



Epistemic uncertainties and natural hazard risk assessment – Part 2: What should constitute good practice?

Keith J. Beven^{1,2}, Willy P. Aspinall³, Paul D. Bates⁴, Edoardo Borgomeo⁵, Katsuichiro Goda⁷, Jim W. Hall⁵, Trevor Page¹, Jeremy C. Phillips³, Michael Simpson⁵, Paul J. Smith^{1,6}, Thorsten Wagener^{7,8}, and Matt Watson³

¹Lancaster Environment Centre, Lancaster University, Lancaster, UK

²Department of Earth Sciences, Uppsala University, Uppsala, Sweden

³School of Earth Sciences, Bristol University, Bristol, UK

⁴School of Geographical Sciences, Bristol University, Bristol, UK

⁵Environmental Change Institute, Oxford University, Oxford, UK

⁶European Centre for Medium-Range Weather Forecasting, Reading, UK

⁷Department of Civil Engineering, Bristol University, Bristol, UK

⁸Cabot Institute, University of Bristol, Bristol, UK

Correspondence: Keith J. Beven (k.beven@lancaster.ac.uk)

Received: 5 July 2017 – Discussion started: 21 August 2017

Revised: 22 January 2018 – Accepted: 8 March 2018 – Published: 24 October 2018

Abstract. Part 1 of this paper has discussed the uncertainties arising from gaps in knowledge or limited understanding of the processes involved in different natural hazard areas. Such deficits may include uncertainties about frequencies, process representations, parameters, present and future boundary conditions, consequences and impacts, and the meaning of observations in evaluating simulation models. These are the epistemic uncertainties that can be difficult to constrain, especially in terms of event or scenario probabilities, even as elicited probabilities rationalized on the basis of expert judgements. This paper reviews the issues raised by trying to quantify the effects of epistemic uncertainties. Such scientific uncertainties might have significant influence on decisions made, say, for risk management, so it is important to examine the sensitivity of such decisions to different feasible sets of assumptions, to communicate the meaning of associated uncertainty estimates, and to provide an audit trail for the analysis. A conceptual framework for good practice in dealing with epistemic uncertainties is outlined and the implications of applying the principles to natural hazard assessments are discussed. Six stages are recognized, with recommendations at each stage as follows: (1) framing the analysis, preferably with input from potential users; (2) evaluating the available data for epistemic uncertainties, especially when they might lead to inconsistencies; (3) eliciting information on sources

of uncertainty from experts; (4) defining a workflow that will give reliable and accurate results; (5) assessing robustness to uncertainty, including the impact on any decisions that are dependent on the analysis; and (6) communicating the findings and meaning of the analysis to potential users, stakeholders, and decision makers. Visualizations are helpful in conveying the nature of the uncertainty outputs, while recognizing that the deeper epistemic uncertainties might not be readily amenable to visualizations.

1 Introduction

Part 2 of this paper constitutes a discussion of some of the issues raised by the review of different natural hazard areas in Part 1 (Beven et al., 2018), with a view to addressing the question of what should constitute good practice in dealing with knowledge related uncertainties in natural hazards assessment. For good epistemic reasons, there can be no definitive answer to the question, only a variety of views. Thus, what follows should be considered as an opinion piece, discussing some of the precepts held by the authors or opinions expressed by others elsewhere, any of which could be subject to revision in the future. However, we would argue that

an open discussion of the issues is valuable in itself at this stage.

There are, perhaps, a number of things on which we can agree:

- that epistemic uncertainties are those that are not well determined by historical observations, including where those observations are used in the formulation and conditioning of predictive models or simulators;
- that they arise in all stages of risk analysis which, in Part 1 of this paper and elsewhere (Rougier et al., 2013), have been characterized in terms of assessing Hazard, Footprint, and Loss;
- that they might not have simple statistical structure and may not be bounded, complete, or exhaustive, leaving open the potential for future surprise;
- that any analysis will be conditional on the assumptions about the hazard model, the footprint model, and the loss model components; and about the nature of the errors associated with such model components.

Clearly, at each analysis stage there is the potential for different sets of such assumptions or methodological choices by the analyst and, since these decisions are necessarily not well determined by historical data (and there may be an expectation of future change or surprise), there can be no single right answer. Thus, an evaluation of the potential impacts of feasible sets of assumptions and the communication of the meaning of the resulting uncertainties to users of the analysis should be important components of good practice. In what follows we discuss a conceptual framework for dealing with epistemic uncertainty in natural hazards risk assessments, and express some opinions on what might constitute good practice in framing an analysis and the communication of the results to users.

2 A framework for good practice in natural hazards modelling

We suggest that a general framework for natural hazards modelling should include the following steps:

1. Establishing the purpose and framing the nature of the risk analysis to be undertaken in terms of the hazard, footprint, and loss to be considered.
2. Evaluating the information content of the available data.
3. Eliciting opinions about sources of uncertainty and the potential for future surprise.
4. Choosing a methodology and defining workflows for applying the method correctly and accurately.

5. Assessing whether a decision is robust to the chosen assumptions through some form of sensitivity analysis.

6. Communicating the meaning of the uncertainty analysis through a condition tree and audit trail, and visualising the outcomes for effective decision-making.

The following discussion is structured following these six steps.

2.1 Framing the nature of the risk analysis

The first, essential step when framing a risk analysis in natural hazards is to establish a defined purpose for the analysis. This is usually resolved by discussion with the “problem owner” or stakeholder group who will make use of the findings, therefore their requirements should determine, in the main, what hazards, magnitudes, footprints and losses are to be considered. In most cases, this also involves bounding the problem, i.e. differentiating between what is to be included and what is to be considered beyond the analysis. Often, these latter decisions will be based on expert judgement, either informally or as a formal elicitation exercise. Historically, the framing of many risk analyses is already institutionalized as a set of rules or statutory requirements. In several countries there are defined workflows for different types of risk, with some similarities of approach but differences in detail. Framing the problem has become more difficult with a greater recognition that the risk might not be stationary into the future. Future change is one of the most important sources of epistemic uncertainty in many natural hazards, especially for floods, droughts, landslides, and wind storms.

When analysing natural hazards, we are always interested in more extreme events, and the analysis of the risk of such events for use in decision making is made much simpler by considering that different (uncertain) components of the risk can be defined in terms of probability distributions. This effectively allows for a situation wherein all epistemic uncertainties that are recognized can be treated analogously to aleatory uncertainties. This is usually the form of output of an expert elicitation exercise, for example (see Sect. 2.3 below). Such an approach allows for the full power of statistical theory to be applied, since the analysis is based on the manipulation of probabilities, but as a result, the risk assessment depends heavily on distributional assumptions that are made, notably in the tail behaviour of the extremes (of both magnitudes and potential losses). Since samples of extreme events are, by their very nature, generally small then there has to be some epistemic uncertainty about the appropriate distributions to be applied, even if some countries have institutionalized certain distributions as a convenient standard practice in this respect, at least for magnitudes (see Sect. 3 below). Footprints and losses are often treated more deterministically (as discussed, for example, in Rougier et al., 2013).

Natural hazard risk assessments have muddled along for many decades on this basis without undue criticism, except from communities that get impacted because little or no protection is provided, or because any protection employed is subsequently deemed to have been excessive. In such situations, the assessment, and the analyst, can always be defended by invoking the capricious nature of reality. This is because the magnitude of a particular event is usually evaluated in terms of its frequency of occurrence (as a “return period” or “annual exceedance probability”); then, if a new event comes along that is bigger than that estimated as the reference event for a risk analysis, it has, by definition, a lower probability of exceedance. It may be a surprise if two events of low probability occur in quick succession but, even then, there is always a small but finite statistical probability of that occurring under the assumptions of the analysis. Effectively, the analyst cannot be wrong. Post hoc, of course, the new event(s) can also be used to revise the risk analysis, including potentially changing the distribution or model used in the analysis. This has, perhaps, been one reason why there has been little pressure to treat epistemic uncertainties more explicitly when, again for good epistemic reasons, it is difficult to know exactly what assumptions to make. It is also a reason why aleatory uncertainties, commonly defined as irreducible uncertainties in the literature, might be reducible as more information becomes available.

2.2 Evaluating the information content of the available data

The assessment of the potential for future natural hazard events frequently involves the combination of model outputs with data from historical events. Often, only the simplest form of frequency analysis is used where the model is a chosen distribution function for the type of event being considered and where the data are taken directly from the historical record. Such data are often used uncritically (since they are generally the only information available to condition local uncertainty estimates) but it is important to recognize that both model and data will be subject to forms of epistemic uncertainty. In some cases this might lead to the situation that either the model or the data might be “disinformative” in assessing the future hazard. A frequency distribution that underestimates the heaviness of the upper tail of extreme events, for example, might lead to underperformance of any protection measures or underestimation of the zone at risk.

