[General reply from the authors]

We would like to thank the reviewer for taking the time to review our manuscript. We highly appreciate his suggestions and comments, which are helpful in improving the manuscript. 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]

[R1.1] This article assesses the impact of 12 interventions on flood water levels in the Waal River, including estimation of uncertainty. It goes further than most other studies in 1) estimating uncertainty in uncertainty attributable to the methods used, 2) discussing the potential to approximate the uncertainty, in ways that make the results of the analysis more accessible to a broader audience. The article is well written and clear. The issues I would like to highlight are relatively minor - clarifying some claims rather than fundamentally questioning them. I do suggest some small extensions to the analyses to help in clarifying these claims.

[Reply to R1.1] We are grateful for the kind words, and the acknowledgement of the reviewer that our efforts to quantify uncertainty by approximation makes uncertainty quantification more accessible to a broader audience.

[R1.2] I would rather that the abstract avoid emphasising that "relative uncertainty" is a "newly introduced" metric. As the authors do acknowledge in the main text, it is closely related to the coefficient of variation. If this metric is kept as a contribution, it would require greater discussion of why is preferred over the CV. The greater contribution here, in any case, is that relative uncertainty is identified as a key parameter to obtain approximations of uncertainty in this type of problem.

[Reply to R1.2] We agree with the reviewer that the introduction of relative uncertainty as a metric is not the key contribution of this paper, rather that the relative uncertainty is an effective metric to discuss the uncertainty of engineering interventions, and that this should be reflected in the abstract. In the abstract, we wrote:

“ (...) We show that the uncertainty of an intervention can be adequately described by the newly introduced `relative uncertainty' metric, defined as the ratio between the confidence interval and the expected effect. (...)”

We changed this sentence to:

“(...) We identified the relative uncertainty, defined as the ratio between the confidence interval and the expected effect, as a metric to compare uncertainty between different interventions. (...)”

[R1.3] p9 L25 "We would need to obtain both the model output probability distributions, as well as the covariances between the two distributions" This is not entirely true - most of the parameters would be identical in both states of the river, so the difference can be calculated directly as a single distribution - covariances probably do not need to be calculated. The authors do end up doing this with the "union stochastic parameter space" (p11), so it seems worthwhile to clarify that you do not explicitly estimate covariances. It is still correct that this is probably not tractable for 12 interventions.

[Reply to R1.3] The reviewer is correct in noting the covariances are not explicitly computed (or estimated). To be clear, we are not referring to the covariances between the stochastic parameters (which are in our case identical between the two states of the river), but the covariances of the distributions of the computed water levels. These should always be considered, either explicitly (if done analytically) or implicitly. Otherwise, the uncertainty of the intervention will be needlessly overestimated. Furthermore, the problem is *analytically* intractable because of the numerical models involved, regardless of any covariance.

We wrote [p9, L23-L27]:

"To compute differences between these two (stochastic) states of the river system, we would need to obtain both the model output probability distributions, as well as the covariances between the two distributions (Berends et al., 2018). This problem is not analytically tractable, but can be solved numerically by Monte Carlo simulation (MCS; Metropolis, 1984; Stefanou, 2009). However, MCS is not practically feasible; (...)"

We changed these sentences to:

"To analytically compute differences between these two (stochastic) states of the river system, we would need to obtain both the model output probability distributions, as well as the covariances between these two distributions. This problem is analytically intractable, but can be solved numerically by Monte Carlo simulation (MCS; Metropolis, 1984; Stefanou, 2009), while accounting for covariance by sampling from a union stochastic parameter space (Berends et al., 2018). In the numerical approach, covariances are not explicitly computed.

However, MCS is not always practically feasible or desirable. (...)"

[R1.4] p15 To support the claim that values are relatively constant, it would be useful to show the range of E(Ur90) over the length of the river, either in Figure 7 or Table 2 E(range(Ur90)) would also be useful, given it is available, i.e. the range of Ur90 over the length of the river, averaged over all realisations There is enough variation there assuming relative uncertainty is constant seems like it would still be a substantial simplification.

[Reply to R1.4] The first suggestion of the reviewer is to add the range of E(Ur;90) to either Fig. 7 or Table 2. In response, we will add the standard deviation of E(Ur90) to Table 2 to explicitly show the variation along the channel. In the text, we wrote [p14, L14-L16]:

"Next, we compute the relative uncertainty (Ur;90) for each intervention and intensity. Since Ur;90 is stochastic, we first reduce this by calculating the expected value (E(Ur;90)) for each point along the river (Fig. 7). Results show that values for E(Ur;90) are fairly constant over the river length up to the location of the maximum flood level decrease (xm) for all interventions."

