Response to Reviewer #1 Comments

We have provided the point-by-point responses to the comments. Author responses are in **bold** and *italic.* The changes to the manuscript are highlighted in yellow. The line numbers mentioned in this document refer to the clean version of revised manuscript.

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:

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

Corrected. We have modified the 'section' with the capital S.

Line 190: 'As presented in the Equation (2) of Section 2.2,…'

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

Corrected. We have modified the 'section' in the caption of this figure with the capital S.

Figure 4: 'Details in Section 4.2 and Table 4,…'

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

Corrected. We modified the citation format of the references here and remove the brackets.

Line 278: 'In this study, we implemented the same HRUs discretization approach described in Salwey et al. (2024), which uses three equal classes of slope and accumulated area, catchment boundaries as well as a 2.2-km input grid.'

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

Corrected. We have modified the 'section' with the capital S. We also have carefully checked throughout the manuscript and also corrected in Line 379, Line 520,

Line 280: 'This is consistent with national climate projection data, detailed in Section 3.3…'

Line 379: 'described in Section 3.3'

Line 520: 'More details are discussed in Section 5.2.'

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

Corrected. We modified the citation format of the references and remove the brackets. We also have carefully checked about this throughout the manuscript.

Line 367: 'Two calibration approaches, namely (a) catchment by catchment and (b) nationally-consistent calibration, were used to calibrate the coupled model following the study from Lane et al. (2021).'

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

In our study, the groundwater grids need to extend beyond the catchment boundary to ensure accurate simulations, as our model uses a no-flow boundary condition that prevents groundwater flow from crossing the simulation boundary. If the groundwater grid boundary is too close to the catchment boundary, groundwater may accumulate at the edges of catchment boundary, potentially affecting results. To address this, we determined a larger groundwater gridded simulation area, extending beyond the catchment boundary in all directions. This extended area, referred to as the buffer zone, provides a buffer between the groundwater grid's simulation domain and the catchment boundary.

Explanation of the buffer zone has been added in Section 2.3 in Lines 210-214.

'In our study, we assumed that no water can move and leave the groundwater system across the boundary, since no-flow lateral boundary condition is adopted in the groundwater model. To reduce the effects of this no-flow boundary condition and allow for inter-catchment groundwater exchange, the groundwater simulation domain is extended beyond the catchment boundary in all directions (Figure 3b). This expanded groundwater gridded simulation area is referred to as the buffer zone in our study (light blue grids in Figure 3b and 3c). Absence of the buffer zone could lead to the potential buildup of water in the adjacent cells of the lateral boundaries due to the adoption of the no-flow boundary condition.'

We also add Figure 3 in the revision to better demonstrate this.

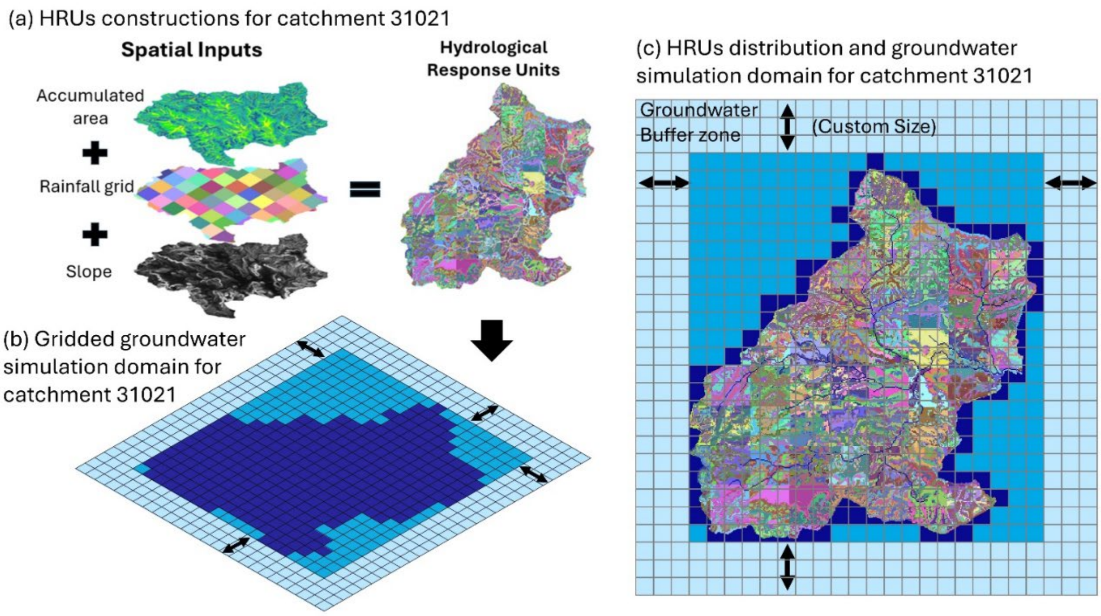

Figure 3: The DECIPHeR-GW coupling and spatial interaction from DECIPHeR Hydrologic Response Units (HRUs) to groundwater model grid cells for one example catchment Welland at Ashley 31021. (a) the HRUs constructions process for catchment 31021; (b) the gridded groundwater simulation domain for catchment 31021. (c) DECIPHeR-GW coupling and spatial interaction between HRUs and groundwater grids.

7) 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 50 m grid elevation data used for delineating HRUs is the same as the data utilized in previous DECIPHeR papers (Coxon et al., 2019; Lane et al., 2021).

The precipitation and evapotranspiration grid data used in this study differs from those in earlier studies. For this paper, we employed higher-resolution gridded climate data (2.2 km) compared to the 5 km grid used in the previous papers (Coxon et al., 2019; Lane et al., 2021). The use of different rainfall input grids will result a different delineation of HRUs in the model. Therefore, for fair comparison, we re-run the DECIPHeR model with 2.2 km gridded precipitation and evapotranspiration data as the benchmark runs to compare with the simulation results from DECIPHeR-GW model.

We clarified this in Lines 272-274.

'For the surface water component, a 50 m gridded digital elevation model (Intermaptechnologies, 2009) (also used in Coxon et al. (2019); Lane et al. (2021)) was adopted as the basis for the Digital Terrain Analysis to build the river network and define the HRUs across all England and Wales catchments.

Line 277-281: 'In this study, we implemented the same HRU discretization approach described in Salwey et al. (2024), which uses three equal classes of slope and accumulated area, catchment boundaries as well as a 2.2-km input grid. This is consistent with national climate projection data, detailed in Section 3.3 and higher resolution input data compared to other previous studies using DECIPHeR (Coxon et al., 2019; Lane et al., 2021).'

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

This calibration approach, which uses the same set of parameters for identical soil and lithology types, may have limitations. As noted in lines 608-621 of the discussion, hydrogeologic properties like transmissivity (T) and specific yield (Sy) can vary even within the same Chalk aquifer. In fact, the Chalk is known for its inter-catchment variations and variability across different sections of the aquifer (Allen et al. 1997). However, in hydrological modelling, it is generally assumed that catchments with similar soil textures and geological conditions will exhibit comparable rainfall-runoff responses. This concept underpins the use of Hydrological Response Units (HRUs) and the Model Parameter Regionalization (MPR) method (Samaniego et al., 2010).

In the absence of detailed consistent data and clear correlations between soil texture, geology, and model parameters, we believe this approach is the best available calibration method. It provides a useful reference for calibrating ungauged catchments.

