

In this letter we summarise how we have revised the article based on the feedback of the reviewers and community. Elaborate responses to each point raised by the reviewers are included in the following pages.

1. Introduction

- We have included more explicit references to already ongoing debates on water models, in response to reviewer #1 and #3, including the fit-for purpose discussion.
- We have added a discussion on why it is necessary to discuss the influence of water models in particular, especially in relation to their connection with water infrastructure and the role of water in society, as well as the few articles that exist on the topic.

2. Defining models

- We have elaborated on work done by Science-Technology and Society scholars specifically to understand the influence of technology and knowledge, including the work of Demeritt and Stengers in response to reviewer #2
- We have made clearer in the revised article that the political nature of modelling is always an interplay of actors (commissioners, scientists, modelers, politicians, other stakeholders) and the model itself, in response to reviewer #1 and #3.

3. Methodology

- In response to all reviewers, we have clarified why we find the combination of a narrative review and the ROSES method for review useful in this case, despite the large number of articles that have been excluded.
- In response to reviewer #3, we have more clearly explained why certain articles are included and others excluded.
- We have adjusted the query to include hydraulic and groundwater models too
- We have reviewing the articles suggested by the three reviewers and community, to identify how/if they can be included in the review. We are thankful for the additional input.
- We have clarified how the articles were analysed and how the four overarching themes and 13 topics were found.

4. Results

- We have updated the results based on the additional articles that will be included in the review. We have indicated which reviewed documents are PhD dissertations.

5. & 7. Discussion and conclusion

- We have streamlined the discussion to avoid repetition between the conclusion and discussion, based on feedback for reviewers #2 and #3.

6. Call for power-sensitive modelling

- A very important element of our revision is to add examples to clarify and exemplify the considerations that can guide power-sensitive modelling, in response to reviewers #2 and #3, especially referring to existing work.

Point-by-point response to the reviews including a list of all relevant changes made in the manuscript can be found on the next page. We thank the reviewers for their feedback and for helping to improve the manuscript.

Reviewer 1

Reviewer: *Because the authors seem to be falling into the trap of talking about models as if they were responsible for outcomes – and as I have written several times before, decision makers are only too happy to transfer responsibility to the models. But that is not how (in general) it works. With the exception of some academic studies (such as many of those cited here) models are conditioned for a purpose. The power relationship here is not at all in the model, the model is only a consequence of making assumptions, so the important factor is who decides on the assumptions. This will be constrained initially by how a project is framed (the political as used by the authors, which may well reflect the vested interests of whoever is doing the commissioning or decision making) and then by how the modeller interprets the brief (including how much time is costed when bids are competitive). Of course, in a sociohydrological model, then the relative importance of different stakeholder groups might be a factor that needs to be included – but that again comes down to whether the assumptions made are appropriate for the purpose.*

Answer: We thank Keith Beven for the review, and for pointing out many commonalities in understanding the political processes of developing and using models while challenging our argument that the model itself is not neutral. We find this challenge valuable because it is a common point of alienation between hydrologists and critical water scholars that we wish to constructively engage with in this article.

We here summarise how we will deepen our argument in the revised article:

We first emphasise that the non-neutrality of models and their power is not necessarily a negative trait, and that a model is not all-powerful. Its power manifests in different ways, some very visible and some not, and sometimes successful and sometimes less. For instance, models can legitimise and/or challenge policy projects; they provide certain pathways for action and potentially exclude others, and they reify assumptions about the world. Even if a model is only used in academics it still informs the way we think about those water-related issues. Models can make it challenging to question certain assumptions in a model, especially as choices are encoded in the model and become “black boxed”, which makes it increasingly difficult to unravel, especially for non-experts. They can also work in an in- or excluding way, including through certain jargon and language, or certain technology and visualizations of the results used.

Second point is that models are not just a partial, but also a specific representation of reality. People involved in the modelling process might have different ideas of what variables, boundaries, scales etc. are relevant or not (depending on their knowledge, values and experiences) in light of the defined purpose of the model. So, whether a model is “fit for purpose” will be assessed differently by different actors. Modelling involves many (conscious and unconscious) choices and assumptions, which are the product of – and in turn influence -the interplay between different actors (commissioners, users, modelers and affected stakeholders). Each bring in different expertise, world views and ideas for the future. This interplay is enabled or constrained by technology and the modelling process.

Third, not all stakeholders have the same ability to influence the modelling choices or use the modelling outcomes. It is therefore that we wish to reconnect both model developers and users with the models so that they recognize the possibilities that the model enables and those that it forecloses and to engage with this constructively.

In the attached file we share our answer to the other points made by the reviewer.

Reviewer: *The confounding of this distinction starts in the section on Defining Models which suggests that a model is anything that takes an input to produce an output. There is nothing here on purpose, fitness-for-purpose, responsibility or explicitly defining the condition tree of assumptions (including*

considerations of uncertainty) in any model application. Indeed, the study cited by Ramsey (2009) (L165ff) is an example of where the assumptions used made the model not fit-for-purpose (see the discussion of fit-for-purpose in Beven and Lane, 2022). The later example of Andersson et al (2004) is also an example of an application where the model was not fit for purpose. So is not part of the issue here the question of how to get consensus on what might be fit-for-purpose (depending on the purpose). It could be useful here to bring in the idea of the condition tree of assumptions as a way of improving this process – both to get participatory agreement and understanding (and an audit trail for later reconsideration). See for example, Beven et al. CIRIA 2014; or the Page et al., HESS 2023 toolbox). This is a way of getting the more open/explicit approach suggested in Section 4.1.3 and in the conclusions.

Answer: We agree that purpose and fitness-for-purpose are important considerations in the practice of building and using models that we will discuss more explicitly in the revised article. The contestations about the purpose of a model and how that translates into the model choices, and the fitness-for-purpose of a model show the political nature of the model and the modelling process itself. We like the idea of an audit trail, to which we may add our aforementioned political considerations. For instance, what a power-sensitive modelling approach could contribute to the principles you share ‘For a Turing-liketest for model plausibility/(in)validation and fitness-for-purpose’ in Beven and Lane (2022: pp 7) is that it could help to guide additional questions such as ‘whose purpose’ is served, and for what? (related to point 1). And what is relevant expertise to take into account, and does this only concern environmental modellers, or also other types of knowledge? (related to point 7).