We changed these sentences to refer to the new information in Table 2:

"Next, we compute the relative uncertainty (Ur;90) for each intervention and intensity. We first marginalise the uncertainty in Ur;90 (which is due to CORAL estimation) by calculating the expected value (E(U r;90)) for each point along the river (Fig. 7). Figure 7, as well as the relatively small standard deviations in Table 2, show that values for E(Ur;90) are fairly constant over the river length up to the location of the maximum flood level decrease (x m) for all interventions."

The second suggestion of the reviewer, is to show E(range(Ur;90)), as well. This would show the variation in the estimation uncertainty along the channel. In figure 8, we visualized the estimation uncertainty at x_m. We feel that including variation along the channel as well, would conflate the estimation uncertainty with the variation along the channel and as such not provide meaningful additional information, while making the figure more difficult to understand.

[R1.5] p16 Figure 8 "FLPSMOOTH is significantly more uncertain" The use of the word "significantly" suggests this is a statistical statement. It might be worth putting a p-value on this, given you have uncertainty on the uncertainty. Visually, it seems high intensity FLPSMOOTH could be significantly more uncertain (p=0.05), but that low intensity is comparable to groyneflow or minemblow, also confirmed by Figure 7. L6-14 could also comment on statistical significance of differences in uncertainty.

[Reply to R1.5] Here, it was not our intention to provide a statistical significance of differences between the distributions, given that that the answer from any statistical test used to obtain a p-value would be conditional on both the intervention subsample size and the number of MCMC steps used in the approximation. To prevent confusion, we removed the word 'significant'.

We wrote [caption Figure 8, p 16]:

"(…) showing that /flpsmooth/ is significantly more uncertain than other interventions."

We changed this sentence to:

"(…) showing that /flpsmooth/ is more uncertain than other interventions."

[R1.6] p17 Figure 9 This figure is illustrative of the interpolation method, but currently only hints at how good the interpolation method would be. You have two sources of information on this: 1) the uncertainty in MCI90 and E(dH) 2) the variation in these values along the river The error bars for (1) in the figure do suggest that there is quite a bit of uncertainty involved in the interpolation. I would suggest showing a ribbon around the line encompassing all potential lines that could be interpolated using 1 and 2. This doesn't detract from the utility of using linear interpolation, but would provide a better sense of the confidence one should have in using this simplification. If the uncertainty due to using CORAL is too large, perhaps more runs might also help make your point more effectively.

[Reply to R1.6] We agree with the reviewer that there is quite a bit of uncertainty involved in the interpolation and that multiple lines could be drawn between the high and low intensities. In fact, showing a ribbon around the line would still not resolve all uncertainty, as this still is conditional on the assumption of a linear model (i.e. a ribbon would show parameter uncertainty of the interpolation model, but not the model uncertainty of the interpolation model structure).

The main purpose of this figure, as with Figure 10 (see reviewer comment R1.7), is to show how uncertainty affects design decisions. We believe that introducing more complexity, by showing every possible line that could be drawn (under the assumption of the linear model), would detract from this message. On the other hand, it should be clearly stated that the linear line is, at best, a first approximation. We made two changes in the manuscript to reflect this.

In the results section we wrote [p16, L16 – p17, L1]:

"By linearly interpolating between the two low and high intensities, we can estimate the uncertainty for a given expected effect for each intervention (Figure 9)."

We changed this sentence to:

"By linearly interpolating between the expected values of the low and high intensities, we obtain a first approximation of the uncertainty for a given expected effect for each intervention (Figure 9)."

In the discussion we wrote [P18, L25-28]:

"In studying the effect of uncertainty on intervention design (figures 9 and 10) we linearly interpolated between the high intensity and low intensity interventions. The assumption of linearity is reasonable given the approximately constant Ur;90 between the two intensities. However, in practice this should be only seen as a first approximation. Multiple design iteration would be necessary to arrive at a design that meets the requirements."

We changed and expanded upon this section:

"There are several simplifications done in this paper that merit discussion for a more practical application. First, the linear interpolation between low intensity and high intensity interventions (figures 9 and 10) was used to obtain a first approximation of the uncertainty for intermediate intensity interventions. These figures may be used to illustrate what confidence intervals may be expected for a given range of relative uncertainties, but should not replace a full probabilistic analysis. "

We further note that the variation alongside the river is no source of information for Figure 9, as this is strictly based on model prediction as x_m (the location of maximum effect). To clarify this we changed the caption of Figure 9 from:

"The expected effect (E(ΔH)) against the 90% confidence interval, (…)"

To

"The expected effect (E(ΔH)) at x_m against the 90% confidence interval, (…)"

[R1.7] p17- exceedance probabilities If I understand correctly, this now linearly interpolates exceedance probabilities rather than relative uncertainty. It's not clear why this is a permissible inference. Are there characteristics of the distributions that allow this jump? What makes (all!) exceedance probabilities linear if relative uncertainty is constant with regard to expected effect? I assume at least some kind of symmetry comes into play? Even if the assumption of linearity doesn't end up holding completely, I would be in support of keeping this analysis in the paper, with some representation of the uncertainty in uncertainty involved (as for Figure 9). I think it's very important to be able to say something about the extremes of the distribution, not just about the size of the uncertainty.

