Telteu et al., 2021, under review GMD: Understanding each other's models: a standard representation of 16 global water models to support improvement, intercomparison, and communication
https://gmd.copernicus.org/preprints/gmd-2020-367/

Thank you very much for your comments and recommendations.
The original comments of reviewers are in black color and indicated by “RC”.
Replies by the authors are indicated by “AC” and colored in blue.
Summary of our changes:

- We added the number of 16 global water models (16) analyzed, in this study, in manuscript´s title, to avoid confusion.
- We changed the manuscript’s structure:
    1. Introduction
    2. Modeling approaches and terminology used in global water modelling
    3. Key characteristics of 16 global water models included in the study
    4. Creating the standard writing style of model equations
    5. Similarities and differences among 16 global water models
    6. Number of water flows, water storage compartments, and human water use sectors included in 16 GWMs
    7. Potential future research of 16 global water models
    8. Recommendations for future multi-model intercomparison projects and extended assessments
    9. Conclusions
- We deleted the repeating information and we revised statements.
- We added two figures to visualize the number of GWMs that simulate vertical and lateral water balance in the ISIMIP2b framework (Figures 1 and 2).
- We revised section 2. New section 2 has two subsections. The subsection 2.1 presents different modelling approaches in global water modelling necessary to understand similarities and differences among 16 global water models. The subsection 2.2 presents definitions used in global water modelling.
- We revised and combined old information from old subsection 3.1 and 3.3 in a new section 3 (with four new subsections). In the new section 3, we present key characteristics of 16 global water models, analyzed in the present study. The subsection 3.2 became the new section 4 and presents information about our approach in creating a standard writing style.
- We moved information from old section 4, beginning of section 4, subsection 4.1 and subsection 4.2, as well as Table 11 to the supplementary information, to streamline the manuscript (please see Table S97).
- We revised section 6. The new section 6 has one section to present number of water flows, water storage compartments, and human water use sectors included in the 16 analyzed GWMs.
- The old subsection 6.4 (Potential future research in global hydrological modeling) became the new section 7, with the title: Potential future research of 16 global water models.
- The old subsection 6.3 (Recommendations for multi-model intercomparison projects) became the new section 8 (Recommendations for future multi-model intercomparison projects and extended assessments).
- We revised the list of references.

Please see below our answers to your comments and recommendations.

Answers for Reviewer #1: Wouter Knoben

RC: Summary posted online by **Wouter Knoben**

The authors have written down the model code that exists in 16 global models using standardized terminology (in the Supporting Information). This facilitates comparison between the models. The authors qualitatively compare the models in great depth and summarize this information in tables in the main manuscript. The manuscript also covers a variety of other topics: “typical” model setups in the Global Hydrologic Modeling, Land Surface Modeling and Dynamic Global Vegetation Modeling communities, a general overview of earth system models and known deficiencies, and lessons learned from the ISIMIP2b model intercomparison project, which the 16 models were part of.

To start, let me say that the work shown in the Supporting Information (SI) is impressive. I know that standardizing model code into a single format is not easy and doing this for 16 models of the complexity typical of Earth System Models is no small feat. I expect that the SI to this manuscript can become a valuable resource for Earth System modelers. Unfortunately, I also need to say that I found this paper difficult to review for two reasons.

First, the discussion of model differences and similarities is based on the standardized description of model code in the SI but it is virtually impossible for any reviewer to factually check this information. There is simply too much of it. Consequently, the reader needs to trust that this information is correct; which they may do if the process used to generate the standardized code is transparent and robust. The description of the method used to standardize the model code is currently limited to section 3.2 (some definitions, a paragraph on the actual process used and a description of which subscripts and superscripts are used) and section 6.1 (more definitions). I think the paper needs to be more descriptive of the methodology used to standardize the models’ equation and of the ways in which the authors ensured that the descriptions in the SI match the actual code in the models.

Second, I think this paper may be trying to do too many things at once. As far as I can tell, the paper covers three general themes (with some overlap between them):

Introducing a standardized way of writing ESM code, as evidenced by the manuscript’s title, the amount of work spent on creating the SI, section 3.2 and 6.1 (method for standardizing equations), and the lengthy discussion of model similarities and differences in section 5.

Providing a general commentary on the state of, and challenges associated with, global hydrologic modelling, as evidenced by section 2 (typical model use in different modelling communities and confusion about terminology), section 4 (general history and challenges with global hydrologic modelling), section 6.2, Table 11 and its submission as a “review and perspective” paper.

Laying the groundwork for a follow-up ISIMIP2b paper by describing the models and process of this MIP, as evidenced by the introduction, sections 3.1 and 3.3, section 5, and sections 6.3 (lessons learned from the MIP) and 6.4 (future work planned by MIP contributors).

I think any of these themes can be a good contribution to GMD but combining all three into a single paper seems to me to be too much. The manuscript is currently a bit haphazard in its organization, it was sometimes unclear to me how sections related to one another and due to the extremely broad scope I think none of the three themes get the amount of attention and detail they need to be convincing. What I missed for the 1st item was a detailed description of how the standardized writing scheme was developed, its strengths and weakness, procedures used to robustly translate model code, applicability to models outside this set of 16, a discussion of the implications of the discovered similarities and differences for ensemble modelling and model intercomparison, etc. What I missed for the 2nd item was a discussion of a considerable number of existing commentaries on this topic (some suggestions below) and a discussion of the information presented in Table 11. What I missed for the 3rd item is a more in-depth description of the MIP, established procedures, etc. Given that the manuscript is already just a bit shy of 1000 lines of actual text, I doubt there is space to fully cover all three themes. I would therefore strongly recommend clearly defining the scope of the paper and streamlining/modifying the text accordingly.

I have added various comments as annotations to the uploaded .pdf in the hopes that they are helpful to the authors in clarifying the text.

Kind regards,

Wouter Knoben

Possibly relevant literature

Archfield, S. A., et al. (2015), Accelerating advances in continental domain hydrologic modeling, Water Resour. Res., 51, 10078– 10091, doi:10.1002/2015WR017498.

Bierkens, MFP (2015) Global hydrology 2015: State, trends, and directions. Water Resour Res 51:4923-4947. https://doi.org/10.1002/2015WR017173

Clark MP, Fan Y, Lawrence DM, et al (2015a) Improving the representation of hydrologic processes in Earth System Models. Water Resources Research 51:5929–5956. https://doi.org/10.1002/2015WR017096

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D. (2017): The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, Hydrol. Earth Syst. Sci., 21, 3427–3440, https://doi.org/10.5194/hess-21-3427-2017

Gleeson, T., Wagener, T., Döll, et al. (in review, 2020) HESS Opinions: Improving the evaluation of groundwater representation in continental to global scale models, Hydrol. Earth Syst. Sci. Discuss. [preprint], https://doi.org/10.5194/hess-2020-378

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48, W08301, doi:10.1029/2011WR011044

AC: Thank you very much for your recommendations and the general advice – we greatly appreciate it and tried to consider it as much as possible during the review. The present study is a review and perspective paper on 16 global water models (GWMs) that provide simulations for ISIMIP2b. We do not make a review on Earth System Models (ESMs). Some GWMs, mainly, the land surface models (LSMs), represent a part of ESMs.

We changed the paper structure, based on the recommendations received.

Our new paper structure is:

1. Introduction
2. Modeling approaches and terminology used in global water modelling
3. Key characteristics of 16 global water models included in the study
4. Creating the standard writing style of model equations
5. Similarities and differences among 16 global water models
6. Number of water flows, water storage compartments, and human water use sectors included in the 16 GWMs
7. Potential future research of 16 global water models
8. Recommendations for future multi-model intercomparison projects and extended assessments
9. Conclusions

The first part of the article (sections 2 to 6), the review part, facilitates a better understanding of how the models work and shows similarities and differences among the models to explain better their various results. The second part (sections 7 and 8), the perspective part, presents current and future modelling potential. This article presents comprehensively decisions made in the ISIMIP2b global water community to simulate the impact of climate change on freshwater systems.

We consider that we need to provide information on communities' goals, terminology used, the modeling experiment to facilitate a better understanding of how these models work. Therefore, all these issues are connected and need to be presented in a single article. The present study offers time savings in searching for terminology and model structures. Firstly, this study would be interesting mostly for modellers and readers analysing multi-model data who want to check why specific differences in simulation results occur. Secondly, our audience with little or no knowledge on GWMs needs a single article that presents an overview of GWMs.
We revised the title to fit better with our main purpose: *Understanding each other's models: a standard representation of 16 global water models to support intercomparison, improvement, and communication.*

Specific comments Reviewer #1:
RC, line 82: It may help the reader to include a one-sentence summary of what is actually done in this study at the start of this paragraph, to better connect the knowledge gap identified in the previous paragraph and the study goals discussed in the highlighted sentence. Currently the "how" these goals will be achieved seems to be missing.
AC: We have introduced a summary of what we have done in this study.
The revised statement is (please see section 1 Introduction):
In this complex scientific context, the present study represents a step forward to increase understanding of process representation and inter-model differences within one large MIP, specifically, ISMIP – the Inter-Sectoral Impact Model Intercomparison Project (Frieler et al., 2017). We assessed the equations applied by 16 state-of-the-art global water models (GWMs) to simulate the vertical and lateral water balance, human water use sectors, and desalination on the global scale. We created a standard writing style of these equations to identify similarities and differences among models. Thereby, the global water community has through this study an overview of the model structures and the basis required to interpret various model results, to design future experiments on how model equations, model configurations, and model parameter values influence the model outputs.
In summary, our three main goals are:
- *to provide a better understanding of how 16 state-of-the-art global water models are designed;*
- *to show similarities and differences among them, based on their equations;*
- *to underline future research potential in global water modeling.*

RC, line 101: Differences in modeling approaches: In my opinion the different communities are moving ever closer and the dividing lines between them are somewhat vague. This section may be helped by including a few examples of models that are considered LSMs, GHMs and DGVMs in the corresponding paragraphs.
A: We present some examples in section 3.
We consider that these communities collaborate, but they have different research questions and focus on specific hydrological processes. Therefore, these communities need a strong interaction to improve the models.

RC, line 101: I think this section may be slightly too succinct to clearly describe the differences and similarities between the three communities, especially if the target audience is students or researchers who may have experience with at most one of these communities. It is difficult to offer concrete suggestions but perhaps section 2 in Archfield et al. (2015; https://doi.org/10.1002/2015WR017498) can offer inspiration. That section presents a similar comparison between three large-domain modelling communities but provides substantially more detail.

AC: In this section, we present the modelling approaches of climate, global hydrological, and vegetation communities, as well as the link between these communities by using one example. This example is understandable for readers without knowledge in one of these communities, mainly, students, because it connects them with reality, they can imagine and identify the connection between processes. Readers receive an introduction about the different modelling approaches. This facilitates a better understanding of the different model structures, similarities and differences among models (section 5).
Thank you for your recommendation! Archfield et al., 2015 in section 2 presented a comparison between modelling approaches in catchment hydrology, global water security, and land surface modeling communities, but does not connect these communities. They also present development trajectories for these communities.
We cited Archfield et al., 2015 in the present study, section 2.1, to direct readers towards more details on global hydrological and land surface communities.
Our revised statement is (please see subsection 2.1):
However, these three communities focus on specific hydrological and atmospheric processes, as well as anthropogenic impacts. These key aspects are important for their specific research leading to different modelling approaches, specific evaluation studies of model performance (Archfield et al., 2015), and different field-specific meanings of terminology used (Beven and Young, 2013). Thus, combining the expertise in their key aspects would create a strong synergy and improve the models of these communities, but for this goal, they have to interact with each other, identify their similarities and differences and share experiences. They need to undertake joint experiments, present and discuss their results, discuss how they influence and depend on each other, and how water modeling can be improved (Cucchi et al., 2020).