Reviewer: *I would argue that the political in the sense used by the authors is not really the “broad influence of the models” that they suggest but rather the whole framing of the decision making process, within which the model is only a tool. This can often be seen in how historical legal constraints on water resources management dominate any attempts to achieve sustainability of use (for either human use or biodiversity) regardless of whatever model might be used. I accept, of course, that models are not necessarily neutral in this process, even if most modellers that I know of will try to do the best they can given the data they might have available). The assumptions can certainly reflect the power of requiring a certain outcome (a nice example, in the inquiry into the Sellafield Rock Experiment Facility as the first stage in the UK nuclear waste disposal strategy, where the two opposing sides used quite different assumptions in assessing groundwater flow pathways).*

Answer: We agree that models are connected to a wider political process of managing water. But models are not just tools because they encode a certain mindset of how to understand and manage water as described above. In fact, the very notion that they are just tools, which is very prominent in water management, obscures the many choices in structuring understanding and management of water (e.g. the optimisation paradigm further discussed below) that are inscribed in the models. In that sense they are even more politically powerful in pushing a particular discourse because they are portrayed as being neutral. The Sellafield example is a good one, although we are also interested in the more nuanced (everyday) workings of model politics.

Reviewer: *The idea of purpose also interacts with framing – this is evident in the examples of the Rhone and Seine in Section 4.1.2. These projects were tasked with different purposes when the projects were commissioned – it is too simple to say that this was the result of contrasting world-views of geographers and engineers. It was actually the result of upstream decisions about commissioning the projects before any model or modeller was even involved (actually that might not be entirely correct, as I believe there was initially a commission project at least on the Seine to consider what might be technologically possible - but more background detail would be needed to make that point). It is again the wider political framework that is determining outcomes here – not the model (though, yes, as above for the Seine, in some*

cases the available model technology might feedback to how projects are defined, but that is not what you are saying here. It would be an interesting study in itself – many of the commercial SHE model projects were of that type, for example).

Answer: We shall make clearer in the revised article that the political nature of modelling is always an interplay of actors (commissioners, scientists, modelers, politicians, other stakeholders) and the model itself. Bouleau (2014) describes in the article on the Seine and the Rhone how management-cultures shape how models are designed, and in turn how models shape their environments. It is thus both the (also historical) wider political framework that is described in this article, and how it manifests in and through the model, as well as the influence from the models themselves. Especially the articles reviewed under section 4.3 ‘Modelling and real-world impact’ showcases this interaction.

Reviewer: *L265 – you put ‘optimal’ in commas but without further comment when there has long been discussion in hydrological modelling critical of the concept of optimality, either in model outputs or in decision making (e.g. my 2006 equifinality paper and earlier). So yes, modeller choices matter and make a difference but these are old examples now – better if multiple model outputs are considered (model ensembles are now widely used) in a context of assessing fitness-for-purpose within the limitations of uncertainty.*

Answer: We shall explain the commas in the article. What we add is that even with ensemble models the political charge of models does not go away – we will take up this point in an expanded discussion. Although these are old discussions, and perhaps an old example, the concept of optimality is still alive and is relevant to discuss in the context of the review. Also, few scientific articles describe the impact of this modelling practice in a case study. More openness on the effects of these modelling choices could contribute to changing these practices.

We also fully agree that it is indeed better when multiple sources of knowledge, including models, are used for decision making. This also hints at the different influences different models can have.

Reviewer: *L297. – yes, but there was a history to that in a simplified interpretation of the results of a CEH study of the impacts of climate change on the Severn and Thames catchments. It was conditional on the climate scenarios chosen and available at that time (and so should have been subject to revision). It is not what the models or modellers of that study said but, for simplicity, (and cheapness in low budget applications) it has persisted. Another example of the dominance of the wider political framework over the model.*

Answer: We are grateful for these further insights. Still, whatever wider political economy and modellers’ choices played a role, they are now encoded in that particular model and not so easy to unpack (hence your insider knowledge is so valuable).

Reviewer: *Section 4.3.1. OK, so models can be used in ways to support vested interests but that again is a political issue, not a model-specific issue (as in the Sellafeld example above where the opposing side effectively won the case by demonstrating the huge potential uncertainty in model results. Similar model, different assumptions. So why is the comment about models being value neutral just thrown in at the end without further discussion? That is a claim made for political ends, not a feature of the model.*

Answer: This too we shall make clearer in the revised article. Models have characteristics that make them have specific and differentiating effects for different stakeholders (through the model outcomes), and allow different people to participate in the modelling process, use the model and model outcomes differently.

If models are purported as being value neutral when in fact they are not then the models are implicated in this political maneuvering.

Reviewer: L335. *In respect of groundwater models there are also examples of studies showing that different experts come up with different conceptual models (e.g. Refsgaard conceptual groundwater model paper), but, more important here, are the post-audit studies of Bredehoeft and Konikow (1992) and Anderson and Woessner (1992) in a special issue of AWR who showed that nearly all groundwater models did not prove to be correct with hindsight, but mostly because the projected boundary conditions had not proven correct (ie. the assumptions rather than the model again). That has put people off doing post-audit analyses for both hydrological and groundwater models ever since..... (even though that could be an invaluable learning process).*

Answer: Thank you very much for bringing this up. We agree with the importance of post-audit of models, and doing that together with affected stakeholders. We would see the assumptions as part of the model, not something outside of the model.

