Reply to RC1 (Wouter Knoben)

Paper summary

This paper summarizes the outcomes of a large effort to create graphical representations of the inner workings of the 16 models that participated in ISIMIP2b. The result is a pair of diagrams, where the first shows in a 3-dimensional way (A) how the model discretizes the vertical domain (e.g. snow, canopy, soil, groundwater), (B) which lateral (surface) components it includes (e.g. lakes, reservoirs, wetlands), and (C) which human water use sectors are included (e.g. agriculture, livestock, industry). If a model doesn't include a given component, that component is greyed out. The paper briefly describes the process that led to the creation of these diagrams, possible ways in which the diagrams could be used, and some thoughts about creating these diagrams for models not included in ISIMIP2b.

Review summary

I think this paper is an interesting and timely contribution, and I expect this process was anything but easy. Better understanding which models to use when, where and for what purpose is critical for actionable decision making, and these diagrams might be helpful in both conversations between model developers and model users, as well as outline experiments that would lead to such better understanding. The graphics are clean and the paper is generally easy to read. However, I think some important information is missing and I think the paper needs to be revised before it can be published.

Answer: Thank you for your time to review the manuscript and for providing your comments and suggestions. We will reply to the referee comment, indicated by **R1** (in black)**,** by our answer indicated by **Answer** (in green), and corresponding actions, indicated by **Action** (in blue) and textual changes in *italic font*:

RC1: Major comments

I have several major comments, based on my reading of the paper and multiple line-by-line comments that can be found in the attached .pdf:

RC1: 1. The main methodology used in this paper seems to be that everyone involved went through a long process of deliberations about how the final diagrams should look. I'm not very familiar with how such social processes are typically documented and described in journal articles, but the current description of it in the paper is very short: there is almost nothing about the process beyond its outcomes. However, these discussions lie at the heart of the resulting diagrams and I think more description of how they were organized, which stakeholders and backgrounds were present, how different points of views and needs were balanced etc. is needed. I think the paper in particular needs more information about how the main trade-offs between accuracy/detail and aesthetics/clarity were made, and why the resulting two diagrams are seen as the right balance between these different things. Are there transcripts of the conversations that were had?

Answer: Thank you for the valuable comment. The whole process, or the study-design from beginning to the finalization of the diagrams was not specifically pre-designed but was a

community-based effort that was initiated and maintained by (a group of) key persons (who changed over time). In such a large group of contributors felt that there needs to be a steering group who determines the way forward, and here the coordinators of the global water sector of ISIMIP played a big role. At some stage in the development, the funding for the graphics designer ran out but there was still sufficient momentum in the project team to drive towards completing the final set of diagrams.

We agree that in the current manuscript there is a lack of detailed description of the process. This is largely because the overall aim of the paper is to present the diagrams. Describing the process of producing the diagrams is a secondary aim, which we included as we anticipated that readers would find it interesting to know about the stages we went through to produce the diagrams. However, the process of production was not intended, nor planned, to be a piece of social science research, so we did not involve social scientists to document the process. In hindsight, this was a missed opportunity. Rather, the focus of the activity was to produce a set of diagrams amongst a set of modeling experts, which we achieved. While there are no transcripts of the discussions, there are hand-written notes and email conversations, so the process *was* documented. We appreciate your comment, and also the comments of Referee #2, both which we believe will help to achieve our secondary aim of describing the process in a way that will be of interest to readers. This has motivated us to refer back to our notes and emails and describe the whole process in more detail, to provide recommendations to others who may like to embark on similar efforts.

Action: We describe the whole process in much more detail. This includes a generation of a visual time-line with milestones (e.g. conferences/workshops), meetings, interactions of the steering group with the graphics designer, stakeholder interactions, review rounds of diagram drafts and brief summaries of the results of each interaction round, but also of the manuscript. It also includes information, which software has been used for the different parts of the discussion. We also would like to provide at least some of the diagram drafts in the Supplementary material to document the development. Furthermore, we further elaborate about some of the difficulties that arise when so many modeling groups and persons/opinions arise. That also includes some metrics as numbers of groups per model involved, and in particular difficulties when responsible persons change positions over time or leave academia. We believe that this information and shared experience will be helpful for other communities that intend a similar activity.