RC: Line 109: comment on: “The global hydrological community is focused on surface hydrologic processes, primarily river flow simulation and its daily to century-scale changes”. This may be a bit confusing. Even though river flow simulations are a primary focus for GHMs, I doubt there are many GHMs that do not include any subsurface processes. Besides, global groundwater availability is an active field of study. See e.g. Gleeson and co-authors and references therein:
- Nature, 2012, https://doi.org/10.1038/nature11295
- Nat. Geo., 2016; https://doi.org/10.1038/ngeo2590
- HESS, under review, https://doi.org/10.5194/hess-2020-378
AC: Thank you for your comment and references. You are completely right, all the models do represent the hydrological processes in soil (subsurface runoff), but five GWMs do not have a specific representation of groundwater. Please see the next reply for a modification of the manuscript.
RC: Line 109: “daily”: Is this typical? Such models need to be run at sub-daily timescales to properly capture diurnal variations which can significantly impact simulations of e.g. evaporation, transpiration and snow melt.
AC: For most GHMs, daily time steps are still state-of-the-art, whereas for LSMs, sub-daily time steps are standard. Our revised statement reflecting this and your previous comment is: *The global hydrological community focuses primarily on surface water and groundwater availability, its human interference, and their daily to century-scale changes.*

RC: Line 132: I think this section is critical but in my opinion in its current shape it is not as helpful as it could be. The reader currently gets two examples and is left to wonder how many other occurrences of such ambiguity exist but the paper offers no further guidance.
What may be helpful is to divide the modeling chain into distinct elements and go through each of them systematically, and do the same for the main states and processes that each community considers. I.e. the elements "input variables, state variables, parameters,

constants, and output variables" mentioned on line 182. Brunner et al. (2018; https://doi.org/10.1002/hyp.13227) could possibly be helpful.
A: Thank you very much for the suggestion to restructure this section. In our new section 2.2, we define terms used in global water modelling, necessary to understand how global water models work. We moved the paragraph on *climate forcings* as a note of Table S86 because it is necessary to clarify this for our readers. We defined other terms in Tables S84. We believe that we have reached now a more systematically description of the modelling chain.

RC: Line 138: In this case is "active" considered to be a synonym of "dynamic"? If so, that may need to be clarified. Which community uses which definition: *dynamic vegetation* and *active vegetation*?
AC: Many scientists from both communities are using both expressions: We made the decision, in the present study, to separate the *active vegetation* from *dynamic vegetation*. Yes, *active* is synonym of *dynamic* but it describes a different process in this study. In this study, we use *active vegetation* to highlight if models include the photosynthesis scheme in their structure and if they have the ability to simulate actively changes in vegetation, in an area, because of changes in the CO_2 concentration, air temperature, and precipitation. We use *dynamic vegetation* to define changes in vegetation from one geographical area to another because of competitive and biogeographical processes determined by climate change (geographical distribution of plants) or human activities.

RC: Line 144: This seems relevant information, but is the section on definition of terms the best place for it? It doesn't fit very well with the rest of the contents of this section
AC: Thank you. We moved this information to subsection 4.3.

RC: Line 151: Also relevant but it lacks a connection to the differences in terminology that this section is supposed to be about. Perhaps adding a sentence like "For successful interaction and collaboration, being aware of differences in vocabulary and potentially agreeing on a list of definitions are a necessary step." can add this connection
AC: Thank you for your recommendation. We revised this paragraph and moved it to our new subsection 2.1.
In our new subsection 2.1, we present different modelling approaches.
Our revised statement reads:
Thus, combining the expertise in their key aspects would create a strong synergy and improve the models of these communities, but for this goal, they have to interact with each other, identify their similarities and differences and share experiences. They need to undertake joint experiments, present and discuss their results, discuss how they influence and depend on each other, and how water modeling can be improved (Cucchi et al., 2020).
In our new subsection 2.2, we present definitions used in global water modelling.

Thanks to your recommendation, we end subsection 4.3 (where we present definitions used in the present study and challenges found in defining the analyzed variables) with the next paragraph:
In summary, in global water modelling, we need to be aware of differences in vocabulary. A widely accepted list of definitions would avoid confusion and facilitate successful interaction and collaboration. Furthermore, we need to clarify hydrological terms to peers from other disciplines, stakeholders, and a general audience (Brunner et al., 2018) to facilitate easier communication, understanding, and analysis.

RC: Line 185: “Parameters may change in space, but do not change in time”: I'm not sure this is entirely correct. Perhaps it depends on the definition of "parameters". I'm thinking of

something like a Leaf Area Index which is often thought of as a parameter but also changes in time for deciduous trees, or snow albedo which changes as a snow pack matures. I assume something like "maximum canopy water storage" (next sentence) will be time-variant in the vegetation modelling community as vegetation matures or changes. I suggest to remove the highlighted sentence.
AC: We deleted the sentence.

RC: Line 189: “Some processes are parameterized, meaning that their values are precisely marked in the computer code and are not calculated by the model itself.”
This needs more clarification, partly because the term "parametrization" is also used to refer to the equation that describes a process. For example, "stomatal resistance in model X is parametrized with either Ball-Berry or Jarvis formulations".
Does this sentence mean that sometimes model processes are set at fixed values? E.g. "percolation is assumed to occur at a fixed rate of x mm d-1".
AC: Thank you for pointing out this inconsistency. We deleted this sentence. We included subsection 2.2 where we define the terminology used in global water modelling.

RC: Line 193: “Ultimately, a model also uses constants, properties of the model that do not change in space and time“ I have always thought "constants" refers to physical constants such as the freezing point of water or gravitational acceleration. Is this what is meant by "properties of the model"?
AC: A model describes a hydrologic system and in this case, constants are properties or characteristics of the model: density of water at 0°C. We revised our text on constants, please see section 2.2.
A global water model describes the dynamic behavior of a hydrological system that includes input variables, state variables, parameters, constants, and output variables (Bierkens and van Geer, 2007). State variables define how much water is in a compartment or storage at the beginning of the simulation, and can change in space and time, for example, canopy water storage. Their variation is caused by a variation of the input variables, for example, precipitation. State variables are related to the input variables and output variables through parameters, for example, infiltration capacity of the soil. Parameters and coefficients represent numbers that describe a particular characteristic of reality, of the model, of the catchment area or flow domain. Some examples are runoff coefficient, soil porosity, hydraulic conductivity of different soil horizons, maximum soil water storage, maximum canopy water storage, mean residence time in the saturated zone, surface roughness, and vegetation properties (Beven, 2012). A model also uses physical and mathematical constants meaning characteristics of the model that do not change in space and time such as catchment area. Physical constants are physical quantities that can be measured and have a constant value in time, for example, the density of water at 0°C, the density of ice. Mathematical constants cannot be measured, but can be calculated and have a fixed numerical value, for example, $e = 2.718\ldots$, $\pi = 3.142$, $i2 = -1$. Ultimately, output variables vary in space and time, for example, streamflow in a river catchment.
Thus, a water global model includes many equations written with a programming language in a model code to simulate freshwater systems. During simulations, many parameters receive specific values because they cannot be measured everywhere, therefore, they are calibrated or tuned to attain the best match between simulated and observed data. The final steps of a simulation are to validate simulated and observed data, to find out how well they fit, and to evaluate the simulated results through analysis and visualization.

RC: Line 194: It seems to me that the description of steps taken only starts here. The text before this point in this section may be better placed in the section about (ambiguity of) definitions.
AC: We moved the text to the subsection 2.2.

RC: Line 200: Adding to my comment above, it would help me if "parametrization" and parametrized" were to be strictly defined in the definition section. I'm currently not quite sure how to interpret this part of the sentence.
AC: Thank you. We included the definition in our new subsection 4.3.
Our revised statement reads: *We define parameterization as changes of model parameter values (Samaniego et al., 2010).*

RC: Line 206: This must have been an enormous amount of work and I congratulate the authors for completing it. However, these tables are the resource upon which the remainder of the manuscript is based and they are practically impossible to review without re-doing the authors' work.
Did the description and translation of these model equations include some form of quality control to ensure that what's in the tables matches the model code? Are there other considerations that can help the reader have confidence that these tables accurately reflect each models' code and thus that the remainder of this manuscript. is well-founded?
AC: Thank you for raising this concern. We explained in a new paragraph from subsection 4.5 that the equations from the supplementary information have been checked by modelling teams. This was reached with many revision rounds until all authors agreed on the representation of the models in the equations. Nevertheless, many of these equations were published in peer-review articles mentioned in Table 12, although with a different notation as the one used in this study.
The revised statement from subsection 4.5 reads:
In the next step, modelling teams created and provided the model equations, used to provide simulations for ISIMIP2b, according to the generated lists. Each modelling team, involved in this study, internally checked and reviewed its model, based on the model code and peer-review articles mentioned in Table 12 or only on the peer-review articles on model description mentioned in Table 12. In some cases, modelling teams provided the equations using our standard writing style and symbols presented in subsection 4.4, while in other cases using their specific writing style. Therefore, the modelling teams checked the model equations on their plausibility.

RC: Lines 214 215: to add “for”.
AC: Thank you. Our new statement is: *We selected “S” to describe water storage, “P” to describe everything connected to precipitation, “E” for everything related to evaporation, “R” for everything related to runoff, “Q” for everything related to streamflow and outflow, and “A” for water abstractions.*

RC: Line 259: “Five models (DBH, JULES-W1, Mac-PDM.20, VIC, and WAYS) do not use any river routing scheme for the ISIMIP2b; therefore, they do not compute streamflow.”
Should this be "aggregated streamflow" or "routed streamflow"? Local (within-grid) streamflow can still be computed for these models as R = Rs + Rsb.
AC: Thank you for pointing this out. In Table S84, we presented equations for water stocks and flows. In this study, we use total runoff for “aggregated streamflow” and streamflow for “routed streamflow”, according to the ISIMIP2b simulation protocol. Total runoff refers to the total amount of water that runs-off the grid-cell, either over the soil surface, or from the

subsurface (lateral flow). In some studies, the streamflow is converted to runoff by dividing the streamflow values with the area upstream of the gauging station (for example, the area upstream of station according to the DDM30‘ river network Döll and Lehner, 2002). “Streamflow”, as mentioned in the sentence you are referring to, refers to the volumetric flow rate of water through a river cross-section. The streamflow is transfer through a channel to the ocean or to an inland sink.

RC: Line 267: “CWatM calibrates monthly or daily streamflow for 12 catchments using the Distributed Evolutionary Algorithms in Python (DEAP) approach (Burek et al., 2020), while WaterGAP2 uses a very simple basin-specific approach to match long-term mean annual observed streamflow at the outlet of 1,319 gauged hydrological stations. It considers runoff as a nonlinear function of soil moisture and uses a runoff coefficient and two correction factors to calibrate the simulated and observed streamflow (Müller Schmied et al., 2014; Müller Schmied et al., 2021).”
For all of these, how are the results of local calibration (to e.g. 12 catchments, 1319 gauges or all GRDC observations) used to inform parameter values for global simulations? Is some form of parameter regionalization used?
AC: Indeed, the calibrated parameters have been regionalized. For WaterGAP2, one of the three calibration parameters is regionalized using a multiple linear regression approach considering a number of basin descriptors (Müller Schmied et al., 2021, their Sect. 4.9.2). For CWatM, an evolutionary algorithm with KGE as objective function was applied and WFDEI meteorological data were used as forcing (Burek et al., 2020). We cited in the manuscript Burek et al., 2020 and Müller Schmied et al., 2021 for detailed information.

