Civil Engineering
University of Victoria
PO Box 3055, EOW 341
3800 Finnerty Road
Victoria BC V8P 5C2 Canada

November 17, 2020

Dear Editor Alberto Guadagnini,

We appreciate the suggestions and opinions from four different reviewers summarized below. Based on each of these we have reflected at depth both on the manuscript and the broader effort towards analysing and modeling groundwater at large-scales. First we provide an overview of our response to the referees and then below we respond to each of the individual and specific referee comments in blue text with **specific changes to the manuscript highlighted in bold blue text**.

These reviews were useful in improving the manuscript in a few substantial ways:

1) **Better incorporating the insights and lessons learnt from regional groundwater models.** Even though this was our intention as described below in our response to RC1, comments especially by RC1 and RC3 highlighted that we did not do a good job of incorporating the insights and lessons learnt from regional groundwater models - indeed, most of the co-authors on this opinion piece have been involved in the development of regional groundwater models and are keenly aware of both their utility and challenges. In the revised version of the manuscript, we more clearly highlighted this foundational knowledge from the world of regional modeling by organizing this information into the new Section 2.1 ('Synergies between regional-scale and large-scales'), and how these lessons can be translated to the continental and global scales (Section 2.2, 'Differences between regional-scale and large scales') .

2) **Importance of geology and heterogeneity.** Two referees (RC1 and RC3) both mention that we did not adequately describe the crucial role of geology in determining groundwater dynamics. We agree with this so we have added some additional text in Section 1. Primarily we refer to a manuscript Condon et al. (in preparation; near submission) that focuses on the question of hydrogeological conceptualization at large scale and that includes a comprehensive discussion on how to incorporate geology and heterogeneity. In order to not be unnecessarily repetitive, we do not further expand on the importance of geology and heterogeneity in this manuscript under consideration for HESS.

While we respect the views of all colleagues in our field and deeply appreciate the effort each reviewer invested in order to review our submission, we seem to be at odds with a few important general sentiments that can be found across these reviews. These differences can be summarized by these two points:

1) **Value of continental to global-scale models.** RC 1-3 each in different ways question or doubt the overall purpose of groundwater representation in continental to global scale models (herein called 'large-scale'). Collectively, we see crucial scientific and sustainability challenges that are best addressed by large-scale models that are summarized by Section 1 which is entitled "Why and how is groundwater modeled at continental to global scales?' We spent significant effort in clarifying and refining these

arguments, and don't see any significant way of improving them, so we may be left at odds with these reviewers. Since this is an Opinion paper, we argue that this is an important difference in opinion that should be considered in the broader scientific dialogue.

For clarity we can restate the thrust of our argument for the large-scale groundwater modelling:

- To assess and inform on the global state of groundwater resources under climate change and human water use.
- To assess and compare the potential and efficacy of strategies and policies towards groundwater sustainability.
- To correctly incorporate the dynamic interactions between groundwater, surface water and land surface processes in earth system models which are critical to accurate simulations and ultimately predictions. Note: due to its inertia, the inclusion of groundwater is likely to improve seasonal atmospheric forecasts as well as streamflow predictions across large basins and continents.

The first two points have to do with global change, global sustainability and global sustainability governance. Many accept the utility of global economic analyses, global biodiversity reports and global comparative studies on social wellbeing, so it is unclear to us why groundwater sustainability is only a local to regional problem that cannot be considered as a global issue?

2) **Focus on model evaluation rather than model development.** RC1 generally suggests it might be more fruitful to focus on alternative approaches to developing continental to global models that we refer to as 'a patchwork approach' in Section 1. While we agree that this is an interesting and potentially useful idea, we see at least four arguments (summarized in Section 2) that make this a challenging approach. More generally, we argue that method development has been the focus of other review papers (Bierkens, 2015; Condon et al., in prep.) and Opinion papers (Fan et al., 2020). We feel that a significant missing niche that this paper fills is a forward-looking synthesis of approaches to model evaluation - this is a tighter and hopefully more impactful scope. Following up on a suggestion of RC1, we did add a few sentences to Section 1 about this nascent idea of 'patchwork modeling'.

