Epistemic uncertainties and natural hazard risk assessment – Part 1: A review of the issues

K. J. Beven\textsuperscript{1,2}, W. P. Aspinall\textsuperscript{4}, P. D. Bates\textsuperscript{5}, E. Borgomeo\textsuperscript{7}, K. Goda\textsuperscript{3}, J. W. Hall\textsuperscript{7}, T. Page\textsuperscript{1}, J. C. Phillips\textsuperscript{4}, J. T. Rougier\textsuperscript{6}, M. Simpson\textsuperscript{7}, D. B. Stephenson\textsuperscript{8}, P. J. Smith\textsuperscript{1}, T. Wagener\textsuperscript{3,9}, and M. Watson\textsuperscript{4}

\textsuperscript{1}Lancaster Environment Centre, Lancaster University, Lancaster, UK
\textsuperscript{2}Department of Earth Sciences, Uppsala University, Uppsala, Sweden
\textsuperscript{3}Department of Civil Engineering, Bristol University, Bristol, UK
\textsuperscript{4}School of Earth Sciences, Bristol University, Bristol, UK
\textsuperscript{5}School of Geographical Sciences, Bristol University, Bristol, UK
\textsuperscript{6}School of Geographical Sciences, Bristol University, Bristol, UK
\textsuperscript{7}School of Mathematics, Bristol University, Bristol, UK
\textsuperscript{8}Environmental Change Institute, Oxford University, Oxford, UK
\textsuperscript{9}Department of Mathematics and Computer Science, Exeter University, Exeter, UK
\textsuperscript{9}Cabot Institute, University of Bristol, Bristol, UK

This discussion paper is/has been under review for the journal Natural Hazards and Earth System Sciences (NHESS). Please refer to the corresponding final paper in NHESS if available.
Abstract

Uncertainties in natural hazard risk assessment are generally dominated by the sources arising from lack of knowledge or understanding of the processes involved. There is a lack of knowledge 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 that are made for risk management, so it is important to communicate the meaning of an uncertainty estimate and to provide an audit trail of the assumptions on which it is based. Some suggestions for good practice in doing so are made.

1 Introduction

With the increasing appreciation of the limitations of traditional deterministic modelling approaches, uncertainty estimation has become an increasingly important part of natural hazards assessment and management. In part, this is a natural extension of the evaluation of frequencies of hazard in assessing risk, in part an honest recognition of the limitations of any risk analysis, and in part because of the recognition that most natural hazards are not stationary in their frequencies of occurrence. Non-stationarity might result as a consequence of the intrinsic stochastic evolution of natural systems; a volcano can exhibit multiple types of eruption activity, occurrence of a debris flow might depend on a very local rainfall event. It might also result from climate change and other anthropogenically induced changes (Rougier et al., 2013; Hirsch and Archfield, 2015). Figure 1 shows some statistics on the publication of papers concerned with uncertainty assessment for different types of natural hazards. While these show an in-
crease over the past 15 years, some hazard types return zero results, and a search on “epistemic uncertainty” AND “natural hazards” returned zero results on Web of Science.

There is a growing practice of recognizing different types of uncertainty in risk assessments (Hoffman and Hammonds, 1994; Helton and Burmaster, 1996; Walker et al., 2003; van der Sluijs et al., 2005; Refsgaard et al., 2006, 2007, 2013; Beven, 2009, 2012, 2013; Warmink et al., 2010; Rougier and Beven, 2013, 2014; Beven and Young, 2013). In particular, since the time of Keynes (1921) and Knight (1921) it has been common practice to distinguish between those uncertainties that might be represented as random chance, and those which arise from a lack of knowledge about the nature of the phenomenon being considered. Knight (1921) referred to the latter as the “real uncertainties” and they are now sometimes called “Knightian uncertainties”. While Knight’s thinking pre-dated modern concepts and developments in probability theory (e.g. de Finetti, 1937, and others), the distinction between probabilistic and knowledge uncertainties holds.

