

Believe it or not? The challenge of validating large scale probabilistic risk models

Paul Sayers^{1,a}, Rob Lamb², Mike Panzeri³, Hayley Bowman⁴, Jim Hall⁵, Matt Horritt⁶, Edmund Penning-Rowsell⁷

¹ Partner, Sayers and Partners, Fellow, Environmental Change Institute, University of Oxford.

² JBA Trust and Lancaster Environment Centre, Lancaster University

³ HR Wallingford

⁴ Environment Agency, UK

⁵ Professor and Director, Environmental Change Institute, University of Oxford

⁶ Horritt Consulting

⁷ Flood Hazard Research Centre, Middlesex University

Abstract. The National Flood Risk Assessment (NaFRA) for England and Wales was initially undertaken in 2002 with frequent updates since. NaFRA has become a key source of information on flood risk, informing policy and investment decisions as well as communicating risk to the public and insurers. To make well informed decisions based on these data, users rightfully demand to know the confidence they can place in them. The probability of inundation and associated damage however cannot be validated in the traditional sense, due the rare and random nature of damaging floods and the lack of a long (and widespread) stationary observational record (reflecting not only changes in climate but also the significant changes in land use and flood defence infrastructure that are likely to have occurred). To explore the validity of NaFRA this paper therefore provides a bottom-up qualitative exploration of the potential errors within the supporting methods and data. The paper concludes by underlining the need for further research to understand how to robustly validate probabilistic risk models.

1 Introduction

Over recent years risk analysis (reflecting a combination of probability and consequence) has become common place in support of flood management decisions in England¹. Building upon the National Assets at Risk² studies in the late 1990s the National Flood Risk Assessment (NaFRA) was first undertaken by the Environment Agency in 2002 to provide a quantified assessment of national economic risk of flooding from rivers and the sea, and its regional distribution³. Frequent updates of NaFRA (based on progressive improvements to the methods and data) have taken place since^{4,5}. During this time the use of the results has also changed. The NaFRA products are increasingly used to support activities across flood risk management, including helping the Environment Agency report against national outcome measures, explore national scale long term investment scenarios⁶, and communicate flood risk to the public (providing the data behind the Environment Agency's "What's In Your Back Yard" (WIYBY) service, including a "Risk of flooding from rivers and sea" map since December 2013 and as open data since December 2014).

The continued increase in use of the NaFRA products, has been accompanied with greater pressure to understand if the results are 'fit for purpose' – recognising one 'golden rule' of good flood risk management is to communicate both risk and uncertainty⁷. Penning-Rowsell, 2014 for example suggests the national estimate of risk derived through NaFRA significantly overstates the real risk⁸.

This paper further describes the difficulties associated with the validation of the NaFRA results, the sources of potential error and presents a new validation framework to guide an on-going process of validation.

2 The validation challenge

The inability to satisfactorily 'validate' the national assessment of flood risk reflects two issues: (i) **validation is decision-dependent**. This is because a decision maker's expectation for tolerable uncertainty, at a particular scale or in a particular flood characteristic (depth, velocity, probability etc), is ultimately related to value judgements made about their relative importance^{9,10} and (ii) **flood risk is not directly or fully observable**. Given that damaging flood events are rare and flood systems (the climate, the defences, the properties etc) are not stationary, it is difficult, if not impossible, to determine the accuracy of an estimate through comparison to objective measured data alone.

The 'validation' of a risk assessment must necessarily be considered as a process of structured reasoning about the level of confidence needed to support a particular decision and the credibility of the assessment of risk in that context. For example, a user of the national aggregation of risk, which was the intended use of the assessment may be able to rely on local errors cancelling out (subject to an understanding about whether or not such errors exhibit any particular biases); an individual homeowner, insurer or local planning officer faced with an obviously implausible map of modelled flood risk may not consider the information reliable enough to be useful. Because of this variation in use, any attempt to provide a universal definition of what is "good enough" would

^a Corresponding author: paul.sayers@sayersandpartners.co.uk

flounder. In addition to the difficulty associated with a varying notion of validity, any process of validating probabilistic flood risk estimates also has to overcome a number of underlying challenges:

- Lack of ‘risk’ observations:** Flood risk (a function of both probability and consequence) can only ever be estimated and not fully measured directly. This reflects natural variability in climate (and the need to include rare physical circumstances that may never have been observed) but also stochastic behaviour in other factors that determine flood risk (for example, the chance of breach or blockage) and changing nature, of the flood system itself (for example changes in defences and receptor exposure and vulnerabilities). Consequences can be observed, but not a combination of consequences and probability.
- Observed behaviour during frequent storm events may not be replicated in extreme cases:** System risk models, and the individual model components they comprise, can be compared (or even calibrated) against more frequent, observed events. The process of calibration cannot be relied upon to ensure reasonable performance during very extreme events where no observations exist and the behaviour of the system may be fundamentally changed (for example, a single, or even multiple, breach(es) may occur, or storms may be more widespread or persist for longer than historically observed and lead to a non-linear, modelled, escalation in impacts – as first recognised in the wake of the Asian Tsunami, 2004).
- Heterogeneity of real flood systems:** Perhaps more so than in more others countries the physical characteristics of the individual flood systems across England vary significantly. Equally, the quality of the data available within each so-called Flood Risk Management System (FRMS) varies between systems and within each system. It is therefore difficult to assemble benchmark tests (against which the national analysis can be compared) that cover all physical settings (catchment types, defence systems, receptors etc.) and data quality combinations that may be encountered. This, of course, also means that a national-level methodology for risk assessment may not fit local circumstances.
- Variation in the weakest links:** An assessment of the probability of flooding and the associated impacts requires all important aspects of the flooding system (the sources, pathways and receptors¹) to be represented. The important processes will however vary from one location to another. For example, using a model that focuses solely on the hydrology and channel hydraulics whilst ignoring the