Similarly, any data used to condition the risk might not always be sufficiently certain to add real information to the assessment process. In the context of conditioning rainfall-runoff model parameters, for example, Beven et al. (2011) and Beven and Smith (2015) have demonstrated how some event data suggest that there is more estimated output from a catchment area in northern England than the inputs recorded in three rain gauges within that catchment. No hydrological

model that maintains mass balance (as is the case with most rainfall-runoff models) will be able to predict more output than input, so that including those events in conditioning the model of the catchment area would lead to incorrect inference (at least for catchments with no significant groundwater storage). Why the data do not satisfy mass balance could be because the rain gauges underestimate the total inputs to the catchment, or that the discharge rating curve (which relates measured water levels to discharge at an observation point) overestimates the discharges when extrapolated to larger events.

Other examples come from the uncertain identification of inundated areas by either post-event surveys or remote sensing (e.g. Mason et al., 2007), and observational data on the evolution of tropical cyclones (e.g. Hamill et al., 2011). These are just some examples of the more general problem of assessing the information content of data that is subject to epistemic uncertainties. Normal statistical methods of assessing information content (e.g. entropy, Fisher or Godambe measures, see Chandler, 2014) may not be appropriate when the error characteristics have complex structure, are non-stationary in time or space, or are physically inconsistent (Beven and Smith, 2015). Thus, as part of good practice, it is important to consider the meaning of available data carefully and critically, and recognize that not all data might be useful in contributing information to an analysis.

2.3 Quantifying uncertainties with expert judgments

Expert judgement is one of the main ways of trying to take account of epistemic uncertainties in natural hazards risk assessment. This may involve obtaining information by elicitation from knowledgeable experts about what assumptions are valid or appropriate for the problem at hand, what form any conceptual model or models should take, and which essential factors or parameters need to be included in any numerical analysis of uncertainties. In a more restricted sense, an elicitation may be conducted solely for quantifying parameter uncertainties as probability density distributions (e.g. PDFs) – which is the specific focus here.

However, in the context of almost any scientific elicitation, expert contributors generally utilize quick mental heuristics to make judgments under uncertainty. Such judgments can be affected by cognitive biases, such as “anchoring”, i.e. settling on an initial evaluation but then failing to adjust sufficiently for associated uncertainties. The existence of instinctive biases underlines the need for experts to be instructed about the potential for biases to influence judgments, highlighting their nature and how to minimize their impact. When a group of experts is used, their inputs need to be combined into a joint finding, and this will require the assistance of a facilitator, who ensures a structured procedure or methodology is followed. The facilitator can pool judgments mathematically, e.g. with an algorithmic approach (see below) or they can seek to derive an uncertainty distribution by helping

the expert group work towards consensus by discussion, the so-called “behavioural” approach. There are strengths and weaknesses with either; for a broader consideration of expert elicitations *sensu lato* and a comprehensive discussion of attendant issues, such as cognitive biases and how to reduce them, the recent report by RWM (2017) should be consulted.

For natural phenomena and in particular for rare or extreme events, experts can be asked to assign distributional estimates to the probabilities of occurrence or timing of events, or possibilities of potential outcomes, often using an algorithmic approach. Such estimates will be necessarily judgement-based and, while they may be epistemically incomplete in some regards, good experts will base their judgements on a comprehensive synthesis of available evidence and their tacit knowledge. When epistemic uncertainties dominate in natural hazards assessment, some individual experts will sometimes be surprised by events or outcomes and, in consequence, their judgments could appear inaccurate or uninformative *post hoc*. To mitigate such situations, it is therefore important to include the judgments of as many experts as possible. This, in turn, can raise questions about the statistical independence of experts, many of whom may have similar backgrounds and training, and about whether more weight should be given to the judgements of some experts relative to others, and about how this should best be done. For instance, different levels of experience may be an important factor when it comes to expert judgement performance.

Cooke (2014) gives a good recent summary of some of the issues involved in acquiring and pooling expert scientific judgments for uncertainty quantification as probability distributions (see the extensive supplementary material to that paper in particular). He points out that both Knight (1921) and Keynes (1921) suggested that the use of probabilities elicited from experts might be a working practical solution for dealing with these types of “real” uncertainties. A variety of methods have been proposed for assessing the value of experts and combining their judgements in an overall risk assessment (see Cooke, 1991; O’Hagan et al., 2006; and Aspinall and Cooke, 2013). Cooke (2014) includes a review of analyses that have been carried out, seeking to validate assessments conducted with the classical model structured expert judgment (SEJ) method (Cooke, 1991). This appraisal of applications, in a variety of specialisms, includes some for natural hazards (see Cooke and Goossens, 2008; Aspinall and Cooke, 2013; Aspinall and Blong, 2015).

A recent application of the Classical Model SEJ has provided an unprecedented opportunity to test the approach, albeit in a different field. The World Health Organization (WHO) undertook a global study involving 72 experts distributed over 134 expert panels, with each panel assessing between 10 and 15 calibration variables concerned with food-borne health hazards (source attribution, pathways and health risks of food-borne illnesses). Calibration variables drawn from the experts’ fields were used to gauge performance and to enable performance-based scoring combinations of their

judgments on the target items. The statistical accuracy of the experts overall was substantially lower than is typical with a Classical Model SEJ, a fact explained by logistical constraints imposed on the WHO elicitation process. However, viable performance-based weighted combinations were still obtained from each panel, based on individuals’ statistical accuracy and informativeness measures, determined from calibration variables. In this case, in-sample performance of the performance-based combination of experts (the “performance weights decision maker” PW DM) is somewhat degraded relative to other classical model SEJ studies (e.g. Cooke and Coulson, 2015) but, this said, performance weighting still out-performed equal weighting (“equal weights decision maker” EW DM) (Aspinall et al., 2016).

Because a large number of experts assessed similar variables it was possible to compare statistical accuracy and informativeness on a larger dataset than hitherto (Aspinall et al., 2016). For certain food-borne health hazards, some regions of the world were considered interchangeable, and so a panel could be used multiple times. Also, many experts participated in several distinct panels. For these reasons, any statistical analysis of results that considers the panels as independent experiments is impossible, and so proper out-of-sample analysis was infeasible. That said, this extensive study has provided new perspectives on the efficacy of SEJ. Most significant in this data set was the negative rank correlation between informativeness and statistical accuracy, and the finding that this correlation weakens when expert selection is restricted to those experts who are demonstrated by the Classical Model empirical calibration formulation to be more statistically accurate (Aspinall et al., 2016). These findings should motivate the development and deployment of enhanced facilitator and expert training, and advanced tools for remote elicitation of multiple, internationally dispersed panels – demand for which is growing in many disciplines (e.g. low probability high consequence natural hazards; climate change impacts; carbon capture and storage risks).

As noted above, consulting experts via a structured approach is one way of deciding on the appropriate methodology and on sets of assumptions for probabilistic assessment of a particular problem which, with associated probabilities, will define the workflow for the analysis (see next Section). In many instances, event probabilities can be difficult to estimate because of insufficient or unreliable data, and even informed reasoning can involve some subjectivity. This being the case, expert judgments are perhaps sometimes more analogous to betting odds than empirical probabilities (even if assessed over multiple experts). However, such “odds” can follow the axioms of probability – the original formulation of Bayes theorem was expressed in terms of defining the odds for a given proposition or hypothesis (Bayes, 1763) – and if accompanied by some expression of quantified uncertainty, they can be treated as probabilities for the convenience of analysis (as is the case for most fitted probability distributions in natural hazards assessments).

2.4 Choosing a methodology and defining workflows for applying a method correctly

The review of Part 1 of this paper shows that some differences in practice exist between different hazard areas, in part depending on the availability of data and the potential for useful forecasting as well as simulation. It is helpful to distinguish three types of uncertainty analysis (e.g. Beven, 2009).

The first is a forward analysis where the outputs depend entirely on propagating prior assumptions about the sources of uncertainty, albeit that those prior assumptions might be derived from historical sequences. Risk assessments of droughts, dam safety, landslides, ground motion from earthquakes, and tsunamis tend to be of this type. When trying to draw inferences about future hazard occurrences it can be difficult to define those prior assumptions so that, in such an analysis, the decisions about how far those assumptions truly reflect potential sources of epistemic uncertainty (for example in climate change projections) are paramount. This is, necessarily, an exercise in expert judgement, which may be formalized in the type of expert elicitation exercise discussed earlier.

The second form of analysis involves conditioning the prior estimates of uncertainty for a simulation model on observational data. Flood inundation maps (using historical flood outlines and discharge estimates), the inversion methods used for identifying earthquake ruptures, and source terms for ash cloud simulations are of this type. In general, such methods will help to constrain model uncertainties, but will be dependent on both the range of models considered and the way in which they are evaluated relative to the available observations. A number of conditioning methodologies are available including formal Bayes methods (Bernardo and Smith, 2009); Bayes linear methods (Goldstein and Wooff, 2007); approximate Bayesian computation (ABC, Vrugt and Sadegh, 2013; Nott et al., 2014) and generalized likelihood uncertainty estimation (GLUE, Beven and Binley, 1992, 2014; Blazkova and Beven, 2009). The latter can make use of formal and informal likelihood measures and limits of acceptability, as well as alternatives to Bayes rule in combining different model evaluations.