In fact, an argument can be made that all sources of uncertainty can be considered as a result of not having enough knowledge about the particular hazard occurrence being considered: it is just that some types of uncertainty are more acceptably represented in terms of probabilities than others. In current parlance, there are the “aleatory uncertainties”¹ while the Knightian real uncertainties are the “epistemic uncertainties”². Aleatory uncertainties represent variability, imprecision and randomness, or factors that can be modelled as random for practical expediency, which can be represented as forms of noise within a statistical framework. Within epistemic uncertainties it is possible to subsume many other uncertainty concepts such as ambiguity, reliability, vagueness, fuzziness, greyness, inconsistency and surprise that are not easily represented as probabilities. This distinction is important because most methods of decision making used in risk assessments are based on the concept of risk as the product of

¹From the Latin “alea”, meaning a die or game of dice.
²From the Greek “ἐπιστήμη”, for knowledge or science.
a probability of occurrence of an event and an evaluation of the consequences of that event. If there are important uncertainties in the assessment of the occurrence that are not easily assessed as probabilities, or if there are significant epistemic uncertainties about the consequences, then some other means of assessing risk decisions might be needed. Given lack of knowledge, there is also plenty of opportunity for being wrong about the assumptions used to describe sources of uncertainty, or having different belief systems about the representations of uncertainties, which is sometimes referred to as ontological uncertainty\(^3\) (e.g. Marzocchi and Jordan, 2014; Beven, 2015). Epistemic and ontological uncertainties are also sometimes referred to as “deep uncertainties”, including in risk analysis and natural hazards (e.g. Cox, 2012; Stein and Stein, 2013).

For example, the results of a natural hazards assessment will often depend on the outputs of a model or simulator. This may be a stochastic simulator of frequencies, or a deterministic simulator of the footprint of impact of the hazard. Simulation models are always approximations of the complex real system but most environmental modellers are pragmatic realists in their approach to modelling (Beven, 2002). Thus, they intend the variables in a simulation to represent some real world quantities while having the pragmatic understanding that the simulator will necessarily be subject to simplifying assumptions and approximations.

Those simplifying assumptions imply that the simulator outputs will be, to a greater or lesser extent, uncertain representations of the real world system when compared to observations in model evaluation. The resulting residuals will be subject to both aleatory, epistemic and ontological uncertainties: in the process representations; in the parameters; in the boundary conditions; and in the historical data that might be used to calibrate or validate the outputs from a simulator (e.g. Beven, 2009). The structural error of a simulator might not be aleatory in character, but a form of bias with non-stationary characteristics that might be difficult to represent by a simple model discrepancy function (as suggested by Kennedy and O’Hagan, 2001) or reification structure.

\(^3\)From the Greek “\(\delta\nu\gamma\epsilon\)” meaning “being or that which is”, the present participle of the verb “\(\epsilon\nu\mu\)”, to be.
(see Goldstein and Rougier, 2003). If the different sources of epistemic uncertainty that underlie a hazard assessment are oversimplified then it might result in quite incorrect inferences. An important class of such uncertainties is where measurement accuracies and biases in the historical record have changed over time in poorly defined ways. Sometimes, knowing what measurement technologies were used might allow some estimate of the changing uncertainty to be made. In other cases, such as the changing in water level to discharge rating curves after major events, the changes might be quite arbitrary and difficult to characterize (e.g. Westerberg et al., 2011; McMillan and Westerberg, 2015). Similarly in the instrumental records of earthquakes, completeness and detection sensitivity can be subject to abrupt temporal changes in seismograph network coverage or configuration (Ogata and Katsura, 1993; Utsu, 2002; Kagan, 2003; Woessner and Wiemer, 2005) while the reliability of temperature records in assessing climate change has been a particular source of controversy (e.g. Jones et al., 2011).

And yet, even where there is a significant historical database of events, then the estimates of magnitudes for low probability events will still be highly uncertain. In the case of floods, for example, where there may be decades of data available, and sometimes over 100 years, stochastic experiments with different frequency distributions suggest that to get a good estimate with low uncertainty of an event of a given frequency requires observations for a period of at least ten times the return period for that frequency (i.e. 1000 years of record for an annual exceedance probability of 0.01). Decades of data are not then sufficient and it should be good practice to assess the associated uncertainty.

Such an analysis is, of course, assuming that the occurrence of extreme events is stationary in time, something that has been queried for floods (e.g. Koutsoyiannis, 2003, 2013; Wilby et al., 2008), for rainfall episodes (e.g. Marani and Zanetti, 2015), for earthquakes (Hakimhashemi and Grünthal, 2012), for eruptions (Mendoza-Rosas and De la Cruz-Reyna, 2008; Deligne et al., 2010), for storms (Mailier et al., 2006; Vitolo et al., 2009, and more generally in the context of risk (Serenaldi, 2015). The potential for non-stationary statistics due to grouping of events through some more...
complex stochastic simulator, or due to environmental change is then an additional source of epistemic uncertainty in the risk analysis (Mumby et al., 2011; Koutsoyiannis and Montanari, 2012; Beven, 2015).