RC: Line 288: “Global water models were developed from the earliest land surface models created by Manabe (1969), Freeze and Harlan (1969), and Deardorff (1978). These first land surface models simulated the terrestrial water cycle by considering vegetation processes, evaporation, soil moisture, and snow cover. Later on, Dooge (1982) identified the two major challenges of global hydrology: scaling and parameterization. Eagleson (1986) declared the necessity of global-scale hydrology. Inevitably, during the 1990s, the first global hydrological models were developed (Alcamo et al., 1997; Vörösmarty et al., 1998, Arnell, 1999). Over the years, many models have been developed and improved and many studies have been done to assess freshwater resources on the global scale (Bierkens, 2015)”.
This seems oddly placed here and may fit better in the introduction of this manuscript.
AC: We revised this information and moved it to the supplementary information, before Tables S84 (with definitions used in the present study), to streamline the manuscript.

RC: Line 295: This seems to repeat parts of section 3.1.
AC: Thank you. We deleted this information from section 4. In our new section 4, thanks to your recommendations, we describe our method – how we created the standard writing style of model equations.

RC: Line 303: “For smaller catchments, the results are often not reasonable (e.g., Beck et al., 2016) and require some corrections (eventual post-process) due to inaccurate input data, spatial heterogeneity, and the lack representation of some hydrological processes, for example, capillary rise, artificial transfers, and pond development (Döll et al., 2003; Hunger and Döll, 2008).”
This may be the case but I'm not quite sure if the introduction to the review section is the right place to mention this issue. Is there a section about "Common weaknesses shared by all models" where this might be more appropriately placed?

AC: Thank you. We consider that this information provides explanation regarding the real application of the models and the default scale of 0.5° x 0.5°, used for the 16 global water models in ISIMIP. We moved this information to the new section 3.

RC: Line 305: "Hattermann et al. (2017) highlighted the role of global and regional water models. Global water models assess the large-scale impacts of climate change and its variability, while regional water models assess the small-scale impacts that are specific to a particular river, catchment, or region. Gosling et al. (2017) underlined that the global and regional water models share many similarities regarding runoff simulation results and their conceptual approach to model development, although the GWM results vary more than regional water results." This also seems oddly placed here and better suited to the introduction. At this point (for me) it is well-established that the manuscript is about GWMs and the sudden mention of regional water models threw me off a bit.
AC: Thank you. We revised this information and moved it to the supplementary information, before Table S84, to streamline the manuscript.

RC: Line 310: "Ultimately, GWMs have faced many challenges in selecting a good method to estimate water storage compartments, water flows, and human water use sectors. Some of these are presented in the following subsections."
AC: We deleted this sentence.

RC: Now I reach this sentence, I think I understand this section better. From the section title "Review of the global water models included in the study" I expected this section to provide an overview of differences and similarities between the GWMs. Instead, this section and its subsections seem to list several "Common challenges associated with global water models". I would suggest to change the section title to something like that to manage reader expectations. I would also suggest to streamline this section and keep the meat of the content in the subsections.
AC: Thank you again for helping to streamline the manuscript. We moved the information on the history of global water modelling and the old subsections 4.1 and 4.2 to the supplementary information, to streamline the manuscript. We moved some information from the old subsection 4.3 to the new section 8.

RC: Line 312: Models
AC: We have corrected the word.

RC: Line 315: "For example, Wartenburger et al. (2018) concluded that the values of actual land evapotranspiration are affected by the methods used to estimate evapotranspiration, number of soil layers, model structure, and uncertainties in the climate input datasets. Reviewer suggested to write "simulated" land evapotranspiration.
AC: Wartenburger et al. (2018) called this process "actual land evapotranspiration".
Our revised statement reads: *actual (simulated) land evapotranspiration*.
We moved this information to the supplementary information to streamline the manuscript (before Table S84).

RC: Line 335: I'm confused by the use of "nevertheless" here. The findings listed above seem like excellent reasons for evaluation by increasing our understanding of models' strengths and weaknesses. I'd suggest to remove "nevertheless" and move this sentence to the start of this subsection to introduce that model evaluation is needed.

AC: Thank you. We moved this information to the supplementary information, before Table S84, to streamline the manuscript.

RC: Line 340 – 341: Needs brackets, This too
AC: We added the brackets.

RC: Line 342: Should this be "existent" or "present"?
AC: We corrected the word, thank you. It should be "existent".

RC: Line 350-352: „Nevertheless, studies on water scarcity and their results are affected by their methodology, definitions, and assumptions"
In summary, Some references would be good here.
AC: Thank you. We deleted this information, to streamline the paper.

RC: Line 369: This is more commonly referred to as the "Rainfall-Runoff Modelling Toolbox" or RRMT. Also, I went to check the references to confirm I had this right but Wagener et al (2001) seems to be missing from the reference list.
AC: Thank you for pointing out this mistake, we added Wagener et al., 2001 in the reference list.
RC: Line 371 – 372: to add "Structure for" + Modeling
AC: We corrected the sentence.

RC: Line 387: Upon reaching the end of section 4 I'm somewhat uncertain about its purpose in the main text. This information is very specific in some cases (section 4.1) and very general in others (section 4.3 is essentially a list of uncertainties, modeling frameworks and uncertainty estimation approaches). Based on section 3 (description of the work done to standardize model equations into a single document) I had expected section 4 to dive deeply into what can be learned from the material described in section 3, but this seems to happen in section 5 (which I have not read yet). I would strongly encourage the authors to reflect on what the purpose of the current section 4 is and whether that purpose is achieved. In case it helps, what I think section 4 tells me is "modeling the real world accurately is difficult because study conclusions are strongly dependent on models, methods and assumptions, but there are ways you can try to quantify this uncertainty." It's not clear to me why I need to have the information presented here in order to understand section 5, or how it connects to section 3. From my point of view a careful selection of the material in section 4 could be moved to the introduction but much of it might be removed all together without harming the manuscript and improving the flow between sections 3 and 5.
AC: Thank you for raising these concerns. We restructured the manuscript. In our new section 4, we present our approach – creating the standard writing style of model equations. We present in our new section 8 a list of modeling frameworks that could be tested by global water modelling.

RC: Line 391: This section is quite dense and unless one is looking for specific information, it is a bit difficult to retain much of the information here. I'm wondering if this might not be neatly summarized in a sketch of the hydrologic cycle that includes all processes present in the models and a note next to each process with the number of models that include this process.
AC: Thank you for your recommendation. We included two new figures on number of GWMs that simulate vertical and lateral water balance (please see Figures 1 and 2).

RC: Line 396: Probably more correct to call this "change in canopy water storage".
A: We have corrected the sentence.

RC: Line 408: "Generally, prescribed vegetation ignores the decisive interaction between vegetation and runoff as well as interactions between the atmosphere and Earth's surface". Can it be clarified what both of these interactions are? This is not clear to me as someone with little background in vegetation modeling.
AC: The revised statement reads (please see subsubsection 5.1.1): *Generally, prescribed vegetation ignores the decisive interaction between vegetation and runoff and interactions between the atmosphere and Earth's surface, partly presented in section 3.2 (Gerten et al., 2004; McPherson, 2007; Nicholson, 2000).*
.
RC: Line 419: Has this acronym been defined yet?
A: PFTs represent plant functional types. We explained the acronym in the text.
Our revised statement reads (please see subsubsection 5.1.1):
Generally, it was found that simulations depend on the number of plant functional types (PFTs) prescribed or defined in the model and on the processes used to estimate plants' ability to adapt, acclimate, and grow in new environmental conditions (Sitch et al., 2008).

RC: Line 423: "GHMs use the degree-day method to compute snow accumulation and snowmelt, while LSMs use the energy balance method " - typically
AC: We added the word: "typically".

RC: Line 432: "Further, seven models differentially estimate snow under the canopy (Table S10). " Meaning that these models use different methods/equations to estimate snow under canopy?
AC: yes, we mentioned Table S10 for more information about computation of "snow under the canopy".

RC: Line 434: "MATSIRO is the only model that distinguishes between sublimation on snow-covered ground and snow-free ground". Perhaps this is a really simple question, but if the surface is snow-free, what's being sublimated?
AC: Thank you, there was the word "evaporation" missing: The revised statement reads: *MATSIRO is the only model that distinguishes between sublimation and evaporation on snow-covered ground as well as evaporation on snow-free ground.*

RC: Line 435: "Snow layers vary between 1 (most of the GHMs) and 12 (CLM5.0; Tables 7 and 8)." Are the number of layers fixed (but with time-varying depth) or can the number of layers vary in time (for computational efficiency)?
AC: The numbers of layers are fixed.
The revised statement reads: *The number of snow layers is fixed and it varies among the 16 GWMs between 1 (most of the GHMs) and 12 (CLM5.0; Tables 7 and 8). Most of the GWMs present no upper limit for snow storage (Tables S48 – S51).*
RC: Line 436: "Hydrologically, this includes an unsaturated zone." This seems to imply that it also includes something else. What is that?
AC: The revised statement reads: *Hydrologically, this includes the unsaturated zone or vadose zone, the part of Earth between the land surface and the top of the phreatic zone (water table).*

RC: Line 438: "Overall, 10 models consider initial infiltration as inflow of the soil storage, while 3 models (H08, JULES-W1 and WAYS) consider throughfall (Table S14). Mac-

PDM.20 considers total precipitation as inflow of soil storage (Table S14).” Is there an effective difference between these three approaches or is it just the name that changes?
AC: Infiltration, throughfall, and total precipitation have different values. Total precipitation represents input data for some GWMs. GWMs compute differently infiltration (Table S25) and throughfall (Table S5).
The revised statement reads: *Overall, 10 models consider initial infiltration as inflow of the soil storage, while three models (H08, JULES-W1, and WAYS) consider throughfall (Table S14). Mac-PDM.20 considers total precipitation as inflow of soil storage (Table S14). Thus, infiltration, throughfall, and total precipitation have different values among 16 models because the models compute infiltration and throughfall differently, while total precipitation represents the input data for some models.*

RC: Line 452: “The mHM model has one more bucket between the soil storage and groundwater storage named “unsaturated storage” representing the source for interflow and groundwater recharge. “
Is this correct? Wouldn't one typically expect the unsaturated zone to be above the saturated zone, or does mHM differentiate between tension storage that fills first and remaining porosity that drains under gravity?
AC: The Reviewer is right. The unsaturated (soil) zone (UZ) storage is located below the last soil horizon (n horizons possible) and on top of the saturated storage (i.e., groundwater proxy storage). This conceptualization is based on the HBV model structure. In this respect, mHM differs from HBV parameterization on the number of soil layers (HBV has only one) and the parametrization of the soil infiltration and percolation rates, which are based on parameterizations of Brooks and Corey, 1964 and Twarakavi et al., 2009. From UZ, three fluxes are possible: (1) percolation to the saturated storage (groundwater proxy), (2) quick interflow and (3) fast interflow, both contributing to surface runoff component and routing scheme. mHM has no tension storage parameterization. The groundwater storage generates in turn the based flow and, if activated, deep groundwater percolation (leaking basins) in case that a karstic system is present.