Because epistemic uncertainties are involved, including the potential for non-stationary bias and error characteristics and unknown commensurability errors between observed and predicted variables, it might not always be good practice to use formal statistical likelihood measures in such evaluations. Indeed, epistemic uncertainties make it difficult to test such models as hypotheses in rigorous ways, and may mean that multiple different model structures might all be consistent with the available observations (e.g. Beven, 2002, 2006, 2012). There may also be issues of whether the available models are fit-for-purpose when compared with observations, even when the uncertainty associated with those observations is taken into account. We should certainly be wary of using overly simple statistical analyses without test-

ing the validity of the assumptions made, or of disregarding important sources of uncertainty (such as is often done in flood risk analysis for example, where uncertainty in historical flood magnitudes is generally neglected). There will also be aspects of an analysis of epistemic uncertainties that may not be amenable to such conditioning, for example assumptions about potential future climate scenarios or other future boundary conditions. These will depend on the type of expert judgment or elicitation in a way similar to a forward uncertainty analysis.

The third form of uncertainty analysis can be used when the interest is in forecasting a hazard into the near future and when observables are available in real time to allow the use of data assimilation to constrain prediction uncertainties. A variety of data assimilation methods are available, from the variational methods commonly used in weather forecasting, to ensemble Kalman filters and particle filters. Such methods are used in real time forecasting of floods, ash clouds and wind storms (see Part 1 of this paper). It is perhaps instructive in the context of a discussion of epistemic uncertainties that in generating an ensemble of future weather forecasts, singular vector techniques are used to choose more extreme perturbations in formulating the members of the ensemble, so as to stand a greater chance of bracketing the potential range of future weather over the lead time of a few days. The ensemble members should therefore be considered to be of unknown probability, even if the outputs are sometimes interpreted in probabilistic ways (such as in decisions about alert status in the European Flood Awareness System based on simulated river flows forced by ECMWF ensemble predictions of rainfalls when compared to a long term historical reanalysis, see Pappenberger et al., 2013).

All of these forms of uncertainty analysis strategies and workflows are subject to epistemic uncertainties. The first form of forward analysis depends wholly on how they are incorporated into prior assumptions, the second allows for some conditioning of outcomes on available data, but cannot easily allow for differences in the future, the third can probably allow best for incorrectly specified assumptions, because of the potential for continuous updating and a limited time frame for the forecasts. However, in all cases it will be good practice to assess the sensitivity of the outcomes to a range of feasible assumptions (e.g. Saltelli, 2002; Tang et al., 2007; Saltelli et al., 2008; Pianosi et al., 2015; and Savage et al., 2016).

2.5 Assessing whether a decision is robust to the chosen assumptions

The primary reason for making uncertainty assessments for evaluating risk in natural hazards is because taking account of uncertainty might make a difference to the decision that is made (e.g. Hall and Solomatine, 2008; Rougier and Beven, 2013; Hall, 2003; and Simpson et al., 2016). For many decisions a complete, thoughtful, uncertainty assessment of risk

might not be justified by the cost in time and effort. In other cases, the marginal costs of such an analysis will be small relative to the potential costs and losses, so a more complete analysis using more sophisticated methods would be justifiable, including using expert elicitations in defining the assumptions of the relevant workflow.

Formal risk-based decision making requires probabilistic representations of both the hazard and consequence components of risk, i.e. an assumption that both hazard and consequences can be treated as aleatory variables, even if the estimates of the probabilities might be conditional and derived solely from expert elicitation or estimates of odds. The difficulty of specifying odds or probabilities for epistemic uncertainties means that any resulting decisions will necessarily be conditional on the assumptions (as discussed, for example, by Pappenberger and Beven, 2006; Beven, 2009; Sutherland et al., 2013; Rougier and Beven, 2013; and Juston et al., 2013).

However, it also leaves scope for other methodologies for uncertainty assessment, including fuzzy possibilistic reasoning, Dempster–Shafer evidence theory, Prospect Theory and Info-gap methods (see Shafer, 1976; Kahneman and Tversky, 1979; Halpern, 2003; Hall, 2003; Ben-Haim, 2006; and Wakker, 2010). There is some overlap between these methods, for example Dempster–Shafer evidence theory contains elements of fuzzy reasoning and imprecise reasoning, while both Prospect Theory and Info-Gap methods aim to show why non-optimal solutions might be more robust to epistemic uncertainties than classical risk based optimal decision making. There have been a few applications of these methods in the area of natural hazards, for example: Prospect Theory to seismic design optimization (Goda and Hong, 2008); and Info-Gap Theory to flood defence assessments (Hine and Hall, 2010), drought assessments in water resource management (Korteling et al., 2013), and earthquake resistant design criteria (Takewaki and Ben-Haim, 2005). All such methods require assumptions about the uncertainties to be considered, so they can be usefully combined with expert elicitation.

By definition we cannot distinguish between different choices about how to represent epistemic uncertainty through comparison with observations. However, we can always test how much it matters if we make different assumptions, if we change boundary conditions, or if we include the potential for data to be wrong. While most global sensitivity analysis approaches assume that we can define some probability distribution to characterize the potential variability of the inputs into the analysis, we might still gain useful information concerning whether the epistemic uncertainty related to an individual input might even matter from such a sensitivity analysis (see, for example, the suggestion of using conditional risk exceedance probability curves in Rougier and Beven, 2013). We now have formal approaches to sensitivity analysis that enable us to include discrete choices, e.g. different probability distribution functions, imprecise probabilities, or process representations (e.g. Hall, 2006; Baroni and Tarantola, 2014;

and Savage et al., 2016) or that allow us to explore the impact of distributions that are practically unconstrained due to a lack of observations (e.g. Prudhomme et al., 2010; Singh et al., 2014; Almeida et al., 2017). Alternatively, we can directly select an approach that attempts to find robust decisions in the presence of poorly bounded uncertainties (e.g. Bryant and Lempert, 2010; Steinschneider et al., 2015).

However, while there are important advantages in treating a risk analysis or reliability problem as a problem in assessing probabilities, this might lead to a degree of confidence in the outcomes that may not be justified. That will especially be the case when a statistical analysis is too simplified, for example in fitting a simple likelihood function to the residuals of a dynamic simulation (Beven and Smith, 2015; Beven, 2016) or fitting a magnitude-frequency distribution to historical data without allowing for the uncertainty in the data, the uncertainty in tail behaviour, the possibility of multiple underlying drivers in the hazard, or the potential for joint hazard occurrences. Asymptotic extreme value distributions are often applied to small samples without consideration of the potential for non-stationarity and dependencies in occurrences, or the resulting uncertainty in the frequency estimates of extremes. These are (common) examples of poor practice, especially where the assumptions are not checked for validity.

Thus the first stage in assessing robustness in this respect is to check the validity of the assumptions, wherever that is possible. Where model errors can be evaluated against historical data, they should be checked for consistency, stationarity, and residual structure. Where there is a cascade of nonlinear model components, the effect of each component on the structure of the output errors should be assessed (Gaussian input errors will not remain Gaussian when processed through a nonlinear model structure, and the resulting non-stationarity in error characteristics might be difficult to represent in any likelihood evaluation or conditioning exercise). In such cases a more formal, sophisticated, analysis will be justified with at least a sensitivity analysis of the outcomes to different assumptions when assessing robustness.

It is clear from discussions amongst the authors and in the refereeing of this paper, that opinions vary as to whether a more sophisticated statistical analysis is already sufficient. It is certainly more difficult to assess the impact of analysing the probabilities as if they were incomplete. This leaves open the potential for future surprise when the epistemic uncertainties are such that the probabilities are not well defined. Falling outside historical data support, a surprise event is intrinsically non-probabilistic (and might not be easily be foreseen or even perceived by expert elicitation. This is epistemic uncertainty as a form of “unknown unknowns” (or very hard to evaluate unknowns). However, the potential for some form of surprise event is often envisaged, which suggests that where this might have catastrophic consequences it should be made an explicit part of a robust analysis.

Since surprise does not enter easily into a risk-based decision making framework, the question is how to allow for

incompleteness? Techniques might include an extension of the expert elicitation approach to consider the unexpected; a form of Info-Gap analysis, looking at impact of more extreme events and the opportuneness benefit if they do not occur (Ben-Haim, 2006); an evaluation of the most extreme event to be expected, even if of unknown probability; or a type of factor of safety approach based on costs of being precautionary. All of these approaches are associated with their own epistemic uncertainties, but addressing the question of surprise and whether the probabilities of a risk analysis are properly defined or incomplete is, at least, a step in the direction of good practice.

Potential climate change impacts on future risk represent one example where the problem of incomplete uncertainties and discrete choices is central to the analysis. Here, we can make assumptions and give different projections (e.g. from different Global Circulation Models, or from different down-scaling procedures, or from different scenarios or ways of implementing future change factors) equal probability. But this is a case where the range of possibilities considered may not be complete (e.g. Collins et al., 2012) and climate change might not be the only factor affecting future risk (Wilby and Dessai, 2010). We could invoke an expert elicitation to say whether a particular projection is more likely than another and to consider the potential for changes outside the range considered, but, as noted earlier, it can be difficult sometimes to find experts whose views are independent of the various modelling groups. It is also not clear whether the modelling groups themselves fully understand the potential for uncertainties in their simulations, given the continuing computing constraints on this type of simulation.

