Reply to Reviewer #1 Comments

We have organized the reviewer's comments into categories and provided corresponding responses. Author responses are in **bold** and *italic.*

Review comment: **DECIPHeR-GW v1: A coupled hydrological model with improved representation of surface-groundwater interactions** by Yanchen Zheng et al.

This manuscript by Yanchen Zheng et al. presents a new coupled hydrological model called DECIPHeR-GW v1, which has a specific focus on enhanced representation of surface-groundwater interactions. The model couples two previously published models: an HRU-based hydrological model (DECIPHeR, Coxon et al., 2019) and a 2D groundwater model (Rahman et al. 2023). The coupling results in feedbacks between receiving recharge, simulating surface-groundwater interactions and returning groundwater levels and – discharges, of which the latter is then again incorporated in the river routing of the HRU base model. These interactions are all based on three interaction scenarios: groundwater head below bottom of the root zone, groundwater head is within the root zone, and groundwater head is higher than the topography. The aim of this study was to develop a coupled version, that is computational efficient even at large scales and able to represent the surface-groundwater interactions with high skill.

The calibration and validation was done on 669 catchments and 1804 groundwater wells. While the calibration was solely focused on streamflow data as the objective, the groundwater observations were used to evaluate the internal dynamics of the coupled model. The coupled model improved the simulation results in groundwater-dominated catchments, however strongly human influenced catchments remain challenging. Overall, the coupled model seems to produce robust streamflow simulations thanks also due to the incorporation of the temporal dynamics of groundwater levels and outperforms the original DECIPHeR model in catchments with minor human influence.

The manuscript is well written and easy to follow, the additional extensive supplement provides the reader with even more information, where of interest.

Thank you for your comments. We greatly appreciate your thoughtful and positive feedback on our paper.

The following points, remarks and questions are mostly raised for further clarifications, no major comments:

Line 165 capital S for section 2.2 (or check to keep consistency)

Figure 3, description, capital S for section 4.2 (or check to keep consistency)

Line 243 Citation does not need to be in brackets I believe

Line 244 capital S for section 3.3 (or check to keep consistency)

Line 324 Citation does not need to be in brackets I believe

Thank you for identifying these typographical errors in our paper. In the revision, we will review the text to ensure that every mention of 'Section' starts with a capital 'S'. We will also ensure proper citation references are used throughout.

Line 185 what buffer zone was defined for the demonstrated model?

The buffer zone is defined as the groundwater simulation grids that lie outside the catchment boundary. A buffer zone is necessary as the groundwater model adopts a no-flow boundary condition and therefore if the groundwater grid boundary is too close to the catchment boundary, the potential accumulation of groundwater flow at the boundary might affect the simulation results. We define the size of the buffer zone used in our analysis in Section 3.2. In the revision, we will ensure we better clarify the buffer zone and its size.

Line 236 50m gridded elevation map mentioned to define HRUs, does this differ from the original DECIPHeR model? Or where there in general specific changes (besides the parameters listed in 3.4) done for this version of DECIPHeR(-GW) presented here compared to the original DECIPHeR model?

The 50m grid elevation data used for delineating HRUs is the same as the elevation data used in the previous DECIPHeR model (Coxon et al., 2019; Lane et al., 2021). However, the precipitation and evapotranspiration data used in this analysis are different to those used in earlier papers using the DECIPHeR model and we use a different HRU setup with higher-resolution gridded data (2.2 km) compared to the 5 km grid used in previous papers (Coxon et al., 2019; Lane et al., 2021).

We will clarify this in the revision.

Line 300-302 could there be a potential pitfall doing the calibration like that?

We agree that this could be a pitfall in some cases. For instance, even within the same Chalk aquifer, hydrogeologic properties like transmissivity (T) and specific yield (Sy) may vary.

However, in hydrological model simulations, we typically assume that units or catchments with the similar soil texture and geological conditions are expected to exhibit comparable hydrological rainfall-runoff response processes. This is a way of conceptualizing the catchment underlying surface and subsurface conditions. It forms the theoretical foundation and significance behind using Hydrological Response Units (HRUs) and the Model Parameter Regionalization (MPR) method (Lane et al., 2021; Samaniego et al., 2010) in hydrological modelling.

Therefore, in the absence of detailed data and a clear relationship between soil texture, geological conditions, and model parameters, we believe that using the same model parameters for the same soil texture and geological conditions is the best calibration approach we can adopt and provides a reference for calibrating ungauged catchments.

Line 325 what is the benchmark model?

