Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., author comment AC1
https://doi.org/10.5194/gmd-2022-59-AC1, 2022

Author Response to Referee Comments on gmd-2022-59

Danielle S. Grogan et al.

Author comment on "Water balance model (WBM) v.1.0.0: a scalable gridded global hydrologic model with water-tracking functionality" by Danielle S. Grogan et al., Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2022-59-AC1, 2022

Author responses in bold, and original referee comments reproduced here for context.

Reviewer #1

We would like to thank Reviewer #1 for their very thorough commentary, and this has significantly improved the paper.

The model description paper of Grogan et al. describes an Open Source version of the model WBM. Overall, thy provide a well-structured summary of the model. I like to also highlight the availability of data and model source code. However, I also think the manuscript requires some clarifications to be a helpful addition to the scientific community.

Foremost, the abstract and introduction provide no indication of why the model is relevant and how its result already has or will contribute to our scientific knowledge. It is also unclear how this model differs from the vast collection of other global hydrological models. What are the features that make it unique? Why should I be interested as a potential user and scientist to have a closer look? What are the current challenges?

Additional notes:

Previous GMD guidelines stated that the model version needs to be noted in the manuscript title. Please check if that is still the case.

You are correct, we will add the version number to the title.

11: what does long mean? Maybe instead, refer to the first published version in year X

We will make this change to clarify that WBM has a publication history since 1989.

12: So, the previous versions have not included it, and this is a new feature?

While the tracking features have been used in prior publications such as Grogan et al. 2017 and Zuidema et al. 2020, the tracking module has not been described

generally until this publication.

14: I do not think it is necessary to refer to the GitHub link in the abstract. Please instead describe what makes WBM unique and why it is useful. I am halfway into the abstract and still have no idea why I should care about the model

15: Remove unnecessary technical detail in the abstract.

16-17: Ok, so what have you learned? What is the model able to do? Why should I care as a scientist and possible user?

We will update the abstract to remove the GitHub link, reduce technical detail and expand upon what we learned, what the model can do, and why it is useful.

17 - 22: Ok, so this is really interesting, but the sentences are long. If this is a unique feature of this model, it should be stated. In what new ways can we perform experiments with that model that are not possible with other models? After reading only the abstract, it is still unclear why I should care about this model and how it has maybe already contributed to science and will continue to be of interest. What are the scientific questions that it is designed to answer or will enable us to answer in the future? How does it differ from other models? How well or bad does it perform overall / compared to other models? What is the spatial resolution?

The sentences do not appear to us to be overly long, but we will shorten if the Editor advises this. We will add to the abstract some content on what the model can do, its contribution, the ability to capture the hydrological cycle and the spatial resolution.

Introduction: I think you provide an excellent summary of what has been developed. However, I wonder if that should be condensed to a table instead. Half of the test is just references. Also, it would be nice to focus more on why we build these global models and what kind of questions they are supposed to answer, and what they can't do. There are obvious limitations, and people have been criticizing them a lot (sometimes fairly, sometimes not); because of that I think it is essential to highlight the ongoing discussion of what they are and what scientific insights we gained. And specifically, what the remaining challenges are - possibly hinting on your model? How is it different from all the literature that you are outlining?

We appreciate the reviewer's constructive feedback regarding the framing we present in the introduction. While we agree with many of these comments, we do not feel the text should be converted to a table as this does not represent a comprehensive and detailed review of GHMs. Putting it into a table may give that impression.

We will add to the introduction a paragraph discussing some recent model intercomparisons (GHMs with a focus on human processes) to address what they do well and do not do well. We will also include text on how WBM fits into the milieu of other GHMs. For example:

Many of the GHMs were developed to address questions at global and continental scales and these models have been designed to capture the macro-scale behavior of the water cycle in both the natural and human systems (Telteu et al., 2021). In an assessment of six GHMs Zaherpour et al., (2018) found these models did not perform well during low runoff periods and they tended to overestimate mean annual runoff and discharge. Veldkamp et al., (2018) evaluated five of those models and the inclusion of human impacts in these

hydrological models greatly improved river discharge estimates and in most cases lowered estimates of river flow. The human influence on the hydrological system is still in need of development in GHMs and Wada et al., (2017) included better representation of regional water management, co-evolution of the human-water system and improved human water management information as some of the areas to focus on for improvement in hydrological modeling. A large challenge for macro-scale hydrological modelers is to better capture the human decision-making around water movement, use, and consumption and one direction is via linking models from the social sciences to our hydrological models.

The model described in this paper, the Water Balance Model, captures all the major land surface water stocks and fluxes with a focus on human alterations of the water cycle. A significant contribution of this version model is the ability to track water depending on its source or use through the entirety of the system. When compared to global river discharge from similar GHMs, the WBM tends to fall within the range of these other models (Dai and Trenbeth, 2002, van Beek et al. 2011). The consistency with other models of global hydrology, with the addition of component tracking makes it possible to more deeply evaluate the generative processes that lead to the predicted hydrologic fluxes.

These are all questions that can be touched upon in the abstract.

Fig.1: This is very helpful. Could you add the timescales on which these fluxes and storages are simulated?

We will expand the caption to provide additional details including time scales.

Table 1: I think this can be moved to the supplement.

Yes, we agree that the table can be moved to the supplement.

199: This documentation should be appended as supplemental material or uploaded somewhere to provide a doi. If the GitHub repository is lost, this link is not really helpful. This is also the case in various other places in the manuscript.

Yes, we will include the GitHub documentation as part of the supplement (provided this falls within the journal's guidelines).

Fig3: The y-axis is different on the plots and thus confusing. Also, the quality does not seem to be high. Not much to see when zooming in.

We will adjust the figures to a common y-axis range and improve the resolution.

Please also add a comparison to other global models. If it performs worse, state why the model's unique features are still useful.

In response to reviewer's requests, we will be adding a new table to any revision of the manuscript similar to that of Table 6 that compares WBM discharge estimates to those of previous GHMs. Though this table is still in draft form, we note that WBM's estimate of global discharge is in line with other's estimate over the same time period: about 40,000 km^3 yr^{-1} exorheic discharge and 2,000 km^3 yr^{-1} endorheic discharge. These estimates are consistent with those of Sutanudjaja et al. (2018) over the same time period (2000 to 2010), and a bit higher than a variety of studies that modeled epochs between 1960 and 2000, which generally coalesced around 36,000 to 39,000 km^3 yr^{-1}.

Further, I was expecting to see something like Fig 6 and 5 here. Maybe move Fig. 3 to the supplement and refer to the result section.

Figure 3 is part of a section (3.1) which summarizes a number of published evaluation studies using previous versions of WBM. The range of variation in climate drivers, which affect hydrological results, is well known and we do not feel this paper is the place to explore these issues.

Fig.4: Please refer to Table 6.

We will refer to Table 6 in the Fig 4 caption.

Table6: Please add the model name. Is that the absolute difference to the simulation that you are showing? Or the absolute value? The description text is confusing on this matter.

The values presented in Table 6 represent the absolute values of each study's estimate of global irrigation withdrawals. We will adjust the caption as we agree the wording is misleading. We will also add the model names to each row in the table where applicable.

References:

van Beek, L.P.H., Wada, Y., Bierkens, M.F.P.: Global monthly water stress: 1. Water balance and water availability. Water Resources Research 47. https://doi.org/10.1029/2010WR009791. 2011

Dai, A., & Trenberth, K. E.: Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. Journal of hydrometeorology, 3(6), 660-687.

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5-arcmin global hydrological and water resources model, 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018, 2018.

Telteu, C.-E., Müller Schmied, H., Thiery, W., Leng, G., Burek, P., Liu, X., et al. (2021). Understanding each other's 1390 models: an introduction and a standard representation of 16 global water models to support intercomparison, improvement, and communication. Geoscientific Model Development, 14(6), 3843–3878. https://doi.org/10.5194/gmd-14-3843-2021

Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Müller Schmied, H., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, Environ. Res. Lett., 13, 055008, https://doi.org/10.1088/1748- 9326/aab96f, 2018.

Wada, Y., Bierkens, M. F. P., de Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., Konar, M., Liu, J., Müller Schmied, H., Oki, T., Pokhrel, Y., Sivapalan, M., Troy, T. J., van Dijk, A. I. J. M., van Emmerik, T., Van Huijgevoort, M. H. J., Van La- nen, H. A. J., Vörösmarty, C. J., Wanders, N., and Wheater, H.: Human–water interface in hydrological modelling: current status and future directions, Hydrol. Earth Syst. Sci., 21, 4169–4193, https://doi.org/10.5194/hess-21-4169-2017, 2017.

Zaherpour, J., Gosling, S. N., Mount, N., Müller Schmied, H., Veldkamp, T. I. E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J., and Wada, Y.: Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts, Environ. Res. Lett., 13, 065015, https://doi.org/10.1088/1748-9326/aac547, 2018.

Reviewer #2

This manuscript presents the UNH Water Balance Model, a hydrology model that has been developed over several decades, but released publicly (open source code) for the first time.