RC: Line 454: “LPJmL was adjusted, and the water from the uppermost soil layers is considered to contribute to surface runoff if excess of storage is calculated according to the infiltration or percolation rates, which depend on soil type. LPJmL routes, what was previously lateral runoff, from “layer 0” (first 20 cm), as surface runoff.” How and why?
AC: Thank you very much for your question. In LPJmL, the lateral runoff from layer 0 (uppermost layer) was rerouted to surface runoff to fit the ISIMIP2b protocol. LPJmL computes surface runoff (on soil surface) and lateral runoff (in all soil layers), for ISIMIP2b, identically to how they are computed in the original model version. However, lateral surface generated in layer 0 is added to the surface runoff output variable instead of to the lateral runoff output variable.

RC: Line 461: JULES-W1 also uses a “zero-layer” scheme that does not use explicit model layers to represent snow, instead adapting the topsoil level to represent lying snow processes (Best et al., 2011). Is this the correct word to use here?
AC: yes, it is, according to Best et al., 2011. The revised statement reads:
JULES-W1 also uses a “zero-layer” scheme that does not use explicit model layers to represent snow, instead adapting the topsoil level to represent existent snow processes. In the original “zero-layer”, snow scheme has a constant thermal conductivity and density. Bulk thermal conductivity of snow on the surface layer decreases due to both the increased layer thickness and the different conductivities of snow and soil. Surface energy balance and heat

flux between the surface layer are controlled by insulation factors and layer thickness. (Best et al., 2011).

RC: Line462: “In the DBH model, runoff is generated directly when soil layer is saturated, or is generated when rainfall intensity is larger than the infiltration rate estimated with the Green–Ampt method (Tang et al., 2006).” - “a” soil layer.
AC: Thank you, we corrected the sentence.

RC: Line 475: “Some GWMs compute vertical water movement in unsaturated soils by applying the Richards equation (Richards, 1931; e.g., CLM4.5, CLM5.0, CWatM, JULES-W1, MATSIRO, ORCHIDEE, VIC). However, the Richards equation might be not relevant for the models that have one soil layer.” Can it be clarified why not or can a reference be added that supports this?
AC: Mainly, the Richards’ equation describes the flow of water in an unsaturated porous medium due to the actions of gravity and capillarity. This is a very complex process. It cannot be very well simulated by models with one soil layer and without capillary rise. We provided a reference for this.
The revised statement reads: *However, the Richards equation may not be relevant for the models that have one soil layer because of its complexity and of missing capillary rise (Lee and Abriola, 1999; Farthing and Ogden, 2017).*

RC: Line 482–483: “Soil column configuration. Number of soil layers ranges between 1 (H08, MPI-HM, and WaterGAP2) and 25 (20 soil layers + 5 bedrock layers: CLM5.0), while total soil depth is between 1 m (H08) and 49.6 m (CLM5.0; Tables 7 and 8). ” Are the numbers in this section a reflection of modeling decisions made for ISIMIP2b or are the models hard-coded to use these numbers of layers?
AC: Thank you for your comment. No, these values are not prescribed in the ISIMIP2b simulation protocol. Some GWMs are hard-coded to use these numbers of layers, for example, CLM5.0. We mentioned, in this study in the lines 482–483, the Tables 7 and 8.

RC: Line 488: “Groundwater storage, beneath the soil water storage compartment, receives water from seepage and groundwater recharge.” Are these two terms for the same process or do they refer to distinctly different processes?
AC: GWMs use the term groundwater recharge when they include in their structure a groundwater compartment (recharge of the groundwater compartment), while seepage when they do not include in their structure a groundwater compartment. Seepage is the amount of water that leaks at the bottom of the soil storage. ISIMIP2b relates seepage with groundwater recharge for the models that do not include groundwater storage, supposing that this water would reach groundwater storage if it would exist (JULES-W1 and LPJmL).
The revised statement reads: *Groundwater storage, beneath the soil water storage compartment, receives water from drainage (e. g., MPI-HM) or aquifer recharge (e. g., CLM4.5) or groundwater recharge (e. g., WaterGAP2) (Tables 9 and 10). In ISIMIP2b, two models (JULES-W1 and LPJmL) consider the water excess from the bottom soil layer as seepage and relate this variable with groundwater recharge because they do not have a groundwater compartment.*

RC: Line 489 – 490:“Groundwater storage – Hydrologically, it includes the saturated zone or phreatic zone.” This may another definition that varies between communities. I have often seen groundwater storage be used to indicate a deep aquifer, whereas the saturated and unsaturated zones are placed on top of this aquifer and are where evaporation etc. takes place.

AC: The revised statement reads: *In GWMs, groundwater compartment simulates hydrologically the saturated zone or phreatic zone (WaterGAP2) or an unconfined aquifer (CLM4.5).*

RC: “It loses water through capillary rise, groundwater runoff, and abstraction for human water use”. Maybe I'm missing it but it seems that text here does not include any mention of one of the models simulation groundwater abstraction.
AC: Thank you. The revised statement reads: *It loses water through capillary rise, groundwater runoff, and groundwater abstraction for human water use.*

RC: Line 500: Suggest to move “between 1 and 13” this directly after "... in time
A: Thank you. We revised the sentence.

RC: Line 562-563: “River storage fills with water through flows above and below the ground. It loses water through streamflow, evaporation, channel transmission, and water abstraction for human water use. Five models (DBH, JULES-W1, Mac-PDM2.0, VIC, WAYS) do not include river storage for ISIMIP2b simulations, because of computational and resource constraints, nor do they compute streamflow”.
Does this mean routed streamflow, i.e. catchment-aggregated values?
AC: yes. In this study, we distinguish between runoff and streamflow. We defined these terms in Table S84 (please see our answer above).

RC: Line 596: “Human water abstraction represents the sum of the water consumed by humans, evaporative and speculative water losses (named water consumption), and water returned to the groundwater or surface water compartments (named return flow, being the part of the water not consumed)”: I'm not familiar with the term speculative water loss - can a definition be added?
AC: Thank you, indeed it is not a common term. The revised statement reads: *Human water abstraction represents the sum of the water consumed by humans, evaporative water and other water losses (named water consumption), and water returned to the groundwater or surface water compartments (named return flow, being the part of the water not consumed).*

RC: Line 601: “Irrigation water demand (potential irrigation water abstraction) is computed by three models (Table S52).” Where does this demand come from in the other models? Is it some constant or prescribed value?
AC: Thank you for your comment. Nine models compute irrigation water demand for ISIMIP2b (please see Table S52 and Tables S53 and S56 for sub-components of irrigation water demand). We corrected and integrated this information into Table S52. Seven models do not compute the human water use sectors. Our revised statement reads (please see subsubsection 5.2.1): *Irrigation water demand (potential irrigation water abstraction) is computed by nine models (Table S52).*

RC: Line 660: „**5.4 Examples of how parameterization can differ between GWMs**“ This section seems a bit sparse. The reader can easily see that different models use different parametrizations by looking at the Supporting Information and I am not entirely sure that merely repeating that information here is very meaningful. Perhaps this section can be improved by adding some form of quantification of the impact of these different parametrizations, to give the reader an indication of how much this can matter. This can be limited to highlighting a few (extreme) examples found in ISIMIP2b.
AC: Thank you for your recommendation. We moved this subsection to the supplementary information. We decided to keep this information because for beginners in global water

modelling these examples are useful to understand how parameterization can differ among 16 GWMs. We did not quantify the impact of these different parametrizations in ISIMIP2b. This is beyond the scope of the present study.

RC: Line 689: How is the number of water storage compartments calculated? Is this a count of the number of state variables (e.g. canopy water, soil water, etc) or a count of the number of layers that each state variable is discretized into?
AC: As mentioned in lines 686-687, we counted the number of water flows, storage compartments, and use sectors for each model participating in ISIMIP2b. Layers are not explicitly considered. Additionally, we have provided the figures 5 and 6, which show the "number of water flows, number of water storage compartments, number of human water use sectors". In the present study, we have provided definitions in Table S84 for each water flow, water stock, and human water use sector.
The old subsection 5.5 became section 6. In our new section 6, we describe our method.
Our revised statement reads:
One way of showing the model structures is to count the number of water flows, compartments, and human water use sectors included in each model participating in ISIMIP2b. For example, a model includes three water compartments if it computes canopy storage, soil storage, and snow storage. In this section we want to increase readers' awareness of model structures and offer the readers a final overview of how the models work, and how many water storage compartments, flows, and human water use sectors are included in their structures. We consider that two issues are useful to interpret model results, first, knowing model structures, and, second, identifying the effect of model structures on model results. However, the present study is focused only on the first issue, respectively, knowing model configurations needed to interpret various model results.

RC: Line 690: "LSMs and DGVMs have a relatively smaller number of processes (in this count), but each process is simulated in a more sophisticated way or has a physically based representation". This seems (in a more sophisticated way) a bit subjective. Suggest to rephrase.
AC: The revised statement reads: *LSMs and DGVMs have a relatively smaller number of processes (in this count and in this study), but each process has a mechanistic interpretation.*

RC: Line 690: "processes"
A: We corrected the sentence.

RC: Line 704: "6.1 Challenges in making this intercomparison study" This title can be more specific, seeing how this section nearly exclusively focuses on terminology. It also reads more like a method section to me, a precursor needed to introduce the standard writing style from section 3.2. It might make sense to move this section.
RC: Lines 710 – 716: This seems like it should be part of the definition or methods sections.
A: Thank you. We deleted this title and we restructured the manuscript. We created section 4 to describe our method and we created subsection 2.2 to present definitions used in global water modelling.

R: "Simulating the terrestrial water cycle on the global scale involves many challenges. These various challenges were identified by reviewing articles published by the climate, global hydrological, and vegetation communities (Table 11). The challenges have been classified according to the 23 unsolved problems in hydrology (UPH), identified by Blöschl et al., 2019, to harmonize the efforts of the global and catchment hydrological communities. These

challenges can generally be overcome through the development of new datasets, innovative and creative collaboration among communities, and investment in technical infrastructure." 6.2 Challenges in global hydrological modeling, Line 736 – 742: I agree that there are many challenges with global hydrologic modelling but I'm not entirely sure how this section connects to the rest of the paper and specifically the standardized writing scheme and the qualitative description of differences and similarities between models.
AC: Thank you. We moved this information to the new section 8. We moved Table 11 to the supplement (please see Table S97), to streamline the manuscript.
Our revised statement reads:
Certainly, simulating the terrestrial water cycle on the global scale involves many challenges, as we presented in this study. Other challenges have also been synthesized by reviewing articles published by the climate, global hydrological, and vegetation communities and have been classified according to the 23 unsolved problems in hydrology (UPH) identified by Blöschl et al., 2019 (Table S97). In summary, these challenges can generally be overcome through innovative and creative collaboration among communities and investment in technical infrastructure.

