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

Yanchen Zheng[1], Gemma Coxon[1], Mostaquimur Rahman[2], Ross Woods[2], Saskia Salwey[1], Youtong Rong[1], Doris E Wendt[1]

[1]School of Geographical Sciences, University of Bristol, Bristol, BS8 1SS, UK
[2]School of Civil, Aerospace and Design Engineering, University of Bristol, Bristol, BS8 1TR, UK

Correspondence to: Yanchen Zheng (yanchen.zheng@bristol.ac.uk)

Abstract. Groundwater is a crucial part of the hydrologic cycle and the largest accessible freshwater source for humans and ecosystems. However, most hydrological models lack explicit representation of surface-groundwater interactions, leading to poor prediction performance in groundwater-dominated catchments. This study presents DECIPHeR-GW v1, a new surface-groundwater hydrological model that couples a Hydrological Response Units (HRU)-based hydrological model and a two-dimensional gridded groundwater model. By using a two-way coupling method, the groundwater model component receives recharge from HRUs, simulates surface-groundwater interactions, and returns groundwater levels and groundwater discharges to HRUs, where river routing is then performed. 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. 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. 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. Furthermore, our model reproduces temporal patterns of the groundwater level timeseries, with more than half of the wells achieving a Spearman correlation coefficient of 0.6 or higher when comparing simulations to observations. 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.

1 Introduction

Groundwater systems are a vital component of the hydrologic cycle connecting recharge zones and discharge, and facilitating complex interactions between the surface and sub-surface (Kuang et al., 2024; Gleeson et al., 2016; Giordano, 2009). As the main freshwater storage component of the hydrologic cycle (Aeschbach-Hertig and Gleeson, 2012), groundwater systems support baseflow levels in rivers (Miller et al., 2016; Gleeson and Richter, 2018) and provide key water supplies for industry, agriculture, and public use, especially during droughts (Famiglietti et al., 2011; Siebert et al., 2010; Giordano, 2009). As such, they are a critical resource for people, economies and the environment (Loaiciga and Doh, 2024) and play a vital role in water management. Often, groundwater models support groundwater management decision-making for local (Wang et al., 2019; Wendt et al., 2021a), national (Dobson et al., 2020; Lee et al., 2007), continental (Rama et al., 2022; Condon and Maxwell, 2015), and global scales (De Graaf et al., 2019; Turner et al., 2019; Gorelick and Zheng, 2015).

Groundwater systems and their interactions with surface water form an active component of the hydrologic water cycle, which can have significant effects on climate, surface energy and water partitioning (Gleeson et al., 2021; Kuang et al., 2024). The importance of representing surface-groundwater water interactions in hydrological models is widely acknowledged (Gleeson et al., 2021; Condon et al., 2021; Bierkens et al., 2015; Clark et al., 2015), especially under the influence of climate change and intense anthropogenic activities (Benz et al., 2024; De Graaf et al., 2019; Condon and Maxwell, 2019). Neglecting these important surface-groundwater water interactions may lead to unrealistic partitioning of precipitation between runoff and other water balance terms, such as significant evapotranspiration biases (Famiglietti and Wood, 1994; Condon and Maxwell, 2019), causing inaccurate prediction of the hydrologic states and fluxes (Naz et al., 2022; Wada et al., 2010). Gnann et al. (2023) demonstrated strong disagreement among many models in describing groundwater recharge, indicating potential errors in estimating the contribution of groundwater to evapotranspiration and streamflow. 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.

To counter these problems, there has been a growing interest in integrating groundwater models into hydrological models, accompanied by notable progress in groundwater modelling analysis and evaluation at various scales (Gleeson et al., 2021; Condon et al., 2021). A variety of coupled surface-groundwater water models has emerged across different scales (summarized in Table S1). Examples at regional scale include SWAT-MODFLOW (Bailey et al., 2016), TopNet-GW (Yang et al., 2017), mHM-OGS (Jing et al., 2018), CWatM-MODFLOW (Guillaumot et al., 2022), GSFLOW-GRASS (Ng et al., 2018), JULES-GFB (Batelis et al., 2020), SHETRAN (Ewen et al., 2000), CLSM-TOPMODEL (Gascoin et al., 2009), CaWaQS3.02 (Flipo et al., 2023), ORCHIDEE (Guimberteau et al., 2014), HydroGeoSphere (Ala-Aho et al., 2017; Brunner and Simmons, 2012) etc.; at the continental scale, such as ParFlow (Maxwell et al., 2015), ParFlow-CLM (Naz et al., 2022); and at the global scale, models like GLOBGM (PCR-GLOBWB-MODFLOW) (Verkaik et al., 2022; De Graaf et al., 2017), WaterGAP2-G^3M (Reinecke et al., 2019; Müller Schmied et al., 2014). The configuration of these models are tailored to their specific purpose and simulation objectives, with each adopting distinct and diverse methodologies for coupling groundwater models. These coupling methodologies range from more simple conceptual approaches to highly sophisticated fully physical-based coupling techniques.

Many conceptual coupled models employ simplified groundwater representations. For example, groundwater is described as a linear reservoir or additional storage (Yang et al., 2017; Gascoin et al., 2009; Guimberteau et al., 2014; Müller Schmied et al., 2014), receiving groundwater recharge and discharging into a river within the same grid cell or other computation unit. These models typically compute time-series of groundwater storage rather than groundwater hydraulic heads. Although representing groundwater as a water storage could enable global-scale assessment of groundwater resources and stress (Turner et al., 2019; Wada et al., 2014; De Graaf et al., 2014), the absence of groundwater hydraulic heads simulations may hinder effective local and regional water resource management (White et al., 2016; Gorelick and Zheng, 2015). Moreover, lateral groundwater flow between grid cells or surface-groundwater interactions is critical as absent lateral flows result in large inaccuracies (Ferguson et al., 2016; Fleckenstein et al., 2010; Xin et al., 2018; Wada et al., 2010). In contrast, some physically-based coupled models integrate three-dimensional (3D) coupled surface-groundwater flow models (Ewen et al., 2000) or adopt pseudo 3D diffusivity equation (Flipo et al., 2023), two-dimensional (2D)/3D Richard's equation (Maxwell et al., 2015; Naz et al., 2022; Brunner and Simmons, 2012; Ala-Aho et al., 2017) to simulate the groundwater flow. Yet, such complex model structure significantly increases numerical complexity and computation time (Jing et al., 2018; Gleeson et al., 2021), resulting in many coupled models remaining uncalibrated or requiring extensive computation time for calibration and validation (Reinecke et al., 2019; Verkaik et al., 2022; Ewen et al., 2000; Maxwell et al., 2015; Naz et al., 2022). Calibrating these models within a stochastic framework is computationally infeasible, leading to significant uncertainty in simulation results, which further hinders an application in large-scale simulations and water management.

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. Section 3 and 4 demonstrate the implementation to 669 catchments in England and Wales and its calibration and evaluation results against large sample of streamflow and groundwater level observations. Discussion of advantages as well as potential future model developments are summarized in the last section.

2 The DECIPHeR-GW model

2.1 Rationale

Our main aim was to develop a coupled hydrological model that represents surface-groundwater interactions whilst maintaining computational efficiency. To achieve this, we coupled a hydrological model (DECIPHeR) with a large-scale 2D groundwater model that have both been applied at national scales (Coxon et al., 2019; Rahman et al., 2023). Both models are described below, note that more details can be found in their respective papers.

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). 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). The model's open-source nature and its flexible model structure facilitated the opportunity to develop new modules of hydrological processes, i.e., groundwater representations. 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.

The large-scale groundwater model utilized in this paper is developed by Rahman et al. (2023). This 2D gridded model employs a transient groundwater flow equation for numerical groundwater flow simulation. Their study presents the first development of a numerical groundwater flow model for large-scale simulations using local hydrogeological information. 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 omits 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). Moreover, this groundwater model also has relatively low requirements of input data and model parameters. Besides open-access data like geology and topography, the model needs groundwater recharge data as inputs, which can typically be derived by a land surface hydrological model. This low data requirement facilitates coupling this groundwater model with other hydrological models.

2.2 DECIPHeR-GW model structure

The new coupled DECIPHeR-GW model fully integrates the DECIPHeR and the groundwater models, as shown in Figure 1, which consists of the HRU-based surface water model component and the 2D gridded based groundwater model. At each time step, the groundwater model receives recharge values (Q_{RC}) from the surface model component, i.e., the root zone storage (S_{RZ}) at HRU scale, simulates surface-groundwater interactions, and passes the derived groundwater head (H_{GW}) and groundwater discharge (Q_{GWDS}) back to HRUs for the river routing.

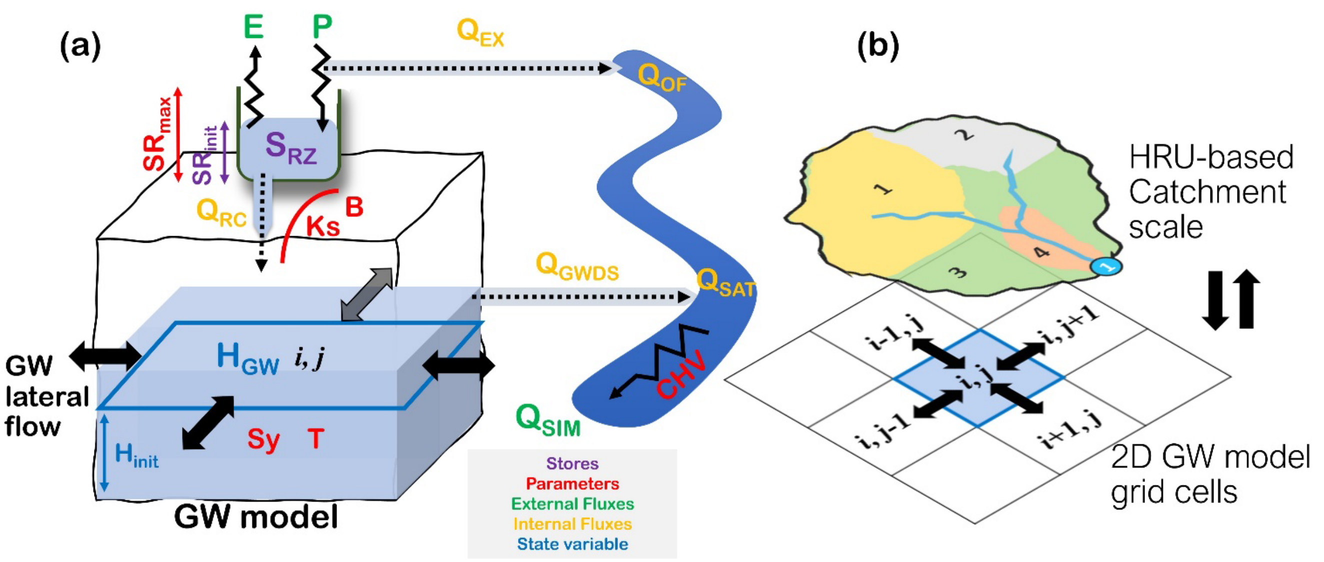


Figure 1: Schematic view of (a) the DECIPHeR-GW model structure and (b) spatial interaction between DECIPHeR HRUs and groundwater model grid cells.

The surface water component (e.g., S_{RZ}) as well as the river routing module of the coupled model were taken from the hydrological model DECIPHeR (Coxon et al., 2019). The root zone store is the main surface water component in the coupled model, which directly interacts with precipitation (P) and evapotranspiration (ET), with a maximum storage determined by the model parameter SR_{max}. At each time step, precipitation is added to S_{RZ}, and the actual evapotranspiration (ET) is calculated and removed directly from the root zone. Equation (1) was used to derive the actual evapotranspiration (ET) for each HRU, which depends on the potential evapotranspiration rate (PET) and the saturation level of the root zone storage.

$$ET = PET \cdot (S_{RZ}/SR_{max}), \tag{1}$$

SR_{init} represents the initial root zone storage for each HRU, which requires initialization at the beginning of the simulation. Previous studies (Coxon et al., 2019; Lane et al., 2021) have shown that this parameter exhibits low sensitivity to the model results. Consequently, SR_{init} is initialized as half of the SR_{max} in this study instead of behaving as a model parameter for calibration. 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.

