Throughout this response, the reviewer's text is presented in black, our response in blue, and the proposed revisions in green.

This article provides an evaluation of the consequences of the lack of a representation of reservoir coordination within a multi-reservoir system when simulating flood and drought events in large-scale hydrological models. The model Water Balance Model simulates a multi-reservoir system in USA. The model includes the representation of each reservoir operation policy (using predefined parameters according to each reservoir purpose) but it does not represent the coordination between reservoirs. The global sensitivity analysis Method of Morris is used to assess the effect of the parameterization to the model outputs. Authors conclude that the representation of reservoir policies independently is not enough and that, in addition, we need to capture reservoir coordination in large-scale models to properly simulate flood and drought effects.

We thank the reviewer for their clear understanding of our paper, and more broadly for their thoughtful and comprehensive review.

The article is well written and structured. The Introduction makes a good review of the hydrological impacts of multi-reservoir systems and previous attempts in representing reservoir systems in hydrological models.

We thank the reviewer for their kind words.

The methodology is well defined with the exception a few aspects that need further explanation. The article does not explicitly say where the parameters of the reservoir rules (Table 1) come from. Moreover, the authors do not specify the parameter ranges. If the values in Table 1 were obtained by calibration in a previous work, the authors could show the ranges applied in that calibration or just reference that work. If there is no previous calibration, how the predefined values of the parameters produce a good agreement between observed and simulated storages and releases (e.g. Figure 5) in normal climate conditions?

Thanks for pointing out our lack of explanation of where the parameters of the reservoir rules come from. In a revised version, we will insert in Section 4.2 the following paragraph..

The general form of the reservoir rule was first presented by Proussevitch et al. (2013) and validated using the GRanD database (Lehner and Liemann, 2011). Variants of this rule have been used with a daily time step on the Niger river basin (Oyerinde et al., 2016), and with large-scale assessments using WBM (Grogan et al, 2015; 2017; Zaveri et al, 2016; Liu et al., 2017). The fine-tuning of the parameters when establishing this version of the rule was made using a set of 22 large North-American and Eurasian reservoirs in offline mode., including the two largest reservoirs in the USRB (Palisades and American Falls, daily release NSE 0.70 and 0.60 respectively). Similar to what happens when a reservoir rule that classifies reservoirs by purpose is used in a large-scale model, we did not fine-tune the rule to each reservoir. This allows us to use the reservoir rule in conditions that are similar to what is done in most state-of-the-art hydrological models.

Concerning the obtention of parameter ranges from calibration: we commend the reviewer for their rigor but would like to point out that in the literature on reservoir release rules for large-scale hydrological models (discussed in depth in the Introduction), parameter ranges are not given and instead, single values are suggested. In our diagnosis of these rules, we choose to follow a similar methodology.

The results and discussion are also clear and well structured but there is a lack of discussion of how the methodology applied here can be used by others. I was wondering if this could be done using a different model where the parameters of the reservoir rules are unknown and need to be obtained by calibration.

The goal of this paper is not to propose a diagnosis framework on what constitutes a "good enough" representation of reservoir operations, but to shed light on what unintended consequences of not representing coordination can be. It should be read as a diagnosis of the kind of effects that can emerge with existing representations, if used within large-scale hydrological for flood and drought assessments.

The revised version will insert a subsection in the methodology that clarifies our rationale in the design of the diagnostic analysis, and how the Method of Morris is only part of the analysis. This will also clarify that we are not trying to implement an approach to be used (although we welcome others to carry their own diagnostics) but rather, we want to make a point about reservoir representations in hydrological models.

Lastly, I think that the paper needs further and clearer discussion on why the lack of representation of reservoir coordination is most likely to be the main reason of this failure to simulate flood and drought events.

We agree with this comment. We have written a detailed answer to the same comment by Reviewer's #1 and rather than offering a boiled-down version, we believe it is better to refer to it.

In conclusion, this paper makes a relevant contribution to the growing discussion around the representation of reservoir systems in hydrological models and it has clear practical implications. The authors provide practical recommendations and possible solutions. While the representation of reservoir coordination is still very difficult to implement in models, this study highlights its importance and the need to, at least, consider this limitation when modelling catchments containing reservoir systems under extreme conditions.

We would like to thank the reviewer for their kind words and accurate assessment.

Other comments:

- While the authors provide a justification for the 10% range used for the sensitivity analysis, in my opinion, it would be interesting to also show what range of variation around base values should be applied to properly simulate any of the drought and flood events.

As we discuss above, the intent of this study is not to correct the errors observed but rather, to diagnose them as an artifact of reservoir rule representations using a state-of-the-art

assessment model example in an institutionally complex reservoir cascade. For this diagnosis, the 10% range is adequate in showing how the parametric effects across the reservoir cascade in the USRB are highly complex, interdependent, and dynamic.

- Page 15, Lines 17-19: What if the releases were represented as cumulative releases, the sensitivity would be as consistent as for the storage?

This is correct. However, if we used cumulative releases then at Jackson Lake we would have the exact same and opposite sensitivity indexes compared to storage. This information is captured in the current results.

- Page 21, Lines 2-4: Could you please provide with a possible explanation to this unexpected result?

Lines 5-12 provide this explanation. We will make the link with lines 2-4 explicit in the revision, so it is clear to readers (including reviewers).