AUTHOR'S RESPONSES TO REFEREE #1

These are the Authors' replies to comments from Dr. Charles Rougé (Referee #1), received and published on 14 June 2019. We use blue colour for our replies and black colour for Referee's comments.

RESPONSES:

Firstly, we want to sincerely thank Referee #1 for the remarks and recommendations which will undoubtedly improve the quality and scope of the paper.

My main concerns are about 1) the lack of specifics and metrics for the hydrological (and water resource) model validation, 2) the lack of critical analysis associated with the production of risk indicators associated to very high magnitude, very low probability events when only having a comparatively short flow record in hand, and 3) how results can be made meaningful for risk and resilience planning in a changing world.

I would advise authors to prioritise these aspects when preparing a revised version of this manuscript. A general remark is that authors seem keen to follow governmental guidelines and engineering practice, but the purpose of research should be to inform and improve these instead. For instance, research can do so by pointing out the limitations of engineering regulations and practice as a first step towards improving them. There is scope for authors, by engaging in a critical reflection of the assumptions they had to make, to shed light on the limitations of the current guidelines, and on which assumptions are critical. This would be a well-thought-out research publication with real world consequences for engineering practice. This would require little extra modelling work, even though a basic sensitivity analysis would be a low-investment, high-reward endeavour in that it would highlight the parameters that have an outsized influence on the final risk indicators (and this would ease their discussion of which assumptions are critical).

We will do our best to address the issues raised by Referee #1.

1) Calibration / validation is not very precise on how much data there is to calibrate / validate against (how many years?), how the different periods are divided, or what the calibration criteria are. In particular, the quality of the fit (measured with Nash-Sutcliffe or Kling-Gupta) should be disclosed.

The main reason we haven't included a detailed description of the parameters driven the calibration/validation of the hydrological model is that we didn't want to overwhelm the reader with too much information. However, remarks of Referee #1 are very appropriate and will be taken into account when presenting the calibration in Section 5.2.1.

Figure 6: I understand this is simulated vs. observed daily flow from January 1st 2011 to December 31 2015. This should be made clearer, e.g. by inserting specific dates on the x-axis instead of having year labels at the middle of each year. Also on the same figure: why are the calibration results presented only for 2011-2015, a short five-year period? I understand the choice for figure presentation, but I would assume that the model is calibrated / validated against a longer period, and that should be made clear in the text.

Figure 6 was displayed that way to be clearly readable, but also to present the hydrological data that was available for this article. For instance, the Barco de Ávila gauging station contains only 2 periods with valid data:

- From 01/01/1971 to 01/07/1989.
- From 01/10/2011 to 30/09/2015.

Indeed, the x-axis in Figure 6 starts the 1st of October 2011 (first date of the 2011-2015 data period for the Barco de Ávila station) and ends the 30th of September 2015 (last date of hydrological data). This will be made clearer in the revised version.

More importantly maybe, it is not clear how the peak magnitudes in the model and observations match, and this is the crucial part of the flow for dam safety. Would it be possible to plot the differences between the two for days with flows have a certain threshold? Or for annual maxima (since those are the days used to derive the Gumbel distribution)?

The authors agree that this analysis would benefit the clarity of the paper. The revised version will contain an evaluation of the relation between the discharge peak magnitudes in the model and in the observation series.

Likewise, the most important component of the hydrological model validation is whether the behaviour at high storage is reproduced for events during the historical record. In 2001 in particular, simulated reservoir levels are higher than observed levels: authors should understand why that is and what may be the consequences for their model-based risk assessment.

Results shown in Figure 7 do not correspond to the correct version of the model used. Instead, they correspond to a version where the seasonal maximum storage limitation was not yet implemented in the model and hence, they are invalid. The authors want to excuse for this mistake and will update a new figure with the correct series, in which the 2001 simulated levels do not exceed the observed ones.

2) The lack of depth of the analysis of the relationships between modelling assumptions, the uncertainties associated with them, and modelling outcomes is not limited to the calibration of the hydrological and water resource models.

