Supplement of Geosci. Model Dev., 18, 4247–4271, 2025
https://doi.org/10.5194/gmd-18-4247-2025-supplement
© Author(s) 2025. CC BY 4.0 License.

Geoscientific Model Development
Open Access
EGU

Supplement of

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

Yanchen Zheng et al.

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

The copyright of individual parts of the supplement might differ from the article licence.

Text S1. Summary of existing surface-groundwater coupled models

Table S1 summarizes a variety of current groundwater-surface water coupled models across different scales.

5

Table S1. Summary of existing surface-groundwater coupled models across different scales.

Models	Application domain	Coupling method	Model structures	Surface-groundwater interactions	Groundwater Recharge	Lateral flow	Groundwater head	Groundwater use	Evaluation against groundwater data	Model calibration	Computation time	Reference
SWAT-MODFLOW	semi-arid region, Sprague River Watershed (4100 km^2), Oregon, USA	Two way coupling	SWAT, HRU-based;MODFLOW, grid cells	Represented	pass SWAT HRU-calculated deep percolation as recharge	Represented	Represented	Not represented	Yes; compare with observed average groundwater discharge	Yes	The simulation requires approximately 11 h for the 34-year period simulation period for their study basin.	Bailey et al. (2016)
TopNet-GW	Pareora catchment, Canterbury region, New Zealand (530 km^2)	Groundwater is represented as an additional storage	Catchment scale; Add additional layers of groundwater stores	Represented	No	Not represented	Not represented	Not represented	No	Yes	Not mentioned	Yang et al. (2017); (Griffiths et al., 2023)
mHM-OGS	central European mesoscale river basin – Nägelstedt (850 km^2)	Two way coupling	Grid-based; OGS acts as a plug-in to the original mHM modeling framework	Not represented	mHM simulated recharge	Not represented	Represented	Not represented	Yes; match with long-term mean of observed groundwater levels from 19 wells.	Yes	Not mentioned	Jing et al. (2018)
CWatM-MODFLOW	Seewinkel region in Austria (573 km^2), Upper Bhima basin in India (basin area is 46000 km^2)	Two way coupling	Both grid-based.1km for CWatM, 100-250m for MODFLOW	Represented	CWatM simulates the groundwater recharge	Represented	Represented	Represented, compare the model with and without irrigation	Yes, compare against 62 and 351 boreholes for Seewinkel and Bhima	Yes	Only the GW model simulation for Bhima basin at a 100m resolution lasts around 2 d versus 9 h at a 250m resolution.	Guillaumot et al. (2022)
GSFLOW-GRASS/GSFLOW	3 catchments, areas range from 12.5 to 3723 km^2	Two way coupling	PRMS HRUs scale and grid-based MODFLOW	Represented, especially groundwater interacted with streams and lakes	Represented	Represented	Represented	Not represented	No	Uncalibrated for the test cases, but can be calibrated after proper setup.	The GSFLOW runtime is sensitive to the MODFLOW resolution rather than the surface domain resolution	Markstrom et al. (2008); (Ng et al., 2018)

Models	Application domain	Coupling method	Model structures	Surface-groundwater interactions	Groundwater Recharge	Lateral flow	Groundwater head	Groundwater use	Evaluation against groundwater data	Model calibration	Computation time	Reference
JULES-GFB	Two synthetic experiments and 6 neighbouring UK catchments	Two way coupling	Grid-based; JULES is the land surface model	Represented	Represented	Represented	Represented	Not represented	No	Yes	Not mentioned	Batelis et al. (2020)
SHETRAN	Lambourne catchment over a Chalk aquifer	Two way coupling	Grid-based;	Represented as a sine curve, with recharge seasonality, mean annual recharge serving as input variables	Represented	Not represented	Represented	Represented	No	Yes	Not mentioned	Ewen et al. (2000); (Parkin et al., 2007)
CLSM-TOPMODEL	The Somme River catchment in France (5566 km^2)	Groundwater is represented as a storage reservoir	Grid-based;	Not represented	Represented	Not represented	Not represented	Not represented	No	Yes, calibrated with observed streamflow	Not mentioned	Gascoin et al. (2009)
CaWaQS 3.02	The Seine River basin (~76000km^2)	GW is coupled as a module, based on the pseudo-3-D diffusivity equation	Grid-based;	Not represented	Represented	Not represented	Represented	Represented	Yes, compare with 340 groundwater head observation time series	Yes, calibrated in 384 river stations	Not mentioned	Flipo et al. (2023)
ORCHIDEE	Five major sub-basins in Amazon	One way coupling, groundwater is represented as a slow reservoir	Grid-based;	Not represented	Represented	Not represented	Not represented	Not represented	No	Yes	Not mentioned	Guimberteau et al. (2014)
HydroGeoSphere	A montane catchment (3.2 km^2) in the Scottish Highlands	Two way coupling	Grid-based;	Represented, the interaction is simulated by using a 2D and the 3D form of Richards's equation for variably saturated flow	Represented	Represented	Represented	Not represented	Yes	Yes, calibrated against streamflow, evapotranspiration and also groundwater levels	Calibration time for this case study is 2-3 days.	Brunner and Simmons (2012); (Ala-Aho et al., 2017)

Models	Application domain	Coupling method	Model structures	Surface-groundwater interactions	Groundwater Recharge	Lateral flow	Groundwater head	Groundwater use	Evaluation against groundwater data	Model calibration	Computation time	Reference
ParFlow-CLM	Continental scale, Europe domain at 3km resolution (1544*1592 grid cells)	Two way coupling	Grid-based; ParFlow is coupled to CLM	Represented	Represented	Represented	Represented	Not represented	Yes, compare the water table depth anomalies	No, uncalibrated	Not mentioned	Naz et al. (2022)
ParFlow	Continental US	Two way coupling	Grid-based; ParFlow is an integrated subsurface and surface hydrologic model	Represented	Represented	Represented	Represented	Not represented	Yes, compare simulated and observed hydraulic head	No, uncalibrated	The entire simulation utilized 2.5M core hours of compute time, less than 1 week of wall-clock time (~ 150 h)	Maxwell et al. (2015)
WaterGAP2.2	Global	Groundwater is represented as a storage	Grid-based;	Not represented	Represented, the groundwater recharge is calculated as a fraction of runoff from land	Not represented	Not represented	Represented	No	Yes, calibrated against mean annual discharge of the 1319 basins	Not mentioned	Müller Schmied et al. (2014)
WaterGAP2-G^3M	Global	One way coupling, the model can only simulate the GW discharge to surface water, but not recharge from surface water to GW.	Grid-based;	Represented	Not represented	Represented	Represented	Not represented	Yes	No, uncalibrated	G3M can compute the presented steady-state solution on a computer with four cores in about 30min.	Reinecke et al. (2019)
GLOBGM (PCR-GLOBWB-MODFLOW)	Global	One way coupling	Grid-based;	Represented	Represented	Represented	Represented	Represented	Yes	No, uncalibrated	The daily simulation for the years 1958-2015 requires around 4.5 months.	De Graaf et al. (2017); (Verkaik et al., 2022)

Text S2. The spatial mapping conversion of the state and flux variables across HRU scale and groundwater grid scale in the coupled model

Since the recharge is calculated at the HRU scale based on the position of groundwater head, the storage and the depth of root zone (Figure 2 and equation (2) of the main paper), the gridded groundwater head needs to be transferred from grid cell to HRU scale. We adopt the area-weighted method to convert this state variable groundwater head. In Figure S1, H_1, H_2,…, H_4 represents the groundwater heads at grid scale. The green shaded HRU is divided into four parts by these four groundwater grid cells. Each grid cell's groundwater head value was weighted by the proportion of its overlayed area within the HRU boundary ($\frac{A_i}{A_{HRU}}$). The HRU-scale groundwater head is then calculated as the sum of these area-weighted values ($\frac{A_i}{A_{HRU}} * H_i$) across all grid cells within the HRU. This HRU-scale groundwater head is subsequently used to compare with the topography and the bottom of the root zone for calculating the recharge.