As we state in line 335: “Kroepsch (2018) and Sanz et al. (2019) discussed how groundwater models can be used to legitimise policies even if there is limited information available.” It is telling that there are few post audit analyses done for these kind of projects indeed, and a point of attention for our joint project to improve models. We would call for a power-sensitive post-audit, both to learn how to improve the representation of the physical – including the boundary conditions as you point out, as well as to reflect on the influence the model had and whether this was constructive in the decision making process. We will add a reflection to this end in the discussion (section 4.5).

Reviewer: Section 4.3.2. *“When modelling is presented as a neutral scientific process legitimacy given by external consultants.” These are surely not the same thing as is being implied here. The consultants may well have been trying to do the best job possible given the data available, independently of any vested interest (you do not provide any evidence to the contrary). The model, given better data might well have been a better representation of the system. The model can be neutral in that respect; even if the way it is used might not be. The question again is whether that implementation should have been considered fit-for-purpose for the decisions being made. “Framing their actions as illegal...” also surely has nothing to do with the model?*

Answer: To us it’s again about the model’s influence in these situations. Other comments made above apply.

Reviewer: L363 *“The decision over water allocation was eventually enforced through influence at the highest political level, the President of Mexico. Jensen (2020) also confirmed that the power of high-level decision makers plays a key role. In the case of the Mekong, the author showed there is a certain saturation in knowledge developed by models, and there is a clear limitation in their impact as governments were unwilling to build on these insights.” But exactly! That is not the same as you then go on to say “The previous examples show how models can work exclusively...” when their outputs are rejected or ignored????? The models are not working exclusively in support of the chosen solution surely? And what you discuss in the remainder of this section is how a model might be useful as a tool in framing good practice (as was also the case of Pickering). But it is again not the model as such (which might still be too much of a simplification or lacking in data), but the way the model is used.*

Answer: We will rephrase this section to strengthen the consistency of the argument. Indeed, the influence of models is relative: models will not automatically produce outcomes. However, models and model outcomes do make certain usages possible and disable others.

Reviewer: *Section 4.4.1 You mention the importance of scale here but not the importance of visualisations in how local people can interact with model outputs (as in the models of everywhere concepts of Beven, HESS 2007, Beven et al., JRBM 2014; and Blair et al., EMS 2019).*

Answer: That is indeed an omission in the article, and thank you for pointing it out. How the output of a model has influence, for instance through visualisations, graphs, scales or perhaps interactive dashboards, has not specifically come up in the articles we reviewed. It is interesting indeed, as there are ample articles that point to the politics of maps, or for instance the influence of a certain use of colour.

Reviewer: *Avoid models that are overly complex? Over-complex with respect to what? To the problem at hand or to the understanding of stakeholders. The second should surely not override the adequate complexity of the first?*

Answer: The advice of Opitz-Stapleton and MacClune (2012) to reduce model complexity is in relation to connecting and discussing the model with stakeholders. There are cases in which researchers have chosen to work with simpler models to enable stakeholders to engage in the analysis, for instance also suggested by Srivinasan et al. (2018) of which the article is included in the review. In the revised version we will clarify this point.

Reviewer: *Having said that, the power relationships of this type of co-evolutionary modelling are definitely an issue, but as some of your example studies have shown not necessarily insurmountable if the will is there. But many of the problems you have identified so far are a result of the imposition of power structures, regardless of the model or its results.*

Answer: Comments above apply. Those power structures are inscribed in the models.

Reviewer: *Section 4.4.2. Consideration of assumptions about uncertainty (including epistemic uncertainties) in modelling is critical to model evaluation and fitness-for-purpose but this is only really mentioned in this Section in relation to an application of SWAT (widely applied – it is free to use - but which elsewhere has been shown to be not fit-for-purpose for the type of application described here, even allowing for uncertainty – see Hollaway et al., JH, 2018).*

Answer: The main point we want to show is that the characteristics of the model influence the interaction between the model, model users and affected stakeholders. Each stakeholder will make a different assessment of the fitness-for-purpose of the model and the treatment of the uncertainties. That process makes the model political.

Reviewer: *What do the authors mean here by “As modelling inhibits more uncertainty than measurement” (L431)? The model here is being used to predict years ahead so should surely be more uncertain than any available measurements (that cannot in any case be made in the future).*

Answer: The sentence refers to the summary of Wardropper et al. (2017) on how inherent uncertainty in the Soil and Water Assessment Tool (SWAT) application to the Yahara Watershed in Wisconsin (USA) influenced the development and implementation of a water quality management program. It is indeed a redundant sentence, which we have adjusted into: “The authors questioned how the inherent uncertainty in this approach affected people in the watershed.”

Reviewer: *L447 Pickering – “ultimately played a key role in shaping flood management strategy in the area.” Well yes – but you should then perhaps finish the story. It showed that the NFM strategy preferred by local stakeholders would not protect the properties at risk. The fact that the EA had also been involved in the process then meant that they invested £1.5m in a concrete flood detention basin, despite the cost-*

benefit for the scheme being considered too low compared with other sites. That was then a political decision.

Answer: We appreciate these further insights. Again, different models were implicated here in developing various intervention scenarios and cost benefit calculations – irrespective of the final political decision made. We don't claim that models make decisions, but that they are implicated (sometimes more sometimes less) in these decisions. What interests us here is that the original cost benefit model was challenged by an alternative model, even if this didn't suggest the interventions that were ultimately made (so it's more about the political power of the first model being challenged here).

Reviewer: L459. *“they conclude that models assuming that residents are well informed and have shared understandings of the water supply system might lead to an oversimplification of sociohydrological dynamics in a given location, and that more local involvement could mitigate this”. OK (though not clear what type of quantitative model you might mean here) but is that just not another example of poor assumptions/poor practice/not fit-for-purpose/design for vested interest – ie. there is again surely a need to distinguish between the model and the way in which it is used.*

Answer: Comments above apply. It's about the encoding of assumptions in models and the obscuring of the encoder.