Overall, we have tried in earnest to make changes to the manuscript in response to each of the Referees. We made a number of small to moderate changes to the manuscript and have responded to each of the Referee comments and suggestions below. We note that we were surprised by the tone of RC2 and that in our reply we focus on keeping this as a scientific debate - we are happy to diligently address serious comments with evidence.

In the end this is an Opinion piece so it is not surprising that some do not agree - RC3 says our "thoughts are controversial and I would say that many of us (hydrogeologists) would not share them." But we stand strongly behind this paper. We argue it is a good, worthwhile and timely contribution that will hopefully nudge various scientific communities to improve the evaluation of groundwater representation in continental to global scale models - even if our thoughts seem controversial today.
We hope that you will consider our manuscript for HESS.

Sincerely,

Tom Gleeson
Associate Professor and President's Chair,Civil Engineering, University of Victoria

RC1 - Keith Beven (Referee) received and published 1 September 2020: Is evaluation of a global model the right approach?

I understand the requirement to have improved groundwater modelling capability in earth system science but I am really not sure that this paper is advocating a correct strategy for achieving that. It starts from the viewpoint of a community of global ground-water modellers that there are global groundwater models available that need to be evaluated (with a view to improvement). As such it completely ignores the experience of what are here called the regional hydrogeologists in implementing operational groundwater models (with all their difficulties of conceptual models of the geology, spatial heterogeneities in transmissivities, fracturing, disconnections between layers and local confinement, patterns of (sometimes unlicensed) abstractions, etc etc). There is, for example, no mention of the Danish National Water Resources Model that has tried to do this at a national scale (and even then run into scale, conceptualisation, and parameterisation problems). So, in that groundwater is framed by local geology, which can vary at below the global groundwater model grid scale, it would seem to be much more productive if the approach to the global problem was to provide a portal to make use of that regional information much more directly than the portal for model evaluation suggested here.

We deeply appreciate Prof. Beven's thoughtful commentary and suggestions - these provided many opportunities to reflect again on past, present and future large-scale modeling that includes groundwater parameterizations. Prof. Beven is making the point that no global model will have the same level of detail that local models have. That is true, however, the solution is not that we therefore need to reproduce all the local information in the global model. We do not need to reproduce the world at hyper-resolution to address some of the questions we want to answer at the large scale. The question is rather what aspects of the local system have to be reproduced and what scale/resolution is required to do so.

We agree that the experience of regional hydrogeologists in implementing operational groundwater models is very important and was under-represented in the manuscript. A number of the coauthors have directly developed (e.g. Fogg, Cuthbert, Bierkens, Zipper, Gleeson, Bresciani) or evaluated regional groundwater models (Hill), or are close project collaborators of people who do. So even though we did not emphasize this experience proportionally, we bring this important perspective to the problem of evaluating large-scale models as well.

In Section 2 (Current Model Evaluation Practices) we had a final paragraph highlighting the evaluation of regional-scale groundwater models, as well as a list of five ways that evaluation of describing how regional-scale and large-scale model evaluation practices are different. In order to provide greater emphasis on regional models **we moved this paragraph on regional scale groundwater model evaluation forward to the beginning of Section 2 and expanded on it as the new section '2.1 Learning from regional-scale modeling'**.We also clarified the important differences between large-scale and regional scale models in the new Section 2.2 Adding subsections to Section 2 also helps with the overall flow of this section.

We agree that the Danish National Water Resources Model is a good example of such an effort as are numerous regional models by the USGS starting with the Regional Aquifer System Assessment (RASA) program in the 1970's. We feel it is beyond scope to mention or describe individual models or modeling programs, but rather more important to point towards efforts to synthesize regional models and learnings about these models. With this in mind we **now referenced the Rossman & Zlotnik (2014) review in Section 2, which is a useful synthesis of regional models across the western United States.** To our best knowledge there is no other similar synthesis of regional models for other regions of the world. As a side note, we have referenced Troldborg et al. (2007) and Refsgaard et al. (2007) which are both related to but not directly about the Danish National Water Resources Model.

Finally, we note that we fully agree that a portal to make use of that regional information would be highly useful. This is exactly what we meant by including "regional-scale models that meet the standards described above and could facilitate inter-scale comparison (Section 3.2)" in our list of desired ingredients in our envisioned 'Groundwater Modeling Data Portal".

