Reply to Reviewers' Comments

Key:

> Reviewer comment.

Response.

Reply to Reviewer #2

> General Comments:
>
> This work uses the SCEUA automatic calibration platform to calibrate surface water flows predicted over a number of watersheds in the UK. The SHETRAN distributed model is used as the simulation platform, run daily and compared to daily streamflow values. While I appreciate the goal of arriving at the right answers for the right reasons, the analysis to establish the physical basis for the calibrated values is insufficient to establish this. Furthermore, the approach used to conceptualize each model is very similar if not identical to the watershed models used to contrast the work done here. I think more work is needed along with revisions to the manuscript to establish these points. Comments are below.

We gratefully thank the Reviewer for the evaluation of our work and also for the constructive comments, the corrections and suggestions provided will certainly help to improve the manuscript to make it ready for publication. These are discussed below.

> * what are the limitations and biases in the shallow subsurface? The authors state that a single aquifer unit 20m in thickness was used; this is only shallow surficial, unconfined groundwater. The fixed 20m thickness has an influence on the transmissivity estimates. The authors should explore this sensivity with some number of representative watersheds to ensure this does not impart bias.

We thank the Reviewer for this valuable comment and suggestion. We agree that the fixed 20 m aquifer thickness represents a simplification of the shallow subsurface and may influence the transmissivity estimates. In response, we will explore this sensitivity further by running a set of representative catchments with an increased aquifer thickness of 50 m. We will assess the impact on model performance (e.g., NSE) as well as on the resulting transmissivity values. The outcomes of this sensitivity analysis will be included in the revised manuscript to evaluate any potential bias introduced by the fixed aquifer thickness.

> *The authors state that some watersheds were modeled with 1km columns while others were modeled with 5km columns. This is a substantial and seemingly arbitrary adjustment in resolution. What are the limitations of the grid resolution switch for the 16 catchments run at 5km? Did the authors conduct a sensitivity study on resolution to establish these final

values? Given that topography will be smoothed out substantially at 5km, a sensitivity study to ensure that results might be transferrable between these different resolutions is needed.

We thank the Reviewer for this comment. Sixteen catchments were run at a 5 km grid resolution because they were too large to be simulated feasibly at 1 km within the autocalibration framework. For example, the 54057 catchment (Severn at Haw Bridge) has a catchment area of 9895 km^2 and a single simulation at 1km resolution took approximately 2 days to complete. Consequently, using a coarser 5 km grid was the only practical option for including these catchments in the national-scale analysis.

We acknowledge that moving to 5 km resolution smooths topography and subsurface structure and may affect hydrological processes, and we agree that this introduces limitations. While a full resolution-sensitivity analysis across all large catchments was not feasible due to computational constraints, we will add discussion to the revised manuscript outlining these limitations explicitly and explaining how grid resolution may influence the transferability and comparability of results between the 1 km and 5 km simulations. We will also include a summary of simulation times to clarify the computational constraints underpinning this choice.

*It is well known the groundwater does not follow surface topographic divides and aquifer systems connect watersheds laterally. It appears that each of these watersheds is modeled independently which does not allow for groundwater import and export between watersheds along with regional groundwater flow. Can the authors show that lateral groundwater flow does not impact their results?

We thank the Reviewer for raising this important point. We recognise that groundwater systems do not always follow surface topographic divides and that lateral groundwater flow between neighbouring catchments can occur. In this study, each of the 698 catchments was modelled independently; however, the catchment boundaries used are those defined by the National River Flow Archive (NRFA), which has undertaken detailed assessments to align surface-water and groundwater boundaries wherever possible. According to the NRFA documentation, only 7 of the 698 catchments have known mismatches between surface and groundwater divides.

We agree that lateral groundwater exchanges may occur in some locations, but given the national scope and the independence of the model runs, it is not possible to explicitly simulate cross-catchment groundwater flow or quantify its effect within this framework. To address this limitation transparently, we will include in the appendix a list of the catchments where the NRFA identifies discrepancies between surface and groundwater boundaries and will discuss the implications of these limitations in the revised manuscript.