Reviewer: L477 *“which raises questions about the responsibility and accountability of those making and using models” Well yes, and that is the problem I have with most of your discussion since you are placing responsibility on the “model” and not on those who use and misuse them (not necessarily even the modeller, but even more so those who commission studies and/or use the results for decision making as you have demonstrated).*

L606 *“how quantitative models may help to foster transformative pathways towards more just and equitable water distributions.” But why put the focus on the model here? What is needed is the political will for more just and equitable water distribution. Given that will, everything else would probably follow, but I do not see how you expect models (or modellers come to that) to influence the neoliberal capitalism or centralised communist systems that prevail in most countries where sustainable and equitable water issues are important. Look at the UK – we do not have the problems of degrading water quality because of any modelling or modeller issues. And announced this week is a relaxation of the rules on new housing developments in respect of water quality. A relaxation of rules in designating water quality categories is also expected now that we do not have to conform to the EU WFD standards.*

Underlying many of the examples you provide is exactly that political framework, outside the control of modellers and their models.

Answer: As explained above we see the model as being political as its characteristics and outcomes enable and constrain uses. It is the interaction between the commissioners, model, model makers and users and stakeholders that leads to policy decisions and implementation. In the case of the relaxation of water quality rules, there are many groups that will engage with trying to protect water quality in the UK. Modelling may play a role in this too.

Our call for power sensitive modelling is indeed a call for those who are part of modelling processes – including commissioners, modellers, and users, to become more critical or sensitive but not only to the context within which they use their models, but also through an understanding that there is an interaction between the way a model is made and how it functions in that context. It is therefore that we think that a power-sensitive approach could be useful.

Reviewer: *So while I appreciate the sentiments that lie behind the paper, I think you have got the framing wrong. Most of the power issues in moving towards better water management have very little to do with models or modellers – they are political (in the sense of the authors' use). So to say that models need to become power sensitive is not correct. It might be better put that modellers need to be more critical or sensitive to the context within which they use their models – both in terms of the framing of a project and whether a model is fit-for-purpose within such a project, but it is often the case that there is much that is outside their control (the legal basis of water rights in a country, the details of a project commission, etc). You give examples of where model outputs were rejected because they were in conflict with a desired outcome – that is surely a bigger problem of power.*

Answer: Yes, but important to us is again the encoding of power sensitivity in models (which might require ensembles of models). We shall try and make that clearer and that we indeed argue for a reconnection of the responsibility of model developers and users to models.

Reviewer: *So my apologies for these extended comments but it is something I have thought about over a long period of time. I think there are ways ahead (I have suggested some partial solutions in the past such as the condition tree / audit trail of assumptions, models of everywhere as ways of facilitating interaction and criticism, better evaluation of fitness-for-purpose and consideration of model uncertainties) but I think this paper needs to be reformulated much more about power in the political framing of the modelling process than assigning the responsibility to power sensitive models.*

Answer: We are grateful for this interaction which we believe will result in a refinement of our arguments as detailed above.

Reviewer: *Some other points*

The choice of papers analysed seems incomplete. For the once case study that I know something about (Pickering) some papers are included (the papers by Lane et al. and Landström et al), but others are not (particularly Lane et al. 2011, Doing flood risk science differently in TIBG, and the papers on Pickering by Sarah Whatmore). Were these considered as having too much overlap or is it an indication that the methodology could have retrieved other relevant papers?

Answer: In this case, it was the overlap between those papers that prevented us from citing all of them. As we aim to develop a growing database, we will make sure all papers on this project are included in the analysis.

Reviewer: *The Morgan and Morrison 1999 paper does not appear in the references.*

Answer: Thank you for pointing this out. We have added it.

Reviewer: *The Pielke Jr. reference appears twice.*

Answer: Thank you for pointing this out, we have deleted one reference.

Answer to reviewer 2

- 1 My apologies that this review is a little late. This paper needed quite some thought. I find its goals useful and there is some very useful material therein that the hydrological modelling community should be aware of. I should be honest about a few of my own positions with respect to this work. I don't particularly like this kind of methodology – finding literature this way often means that very interesting papers that perhaps frame themselves differently get lost – and because they are not found other kinds of framing that exist are left hidden. Second, I apologise that there is quite some reference to my own work in my report but I have been working on this topic for almost 20 years. Some of the ideas the authors have could and have been developed a lot more (and the work I refer to, notably Lane 2012, might help them access the work of others besides me). In especially the Discussion, I felt that the authors did not really show as much awareness as might be ideal of some of the dimensions that they address (mental images, framing) and which have already been discussed in relation to hydrology. There could be a lot more useful Discussion and crucially, perhaps my key recommendation, some much deeper thinking regarding what a power-sensitive modelling approach might look like and how it might come about (which itself would have to be a political act).

The detailed comments identify some of these issues that could be developed. As major comments though, and in addition, I think the following need some treatment in the paper.

Answer

Thank you for your supportive comments in relation to the aims of the review.

Regarding the methodology, we are aware we can never present a complete overview of all papers that engage with the influence of modelling in water, and reflect on this in the methodology as well. It is also challenging that different disciplines write about the topic with different jargons, and also that reflections on the influence of models and modelling are sometimes hidden in papers. We have therefore decided to make a start with collecting papers, combining two methods; a narrative review in which we identified papers that we were already aware of, and in addition did a systematic review to identify others. Also, we specifically have opted for HESS that has an open review process to invite people to share more papers.

We will enrich the discussion with references to papers that have addressed the question of what power-sensitive modelling might look like.

- 2 1. Models are themselves political objects – they contain intellectual property that can come to have commercial value – the history of how flood inundation modelling went from 1D to 2D in the UK was about primarily the protection of vested interests in intellectual property (for example, see one of the sections in Lane, S.N., November, V., Landström, C. and Whatmore, S.J., 2013. Explaining rapid transitions in the practice of flood risk management. *Annals of the Association of American Geographers*, 103, 330-42) – this is similarly reflected in James Porter's fascinating work about how flood maps get made (which is essentially based upon flood inundation modelling) and the negotiation between modellers/mappers and regulatory agencies (Porter, J. and Demeritt, D., 2012. *Flood Risk Management, Mapping and Planning: The Institutional Politics of Decision-Support in England*. Environment and Planning

Answer

Indeed, the economic dynamic of modelling has not been explicitly touched upon in the articles reviewed, and we acknowledge it is an important dynamic/driver in relation to the influence of models. We will update the review based on additional articles we have received and that will include a mention of this either as a separate section or in the comments, based on the outcomes of the review.