Both Keynes and Knight were economists working in situations where human activity forms an essential part of the system (and a decade before Kolmogorov’s axioms formalised the concept of probability). The assessment of natural hazard risk is similar, both through anthropogenic influences on the occurrence and footprint of the hazard and but also through the impact on the potential consequences. Financial or economic assessments of consequence are also subject to epistemic uncertainties resulting from the limited availability of suitable data, due in part to the scarcity of event records and uniqueness of place and process. Moreover risk analysis is often motivated by the desire to make decisions regarding mitigation. This requires the assessment of consequences in a system subject to unknown future change; and where, for example, the mitigation measures will influence the future development of the system (for example building defences tends to increase development and the value of what is at risk, Di Baldasarre et al., 2013; Viglione et al., 2014). Under such feedback along with other societal changes any consequences may be considered indeterminate and hence introduce significant epistemic uncertainty into any risk assessment.

2 Dealing with epistemic uncertainties

Epistemic uncertainties arise in all parts of a risk assessment. We are limited in our knowledge of how best to represent processes in complex domains, even if we think we have a good basic understanding of the physics involved. We are limited in our knowledge of how to specify the parameters of those process representations in applications to particular locations or over specific time frames. We are limited in our knowledge of the forcing boundary conditions. We may not always properly understand how measured variables relate to variables in a simulator (the commensurability issue, e.g. Beven, 2006, 2012). There may also be issues that limit simulator accuracy but
which have not yet been recognized (the “unknown unknowns” in the famous phrase of Donald Rumsfeld, but which have been recognised since at least the times of Plato). Those knowledge limitations might result in complete surprises with significant impact (the “black swans” of Nassim Taleb). In his recent book on Anti-fragility, Taleb (2012) argues that society should recognize the potential for such surprise events in managing risk, particularly in the event that the consequences might be catastrophic (as in the 2008 financial collapse). Recent natural hazard examples might include the levée failures associated with Hurricane Katrina or the siting of the Fukushima nuclear power plants along a coast where past destructive tsunamis had occurred.

But as noted earlier, in any formal assessment of risk from a natural hazard it is generally a requirement to specify the probability of occurrence of an event. If the various types of epistemic uncertainties are the result of a lack of knowledge, then it will be hard to quantify one’s epistemic uncertainty, harder still to represent it in the form of a probability distribution over possible outcomes. In general, any quantification will depend on the judgements of experts in a particular natural hazard area but even experts find it difficult to estimate probabilities for sources of epistemic uncertainty with any degree of confidence (the probabilities might be indeterminate, Levi, 2000; Hajek and Smithson, 2012). So what then to do?

Some statisticians have argued that the formal axiomatic framework of probability is the only way to consider representing uncertainty (e.g. O’Hagan and Oakley, 2004), even if those probabilities will only be conditional on current knowledge. Those probabilities might be informed by taking expert advice on what range of potential surprises might be possible with a view to providing at least an estimate of the range of probabilities for potential outcomes (Cooke, 1991; O’Hagan et al., 2006; Aspinall, 2010; Aspinall and Cooke, 2013). Rather than a probability distribution, this might result in a representation as imprecise probabilities or a p box (e.g. Walley, 1991, 2000; Levi, 2000; Rubio et al., 2004; Ferson et al., 2007; Hall et al., 2007; Karanki et al., 2009). Methodologies for expert elicitation are now well established, but it is known that experts tend to underestimate the potential uncertainties arising from lack of knowledge.
and not always to consider sources of uncertainty outside their range of direct experience. It has been suggested, for example, that this has led to an underestimate of the potential uncertainties associated with climate change predictions (e.g. Cooke, 2014).

A second strategy is to assume that all epistemic uncertainties can be treated within a statistical framework as if they were aleatory in nature. Some of the foundations for modern Bayesian statistics lie in treating uncertainties in this way for decision making (Cox, 1946; Savage, 1954; Lindley, 1971; de Finetti, 1974). When a simulator is involved in the analysis, in general this involves adding uncertainty to the outputs of the simulator, conditional on an assumption that the simulator is correct. The error or discrepancy is treated as an additional (often additive) stochastic component that might have components of bias, heteroscedasticity and covariation of uncertainties in time and space, including the identification of functions to correct for the limitations of the simulator (Kennedy and O’Hagan, 2001; Goldstein and Rougier, 2009). Clearly the identification of the stochastic model for the errors is easiest where there is a historical database against which the simulator predictions can be compared (see Hall et al., 2011, for a flood inundation example). An example application is the stochastic downscaling with bias corrections that is used to estimate changes in precipitation statistics for evaluating the effects of climate change on extreme events in future decades. The climate simulators (even using dynamic downscaling to finer grid scales) may not reproduce the historical rainfall statistics, in part for reasons of epistemic uncertainty including the limitations of current climate simulators. Bias correction and stochastic downscaling allows the historical statistics to be matched more closely, with some uncertainty (e.g. Ning et al., 2012; Chen et al., 2013; Addor and Seibert, 2014; Rufault et al., 2014). An additional epistemic uncertainty is then introduced, however, in the necessary (but very doubtful) assumption that both the simulator form and the corrections that apply for the historical period will also hold in future (Hall, 2007; Ho et al., 2012; Tye et al., 2014), or that some reasoning for how the uncertainties might change in future can be defined (Rougier, 2007). An extension of that approach is to examine a number of different plausible scenarios for the assumptions of an analy-
sis and examine the conditional risk exceedance probabilities across all the scenarios (Rougier and Beven, 2013).