In lines 608-621 of the discussion, we have discussed this point and how it could be further improved in future.

Line 608-621: 'Parameterizing surface-groundwater coupled models across large scales and diverse geological types remains challenging due to the difficulty in accurately representing geological heterogeneity (Gleeson et al., 2021; Condon et al., 2021). In our study, groundwater level simulations are highly dependent on hydrogeological parameters (i.e., T and Sy; see sensitivity analysis in Figure S12). Although we have attempted to capture the complexity of geological conditions by using different parameter ranges across 5000 simulations for a total of 101 lithology types, parameters for the same lithology type can only be assigned the same set of values for one simulation. In reality, parameters such as T can vary significantly even within the Chalk aquifer (Allen et al. 1997). A recent study presented a three-dimensional geological digital representation model of Great Britain using extensive geological maps and borehole data (Bianchi et al., 2024). They developed a national-scale groundwater model of Great Britain (BGWM) using this detailed geological data to consider the heterogeneity characteristics of aquifers, demonstrating its capability to accurately simulate groundwater dynamics. Griffiths et al. (2023) developed a method to estimate the initialized groundwater model parameter set using national-scale hydrogeological datasets to improve the parameterization of New Zealand's national groundwater model. Adopting more accurate and detailed geological data and advanced sampling methods to parametrize the model could be another direction of further improving the model performance (Hellwig et al., 2020; Henriksen et al., 2003; Westerhoff et al., 2018).'

9) Line 325 what is the benchmark model?

We clarify the concept of the benchmark model in Lines 347-350 of Section 3.5. We use the DECIPHeR model introduced by Lane et al. (2021) as the benchmark model, which employs the Multiscale Parameter Regionalization (MPR) method to parameterize the model's parameters.

As noted in our previous responses, since the DECIPHeR-GW model uses a different resolution (2.2 km) for the input rainfall and evapotranspiration grid data, we re-ran the DECIPHeR model from Lane et al (2021) with the same 2.2 km gridded data to ensure consistency with the simulations from the DECIPHeR-GW model. These simulations using the 2.2 km gridded data are referred to as the benchmark model runs.

Lines 347-350: 'In this study, we set up the simulations for 669 catchments using the DECIPHeR model introduced by Lane et al. (2021) as the benchmark model for comparison with the DECIPHeR-GW model. The DECIPHeR model in Lane et al (2021) employs the Multiscale Parameter Regionalization (MPR) method to parameterize model parameters while maintaining the original DECIPHeR model structure (Coxon et al., 2019) without groundwater representation.'

10) 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 model 5,000 times for each catchment with different parameter sets. We then applied both two calibration methods separately to identify the corresponding best-performing parameters. Each method is optimized for its specific application, and the parameter values are saved separately for use in their respective contexts. We have added clarifications about this in lines 367-368:

'Two calibration approaches, namely (a) catchment by catchment and (b) nationally-consistent calibration, were used to calibrate the coupled model following the study from Lane et al. (2021). These two calibration methods are applied separately to identify the corresponding best-performing parameters, with the parameter values saved for their respective applications.'

11) 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 think the overestimation of streamflow in central and southeastern England is due to our coupled model not accounting for human water interactions, especially surface and groundwater abstraction. These regions, with low rainfall, dry climates, and high urbanization, heavily rely on surface water and groundwater, leading to significant abstraction. However, our model only considers natural streamflow and groundwater movement, ignoring water consumption, resulting in overestimated streamflow.

While wastewater discharge in urban areas affects river flow (Coxon et al., 2024), incorporating it will increase streamflow. We emphasize the importance of accounting for all human activities, including abstraction and wastewater discharge, in the coupled model for more accurate streamflow simulations. We clarify this point in the revisions and checked throughout the paper.

Lines 399-401: 'However, the coupled model still tends to overestimate streamflow in some catchments in central and southeast England, which could be due to human activities such as surface water and groundwater abstractions (Salwey et al., 2023; Wendt et al., 2021; Bloomfield et al., 2021).'

Clarifications have also been made in lines 448-453:

'In the Thames at Kingston River basin (catchment ID: 39001) where surface water and groundwater abstractions are prevalent, the coupled model tends to overestimate flows particularly during the dry periods (Figure 6f). Wastewater returns from sewage treatment works are also common in these regions and could influence streamflow (Coxon et al., 2024), potentially contributing to the decline in KGE performance. This decline in performance indicates the challenge of simulating flows in heavily human impacted catchments and underscores the need to enhance the representation of human-water interactions in the hydrological model.'

12) 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?

It is indeed possible to explore using different weighting schemes in the national calibration. However, the primary focus of this paper is to evaluate whether the performance of the coupled model improves after incorporating the groundwater module, compared to the benchmark DECIPHeR model. Therefore, we adopted the same parameter calibration methods previously used in DECIPHeR calibrations (Lane et al., 2021; Salwey et al., 2024).

Our results indicate that while the national-consistent calibration method shows reduced performance overall, the coupled model still outperforms the benchmark model in approximately 50% of groundwater-dominated catchments.

We agree that investigating alternative weighting approaches for parameter calibration in national-consistent method is a promising direction for future research. However, this is not the focus of our current paper. In this revision, we have added the clarification about this issue.

Lines 408 – 413: 'By assigning equal weights to all catchments, the model parameters for groundwater-dominated catchments might not be constrained properly under the national-consistent approach, leading to reduced performance in those areas. However, 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 (Figure 5f). Future work is suggested to explore alternative weighting approaches to enhance parameter calibration, instead of equal weighting.'

13) 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)?

We have revised to clarify the KGE of 0.85 we refer to here is the performance from the benchmark DECIPHeR model under the catchment-by-catchment (CBC) calibration method for catchment 39001. To make it easier for readers to compare the model performance, we have added an additional column in Table 3 to present the benchmark model's KGE performance under the national-consistent (NC) calibration method.

The KGE values for the NC method are lower compared to the CBC method, because the CBC method identifies the best set of parameters from all 5000 simulations, while the NC method selects a set of parameters that aims to optimize the simulation results for all catchments. This is explained in Section 4.1, lines 402-405.

Regardless of the calibration method (CBC or NC) used, our coupled model can produce better results in groundwater-dominated catchments with minor human influences, such as in catchment 39028. However, in the paper, we were trying to highlight that for this catchment 39001, a catchment with significant human activities, the DECIPHeR model performs better than the DECIPHeR-GW coupled model under the CBC method. Revisions and clarifications are as follows:

Lines 433–440: 'Especially in the groundwater-dominated chalk catchment (39028), characterized by small net loss from abstractions and discharges (minor human influences) and essentially a natural baseflow-dominated flow regime, the streamflow hydrograph simulations from the coupled model significantly improve and fit well compared to observations (Figure 5e), with the KGE metric increasing almost twofold compared to the benchmark under both catchment-by-catchment and national-consistent calibration methods (showed in Table 4). In addition, under catchment-by-catchment calibration method, the coupled model performed well for other aquifer types, as shown by the results from a limestone catchment 31021 (Figure 6d) and sandstone catchment 54044 (Figure 6c), with KGE values exceeding 0.80.'

