



Global hyper-resolution modeling of historical and future groundwater dynamics

Barry van Jaarsveld¹, Niko Wanders¹, Nicole Gyakowah Otoo¹, Edwin H. Sutanudjaja¹, Jarno Verkaik², Daniel Zamrsky^{1,3}, and Marc F. P. Bierkens¹

¹Department of Physical Geography, Utrecht University, Princeton laan 8a, Utrecht, the Netherlands

²Groundwater and Water Security, Deltares, Utrecht, the Netherlands

³Department of Water Resources and Ecosystems, IHE Delft Institute for Water Education, Delft, the Netherlands

Correspondence: Barry van Jaarsveld (a.s.vanjaarsveld@uu.nl)

Received: 10 February 2026 – Discussion started: 20 March 2026

Revised: 20 August 2026 – Accepted: 20 August 2026 – Published: 7 September 2026

Abstract. The sustainable management of global groundwater resources is a key societal challenge and is central to the Sustainable Development Goals. The localized dynamics of groundwater abstraction, topography, and surface-water interactions, as well as the sensitivity of groundwater-dependent ecosystems, call for high-resolution information to support effective groundwater management. At the same time, groundwater observations are very limited and concentrated in a few regions, rendering large parts of groundwater resources ungauged. To address limited observations and coarse global models, we applied the global groundwater model GLOBGM (v1.1) to simulate past and future groundwater heads and water table depth at 30 arcsec (~ 1 km) on a monthly time step. Model calibration improved mean bias in water table depth predictions from -4.8 to 3.6 m compared to GLOBGM v1.0, with depth-weighted bias reduced from 34.2 to 32.5 m across $34\,800$ observation wells. Groundwater dynamics are simulated for a historical reference period (1960–2019) to support model evaluation and attribution of observed impacts to climate variability and change. Baselines (1960–2014) and three combined socioeconomic-climate scenarios (2015–2100; SSP1-RCP2.6, SSP3-RCP7.0, SSP5-RCP8.5) are simulated with five global climate models, supporting detection and impact assessment of future change. Validation against monthly observations yielded skillful predictions ($KGE-NP_{skill}$) in approximately 60 % of deep wells (> 60 m) and 83 % of shallow to intermediate wells (0–60 m). When validated against annual observations, 66 % of deep wells (> 60 m) and 71 % of shallow to intermediate wells (0–60 m) were skillful. Simulations are further bias corrected using a machine learning approach. Historical trend analysis (1960–2019) accurately reproduced known groundwater depletion regions such as the U.S. High Plains, Arabian Peninsula, and Indo-Gangetic Plain, while also identifying rising water tables in northern latitudes and Arctic regions, which are linked to climate-driven recharge changes. Future scenario-based simulations suggest rising water tables for most continents in the next century, with Europe being a notable exception. However, known regions of groundwater depletion are expected to persist. Regions that disagree with observations or show reduced reliability are mapped, and quality assurance flags are provided to guide the appropriate use and interpretation of the results. The resulting data set offers high-resolution information to assess groundwater dynamics for the past and future, supporting improved global water resource management and climate impact assessments.

1 Introduction

Groundwater is the most abundant source of fresh liquid water; globally, 98 % of accessible fresh liquid water is in the form of groundwater (Döll et al., 2012; Gleeson et al., 2020). People increasingly rely on these resources to supplement surface water abstractions to meet total water demands (Custodio, 2010). Groundwater is vulnerable to overexploitation when abstraction rates persistently exceed natural groundwater recharge – the process by which surface water and precipitation infiltrate the soil to replenish aquifers. Driven by population growth and associated increases in water demand for agriculture, domestic, and industrial purposes, groundwater abstraction has increased significantly in recent decades, ultimately causing a decline in groundwater storage in many regions (Bierkens and Wada, 2019; Niazi et al., 2024; Wada and Bierkens, 2014; Kuang et al., 2024; Rodell et al., 2018; Döll et al., 2012; Leijnse et al., 2024). Besides anthropogenic pressures, groundwater is sensitive to decadal and annual climate variability and its influence on groundwater recharge (Taylor et al., 2013; Moeck et al., 2020b). These effects are complex, vary geographically, and are mediated by interactions between topography, vegetation, underlying geological properties, and surface water-groundwater exchange (Moeck et al., 2020b; Maples et al., 2019). For example, groundwater in deep confined aquifers with low permeability is relatively insensitive to changes in short-term recharge dynamics, whereas groundwater in highly permeable shallow aquifers with vegetation that facilitates exchanges across the land-atmosphere continuum is more tightly coupled to recharge variability (Istedt et al., 2016; West et al., 2022; Maples et al., 2019). Consequently, the future of groundwater storage depends on complex interactions among human water demands, climate, and ecosystems (Vörösmarty et al., 2000; Wada and Bierkens, 2014).

Overexploitation of groundwater causes water table depths to decline, a consequence of groundwater abstraction evident in many regions, especially in arid and dryland areas (Jasechko and Perrone, 2021; Jasechko et al., 2024). This reduces groundwater's reliability as a water source, given the physical and economic constraints associated with abstracting deeper reserves (Perrone and Jasechko, 2019). Since half of global river flows are groundwater-sustained, deepening water tables threaten downstream surface water supplies (de Graaf et al., 2024; Xie et al., 2024; Bierkens et al., 2021) and may increase vulnerability to hydrological drought (Wada et al., 2013; Vega-Briones et al., 2024). Reduced reserves also cause land subsidence, accelerate sea level rise, and degrade groundwater quality (Kuang et al., 2024; Bierkens and Wada, 2019). Deeper water table depths also influence the composition and functioning of groundwater-dependent ecosystems and the services they provide (Saccò et al., 2024; Fan et al., 2017; Otoo et al., 2025). Ultimately, groundwater forms an important component of global socio-ecological

systems and the Earth system as a whole (Huggins et al., 2023).

Sustainable management of the world's remaining groundwater reserves is a key societal challenge and is cited in several Sustainable Development Goals (Garrick et al., 2017; UN Water, 2022). Policymakers must implement adaptation and mitigation strategies to secure water availability by maintaining groundwater levels within safe operating limits. However, this requires reliable estimates of current and future groundwater volumes. Relying solely on observed groundwater data is insufficient given the severe limitations in global spatial and temporal coverage (Condon et al., 2021; UN Water, 2022). For example, the current state-of-the-art dataset contains 180 000 groundwater time series from 41 countries, with over 90 % of these time series occurring in North America, Australia, or Europe (Bäthge et al., 2026). To overcome this challenge and provide global-scale information, water managers rely on physically based numerical global groundwater models (Condon et al., 2021). However, until now, most groundwater flow models have had limited spatial extent, allowing regional-scale analyses ($\leq 1\,000\,000\text{ km}^2$) at most (Zamrsky et al., 2025).

Groundwater models of global extent were first formulated in the past decade (Fan et al., 2013), with more recent efforts enabling transient simulations (Reinecke et al., 2019; Costantini et al., 2023; de Graaf et al., 2019). However, these models provide information at spatial resolutions that are often too coarse to meet the demands of policymakers and stakeholders, who require finer-scaled information (Condon et al., 2021; Bierkens et al., 2015). To overcome this limitation, spatial resolution must be enhanced to make global models more locally relevant and actionable (Bierkens et al., 2015; Wood et al., 2011). To this end, hyper-resolution groundwater models have been highlighted for their potential to include climate, vegetation, and anthropogenic influences on groundwater at appropriate spatial resolutions for past and future (Bierkens et al., 2015). This is especially relevant, as hyper-resolution models can define current water scarcity hotspots, project future hotspots, and support the planning of adaptation strategies (Leijnse et al., 2024; Bierkens, 2015). Recent work demonstrates that global groundwater models at $\sim 1\text{ km}$ spatial resolution are possible with today's high performance computers (Verkaik et al., 2024; Kollet et al., 2026). This technical development offers the opportunity to generate global hyper-resolution estimates of past and future groundwater reserves at scales relevant to policymakers.