There are also other dangers with this approach. By assuming that (after bias correction and other, generally simple, transformations) all sources of uncertainties can be treated as aleatory with known distribution, this method will generally overestimate the information content of the historical data (see, for example, Beven, 2012; Beven and Smith, 2014). Use of simple aleatory error based likelihoods or probabilities does not allow enough potential for surprise from arbitrary rather than aleatory future occurrences (Beven, 2015). Arbitrary variation suggests that, at least for some restricted period of time we are interested in, there may be no clear, stationary, distribution of occurrences. This is particularly the case where an extreme or catastrophic impact might be the result of a particular combination of events that has not been seen before, such as the multiple ruptures in the Tōhoku earthquake that produced the Fukushima tsunami; this event, and the potential for similar rare, extreme compound tectonic failures in certain subduction zones has been termed an “earthquake supercycle” by Herrendörfer et al. (2015).

A further way of allowing for epistemic uncertainties is to accept that it may not be possible to estimate the probabilities of future events with any degree of certainty and find some other way of protecting society against future events. Expert elicitation might reveal no consensus on representing potential future outcomes as probabilities or possibilities. This means that the methods of risk-based decision theory cannot be used. Instead, it will be necessary to act in a way that is precautionary or robust to unknown future occurrences. This might be based on a sensitivity analysis of vulnerability to future extremes (such as that for future floods in Prudhomme et al., 2010), or the type of Info-Gap methodology for decision making proposed by Ben-Haim (2006, see also for example, Hine and Hall, 2011, in a floods context). The Info-Gap approach also allows the “opportuneness” benefits of the non-arrival of a damaging event to be taken into account, but is dependent on assuming that the outputs from the simulator being used are an adequate representation of the real system. While past performance remains
the best indicator of simulator adequacy, we should still be wary of inferring that this
will apply to future risk (e.g. the dependence of simulator success on input realisation
demonstrated by Blazkova and Beven, 2009; and the use of different ensembles of
behavioural models for different flood risk zones in the same flow domain in Pappenberger
et al., 2007).

There may also be good legal reasons to be precautionary if there is some concern
with expecting the unexpected resulting from epistemic uncertainties. Western soci-
eties increasingly seek to place blame following natural hazard disasters. Thus being
explicit about the expected uncertainties is a means of protection against blame, at
least provided that the assumptions that are used in representing the sources of un-
certainty and the way that they are conditioned by data can be justified.

The most famous recent case of this type is the legal case resulting from the fa-
tal L’Aquila earthquake in Italy in 2009. In 2012, seven persons associated with or
members of the National Commission for the Forecast and Prevention of Major Risks
were each convicted of multiple manslaughter in relation to some of the deaths in the
earthquake, having been charged with failing in their public duty to ensure a natural
disaster was avoided. While the court understood the infeasibility of predicting a major
earthquake, it was argued that a calming statement issued by one of the seven, a civil
servant, seemingly acting as a spokesperson for the Commission, had led some peo-
ple to decide to return to their homes even while the seismic unrest continued, when
otherwise they might have been more precautionary. In 2014, the six scientists, but not
the civil servant who chaired the committee, were acquitted on appeal, but may yet
face further prosecution.

3 Epistemic uncertainty and eliciting expert judgement

Expert judgement remains, however, one of the main ways of trying to take account of
epistemic uncertainties in natural hazards risk assessment. Experts can be asked to
assign estimates of the probabilities or possibilities of potential outcomes. Such esti-
mates will be necessarily judgement-based only and may be incomplete. When epistemic uncertainties dominate in natural hazards assessment it is evident that some individual experts will sometimes be surprised by events or outcomes, and in some case the judgements of some experts may be quite inaccurate or uninformative. In such situations it is therefore important to include the judgements of as many experts as possible which raises questions about the independence of experts (who may have similar backgrounds and training, even if different levels of experience) and about whether or how more weight should be given to the judgements of some experts relative to others.

