

Reviewer 1:

This study did a thorough evaluation of the SHOWER model for modeling groundwater responses to water management scenarios in real catchments in groundwater-rich areas. I found the modeling approach to be robust, well-documented and technically sound. I believe it is exemplary work of interest and value for HESS readers. I have two main comments.

We thank the reviewer for their positive response to our work and welcomed the two main comments, see below.

My first comment: Clarify the knowledge gap. In the introduction, particularly around page 2, lines 40-75, and in Table S1, a variety of existing modeling approaches and limitations are introduced. At this point, it sounds like the paper simply combines and evaluates a groundwater model, rainfall-runoff model, and water management practices model, with a calibration including of management interventions in real catchments. If other models are already doing these, either individually or in combination, then the specific novel advancement of this work should be clearer. I suggest adding a little more background on specific models or cases that do similar things, potentially including those mentioned in Table S1 or models like SWAT-MODFLOW and ParFlow used in other areas. Then, describe more clearly how the SHOWER model and/or the analyses in this study go beyond the previous work to fill a specific knowledge gap (perhaps something related to water management and droughts in real catchments).

We appreciate the concerns raised by the reviewer and believe that we can address this by clarifying what can be modelled well using the current modelling tools available and what we think is missing given the highly managed groundwater systems in the UK. For example, there is limited water management influence included in most of the listed rainfall-runoff models in lines 42-44 and most of the recent modelled water management influence focuses on surface water abstractions, for example in Rameshwaran et al. (2022) or on reservoirs by Salwey et al. (2024). Groundwater is assumed to be largely uninfluenced by abstractions, responding as it would do in a natural system, which is not the case for many regions in the UK particularly in drought conditions (Wendt et al. 2020).

We intended to introduce SHOWER to fill this research gap as it has a simple rainfall-runoff structure that includes (drought) management strategies for both surface water and groundwater stores and fluxes. Another key advancement is that we incorporated different groundwater model options to better simulate groundwater flow in large aquifers in the UK that are used for drinking water supply. With the three different options, we could adapt baseflow generation to account for karstic, fracture or largely

porous flows instead of using a linear baseflow release that is currently implemented in other rainfall-runoff models. These adaptations have shown improvements in groundwater-rich areas in the UK and abroad (Wendt et al. 2021; Stoelzle et al. 2015).

To this end, we have rephrased the second (L42-59) and last paragraph (L84-86) in the introduction to clarify the research gap:

L42-59: The range of hydrological models that is used to inform water management decisions and drought policies in England and Wales is however primarily focused on surface water in unmanaged or ‘near-natural’ conditions, such as Grid-to-Grid (Bell et al., 2007), GR4J/GR6J (Coron et al., 2017), JULES-GB (Batelis et al., 2020), Qube (WHS, 2024). A full overview is available in Environment Agency (2023). Recent advances in hydrological modelling have addressed the lack of management interventions by introducing long-term average and monthly varying surface water abstractions and discharges (Coxon et al. (2019) in DECIPHeR and Rameshwaran et al. (2022) in Grid-to-Grid, respectively) and by adding reservoirs (Salwey et al. (2024) in DECIPHeR and Hughes et al. (2021) in SHETRAN). While surface water processes are typically well-represented in these models, groundwater representation is often simplified. Groundwater is assumed to be largely uninfluenced by abstractions and therefore models typically release groundwater storage as baseflow. Although this is the behaviour we would observe in a natural system, this is not the reality for many regions in the UK where a large proportion of the groundwater is abstracted (BGS, 2024). Additionally, the linear approximation to generate baseflow in hydrological models often results in large errors during floods and droughts in groundwater-rich areas (Smith et al., 2019; Hannaford et al., 2023a). There are a handful of groundwater model options are setup for UK (regions), which vary in complexity. These range from a lumped catchment model approach (Aqumod (Mackay et al., 2014)) to spatially distributed groundwater level modelling with either only groundwater levels (Rahman et al., 2023) or a combination of levels, flows (Zheng et al., 2025), and averaged abstractions (Lewis et al., 2018; Bianchi et al., 2024). However, similar to the range of surface water models, none of these groundwater models includes dynamic abstractions, management interventions or the option to include a drought policy to support decision-making (see Supplementary Material Figure and Table S1 for more details).

