REVIEWER 1

Reviewer: A review of the issues relating to epistemic uncertainty is very welcome. However, as might be anticipated from a paper with as many as 13 listed authors, it is very disjointed.

Clearly, individual authors have contributed their own sections and paragraphs, relating to the particular natural hazards with which they are most familiar and competent.

However, the most interesting questions are those which relate to understanding why particular methods are suitable and have been adopted for some natural hazards, but not others.

To address such questions requires much more author interaction and dialogue than have taken place during CREDIBLE meetings. Having attended several such meetings, the outcome comes as little surprise, if rather disappointing.

To resolve the inter-hazard questions would require a major revision to this paper. This might be asking too much of the authors at this stage. What a reader should expect is that the paper be more comprehensible and read as if it were one article, rather than a patchwork of sections. I suggest that the senior author(s) revise the paper to make it more cohesive.

Response: We accept the critical comment by the reviewer and have restructured the paper to have a more coherent cross-hazard structure, and trying to avoid the impression of a collection of non-integrated sections.

REVIEWER 2

Reviewer: The authors attempt to present a comprehensive review about the role of epistemic uncertainties in natural hazard risk assessment. Before publication, however, substantial work needs to be carried out to improve the paper.

My major criticism is that the authors have not clearly defined the audience of the manuscript. NHESS is an interdisciplinary journal read by all different types of earth scientists, practitioners and decision makers. Topics include, among others, hydrology, geology, climate sciences, geomorphology. Yet the paper fails to be accessible for these audiences, for several reasons: 1. several specialist concepts are only briefly introduced, if all. For example, bias correction or downscaling are rather complex issues, but are not even explained in two lines. Even references to state-of-the-art overviews are missing. 2. also general statistical or risk assessment concepts are not well explained. For instance, the SEJ approach is presented, but not explained. 3. The paper tries to be comprehensive, but as a result often merely paraphrases issues rather than guiding the reader about problems and issues. It is often lost in detail, doesn't see the woods for the trees. 4. The language is often not appropriate. For

instance the term simulator is widely used in the statistical literature, but it is basically unknown to people using these simulators. They would write of dynamical models, climate models, hydrological models etc. But still the term has not been clearly introduced in the paper. This (and similar issues) make the paper difficult to read for non-experts. All in all one sometimes has the impression that the authors simply want to impress the reader about how much work they have done in the field rather than providing a useful account of the issues.

Related to the previous point: given the scope of the journal and the experience of the authors, I find it slightly irritating that at least one third of the references are self citations - many key publications from the individual disciplines are missing. For instance in climate science there is a whole bunch of key papers written by, e.g., Reto Knutti, Claudia Tebaldi, or David Stainforth, which has not been cited. Further examples will be listed below. I would guess that the same holds for other disciplines.

Again, related to the issues in I: <u>often, the manuscript does not attempt to bring across the most relevant points</u>. In one particular case I find this dangerous and misleading: the author at several places state the role of unknown unknowns, but fail to work out and highlight the associated problems. They nonchalantly state that "modellers are pragmatic realists" who know that models are simplified representations of reality, and at other places mention the possibility for surprises. They even state that experts often underestimate the uncertainties about issues not directly related to their own field of research. So they claim that the assumptions made in developing such simplified models need to be communicated to stakeholders. But I think this is a dangerously naive view of how stakeholders perceive science. In many cases, surprises are really to be expected, and scientists, if thinking clearly about their work, know that their uncertainty assessments are basically useless because the assumptions are overly simplifed. But still, they sell these results to stakeholders, mentioning the uncertainties, and wrap them in nice scientific parlance. **From my own experience with stakeholders I know that they are often completely unaware of the weak basis our predictions are based on, because they lack the scientific knowledge of really grasping the seriousness of our model deficiencies**. Here I would really urge **to add some real world examples**, where predictions have utterly failed because models where too simple. A famous example is the recent financial crisis, but most likely the authors know much better examples from natural hazards. I found none, because in all the examples listed - Katrina, Aguila, Fukushima - it was about wrong management decisions.

Actually, the authors themselves fall into the trap I described above. I found several examples where they lay out specific examples and then fail to properly describe the associated epistemic uncertainties because they lack the specialist knowledge. For instance, the authors describe downscaling and bias correction, but place naive statements such as "to correct for any bias predicted and observed values" because simulator variables are not commensurate with observations. Bias correction is a typical example of a wide spread technique that is applied without much understanding of the underlying climate and

climate model errors such that bias corrected multi model ensemble projections might provide severely ill-designed uncertainty assessments which might lead to wrong adaptation decisions.