representation of defence performance may provide credible results in some areas, whereas the results in others may contain profound errors (in this case, for example, in floodplains defended by raised defences). The credibility of the weakest element of the system model is also important. Consider for example a defended floodplain. A system model of a defended floodplain that combines a fully hydrodynamic model (established using a highly accurate representation of floodplain topography) but linked to a simplified representation of breach location and growth is likely to be as ‘wrong’ as a simplified inundation model linked to a detailed (more accurate) assessment of breach location and growth.

3 Overview of the NaFRA analysis

The system risk analysis that supports NaFRA is based upon a model called ‘RASP’. The RASP combines a probabilistic description of the source (fluvial and coastal storm loads) and pathway (defence performance) with a simplified inundation analysis and an assessment of direct impacts – Figure 1. More detail on the RASP methods can be found in^{4,5,11}.

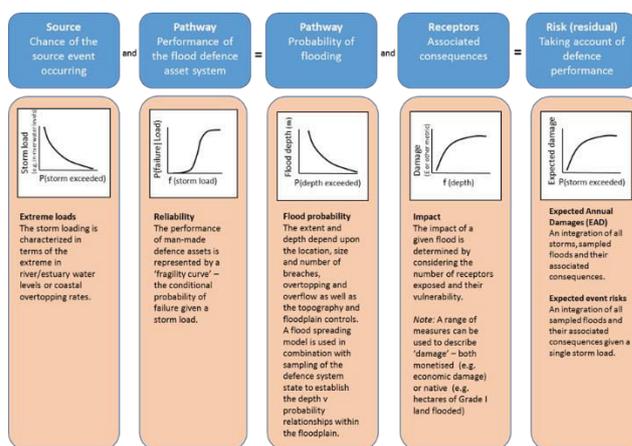


Figure 1 Overview of the main components of the RASP system risk analysis method that underlies the National Flood Risk Assessment (adapted¹¹)

4 Understanding the nature of errors

As with all models, RASP is subject to error. Inaccuracies in data (introduced, for example, through inaccuracies in measured defence crest levels or the land use in at-risk areas) and the performance of individual model components (introduced, for example, through simplifications of the breach processes) lend themselves to direct inclusion within a local or global sensitivity analysis to determine their influence on the assessment of

risk¹³. These errors are referred to here as 'quantifiable knowledge uncertainty'. Model structure errors (introduced, for example, through the assumption that the floodplain is independent of the river channel, or the exclusion of blockage as a failure mode for bridges) do not lend themselves to traditional sensitivity analysis and are more difficult to assess. Such errors are referred to here as 'gross potential errors'. The distinction between these two sources of error is important. Consider, for example, a detailed analysis that represents all of the quantifiable uncertainty in data and model parameters. The resulting distribution of damage for a particular urban floodplain appears complete and may be interpreted as representing the 'whole truth'. In reality the basis of the model is 'wrong' – because there is a high chance that during an extreme summer flood a low arched bridge downstream of the urban area (a structure that is not represented in the model) will be blocked by a caravan from the upstream park (a debris recruitment process not represented in the model). As a result, the system model provides only a partial truth and the 'true' risk lies elsewhere. This distinction is shown schematically in Figure 2.

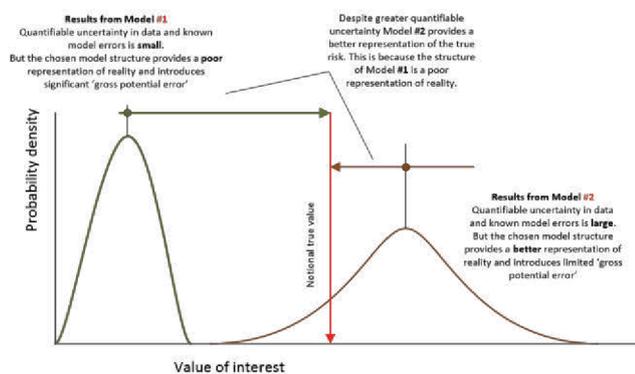


Figure 2 Gross Error Potential and Quantifiable Knowledge Uncertainty both need to be understood to gain confidence in any assessment of risk

A number of studies have therefore been undertaken to explore the potential sources of error within NaFRA^{12,13,14,15,16}. These studies attempt to quantify uncertainties that may have a significant influence on the credibility of the NaFRA results including model structure errors (such as the lack of interaction between the floodplain and river channel), model component uncertainties (reflecting the simplification of the floodplain process used) and data uncertainties (particularly around basic data such as crest level and the location and type of property that maybe exposed to flood waters).