This study aims to produce the first hyper-resolution monthly estimates of global groundwater storage, incorporating the impacts of climate change and human activities across past and future periods. To this end, using scenarios and global climate models (GCMs) from the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP), monthly groundwater heads and water table depths are simulated at 30 arcsec ($\sim 1\text{ km}$ at the equator). We use the ISIMIP3a climate forcing to simulate groundwater reserves for the his-

torical reference period (1960–2019), which can be used for detection and attribution of past climate and socioeconomic change impacts on groundwater reserves. To project future trends, we employ ISIMIP3b inputs from a suite of five GCMs, covering both the baseline (1960–2014) and future periods (2015–2100) under combined socioeconomic and climate change scenarios (SSP1-RCP2.6, SSP3-RCP7.0, and SSP5-RCP8.5). These simulations are analyzed by calculating long-term trends, providing a global initial assessment of historical shifts in the global groundwater system and how it may change in the future.

2 Methods

2.1 The GLOBGM v1.1 model

Monthly estimates of groundwater heads and water table depths were simulated using GLOBGM v1.1, a global 30 arcsec, two-layer transient, MODFLOW-based groundwater flow model. GLOBGM is the successor to the 5 arcmin version originally developed by de Graaf et al. (2017). For brevity, we refer to Verkaik et al. (2024) for a detailed description of model inputs, parametrization, and run-time configurations. Briefly, GLOBGM is based on a prototype version of MODFLOW6, which is parallelized using message-passing interface and makes use of unstructured grids. Independent unstructured grids are distributed over three continental-scale groundwater models – Afro-Eurasia, the Americas, Australia – and one remaining model for the islands (Fig. 1). Each of the four groundwater models is partitioned into non-overlapping sub-models that are coupled within the MODFLOW linear solver. The two-layer configuration represents unconfined, confining, or confined aquifer systems (Verkaik et al., 2024). For cells where a confining layer is present, the top model layer represents the confining layer, and the bottom layer is the confined aquifer; in regions where a confining layer is absent, the bottom model layer represents the unconfined aquifer. GLOBGM employs an offline coupling approach whereby 5 arcmin resolution local runoff, groundwater abstraction, and groundwater recharge from PCR-GLOBWB2 (Sutanudjaja et al., 2018) are resampled to match GLOBGM's 30 arcsec grid.

GLOBGM v1.1 builds upon GLOBGM v1.0 (Verkaik et al., 2024), and in contrast to GLOBGM v1.0, a dynamic drainage elevation routine was implemented to better represent permanent wetland areas and groundwater-surface water exchange. To address a major source of uncertainty in global modeling, the updated version also incorporates a new downscaling method for groundwater recharge using observed data (Reinecke et al., 2024; Gnann et al., 2023; Bierkens and Wada, 2019). The treatment of river discharge and abstraction is consistent with GLOBGM v1.0. River discharge from PCR-GLOBWB2 was converted to 30 arcsec by accumulating it through the 30 arcsec HydroSHEDS river drainage network (Lehner et al., 2008; Verkaik et al., 2024), which is

then converted to surface water levels using Manning's equation. The 30 arcsec surface water levels thus obtained were subsequently used to parameterize the MODFLOW river and drainage boundary conditions; GLOBGM then computes groundwater-river exchange as a head-dependent flux based on simulated heads, stage, and conductance. Groundwater abstraction was resampled to 30 arcsec using a nearest neighbor algorithm.

2.1.1 Dynamic Drainage Elevation

In the dynamic drainage elevation routine, each cell's drain elevation adjusts based on soil saturation. As soils become wetter, the drainage elevation rises closer to the surface; as they dry, it lowers toward the base elevation. Discharge occurs only when the hydraulic head exceeds this dynamic drain level (Eq. 1). By coupling drain elevation to saturated area fractions, this approach maintains groundwater levels closer to the surface in wetland areas, thereby reducing unrealistic head fluctuations and highly intermittent discharge events. To simulate variations in the saturated area fraction, the improved ARNO scheme (Todini, 1996; Hagemann and Gates, 2003), an integral part of PCR-GLOBWB2, identifies areas subject to surface runoff.

$$f_{\text{sat}} = 1 - \left(\frac{W_{\text{max}} - W}{W_{\text{max}} - W_{\text{min}}} \right)^{\frac{\beta}{\beta+1}} \quad (1)$$

where W_{max} is the maximum water storage, W_{min} is the minimum water storage, W is the actual water storage, and β is a dimensionless shape parameter in PCR-GLOBWB2.

2.1.2 Groundwater Recharge Downscaling

Accurately predicting groundwater recharge remains a challenge for current global hydrological models (Reinecke et al., 2024; Gnann et al., 2023; Bierkens and Wada, 2019). Previous studies have shown that PCR-GLOBWB2, to which GLOBGM is coupled, tends to underestimate recharge rates compared with observations (Berghuijs et al., 2024). The bias is most pronounced in arid or semi-arid regions where recharge rates are low, intermittent, and unpredictable (Berghuijs et al., 2024). Furthermore, GLOBGM v1.0 predicts groundwater heads that are generally lower than in situ observations and other large-scale groundwater models (Reinecke et al., 2024; Verkaik et al., 2024). To address these challenges, recharge supplied to GLOBGM v1.1 was downscaled and corrected using observed groundwater recharge data.

A Generalized Additive regression approach using pyGAM (Servén et al., 2018) was employed to produce a long-term average 30 arcsec recharge field ($\text{GWR}_{\text{observed}}$) from known drivers and covariates (Table 1). This approach has previously been applied in groundwater modeling contexts (Mohan et al., 2018; Zamrsky et al., 2025). In contrast to these studies, land cover was not explicitly included

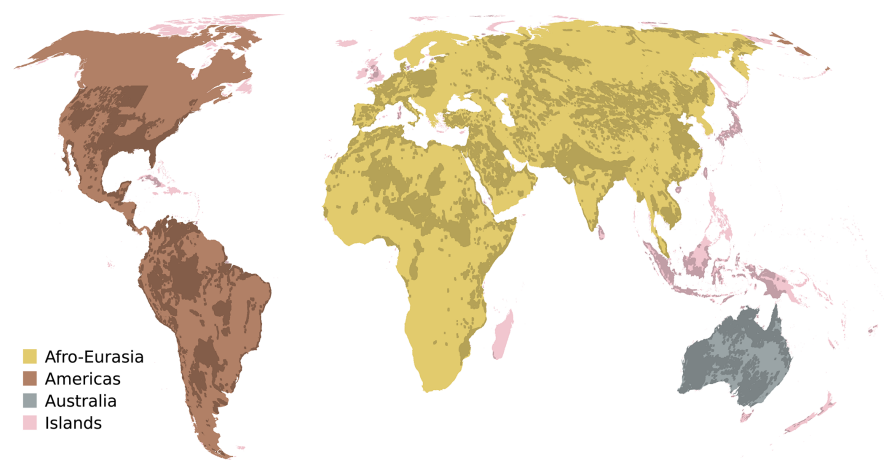


Figure 1. The four groundwater models that constitute GLOBGM, a global groundwater model. Dark overlays indicate the presence of a confining layer.

Table 1. Response and predictor variables used for groundwater recharge downscaling.