(it is interesting that "geology" as such does not appear in the text – only in authors affiliations – we have to infer it from "conceptualisations").

This is a useful and interesting observation and reflection, and Prof. Beven is correct that the crucial role of 'geology' on groundwater systems is somewhat obfuscated under 'conceptualizations' in the Opinion manuscript. As noted by the author affiliations, many of the coauthors identify as, and have foundational training in geology or earth science. In addition, we have been involved in various efforts such as developing the global permeability maps based on a compilation of geology maps and data for large-scale groundwater analysis and modeling (Gleeson et al. 2011; Fan et al. 2015; Fan et al. 2020).

An earlier version of this manuscript entitled "Groundwater representation in continental to global hydrologic models: a call for open and holistic evaluation, conceptualization and classification" (also available on as Version 1 of this manuscript on EarthArXiv) had an entire section entitled "SYSTEM CONCEPTUALIZATION" that we are removed from the HESS Opinion submission to focus the manuscript on model evaluation; we plan a follow-up manuscript on the role of different geological environments and different approaches to conceptualizing the key geological elements of these geological environments. In sum, we fully agree that geology is crucial to groundwater analysis and modeling at all scales, we feel it is out of scope of this manuscript to focus specifically on geologic differences in groundwater systems in different regions or more generally.

To acknowledge the critical role of geology, we added a sentence about this as a possible future direction in large-scale models that represent groundwater in Section 1. We primarily refer to Condon et al. (in prep) which contains much more detail, as described above.

There will, of course, be gaps in the global coverage where there are important groundwater bodies but where no regional or local models are available and data is poor. Certainly in those situations we would need to resort to expert elicitation in creating a suitable model to make the coverage more complete. But that is a different problem. Because bringing in regional hydrogeologist expertise to evaluate the global modelling (as suggested) would seem to be doomed to failure. The global grid scale and variation in parameters is too crude.

We agree that expert elicitation will be challenging because of many differences such as the technical ones mentioned here (grid scale and parameter variation) but also perceived or real differences in perspectives, priorities and backgrounds between regional-scale and large-scale modelers. But we think that this challenge could be potentially worth it but for the pure scientific advances but also the potential new synergies between these two communities. As described above, many of us move between and have strong connections in both communities, so we can be the bridges to help make these connections. Scale and commensurability remain crucial scientific challenges so these efforts can hopefully contribute to these challenges.

This type of direct comparison using expert elicitation as Prof. Beven is suggesting here would be useless. Such an expert elicitation exercise has to be done in the context of the purpose of the global model. What is the model for and what local properties have to be preserved to achieve this. Clearly we would not propose to elicit a hillslope hydrologist regarding her specific knowledge of a particular hillslope to evaluate a catchment scale model (the only answer could be that the model has to be rejected). However, we might reasonably ask her to suggest what variability we should expect across all hillslopes in a region (or something like that).

Therefore in Section 3.2 we added this text:
"Expert elicitation also has a number of challenges including: 1) formalizing this knowledge in such a way that

it is still usable by third parties that did not attend the expert workshop itself; and 2) perceived or real differences in perspectives, priorities and backgrounds between regional-scale and large-scale modelers.

The generation of recharge rates and evapotranspiration rates when the water table is near the surface just cannot be properly represented when the grid scale cannot reflect the local variations in topography, but there is nothing here about evaluating such fluxes (and getting such boundary conditions right is surely rather important.... Or is that being left to the global land surface modellers rather than the global groundwater modellers?).

The scale dependency of recharge rates and evapotranspiration is important and many large-scale models are too coarse to correctly resolve groundwater-surface water interaction and the interaction with evaporation unless subgrid-processes are incorporated.

Some advancements strongly depend on increasing the resolution of global models as extensively discussed Fan et al. (2019, WRR). We are not disputing this at all. However, we still need to find a hierarchical strategy to evaluate whether global models are good enough for their purpose, or whether they lack key behaviours that need to be preserved. These could be included either through increased resolutions or through statistical sub-grid representations.

The land surface models that calculate diffuse and concentrated groundwater recharge have sub-grid parameterizations to account for un-resolved fractions of saturated soils (TOPMODEL, ARNO-scheme) and the two-way coupled version of our global groundwater model accounts for groundwater discharge through small streams, seeps and springs by a subgrid parameterization. Furthermore, resolutions are increasing (look at PARFLOW across the U.S and Europe; we have recently been running 1 km transient 2 layer MODFLOW 6 for the globe for 50 years within 1 day), so these issues will become less important in the near future.