Response to the general comments: While accepting the criticism of the referee, it seems that there was some misunderstanding of what the paper was attempting to achieve. It was not intended as a comprehensive review. That would be impossible given the large number of papers published across disciplines. It was much more an opinion piece (a review of the issues as the title suggests) intended for all the readership of NHESS who might be interested in the impacts of epistemic uncertainties. Certainly we assumed some acquaintance with existing techniques, while providing citations to be followed up as necessary, but evidently too much for this referee.

There seem to be some points presented here as criticisms (eg bias corrections, communication with stakeholders, surprises) where the referee seems only to be reinforcing what we are trying to convey in the paper.

The revised manuscript broadens the cited literature to reflect key papers, including those suggested by the referee, but making it clear that this is as much an opinion piece as a review. We have also tried to reference more examples as suggested.

We have also addressed the following points of detail in the revised text

Reviewer: page 7330, lines 4 to 18: here, human reflexive uncertainty should be discussed. In particular in climate sciences, it is only possible to provide probabilities (if at all) conditional upon emission scenarios, because the evolution of climate in the future will influence mitigation decisions in an essentially unpredictable way.

But conditional scenarios are already presented as a way of dealing with epistemic uncertainties, including the general impossibility of associating them with a probability (see also Rougier and Beven, 2013)

Reviewer: page 7341, lines 21-28: in the last ten years there have been several now classical review papers about statistical downscaling and bias correction which have not been cited here (e.g. Fowler et al., Int. J. Climatol, 2007; Maraun et al., Rev. Geophys., 2010). Also the body of critical papers about bias correction is growing fast, but none have really been cited here.

Additional references have been added

Reviewer: page 7349, line 22: I think the paragraph about sensitivity studies is key, also in the light of the issues I discussed above about surprises. For instance, recent papers showed that our current generation regional climate models might not be fit for purpose to simulate heavy summertime extremes (e.g., Kendon et al., Nat. Clim. Change, 2014; Meredith et al., J. Geophys. Res., 2015). These problems have been revealed by sensitivity studies and highlighted that traditional uncertainty assessments would have utterly failed.

Exactly – and thanks for the additional references

Reviewer: page 7351, line 18: there are no projections on decadal scales, only predictions. On such short time scales, scenario uncertainties don't play any relevant role yet, i.e., one really produces actual predictions.

Reworded

Reviewer: page 7352, lines 20-22: such sentences need much more emphasis!

OK

Reviewer: page 7353, lines 1-2: remember, there is no free lunch. If there is a fundamental problem, also sophisticated Bayesian approaches will not provide a way around.

Again, this reinforces what we are trying to say

Reviewer: page 7353, line 15: grammar, verb is missing

Corrected

Reviewer: page 7356, line 3: arise, not arises (refers to issues)

Corrected

Reviewer: page 7357, line 7-17: this paragraph needs a much more critical stance, as discussed above. I fear the approach laid out here simply helps to avoid liability, but does not improve the basis for decision makers.

Reviewer: page 7357, line 18-: this paragraph is one example (of several) where the authors are lost in detail. It is completely inaccessible to non-experts. Please provide explanations or delete it.

Reviewer: page 7358, line 17-: this paragraph is too vague

Reviewer: page 7360, line 19-21: too vague, and even misleading (see discussion above).

Reviewer: page 7361, line 6-13: surprises are not just possible, but very likely! The whole paragraph is too vague and uncritical.

See completely revised text - These comments have been taken into account

EDITOR

The invited reviewers for this paper were all people who are qualified to comment on generic methodological aspects. You have doubtless read the two

existing reviews already, and seen that they both express disappointment with the paper. Reviewer 1 finds it disjointed, and expresses a view that it would have been useful to understand which approaches might be more or less suitable for different hazard areas (and why). This reviewer was chosen for their "landscape-scale" appreciation of the subject matter. Reviewer 2 was asked specifically to comment on aspects relating to climate and related issues: this reviewer is also very critical, and adds some detail to the concerns raised by Reviewer 1. **Specific concerns here are that the paper is not accessible to a wide readership** (this links with the "disjointed" concern from Reviewer 1) because, for example, technical or subject-specific terms are not defined clearly for the benefit of those who are unfamiliar with them. **This reviewer also notes that the paper overemphasizes the authors' own work to an inappropriate degree**. I strongly endorse this view, and note that a similar comment was made in relation to the companion paper.