Consideration of climate change risks in these contexts has to confront a trio of new quantitative hazard and risk assessment challenges: micro-correlations, fat tails and tail dependence (e.g. Kousky and Cooke, 2009). These are distinct aspects of loss distributions which challenge traditional approaches to managing risk. Micro-correlations are tiny or local correlations that, as individual factors, may be harmless, but very dangerous when they coincide and operate in concert to create extreme cases. Fat tails can apply to losses whose probability declines slowly, relative to their severity. Tail dependence is the propensity of a number of extreme events or severe losses to happen together. If one does not know how to detect these phenomena or relationships, it is easy to not see them, let alone cope with or predict them adequately. Dependence modelling is an active research topic, and methods for dependence elicitation are still very much under development (e.g. Morales et al., 2008).

It is hard to believe that current natural hazard and climate models are not subject to these types of epistemic uncertainties. In such circumstances, a shortage of empirical data inevitably requires input from expert judgment to determine relevant scenarios to be explored. How these behaviours and uncertainties are best elicited can be critical to a decision process, as differences in efficacy and robustness of the elic-

itation methods can be substantial. When performed rigorously, expert elicitation and pooling of experts' opinions can be powerful means for obtaining rational estimates of uncertainty.

2.6 Communicating the meaning of uncertainty analysis and visualising the outcomes for effective decision-making

Given all the assumptions that are required to deal with epistemic uncertainties in natural hazard and risk analysis, there are real issues about communicating the meaning of an uncertainty assessment to potential users (e.g. Faulkner et al., 2007, 2014). Good practice in dealing with different sources of uncertainty should at least involve a clear and explicit statement of the assumptions of a particular analysis (Hall and Solomatine, 2008). Beven and Alcock (2012) suggest that this might be expressed in the form of condition trees that can be explicitly associated with any methodological workflow. The condition tree is a summary of the assumptions and auxiliary conditions for an analysis of uncertainty. The tree may be branched in that some steps in the analysis might have subsidiary assumptions for different cases. The approach has two rather nice features. Firstly, it provides a framework for the discussion and agreement of assumptions with experts and potential users of the outcomes of the analysis. This then facilitates communication of the meaning of the resulting uncertainty estimates to those users. Secondly, it provides a clear audit trail for the analysis that can be reviewed and evaluated by others at a later date.

The existence of the audit trail might focus attention on appropriate justification for some of the more difficult assumptions that need to be made; such as how to condition model outputs using data subject to epistemic uncertainties and how to deal with the potential for future surprise (see Beven and Alcock, 2012). Application of the audit trail in the forensic examination of extreme events as (and when) they occur might also lead to a revision of the assumptions as part of an adaptive learning process for what should constitute good practice.

Such condition trees can be viewed as parallel to the logic trees or belief networks used in some natural hazards assessments, but focussed on the nature of the assumptions about uncertainty that leads to conditionality of the outputs of such analyses. Beven et al. (2014) and Beven and Lamb (2017) give examples of the application of this methodology to mapping inundation footprints for flood risk assessment.

In understanding the meaning of uncertainty estimates, particularly when epistemic uncertainties are involved, understanding the assumptions on which the analysis is based is only a starting point. In many natural hazards assessments those uncertainties will have spatial or space-time variations that users need to appreciate. Thus visualization of the outcomes of an uncertainty assessment has become increasingly important as the tools and computational resource available

have improved in the last decade and a variety of techniques have been explored to represent uncertain data (e.g. Johnson and Sanderson, 2003; MacEachren et al., 2005; Pang, 2008; Kunz et al., 2011; Friedemann et al., 2011; Spiegelhalter et al., 2011; Spiegelhalter and Reisch, 2011; Jupp et al., 2012; Potter et al., 2012).

One of the issues that arise in visualization is the uncertainty induced by the visualization method itself, particularly where interpolation of point predictions might be required in space and/or time (e.g. Agumya and Hunter, 2002; Couclelis, 2003). The interpolation method will affect the user response to the visualization in epistemically uncertain ways. Such an effect might be small but it has been argued that, now that it is possible to produce convincing virtual realities that can mimic reality to an apparently high degree of precision, we should be wary about making the visualizations too convincing so as not to induce an undue belief in the model predictions and assessments of uncertainty in the potential user (e.g. Dottori et al., 2013; Faulkner et al., 2014).

Some examples of visualizations of uncertainty in natural hazard assessments have been made for flood inundation (e.g. Beven et al., 2014; Faulkner et al., 2014; Leedal et al., 2010; Pappenberger et al., 2013); seismic risk (Bostrom et al., 2008); tsunami hazard and risk (Goda and Song, 2016); volcanic hazard (Marzocchi et al., 2010; Wadge and Aspinall, 2014; Baxter et al., 2014); and ice-sheet melting due to global temperature change (Bamber and Aspinall 2013). These are all cases where different sources of uncertainty have been represented as probabilities and propagated through a model or cascade of models to produce an output (or outputs) with enumerated uncertainty. The presentation of this uncertainty can be made in different ways and can involve interaction with the user as a way of communicating meaning (e.g. Faulkner et al., 2014). But, as noted in the earlier discussion, this is not necessarily an adequate way of representing the “deeper” epistemic uncertainties, which are not easily presented as visualizations (Spiegelhalter et al., 2011; Sutherland et al., 2013).

3 Epistemic uncertainty and institutionalized risk assessments

The framework for good practice discussed in the previous section allows for considerable freedom in the choice of methodology and assumptions, but it has been suggested that these choices must be justified and made explicit in the workflow, condition tree, and audit trail associated with an analysis. This contrasts with an approach that has been common in many fields of natural hazards assessment where those choices have been institutionalized, with specified rules usually based on expert judgement, to provide science-informed planning for dealing with potentially catastrophic natural hazards.

One example of such an institutionalized approach is in the use of estimates of annual exceedance probabilities that are used for natural hazards planning in different countries. In the UK, frequency-based flood magnitude estimates are used for flood defence design and planning purposes in rather deterministic ways. For fluvial flooding, defences are designed to deal with the rare event (annual exceedance probability of less than 0.01). The footprint of such an event is used to define a planning zone. The footprint of a very rare event (annual exceedance probability of less than 0.001) is used to define an outer zone. Other countries have their own design standards and levels of protection.

While there is no doubt that both deterministic and frequency assessments are subject to many sources of epistemic uncertainty, such rules can be considered as structured ways of dealing with those uncertainties. The institutionalized, and, in some cases, statutory, levels of protection are then a political compromise between costs and perceived benefits. Such an approach is well developed in earthquake engineering in assessing life cycle costs and benefits, at least for uncertainties that can be defined as probability distributions (e.g. Takahashi et al., 2004). Flood defence is also an example where the analysis can be extended to a life cycle risk-based decision analysis, with costs and benefits integrated over the expected frequency distribution of events (Sayers et al., 2002; Voortman et al., 2002; Hall and Solomatine, 2008). In the Netherlands, for example, where more is at risk, fluvial flood defences are designed to deal with an event with an annual exceedance probability of 0.0008, and coastal defences to 0.00025. In doing so, of course, there is an implicit assumption of stationarity of the hazard when based only on the analysis of current and historical data.

There have also been practical approaches suggested based on estimating or characterising the largest event to be expected in any location of interest. Such a maximal event might be a good approximation to very rare events, especially when the choice of distribution for the extremes is bounded. In hydrological applications, the concepts of the probable maximum precipitation and probable maximum flood have a long history (e.g. Hershfield, 1963; Newton, 1983; Hansen, 1987; Douglas and Barros, 2003; Kunkel et al., 2013) and continue to be used, for example, in dam safety assessments (e.g. Graham, 2000, and Part 1 of this paper). In evaluating seismic safety of critical infrastructure (e.g. nuclear power plants and dams), there have been some who would move away from probabilistic assessment of, say, future earthquake magnitudes, preferring the concept of a deterministic maximum estimate of magnitude (McGuire, 2001; Panza et al., 2008; Zuccolo et al., 2011). These “worst case” scenarios can be used in decision making but are clearly associated with their own epistemic uncertainties and have been criticized because of the assumptions that are made in such analyses (e.g. Koutsoyiannis, 1999; Abbs, 1999; Bommer, 2002).

An important problem with this type of institutionalized analysis is that sensitivity to the specified assumptions is

rarely investigated and uncertainties in such assessments are often ignored. A retrospective evaluation of the recent very large Tōhoku, Japan, earthquake indicates that the uncertainty of the maximum magnitude in subduction zones is considerable and, in particular, it is argued that the upper limit in this case should have been considered unbounded (Kagan and Jackson, 2013). However, this reasoning has the benefit of hindsight. Earlier engineering decisions relating to seismic risk at facilities along the coast opposite the Tōhoku subduction zone had been made on the basis of work by Ruff and Kanamori (1980), repeated by Stern (2002). Stern reviewed previous studies and described the NE Japan subduction zone (Fig. 7 of Stern, 2002) as a “good example of a cold subduction zone”, denoting it the “old and cold” end-member of his thermal models. Relying on Ruff and Kanamori (1980), Stern re-presented results of a regression linking “maximum magnitude” to subduction zone convergence rate and age of oceanic crust. This relationship was said to have a “strong influence... on seismicity” (Stern, 2002), and indicated a modest maximum moment magnitude (M_w) of 8.2 for the NE Japan subduction zone. It is not surprising, therefore, that such authoritative scientific sources were trusted for engineering risk decisions.

