Editorial decision

Dear Authors, I would thank you for your patience, as well as careful addressing the comments made by the reviewers. On the recommendation of the reviewers, I would like to recommend the paper for publication. However, some minor suggestions have been made to strengthen the link of the work presented to the large scale studies that consider reservoir operations and the challenges faced. I would think these are reasonably straightforward clarifications that contribute to underlining the novel contribution the paper makes. I am sure you will be able to provide these, following also the quite specific suggestions of the reviewer. In your response please indicate if and how these minor comments have been addressed. I will then review those, before the final recommendation for publication.

Regards and all the best wishes for 2021

Dear Editor,

Thank you for carefully assessing our paper. We have provided responses to the reviewer's suggestions and have added clarifications where appropriate to our manuscript.

We wish you all the best for 2021.

Responses to comments from reviewer #3

Authors have improved the paper with clear and direct statements about the contribution of the paper. Based on the response, I disagree with the authors that the lack of coordination was because researchers did not know any better, but because of a data availability and computational burden this was deemed an acceptable assumption for the time being. I agree with the authors that the quantification of such assumption is timely and of high interest. In view of this clarification, more specific recommendations are provided below to clarify i) why coordination was not ignored but rather seen as a computational tradeoff which is key for further recommendations (not solutions), ii) how the cost of this computational tradeoff is a timely scientific contribution, and iii) that the experimental approach needs to make the link with science questions typically addressed with such large studies rather than only watershed-scale flood and droughts.

Thank you for your time invested in re-assessing our work, and for providing new insightful comments. We really appreciate your time and effort.

Note that in what follows, we use page and line numbers from the new marked manuscript (track changes pdf file).

We agree that the lack of coordination in hydrological models' representations of multi-reservoir operations is non-trivial and addressing the issue does pose significant potential computational challenges. However, science (whatever the discipline) is an evolutionary learning process that must carefully contront when assumptions that were reasonable in previous studies need to be carefully re-examined. The trade-offs between computational demands and model accuracy are also themselves rapidly evolving with new emerging computational architectures, new software development services, and continual innovations in hydrological models' representations of coupled human-natural processes.. In fact, there is no evidence from our careful review of the literature that there has been a broad and consistent effort to revisit the assumption of coordination in multi-reservoir systems. Moreover, the evidence of computational demands precluding future engagement with the issue is not substantiated by our highlighted discussion of the rapidly changing rule-forms or embedded optimization strategies that have already emerged to date. The end of the second paragraph (p 2 L 20-24) cites the two studies that in our understanding, concern themselves with errors introduced by common (non-coordinated) representations of reservoir operations. None of these use the word "coordination", and only Masaki et al (2017) discusses trade-offs between computational efficiency and detailed representations of natural and social systems, before discussing how more accurate representations are needed.

As for the clarifications asked here, we would like to clarify that:

i) There is no substantive evidence in the literature on reservoir representations that the trade-offs between accuracy and computational costs have been carefully evaluated before deciding in any of the many papers that propose ever-more sophisticated release representations. Throughout the paper we prefer to remain agnostic as to why coordination has not been represented even as model sophistication increased. There are in fact many issues that can be considered on equal footing with computational demands such as increased data requirements, challenges in choosing appropriate representations of state-aware adaptive operations, issues associated with conflicting operational objectives, challenges in abstracting the importance of exogenous information such market-based energy prices, etc. Instead, we focus on describing the unintended consequences from this approximation using the WBM illustrative example in collaboration with the model's lead developers.

ii) We appreciate the reviewer's focus on the trade-off between computational cost and model accuracy, and again, we fully agree there is one. However, we are not quantifying it in this paper and are wary of conjecture on the topic. To quantify this trade-off we would need to propose alternative reservoir rules with different levels of coordination, and run the full WBM with them to compare both model behavior and computational cost. Our core supported contribution is the quantitative diagnostic illustration of the importance and unintended consequences of not capturing coordination in complex multi-reservoir systems.