Lines 453–455: 'Meanwhile, it's interesting to see that the benchmark model produces better simulation results for a catchment with significant human activities, such as the Thames River basin, with a KGE of 0.85 under catchment-by-catchment calibration method, despite not accounting for either groundwater or human-water interactions.'

14) 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. We are currently working on developing this model to incorporate the impacts of human influences. However, the current paper is focused primarily on the coupling of surface water and groundwater models, with an emphasis on evaluating the model's performance when only natural groundwater processes are considered. As outlined in Section 5.2 of the Discussion, our next step is to integrate human-water interactions into the coupled model. This expanded analysis will be developed further and presented in a separate publication.

Reference:

Bloomfield, J. P., Gong, M., Marchant, B. P., Coxon, G., and Addor, N.: How is Baseflow Index (BFI) impacted by water resource management practices?, Hydrol. Earth Syst. Sci., 25, 5355-5379, 10.5194/hess-25-5355-2021, 2021.

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.

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.

IntermapTechnologies: NEXTMap British Digital Terrain 50m resolution (DTMlO) Model Data by Intermap, NERC Earth Observation Data Centre [dataset], available at: http://catalogue.ceda.ac.uk/uuid/f5d41dbll70f41819497d15dd8052ad2 (last access: 3 June 2019), 2009.

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., Singer, M. B., and Hutton, C.: National-Scale Detection of Reservoir Impacts Through Hydrological Signatures, Water Resources Research, 59, e2022WR033893, https://doi.org/10.1029/2022WR033893, 2023.
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.
Wendt, D. E., Bloomfield, J. P., Van Loon, A. F., Garcia, M., Heudorfer, B., Larsen, J., and Hannah, D. M.: Evaluating integrated water management strategies to inform hydrological drought mitigation, Nat. Hazards Earth Syst. Sci., 21, 3113-3139, 10.5194/nhess-21-3113-2021, 2021.

Response to Reviewer #2 Comments

We have provided the point-by-point responses to the comments. Author responses are in **bold** and *italic.* The changes to the manuscript are highlighted in blue. The line numbers mentioned in this document refer to the clean version of revised manuscript.

Title: DECIPHeR-GW v1: A coupled hydrological model with improved representation of surface-groundwater interactions

Summary

This study presents a coupled hydrological modeling framework integrating the DECIPHeR land surface hydrological model with a 2D groundwater model. Applied to 669 watersheds across England and Wales, the coupled model (DECIPHeR-GW) demonstrates improved streamflow simulations, particularly in areas with strong groundwater-surface water interactions.

DECIPHeR-GW features an HRU-based structure that feeds a gridded groundwater model, allowing for dynamic water exchange based on water table and root zone elevations. The model incorporates six key parameters for stochastic calibration (including soil and aquifer properties) and enables simulations across large domains.

Evaluation against observations shows enhanced temporal variability and streamflow magnitude compared to the uncoupled model. While generally successful, the study acknowledges challenges in watersheds significantly impacted by human activities.

Positive Aspects

- The manuscript shows great clarity and organization, making it highly readable and accessible. The manuscript effectively guide the reader through the study's objectives, methodology, results, and discussion.
- The presentation of results is good, with a clear and concise narrative that effectively conveys the key findings. The authors interpret the results, providing insightful discussions on their implications and limitations.
- The figures and tables are well-designed and informative, effectively illustrating the key findings and supporting the conclusions.
- The supplementary information is also valuable and well-presented, providing important details and supporting data that helps in the overall understanding of the study.

Thank you very much for your comments. We are delighted to see your positive feedback on our manuscript.

General Comments:

1) The manuscript presents a coupled land-surface and groundwater model. While the importance of incorporating groundwater is recognized, a more focused research question is needed. The authors should clearly articulate how their approach differs from existing coupling methods, highlighting the novelty of their model. Additionally, a discussion on the positioning of their model within the spectrum of simplified to physics-based groundwater representations is needed.

It is important to clarify that the primary objective of this study is not to propose a novel coupling methodology but to develop and describe a new coupled model.

In our coupled model, the coupling method, including the mapping and transformation of variables between HRU-scale surface components and gridded groundwater systems, is adapted from existing approaches used in coupling SWAT and MODFLOW. The key novelty of our coupling method lies in the introduction of three dynamic scenarios to simulate the surface-groundwater interactions.

We highlight and clarify our objectives and novelty in Lines 88-92.

'This paper proposes a coupled hydrological model DECIPHeR-GW with a specific focus on enhancing the representation of surface-groundwater interactions whilst maintaining computational efficiency for national or large-scale modelling applications. This study presents the first attempt to couple the DECIPHeR HRU-scale model with a new 2D gridded groundwater model and expands the diversity of coupling approaches available for integrating HRU-scale surface models with grid-based groundwater models. The novelty of our coupled method lies in the introduction of three dynamic scenarios to simulate the surface-groundwater interactions. These scenarios adjust recharge fluxes based on root zone saturation and groundwater head positions. We discuss the rationale behind coupling DECIPHeR and the 2D gridded groundwater model in Section 2 and provide detailed descriptions of the coupled model structures.'

A discussion on the positioning of our model within the spectrum of simplified to physics-based groundwater representations has been included in Lines 545-553 in discussion section.

'Furthermore, our groundwater component provides groundwater simulation results that compare well to observations with high computational efficiency. Some relatively simplified models only produce groundwater storage (Yang et al., 2017; Guimberteau et al., 2014; Griffiths et al., 2023; Müller Schmied et al., 2014), while some models adopted a lumped groundwater model structure, failing to capture spatial variability of groundwater distribution (Yeh and Eltahir, 2005; Gascoin et al., 2009; Ejaz et al., 2022). Our model provides simulated groundwater level at grid scale, facilitating model validation against groundwater observations and producing the spatial groundwater distribution. Although our 2D groundwater model omits

vertical water movement, it is structurally simpler compared to more complex 3D models (Bailey et al., 2016; Ewen et al., 2000; Maxwell et al., 2015; Naz et al., 2022), making it better suited for large-scale simulations and allowing for multiple model calibrations.'

2) The manuscript emphasizes the model's scalability. However, a more detailed discussion on the potential scale mismatch between the regional land surface model and the large-scale groundwater model is required. Specifically, the authors should clarify what type of groundwater flow represented at the 1 km grid scale, considering the local-scale flows discharging into streams. The manuscript should address how these different scales are reconciled within the model.

In this revision, we have added more details about the size of HRUs across our study catchments in Lines 277-284. The average size of the generated HRUs across all study catchments is 0.31 km², which is the comparable scale with the 1 km groundwater grids. Therefore, scale mismatch is not a significant concern, as the scales are relatively similar. We apply the existing mapping methods from coupling SWAT and MODFLOW to transfer and convert the variables and fluxes across the HRU scale and grid scale.

Lines 277-284: 'In this study, we implemented the same HRU discretization approach described in Salwey et al. (2024), which uses three equal classes of slope and accumulated area, catchment boundaries as well as a 2.2-km input grid. This is consistent with national climate projection data, detailed in Section 3.3 and higher resolution input data compared to other previous studies using DECIPHeR (Coxon et al., 2019; Lane et al., 2021). The average size of the generated HRUs across all study catchments is 0.31 km², with HRU areas ranging from the largest 3.55 km² to the size of one DEM grid cell (0.0025 km²).