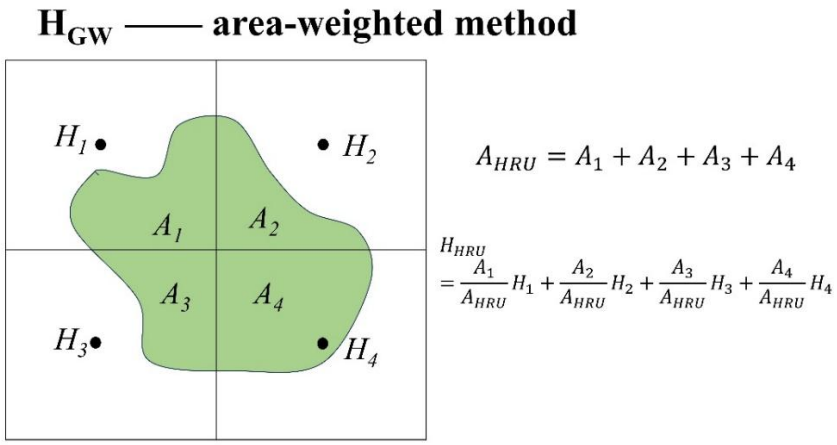

Figure S1: Schematic demonstrating the conversion of the state variables, i.e., groundwater heads (H_{GW}), from the groundwater grid scale to HRU scale. The grid cells are the groundwater model domain, while the green shaded area represents the HRU.

After the HRU-scale recharge is derived, this flux variable must be converted to the grid scale to serve as an input for the groundwater system component (Figure S2). We followed the principle of water volume conservation. The total recharge volume for each groundwater grid is obtained by summing the products of the recharge volume of each HRU ($A_{HRUi} * Q_{RCHRUi}$) and the proportion of its overlap with the groundwater grid ($\frac{A_i}{A_{HRUi}}$). This total recharge volume is then divided by the area of the groundwater grid (A_{GW}) to determine the recharge flux ($Q_{RC(GW)}$) for that grid.

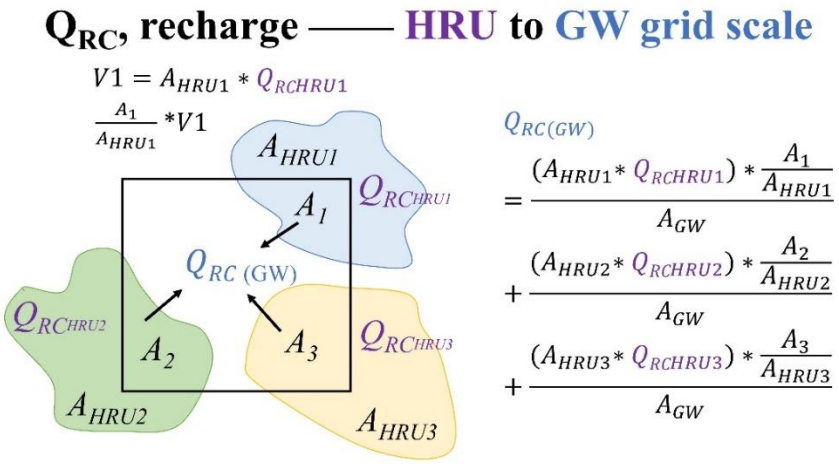

Figure S2: Simplified diagram for illustrating the conversion of HRU scale recharge flux to the groundwater grid scale.

Similarly, the grid-scale groundwater discharge is generated from the groundwater components, and this flux needs to be transferred to HRU scale for giving back to the surface water components for river routing. Figure S3 details our specific approach to achieve this conversion. To conserve mass, we assume that the groundwater discharge of each groundwater grid will be completely drained by the overlapping HRUs. If the groundwater grid is overlayed with multiple HRUs, the total groundwater discharge volume from the grid cell will be distributed among these HRUs according to their respective area proportions within the grid cell. Taking Figure S3 as an example, the groundwater grid 1 overlaps with two HRUs, with the total area of their overlap being the sum of $A_{1\text{-}1}$ and $A_{2\text{-}1}$. Thus, HRU1 can obtain the discharge contributed by groundwater grid 1 by multiplying the HRU1's overlayed area proportion ($\frac{A_{1-1}}{A_{1-1}+A_{2-1}}$) and the volume of groundwater discharge volume for grid 1 ($A_{GW} * Q_{GWDS1}$). As groundwater grid 2 overlaps exclusively with HRU1, all the groundwater discharge volume from this grid will be allocated to HRU1 ($A_{GW} * Q_{GWDS1} * \frac{A_{1-2}}{A_{1-2}}$). For HRU1, the HRU-scale groundwater discharge flux ($Q_{GWDS(HRU)}$) is determined by summing the proportionally distributed discharge amounts from all overlapping groundwater grid cells and then dividing by the area of the HRU1.

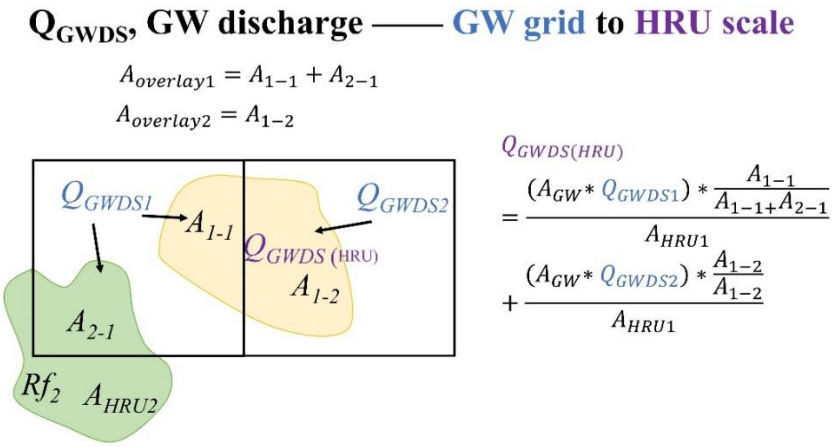

Figure S3: Diagram illustrating the conversion of groundwater discharge flux from grid scale to the HRU scale.

Text S3. Water balance checks

Three different categories of water balance checks were implemented to verify that mass is conserved in the coupled model. The first type of water balance checks were employed to ensure the correct transfer of recharge (Q_{RC}) and groundwater discharge (Q_{GWDS}) fluxes between the HRU scale and groundwater grid cells. To be specific, at each time step, we calculated and ensured that the relative error between the recharge volume obtained by multiplying the recharge value at the HRU-scale with the corresponding HRU area and the volume that are converted from the grid-scale recharge is less than 10^{-6}.

Secondly, for the surface component of the coupled model, we conducted the water balance checks for each HRU at every time steps, as shown in the Equation (1). The principle of Equation (1) is conservation of mass of the surface component. The water entering the surface component (via precipitation P) and stored in the catchment (initialized SRZ_{init}) must balance with the sum of the outputs, which include actual evapotranspiration (ET), recharge leaked to the groundwater system (Q_{RC}), surface excess flow (Q_{EX}) and water retained in the catchment at the last time step (SRZ_{end}).

Equation (2)-(3) present the water balance checks for the groundwater component of the coupled model. We adopted separate water balance check equations for surface and groundwater components because their corresponding fluxes and variables are at different scales (i.e., HRU-scale against grid-scale), and also at different domains (i.e., catchment domain and groundwater grid domain). The groundwater balance checks were performed to ensure that the accumulated recharge volume across all time steps minus the accumulated groundwater discharges volume equals the changes in groundwater heads. Since deriving the groundwater heads involve solving the large sets of algebraic equations, which usually uses the iterative techniques to obtain an approximation solution with specified tolerance (Reilly and Harbaugh, 2004). For example, we adopt the preconditioned conjugate gradient solver to compute the groundwater heads matrix, configured with 1500 iteration maximum number and a tolerance of 10^{-6}m. Careful mass balance check is critical to ensure the physical meaning

of the mathematical solution. In this study, the water balance error $error_{groundwater}$ was calculated by dividing the long-term groundwater balance deviations with the accumulated recharge volume. If this ratio is less than 1%, then we assumed that the groundwater component is water balanced.

$$wbal_{surface} = P - ET + (SRZ_{init} - SRZ_{end}) - Q_{RC} - Q_{EX}, \quad (1)$$

$$wbal_{groundwater} = \sum_{t=1}^{n} Q_{RC} * A_{GW} - \sum_{t=1}^{n} Q_{GWDS} * A_{GW} - (H_{end} - H_{init}) * S_y * A_{GW}, \quad (2)$$

$$error_{groundwater} = \frac{wbal_{groundwater}}{\sum_{t=1}^{n} Q_{RC}}, \quad (3)$$

where, $wbal_{surface}$ indicates the water balance deviations for the surface component, while $wbal_{groundwater}$ denotes the water balance deviations for the groundwater component. *P*, *ET* is the precipitation and actual evapotranspiration, respectively. SRZ_{init} and SRZ_{end} represents the root zone storage at the first and last time step. Q_{EX} is the excess flow to the river channel. Q_{RC} is the recharge value leaked to the groundwater system, whereas Q_{GWDS} denotes the groundwater discharges. *n* is the total time step. H_{init} and H_{end} represents the groundwater heads at the first and last time step. A_{GW} is the area of the groundwater simulation domain. S_y is the specific yield.