This said, MacCaffrey (2008) had pointed out, before the Tōhoku earthquake, that the history of observations at subduction zones is much shorter than the recurrence times of very large earthquakes, suggesting the possibility that any subduction zone may produce earthquakes larger than M_w 9. In respect of the Ruff and Kanamori relationship, epistemic uncertainties (as discussed in Sect. 2.5 above) were unquestionably present, and almost certainly large enough, to undermine the robustness of that regression for characterising a long-term geophysical process. Thus, epistemic uncertainties for any purported maximum event should be carefully discussed from both probabilistic and deterministic viewpoints, as the potential consequences due to gross underestimation of such events can be catastrophic (e.g. Goda and Abilova, 2016).

These deterministic maximal event approaches do not have probabilities associated with them, but can serve a risk averse, institutionalized role in building design or the design of dam spillways, say, without making any explicit uncertainty estimates. In both of these deterministic and probabilistic scenario approaches, the choice of an established design standard is intended to make some allowance for what is not really known very well, but with the expectation that, despite epistemic uncertainties, protection levels will be exceeded sufficiently rarely for the risk to be acceptable.

Another risk averse strategy for coping with lack of knowledge is in the factors of safety that are present in different designs for protection against different types of natural hazard, for example in building on potential landslide sites when the effective parameters of slope failure models are subject to significant uncertainty. In flood defence design, the concept of “freeboard” is used to raise flood embankments or other

types of defences. Various physical arguments can be used to justify the level of freeboard (see, for example, Kirby and Ash, 2000) but the concept also serves as a way of institutionalising the impacts of epistemic uncertainty.

Such an approach might be considered reasonable where the costs of a more complete analysis cannot be justified, but this can also lead to overconfidence in cases where the consequences of failure might be severe. In such circumstances, and for managing risks, it will be instructive to make a more detailed analysis of plausible future events, their impacts and consequences.

4 Epistemic uncertainty and scientific practice

How to define different types of uncertainty, and the impact of different types of uncertainty on testing scientific models as hypotheses has been the subject of considerable philosophical discussion that cannot be explored in detail here (but see, for example, Howson and Urbach, 1993; Mayo, 1996; Halpern, 2003; Mayo and Spanos, 2010; Gelman and Shalizi, 2013). As noted earlier, for making some estimates, or at least prior estimates, of epistemic uncertainties we will often be dependent on eliciting the knowledge of experts. Both in the Classical Model of Cooke (1991, 2014) and in a Bayesian framework (O’Hagan et al., 2006), we can attempt to give the expert elicitation some scientific rigour by providing empirical control on how well the evaluation of the informativeness of experts has worked. Empirical control is a basic requirement of any scientific method and a *sine qua non* for any group decision process that aspires to be rational and to respect the axioms of probability theory.

Being scientific about testing the mathematical models that are used in risk assessments of natural hazards is perhaps less clear cut. Models can be considered as hypotheses about the functioning of the real world system. Hypothesis testing is normally considered the domain of statistical theory (such as the severe testing in the error statistical approach of Mayo, 1996), but statistical theory (for the most part) depends on strongly aleatory assumptions about uncertainty that are not necessarily appropriate for representing the effects of epistemic sources of uncertainty (e.g. Beven and Smith, 2015; Beven, 2016). Within the Bayesian paradigm, there are ways of avoiding the specification of a formal aleatory error model, such as in the use of expectations in Bayes linear methods (Goldstein and Wooff, 2007), in approximate Bayesian computation (Diggle and Gratton, 1984; Vrugt and Sadegh, 2013; Nott et al., 2014), or in the informal likelihood measures of the GLUE methodology (Beven and Binley, 1992, 2014; Smith et al., 2008). It is still possible to empirically control the performance of any such methodology in simulating past data, but, given the epistemic nature of uncertainties this is no guarantee of good predictive performance in future projections. In particular, if we have determined that some events might be disinformative for model calibration purposes, in

forecasting or simulation we will not then know if the next event would be classified as informative or disinformative if the observed data were made available, with important implications for prediction uncertainties (Beven and Smith, 2015).

In this context, it is interesting to consider what would constitute a severe test (in the sense of Mayo) for a natural hazard risk assessment model. In the Popperian tradition, a severe test is one that we would expect a model could fail. However, all natural hazards models are approximations, and if tested in too much detail (severely) are almost certain to fail. We would hope that some models might still be informative in assessing risk, even if there are a number of celebrated examples of modelled risks being underestimated when evaluated in hindsight (see Part 1 of this paper). And, since the boundary condition data, process representations, and parameters characteristic of local conditions are themselves subject to epistemic uncertainties, then any such test will need to reflect what might be feasible in model performance conditional on the data available to drive it and assess that performance. Recent sensitivity studies of Regional Climate Models, for example, have suggested that they cannot adequately reproduce high intensity convective rainstorms and consequently might not be fit-for-purpose in predicting conditions for future flooding induced by such events (e.g. Kendon et al., 2014; Meredith et al., 2015). High resolution, convection-resolving models can do better (e.g. the CRM2 model at 2.2 km resolution, reported by Ban et al., 2014), but will still be subject to epistemic uncertainties in the representation of the boundary conditions both at the edge of the domain (from simpler larger scale atmospheric models) and in representing the hydrology of the land surface (from simple land surface parameterizations).

Some recent applications within the GLUE framework have used tests based on limits of acceptability determined from an assessment of data uncertainties before running the model (e.g. Liu et al., 2009; Blazkova and Beven, 2009). Such limits can be normalized across different types and magnitudes of evaluation variables. Perhaps unsurprisingly, it has been found that only rarely does any model run satisfy all the specified limits. This will be in part because there will be anomalies or disinformation in the input data (or evaluation observations) that might be difficult to assess a priori. This could be a reason for relaxing the severity of the test such that only 95 % of the limits need be satisfied (by analogy with statistical hypothesis testing) or relaxing the limits if we can justify not taking sufficient account of input error. However, in modelling river discharges, it has been found that the remaining 5 % might be associated with the peak flood flows or drought flows which are the characteristics of most interest in natural hazards. Concluding that the model does not pass the limits test can be considered a good thing (in that we need to do better in finding a better model or improving the quality of the data, or making a decision in a more precautionary way). It is one way of improving the science in situations where epistemic uncertainties are significant.

This situation does not arise if there are no conditioning observations available so that only a forward uncertainty analysis is possible, but we should be aware in considering the assumptions that underlie the condition tree discussed earlier, that such a forward model might later prove to be falsified by future observational data. And, if we cannot argue away such failure, then it will be necessary to seek some other methodology for the risk assessment that better accommodates gaps in our knowledge or understanding.

5 Conclusion and summary of recommendations

In assessing future risks due to natural hazards it is generally necessary to resort to the use of a model of some form, even if that is only a frequency distribution for expectations of future magnitudes of some hazard. Even in that simple case, there will be limitations in the knowledge of what distribution should be assumed, especially when the database of past events is sparse. For risk-based decision-making the consequences of events must also be modelled in some way and these, equally, are liable to be subject to uncertainties due to limited knowledge. Even though our simulation may be shown to match well a sample of past data, they may not perform adequately in future because of uncertainty about future boundary conditions and potential changes in system behaviour. All of these (sometimes rather arbitrary) sources of epistemic uncertainty are inherently difficult to assess and, in particular, to represent as probabilities, even if we do recognize that those probabilities might be judgement-based, conditional on current knowledge, and subject to future revision as expert knowledge increases. As Morgan (1994) notes, throughout history decisions have always been made without certain knowledge, but mankind has mostly muddled through.

But, this rather underplays the catastrophic consequences of some poor decisions (including the recent examples of the L'Aquila earthquake in Italy, and the Fukushima tsunami in Japan), so there is surely scope for better practice in future in trying to allow for: all models being intrinsically imprecise; most uncertainties being epistemic (even if some might be treated as if aleatory); all uncertainty estimates being conditional; a tendency for expert elicitations to underestimate potential uncertainties, and the potential for future surprises to happen. Epistemic uncertainty suggests that we should be prepared for surprises and should be wary of both observational data and models which, while they might be the best we have, might not be necessarily fit for purpose in some sense. Surprises will occur when the probability estimates are incomplete, when the distribution tails associated with extremes are poorly estimated given the data available, or where subtle high dimensional relationships are not recognized or are ignored.