The authors provide a literature review of the model history and evaluate the current model performance against river discharge observations and irrigation water supply requirements. The WBM performs better across North America and Europe in terms of discharge and irrigation water supply, but relatively poorly across Asia and portions of South America.

The manuscript then provides examples of regional simulations including the Indus River watershed as well as the Wyoming headwaters contributing to river flow to the Colorado, Colombia and Mississippi River. This provides an opportunity to demonstrate the novel water component tracking functionality that enables identification of source regions and source stocks for river discharge and irrigation water supply.

The tracking feature appears to be a significant advance in river flow diagnostics and is capable of determining source spatial regions, and source water components (precip, agriculture, groundwater). This should be a valuable tool for land management and government policy makers. Finally, the authors provide an overview of WBM run-time instructions describing the necessary input data and setup scripts to perform a simulation.

This reviewer was impressed with the breadth of the manuscript included 1) a literature performance review, 2) multi-domain simulations with emphasis on the new diagnostic tracking functionality and 3) an overview of a model setup.

We would like to thank Reviewer #2 for their kind words and thorough commentary on the paper. It will significantly improve the paper.

This reviewer would have liked much more discussion devoted to WBM performance.

We will generate a more detailed assessment of river discharge performance and how this model compares to other similar models. In response to this comment and others, we will be adding a section in the discussion regarding WBM performance with respect to observations and other global hydrologic models.

The WBM had a strong high bias in simulating global irrigation water supply as compared to other studies (Table 6). This apparently was caused by a systematic underestimation of discharge for the China/Asia region, but very little discussion (only a mention regarding better parameters are needed) was devoted to this topic. Whereas the authors provided a comparison against similar hydrology models in terms of simulated irrigation supply, no comparison was provided for discharge rates against other models for better context and perhaps lead to a discussion of what model components or parameters are most in need of improvements. This reviewer would have liked more justification or explanation to describe the skill of WBM such that a new user could avoid certain regions or pay special attention to parameters which are poorly constrained. Also an inclusion/reference of a model tutorial would be helpful for new users to begin interacting with the WBM.

The reviewer, along with others, identified a greater need to focus on WBM model performance in our manuscript, which we plan to accommodate in any revision of the manuscript. We have compiled a table of prior GHM estimates for total exorheic and endorheic river discharge. Though this table is still in draft form, we note that WBM's estimate of global discharge is in line with other's estimate over the same time period: about 40,000 km^3 yr^{-1} exorheic discharge and 2,000 km^3 yr^{-1} endorheic discharge. These estimates are consistent with those of Sutanudjaja et al. (2018) over the same time period (2000 to 2010), and a bit higher than a variety of studies that modeled epochs between 1960 and 2000, which generally coalesced around 36,000 to 39,000 km^3 yr^{-1}. We will also build out our comparison of model fit statistics across space presented in Figure 6 to include discussion about common causes for model regional misfit by GHMs.

The reviewer also suggested the development of a full tutorial. While we realize that this may indeed be helpful for some users, we would like to see if the instruction manual we prepared and released through our GitHub repository is found to be lacking by the community before we endeavour to invest in the development of a full tutorial, which we understand to be quite time intensive.

Line 66: Very nice explanation of the value of this source component tracking feature in this paragraph.

We thank the reviewer for this kind assessment.

Line 85: Global models come an out-of-the-box setup of preferred sub-model structures and parameterizations. There is a table devoted to the key default parameters (Table 2), but no discussion of the optimization process that is required for regional simulations, or representation of the range of parameters to make these regional simulations perform well. The authors do present some discussion and results based upon the contribution of uncertainty due to the climate forcing (Figure 3), but missed an opportunity to discuss the contribution of parameter uncertainty in the manuscript.

Several reviewers expressed an interest in seeing greater detail presented regarding calibration and parameterization strategies, and we would like to provide this general response. Any revision to our manuscript will provide additional detail needed to form a baseline understanding of the parameterization strategy for the model; however, we acknowledge at the outset of this response that there is to date untapped potential for more rigorous evaluation of uncertainty quantification using WBM.

First, following comments from Reviewers 2 and 5, we think that building out Table 2 to include a greater cross-section of parameters commonly adjusted in regional studies is appropriate. We note that this table will be redundant to the WBM_Usage_and_Input_Reference.xlsx spreadsheet on the WBM GitHub page; however, we agree that providing a subset of this reference within the manuscript will improve the readability. Any revision to this table will include default values, reasonable ranges, parameter description, and important citations where applicable. We note that to the extent possible, we have relied on structuring the model consistent with empirically meaningful parameters. As such, values presented in Table 2 will often reflect syntheses of field observations and uncertainty as characterized therein. Other model parameters are more synthetic and have less direct connection to field observation. Many of these parameters have been evaluated through calibration exercises over the years in studies summarized in Section (3.1). The reasonable ranges to be included in any revision to Table 2 will be based on what the authors consider as appropriate starting points for parametric uncertainty analyses based on a

combination of prior experience and physical meaning.

Previous work to calibrate WBM has generally leveraged manual calibration, with several instances of more rigorous calibration attempts. Parameterization of core WBM components evolved through iterative attempts to capture response in both global and regional contexts. Generally, it has been found that parameterization schemes as represented by the default parameter values in WBM_Usage_and_Input_Reference.xlsx reflect reasonable compromises that adequately represent discharge time-series in global simulations, as well as regional contexts focused on temperate, humid, and modestly developed watersheds. We plan to highlight that uniform parameterization can be applied to unique watersheds to capture non-default response (Samal et al. 2017, Zuidema et al. 2018), or that spatially varying parameterizations can capture more finely resolved nuance in watershed properties (Zuidema et al. 2020).

How modular is WBM in terms of testing particular hypotheses about hydrology and competing methods, etc. ? There is some brief description at the very end of turning flags on/off, but no specifics in how this influences representation of hydrology. Given the source tracking capability it would be interesting to test the impact of certain model assumption/hypotheses.

We agree with the reviewer that combined with WBM's source tracking capability, varying the processes represented by the model provide for powerful tools to address hypotheses in regional and global hydrology. Generally, this strategy underlies much of the development that has gone into recent work that leverages WBM, including Grogan et al. (2017), Zuidema et al. (2020), Rougé et al. (2021). We have added new optional functionality throughout the history of development in part because it captured important processes that improved representations. While we acknowledge that such experimentation makes for compelling analyses that show the importance of capturing anthropogenic processes in GHMs (Veldkamp et al. 2018), we consider repeating such experimentation here to be beyond the scope of this manuscript.

Model Description:

Line 140: I am assuming the representation of snow is considered single-layer, and does not include multiple layers. Things like snow properties and albedo are not explicitly taken into account. Also insolation and aspect are not considered within the melting term? Care to comment how this might influence your snow source? Has this ever been validated against gridded snow data sets?

The snow model is a simple single-layer approach with elevation bands providing within grid cell variability (description starts on line 115 in the manuscript). Grogan et al. (2020) corroborated daily model predictions of snow water equivalent to 1034 observation points in the US Northeast. We will update the text to explicitly mention the single layer representation and estimates of the quality of fit to snow water equivalent.

Line 200: "Actual evapotranspiration (AET) from naturally vegetated land areas is a function of the PET, soil moisture, and soil properties."

So rooting depth is not taken into account? What function or purpose is setting the soil moisture pool depth to that of the rooting depth? I assume it's a single layer soil subsurface then?

In response to this question and direct requests from other reviewers, we plan

to include details of the soil water balance calculation in any revision of the manuscript. AWC is a difference between soil field capacity (porosity that can hold water) and wilting point (minimal porosity that plants cannot extract water from). Because it is a relative metric (m/m), it needs to be multiplied by the rooting depth to have it in mm.

Line 210: "While AET from other land cover types (e.g., forest or grassland) can be parameterized and simulated, no published study has yet used this option of WBM. Actual evapotranspiration from other consumptive water uses are described below in Section 2.2.5."

Are the default parameters provided for forest/grassland types to calculate AET, or does the user have to provide them? If this option has never been used, how does the model treat forest/grassland if run globally – which includes forest and grassland cover?

WBM parameters that control all hydrological cycle processes have default values that a user can set and overwrite. This public version of WBM has over 250 of them (see WBM_Usage_and_Input_Reference.xlsx in the GitHub repository). Regarding this particular comment, default parameters that control AET were inherited from Federer et al. (2003). Specifically it includes a flag from the Hamon PET model, and optimized soil drying parameters for the best estimate of Global river runoff and discharge. The default AET settings do not use specific types of land cover and treat land as a generic land cover type, but the user can optionally set AET parameters for each land cover type. Using default settings in WBM is equivalent to treating the land-surface as covered by a reference vegetation (Allen et al. 1996).

Line 275: The term "Shallow groundwater pool" is not used in Figure 1. I assume this is the same thing as "groundwater recharge pool"? If so, make sure the terminology is consistent.

Yes, and thank you for pointing out the need for this correction.

Do the grid cells communicate for surface runoff and subsurface discharge or does this get routed directly to river transport model?