Text S4. Groundwater level data quality control process

Groundwater level observation data from 3,888 groundwater wells in England and Wales were collected from the Environment Agency's groundwater monitoring network database, encompassing various temporal resolutions, temporal coverages, and data quality considerations. There are 3005 wells providing manually measured data (referred to as 'Dipped data') at daily or monthly intervals, and 883 wells provide automatically 'Logged data' recorded by pressure transducers at sub-daily scales. Moreover, 395 wells have both types of data available. The locations of these wells are depicted in Figure S4a. The temporal coverage ranges from a minimum of 4 years with non-continuous observations to a maximum of 188 years, with a median of 41 years (Figure S4c). Several quality control steps have been applied to the measured groundwater level data before used for the model evaluation, the workflow is shown in the Figure S4b.

Step 1: Given that the data quality indicators are also available with the timeseries data in the database. Only data labelled with 'Good' and 'Estimated' are selected for analysis.

Step 2: We exclude the data with z-score > 3.5 to remove some outliers in the groundwater level timeseries.

Step 3: We prefer to use the manual measured Dipped data in the analysis and also because we run the model at the daily scale. In cases where both types of data are accessible, we prioritize the use of Dipped data. If only Logged data is available for the well, we aggregate the sub-daily data to the daily scale.

Step 4: Wells with records spanning shorter time periods are excluded. Records must contain over 100 measurements and span a minimum of 8 years to be included in the analysis.

Step 5: Only the groundwater wells intersecting with our selected study catchments are included.

As a result, daily groundwater level measurements from a total of 1804 groundwater wells were used for the model evaluation.

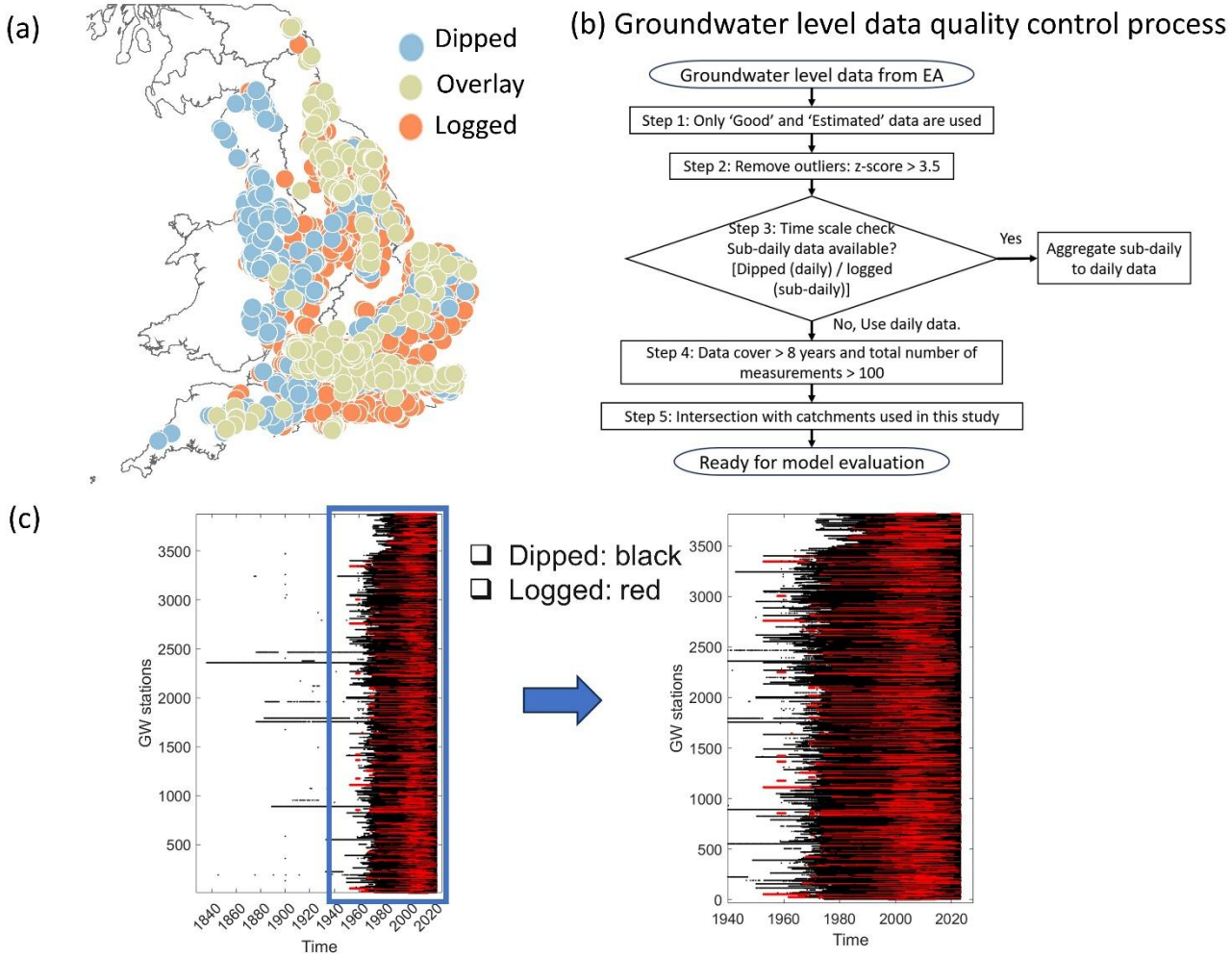

Figure S4: (a) Map of 3888 groundwater wells, (b) groundwater level observations quality control process and (c) the temporal coverage of the groundwater level data.

Text S5. Ks and B parameter range for different soil texture classes

Table S2 presents the lower and upper ranges for saturated hydraulic conductivity (Ks) and pore size distribution index (B) model parameters. These ranges are determined based on the studies from (Clapp and Hornberger, 1978; Rawls et al., 1982), which provides the field measured Ks and B soil sample values ranges for different soil texture classes.

Table S2. The lookup table of Ks and B parameter ranges for different soil texture classes.

No_soil_class	Pore size distribution index (B, -)		Saturated hydraulic conductivity (Ks, m/d)		Soil texture
	B_lower	B_upper	Ks_lower	Ks_upper	
1	0.01	1.93	0.504	15.206	Sand
2	0.01	2.27	0.147	13.507	Loamy sand
3	0.01	1.92	0.062	2.995	Sandy loam
4	0.01	1.72	0.016	0.622	Silt loam
5	0.01	1.79	0.032	0.600	Loam
6	0.01	1.38	0.010	0.544	Sandy clay loam
7	0.01	1.21	0.004	0.147	Silty clay loam
8	0.01	0.99	0.006	0.212	Clay loam
9	0.01	1.93	0.003	0.187	Sandy clay
10	0.01	0.77	0.002	0.089	Silty clay
11	0.01	0.90	0.001	0.111	Clay

Text S6. T and Sy parameter range for different lithological classes

The lower and upper bounds of Transmissivity (T) and specific yield (Sy) for different lithological classes in England and Wales are presented in the Table S3. These parameters ranges are determined based on the study from Rahman et al. (2023). To be specific, the lower and upper bounds of Transmissivity (T) are taken from their Table S1. In cases where certain lithology types lack specific ranges, the lower and upper bounds are calculated by dividing or multiplying the estimated value by 50%. As for the specific yield (Sy), the lower and upper bounds are determined by increasing or decreasing an order of magnitude from the estimated value. The distribution of all these 101 lithological classes is presented in Figure S5.