In this paper, a series of six stages has been suggested as a way of structuring good practice and conveying the meaning

of uncertainty estimates in natural hazards assessments in future. The recommendations associated with each of these six stages can be summarized as follows:

1. *Framing the analysis, preferably with input from potential users.* This should aim to identify key sources of uncertainty, those which might be treated as if they are aleatory and those for which information is lacking and that might therefore have the potential to produce future surprises.
2. *Evaluating the available data.* Epistemic uncertainties associated with data should be considered more explicitly, especially when they might lead to inconsistencies with the fundamental concepts of a model or analysis (e.g. the potential for disinformation).
3. *Eliciting information on sources of uncertainty from experts.* When sources of uncertainty are identified as subject to a paucity or lack of knowledge then expert elicitation is a potentially useful way of constraining the analysis. This requires a structured approach to avoid the underestimation of uncertainty and bias in estimates.
4. *Defining a workflow.* Given the information gained from steps 1 to 3, the choice of methodology should be considered carefully and implemented as a workflow that will give reliable and accurate results. It is important not to make overly simplistic assumptions, particularly when epistemic uncertainties contain the potential for future non-stationarities or surprises.
5. *Assessing robustness to uncertainty.* The sensitivity of the outcomes to assumptions made in the workflow should be assessed, including the impact on any decisions depending on the analysis.
6. *Communicating the findings and meaning of the analysis to potential users, stakeholders and decision makers.* Communication should involve a summary of assumptions (a condition tree) and visualizations helpful in conveying the nature of the uncertainty outputs, while recognizing that the deeper epistemic uncertainties might not be amenable to visualizations.

How to deal with epistemic uncertainties in all areas of natural hazards remains a difficult problem and requires further research in trying to define good practice. This is particularly the case for methods of assessing the models that are used in natural hazard risk assessments and their fitness-for-purpose in a scientifically rigorous way (e.g. Beven, 2018). Adopting the six stages suggested here would represent one beneficial step in the direction of good practice.

Data availability. No data sets were used in this article.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue “Risk and uncertainty estimation in natural hazards”. It does not belong to a conference.

Acknowledgements. This work is a contribution to the CREDIBLE consortium funded by the UK Natural Environment Research Council (Grant NE/J017299/1). Thanks are due to Michael Goldstein, the referees, and guest editor for comments on earlier versions of the paper.

Edited by: Richard Chandler

Reviewed by: Mike Poole and one anonymous referee

References

- Abbs, D. J.: A numerical modeling study to investigate the assumptions used in the calculation of probable maximum precipitation, *Water Resour. Res.*, 35, 785–796, <https://doi.org/10.1029/1998WR900013>, 1999.
- Agumya, A. and Hunter, G. J.: Responding to the consequences of uncertainty in geographical data, *Int. J. Geogr. Info. Sci.*, 16, 405–417, 2002.
- Almeida, S., Holcombe, E. A., Pianosi, F., and Wagener, T.: Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change, *Nat. Hazards Earth Syst. Sci.*, 17, 225–241, <https://doi.org/10.5194/nhess-17-225-2017>, 2017.
- Aspinall, W. and Blong, R.: Volcanic Risk Management, Chapter 70 in: *The Encyclopedia of Volcanoes*, edited by: Sigurdsson, H., Houghton, B., McNutt, S., Rymer, H., and Stix, J., 2nd Edition, Academic Press ISBN 978-0-12-385938-9, 1215–1234, 2015.
- Aspinall, W. P. and Cooke, R. M.: Expert Elicitation and Judgement, in: *Risk and Uncertainty assessment in Natural Hazards*, edited by: Rougier, J. C., Sparks, R. S. J., and Hill, L., Cambridge University Press, Chapter 4, 64–99, 2013.
- Aspinall, W. P., Cooke, R. M., Havelaar, A. H., Hoffmann, S., and Hald, T.: Evaluation of a Performance-Based Expert Elicitation: WHO Global Attribution of Foodborne Diseases, *PLoS ONE*, 11, e0149817, <https://doi.org/10.1371/journal.pone.0149817>, 2016.
- Bamber, J. and Aspinall, W. P.: An expert judgement assessment of future sea level rise from the ice sheets, *Nat. Clim. Change*, 3, 424–427 <https://doi.org/10.1038/nclimate1778>, 2013.
- Ban, N., Schmidli, J., and Schär, C.: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations, *J. Geophys. Res.-Atmos.*, 119, 7889–7907, 2014.
- Baroni, G. and Tarantola, S.: A general probabilistic framework for uncertainty and global sensitivity analysis of deterministic models: a hydrological case study, *Environ. Model. Softw.*, 51, 26–34, 2014.
- Baxter, P. J., Searl, A., Cowie, H. A., Jarvis, D., and Horwell, C. J.: Evaluating the respiratory health risks of volcanic ash at the eruption of the Soufrière Hills Volcano, Montserrat, 1995–2010, in: *The Eruption of Soufrière Hills Volcano, Montserrat from 2000 to 2010*, edited by: Wadge, G., Robertson, R. E. A., and Voight,

- B., *Memoir of the Geological Society of London*, 39, 407–425, 2014.
- Bayes, T.: An essay towards solving a problem in the doctrine of chances, *Phil. Trans. Roy. Soc. Lond.*, 53, 370–418, 1763.
- Ben-Haim, Y.: *Info-gap decision theory: decisions under severe uncertainty*, Academic Press, 2006.
- Bernardo, J. M. and Smith, A. F. M.: *Bayesian Theory*, Vol. 405, Wiley, Chichester, 2009.
- Beven, K. J.: Towards a coherent philosophy for environmental modelling, *Proc. Roy. Soc. Lond. A*, 458, 2465–2484, 2002.
- Beven, K. J.: A manifesto for the equifinality thesis, *J. Hydrology*, 320, 18–36, 2006.
- Beven, K. J.: *Environmental Models: An Uncertain Future?*, Routledge, London, 2009.
- Beven, K. J.: Causal models as multiple working hypotheses about environmental processes, *Comptes Rendus Geoscience, Académie de Sciences, Paris*, 344, 77–88, <https://doi.org/10.1016/j.crte.2012.01.005>, 2012.
- Beven, K. J.: EGU Leonardo Lecture: Facets of Hydrology – epistemic error, non-stationarity, likelihood, hypothesis testing, and communication, *Hydrol. Sci. J.*, 61, 1652–1665, <https://doi.org/10.1080/02626667.2015.1031761>, 2016.
- Beven, K. J.: On hypothesis testing in hydrology: why falsification of models is still a really good idea, *WIRES Water*, 5, e1278, <https://doi.org/10.1002/wat2.1278>, 2018.
- Beven, K. J. and Alcock, R.: Modelling everything everywhere: a new approach to decision making for water management under uncertainty, *Freshwater Biol.*, 56, 124–132, <https://doi.org/10.1111/j.1365-2427.2011.02592.x>, 2012.
- Beven, K. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, 1992.
- Beven, K. and Binley, A.: GLUE: 20 years on, *Hydrol. Process.*, 28, 5897–5918, 2014.
- Beven, K. J. and Lamb, R.: The uncertainty cascade in model fusion, in: *Integrated Environmental Modelling to Solve Real World Problems: Methods, Vision and Challenges*, edited by: Riddick, A. T., Kessler, H., and Giles, J. R. A., Geological Society, London, Special Publications, 408, <https://doi.org/10.1144/SP408.3>, 2017.
- Beven, K. J. and Smith, P. J.: Concepts of Information Content and Likelihood in Parameter Calibration for Hydrological Simulation Models, *ASCE J. Hydrol. Eng.*, 20, A4014010, [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000991](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000991), 2015.
- Beven, K., Smith, P. J., and Wood, A.: On the colour and spin of epistemic error (and what we might do about it), *Hydrol. Earth Syst. Sci.*, 15, 3123–3133, <https://doi.org/10.5194/hess-15-3123-2011>, 2011.
- Beven, K. J., Leedal, D. T., and McCarthy, S.: Framework for assessing uncertainty in fluvial flood risk mapping, CIRIA report C721/2014, available at: http://www.ciria.org/Resources/Free_publications/fluvial_flood_risk_mapping.aspx, 2014.
- Beven, K. J., Almeida, S., Aspinall, W. P., Bates, P. D., Blazkova, S., Borgomeo, E., Freer, J., Goda, K., Hall, J. W., Phillips, J. C., Simpson, M., Smith, P. J., Stephenson, D. B., Wagener, T., Watson, M., and Wilkins, K. L.: Epistemic uncertainties and natural hazard risk assessment – Part 1: A review of different natural hazard areas, *Nat. Hazards Earth Syst. Sci.*, 18, 2741–2768, <https://doi.org/10.5194/nhess-18-2741-2018>, 2018.
- Blazkova, S. and Beven, K.: A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, 45, W00B16, <https://doi.org/10.1029/2007WR006726>, 2009.
- Bommer, J. J.: Deterministic vs. probabilistic seismic hazard assessment: an exaggerated and obstructive dichotomy, *J. Earthq. Eng.*, 6, 43–73, 2002.
- Bostrom, A., Anselin, L., and Farris, J.: Visualizing seismic risk and uncertainty, *Ann. New York Acad. Sci.*, 1128, 29–40, 2008.
- Bryant, B. P. and Lempert, R. J.: Thinking inside the box: A participatory, computer-assisted approach to scenario discovery, *Technol. Forecast. Social Change*, 77, 34–49, 2010.
- Chandler, R. E.: Classical Approaches for Statistical Inference in Model Calibration with Uncertainty, Chapter 4 in: *Applied Uncertainty Analysis for Flood Risk Management*, edited by: Beven, K. J. and Hall, J. W., Imperial College Press, London, 2014.
- Collins, M., Chandler, R. E., Cox, P. M., Huthnance, J. M., Rougier, J. C., and Stephenson, D. B.: Quantifying future climate change, *Nat. Clim. Change*, 2, 403–409, 2012.
- Cooke, R. M.: *Experts in uncertainty: Opinion and Subjective Probability in Science*, Oxford University Press: Oxford, 1991.
- Cooke, R. M.: Messaging climate change uncertainty, *Nat. Clim. Change* 5, 8–10 <https://doi.org/10.1038/nclimate2466>, 2014.
- Cooke, R. M. and Coulson, A.: In and out of sample validation for the classical model of structured expert judgment, *Resources for Future*, Washington DC, 2015.
- Cooke, R. M. and Goossens, L. L.: TU Delft expert judgment data base, *Reliab. Eng. Syst. Safe.*, 93, 657–674, 2008.
- Couclelis, H.: The certainty of uncertainty: GIS and the limits of geographic knowledge, *Trans. GIS*, 7, 165–175, 2003.
- Diggle, P. J. and Gratton, J.: Monte Carlo Methods of Inference for Implicit Statistical Models, *J. Roy. Stat. Soc. B*, 46, 193–227, 1984.
- Dottori, F., Di Baldassarre, G., and Todini, E.: Detailed data is welcome, but with a pinch of salt: Accuracy, precision, and uncertainty in flood inundation modeling, *Water Resour. Res.*, 49, 6079–6085, 2013.
- Douglas, E. M. and Barros, A. P.: Probable maximum precipitation estimation using multifractals: application in the Eastern United States, *J. Hydrometeorol.*, 4, 1012–1024, 2003.
- Faulkner, H., Parker, D., Green, C., and Beven, K.: Developing a translational discourse to communicate uncertainty in flood risk between science and the practitioner, *Ambio*, 16, 692–703, 2007.
- Faulkner, H., Alexander, M., and Leedal, D.: Translating uncertainty in flood risk science, Chapter 24 in: *Applied Uncertainty Analysis for Flood Risk Management*, edited by: Beven, K. J. and Hall, J. W., Imperial College Press, London, 2014.
- Friedemann, M., Raape, U., Tessmann, S., Schoeckel, T., and Strobl, C.: Explicit modeling and visualization of imperfect information in the context of decision support for tsunami early warning in Indonesia, in: *Human Interface and the Management of Information. Interacting with Information*, 201–210, Springer Berlin Heidelberg, 2011.
- Gelman, A. and Shalizi, C. R.: Philosophy and the practice of Bayesian statistics, *Brit. J. Math. Stat. Psy.*, 66, 8–38, 2013.
- Goda, K. and Abilova, K.: Tsunami hazard warning and risk prediction based on inaccurate earthquake source parameters, *Nat. Haz-*