Cooke (2014) gives a good recent summary of some of the issues involved in expert elicitation (see the extensive supplement to that paper). He points out that both Knight (1921) and Keynes (1921) suggested that the use of elicited expert probabilities might be a working practical solution to 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., 2004; Aspinall and Cooke, 2013). Cooke (2014) includes a review of post-elicitation analyses that have been carried out seeking to validate assessments conducted with the Classical Model Structured Expert Judgment (SEJ) (Cooke, 1991). This appraisal of applications in a variety of fields 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 study involving 72 experts distributed over 134 expert panels, with each panel assessing between 10 and 15 calibration variables concerned with foodborne health hazards (source attribution, pathways and health risks of foodborne 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 operational limitations in the WHO global elicitation process. However, based on these statistical ac-
accuracy and informativeness measures on the calibration variables, performance-based weighted combinations were formed for each panel. 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 performance weighting still out-performed equal weighting (“Equal Weights Decision Maker” EW DM) (Cooke et al., 2015).

Because a large number of experts assessed similar variables it was possible to compare statistical accuracy and informativeness on a larger dataset than hitherto (Cooke et al., 2015). For certain foodborne 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 out-of-sample analysis was infeasible.

This extensive study has provided new perspectives on the efficacy of SEJ (Cooke et al., 2015). 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. These findings should motivate the development and deployment of enhanced elicitor 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).

4 Epistemic uncertainty, the “maximum event” and factors of safety

In several fields of natural hazards assessment there have been practical approaches suggested based on estimating the maximum event to be expected in any location of interest. Such a maximum might be a good approximation to very rare events, es-
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, Paper 2). In evaluating seismic safety of critical infrastructures (e.g. nuclear power plants and dams), there has been a recent movement away from probabilistic assessment of earthquake magnitudes to the concept of a deterministic maximum estimate of magnitude (McGuire, 2001; Panza et al., 2008; Zucollo 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 criticised because of the assumptions that are made in such analyses (e.g. Koutsoyiannis, 1999; Abbs, 1999; Bommer, 2002).

Sensitivity to such assumptions is rarely investigated and uncertainties in such assessments are generally ignored. A retrospective evaluation of the anticipated very large earthquake in Tōhoku, Japan, indicates that the uncertainty of the maximum magnitude in subduction zones is considerable and in particular, the upper limit should be considered unbounded (Kagan and Jackson, 2013). This conclusion, however, 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; Fig. 5b), and indicated a modest maximum magnitude of 8.2 Mw for the NE Japan subduction zone. It is not surprising that these authoritative scientific sources were trusted for engineering risk decisions.
Moreover there was no associated uncertainty analysis for the Ruff and Kanamori relationship and later, but before the Tōhoku earthquake, MacCaffrey (2008) pointed out 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 magnitude 9 Mw. Thus, epistemic uncertainties for the 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.

One defence that can be offered this type of analysis is that, with appropriate rules based on expert judgement, it can provide a formalized way of elaborating science-informed planning for dealing with potentially catastrophic natural hazards. In this, it can be considered to be similar to the institutionalised annual exceedance probabilities that are used for planning in different countries. Thus, in the UK, frequency-based flood magnitude estimates are used for flood defence design and planning purposes. 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 institutionalised, and, in some cases, statutory, levels of protection are then a political compromise between costs and perceived benefits. The flood defence example is one where the analysis can be extended to a full risk-based decision analysis, where costs and benefits can be 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.
Deterministic maximum event approaches are not associated with a probability, but can serve a similar, risk averse, institutionalised 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 the epistemic uncertainties, protection levels will be exceeded sufficiently rarely for the risk to be acceptable.

Another risk averse strategy to 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 simulators 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 such an approach can also lead to overconfidence in cases where the consequences of failure might be high impact. It such cases it will be instructive to make a more detailed analysis of plausible future events and their consequences in managing the risk.