L84-86: SHOWER builds on the lumped socio-hydrological model introduced in Wendt et al. (2021) and can simulate groundwater levels, baseflow and reservoir levels for different hydrogeological conditions under different drought management strategies, by applying different methods to coordinating reservoir and groundwater abstractions.

Additionally, we have revised Table S1 to reflect the examples in the text better (see revised manuscript).

My second comment: Highlight results beyond evaluations. The study is very heavy on the technical aspects, and the results are essentially model evaluations, without highlighting further scientific theories or comparisons tested. This is evident in the title (“Evaluation of...”), as well as the methods and results section headers which only go to calibration and evaluation. I suggest bringing the results beyond model evaluation, uncertainties, and parameter sensitivities more prominence. As an idea for this, you could state and test a scientific hypothesis about the groundwater and water management interactions, like a case study in the real catchments to show a scientific application of the model. The paper is already set up with different management scenarios that are tested to see their hydrologic impacts on droughts compared to baseline (e.g., Figures S9 to S11, lines 17-18 in the abstract, and lines 455-458 in the conclusion), so it is possible that no new analyses need to be done, just reframing. Perhaps you could create a hypothesis in the last introduction paragraph (Third,...) about different management scenarios affecting things like drought duration and deficit in different geologies, and then put all these findings into a 3rd results section with a header clearly beyond evaluations (like “3.3. Management scenarios and drought impacts”). This could help show the theoretical advancements in real catchments to potential users of the model.

We greatly appreciate the suggestions provided by the reviewer, in previous versions of the manuscript we had contemplated a similar setup although we moved away from this idea given that the influence of drought management scenarios is already investigated using the Global Sensitivity Analysis (GSA). The aim of the GSA is to investigate 1) changes in parameter sensitivity of discharge and groundwater storage over time, 2) the overall sensitivity of discharge and groundwater to management scenarios and 3) implications for droughts when applying management scenarios (Lines 169-171).

We have clarified this in the abstract (Lines 9-11 and 14-16) to emphasise the focus of the GSA on these three aspects and highlight primary results of the response-based analysis.

A generic application of the drought management strategies to the case studies seemed therefore double and not the right fit, as water managers in the case study areas are likely to apply various combinations of the modelled drought management scenarios at the same time, whereas (for testing purposes) we applied these scenarios one at the time. Plus, the existing significant pieces of work, the GSA and case study application, make a lengthy paper and therefore we aimed to reduce any duplications.

However, we do appreciate the suggestion and acknowledge that the current result section of the GSA covers the first and second analysis in much detail but it did not elaborate much on the effect of drought management scenarios on surface water and

groundwater much. Instead of creating the suggested additional header at the end of the result section, we have integrated these suggestions within section 3.1 adding to the third component of the GSA (L304-318). We elaborated in more detail on the consequences of the identified leverage in Figure 5, because two quite different aspects of leverage of the integrated scenarios are evident (particularly the conjunctive use scenario), which resulted in quite different consequences for groundwater drought characteristics in the three groundwater modules (now in S9-S11 in Supplementary material). The revised paragraph now links the changed drought characteristics to different geologies and management strategies (see also the first point of Reviewer 2).

L304-318: The leverage of the modelled drought management scenarios is also reflected in the groundwater drought characteristics with drought deficit, i.e. the intensity of drought events being most sensitive (Figure 6). This intensification of groundwater drought events is largely due to an overall lower groundwater level in the karstic and porous module (Figure 5) following from the conjunctive use scenario. Detailed results highlight the strong negative difference between the baseline and the integrated management scenario (S9-S10 in Supplementary Materials). The overall influence on drought duration is positive, meaning shorter droughts, for most scenarios in the porous and fractured module. In the karstic module (and to some extent in the porous module) a larger spread of drought durations is found, which also reflects the increased sensitivity to groundwater demand under drought conditions for this specific aquifer type (Figure 6). Again, the largest differences are found for the conjunctive use scenarios compared to the trigger-driven scenarios (Figure S9-S11 in Supplementary Materials).

The distinct difference in leverage in fractured module between the trigger-driven and integrated drought management scenarios is mostly reflected in the drought frequency, as drought intensity and duration follow the same pattern -but less strongly negative/positive compared to the other groundwater modules. The smaller differences compared to baseline might be due to the overall higher groundwater storage levels in the integrated scenarios (Figure 5). The increased water supply and reduced water demand scenarios increase/decrease drought frequency respectively. Integrated scenarios result in a large range in drought frequency with maintaining hands off flow scenarios slightly reducing overall frequency (Figure S11 in Supplementary Material).