Recharge Q_{RC} from the root zone storage is computed by implementing the non-linear equation from Famiglietti and Wood (1994), which takes into account the soil hydraulic properties and the storage capacity of the root zone (Equation (2)). In our coupled model setup, recharge is driving the groundwater model component.

$$Q_{RC} = K_s[\frac{S_{RZ}}{SR_{max}}]^{\frac{2+3B}{B}}, \tag{2}$$

where K_s is the saturated hydraulic conductivity (m/time step), and B is the pore size distribution index (dimensionless).

The groundwater model component was developed by Rahman et al. (2023), which uses a transient groundwater flow equation in two spatial dimensions (Equation (3), Figure 1b). The finite difference approximation is used to discretize Equation (3) and an implicit approach is employed to solve it. A no-flow lateral boundary condition is implemented in the model. Spatially, the model domain can be discretized using a user-defined uniform grid according to the topography. With the input of recharge (Q_{RC}), groundwater initial head (H_{init}) and hydrogeology (i.e., transmissivity T and specific yield Sy) data, gridded groundwater heads (H_{GW}) can be calculated at each time step through solving large sets of linear equations.

Whenever modelled groundwater head exceeds the topography, groundwater discharge (Q_{GWDS}) is calculated using Equation (4). 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). Given the high sensitivity of groundwater head simulation to hydrogeological data (Rahman et al., 2023), transmissivity (T) and specific yield (Sy) are selected as model parameters for calibration in the coupled model.

$$S_y \frac{\partial h}{\partial t} = \nabla(T\nabla h) + R, \tag{3}$$

$$Q = S_y \times (h - h_{top}), \tag{4}$$

where S_y is specific yield (-), h is the groundwater head (m), t is time, T is transmissivity (m^2/time step), R is the potential recharge rate (m/time step) and h_{top} is the topographic height (m).

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 SR_{max}, 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).

Table 1. Overview of model stores, fluxes, state variables and parameters. (mAOD in this table stands for metres Above Ordnance Datum, i.e. sea level)

Category	Name	Meaning	Unit
Stores	S_{RZ}	Root zone storage	m
	SR_{init}	Initial root zone storage	m
Internal fluxes	Q_{EX}	Saturated excess flow	m/time step
	Q_{OF}	Overland flow	m/time step
	Q_{RC}	Recharge flow	m/time step
	Q_{GWDS}	Groundwater discharge	m/time step
	Q_{SAT}	Saturated flow	m/time step
External fluxes: input	P	Precipitation	m/time step
	ET	Actual evapotranspiration	m/time step
External fluxes: output	Q_{sim}	Simulated discharge	m/time step
State variable	H_{init}	Initial groundwater head	m (mAOD)
	H_{GW}	Groundwater head	m (mAOD)
Model parameters	SR_{max}	Maximum root zone storage	m
	Ks	Saturated hydraulic conductivity	m/time step
	B	Pore size distribution index	dimensionless (-)
	CHV	Channel routing velocity	m/time step
	T	Transmissivity	m^2/time step
	Sy	Specific yield	dimensionless (-)

2.3 Surface-groundwater interactions

To represent dynamic surface-groundwater interactions, three scenarios (as shown in Figure 2a, b and c) have been implemented in the coupled model setup. At each time step, the position of the groundwater head and root zone storage determines the occurrence and the amount of recharge. For example, if the groundwater head is below the bottom of the root zone (Figure 2a), we assume that recharge occurs, leaking from the root zone storage to the groundwater system after removing the actual evapotranspiration. As presented in the Equation (2) of Section 2.2, the value of recharge depends on the soil texture and the saturation level of root zone storage. The recharge was set not to exceed the root zone storage S_{RZ}. The bottom of root zone is defined as the topography H_{topo} minus the depth of the root zone D_{RZ}. The root zone depth is estimated using Equation (5) according to previous studies (Wang-Erlandsson et al., 2016; Lane et al., 2021).

$$D_{RZ} = \frac{SR_{max}}{porosity}, \tag{5}$$

If the groundwater head reaches the bottom of the root zone but below the topography (Figure 2b), we assume no exchange of water takes place between the surface and groundwater system in this case (i.e., no recharge). In the last scenario, if groundwater head exceeds the topography (Figure 2c), groundwater discharge is generated (no recharge). Groundwater discharge is subsequently passed to the HRUs as the saturated flow and added to the nearest river channel for river routing.

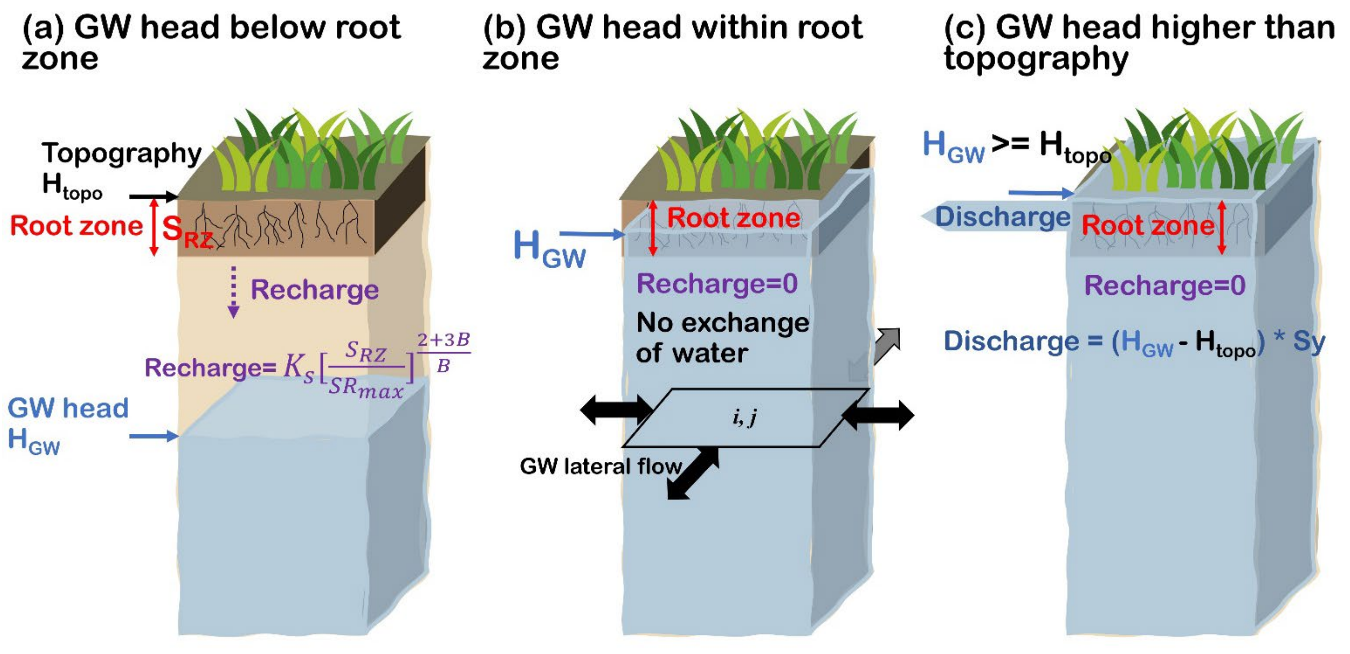


Figure 2: Schematic model set up of surface-groundwater interactions under three scenarios: (a) groundwater head is below the bottom of the root zone; (b) groundwater head is within the root zone; and (c) groundwater head is higher than the topography. The colour coding of the text is as follows, red indicates the root zone, purple represents recharge, and blue denotes the modelled groundwater heads.

In all three scenarios, the root zone storage receives rainfall and actual evapotranspiration is subtracted as usual at every time step (Equation (1)), regardless of the movement of the groundwater heads. Whenever root zone storage is full, any rainfall excess is generated as overland flow and then added to the river channel.

Given that we build and run the coupled model for each catchment, the groundwater model gridded domain needs to be first determined according to the catchment boundary before the simulations. 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. The groundwater grids and buffer zones outside the catchment boundaries do not incorporate or consider HRUs, which are exclusively confined within the catchment boundaries. 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.

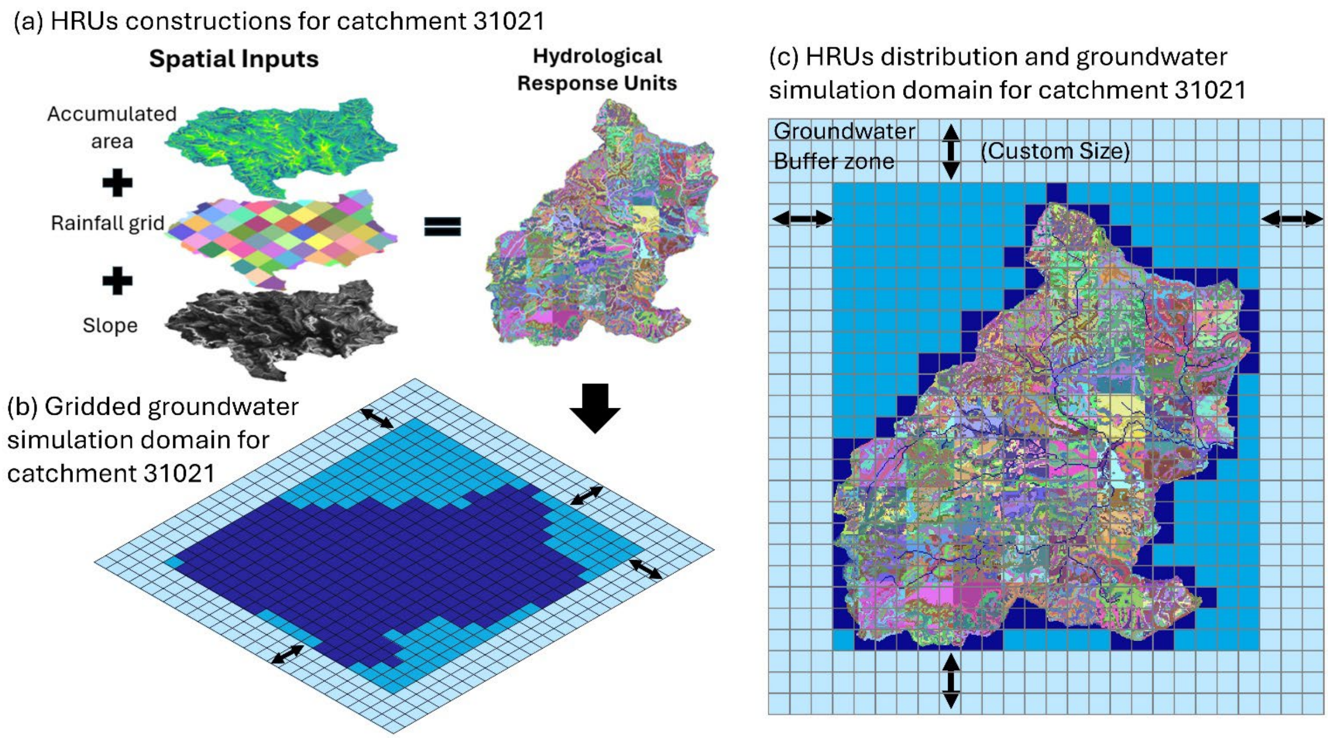


Figure 3: The DECIPHeR-GW coupling and spatial interaction from DECIPHeR Hydrologic Response Units (HRUs) to groundwater model grid cells for one example catchment Welland at Ashley 31021. (a) the HRUs constructions process for catchment 31021; (b) the gridded groundwater simulation domain for catchment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

The recharge, groundwater discharge fluxes as well as the state variable groundwater head need to be transferred between surface water component HRUs and gridded groundwater cells. To address this spatial scale discrepancy between variables, a model mapping scheme is adopted, which follows a similar procedure to coupling the HRU-based SWAT model and gridded groundwater model MODFLOW (Bailey et al., 2016). For a given HRU, the proportion of its area overlapped by different grids is needed for transferring variables from HRUs to grids. Conversely, to transfer variables from grids to HRUs, the proportion of each grid cell area that is occupied by different HRUs is needed. Both these proportions are calculated as the weighting matrix at the beginning of the simulation and stored for transferring variables at each time step. Detailed model mapping methods and the schematic figures can be found in Text S2 and Figure S1-S3. Water balance checks were implemented to verify conservation of mass in the coupled model (See Text S3 of the supporting information).