Type	Variable	Units	Resolution	Reference
Response	Observed Recharge	m yr^{-1}	Point Data	Moeck et al. (2020a)
Predictor	Aridity Index	–	30 arcsec	Zamrsky et al. (2024)
	Clay, Sand, & Silt Fraction	%	30 arcsec	Poggio et al. (2021)
	Potential Ref. Evaporation	m yr^{-1}	30 arcsec	Brun et al. (2022b)
	Precipitation	m yr^{-1}		
	Recession Coefficient	–	30 arcsec	Verkaik et al. (2024)
	Slope	–		
	Elevation	m		
	Hydraulic Conductivity	m d^{-1}		
	PCR-GLOBWB2 Recharge	m yr^{-1}	5 arcmin	Sutanudjaja (2024)

as a predictor here; instead, the 5 arcmin PCR-GLOBWB2 recharge was included to capture land cover-associated residuals by proxy. In Appendix A, we elaborate on the performance of the pyGAM model in predicting recharge. To prevent unrealistic recharge rates, $\text{GWR}_{\text{observed}}$ was further processed to ensure it did not exceed local precipitation rates (Brun et al., 2022a, b). Finally, to calculate a correction factor (GWR_{cf}), $\text{GWR}_{\text{observed}}$ was compared with PCR-GLOBWB2 5 arcmin recharge (bilinearly interpolated to 30 arcsec) from a GSWP3-W5E5 forced naturalized run (Eq. 2). For the steady-state simulation, the 5 arcmin PCR-GLOBWB2 recharge was interpolated to the 30 arcsec GLOBGM v1.1 grid via bicubic interpolation and corrected (Eq. 3). For the transient simulations, for each time step, PCR-GLOBWB2 was interpolated and corrected (Eq. 3) and, to avoid unrealistic estimates, recharge values were constrained by local precipitation rates (Eq. 4).

$$\text{GWR}_{\text{cf}} = \frac{\text{GWR}_{\text{observed}}}{\text{GWR}_{5\text{arc}'}}$$

(2)

where $\text{GWR}_{\text{observed}}$ is the observed recharge field and $\text{GWR}_{5\text{arc}'}$ is the 5 arcmin recharge from PCR-GLOBWB2 interpolated to 30 arcsec.

$$\text{GWR}_{\text{corrected}} = \text{GWR}_{5\text{arc}'} \times \text{GWR}_{\text{cf}}$$

(3)

where GWR_{cf} is the correction field and $\text{GWR}_{5\text{arc}'}$ is the 5 arcmin recharge from PCR-GLOBWB2 interpolated to 30 arcsec.

$$\text{GWR}_{\text{corrected}} = \begin{cases} P_{\text{month}} & \text{if } \text{GWR}_{\text{corrected}} > P_{\text{month}} \\ \text{GWR}_{\text{corrected}} & \text{if } \text{GWR}_{\text{corrected}} \leq P_{\text{month}} \end{cases}$$

(4)

where $\text{GWR}_{\text{corrected}}$ is the corrected recharge field and P_{month} is the local precipitation for the month.

A multiplicative correction factor was chosen to avoid negative recharge values. The multiplicative correction factor, fitted on long-term mean recharge estimates, produced unrealistically high recharge rates when applied directly at the

monthly time step during the transient simulation, particularly during wet months. To prevent the numerical instabilities and convergence failures that arose from these excessive fluxes, an upper bound was imposed on the corrected monthly recharge. Monthly precipitation was selected as this constraint, implicitly assuming that precipitation excess can be transferred to recharge within a single monthly time step. This assumption holds in many climatic settings but is conservative in arid environments, where dry-month precipitation is negligible and the correction is suppressed. An annual precipitation bound would relax this suppression in dry months, but permit monthly recharge fluxes that exceed physically plausible rates which would redistribute recharge in time rather than correcting it. We chose the monthly constraint to preserve numerical stability and plausible monthly fluxes, at the cost of dry-month recharge in water-limited regions.

2.2 Calibration, spin-up and initial conditions

GLOBGM v1.0 was uncalibrated, and initial conditions were calculated using a method inherited from its predecessor, PCR-GLOBWB2 (Verkaik et al., 2024; de Graaf et al., 2019). This process involved simulating steady-state groundwater heads under naturalized conditions, followed by a transient spin-up period in which the first year of forcing was repeated back-to-back. However, uncalibrated models prove less accurate than calibrated counterparts, such as the CONUS groundwater and global-scale inverse models (Verkaik et al., 2024; Zell and Sanford, 2020; Fan et al., 2013). Furthermore, spurious trends were evident in GLOBGM v1.0 simulations (Verkaik et al., 2024). To address these limitations, GLOBGM v1.1 presented here was calibrated, and an alternative spin-up procedure was implemented.

2.2.1 Calibration

To improve the accuracy of groundwater heads and water table depths simulated by GLOBGM v1.1, hydraulic conductivity, anisotropy, and entrance resistance parameters were calibrated using the steady-state solution. For anisotropy and entrance resistance, a range of global prefactor adjustments was assessed (Table 2). While a broader range of prefactors was possible, adjustments were limited to within two orders of magnitude. This constraint prevented significant departures from the original model parameterization, which was necessary given the spatially concentrated distribution of observations. Previous calibration attempts on the 5 arcmin predecessor reported limited sensitivity to variations in hydraulic conductivity when applied as a global prefactor to the entire domain (Otoo et al., 2025). Therefore, to incorporate regional sensitivities to changes in hydraulic conductivity, different combinations of prefactors were assessed (Table 2), which varied by lithological class (Hartmann and Moosdorf, 2012). In total, 162 permutations were assessed.

Table 2. Parameters and prefactors used during calibration. Lithologies are from the GLiM dataset (Hartmann and Moosdorf, 2012).

Parameter	Lithology Class	Candidate values
Ks	Fine & coarse unconsolidated	0.1–1.0–10.0
	Fine & coarse	0.1–1.0–10.0
	Carbonate	0.1–1.0–10.0
Anisotropy	Consistent across all lithological classes	0.1–1.0
Entrance resistance	Consistent across all lithological classes	0.1–1.0–10.0

Candidate calibration values were assessed by running the steady-state solution for the historical reference simulation (Sect. 2.3) under pristine natural conditions (i.e., unaltered by anthropogenic influence on surface water and groundwater dynamics). Note that updates regarding groundwater recharge downscaling and saturated area fraction adjustments were included prior to calibration (Sect. 2.1.1 & 2.1.2). To select optimal calibration settings, long-term averaged groundwater table depths from observations that overlapped with the forcing period (i.e., 1960–2019) were used, obtained from IGRAC’s Global Groundwater Monitoring Network (IGRAC, 2024). Given that the steady-state solution of GLOBGM v1.1 was forced with data representative of the hydrological cycle in a pristine state, observations were filtered to exclude regions with substantial groundwater abstraction. Specifically, observations were filtered to exclude known hotspots for water scarcity (Leijnse et al., 2024), as these regions likely experience extensive groundwater abstraction. Including observations from groundwater abstraction hotspots would confound the effects of groundwater abstraction with model error and skew calibration toward correcting for depletion. Using the filtered dataset of 34 800 observation wells, calibration permutations were evaluated by the mean bias over all locations (Bias_{WTD}; Eq. 5).

$$\text{Bias}_{\text{WTD}} = \text{WTD}_{\text{obs}} - \text{WTD}_{\text{sim}} \tag{5}$$

where WTD_{sim} is the simulated water table depth and WTD_{obs} is the observed water table depths.

To enable a direct comparison between model simulations and observations, we used the water table depth outputs from GLOBGM v1.1, derived from heads and surface elevation. Note that due to this conversion and Eq. (5), a positive bias indicates that simulated water table depths are shallower than observed, whereas a negative bias indicates that they are deeper. For unconfined aquifers, where no confining layer is present, observations were assigned to the bottom model layer. Where a confining layer was present in the model, observations were assigned to the correct model layer based on the reported filter. If filter depth was not reported,

model layers were assigned using an approach that evaluates the correlation between remotely sensed soil moisture data and simulated groundwater heads, as done by Verkaik et al. (2024). More specifically, if the observed groundwater time series, when expressed as groundwater heads, correlated well (> 0.5) with soil moisture (ESA CCI SM; Dorigo et al., 2017), it was assigned to the upper model layer (i.e., confining layer); if no correlation was found, it was assigned to the bottom model layer (i.e., confined aquifer).

We selected the optimal calibration set by prioritizing the accuracy of shallower estimates (< 60 m), as these reserves are more relevant for human use and for sustaining ecosystems through baseflow contributions to surface water bodies and groundwater-dependent ecosystems (de Graaf et al., 2019; Reinecke et al., 2024; Fan et al., 2013). Mean absolute bias per depth class (< 0 , 0–5, 5–10, 10–20, 20–60, > 60 m) was calculated. The mean of the per-class absolute bias values was used as the overall performance score for each permutation. Calibration permutations were rank ordered by average depth weighted absolute bias, and the top ten were identified.

