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

The manuscript presents a new large scale reservoir operations model with generic operating rules associated with the reservoir main operational purpose such as flood control or irrigation, or default. The reservoir model stands out from equivalent models in that the releases are decided daily based on the daily storage level, shapes with combined log and exponential curves that accelerate the release in times of floods when close to full capacity and slows down the release in times of droughts with differences in the thresholds and propensity to release and store based on the purpose of the reservoir. The overall release is scaled by the long term mean annual flow. The model is implemented at high resolution (.6 km, daily time step) over the Upper Snake River Basin, which is a snowmelt driven basin. The method of Morris is used to identify the reservoir release parameters that tend to be most influential in the reservoir release and storage variations throughout an 8 year period. Upstream reservoirs are used to evaluate the approach while downstream reservoirs are used to evaluate the impact of upstream reservoirs. A flood and drought events are evaluated with respect to observed operations to categorize the error associated with the lack of representation of reservoir coordination. Authors conclude that reservoir coordination is needed to represent flood and drought in typical reservoir models, and that optimization of rules with foresight would help in this endeavor. All simulations were performed on very high performing computational resources taking 2 days for 8 year simulation over the Upper Snake River Basin.

Thanks for this detailed and accurate account of our work. One (very minor) clarification being that each simulation does not take 2 days (see line 9 on page 13).

The subject is very interesting for the HESS community and the manuscript is well written but there are a number of concerns that would need to be addressed before consideration for publication. The main concerns are about the two (great) highlights of the paper : the new model and the time sensitive analytics; i) the manuscript presents a new large scale model, with a very interesting concept for the releases that is however not enough evaluated and discussed, and ii) the approach to quantify the contribution of reservoir coordination to better represent floods and droughts needs to be improved – it is based on inference statements and the model could be modified to include information about upstream reservoir release to demonstrate the point about coordination.

Thanks for this comment.

(i) Regarding the general concern on the model, we would like to point out that the hydrological model is not new, and neither is the rationale for the release rule. In fact, WBM is a well established representative of the broader class of large scale hydrological assessment models that are being used in regional to global applications, as highlighted in p2 lines 19-22 and p5 lines 18-29. As for the release rule itself, it has also been used with WBM in the past, and its precise provenance will be spelled out in Section 2.4 in a revised version.

What is more, the release rule we use is not put forward for its novelty but for being a state-of-the-art representative of the emerging types of rules that are being employed as reviewed in our Introduction. We also insist on this fact when we detail the rule in Section 2.4 (see p. 11 lines 8-15)

(ii) It is not our intent to precisely quantify the exact contribution of reservoir coordination to overall prediction errors, but rather to use an institutionally complex multi-reservoir cascade that is known to exploit high levels of coordination to illustrate qualitative artifactual behaviors that can emerge from the absence of coordination in how large-scale hydrological models abstract major storages.

This being said, we agree that a new kind of release rule for large-scale hydrological models could take into account coordination. The goal of the paper is not to propose such a rule and demonstrate it, but to diagnose the issue to motivate future efforts in the research community. We will clarify that in a revised version, in the methodology and in abstract or introduction.

Minor feedbacks are that the reference to typical reservoir model is misleading and the analytics with the method of Morris is very hard to follow.

We address the reviewer's concern within their point 1) below. As for the analytics, we appreciate the feedback from the 3 reviewers and will make clarifications to interpretability as discussed in our responses here and for Reviewers 1 and 2.

1) Reference to typical reservoir operations model seems misleading. At the scale of the Upper Snake River Basin, typical reservoir operations models have a nodal architecture and represent accurate reservoir operating rules that can be revisited in optimization mode and especially in forecast mode to mitigate reservoir and drought events. The manuscript here refers to very large scale spatially distributed reservoir models that have been developed initially to be fully coupled with hydrology model and research land-surface-atmosphere interactions. Those models are typically applied over multiple independent large river basins. I would suggest to not refer to typical reservoir model where most of the community understand reservoir models where rules can be optimized and are applied to one basin at a time. Please refer to large scale distributed reservoir model or equivalent differentiation from nodal operational reservoir models.

We agree that had the goal been to propose the most accurate operating rules possible for this basin, we would have chosen very different rule forms. However, the goal of our study is to diagnose the implications for how state-of-the-art large scale hydrological assessment models represent reservoir operation rules. We understand that this points needs to be made even more explicit, and the revision will add a subsection to the methodology that will point this out.