A central observation regarding this analysis is that risks to the dam are fully dependent on the existence of rare events whose probability of occurrence is extrapolated by fitting a predetermined distribution to short (< 100 data points) annual maxima time series. This has several consequences:

(i) The magnitudes of very long return period events will almost certainly be very sensitive to the parameters of the fitted distribution (not to mention the fact that other distributions than Gumbel's exist), so it seems compulsory to quantify the uncertainty on the Gumbel distribution's parameters, and understand the consequences of that uncertainty on the results.

The authors agree with the remarks of Referee #1. A sensitivity analysis of the parameters would highlight how results are dependent on the pre-defined choices made. However, it is important to understand the computation cost with which we are dealing: a complete simulation, from the definition of the Gumbel distribution to the calculation of the dam risk, has an average duration of 24 hours. Thus, the computation duration of a sensitivity analysis (which entails several simulations for each case) applied to the ensemble of the 163 climatic models used would be incompatible with the publication deadlines imposed by the NHESS journal. However, the authors suggest performing this sensitivity analysis to the Base Case (present situation) and analyse the effect on its risk results; this would give an idea on how the other cases would react.

(ii) Similarly, the choice of data (annual maxima vs. peak over threshold) may influence the estimates and therefore, risk indicators.

In this case, the annual maxima method has been arbitrary chosen among the different methods available. This has been selected since is a well know technique worldwide. Please refer to the previous author's comment for the convenience of applying a sensitivity analysis.

(iii) The formula linking hourly and daily intensity (equation (3)) is climate- and location dependent, and the study explores future climates for which the parameters of this formula may change.

This is one of the main limitations of working with daily precipitation data: it is difficult to establish IDF relations when no sub-daily data is available. Thus, in order to deal with this issue, the option chosen was to rely on pre-defined formulations.

Indubitably, the study could benefit from a more detailed analysis capable of producing a time-dependent relation for each climate model. However, this exceeds the scope of the paper.

Nonetheless, a clarification of these issues and a justification of the method chosen will be included in the revised version of the paper.

Authors could improve the value of their manuscript by showing how accounting for uncertainty in the estimates could change failure probabilities. Similarly, while readers can only appreciate the inclusion of gate performance indicators, the consequence of the assumptions of deterioration in performance on the overall results should be clarified (in other words, how does deterioration in gate performance affect future probabilities of failure?).

The effect on risk of gate performance deterioration is displayed in Figure 13. In this figure, the effect of each risk component (Previous pool level, Gate performance, Floods and Social consequences) has been isolated. It is however true that no clear explanation is presented in the text. This will be amended in the revised version of the paper.

3) Maybe it is because the visuals are not conclusive (they often tend to show a risk increases but individual models / scenarios are all over the place), but it is not clear what the results mean for reservoir managers and planners. I would suggest for authors to present the ensemble of climate projections they use as... an ensemble, by deriving mean / uncertainty (e.g. standard deviation?) for each emission scenario and each period. Uncertainty analysis on the parameters that influence risk indicators calculation (see remarks (2)) would make these estimates of ensemble mean and standard deviation more robust and enable them to present their results in a way that can inform decision-making.

We agree this is a key and complex aspect in the exploitation of such results. Although a certain general increase of the risk can be extracted from the results, it is difficult to directly define unequivocal recommendations for dam owners and managers. Different factors play important roles when assessing risk management action plans: Are risk acceptable in present situation? And in future scenarios? What are the risk reduction measures envisaged? How long should we wait until we implement them? What is the efficiency of each of these measures? What criteria should we follow to prioritize them? These are relevant questions that can be mentioned (but not resolved) in the paper. Thus, we will make sure that a more complete overview of the problem is introduced, which will help contextualize the usefulness of such approach.

It is worth mentioning that this is a line of research that the authors are currently following: comprehensive decision-making support based on future changes in dam risk. We invite Referee #1 and readers in general to track the authors’ supplementary articles that explain next steps of the overall methodology and that are under review in other journals.