We constructed and operated the gridded groundwater model based on the topography data at 1 km spatial resolution, which is the comparable scale with the size of HRUs.'

We also have clarified this in the abstract (Lines 16-17).

'Our new coupled model was set up at 1 km spatial resolution for the groundwater model, and the average size of the surface water HRUs was 0.31 km².'

In addition, we added a figure to show the spatial distributions of HRUs and groundwater grids to provide more clear visualizations of the scales. Please see Figure 3 below in the response to comment 14.

3) The manuscript highlights computational efficiency as a key advantage of DECIPHeR-GW. A more thorough discussion on the model assumptions that contribute to this efficiency is needed. The authors should explicitly state which processes are simplified or neglected in both the land surface and groundwater components. A comparison of the model's assumptions to those of other computationally expensive models, particularly those that incorporate fine-

resolution environmental data and capture land surface heterogeneity, would provide valuable context.

In Lines 115-117, we included the reasoning behind why the surface component DECIPHeR is computationally efficient.

‘Moreover, with its river basin auto-build function, HRU-based grouping of similar landscapes, and simple model structure that excludes complex land surface fluxes, the DECIPHeR model can simulate multiple model runs for calibration and sensitivity analysis against observational data at national-scales.’

In Lines 122-124, we clarified why the groundwater model is computationally efficient and also the processes that are simplified and omitted.

‘The advantage of this model is its capability to simulate groundwater hydraulic heads, enabling groundwater resources assessment and management. This groundwater model omits river channel representation and simulates only groundwater flow movements between grids. Additionally, the model operates in two dimensions using 2D hydrogeological data and ignores vertical water movement. These prioritisations ensure the model is computationally efficient, facilitating multiple simulations for both calibration and evaluation against groundwater level observations or a model parameter sensitivity analysis, as presented in Rahman et al. (2023). This high computational efficiency is critical, as many existing large-scale coupled models are published in an uncalibrated state due to high computational costs (Maxwell et al., 2015; Reinecke et al., 2019; Naz et al., 2022; Verkaik et al., 2022).’

A comparison of the model's assumptions to those of other computationally expensive models has been added to lines 580-590 in discussion section.

‘Our groundwater component omits vertical water flow and river representation, requiring only two subsurface hydrogeological properties. Our model may encounter challenges in regions with significant vertical hydrogeological variability, requiring more additional test in future work for these regions to ensure accuracy. In contrast, some complex 3D groundwater models need to discretize aquifers vertically and include specialized modules for river simulation (Bailey et al., 2016; Ewen et al., 2000; Ng et al., 2018; Maxwell et al., 2015), demanding finer-resolution hydrogeological data to capture land surface heterogeneity and higher computational costs. Currently, lots of existing coupled surface-groundwater models either cannot perform or require excessive time for calibration due to high computational costs (Ng et al., 2018; Parkin et al., 2007; Naz et al., 2022; Reinecke et al., 2019), which limits the ability to assess uncertainty in presented results and hinder future model applications. The computational efficient feature of our proposed model allowed us to calibrate it against extensive observed data, including 669 streamflow gauges and 1804 groundwater wells, thereby providing reliable results for future application.’

Specific Comments

- **Abstract:**

4) **[19]** A more specific description of the catchment characteristics would help the understanding of the study area.

Revised. We have added more catchment characteristics description of our study catchments in Lines 19-21.

'The model provides streamflow simulation with a median KGE of 0.83 across varying hydro-climates, such as wetter catchments with a maximum mean annual rainfall of 3577 mm/year in the west and drier catchments with minimum 562 mm/year in the east of Great Britain, as well as diverse hydrogeological conditions including chalk, sandstone and limestone.'

5) **[20]** Please specify the variable being analyzed in this context.

Clarified. The variable being analyzed here is streamflow.

Lines 21-23: 'Higher KGE values are found particularly for the drier chalk catchments in southeast England, where the average KGE for streamflow increased from 0.49 in the benchmark DECIPHeR model to 0.7.'

6) **[23-24]** The abstract currently does not explicitly establish a strong connection between computational efficiency and large-domain simulations as a significant challenge in traditional land surface modeling. Consider incorporating some of the performance measurements described in the Discussion.

- Consider including a brief statement in the abstract regarding the spatial resolution of both the land surface and groundwater components of the coupled modeling framework to provide further context for the reader.

We included a brief statement of the spatial resolution of the surface and groundwater components in the coupled model and provided quantitative metrics to highlight its high computational efficiency in the abstract.

Lines 14-18: 'Depending on the storage capacity of the surface water model component and the position of the modelled groundwater level, three scenarios are developed to derive recharge and capture surface-groundwater interactions dynamically. Our coupled model was set up at 1 km spatial resolution for the groundwater model, and the average size of the surface water HRUs was 0.31 km². The coupled model was calibrated and evaluated against daily flow timeseries from 669 catchments and groundwater level data from 1804 wells across England and Wales.'

Lines 25-29: 'Simulating 51 years of daily data for the largest catchment, the Thames at Kingston River Basin (9948 km^2), takes approximately 17 hours on a standard CPU, facilitating multiple simulations for model calibration and sensitive analysis. Overall, this new DECIPHeR-GW model demonstrates enhanced accuracy and computational efficiency in reproducing streamflow and groundwater levels, making it a valuable tool for addressing water resources and management issues over large domains.'

- **Introduction:**

7) **[46-49]** It would be beneficial for the authors to explicitly articulate the specific aspects of groundwater representation in existing models that are challenging or form the basis of their research hypothesis.

We summarised this in Lines 50-55.

'Moreover, many hydrological models across regions and countries globally struggle to reproduce streamflow dynamics in groundwater-dominated catchments (Massmann, 2020; Coxon et al., 2019; Badjana et al., 2023; Mcmillan et al., 2016; Lane et al., 2019; Hartmann et al., 2014) due to either oversimplified groundwater processes (Yang et al., 2017; Guimberteau et al., 2014; Gascoin et al., 2009) or complex groundwater components that are challenging to calibrate at large scales (Maxwell et al., 2015; Ewen et al., 2000; Naz et al., 2022), leading to difficulties in predicting and managing water resources in these regions.'

More details about the challenges in existing model's groundwater representation are summarised later in the introduction in Lines 69-85.

8) **[80-86]** A dedicated section discussing the novelty of the proposed approach would add more to the relevance of the study. This section should clearly differentiate the current methodology from previously mentioned modeling approaches, highlighting the unique contributions and advancements of the presented work.

- The inclusion of a dedicated paragraph discussing the scales of the modeling framework is recommended. This paragraph should address potential scale mismatches between the land surface and groundwater components, and how these differences are addressed within the model.

We have emphasized the novelty of our work in the last section of introduction (Lines 88-92) by distinguishing our methodology from previously modelling approaches, highlighting the unique contributions and advancements of our work. See our revisions in response to comment 1. As for the second comments, please see response to the second general comments about the scalability.

- **The DECIPHeR-GW model**

9) **[93-101]** To enhance clarity, it would be beneficial to include a more detailed description of the HRU construction process. This would clarify how the domain is discretized and how this discretization may influence the representation of key hydrological processes.

When referring to "previous studies" in line 97, please specify whether this refers specifically to the DECIPHeR model or to land surface models in general.

In line 101, it would be helpful to elaborate on the specific requirements for large-scale simulations, as defined by the authors.