RC: "Wagener et al., 2020 well described the hydrological knowledge gaps as hydrologic lions, similar to the knowledge gaps of medieval maps represented as lions. They proposed focusing hydrological research on openly shared perceptual models, inclusion of metadata for each hydrologic study (e.g., location and time period covered by a study), and effective knowledge accumulation. In addition to these statements, we also propose focusing on effective collaboration that starts with effective wish lists, including specific research questions, goals to answer these questions, methods to achieve the goals, datasets to be used, and tasks to be done.", Line 467 – 752: Agreed, but again I'm not fully sure how this connects to the standardized writing scheme and the model descriptions that make up the rest of the paper. This reads like a recommendation based on doing a MIP, not one based on writing model equations in a standard way.
AC: We moved this information, on suggestions about how water modelling could be improved to the new section 8 (Recommendations for future multi-model intercomparison projects and extended assessments). We consider that Wagener et al., 2020 present some suggestions to improve water modelling. Our statement from above presents some suggestions to improve water modelling. Our statement is based on our experience achieved in doing the present study. We have started our present study using "wish lists" with the water stocks, water flows, human water use sectors that we would like to analyze. The "wish lists" are presented in Tables 1 to 5. This decision was made because of the models' complexity and we decided to share this insight here.

RC: Line 747: (2020)
AC: We have added brackets.

RC: Line: 758 – 759: "The review of GWMs, presented in section 4, highlights the need to design hydrological inter-model comparison studies by nominating models or research questions according to some specific criteria, for example,….." This statement may benefit from referencing e.g. Gupta et al. (2007, 10.1002/hyp.6989), Clark et al. (2011, 10.1029/2010WR009827) and/or Gupta et al. (2012, 10.1029/2011WR011044) who outline this need also and follow up with suggestions of how to make this possible.
AC: Thank you. We cited the recommended studies.

RC: Line 835: What I have missed so far is the following:

1. The numerical implementation of each model's code. This is known to cause (potentially large) differences between models (see e.g. Clark & Kavetski, 2010, 10.1029/2009WR008894) and is one the reasons for development of multi-model comparison frameworks (like those mentioned on line 371).
AC: Thank you for your recommendation. While we agree that the numerical implementation can be a reason for differences in model output, this specific item is beyond the scope of this study. Thanks to your comment, we suggest this research goal for a future study. In the introduction and conclusion, we mentioned our main goal to present similarities and differences among GWMs, on how they simulate the water cycle to facilitate a better interpretation of their various model results and have a better understanding of how these 16 GWMs work.
The new statement reads (please see section 8): *Another future study might focus on the numerical implementation of each model code.*

RC: A discussion about the impact of these differences and similarities between models. Possible questions to consider: What do the findings in this paper mean for model comparison studies and ensemble studies? Do these models form an unbiased ensemble or are they all lacking the same processes? Given the difficulty in finding commonalities between model terms, can we be sure that these things that now carry the same name actually represent the same process and correspond to the same process in reality?
AC: Thank you for your comment.
We mentioned that these similarities and differences found among 16 GWMs, based on their model equations, facilitate a better interpretation of their various results. Many studies present various model results, but they do not interpret these results by connecting them with the applied equations. This study analyses the equations of these 16 GWMs to simulate hydrological processes. We can be sure that these similarities and differences exist because we present model equations in the supplementary information that support our statements. Modelling teams developed and checked if these model equations correspond with model codes and / or peer-reviewed articles mentioned in Table 12. Some models are lacking the same processes, mentioned in section 5, but they include or exclude other processes resulting in different model structures that determine different model results. Some models simulate the same process by applying the same equation, but they have different parameter values contributing to different model results. This highlights our finding, mentioned in section 8, to make many experiments on parameterizations, equations applied, and model results.
We introduced the new statement, at the beginning of section 5.
Several studies highlighted the need to understand better modeling approaches, model structures, model equations, and similarities and differences among models (Zhao et al., 2017; Veldkampet al., 2018; Schewe et al., 2019). Therefore, in this section, we present some similarities and differences among 16 GWMs in simulating the terrestrial water cycle. This information enables us the interpretation of the different model results found in some model comparison and ensemble studies (Zaherpour et al.,2018; Wartenburger et al., 2018; Scanlon et al., 2019), as well as those by Gudmundsson et al., 2021; Reinecke et al., 2020; and Pokhrel et al., 2021. This information also strengthens our understanding of how these models work. Briefly, the 16 analyzed GWMs include in their structure similar hydrological processes, but they have different model structures.

RC: Line 836: I had expected some discussion about the standardized way of writing model equations. Some questions that come to mind are:
- Does this method have any weaknesses?
- How adaptable is this to models outside the set of 16 used here?
- How would a reader best go about writing their model in these terms?

AC: Thank you for your valuable recommendations.
We introduced new statements on our approach in section 9.
Our new statements are:
In summary, we mention that our approach was affected by models' complexity and is limited to eight water storage compartments and their flows, desalination, and five human water use sectors mainly, because of models' complexity. We conclude that the standard writing style of the equations is useful and necessary for finding similarities and differences among models for each water storage, human water use sector, and water flow. In addition, it can be leveraged for explaining the different model outputs, for classification of the models based on cluster analysis, and for selecting the right model for the right application. It can also be used for drawing a standard schematic visualization of the water cycle, for describing models on ISIMIP and ISIpedia platforms (the open climate-impacts encyclopedia, a part of the ISIMIP, https://www.isipedia.org/), and for understanding how models work. Other modelling teams can apply, in their studies, our lists with water storage compartments, flows, and human water use sectors and the symbols presented in the supplementary information. They can follow our steps in creating a standardized writing style of model equations and they may be aware of some challenges that could encounter. This study represents a roadmap in finding similarities and differences among models. However, it should be noted that these equations are available only for model versions used for ISIMIP2b.

RC: "We consider the simulations provided by the ISIMIP2b global water models to represent good hypotheses of our water future and based on them we can make decisions." Lines 886-887: Why? This study makes no mention at all of the simulations these models provide nor of the accuracy of the simulations of individual models.
AC: We need to estimate our water future for a better water management. The ISIMIP2b simulations are made with the analyzed 16 GWMs (Reinecke et al., 2020; Pokhrel et al., 2021, Gudmundsson et al., 2021) and they represent good hypotheses of our water future because they are provided by the 16 state-of-the-art GWMs. In the end, we deleted this sentence.

AC: References:
Beven, K: Rainfall-Runoff Modelling – The Primer, 2nd ed, Lancaster University, UK, Wiley-Blackwell, A John Wiley & Sons, Ltd., Plublication, 2012.
Beven, K., and Young, P.: A guide to good practice in modeling semantics for authors and referees, Water Resour. Res., 49, 5092–5098, https://doi.org/10.1002/wrcr.20393, 2013.
Bierkens, M.F.P., van Geer, F. C.: Stochastic Hydrology, Department of Physical Geography, Utrecht University, 237, http://www.earthsurfacehydrology.nl/wp-content/uploads/2012/01/Syllabus_Stochastic-Hydrology.pdf, 2007.
Brooks, R. H., Corey, A. T.: Hydraulic properties of porous media, Hydrology papers, Colorado State University Fort Collis, Colorado, 1964.
Burek, P., Satoh, Y., Kahil, T., Tang, T., Greve, P., Smilovic, M., Guillaumot, L., Zhao, F., and Wada, Y.: Development of the Community Water Model (CWatM v1.04) – a high-resolution hydrological model for global and regional assessment of integrated water resources management, Geosci. Model Dev., 13, 3267–3298, https://doi.org/10.5194/gmd-13-3267-2020, 2020.
Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., and Buontempo, C.: WFDE5: bias-adjusted ERA5 reanalysis data for impact studies, Earth Syst. Sci. Data, 12, 2097–2120, https://doi.org/10.5194/essd-12-2097-2020, 2020.

Döll, P. and Lehner, B.: Validation of a new global 30-min drainage direction map, J. Hydrol., 258, 214–231, https://doi.org/10.1016/S0022-1694(01)00565-0, 2002.
Gudmundsson, L., Boulange, J., Do, H. X., Gosling, S. N., Grillakis, M. G., Koutroulis, A. G., Leonard, M., Liu, J., Müller Schmied, H., Papadimitriou, L., Pokhrel, Y., Seneviratne, S. I., Satoh, Y., Thiery, W., Westra, S., Zhang, X., Zhao, F.: Globally observed trends in mean and extreme river flow attributed to climate change. Science 371 (6534), 1159-1162, doi: 10.1126/science.aba3996, 2021.
Lee, D. H. and Abriola, L. M.: Use of the Richards equation in land surface parameterizations, Journal of Geophysical Research – Atmospheres, 104, D22, 27519–27526, https://doi.org/10.1029/1999JD900951, 1999.
Farthing, M. W., Ogden, F. L., Numerical Solution of Richards' Equation: A Review of Advances and Challenges, Soil Science Society of America Journal, 81, 6, 1257 – 1269, https://doi.org/10.2136/sssaj2017.02.0058, 2017.
Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., Schumacher, M., Shadkam, S., Telteu, C.-E., Trautmann, T., and Döll, P.: The global water resources and use model WaterGAP v2.2d: model description and evaluation, Geosci. Model Dev., 14, 1037–1079, https://doi.org/10.5194/gmd-14-1037-2021, 2021.
Twarakavi, N., K. C., Sakai, M., and J. Šimůnek: An objective analysis of the dynamic nature of field capacity, Water Resources Research, 45, W10410, 1–9, https://doi.org/10.1029/2009WR007944, 2009.
Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A., Liu, J., Papadimitriou, L., Schewe, J., Müller Schmied, H., Stacke, T., Telteu, C.-E., Thiery, W., Veldkamp, T., Zhao, F., and Wada, Y.: Global terrestrial water storage and drought severity under climate change, Nat. Clim. Chang, https://doi.org/10.1038/s41558-020-00972-w, 2021.
Reinecke, R., Müller Schmied, H., Trautmann, T., Andersen, L. S., Burek, P., Flörke, M., Gosling, S. N., Grillakis, M., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Thiery, W., Wada, Y., Yusuke, S., and Döll, P.: Uncertainty of simulated groundwater recharge at different global warming levels: a global-scale multi-model ensemble study, Hydrol. Earth Syst. Sci., 25, 787–810, https://doi.org/10.5194/hess-25-787-2021, 2021.

Answer to Reviewer #2

RC: General comments:

1. The structure of the manuscript really needs some work and to be reconsidered.

The organisation of this manuscript doesn't seem to have been thought out thoroughly before writing. There is a lot of repetition in the sections. There is no real flow to the manuscript, which makes it difficult to follow and to see clearly what has been done, in what context, and what the outcome was.

My recommendation would be to rethink the structure of the manuscript, and clearly define each of the sections with clear headings. I think the review of the literature and models would be better if it came before the discussion of the creation of the standardised writing style, and this would set out the context and the issues that frame the need for the study.

AC: Thank you very much for your recommendations and your advice. We rethought the structure of the manuscript in order to exclude the unnecessary repetitions. We created new sections and decided on new titles.

