My name is Mike Poole and I am a Senior Research Manager at Radioactive Waste Management Limited (RWM) and I was asked by the editor if I would provide high-level comments on this Natural Hazards and Earth System Sciences journal paper (from here on 'The Paper').

My experience relevant to the review is twofold:

a) The requirement for RWM to calculate potential risks to future populations from buried radioactive waste over geological timescales requires quantification and treatment of many disparate uncertainties, and is in many ways analogous to the hazard risk assessments considered in The Paper.

b) I have recently written a report explaining in detail RWM's strategy for handling uncertainty [1], published earlier this year, so it is interesting to compare notes.

The Paper covers the issues that I would expect to see covered, from the need to build a model of the system, including consequences, and carry out a probabilistic analysis of some sort. The need to rely fairly extensively on expert judgement and the importance of effective elicitation methods is also clear. The Paper also recognises the challenges in communicating issues related to risk, assessment of risk, and uncertainty – although perhaps the section recognising the importance of good visualisation methods, could use some, ahem, pictures!

I would like to focus my comments on four aspects of The Paper in particular, which I will state up front because they interrelate to some extent:

1. The extent to which it is sensible to distinguish epistemic and aleatory uncertainty.
2. Probabilistic vs non-probabilistic methods for quantifying uncertainty.
3. The role of expert elicitation and methodology used.
4. The extent to which it is necessary to make assumptions.

1. Epistemic and aleatory uncertainties.

At the point of developing a model for assessing risks, for practical purposes, I think the distinction between aleatory uncertainty and epistemic uncertainty is moot. Such a model needs to account for all sources of uncertainty on the same footing, irrespective of type. When considering the results from a model, e.g. which parameters the performance measure is sensitive to, I accept that it may then become important which uncertainties are epistemic, and therefore potentially reducible.

The Paper says that epistemic uncertainties can be inherently difficult to represent as probabilities – I would say all uncertainties can be represented as probability distributions, the difficulty is more that people struggle to do this (hence the need for elicitation methods) and that on some occasions the probability distribution (if unbiased) is so wide that it is describing virtually no knowledge, so is therefore unhelpful.

2. Probabilistic and non-probabilistic methods

I found S2.5 of The Paper rather difficult to follow. It introduces the spectre of non-probabilistic methods for quantifying uncertainty, and talks rather too much about assumptions (see 4 below). I think a paper on good practice could offer more on the relative merits of different methods, but I detect some fence-sitting in this discussion as it unfolds. I am not an expert in non-probabilistic methods, but have a deep mistrust of them for the following reason.

I presume the reason they were devised is largely because people have great difficulty estimating probabilities without bias. But probability is mathematically the correct way to represent uncertainty. If any alternative model is used, then this must add additional significant model uncertainty to the overall system. It seems to me therefore that for such an alternative model to prove useful, people would need to be better at estimating the

required quantities for that model (whatever they are), than they are at estimating probabilities, by more than this additional model uncertainty. I am sceptical that this is ever true, therefore I am a strong advocate of a probabilistic approach, particularly because I believe that the difficulty in estimating probabilities without bias can be overcome by effective elicitation methods.

3. Expert elicitation

Good expert elicitation methods are essential when expert judgement is the only way to quantify uncertainty in model parameters. The Paper could be clearer whether elicitation is the process of quantifying uncertainty in a parameter in the form of a probability density function (PDF) by experts only, or includes eliciting other information from experts to e.g. what assumptions to make, conceptual models etc. It's the former that I am interested in commenting on here. (I've avoided the term elicitation in [1] and referred to the process throughout for clarity as 'uncertainty quantification by expert judgement'.)

The Cooke methodology cited in The Paper is an excellent method for combining PDFs from different experts taking into account both their knowledge and their skill in quantification of uncertainty, but a concern is that few experts actually contribute to the combined PDF because most are very poor at the latter, which isn't great use of resources. It is probably the best method for mathematically combining input from multiple experts, but unless significant steps are taken to minimise experts' bias, that will be the best of a bad bunch – maybe that's why on P5 The Paper recommends using as many experts as possible. Training and good facilitation can help here and I strongly support the comment on P6 of The Paper about the development of training methods for experts and facilitators alike – this is necessary.

I think an effective elicitation process should consist of three things:

1. People use quick mental heuristics to make decisions under uncertainty, which are subject to significant cognitive biases, usually resulting in over-confidence. There is therefore a need for an initial training exercise which highlights these natural biases. I have developed one using weather statistics which works effectively (see S6.2 of [1]).
2. The process of an expert generating a PDF needs to follow a structured approach designed to avoid these cognitive biases, especially 'anchoring' on an initial guess and failing to adjust enough for uncertainty. This is essentially done by asking the question in an indirect way ensuring that the expert considers extreme values (of the PDF) first (see S6.3 and S6.4 of [1]). Just asking experts to give ranges or properties of the distribution directly will not avoid bias.
3. If a group of experts are used, their inputs need to be combined. This can be done mathematically e.g. by the Cooke method, or the PDF can be generated by the group together by discussion aiming for consensus. There are pros and cons of both; we have tended to do the latter (see S6.4 of [1]).

I think The Paper would benefit from a paragraph on cognitive biases (see S5.1 of [1]) as the reason for requiring a formal elicitation methodology at the beginning of S2.3.

4. Assumptions

The Paper uses the word 'assumptions' far too often, in my view, and I don't think phrases like 'assumptions about the uncertainty' (e.g. P9) actually make sense – an assumption is something you make to avoid quantifying uncertainty, it has no place in the process of doing so. If authors really believe they are making assumptions about uncertainty instead of quantifying it, that doesn't leave us in a good place.

Clearly, there is a need for some assumptions in defining scenarios, and building a conceptual model for a given scenario, but care needs to be taken mixing assumptions with a probabilistic approach for two reasons. Assumptions exclude uncertainty, meaning the

results of any probabilistic analysis on the remaining uncertainty are conditional on the assumptions being correct, if the majority of uncertainty is excluded by assumptions, the probabilistic analysis may not prove very helpful. If very conservative assumptions are used for a subset of parameters in a model as an alternative to quantifying uncertainties, this can cause an otherwise probabilistic assessment to be biased, which again may reduce its usefulness (see S4.3 of [1]).

P3 of The Paper cautions about making assumptions about the form of probability distributions, and is right to highlight this as an issue. In my view when using expert judgement to quantify uncertainty in a parameter, it is best to not to constrain the shape of the distribution. Therefore, I would recommend using a general cumulative distribution – that is a PDF where the cumulative distribution function (CDF) is piecewise linear – which allows the expert to put as much or as little structure on the PDF (particularly in the tails) as he or she is able to. Another important issue is the correct choice of scale, linear or logarithmic. These issues are discussed further in S2.3, S2.4, S6.3 and S6.4 of [1].

I hope these comments are useful and am happy to discuss them if that would be helpful.

Mike (mike.poole@nda.gov.uk)

[1] RWM, *Geological Disposal: Methods for Management and Quantification of Uncertainty*, NDA/RWM/153, 2017. This report is available here: https://rwm.nda.gov.uk/publication/methods-for-management-and-quantification-of-uncertainty/