We have added more details about how the study area is discretized and delineated as well as the HRU construction process. We clarified that more HRUs can represent more detailed hydrological processes but can increase run times. The spatial resolution of HRUs is typically user-defined. Modified in Lines 103-112.

'DECIPHeR is a flexible modelling framework (Coxon et al., 2019), which has been implemented across various locations (Shannon et al., 2023; Dobson et al., 2020). The DECIPHeR model has an auto-build function in the digital terrain analysis (DTA) that defines river basin boundaries based on the downstream gauge. Each river basin is constructed and run independently. After the river basin has been delineated, hydrologically similar points with identical climatic inputs (e.g., rainfall, evapotranspiration) and landscape attributes (e.g., geology, land use, soil, slope) are grouped into hydrological response units (HRUs). Each HRU, as the main spatial element, is considered as an independent model store. All HRUs can have different spatial inputs and model parameter values to represent diverse and localized processes. The simplest setup uses one HRU per river basin, while the most complex uses one HRU per DEM grid cell. The spatial resolution of HRUs is typically user-defined, see the full description of DECIPHeR model structure and evaluation results for Great Britain in Coxon et al. (2019).'

The 'previous studies' refers to the DECIPHeR model. Clarified in Line 112.

'Previous studies on the DECIPHeR model have shown that model performance in groundwater-dominated regions can be inadequate, underscoring the need to enhance surface-groundwater interactions (Coxon et al., 2019; Lane et al., 2021).'

Our requirement for large-scale simulations is the ability to perform multiple model runs to meet the needs of model calibration and sensitivity analysis. Revised in Lines 115-117.

'Moreover, with its river basin auto-build function, HRU-based grouping of similar landscapes, and simple model structure that excludes complex land surface fluxes, the DECIPHeR model can simulate multiple model runs for calibration and sensitivity analysis against observational data at national-scales.'

10) **[106-113]** While other large-scale coupled models can be computationally expensive due to the inclusion of detailed processes (such as vertical water

movement), it is unclear how DECIPHeR-GW balances computational efficiency with process representation.

If computational demands and input data requirements are reduced, it is essential to clearly describe which hydrological processes are simplified or omitted, and how these simplifications are compensated for through the calibration process.

In this revision, we have explained why the surface component DECIPHeR is computationally efficient. We also clarified why the groundwater model is efficient and the processes that are simplified. Please see the response to the third general comments.

To better parameterize the model, we link groundwater model parameters, i.e., transmissivity and specific yield, with lithology conditions. For each lithology type, the transmissivity and specific yield will be sampled in a range for 5000 times for model calibration. We have showed a good performance for both groundwater and discharge.

In general, lateral flows are considered the prominent type of flow compared to vertical flows in many aquifers in the UK, particularly in the top of the aquifer (Bianchi et al., 2024). However, in some regions where aquifers have significant vertical hydrogeological variability, vertical flow can be more prevalent compared to lateral flows due to enlarged fissures (Allen et al., 1997). More detailed studies are needed to evaluate the accuracy of our model in these regions. Alternative approaches consist of increasing the grid-discretization and additional input data to improve representation (Bianchi et al., 2024). The detailed tests fall beyond the scope of this paper and will be a future work.

We have added the potential drawback of our model without considering vertical flow in lines 580-583 of discussion.

Line 580-583: 'Our groundwater component omits vertical water flow and river representation, requiring only two subsurface hydrogeological properties. Our model may encounter challenges in regions with significant vertical hydrogeological variability, requiring more additional test in future work for these regions to ensure accuracy.'

11) **[120]** Given that the variable Qex in Figure 1 may represent recharge rates exceeding infiltration capacity, it is important to discuss whether the model considers the potential for saturated overland flow.

Our model does not account for infiltration capacity; however, it does consider saturated overland flow, which occurs when the root zone reaches its maximum storage capacity. We have explained this in Lines 151-153.

'Once the root zone storage is full, excess rainfall is generated as saturated excess flow (Q_{EX}), which is considered as the saturated overland flow (Q_{OF}), and then added to the river channel for river routing. The coupled model does not consider infiltration capacity.'

12) **[147]** Does the hydrological model allow for two-way interactions between river routing and the HRUs, enabling water from the river to contribute to aquifer recharge?

Our coupled model is unable to simulate direct river water contributing to aquifer recharge. We have added the explanation and clarifications about this in Line 167-169.

'The groundwater discharge is passed back to the HRUs as the saturated flow (Q_{SAT}) and added to the nearest river channel for river routing. The surface component from DECIPHeR does not directly account for water flow from river to HRUs and the groundwater model lacks explicit river channel representation, thus the coupled model does not capture river water contribution to aquifer recharge. Instead, aquifer recharge is accounted for via the root zone (see also Figure 2).'

13) **[154]** Please provide further details on how the parameterization of the groundwater grid is connected to the characteristics of the overlying HRUs. Does each HRU has a set of soil parameters and those are weighted average to parametrize the groundwater grid?

The model parameters for the surface water and groundwater components are at different scales, and each is prepared independently without the need for conversion. For instance, the groundwater component parameters are at 1km grid scale, the parameterization of their values relies on lithology types, and soil texture information are not required. Soil texture parameters determine the maximum root zone storage (SRmax), saturated hydraulic conductivity (Ks), and pore size distribution index (B). These are defined at the HRU scale. More details can be found in Table 3. Clarified in Line 177-182.

'The overview of all model stores, fluxes, state variables and model parameters are summarized in the Table 1. There are six model parameters in the coupled model that can be sampled or set to default values. The model parameters for the surface water and groundwater components are at different scales, and each is prepared independently. The parameters SRmax, Ks, B and CHV, control the surface water model component (including recharge and river routing), are at HRU or catchment scale, which needs soil texture and land use information for determining their parameterization. Parameters T and Sy, which govern the groundwater flow simulation, are determined by a lithology map that matches the spatial resolution of the groundwater grids. Details of the river routing approach can be found in Coxon et al. (2019).'

14) **[187]** How are the HRUs over the buffering zone updated within the model if they belong to a different catchment?

In our current setup, HRUs are confined within catchment boundaries and do not extend into the buffer zone. The coupled model operates at the river basin scale, defining the groundwater simulation domain to include all groundwater grids covering the catchment and an additional buffer zone. This domain, particularly the buffer zone, extends beyond the catchment boundaries and excludes HRUs. We have clarified this in Lines 210-215.

'To reduce the effects of this no-flow boundary condition and allow for inter-catchment groundwater exchange, the groundwater simulation domain is extended beyond the catchment boundary in all directions (Figure 3b). This expanded groundwater gridded simulation area, referred to as the buffer zone in our study (light blue grids in Figure 3b and 3c), ensures adequate separation between the groundwater grid's simulation boundary and the catchment boundary. Because of adopting the no-flow boundary condition, absence of this kind of buffer zone could lead to the potential buildup of water in the adjacent cells of the lateral boundaries. The groundwater grids and buffer zones outside the catchment boundaries do not incorporate or consider HRUs, which are exclusively confined within the catchment boundaries.'

In addition, since several comments are related to this (comment 15, 17), we added one figure showing the spatial mapping of HRUs and groundwater simulation domains to better demonstrate this.