As for the reviewer's concern with the geographical scale of our study, we would like to point out that release rules such as they are encoded in large-scale hydrological models must not lead to large qualitative errors in mid-size basins such as the USRB. Indeed, larger-scale studies would contain a set of basins that are not hydrologically connected with 1) a number of mid-size basins, and 2) large basins containing several headwater basins such as the USRB. Therefore, insights from a diagnostic approach on the USRB will be relevant to large-scale assessments if we apply the release rule in the same way (i.e., without over-parameterizing it).

For these reasons, the release rule is applied here as it would be within a large-scale hydrological model, that is, without fine-tuning the parameters to each individual reservoir. This will also be made explicit in the revision.

2) A new large scale reservoir operations model : please provide more details

We would like to clarify that our paper is not focused on a "large scale reservoir operation model". In fact our point is quite the opposite: pointing out the limitations of standard rule-based representations in hydrological models that consider and parameterize reservoirs separately as reviewed in detail in our Introduction.

Revision will explicitly remind the reader that the reservoirs are not coordinated in the model, but only in real-life operations. We will provide evidence of historical operations in the period of record.

- what is the river routing process for this high spatial resolution and daily time step? A recommendation in the introduction is not to aggregate reservoir storage but many reservoirs have less than 2 days in travel time. How does the reservoir model decision release algorithm adjust stability?

Thanks for this. We reply to each sentence separately and in order:

1) Similar to the above, we would like to point out that the release rule presented here does not consider what happens downstream, where the river routing is done by the hydrological model (WBM).
2) Likewise, we are not trying to issue a recommendation of ours when observing in the introduction that Shin et al (2019) recommend not aggregating reservoir storages. That second paragraph in the introduction sets the context, which is the evolution of hydrological models towards use in ever better-resolved models. Our work looks at what happens with a model that is relatively highly-resolved, but far from hyperresolution. At the resolution we use though, it makes little sense to aggregate storages because they are less than two days apart: this would negate the sought advantages of a higher resolution. As we understand the Shin et al paper, the search for better-represented dynamics is the reason for their recommendation.
3) We apologize but we are not sure what you mean by "adjust stability" here. This said, the reservoir rule mechanisms are completely exposed in Section 2.4, so we invite the reviewer to check whether the answer to their question is positive and negative.

- How are the 6 parameters initialized? Are the necessary data widely available? What are the assumptions?

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.

As for the assumptions, we would like to point out that the key one, common to release rules for large-scale hydrological models is that by construction, each reservoir gets separate parameters that do not depend explicitly on the behavior of reservoirs upstream. This will be made explicit in Section 2.4 where the release rule is introduced.

- Evaluation of the smoother release curves with other models. In other equivalent models that are cited (Hanasaki, Doell, Biemans, Voisin, etc) , releases are decided daily based on reservoirs minimum and maximum capacities, minimum environmental flow and tend to follow monthly storage and releases targets with no foresight, but using long term mean monthly inflow, which also tends to be regulated or natural flow depending on the models. What is the improvement for those rules? The obvious features are the changes in release rates - how does it improve the flow representation in general?

The reviewer cites four release rules from four papers (Hanasaki et al (2006), Doll et al (2003), Biemans et al (2011), Voisin et al (2013)) that have release rules that decide a monthly release target in order to analyse outputs with a monthly time step. Given this, it makes sense for the rules to be based on monthly parameters.

Yet, as detailed in our introduction (p3 lines 15-35) the transition of applications to include flood concerns means that shorter timescales have to be considered. The smooth monthly curves are not appropriate for finer time scaled extremes (floods) and for large storages those extremes carryover in the effects on water operations for droughts. So in short, there is tension and difficulty in resolving floods and droughts in complex cascades like USRB using standard rule forms and assumptions of independence between reservoirs.

The revision will clarify that the version of the Hansaki et al rule that is used for flooding is modified by Mateo et al (2014) at p3 line 16 to be used with daily time step. That sentence will be amended in that regard.

Please note that we do not try to use a “best” rule but a rule that shares some key characteristics with state-of-the-art available rules (and we would urge the reviewer to refer to newer representations, e.g. see references in p3, lines 30-35, rather than those they cite here).

- reservoir coordination. Note that the use of a rolling past 20-year of mean monthly regulated inflow provides a minimum of reservoir coordination mostly during extreme events. “Some” coordination is represented through the use of mean monthly regulated flow and also the allocation of water demand to a number of reservoirs based on how full they are. This feature is not present in this model representation, and would likely not drastically change extreme events. Yet it does represent “coordination” around releases and other water management performance metrics than flow and storage, rather coordination on meeting basin-scale water demand. There were statement throughout the paper saying that there was no coordination at all, which seemed then inaccurate and should be clarified.