3 Model implementation and evaluation across England and Wales

3.1 Study area and catchments selection

To test our new coupled model, we apply DECIPHeR-GW over a large sample of catchments across England and Wales. Extensive and high-quality open source hydro-climate and geological data are available in England and Wales, such as the CAMELS-GB dataset (Coxon et al., 2020), along with a large amount of groundwater level observations (Environment Agency, 2023), making it highly suitable for testing and evaluating our coupled model. Also, Great Britain exhibits a wide diversity in hydrogeology with units ranging in age from Pre-Cambrian (Allen et al., 1997), resulting in a wide variety of aquifer types (Figure S5). This allows us to test the robustness of the coupled model under a range of hydrogeological conditions modelling for the three principal aquifers: Chalk, Permo-Triassic sandstone and Jurassic limestone (Allen et al., 1997). The Chalk aquifer, notably distributed in the south-east of England, is highly permeable, where catchments are connected to a wider regional groundwater system, resulting in inter-catchment subsurface flows (Allen et al., 1997; Oldham et al., 2023). Despite the vast range of hydrological models applied to this region (Coxon et al., 2019; Lane et al., 2019; Lane et al., 2021; Hannaford et al., 2022; Lees et al., 2021; Bell et al., 2007; Ewen et al., 2000; Seibert et al., 2018; Lewis, 2016), deficiencies in model performance persist for these groundwater flow-dominated catchments. Thus, we test our coupled model over England and Wales, with the aim of improving model performance in these groundwater-dominated regions through better representation of surface-groundwater interactions.

We selected 669 catchments from all river records in the National River Flow Archive (NRFA) across England and Wales to evaluate the coupled model and represent a variety of hydro-climate characteristics, which ensures the robustness and generalizability of our results. All catchments are shown in the Figure 4a-c that are selected based on the following data criteria. Note that catchments in Scotland were excluded from our analysis due to lack of access to hydrogeological data.

First, to ensure robust calibration, only catchments with over 20 years of observed data within the calibration period spanning from 1980 to 2010 were selected. The model was configured to run from 1970 to 2020 based on data availability, capturing a broad range of climate conditions during this period. The initial 10 years served as a warm-up period, with calibration performed from 1980 to 2010, followed by model evaluation in the subsequent years. Secondly, we excluded catchments that are affected significantly by reservoirs as the coupled model does not incorporate the reservoir operating rules. Using a suite of hydrological signatures we identified 25 catchments where reservoirs were having a significant impact on the water balance or flow variability and excluded these from our sample (Salwey et al., 2023). Thirdly, catchments with runoff coefficient (calculated as the ratio of mean annual discharge and mean annual precipitation) greater than 1 were also excluded from the analysis due to potential issues with data quality, missing rainfall data or substantial human-water interactions that we didn't consider in this coupled model.

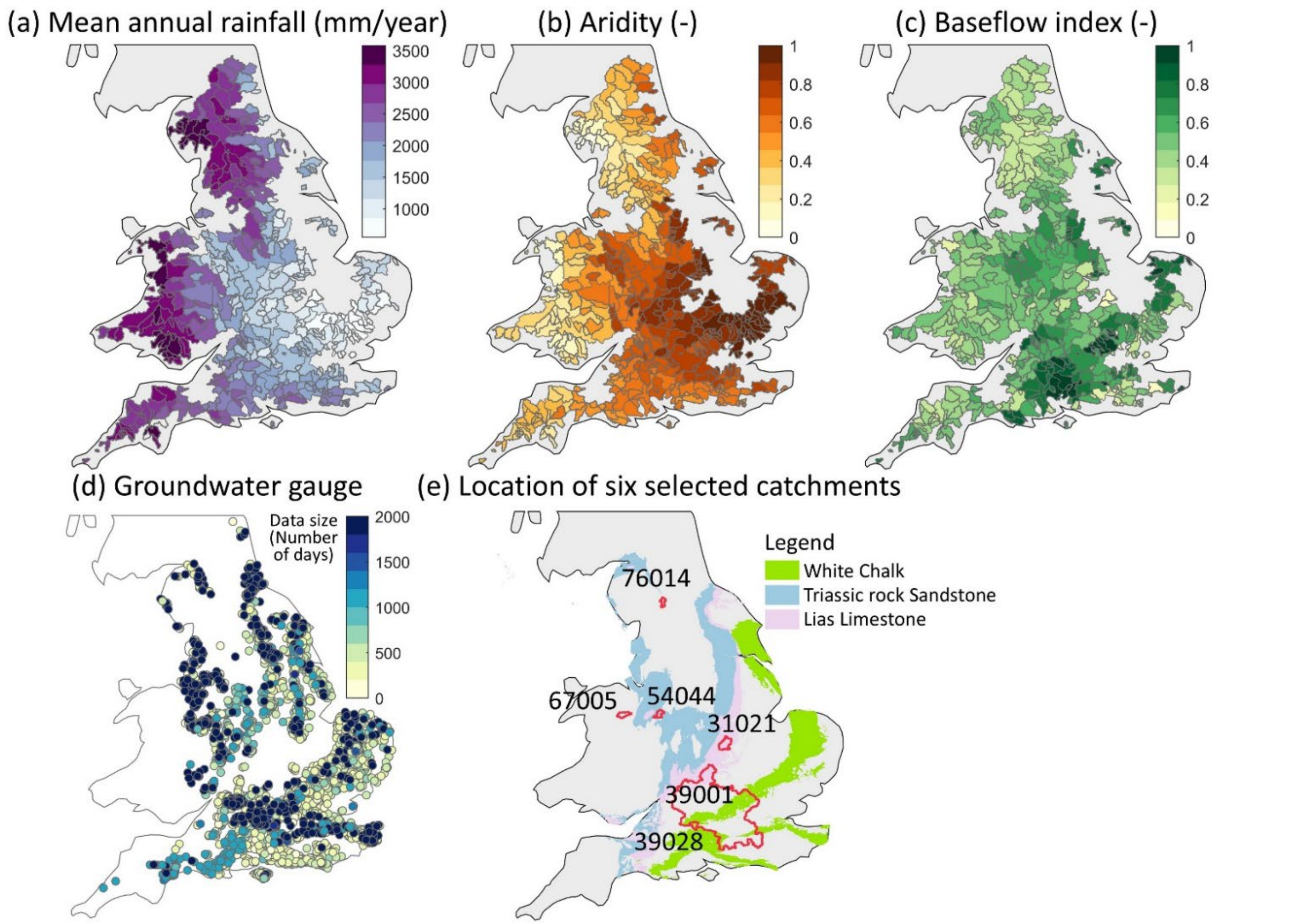


Figure 4: Hydro-climate, geology, and available groundwater well locations of 669 catchments used in this study. (a) Mean annual rainfall (mm/year), (b) Aridity (-), (c) Baseflow index (-), (d) The locations of 3888 groundwater wells collected in this study, and (e) The locations of six selected catchments (Details in Section 4.2 and Table 4). The hydrogeological properties map in the background (this figure contains British Geological Survey materials © UKRI 2020) highlights highly productive aquifers, including white Chalk, Triassic Sandstone, and Lias Limestone.

3.2 Surface water component and groundwater model configuration

For the surface water component, a 50 m gridded digital elevation model (Intermap Technologies, 2009) (also used in Coxon et al. (2019); Lane et al. (2021); Coxon et al. (2019)) 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. Headwater cells were extracted from Ordnance Survey river layers (Ordnance Survey, 2023) and then routed downstream along the steepest slopes in the catchment to create the river network for the coupled model. Defining HRUs is a critical step in the application of surface water component, because these HRUs act as an individual model store with different spatial inputs and model parameter values. 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. The groundwater model simulation domain is defined by grids overlaying the catchment boundary and the buffer zone. Text S11 and Figure S14 in supporting information provide details of how we determined a buffer zone size, which resulted in a 3 km buffer zone around the catchment boundary to reduce the impact of no-flow boundary conditions. Future users can adjust this buffer value as needed. We used the long-term steady-state simulated groundwater heads from Rahman et al. (2023) as the initial condition for the groundwater model to ensure the model achieves a stable and reasonable operational state as quickly as possible. 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, 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)

3.3 Input and evaluation datasets

Daily precipitation, potential evapotranspiration (PET), streamflow and groundwater level data were used to run and evaluate DECIPHeR-GW. For the input data, this study uses the observation-based gridded daily precipitation and PET data derived from HadUK-Grid, a newly produced dataset providing gridded climate observations for the UK at a spatial resolution of 1km (Hollis et al., 2019). Daily precipitation data from HadUK-Grid, available from 1891-present, is derived from the Met Office UK rain gauge network, which is quality controlled and then inverse-distance weighted interpolation is applied to generate the daily rainfall grids. Daily PET data, available from 1969-2021, is calculated using the Penman-Monteith equation with climate variables obtained from HadUK-Grid (Robinson et al., 2023). To align with the existing model setup and the grid used for national climate (Robinson et al., 2021; Lane and Kay, 2022; Salwey et al., 2024), these climate variables were upscaled to a 2.2-km grid for hydrological simulations.

To evaluate river flows generated in DECIPHeR-GW, daily observed streamflow data sourced from NRFA were used to calibrate and evaluate the model performance. The modelled groundwater levels are evaluated using groundwater level observation data from the Environment Agency's groundwater monitoring network database (Enviroment Agency, 2023). The groundwater level observations for a total of 3888 groundwater wells in England and Wales were collected, which covers a variety of temporal resolution and coverage with varying levels of data quality. Before using these in model evaluation, several quality control steps were applied to the measured groundwater level data, as illustrated in Figure S4b. Details of data quality control are provided in the supporting information (Text S4). There are 3005 wells providing manually measured data ('Dipped data') at either daily or monthly intervals, while 883 wells offer automatically 'Logged data' recorded by pressure transducers at sub-daily scale. Furthermore, there are 395 wells where both types of data are available (see the locations in Figure 4d and

Figure S4a). The temporal coverage varies significantly, with a median of approximately 41 years and the shortest period being just 4 years of non-continuous observations (Figure S4c). After the data quality control, data from 1804 groundwater wells were used for the model evaluation.

3.4 Model parameters

A total of six model parameters need to be calibrated to run the coupled model. Parameters SR_{max} and CHV were already included in the DECIPHeR model structure. For the coupled model, we sampled these two model parameters using the same method following Lane et al. (2021). Specifically, SR_{max} is sampled by adopting the multiscale parameter regionalization (MPR) strategy, which was first estimated at the high resolution based on the geophysical data and the transfer function, and then upscaled to the HRU scale. The channel routing parameter CHV, which is not associated with spatial fields, was not parameterized using MPR and calibrated through random sampling instead. Details about the sampling method of these two model parameters can be found in the work from Lane et al. (2021).

In addition to the two mentioned above model parameters, we have introduced 4 new model parameters in the coupled model, i.e., saturated hydraulic conductivity (Ks) and pore size distribution index (B), which interact with the surface water components, and transmissivity (T) and specific yield (Sy), which drive groundwater flow. We use representative ranges of saturated hydraulic conductivity (Ks) and pore size distribution (B) from various soil texture measured from a large sample of soil from Clapp and Hornberger (1978); (Rawls et al., 1982). Maps of soil surface properties (porosity, percentage sand, silt and clay) at a 50m raster were sourced from Lane et al. (2021) for deriving the root zone depth and soil texture classification. Soil texture is classified based on the United States Department of Agriculture (USDA) criteria. Ks and B values were sampled in the corresponding range for each soil texture classification using Monte-Carlo method at the high resolution map (50m raster) of soil texture, and then the geometric mean was calculated for upscaling to the HRU scale for calibration.