2.2.2 Spin-up and initial conditions

We aimed to derive a set of initial groundwater head states for the simulation start year (i.e., 1960), such that initial conditions resemble the observed data available for 1960. First, the steady-state estimates of groundwater heads under pristine natural conditions were obtained. These steady-state results served as the starting point for a transient simulation for the year 1960, without abstraction, for 75 iterations back-to-back. Subsequently, a second transient simulation for the year 1960, with abstraction activated, was conducted for 50 iterations, allowing simulated groundwater levels to approach observed levels. For this spin-up protocol, GLOBGM v1.1 was forced with data relevant to the historical reference simulation described in Sect. 2.3. Similar to the calibration methodology, observed water table depth data were obtained from IGRAC's Global Groundwater Monitoring Network. Observations were filtered for the year 1960 (375 observation wells). At the end of each iteration, Bias_{WTD} (Eq. 5) and percentage bias (PBIAS, Eq. 6) were calculated. Simulated and observed water table depths were assigned to the correct model layer as done in Sect. 2.2.1. Our goal was to minimize the difference between modeled and observed groundwater depths of the starting year (i.e., 1960), using the convergence of Bias_{WTD} and PBIAS as indicators of stability. To assess the evolution of head estimates over the model domain, the average groundwater head for each sub-model was calculated (HDS; Eq. 7).

$$\text{PBIAS}_{\text{WTD}} = \frac{\sum (\text{WTD}_{\text{sim}} - \text{WTD}_{\text{obs}})}{\sum \text{WTD}_{\text{obs}}} \times 100 \quad (6)$$

where WTD_{sim} and WTD_{obs} denote the simulated and observed water table depths, respectively.

$$\overline{\text{HDS}} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \text{HDS}_{i,j} \quad (7)$$

where $\text{HDS}_{i,j}$ is the groundwater head at the grid cell in row i and column j , n and m are the numbers of rows and columns, respectively.

Achieving absolute dynamic equilibrium in a global groundwater model at the hyper-resolution is computationally prohibitive and would require thousands of iterations (Verkaik et al., 2024). Furthermore, given the significant structural uncertainties in global groundwater models, further refinement of the initialization is expected to provide negligible benefit (Reinecke et al., 2024). Therefore, groundwater heads estimated at the final iteration were used as initial conditions for all subsequent simulations presented herein. To account for potential spin-up issues in the final dataset, a procedure was developed to identify locations and time periods where the model is likely not in dynamic equilibrium. Focusing on the historical reference simulation, the valid month for each grid cell was identified, defined as the month after which the model is considered to be in a spun-up state. Non-equilibrium states were identified using a temporal correlation analysis. This approach relies on the fact that in areas affected by spin-up artifacts, hydraulic heads undergo a systematic adjustment that overpowers natural groundwater dynamics, and the magnitude of this adjustment decreases over time until the spin-up effect dissipates. To quantify this, the absolute difference in hydraulic heads between consecutive months was calculated (Eq. 8).

$$\Delta \text{HDS}_t = |\text{HDS}_t - \text{HDS}_{t-1}| \quad (8)$$

where HDS_t is the hydraulic head (in meters) at time step t , and ΔHDS_t represents the absolute month-to-month change in hydraulic head.

As a first pass, to filter out locations where groundwater heads are static, if the mean absolute difference of the entire simulation was less than 0.001 m, the cell was considered static and the entire simulation period was considered valid. For the remaining cells, a 12-month rolling window was applied to compute the mean difference and Pearson correlation coefficient between absolute differences and time (Eq. 9). A strong negative correlation indicates that the differences are decreasing, which signals incomplete spin-up. In contrast, a weak correlation suggests the model responds to dynamic forcing rather than initialization artifacts. The start valid date was determined by comparing the correlation coefficient, r_t , against a threshold of -0.2 . If the correlation in the first window was greater than -0.2 , the cell was marked as valid for the entire simulation. However, if the initial correlation indicated a strong negative trend, subsequent years were scanned to identify the first instance where r_t rose above -0.2 , designating the valid simulation time step. To prevent false negatives caused by numerical noise, if the

mean change ($\mu_{\text{HDS}} = \frac{\mu_{\Delta\text{HDS}_t}}{\tau}$) in head within the window was less than 0.001 m (i.e., the convergence criterion), the window was also marked as valid regardless of the correlation value.

$$r_t = \frac{\text{cov}(\Delta\text{HDS}, T)}{\sigma_{\Delta\text{HDS}} \cdot \sigma_T} \quad (9)$$

where r_t is the Pearson correlation coefficient at time step t , T represents the vector of time indices within the rolling window, $\text{cov}(\Delta\text{HDS}, T)$ is the covariance between head differences and time, and $\sigma_{\Delta\text{HDS}}$ and σ_T denote the standard deviations of head differences and time indices, respectively.

2.3 Simulations

Groundwater abstraction, recharge, and river discharge from the HYPFLOWSCI6 (HYdrological Projection of Future gLocal Water States with CMIP6) dataset were used as forcing for the simulations presented here and processed for MODFLOW according to Sect. 2.1 (Sutanudjaja, 2024; Kuzma et al., 2023). The HYPFLOWSCI6 dataset represents the integration of PCR-GLOBWB2 with CMIP6 climate forcings to project future water stress under three combined socioeconomic-climate scenarios: SSP1-RCP2.6, SSP3-RCP7.0, and SSP5-RCP8.5 (Sutanudjaja, 2024; Kuzma et al., 2023). Following the HYPFLOWSCI6 simulation framework, simulations were divided into two parts: (1) historical reference simulation and (2) scenario-based simulations. For the historical reference simulation, GSWP3-W5E5 forcing was used for the period 1960–2019 (Lange et al., 2021). For the scenario-based simulations the baselines (1960–2014) and future periods (2015–2100) for three SSP-RCP scenarios and five bias corrected GCMs were used (Table 3; Lange and Büchner, 2021). The choice of the specific five GCMs ensures the computations are feasible while still representing the spread of the full CMIP6 ensemble (Reinecke et al., 2026).

The simulations were initiated from the best-fit set of initial conditions as described in Sect. 2.2.2. These simulations include an additional spin-up year to avoid sudden jumps between initial conditions and the simulation start year. All simulations were conducted on Snellius, the Dutch national supercomputer. The MODFLOW6 prototype used in GLOBGM requires substantial pre-processing of the PCR-GLOBWB2 inputs into exactly one binary file per unstructured grid. In addition, post-processing is required to convert the binary outputs created by MODFLOW6 into data formats more commonly used by the community. To streamline the multi-model simulations, the pre-processing, MODFLOW6 simulation, and post-processing steps were collected into a single Snakemake workflow constructed to fit the HPC infrastructure on Snellius (Mölder et al., 2021). Satisfactory simulation times were achieved when running GLOBGM v1.0 on 12 compute nodes (Verkaik et al., 2024). For the historical reference simulation (1960–2019), the total run time

was approximately 21 h. For a single GCM simulation, with its baseline (1960–2014) and three future runs (2015–2100), the total simulation time was approximately 106 h. These simulations in total required 530 h on 2304 AMD EPYC 9654 CPUs, and consumed 2.05 MWh. Based in the Netherlands, this has a carbon footprint of 767.01 kg CO₂e, which is equivalent to 1.3 flights from New York City to San Francisco (calculated using <http://green-algorithms.org>, last access: 29 May 2026; Lannelongue et al., 2021).