Our new structure is:

1. Introduction

2. Modeling approaches and terminology used in global water modelling
3. Key characteristics of 16 global water models included in the study
4. Creating the standard writing style of model equations
5. Similarities and differences among 16 global water models
6. Number of water flows, water storage compartments, and human water use sectors included in the 16 GWMs
7. Potential future research of 16 global water models
8. Recommendations for future multi-model intercomparison projects and extended assessments
9. Conclusions

RC: 2. The actual substantial work that has been done, i.e. the standardising writing style for all of the global water models, needs more of a spotlight.
The description of the methodology used to create the standard writing style for the models lacks detail. This is the part of the manuscript that is substantial and novel, and what is being presented as the new contribution to the current knowledge and the hydrological modelling community. What I believe is missing is a detailed description of how you have standardised the model code. You mention in your manuscript that it was very difficult to find the similarities between the models especially between the different terms used. Therefore, without a clear methodology section, it is difficult for the reader to verify what you have done. It would be good to have a discussion of what the strengths and weaknesses of your approach are, and what are the issues/difficulties that you encountered along the way and how you managed to overcome these. Do you intend your method to be used beyond the models that are a part of ISIMIP2? Is this work intending to set out a roadmap, of sorts, for future work of a similar nature? I think to fix this issue, and to link with the previous comment, a dedicated methods section should be added to the manuscript.
AC: Thank you very much for your recommendations. We created section 4 to describe our approach to create a standard writing style for the GWMs. We mention that we did not standardized the model codes, instead, we standardized the writing style of the equations applied by 16 GWMs to simulate eight water stocks, water flows, desalination, and five human water use sectors, mainly, because of models' complexity.
Our revised statement reads:
In summary, our three main goals are:
• to provide a better understanding of how 16 state-of-the-art global water models are designed;
• to show similarities and differences among them, based on their equations;
• to underline future research potential in global water modeling.
Therefore, in our new section 4, we describe the steps followed to standardize the writing style of model equations, present challenges encountered. In conclusions (section 9), we explain how other modelling teams can use our approach.

RC: 3. This manuscript is trying to do far too more.
Your manuscript covers a lot of different areas. It covers the standardising of the model code, tried to review the literature of multi-model intercomparisons, and tried to look at the challenges that are face by the global hydrological modelling research community and attempts to provide recommendations for the future. I believe that each of these elements are very important, but for one paper to tackle each of these seems quite excessive. I think the review of the literature is valuable, but it should form part of the introduction to the work, and 'set the scene' for the substantial work that has been carried out. I also think that you should consider cutting the manuscript down substantially.

Section 5 is particularly long and it is almost impossible as a reader to take in all the information that is given. It would be much better to present these similarities and differences in the model in a more visual way. This section would also benefit from having less description of the models and more of a discussion of the implications of the similarities and differences that are found, e.g. do all the models seem to miss certain types of processes/model components?

Below are some specific comments that I had about each of the sections.

AC: Thank you very much for your recommendations. We rethought and changed the manuscript structure. In the new section 3, we present the models analyzed in this study. We moved challenges faced by global water modelling (Table S97) to the supplementary information because we consider that this table harmonizes the efforts of the global and catchment hydrological communities. We believe our readers need to be aware of existing challenges in water modelling to understand better different modelling approaches.

Specific comments:

RC: Abstract:

On page 2 line 53-55: "Our results highlight that the predictive uncertainty of GWMs can be reduced through improvements of the existing hydrologic processes, implementation of new processes in the models, and high-quality input data." – there is no mention in the abstract of how these kinds of conclusions are going to be made, what tests on the models performance are done, etc.

AC: We have mentioned in the abstract in the Lines 44-46: "This study provides a comprehensive overview of how 16 state-of-the-art GWMs are designed. We analyze water storage compartments, water flows, and human water use sectors included in 16 GWMs that provide simulations for the Inter-Sectoral Impact Model Intercomparison Project phase 2b (ISIMIP2b)."

Our revised statement reads:

Ultimately, we consider that similarities and differences found among the models analyzed in this study enable us to reduce the uncertainty of multi-model ensembles, to improve the existing hydrological processes, and to integrate new processes.

Our statement is based on our analysis, made through this study, on model structures and model equations applied to simulate hydrological processes. We identified similarities in hydrological processes among 16 GWMs (section 5). Our general conclusion is that the uncertainty of GWMs can be reduced through improvements of the existing hydrologic processes, implementation of new processes in the models, and improvement of the input data. We will have better results if we improve the models. For our general conclusion, we consider that we do not need tests on model performance.

RC: Introduction:

The first paragraph is far too long. It would benefit from being broken into smaller parts. There is no clear focus either. It would be great if you could define clearly straight away what the paper is going to be about and the general context its in.

Very long intro, consider streamlining, and concisely summarising previous MIPs, the usefulness of MIPs, and what has been learned from previous work. Use this as an opportunity to frame the context of the work that will follow.

Maybe more information about hydrological MIPs would be useful. Page 3 Line 84 is the first time that you mention you are going to be comparing global water models.

Page 3, line 83-85: use this as an opportunity to clearly state what exactly has been done in the work and how you have gone about achieving that. State the clear questions that have been identified from the literature review and what the gaps in the current knowledge are.

AC: Thank you very much for your constructive comment. We realized the present study through a model intercomparison project (MIP) – ISIMIP2b. Generally. MIPs have been designed to inter-compare models. However, they focused on evaluating models' performance in the past and models' agreement for the future. Few of them focused on the interpretation of various model results. We need to analyze model structures and equations and discover similarities among models to interpret various model results.

We broke the first paragraph into smaller parts, each part has its purpose. We present the scientific context and the main aims of this study. We included a new paragraph:

In this complex scientific context, the present study represents a step forward to increase understanding of process representation and inter-model differences within one large MIP, specifically, ISMIP – the Inter-Sectoral Impact Model Intercomparison Project (Frieler et al., 2017). We assessed the equations applied by 16 state-of-the-art global water models (GWMs) to simulate the vertical and lateral water balance, human water use sectors, and desalination on the global scale. We created a standard writing style of these equations to identify similarities and differences among models. Thereby, the global water community has through this study an overview of the model structures and the basis required to interpret various model results, to design future experiments on how model equations, model configurations, and model parameter values influence the model outputs.

In summary, our three main goals are:

- *to provide a better understanding of how 16 state-of-the-art global water models are designed;*
- *to show similarities and differences among them, based on their equations;*
- *to underline future research potential in global water modeling.*

Section 2:

RC: Page 3, line 96-100: "The terrestrial water cycle is simulated globally by three different communities that have developed three types of models: (i) the climate community has developed land surface models (LSMs); (ii) the global hydrological community has developed global hydrological models (GHMs); (iii) the vegetation community has developed dynamic global vegetation models (DGVMs). These communities interact with each other, but generally focus on specific hydrological processes that are important for their research and are presented in the following subsections."

I think these communities are being brought together more in recent years and the boundaries between them are becoming blurred.

AC: We consider that these communities collaborate, but they have different research questions and focus on specific hydrologic processes. Therefore, these communities need a strong interaction to improve the models.

RC: 2.1. differences in modelling approaches:

Some examples of LSMs, GHMs and DGVMs along with the descriptions would be beneficial for the reader.

AC: Thank you for your comment. We changed the purpose of subsection 2.1.

In subsection 2.1, our new goal is to explain the different modelling approaches existent in global water modelling. We consider that the reader needs to understand the different research aims existent among climate community, global hydrological community, and vegetation community. Being aware of these differences among communities, the reader will better understand the difference between a land surface model (LSM), a global hydrological model (GHM), and a dynamic global vegetation model (DGVM). Furthermore, the reader will be able to understand better why models have different structures and why models apply different equations. Therefore, according to the new paper structure, we present examples in section 3, where we present the key characteristics of the models used in the present study.

RC: Page 4, line 109: “The global hydrological community is focused on surface hydrologic processes, primarily river flow simulation and its daily to century-scale changes“ you say daily, but many GHMs run at sub-daily time steps.
AC: We wrote the main goal of GHMs which are typically run with daily time steps. For most GHMs, daily time steps are still state-of-the-art, whereas for LSMs, sub-daily time steps are standard. Also in reference to referee 1, the revised statement reads:
The global hydrological community focuses primarily on surface water and groundwater availability, its human interference, and their daily to century-scale changes.

RC: The interchangeable use of LSMs and the climate community to me becomes confusing. Consider choosing one and sticking with it throughout the text.
AC: Thank you for your comment. We revised subsection 2.1. Our new goal is to describe the research purpose of each community that is performing water modelling and to mention the models that each community has developed to study their specific research questions. Therefore, we consider it necessary to emphasize the link between the climate community and LSMs.

RC: 2.2. ambiguity of terminologies used in hydrological modelling
This section is definitely a good addition to this manuscript and useful for the read. However, this section lacks a real structure and focus. There are only 2 examples give (active vegetation and dynamic vegetation). This would potentially benefit with some smaller subsections with more examples of terminology ambiguities and how they are dealt with in the study.
AC: We changed the goal of this subsection, based on the recommendations received to restructure the manuscript and to define terms used in water modelling (please see comments and answers of referee #1). Because ambiguity of terminologies used in water modelling could be discussed in several studies (e. g., please see the example received from referee #1: Brunner et al., 2018), the new goal of subsection 2.2 is to present definitions used in global water modelling. We define terms used in water modelling, necessary to understand how the 16 analyzed global water models work. We believe that we have reached now a more systematically description of the modelling chain.

RC: “In the end, because of differing and complementary perceptions and details of their models, it is important that these communities interact, identify their similarities and differences, share experiences, learn from different experiments, undertake joint experiments, present and discuss their results, and discuss how they influence and depend on each other and how hydrological modeling can be improved. Therefore, collaboration among these communities will result in new multi-model intercomparison projects and multi-model ensembles that will facilitate new analyses, comparisons, understandings, and improvements.” The final paragraph doesn’t really fit with the rest of the information in the section. Again, consider restructuring and redefining what it is that you want to accomplish with this section.
AC: We improved and moved the statement to subsection 2.1. In the new subsection 2.1, we present differences in water modeling approaches.
Our new revised statement reads:
On global scale, the terrestrial water cycle is simulated by three different communities that have developed three types of models: (i) the climate community has developed land surface models (LSMs); (ii) the global hydrological community has developed global hydrological models (GHMs); (iii) the vegetation community has developed dynamic global vegetation models (DGVMs). However, these three communities focus on specific hydrological and atmospheric processes, as well as anthropogenic impacts. These key aspects are important for their specific research leading to different modelling approaches, specific evaluation studies

of model performance (Archfield et al., 2015), and different field-specific meanings of terminology used (Beven and Young, 2013). Thus, combining the expertise in their key aspects would create a strong synergy and improve the models of these communities, but for this goal, they have to interact with each other, identify their similarities and differences and share experiences. They need to undertake joint experiments, present and discuss their results, discuss how they influence and depend on each other, and how water modeling can be improved (Cucchi et al., 2020).

RC: "Generally, it is recommended to include this process in models because elevated CO_2 concentrations cause physiological and structural effects on plants and indirectly influence runoff and evapotranspiration over a geographical area. The physiological effect reduces the opening of leaf stomata because less water is needed to assimilate carbon, leading to decreased transpiration and, indirectly, increased runoff. The structural effect or fertilization effect causes an increase in plant growth and leads to increased transpiration per unit area and, indirectly, a decreased runoff (Gerten et al., 2014). However, Singh et al. (2020) demonstrated that increased leaf area under elevated CO_2 concentrations (structural effect) might counterbalance the increased water use efficiency (physiological effect)."
Page 5, line 144-149. Not sure if this information fits well in this section. Supposed to be discussing the ambiguities of terminology but instead are talking about why its important to include active vegetation in a model?
AC: Thank you. We moved this paragraph to the supplementary information, to streamline the manuscript.