5 Epistemic uncertainty and disinformative data

The assessment of the potential for future natural hazard events frequently involves the combination of outputs from a simulator with data from historical events. Often, only the simplest form of frequency analysis is used where the simulator is a chosen distribution function for the type of event being considered, and where the data are taken directly from the historical record. Both simulator and data will be subject to forms of epistemic uncertainty to the extent that either might be “disinformative” in assessing the future
hazard. A frequency distribution that underestimates the heaviness\(^4\) 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 simulator 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 the catchment for many events. No simulator that maintains mass balance (as is the case with most rainfall–runoff simulators) will be able to predict more output than input, so that including those events in conditioning the simulator would lead to incorrect inference. 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 identification of inundated areas by either post-event surveys or remote sensing (e.g. Mason et al., 2007), and the effects of extreme conditions such as predicting the evolution of tropical cyclones (e.g. Hamill et al., 2011).

6 Epistemic uncertainty as future scenarios

What else can we do to consider potential epistemic uncertainties given that we lack the ability to describe them adequately or reliably with some probability distribution? One strategy that can in principle always be applied is sensitivity analysis (Saltelli, 2008).

\(^4\)A heavy or fat tail is a property of probability distributions exhibiting extremely large kurtosis particularly relative to the ubiquitous normal, or lognormal, distributions which are examples of thin tail distributions. The term “fat tail” is a reference to the tendency of a distribution to have more observations in the tails than normal or lognormal distributions.
2002; Tang et al., 2007; Saltelli et al., 2008; Pianosi et al., 2015) or scenario discovery (van Notten et al., 2005; Bryant and Lempert, 2010). How much does it matter if we make different assumptions, if we change boundary conditions, if we include the potential for data to be wrong etc.? While most Sensitivity Analysis approaches assume that we can define some probability distribution to characterize the potential variability of the inputs into our simulators, we might still gain useful information regarding whether uncertainty of an input might even matter from such an analysis. There are also formal ways to test the impact of discrete choice, such as the resolution level of our simulators (Baroni and Tarantola, 2014). We can, of course, make a further assumption and give different projections (from different Global Circulation Models, or downscaling procedures, or ways of implementing future change factors) equal probability but this is a case where the range of possibilities considered may not be complete. 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 who are independent of the various modelling groups. It might be better to consider these projections as only one ensemble of future possibilities within a wider sensitivity analysis or scenario discovery approach (e.g. Bryant and Lempert, 2010; Prudhomme et al., 2010; Singh et al., 2014) while remembering that climate change might not be the only factor affecting change in future hazard (Wilby and Dessai, 2010).

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 (Kousky and Cooke, 2009). These are distinct aspects of loss distributions which challenge traditional approaches to managing risk. Micro-correlations are negligible correlations which may be individually harmless, but very dangerous when they coincide and operate in concert. Fat tails, noted above, 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, 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 (Morales Napoles et al., 2008). It is hard to believe that the current climate models are not subject to these types of epistemic errors. In such circumstances, 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 elicitation methods can be substantial. When performed rigorously, expert elicitation and pooling of experts’ opinions can be powerful means for obtaining rational estimates of uncertainty.

7 Epistemic uncertainty and fitness-for-purpose

One of the implications of epistemic error is that the simulators used in natural hazards assessment might not just be uncertain, but may not be fit-for-purpose in making the predictions or projections that might feed into decision making processes. Since most decisions are concerned with future occurrences, this might not be only because of the failings of the simulator itself, but also because of assumptions about the nature of future boundary conditions (e.g. the post-audit analyses of Konikow and Bredehoeft, 1992). There is some suggestion that this could be the case for the current generation of climate simulators (e.g. Suckling and Smith, 2013), even for temperature projections on a decadal scale. For precipitation extremes, as required for the assessment of future floods and droughts, the situation is likely to be less satisfactory, even though there have been studies published that have attempted to attribute part of the severity of past flood events to anthropogenic effects (e.g. Pall et al., 2011, but see also Hulme et al., 2011).

Uncertainty assessments are designed to compensate for the fact that we expect such simulators to be approximations to the complexity of the real world system. In climate modelling, this is intrinsic to the modelling process in that there is no expectation that simulator variables at the grid scale will be commensurate with local historical
observations. Thus a downscaling process is necessary, either by nesting a finer grid simulator within the global simulator, or by a stochastic correction simulator. For the latter, this will involve using the historical observational record not only to represent local uncertainty but also to correct for any bias between predicted and observed values. Those bias corrections will then be carried over to future projections, even though the dynamics are predicted as changing in future. In the UK Climate Projections 2009 project (UKCP09), downscaling has been implemented for every 5 km grid square in the country, allowing a stochastic weather generator to produce future realisations of temperatures and precipitation. UKCP09 has been used, for example, in assessing future floods in the UK (Kay and Jones, 2012) and similar studies have been carried out in many other countries.