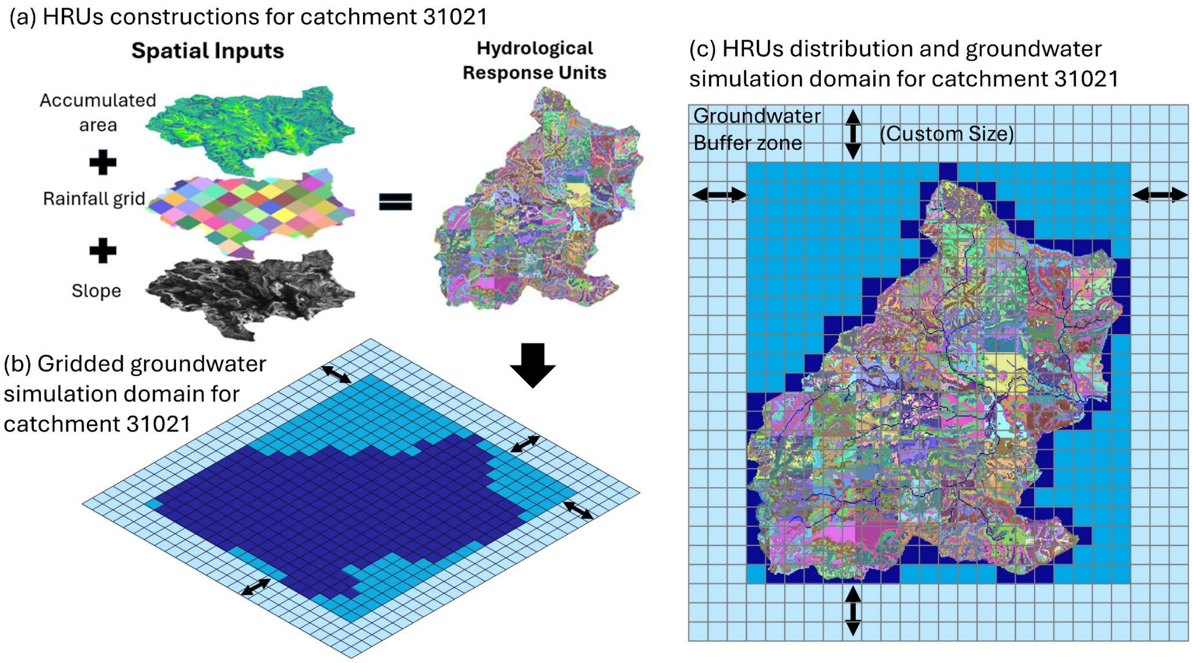

Figure 3: The DECIPHeR-GW coupling and spatial interaction from DECIPHeR Hydrologic Response Units (HRUs) to groundwater model grid cells for one example catchment Welland at Ashley 31021. (a) the HRUs constructions process for catchment 31021; (b) the gridded groundwater simulation domain for catchment 31021. (c) DECIPHeR-GW coupling and spatial interaction between HRUs and groundwater grids.

15) [191] How are the systems of watersheds connected, considering that each watershed has its own groundwater model?

Is the connection solely through river routing, or do the buffering zones of adjacent watersheds also interact or overlap?

Our coupled model currently runs at river basin scale. Similar to other widely-used hydrological models, such as SWAT model, the digital terrain analysis (DTA) of our surface water component DECIPHeR model delineates catchments using the downstream gauge and clips the groundwater grid for the simulation domain. Each river basin is configured individually, rather than modelling the entire continent or nation and does not account for the exchange of hydrological variables, such as groundwater flow, between neighbouring river basins. Yet, within each river basin, we do consider the exchange of hydrological variables, such as groundwater flow, across catchments. The buffer zones of adjacent river basins may overlap geographically, but they do not interact with each other. We clarified this in Lines 215-220.

'Users can customize the size of buffer zone according to the modelling objective. Details on how to determine the appropriate buffer zone size for our analysis are provided in Section 3.2. Note that the coupled model is currently designed to run each river basin individually, without accounting for the exchange of hydrological variables, such as groundwater flow, across river basins. Within each river basin, we do consider the exchange of hydrological variables across catchments. While buffer zones of adjacent river basins may overlap geographically, they remain hydrologically independent and do not interact.'

We have noted this in the discussion (Lines 622-631) as an limitation. This could be a potential area for improvement in future work.

'Since our coupled model retain the digital terrain analysis (DTA) configuration of the DECIPHeR model (Coxon et al., 2019; Lane et al., 2021) and currently operates at the river basin scale, each river basin is configured and run individually, rather than modelling the entire continent or nation. There is no consideration of hydrological variable exchanges, such as groundwater flow, across river basins. Additionally, this setup can result in inaccuracies for small, isolated catchments, as groundwater grids outside the boundaries lack HRUs distribution and do not receive rainfall or recharge. The fixed buffer zone makes up a relatively larger proportion in small catchments compared to larger ones, which may explain the model's poor performance in these small and isolated catchments. To address these issues, we recommend improving the DTA model setup in future research by configuring the model for the entire continent or region, simulating all HRUs and associated groundwater grids simultaneously at each time step. This will ensure accurate rainfall and groundwater recharge computations across the study area and better represent inter-catchment flow dynamics.'

- **Model implementation and evaluation across England and Wales**

16) **[202-205]** To enhance clarity and reproducibility, a detailed description of the meteorological, soil properties and elevation data used in the study would be beneficial. This should include information on the temporal and spatial resolution of the data, as well as a list of the variables used to parameterize the model.

We added a table in the Section 3.2 to provide more detailed information of all the meteorological, hydrological, soil properties and elevation data used in the study. Modifications have been made in Lines 289-292.

'A detailed description of all the topography, hydro-climate, land use, soil texture and hydrogeology variables, that are used for model configuration, inputs, parameterization and evaluation, are summarised in Table 2. The following Sections 3.3 and 3.4 introduce more details about the model input and evaluation datasets, and model parameterization.'

Table 2. Detailed descriptions of the topography, hydro-climate, land use, soil texture, and hydrogeology variables used for model configuration, inputs, parameterization, and evaluation in this study.

Category	Variables and dataset	Spatial resolution and coverage	Temporal resolution and coverage	Description	Sources and references
Topography	Digital elevation model (DEM)	50 m gridded	-	Inputs for Digital Terrain Analysis to generate the river network and define HRUs across study area	(Intermap Techologies, 2009)
Climate	Precipitation	2.2 km gridded	Daily timeseries, 1970-2020	Model inputs	(Hollis et al., 2019)
	Potential evapotranspiration (PET)	2.2 km gridded	Daily timeseries, 1970-2020	Model inputs	(Robinson et al., 2023)
Hydrology	Streamflow	669 river gauges	Daily timeseries, 1970-2020	Model evaluation	UK National River Flow Archive
	Groundwater level	3888 groundwater wells	Varied temporal resolution and coverage	Model evaluation	(Environment Agency, 2023)
Land use	Land use map	50 m gridded	-	Basemap for estimating the model parameter SRmax	Derived from reclassifying the UKCEH Land Cover Map (Lane et al., 2021; Rowland et al., 2017)
Soil texture	Sand, silt and clay percentage	50 m gridded	-	Basemap for deriving the root zone depth and soil texture classification and estimating the model parameter Ks and B	LandIS national soils map for England and Wales (Lane et al., 2021)
	Porosity	50 m gridded	-	Basemap for deriving the root zone depth and estimating the model parameter SRmax	Maps of porosity were sourced from (Lane et al., 2021)
Hydrogeology	Initial groundwater heads map	1 km gridded	-	Long-term steady-state simulated groundwater heads from Rahman et al. (2023) as the initial condition for the groundwater model	(Rahman et al., 2023)
	Digital geological map for lithology type	1:625000 map scale	-	Lithological classes described in this map used for estimating the Transmissivity (T) and Specific yield (Sy)	(British Geological Survey, 2010; Rahman et al., 2023)

17) **[244]** The manuscript would benefit from a more detailed description of the spatial resolution and configuration of the HRUs. The inclusion of a figure illustrating the HRU distribution within the study area would significantly help.