RC: 3.2. Steps taken to realise the standard writing style of model equations
The first paragraph is all good information, but it might be better placed in the ambiguity of terminology. That would allow for a more description and definition of the terms that you have used.
AC: Thank you. We moved the paragraph to subsection 2.2, where we define terms used in global water modelling.

RC: "Parameters may change in space, but do not change in time. Parameters and coefficients represent numbers that describe a particular characteristic of reality, of the model, of the catchment area or flow domain such as runoff coefficient, soil porosity, hydraulic conductivity of different soil horizons, maximum soil water storage, maximum canopy water storage, and mean residence time in the saturated zone (Beven, 2012). Some processes are parameterized, meaning that their values are precisely marked in the computer code and are not calculated by the model itself."
Page 6, line 191: could give some examples of parameterisation methods/techniques, include some relevant references.
Page 6, line 186 and line 189: examples would be beneficial for the reader.
AC: Thank you. We revised the subsection 2.2. We define terms used in water modelling, necessary to understand how the 16 analyzed global water models work. We believe that we have reached now a more systematically description of the modelling chain.
Our revised statements are: *Parameters and coefficients represent numbers that describe a particular characteristic of reality, of the model, of the catchment area or flow domain. Some examples are runoff coefficient, soil porosity, hydraulic conductivity of different soil horizons, maximum soil water storage, maximum canopy water storage, mean residence time in the saturated zone, surface roughness, and vegetation properties (Beven, 2012).*
Thus, a water global model includes many equations written with a programming language in a model code to simulate freshwater systems.

We define parameterization as changes of model parameter values (Samaniego et al., 2010). We present other definitions of water flows, stocks, desalination, and human water use sectors, used in this paper work, in the supplementary information.

RC: "3.2. Steps taken to realise the standard writing style of model equations"
I think this section is missing something about how you checked and made sure that the translation of the model code was in fact correct and robust. And provide this for the reader so that they have a way to validate what you have done.
AC: We agree that an automated way of translating model code into standardized equations would be a very useful tool, especially in terms of reproducibility. However, such a method does not exist currently and would most probably require the modellers to rewrite parts of the model code - which is way beyond the scope of the study. Thus, our method to ensure correct and robust equations is the same as in all model documentations: the modelers themselves checked it repeatedly and furthermore the plausibility of the equations were evaluated by the authors compiling the tables.
Therefore, we restructured the manuscript. We created section 4 where we describe our approach. We mention that the equations from the supplementary information have been checked by modelling teams. This was reached with many revision rounds until all authors agreed on the representation of the models in the equations. Nevertheless, many of these equations were published in peer-review articles mentioned in Table 12, although with a different notation as the one used in this study.
The revised statement is, subsection 4.5:
In the next step, modelling teams created and provided the model equations, used to provide simulations for ISIMIP2b, according to the generated lists. Each modelling team, involved in this study, internally checked and reviewed its model, based on the model code and peer-review articles mentioned in Table 12 or only on the peer-review articles on model description mentioned in Table 12. In some cases, modelling teams provided the equations using our standard writing style and symbols presented in subsection 4.4, while in other cases using their specific writing style. Therefore, the modelling teams checked the model equations on their plausibility.

3.3. Key characteristics of the global water models
Extremely lengthy description of the models. All of this information is given in the tables and in the supplement. I know that a description like this is needed, but consider cutting this text back to make it more readable. Very easy to start getting a bit lost.
A: Thank you for your comment. We now improved the text based on the valuable comments received during the review process. In our new section 3, we created subsections to present key characteristics of 16 GWMs. We consider it necessary to provide this information because these characteristics facilitate a better understanding of how the 16 global water models work.

RC: "CWatM calibrates monthly or daily streamflow for 12 catchments using the Distributed Evolutionary Algorithms in Python (DEAP) approach (Burek et al., 2020), while WaterGAP2 uses a beta function for the calibration of 1,319 gauged hydrological stations considering runoff as a nonlinear function of soil moisture. WaterGAP2 uses a runoff coefficient and two correction factors to calibrate the simulated and observed streamflow (Müller Schmied et al., 2014). "
Talk about using 12 catchments and 1319 gauging stations for calibration, but then how is this used? Is there some method of regionalisation of the parameters?

AC: We mentioned the Distributed Evolutionary Algorithms in Python (DEAP) approach for CWatM and we have now provided additional information for WaterGAP2. We also provided references.
Please see our answer to referee #1. Thank you.
Indeed, the calibrated parameters have been regionalized. For WaterGAP2, one of the three calibration parameters is regionalized using a multiple linear regression approach considering a number of basin descriptors (Müller Schmied et al., 2021, their Sect. 4.9.2). For CWatM, an evolutionary algorithm with KGE as objective function was applied and WFDEI meteorological data were used as forcing (Burek et al., 2020). We cited Burek et al., 2020 and Müller Schmied et al., 2021 for detailed information.

RC: Section 4: Not sure if 'review' is the best word here. Makes the reader think you are going to talk through the strength/weaknesses and similarities/dissimilarities of the models. First paragraph may be better in the introduction. Could switch out some of the information in there already and change it for this. Some of the info in the introduction isn't really relevant to the state of global hydrological models/water models.
AC: Based on the recommendations received, we restructured the manuscript and revised section 4. In the new section 4, we describe our approach and what steps we followed to realize the standard writing style of model equations. Therefore, we moved information about models' strengths and weaknesses to the supplementary information, to streamline the manuscript. We improved and moved the first paragraph, with information on the history of global water modelling, to the supplementary information, to streamline the manuscript.

RC: 4.1. evaluation of global water model to observations
This section feels more like a list of different studies and what they found. Not really sure how much value it adds to the manuscript. A review of what has been done on these models is important, but this section could be more concise to cut down some text. Some of this information could be incorporated into different sections. Maybe this review of studies should come earlier in the manuscript? Maybe put it as section 3 instead?
AC: Thank you. We moved this information to the supplementary information, to streamline the manuscript.

RC: 4.3. uncertainties in global water models
"Multi-model intercomparison studies showed a significant variation in the model results. One explanation could be that global hydrological modeling imposes uncertainties from forcing data, model parameters, processes included or excluded, and numerical algorithms used. Additionally, each modeling group has a different model development concept and purpose." – Page 11, line 352-354 need a few references to some of these studies.
Again, feels too much like a list of studies and what they found. I think the paper would be a lot more useful if these findings were written in a way that makes the work that is being presented make more sense.
AC: Thank you. We moved this information to the supplement, to streamline the manuscript.

RC: Page 12, line 366-367: "therefore, these is a need to better understand the models' structure complexity, their equations, and their approaches, and to improve the quality of the input data" this is a good example of explaining how the things that you have written about from the literature review link back to the work that you have done!!
AC: Thank you! We revised the paragraph and introduced it at the beginning of section 5. Our revised statement reads:

Several studies highlighted the need to understand better modeling approaches, model structures, model equations, and similarities and differences among models (Zhao et al., 2017; Veldkampet al., 2018; Schewe et al., 2019). Therefore, in this section, we present some similarities and differences among 16 GWMs in simulating the terrestrial water cycle. This information enables us the interpretation of the different model results found in some model comparison and ensemble studies (Zaherpour et al.,2018; Wartenburger et al., 2018; Scanlon et al., 2019), as well as those by Gudmundsson et al., 2021; Reinecke et al., 2020; and Pokhrel et al., 2021. This information also strengthens our understanding of how these models work. Briefly, the 16 analyzed GWMs include in their structure similar hydrological processes, but they have different model structures.

RC: Page 12, line 369: this is more commonly referred to as Rainfall-Runoff Modelling Toolbox
(RRMT).
AC: We corrected the sentence.

RC: Could be good to add some insight into what has been learned from these different methodologies created by the catchment community.
page 12, line 373-382: talking about the different approaches to parameterisation. This is the more detailed description of parameterisation approaches that would have been beneficial in section 3.2. maybe a little bit repetitive? I would consider moving some text around and combining sections to streamline the manuscript.
Talking about the methods that have been developed by research communities (i.e. flexible frameworks and parameterisation methods) - is this really about the uncertainties, or is this a way of dealing with the uncertainties?
Talking about the catchment modelling community at the end of the section. It might be good to introduce the catchment modellers somewhere formally to talk about how they are a part of the model development (Archfield et al. (2015), doi:10.1002/2015WR017498) good example of a paper that has a similar discussion.
AC: Thank you for your recommendations. We revised and moved lines 369 – 387 to section 8, where we present recommendations for future multi-model intercomparison projects and extended assessments.
Please see our response to Reviewer #1: We added Archfield et al. (2015) in subsection 2.1, where we present three communities that simulate terrestrial water cycle. These communities focus on specific hydrological and atmospheric processes and anthropogenic changes that are important for their specific research resulting in different modelling approaches.

RC: Section 5:
Page 12, line 389: change this from ‘among models’ to ‘models used in this study’ or something else along these lines. Currently this reads ambiguously. Make it clear that you are talking about the ISIMIP models.
AC: Thank you for your recommendation. We have corrected our statement.

RC: 5.1. Similarities and differences in simulating the water storage compartments.
For me this section is just too long and too dense. There is too much information, and it is nearly impossible as a reader to retain much of it. Unless you were reading this to look for specific information, it is too detailed. Consider cutting this text back substantially. Also consider thinking of a way to show this visually.
This is mostly just a description of the hydrological cycle, and then which of the models represents these features and how. Not sure how valuable all of this information is to a reader.

AC: Thank you for your comment. We realized this study within the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP) framework. ISIMIP2b was designed to inter-compare impact models included in the global water sector. Modelling teams of the ISIMIP2b global water sector realized many peer-review studies to evaluate models' performance in the past and models' agreement for the future and obtained different model results (e. g., Reinecke et al., 2020; Pokhrel et al., 2021, Gudmundsson et al., 2021). Consequently, making these studies, we identified the need to analyze model structures and equations and to discover similarities among models to be able to interpret different model results and better understand how these 16 GWMs work.
Therefore, we consider that the present study represents a step forward to:
- interpret different model results;
- increase understanding of process representation and inter-model differences;
- find new ways to improve the models;
- find new ways to make an intercomparison study;
- design future experiments on how model equations, model configurations, and model parameter values influence the model outputs.

The present study presents some similarities and differences among 16 GWMs that are necessary to interpret the different model results from Reinecke et al., 2021, Pokhrel et al., 2021, Gudmundsson et al., 2021, as well as of future ISIMIP2b studies.
For a better understanding of how similar or different are 16 GWMs, we included Figures 1 to 6 with the number of analyzed GWMs that simulate water stocks, water flows, and human water use sectors.
Yes, this section represents a description of the hydrological cycle because we have analyzed 16 GWMs based on how they simulate the water cycle, as we mentioned in introduction. The 16 GWMs include six land surface models (LSMs), nine global hydrologic models (GHMs), and one dynamic global vegetation model (DGVM). These 16 GWMs have been developed by three communities that focus on specific hydrological and atmospheric processes and anthropogenic changes that are important for their specific research resulting. Please see our comment above.

RC: 5.2. similarities and differences in simulating human water use sectors. Same comments as section before.
AC: Please see our comment above.

RC: 5.4. examples of how parameterisations can differ between GWMs
The opening for this section is repetitive of information previously.
Stating that the parameterisation and model structure lead to differing results. This is repeating previous points, and maybe stating the obvious? This section would benefit from a few examples of different parameterisation methods that have been used by a few of the models, and how much of an impact that has had on their simulations. I think this would make this section more valuable to the reader. As it is, I'm not sure of its relevance or impact.
AC: Thank you for your recommendations. We moved this information to the supplement, to streamline the manuscript. Please see our response to Reviewer #1.