5 Validation approach

Although a definitive statement on the importance of different sources of error and uncertainties within NaFRA in the context of the multiple decisions it now supports is

not possible, a number of studies have been undertaken that provide useful insights. The validation process promoted here synthesizes these previous studies and presents a 'top-down' process (focused on the credibility of the aggregated risks at a national scale) and a 'bottom-up' process (focused on the local credibility of individual datasets, model components and model structures used within the analysis) - Figure 1. The top-down approach compares (i) the nationally aggregated risk as estimated through NaFRA with an actuarial analysis of historical events, (ii) selected results from the NaFRA analysis with local, more detailed, applications of the same underlying methods, and (iii) explores the potential for comparison with alternative system models. The bottom-up approach explores the performance of the individual deterministic component parts of the national risk model, the probabilistic simulation framework and the supporting data and model parameters. This is done to provide disaggregated information on the ability of each component to represent specific aspects of the flooding system and hence where potential opportunities for improvements may lie.

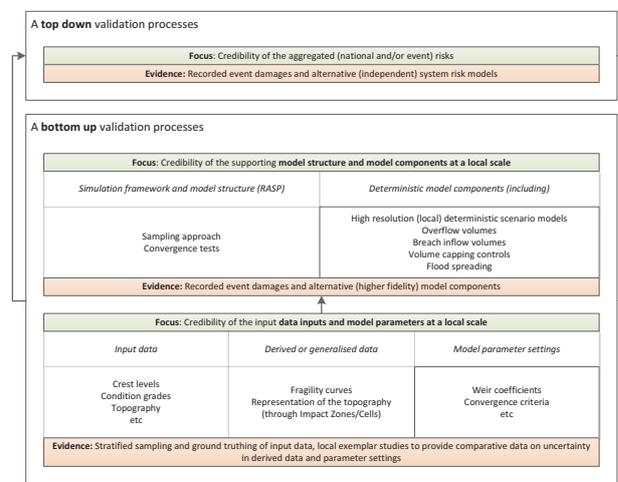


Figure 3 Developing a structured understanding of the confidence within flood risk system models requires a combination of top down and bottom up perspectives

6 Review of potential errors

The following Tables provide a synthesis of the findings from these studies. Table 2 focuses on model structure errors. Tables 3 and 4 focus on the individual model components used and the supporting data.

Table 2 Model structure errors that may have a significant influence on the credibility of the NaFRA results (adapted^{12, 15, 16})

Assumption	Description
Full dependence of single storm loading across a flood area.	Hydraulic loading conditions are assumed to be fully dependent (equal to return period) along the boundaries of any given floodplain. In situations where the boundary loads respond to the same storm events this is likely to be a reasonable assumption. In coastal areas (where local bathymetry and orientation vary markedly over short lengths), estuaries and confluences, where joint probability issues arise this is likely to be a much more significant assumption which may introduce bias ^{17,18} . No consideration is given to the condition of temporally sequenced or clustered events (and loading conditions that are known to be important).
Independence of storm loading between flood areas	Each Flood Area is assumed independent of all others and hence a change in one Flood Area (a breach or defence raising for example) has no influence on loads elsewhere. This is clearly untrue in some areas where upstream and / or downstream behaviour can have a marked impact on flooding elsewhere. This assumption also makes understanding the event risk regionally or nationally difficult. Some attempts have been made to overcome these issues through the inclusion of a basic feedback mechanism ¹⁹ and the potential for a more rigorous consideration of the spatial dependence issues ²⁰⁻²¹ .

Independence of strength between defence sections	The performance of one defence asset is assumed to be independent from any other (i.e. the performance – breach or not - of one defence does not influence the strength of its neighbour). This presents challenges in defining an independent defence length within a single long(er) defence and any interactions in behaviour that may exist (for example at transitions between structures, or when a neighbouring defence is breached and breach flows may compromise the rear face/toe of an adjacent defence) ²⁷ .
Single defence line and performance of point assets	Only a single (primary line) of linear raised assets (and off-line flood gates etc.) are treated probabilistically. Secondary line of raised defences and in-line barriers, pumps and sluices are assumed to perform to rule (i.e. a secondary defence is assumed 'unbreachable' and a barrier is assumed to operate as expected). In some settings both of these assumptions can be significant. Separate model runs are presently required to reflect the possible system states (operating / not operating) of barriers and pumps (with loading conditions that represent the different system states).
Interaction of floodplain flows with river/sea	Within the model the exchange of water from the river/sea to the floodplain is one way (channel to floodplain), whereas in reality there is often a significant interaction.
Breach initiation and probability of failure considered separately	The process of breach initiation is considered from the assessment of failure probability. In the fluvial case for example this means that when a defence is sampled as 'failed' the associated breach is assumed to occur when the water levels first rises above the breach invert level. This simplification and could influence inflows to the floodplain significantly.
Spatial location of breaching	Flood extent and depth can be sensitive to breach location. Although a common assumption even within detailed analysis, the basic assumption made that breaches occur at the centre of the defence section is a limitation.
Sampling framework	The simulation framework controls the way in which the model is structured, how the individual deterministic runs are selected (the sampling approach) and how convergence is defined (through automated convergence controls that determine when sufficient runs have been undertaken). Alternatives to the approach used within RASP do exist but understanding the benefits such alternatives may provide has been difficult to gauge.
Representation of receptor impacts on a 50m grid	A pre-process creates a composite impact curve relating flood depth to damage for each Impact Cell (typically a 50m grid). This assumes the flood level is unchanging across the Impact Cell (not depth as that is based on the finest resolution of the underlying DTM)
Representation of receptor damage based solely on depth	Within the RASP engine the flood depth v damage curves as set out in the Multi-Coloured Manual ²⁷ are used to translate flood depth to damage taking account of residential or non-residential property. In recent years a post process has been applied that uses the estimated annual probability of flooding to directly determine an annual average damage. This is done using the Weighted Annual Average Damage (WAAD) methodology. The uncertainty introduced through the WAAD approach is unknown but likely to be significant.
Each receptor is treated as an independent entity	It is widely recognised that risks cascade and escalate as they do so. This is particularly the case through infrastructure network but may also be true within communities. The RASP analysis undertaken to date ignores these impacts.