WBM assumes that surface runoff and "subsurface discharge", which we call baseflow, do not interact on the landscape, but both contribute to river flow as independent flow-paths. We will make this distinction clearer in the description of the model.

The term "unlimited unsustainable groundwater source" seems confusing. Is there a better way to describe this fossil ground water?

From the model perspective the water users are tapping into a pool of water that is not explicitly recharged and therefore it is unsustainable, even though this pool is not finite. The concept and the term "unsustainable" has been used in our previous published work (Grogan et al., 2017; Liu et al., 2017).

Several terms for this concept are used in the literature depending on the application, and we make particular reference to the useful table of definitions in Bierkens and Wada (2019). These different terms have been acknowledged in the text (starting on line 289). We selected our terminology as an adaptation to the "physically non-sustainable groundwater use or groundwater depletion" defined by Bierkens and Wada (2019), and our conceptualization is consistent with that use. Because we define a flux of water that satisfies demand unmet by

local resources (i.e. non-sustainable groundwater use), this implies a pool of water from which this water is drawn, which we do not represent explicitly, that is best identified in our use as a tracking descriptor as "non-sustainable groundwater". We adopt "unsustainable" as equivalent to "non-sustainable" because the word "unsustainable" is defined in both the Merriam-Webster and Cambridge dictionaries and "non-sustainable" is not. While we do agree with the reviewer that it is the flux or the use of groundwater that is unsustainable, it is illogical to describe subsequent fates or flow paths abstracted unsustainably as "use". For instance, in describing the fractions of primary water components that drain to the ocean, it is awkward to describe a fraction of this discharge as "unsustainable groundwater use", when we are clearly treating it as representing a fraction of flow within the river system.

We are also unsatisfied with the alternative terminology posed (fossil groundwater) by the reviewer. Again, following definitions from Bierkens and Wada (2019), some terms imply knowledge of the age of recharge to the aquifer that we do not characterize in typical WBM simulations (e.g. recharging 12,000 years before present (Jasechko et al. 2017) in the case of fossil groundwater). Therefore, we prefer terminology that implies no assumptions of the era of recharge.

We acknowledge that the terminology that we have been using in this and our prior papers is not ideal, though it has been deliberated extensively. However, we view it as satisfactory for the time-being and for the purposes we are using it. We plan to clarify our definition of unsustainable groundwater in any revision of the manuscript.

Line 550-555: Here you mention spin up time in the context of water source tracking, but I feel this discussion could come much earlier when describing the model dynamics and features themselves.

We will be adding some further detail to the General Overview (Section 2.1), which will provide a brief introduction to typical practices in model spin-up.

Table 4: Please define 'relict' here.

We agree with the reviewer that the definition of relict water found on Line 856-857 comes too late in the manuscript and should be moved to the area around Table 4.

Because the model is becoming open source, the presumption is to allow for a wider user community. Do you provide a tutorial to familiarize the user with WBM? The authors provide a four-step description towards the end of the manuscript, plus a reference to an instruction manual, but a tutorial would be a great advance.

As stated earlier in our response, we will assess if the instruction manual released through our GitHub repository is found to be lacking by the community.

Model Validation:

Line 620: I recognize this section is devoted to a summary of WBM literature, but it is difficult to evaluate the skill of the model without the context of comparing against other similar hydrology models. There must be some model hydrology intercomparison studies to show here for global river discharge. Certainly Figure 3 might benefit for a comparison against other models.

The reviewer, along with others, identified a greater need to focus on WBM model performance in our manuscript, which we plan to accommodate in any revision of the manuscript. We have compiled a table of prior GHM estimates for total exorheic and endorheic river discharge that shows how this parameterization of WBM results are consistent with other estimates of global discharge during the same model epoch of 2000 to 2010. We will also build out our comparison of model fit statistics across space presented in Figure 6 to include discussion about common causes for model regional misfit by GHMs.

Line 624: What sort of parameter calibration is performed here? Hand tuning, or more formal DA approaches? Are these parameters available for the user?

Throughout the development history of WBM, default values for most parameters have been established through a combination of manual calibration, pre-calibration approaches such as Generalized Likelihood Uncertainty Estimation, and from literature review of applicable properties. The parameter set used here reflects those default values, and no separate calibration was performed for this manuscript. The revised manuscript will build out Table 2 to include a more complete listing of commonly adjusted parameters during regional calibration exercises, and their physically plausible ranges. For a complete listing of parameters, their units, complete descriptions of their behavior and interaction with the model, the reviewer is referred to the WBM_Usage_and_Input_Reference.xlsx spreadsheet on the WBM GitHub page.

Section 3.2.1: Table 2 is good, but a physical definition for the parameters should be stated in the table and not just in text. Perhaps a summary table of parameter values for the literature review manuscripts performed at different regions/resolutions, in addition to the default values.

In any requested revision of the manuscript, we plan to expand the table to include definitions for each parameter, reasonable ranges of parameter values based on literature review or prior experience, units, and parameter definitions. This information will be provided as a subset of the most important and common parameters adjusted, a complete listing of which can be found in the WBM_Usage_and_Input_Reference.xlsx spreadsheet available in the WBM GitHub repository.

Line 690: "Above, we reviewed previously-published WBM validations. As none of the prior versions of WBM code have been released open source, it is important to validate the exact model structure in this first open-source release. "

Was there significant mechanistic changes to these pre-release versions? Was adding the tracking capability the only significant difference from this official release version? From Table 7 it seems like you add some new functionality from previous versions: "Added rainfed agriculture, other land cover types, inter-basin transfers, domestic and livestock water demand", but you don't mention that here, and it seems to the reader that the only change is making it open source, when there are some structural changes.

We propose the following suggestions to clarify any confusion over the collection of processes represented in this release of WBM, and prior studies that we reviewed in Section 3.1. First, to address your question directly, yes, there were 'significant mechanistic changes' compared to many of the prior studies reviewed in Section 3.1. Many of these studies targeted specific research questions requiring new, unique, and narrowly applicable process representation. The open-source release presented here, represents a core trunk of process representation that we consider is widely applicable by itself,

and is a useful starting point or baseline from which deviations in the code needed to accommodate future unique process representation can be described. Therefore, the most significant mechanistic changes are the removal of processes that we consider one-off or too narrowly corroborated by empirical data to distribute as having validity over global scope.

We propose that any revision to the manuscript should more clearly spell out how the core functionality included in WBM (v1.0.0) reflects this core of functionality common to preceding studies (including tracking), despite unique process representation in some of the referenced previous work. Furthermore, we also propose that WBM (v1.0.0) be added as a separate row to Table 7, with an entry that reflects the relationship between the released version and prior work.

Table 6: Appreciated how the WBM irrigation withdrawal estimate was put in the context of other studies. Would it be possible to construct something similar for global discharge? Especially since the author attributes the high irrigation withdrawal estimate (China) to relatively high discharge rates in the Asia domain, it seems like it would be worthwhile to hone in on discharge biases, and diagnose where and what location these are occurring.

Yes, we will be adding an additional table and brief discussion of global discharge estimates from this and other global modeling studies in any requested revision to the manuscript. Regarding your specific comment about a higher irrigation withdrawal prediction than other prior work in China, we would like to point the reviewer to Figures 5c and 6c, which illustrate low biases in discharge in China and southeast Asia.

Line 800: "Global discharge is dominated by rain over most of the globe, with snowmelt an important contributor at the poles, and both glacier runoff and unsustainable groundwater important regionally."

"Snowmelt is an important contributor at the poles": That seems like an oversimplification. Seems that Figure 8 shows there is significant snowmelt contributions well down into the northern mid-latitudes especially for mountainous terrain such as the Rocky Mountains and the Himalayas where a larger population rely on snow runoff for irrigation etc. This should be mentioned.

Thank you for this comment. We propose to modify the text in this section to read as follows:

Global discharge is dominated by rain over most of the globe, with snowmelt an important contributor at high latitudes and high altitudes, with both glacier and unsustainable groundwater important regionally.

Figure 8: Glacier run-off seems to be unrealistic in the SouthWest and MidWest US where none should be occurring. Glaciers are apparently determined solely by snow input and melting algorithm and not prescribed like land cover types?

The Glacier Runoff panel in Figure 8 shows the fraction of glacier water found in river discharge. There are glaciers found in the Colorado and Mississippi Rivers and the meltwater travels downstream. That water is used for irrigation and there will be non-zero fractions of glacier water found throughout these regions.

Figure 11: Perhaps outline the watershed domain of the Indus river in this figure within the larger figure 10 for better reference and perspective.

Thank you for this suggestion, we will include the requested outline in any revision of the manuscript.

Figure 11b: There are no units in this figure. Why is there such a large discontinuity between months 12 and month 1? Why does glacier runoff have such a strong contribution upstream at point A, yet such a small contribution downstream at point B?

We will include units in the figure in any revision of the manuscript.

The plot reflects the large differences in flow between the months of December and January. We note that there are similarly large differences in flow between the months of November and December. Most of the water at the river mouth during December reflects negligible water withdrawals upstream for irrigation as there is no or little irrigation in the month of December.