Minor Comments

Title: Consider updating the title to clarify the more novel scientific findings. Evaluation of a model in itself does not suggest the scientific advancement of the work to me, and I think it has the potential to be more generally impactful. Perhaps something like “Socio-hydrological model reveals how water resource

management affects drought duration and deficit in real groundwater-rich catchments”

We appreciate the bolder title suggestion emphasising some of the findings of the global sensitivity analysis, but given that the evaluation is central to the manuscript and presentation of SHOWER, we prefer to keep these aspects in the title.

Page 2, lines 29-34. Consider condensing this to basically say highly managed groundwater systems are present around the globe, rather than elaborating on the individual regions. Though it’s fine with me if you prefer to keep it as-is.

We kept the examples as is.

Page 20, lines 360-370. I’m not sure what the scientific value of showing that the SHOWER model had similar or better discharge performance to other models is, particularly if those models were simulated using different approaches for factors like input data, time periods, observations, or calibrations. I suggest condensing this.

We have rephrased second paragraph (L385:394) in the discussion to emphasise how representing groundwater improves modelling results. It is important to inform readers of similar modelling performances, but also to highlight the additional potential of testing management strategies with SHOWER.

L385-394: The SHOWER performances reported in this paper are comparable or exceeding results of other hydrological models for the same catchments. The largest improvements (measured in KGE) are found for the surface water-dominated G2G model (Hannaford et al., 2023b) where SHOWER improves the negative (Pang) and average (Derwent) KGE values. Other rainfall-runoff models, such as GR4J and GR6J yield similar results for their selected best runs in the Pang and Derwent catchments (Hannaford et al., 2023a), although the authors indicate that groundwater-dominated catchments are problematic to model well. The recent coupled DECIPHeR-GW v1 model results are similar to those of SHOWER (Zheng et al., 2025), showing that adding the elaborate groundwater representation in this new DECIPHeR version improves model results for these areas (Coxon et al., 2019; Lane et al., 2021). Even for similar modelling performances, a key advantage of SHOWER is that it explicitly accounts for groundwater and surface water abstractions and reservoir influence, which introduces the possibility of testing the impact of management strategies, which is not possible using the previously mentioned models.

Page 22, lines 424-425. I suggest rewording as something more direct, like “However, this simplification creates the opportunity for modelers with insufficient time and computing resources for more expensive models to be able to calibrate the model and explore results in detail.”

We appreciate the suggested rephrasing and have revised this into the following:

L452-454: However, this simplification creates the opportunity to explore results droughts and management impact in more detail prior to investing in expensive detailed groundwater models.

Page 23, lines 455-458. I really like this conclusions statement. It shows the scientific implications and exciting capabilities of the work well.

I wish the authors the best with this manuscript and their future endeavors.

We thank the reviewer for their constructive comments.

Reviewer 2:

The authors present the evaluation of a modelling tool for drought management that incorporates groundwater as well as abstractions and water management. The model evaluation is quite thorough and this is described in much detail. I think the manuscript is suitable for HESS and will provide a valuable contribution after some minor revisions.

We thank the reviewer for their positive views on the work and the constructive suggestions.

My main issue is with the fact that the main aims and contribution are not very clear. The manuscript reads more like a report on a modelling tool and the main contribution to the existing literature on socio-hydrology/groundwater modelling, is not highlighted very well. What is the main contribution, is it the integration of groundwater or the integration of management scenarios? The results and discussion focus on general model performance, and there is not much focus on the management scenarios and how they influence water availability and drought characteristics. Especially given the title of the manuscript and the emphasis on the model as a tool for groundwater managers and to “evaluate the impact of groundwater abstraction strategies on hydrological droughts”, I would expect more discussion on how well the management actions and socio-hydrological feedbacks are captured and some results on how the different management strategies influence droughts.

We appreciate the comment that the manuscript is quite model-heavy at the moment. This is likely a consequence of us attempting to not lengthen the manuscript unnecessarily. In the revised document, we have rephrased section 2.1 that introduces the management strategies (lines 130-154).