Table 3 Model component uncertainties that may have a significant influence on the credibility of the NaFRA results

Limitation/Assumption	Description
Inflow volumes (onto the floodplain)	They are a function both of the storm load (hydrology, sea level and wave overtopping calculations) and the assumed performance of the defence (breach or not). Sources of error include both the external load calculations, potential errors or omissions in the breach initiation, evolution, location and timing as well as the assumption of unconstrained (undrowned) inflows (because of the lack of interaction between the sea/in-river and the floodplain – discussed in Table 2).
Breach growth	The final breach width and invert is estimated using approximate rules. For some defence types (fluvial earth embankments) the science is now maturing to allow the potential for better estimates to be made ²⁴ .
Flood spreading	The RFSM is used to spread inflow floodwaters across the floodplain ²⁵ . The RFSM compares poorly to fully hydrodynamic models in complex settings. Recent studies have recommended updating the RFSM with a time-stepping approach (recently developed as a standalone model, although without a quantified cost-benefit justification to do so in comparison with other more physically-based flood models).

Table 4 Data uncertainties that may have a significant influence on the credibility of the NaFRA results

Limitation/Assumption	Description
Storm loading conditions	Extreme values of peak water levels and mean overtopping are used to represent storm loads. There is no ability to vary the loading conditions through an event to take direct account of the local hydrograph shape (beyond fairly crude embedded assumptions).
Defence fragility	Generic fragility curves have been developed for a series of 'defence types'. These generic curves are often used rather than site specific ones. The importance of this is unknown but is generally perceived that the fragility curves in use prior to 2014 overstated the fragility of certain defence types; an update was made in 2014 ²⁷ addressed perceived weaknesses in the fragility curves and produced a new set that is currently in use although this remains an issue under review.
Defence location and geometry	Inflow location is assumed to occur only at an asset mid-point; the only geotechnical aspect within the breach function for fluvial and tidal assets is whether the asset is hard or soft. Coastal assets use a breach length that is fixed and defined by structure type. Equally condition grades, crest levels, ground levels etc are represented through a single value. An on-going study to create a Continuous Defence Line (CDL) may help gather the underlying information to enable future method changes.
Floodplain topography and pathways	Floodplain topography within the NaFRA products is typically coarsely represented (regular 50 metre grid resolution). Finer resolutions and the inclusion of the connections (through embankments etc) are possible within the RFSM, and often applied with regional application of the Modelling Decision Support Framework 2 (MDSF2) ¹⁹ .
Representation of receptors	The National Receptor Dataset (NRD) is used within NaFRA to identify the location and type of properties that may be damaged in a flood event. Although significantly improved over recent years, considerable questions remain in terms of both over-representing some property types and under representing others. This is discussed in more detail in the following section.

7 Performance of individual model components

In river water levels: NaFRA uses extreme water levels provided from detailed local models (run assuming defences and barriers, where they exist, perform perfectly). In the absence of local model results, estimates are taken from the national flood zone mapping, which

evolved over a number of years from a simple “indicative” assessment of areas that may be at risk of flooding (without accounting for flood defences), to a patchwork of more or less detailed modelling of areas at risk of flooding²⁸. HR Wallingford (2006) includes some assessment of uncertainty in the fluvial load calculations and suggests that a “good” estimate of the 1:1000 year water level in the Thames Estuary (reflecting the assumed quality of the local data sources) has an associated error of $\pm 0.35\text{m}$ ²⁹. Errors in more complex settings (such as steep catchments) are likely to be considerably greater.