- 2A A). There is a tendency to reduce politically sensitive modelling to the interface between models, modellers and society – and to overlook the politics of modelling itself and how this also needs to become

more politics sensitive. In turn this would sit this paper also in wider Science-Technology-Studies perspectives on power within scientific practice. Politics exists within modelling communities and this can have a profound impact on how modelling is done. Politics can constrain what is modelled and how. This needs more consideration and discussion.

Answer

Thank you for bringing this up. The politics of science in relation to models has been included in the review, especially in sections 4.1 and 4.2 on ‘Mental models and policy projects’, and ‘The influence of modellers’ choices’. Where possible, we will elaborate based on the additional papers we will add in the review. Also, we agree it can be more specific in the discussion, as well as in the call for power-sensitive modelling and will update the article accordingly.

- 3 2. If we accept the hypothesis that power is imbricated with and within models of the water environment (which I do), the question becomes what to do. The paper has very little on what this means for practice. A starting point may be the obvious point that we should not make recourse to some kind of naïve goal that we should exclude power and politics from modelling such that models give us access to some kind of higher, truer knowledge and models can then reign supreme ; that would simply be an internal contradiction in that it would imply the transfer of power to models and modellers (that is itself be a political statement). So, if power and politics cannot be excluded what would a more democratic form of modelling mean in practice. What do we mean by “democratic”? How might different kinds of politics (e.g. majoritarian versus minoritarian) lead to models being used, power-sensitively, in different ways? What are the ways of using models that allow those excluded from decision making to become included in decision-making through the use of models. When I got to the end of the discussion I felt short-changed that after all the examples this hadn’t been thought through.

Answer

Thank you for bringing up the question: what now? This paper is part of ongoing efforts to understand the power related to modelling, and other articles are planned that engage with the question: how do we practically engage with the power of models? We base this work on an understanding that each model and modelling process is different and that this requires a flexible approach.

What we will do in the current article is to provide a further explanation in relation to the four elements (the mental models and policy projects, the influence of modellers choices, the impacts models have, and engaging with non-modellers) more explicitly, as well as reflect more on practical examples from the reviewed papers on how to engage in a power-sensitive way.

- 4 3. When I wrote the 2014 paper published in HESS I should confess that this was motivated by an extreme skepticism (that I still hold strongly) of socio-hydrology and the focus of the Panta Rhea decade; this included the appearance of diagrams coupling models of people to models of the hydrological cycle that reminded me of those the Dick Chorley and others were advocating in the 1960s and 1970s. Yet again, we as hydrologists seemed to be bolting people onto the natural environment with very little clue as to actually what the “socio-“ really means. This was ironic because at the same time there was much interest from social scientists, notably Linton, Budd etc., in the Hydrosocial cycle which had, in my view, a much deeper understanding of how people and water are truly coupled. Now, I regret leaving in Lane (2014) this argument a little too subtle. But, I think the distinction between socio-hydrology and the Hydrosocial cycle does actually matter to this article because the need for political sensitivity reflects (for ethical reasons, at the very least) the need for a much more nuanced understanding of how hydrological models fit into the world. This is lost at the moment in this article. It probably does not need much discussion added but the notion of a hydrosocial cycle and the mutual influences between people and water strengthen the argument that we should also take a political perspective for how hydrological modelling is done.

Answer

Thank you for bringing this up, and we fully agree with the importance of socio-political variables in understanding water systems. However, we did not explicitly focus on comparing socio-hydrology and hydrosocial studies because that could (and perhaps should) very well be a new paper. Nevertheless, we acknowledge that these different ontologies are important to mention, not in the least as the hydrosocial cycle acknowledges the political as a transformational force capable of changing landscapes. We will dedicate time to engage with these different ontologies in the introduction and discussion.

Minor comments

- 5 L36 – but also by the situated nature of the modeller themselves – and the work they do to come to trust their own models – see Lane, S.N., 2012. Making mathematical models perform in geographical space(s). Chapter 17 in Agnew, J. and Livingstone, D. Handbook of Geographical Knowledge. Sage, London
This is a very good point; we will integrate this in the article.
- 6 L36-7 – your references here, unless I am mistaken, largely support the counter-factual – do you have any better references of modellers themselves presenting their models as neutral tools? Without this there is a risk that you are setting up a “straw doll”.
Indeed, often these positions are not made explicit, which is also shown through the literature review we did. Neutrality is often assumed and non-neutrality needs to be actively proven to challenge the dominant paradigm. Hence the chosen references.
- 7 L37-8 – I don’t agree with this as it is written – we may think models can travel easily between places but much modelling in practice is about making them work in particular settings. There are studies of hydrological modelling in practice – see for instance Landstrom, C., Whatmore, S.J. and Lane, S.N., 2011. Virtual Engineering: computer simulation modelling for UK flood risk management. Science Studies, 24, 3-22
A challenge in our article is that we cover many different models, at different moments of their development. Some require a lot of work to make them fit, and others travel more easily based on how they are designed and what they do. We will tweak the sentence: “An additional challenge is that models become increasingly complex and some travel easily between places of application.”
- 8 L63 – the notion of power sensitive modelling is interesting – but I think the paper is missing a wider link here into science technology studies that argues that power is an inevitable (and sometimes malign) component of any piece of scientific investigation. Isabelle Stengers has treated this generally (her book “Une autre science est possible” is probably the simplest entry point); there is then work around any one of a number of disciplines exploring this as well as some very specific to modelling of the environment (Demeritt, 2001 is a classic in this sense). I appreciate that the paper is about water but it is not well situated with respect to the wider STS work on the demonstrated influence of power upon scientific practice (and the very obvious point that the moment you advocate a model as neutral and therefore *the* valid basis for decision-making, you immediately transfer power to the model and the modeler, that is you make a political decision). Some of these examples also could be used to develop the argument around L90 a bit more – where it is under-developed in terms of what we already know about the relationship between power and modelling.
This comment strongly relates to the second point you make (or 1A), and we agree that, through the review, we have focused mostly on the influence of models in practice, and less on power in scientific practice. We will bring this out more explicit in section 2, defining models, as well in the discussion, and also bringing out these points through the STS work we have included, such as Knorr-Cetina, 1999, Bijker, 2017; Bijker et al., 1987; Latour, 2000; Latour and Woolgar, 1986; MacKenzie and Wajcman, 1999.
- 9 Shackley, S., Risbey, J., Stone, P., and Wynne, B.: Adjusting to policy expectations in climate change modeling – An interdisciplinary study of flux adjustments in coupled atmosphere-ocean general circulation models. Climatic Change, 43, 413–454, 1999
Demeritt, D.: The Construction of Global Warming and the Politics of Science, Annals of the Association of American Geographer, 91, 307–337, 2001
Lahsen, M.: Seductive simulations? Uncertainty distribution around climate models, Social Studies of Science, 35, 895–922, 200B.

