[General reply from the authors]

We thank the reviewer for his time spend reviewing our manuscript and his suggestions for improvement. Below we have replied to the various comments made by the reviewer. For ease of reading we have colour-coded our responses:

Colour coding of response:
Black - Reviewer comment
Blue - Author reply
Green - Proposed change in manuscript

[Replies to reviewer comments]

[R2.1] Overall the paper is well structured, well written and quite clear. However, I have a number of reservations regarding the manuscript. Several of these overlap with the points raised by Joseph in his comments, so I won't discuss those here in any further detail.

[Reply to R2.1] We thank the reviewer for his kind remarks. We acknowledge that some comments overlap with the first review, and will refer to our reply to the first review if applicable.

[R2.2] My main reservation is regarding the contribution that the paper is trying to make. In the abstract the authors in the second sentence mention the role of model-based analyses in supporting decision making, while ending the abstract with claims about model-based decision making. In the middle of the abstract, the authors mention a new metric 'relative uncertainty'. Next, they argue that using this new metric they can provide insight into the uncertainty about the effects of a variety of flood risk reduction intervention.

I have two reservations here. First, regarding the claims related to decision making. In my view, these claims are not well developed in the paper. Outside of the introduction and conclusion, the term itself only appears once. Moreover, if I look at the presented results and their visualization, I doubt these would be used by a decision maker or even someone directly advising a decision maker. Rather, the presented analyses are useful for people working on the design of flood risks reduction strategies, while some of the results might indirectly be used in decision making. As such, I would suggest removing the term from the title as well as tone done any other claims made in the abstract and introduction. Alternatively, the authors would need to expand their discussion in the main text on how providing insights into the range of uncertainty about expected effects of measures can assist decision making on flood risk. This, however, requires discussing notions such as robustness (see e.g., McPhail et al., 2018) and flexibility, as well as discussions on well characterized uncertainty (i.e., you have a meaningful pdf) and Knightian or deep uncertainty (i.e., for whatever reason you don't have a meaningful and uncontested pdf).

[Reply to R2.2] We acknowledge that our use of the term "decision making" was intended to be narrower, and specific to decisions regarding the design of interventions. As the reviewer might well know, the established procedures for the design and approval of river interventions in The Netherlands (in Dutch, "Rivierkundig beoordelingskader") is highly model-based. It was inspired by this context that we intended to frame our work. We think it is to this the reviewer refers when he says "Rather, the presented analyses are useful for people working on the design of flood risks reduction strategies, while some of the results might indirectly be used in decision making."
So rather than being used directly in decision making, these models, and the extension offered by our work, would assist in decision making.

To avoid a too liberal use of the term 'decision making' in the context of the literature suggested by the reviewer, we made the following modifications.

The title was:
"Uncertainty quantification of flood mitigation predictions and implications for decision making."

We changed it to:
"Uncertainty quantification of flood mitigation predictions and implications for interventions."

In the abstract we wrote:

"In general, we conclude that the uncertainty of model predictions is not large enough to invalidate model-based decision making, nor small enough to neglect altogether. Instead, uncertainty information can be used to improve intervention design and enrich the decision making process."

We changed this to:

"In general, we conclude that the uncertainty of model predictions is not large enough to invalidate model-based intervention design, nor small enough to neglect altogether. Instead, uncertainty information is valuable in the selection of alternative interventions."

[R2.3] My second reservation is with the new metric itself. As also indicated by Joseph, this metric is closely related to the coefficient of variation. Moreover, the more theoretical discussion of this new metric is confined to one short paragraph around equation 5. Too play devils advocate: what is the merit of publishing a paper whose only contribution is a single equation closely related too an already established metric? The case study mainly serves to establish the value of this metric, while the models are taken from earlier work. If you insist on having the metric as a key contribution, than a comparative perspective would be more appropriate. So, what other metrics already exist that could serve a similar function? Classic robustness metrics as well as the coefficient of variation would be logical candidates. How is this metric different from these, and what are the relative merits of the new metric relative to the others?

[Reply to R2.3] We agree that the key contribution is not the relative uncertainty. In response to remark **[R1.2]** we have removed the focus on this metric.

[R2.4] Basically, both reservations have to do with how the authors currently position their work. I am having profound reservations regarding the first two paragraphs on page 3. First, the authors cite Pinter on fuzzy math, suggesting that not providing explicit uncertainty quantification leaves decision makers free to interpret the uncertainty in any way they like. My argument would be that the converse also often happens. As for example elaborated nicely in Pilkey and Pilkey-Jarvis (2007), a lot of quantitative analysis for supporting decision making becomes useless arithmetic that is used strategically. In a different line of literature, researchers working on post normal science often claim that many models have not even one significant digit. That is, models give only a false sense of precisions. A third line of research, as exemplified by Sarewitz (2004) shows how more research and increasing efforts to quantify uncertainty can often make environmental controversies worse. In short, the relationship between models, model results, uncertainty about models results and decision making processes is quite a bit more complicated than the authors seem to think. It would benefit the manuscript it the claimed contribution is better positioned relative also to these strands of literature.

[Reply to R2.4] As motivated in **[Reply to R2.2]**, we did not aim to focus on the decision-making process. The authors are aware of the post-normal science discussion, and concede that quantitative information can be used strategically, especially if there is too much focus on technical uncertainties to create superfluous knowledge which is not relevant to the policy process (see e.g. Warmink et al., 2017). We would not argue for the results of such uncertainty quantification (or model results in general) to take the centre stage in decision making, but merely to provide valuable information to assist in that process. We appreciate the suggestions of the reviewer to embed our work in the suggested literature, and will rewrite the paragraph [P3, L3-L10] considering these suggestions and the intended scope of our work (cf. **[Reply to R2.2]**).

Warmink et al. 2017: (P4592) in Water Resour Manage (2017) 31:4587–4600. DOI 10.1007/s11269-017-1767-6.

Minor remarks

[R2.5] Might it not be more convenient to show the interventions (2.2.1-2.2.6) in a table?

[Reply to R2.5] Our choice for subsections instead of a different format (e.g. a 'box') was mainly motivated by the limitation of the NHESS template rules. If approved for final typesetting, we will confer with the journal whether a different format is possible.

[R2.6] What is the runtime of the detailed model for a single run?

[Reply to R2.6] The runtime of a single model run was approximately 2.5 hours. We added this information to section 2.2.

[R2.7] Page 18, line 17, "for new ones greatly the unexplained" I guess some words are missing here

[Reply to R2.7] The missing word is "increased". For our answer we refer to our reply to **[R1.14]**

[R2.8] Page 17, line 4, partly overlapping with Joseph's comment, but what justifies the linear interpolation?

[Reply to R2.8] We refer to our answer to **[R1.7]**