However, we agree that this trade-off will be important to carefully consider going forward, so we inserted the following sentence in the discussion (P 29 L 21-22)

"Approaches to address this need will have to contend with trade-offs between the quality of multi-reservoir operations modeling, computational costs, and data availability (Masaki et al., 2017)."

iii) Our introduction already addresses this via a detailed overview of the literature: how large-scale hydrological models have evolved (both in scope and through their representation of reservoir operations) and where that modeling is going (towards high-resolution models able to forecast and monitor water extremes and their consequences). In that sense, we explain how a detailed diagnostic of commonly used assumptions is timely, before we start seeing sophisticated models that make flawed predictions of the future (on this, please refer to our discussion, especially the paragraph starting P 28 L 31)

Since this paper is not a review, our survey of the literature is focused on the topic at hand. We see no specific evidence we should broaden the scope of our paper / literature review.

Overall:

[Approach to quantity the contribution of reservoir coordination needs improvement]; Authors have clarified the scope with clear statements.

[Conclusion is not novel and forward recommendations are not provided]; Scientific literature should include contributions that challenge previous approaches and results. From the present analytics, it is clear that high resolution model should invest in representing coordination between reservoirs, (alongside better operating rules as well), and evaluation approaches going beyond goodness-of-fit.

Thanks for this assessment.

Most of the analytics focus on how floods and droughts, while large scale studies typically focus more on energy-water-food nexus questions, flows of commodities and virtual flows across regions. From this paper, I would not conclude that previous large scale analyses on energy-water-food are completely wrong, and typically there are disclaimers around extreme events. And it is fair and supported by this analytics to say that coordination between reservoirs might be a strategic research area in large scale studies to address extreme events.

Our study does not pretend to falsify any previous large-scale hydrological modeling effort that dealt with water extremes in some way. As noted in (iii) above, we simply caution against using sophisticated but non-coordinated reservoir representations in studies that account for water extremes.

Accordingly, as noted in more detail below, most of the suggestions are around:
- clarification that operational water models exist and should be used for watershed-scale flood and drought studies

We agree that operational water models exist, but they are not commonly used in large-scale hydrological models and there are good reasons for this (e.g., data availability issues, refer in particular to discussion paragraph starting P 29 L 21). Therefore, these operational models are clearly out of scope here.

- the justification of the USRB as a case study would benefit from including links to energy-water-food questions and expand beyond flood and droughts.

We openly cite energy-food-water questions in our presentation of the USRB (see P 5 L 15). What is more, water scarcity is explicitly linked with water use for irrigation (see Section 4.3)

- a couple recommendations were provided in the discussion section however they represent the same computational and data challenges that led to “ignoring reservoir coordination as an accepted computational tradeoff” until more data are available and modeling platforms allow to run forward looking simulations with objective functions going beyond watershed-scale interest.

Thanks for these.

All pages and lines are based on the marked document which included the responses to reviewers. The actual marked change manuscript started on page 16 (out of 58).

1) Conclusion and recommendations

L17, L26 and conclusion. “there remains opportunities for research to determine which aspects of human management are most urgent to integrate”. I cannot agree more. The conclusion of the paper is relatively generic and a clear statement of the authors concluding that based on their analytics, they identified reservoir coordination as a next priority for research development would strengthen the way this paper can be cited. The conclusion already mentions “high resolution modeling” and “need support from more than water resources modelers" but was not as direct on the emphasis on coordination between reservoirs. While not discussed in this paper, more accurate flows or and more accurate reservoir operations could also be needed.

Thanks for this insight.

2) Choice of the basin and choice of the models to support large scale science questions: P18L11-12: the USRB is presented as a good case study because of its 128 reservoirs and water management known to address both flood and drought events. This is contrasting with large scale studies that have often focused on overall water availability for different water users and regional flow of commodities (electricity, food). The paper might benefit from linking the USRB case study with known regional energy-water-food virtual flows which might be challenged by the way flood and droughts are represented. This would provide a more direct link with large scale studies.

Please note that the third paragraph of the introduction, discussed in your comment, comes after the framing from the two previous paragraphs.

3) Framing
P19, L26. "It is worth noting that all the reservoirs representations discussed above do not account for coordination […] to date there has been not carefully designed diagnostic". While authors also discuss it in the discussion, it would be good here to say why coordination had not been addressed before. The lack of data and the computational needs to represent coordination have been a roadblock. I do not think that this is fair to say that "coordination was assumed non-existent" rather coordination was ignored due to limited data availability and overall treated as a computational trade off. There has been to date no diagnostic quantifying the cost of this computational tradeoffs. It is timely to quantify it in order to inform research priorities as the community advances in high resolution modeling.