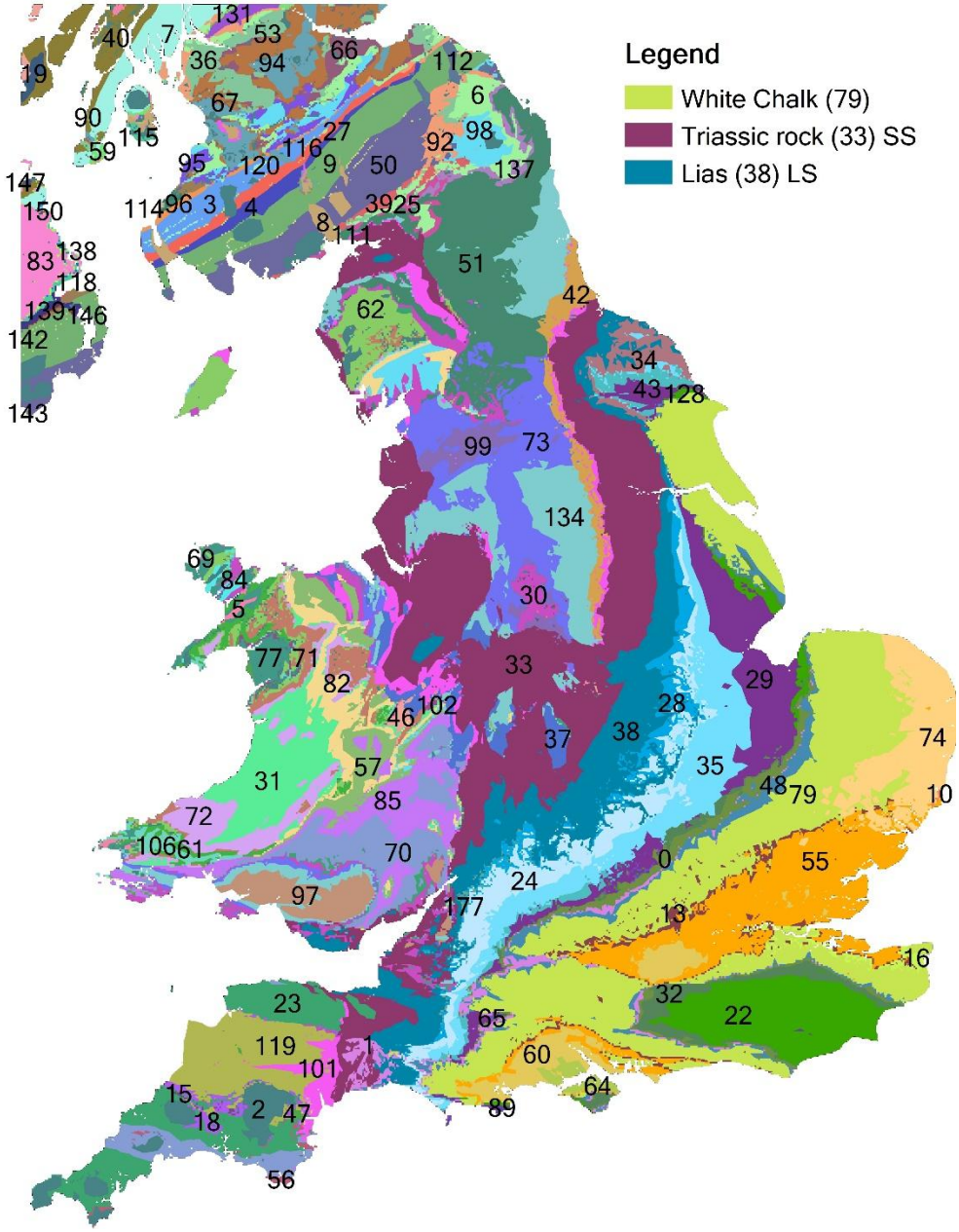

Figure S5. Hydrogeological properties lithology type map produced by the British Geological Survey. Each number on the map corresponds to a specific lithology type as detailed in Table S3.

Table S3. Transmissivity (T) and specific yield (Sy) lower and upper bounds for different lithological classes in England and Wales.

No.	Lithological class index in Figure S5	Transmissivity (T, m^2/d)		Specific yield (Sy, -)		Lithology name
		T_lower	T_upper	Sy_lower	Sy_upper	
1	0	25	75	0.00173	0.1	Gault clay
2	1	1.24	1565	0.003	0.1	Upper greensand
3	2	0.1	1	0.00004	0.004	Igneous
4	3	0.7	2.1	0.01	0.1	Kirkcolm
5	4	0.8	2.4	0.01	0.1	Shinnel Glenlee
6	5	0.8	2.4	0.01	0.1	Neoproterozoic
7	6	0.7	2.1	0.01	0.1	Inverclyde
8	7	0.1	1.4	0.0005	0.05	Southern highland
9	8	0.7	2.1	0.01	0.1	Stewertry
10	9	50	200	0.0002	0.02	Gala
11	10	390	600	0.0075	0.1	Neogene
12	101	2	50	0.014	0.1	Permian
13	13	40	700	0.0005	0.05	Lambeth group
14	15	0.2	45	0.006	0.1	Carboniferous
15	16	45	135	0.0001	0.01	Thanet sand
16	18	0.5	3.5	0.006	0.1	Teign valley
17	19	0.3	0.9	0.00004	0.004	Appin_group
18	22	25	75	0.000014	0.0014	Wealden group
19	23	10	51	0.005	0.1	Upper_devonian
20	24	105	1400	0.002	0.1	Great Oolite
21	25	10	1000	0.005	0.1	Border group
22	27	0.5	1.5	0.001	0.1	Crawford_moffat
23	28	3	11000	0.00005	0.005	Inferior Oolite
24	29	0.05	0.15	0.002	0.1	Ampthill clay
25	30	1	45	0.002	0.1	Dinantian_rock
26	31	1	3000	0.0008	0.08	Landovery
27	32	33	3400	0.00006	0.006	Lower greensand
28	33	1	275	0.001	0.1	Triassic_rock
29	34	8.5	25.5	0.02	0.1	Ravenscar
30	35	0.05	0.15	0.006	0.1	Oxford clay
31	36	1	45	0.001	0.1	Dinantian
32	37	150	450	0.000064	0.0064	Warwickshire
33	38	9	39	0.00015	0.015	Lias
34	39	5	15	8E-08	0.000008	Riccarton
35	40	5	15	8E-08	0.000008	Argyll

36	42	150	450	0.005	0.1	Zechstein
37	43	695	700	0.00005	0.005	Corallian
38	134	1.6	43	0.000004	0.0004	Pennine coal
39	46	0.3	1.1	0.0001	0.01	Metasedimentary
40	47	2	4000	0.0005	0.05	Ecocene
41	48	380	1500	0.0015	0.1	Grey chalk
42	98	10	51	0.005	0.1	Silurian
43	50	0.5	1.5	0.002	0.1	Hawick
44	51	1	3	0.018	0.1	Yoredale
45	53	25	75	0.018	0.1	Clackmannan
46	62	0.3	1.1	0.0005	0.05	Ordovician
47	55	10	522	0.0005	0.05	Thames
48	56	10	51	0.0005	0.05	Devonian
49	57	10	51	0.005	0.1	Ludlow
50	59	17.5	52.5	0.0002	0.02	Arbuthnott_Garvock
51	60	17	124	0.005	0.1	Bracklesham
52	61	0.3	1.1	0.0005	0.05	Llanvirn
53	64	18	775	0.005	0.1	Solent
54	65	7.5	22.5	0.000002	0.0002	Portland
55	66	400	1200	0.0002	0.02	Strathclyde
56	67	0.2	245	0.001	0.1	Silesian
57	69	2.5	150	0.001	0.1	Cambrian
58	70	25.5	76.5	0.001	0.1	Lower_Devonian
59	71	0.3	1.1	0.0005	0.05	Caradoc
60	72	0.3	1.1	0.0005	0.05	Ashgill
61	73	43	43	0.00002	0.002	Millstone grit
62	74	238	772	0.0075	0.1	Quaternary
63	77	45	135	0.0006	0.06	Upper_cambrian
64	79	380	1500	0.0015	0.1	White chalk
65	82	7.5	22.5	0.005	0.1	Wenlock
66	83	100	300	0.00001	0.001	Paleogene
67	84	0.7	2.1	0.01	0.1	Metamorphic
68	85	10	51	0.005	0.1	Pridoli
69	89	300	500	0.0003	0.03	Purbeck limestone
70	90	4	53	0.0005	0.05	New red sandstone
71	92	25	75	0.00005	0.005	Stratheden
72	94	250	750	0.01	0.1	Scottish_coal_measures
73	95	25.5	76.5	0.0005	0.05	Lanark
74	96	0.25	0.75	0.0005	0.05	Tappins
75	97	5.5	16.5	0.000007	0.0007	South Wales upper coal

76	99	40	120	0.002	0.1	Bowland High
77	102	0.3	1.1	0.0005	0.05	Tremadoc
78	106	0.25	0.75	0.0005	0.05	Arenig
79	109	0.25	0.75	0.0005	0.05	Highland_border_complex
80	111	25.5	76.5	0.0005	0.05	Appleby
81	112	25.5	76.5	0.0005	0.05	Reston
82	114	45	135	0.0006	0.06	Cambrian_Ordovician
83	115	4.59	75	0.0001	0.01	Mercia mudstone
84	116	0.25	0.75	0.0005	0.05	Patrick_Glenwhargen
85	118	0.25	0.75	0.0005	0.05	Leadhills
86	119	5	15	0.001	0.1	Holsworthy
87	120	0.7	2.1	0.01	0.1	Black craig
88	128	275	825	0.0015	0.1	Hunstanton
89	131	17	51	0.0002	0.02	Strahmore
90	137	49	157	0.01	0.1	Fell sandstone
91	138	33	3400	0.003	0.1	Hibernian greensand
92	139	93	410	0.00001	0.001	Sherwood sandstone
93	141	0.5	1.5	0.001	0.1	Armagh
94	142	0.5	1.5	0.001	0.1	Belfast
95	143	0.5	1.5	0.001	0.1	Tyrone
96	146	0.5	1.5	0.001	0.1	Enler
97	147	0.5	1.5	0.001	0.1	Leitrim
98	149	0.5	1.5	0.001	0.1	Holywood
99	150	0.5	1.5	0.001	0.1	Cross_slieve
100	151	0.5	1.5	0.001	0.1	Red bay
101	177	1.8	600	0.0003	0.03	Bridport sand

Text S7. Spatial maps of model performance

Figure S6 presents the spatial maps of model performance during calibration period using two calibration approaches. Spatial maps of coupled model performance for KGE correlation and flow duration curve (FDC) components are presented in Figure S7 under two calibration approaches (catchment-by-catchment and national consistency) for both the calibration and evaluation periods.