The data that supports NaFRA is however derived from UK industry-standard methods published in the Flood Estimation Handbook (FEH)³⁰ and associated data sets. These combine extreme value analysis of historical measurements of river flows and rainfall with empirical models based on catchment typology and other geographical data to produce estimates of the marginal distribution of peak river flows at any location on a river. The FEH is designed as a guide for experts to select an appropriate analytical method based on detailed assessments of data and hydrological knowledge of a study area. There is the potential for significant errors through mis-specification of the hydrological analysis. Kjeldsen (2014) studied uncertainty about a 1/100 AEP design flow estimate using the FEH ‘index flood’ methods and found that the 95% confidence interval at an ungauged site using donor transfer will be of the order of minus 55% to plus 125%³¹. The implications of this uncertainty in hydrological analysis on the load levels at a given location within the national analysis is unclear, but a pragmatic approach is taken whereby local Environment Agency staff review the loading conditions and are able to make modifications if considered appropriate.

Fragility: The use of fragility curves for representing the performance of flood defences has its origins within the US³². This work has been extended within the UK and Europe over the past decade, see for example^{1,4,33,34,35}. The system models use the fragility curves to define the likelihood of different defence system states. Multiple defence system states (and hence flood extents) are simulated for any given hydraulic loading event (e.g. extreme river flow or coastal overtopping). The models consider a wide range of hydraulic loading events and combine information on the likelihood of the hydraulic loads with information on the likelihood of defence failure to evaluate the overall flood risk. The NaFRA 2002 considered the impact of uncertainty in the fragility estimates to be propagated through to the assessed risk, suggesting an impact of between 0.5 to 2 times the central estimate of EAD³. Consideration of uncertainty in the defence performance (given load) is no longer represented in the risk assessment.

Breaching: If a defence is sampled to be in a breached state, the dimensions of the breach is estimated using a

simplified approach based on the return period of the flood event. To date no detailed comparison has explored the credibility of these equations in the context of national risk estimates. Although they are widely considered to be a significant simplification and not appropriate for local studies or studies that require the inflow hydrograph to be estimated. The representation of breach processes can clearly be improved within RASP, for example van Damme 2012 introduced a rapid breach model that is expressly designed for use within flood risk models³⁶; but the associated impact on the national risk estimates remains difficult to determine.

Inundation simulation: The RFSM (introduced earlier) is exceptionally computationally efficient at the expense of a full representation of the physical processes. Environment Agency (2010) compares the results from a wide range of models for a series of tests and (unsurprising) concludes that those that solve the full SWEs with shock-capturing schemes are most accurate for use, especially where breach simulation is required³⁷. No evidence was presented on the influence this improved representation would make if implemented at a national scale or the associated cost of doing so. Arguments of physical reasoning however provide a compelling case for replacing the RFSM with a dynamically linked floodplain and channel model (particularly in undefended areas). This dynamic link is important for a number of reasons. Most notably the loss of water from the river channel to the floodplain from upstream breaches is reflected with a lowering of the downstream river-channel water level (the river-channel/floodplain feedback mechanism). This would also enable breach models to provide the 1D/2D link. Despite the undoubted improvement this would provide at a local scale there is legitimate debate, given limited resources, around where improvement priorities lie and the nature of the national scale analysis tools.

Estimating damage given flooding: The estimate of damage relies upon the National Receptor Dataset (NRD) to determine the location and type of property (residential or non-residential broken down by a series of sub-classes) and flood damage relationships from the Multi-coloured Manual (MCM) to convert an estimate of the flood hazard within an Impact Cell to pounds damage. Both of these have the potential for introducing significant error. For example, Figure 5 highlights the potential difficulty in using receptor data for assessing risk at national scale when it is not practical to ground truth each point and continued efforts to implement ‘intelligent automated rules’ to avoid incorrect or misleading classifications will be important.

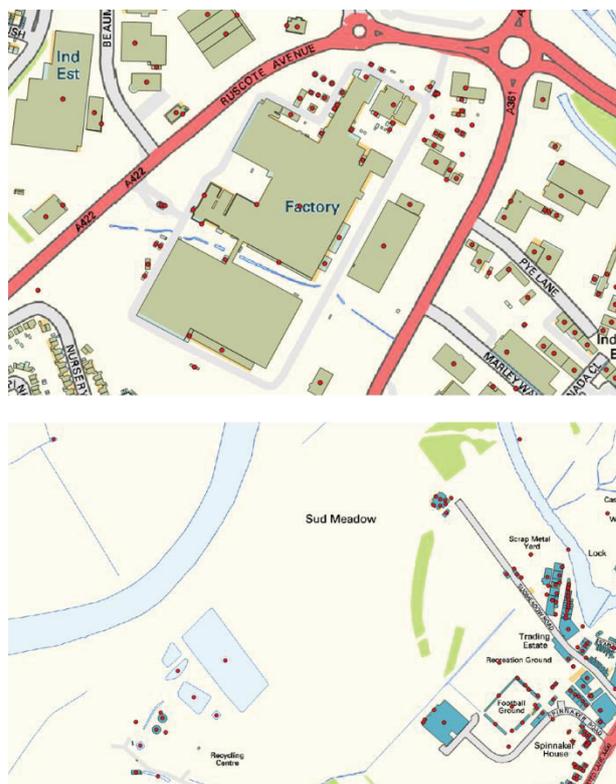


Figure 5 Non-residential NRD points and correspondence to buildings. In Banbury (top), a single factory site is associated with multiple NRD points. In Gloucester (bottom), many NRD points on the floodplain are not associated with any property of interest in the context of NaFRA³⁸

8 Emerging framework to support the validation of probabilistic risk estimates

In recent years a number of advances have been made in how to approach the comparison of probabilistic outputs³⁹. The steps below extend past studies to more specifically focus on probabilistic risk models. The emerging guiding principles of this process are outlined below:

Step 1 – Be clear on how the estimates are intended to be used

1.1 Ensure the intended use is well described and understood - Ensuring that all decision relevant questions to be answered are clearly set out together with the intended use and known limitations – is a crucial first task. The validation process should then be designed to respond (as a minimum) directly to these questions.