RC: 5.5. How many water flows, water storage compartments, and human water use sectors are included in the GWMs?
How are you defining these compartments here? Are you counting all of the different processes etc, or are you counting the layers that are in the model structure? I think this is a bit confusing.
Telling the reader the number of compartments that each model has might not be that informative? It might be better to say what the effect of these different compartments are? Are

models with more/less better/worse? What is the effect?
AC: Thank you for your comment. We mentioned in the text that we counted the water compartments. The old subsection 5 became the new section 6. This section increases readers' awareness of model structures and enables the readers to have a better understanding of how models work. It offers a final overview of how many water stocks, flows, and human water use sectors are included in the models. We consider that both issues are useful to interpret model results, first, knowing model structures, and, second, identifying the effect of model structures on model results. The present study focuses on the first issue – knowing model structures. We consider that the effect of different water compartments on model results represents the goal of a future study. We wrote this in the new section 8.
The new statement reads: *One way of showing the model structures is to count the number of water flows, compartments, and human water use sectors included in each model participating in ISIMIP2b. For example, a model includes three water compartments if it computes canopy storage, soil storage, and snow storage. In this section, we want to increase readers' awareness of model structures and offer the readers a final overview of how the models work, and how many water storage compartments, flows, and human water use sectors are included in their structures. We consider that two issues are useful to interpret model results, first, knowing model structures, and, second, identifying the effect of model structures on model results. However, the present study is focused only on the first issue, respectively, knowing model configurations needed to interpret various model results.*
Our new statement reads (please see section 8): *Another future study might focus on the numerical implementation of each model code.*

RC: Section 6:
6.1. challenges in making this intercomparison study
This section pretty much just talks about the challenges of terminology. Maybe it would be better if it was moved to the other part in section 3 where the challenges of terminology is introduced. I think quite a lot of this would be better as a part of a methods section. You have identified all the challenges and problems that you have come across and then explained how you overcame them. This is important in the defining of the standardised model equations. Consider moving all of this into a methods section.
AC: Thank you for your recommendations. We created section 4 to present our method and definitions used in the present study.

RC: 6.2. Challenges in global hydrological modelling.
Not sure what this has added to the article? Its very brief and doesn't really give any information to the reader. Consider removing. I think you are trying to add too much into this paper. Not sure how this is important. Yes there are many challenges in global hydrological modelling, but these should be the topic of focus of another piece of work all together.
AC: We moved this information to the new section 8, because we consider that the readers need to be aware of some challenges in global water modelling. We moved Table 11 to the supplementary information because it synthesizes challenges in water modelling that are useful to understand better water modelling.

RC: 6.3. recommendations for multi-model intercomparison projects. I think this whole section could be removed. Here I think you are trying to give recommendation for what MIPs should do in the future. However, the work that you have done and have presented is a dataset of standardised model equations. Again I think this is an example of where this paper is being over ambitious with its content.
Page 24, line 764-772: this is a good bit of this section. This is a good way of recommending to model developers etc how they can make their models more transparent. This also sort of

creates a roadmap for how it would be possible to standardise models of the future. This section could be given a new heading, and make it more clear that this is the recommendation that is being made so that the work that you have done (i.e. standardising model code) can be reproducible in the future.
Page 23, line 773-775: here you have identified what this paper is really lacking and what would make this a lot more palatable for the reader. Also easier to read and follow in general.
AC: Thank you for your recommendations. In our new section 8, we present our « learning by doing » experience and we want to share this. We also present our recommendations for future MIPs. There are many challenges in harmonizing information about model structures and in making a dataset of standardized model equations. We consider that this information is very useful for future studies on water modelling. Generally, before starting a study, it is very useful to know how others have done a standard writing style of model equations and what challenges have encountered. This also saves time in doing a similar study. Therefore, we decided to keep this information in the present study and share our experience with our readers.

RC: Page 23, line 753-757:" We encourage communities to write and convey clear, simple, and understandable texts for large audiences. We consider that simplicity improves communication, and communication starts with a common language, the same words having the same meaning for the sender and the receiver. Theoretically this is possible, but in practice, there are some discrepancies among scientists (highlighted in subsection 2.2), as well as between scientists and stakeholders, as revealed by Sultan et al. (2020)."
this is good info. This shows that there is a good motivation for the work that you have done. This should go into the introduction or in a section where you are talking about the different research communities.
AC: Thank you very much for your recommendations. The old subsection 6.3 became the new section 8. We decided to keep this info in section 8 because we encourage communities to write and convey clear, simple, and understandable texts for large audiences, future MIPs and studies.

RC: Section 7:
You say that this was done to find the similarities and differences, which you have done and presented. I think what is lacking is a flavour of effect/impacts that this has.
AC: We consider that the effects / impacts represent the goals for future studies that need to include the model results for a better understanding of these effects. This is beyond the scope of the present study. We mention our goals in Introduction.
In summary, our three main goals are:
- to provide a better understanding of how 16 state-of-the-art global water models are designed;
- to show similarities and differences among them, based on their equations;
- to underline future research potential in global water modeling.

RC: "We consider the simulations provided by the ISIMIP2b global water models to represent good hypotheses of our water future and based on them we can make decisions."
Page 27, line 886-887: the way you have rounded off this paper makes no sense, as you have presented no information about the performance of the simulations of the models themselves.
AC: Please see below our response to Reviewer #1. Thank you.
We need to estimate our water future for a better water management. The ISIMIP2b simulations are made with the analyzed 16 GWMs and they represent good hypotheses of our water future because they are provided by the 16 state-of-the-art GWMs.
In the end, we deleted this statement.

SC: Comment posted online by Charles Vörösmarty.

In the opening paragraph to section 4 (Review....), the authors present a short summary of some key developments in global water modeling. They also state on lines 290-92 "....Dooge (1982) identified the two major challenges of global hydrology: scaling and parameterization. Eagleson (1986) declared the necessity of global-scale hydrology. Inevitably, during the 1990s, the first global hydrological models were developed (Alcamo et al., 1997; Vörösmarty et al., 1998, Arnell, 1999)."

While I realize the purpose of the current work is not to present an exhaustive review, the authors' statement assigning the decade of the 1990's to the development of the first such models is historically incorrect. These models, as well as essential inputs, calibration/validation data sets, and modeling application studies were in fact first developed during the 1980s, motivated in no small measure by the proposals made by Dooge and Eagleson. I provide a short list of publications that support this assertion. All of these pre-date, and some substantially, the first paper in the list which appeared in the late 1990s (Alcamo et al. 1997--which I note parenthetically appears not to have been published in the peer-reviewed literature).

In keeping with the comment of Dooge on calibration and scaling I believe that the paper by Federer et al. 1996 might be particularly relevant to cite. It is also important to note that without subdstantial effort to create digital archives for calibration and validation data, the community's progress toward a global-scale capability would like have languished for quite some time. For this reason, I include in the list a global hydrological data compendium that was broadly adopted by the community for this purpose after it was made available in 1996. It might also be noted that the first global-scale application study of the impact of hydraulic engineering (i.e., on dams and reservoirs) was published in 1997; an absolute requirement was the use of these first generation models and their supporting digital hydrologic data archive.

I would anticipate that the authors to be kind enough to acknowledge this shortcoming.

Gildea M.P., B. Moore, C.J. Vörösmarty, B. Berquist, J.M. Melillo, K. Nadelhoffer, and B.J. Peterson (1986). A global model of nutrient cycling: I. Introduction, model structure and terrestrial mobilization of nutrients. In: Correll, D. (ed.), Watershed Research Perspectives. Smithsonian Institution Press, Washington, D.C.

Vörösmarty, C.J., M.P. Gildea, B. Moore, B.J. Peterson, B. Berquist, and J.M. Melillo (1986). A global model of nutrient cycling: II. Aquatic processing, retention, and distribution of nutrients in large drainage basins. In: Correll, D. (ed.), Watershed Research Perspectives. Smithsonian Institution Press, Washington, D.C.

Vörösmarty, C.J., B. Moore, M.P. Gildea, B. Peterson, J. Melillo, D. Kicklighter, J. Raich, E. Rastetter, and P. Steudler (1988). A global, georeferenced model of hydrology and water quality applied to the Amazon Basin. In: Nittrouer, C. and D. DeMaster (eds.), Proc. of AGU Chapman Conference on the Amazon Dispersal System. Charleston, SC.

Vörösmarty, C.J., B. Moore, M.P. Gildea, B. Peterson, J. Melillo, D. Kicklighter, J. Raich, E. Rastetter, and P. Steudler (1989). A continental–scale model of water balance and fluvial transport: Application to South America. Global Biogeochemical Cycles 3: 241-65.

Vörösmarty, C.J., B. Fekete, and B.A. Tucker (1996). River Discharge Database, Version 1.0 (RivDIS v1.0), Volumes 0 through 6. A contribution to IHP-V Theme 1. Technical Documents in Hydrology Series. UNESCO, Paris.

Federer, C.A., C.J. Vörösmarty, and B. Fekete (1996). Intercomparison of methods for potential evapo-transpiration in regional or global water balance models. Water Resources Research 32: 2315-21.

Vörösmarty, C.J. K. Sharma, B. Fekete, A.H. Copeland, J. Holden, J. Marble, and J.A. Lough (1997). The storage and aging of continental runoff in large reservoir systems of the world. Ambio 26: 210-19.

AC: Thank you very much for your constructive comment and recommendations, as well as for compiling a list of papers relevant for the history of GWMs development. We revised our statement. We moved the paragraph to the supplementary information to introduce the definitions in water modelling and streamline the new paper structure.

Our revised statement reads:

Global water models (GWMs) were developed from the earliest land surface models (LSMs) created by Thornthwaite and Mather (1957), Manabe (1969), Freeze and Harlan (1969), and Deardorff (1978). These first land surface models simulated the terrestrial water cycle by considering vegetation processes, evaporation, soil moisture, and snow cover. Later on, during the 1980s, the first global hydrological model (GHMs) was developed, with a spatial resolution of 0.5°x0.5°, accompanied by its essential inputs, calibration and validation datasets, and modeling application studies, thereby, emphasizing the necessity of global-scale hydrology (Vörösmarty et al., 1989). Dooge (1982) identified the two major challenges of global hydrology, scaling and parameterization, and concluded that a global scale model requires prudent simplification (Dooge, 1986).

In the 1990s, LSMs were intensively improved (Sellers et al., 1997), other GHMs were developed (Yates, 1997; Arnell, 1999), experiments on parameterization were done (Federer et al., 1996), and a global hydrological data compendium was made publicly available for model calibration and validation (Vörösmarty et al., 1996). Furthermore, Vörösmarty et al., 1997 assessed globally the impact of hydraulic engineering (i.e., on dams and reservoirs) on the river systems. This decade is remarkable through the community's progress toward a global-scale capability.

In the 2000s, other studies appeared on the hydrological calibration (Nijssen et al., 2001; Doll et al., 2003), human impact schemes (Alcamo et al., 2003; de Rosnay et al., 2003), and vegetation dynamics, including CO2 fertilization effects (Gerten et al., 2007). Ultimately, over the years, many global water models have been developed and improved and many studies have been done to assess freshwater resources on the global scale (Bierkens, 2015).