We explain the concept of the benchmark model runs on Lines 305–310 of Section 3.5. The benchmark model runs use the DECIPHeR model introduced by Lane et al. (2021), which employs the Multiscale Parameter Regionalization (MPR) method to parameterize the model's parameters. As noted above, the benchmark model runs are calibrated and evaluated using the same method as the coupled model. We will further clarify this point in the revised version.

Line 325 was the national calibration done on top of the catchment calibration, or both separate and the parameter values saved for the specific use of the model (e.g. national vs catchment runs)?

The catchment-by-catchment and national-consistent simulations are two separate model parameter calibration methods. In this study, we ran the coupled and benchmark model 5,000 times for each catchment with different parameter sets. Then, we sequentially applied the two calibration methods to determine which set of parameters provided the best simulation. Each

calibration method is optimized according to different application needs, and the parameter values are saved separately for use in their respective applications. We will clarify this in the revised paper.

Line 357 any educated guess what are the driving factors are in the model for the overestimated streamflow locations? Or how they could be changed to include for example the waste water discharges mentioned (or other human influences)?

We believe that the overestimation of streamflow in these catchments in central and southeastern England is primarily due to the fact that our current model does not account for surface and groundwater abstraction. In southeastern England, rainfall is relatively low and urbanization is high, thus the demand for water is significant. Including surface and groundwater abstractions would reduce the simulated streamflow and are an important next step for the hydrological modelling. Wastewater discharges are also an important component of the streamflow in urban regions (Coxon et al., 2024) but would increase the simulated model flows rather than decrease them. We will re-phrase these sentences to clarify this in the revisions and thank you for pointing this out.

Line 365 Would there also be an option to not use equal weights? E.g. including a sort of ratio weight for different catchment sizes included in the national calibration?

Thank you for your suggestions. It would certainly be an option to use different weights in the national calibration. However, in this paper, our primary focus is to compare whether the coupled model's performance improves after incorporating the groundwater module compared to the benchmark DECIPHeR model. Therefore, we used the consistent parameter calibration methods as previously applied in calibrating DECIPHeR (Lane et al., 2021; Salwey et al., 2024). According to our results, despite the reduced performance with the national-consistent calibration method, the coupled model still outperforms in approximately 50% groundwater-dominated catchments compared to the benchmark model.

We believe that exploring different weighting approaches for parameter calibration in national-scale simulations would be a valuable research topic. In our future research, we will consider optimizing different weighting schemes to make national parameter calibration more reasonable. We will include this in the revised paper.

Line 405 is the KGE of 0.85 referring to the model that after the national calibration or the catchment only? And how would they differ (also in relation to the benchmark model)?

This KGE of 0.85 refers to the performance of the benchmark model using the catchment-only calibration method for catchment 39001. We will clarify this in the revised paper. For catchment 39001, the KGE value under the national-consistent calibration method in the benchmark model is 0.61. To provide more detailed KGE comparison between coupled and benchmark model as well as two calibration methods, we will include an additional column in Table 3 to present the benchmark model's KGE performance under the national-consistent calibration method in the revised paper.

Line 410 could there be structural components that could be added that represent the human influences? (maybe more for a future study)

Thank you for your comments. Yes, we are currently working on developing this model to account for the impact of human influences. In the current paper, our focus is primarily on the coupling of surface water and groundwater models, specifically evaluating the model's performance when only natural groundwater processes are considered. Our next step is to integrate human-water interactions into the coupled model. This analysis will be developed further and presented in a separate publication.

References:

Coxon, G., Freer, J., Lane, R., Dunne, T., Knoben, W. J., Howden, N. J., Quinn, N., Wagener, T., and Woods, R.: DECIPHeR v1: dynamic fluxEs and connectivity for predictions of HydRology, Geoscientific Model Development, 12, 2285-2306, 2019.

Coxon, G., McMillan, H., Bloomfield, J. P., Bolotin, L., Dean, J. F., Kelleher, C., Slater, L., and Zheng, Y.: Wastewater discharges and urban land cover dominate urban hydrology signals across England and Wales, Environmental Research Letters, 19, 084016, 2024.

Lane, R. A., Freer, J. E., Coxon, G., and Wagener, T.: Incorporating uncertainty into multiscale parameter regionalization to evaluate the performance of nationally consistent parameter fields for a hydrological model, Water Resources Research, 57, e2020WR028393, 2021.

Salwey, S., Coxon, G., Pianosi, F., Lane, R., Hutton, C., Bliss Singer, M., McMillan, H., and Freer, J.: Developing water supply reservoir operating rules for large-scale hydrological modelling, Hydrol. Earth Syst. Sci., 28, 4203-4218, 10.5194/hess-28-4203-2024, 2024.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, https://doi.org/10.1029/2008WR007327, 2010.