1.2 Define the nature and scale of the outputs required – Ensuring effort is devoted to validating the most important outputs in the context of the decisions being supported. This should include both the primary uses (e.g. NaFRA is primarily intended to support national scale decisions) and secondary uses (e.g. the access to NaFRA via open-data products now means outputs are often used at much more local scale than initially

intended). Being clear on specific scales and outputs that will form part of the validation process focused the validation effort and ensures the results provide meaningful insights.

Step 2 – Set practical expectations of the level of confidence required

2.1 Define validation objectives– Defining what is ‘good enough’ is extremely difficult as the degree of uncertainty that is acceptable will vary according to the specifics of the decision being made. Despite this, clear validation objectives are needed to structure the validation process. But, rather than expressing such objectives in quantified limits (as would traditionally be used in validating a river flow model for example) the validation of probabilistic risk models requires objectives couched in terms of ‘fitness for purpose’. How best to do this, in both qualitative and/or quantitative terms, presents significant challenges but nonetheless is a prerequisite to any serious attempt at validation.

Step 3 – Understand the linkage between real world and model processes

3.1 Map out important ‘real world’ processes - A critical view of the physical processes and relationships that contribute to flood events helps ensure important processes are not ignored or misrepresented because of the way the model works.

3.2 Map out the ‘model’ processes - Understanding how the model represents the individual physical processes and how the model structure brings these together to capture the real world, whole system, behaviour. In doing so, distinguish between the various elements of model components, model structure and data. Set out the interactions between data and model components to understand the dependencies within the analysis approach.

Step 4 – Undertake and record calibration and comparison activities

4.1 Calibration – Record all attempts to calibrate the individual component models (and the corrections made to model parameters) and model response. Can these factors be considered to hold true for more extreme (un-calibrated) events?

4.2 Comparisons – Record all attempts to compare the model outputs for the events of interest. This may include application of the approach to benchmark datasets and inter-comparison of alternative modelling approaches (including model component comparison and comparison of results from alternative system models). In doing so, use well defined comparable observations

where these exist and recognise the difficulties, and inevitable errors, in such observations.

Step5 - Assess uncertainty and identify key sensitivities

Whilst a comprehensive, quantified uncertainty analysis of models used for risk assessments such as NaFRA is a formidable challenge⁴⁰, a pragmatic framework for uncertainty assessment is proposed in Figure 6. This comprises five sub-tasks as outlined below.

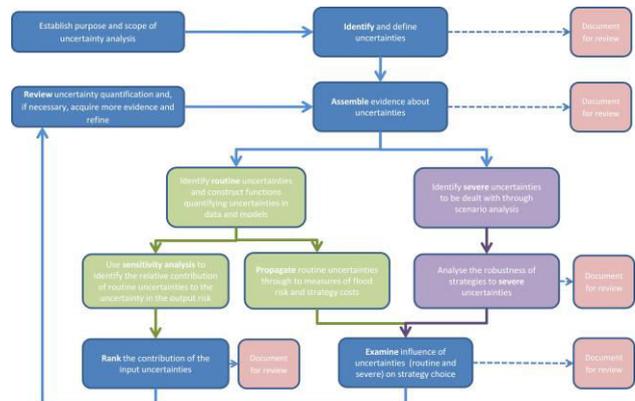


Figure 6 Pragmatic framework for uncertainty assessment ^{41,42}.

5.1 Assess uncertainty – Record any attempt to formally assess uncertainty (using scientifically defensible approaches). But also accept structured expert argument (where judgements are transparent and available for challenge) as a valid contribution where clearly supported by evidence (for example, in the absence of sufficient measurements or alternative models that can be considered to approximate the "truth", expert interpretation is required – for example applying engineering judgement to alternative breach models or fragility inputs).

5.2 Determine sensitivities - Record any attempt to determine key sensitivities and provide real world arguments to explain the model behaviour.

5.3 Provide confidence statements – Explicitly include confidence statements (either qualitatively or quantitative) and the evidence upon which they are based.

5.4 Be clear on deficiencies – Acknowledge gaps in the data and approach.

5.5 Make recommendations – Use the evidence to enable future improvements to be targeted. Focus on the performance of the NaFRA outputs within important physical settings, enabling quick win improvement

opportunities to be identified (for example suggesting improvements in both undefended and defended areas).