Transmissivity (T) and specific yield (Sy), as the parameters of groundwater component, needed to align with its gridded structure, which is set at 1 km grid resolution for parameter input. Following Rahman et al (2023), these parameters can be obtained from the representative ranges for different lithology classes based on extensive literature review and reports for England and Wales (Allen et al., 1997; Jones et al., 2000). The 1:625000 scale digital geological map of the United Kingdom developed by the British Geological Survey (BGS) is used for providing the lithology classes at 1 km grid resolution. By adopting this lithology map and the lookup table from Rahman et al. (2023), the parameter values of T and Sy can be sampled through Monte-Carlo method for every 1 km grid cell. Table 3 summarizes the functions, parameter ranges and catchment attributes data used in this study for sampling the model parameters. The lookup tables for linking Ks, B with soil texture class and T, Sy with lithology types as well as the detailed parameter ranges are provided in Table S2 and Table S3 in the supporting information. Since the model parameters are linked with the soil and lithology types, catchments with the same spatial attributes will be calibrated with the same set of model parameters.

Table 3. Model parameters range, transfer functions and catchment attributes data used in this study

Parameter	Parameter description (Unit)	Catchment attribute data/ Sampling method	Transfer function/ Parameter Range
SR_{max}	Maximum root zone storage (m)	Porosity (p) and land use (u). Global parameters are constrained using the root depth associated with different land uses.	$SRmax = g_1 \cdot p \cdot \begin{cases} g_2, & u=1 \\ g_3, & u=2 \\ g_4, & u=3 \\ \vdots & \\ g_{11}, & u=10 \end{cases}$ g_1 is the scaling factor. $g_2 \sim g_{11}$ are the estimated root zone depths for different land use types. Details see Lane et al. (2021).
CHV	Channel routing velocity (m/time step)	Random sampling from the lower and upper bound according to previous applications (Coxon et al., 2019; Lane et al., 2021)	[100, 4000]
Ks	Saturated hydraulic conductivity (m/time step)	Surface soil texture (*sc*) based on the percentage sand, percentage clay and percentage silt; Lookup table from (Clapp and Hornberger, 1978; Rawls et al., 1982) linking *Ks* field measured representative values range according to soil texture	$Ks = \begin{cases} g_{12}, & sc=1 \\ g_{13}, & sc=2 \\ g_{14}, & sc=3 \\ \vdots & \\ g_{22}, & sc=11 \end{cases}$ Ks values range for each soil texture class is presented in Table S2.
B	Pore size distribution index (-)	Same with Ks, lookup table linking B field measured representative values according to soil texture (*sc*)	$B = \begin{cases} g_{23}, & sc=1 \\ g_{24}, & sc=2 \\ g_{25}, & sc=3 \\ \vdots & \\ g_{33}, & sc=11 \end{cases}$ B values range for each soil texture class is presented in Table S2.
T	Transmissivity (m^2/time step)	Lithology types (*lt*); Lookup table from (Rahman et al., 2023)	$T = \begin{cases} t_1, & lt=1 \\ t_2, & lt=2 \\ t_3, & lt=3 \\ \vdots & \\ t_n, & lt=n \end{cases}$ T values range for each lithology type is presented in Table S3. *n* is the total number of lithology types.
Sy	Specific yield (-)	Lithology types (*lt*); Lookup table from (Rahman et al., 2023)	$S_y = \begin{cases} s_1, & lt=1 \\ s_2, & lt=2 \\ s_3, & lt=3 \\ \vdots & \\ s_n, & lt=n \end{cases}$ Sy values range for each lithology type is presented in Table S3. *n* is the total number of lithology types.

3.5 Model calibration and evaluation

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. The objective is to utilize these simulations as a benchmark to evaluate the performance of the coupled model after implementing the groundwater processes representation. Note that these benchmark model runs are calibrated and evaluated using the same method with the coupled model as described below.

We use non-parametric KGE metrics (Pool et al., 2018) to calibrate and evaluate the model results, which comprises three components accounting for the errors in mean flow, flow variability and the correlation between observed and simulated flow. This non-parametric KGE is proposed to avoid overfitting to particular hydrograph elements. In contrast to the parametric KGE (Gupta et al., 2009), this metric incorporates the difference between Flow Duration Curve (FDC) to indicate variability instead of standard deviation and employs Spearman correlation in place of the Pearson correlation coefficient.

Both coupled and benchmark model was calibrated and evaluated across all 669 catchments by running 5000 simulations in each catchment (i.e., each of the 5000 regionalization of parameters g_1-g_{33}, t_1-t_n, s_1-s_n mentioned in Table 3 is used for all catchments). The model simulates the period from 1970 to 2020 at daily time step. Simulations from 1970 to 1979 were treated as a warm-up period, and the non-parametric KGE was calculated separately for the calibration period from 1980 to 2010 and the evaluation period spanning from 2011 to 2020. These periods were selected as a suitable test for the model, encompassing a variety of climatic conditions to showcase its capability to reproduce major national-scale hydrological extremes, including floods in 2007, 2015, and 2019, as well as droughts in 1984, 2003, 2011 and 2018. 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. The first catchment by catchment calibration is to find the best performing simulation (maximum KGE across 5000 simulations) and its corresponding parameter sets for each catchment. The second nationally-consistent calibration scheme enables us to identify the best national model parameter sets across all catchments. The median KGE across all catchments is calculated for each simulation and the nationally-consistent approach selects the simulation with the highest median KGE. The second calibration approach is beneficial for national model parameter regionalization, offering valuable insights on model parameter selection for model application in ungauged catchments. In contrast, the first calibration method demonstrates the optimal performance achievable by our coupled model. For the national-consistent calibration approach, following Lane et al., (2021), catchments with maximum KGE values below 0.3 in the first calibration method (catchment by catchment) were excluded from the median KGE calculation. This exclusion avoids catchments where the model structure was not suitable, while retaining as many catchments as possible.

Furthermore, modelled groundwater levels are assessed using a large sample of groundwater level observations from 1804 wells in England and Wales (described in Section 3.3) for the model evaluation. Due to the scale discrepancy between the 1 km grid scale simulated groundwater level and point-scale observations of specific wells, we use the Spearman correlation coefficient to quantify the ability of the coupled model in reproducing the temporal correlation and don't calculate the bias.

4 Results

4.1 Overall model performance across catchments

Figure 5a presents the non-parametric KGE values of the simulated streamflow for the coupled model across 669 streamflow gauges during the evaluation period. The calibration results, which are consistent with evaluation results, are detailed in the Supporting Information file (Figure S6). Using the catchment-by-catchment calibration method (Figure 5a-d), overall, the coupled model performs well in simulating streamflow across catchments, with a median KGE of 0.83, and most catchments (81%) achieving 0.7 or higher. Figure 5b illustrates the KGE differences between the coupled model and benchmark runs by using DECIPHeR. Approximately 70% of the catchments exhibit KGE differences of 0.1 or less between the coupled and benchmark models, indicating that the coupled model achieves comparable results with those of the benchmark model.

Notably, the coupled model demonstrates better performance in groundwater-dominated chalk catchments with baseflow index > 0.75 (blue dots in Figure 5b), where the average KGE improves from 0.49 with the benchmark model to 0.70. In the southeast's chalk region, the coupled model achieves KGE improvements exceeding 0.35 in 20 catchments, with 6 catchments showing improvements greater than 1. In contrast, the benchmark model performs slightly better in the western regions of England and Wales (indicated by orange dots in Figure 5b), where catchments are wetter with mean annual rainfall exceeding 1500 mm/year, achieving a median KGE around 0.88. Nevertheless, the coupled model still maintains a median KGE of 0.80 for these wetter catchments. The comparison of the KGE bias component between two models, as displayed in Figure 5c and 5d, further confirms that the coupled model improves the reproduction of the water balance for these groundwater-dominated catchments in the southeast, particularly those in the Thames River basin. 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., 2021b; Bloomfield et al., 2021).

As expected, a performance drop is observed in the national-consistent calibration strategy (Figure 5e-f), since the parameterization is not optimized for individual catchments. Compared to the catchment-by-catchment calibration, approximately 50% of catchments experienced a decline of less than 0.1 in KGE for the coupled model, whereas 64% experienced a decline for the benchmark. The decrease in KGE scores is primarily concentrated in the southeast of England, echoing findings of Lane et al. (2021). This might be attributed to the method used for catchment selection in the national regionalization process. Groundwater-dominated catchments with baseflow index > 0.75 account for less than 10% of the total catchments calibrated in this study. 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.

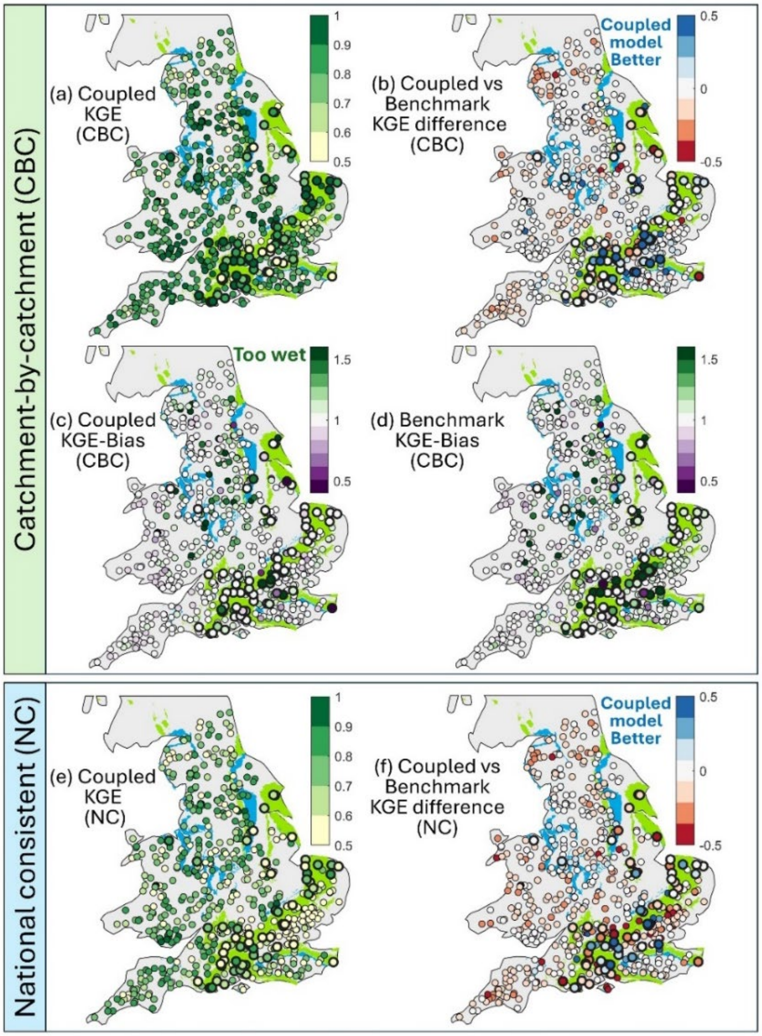


Figure 5: Spatial maps of model performance using two calibration approaches (a) Catchment-by-catchment (CBC) and (e) National-consistent (NC), the non-parametric KGE differences between the coupled model and the corresponding DECIPHeR benchmark runs (b, f), and the bias component of KGE for the coupled model and benchmark runs under CBC approach (c, d). The maps for other KGE components are provided in the supporting information (Figure S7). Each dot represents the performance at a river gauge during the evaluation period. Model performance maps for the calibration period are provided in the supporting information (Figure S6). The scatter dots for groundwater-dominated catchments (baseflow index > 0.75) were labelled with larger dots and outlined with thicker borders. The background of the maps highlights the areas of high productivity in aquifers (this figure contains British Geological Survey materials © UKRI 2020). Light green represents highly productive aquifer (fracture flow), while blue indicates the intergranular flow of a highly productive aquifer.