It is also important to note the formerly clear boundaries between land surface modellers and groundwater modellers are dissolving so this classification is less useful now. Each model has different strengths, capacities and approaches to solving for water fluxes and stores below, at and above the land surface.

Finally we note the recharge rates and evapotranspiration are mentioned in Section 2.3 and included in Table 1. Since these are standard fluxes for evaluation of many large-scale models, it does not seem necessary to describe them in detail in Section 2.3.

The paper recognises the issues of commensurability, but has no suggestion for how to take that into account (except for the use of “signatures” but it is then not explicit about how that might actually work).

This was a useful critique - we did spend much more time recognizing the issues of commensurability than directly taking it into account. We note that in Section 3.2 we describe how inter-scale model comparison could use variables routinely simulated in regional-scale models such as baseflow or recharge to evaluate large-scale models. “In this way, the output fluxes and intermediate spatial scale of regional models provide a bridge across the “river of incommensurability” between highly location-specific data such as well observations and the coarse resolution of large-scale models.”

In our minds, a number of the future priorities outlined in Section 3.1 can partially address or at least improve upon some of our current commensurability challenges (Section 2) - but this did not make it from our minds to the manuscript! For example, using data that is more consistent with the scale modelled grid resolution will reduce the commensurability challenges. **We therefore added this text to Section 3.1:**
“Using data (such as baseflow, land subsidence, or the spatial distribution of perennial, intermittent, and

ephemeral streams) that is more consistent with the scale modelled grid resolution will hopefully reduce the commensurability challenges."

And some of the suggestions for "evaluation" seem to me to be rather circular (see comments on manuscript).

In the attached we have also responded to each of the detailed manuscript suggestions as comments in the pdf.

And then there is karst. This is rather important in some parts of the world. It does get just a passing mention in the text (through the paper of Hartmann et al., 2017) but there is no discussion of how this might fit into a global model based on PDE continuum approaches.

We agree that karst is crucial to consider in groundwater modeling and one of the co-authors (Andreas Hartmann) is an emerging leader on new approaches to karst in large-scale models. Here too, there is significant overlap with Condon et al. (in prep., so we feel that it is not appropriate to add significant text about model development to this manuscript that is focused on model evaluation. Therefore, **we added a reference to karst in Section 1** and in this reply we give a more fulsome discussion of the role of karst in large-scale modeling:

1. When regional balances are considered, equivalent porous medium approaches that do not explicitly consider karst have shown to perform not too bad (Scanlon at el, 2003 JoH), which may justify disregarding karst heterogeneities/discontinuities to some extent.

2. A new paper using WoKaS database (Hartmann et al., 2020) shows that the delay of the recharge signal through karstic groundwater dynamics has to be considered at the seasonal scale while Hartmann et al. (2017) showed that recharge is a very sensitive variable that large scale models need to account for.

3. For recharge and for groundwater, karstic discontinuities are important to groundwater quality (not the focus of this manuscript). Global water quality data sets, when finally compiled, will help to evaluate how much the lack of fast flow representation affects large-scale model simulations (Hartmann, 2016).

4. Regional applications of karst specialized groundwater models (MODFLOW-CFP, Reimann and Hill, 2009, among others) that provide combined PDE applications for laminar Darcy flow in the karst matrix and turbulent flow in a discrete conduit network may provide some directions for future improvements.

The paper is the outcome of a workshop on global groundwater modelling but for all the expertise available it seems to me to be wrong about how to approach the problem (perhaps because the expertise of operational groundwater modellers was not that well represented). Evaluation of the type of global groundwater models being suggested is not really the issue. For all sorts of reasons we can expect that they will be too crude and too approximate and will not make best use of local information where that is available (even to the point of local rejection). Global groundwater is an aggregation of regional and local groundwater systems with all their different geological and other characteristics. If the problem of using that regional and local information directly is only computational, then ways could be found of simulating the responses more efficiently (not necessarily using a coarser grid, but, for example, perhaps using machine learning). Where a regional or local model cannot be used because of bureaucratic reasons, then it will be necessary to construct a simulator in the same way as for a data-sparse area, but again without data there can be no evaluation (the expert elicitation will already have been used in the construction). I would suggest, therefore, that this paper needs more, and deeper thought, and should not necessarily have a starting point of here is global groundwater model how do we evaluate it, but rather here are all the important aquifers worth representing, how should their response be best represented (which might of course be locally/regionally a PDE continuum model – or not)?

