in operation at the time. This analysis should involve a comparison of the locations and extents of flooding forecast against any available reconnaissance data. Any changes to flood defence configurations since the event will need to be considered in this analysis.
- **Near Miss Analysis:** Evaluating how well the system would have performed during near miss events is perhaps as important as evaluating its performance for actual flood events. Near miss events are defined as occasions when flooding very nearly

happened, but did not. These types of events are very important with respect to the wave overtopping component of the system. Wave overtopping models are notoriously uncertain and this is particularly the case with respect to the prediction of small volumes of overtopping, i.e. those that would be occurring in the initiation stages of an event and more frequent flood events. Near miss events provide a good indication of the transition point between no flooding and flooding. Ensuring that the overtopping forecasts are not overly conservative at the point of the initiation of flooding is very important. The criticism of many previous forecasting systems has been that overtopping is forecasted to occur much more regularly than it has in reality.

- **Frequency Analysis:** A long-term frequency analysis evaluating the annual rate of Flood Warnings and Alerts that would have been issued had the forecasting system been in operation historically can also provide important information on its performance. For instance, if the long-term performance testing indicates that Alerts and Alarms would have been issued several times a year, and in fact flooding only occurs very rarely, then clearly system improvements are required. This type of long-term performance testing is particularly important for systems that include wave overtopping.

5. SYSTEM REFINEMENT

All of the above methods (which are by no means exhaustive) should be used to evaluate and refine the performance of any regional coastal flood warning system before it is put into operation. It is inevitable that the performance testing will reveal that early versions of the system would not have performed as well as desired if they had been in operation historically, and that further calibration and validation is therefore required, involving iterative testing and modifications of the following nature:

- **System de-bugging**: Given the complexity of assimilating all of the data and model components used in the forecasting system, the possibility of system bugs is high. Before any modifications are made to the individual components of the system, it is important to ensure that there are no simple coding errors causing the problems. Errors may have occurred, for instance, in the production of the look-up tables or the system assimilation algorithms.
- **Model Component Checks**. It should be the case that all of the individual regional models used in the development of the system (i.e. wave transformation, wave overtopping, flood inundation) have been fully calibrated and validated before the results have been integrated into the system. However, if the performance testing reveals major issues within the system that are not related to system bugs, then a review of the individual modelling components is required. More often than not, it is likely that the issues lie with the wave overtopping models and re-schematisation of these may be required.
- **Refinement of Alert and Alarm thresholds**. All flood forecasting systems require the assignment of Alert and Alarm thresholds for key variables of sea-level and wave overtopping. For sea-level thresholds these are normally defined based on surveyed levels representing the ground level/water level at which time flooding will commence. From this perspective, there is generally little merit in modifying these threshold levels to improve system performance. If the sea-level thresholds do not appear correct, this is probably an indication that something else is wrong with the

system. There is perhaps more flexibility with respect to the thresholds set for overtopping discharge. These thresholds are more subjective than sea-level thresholds and are often based on recommendations provided in the EurOtop manual. It is often possible to improve the performance of a system by simply testing variations in terms of overtopping thresholds. However, care is required to ensure that these modifications are not simply a way to mask issues in the underlying overtopping models.

- **Identification of biases in archive forecast data**. Every iteration of the regional forecast system will ultimately be driven by the input forecasts coming from the CS3X and WAVEWATCH III models. Therefore, any errors in these forecasts will carry through the system. The iterative system refinement processes proposed herein should be able to identify and account for any systematic biases in these forecasts (e.g. a tendency to over or under predict sea-level or offshore wave properties). However, this is complicated by all of the issues identified in Section 4.1. In particular, different systematic biases may exist for different model versions, complicating the identification of any one trend to account for. Nevertheless, it is important to consider this element and review Met Office and NOC annual validation reports available on-line as part of the system performance testing and refinement stage (Met Office, internet).

Modification and refinement of the forecasting system should be undertaken in full partnership with the operating authority, ensuring that: the methods used are fully understood; the levels of uncertainty and anticipated performance are transparent, and; a mechanism for future system monitoring and performance review is put in place. With respect to the later, a strategy should be put in place for the collection of relevant reconnaissance data to allow annual or post event performance review to be undertaken.

6. CONCLUSIONS AND RECOMMENDATIONS

This paper outlines a general procedure for the development of regional coastal flood forecasting systems, including new recommendations and methods for long-term performance testing. Whilst some of the performance testing measures recommended above have been used regularly for fluvial forecasting systems they have yet to be employed in the coastal environment. This omission is related to a perception that relevant data are not available for this type of coastal analysis and also a lack of appropriate software and tools. As discussed above, up to 20 years of archive forecast data are now available from the UK Met Office and NOC (covering both Ireland and the UK) that can be used to drive coastal forecasting performance testing. As well as recommendations with respect to Probability of Detection and False Alarm Rate analysis, this paper outlines additional qualitative approaches that can be used to further evaluate and refine system performance, particularly with respect to wave overtopping.

The software package ForeCoast, recently developed by JBA and Plan B UK, has specifically been designed to foster the system performance testing methods described herein. This software, in combination with the archive forecast data from the Met Office and NOC, can be used to test and refine the performance of any coastal flood forecasting system, new or old, even if it has not be developed by JBA. ForeCoast, and the methods described herein, are currently being tested as part of the construction of 5 new regional coastal flood forecasting systems under development by JBA, on behalf of the Environment Agency for England and Wales and the Scottish Environmental Protection Agency. It is recommended that following these studies, and the lessons learned from them and others, new best practise guidelines are penned for the development of regional coastal flood forecasting systems. It is hoped that these guidelines can draw on, and further develop, many of the concepts discussed in this paper.

7. REFERENCES

EurOtop (2010) "Wave Overtopping of Sea Defence and Related Structures: Assessment Manual", Overtopping Course Edition, November 2010. HR Wallingford.

Environment Agency (2007) Operational Instruction 145_07 Real time model development.

UK Met Office (internet) http://www.ncof.co.uk/Operational-Ocean-Forecasting.html