2.4 Quality Assurance, Validation and Bias Correction

2.4.1 Quality Assurance

The hydrogeological schematization in GLOBGM targets large sedimentary aquifer systems. In regions dominated by karst, permafrost, or mountainous terrain, the simulations have been shown to be relatively less accurate due to incomplete process representation (Verkaik et al., 2024; de Graaf et al., 2017; Gädeke et al., 2020). Therefore, a map of regions where predictions are likely less reliable or fall outside the range of applicability of GLOBGM v1.1 is provided. Regions dominated by karst aquifer systems were obtained from the World Karst Aquifer Map (WHYMAP WOKAM; Chen et al., 2017) and rasterized to match the grid resolution used in GLOBGM. The Global Permafrost Zonation Index Map was used to obtain regions dominated by permafrost, using a threshold value of 0.9, which corresponds to areas with permanent and continuous permafrost (Gruber, 2012; Ford and Frauenfeld, 2016). To identify mountainous regions, the standard deviation of slope within each GLOBGM grid cell was utilized. Grid cells within the top 1 % of topographic heterogeneity, characterized by slope standard deviation values exceeding 77 (dimensionless), were delineated as mountainous terrain.

2.4.2 Validation

The historical reference simulations (Sect. 2.3) were validated against observations obtained from IGRAC's Global Groundwater Monitoring Network (IGRAC, 2024). Observations were filtered to include only those wells covering 1965–2019 with a minimum of 8 years of data; observations prior to 1965 were excluded from validation to avoid spin-up artifacts (Sect. 2.2.2). To match the observed data to the correct model layer, the same procedure relying on correlations between soil moisture and simulated water table depths was used as in the calibration (Sect. 2.2.1) and initial conditions (Sect. 2.2.2) and if multiple wells were located within a single grid cell, the one with the most data was retained.

To assess the accuracy of the simulations, the non-parametric Kling-Gupta Efficiency (KGE-NP; Eq. 10; Pool et al., 2018) was calculated between observed and simulated groundwater heads. Validation focused on groundwater heads, not water table depth, to avoid numerical instability in

Table 3. Climate forcing datasets and future scenarios for groundwater simulations with GLOBGM v1.1.

Simulation	Forcing	Period	Description
Historical Reference	GSWP3-W5E5 (Lange et al., 2021)	1960–2019	Reanalysis-based reconstruction which allows validation and impact attribution against observed climate records.
Future Scenarios	GFDL-ESM4 (Dunne et al., 2020)	1960–2014	Baseline simulation.
	IPSL-CM6A-LR (Boucher et al., 2020)	2015–2100	SSP1-RCP2.6: Global warming of 2 °C, radiative forcing of 2.6 W m ^{−2} in 2100, optimistic sustainable development with low barriers to mitigation and adaptation.
	MPI-ESM1-2-HR (Mauritsen et al., 2019)		SSP3-RCP7.0: 3.6 °C global warming, 7.0 W m ^{−2} radiative forcing in 2100, business as usual world with regional rivalry and medium-high emissions pathway.
	MRI-ESM2-0 (Yukimoto et al., 2019)		SSP5-RCP8.5: 4.4 °C global warming, 8.5 W m ^{−2} radiative forcing in 2100, pessimistic scenario with high economic growth and high fossil fuel dependence.
	UKESM1-0-LL (Sellar et al., 2020)		

the KGE-NP bias term caused by small denominators. KGE-NP values range from $-\infty$ to 1.0.

$$\text{KGE-NP} = 1 - \sqrt{(r_s - 1)^2 + (\alpha_{\text{NP}} - 1)^2 + (\beta - 1)^2} \quad (10)$$

where r_s is the Spearman rank correlation between the simulated and observed time series; α_{NP} is the variability ratio, calculated from the relative flow duration curve as $1 - \text{MAE}(\text{FDC}_{\text{sim}}, \text{FDC}_{\text{obs}})$, where FDC represents the normalized flow duration curves; and β is the bias ratio, defined as the ratio of the mean of simulations to the mean of observations.

Thereafter, KGE-NP scores were normalized into a skill score ($\text{KGE-NP}_{\text{skill}}$) following the approach of Knoben et al. (2019) (see Appendix B). A skill score greater than 0 indicates that the model outperforms the benchmark, while a score of 1 indicates perfect performance. This normalization is necessary because, unlike the standard KGE, the KGE-NP benchmark threshold is not a universal constant and depends on the variability of the observed time series. By normalizing to the skill score, we recover a threshold of zero that carries the same interpretation across all wells and is therefore suitable for global-scale evaluation and inter-model comparison.

$$\text{KGE-NP}_{\text{skill}} = \frac{\text{KGE-NP} - \text{KGE-NP}_{\text{bench}}}{1 - \text{KGE-NP}_{\text{bench}}} \quad (11)$$

where $\text{KGE-NP}_{\text{bench}}$ is the KGE-NP score obtained by the mean flow benchmark, where $\text{HDS}_{(t)} = \overline{\text{HDS}}_{\text{obs}}$ for all time steps.

As groundwater observations are sparse and unevenly distributed, with most observations occurring in North America, Europe, and Australia (Fig. C1; B  thge et al., 2026), we additionally use 0.25   GRACE/GRACE-FO total water storage anomalies (GRACE/GRACE-FO TWSA; Li and Kusche, 2026) as a coarse scale validation of the simulations presented here. We calculated monthly (5 arcmin) simulated total water storage by replacing the groundwater component

of the 5 arcmin PCR-GLOBWB historical reference simulation outputs with GLOBGM estimates. GLOBGM groundwater storage was estimated from the hydraulic heads and storage coefficients (Verkaik et al., 2024). In order to match GRACE TWSA, simulated total water storage was converted to anomalies (GLOBGM TWSA) relative to the 2004–2009 mean and coarsened to 0.25  . Simulated and GRACE TWSA have been compared in several global model evaluations, most of which estimate trends separately and quantify agreement based on sign and magnitude agreement (Scanlon et al., 2018; Tiwari et al., 2025; Felfelani et al., 2017). However, this does not reveal whether the trends differ significantly (Gelman and Stern, 2006) or account for GRACE TWSA uncertainty.

Therefore, to determine whether GLOBGM TWSA is significantly different from GRACE/GRACE-FO, we look at the difference time series ($\Delta\text{TWSA} = \text{TWSA}_{\text{GRACE}} - \text{TWSA}_{\text{GLOBGM}}$) and calculate the trend using the Theil-Sen estimate (β_{Δ}) and significance ($p_{\beta_{\Delta}}$) using the modified Mann-Kendall test (Hamed and Ramachandra Rao, 1998). The influence of GRACE/GRACE-FO uncertainty on β_{Δ} is incorporated through Monte Carlo sampling from which we calculate the 95 % confidence interval ($\text{CI}(\beta_{\Delta})$). In addition, we determine the significance of the GRACE/GRACE-FO TWSA Theil-Sen slope (p_{GRACE}). These statistics are then used to differentiate between regions which are stable and in agreement, trends and magnitudes agree, model too dry, model too wet, and sign disagreement (Table 4).

2.4.3 Machine Learning Bias Correction

To correct for systematic bias in simulated water table depths, a machine learning-based bias (ML bias) correction approach was implemented. A Light Gradient-Boosting Machine model (Ke et al., 2017) was trained to predict mean annual bias (Eq. 12) in observed groundwater heads from

Table 4. Classification of grid-scale agreement between GRACE/GRACE-FO and GLOBGM terrestrial water storage trends. p_{GRACE} : significance of GRACE/GRACE-FO TWSA Theil-Sen slope. p_{sim} : significance of GLOBGM TWSA Theil-Sen slope. $\text{CI}(\beta_{\Delta})$: 95 % confidence interval of the difference time-series Theil-Sen slope. $p_{\beta_{\Delta}}$: significance of the difference time-series Theil-Sen slope.