Sundberg, M.: The Everyday World of Simulation Modeling: The Development of Parameterizations in Meteorology, Science Technology and Human Values, 34, 162–181, 2009

Bryse, K., Oreskes, N., O’Reilly, J., and Oppenheimer, M.: Climate change prediction: Erring on the side of least drama?, Global Environmental Change, 23, 327–337, 2013.

Answer

Thank you for these references. We had initially chosen to keep the introduction and definition of models short, and focus on showcasing and learning from the “water”-articles to acknowledge the specific approaches and culture that define how models are used and discussed in the water sector. We also see the importance of bringing in lessons from other fields, and will dedicate space for this in the discussion.

10 L84-88 See Lane (2012) op cit.

Answer

Thank you for the reference.

11 L108 – water is not just about hydrology but also hydraulics – is there a reason you did not also look at hydraulics. Note also that this way of searching likely misses more specific model applications or where models are not really referred to as “models” even if modelling is an integral part of what is being done. This is true, and we will redo the query to include hydraulic models and other types (hydro-dynamic for instance) by adapting the query to:

Answer

TITLE-ABS-KEY ("water model*" OR "hydr* model*") AND TITLE-ABS-KEY (justice OR equit* OR politic* OR ethic*).

(including 307 articles instead of 293, part of which can be explained by including now the 2nd half of 2022. We will also include the articles published in 2023, resulting in 323 articles included in the systematic review)

12 L115 – but isn’t this also an important result in the context of this paper?

Answer

Yes, it is indeed an extremely important outcome! There is very little written, unfortunately. We will highlight this more explicitly in the article as outcome.

13 L124 – clarify what you mean by “narrative style”

Answer

We have changed this to ‘narrative review’ to clarify

Answer to the reviewer 3

Reviewer: First of all, I have to apologize for being incredibly late with my review.

Then to the content of this review itself: I am afraid that my conclusion is that I think this paper is not ready yet for further review after revision. I find the topic of the paper crucial for the hydrological community, if only – but not restricted to – because the hydrological community has recently discovered that human agents matter in how water flows are shaped and that hydrological models have a role to play in the world. The socio-hydrological ideas and the HELPING initiative are signs of this, respectively, and each with some rather major issues on what they actually entail in theoretical and practical terms, if you ask me. Having said that, this paper is not the input I think we need to enhance the debate on such issues. I have three main reasons to defend this (harsh) claim about the paper, each of which I will detail below:

1. The method is arbitrary and not transparently displayed
2. The content discussion tends to the superficial and does not engage with crucial issues
3. The conclusions are not supported by the review itself and include some strange aspects

Answer: Thank you for your comments. We will build on your suggestions to improve the manuscript, especially regarding an elaboration on the method we used, to deepen the discussion and strengthen the conclusion - including clarifying the aim and outcomes of the review. Please find our answers to your specific comments below.

Reviewer: Method

What (some of) the authors did was search for material with key words in a database and write a text. One could use specific names for this, but I do not see anything structurally different from my simple positioning of the method. That method is in itself obviously the basis for most literature reviews, but I would have liked to see much more detail about the choices – as the search itself is shaping the data (the resulting texts). The choices are described, not explained and not detailed.

Answer: We will elaborate on the methodological approach used in the literature review, specifically the ROSES method for environmental sciences (<https://www.roses-reporting.com/>), and in the thematic analysis including further explaining the choices we have made in terms of the query and the selection of texts. We have started with a review of articles that analysed elements of how models have influence, that were familiar to us as good examples, or that came up in a general search. We analysed the keywords listed in these articles to identify if a common query could be distilled, which proved to be impossible due to different jargon or different approaches to the analysis of model-use in the articles. We have then tried different queries based on keywords such as ‘influence’, ‘power’, ‘values’, ‘reflexivity’, ‘accountability’, and ‘responsibility’, or specific theories that engage with deconstructing models and their role in society, including ‘Science and Technology Studies’, or ‘Social Construction of Technology’. None of the queries yielded sets of articles that analyze the socially and ecologically differentiating effects of the use of models. This testing phase helped us to define the final query with alternative words including ‘justice’, equity, politics or ethics, that resulted in the articles we reviewed in the article. We accepted that the results of the query included many articles that did not answer to our research question, based on experience with the other queries, or wish to avoid biases, as well as the interdisciplinary nature of the research, and because the query did help us to identify articles on the influence of models that were unfamiliar to us.

In the revised version of the article we will explain more about the literature review methodology and process and why the majority of the 300 articles did not qualify.

Reviewer: The step to reduce the number of texts from about 300 to about 20 suggests to me that the keyword search was actually not very useful. How somehow magically afterwards some 30 other papers

were found to be added remains obscure as well. The search did not find them – other type of text perhaps, but see below on that? I am quite certain that I would come up with a different selection of texts if I would follow the method as the authors describe – assuming the method description would provide the details I would need to replicate anything.