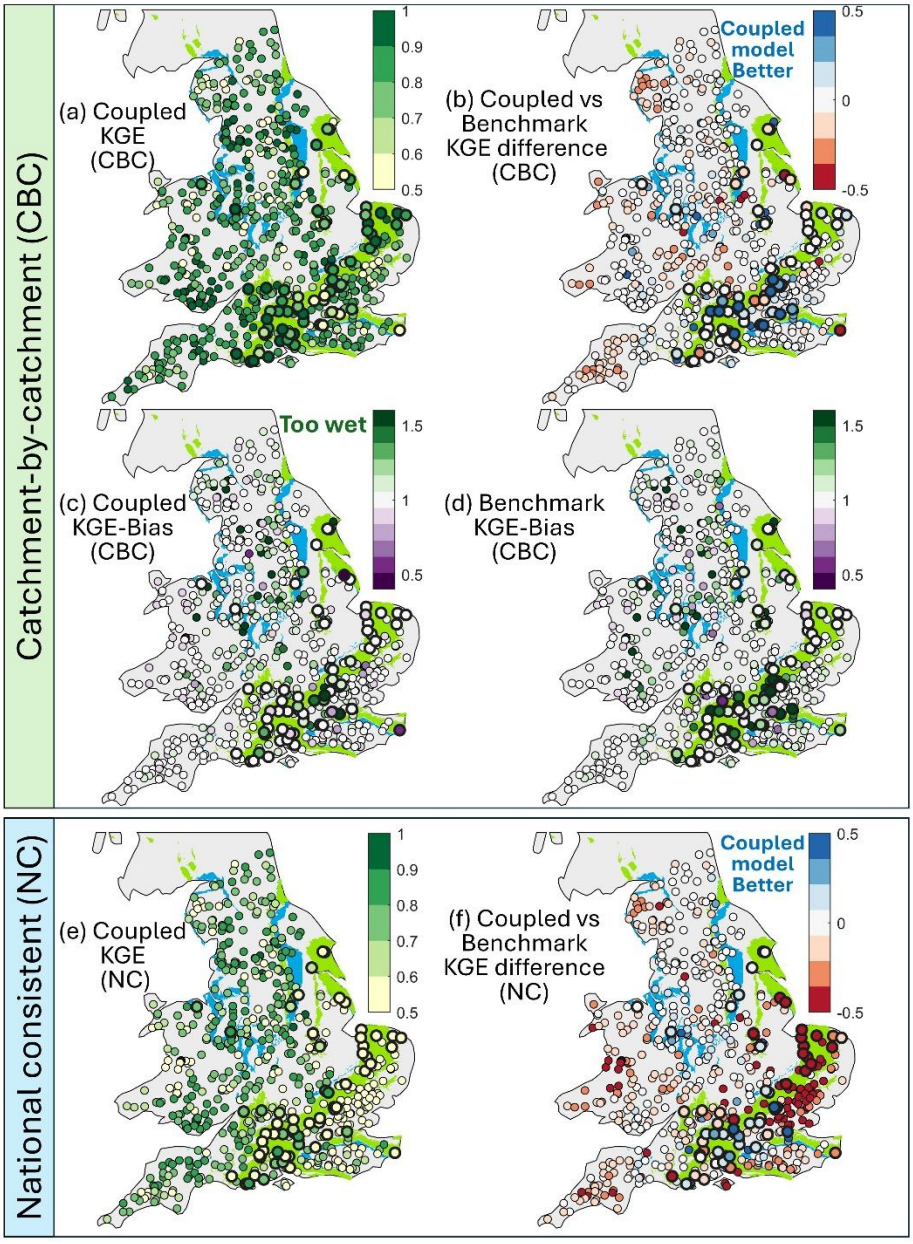

Figure S6: Spatial maps of model performance using two calibration approaches (a) Catchment-by-catchment (CBC), (d) national-consistent (NC), the non-parametric KGE differences between the coupled model and the corresponding DECIPHeR-MPR benchmark runs (b, e), and the bias component of KGE for the coupled model and benchmark runs under catchment-by-catchment approach (c, f). Each dot represents the performance at a river gauge during the calibration period. The scatter dots for groundwater-dominated catchments (baseflow index > 0.75) were outlined with thicker borders. The background of the maps highlights the areas of high productivity in aquifers. Light green represents the fractures flow highly productive aquifer, while blue indicates the intergranular flow highly productive aquifer.

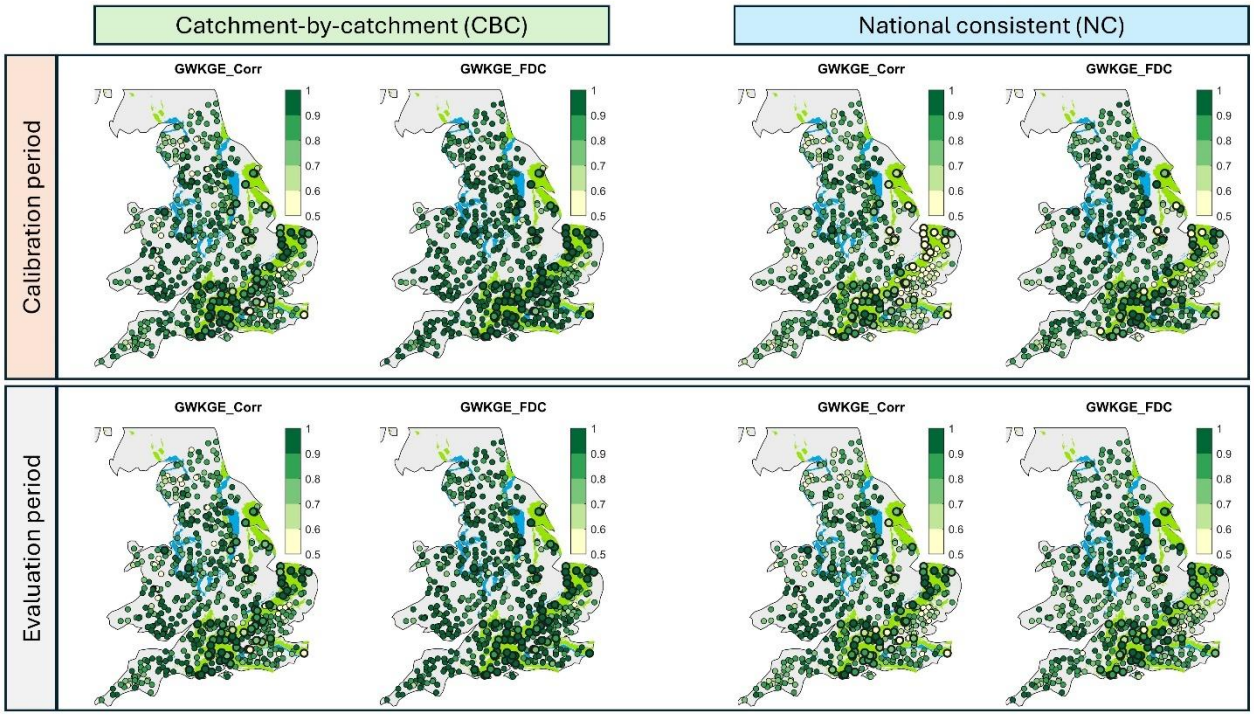

Figure S7: Spatial maps of the coupled model performance for KGE correlation and flow duration curve (FDC) component under two calibration approaches (catchment-by-catchment/national consistent) during calibration/evaluation period.