Category	Decision criteria					Interpretation
	p_{GRACE}	p_{sim}	$\text{CI}(\beta_{\Delta})$	$p_{\beta_{\Delta}}$	Sign relation	
1 Stable agreement	≥ 0.05	–	incl. 0	–	–	GRACE shows no significant trend and no resolvable difference from the model.
	≥ 0.05	–	excl. 0	≥ 0.05	–	Interval resolves a difference, but it is not statistically systematic.
2 Trend and magnitude agreement	< 0.05	–	incl. 0	–	–	GRACE trends significantly and the model reproduces it within GRACE uncertainty.
	< 0.05	–	excl. 0	≥ 0.05	–	Interval resolves a difference, but it is not statistically systematic.
3 Model too dry	< 0.05	< 0.05	excl. 0	< 0.05	same, $\beta_{\Delta} > 0$	Both series trend the same way, but the model depletes too fast or wets too slowly.
4 Model too wet	< 0.05	< 0.05	excl. 0	< 0.05	same, $\beta_{\Delta} \leq 0$	Both series trend the same way, but the model depletes too slowly or wets too fast.
5 Sign disagreement	< 0.05	< 0.05	excl. 0	< 0.05	opposite	Both trends are significant and run in opposite directions.
6 Model misses trend	< 0.05	≥ 0.05	excl. 0	< 0.05	–	GRACE establishes a significant trend the model does not reproduce.
7 Spurious model trend	≥ 0.05	< 0.05	excl. 0	< 0.05	–	The model trends significantly in a direction GRACE does not support.

historical reference simulations (see Sect. 2.3). Observations were obtained from IGRAC's Global Groundwater Monitoring Network (IGRAC, 2024) and included all available annual observation wells (96 799) that overlapped with the simulation period; if multiple wells were located within a single grid cell, the one with the most data was retained. As features, a select number of GLOBGM input parameters (steady-state water table depth, drain conductance, elevation, slope, model layer, recharge, abstraction) and Google's AlphaEarth Foundations geospatial embeddings (Brown et al., 2025) were used. In addition, absolute latitude was added as a proxy for aridity. Observations were split into training (80 %), validation (10 %), and test (10 %) sets using K-means clustering ($n = 300$) on normalized geographic coordinates, thereby ensuring spatial independence and preventing spatial autocorrelation from inflating performance metrics. Gain-based feature importance scores were calculated for each stage separately, and the importance scores were normalized to enable comparison between the stages.

$$\overline{\text{Bias}}_{\text{HDS}} = \frac{1}{n} \sum_{i=1}^n (\text{HDS}_{\text{sim},i} - \text{HDS}_{\text{obs},i}) \quad (12)$$

where HDS_{sim} is the simulated groundwater head and HDS_{obs} is the observed groundwater head.

We implemented a two-stage ensemble: (1) a binary classifier to predict whether bias is positive ($p_{\text{sign}} = P(\overline{\text{Bias}}_{\text{HDS}} > 0)$), and (2) a magnitude regressor to predict absolute bias magnitude. For the binary classifier, the bias target was converted into a binary classification target using the indicator function $y_{\text{sign}} = \mathbb{I}(\overline{\text{Bias}}_{\text{HDS}} > 0)$, and was trained using a binary log-loss metric. Although trained on the log-loss metric, performance is also reported in terms of the proportion of observations where the sign was correctly predicted. For the magnitude model, the target is the log-transformed absolute bias ($y_{\text{mag}} = \log(1 + |\overline{\text{Bias}}_{\text{HDS}}|)$) and is evaluated using RMSE. For the final bias prediction, the predictions of sign probability and magnitude were combined (Eq. 13). When the sign classifier is uncertain ($p_{\text{sign}}(\mathbf{x}) \approx 0.5$), the coefficient $s(\mathbf{x})$ approaches zero and the predicted bias is attenuated toward zero, so no full correction is applied where the direction of the bias is ambiguous.

$$\begin{aligned} \hat{b} &= s(\mathbf{x}) \cdot \hat{m}(\mathbf{x}) \\ &= [2 p_{\text{sign}}(\mathbf{x}) - 1] \cdot \hat{m}(\mathbf{x}) \end{aligned} \quad (13)$$

where $s(\mathbf{x})$ is the prediction from the sign classifier (stage 1) and $\hat{m}(\mathbf{x})$ is the prediction from the magnitude regressor.

The trained two-stage model was then used to predict the bias for all grid cells in the model domain, which was then applied to bias-correct GLOBGM v1.1 groundwater heads and water table depths described in Sect. 2.3. To prevent overcorrections and preserve original model physics, an asymmetric ramp-factor correction scheme is implemented (Eq. 14). This approach applies a damping ramp exclusively when the predicted correction would raise the simulated head toward the land surface. In cases where the correction lowers the simulated head, no physical land-surface boundary is at risk of being breached, and thus the full correction is applied without restriction. Finally, the correction was applied to the simulated heads and water table depths (Eqs. 15 & 16).

$$\omega = \begin{cases} 0, & \text{if } \min(\text{WTD}_{\text{sim}}) \leq d_{\text{protect}}, \\ \frac{\min(\text{WTD}_{\text{sim}}) - d_{\text{protect}}}{w_{\text{ramp}}}, & \text{if } d_{\text{protect}} < \min(\text{WTD}_{\text{sim}}) < w_{\text{ramp}}, \\ 0.9, & \text{if } \min(\text{WTD}_{\text{sim}}) \geq w_{\text{ramp}}, \end{cases}$$

$$\Delta d_{\text{correction}} = \begin{cases} \hat{b}\omega, & \text{if } \hat{b} < 0, \\ \hat{b}, & \text{if } \hat{b} \geq 0. \end{cases} \quad (14)$$

where d_{protect} is a protection depth below which no correction is applied (0.1 m) and w_{ramp} is the ramp width over which the correction scales from 0 to 0.9 (2 m).

$$\text{HDS}_{\text{corrected}}(t) = \text{HDS}_{\text{sim}}(t) - \Delta d_{\text{correction}} \quad (15)$$

$$\text{WTD}_{\text{corrected}}(t) = \text{WTD}_{\text{sim}}(t) + \Delta d_{\text{correction}} \quad (16)$$

2.5 GLOBGM v1.1 in comparison to previous modeling efforts

To assess model performance in relation to previous global modeling efforts, we benchmarked GLOBGM against existing global groundwater models. This allows us to assess the model skill relative to the current state-of-the-art and identify where improvements have been achieved. We compared the steady state GLOBGM v1.1 results with those of Fan et al. (2017), de Graaf et al. (2017), Reinecke et al. (2019) and Verkaik et al. (2024). For the transient simulation, GLOBGM v1.1 was compared to Verkaik et al. (2024) and de Graaf et al. (2017).

To compare inter-model performance for steady state estimates, we calculated the absolute bias between observations and simulated water table depth. However, for the transient simulations we used the KGE-NP_{skill} with reference to the mean flow benchmark. Given that the models used for comparison are at the 5 arcmin (de Graaf et al., 2017; Reinecke et al., 2019) and 30 arcsec resolution (Fan et al., 2017; Verkaik et al., 2024), to compare the models in a pragmatic manner, we calculated the validation metrics on the observations described in Sect. 2.4.2, which are geared towards the validation of a 30 arcsec grid. For these comparisons, GLOBGM v1.1 is represented by its raw uncorrected outputs and for transparency, we also report the optional ML-corrected product (ML-GLOBGM v1.1).

2.6 Analysis

To provide insights into the spatial distribution of groundwater at the 30 arcsec resolution and highlight the benefit of increased resolution the long-term average annual groundwater table depth was calculated at the global scale (Eq. 17). For visualization purposes, a “from ground level downwards” view of groundwater table depths is provided; for locations where a confining layer exists, only the top model layer is visualized.

$$\overline{\text{WTD}}_{i,j} = \frac{1}{12N} \sum_{k=1}^N \sum_{m=1}^{12} \text{WTD}_{i,j,k,m} \quad (17)$$

where $\overline{\text{WTD}}_{i,j}$ is the average WTD at grid point (i, j) , N is the total number of years, m is the month index, and $\text{WTD}_{i,j,k,m}$ is the WTD at grid point (i, j) in month m of year k .