A search strategy with the keywords runs the risk that modelling applications that may have been of interest to the review – for example how stakeholders are involved or how models are used in practices – without using the specific terms of interest of the authors of this text are not found. One can never be certain in a review to have found everything relevant, but one could do a better job to stretch the limits of available texts using approaches like citation analysis. Do these texts/authors cite each other, do they refer to similar texts – or not? These authors bring the texts together, but do the texts agree?

Answer: Thank you for the reflections on the methodology.

In terms of search strategy: We agree that the query indeed provided few results, showing the very limited number of studies on the socially and ecologically differentiating effects of the use of water models. We have used a hybrid approach to the collection of articles for three main reasons: From experience we know that articles that discuss the implementation of models in practice, or that discuss stakeholder involvement, do not necessarily discuss or reflect on the influence of models. Secondly, we wanted to avoid a bias in the collection of articles by only collecting it through general searches or only articles we were aware of. Also, the chance that papers, that are collected through a query on the topic of the influence of models, are not eligible is high; We are dealing with different jargon and words with multiple meanings (such as power), as well as a tendency to refer to the potential influence of models without elaborating or questioning it. In the article we will elaborate on this more.

We will clarify better in the revised article why the methodology was chosen, and how the papers were selected in both the narrative review and through the systematic review.

Reviewer: It would have been useful to specify much more about the papers that were finally used. Regional coverage of cases and authors, years of publication, journals, key words, type of water that is modelled, type of model used, perhaps even which equations were used – all these seemingly administrative items can already be revealing about what the community of authors does represent.

Answer: That is a good suggestion. We will include more information about the papers in a table in the supplementary material.

Reviewer: The texts that are finally used include both papers and PhD theses. Is that actually allowed – or perhaps less strict “advisable”? It does include some issues of how much papers can include in their word count compared to a thesis, right?

Answer: We do acknowledge there are different scientific practices on whether it is acceptable to refer to PhD theses and that it is a very valid discussion. For this article we have made the decision to include them as some of them are not published as articles, yet represent in-depth research on models. We therefore do see the value of including them in the analysis, but we will emphasize in the review which texts are the PhD theses.

Reviewer: How the narrative aspects were defined (on their content see below) is obscure. The reader must trust the authors that what they have selected makes sense, including which text is used for which aspect. I happen to know a few of the texts, and my selection of themes and the link to texts would have been different. A sample of $n=1$ is weak (yes, I know) but I would have used something like text analysis in combination with the admin-indicators I mentioned earlier to make my selection process of themes transparent.

Answer: We will elaborate on how the themes and subthemes emerged from the papers. In brief, we first coded the articles using an inductive or ground-up approach (see for instance Linneberg and Korsgaard,

2019). We derived the codes from the articles based on the question of what elements in the papers related to how water models and their modelling processes have influence. Through different iterations of the codes we combined them in thirteen themes/codes that we subsequently grouped in four overarching themes.

We agree that the selection of the themes is shaped by our own reading and analysis of the articles as well as our experience in modeling and research modeling practices. In revising the paper, we will be more explicit about the selection of the themes and discuss the limits of our methodology.

Reviewer: Content

The Introduction is very general and surprisingly low on references. Section 2 does include a few more, and might be integrated into the Intro. Are some of the references in 1 and 2 perhaps also feasible inputs for the review process itself?

Answer: We have chosen for a general introduction to share why we write the article, and a more specified section that discusses our definition of a model to share with the readers our understanding of what models are. We think that discussing the reason of writing the article (intro) should not be conflated with section 2 (definition of models) as these are two distinct topics, but we can enrich the introduction and link to more existing literature.

Reviewer: The definition of what a model is to the authors is very broad. I would argue that the definition holds for almost any theory. I would actually agree that hydrological models are indeed theoretical claims, but that does not mean I would use such a generous definition. A broad use of terminology might actually obscure that the quantification aspect of models might be quite important – including differences between types of models, the actual equations that are used, the temporal and spatial coverage and steps, etcetera. This would also allow including the notion that many hydrological signals can be mimicked in many models – up to the point that rainfall can be estimated from a groundwater model, something that was obviously not the idea of that model when designed. The realization within the hydrological community that anything can be modelled has actually created a debate within that same community, perhaps not yet in the terms of the authors, but possibly of relevance for them.

Answer: Our choice of a rather broad definition of models is motivated by the wish to allow for discussions between different disciplines. More specific definitions are possible, but we found them less helpful in our interdisciplinary analysis. We have taken this broad starting point in combination with a review of articles to derive what elements can be important when understanding and engaging with the influence of models. This also includes the quantification aspect that we discuss in section 2, and also in section 4.3.1 Naturalising and legitimizing world views through models. We agree this can be made more explicit.

Regarding the reference to the debates in the hydrological community. We realize there are many debates ongoing on modelling in general as well as on specific models. We aspire to constructively contribute to these by drawing attention to the potential power of models and the role of the commissioner, modellers, users and the potentially affected, which is little debated in scientific literature.

Reviewer: The broad definition and some of the contents mentioned (see below) would have actually allowed the authors to include discussions within the hydrological community on building links with other fields, on models as I already mentioned, and comparable more interdisciplinary awareness in the hydrological community to position their model debate. Work by the group of Dr. Van Loon comes to my mind, to name just one example. I do agree that models are worth discussing, but let's not argue that they are special in terms of supporting the powerful, offer limited world views and the like. Research efforts have the tendency to do the latter, with only a few having an answer to the former.