The question is how far, given the assumptions and epistemic uncertainties involved in producing such projections, we should recognise the possibility that they might not be fit for purpose. Similar issues arise in other areas of natural hazards, particularly where understanding and knowledge are still being gained about how systems work because of difficulties in observation, the rarity of occurrences, or the uniqueness of local circumstances (e.g. the near field processes in the formation of ash clouds; the initiation and representation of flow processes in lahars and debris flow; recurrence of earthquakes; the variety of ground movement simulators used in assessing the impacts of earthquakes; the role of roots in shallow landslides and debris flows). In such cases we should be wary of putting too much faith in any probabilistic assessments of potential outcomes, and recognise the potential for future surprise.

Another analogous example is in performing site-specific seismic hazard assessments for safety-critical facilities, especially in areas of low- to moderate seismicity (Aspinall, 2013). Whilst there may be just about sufficient historical earthquake data to characterize magnitude-recurrence activity rates at a national or regional scale, sparseness or total absence of data at the local scale can significantly inflate epistemic uncertainty in critical source zones within a conventional probabilistic hazard assessment.
model. In such circumstances, recourse may be made to Bayesian data updating tech-
niques, but these are not (yet) widely applied in that domain.

With the vintage of many site-specific seismic hazard assessments in the UK nuclear
industry becoming decades old, the same notion (Bayesian updating) offers one way
of re-assessing the original studies in the light of newer data, new theories and new
empirical evidence from elsewhere (Woo and Aspinall, 2015). At the very least, this
approach could serve to reduce “unfitness-for-purpose” of such assessments.

8 What should constitute good practice in dealing with epistemic
uncertainties?

The above discussion reveals that, as might be expected, there are no generally agreed
methods of dealing explicitly with arbitrary epistemic uncertainties (for good epistemic
reasons of course). It is worth remembering that, in representing lack of knowledge
or understanding, epistemic uncertainties are inherently reducible by the collection
of additional data and improved process representations. The first consideration in any
assessment should, therefore, a focus on available data and its interpretation. There
are clearly limitations to this, in particular for rare, high impact, natural hazard events
when the time scale of the observational record is limited with respect to the occur-
rences of extreme events, while longer historical records are often associated with
their own epistemic uncertainties that are not now easily reducible from the present.

So we are consequently often in the situation of making the best of the information,
often ambiguous, that we have about the nature of epistemic uncertainties. There are
approaches based on making some attempt to assign probability distributions to them,
including imprecise probabilities, p boxes or fuzzy possibility measures and the use
of expert elicitation. There are approaches based on sensitivity analysis or scenarios
based on different assumptions. There are approaches to being robust to future out-
comes or being precautionary in planning for the future.
The outputs from any such analysis will of course then depend on the assumption that are made, and where some historical data are available to condition those estimates, how the probabilities or possibilities are modified on the basis of the data. This is clearly a situation which can be put into the structure of a logic tree (e.g. Newhall and Hoblitt, 2002; Bommer et al., 2005; Bommer and Scherbaum, 2008; Marzocchi et al., 2012; Delavaud et al., 2012) or Bayesian belief network; (Dlamini, 2010; Chen and Pollino, 2012; Aspinall and Woo, 2014). Both have been used widely in natural hazards assessment (e.g. for seismic hazard assessments in both the US and Europe\(^5\)).

However, as already noted there are dangers in applying Bayesian statistical theory, particularly in using a simple error model and associated likelihood function to represent epistemic uncertainties (see the discussion in Beven, 2009, 2012 and Rougier and Beven, 2013). In addition, potential events (at least those that can be thought of) need to be assigned prior probabilities, but these will need to be based on expert judgement, which might be prejudiced against the unexpected, especially where there is a vested community interested in a particular modelling framework. And, of course, there is no guarantee that the estimated probabilities should be considered complete (we may not think of all the potential possibilities that might occur)\(^6\). In these situations there are real issues about communicating the meaning of an uncertainty assessment to potential users (e.g. Faulkner et al., 2007, 2013).

Thus 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. Beven and Alcock (2012) suggest that this might be expressed in the form of condition trees. The condition tree for any particular application 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.

---


\(^6\)More precisely, it is not possible to construct a probability space if the set of all envisaged events do not consist of all possible events in the Borel space.
approach has two rather nice features. Firstly, it provides a framework for the discussion and agreement of assumptions with 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 simulator 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 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 (2016) give examples of the application of this methodology to the flood risk mapping problem. The (seismic) hazard assessment Bayesian updating concept, mentioned above, could fulfil a similar role.