9 Conclusions

This paper highlights the ‘validation’ of probabilistic large scale estimates of risk as a process of gaining confidence and understanding through multiple strands of inquiry. The important data, model and model structure uncertainties that currently influence NaFRA are presented together with an emerging framework of validation.

The validation framework highlights that assessing the uncertainty in a modelled estimate of risk requires more than simply an assessment of data and model parameter uncertainties; understanding the credibility of the model structure is equally, if not more, important. Doing so however is problematic and will continue to be a focus of future efforts, particularly through forensic analysis of observations and modelled outputs and by comparing results from comparable alternative system models. This is a process that remains an on-going challenge.

10 References

1. Sayers PB; Hall JW; Meadowcroft IC (2002). Towards risk-based flood hazard management in the UK. *Civil Engineering* 2002, 150(5), 36-42.
2. MAFF (1998). National Assets at Risk (NAAR). Authors Deakin R, Burgess, Samuels, P, Sayers P, Chatterton J. A report prepared by Halcrow for MAFF (David Richardson)
3. Environment Agency (2002). National Flood Risk Assessment (NaFRA). Authors Sayers PB, Panzeri M, Meadowcroft, I. A report prepared by HR Wallingford for the Environment Agency.
4. Hall J; Dawson R; Sayers P; Rosu C; Chatterton J; Deakin R (2003). A methodology for national-scale flood risk assessment. *Proceedings of the Institution of Civil Engineers - Water & Maritime Engineering* 2003, 156(3), 235-247. Robert Carr Prize
5. Gouldby BP, Sayers PB, Mulet-Marti J, Hassan M, Benwell D (2008) A Methodology for Regional Scale Flood Risk Assessment *Proceedings of the Institution of Civil Engineers: Water Engineering*, June 2008
6. Environment Agency (2014). Long term investment scenarios (for flood and coastal risk management). Author Meadowcroft. Published by the Environment Agency, London.
7. Sayers P B, Galloway Gerry, Penning-Rowse Edmund, Shen F, Wen K, Chen Y, Le Quesne T (2014). Strategic flood management: ten ‘golden rules’ to guide a sound approach. *Journal: International Journal of River Basin Management*
8. Penning-Rowse, E. (2014). A realistic assessment of fluvial and coastal flood risk in England and Wales. *Transactions of the Institute of British Geographers*, 2015, 40(1), 44-61
9. Hall, J.W. (2014) Editorial: steps towards global flood risk modelling. *Journal of Flood Risk Management*, 7(3): 193-194.
10. Sayers, P., Galloway, G., & Hall, J. (2012). *Robust decision making under uncertainty - Towards adaptive and resilient flood risk management infrastructure*. London: Thomas Telford
11. Sayers PB and Meadowcroft IC (2005). RASP - A hierarchy of risk-based methods and their application. *Proceedings of the 40th Defra Conf. of River and Coastal Management*.
12. Environment Agency (2013). NaFRA Method Improvements Task 3.6 – Future Benchmarking. Authors Gouldby, B., Lhomme, J., McGahey, C. and Panzeri, M.
13. Gouldby B, Sayers PB, Panzeri MC, Lanyon J (2010) Development of application of efficient methods for the forward propagation of epistemic uncertainty and sensitivity analysis within complex broad-scale flood risk system models. *Canadian Journal of Civil Engineering*; Mar2010, Vol. 37 Issue 3, p496. Special Issue on Hydrotechnical Engineering.
14. McGahey, C Sayers, P.B Dunning P and Neve P (2006). Exploring the sensitivity of RASP HLM+ to variations in input data and model parameters. A report by HR Wallingford for the Environment Agency.
15. Environment Agency (2011). NaFRA Confidence Method Development – Final. Authors Sayers P.B, Lamb, R Meadowcroft, I. A report by JBA/Sayers and Partners published by Environment Agency
16. Environment Agency (2011) NaFRA Confidence Route Mapping – Final Authors Sayers P.B, Lamb, R Meadowcroft, I. A report by JBA/Sayers and Partners published by Environment Agency
17. Lewis, M., Horsburgh, K., Bates, P. and Smith, R. (2011) Quantifying the uncertainty in future coastal flood risk estimates for the UK. – *Journal of Coastal Research*: 870 – 881
18. Wyncoll, D and Gouldby, B. (2012) Application of a multivariate extreme value approach to system flood risk analysis – *Journal of Flood Risk Management* 8 (2), 145–160
19. Mc Gahey, C, Sayers, P.B, Wills, M. Lanyon, J. Weisgerber, A. Surendran, S., Baylis, A., Linford, T. (2010). Modelling and Decision Support Framework 2 (MDSF2) - Initial Applications and Lessons Proc. Of the FCM10 Conference. Environment Agency
20. Wyncoll, D. and B. Gouldby (2012). Application of a multivariate extreme value approach to system flood risk analysis. *Journal of flood risk management*
21. Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., Batstone, C. (2010). A new method to assess the risk of local and widespread flooding on rivers and coasts. In: *Journal of Flood Risk Management*. 3, 4, p. 323-336. 14 p.
22. Vrouwenvelder, A.C.W.M. (2006). Spatial effects in reliability analysis of flood protection systems. - *IFED Forum 2006*, Lake Louise, Canada.
23. Penning-Rowse, E., Johnson, C., Tunstall, S., Tapsell, S., Morris, J. and Chatterton, J., Green C. (2005). *The Benefits of Flood and Coastal Risk Management: A manual of Assessment Techniques (The Multi-coloured Manual)*. Flood Hazard Research Centre, Middlesex University.
24. Hassan, M., M. Morris, G. Hanson, and K. Lakhal. 2004. Breach formation: Laboratory and numerical modeling of breach formation. In *Proc. Dam Safety 2004*, Association of State Dam Safety Officials (ASDSO), CD-ROM. Lexington, Ky.: Association of State Dam Safety Officials.