L130-154: Water management impact is modelled using four separate drought management scenarios that are compared to a baseline scenario, which influence all fluxes in the black box in Figure 1. In the baseline scenario there are no water management interventions and surface and groundwater water demand are simply abstracted from the reservoir and groundwater storage. The other four drought management scenarios were defined in Wendt et al. (2021) and represent common drought management practices in the UK. The first is to increase groundwater supply, using more of (underused or old) licenced groundwater boreholes and the natural storage buffer that aquifers provide. The second is to reduce water demand, which often starts early with a media campaign to stimulate lesser water use by the public. Severe measures can, however, restrict water use for commercial or non-essential public water use. These threshold-based scenarios (following drought triggers) depend on the severity of a meteorological and/or hydrological drought. Measures are often introduced

gradually and their severity increases depending on thresholds for precipitation, discharge, reservoir or groundwater storage levels that are related to historical drought events (for details see Wendt et al. (2021)). The next two scenarios apply regardless of a defined drought and are integrating surface water and groundwater use. The third scenario manages surface water and groundwater in conjunction depending on which resource has a higher availability at a time. For example, in areas with large groundwater storage, more groundwater is used compared to surface water and vice versa for areas with shallow groundwater storage. In practice, this requires high flexibility in management operations. The last scenario aims to preserve ecological minimum flows in rivers by reducing surface water abstractions. Water is taken from groundwater instead. We have modelled the threshold-based scenarios using average thresholds for precipitation, discharge, reservoir and/or groundwater levels, following Wendt et al. (2021). For the first scenario, we increased groundwater use whereas in the second scenario both surface water and groundwater are reduced equally. The third scenario integrates water storage and takes water from the highest store (either groundwater or reservoir storage) to meet water demand. This represents a non-restrictive application of conjunctive use practices. The last scenario maintains a threshold (Qbeco) for the ecological minimum flow from baseflow (plus the release flow from an upstream reservoir, if present) and groundwater demand is ceased when this threshold is reached. In the case of water demand exceeding reservoir and groundwater storage, water can be complemented by imported water as either a fixed share or conditionally on (reservoir or groundwater) water levels (Qimp and GWimp).

Additionally, we have linked the impact of the drought management strategies in Figure 5 to the consequence for drought characteristics (Figure 6) in the revised section 3.1 (see also comment 1 of Reviewer 1). This revised section discusses in more detail how the management actions translate to drought characteristics for the three groundwater modules.

Lastly, we have added to the discussion in section 4.2 (L428-433) to further elaborate on the influence of management scenarios.

L428-433: The notable influence of these integrated water management strategies on drought characteristics (Figure 6) is encouraging, but further research regarding the effectiveness of strategies is needed. Simulations indicate shorter drought durations by applying integrated management strategies with a mix of both intensified and relieved drought intensity particularly in the karstic and porous groundwater module. Additionally, the impact of demand measures that are widely applied in England, are most effective in a fast-responding shallow (fractured) groundwater module but less so in other groundwater modules, which emphasises the wider need for research.

Linked to this, I would like to see a bit more detail on how the management strategies are implemented in the model. Also, it would be good to see the socio-hydrological processes represented in more detail in figure 1. In its current form. It looks like the model is an extensive hydrological model but not a socio-hydrological model, with management just included as scenarios, rather than feedbacks between the hydrology and decision making.

I think this is just a matter of changing the focus of the results and rewriting the introduction and discussion so that it better highlights what the main contribution of this manuscript is.

We agree that more detail was needed in section 2.1 (methods) to introduce the drought management strategies and later in section 3.1 how these scenarios are reflected in the results. We believe that the model is an extensive hydrological model that can be used to understand management feedbacks in the hydrological cycle. For example, when maintaining low flows, groundwater use is restricted only when flows fall below a certain threshold. This conditional use of groundwater storage results in higher flows and less groundwater use, with a larger dependency on stored surface water or imported surface water. These management strategies are indeed defined based on common drought management strategies in Wendt et al. (2021). We have addressed this in the revised section 2.1 L130-154 (see comment 1 above).

The manuscript could do with a grammar check and some clearer language. Some examples:

- **Line 71-74, it is not entirely clear to me what will be compared with what.**

These sentences are introducing the global sensitivity analysis that goes beyond a one-to-one comparison with historical observations. The sentences are now revised:

L78-81: We need an alternative approach that evaluates the models' ability to reproduce historical observations in a managed environment and examines the model's consistency in input–output response with our understanding of each catchment (i.e. the perceptual model) (Wagener et al., 2022)

- **Line 129: “These scenarios are modelled the following.”**

We revised this into L146: We have modelled ..