RC1: 2. I have some concerns about some of the phrasing in this paper, and how that relates to the wider context of modeling capabilities as well as extensions of these diagrams. The paper is quite clear about the fact that the term for the complete diagram (i.e., the "'ideal' model") is not meant to be seen as a statement that describes how the ideal Earth system model looks. I believe that if this is so, then simply using a different term is more appropriate. I've suggested "ISIMIP-complete" in the comments but I think anything that avoids the implicit message that this is how an ideal model looks is better than what is currently used.

This becomes particularly important in the discussion section of the paper. Here the authors discuss these ISIMIP models and the resulting "ideal" model diagram as what is currently feasible within the scientific community. The word "ideal", in my opinion, implies much more than is justified here. The community as a whole has larger modeling capabilities than what is shown by this specific subset of models, and what can feasibly be done by the community extends beyond what this "ideal" model diagram shows. I think it is important to be honest

about limitations in our models (and these diagrams do a good job of giving high-level overviews of what specific models can and cannot do), but I think it is equally important to not undersell what is currently feasible if the community were to integrate all the separate bits of expertise in a coherent way. I don't think this is necessarily a discussion that needs to be had in this paper, but I do think it is important to acknowledge that the word "ideal" implies certain things, no matter how often the paper says that that is not the way the reader should interpret the word - particularly if these diagrams are partially meant to facilitate discussion with stakeholders who possibly don't have much personal modeling experience or clear overviews of the current-state-of-the-art of environmental modeling. Using a different term than "ideal" completely avoids all of this.

Finally, I think the phrase "ideal" model limits the ability to extend this diagram beyond what it currently includes. The authors list multiple aspects of environmental modeling that are not included by any of the models in ISIMIP2b and thus are not included in the "ideal" diagram. What happens when a model is included that introduces a new capability? Will this lead to the "ideal v2" or "slightly more ideal" diagram? A more version-y phrase would be more extensible in such future scenarios. This would also be more in line with the GMD requirement to include specific version numbers in the titles of a number of manuscript types (I'm aware that this is not required for review and perspective papers, but that doesn't make it a bad idea in general).

Answer: Thank you for pointing out this issue so clearly, and thank you for the suggestions for improvement. We fully agree. We had some very interesting and detailed discussions on what words to use to name the 'ideal' model when drafting our first submission - we acknowledge the issues of referring to it as an "ideal" model, which is why we were careful to caveat our use of the term. However, you make some very valid and well argued points about how this term can be (mis)interpreted, especially by a non-modeller, so we agree the most straightforward thing to do is to not refer to it as "ideal". This also avoids the "slightly more ideal" situation that will no doubt arise when new modeling capabilities in the community are realized.

Action: Throughout the manuscript, we avoid using the term "*ideal*" and focus on the relation to ISIMIP2b. Hence, we will use the term "*ISIMIP2b-complete*" instead of "*ideal*".

RC1: 3. I think the paper could also use a bit more text on some of the more practical concern about modeling capabilities that go beyond what's currently in the diagrams. How easy will it be to adapt the JSON tool and the diagrams themselves with new fluxes or other relevant information? Will there have to be a new design process to avoid cluttering what is currently there? Will new trade-offs between accuracy and aesthetics need to be made?

Answer: Thank you again for the very helpful comment. There is no doubt that as more and more processes are added, the ISIMIP2b-complete diagram will become more cluttered. This will not happen overnight but it is our hope that in the future the models include a more complete representation of the hydrological cycle - at this point, it would likely be necessary to re-draw the underlying diagram and make significant changes to the JSON file. For the foreseeable future, however, the JSON tool will fulfill its intended purpose because it is fairly trivial to add one or two new components and/or move elements around into areas where there is more space.

Action: We better document the JSON Tool (within the code and in the readme file) and improve readability. Furthermore, we add the intention of the tool generation into the process description. Finally, we reflect in the recommendations how such a tool can be reused for similar exercises.