25. Sayers, P.B and Marti, J (2006) RFSM - Rapid Flood Spreading Method - Initial Development. A report for the Environment Agency Thames Estuary DT4.
26. Lhomme, J.and Sayers, PB. and Gouldby, B.P. and Samuels, P.G. and Wills, M. and Mulet-Marti, J. Recent development and application of a rapid flood spreading method. In: FLOODrisk 2008, 30 September - 2 October 2008, Keble College, Oxford, UK. (2008)
27. Environment Agency Report. Reliability of flood defences: Improved generic fragility curves for use in flood risk analysis. Authors Topple, A. & Simm J.
28. Porter, J. and Demeritt, D. (2012) Flood Risk Management, Mapping and Planning: The Institutional Politics of Decision-Support in England, Environment & Planning A. DOI: 10.1068/a44660
29. HR Wallingford (2006). Defence Overflowing and Breach Inflow Volumes: Explicitly accounting for Uncertainty and Enabling Defence Specific Estimates. Development of the IA 8/10 System Model. HR Wallingford report for the Environment Agency Thames Estuary 2100 studies - IA8/10 Technical Note 12
30. Robson, A.J. and Reed, D.W., 1999. Flood estimation handbook. Institute of Hydrology, Wallingford. Various updates since
31. Kjeldsen, T.R. (2014), How reliable are design flood estimates in the UK?. Journal of Flood Risk Management. doi: 10.1111/jfr3.12090
32. USACE (1996). Risk-based Analysis for Flood Damage Reduction studies. Engineer Manual EM 1110-2-1619, U.S. Army Corps of Engineers, Washington, D.C.
33. Environment Agency (2007). Report FD2318_5926_TRP. Authors Sayers P, Simm J, Buijs F, Flikweert, J A report by HR Wallingford for the Environment Agency.
34. Simm, J., Gouldby, Sayers, P.B, Flikweert, J, Wersching S and Bramley M (2009). Representing fragility of flood and coastal defences: Getting into the detail. Flood Risk Management: Research and Practice. P. Samuels, S. Huntington, W. Allsop and J. Harrop, Taylor and Francis Group, London, UK.
35. Vorogushyn, S., B. Merz and H. Apel (2009). "Development of dike fragility curves for piping and micro-instability breach mechanisms." Nat. Hazards Earth Syst. Sci. 9(4): 1383-1401
36. Van Damme, M., Morris, M.W., Borthwick, A.G.L., and Hassan, M.A.A.M. ,2012, A rapid method for predicting embankment breach hydrographs, Floodrisk2012, Rotterdam, The Netherlands (nominated for the paper award)
37. Environment Agency (2010). Benchmarking of 2D Hydraulic Modelling Packages, Environment Agency Report SC080035/SR2, Bristol. A report by Pender et al, Heriot Watt University.
38. Sayers, P.B., Horritt, M. S., Penning-Rowsell, E., and McKenzie, A. (2015). Climate Change Risk Assessment 2017: Projections of future flood risk in the UK: Appendix G Validity of present day and future risks. Pages 125. Sayers and Partners LLP report for the Committee on Climate Change.
39. Environment Agency (2010). Validation of probabilistic flood models. Environment Agency Report SC090008.
40. Beven, K., and Lamb, R. (2014). The uncertainty cascade in model fusion. In Riddick, A. T. Kessler, H. and Giles, J. R. A. (eds) Integrated Environmental Modelling to Solve Real World Problems: Methods, Vision and Challenges. Geological Society, London, Special Publications, 408, <http://dx.doi.org/10.1144/SP408.3>
41. Hall, J.W. and Solomatine, D. (2008) A framework for uncertainty analysis in flood risk management decisions. International Journal of River Basin Management, 6(2): 85-98.
42. Sayers P B, Galloway Gerry, Penning-Rowsell Edmund, Shen F, Wen K, Chen Y, Le Quesne T (2013). Flood Risk Management: A strategic approach. Published in English by UNESCO, Paris in May 2013 (ISBN 978-92-3-001159-8) and in Chinese by Water Publishing in October 2012 (ISBN:978-7-5170-0201-7). Published in association with WWF, the General Institute of Water Design and Planning, Beijing (GIWP) China and Asian Development Bank (ADB).

11 Acknowledgements

The ForUM (Grant NE/M008851/1) provided the support funding for this work and is gratefully acknowledged together with the support and useful insights are all co-authors and their organisations.