Point A is located in the foothills of the Himalayas and a large portion of the basin upstream is covered by glacier. Downstream of Point A is the Indo-Gangetic Plain, which contains extensive irrigated agriculture, and large amounts of the glacier water are lost to evapotranspiration. The basin outlet (Point B) includes drainage from several eastern tributaries originating in India. These tributaries are supported predominately by rain water, and from extraction of unsustainable groundwater.

4.2 Return flow tracking

Line 867: I feel like this "relict" and "pristine" distinction should be made earlier, because it is used in an earlier table.

We agree with the reviewer that the definition of relict water found on Line 856-857 comes too late in the manuscript and should be moved to the area around Table 4.

Return water diagnostics and source water diagnostics and tracking seem to be some of the most useful components of the system. Would it be easy to port these 'diagnostics' modules to other hydrology models or is it baked into the system?

The code is deeply integrated into the model, however, the concepts and techniques could be adopted by other modeling groups and are very similar to the techniques used in HBV (Stahl et al. 2017, Weiler et al. 2018).

Figure 12 and 13: Care to comment on why the irrigation return flow waters are so high in the Northern India region? It might help the reader contextualize this diagnostic, and make for interesting discussion of what makes that region so unique.

We will make the following addition to the caption:

The large fraction of irrigation return flow in northern India and Pakistan results from this region being one of the most intensively irrigated regions in the world. Water is used and reused multiple times with low irrigation efficiency (Zaveri et al., 2016).

Figure 14: These plots are interesting and potentially very valuable. Couple questions, in Figure 14b could the color scale be changed to emphasize the 0 to 0.2 discharge fraction? A large part of the watershed domain seems to be the same color of green that is hard to distinguish at all.

Could you not color Wyoming such that the river network can be seen easily? It doesn't seem to be a need to color Wyoming dark blue since it is the headwater region. Also in Figure 14a is this distribution pattern mostly driven by snowmelt? It doesn't look like the distribution is picking up the summer monsoon season where a larger fraction of the water should derive from Arizona and New Mexico. I suppose this depends if the climate forcing captures the SouthWest monsoon.

We did not color Wyoming separately from the rest of the grid cells. All small rivers in Wyoming have 100% discharge originating in the state and therefore they are colored in the darkest blue. In a few cases where rivers enter into Wyoming, such as the North Platte River, they reduce the fraction of Wyoming-sourced water and lighten the color. As more Wyoming-sourced water is added to those rivers, the colors will get progressively darker. We will try to make a clearer figure in any revised manuscript.

Regarding Figure 14a, yes, the large peak in Fig. 14a is a result of snowmelt.

Model Code development section: Although I do find this section interesting it may be better suited for the appendix, such that the reader can immediately go from the results to the discussion. You have provided steps, and there is access to an instruction manual, but do you also provide a simple user tutorial for a cut down domain and provide the input files so the user can familiarize themselves with the steps?

Another reviewer expressed appreciation that this section was included. Therefore, we will consider where this section is best located, in the main text or in the supplement, and we encourage the Editor to provide an opinion. The tutorial request was discussed above.

Discussion:

The tracking capabilities of this model which can attribute discharge source regions, or impact from agriculture on discharge are a great feature and worthy of discussion. The authors refer to the modular nature of the model in order to toggle on/off different features, but they don't provide a test case of this, where for a single experiment, different model structures/assumptions/parameters are switched up to identify the model sensitivity. Just a suggestion.

We do provide a variety of simulation runs showing the breadth of the model's capabilities, and more detailed analyses can be found in the manuscripts cited that make specific use of this functionality (Grogan et al. 2017, Zuidema et al. 2020). We feel that any additional sensitivity analyses using tracking functionality would detract from the focus of documenting the WBM v1.0.0 model, and would be best suited to separate future work allowing us to keep the manuscript to a manageable length.

I was certainly expecting at least more discussion concerning model performance related to river discharge and irrigation water withdrawal as covered in the first part of the manuscript. It was a bit concerning that the WBM was a bit of an outlier when estimating global irrigation water withdrawal, and this was partially attributed by the authors to low discharge rates across China and Asia. Very little to no discussion or explanation was provided for this. The discharge rates seemed to perform relatively well in other regions includes North America, but it was difficult to contextualize given no comparison in performance was provided from other hydrology models.

In any requested revision of the manuscript, we will be drawing out our discussion of model performance as suggested by the reviewers. We expect to

provide a greater number of efficiency measures than the index of agreement, also suggested by several reviewers. We will be adding an additional table, structured similarly to Table 6, that compares WBM's estimate of global discharge to other macro-scale models, which shows WBM's estimate of global discharge to be consistent with other models.

To highlight some of the key points in the discussion of any text revision, we will note that input data dictates the model's ability to capture observed discharge response (Figure 3). The dataset that we use in this analysis (MERRA2, Gelaro et al. 2017) is a reanalysis product that draws on large observational datasets available for the global north, and better simulation performance in these regions likely reflects data available for reanalysis.

References:

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., others, 1998. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. FAO, Rome 300.

Bierkens, M. F., and Wada, Y. (2019). Non-renewable groundwater use and groundwater depletion: a review. Environmental Research Letters, 14(6), 063002.

Federer, C. A., Vörösmarty, C., & Fekete, B. (2003). Sensitivity of annual evaporation to soil and root properties in two models of contrasting complexity. Journal of Hydrometeorology, 4(6), 1276-1290.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate 30, 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Grogan, D. S., Wisser, D., Prusevich, A., Lammers, R. B., and Frolking, S.: The use and re-use of unsustainable groundwater for irrigation: a global budget, Environ. Res. Lett., 12, 034017, https://doi.org/10.1088/1748-9326/aa5fb2, 2017.

Grogan, D. S., Burakowski, E. A., and Contosta, A. R.: Snowmelt control on spring hydrology declines as the vernal window lengthens, Environ. Res. Lett., 15, 114040, https://doi.org/10.1088/1748-9326/abbd00, 2020.

Jasechko, S., Perrone, D., Befus, K. et al. Global aquifers dominated by fossil groundwaters but wells vulnerable to modern contamination. Nature Geosci 10, 425–429 (2017). https://doi.org/10.1038/ngeo2943

Liu, J., Hertel, T. W., Lammers, R. B., Prusevich, A., Baldos, U. L. C., Grogan, D. S., and Frolking, S.: Achieving sustainable irrigation water withdrawals: global impacts on food security and land use, Environ. Res. Lett., 12, 104009, https://doi.org/10.1088/1748-9326/aa88db, 2017.

Rougé, C., Reed, P., Grogan, D., Zuidema, S., Prusevich, A., Glidden, S., Lamontagne, J., and Lammers, R.: Coordination and Control: Limits in Standard Representations of Multi-Reservoir Operations in Hydrological Modeling,

https://doi.org/10.5194/hess-2019-589, In review.

Samal, N. R., Wollheim, W., Zuidema, S., Stewart, R., Zhou, Z., Mineau, M., Borsuk, M., Gardner, K., Glidden, S., Huang, T., Lutz, D., Mavrommati, G., Thorn, A., Wake, C., and Huber, M.: A coupled terrestrial and aquatic biogeophysical model of the Upper Merrimack River watershed, New Hampshire, to inform ecosystem services evaluation and management under climate and land-cover change, 22, 18, https://doi.org/10.5751/ES-09662-220418, 2017.

Stahl, K., Weiler, M., Freudiger, D., Kohn, I., Seibert, J., Vis, M., Gerlinger, K., and Böhm, M.: Final report to the International Commission for the Hydrology of the Rhine basin (CHR), 146, 2017.

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5-arcmin global hydrological and water resources model, 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018, 2018.

Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Müller Schmied, H., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, Environ. Res. Lett., 13, 055008, https://doi.org/10.1088/1748- 9326/aab96f, 2018.

Weiler, M., Seibert, J., and Stahl, K.: Magic components—why quantifying rain, snowmelt, and icemelt in river discharge is not easy, 32, 160–166, https://doi.org/10.1002/hyp.11361, 2018.

Zaveri, E., Grogan, D. S., Fisher-Vanden, K., Frolking, S., Lammers, R. B., Wrenn, D. H., Prusevich, A., and Nicholas, R. E.: Invisible water, visible impact: groundwater use and Indian agriculture under climate change, Environ. Res. Lett., 11, 084005, https://doi.org/10.1088/1748-9326/11/8/084005, 2016.

Zuidema, S., Wollheim, W., Mineau, M. M., Green, M. B., and Stewart, R. J.: Controls of Chloride Loading and Impairment at the River Network Scale in New England, 47, 839–847, https://doi.org/10.2134/jeq2017.11.0418, 2018.

Zuidema, S., Grogan, D., Prusevich, A., Lammers, R., Gilmore, S., and Williams, P.: Interplay of changing irrigation technologies and water reuse: example from the upper Snake River basin, Idaho, USA, 24, 5231–5249, https://doi.org/10.5194/hess-24-5231-2020, 2020.