We thank the reviewer for this comment. It underlines that there are different interpretations of what coordination means. In this paper, we focus on active forms of coordination by human operators, where the dynamics across space and time in release decisions occur in a way that cannot be explained by the immediate or short-term hydrological or climate conditions. A key feature of coordination as understood in this paper is that upstream reservoirs react to and anticipate downstream water issues.

By contrast, the reviewer seems to focus in this comment on two forms of passive coordination, in which reservoirs adjust to varying on-site conditions that are typically imposed on downstream reservoirs by upstream operations. These conditions unfold concurrently to the release decision (inflows influenced by upstream water management, or at-site withdrawals).

We will clarify this difference between our paper’s focus (active coordination) and what is common in release rules in large-scale hydrological models (passive coordination) in the introduction (in the review of the literature) and possibly in the abstract.

Side note on inflows as a means of coordination: the reservoir model uses long-term (5 year) mean (regulated) inflow, and the paper illustrates that this may not be sufficient to represent coordination (we will clarify in the revision what “long-term mean” means in Section 2.4 and Figures 3 and 4) .

- evaluation of the model and transfer to other regions: âA ˘ c whether the coordination ´ between reservoirs was represented or not, how does it affect the vulnerability metrics at the scale of the basin, which is what those models were initially developed for?

We thank the reviewer for keeping their eye on the end-goal here. We agree that large-scale hydrological models are increasingly used to assess water vulnerability. Vulnerability metrics are useful when the difference between a situation and the other is quantitative, but this work is striving to highlight qualitative differences: an adverse event vs. its absence.

This being said, we agree that it is important to put the results into context. For instance, we do explain the consequences of losing control of regulating downstream releases at American Falls in a low-flow event (p 16 lines 29-32). We will add a sentence clarifying that emptying American Falls instantly disrupts irrigation schedules and forces farmers to watch their crops wither.

As for the 2011 flood event, we will add that the peak simulated release at Palisades in Spring 2011 is 50% higher than historical daily release at any point in the last 40 years. The historical maximum corresponds to the 1997 flood, which led to and six counties being declared in a state of disaster and delivered over $11 million in relief by the federal U.S. government (https://www.weather.gov/safety/flood-states-id)

âA ˘ c Most of those models have been developed for application to a wide range of ´ climatology conditions. The model here is applied to a relatively very small basin for its kind. If this manuscript will be used as reference for this large scale reservoir models, it should be either evaluated with respect to other generic rules, or the applicability to larger regions and very different regions should be presented.

Thanks for this comment. Indeed, we are using an example of generic rule that shares key common characteristics with other generic rules; these rules are usually applied to larger hydrological areas.

As explained in a previous comment, it is important to point out what the absence of coordination may lead to if using these rules in headwater basins that would be part of larger-scale studies with large-scale hydrological model.

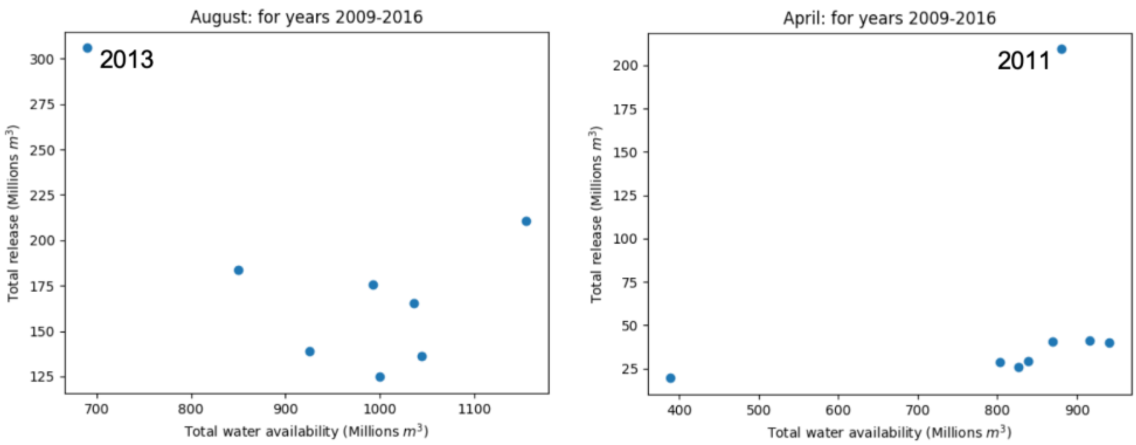
We will clarify this reason for zooming in to a midsize basin either in the introduction or in the methodology section.