To analyze how groundwater resources have changed in the past and may change in the future, we estimated trends in water table depth using the Theil-Sen estimator (Sen, 1968; Theil, 1950) so that the slope for each cell was calculated along the time dimension as the median of all pairwise slopes between distinct time points (Eq. 18). We tested the slopes for each grid cell using the Mann-Kendall test (Kendall, 1975). When presented as a raster, slopes less than 0.001 m yr^{-1} , not statistically significant at the 95 % confidence level, or both were masked. For visualization purposes, only the bottom layer of the model is presented, which represents fully unconfined aquifers or confined aquifers.

$$m_{i,j} = \text{median} \left\{ \frac{\text{HDS}(t_l, i, j) - \text{HDS}(t_k, i, j)}{t_l - t_k} \mid \forall k < l \right\} \quad (18)$$

where $m_{i,j}$ is the median slope at grid point (i, j) , $\text{HDS}(t_k, i, j)$ is the value of HDS at time t_k and grid point (i, j) . The median is taken over all pairs of time indices $k < l$.

In addition, we present the slopes of mean water table depth aggregated by spatial units. The spatial units are combined water provinces (Straatsma et al., 2020) and WHYMAP aquifer dataset (Chen et al., 2017). We first computed, for each spatial unit and each year, the area-weighted mean water table depth over all cells within that unit, and then applied Eq. (18) and the Mann-Kendall test to aggregated time series. For the spatially aggregated representation, slopes less than 0.001 m yr^{-1} , not statistically significant at the 95 % confidence level, or both were masked. To complement the trends, we calculated area-weighted mean water table depth anomalies, relative to 1965, over the simulation period aggregated by continent.

3 Results

3.1 Calibration, Spin-up and Initial Conditions

3.1.1 Calibration

GLOBGM proved sensitive to adjustments in hydraulic conductivity, anisotropy, and entrance resistance (Table 5 & Fig. 2e). Among the top ten calibration combinations, candidate 1 achieved the lowest mean absolute bias overall, for the 0–5 m and 10–20 m categories, and the second lowest in the 5–10 m category (Table 5). Consequently, a prefactor of 0.1 for all parameters was chosen as the best set. The selection of these prefactors correspond to the minimum values tested, which suggests that smaller prefactors may have provided better performance. However, the range of calibration factors was intentionally constrained within two orders of magnitude to avoid significant deviations from the original parameterization and prevent overfitting.

Greater variance among permutations was observed for locations where simulated water table depths are deeper than observed. This suggests that calibration influenced over-estimated depths more significantly than under-estimated ones (Fig. 2e). Varying calibration prefactors tended to shift entire distributions rather than simultaneously correcting outliers at both ends (Fig. 2e). As a result, the chosen prefactor set improved the accuracy of over-estimated depths but increased the error for under-estimated depths. However, this set maintained the best overall bias, as gains at the negative end of the distribution were offset by penalties at the positive end (Table 5 & Fig. 2e). Compared to the previous uncalibrated GLOBGM v1.0 (Verkaik et al., 2024), the chosen calibration settings display marked reductions in overall depth weighted bias (Table 5). This is especially prevalent for shallower water table depths (< 20 m; Table 5). For wells between 20 and 60 m calibration has little effect, whereas for wells deeper than 60 m calibration increased bias (Table 5). In comparison to GLOBGM v1.0, unweighted bias shifts from a tendency to overestimate water table depths (−4.8 m) to underestimation when calibrated (3.6 m) (Table 5 & Fig. 2e). Comparing the uncalibrated GLOBGM v1.1 with v1.0 reveals how the dynamic elevation routine and groundwater recharge downscaling affect the model's accuracy. Compared to GLOBGM v1.0, the uncalibrated v1.1 predictions were closer to the ideal bias score of zero, by a magnitude of 1.6 m, again centered around improvements in simulations that were deeper than observed (Table 5 & Fig. 2e).

3.1.2 Spin-up and Initial Conditions

The objective of the first transient spin-up step was to achieve dynamic equilibrium for the year 1960, whereas the second step incorporated abstraction to reproduce expected water table depths. The key assumption is that initiating the simulation from a naturalized steady state solution and simulating

1960 iteratively, without abstraction, would yield water table depths which are shallower than observed; therefore, by activating abstraction in the second step, water table depths would approach observed values. During the first step, the change in average groundwater heads between successive iterations decreased – suggesting the model approached dynamic equilibrium. The correlation-based analysis aimed at detecting regions where the model was not properly spun up reveals that the majority of grid cells were correctly initiated at the start of the simulation. Regions that were incorrectly initiated were sparse and geographically confined. Most regions reached adequate spin-up within the first 5 years, with 99.7 % of the domain deemed valid after the first year, 99.8 % after the second year, 99.9 % thereafter, and 99.99 % by 1965 (Fig. 3b).

3.2 Quality Assurance, Validation and Bias Correction

3.2.1 Quality Assurance

Previous evaluations of the 5 arcmin predecessor of GLOBGM v1.1 have revealed that performance degrades in regions where geology is dominated by karst aquifers, permafrost, or mountainous terrain (Fig. 4) due to incomplete process representation. GLOBGM performs better in sedimentary than in karst basins, presumably because the Darcian flow assumption is questionable in karst settings (de Graaf et al., 2017; Verkaik et al., 2024). In mountainous regions, GLOBGM models groundwater heads within mountain blocks but does not parametrize water bodies in hillslopes and smaller alluvial mountain valleys, since the hydrogeological schematization is too coarse. Moreover, it is assumed that the secondary permeability of fractured hard rock in mountainous regions can also be represented by the principles of Darcian groundwater flow (Verkaik et al., 2024). Research into the applicability of PCR-GLOBWB2 to simulate the hydrology of permafrost regions shows that it does not capture the dynamics of permafrost with high fidelity due to its simplified representation of ice melt (i.e., degree-day-factor) and lack of dynamic permafrost schemes, which leads to underestimated discharge and, consequently, groundwater recharge (Gädeke et al., 2020). Moreover, it is challenging to define groundwater resources in areas where soil is frozen up to many tens of meters deep (Diak et al., 2023; Cochand et al., 2019). These masks are provided, together with the simulations, to inform users of regions where GLOBGM v1.1 predictions are likely less reliable.

3.2.2 Validation

When validating on the uncorrected simulations, monthly scores (74.7 % $KGE-NP_{skill} > 0$) are generally higher than annual (67.0 % $KGE-NP_{skill} > 0$) scores (Fig. 5a, e). This trend is also evident in the constituent components of KGE-NP, where correlations are generally higher when validated at monthly scales (Fig. 5c, g). Variability and bias show no

Table 5. Comparison of the top 3 calibration candidates (ordered by mean absolute depth-weighted bias) against the reference GLOBGM v1.0 benchmark. Parameter settings are dimensionless multipliers applied to the default values. Values in bold indicate the chosen calibration settings.

Model	Parameter Settings					Mean Absolute Bias by Depth Class (m)							
	Ks (Fine+Coarse)		Carbonate	Anisotropy	Entry Res.	Mean Bias	< 0	0–5	5–10	10–20	20–60	> 60	Mean
	Uncon.	Consol.											
Candidate 1	0.1	0.1	0.1	0.1	0.1	3.6	37.3	9.6	12.0	15.6	27.2	93.3	32.5
Candidate 2	0.1	0.1	0.1	1.0	0.1	3.3	37.7	9.7	11.9	15.4	27.0	93.1	32.5
Candidate 3	0.1	1.0	0.1	0.1	0.1	1.1	37.6	9.8	12.3	15.9	27.2	92.7	32.6
Uncalibrated	–	–	–	–	–	–1.6	40.1	11.8	14.7	18.3	26.5	85.7	32.9
v1.0	–	–	–	–	–	–4.8	41.3	13.6	17.1	20.4	27.3	85.6	34.2