Response: See comments following editors final comment below

Editor: To the reviewers' reports, I add here some detailed notes of my own. Thus: given that the aim of this NHESS special issue is to showcase new approaches for estimating risk and uncertainty in natural hazards, any review articles will ideally go some way beyond the "standard" level of review in this area. I must confess that although the present submission makes some very valid points, much of the material feels rather familiar and to echo points that have been made repeatedly elsewhere, albeit often in hazard- or discipline-specific outlets. Given the very substantial effort that has been made in the last few years to bring communities together and transfer ideas between disciplines, in the UK in particular, I am disappointed that **the current submission misses an opportunity to "set the record straight" by articulating valid viewpoints that have perhaps received limited exposure in natural hazards communities**. Indeed, at some level I don't think the contribution is really a "review" so much as an **"opinion piece"** - and I have some doubts as to whether all of the authors have *really* signed up to all of the opinions expressed therein (certainly, there are some statements in the paper that surprise me when I look at the author list). As with the companion paper: **if this article is to deliver on what it promises, it needs to be much better balanced and to show considerably more awareness of other relevant literature**.

There are implicit criticisms of probabilities as being potentially inappropriate for representing different types of uncertainty: but in all cases (for example the material in lines 82-2) these are predicated on an interpretation of "probability" to which a Bayesian, for example, would not subscribe. Thus, many NHESS readers will interpret probabilities in the way that the authors imply: but for a review of this nature there is an obligation, I think, to acknowledge and articulate clearly the alternative viewpoint in which probabilities are used merely as a calculus to represent one's knowledge about the state of the world. I do not for a moment disagree that other uncertainty concepts may be useful (as indicated in lines 82-85); but it must be made absolutely clear, with appropriate explanation, that the arguments against the use of probability are perhaps more precisely regarded as arguments against the *interpretation* of probabilities that many hazard scientists are familiar with: the problem is not necessarily with

the probability calculus per se.

Response: See comments following editors final comment below

Editor: There are two or three places in the paper where rather sweeping statements are made about the disadvantages and (implied) inapplicability of specific techniques, but where the fundamental problem seems to be with the implementation rather than the underlying concept itself. For example:

Lines 205-206 "even experts find it difficult to estimate probabilities for sources of epistemic uncertainty with any degree of confidence". This is a standard criticism levelled at probabilistic elicitation exercises in the natural hazards community. However, the problem arises at least in part because natural hazards experts invariably are not trained in how to interpret probabilities in such situations, and elicitation exercises also are often carried out by those lacking appropriate technical (i.e. mathematical and statistical) skills and awareness. I am also aware that these kinds of techniques *are* routinely used in other application areas, and I believe a considerable amount of work has been done on the elicitation of probabilities in such a way as to be relevant for the subsequent uncertainty analysis (this is, however, an area in which I was hoping to obtain additional input from the third reviewer). It is certainly challenging, and in general it requires a skilled and experienced facilitator who knows what questions to ask and how to ask them (as, indeed, you acknowledge on lines 379-382): in non-critical situations therefore, the costs of such an approach might be deemed to outweigh the benefits. But in this kind of review article, you have a responsibility to paint a balanced picture and to ensure that the *narrative* (rather than just the occasional parenthetical remark) is faithful to this picture.

Response: We hope that the completely revised structure allows this more balanced presentation, but we remain convinced that current (good, skilled) practice is overconfident in its inference.

Editor: Lines 262-264 "Use of simple aleatory error based likelihoods or probabilities does not allow enough potential for surprise from arbitrary rather than aleatory future occurrences": again, this confuses concept with implementation. I agree 100% with the statement as written; but the implied conclusion (that the problem is with the use of likelihoods or probabilities for aleatory uncertainties) does not follow. In my view the key word in the quotation above is "simple". This is particularly relevant given that it follows on from a discussion of stochastic downscaling with bias corrections. Bias-correction approaches are mostly jaw-droppingly naive, and there is plenty of literature around that not only makes this point but also highlights the existence of much more sensible downscaling approaches that address many of the concerns (why are there no citations to papers by, for example, Douglas Maraun and co-workers?). Of course, you can never rule out the "black swan" (I hate that expression) - there's a really nice example in one of Stuart Coles' papers, involving a flash flood in Venezuela that was off the scale by comparison with anything in the historical record. But you can certainly improve by one or two orders of magnitude on "typical" current practice: just ensure that the work is done by, or in collaboration with, people who have the skill set and required

training to make a decent job of it. In terms of handling epistemic uncertainties in natural hazards, lack of appropriate skills is at least as big a problem as anything else that is mentioned in the paper!