3) Evaluation of the contribution of reservoir coordination – artifact of the model? - the main assumption is that the daily releases are based on storage only. All other equivalent models used an estimate of the expected monthly inflow. The main conclusion of the paper is that the coordination between reservoirs should be represented. While I do believe in this conclusion, it seems that the reference to “typical reservoir model” is not justified if the monthly inflow (proxi for foresight without forward running all the models involved) is not represented at all like in other models. My recommendation would be to modify the experiment to evaluate perhaps incremental and simple levels of coordination (aka adding inflow as parameter for the decision release, or a proxi for inflow) to complement the interpretation of the results and provide more quantifiable statements.

We do agree that models that schedule release over a monthly time step also use inflows over a monthly time step. Then, release is naturally a function of available water (beginning-of-month storage plus monthly inflow) makes more sense than determining release as a function of storage alone. Likewise, daily release decisions should be a function of beginning-of-day storage and daily inflows. This rationale of taking the same time step for inflows and releases is common to most release rules for large-scale hydrological models, up to the most recent rules (e.g. Yassin et al, 2019, in this journal). We will precise in Section 2.4 that the rule is separately implemented at the hourly time step in the model, to assimilate inflow into outflow calculations, in order to produce a total release at the daily time step in the model.

Beyond this theoretical reasoning, we conducted another check for added confidence that the results are not an artefact of using monthly expected inflows. In the two events examined in this paper, we produced the figure below, which plots historical monthly release (y-axis) as a function of available water (storage + monthly inflows) for all years of the modeled period 2009-2016 at the most upstream reservoir (Jackson Lake). Results show that historical operations could not be replicated simply by incorporating inflows (even monthly inflows) to the release rule.

For the drought in August 2013 (Section 4.3; left panel) as for the flood in April 2011 (Section 4.4; right panel), the release response is unusual. In the 2013 drought, the reservoir has already been emptied to rescue American Falls levels downstream but releases are still the highest in 8 years, whereas in the 2011 flood, releases are around 8 times as high whereas water availability is similar for 6 of the other 7 years.



We will insert the above figure in the revised version, assorted with an explanation that highlights how the anomalous releases (in August 2013 and April 2011 respectively) correspond to coordination that could not be replicated by simply considering expected monthly inflows.

4) Evaluation of the contribution of reservoir coordination during extreme events I found it extremely hard to follow the text and interpretation of the drivers of the release (annual flow versus objective of this reservoir or upstream reservoirs, and shape of release) by just looking at the figures. Most of the text describes the observed operations and coordination and how the model does not capture it. It is unclear how the method of Morris helps with the interpretation during extreme events. While the visualization is very nice to show the data, it seems that those figures could go in the supplemental material and another figure that compiles those time series and support the text would help.

Thanks for this comment that will help us clarify our manuscript. We believe that we were not explicit enough in explaining the rationale for the method of Morris.

In a revised version, we will explain in Section 3 (methods) that we use the method of Morris mainly to identify temporal “signatures”, i.e. combinations of variables that are concurrently dominant. Then we can study the evolution of these temporal “signatures” through time and through the reservoir cascade, to gain insights into how the model operates.

5) Overall discussion and recommendation versus computational resources needs The authors conclude that foresight should be represented, which is also very sound. Yet the computational resources brought forward for such a relatively small basin are huge which decrease the feasibility at a continental or global scale. Optimization also bring other uncertainties and more computational needs. While the authors seem to indicate that this is what we should do, those were actually drawbacks and motivation for developing those large scale generic models. The recommendation is confusing and perhaps the authors could provide a clarification on new model performances to make it possible now? Please also note that nodal models that typically support reservoir operation optimizations do not provide the spatially distributed feedback into the hydrology model to represent hydrology-land-surface-atmosphere interactions. Maybe the authors meant that we need different types of large scale reservoir models? This would be very sound – just need to be clearer about recommendations then.

We fully agree with the reviewer that there is a research challenge ahead in terms of incorporating human complexity into large-scale hydrological models. We recognize that a first step in addressing this challenge is to signal the consequences of ignoring it, and that is what our paper is trying to do. Clearly, it does not try to provide an easy fix, and while the discussion outlines some ways forward, we do not pretend solutions are going to be easy to design and implement.