9 Visualising uncertain outcomes

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 visualisation 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; Friede-
One of the issues that arises in visualisation is the uncertainty induced by the visualisation method itself, particularly where interpolation of point predictions might be required in space and/or time (e.g. Aguyama and Hunter, 2002; Couclelis, 2003). The interpolation method will affect the visualisation in epistemic 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 visualisations too good so as not to induce an undue belief in the simulator predictions and assessments of uncertainty in the potential user (e.g. Faulkner et al., 2014; Dottori et al., 2013).

Some examples of visualisations 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); 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, a simulator or cascade of simulators to produce uncertain outputs. The presentation of 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 visualisations (Spiegelhalter et al., 2011).

10 Epistemic uncertainty and decision making

A 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; Hall, 2013; Rougier and Beven, 2008).
For many decisions a complete, thoughtful, uncertainty assessment of risk might not be justified by the cost in time and effort. Any simplified assessment can then be recorded in the condition tree for such an analysis as part of good practice. 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 will be possible, including using expert elicitations in defining the assumptions of the relevant condition tree.

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. The difficulty of specifying probabilities for epistemic uncertainties means that any resulting decisions will necessarily be conditional on the assumptions. Hence the importance of having a well-defined condition tree and audit trail as part of good practice, in both agreeing assumptions and communicating the meaning to uncertainties to decision makers and in setting the framework for any expert elicitation process. Some of these issues are discussed by Pappenberger and Beven (2006); Sutherland et al. (2013) and Juston et al., 2013).

It also leaves scope, however, for other methodologies, 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; 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 just a few applications of these methods in the area of natural hazards, for example: Info-Gap theory to flood defence assessments (Hine and Hall, 2011); 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 can be usefully combined with expert elicitation.

11 Epistemic uncertainty and doing science

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 some empirical control on the 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 (see Beven and Smith, 2014; Beven, 2015). 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 Generalised Likelihood Uncertainty Estimation
(GLUE) methodology (Beven and Binley, 1992, 2013; Smith et al., 2008). It is still possible to empirically control the performance of any such methodology in simulation of past data, but, given the epistemic nature of uncertainties this is no guarantee of performance in future projections. In particular if we have determined that some events might be disinformative for model calibration purposes, we will not know if the next event would be classified as informative or disinformative if the observed data were available (Beven and Smith, 2014).

In this context, it is interesting to consider what would constitute a severe test (in the sense of Mayo, 1996) 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 certainly likely 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 Paper 2). 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 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 normalised 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. In modelling river discharges, however, it has been found that the remaining 5% might be associ-
ated 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). It is one way of improving the science in a situation 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 of 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 gives greater allowance for our lack of knowledge.

12 Summary and conclusions

In assessing future risk due to natural hazards it is generally necessary to resort to the use of a simulator (or 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 to the knowledge of what distribution should be assumed, especially when the past database is sparse. For risk-based decision-making the consequences of events must also be modelled in some way and are equally subject to uncertainties due to limited knowledge. Even if our simulations can be shown to match past data, they may not perform so well 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 recognise 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 muddled through.
But, this rather underplays the catastrophic consequences of some poor decisions (including some of the recent examples noted earlier), so there is surely scope for better practice in future in trying to allow for all models being wrong, all uncertainties being epistemic (even if some might be treated as if aleatory), all uncertainty estimates being conditional, all expert elicitations tending to underestimate potential uncertainties, and the potential for future surprise. Because epistemic uncertainty also implies the possibility of future surprise, and while this might then instigate a revision of the associated risk, it also suggests that we should be prepare for surprises and should be wary of both observational data and simulators that might be the best we have available but which might be disinformative and not necessarily fit for purpose. Surprises will occur when the probability estimates are incomplete, where 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.

This then raises issues about the meaning of uncertainty estimates and how they might be interpreted by potential users, stakeholders and decision makers (e.g. Sutherland et al., 2013). Visualisations can be helpful in conveying the nature of uncertainty outputs but the deeper epistemic uncertainties might not be amenable to visualisations. How to deal with epistemic uncertainties in all areas of natural hazards requires further research in trying to define good practice, particularly in assessing the models that are used in natural hazard risk assessments in a scientifically rigorous way. The type of condition tree, audit trail and model hypothesis testing suggested here represent one step in that direction. Part 2 of this paper discusses applications of these concepts to different natural hazard areas in more detail.

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 for comments on an earlier draft of the paper.
References


Figure 1. Numbers of papers published since the year 2000 according to a search on Web of Science using “uncertainty estimation” together with various natural hazard descriptors. Note that several return zero results.