- **Line 131, what is meant by “integrates water demand between the stores”?**

We have revised this sentence to emphasise that waters storage is integrated and water demand can be met by either surface water or groundwater.

L148: The third scenario integrates water storage and takes water from the highest store (either groundwater or reservoir storage) to meet water demand.

- **Line 391: “ Simulated discharge and groundwater results show a significant influence on simulated discharge and groundwater”.**

We have revised this into:

L415-416: Discharge and groundwater results are significantly altered due to the management scenarios, particularly during drought periods.

- **Line 435: \leq should be \geq ?**

Correct, this is now changed into ' \geq ' meaning larger than 1000km²(L463)

- **Line 447: “We also identified (un)influential parameters in the GSA that aided calibration when applying SHOWER to three case study areas in the UK.” Do you mean the parameters aided calibration or that the fact that you identified these parameters has aided the calibration? How has this helped the calibration? This wasn't clear to me from the methods or results section.**

The GSA shows which parameters are most influential and that supports calibration of the model, as model results will be mostly affected if these values are changed. Uninfluential parameters point towards (coded) independencies, which confirm the selected model structure and the GSA can be used to identify whether the code works correctly testing the consistency of the model (L192:194 in revised manuscript).

These are just a few examples, the whole manuscript would benefit from a thorough grammar check, because there were many sentences that are difficult to understand.

We thank the reviewer for their careful reading and ensured to revise the spelling and grammar in the revised manuscript (see also the track changes document).

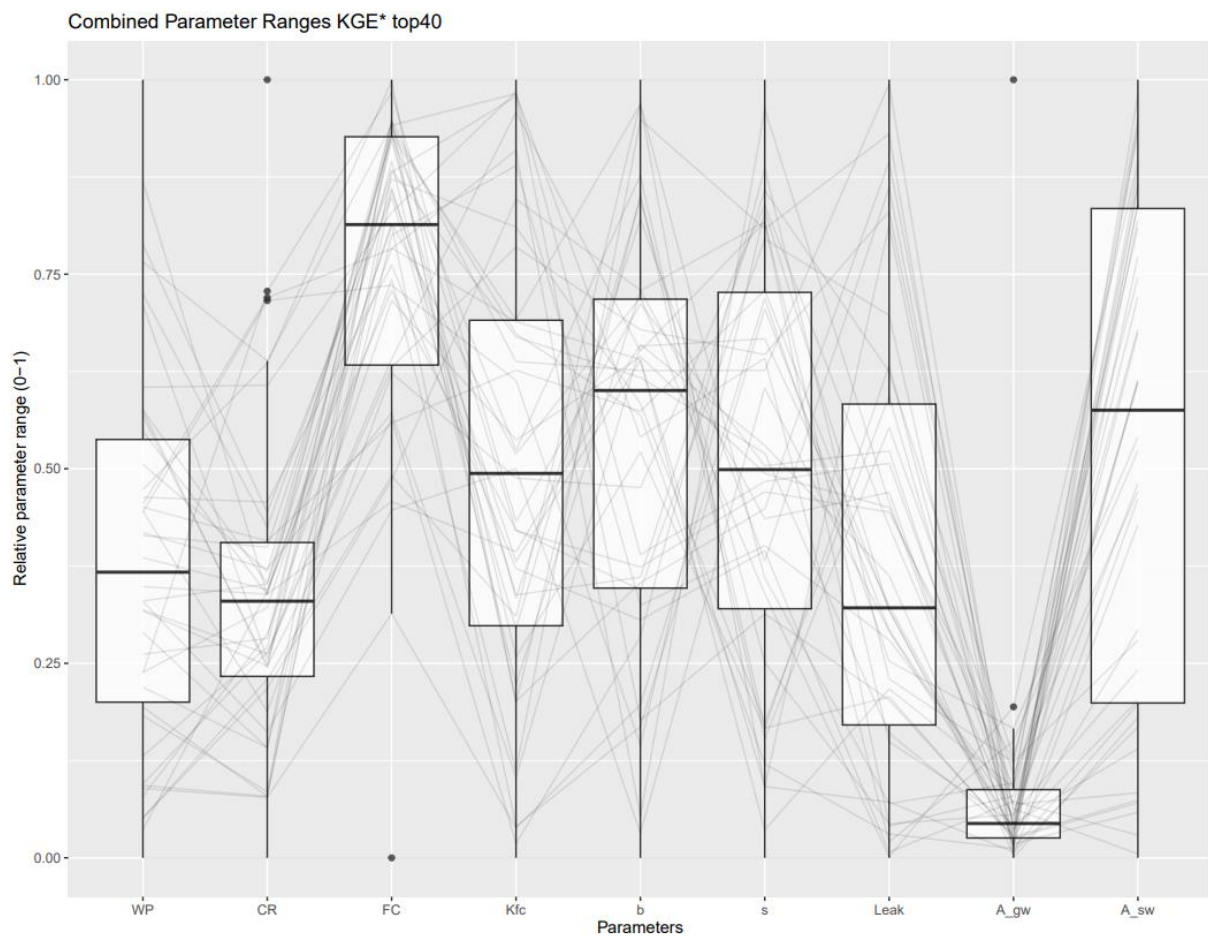
In Lines 229-230 the authors talk about a step-change around 50-60 simulations, however the figures in the supplement that are referred to seem to indicate this is around 40-50 simulations. In addition, it is not clear to me what is meant by 40-50 simulations. Are these just the 40 best performing parameter combinations? In that case, what does this step change mean? Is it just a strange artefact, or is it related to some change in the parameter values? The relevance of this step change is not clear to me.