Reviewer #3

General comments

The paper describes the WBM global hydrological model in detail. WBM is one of the earliest global hydrological models which contributed to the establishing the field of global hydrology. This paper provides the full description of the model together with the development history which will be quite useful for the modeling community. In particular, the water source tracking function is novel and very interesting. The paper is well prepared and mostly very readable. I have only minor technical comments.

We would like to thank Reviewer #3 for their time in providing thoughtful comments on this paper.

Specific comments

Line 189 "Soil moisture balance calculations for natural landcovers are fully described in (Wisser et al., 2010a) and crop landcovers in (Grogan, 2016).": Better to show the essence here because soil moisture balance calculation is the most fundamental function of any hydrological models.

We can add a brief introduction of the soil moisture balance calculations that are present in the model's documentation in the main body of the manuscript. We have been reluctant to add to the length of the manuscript, and this functionality is well documented elsewhere; however, we do concur that it is fundamental to this hydrologic model.

Line 270 "PyGEM's standard output format is not gridded; rather, post-processed PyGEM output is required as input for WBM (Prusevich, et al., 2021).": How frequently is the glacier fraction updated (e.g. daily, monthly, annually)?

Thank you for catching that. Glacier water (glacier runoff) is updated at monthly timesteps in the model in accordance with the source PyGEM glacier point data. Some PyGEM variables such as glacier volume and area are updated at annual timesteps so WBM also updates those layers at the annual timestep. We can update the text to clarify details on glacier water input in any revision of the manuscript.

Line 358 "Rather, they collect rainwater and surface runoff, storing it on the land surface and preventing it from reaching the rivers system": How are these processes formulated? What are the key inputs and parameters?

Greater detail regarding how these small reservoirs are handled is available in the technical documentation that we plan to include as a supplement to the manuscript if it is consistent with the Journal's practices.

Line 368 "WBM's inter-basin transfer methods were first developed and described in (Zaveri et al., 2016) and described again in (Liu et al., 2017).": Can this inter-basin transfer scheme be applied to global simulations? If so, how the parameters were set (i.e. is such information available)?

An IBT database was developed for specific publications (e.g. Zaveri et al., 2016), however a global version has not been released. This functionality is available to users if they develop tables representing inter-basin transfers with further details available in the technical documentation.

Line 400 "Stream water available for extraction is estimated as 80% of water retained in river and reservoir storage following routing during the previous time-step W^{k-1}, plus the volume, Vstream, represented by flow through the reach during the previous time-step:" A bit hard to read and associate with Equation 26. What is Vstream? Is this representing the available surface water?

We now see this sentence is confusing and we will rework it in an effort to make it clear. V_{stream} is the total volume potentially available for extraction from the prior time-step.

Line 621 "The global simulations described above used a grid cell resolution of 0.5

degrees.”: This should be mentioned in the previous paragraph.

Yes, we will move the resolution to the previous paragraph.

Line 629 “These continental-scale simulations of India used the same 0.5 degree spatial resolution as the global simulations.”: What were the input meteorological data used in these simulations? The performance of river discharge simulation is largely dependent on the quality of input meteorological data (e.g. Hanasaki et al. 2022, HESS).

The Zaveri et al (2016) paper used the Asia-specific APHRODITE climate drivers. We will amend the text to clarify this.

Line 721 “We also calculate the Index of Agreement, d, (Willmott, 1981)”: Why was this indicator chosen? I recall that most of the earlier works used NSE.

For brevity, we selected a single efficiency measure, d, in addition to a measure of model bias. Several researchers have pointed out short-comings of any single efficiency measure including the Nash-Sutcliffe Efficiency (Krause et al. 2005, Knoben et al. 2019), which tends to be insufficiently sensitive to systematic biases. Still, in deference to this comment and other similar comments by other reviewers, we should amend our presentation with reports of additional efficiency measures, likely to include NSE and the Kling-Gupta Efficiency (Gupta et al. 2009).

Line 740 “Despite the global average good agreement, there is significant spatial variability, with lower MBE values across much of South America and East Asia (Figs. 5c and 6c).”: When one looks at the absolute MBE, the performance of river discharge simulations in arid or semi-arid regions always appears to be "good" because the runoff is very small. This needs to be pointed out in the text.

Thank you for pointing this out. We will include this in the text. In addition, we are considering adding a relative bias metric that would highlight any issues with model misfit in arid and semi-arid regions.

Line 750 Figure 6: What is the difference between Figure 5 (c) and 6 (c)? Only the unit is different?

Yes, the figures are showing the difference between daily and monthly metrics as identified by both the units, and in the captions of the two figures. We include both as many researchers running global simulations use monthly outputs in their papers. We also include daily metrics as the model runs at this time step.

References:

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hanasaki, N., Matsuda, H., Fujiwara, M., Hirabayashi, Y., Seto, S., Kanae, S., and Oki, T.: Toward hyper-resolution global hydrological models including human activities: application to Kyushu Island, Japan, Hydrol. Earth Syst. Sci, https://doi.org/10.5194/hess-2021-484, accepted, 2022.

Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrology and Earth System Sciences 23, 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Adv. Geosci. 5, 89–97. https://doi.org/10.5194/adgeo-5-89-2005

Zaveri, E., Grogan, D. S., Fisher-Vanden, K., Frolking, S., Lammers, R. B., Wrenn, D. H., Prusevich, A., and Nicholas, R. E.: Invisible water, visible impact: groundwater use and Indian agriculture under climate change, Environ. Res. Lett., 11, 084005, https://doi.org/10.1088/1748-9326/11/8/084005, 2016.

Reviewer #4

General comments

The paper "WBA: A scalable gridded global hydrologic model with water tracking functionality", by Grogan et al. provides a description of the first open source version of the University of New Hampshire Water Balance Model. The authors chose an approach that combines parts of a "classical" model description – i.e. description of the functionalities and fundamental equations, validation and a selection of case studies – with a literature review of the history of the model, previous studies with WBA and validation of previous model versions. Overall, this structure works really well and the paper is well written, thus, I only have a few minor suggestions.

We thank the reviewer for taking the time to carefully read and review this paper.

Minor comments

1) With respect to the model description, the authors did a very good job at providing a general overview over the basic equations and dependencies in the model without overloading the manuscript with technical details (which is perfectly reasonable given that most WBA components have been used in previous studies and have been well documented). However, it would be extremely helpful if the structure of the section 2.2 could be related to what is shown in figure 1, i.e. that all elements that are shown in the figure are discussed in the model description, preferably even in a way that each element in the figure has a subheading in the text.

We will work on harmonizing the figure 1 labels and the section titles to improve clarity.

2) I find the use of the term "unsustainable ground water" somewhat problematic, since it is not the groundwater itself that is unsustainable but its use e.g. for irrigation. A term that clearly states either which real world pool is represented – e.g. fossil water – or what it constitutes in the model – namely an unlimited water supply to balance demand-supply mismatches – would be more appropriate. Maybe the authors could also add some discussion to section 2.2.2, detailing how the use of this pool affects simulations (especially projections) in regions where fossil ground water is being depleted.

From the model perspective the water users are tapping into a pool of water that is not explicitly recharged and therefore it is unsustainable, even though this pool is not finite. The concept and the term "unsustainable" has been used in our published papers (e.g. Grogan et al., 2017; Liu et al., 2017).

Several terms for this concept are used in the literature depending on the application, and we make particular reference to the useful table of definitions in Bierkens and Wada (2019). These different terms have been acknowledged in the text (starting on line 289). We selected our terminology as an adaptation to the "physically non-sustainable groundwater use or groundwater depletion" defined by Bierkens and Wada (2019), and our conceptualization is consistent with that use. Because we define a flux of water that satisfies demand unmet by local resources (i.e. non-sustainable groundwater use), this implies a pool of water from which this water is drawn, which we do not represent explicitly, that is best identified in our use as a tracking descriptor as "non-sustainable groundwater". We adopt "unsustainable" as equivalent to "non-sustainable" because the word "unsustainable" is defined in both the Merriam-Webster and Cambridge dictionaries and "non-sustainable" is not. While we do agree with the reviewer that it is the flux or the use of groundwater that is unsustainable, it is illogical to describe subsequent fates or flow paths abstracted unsustainably as "use". For instance, in describing the fractions of primary water components that drain to the ocean, it is awkward to describe a fraction of this discharge as "unsustainable groundwater use", when we are clearly treating it as representing a fraction of flow within the river system.

We are also unsatisfied with the alternative terminology posed (fossil groundwater) by the reviewer. Again, following definitions from Bierkens and Wada (2019), some terms imply knowledge of the age of recharge to the aquifer that we do not characterize in typical WBM simulations (e.g. recharging 12,000 years before present (Jasechko et al. 2017) in the case of fossil groundwater). Therefore, we prefer terminology that implies no assumptions of the era of recharge.

We acknowledge that the terminology that we have been using in this and our prior papers is not ideal, though it has been deliberated extensively. However, we view it as satisfactory for the time-being and for the purposes we are using it. We plan to clarify our definition of unsustainable groundwater in any revision of the manuscript.