RC1: 4. Something I have missed in the paper is a description about energy balance calculations. The model inputs suggest that at least some of these models try to explicitly account for energy balance components but neither the diagrams nor the text provide any information about this. Were energy-related state variables simplified away in favour of readability? If so, I think this needs to be discussed in the paper.

Answer: Indeed we have focused only on the water balance to avoid overcomplication. In global hydrology there are Land Surface Models, Global Hydrological Models, Dynamic Vegetation Models and hybrids. Whether the energy balance is solved or not is often determined by the type of model, but there are sometimes mixes. Anyhow, we focus on the hydrological process that are included/excluded in each model (often the result of the energy balance) and we decided to not open the discussion about types of models but call them „global water models“. We feel that this might help especially the audience of this paper. So we think the further discussion about energy balance is beyond the scope of the paper and therefore we will refer to Telteu et al. (2021) where this is included.

Action: We will motivate that we do not focus on the energy balance but hydrological balance and relate further discussion to Telteu et al. (2021) where the models are documented in-depth. However, we will include a table that lists which models are LSM, GHMs and hybrid.

RC1: Minor comments

Please see the attached .pdf. There is some overlap between those comments and the major ones I outlined above.

RC1: L 24: Is the grammar correct here?

Maybe splitting this sentence in two before "while" improves readability.

Answer: Thanks.

Action: The sentences will now read as: “*Elsewhere, the co-creation of diagrams between environmental modellers, design creatives, and policy-makers, has facilitated the generation of infographics and visuals that improve scientific understanding and better contextualise the degree of trust placed in modeling results. Examples of such efforts to balance academic integrity and detail alongside maintaining policy-relevance are the IPCC reports and the EU Green Deal (European Commission, 2023).*“

RC1: L 79: I think this section needs a brief description of the level of depth the model diagrams are meant to convey.

For any given category, I expect there will be a range of different implementations (different sub-processes included, different state variables, different equations) that will further determine how suitable any of the models is for a given purpose. However, when I look at the

diagrams themselves they seem to be mostly binary representations that indicate if a model includes a certain broad category or not. I think the text can be a bit clearer about this.

Answer: Thanks. Indeed we focus here more on the information if a component is included or not.

Action: After this sentence, we add the following sentences: „*The diagrams are aimed at a broad audience so they are kept as simple as possible to assist with communication. To this end, the diagrams focus primarily on describing whether a process is included or not, or stating the number of layers in a specific storage for example, rather than elaborating on the specifics how that process is represented internally within the model. This means that the diagrams do not show the different levels of complexity that exist between models when representing any given process. For more detailed information about how each model represents each process, the reader is referred to Telteu et al. (2021).*“

RC1: Sect. 2.1 This section seems (mostly) a repeat of the introduction and the reason for writing this paper, but does not really introduce any methods. Should this perhaps be moved to the introduction?

Answer: We agree, thanks.

Action: As indicated in our answer to your major comment #1, we will revise the methods section and describe the process in more detail which should help to reduce the repetition of content.

RC1: L 115: I'm not very familiar with how such social processes are typically documented and described in journal articles, but this seems a little on the short side to me. These discussions lie at the heart of the resulting diagrams and I think more description of how they were organized, which stakeholders and backgrounds were present, how different points of views and needs were balanced etc. is needed. Are there transcripts of these conversations?

Answer/Action: Thanks, we agree that a more detailed level of description would be beneficial. Please see our response to your major comment #1.

RC1: L 130: Showing a summary of these feedback rounds and the different iterations of the diagrams would be good. Knowing what was considered as options, what worked, and why would be very helpful to others trying to communicate the functioning of their model(s) to others.

Answer: Thanks, this is a good idea and led us, together with your major comment #1, to the development of a timeline with additional information.

Action: We document a summary of the rounds of interactions in the time-line proposed in the answer to major comment #1.

RC1: Sect. 3: The meteorological inputs listed in Figures 1 and 2 suggest that at least some models are more on the physics-based side of things, and do energy-balance calculations. However, only the mass balance (water storage) is mentioned in the diagrams, the results and the discussion section.