4.2 Performance of simulated flow timeseries

Six catchments were selected to demonstrate the coupled model's ability in reproducing the streamflow timeseries with distinct characteristics, i.e., climate conditions, geology types and levels of human impact (Table 4). Specifically, catchments 76014 and 67005 were selected to evaluate coupled model performance in wet climate (mean annual rainfall > 1200 mm/yr), while 39028 and 39001, differing in human impact, represented dry chalk catchments. Catchment 31021 was chosen for limestone, and 54044 for sandstone. The simulation of the 2-year period for 2010 to 2012 using the calibration period model parameters is presented here for these catchments, as it encompasses diverse hydrological extreme events (Marsh et al., 2013). The evaluation period model parameters exhibit the similar pattern and won't change the herein analysis.

Figure 6 illustrates DECIPHeR-GW results for a wide spectrum of hydrological dynamics, including the wetter catchments in the northwest England and north Wales (Figure 6a, b), as well as the drier catchments in the south-east (Figure 6e, f). 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 6e), 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. The simulated streamflow hydrograph using the national-consistent calibration method also closely aligns with the results from the catchment-by-catchment calibration method, with relatively larger differences in performance observed in groundwater-dominated catchments (Figure 6e).

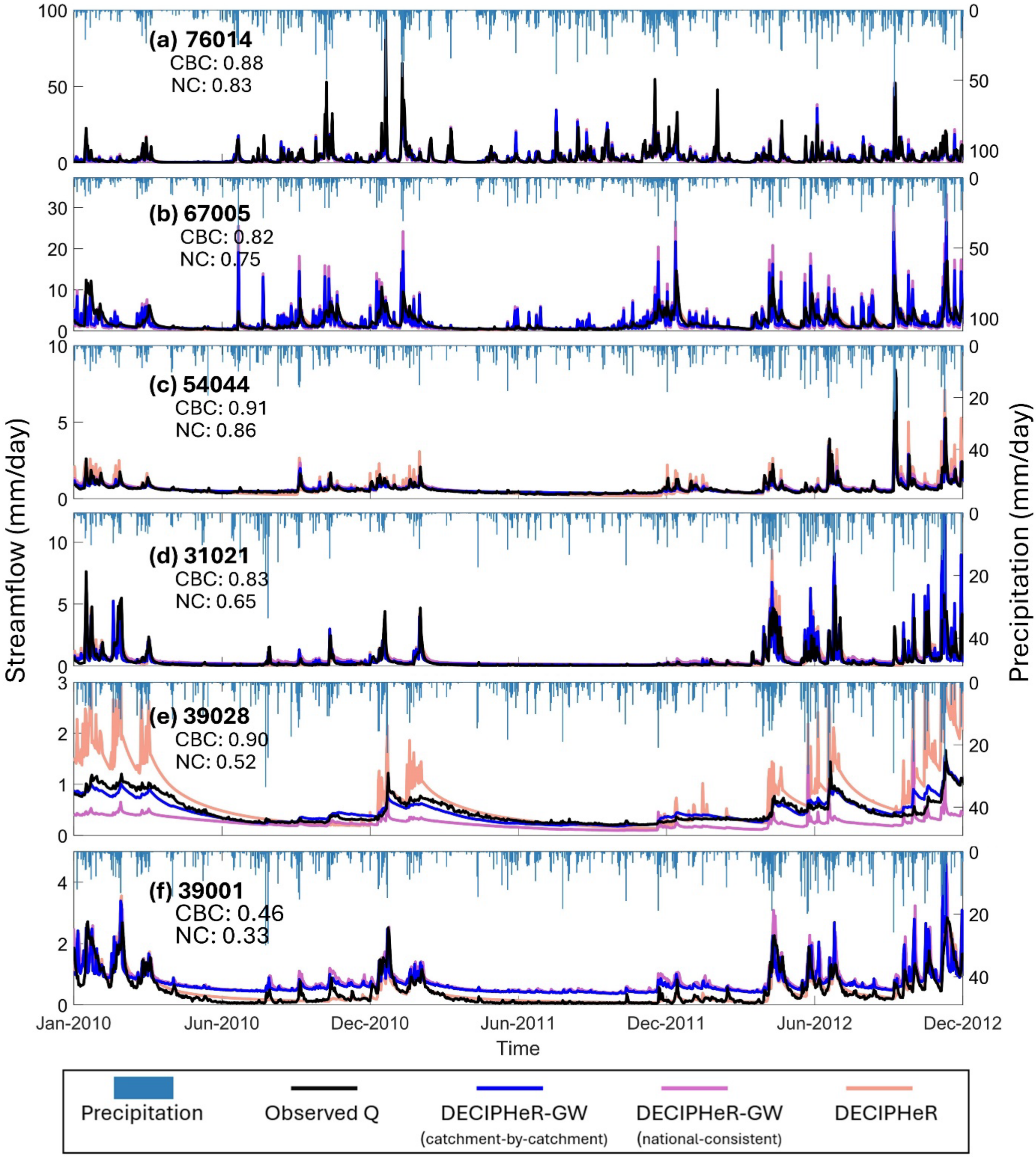
Table 4. Catchment attributes and model performance for the six selected catchments. Their locations are presented in Figure 4e, simulated hydrographs are shown in Figure 6. Baseflow index and aridity are derived from the CAMELS-GB dataset (Coxon et al., 2020). Runoff coefficient is calculated as the mean annual discharge divided by mean annual rainfall. The KGE values presented in this table calculated for calibration periods under catchment-by-catchment (CBC) and national consistent (NC) calibration approaches. Benchmark KGE represents the results from DECIPHeR.

Gauge number	River	Station location	Catchment area (km^2)	Geology type	Mean annual rainfall (mm/yr)	Mean annual PET (mm/yr)	Mean annual discharge (mm/yr)	Runoff coefficient (-)	Baseflow index (-)	Aridity (-)	Coupled model KGE (CBC)	Coupled model KGE(NC)	Benchmark KGE (CBC)	Bench mark KGE (NC)
76014	Eden	Kirkby Stephen	69	No highly permeable bedrock	1514	434	1248	0.82	0.38	0.29	0.88	0.83	0.89	0.83
67005	Ceiriog	Brynkinalt Weir	112	No highly permeable bedrock	1211	477	849	0.70	0.57	0.39	0.82	0.75	0.93	0.92
54044	Tern	Ternhill	93	Sandstone	738	500	280	0.38	0.78	0.68	0.91	0.86	0.82	0.70
31021	Welland	Ashley	247	Limestone	646	508	175	0.27	0.46	0.79	0.83	0.65	0.84	0.75
39028	Dun	Hungerford	101	Chalk	806	505	217	0.27	0.85	0.63	0.90	0.52	0.50	0.23
39001	Thames	Kingston	9948	Chalk	710	508	193	0.27	0.63	0.72	0.46	0.33	0.85	0.61

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. 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. This implies that the benchmark calibration could produce good results, but potentially due to the parameterization that compensates for the absence of these processes representation. Ensuring that model performs well with appropriately structured components is crucial for maintaining both accuracy and reliability (Kirchner, 2006; Gupta et al., 2012).

Furthermore, the simulated streamflow hydrographs for the wetter catchments tends to be flashier than the benchmark simulations (as shown in catchment 67005, Figure 6b). This might be related to the relatively wet conditions of the catchment in combination with the underlaying groundwater system is already saturated or nearly saturated. Once the root zone reaches capacity, runoff is quickly generated as the excess rainfall, leading to a rapid response to precipitation and resulting in more pronounced spikes in the hydrographs. The dynamic variations of these internal variables for this catchment during 2010-2012 are provided in the supporting information (Figure S8). However, for most wet catchments (mean annual rainfall > 1500 mm/year), the coupled model performs well (examples in catchment 76014, Figure 6a), with around 78% of these catchments achieving a KGE greater than 0.7.

A simple model parameter sensitivity analysis (details provided in supporting information Text S10) reveals that the parameters of the surface model component have a greater influence on simulated streamflow hydrographs than on modelled groundwater levels (as seen in Figure S10 and Figure S13). SR_{max}, which determines the maximum root zone storage, plays a crucial role in regulating the flashiness of simulated flows (Figure S10a). Smaller SR_{max} values lead to increased variability in runoff, as runoff is rapidly generated whenever SR_{max} reaches its capacity, causing spikes in the hydrographs due to excess rainfall. Both the B and Ks parameters control the magnitude of recharge, as shown in Figure S10b and c, their effects on simulating streamflow hydrographs are similar, with a relatively greater impact observed for parameter B. Smaller B values lead to reduced recharge, causing the root zone storage to fill up more quickly and resulting in increased overflow and also flashier in streamflow hydrographs. The groundwater related parameters, i.e., T and Sy are intended to control groundwater levels more than streamflow, which is confirmed by this analysis (see Figure S11 and S12). Consequently, this sensitivity analysis indicates that increasing SR_{max} or B values could result in smoother streamflow hydrographs and therefore might improve DECIPHeR-GW's performance in wetter catchments.



480 **Figure 6: Observed and the best simulated streamflow hydrographs using the model parameters from the calibration period for the six catchments across different catchment attributes (shown in Table 4). The best simulated DECIPHeR-GW hydrographs along with their KGE values for both catchment-by-catchment (CBC) and national-consistent (NC) are provided. The DECIPHeR model simulation results (the orange line) presented here are based on the CBC calibration method. To enhance clarity and simplify the**

visuals, the simulation results for the NC calibration method from DECIPHeR are not plotted here, but the KGE metrics for each catchment under this NC calibration method are detailed in Table 4.

4.3 Model evaluations with groundwater levels

We used 1804 groundwater well observations to evaluate grid-scale simulated groundwater levels. In this study, we calibrated the model solely using streamflow data as our objective, while utilizing groundwater observations to evaluate the internal dynamics of the coupled model. Figure 7a-b illustrates groundwater simulations corresponding to the best streamflow simulations under two streamflow calibration methods, i.e. catchment-by-catchment and national consistent. Overall, the groundwater simulation results are generally reliable in capturing the temporal correlation of the observations, particularly in the Chalk region, where over 75% wells achieve Spearman correlation coefficients above 0.6 with a median of 0.77. 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.

Taking catchment 39028 as an example, Figure 7c demonstrates that model performance can vary across 5000 simulations under catchment-by-catchment calibration method. The median Spearman correlation coefficients for different groundwater grids across all simulations in general reach 0.6 or higher. A portion of the groundwater wells has a median Spearman coefficient for groundwater levels exceeding 0.8 (see groundwater well 3, 4 and 5 in Figure 7c), underscoring the model's capability in reproducing the temporal patterns of groundwater variations. Figure 7d presents two examples of simulated groundwater level timeseries against well observations. While these examples are not from the best simulations, they are chosen to demonstrate the model's performance under conditions of both strong and weak temporal correlation.

Figure S12 in the supporting information illustrates the impact of T and Sy model parameters on the groundwater level timeseries for example catchment 39028 (details are recorded in Text S10). Higher T values generally result in lower groundwater levels, which is to be expected as higher transmissivity (T) facilitates quicker lateral flow through an aquifer. In contrast, when Sy is low, the speed of groundwater flow and storage capacity may be reduced, resulting in flashier groundwater levels increasing their variability. Our results confirm the above patterns, showing that higher T leads to decreased groundwater levels and lower Sy leads to greater variability (Figure S12a and b), highlighting the overall agreement and well-representation of physical processes of our coupled model.

Given that poorer temporal correlation observed in some wells, we investigated which factors could contribute, such as short groundwater observation records, low streamflow accuracy in catchments, distance between wells and rivers, and attributes like borehole depth, elevation of wells, and grid elevation contributed to the discrepancies. Our findings point towards key factors, such as borehole depth, river proximity, and streamflow accuracy, which might be affecting the ability to model groundwater levels accurately (see details in Figure S9). We have found lower spearman correlations for wells with deeper boreholes, those closer to the river or the wells with lower streamflow simulation accuracy. This is likely because our groundwater model is 2D without explicit river features representation, which can result in lower performance for wells that are deeper or closer to rivers. More details are discussed in Section 5.2.