See response to previous comments 2, 9, 14 and new Figure 3.

- **Results**

18) **[358]** To further strengthen the analysis, it would be valuable to quantify the correlation between human activity and the model's performance. If these specific catchments are not being monitored, it is important to discuss the potential implications for the calibration process, particularly with regard to the representation of human-induced water abstractions.

We agree that this would be a valuable exercise but to do this properly is beyond the scope of this paper. Our group have recently published papers focusing on the impact of reservoirs (Salwey et al., 2023), wastewater discharges (Coxon et al., 2024) and groundwater abstractions (Wendt et al., 2021; Bloomfield et al., 2021). To date, these analyses have focuses on individual human influences in isolation and to provide a comprehensive analysis of all human influences in combination requires a separate more detailed study.

On reflection, we have removed Figure S8 from the Supplementary Information as it doesn't support the point we wished to make and revised this section of the results. We have added further detail to the discussion on the need for a separate analysis of modelling the impacts of human influences on river flow and when/where this is important. Also, discussion of potential implications for the calibration process have been added.

Revisions in lines 399-401 in the results section:

'However, the coupled model still tends to overestimate streamflow in some catchments in central and southeast England, which could be due to human activities such as surface water and groundwater abstractions (Salwey et al., 2023; Wendt et al., 2021; Bloomfield et al., 2021).'

Revisions in lines 636-649 in the discussion section:

'Given the absence of human influences in the current model version, calibration may lead the adoption of a parameterization that excessively reducing evapotranspiration or lowering groundwater levels through an overly high transmissivity to compensate these human influences, such as water abstractions. The dramatic rise in anthropogenic water use over the last century underscores the need to incorporate these human impacts into hydrological models (De Graaf et al., 2019; Döll et al., 2014; Wada et al., 2017), with significant impacts on river flow demonstrated for catchments across Great Britain from wastewater discharges (Coxon et al., 2024), reservoirs (Salwey et al., 2023) and groundwater abstractions (Wendt et al., 2021; Bloomfield et al., 2021). Many previous models lacked explicit modules for human impacts due to data limitations or relied instead on parameterizations or water use estimation statistics to mimic the human influences (Arheimer et al., 2020; Veldkamp et al., 2018; Sutanudjaja et al., 2018; Müller Schmied et al., 2014; Guillaumot et al., 2022). However, with the increasing availability of observed water abstraction and waste water returns data (Rameshwaran et al., 2022; Wu et al., 2023), it is crucial to integrate additional modules that accurately reflect these influences to ensure precise model parameterization and reliable simulation of internal catchment variables (Dang et al., 2020). In future developments, we aim to improve the overall accuracy and applicability of DECIPHeR-GW for both natural and human-dominated hydrological systems by refining the model to better capture the complexities of human-water interactions.'

19) Figure S8 of the supplementary information appears to be missing units for surface water abstractions, groundwater abstractions, and wastewater discharges.

We have deleted this figure in this revision, see detailed response to comment 18.

20) **[370]** The inclusion of a figure showing the temporal mean water table elevation for the study area would provide insights into the spatial and temporal variability. This would allow for a visual assessment of the water table's consistency with the expected topographic trends.

Since our model operates at the river basin scale, the groundwater grids do not cover the entire UK at the national-scale. In this revision, we added a figure showing the temporal mean of groundwater water table for Thames at Kingston catchment 39001, a groundwater-dominated catchment and one of the largest catchment in our study area. Figure 8 presents the temporal mean simulated groundwater table depth over 1980-2020 for catchment 39001 under the best catchment-by-catchment calibration method. The map shows our simulated groundwater table aligns consistently with topographic trends.

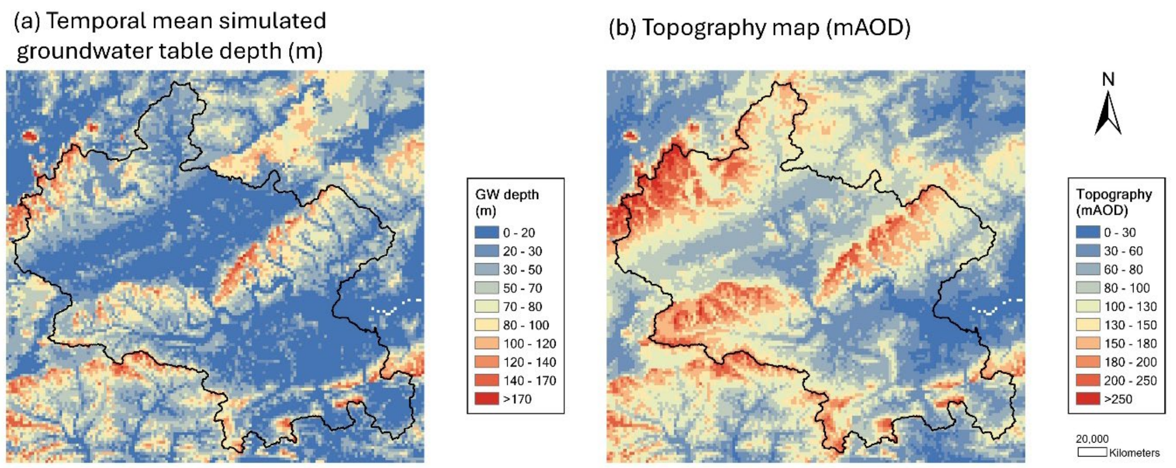

Figure 8: Spatial maps of simulated groundwater table depth for Thames at Kingston 39001. (a) Temporal mean over 1980-2020 of simulated groundwater table depth (difference between local topography and groundwater head in metres below land surface) in catchment 39001 under best catchment-by-catchment calibration method. (b) The topography map for catchment 39001.

This summary has been added to Section 4.3.