Given that the energy balance is critical for e.g. permafrost (stored carbon release), snow and ice formation and melt (disappearing glaciers, changes in seasonal water balance patterns), and vegetation (e.g. migration, growing seasons), I think the paper needs to at least discuss (1) the way the energy balance is calculated in these models, and (2) why none of that made it into the diagrams.

Answer: Indeed, some of the models (mainly Land Surface Models) use an energy-balance approach. As indicated in our answer to your major comment #4, we intend to keep the focus on the hydrological balance in order not to overcomplicate the diagrams and the paper content.

Action: Please see our response to your major comment #4.

RC1: L 136: 1. I think this an important statement and I'm glad to see to repeated throughout the paper. For extra clarity, might it make sense to switch from calling it the "ideal model" to "ISIMIP-complete model"?

2. I think the grammar in this sentence doesn't quite work out. "but not necessarily that this.." doesn't seem to connect to "The term 'ideal' stands for [..]".

Answer: 1: Thanks. Please see our reply to your major comment (2). Thanks.

Action: The sentence now reads as: „*The term "ISIMIP2b-complete" refers to a hypothetical global water model that includes all the fluxes and storages represented in at least one model participating in ISIMIP2b. However, it does not imply that this is the optimal/ideal way of representing the water cycle in a model.*"

RC1: L 140 In Figure 1 I see snow melt, glacier runoff and total runoff in part A of the diagram. In Figure 2 I see that total runoff includes snow melt (I think? The arrow points sideways), glacier runoff, surface runoff, interflow and groundwater runoff.

I'd be interested in the reasoning behind having a separate arrows for snow melt and glacier runoff but lumping all other surface and subsurface fluxes together into "total". The distinction between surface and subsurface flow seems larger to me than the difference between snow/glacier runoff and all other runoff sources.

Related, do I understand correctly from Figure 2 that all these models put snow melt into surface runoff?

Answer: Thanks for the detailed investigation of both figures. The reason for having separate arrows for snowmelt and glacier runoff is because they are outflows of their corresponding storages and therefore process-relevant. For surface runoff (defined in the ISIMIP2b protocol as „Water that leaves the surface layer (topsoil layer) e.g. as overland flow /fast runoff") we intended to draw this process close to the soil surface but not with the intention that it should be interpreted that all snow melt and glacier runoff are ending in surface runoff. The current solution has more a graphical / space reason and we will revise this. Total runoff, on the other hand, is described in the ISIMIP2b protocol as „Total (surface + subsurface) runoff". The inclusion in the diagram has no process-based reason and we can fully understand that this might be difficult to understand especially for readers who are not too familiar with the definitions in the protocol or the outputs.

Action: 1. As the flux „total runoff“ does not have a process-relevant reason and is not related to a specific storage compartment, we delete this flux in the diagrams. 2. To avoid the impression that snow melt and glacier runoff ends in surface runoff, we will staple these fluxes (glacier runoff and snow melt on top of surface runoff).

RC1: L 155: modification of „*less implemented*“ to „*less often implemented*“ as suggested.

Action: Change the sentence to: “Specific processes like capillary rise (included in 5 out of 16 models) and interflow (5 out of 16) are less often included, whereas fluxes like groundwater recharge are represented more frequently (15 out of 16).”

RC1: L 157: I can't fully follow this. Is the reasoning that groundwater tends to move slower than runoff and therefore it's physical incorrect to report groundwater recharge without a time delay compared to quicker fluxes?

Some questions:

1. If LPJmL intended to be coupled to a dedicated groundwater model that gets this recharge flux as input and computes possible lateral movement from that?

2. If we assume a (relatively) steady-state, any recharge into the groundwater would lead to (some) resulting lateral flow, so this lack of a time delay may not be that inaccurate.

All in all, I think a lot of technical detail is missing that would be needed before readers can correctly understand what this sentence is trying to say, and that level of detail seems out of place within the larger paper. Perhaps it's cleaner to simply remove this discussion of seepage, runoff and time steps in a single model.