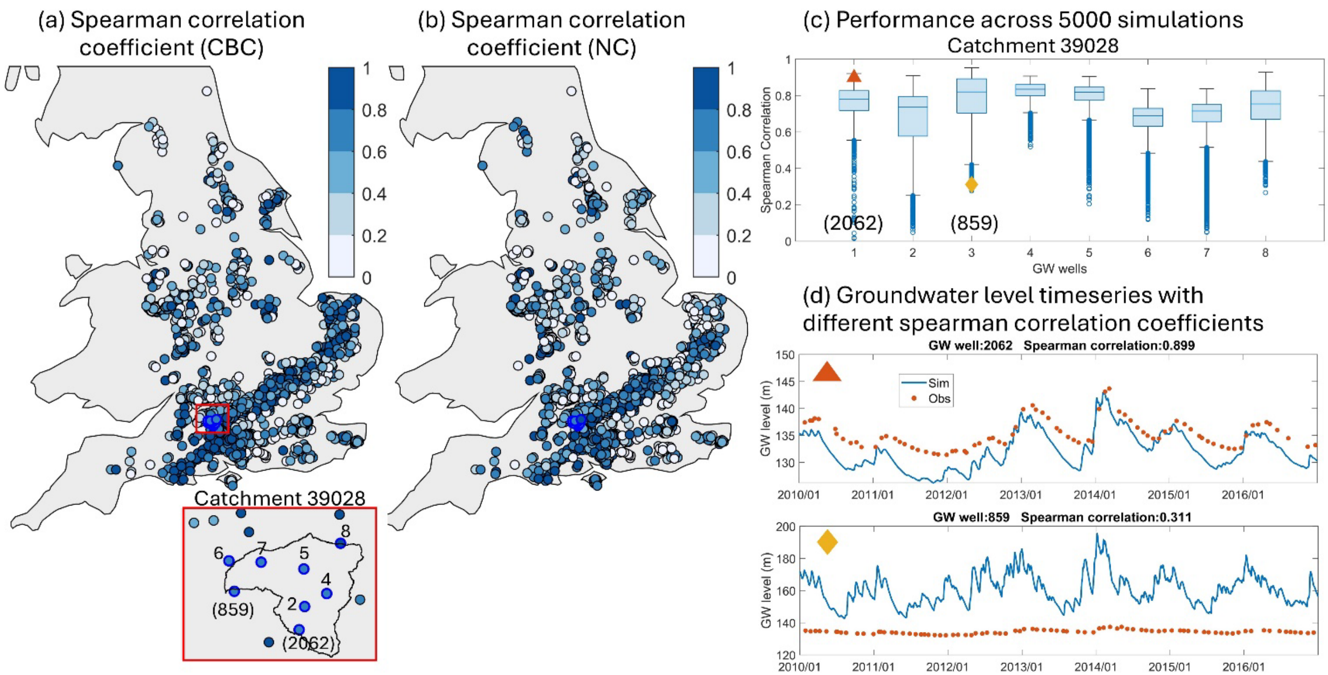


Figure 7: Spatial maps of groundwater level evaluation results. (a) and (b) shows the evaluation results for the simulated groundwater levels under the catchment-by-catchment (CBC) and national-consistent (NC) streamflow calibration methods, respectively. (c) presents the performance of the eight groundwater grids in the Dun at Hungerford catchment (39028) across 5000 simulations under the catchment-by-catchment calibration method. (d) displays the simulated groundwater level time series compared with the observations from two wells, demonstrating cases with strong and weak Spearman correlation coefficients. Example groundwater timeseries shown for two wells at Old School House (GW well 2062) and East Wick Farm (GW well 859).

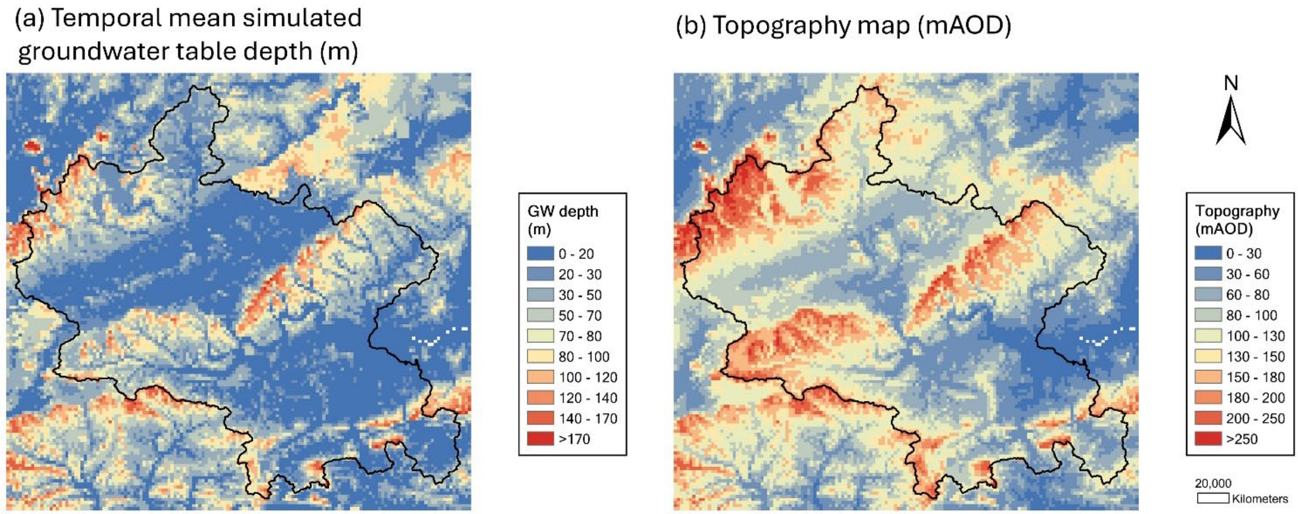


Figure 8: Spatial maps of simulated groundwater table depth for Thames at Kingston 39001. (a) Temporal mean over 1980-2020 of simulated groundwater table depth (difference between local topography and groundwater head in metres below land surface) in catchment 39001 under best catchment-by-catchment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

5 Discussion

5.1 Enhanced performance of DECIPHeR-GW in groundwater-dominant catchments

Based on the evaluation with 669 river flow gauges and 1804 groundwater monitoring sites across England and Wales, our coupled model DECIPHeR-GW v1.0 is able to produce robust streamflow simulations whilst capturing temporal dynamics of groundwater levels. Notably, the model achieves better performance in simulating river flows in groundwater-dominated catchments with baseflow index > 0.75 (Figure 5b), especially simulations for catchments with minor human influence showing significantly higher performance compared to DECIPHeR model. This improvement is most evident in the chalk regions with strong surface-groundwater water interactions, where it reproduces the observed hydrographs (examples in Figure 6e) and enhances hydrological simulation reliability. Moreover, the coupled model also performed well in other aquifer types including sandstone and limestone (Figure 6c and d). Although our coupled model is exhibiting similar or slightly better performance compared to the benchmark model in around 70% of the catchments, the coupled model has a more robust and reliable structure by better representing the groundwater processes. Herein, the coupled model could avoid the unrealistic model parameterisations to compensate for the absence of groundwater representations (Kirchner, 2006; Coxon et al., 2014; Dang et al., 2020). 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 ignores 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. Hence, the well-matched results for streamflow, parameter sensitivity and groundwater levels patterns show the potential of DECIPHeR-GW for future applications especially under climate change.

Additionally, DECIPHeR-GW v1.0 model facilitates a promising tool for water resources management in the southeast England, as existing hydrological models in the UK have faced challenges in accurately simulating streamflow and groundwater heads in these groundwater-dominated catchments. For instance, Lane et al. (2019) assessed four different conceptual hydrological models (TOPMODEL, ARNO/VIC, PRMS, SACRAMENTO) through the Framework for Understanding Structural Errors (FUSE) across over 1000 catchments in England, Wales and Scotland. Their findings revealed these models struggled with simulating biases, standard deviations, and correlations, particularly for the groundwater-dominated catchments in southeastern England. Similar issues have been reported with other models, including Grid-to-Grid (G2G) simulation over 61 Great Britain catchments (Rudd et al., 2017), GR4J application across 303 UK catchments (Smith et al., 2019), SHETRAN performance in 306 UK catchments (Seibert et al., 2018; Lewis, 2016) and SWAT simulation in two medium-scale catchments within the Thames River basin (Badjana et al., 2023). Efforts have been made to improve the groundwater representation in hydrological models like GR6J and PDM (Pushpalatha et al., 2011; Moore, 2007). Yet, models are still unable to accurately capture low flows in some groundwater-influenced catchments, such as those in the eastern Chilterns north of London (Hannaford et al., 2023). Even machine learning models like LSTM, while generally outperforming conceptual models, struggle to accurately simulate streamflow in the groundwater-dominated catchments (Lees et al., 2021). Moreover, most of these models mentioned above cannot simulate the timeseries of groundwater heads, at the same time as producing streamflow timeseries. In this study, our coupled model enables the simulation of inter-catchment subsurface flow and well captures the dynamic surface-groundwater interactions, providing a more precise representation of runoff and groundwater generation process in groundwater-dominated catchments. Consequently, the DECIPHeR-GW model shows potential for future applications, such as in low flow simulation and drought prediction, particularly in groundwater-dominated catchments.

Furthermore, our coupled model is relatively efficient in terms of computational requirements. One simulation over 51-year for the largest Thames at Kingston river basin (9948 km^2) with 27980 HRUs, takes approximately 17 hours to run on a standard CPU, producing simulated streamflow and groundwater level timeseries for all upstream 98 river gauges and 416 groundwater grids simultaneously. A 51-year simulation for the smallest river basin (10 km^2), with 52 HRUs and one river gauge, completes in about one second using a CPU. Future enhancements in computational efficiency of the coupled model can be achieved by employing sophisticated parallel computing techniques. 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.

5.2 Lessons learned from model coupling and ongoing developments

As awareness of the importance of groundwater process-based representation grows, along with the rapid development of groundwater models with a variety of complexity, there is a growing interest in incorporating the groundwater representations into hydrological or land surface models (Gleeson et al., 2021; De Graaf et al., 2017; Maxwell et al., 2015; Irvine et al., 2024; Ntona et al., 2022). When designing coupled models, balancing model complexity with computational efficiency is crucial (Condon et al., 2021; Barthel and Banzhaf, 2016; Henriksen et al., 2003). Therefore, we selected a computationally efficient 2D model (Rahman et al., 2023), which generally yields superior results. However, this model lacks the representation of river network and assumes groundwater above topography is directly discharged to the nearest river, leading to inaccuracies of capturing groundwater dynamics in some low-elevation areas where simulated groundwater levels stay at the surface (see example in supporting information Figure S15). In addition, to achieve a simpler and more efficient structure of the coupled model, we removed the unsaturated zone from the benchmark DECIPHeR model and directly replaced the saturated zone with the groundwater model. This approach is consistent with many existing coupled models that do not account for the unsaturated zone and generally provide robust simulations (Yang et al., 2017; Jing et al., 2018; Reinecke et al., 2019; Müller Schmied et al., 2014; Henriksen et al., 2003). According to our results, while this approach worked well in most catchments, the absence of an unsaturated zone led to flashier hydrographs in some wetter catchments, where the unsaturated zone is critical for storing excess rainfall (Dietrich et al., 2019; Hilberts et al., 2007). Thus, future research are advised to explore and design model structures tailored to their specific needs.

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).

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.

Our study demonstrates the robust performance of the DECIPHeR-GW model in simulating streamflow and groundwater head at a large scale across 669 catchments, highlighting its potential for widespread application in diverse geographical regions. While the model effectively captures natural surface-groundwater interactions, it falls short in accurately representing human influences, particularly in catchments affected by anthropogenic factors like surface/groundwater water abstraction and waste water returns (see example in Figure 6f). 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., 2021b; 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.