We agree with Prof. Beven's thoughtful suggestion of focusing on regional differences. In Section 3 we had this text: "Considering regional differences in model evaluation suggests that global models could in the future consider a patchwork approach of different conceptual models, governing equations, boundary conditions etc. in different regions. Although beyond the scope of this manuscript, we consider this an important future research avenue."

We expand on this idea of a patchwork approach by **adding this text to Section 2:**
"Given the strengths of regional models, a potential alternative to development of large-scale groundwater models would be combining or aggregating multiple regional models in a patchwork approach (as in Zell and Sanford, 2020) to provide global coverage. This would have the advantage of better respecting regional differences but potentially create additional challenges because the regional models would have different conceptual models, governing equations, boundary conditions etc. in different regions. Some challenges of this patchwork approach include 1) the required collaboration of a large number of experts from all over the world over a long period of time; 2) regional groundwater flow models alone are not sufficient, they need to be integrated into a hydrological model so that groundwater-soil water and the surface water-groundwater interactions can be simulated; 3) the extent of regional aquifers does not necessarily coincide with the extent of river basins; and 4) the bias of regional groundwater models towards important aquifers which as described above, underlie only a portion of the world's land mass or population and may bias estimates of fluxes such as surface water-groundwater exchange or evapotranspiration. Given these challenges, we argue that a patchwork approach of integrating multiple regional models is a compelling idea but likely insufficient to achieve the purposes of large-scale groundwater modeling described in Section 1. Although this nascent idea of aggregating regional models is beyond the scope of this manuscript, we consider this an important future research avenue, and encourage further exploration and improvement of regional-scale model integration from the groundwater modeling community.."

We hope this text distills the possibilities, but also the challenge of this approach.

There are many more comments on the manuscript. Please also note the supplement to this comment:https://hess.copernicus.org/preprints/hess-2020-378/hess-2020-378-RC1-C3

In the attached we have also responded to each of the detailed manuscript suggestions as comments in the pdf.

RC3 - Anonymous Referee #3 received and published 8 September 2020: 'a well-written text and that's it'

I am not completely sure about the work. First, it is an "opinion paper", and this is exactly what it is, a recollection of the thoughts of very relevant well-known researchers that gathered together and, after a discussion, put together some thoughts. The problem is that these thoughts are controversial and I would say that many of us (hydrogeologists) would not share them. So, the point I want to make is asking what is the significance of the work.

Thanks for asking this question of the significance. We argue that the significance is providing clear recommendations for the quest to improve the evaluation of large-scale models which are currently the only "coherent scientific framework to examine the dynamic interactions between the Earth System above and below the land surface, and are compelling tools for conveying the opportunities and limits of groundwater resources to people so that they can better manage the regions they live in, and better understand the world around them." (Section 1)

The problem of including groundwater in global models lies on the problem in scales, mostly the temporal ones. This is probably why the authors talk about geological eras, but do not talk about geology. Temporal dynamics are a benefit or a curse, as they run on different timescales that are strongly related to spatial scales of models.

As mentioned above, we were remiss in not focusing more on geology - **we have added this to Section 1**. We agree that temporal dynamics are related to the spatial scales of models.

But also, the amount of information that should be included in the models is size and purpose dependent. And this is also well known. Even the processes that should be considered depend on the size and purpose of the model. Processes that are inherently non-linear at the local scale can be linearized at the regional scale and maybe take as constant (or disregarded) in a global model; a clear example is GRACE, where groundwater is treated in a way that would be considered shameful for all groundwater modelers, but that is capable of producing some quite good results at the scale of a full country (obviously difficult to be used in predictive models). This can hardly be considered the core of a new contribution. And then, for each problem we can use the simplest model (or approach) we can think of, but this depends exclusively on what do we want the model for, as correctly pointed out by the authors. Expert opinion models are fantastic when you have zero data and thus just to get an idea of how to manage resources at some global scale (what the authors call a sustainability-focused purpose), or how to design a global network. But this cannot help you at the local scale (designing a new well). It is the same with water quality, that can be inferred from general geogenic conditions, but again cannot be used to assess the water quality at a point. But my question here, what is new about this that deserves being published?