3) I think it is a good idea to discuss existing model validation in this paper, rather than repeating the respective simulations with the present model version. However, it would be helpful if the authors could detail if and how the present model version differs to the model versions used in the previous studies and how these differences affect the results. Furthermore, with respect to the FrAMES model (component) I find the model validation a bit out of place here, as it is not merely a different version of WBA but a completely different model. I would expect the performance of the implemented functionalities to depend on many other aspects of the model and the forcing data, hence different simulated nitrogen concentrations with WBA. I think it would be sufficient to state, in the model description sec. 2.2.6, that WBA now includes these functionalities based on the parametrizations of the FrAMES model and reference the studies in which FrAMES was validated. However, I think it would be even better if the authors could actually perform the analysis and validate N and temperature in WBA simulations.

The reviewer requested that we provide details of how the current version of the model differs from prior versions of the model. We provide a table describing the evolution of the WBM family of models as Table 7 and discuss this evolution in greater detail. We can add a comment in this section to make the reader aware that there is a section devoted to describing the evolution of the model; however, we decided that bringing such a discussion earlier in the manuscript before the validation section detracted from the flow of the document. The reviewer also asked us to comment on how differences in the current model and

prior implementations of the model could affect results; however, we view this as an impossible request. Each study referenced in this section had specific motivations, input data, and simulated processes. We do not see what utility would be gained by comparing results from different applications of the model, as it would be impossible to attribute differences to model structure, input data, or parameterization without a rigorous and controlled series of experiments. We consider such an experiment beyond the scope of this manuscript, and that this manuscript be an important first series of steps to such an effort.

The older versions of the model, WBMplus and FrAMES, are closely related to the model version described in this paper. Despite the different name, FrAMES is not a completely different model, but it was built from WBMplus with the ability to simulate water quality (including temperature) incorporated into the code. The current WBM version used both as a guide with extensive internal model coding to better allow for human use of water, an improved ease of coding, and increased run time speeds. Verification of equivalent results was paramount when developing the current version of the code. Therefore, the older validation analyses are very much applicable to the current model. However, we will incorporate a statement in the paper to clarify this information to better justify the use of these prior validation exercises.

We would also like to respond to the reviewer's comment or request to repeat simulations of temperature or dissolved inorganic nitrogen (DIN) with WBM v1.0. We debated this extensively in the development of this manuscript and decided that DIN functionality within WBM is limited to regions with ample observational data. We are planning to provide a brief description of the methodology used to parameterize the model following Wollheim et al. (2008) in any revision of the model documentation, and more clearly describe the limitations of this functionality. We find it beyond the scope of this manuscript to apply this functionality to the two simulation domains presented and consider the addition of a third simulation domain more likely to add more confusion than clarity to the manuscript.

Specific comments

Line 22: " ... as well as perform model experiments in new ways". Please, clarify which are these new ways.

Yes, this statement is ambiguous. We will be more specific in the abstract as to what these are: tracking throughout the water system, including tracking attributes such as state of origin.

Lines 55 ff: Evapotranspiration will eventually lead to precipitation and a large fraction of the respective water is even recycled locally. Thus, the statement that only 50% of water is returned to "the system" is somewhat misleading. In contrast, when talking about specific pools in the system a 50% return rate is also often questionable, e.g. in case of fossil water, at least on a centennial timescale.

Without an atmospheric model WBM is not able to assess the degree of local recycling of evaporate. However, we will improve the statement in the text by clarifying that we are referring specifically to direct returns of liquid water to shallow flow paths in local watersheds.

Fig. 1: Would it be possible to make the elements of the figure consistent with subheadings in section 2.2.? For example infiltration is not specifically discussed in the text.

We will work on harmonizing the figure 1 labels and the section titles to improve clarity.

Line 166: Eow is not defined.

Thank you for catching that oversight. E_{ow} is open water evaporation expressed in mm/d and we will add this to the text.

Line 170: "Storm runoff" is this the same as "stormwater runoff"?

The term "storm runoff" provides a useful semantic distinction from "stormwater runoff". We view the collective understanding of stormwater runoff to refer to immediate runoff from impervious surfaces connected directly to water bodies via built infrastructure. While we include such a flux in WBM, we add two other fluxes to this variable in our internal calculations: 1) precipitation incident directly to open water within each pixel, and 2) water that overfills the surface flow pool (R_{exc}, Equation 13).

Line 179: What does WBA do in these grid cells e.g. in case of endorheic basins?

The endorheic basins accumulate water at their "outlets" in a dedicated endorheic lake storage pool and we will update the text to explain these details.

Line 183 f: Is this the only limit on infiltration? Is the state of the soil not taken into account?

Yes, we do not simulate explicit Hortonian runoff. In the case that the soil is already saturated, any throughfall will be split between the quickflow and recharge flow paths. This may present a localized low bias in runoff for extreme precipitation events, but we consider it a suitable simplification for a macro-scale model.

Line 190: It may be helpful to mention that WBA does not have soil layers and does not explicitly represent the vertical flux through the soil or a soil moisture profile.

WBM has a single soil layer. Section 2.2.1 Land surface fluxes discusses the soil layer (defined by rooting depth on line 187) and section 2.2.2 Groundwater discusses the below-soil storage pool (shallow groundwater storage pool). We will correct the labelling in Figure 1 to reflect these terms more accurately. Any revision of the manuscript will provide a brief description of soil water balance within the root zone adapted from previous publications and our technical documentation.

Line 257: How do you justify this default value of 1000 mm, i.e. that the model, in the default mode, has no real limit to the surface storage?

You are correct, the default behavior is to effectively turn the process off. This default keeps model behavior more in line with prior model usage (from early Vorosmarty papers through Grogan et al. 2017), until at such time that we have experience with the parameter to provide better direction to users.

Line 280: I am a bit confused by the unit l/d is that per m^2?

We plan to clarify the text about calculation of baseflow drainage from the shallow groundwater pool. The shallow groundwater pool stores space-averaged groundwater in units of mm. The time-constant drains water from the pool with

units of 1/d, yielding a flux to the stream in units of mm/d. We will adjust the text to clarify this.

Line 490 ff & 503 ff: Is there a lag connected to the return flows?

There is no lag, and we can revise the text in any requested revisions to clarify.

Line 564: I am not familiar with the term "relic water", so I am not sure whether some definition is necessary.

Relict water is defined in lines 854-859 along with a citation to Zuidema et al. (2020) where we first used the term in reference to tracking return flow. In response to other reviewers the definition of relict water found on Line 856-857 comes too late in the manuscript and should be moved to the area around Table 4, and this may help alleviate reader's confusion at this point in the manuscript.

Fig. 2: What is the meaning of the colors? Also why is the down-stream cycle different (no subheadings in "sources" and "water")?

The diagram was designed to show that the same processes are applied to the water as it moves downstream. The second "Local Cycle" box used simplified labels to reduce complexity in the figure and the third cycle is represented by a "…". We will work with the figure to make it clearer and/or add in explanatory text to the caption.

Section 3.1: Could you maybe add a table for a quick overview?

While we acknowledge that there are numerous numeric values and citations presented in these paragraphs, we have reservations about presenting this information as a table. As a table, the citations and numeric values are likely to present a firmer or more concrete assessment of model performance than we are trying to communicate. Much of the nuance and critical aspects of the findings of these studies are presented in the text, and the values are properly contextualized with the text.

Line 611: What about the UDEL climate? In Fig. 3 The R2 looked very promising?

Thank you for pointing out that the UDEL climate was not mentioned. We will correct that oversight.

Fig. 4: Maybe use the same axis for subfigure b and c?

The panels in Figure 4 are from an existing publication, however we will recreate the figures with adjusted axis ranges.

Line 671: While I think it's a good way to use existing validation, I am not sure about the FrAMES model, as the respective formulations lead to a very different outcome in the WBA framework?

The similarities between FrAMES and WBM v1.0's river temperature and in-stream nitrogen modules are essentially identical making the evaluations from the earlier papers appropriate.

Line 725: Could you also include R2 to make it easier to compare the present simulations to those in section 3.1 ?

In response to this and other reviewers' comments, any revision of the manuscript will include additional efficiency metrics beyond the index of agreement.

Line 725: I would be very curious if you could also include an evaluation of the simulated evapotranspiration … maybe against GLEAM data?

Evapotranspiration data is sparse and applies to scales much finer than resolved by the large grid cells used in this paper (0.5 minute cells). The GLEAM data uses a Priestley-Taylor approach and therefore any evaluation would be a comparison between the two PET functions. A comprehensive PET comparison in the WBM context can be found in Vorosmarty et al. (1998).

Fig. 5 & 6: Why is there no Index for the Nile/Indus/Ganges in subfigure c ? Also, would a relative measure make more sense than MBE.

We used Global Runoff Data Centre (GRDC) river flow stations for these maps and the GRDC relies on all nations to contribute their data. In some international basins countries are reluctant to share this data and therefore only a limited time series is available. This is also the case for many stations in Asia. For the Nile, the data range from GRDC covers 1973-1984 while the model runs presented here ranged from 2000 to 2009.