[Reply to R1.7] Since the exceedance levels are directly related to the support points on the (cumulative) density function, the implicit assumption in the linear interpolation is that the transformation of the distribution function for low to high is linear with respect to the expected value. This does not necessarily mean that either distribution should be symmetrical, or even that the distributions share the same distribution. Given that we only have two points (low and high intensity), a linear transformation is the only model we can support. To better clarify the purpose of this figure, as well as the uncertainty involved in the estimation, we made the following modifications:

We wrote [P17, L8-L10]:

"For example, figure 10 (a) shows various exceedance probabilities for /flpsmooth/, linearly interpolated between the low-intensity and high-intensity variants as function of the expected effect. "

We changed this to:

"To illustrate how the exceedance probability can be used to guide intervention design, we linearly interpolated between the low-intensity and high-intensity variants as function of the expected effect (Figure 10). In this figure, the estimation uncertainty is not taken into account. Given that we only have two points (low and high intensity), a linear transformation is the only model we can support. Although we acknowledge that other models are possible, we included this as a first approximation, because it illustrates information available for decision makers."

We added the following to the caption of Figure 10:

"Note that all lines show the expected value of the exceedance level; the estimation uncertainty is not depicted."

[R1.8] p19 Discussion It would be useful to mention that the study only considered individual interventions, whereas it would be possible to combine measures in practice. I assume the combined effect is unlikely to be a linear combination of individual effects, so I don't think this study can be used to support any claims regarding such combinations. It would, however, be useful to highlight it as future work and speculate about possible issues that might crop up.

[Reply to R1.8] We agree with the reviewer that a combination of effects would not be linear and added the following section to the discussion (this follows our modification of P15, L25-L28 in response to **R1.6**):

"Second, the interventions studied in this paper are single archetypes, while in reality designs would likely be a combination of multiple archetypes (e.g. side channel combined with lowering of the surrounding floodplain). Claims regarding the combination of intervention archetypes requires further study."

Technical comments

[R1.9] Table 1 The link with subsections in 2.3 could perhaps be emphasised, e.g. by using consistent ordering and grouping terms

[Reply to R1.9] In the manuscript, the grouping in table one was [Classification error, Main channel Roughness, Vegetation Height, Vegetation density, non-vegetation roughness], while the subsections in 2.3 were [main channel roughness, floodplain roughness, classification error matrix].

We now changed table one such that the order and grouping are identical to the subsections.

[R1.10] p7 L14 "general extreme values distribution" Should be "Generalized Extreme Value distribution"?

[Reply to R1.10] That was indeed intended. We changed ‘general extreme values distribution’ to ‘Generalized Extreme Value (GEV, Weibull variant) distribution’

[R1.11] p7 L20 "a 95% confidence limits at 0.31 m and 1.0 m." drop the "a``

[Reply to R1.11] We dropped the “a”, such that the sentence now reads:

“(..) highly asymmetrical distribution with a mean of approximately 0.58 m and 95% confidence limits at 0.31 m and 1.0 m.”

[R1.12] p14 Table 2 Should "coefficient of variation" be average relative uncertainty E(Ur90)? Otherwise, this raises the question of whether relative uncertainty was indeed needed?

[Reply to R1.12] We corrected the caption to “Expected relative uncertainty along the river (average ± the standard deviation)”

[R1.13] Figure 7 y axis label should read Ur90? (9 is missing). Figure 9 also seems to have some text rendering issues

[Reply to R1.13] Unfortunately, the labels on the y-axis in both figures render correct in our PDF readers (showing Ur90, indeed); so we are not sure how to fix this issue. We will take this up if the manuscript is approved for final typesetting.

[R1.14] p18 L17 "greatly the unexplained variance" Word missing (increased?)

[Reply R1.14]

We wrote [P18, L16-L17]:

“They observed that the removal of existing stochastic elements (e.g., a vegetation polygon) in exchange for new ones greatly the unexplained variance (…) “

We changed this to:

“They observed that the removal of existing stochastic elements (e.g., a vegetation polygon) in exchange for new ones greatly increased the unexplained variance (…) “