We generally concur with the above statements of the referee. What is new here are at least three elements:

- is that never before have the limitations of current evaluation practices for large-scale models (Section 2.3) been distilled;
- nor have a compelling vision for how to holistically improve the evaluation of groundwater represented in large-scale models been proposed (Section 3); and
- finally, our two next steps (Section 4) are forward looking and will hopefully be a useful nudge to multiple scientific communities.

Together, these reasons suggest this is a good, worthwhile and timely contribution that deserves being published.

The authors do not talk about water resources quality at all, and this is a key point in global management.

Yes, this is true.

In Section 1 we have added:
"It is important to note that we do not consider water quality or contamination since large-scale water quality models are in their infancy (van Vliet et al., 2019) even though water quality or contamination is important for water resources, management and sustainability."

I like the sentence "all three strategies (observation-based, model-based, expert-based) should be pursued simultaneously because the strengths of one strategy might further improve others". I fully disagree; more, I do not see any other way of modeling that is not based on: You check all available data, you postulate some potential models using expert opinion and all existing data, and you build the model, including or not uncertainty and model selection criteria, but always based on calibration and model selection criteria. So, I do not sympathize with this idea of having 3 connected but separated strategies. In my opinion they are three faces of the same strategy: do the best you can with what you have and with your client needs.

We understand that RC3 disagrees with our language of '3 connected but separate strategies' and would rather consider them 'three faces of the same strategy'. We acknowledge that our text in Section 3 and 4 (and Figure 1) gave the impression of three separate strategies even though we did write that the three strategies are "potentially mutually beneficial" and later that they "should be pursued simultaneously because the strengths of one strategy might further improve others. For example, expert- or model-based evaluation may highlight and motivate the need for new observations in certain regions or at new resolutions. Or observation-based model evaluation could highlight and motivate further model development or lead to refined or additional hypotheses."

After reading this critique we realised that we needed to clarify that we are not proposing a brand new evaluation method here but separating strategies to consider the problem of large-scale model evaluation from different but highly interconnected perspectives (or 'faces' as RC3 mentions above). All three strategies work together for the common goal of 'improved model large-scale model evaluation' which is what is at the top of the pillars in Figure 1.

Therefore we added this text to the beginning of Section 3:
"We are not proposing a brand new evaluation method here but rather separating strategies to consider the problem of large-scale model evaluation from different but highly interconnected perspectives. All three strategies work together for the common goal of 'improved model large-scale model evaluation' which is what is at the top of the pillars in Figure 1."

Anyway, to summarize, the paper is perfectly fine. It is the contribution of excellent researchers that have their personal views based mainly on their own work. Maybe the contribution will be useful for young researchers, despite some comments/ideas being questionable. I do not recommend publication, because I do not see the point, but if the Editor considers that it should be published, it will be "in present form".

Thank you for summarizing that the paper is perfectly fine. We have done our best effort to improve or clarify that questionable comments or ideas. We hope our arguments above the referee would see the point of publishing the manuscript, especially with all the improvements from it's previous 'present form'.

SC1 - John Ding received and published 26 August 2020: 'Groundwater storage as a quadratic reservoir and the value of new streamflow observations'

In addition to the assumption of a linear reservoir, Q=kS(Lines 144 & 155; Table 1,Row Baseflow ... recession (k)), the groundwater storage need be considered as a quadratic one.