Text S8. Details of internal variables for the wet catchment 67005 during 2010-2012

Figure S8 shows the simulation results of the internal variable changes, including actual PET, root zone storage, recharge, saturated excess flow Q_{EX}, groundwater discharge Q_{GWDS} and also simulated groundwater heads and streamflow, for HRU 11 in catchment 67005 from 2010 to 2012. Based on Figure S8e, it can be observed that since the groundwater system is near saturation, the groundwater head remains above the bottom of the root zone, fluctuating between it and the topography surface. Therefore, under such conditions, no recharge occurs. As soon as the root zone becomes saturated, the saturated excess flow Q_{EX} is generated, which is the main component of the streamflow in this catchment (Figure S8f). According to Figure S8g, it is observed that the simulated hydrograph contains more pronounced spikes compared to the observations.

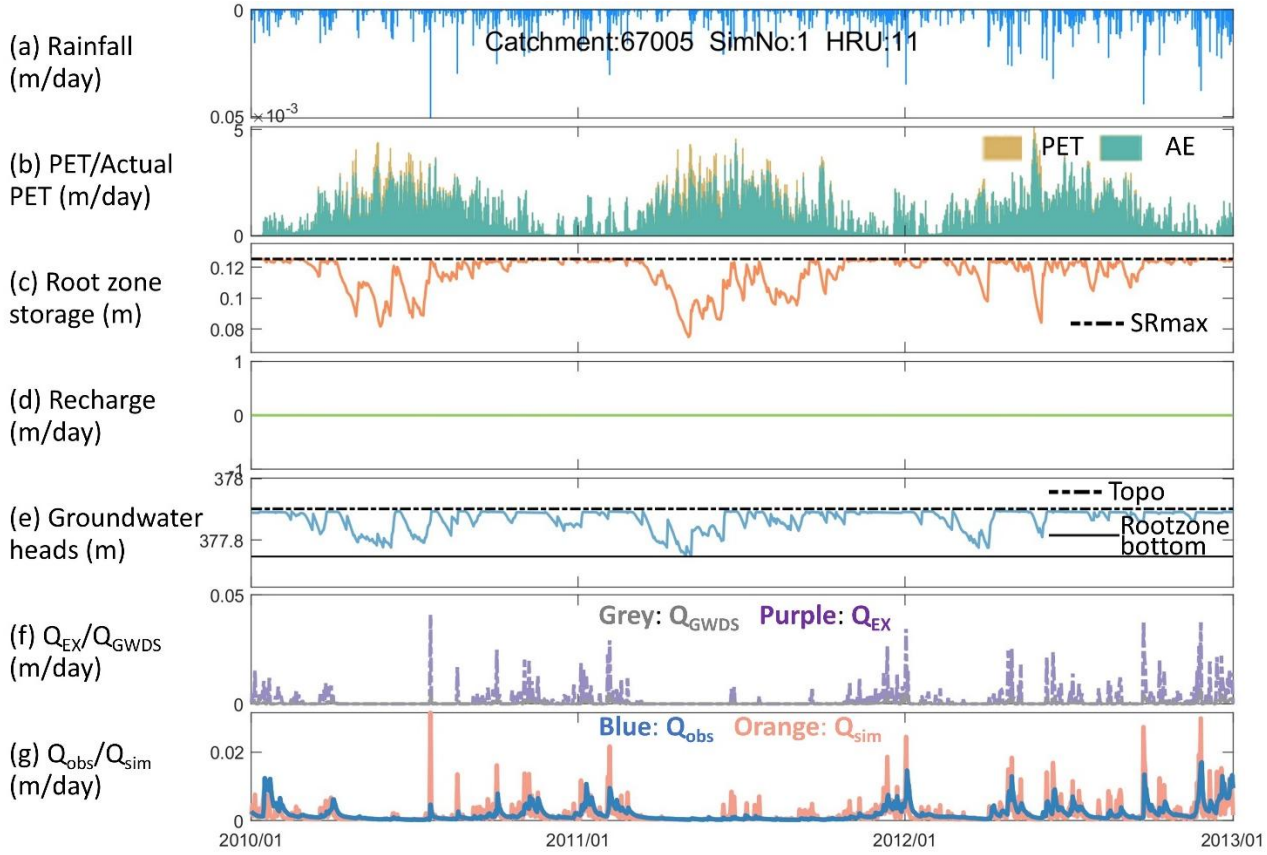

Figure S8: Internal variables for the wet catchment 67005 during 2010-2012.

Text S9 Investigation of factors explaining the poor performance of groundwater simulations

Figure S9 investigated six factors that might explain the poor performance of groundwater simulations, including short groundwater observation records, low streamflow accuracy in catchments, distance between wells from rivers or attributes such as borehole depth, elevation of well, and grid elevation. The results reveal that borehole depth (R=-0.25), river proximity (R=0.19), and streamflow accuracy (R=0.17) are relatively the key factors affecting the groundwater simulation with the p-value < 0.001. This indicates that the coupled model is more challenging to accurately simulate the temporal patterns of groundwater for wells with deeper boreholes, those located closer to the river, or where streamflow simulation accuracy is lower.

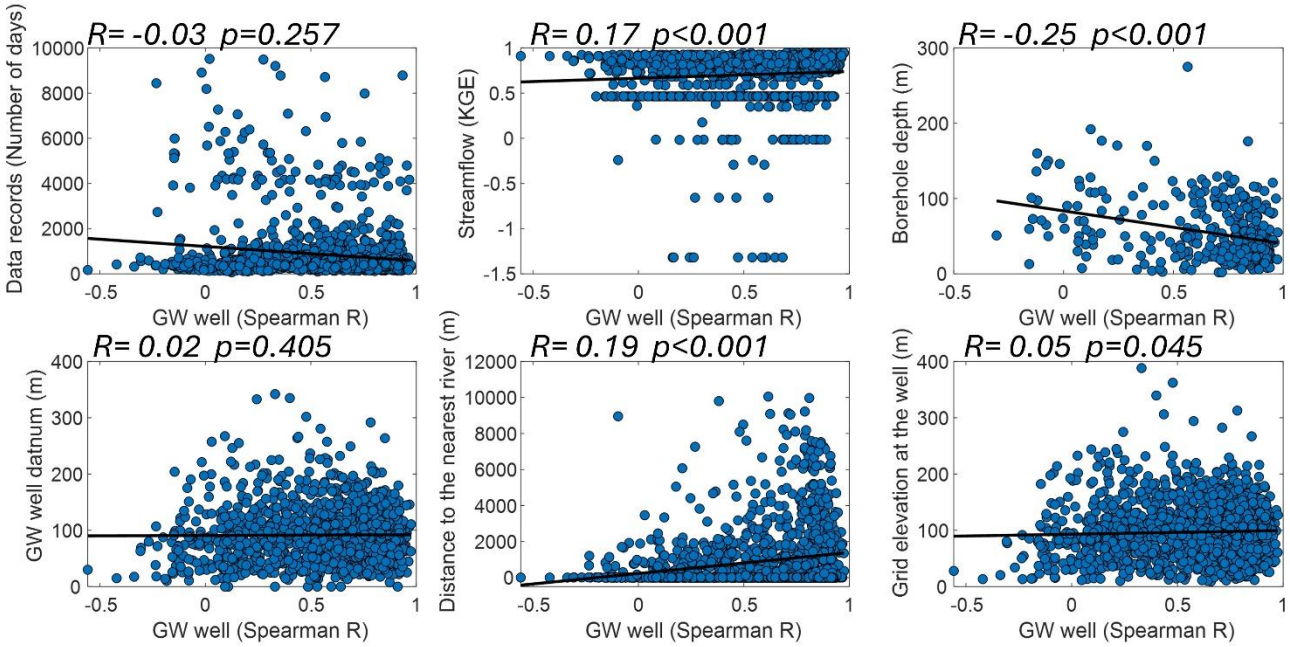

Figure S9: Factors explaining the poor performance of groundwater simulations. X-axis indicates the Spearman correlation coefficients calculated for comparing observed and simulated groundwater levels.

Text S10. Model parameter sensitivity analysis results

To further investigate the sensitivity of model parameters and their impact on simulated flows and groundwater levels, a controlled experiment was conducted in a groundwater-dominated chalk catchment as a case study (catchment ID: 39028, details provided in Table 4). For each model parameter, 1000 simulations were performed by sampling within the specified ranges for the selected parameter while keeping the other parameters at their default values. This approach facilitated a visual assessment of parameter sensitivity and enabled the verification of the model's functional performance, ensuring it operates in a physically meaningful manner.

The CHV parameter, channel routing velocity, showed insensitive in this relatively small catchment, which is consistent with previous studies (Coxon et al., 2019; Lane et al., 2021; Lane, 2021). Therefore, only the effects of the remaining five model parameters on the results are presented here. Overall, model parameters related to surface water components, such as SR_{max}, B, and Ks, exert a greater influence on simulated flows, whereas parameters like T and Sy have a more significant effect on simulated groundwater levels, as illustrated in Figures S10-S13.