Response: In the revision we try to distinguish between skilled and over-simple approaches, but still wish to convey our opinion that even the skilled and conditional representation of epistemic uncertainties needs reconsideration in future.

Editor: Line 491 "disinformative": please define precisely what you mean by this - I have never seen a clearly articulated definition. My best is guess that you think data are "disinformative" if the analysis would be better (in some sense that I don't fully understand and that I would like you to define) without them than with them. However, there is no such thing as negative information. In an ideal world, one would acknowledge explicitly the potential of the data to be incorrect for whatever reason, and would formally incorporate this into the analysis. This might result, for example, in the "disinformative" data values having negligible influence on the results. The thinking here once again seems predicated on the assumption (which is, unfortunately, reasonable in many situations) that the risk assessment is being done using naive and simplistic methods, and by someone who lacks the skills fully to address the problems of combining data and models in a complex situation. It may be judged that it would cost too much to hire somebody to do this kind of work in any particular application, or that it would be too time-consuming; but let's be clear that the problem about "disinformativeness" is not a philosophical problem as the paper seems to suggest: it's a logistical problem about resource availability / allocation. Thus, in lines 505-508 there's a hypothetical example involving mass balance errors in data: this is easily resolved *in principle*, simply by allowing for uncertainty and incorporating the requirement for mass balance formally *and appropriately* into the analysis.

Response: In this case hydrologists in real applications have to make due with past data that cannot be improved by adding resources. Traditionally (even where uncertainties have been estimated) all the data have been used in model calibration and evaluation, despite the fact that some of those data might be consistent with the principles on which the model(s) are based. In some cases it is possible to identify periods of data that are inconsistent in this way and would therefore be disinformative about model inference. The first step in providing a solution is to recognize the problem (which might also arise in other natural hazard domains as noted). It is doubtful if the type of allowance for water balance uncertainty could be simply formulated as suggested across the rather arbitrary errors associated with individual events (especially where storage dynamics introduce complex time structures). However, in the revision we refer only to inconsistencies in data (with disinformation only mentioned in relation to the particular hydrological references).

Editor: Lines 663-665 "there are dangers in applying Bayesian statistical theory, particularly in using a simple error model and associated likelihood function to represent epistemic uncertainties". It's the same thing again: the problem is the "simple", not the "Bayesian statistical theory". You might not be willing to

expend the effort to build a sufficiently realistic representation to overcome the problem: but the danger then is a consequence of your decision, not of Bayesian methodology per se.

Response: Again, see revised structure of the text

Editor: Lines 126-127 "changing in water level to discharge rating curves after major events": something wrong here (or at least the syntax / punctuation is such as to obscure the intended meaning).

Response: reworded

Editor: Lines 260-261 "this method will generally overestimate the information content of the historical data": can you clarify what you mean by this?

Response: This has been clarified in the revised text, it results from the use of overly simple error models

Editor: Lines 278-279: see general comment above about expert elicitation. The current statement here needs to be tempered accordingly. That said, the following sentence about precautionary or robust decision-making is certainly sensible.

Response: agreed

Editor: Lines 297-298 "Western societies increasingly seek to place blame ...": is this true? The only controversial issue of this nature that I can think of is L'Aquila (e.g. in the case of Hurricane Katrina, it sees to me that it was entirely appropriate to question to role of the relevant authorities) - and L'Aquila is the only example that is given. Perhaps you mean: "Following the legal case resulting from the L'Aquila earthquake, there is increased concern among the scientific community about being held responsible for natural disasters if scientific advice is subsequently deemed to have been inappropriate [INSERT PLENTY OF REFERENCES]".

Response: This has been expressed more carefully

Editor: Line 397: I'm not sure that "Paper 2" has been mentioned anywhere previously. In any case, for avoidance of ambiguity (I initially read this as "Graham's second paper") it would be better to write something like "(e.g. Graham, 2000; see also the accompanying paper in this volume)". Similarly line 832.