In the context of model evaluation (Figure 1 and Section 3.1), the importance of acquiring new streamflow data at a project site cannot be overemphasized (Table 1,Row Streamflow).Considering a catchment on a continental scale as either a linear, Q=kS=C1S, or a quadratic storage,Q= (C2S)2, the baseflow can be linearized as follows (Beck et al.,2013, cited by authors; Azmi et al., 2020, SC1 therein; Ding, 2020):logQ(t) = logQ(t0)−C1(t−t0),(1)Or:−1/√Q(t) =−1/√Q(t0)−C2(t−t0),(2)Equations (1) and (2) are based on the logarithmic and NISR (negative inverse square root) transformation of the streamflow Q, respectively. On an ungauged catchment, all these two equations will need to determine their discharge coefficients,C1andC2, are a minimum of three new flow measurements in the field over a period of days or weeks. These will be used to falsify the hypothesis of a linear or a quadratic storage. Direct measurements of the low to mean flow on river on a regional scale are doable, though a logistic and technical challenge (Lines424-427). Legend or hearsay has it, ancient Egyptians measured the Nile River flow by diverting it to a side chamber where the volume of water was measured. Figure 1 shows the data transformation diagram for the log and the negative inverse mthroot (NImR) transform (Santos et al., 2018, SC5 therein; Ding, 2018). Compared to the log transform, differences are small among the first, second and third root of the NImR transform. The NI2R or NISR transform, which is derived for the outflow from a hillslope, maybe considered a representative of these fractional power ones. For application, a catchment can be classified either as linear or quadratic, unless dictated theoretically otherwise. Between the two, the log transformation is a low- to mid-pass filter, and the NISR a low-pass one.

This follow-up will illustrate a comparison of the universal logarithmic and the newNISR transformation as represented, in SC1, by Equations (1) and (2).

Figure 2 summarizes result of a new recession flow analysis for year 1962 for BigSpring near Van Buren, in Missouri, a regional-scale limestone karst having a drainage of 1500 km2.The graph shows a long recession hydrograph from May to August of 1962 (Florea and Vacher, 2006, Figure 2C). The recession literally started one day after the last ofthe major peaks and ended at the start of the next noticeably uptick. It also includes the log and NISR transformation of the observed flow.In their absolute value, from untransformed, to log-, and finally NISR-transformed recession hydrograph, the correlation coefficient improves incrementally; and the recession slope flattens successively by one order of magnitude.

References:

Azmi, E., Ehret, U., Weijs, S. V., Ruddell, B. L., and Perdigão, R. A. P.: Technicalnote: ”Bit by bit”: A practical and general approach for evaluating model compu-tational complexity vs. model performance, Hydrol. Earth Syst. Sci. Discuss.,https://doi.org/10.5194/hess-2020-128, in review, 2020.

Ding, J. Interactive comment on "Technical note: Pitfalls in using log-transformed flowswithin the KGE criterion" by L´eonard Santos et al., Hydrol. Earth Syst. Sci. Discuss.,https://doi.org/10.5194/hess-2018-298-SC5, 2018.

Ding, J. Interactive comment on ”Technical note: ”Bit by bit”: A practical and general approach for evaluating model computational complexity vs. model performance” by Elnaz Azmi et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-128-SC1, 2020.

Florea L.J. and Vacher H.L.: Springflow hydrographs: eogenetic vs. telogenetic karst.Groundwater. 44(3):352-61, 2006

Santos, L., Thirel, G., and Perrin, C.:Technical note:Pitfalls in using log-transformed flows within the KGE criterion, Hydrol. Earth Syst. Sci., 22, 4583-4591,https://doi.org/10.5194/hess-22-4583-2018, 2018.

Thank you for this remark. As stated in Lines 144 and 155, we encourage using groundwater representations that are able to simulate hydraulic heads instead of storages, regardless whether linear, quadratic, or any other non-linear approximation, in order to consider capillary rise of groundwater and groundwater-surface water interactions. But we agree that in respect to model evaluation, non-linear streamflow recession characteristics have a huge potential to evaluate large-scale groundwater models, especially in cases like the karst spring example elaborated in Figure 2 of this comment. We also agree that stream flow and spring flow observations, which have become largely available, allow the evaluation of large-scale models for their performance in simulating linear and non-linear recession characteristics. **We have updated section 3.1 and Table 1 accordingly.**

Figure 1. Streamflow data transformation diagram for a nonlinear storage-discharge function, $Q = (C_n S)^n$.
If $n = 1$, $J_1(Q) = \log Q$; if $n > 1$, $J_n(Q) = -1/Q^{1-1/n}$.
Adapted from Ding, 2018, SC5.