SR_{max}, which governs 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. Conversely, a larger SR_{max}, which indicates a greater root

zone storage capacity, results in a smoother flow hydrograph. Both B and Ks are model parameters that control recharge magnitude, with smaller values of either parameter leading to reduced recharge. As shown in Figure S10b, smaller B values lead to reduced recharge, causing the root zone storage to fill up more quickly and resulting in increased overflow and greater variability in runoff. The performance of Ks and B is generally consistent (Figure S10b and c), though B has a relatively greater impact.

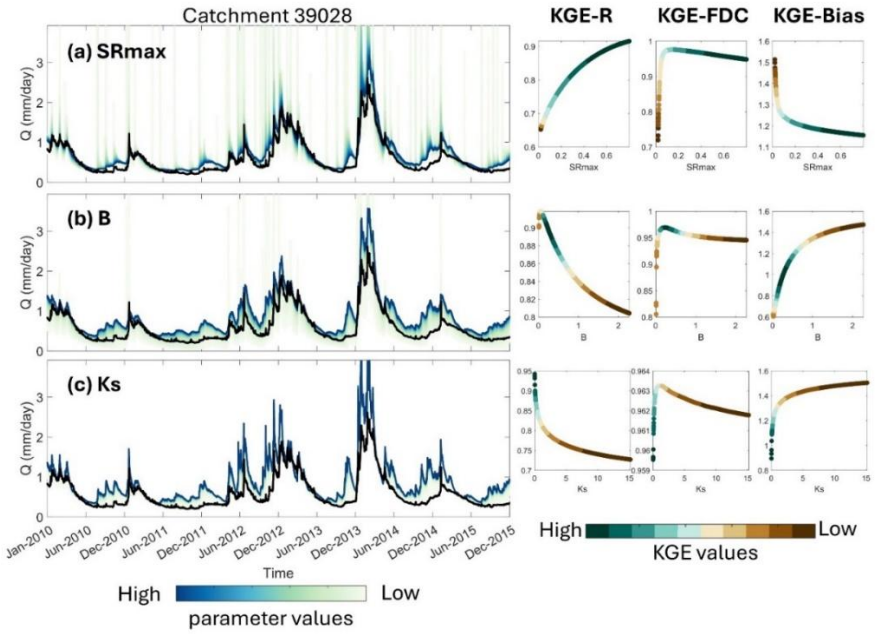

Figure S10: Demonstrating the influence of SRmax, B and Ks model parameters on the hydrograph through one-at-a-time sampling for the Dun at Hungerford (39028). In the left panel, each row illustrates the effects of varying one model parameter on hydrographs while all other parameters are held constant. Hydrographs are color-coded to represent the relative values of the varying parameters, ranging from low (green) to high (blue). In the right panel, the impact of parameter changes on KGE and its three components for simulated flows is illustrated. The effects of T and Sy parameters on the hydrograph are provided in the supporting information (Figure S11).

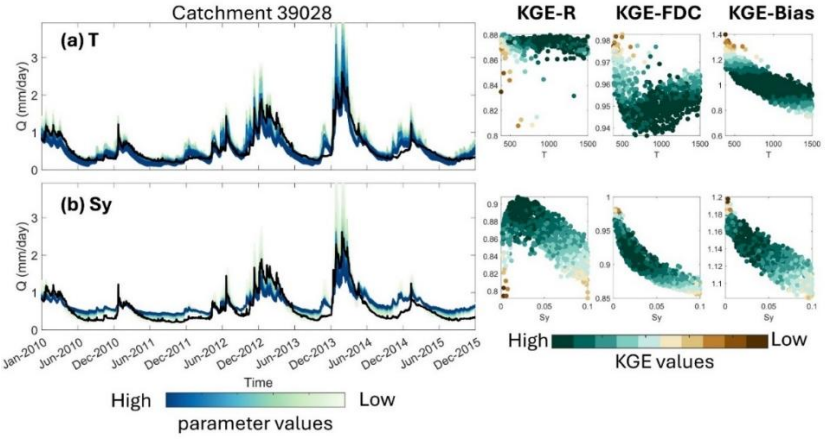

Figure S11: Demonstrating the influence of T and Sy model parameters on the hydrograph through one-at-a-time sampling for the Dun at Hungerford (39028). In the left panel, each row illustrates the effects of varying one model parameter on hydrographs while all other parameters are held constant. Hydrographs are color-coded to represent the relative values of the varying parameters, ranging from low (green) to high (blue). In the right panel, the impact of parameter changes on KGE and its three components for simulated flows is illustrated.

T and Sy are more crucial in controlling the simulation of groundwater levels (Figure S12). If the T is high, groundwater levels are generally lower. This is because a higher T facilitates easier movement of lateral groundwater flow through the aquifer, reducing the accumulation of water and consequently lowering the groundwater levels. Our results, which demonstrate the same pattern that higher T values reduce groundwater levels (Figure S12a), prove our model works effectively. Compared to T, the impact of Sy on groundwater levels is relatively weak. Figure S12b presents that smaller Sy values lead to higher variability in groundwater levels. When Sy is low, the regulation and storage capacity of the aquifer weaken, making groundwater levels more prone to fluctuations and thereby increasing their variability.

These results illustrate the effects of varying individual parameters on the simulation outputs, without accounting for interactions between parameters. This analysis aims to offer a general understanding of how these parameters influence model outputs and to verify that the coupled model is operating properly. Since the testing method is simple and only results for a single geological condition are presented here, the findings should not be overinterpreted. More detailed analysis of the model parameter sensitivity analysis is suggested in future research.

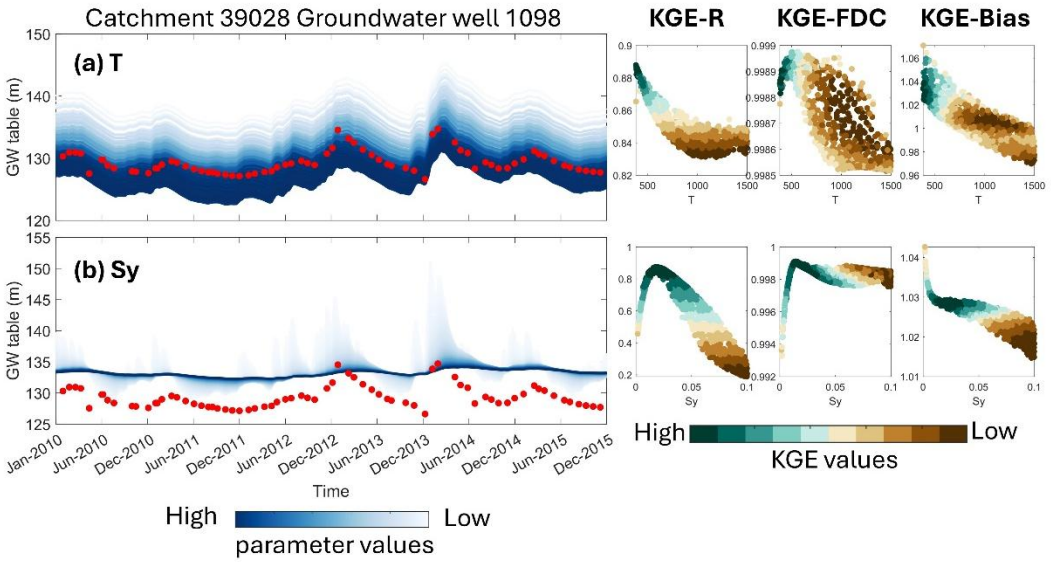

Figure S12: Demonstrating the influence of T and Sy model parameters on the simulated groundwater levels through one-at-a-time sampling for the Dun at Hungerford (39028). The left panel shows how varying each model parameter affects groundwater levels while keeping all other parameters constant. Groundwater levels are color-coded to reflect the relative values of the changing parameters, ranging from low (light blue) to high (dark blue). The right panel illustrates the impact of parameter variations on KGE and its three components for simulated groundwater levels. The effects of the SRmax, B, and Ks parameters on groundwater levels are detailed in the supporting information (Figure S13).

