This manuscript presents an interesting study on root zone storage capacities. However, I have some rather fundamental concerns, namely, using ERA5 data and neglecting parameter uncertainties. Both are straightforward to address, but this would require significant work for new model runs. Below, I discuss these two concerns and list a few minor concerns:

We thank Anonymous Referee # 2 for supporting the relevance of our study. Please find our detailed responses to your concerns highlighted in blue.

Major comments

Use of ERA5 data: there might be good reasons to use ERA5 data. These data are, for instance, a suitable option for global studies. However, one has to be aware of the substantial uncertainties of these data and for the US, the use of 'real' (station-based) observations is preferable, especiallyy for precipitation data (for a detailed analysis see https://egusphere.copernicus.org/preprints/2024/egusphere-2024-864/). From the text, I am not fully sure whether ERA5 data is used for both approaches or only the second one (MCT). Still, both cases would be problematic and I would recommend using station-based data, especially for precipitation, in both cases.

Reply: We thank the referee for their careful review and constructive feedback. We acknowledge the uncertainties associated with ERA5 reanalysis data and their suggestion to consider station observation data instead. In our study, we adopted two distinct approaches to estimate root zone storage capacity. The first method employed a model calibration technique using station observed precipitation and CAMELS streamflow data. The second method utilized MCT with ERA5 data, a leading reanalysis dataset known for its comprehensive spatial and temporal coverage, thereby providing an independent validation of our results.

Since the precipitation, evaporation, and snowmelt data in ERA-5 are all in the same spatial and temporal resolution, if we use ERA-5 evaporation and Caravan precipitation as the Anonymous Referee #2 suggested, it brings in spatial mismatch, which could result in even larger uncertainty. Therefore, we propose to continue utilizing ERA5 for both evaporation and precipitation data throughout our analysis.

While details on how the model has been calibrated are missing, my understanding from reading the manuscript is, that there was one single parameter set for each catchment. This would not be state-of-the-art but one should consider an ensemble of good/acceptable parameterisations. Please explain how parameter uncertainty was considered. If it was neglected, as I am afraid of, I would recommend to use some approach that allows for considering multiple parameter sets.

Reply: We will provide a clearer description of the parameter calibration process and the parameter uncertainties in the manuscript. Specifically, in our study, each catchment underwent extensive parameter calibration where we generated 40,000 different parameter combinations randomly. These combinations were evaluated using the Kling-Gupta Efficiency (KGE) coefficient, and the top 1% performing combinations (equating to 400 parameter sets) were selected. We then averaged these top sets to derive the final parameter set for each of the 497 study catchments. This approach comprehensively addresses parameter uncertainty across our analysis.

Minor comments:

L122: please provide details on how the model was calibrated

Reply: Section 2.2.2 outlines the model calibration process in detail. In the revision, we will enrich this section with additional specifics and clarity regarding the calibration methodology.

L153: this sounds like the soil routine in the HBV model. If yes, a reference to Bergström et al. would be suitable.

Reply: The root zone module constitutes the central component of the FLEX model. In our discussion manuscript, we acknowledged the HBV soil routine and referenced pertinent literature (Lindström et al., 1997; Seibert and Bergström, 2022). However, for this study, we opted to utilize the beta function from the Xinanjiang model (Zhao, 1992) instead of the HBV parameterization.

L169ff: with the focus on soil water storage, I wonder whether it is appropriate to use KGE as the objective function, which is known to focus on performances during high-flow events.

Reply: Firstly, the efficiency of objective functions falls outside the scope of this study. Previous research has extensively discussed this issue (Fenicia et al., 2007; Gupta et al., 2009). In our study, we employed the Kling-Gupta Efficiency (KGE) as our objective function, which is commonly used in hydrological modeling evaluations. KGE assesses the performance across all range of flows, including high-flow, moderate-flow, and low-flow conditions.

L182: Sorry, but I am not sure I fully understand how parameter sets are combined within one cluster. Despite the grouping by Jehn, parameter sets must vary largely within each cluster. If anything, aggregating simulated time series might be more appropriate than averaging parameter values.

Reply: Thank you for bringing this to our attention. To clarify, we did not combine parameter sets within a cluster. Initially, we calibrated parameters individually for each catchment. Subsequently, we computed the average parameter set for all catchments within a cluster. We agree with the reviewer that despite clustering, there are substantial variations in precipitation-runoff dynamics among different catchments. We will ensure to articulate this more clearly in the revised manuscript.

Equations: please avoid multi-letter variable names, these are mathematically incorrect (SM means S times M)

Reply: We will modify these variable names in the revision.

Figures: I find the figures hard to read. Partly this might be due to the small size. In Figure 5 the different axes use a different scale, which is unfortunate.

Reply: In the revised version, we will redraw these figures. The varying scales used in Figure 5 are designed to effectively illustrate the diverse range of root zone storage capacities.

References

Zhao, Ren-Jun. (1992). The Xinanjiang model applied in China. Journal of Hydrology, 135(1–4), 371–381. https://doi.org/10.1016/0022-1694(92)90096-E

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. Journal of Hydrology, 201(1–4), 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

Fenicia, F., H. H. G. Savenije, P. Matgen, and L. Pfister (2007), A comparison of alternative multiobjective calibration strategies for hydrological modeling, Water Resour. Res., 43, W03434, doi:10.1029/2006WR005098.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology, 377(1–2), 80–91.

Seibert, J., & Bergström, S. (2022). A retrospective on hydrological catchment modelling based on half a century with the HBV model. Hydrology and Earth System Sciences, 26(5), 1371–1388. https://doi.org/10.5194/hess-26-1371-2022