Relative bias metrics would provide an additional metric for performance evaluation, but we argue that it would be inappropriate to substitute the absolute measures for relative ones. The absolute values provide some indication of total errors in estimating global discharge, and where errors are located relative to global discharge. We acknowledge that WBM exhibits lower performance in more arid regions, where relative errors would be substantially larger due to both model error and substantially lower discharge. A plot of relative bias would be expected to identify substantial errors in areas that are not contributing significantly to global discharge. However, such a plot would also identify critical simulation errors in regions that are more sensitive to water scarcity. We expect to add a plot in addition to Figure 5 and 6 to illustrate a relative bias metric.

Fig. 8 & 9: Could you do such a figure also for evapotranspiration?

Unfortunately, our tracking system for evapotranspiration is not sufficiently complete to provide equivalent panels to Fig. 8 and 9. We do thank the reviewer for pointing this out, and we have added this to our model development to do list.

Fig. 10: I find the purple and blue colors are very similar, and I am not sure that its only an issue related to my printer.

We printed Figure 10 on an eleven year old HP Color Laserjet CP2025 and the colors are very distinguishable. We will try some other palettes to identify a better color scheme. Perhaps the journal could provide some guidance if they perceive this as an issue.

Line 906 ff: "… published in (Vörösmarty et al., 1989)". I would not use the brackets here.

Yes, that is a typo and it will be corrected.

References

Bierkens, M. F., and Wada, Y. (2019). Non-renewable groundwater use and groundwater depletion: a review. Environmental Research Letters, 14(6), 063002.

Jasechko, S., Perrone, D., Befus, K. et al. Global aquifers dominated by fossil groundwaters but wells vulnerable to modern contamination. Nature Geosci 10, 425–429 (2017). https://doi.org/10.1038/ngeo2943

Grogan, D. S., Wisser, D., Prusevich, A., Lammers, R. B., and Frolking, S.: The use and re-use of unsustainable groundwater for irrigation: a global budget, Environ. Res. Lett., 12, 034017, https://doi.org/10.1088/1748-9326/aa5fb2, 2017.

Liu, J., Hertel, T. W., Lammers, R. B., Prusevich, A., Baldos, U. L. C., Grogan, D. S., and Frolking, S.: Achieving sustainable irrigation water withdrawals: global impacts on food security and land use, Environ. Res. Lett., 12, 104009, https://doi.org/10.1088/1748-9326/aa88db, 2017.

Vörösmarty, C. J. and Iii, B. M.: Modeling basin-scale hydrology in support of physical climate and global biogeochemical studies: An example using the Zambezi River, Surv Geophys, 12, 271–311, https://doi.org/10.1007/BF01903422, 1991.

Vörösmarty, C. J., Moore, B., Grace, A. L., Gildea, M. P., Melillo, J. M., Peterson, B. J., Rastetter, E. B., and Steudler, P. A.: Continental scale models of water balance and fluvial transport: An application to South America, Global Biogeochem. Cycles, 3, 241–265, https://doi.org/10.1029/GB003i003p00241, 1989.

Vörösmarty, C. J., Federer, C. A., and Schloss, A. L.: Potential evaporation functions compared on US watersheds: Possible implications for global-scale water balance and terrestrial ecosystem modeling, Journal of Hydrology, 207, 147–169, https://doi.org/10.1016/S0022-1694(98)00109-7, 1998.

Zaveri, E., Grogan, D. S., Fisher-Vanden, K., Frolking, S., Lammers, R. B., Wrenn, D. H., Prusevich, A., and Nicholas, R. E.: Invisible water, visible impact: groundwater use and Indian agriculture under climate change, Environ. Res. Lett., 11, 084005, https://doi.org/10.1088/1748-9326/11/8/084005, 2016.

Zuidema, S., Grogan, D., Prusevich, A., Lammers, R., Gilmore, S., and Williams, P.: Interplay of changing irrigation technologies and water reuse: Example from the Upper Snake River Basin, Idaho, USA, 24, 5231–5249, https://doi.org/10.5194/hess-24-5231-2020, 2020.

Reviewer #5

Overall comments:

The authors present and describe a new open-source version of the global hydrologic model WBM, emphasizing new capabilities for tracking water sources. The paper is well written, and the overview is fairly comprehensive, including theory, examples, plentiful references to earlier literature, and a discussion of how this open-source version relates to other versions of WBM that have been used over the 3 decades since the first version was created. Not only is the model now open source, but the authors have provided a Singularity container to simplify access/usage. Overall, this is a nice contribution, and I recommend publication after minor revisions.

We appreciate the reviewer's endorsement for this manuscript. We thank the reviewer for making numerous straightforward editorial suggestions that we agree with. In the absence of a direct response, please assume that the correction will be made in any subsequent revision of the manuscript.

Below are listed specific comments keyed to particular line numbers, sections, or equations:

Introduction: I appreciate the overview of applications of GHMs, which seems like a useful entry point for those new to the topic.

Though we appreciate this reviewer's comments on the introduction, any revision of the manuscript will include refinements of the introduction based on other's comments; however, we aim to maintain the general overview we present in the initial submission.

Sec 1.1: the need for water tracking is well motivated here.

69-70 typo?

It is and we will fix it.

86-88 at some point around this section it would be useful to describe how WBM handles gridding. Here an example is given of a fixed-width (120 m) grid, and later examples are noted of lat-lon based grids. Does the user have a choice between these? Does the model account for the variable size of fixed-longitude grid boxes? What happens at the poles? Or does the global configuration exclude very high latitude regions like Antarctica?

We appreciate the reviewer pointing out this important but absent consideration in our model description. We will provide an appropriately detailed description of WBM's grid cell representation in the General Overview section. To briefly address the reviewer's concern, WBM utilizes underlying raster representations using the Geospatial Data Abstraction Library (GDAL) (GDAL/OGR contributors, 2022) and can theoretically use any raster grid types (projections and resolutions) that can be represented by GDAL provided a user enters a flow direction file for those grids. Commonly we have used geographic rasters, and a layer of grid cell area is precomputed for all calculations relying on area using the Haversine formula. Therefore, for geographic rasters, errors in area estimation are most significant close to the poles, where polar projections are more appropriate.

124 and ff: thank you for listing units of each variable.

Eq 3 & 4, should this be P^e? (also, in general, instead of introducing equations with 'defined according to', it can be helpful to say something more descriptive like 'so-and-so depends on temperature T and precipitation rate P according to')

The equations are correct as written. We do not have elevation dependency of precipitation represented in the model and apply a uniform precipitation rate over each pixel. Therefore, we only identify elevation differences in the forms (frozen or rain) of precipitation. We will look to provide more descriptive introductions to variables in our equations as you suggest in any revision of the manuscript.

Eq 4 and other math: if you want text-like typesetting, e.g., the word "if" in eq 4 or a sub/super-script like "max", use \text{if}, W_i^\text{max}, etc. (requires

\usepackage{amsmath})

We will make these formatting revisions as requested.

142 grammar around lapse rates

We see how the list structure of this sentence is a bit awkward, and we will revise to read more smoothly.

158 is Pt a user-defined param? A fixed fraction of P? Calculated in some other way?

P_t represents the throughfall flux calculated in the subsequent equations. We can add an additional equation showing how this variable is calculated in a revised manuscript; however, this equation is found in the technical documentation.

Eq 7 and others: consider the more traditional use of a dot (\cdot) or no symbol at all to represent multiplication, as opposed to an asterisk (which I think traditionally means convolution, even though most modern programming languages use it as a multiplication operator)

Agreed.

212 no cap

eq 13 ff: this is somewhat confusing because the phrase 'immediately moved' suggests a discontinuity but the differential equation suggests differentiability / continuity. Please clarify (maybe via a delta operator in front of R_EXC, which is 1 if volume of retention pool exceeds the threshold and 0 otherwise?)

We agree that equation 13 could be presented more clearly. The reviewer is correct in that a dirac delta would be more appropriate in equation 13 to make it consistent with the conditional formulation of equation 18. The delta would yield an impulse at times when the upper limit of the surface runoff retention pool is exceeded, and the magnitude of the impulse would be equal to the volume in excess of the pool's limit times the time-step length.

250ff isn't there a unit mis-match between W and R in eqs 14-18? And between R and T in 18?

There is indeed a unit mismatch. Each R value in these equations is a flux that is integrated over the timestep and should thus be multiplied by dt. We will make this correction in any revision of the manuscript.

286 unclear what 'stock' means here (it seems to be a standard term with WBM, so please define it before using)

We use the term stock interchangeably with the term pool to denote a control volume. The term stock is widely used in ecological modeling.

Sec 2.2.3: I appreciate the references to papers that describe the routing methods, but it would still be helpful to have a bit more information on linear reservoir routing. For example, does it mean that each grid cell's river discharge output to its neighbor is calculated as a linear reservoir, that is, as a function of river water within the cell? What, briefly, is the basis for assigning a reservoir coefficient? Are these constant or do they depend, for instance, on channel geometry?