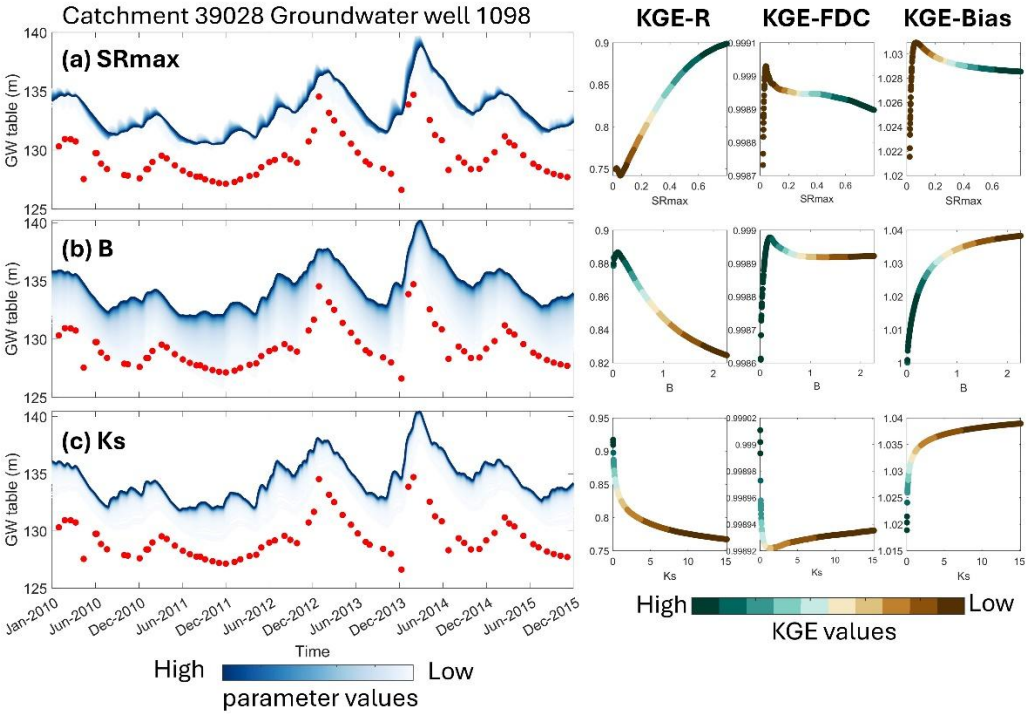

Figure S13: Demonstrating the influence of SRmax, B and Ks model parameters on the simulated groundwater levels through one-at-a-time sampling for the Dun at Hungerford (39028). The left panel shows how varying each model parameter affects groundwater levels while keeping all other parameters constant. Groundwater levels are color-coded to reflect the relative values of the changing parameters, ranging from low (light blue) to high (dark blue). The right panel illustrates the impact of parameter variations on KGE and its three components for simulated groundwater levels.

Text S11. Examples for determining the buffer zone threshold in this study

Figure S14 illustrates the impact of different buffer thresholds on the hydrographs for the 29 km^2 catchment 39049 as an example. Notably, when no buffer threshold is applied or when a large 10 km buffer zone is used, the simulations tend to either overestimate or underestimate the observed hydrographs, respectively. In both cases, the KGE performance is also poor (less than 0.6, as shown in Table S4). These results highlight the importance of setting an appropriate buffer zone. Without a buffer, the model is affected by the no-flow boundary condition, where no water can move across the boundary and leave the groundwater system, leading to the potential buildup of water in the adjacent cells of the lateral boundaries. Conversely, a large buffer zone, where no rainfall or recharge will be considered and generated in the domain, results in an underestimation of the catchment's runoff.

According to the example in Figure S14, as the buffer zone increases from 0 to 5 km, the KGE reaches its optimal value with a 3 km buffer, where the simulated hydrograph matches the observed data well. A 3km buffer zone ensures that the results

are less influenced by no-flow boundary conditions while avoiding an excessively large simulation domain, leading to more reasonable simulation outcomes. Additionally, users can adjust this buffer value as needed.

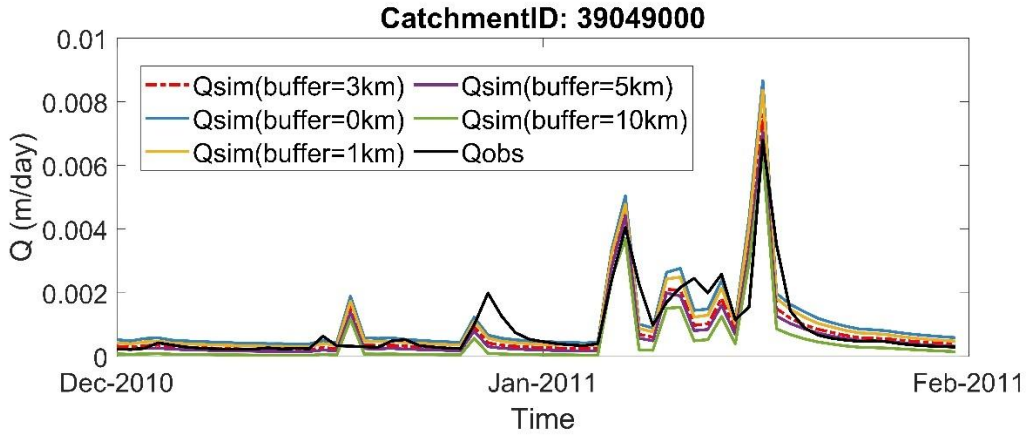

Figure S14: Demonstrating the influence of varying buffer thresholds on the hydrographs. The red dashed line shows the hydrograph with a 3 km buffer that were determined in this study.

Table S4. The KGE values for the simulations with different buffer thresholds

Catchment 39049	Buffer=3 km (used in this study)	Buffer=0 km	Buffer=1 km	Buffer=5 km	Buffer=10 km
KGE	0.73	0.53	0.64	0.71	0.57

Text S12. Examples of poor groundwater level simulations at low-elevation grids due to lack of river representation in groundwater model

Since our groundwater model does not have the represent river grids and assumes that whenever the groundwater level exceeds the topography surface, the portion of groundwater above the surface is converted into groundwater discharge and passed to nearby rivers. In some low-elevation groundwater grids, due to groundwater discharge generated at every time step and the excess water is transferred, causing the groundwater level in that grid to remain at the surface (Figure S15). This does not effectively simulate the dynamic changes in groundwater levels, which is a limitation of our model that needs to be further improved.

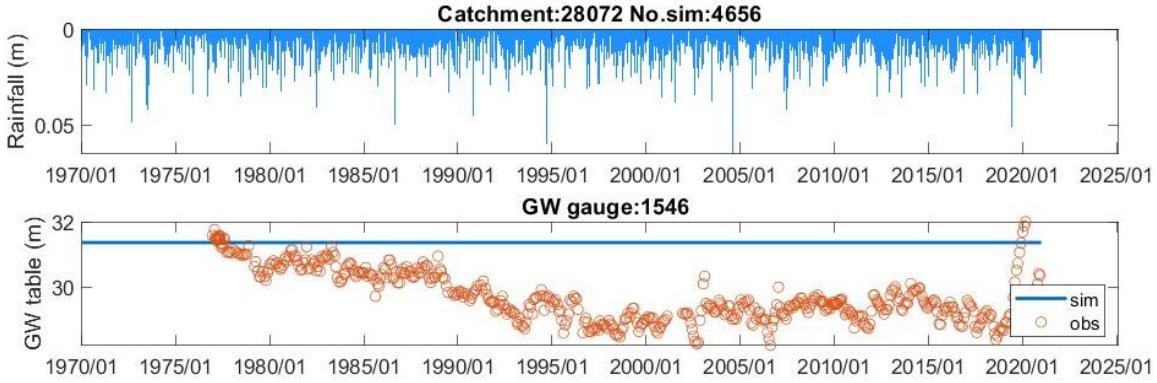

Figure S15: Examples of simulated groundwater levels stay at the topography surface in low-elevation grids.

References

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

Lane, R. A.: National-scale hydrological modelling of high flows across Great Britain: multi-model structures, regionalisation approaches and climate change analysis with uncertainty, University of Bristol, 2021.

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.

Markstrom, S. L., Niswonger, R. G., Regan, R. S., Prudic, D. E., and Barlow, P. M.: GSFLOW-Coupled Ground-water and Surface-water FLOW model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005), US Geological Survey techniques and methods, 6, 240, 2008.

Maxwell, R., Condon, L., and Kollet, S.: A high-resolution simulation of groundwater and surface water over most of the continental US with the integrated hydrologic model ParFlow v3, Geoscientific model development, 8, 923-937, 2015.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Döll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrol. Earth Syst. Sci., 18, 3511-3538, 10.5194/hess-18-3511-2014, 2014.

Naz, B. S., Sharples, W., Ma, Y., Goergen, K., and Kollet, S.: Continental-scale evaluation of a fully distributed coupled land surface and groundwater model ParFlow-CLM (v3. 6.0) over Europe, Geoscientific Model Development Discussions, 2022, 1-29, 2022.

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.

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.

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

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

Reilly, T. E. and Harbaugh, A. W.: Guidelines for evaluating ground-water flow models, DIANE Publishing2004.

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.

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.

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.