- ards Earth Syst. Sci., 16, 577–593, <https://doi.org/10.5194/nhess-16-577-2016>, 2016.
- Goda, K. and Hong, H. P.: Application of cumulative prospect theory: implied seismic design preference, *Struct. Safety*, 30, 506–516, 2008.
- Goda, K. and Song, J.: Uncertainty modeling and visualization for tsunami hazard and risk mapping: a case study for the 2011 Tohoku earthquake, *Stoch. Environ. Res. Risk Assess.*, 30, 2271–2285, <https://doi.org/10.1007/s00477-015-1146-x>, 2016.
- Goldstein, M. and Wooff, D.: *Bayes Linear Statistics, Theory & Methods*, Vol. 716, John Wiley & Sons, 2007.
- Graham, W. J.: Should dams be modified for the probable maximum flood?, *JAWRA J. Am. Water Resour. Assoc.*, 36, 953–963, <https://doi.org/10.1111/j.1752-1688.2000.tb05701.x>, 2000.
- Hall, J. W.: Handling uncertainty in the hydroinformatic process, *J. Hydroinformatics*, 5.4, 215–232, 2003.
- Hall, J. W.: Uncertainty-based sensitivity indices for imprecise probabilities, *Reliab. Eng. Syst. Safe.*, 91, 1443–1451, 2006.
- Hall, J. and Solomatine, D.: A framework for uncertainty analysis in flood risk management decisions, *Int. J. River Basin Manage.*, 6, 85–98, 2008.
- Halpern, J. Y.: *Reasoning about Uncertainty*, Cambridge MS, MIT Press, 2003.
- Hamill, T. M., Whitaker, J. S., Fiorino, M., and Benjamin, S. G.: Global Ensemble Predictions of 2009's Tropical Cyclones Initialized with an Ensemble Kalman Filter, *Mon. Weather Rev.*, 139, 668–688, <https://doi.org/10.1175/2010MWR3456.1>, 2011.
- Hansen, E. M.: Probable maximum precipitation for design floods in the United States, *J. Hydrol.*, 96, 267–278, 1987.
- Hershfield, D. M.: Estimating the probable maximum precipitation, *Trans. Am. Soc. Civil Eng.*, 128, 534–551, 1963.
- Hine, D. and Hall, J. W.: Information gap analysis of flood model uncertainties and regional frequency analysis, *Water Resour. Res.*, 46, W01514, <https://doi.org/10.1029/2008WR007620>, 2010.
- Howson, C. and Urbach, P.: *Scientific Reasoning: the Bayesian approach*, 2nd Edition, Open Court: Chicago, 1993.
- Johnson, C. R. and Sanderson, A. R.: A Next Step: Visualizing Errors and Uncertainty, *IEEE Comput. Graph.*, 23, 6–10, 2003.
- Jupp, T. E., Lowe, R., Stephenson, D. B. S., and Coelho, C. A. S.: On the visualization, verification and recalibration of ternary probabilistic forecasts, *Phil. Trans. Roy. Soc. Lond.*, 370, 1100–1120, 2012.
- Juston, J. M., Kauffeldt, A., Montano, B. Q., Seibert, J., Beven, K. J., and Westerberg, I. K.: Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling, *Hydrol. Process.*, <https://doi.org/10.1002/hyp.9625>, 2013.
- Kagan, Y. Y. and Jackson, D. D.: Tohoku earthquake: a surprise?, *B. Seismol. Soc. Am.*, 103, 1181–1191, 2013.
- Kahneman, D. and Tversky, A.: Prospect theory: An analysis of decision under risk, *Etrica: Journal of the Econometric Society*, 47, 263–291, 1979.
- Kendon, E. J., Roberts, N. M., Fowler, H. J., Roberts, M. J., Chan, S. C., and Senior, C. A.: Heavier summer downpours with climate change revealed by weather forecast resolution model, *Nat. Clim. Change*, 4, 570–576, 2014.
- Keynes, J. M.: *Treatise on Probability*, Macmillan & Co., London, 1921.
- Kirby, A. M. and Ash, J. R. V.: *Fluvial freeboard guidance note*, R&D Technical Report W/187, Environment Agency: Bristol, UK, 2000.
- Knight, F. H.: *Uncertainty and Profit*, Houghton-Mifflin Co. (reprinted University of Chicago Press, 1971), 1921.
- Korteling, B., Dessai, S., and Kapelan, Z.: Using information-gap decision theory for water resources planning under severe uncertainty, *Water Resour. Manage.*, 27, 1149–1172, 2013.
- Kousky, C. and Cooke, R. M.: *Climate Change and Risk Management: Micro-correlations, Fat Tails and Tail Dependence, Resources for the Future*, Washington DC, 2009.
- Koutsoyiannis, D.: A probabilistic view of Hershfield's method for estimating probable maximum precipitation, *Water Resour. Res.*, 35, 1313–1322, 1999.
- Kunkel, K. E., Karl, T. R., Easterling, D. R., Redmond, K., Young, J., Yin, X., and Hennon, P.: Probable maximum precipitation and climate change, *Geophys. Res. Lett.*, 40, 1402–1408, <https://doi.org/10.1002/grl.50334>, 2013.
- Kunz, M., Grêt-Regamey, A., and Hurni, L.: Visualization of uncertainty in natural hazards assessments using an interactive cartographic information system, *Nat. Hazards*, 59, 1735–1751, 2011.
- Leedal, D. T., Neal, J., Beven, K., Young, P., and Bates, P.: Visualization approaches for communicating real-time flood forecasting level and inundation information, *J. Flood Risk Manage.*, 3, 140–150, 2010.
- Liu, Y., Freer, J., Beven, K., and Matgen, P.: Towards a limits of acceptability approach to the calibration of hydrological models: Extending observation error, *J. Hydrol.*, 367, 93–103, 2009.
- MacCaffrey, R.: Global frequency of magnitude 9 earthquakes, *Geology*, 36, 263–266, <https://doi.org/10.1130/G24402A.1>, 2008.
- MacEachren, A. M., Robinson, A., Hopper, S., Gardner, S., Murray, R., Gahegan, M., and Hetzler, E.: Visualizing geospatial information uncertainty: What we know and what we need to know, *Cartogr. Geogr. Inf. Sc.*, 32, 139–160, 2005.
- Marzocchi, W., Sandri, L., and Selva, J.: BET_VH: a probabilistic tool for long-term volcanic hazard assessment, *B. Volcano.*, 72, 705–716, 2010.
- Mason, D. M., Horritt, M. S., Dall'Amico, J. T., Scott, T. R., and Bates, P. D.: Improving river flood extent delineation from synthetic aperture radar using airborne laser altimetry, *IEEE Trans. Geosci. Remote Sens.*, 45, 3932–3943, <https://doi.org/10.1109/TGRS.2007.901032>, 2007.
- Mayo, D. G.: *Error and the growth of experimental knowledge*, Chicago: University of Chicago Press, 1996.
- Mayo, D. G. and Spanos, A.: *Error and Inference*, Cambridge: Cambridge University Press, 2010.
- McGuire, R.: Deterministic vs Probabilistic Earthquake Hazards and Risks, *Soil Dynam. Earthq. Eng.*, 21, 377–384, 2001.
- Meredith, E. P., Maraun, D., Semenov, V. A., and Park, W.: Evidence for added value of convection-permitting models for studying changes in extreme precipitation, *J. Geophys. Res.-Atmos.*, 120, 12500–12513, 2015.
- Morales, O., Kurowicka D., and Roelen, A.: Eliciting conditional and unconditional rank correlations from conditional probabilities, *Reliab. Eng. Syst. Safe.*, 93, 699–710, 2008.
- Morgan, R. P.: A predictive model for the assessment of soil erosion risk, *J. Agric. Eng. Res.*, 30, 245–253, 1994.

