Rebuttal

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 Sellafield 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 Woesnner (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 rephase 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.