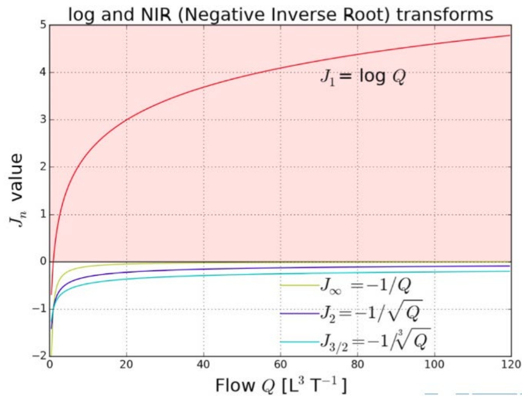
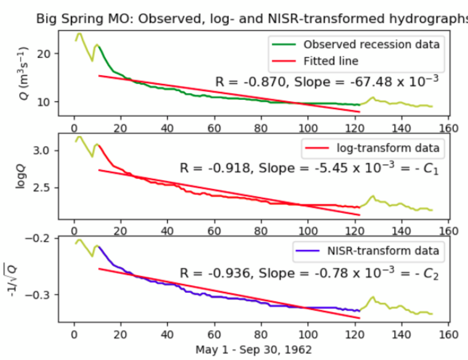


Figure 2. Observed, and log- and NISR-transformed recession hydrographs for Big Spring near Van Buren. Karst aquifer is of limestone, 475 Ma in age, and 1500 km^2 in area (USGS Gage 07067500; Florea and Vacher, 2006).



References

Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, *51*(7), 4923–4947. https://doi.org/10.1002/2015WR017173

Condon, L.E., Stefan Kollet, Marc F.P. Bierkens, Reed M. Maxwell, Mary C. Hill, Anne Verhoef, Anne F. Van Loon, Graham E. Fogg, Mauro Sulis , Harrie-Jan Hendricks Fransen ; Corinna Abesser. Global groundwater modeling and monitoring?: Opportunities and challenges (in preparation)

Fan, Y., Richard, S., Bristol, S., Peters, S., Ingebritsen, S., Moosdorf, N., Packman, A., Gleeson, T., Zaslavsky, I., Peckham, S., Murdoch, L., Fienen, M., Cardiff, M., Tarboton, D., Jones, N., Hooper, R., Arrigo, J., Gochis, D., Olson, J., Wolock, D. (2015) DigitalCrust - A 4D data system of material properties for transforming research on crustal fluid flow. Geofluids. 15: 372-379 doi: 10.1111/gfl.12114

Fan, Y. et al., (2019) Hillslope hydrology in global change research and Earth System modeling. *Water Resources Research*, doi.org/10.1029/2018WR023903

Gleeson, T., Smith. L., Jansen, N., Hartmann, J., Dürr, H., Manning, A.H.,van Beek, R. and A.M. Jellinek (2011) Mapping permeability over the surface of the earth. Geophysical Research Letters, 38, L02401, doi:10.1029/2010GL045565.

Hartmann, A., Liu, Y., Olarinoye, T., Berthelin, R. and Marx, V.: Integrating field work and large-scale modeling to improve assessment of karst water resources, Hydrogeol. J., doi:10.1007/s10040-020-02258-z, 2020.

Hartmann, Andreas, Gleeson, T., Wada, Y., & Wagener, T. (2017). Enhanced groundwater recharge rates and altered recharge sensitivity to climate variability through subsurface heterogeneity. *Proceedings of the National Academy of Sciences*, *114*(11), 2842–2847. https://doi.org/10.1073/pnas.1614941114

Hartmann, A. (2016): Putting the cat in the box: why our models should consider subsurface heterogeneity at all scales, Wiley Interdiscip. Rev. Water, 3(4), 478–486, doi:10.1002/wat2.1146

Reimann, T. and Hill, M. E.: MODFLOW-CFP: A new conduit flow process for MODFLOW-2005, Ground Water, 47(3), 321–325, doi:10.1111/j.1745-6584.2009.00561.x, 2009.

Scanlon, B. R., Mace, R. E., Barrett, M. E. and Smith, B.: Can we simulate regional groundwater flow in a karst system using equivalent porous media models? Case study, Barton Springs Edwards aquifer, USA, J. Hydrol., 276(1–4), 137–158, doi:10.1016/S0022-1694(03)00064-7, 2003.