Lines 492-498: 'The results are highly consistent between the two streamflow calibration methods (Figure 7a and 7b), indicating the coupled model is robust in simulating the groundwater levels. The spatial distribution of temporal mean simulated groundwater table depth over 1980-2020 for Thames at Kingston catchment 39001, a groundwater-dominated and one of the largest *catchment in our study area, is presented in Figure 8, which is based on the best catchment-by-catchment calibration method. The simulated groundwater table depth aligns consistently with topographic trends, confirming that our coupled model also accurately reproduces the spatial variability of the groundwater table.'*

Reference:

Allen, D., Brewerton, L., Coleby, L., Gibbs, B., Lewis, M., MacDonald, A., Wagstaff, S., and Williams, A.: The physical properties of major aquifers in England and Wales, 1997.
Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C. M., Hasan, A., and Pineda, L.: Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation, Hydrol. Earth Syst. Sci., 24, 535-559, 10.5194/hess-24-535-2020, 2020.
Bianchi, M., Scheidegger, J., Hughes, A., Jackson, C., Lee, J., Lewis, M., Mansour, M., Newell, A., O'Dochartaigh, B., Patton, A., and Dadson, S.: Simulation of national-scale groundwater dynamics in geologically complex aquifer systems: an example from Great Britain, Hydrological Sciences Journal, 69, 572-591, 10.1080/02626667.2024.2320847, 2024.
Bloomfield, J. P., Gong, M., Marchant, B. P., Coxon, G., and Addor, N.: How is Baseflow Index (BFI) impacted by water resource management practices?, Hydrol. Earth Syst. Sci., 25, 5355-5379, 10.5194/hess-25-5355-2021, 2021.
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.
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.
Dang, T. D., Chowdhury, A. F. M. K., and Galelli, S.: On the representation of water reservoir storage and operations in large-scale hydrological models: implications on model parameterization and climate change impact assessments, Hydrol. Earth Syst. Sci., 24, 397-416, 10.5194/hess-24-397-2020, 2020.
de Graaf, I. E., Gleeson, T., Van Beek, L., Sutanudjaja, E. H., and Bierkens, M. F.: Environmental flow limits to global groundwater pumping, Nature, 574, 90-94, 2019.
Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T., and Eicker, A.: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites, Water Resources Research, 50, 5698-5720, https://doi.org/10.1002/2014WR015595, 2014.
Ejaz, F., Wöhling, T., Höge, M., and Nowak, W.: Lumped geohydrological modelling for long-term predictions of groundwater storage and depletion, Journal of Hydrology, 606, 127347, 2022.
Ewen, J., Parkin, G., and O'Connell, P. E.: SHETRAN: distributed river basin flow and transport modeling system, Journal of hydrologic engineering, 5, 250-258, 2000.
Gascoin, S., Ducharne, A., Ribstein, P., Carli, M., and Habets, F.: Adaptation of a catchment-based land surface model to the hydrogeological setting of the Somme River basin (France), Journal of Hydrology, 368, 105-116, 2009.
Griffiths, J., Yang, J., Woods, R., Zammit, C., Porhemmat, R., Shankar, U., Rajanayaka, C., Ren, J., and Howden, N.: Parameterization of a National Groundwater Model for New Zealand, Sustainability, 15, 13280, 2023.
Guillaumot, L., Smilovic, M., Burek, P., De Bruijn, J., Greve, P., Kahil, T., and Wada, Y.: Coupling a large-scale hydrological model (CWatM v1. 1) with a high-resolution groundwater flow model

(MODFLOW 6) to assess the impact of irrigation at regional scale, Geoscientific Model Development, 15, 2022.
Guimberteau, M., Ducharne, A., Ciais, P., Boisier, J. P., Peng, S., De Weirdt, M., and Verbeeck, H.: Testing conceptual and physically based soil hydrology schemes against observations for the Amazon Basin, Geosci. Model Dev., 7, 1115-1136, 10.5194/gmd-7-1115-2014, 2014.
Hollis, D., McCarthy, M., Kendon, M., Legg, T., and Simpson, I.: HadUK-Grid—A new UK dataset of gridded climate observations, Geoscience Data Journal, 6, 151-159, 2019.
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.
Maxwell, R., Condon, L., and Kollet, S.: A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3, Geoscientific model development, 8, 923-937, 2015.
Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511-3538, 10.5194/hess-18-3511-2014, 2014.
Naz, B. S., Sharples, W., Ma, Y., Goergen, K., and Kollet, S.: Continental-scale evaluation of a fully distributed coupled land surface and groundwater model ParFlow-CLM (v3. 6.0) over Europe, Geoscientific Model Development Discussions, 2022, 1-29, 2022.
Rahman, M., Pianosi, F., and Woods, R.: Simulating spatial variability of groundwater table in England and Wales, Hydrological Processes, 37, e14849, 2023.
Rameshwaran, P., Bell, V. A., Brown, M. J., Davies, H. N., Kay, A. L., Rudd, A. C., and Sefton, C.: 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, 2022.
Robinson, E. L., Brown, M. J., Kay, A. L., Lane, R. A., Chapman, R., Bell, V. A., and Blyth, E. M.: Hydro-PE: gridded datasets of historical and future Penman-Monteith potential evaporation for the United Kingdom, Earth System Science Data, 15, 4433-4461, 2023.
Rowland, C. S., Morton, R. D., Carrasco, L., McShane, G., Neil, A. W., and Wood, C. M.: Land Cover Map 2015 (25m raster, GB), NERC Environmental Information Data Centre [dataset], 2017.
Salwey, S., Coxon, G., Pianosi, F., Singer, M. B., and Hutton, C.: National-Scale Detection of Reservoir Impacts Through Hydrological Signatures, Water Resources Research, 59, e2022WR033893, https://doi.org/10.1029/2022WR033893, 2023.
Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geosci. Model Dev., 11, 2429-2453, 10.5194/gmd-11-2429-2018, 2018.
Veldkamp, T. I. E., Zhao, F., Ward, P. J., de Moel, H., Aerts, J. C. J. H., Schmied, H. M., Portmann, F. T., Masaki, Y., Pokhrel, Y., Liu, X., Satoh, Y., Gerten, D., Gosling, S. N., Zaherpour, J., and Wada, Y.: Human impact parameterizations in global hydrological models improve estimates of monthly discharges and hydrological extremes: a multi-model validation study, Environmental Research Letters, 13, 055008, 10.1088/1748-9326/aab96f, 2018.
Wada, Y., Bierkens, M. F. P., de Roo, A., Dirmeyer, P. A., Famiglietti, J. S., Hanasaki, N., Konar, M., Liu, J., Müller Schmied, H., Oki, T., Pokhrel, Y., Sivapalan, M., Troy, T. J., van Dijk, A. I. J. M., van Emmerik, T., Van Huijgevoort, M. H. J., Van Lanen, H. A. J., Vörösmarty, C. J., Wanders, N., and Wheater, H.: Human–water interface in hydrological modelling: current status and future directions, Hydrol. Earth Syst. Sci., 21, 4169-4193, 10.5194/hess-21-4169-2017, 2017.

Wendt, D. E., Bloomfield, J. P., Van Loon, A. F., Garcia, M., Heudorfer, B., Larsen, J., and Hannah, D. M.: Evaluating integrated water management strategies to inform hydrological drought mitigation, Nat. Hazards Earth Syst. Sci., 21, 3113-3139, 10.5194/nhess-21-3113-2021, 2021.
Wu, M., Liu, P., Lei, X., Liao, W., Cai, S., Xia, Q., Zou, K., and Wang, H.: Impact of surface and underground water uses on streamflow in the upper-middle of the Weihe River basin using a modified WetSpa model, Journal of Hydrology, 616, 128840, https://doi.org/10.1016/j.jhydrol.2022.128840, 2023.
Yang, J., McMillan, H., and Zammit, C.: Modeling surface water–groundwater interaction in New Zealand: model development and application, Hydrological Processes, 31, 925-934, 2017.
Yeh, P. J. and Eltahir, E. A.: Representation of water table dynamics in a land surface scheme. Part I: Model development, Journal of climate, 18, 1861-1880, 2005.