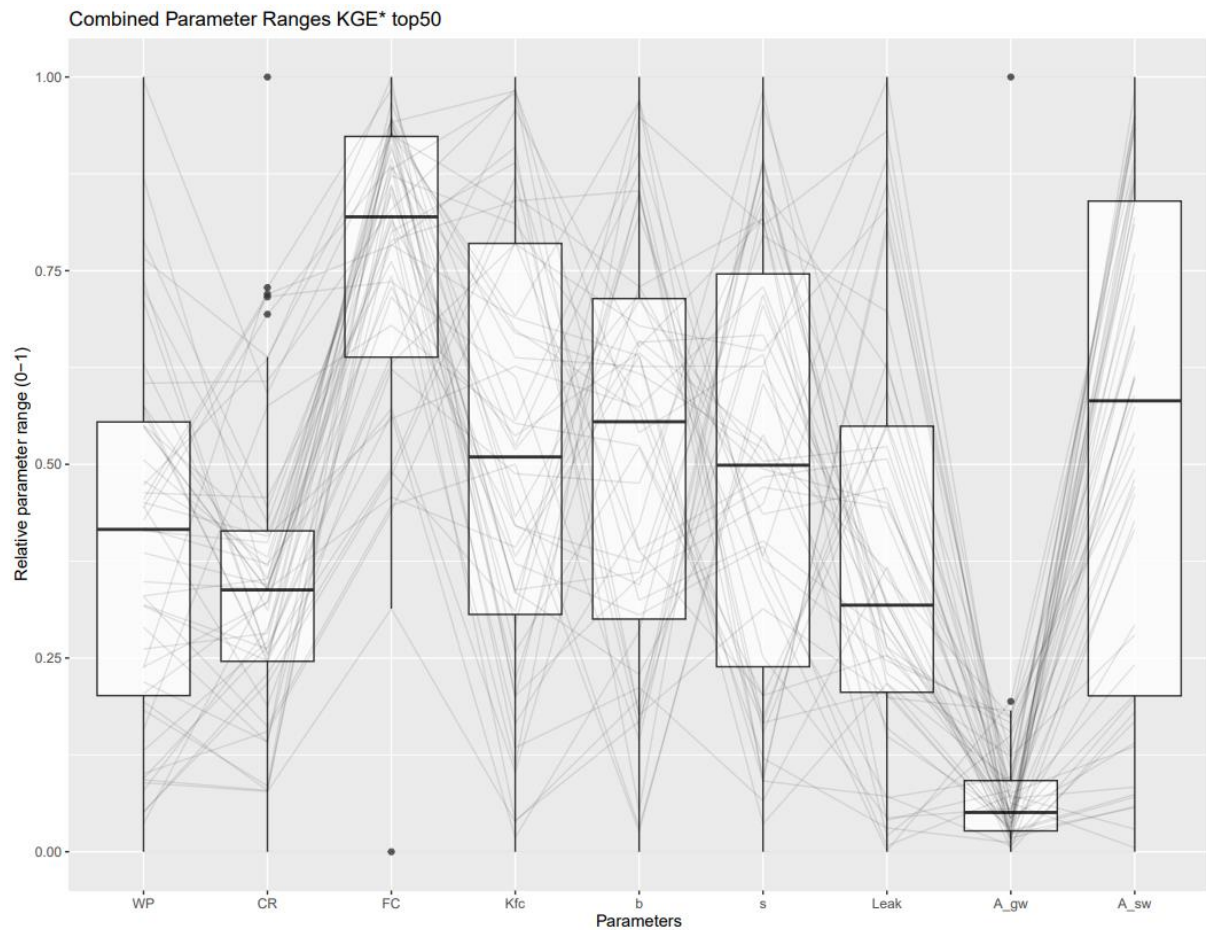
This is a good point raised by the Reviewer that we have investigated whilst doing the analyses -but should have been clearer in the manuscript where a typo adds to the confusion. When doing 10k simulations, finding the 'best' simulations can be achieved in a multitude of ways, which partly depends on the used evaluation criteria. However, defining the exact number of 'best' simulations is arbitrary, as this could be the top 20,