6 Conclusions

DECIPHeR-GW v1.0 is a new coupled surface-subsurface hydrological model that enhances the representation of surface-groundwater interactions and demonstrates good ability in simulating the streamflow and groundwater heads over large model domains. This paper introduces the details of the proposed model structures and its key components. We present an application in England and Wales, where previous hydrological models haven't captured surface-groundwater interactions and have shown poor performance in the south-east of England. Our evaluation against 669 river gauges and 1804 groundwater wells across England and Wales illustrates our coupled model performs well in streamflow simulation, achieving a median KGE of 0.83 across diverse catchments. Additionally, the model accurately captures the temporal patterns of groundwater level timeseries with approximately 56% of the wells showing a Spearman correlation coefficient of 0.6 or higher. More importantly, DECIPHeR-GW presents significant improved results in the drier natural chalk catchments of southeast England, where the average KGE increased from 0.49 in the benchmark DECIPHeR model to 0.7, facilitating a promising tool for water resources management in this region. DECIPHeR-GW is shown to be computationally efficient and capable of being calibrated and evaluated over large datasets of gauges. Being open-source and accompanied by a user manual, DECIPHeR-GW offers researchers an accessible implementation process and could be applied in other regions.

Code availability

The DECIPHeR-GW v1.0 model code (Zheng, 2024a), written in Fortran, is open-source and accessible at: https://github.com/YanchenZheng/DECIPHeR-GW_V1.0. A user manual to guide the researchers to use the model is also provided.

Data availability

The rainfall data (Hollis et al., 2019) is accessible from the CEDA archive (https://archive.ceda.ac.uk/), and the PET data (Robinson et al., 2023) is available from the CEH Environment Data Centre (https://catalogue.ceh.ac.uk/documents/9275ab7e-6e93-42bc-8e72-59c98d409deb). The daily streamflow timeseries are available from the NRFA website (https://nrfa.ceh.ac.uk/), while the groundwater timeseries data is available at https://environment.data.gov.uk/hydrology/explore (last access: 19th April 2023). Simulated flow, groundwater outputs and performance metrics (Zheng, 2024b) of the best model simulations (including both catchment-by-catchment and nationally-consistent calibration) from the DECIPHeR-GW v1.0 model are available at the University of Bristol data repository (https://data.bris.ac.uk/data/), at https://doi.org/10.5523/bris.wt0r1ec81zti2tww4p64fsqr3.

Author contributions

With guidance from GC, MR and RW, YZ led the coupling of the model, implementing the representation of surface-groundwater interactions, model simulations and results analysis. YZ initially drafted the manuscript with significant contributions from GC, MR and RW. SS helped with implementing the Multiscale Parameter Regionalisation (MPR) version of DECIPHeR, while YT provided the technical support on Fortran, set up the debugging mode and also ran simulations on the BC4 system. DW assisted with the selection of the study catchments and design of the output results. All co-authors contributed to the manuscript review and editing.

Competing interests

The authors have no competing interests to declare.

Acknowledgements

GC, DW and YZ were supported by a UKRI Future Leaders Fellowship award (MR/V022857/1). SS was funded by the NERC GW4+ Doctoral Training Partnership studentship from the Natural Environmental Research Council (NE/S007504/1), Wessex Water Ltd and the DAFNI Centre of Excellence for Resilient Infrastructure Analysis (ST/Y003713/1). We greatly appreciate the discussions with Rosanna Lane about the integration of Multiscale Parameter Regionalisation in DECIPHeR. We would like to thank Qidong Fang for his suggestions on the quality control for large sample of groundwater level observations. This work was performed using the computational and data storage facilities of the Advanced Computing Research Centre, i.e., BlueCrystal 4, University of Bristol - http://www.bristol.ac.uk/acrc/.

References

Aeschbach-Hertig, W. and Gleeson, T.: Regional strategies for the accelerating global problem of groundwater depletion, Nature Geoscience, 5, 853-861, 2012.

Ala-aho, P., Soulsby, C., Wang, H., and Tetzlaff, D.: Integrated surface-subsurface model to investigate the role of groundwater in headwater catchment runoff generation: A minimalist approach to parameterisation, Journal of Hydrology, 547, 664-677, 2017.

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.
Badjana, H. M., Cloke, H. L., Verhoef, A., Julich, S., Camargos, C., Collins, S., Macdonald, D. M. J., McGuire, P. C., and Clark, J.: Can hydrological models assess the impact of natural flood management in groundwater-dominated catchments?, Journal of Flood Risk Management, 16, e12912, https://doi.org/10.1111/jfr3.12912, 2023.
Bailey, R. T., Wible, T. C., Arabi, M., Records, R. M., and Ditty, J.: Assessing regional-scale spatio-temporal patterns of groundwater–surface water interactions using a coupled SWAT-MODFLOW model, Hydrological processes, 30, 4420-4433, 2016.
Barthel, R. and Banzhaf, S.: Groundwater and Surface Water Interaction at the Regional-scale – A Review with Focus on Regional Integrated Models, Water Resources Management, 30, 1-32, 10.1007/s11269-015-1163-z, 2016.
Batelis, S. C., Rahman, M., Kollet, S., Woods, R., and Rosolem, R.: Towards the representation of groundwater in the Joint UK Land Environment Simulator, Hydrological processes, 34, 2843-2863, 2020.
Bell, V. A., Kay, A. L., Jones, R. G., and Moore, R. J.: Development of a high resolution grid-based river flow model for use with regional climate model output, Hydrol. Earth Syst. Sci., 11, 532-549, 10.5194/hess-11-532-2007, 2007.
Benz, S. A., Irvine, D. J., Rau, G. C., Bayer, P., Menberg, K., Blum, P., Jamieson, R. C., Griebler, C., and Kurylyk, B. L.: Global groundwater warming due to climate change, Nature Geoscience, 1-7, 2024.
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.
Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., and Famiglietti, J. S.: Hyper-resolution global hydrological modelling: what is next? "Everywhere and locally relevant", Hydrological processes, 29, 310-320, 2015.
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.
Brunner, P. and Simmons, C. T.: HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model, Groundwater, 50, 170-176, https://doi.org/10.1111/j.1745-6584.2011.00882.x, 2012.
Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water resources research, 14, 601-604, 1978.
Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., Hooper, R. P., Kumar, M., Leung, L. R., and Mackay, D. S.: Improving the representation of hydrologic processes in Earth System Models, Water Resources Research, 51, 5929-5956, 2015.
Condon, L. E. and Maxwell, R. M.: Evaluating the relationship between topography and groundwater using outputs from a continental-scale integrated hydrology model, Water Resources Research, 51, 6602-6621, 2015.
Condon, L. E. and Maxwell, R. M.: Simulating the sensitivity of evapotranspiration and streamflow to large-scale groundwater depletion, Science advances, 5, eaav4574, 2019.
Condon, L. E., Kollet, S., Bierkens, M. F., Fogg, G. E., Maxwell, R. M., Hill, M. C., Fransen, H. J. H., Verhoef, A., Van Loon, A. F., and Sulis, M.: Global groundwater modeling and monitoring: Opportunities and challenges, Water Resources Research, 57, e2020WR029500, 2021.
Coxon, G., Freer, J., Wagener, T., Odoni, N. A., and Clark, M.: Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, Hydrological Processes, 28, 6135-6150, https://doi.org/10.1002/hyp.10096, 2014.
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.
Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth Syst. Sci. Data, 12, 2459-2483, 10.5194/essd-12-2459-2020, 2020.

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.

de Graaf, I. E., van Beek, R. L., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F.: A global-scale two-layer transient groundwater model: Development and application to groundwater depletion, Advances in water Resources, 102, 53-67, 2017.

de Graaf, I. E. M., van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Dynamic attribution of global water demand to surface water and groundwater resources: Effects of abstractions and return flows on river discharges, Advances in Water Resources, 64, 21-33, https://doi.org/10.1016/j.advwatres.2013.12.002, 2014.

Dietrich, O., Fahle, M., and Steidl, J.: The Role of the Unsaturated Zone for Rainwater Retention and Runoff at a Drained Wetland Site, Water, 11, 1404, 2019.

Dobson, B., Coxon, G., Freer, J., Gavin, H., Mortazavi-Naeini, M., and Hall, J. W.: The spatial dynamics of droughts and water scarcity in England and Wales, Water Resources Research, 56, e2020WR027187, 2020.

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.

EnvironmentAgency: Hydrology Data Explorer [dataset], https://environment.data.gov.uk/hydrology/explore, 2023.

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.

Famiglietti, J. and Wood, E. F.: Multiscale modeling of spatially variable water and energy balance processes, Water Resources Research, 30, 3061-3078, 1994.

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K., Syed, T. H., Swenson, S. C., de Linage, C. R., and Rodell, M.: Satellites measure recent rates of groundwater depletion in California's Central Valley, Geophysical Research Letters, 38, 2011.

Ferguson, I. M., Jefferson, J. L., Maxwell, R. M., and Kollet, S. J.: Effects of root water uptake formulation on simulated water and energy budgets at local and basin scales, Environmental Earth Sciences, 75, 316, 10.1007/s12665-015-5041-z, 2016.

Fleckenstein, J. H., Krause, S., Hannah, D. M., and Boano, F.: Groundwater-surface water interactions: New methods and models to improve understanding of processes and dynamics, Advances in Water Resources, 33, 1291-1295, https://doi.org/10.1016/j.advwatres.2010.09.011, 2010.

Flipo, N., Gallois, N., and Schuite, J.: Regional coupled surface–subsurface hydrological model fitting based on a spatially distributed minimalist reduction of frequency domain discharge data, Geoscientific Model Development, 16, 353-381, 2023.

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.

Giordano, M.: Global groundwater? Issues and solutions, Annual review of Environment and Resources, 34, 153-178, 2009.

Gleeson, T. and Richter, B.: How much groundwater can we pump and protect environmental flows through time? Presumptive standards for conjunctive management of aquifers and rivers, River research and applications, 34, 83-92, 2018.

Gleeson, T., Befus, K. M., Jasechko, S., Luijendijk, E., and Cardenas, M. B.: The global volume and distribution of modern groundwater, Nature Geoscience, 9, 161-167, 2016.

Gleeson, T., Wagener, T., Döll, P., Zipper, S. C., West, C., Wada, Y., Taylor, R., Scanlon, B., Rosolem, R., Rahman, S., Oshinlaja, N., Maxwell, R., Lo, M. H., Kim, H., Hill, M., Hartmann, A., Fogg, G., Famiglietti, J. S., Ducharne, A., de Graaf, I., Cuthbert, M., Condon, L., Bresciani, E., and Bierkens, M. F. P.: GMD perspective: The quest to improve the evaluation of groundwater representation in continental- to global-scale models, Geosci. Model Dev., 14, 7545-7571, 10.5194/gmd-14-7545-2021, 2021.

Gnann, S., Reinecke, R., Stein, L., Wada, Y., Thiery, W., Müller Schmied, H., Satoh, Y., Pokhrel, Y., Ostberg, S., Koutroulis, A., Hanasaki, N., Grillakis, M., Gosling, S. N., Burek, P., Bierkens, M. F. P., and Wagener, T.: Functional relationships reveal differences in the water cycle representation of global water models, Nature Water, 1, 1079-1090, 10.1038/s44221-023-00160-y, 2023.

Gorelick, S. M. and Zheng, C.: Global change and the groundwater management challenge, Water Resources Research, 51, 3031-3051, 2015.

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.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80-91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resources Research, 48, https://doi.org/10.1029/2011WR011044, 2012.

Hannaford, J., Mackay, J., Ascot, M., Bell, V., Chitson, T., Cole, S., Counsell, C., Durant, M., Facer-Childs, K., and Jackson, C.: Hydrological projections for the UK, based on UK Climate Projections 2018 (UKCP18) data, from the Enhanced Future Flows and Groundwater (eFLaG) project, NERC EDS Environmental Information Data Centre [data set], Environmental Information Data Centre [data set], 10, 2022.

Hannaford, J., Mackay, J. D., Ascott, M., Bell, V. A., Chitson, T., Cole, S., Counsell, C., Durant, M., Jackson, C. R., and Kay, A. L.: The enhanced future Flows and Groundwater dataset: development and evaluation of nationally consistent hydrological projections based on UKCP18, Earth System Science Data, 15, 2391-2415, 2023.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., and Weiler, M.: Karst water resources in a changing world: Review of hydrological modeling approaches, Reviews of Geophysics, 52, 218-242, https://doi.org/10.1002/2013RG000443, 2014.