Answer: With this sentence we wanted to provide the reason for having reported variables for groundwater recharge also for some models that lack an explicit groundwater storage but in this short statement it can be a source for ambiguity in interpretation as you have nicely highlighted with your points. We agree that much more (and model specific) technical detail is required in this work which is not intended.

Action: We remove the sentence in line 156 beginning with „*For example,…*“.

RC1: L162: I don't understand this. Is this meant to say that WaterGAP2 is not like PCR-GLBOWB? In that case, consider changing "and other than" to "in contrast to".

Answer: Thanks for the suggestion.

Action: We change “*and other than*“ to “*in contrast to*“.

RC1: L 164: suggested to change “Less“ to “Fewer“; thanks, we will follow

RC1: L 166: suggested to change “without considering“ to “that do not consider“; thanks, we will follow

RC1: Figure 2: This diagram seems to suggest snow exists to the side of the canopy, but not necessarily on top of it. Is this correct and intentional?

Answer: Thanks. Some models indeed simulate snow at the top of the vegetation, others simulate snow without an interaction of the canopy storage. We intended to have a generic way of figure composition that is at the same time not too detailed. Therefore we drew the snow at the side of the canopy storage.

Action: none

RC1: L 195: Check brackets.

Answer/Action: thanks, this will be corrected in the revised manuscript.

RC1: L 195: Also, how were these updated? Is there a corrigendum to Telteu et al. (2021)?

Answer: There is no corrigendum to Telteu et al. (2021) as there are no errors there. During the process of tool generation, we sent around to the modelling teams a shared online spreadsheet with all the fluxes, storages and water use sectors in columns and the models in rows. We included three columns for each item; one with the status (colour or greyed out) to indicate whether the process was included in the diagram, one with the information that was given in Telteu et al. (2021) and one where the modelling teams should finally confirm if the item is included or not. For some reasons (e.g. as people gained a different opinion on model features over the years; model versions changed slightly between Telteu et al. (2021) or simply due to communication errors between the chain of 16 modelling teams, the coordinator, the graphics designer and back) differences occurred for a few models and a few items.

Action: We touch these discrepancies in the updated description of the progress (see our response to your major comment #1). Furthermore, as we do not intend to confuse readers on the validity of Telteu et al. (2021) we will remove this bit of text.

RC1: L 196: I'd be interested in a discussion about the trade-offs between accuracy and aesthetics in the 3D diagrams. This paragraph seems like the right palce to add that.

Answer/Action: Thanks, we will include this aspect (see our response to your major comment #1).

RC1: L 209: These diagrams will help to give a high-level overview of the differences between models, but they are not enough to accurately answer the "why" question. Doing so would require controlled experiments in e.g. a multi-model framework to isolate the impact of each individual decision that is different between these models.

"[..], we can begin to understand the differences among models at a glance" may be a better phrase to use here.

Answer/Action: Thanks, we fully agree and will modify according to your suggestion.

RC1: L 214: Further examples may be minimum flows for ecology or transportation, and multi-story canopies.

More generally, I think this paragraph supports changing the terminology "ideal" model to something a bit more descriptive of what this diagram actually represents: an amalgamation of the processes currently included in a specific selection of models.

Answer/Action: Thanks, we agree and will change accordingly.

RC1: L 219: I think I slightly disagree with this. This "ideal" diagram can be viewed as a representation of what is currently being done, but what is feasible is a wholly different question. For example, including multi-story canopies isn't currently done in these models, but there are no technical reasons why this isn't feasible. The same goes for permafrost or snow melt (partly) infiltrating into the soil, or any of the other things listed earlier in this paragraph. I think some more nuanced phrasing is required here.

Answer: Thanks, we agree that this could lead to misinterpretation. With the modification of the term „ideal“, the focus will be more on the ISIMIP2b perspective. Further, we will avoid the term „feasible“.

Action: We will rephrase this part to highlight that model development is in progress and that the applied model features are a result of both the current capabilities of the models andalso the demands of the ISIMIP2b modelling protocol. Further, we will modify the term *„environmental*“ in L 220 with *„hydrological*“.