top 40, top 50 or top 100. When identifying a subtle step-change in the performance metric, we decided that it would be most sensible to use the top 50 to show the best simulation results. The identified change in performance is between 40-50 simulations of $0.05 NSE_{log}$ is in fact very small (particularly for the Trent catchment) and not an artifact of parameter values. Upon revisiting this paragraph again, we rephrased the wording to inform readers of the identified slight change in performance that resulted in showing the top 50, rather than the top 5, 20, or 100 simulations.

L245-248: The number of top performing simulations can be defined in multiple ways (e.g. top 100, or top 50 or top 20). We found a slight change in the improvement rate of NSE_{log} around the top 50, particularly in the Chalk simulation (see Figure S2 in Supplementary materials) and therefore settled on using the top 50 simulations for the model evaluation.

We have double-checked the supplementary figures and spread of the parameters in a parallel coordinate plot for the Chalk and Trent (Trent results shown below), but neither indicates a jump / oddity in parameter range. Hence we are more confident this slight change in performance is not due to parameter values.





The numbers in the legend for the relative sensitivity in Figure 3 are not readable.

We have changed the legend to ensure readability in the revised manuscript. Thanks for letting us know.

The data availability statement mentions that the model code is available from DEWendt/SHOWER, but I am unable to find this.

My apologies, as this repository is set to be public and therefore accessible when found. I have added the full link to the data availability section to avoid confusion ([GitHub - DEWendt/SHOWER: Operational Drought modelling tool for based on a socio-hydrological water resource modelling](#)).

References:

Rameshwaran, P., Bell, V. A., Brown, M. J., Davies, H. N., Kay, A. L., Rudd, A. C., & Sefton, C. (2022) Use of abstraction and discharge data to improve the performance of a national-scale hydrological model. *Water Resources Research*, 58, e2021WR029787 <https://doi.org/10.1029/2021WR029787>

Salwey, S., Coxon, G., Pianosi, F., Lane, R., Hutton, C., Bliss Singer, M., McMillan, H., and Freer, J. (2024) Developing water supply reservoir operating rules for large-scale hydrological modelling, *Hydrology and Earth Systems Science*, 28, 4203–4218 <https://doi.org/10.5194/hess-28-4203-2024>

Stoelzle, M., Weiler, M., Stahl, K., Morhard, A., and Schuetz, T. (2015) Is there a superior conceptual groundwater model structure for baseflow simulation? *Hydrological Processes*, 29, 1301–1313 <https://doi.org/10.1002/hyp.10251>

Wendt, D. E., Van Loon, A. F., Bloomfield, J. P., and Hannah, D. M. (2020) Asymmetric impact of groundwater use on groundwater droughts, *Hydrology and Earth System Sciences*, 24, 4853–4868 <https://doi.org/10.5194/hess-24-4853-2020>

Wendt, D. E., Bloomfield, J. P., Van Loon, A. F., Garcia, M., Heudorfer, B., Larsen, J., and Hannah, D. M. (2021) Evaluating integrated water management strategies to inform hydrological drought mitigation, *Natural Hazards and Earth System Sciences*, 21, 3113–3139 <https://doi.org/10.5194/nhess-21-3113-2021>