- Newton, D. W.: Realistic assessment of maximum flood potentials, *J. Hydraul. Eng.*, 109, 905–918, 1983.
- Nott, D. J., Fan, Y., Marshall, L., and Sisson, S. A.: Approximate Bayesian Computation and Bayes' Linear Analysis: Toward High-Dimensional ABC, *J. Comput. Graph. Stat.*, 23, 65–86, 2014.
- O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., and Rakow, T.: *Uncertain Judgements: Eliciting Experts' Probabilities*, Wiley, Chichester, 2006.
- Pang, A.: Visualizing uncertainty in natural hazards, in: *Risk Assessment, Modeling and Decision Support*, edited by: Bostrom, A., French, S. P., and Gottlieb, S. J., 261–294, Springer Berlin Heidelberg, 2008.
- Panza, G. F. M., Kouteva, Vaccari, F., Peresan, A., Cioflan, C. O., Romanelli, F., Paskaleva, I., Radulian, M., Gribovski, K., Herak, M., Zaichenco, A., Marmureanu, G., Varga, P., and Zivcic, M.: Recent Achievements of the Neo-Deterministic Seismic Hazard Assessment in the CEI Region, paper IC2008009, available at: <http://publications.ictp.it> (last access: 15 April 2018), 2008.
- Pappenberger, F. and Beven, K. J.: Ignorance is bliss: 7 reasons not to use uncertainty analysis, *Water Resour. Res.*, 42, W05302, <https://doi.org/10.1029/2005WR004820>, 2006.
- Pappenberger, F., Stephens, E., Thielen, J., Salamon, P., Demeritt, D., J. Andel Schalk, Wetterhall, F., and Alfieri, L.: Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication, *Hydrol. Process.*, 27, 132–146, 2013.
- Pianosi, F., Rougier, J., Hall, J., Stephenson, D. B., Beven, K., and Wagener, T.: Sensitivity Analysis of environmental models: a systematic review with practical workflow, *Environ. Model. Softw.*, 79, 214–232, 2015.
- Potter, K., Rosen, P., and Johnson, C. R.: From quantification to visualization: A taxonomy of uncertainty visualization approaches, in: *Uncertainty Quantification in Scientific Computing*, 226–249, Springer Berlin Heidelberg, 2012.
- Prudhomme, C., Wilby, R. L., Crooks, S., Kay, A. L., and Reynard, N. S.: Scenario-neutral approach to climate change impact studies: application to flood risk, *J. Hydrol.*, 390, 198–209, 2010.
- Rougier, J. and Beven, K. J.: Epistemic uncertainty, in: *Risk and uncertainty assessment for natural hazards*, edited by: Rougier J., Sparks, S., and Hill, L., 40–63, Cambridge University Press: Cambridge, UK, 2013.
- Rougier, J., Sparks, S., and Hill, L.: *Risk and uncertainty assessment for natural hazards*, Cambridge University Press: Cambridge, UK, 2013.
- Ruff, L. J. and Kanamori, H.: Seismicity and the subduction process, *Phys. Earth Planet. Inter.*, 23, 240–252, 1980.
- RWM: Geological Disposal: Methods for Management and Quantification of Uncertainty, Report NDA/RWM/153, available at: <https://rwm.nda.gov.uk/publication/methods-for-management-and-quantification-ofuncertainty/> (last access: 15 April 2018), 2017.
- Saltelli, A.: Sensitivity analysis for importance assessment, *Risk Anal.*, 22, 579–590, 2002.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: *Global Sensitivity Analysis, The Primer*, Wiley, 2008.
- Savage, J. T. S., Pianosi, F., Bates, P., Freer, J., and Wagener, T.: Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model, *Water Resour. Res.*, 52, 9146–9163, <https://doi.org/10.1002/2015WR018198>, 2016.
- Sayers, P. B., Hall, J. W., and Meadowcroft, I. C.: Towards risk-based flood hazard management in the UK, in: *Proceedings of the ICE-Civil Engineering*, 150, 36–42, London: Thomas Telford, 2002.
- Shafer, G.: *A mathematical theory of evidence*, Vol. 1, Princeton: Princeton University Press, 1976.
- Simpson, M., James, R., Hall, J. W., Borgomeo, E., Ives, M. C., Almeida, S., Kingsborough, A., Economou, T., Stephenson, D., and Wagener, T.: Decision analysis for management of natural hazards, *Annu. Rev. Environ. Resour.*, 41, 489–516, <https://doi.org/10.1146/annurev-environ-110615-090011>, 2016.
- Singh, R., Wagener, T., Crane, R., Mann, M. E., and Ning, L.: A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds – Application to a watershed in Pennsylvania, USA, *Water Resour. Res.*, 50, 3409–3427, <https://doi.org/10.1002/2013WR014988>, 2014.
- Smith, P. J., Tawn, J., and Beven, K. J.: Informal Likelihood measures in model assessment: Theoretic Development and Investigation, *Adv. Water Resour.*, 31, 1087–1100, 2008.
- Spiegelhalter, D. J. and Riesch, H.: Don't know, can't know: embracing deeper uncertainties when analysing risks, *Philos. T. Roy. Soc. A*, 369, 4730–4750, 2011.
- Spiegelhalter, D., Pearson, M., and Short, I.: Visualizing uncertainty about the future, *Science*, 333, 1393–1400, 2011.
- Steinschneider, S., McCrary, R., Wi, S., Mulligan, K., Mearns, L. O., and Brown, C.: Expanded Decision-Scaling Framework to Select Robust Long-Term Water-System Plans under Hydroclimatic Uncertainties, *J. Water Resour. Plann. Manage.*, 141, 04015023, 2015.
- Stern, R. J.: Subduction zones, *Rev. Geophys.*, 40, 1–13, 2002.
- Sutherland, W. J., Spiegelhalter, D., and Burgman, M. A.: Twenty tips for interpreting scientific claims, *Nature*, 503, 335–337, 2013.
- Takahashi, Y., Kiureghian, A. D., and Ang, A. H. S.: Life-cycle cost analysis based on a renewal model of earthquake occurrences, *Earthq. Eng. Struct. Dynam.*, 33, 859–880, 2004.
- Takewaki, I. and Ben-Haim, Y.: Info-gap robust design with load and model uncertainties, *J. Sound Vib.*, 288, 551–570, 2005.
- Tang, Y., Reed, P., Wagener, T., and van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, 11, 793–817, <https://doi.org/10.5194/hess-11-793-2007>, 2007.
- Voortman, H. G., Van Gelder, P. H. A. J. M., and Vrijling, J. K.: Risk-based design of large-scale flood defence systems, Delft University of Technology, Faculty of Civil Engineering and Geosciences, 2002.
- Vrugt, J. A. and Sadegh, M.: Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, *Water Resour. Res.*, 49, 4335–4345, 2013.
- Wadge, G. and Aspinall, W. P.: A Review of Volcanic Hazard and Risk Assessments at the Soufrière Hills Volcano, Montserrat from 1997 to 2011, Ch. 24 in: *The Eruption of Soufriere Hills Volcano, Montserrat, from 2000 to 2010: Geological So-*

- ciety Memoirs, Vol. 39, edited by: Wadge, G., Robertson, R. E. A., and Voight, B., Geological Society, London, 439–456, 2014.
- Wakker, P. P.: Prospect theory: For risk and ambiguity, Cambridge University Press, 2010.
- Wilby, R. L. and Dessai, S.: Robust adaptation to climate change, *Weather*, 65, 180–185, 2010.
- Zuccolo, E., Vaccari, F., Peresan, A., and Panza, G. F.: Neo-deterministic and probabilistic seismic hazard assessments: a comparison over the Italian territory, *Pure Appl. Geophys.*, 168, 69–83, 2011.