As noted in our response to the reviewer's first comment, the literature does not give reasons why coordination has not been addressed before, and we do not pretend to settle that matter. This is simply out of scope in our view.

Note we do not state that "coordination was assumed non-existent", nor do we intend to imply it at any point in the paper. We just point out that coordination is not accounted for, and explore potential consequences.

P21L9-12: "all of these characteristics … flood and drought ….dam failure ….climate change .. make the USRB basin a good case study". I found the paragraph misaligned with large scale studies. For dam failure and reservoir operations under extreme events for decision making at the watershed scale, there are a number of reservoir models for that purpose such as RiverWare and MODSIM (https://www.usbr.gov/research/projects/detail.cfm?id=3669). It seems a good basin for its link to energy-water-food research, with a decent number of reservoirs and a size that allows to carry out this computationally intensive diagnostic (as mentioned on P30L20-24 and could be moved earlier). It seems really important to clarify this point about large scale studies else one would still wonder why you did not consider a simplified version of Riverware with and with coordination to address watershed-scale flood and droughts risks.

We disagree with the idea that the paragraph is misaligned with large-scale studies (as already explained in Response 4 to reviewer 3's comments in the first round of revisions).

This being said, we thank the reviewer for their suggestion, and inserted at the end of Section 2.1 (P 5 L 32-33)

"Any unintended consequences from modeling non-coordinated operations would be a note of caution for large-scale studies featuring water infrastructure balancing protection against water extremes with other competing uses".

P34L20 "calibration of individual reservoirs in a cascade is an approximation modelers make at their own risk". Please substitute to "as a computational trade off" or equivalent.

We toned down by deleting "modelers make at their own risk".

We avoided to refer to a computational trade off because we do not quantify it (and nor has anybody else).

P36L30 (difference in storage levels where 2013 could not recover after 2012). While it can be seen as a lack of coordination, the generic rules does not indicate carry-over storage which would drastically impact the way multi year droughts are simulated. I am not sure that this lack of carry over storage should be associated with a lack of coordination, but rather inaccurate reservoir operations.

Here the reviewer explicitly refers to a passage discussing historical levels (L30 and L31 of that version of the manuscript). The passage describes how historical storage levels did not recover at Palisades in the multi-year drought, meaning that the model simulations did not suffer from a non-representation of a carryover storage rule for which there is no evidence in the historical record at the end of the 2012 hydrological year.

To clarify things further in this revision we changed the first phrase of that paragraph (now starting P 22 L 13-16) into:

"Yet, in 2013 historical storage levels at Palisades (yellow line on panel (b)) had not recovered from the exceptional 2012 drawdown due to a combination of low carryover storage and insufficient snowmelt. Palisades reservoir could no longer supply extra water to the Snake River Plain."

Besides addressing the reviewer's question about carryover storage, this emphasises the sentence and those that follow describe historical observations.

Section 4.4 and 4.3 address flood and droughts. It is important to connect them with science question of large scale studies for water-energy-food questions as mentioned earlier.

Water scarcity in 4.3 is linked to irrigation. What is more, flood protection and water supply directly trade-off as reservoir management objectives.

P45L34-P46L1-3: use of watershed scale hydrology models to get more accurate rules. This statement could have been used earlier when presenting the USRB basin, clarifying that such models exist but you use the large scale representation to provide information to large scale studies. For flood and droughts specifically, and watershed -scale decision making for water users only, those more complex models should be used otherwise.

We feel mentioning these models earlier will cause confusion for the reader, since they are out of scope (as expressed earlier).

P46L11-20 (last paragraph in the discussion section): this entire paragraph is about reservoir operations optimization schemes potentially for large scale models – I do not suggest removing it but complementing it with statements about the fact that data availability and computational burden are already a challenge for large studies. And such complex representations would require even more research to develop objective functions that reflect purposes outside of the scale of the watershed and complex water-energy-food interaction with other regions.

Thanks, we agree and inserted a statement that (P 30 L 8-11): "Yet, this approach is also computationally expensive and needs to use offline water balance models to parameterize parsimonious reservoir rules that can be input into large-scale hydrological models."

Thanks again for your comments, which helped to improve the manuscript.