RC1: L 246: I'll repeat my previous comment here because I think it's important. The phrasing here implies that what is being done within this subset of models is representative of what can feasibly be done by the "scientific community" [as a whole]. This is too broad of a statement.

The scientific community's ability to model various sub-systems of what's shown in Figure 1 is (far) beyond what's being done in these models. We have the technical understanding to integrate these capabilities into community frameworks, and I therefore think that what is feasible at this point in time goes beyond the sum of what this specific subset of models can do.

I think it's good to be honest about limitations in our current modeling capabilities, but I think it is equally good to not undersell them either. I would suggest to either include a discussion along these lines (i.e., what capabilities exist that are not included in the models used in this work), or to rephrase certain parts of this paper to be more specific about the fact that what we're looking at in this work is a specific subset of the wider modeling capabilities that exist within the community.

Answer: Thanks for the detailed explanation of your point. We agree that this is too much focused on the models in relation to ISIMIP, ignoring a lot of other modelling approaches. The educational purpose here was more to show the different ways the individual global water models consider / deviate from the (to rephrase) „ideal“ representation, in particular to highlight that there are already in this subset of models quite different ways / levels of complexity available that is probably hidden when an end user just downloads a specific model output variable.

Action: We will modify this part in the sense that we avoid to claim that we judge about the general modelling capability of the scientific community but focus on the subset of the ISIMIP models. With a few examples of e.g., studies that are done in higher spatial resolution for some of the models such as CWatM, or the inclusion of gradient-based groundwater modelling that is done for a few models but not meaningful (i.e. in terms of run-time) for

ISIMIP2b simulations we will provide a certain outlook / wider framing of the modelling capabilities that exists within the community.

RC1: L 252: brackets.

Answer/Action: thanks, will be corrected

RC1: L 258: The first part of this sentence does not logically connect to the second part. Is the "do not" in the first part correct, or is there a "not" missing in the second part?

Answer: Thanks, indeed the word „not“ was missing.

Action: We modify this sentence to „*The new models do not differ significantly in their representation of the water cycle compared to the 2b models. Therefore, the model diagrams presented here need only minor adaptations to accommodate the new models.*“

RC1: L 271: I have some doubts about how well this would work in cases where these other models use different assumptions than what is currently encoded in the "ideal" figure. How adaptable would the JSON-tool and the visualization be for e.g. a model that includes lateral movement of snow (by wind or avalanches), water use for recreation or ecology, river meandering and flood defences, etc.?

Answer: Thanks. Indeed the agreement of a definition of an „ideal“ (ISIMIP-complete) model is a result of discussion inside the community. We intended here to highlight that the idea of generating a common representation of a sector’s model family and the derivation of individual model representations could be used also in other ISIMIP sectors. The JSON-tool works well to support these aims and for the foreseeable future, but when new processes have to be added or it is applied to other sectors, the JSON tool will need to be modified, along with the underlying graphic. We do not claim that the graphical representation (e.g. the background image and the considered components of the water cycle) are so general that it can be directly used for any model/any sector. For example, our framework will not work with black-box models. We followed specific ISIMIP nomenclature and foci of ISIMIP global water sector, e.g. the representation of specific direct human impacts, e.g. water use sectors as defined in the modelling protocol, which means some processes (like river meandering, avalanches, and protection afforded by levees) are missing and we acknowledge that the underlying image behind the JSON-file would need updating to include such processes. The JSON tool provides a very useful starting point, however, and it would mean that modeling teams of other sectors would not have to start from scratch if they wanted to create a visual representation of their models. The JSON-tool acts as a blueprint for the global water cycle and might be taken up from other sectors. We think that if other ISIMIP sectors have a similar action of common graphical representations of their models, a cross-sectoral understanding of the models could be improved, which might also lead in a fostering of cross-sectoral assessments of model outputs. However, these thoughts are not formulated carefully enough.

Action: We will rephrase this section to highlight the thoughts mentioned in the answer to this comment.