Hellwig, J., de Graaf, I. E. M., Weiler, M., and Stahl, K.: Large-Scale Assessment of Delayed Groundwater Responses to Drought, Water Resources Research, 56, e2019WR025441, https://doi.org/10.1029/2019WR025441, 2020.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, Journal of Hydrology, 280, 52-71, https://doi.org/10.1016/S0022-1694(03)00186-0, 2003.

Hilberts, A. G. J., Troch, P. A., Paniconi, C., and Boll, J.: Low-dimensional modeling of hillslope subsurface flow: Relationship between rainfall, recharge, and unsaturated storage dynamics, Water Resources Research, 43, https://doi.org/10.1029/2006WR004964, 2007.

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.

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.

Irvine, D. J., Singha, K., Kurylyk, B. L., Briggs, M. A., Sebastian, Y., Tait, D. R., and Helton, A. M.: Groundwater-Surface water interactions research: Past trends and future directions, Journal of Hydrology, 644, 132061, https://doi.org/10.1016/j.jhydrol.2024.132061, 2024.

Jing, M., Heße, F., Kumar, R., Wang, W., Fischer, T., Walther, M., Zink, M., Zech, A., Samaniego, L., and Kolditz, O.: Improved regional-scale groundwater representation by the coupling of the mesoscale Hydrologic Model (mHM v5. 7) to the groundwater model OpenGeoSys (OGS), Geoscientific Model Development, 11, 1989-2007, 2018.

Jones, H., Morris, B., Cheney, C., Brewerton, L., Merrin, P., Lewis, M., MacDonald, A., Coleby, L., Talbot, J., and McKenzie, A.: The physical properties of minor aquifers in England and Wales. British Geological Survey Technical Report, WD/00/4. 234pp. Environment Agency R&D Publication 68, 2000.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, https://doi.org/10.1029/2005WR004362, 2006.

Kuang, X., Liu, J., Scanlon, B. R., Jiao, J. J., Jasechko, S., Lancia, M., Biskaborn, B. K., Wada, Y., Li, H., and Zeng, Z.: The changing nature of groundwater in the global water cycle, Science, 383, eadf0630, 2024.

Lane, R. A. and Kay, A. L.: Gridded simulations of available precipitation (rainfall + snowmelt) for Great Britain, developed from observed data (1961-2018) and climate projections (1980-2080), NERC EDS Environmental Information Data Centre [dataset], 2022.

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.

Lane, R. A., Coxon, G., Freer, J. E., Wagener, T., Johnes, P. J., Bloomfield, J. P., Greene, S., Macleod, C. J., and Reaney, S. M.: Benchmarking the predictive capability of hydrological models for river flow and flood peak predictions across over 1000 catchments in Great Britain, Hydrology and Earth System Sciences, 23, 4011-4032, 2019.

Lee, J. Y., Yi, M. J., Yoo, Y. K., Ahn, K. H., Kim, G. B., and Won, J. H.: A review of the national groundwater monitoring network in Korea, Hydrological Processes: An International Journal, 21, 907-919, 2007.

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., and Dadson, S. J.: Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models, Hydrol. Earth Syst. Sci., 25, 5517-5534, 10.5194/hess-25-5517-2021, 2021.

Lewis, E. A.: A robust multi-purpose hydrological model for Great Britain, Newcastle University, 2016.

Loaiciga, H. A. and Doh, R.: Groundwater for People and the Environment: A Globally Threatened Resource, Groundwater, 62, 332-340, 2024.

Marsh, T., Parry, S., Kendon, M., and Hannaford, J.: The 2010-12 drought and subsequent extensive flooding: a remarkable hydrological transformation, NERC/Centre for Ecology & Hydrology2013.

Massmann, C.: Identification of factors influencing hydrologic model performance using a top-down approach in a large number of U.S. catchments, Hydrological Processes, 34, 4-20, https://doi.org/10.1002/hyp.13566, 2020.

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.

McMillan, H. K., Booker, D. J., and Cattoën, C.: Validation of a national hydrological model, Journal of Hydrology, 541, 800-815, https://doi.org/10.1016/j.jhydrol.2016.07.043, 2016.

Miller, M. P., Buto, S. G., Susong, D. D., and Rumsey, C. A.: The importance of base flow in sustaining surface water flow in the Upper Colorado River Basin, Water Resources Research, 52, 3547-3562, 2016.

Moore, R. J.: The PDM rainfall-runoff model, Hydrol. Earth Syst. Sci., 11, 483-499, 10.5194/hess-11-483-2007, 2007.

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.

Ng, G.-H. C., Wickert, A. D., Somers, L. D., Saberi, L., Cronkite-Ratcliff, C., Niswonger, R. G., and McKenzie, J. M.: GSFLOW–GRASS v1. 0.0: GIS-enabled hydrologic modeling of coupled groundwater–surface-water systems, Geoscientific Model Development, 11, 4755-4777, 2018.

Ntona, M. M., Busico, G., Mastrocicco, M., and Kazakis, N.: Modeling groundwater and surface water interaction: An overview of current status and future challenges, Science of The Total Environment, 846, 157355, https://doi.org/10.1016/j.scitotenv.2022.157355, 2022.

Oldham, L. D., Freer, J., Coxon, G., Howden, N., Bloomfield, J. P., and Jackson, C.: Evidence-based requirements for perceptualising intercatchment groundwater flow in hydrological models, Hydrol. Earth Syst. Sci., 27, 761-781, 10.5194/hess-27-761-2023, 2023.

OrdnanceSurvey: OS Open Rivers [dataset], available at: https://www.data.gov.uk/dataset/dc29160b-b163-4c6e-8817-f313229bcc23/os-open-rivers, 2023.

Parkin, G., Birkinshaw, S., Younger, P., Rao, Z., and Kirk, S.: A numerical modelling and neural network approach to estimate the impact of groundwater abstractions on river flows, Journal of Hydrology, 339, 15-28, 2007.

Pool, S., Vis, M., and Seibert, J.: Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency, Hydrological Sciences Journal, 63, 1941-1953, 2018.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, Journal of Hydrology, 411, 66-76, https://doi.org/10.1016/j.jhydrol.2011.09.034, 2011.

Rahman, M., Pianosi, F., and Woods, R.: Simulating spatial variability of groundwater table in England and Wales, Hydrological Processes, 37, e14849, 2023.

Rama, F., Busico, G., Arumi, J. L., Kazakis, N., Colombani, N., Marfella, L., Hirata, R., Kruse, E. E., Sweeney, P., and Mastrocicco, M.: Assessment of intrinsic aquifer vulnerability at continental scale through a critical application of the drastic framework: The case of South America, Science of The Total Environment, 823, 153748, 2022.

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.

Rawls, W. J., Brakensiek, D. L., and Saxtonn, K.: Estimation of soil water properties, Transactions of the ASAE, 25, 1316-1320, 1982.

Reinecke, R., Foglia, L., Mehl, S., Trautmann, T., Cáceres, D., and Döll, P.: Challenges in developing a global gradient-based groundwater model (G 3 M v1. 0) for the integration into a global hydrological model, Geoscientific Model Development, 12, 2401-2418, 2019.

Robinson, E., Kay, A., Brown, M., Chapman, R., Bell, V., and Blyth, E.: Potential evapotranspiration derived from the UK Climate Projections 2018 Regional Climate Model ensemble 1980-2080 (Hydro-PE UKCP18 RCM), 2021.

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.

Rudd, A. C., Bell, V. A., and Kay, A. L.: National-scale analysis of simulated hydrological droughts (1891–2015), Journal of Hydrology, 550, 368-385, https://doi.org/10.1016/j.jhydrol.2017.05.018, 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.

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, EGUsphere, 2024, 1-29, 10.5194/egusphere-2024-326, 2024.

Seibert, J., Vis, M. J. P., Lewis, E., and van Meerveld, H. J.: Upper and lower benchmarks in hydrological modelling, Hydrological Processes, 32, 1120-1125, https://doi.org/10.1002/hyp.11476, 2018.

Shannon, S., Payne, A., Freer, J., Coxon, G., Kauzlaric, M., Kriegel, D., and Harrison, S.: A snow and glacier hydrological model for large catchments – case study for the Naryn River, central Asia, Hydrol. Earth Syst. Sci., 27, 453-480, 10.5194/hess-27-453-2023, 2023.

Siebert, S., Burke, J., Faures, J.-M., Frenken, K., Hoogeveen, J., Döll, P., and Portmann, F. T.: Groundwater use for irrigation– a global inventory, Hydrology and earth system sciences, 14, 1863-1880, 2010.

Smith, K. A., Barker, L. J., Tanguy, M., Parry, S., Harrigan, S., Legg, T. P., Prudhomme, C., and Hannaford, J.: A multi-objective ensemble approach to hydrological modelling in the UK: an application to historic drought reconstruction, Hydrol. Earth Syst. Sci., 23, 3247-3268, 10.5194/hess-23-3247-2019, 2019.

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.

Turner, S. W., Hejazi, M., Yonkofski, C., Kim, S. H., and Kyle, P.: Influence of groundwater extraction costs and resource depletion limits on simulated global nonrenewable water withdrawals over the twenty-first century, Earth's Future, 7, 123-135, 2019.

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.

Verkaik, J., Sutanudjaja, E. H., Oude Essink, G. H., Lin, H. X., and Bierkens, M. F.: GLOBGM v1. 0: a parallel implementation of a 30 arcsec PCR-GLOBWB-MODFLOW global-scale groundwater model, Geoscientific Model Development Discussions, 2022, 1-27, 2022.

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5, 15-40, 10.5194/esd-5-15-2014, 2014.

Wada, Y., van Beek, L. P. H., van Kempen, C. M., Reckman, J. W. T. M., Vasak, S., and Bierkens, M. F. P.: Global depletion of groundwater resources, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL044571, 2010.

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.

Wang-Erlandsson, L., Bastiaanssen, W. G., Gao, H., Jägermeyr, J., Senay, G. B., Van Dijk, A. I., Guerschman, J. P., Keys, P. W., Gordon, L. J., and Savenije, H. H.: Global root zone storage capacity from satellite-based evaporation, Hydrology and Earth System Sciences, 20, 1459-1481, 2016.

Wang, J., Jiang, Y., Wang, H., Huang, Q., and Deng, H.: Groundwater irrigation and management in northern China: status, trends, and challenges, International Journal of Water Resources Development, 2019.

Wendt, D. E., Van Loon, A. F., Scanlon, B. R., and Hannah, D. M.: Managed aquifer recharge as a drought mitigation strategy in heavily-stressed aquifers, Environmental Research Letters, 16, 014046, 2021a.

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, 2021b.

Westerhoff, R., White, P., and Miguez-Macho, G.: Application of an improved global-scale groundwater model for water table estimation across New Zealand, Hydrol. Earth Syst. Sci., 22, 6449-6472, 10.5194/hess-22-6449-2018, 2018.

White, E. K., Peterson, T. J., Costelloe, J., Western, A. W., and Carrara, E.: Can we manage groundwater? A method to determine the quantitative testability of groundwater management plans, Water Resources Research, 52, 4863-4882, https://doi.org/10.1002/2015WR018474, 2016.

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.

Xin, P., Wang, S. S. J., Shen, C., Zhang, Z., Lu, C., and Li, L.: Predictability and Quantification of Complex Groundwater Table Dynamics Driven by Irregular Surface Water Fluctuations, Water Resources Research, 54, 2436-2451, https://doi.org/10.1002/2017WR021761, 2018.

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.

Zheng, Y.: DECIPHeR-GW version 1.0 [Software], Zenodo, https://doi.org/10.5281/zenodo.14113870, 2024a.

Zheng, Y.: DECIPHeR-GW v1: A coupled hydrological model with improved representation of surface-groundwater interactions [dataset], https://doi.org/10.5523/bris.wt0r1ec81zti2tww4p64fsqr3, 2024b.