Response: Agreed

Editor: Lines 452-454 "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". I agree with this: the hydrologists are some way ahead of the game because this is a discipline where there is a decades-old culture of thinking stochastically. It's the issue of having the right skill set again: it is quite possible that when other disciplines

have acquired the same skill set, their own risk assessments can be transformed in a way that is currently hard to imagine.

Response: Except that the point here is that those estimates might be quite wrong in ignoring uncertainties, for example, about the relevant frequency distribution.

Editor: Line 457 "annual exceedance probability": whatever that means in a nonstationary climate!

Response: Indeed, but it is allowed to be changed (and usually is after every new flood!!)

Editor: Lines 479-480 "the consequences of failure might be high impact": probably you mean either "the consequences of failure might be serious" or "failure might have a high impact".

Response: good point

Editor: Footnote 4: it seems to me that if a reader doesn't know what a fat tail is, they're unlikely to have encountered the term "kurtosis" before!

Response: good point

Editor: Lines 517-518: "given that we lack the ability" => "if we are unable or unwilling", I think.

Response: good point

Editor: Lines 588-592: this material about the stability of bias corrections is again a criticism of naive approaches. There is nothing (except lack of awareness and / or time) to prevent us from modelling the potential for the bias to change with the dynamics - indeed, some of us do this routinely. Obviously one can never escape from a fundamental assumption that the model structure continues to hold in the future, but if the stationarity assumptions are embedded much more deeply within the model structure - at the level of *physical* rather than empirical relationships, for example - then we gain increased confidence.

Editor: Lines 623-624: I agree with this statement about a possible use of Bayesian updating - it is useful to make this point I think.

Thanks

Editor: Line 654: "assumption" => "assumptions"

Corrected

Editor: Footnote 6 is not necessary, and is needlessly obfuscatory. Nobody needs to know about Borel spaces, they are defined merely so that mathematicians can sleep relatively undisturbed.

Editor: Line 699 "conditionality of the outputs": what does this mean? - Lines 715-718: I would strongly advocate adding

http://www.stats.gla.ac.uk/~adrian/papers/graphics-for-uncertainty-paper.pdf, and references therein, to this list. The paper is under review, I believe.

Editor: Lines 726-729: this material about the resolution of visualisations is in fact merely a modern manifestation of an issue that has been known about for decades: how many decimal places to use in tables of results, what contour spacing to use on a map etc.

Response: True, but it does become a particular problem with visualisations that are increasingly presented as a virtual reality

Editor: Lines 819-823: this issue of whether the next event will be informative or disinformative is easily handled in principle, simply via an appropriate representation of the potential data structure (e.g. by embedding the "information class" of an event as a latent variable). Another example where problems go away if you have better awareness of the possibilities.

Response: We disagree. Such a latent variable requires evidence on which to define the class, but in prediction that evidence is not available a priori – the next event might turn out to have been in an informative class or it might be in a disinformative class a posteriori.

Editor: The bottom line is that, unfortunately, this paper needs a *very* large amount of work to deliver what it promises, at least in the context of a special issue that claims to show-case the state of the art. Some of the less balanced discussions can perhaps be fixed reasonably straightforwardly; but it will take considerably more work to deal with the lack of appropriate citations to the wider literature (noted by Reviewer 2 in the context of the climate literature; I am guessing that the comment applies at least to some other hazard areas as well). It is hard to know how best to deal with this. **Perhaps the path of least resistance would be to reframe the paper (and, maybe, its companion) as more of an opinion piece than a review, and to make absolutely clear that it represents the collective views of the authors and does not attempt to be comprehensive.** If you do this then at least you won't be promising something that you can't deliver in the time available. Nonetheless, even an opinion piece should be balanced, scholarly and even-handed; and the limitations should be clearly acknowledged.

Response: We only ever intended the review to be comprehensive (as far as possible) about the ISSUES, not about research in individual hazard areas. The original paper was structured with this in mind, and was a reflection of the joint opinions about those issues. This leaves us open to criticism from any individual hazard area. More hazard related detail was added in Paper 2 but for

some reason Paper 2 was processed and reviewed before Paper 1, so that it was not seen in the context of Paper 1.

As the editor recognizes there are multiple opinions about how to cope with epistemic uncertainties. This includes the current authorship, but all the authors did indeed sign up to the original manuscript. Going along with the suggestion that it should be presented as more of an opinion piece, we have revised it accordingly to try and make clearer the potential for differences of opinion, the limitations of both simple and skilled statistical approaches, and consequently where we think there are remaining open questions.