Answer: We recognize that in the hydrological community there has been an increased interest in building links with other fields - some of the authors of the manuscript are actively working to this end. We dedicate attention to these developments in section 2 when we discuss models. We are less aware of collaborations that focus on analyzing the socially and ecologically differentiating effects of models and modeling practices themselves. We think there is value in dedicating specific attention to the influence of water models as for instance water infrastructures - shaped by water models - have large ecological and socio-economic impacts, and water models are tools that are frequently used, are often seen as neutral, and as they can be used to support the powerful. This is especially explained in section 4.3 Modelling and real-world impact. We fully agree with you on the importance of discussing models constructively, and in our article we show the value of doing this in a power-sensitive way too. We do not aim to denounce models, but our aim is to explore the potentials of modeling to contribute to more just and equitable water distributions.

Reviewer: This observation of the relevance of other (connected) topics is also clear from the content descriptions. We read about data collection as specific issue, we read about decision on project focus, and – granted – we read about modelling decisions. But why these different aspects can be connected remains unclear. The discussion of the topics does not try to connect the topics, even when several of the texts seem to cover several of the authors’ themes. The topics read like an unevenly distributed shopping list, with an unequal number of texts per topics, with quite superficial summaries of the texts, and again without any cross references between texts and topics.

Answer: We have made an explicit choice to unpack and show the different elements of how models and modelling processes can have socially and ecologically differentiating effects. Based on the articles we do see that the intersection of these elements is an important contribution to the different ways models can have or gain influence, and we will dedicate attention to the interconnection of topics in the conclusion and discussion. The topics are derived from the articles, which explains the difference in length.

Review: In the discussion, suddenly the texts that were rejected for the actual review are used to make a claim about what they discuss or even represent. I am sorry, but that is simply not allowed without a proper review of these texts in itself and providing the reader with the identity of these texts.

Answer: We will elaborate in the methods section why certain texts are not included in the final review, and we will include as annex the list of the articles that we have retrieved through the query and which have not been included in the text.

Reviewer: The conclusion

The conclusions do include quite a number of references. That is at least unusual, but I would say suggests that either the authors do not have a conclusion (we read an extended discussion) or the conclusion does not yet come strongly enough out of the review.

Answer: We will review the section to make more clear for the reader what are key outcomes of the review.

Reviewer: The call for improvement if using models reads like a wish list, without the wished being confirmed as practically possible. The ideas are also rather general and surprisingly de-linked from the aspects that the authors have highlighted in section 4. Again, here the effect of missing possible texts that do discuss interesting uses of models without the terms used by the authors may be seen. Furthermore, a text like Junier (2017) shows quite clearly how relatively good intentions shift within the modelling process – in the sense that the intentions are kept, but the model does no longer align with them. The wish list is nice, but meaningless without much more discussion about implementation in actual practices.

Answer: The call for power-sensitive water modelling is based both on the review of the articles, and on work that is done in relation to the dynamics we identified. It is not a wishlist, but recommendations for doing modelling differently. To clarify this, we will elaborate on the links between the outcomes of the review and the call, and will also clarify the different elements of the call through examples where possible.

Reviewer: One quite strong suggestion at the end is the need to move out of disciplines. This might neglect the less silo-ish nature of the hydrological discipline than the suggestion suggests, but the suggestion also drops out of the sky. The review itself does not clearly prefer this interdisciplinary aspect, so why is it so crucial? The suggestion that the issues with models come from disciplinary focuses is not even mentioned in the review. This suggests that the authors already knew the conclusion before doing the review. In itself the call for interdisciplinarity is not strange, and even done by hydrologists, but it would still need to be related to the review.

Answer: The reason we suggest interdisciplinary collaboration, including between modellers and non-modellers, is that it offers opportunities to create constructive frictions that could be used to allow more conscious decision making in model development as well as facilitate joint learning. Commissioners, modellers and model users do have a role to play here. We suggest this based on several articles that we reviewed, in which authors show that a tunnel vision based on disciplinary world views, norms and values have specific obvious and less obvious effects. This is especially clear in the conclusion of section 4.2 ‘the influence of modeller’s choices, and 4.3 ‘Modelling and real-world impact’. We will revise the conclusion to make sure that the suggestion for interdisciplinary cooperation is better explained based on the review. We will also better explain how disciplines and related world views, norms and values influence modelling based on the reviewed articles, especially in the conclusion of section 4.2 ‘the influence of modeller’s choices, and 4.3 ‘Modelling and real-world impact’.

Reviewer: Furthermore, I find the claim that involving social sciences would solve the issues strange and in need of much more refinement. Are all social scientists equipped for and/or interested in the same issues as the authors? I would argue this is not the case. Why can the observation that many hydrological modellers do things that may be less useful be combined with the claim that collaboration between the general communities of hydrology and social sciences will solve this? The original modellers that did un-useful things will still be member of the hydro-community, right? Why are social scientists in general in the position to teach the hydrologists in general?

Answer: In the conclusion we do call for more interdisciplinary collaboration as a way forward, however we do not suggest that social scientists are ‘in the position to teach the hydrologists’. Our suggestion for interdisciplinary collaboration, which can also be with other-than-social science disciplines, is derived from the review and from our own experience in doing modeling in interdisciplinary settings. It is an invitation for curiosity and joint learning based on the gained awareness from the review that in a modelling process many steps are taken that do have impact on the outcome but these are often not reflected on. Also from the review we learn that models do have the potential to have real-world impacts while often seen as neutral tools, and these potential real-world impacts can be better understood through interdisciplinary collaboration. We will be more explicit about this in the revised version of the paper.

Reviewer: Final remarks

The topic of the text is important, but I think that the evidence that the text brings is not convincingly presented, as I have tried to argue. I would have liked to think that revisions would have been possible to continue the process, but the combination of a weak (description of the) method and a rather unbalanced analysis makes me strongly suggest that the text as is should be rejected.

Please find my handwritten notes on the text in the pdf attached. If there are any questions about remarks (including my handwriting) I am obviously available for further exchange. Again, the topic of this paper deserves further elaboration – but the paper itself needs that elaboration first even more...

Answer: Thank you for your review. We will include your suggestions in the revised manuscript, especially in relation to elaborating on the methodology, clarifying the structure of the analysis, and strengthening the conclusion based on the review. We also remain open for further exchange.