Yes, the linear reservoir scheme calculates reach outflow as a function of water within each pixel, and the release coefficient is a function of estimated celerity and reach length. These details are described in the technical documentation accompanying the model code on our GitHub repository, which can be added as a supplement to the manuscript.

418 does 'scaler' mean 'scaling factor' or 'scalar' or something else?

We will revise the term to scaling factor in any revision of the manuscript.

429-437 Can you elaborate a bit on this treatment and why it is needed?

Because WBM calculates a uniform soil water balance over potentially large pixels, water can be withdrawn in larger pulses when soil moisture falls below the crop depletion factor than is typical in practice. This formulation provides an alternative to more closely approximate the practice of irrigating a rotating subset of crops continuously. We plan to revise the manuscript with an explanation of the utility of this option.

473 capitalization consistency

Sec 2.2.7 source tracking: this section is a bit confusing, partly (I think) because of the wide variety of sources that could be tracked. If I understand right, a user would not normally track ALL of these sources in any given model run, but rather would pick a type of source to track - is that right? Actually, reading forward in section 4, we learn that there are 3 options. It would be helpful to list these options up here in section 2. In addition, a couple of examples would potentially help a lot. They need not be very elaborate, but could be as simple as something like: 'a user interested in X might choose to track sources Y and Z'.

545 again, are these 3 mutually exclusive (i.e., one chooses from among them), or are all 3 tracked simultaneously? (later text suggests the former, but at this point in the text it is not clear)

In any revision of the manuscript, we can provide greater detail in Section 2.2.7 detailing the inter-relationship between the tracking components. First, none of the components (primary, return flow, or land attribute) are mutually exclusive, and a user could implement all three simultaneously if desired. We agree that providing simple examples in this section would help clarify the intended use of this functionality. For example:

A user interested in understanding the role of snowmelt as a component of streamflow downstream of a mountainous region would use primary source component tracking, whereas a user interested in understanding the potential for anthropogenic contaminants to be present in streamflow would use return flow component tracking. Finally, if a user was interested in runoff generated within any political boundary, land attribute tracking could be used. Any combination of the three components could be included in a model initialization and calculated simultaneously. However, the intersection of the three components is not calculated in this circumstance. Therefore, by initializing the model with primary source and return flow component tracking, WBM will not calculate the fraction of (for example) irrigation return flow composed of snowmelt.

557 daily time step: helpful to mention this much earlier.

We concur. A mention of the timestep will be made in the General Overview section in any revision of the manuscript.

Sec 3: helpful to define what you mean by validation (I'm not personally a stickler for semantics, but some would consider the term problematic, and better described by confirmation, testing, or evaluation).

We think the reviewer makes a good point that we will clarify in any manuscript revision. We use the term validation as used in common practice throughout modeling of natural systems, which is better described as corroboration (Oreskes et al. 1994). We will decide to either use an alternative term, or more clearly define our meaning and use of validation in any revision of the manuscript. To remain consistent throughout our responses to reviewer's comments and the draft of the manuscript we have continued our use of the term validation, with an implied meaning more closely aligned with corroboration.

Sec 3.1: are the summarized validation studies performed in conjunction with some kind of calibration / parameter optimization? Or is calibration only used in regional applications? You kind of answer this question around 635-640 but it would be helpful to clarify near the start of this section.

Several reviewers expressed an interest in seeing greater detail presented regarding calibration and parameterization strategies, and we would like to provide this general response. Any revision to our manuscript will provide additional detail needed to form a baseline understanding of the parameterization strategy for the model; however, we acknowledge at the outset of this response that there is to date untapped potential for more rigorous evaluation of uncertainty quantification using WBM.

First, following comments from Reviewers 2 and 5, we think that building out Table 2 to include a greater cross-section of parameters commonly adjusted in regional studies is appropriate. We note that this table will be redundant to the WBM_Usage_and_Input_Reference.xlsx spreadsheet on the WBM GitHub page; however, we agree that providing a subset of this reference within the manuscript will improve the readability. Any revision to this table will include default values, reasonable ranges, parameter description, and important citations where applicable. We note that to the extent possible, we have relied on structuring the model consistent with empirically meaningful parameters. As such, values presented in Table 2 will often reflect syntheses of field observations and uncertainty as characterized therein. Other model parameters are more synthetic and have less direct connection to field observation. Many of these parameters have been evaluated through calibration exercises over the years in studies summarized in Section (3.1). The reasonable ranges to be included in any revision to Table 2 will be based on what the authors consider as appropriate starting points for parametric uncertainty analyses based on a combination of prior experience and physical meaning.

Previous work to calibrate WBM has generally leveraged manual calibration, with several instances of more rigorous calibration attempts. Parameterization of core WBM components evolved through iterative attempts to capture response in both global and regional contexts. Generally, it has been found that parameterization schemes as represented by the default parameter values in WBM_Usage_and_Input_Reference.xlsx reflect reasonable compromises that adequately represent discharge time-series in global simulations. We plan to highlight that uniform parameterizations can be applied to unique watersheds to capture non-default response (Samal et al. 2017, Zuidema et al. 2018), or that

spatially varying parameterizations can capture more finely resolved nuance in watershed properties (Zuidema et al. 2020).

Validation generally: it would be interesting to summarize some of the lessons from testing and validation, it terms of what might be behind systematic under- or over-prediction of discharge. For example, have past validation exercises revealed certain gaps in knowledge, and/or mathematical approximations that would need to be refined in order to improve model performance? This might fit well under Results or Discussion.

This is an important point; however, we feel this would be better treated in a separate paper with a dedicated focus. We currently are working on a project centered around uncertainty quantification and with respect to on-going work focused on the conterminous United States; however, broader evaluations that include WBM should be considered for future work. We would also like to note that in response to other reviewers, we expect to add additional text that contextualizes results from model intercomparison projects that identify common constraints of macro-scale modeling and the importance of including human components into large-scale hydrologic models (Nazemi and Wheater, 2015, Veldkamp et al. 2018, Zaherpour et al. 2018).

Sec 3.2: thanks for differentiating between validation of different versions, and including this section devoted to the open-source version. It's a nice reminder (and demonstration) that testing of models should ideally include the specific code implementation alongside the theory and numerical algorithms.

Eq 31: the first time I read this, my mind immediately went to cancellation of errors -- but then realized that this is actually desirable for a bias metric. You might consider reversing the order of 32 and 31, and introducing the MBE with a phrase like 'in order to measure systematic bias' or something to that effect, so readers don't get hung up on it.

Thank you, we will follow your suggestion.

711 observations per year, or total?

We will clarify that we identify stations with total observations at each station equal to the values cited. These are very loose criteria to maximize spatial coverage of data used.

753 tense

Sec 5: I appreciate the code history and summary of different versions

943 uniformly spaced... in geographic coords? (again, helpful to explain grid set up early in the paper)

Agreed, we use uniformly spaced gridded data, typically in geographic coordinates, but the model does work with any projection recognized by GDAL.

1039 typo

References:

GDAL/OGR contributors: GDAL/OGR Geospatial Data Abstraction software Library., https://doi.org/10.5281/zenodo.5884351, 2022.

Nazemi, A. and Wheater, H. S.: On inclusion of water resource management in

Earth system models - Part 1: Problem definition and representation of water demand, 19, 33–61, https://doi.org/10.5194/hess-19-33-2015, 2015.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. Science, 263(5147), 641-646.

Samal, N. R., Wollheim, W., Zuidema, S., Stewart, R., Zhou, Z., Mineau, M., Borsuk, M., Gardner, K., Glidden, S., Huang, T., Lutz, D., Mavrommati, G., Thorn, A., Wake, C., and Huber, M.: A coupled terrestrial and aquatic biogeophysical model of the Upper Merrimack River watershed, New Hampshire, to inform ecosystem services evaluation and management under climate and land-cover change, 22, 18, https://doi.org/10.5751/ES-09662-220418, 2017.

Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Müller Schmied, H., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, Environ. Res. Lett., 13, 055008, https://doi.org/10.1088/1748- 9326/aab96f, 2018.

Zaherpour, J., Gosling, S. N., Mount, N., Müller Schmied, H., Veldkamp, T. I. E., Dankers, R., Eisner, S., Gerten, D., Gudmundsson, L., Haddeland, I., Hanasaki, N., Kim, H., Leng, G., Liu, J., Masaki, Y., Oki, T., Pokhrel, Y., Satoh, Y., Schewe, J., and Wada, Y.: Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts, Environ. Res. Lett., 13, 065015, https://doi.org/10.1088/1748-9326/aac547, 2018.

Zuidema, S., Wollheim, W., Mineau, M. M., Green, M. B., and Stewart, R. J.: Controls of Chloride Loading and Impairment at the River Network Scale in New England, 47, 839–847, https://doi.org/10.2134/jeq2017.11.0418, 2018.

Zuidema, S., Grogan, D., Prusevich, A., Lammers, R., Gilmore, S., and Williams, P.: Interplay of changing irrigation technologies and water reuse: example from the upper Snake River basin, Idaho, USA, 24, 5231–5249, https://doi.org/10.5194/hess-24-5231-2020, 2020.