*The authors visually compare the transmissivity values produced at the end of the calibration process with BGS aquifers (Fig 11 and 12). Visually, the maps appear to bear no similarity. The other continental scale efforts the authors mention (e.g. Naz et al, Yang et al) start with geologic maps and aquifers to parameterize their model, then adjust parameter values accordingly. 1. Why was this more common approach not taken? 2. What quantitative steps can be taken in the manuscript to show any agreement between the geologically derived transmissivity values and the ones arrived at from this current study?

We thank the Reviewer for this detailed and constructive comment.

Regarding Figures 11 and 12, we believe that the transmissivity comparison in Figure 11 provides significant value to the paper and discussion of the parameters produced. Figure 12, however, presents deep-soil conductivity, not transmissivity, and we agree that both figures can be improved to make the comparisons clearer. We will therefore remake and revise these figures in the updated manuscript to enhance their interpretability and ensure that the analysis is presented more transparently.

1. Why the more common geology-based parameterisation approach was not taken

The Reviewer is correct that many continental-scale modelling studies (e.g., Naz et al., Yang et al.) initialise groundwater parameters directly from geological or aquifer maps before calibration. This *was* the approach taken in Uncalibrated-SHETRAN-UK, which uses hydraulic conductivity values from the *Aquifer Properties Manual* (Allen et al., 1997; MacDonald and Allen, 2001, Lewis, 2016). Autocalibrated-SHETRAN-UK, in contrast, departs from this by calibrating subsurface parameters directly, allowing the model to adjust parameters based on hydrological behaviour rather than relying solely on the mapped geology.

We will make this distinction more explicit in the revised manuscript to avoid confusion.

2. Quantitative steps to show agreement with geologically derived transmissivity

Figure 11 compares transmissivity estimates from point well observations (Allen et al., 1997) with the transmissivity values obtained from this study. We acknowledge that this was not sufficiently clear, and we will improve the figure and text to ensure that the comparison is transparent and easily interpretable.

In addition, to strengthen this part of the analysis we will incorporate:

- scatter-plot comparisons between calibrated transmissivity and observational estimates,
- NSE as a colour dimension, and
- Groundwater level observation comparisons

Together, these revisions will provide a more robust and quantitative assessment of parameter plausibility.

> *It would appear that a major component of the physics based approach would be to generate water table depth values and ET estimates along with streamflow. These values could help characterize change in storage and ET fluxes, which along with streamflow discharge would close the water budget for each watershed potentially limiting the space of equifinality. Given the unconstrained nature of the calibration exercise and simple model configuration, it appears that many solutions might provide the same streamflow estimates. A very thick soil layer could be produced that mimics the aquifer or AE/PE values could change the water budget. These could easily take on unrealistic bounds, detracting from the central theme of the work. The authors should justify the use of just streamflow in calibration. More discussion is needed, perhaps also with sensitivity cases, to ensure that parameter values are not unrealistic.

We thank the Reviewer for this comment. We used streamflow-only calibration to remain consistent with other national-scale studies (e.g., Lane et al., 2019; Lees et al., 2021), none of which incorporate groundwater or ET observations in their calibration procedures. We will clarify this in the revised manuscript.

We agree that relying solely on discharge increases the potential for equifinality. To limit this, only five subsurface parameters were calibrated, each with strict, limiting and realistic bounds (see Appendix C). In addition, the development of the parameters is not entirely unconstrained during the autocalibration as the mass balance is conserved. In the revised manuscript, we will expand the discussion on these limitations and include additional analyses (e.g., comparisons with groundwater levels and transmissivity estimates) to help demonstrate that the resulting parameter values remain plausible.

> *There are no surface water parameters, such as channel width or mannings roughness coefficients included in the analysis or sensitivity. Is this because channel routing was somehow not included in the study or for some other reason?

We thank the Reviewer for this comment and question. Non-calibrated parameters relating to surface water flow, such as the Strickler Coefficent for both overland flow and channel routing are presented in Appendix C. Channel routing is included as the model using diffusive approximation of the St Venant equations (Ewen et al., 2000).

References

Ewen, J., Parkin, G. and O'Connell, P.E. (2000). SHETRAN: Distributed River Basin Flow and Transport Modelling System. ASCE J. Hydrologic Eng., 5, 250-258.