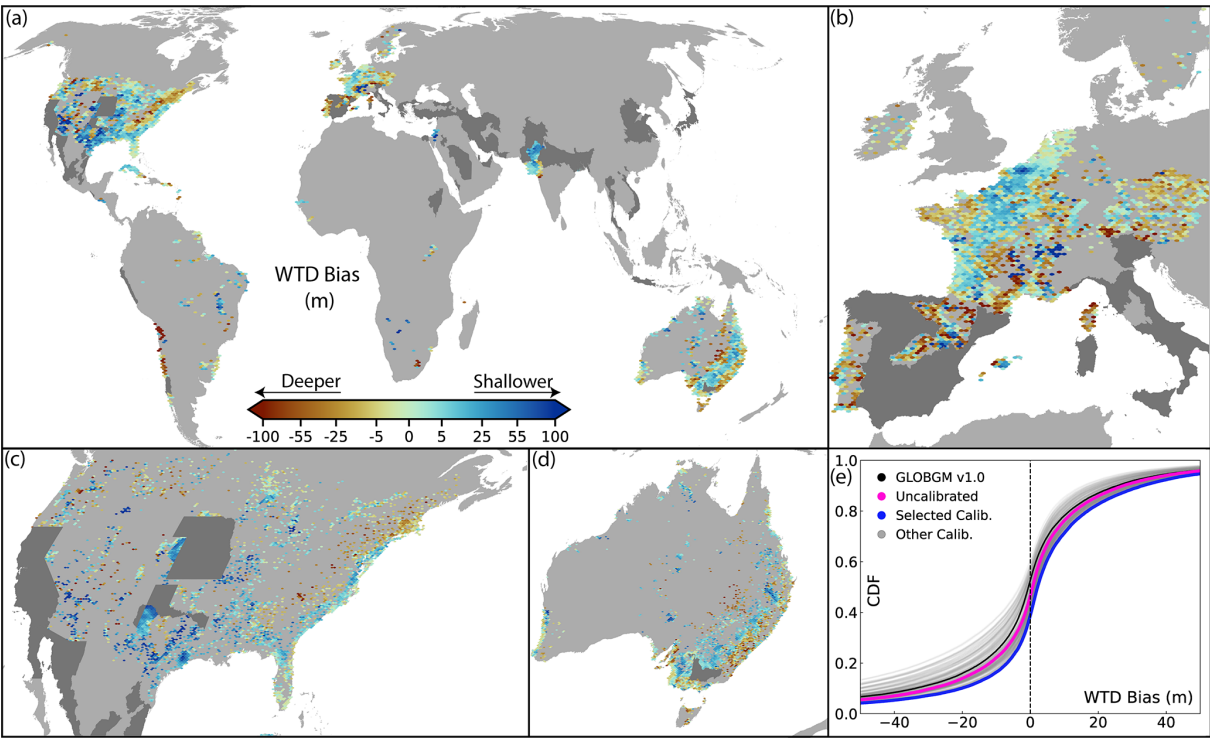


Figure 2. (a) Geographical variation of water table depth bias, for the chosen calibration setting, from 34 800 observation wells that were used to assess different calibration settings. Dark grey polygons indicate regions of known anthropogenic influence on the water system and were excluded from the calibration. With zoomed insets of (b) western Europe, (c) United States, (d) Australia. (e) Cumulative distribution function of the water table depth bias of the 162 calibration permutations.

difference between monthly and annual resolutions (Fig. 5d, e, h, i). The ability of GLOBGM to simulate groundwater heads varies predictably with water table depth. KGE- NP_{skill} scores grouped according to depth categories reveal that the deeper simulations (> 60 & $20\text{--}60$ m) are the worst performing, with approximately 56 %–70 % displaying skillful predictions that improve upon the mean flow benchmark (Fig. 5b, f). Shallower observations ($0\text{--}20$ m) are better reproduced, with 68 %–80 % of locations showing skillful predictions (Fig. 5b, f). This depth-dependent pattern is also

evident in correlation reproduction, where shallower values are more accurately predicted than deeper values (Fig. 5c, g). Groundwater heads tend to be underestimated for observations between $0\text{--}5$ m (Fig. 5e, i). In contrast, groundwater heads are over-estimated for deeper water table depths (Fig. 5e, i). In contrast, the ability to reproduce variability shows no clear correlation with water table depth (Fig. 5d, h).

To determine whether low performance scores at specific locations reflect genuine physical discrepancies or arise

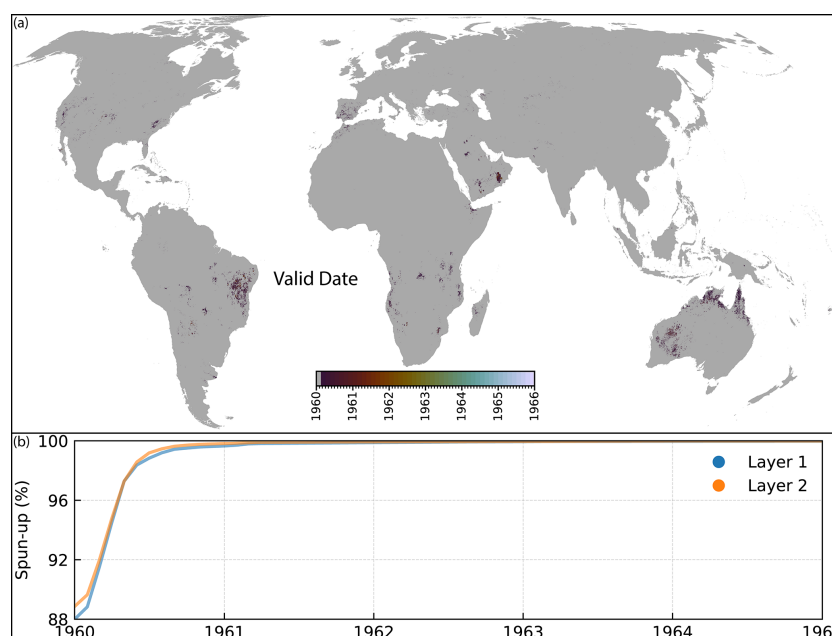


Figure 3. (a) Map showing the first year of valid simulation data per grid cell. (b) Percentage of grid cells in Layer 1 and Layer 2 that are free of spin-up artifacts between 1960 and 1965.

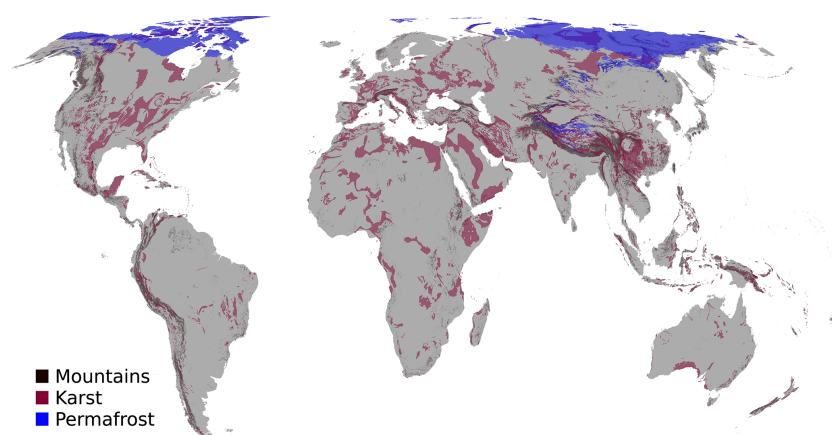


Figure 4. Regions where GLOBGM has process representation limitations due to practical and theoretical limitations; namely, karst aquifers, permafrost and mountainous regions.

from evaluation metrics artifacts and whether performance varies between stable or trending observed time series, we conducted a post-hoc analysis. Where trends are present ($|\text{Theil-Sen slope}| > 0.01 \text{ m yr}^{-1}$) in the observed time series, 83 % (monthly) and 73 % (annual) of the locations are skillful, compared to 78 % (monthly) and 70 % (annual) where observed time series is considered stable ($|\text{Theil-Sen slope}| \leq 0.01 \text{ m yr}^{-1}$; Fig. D1). The overall $\text{KGE-NP}_{\text{skill}}$ score is mainly impacted by small values in the denominator of the bias ratio (β), which is caused by small groundwater heads. Removing these cases ($\mu\text{HDS}_{\text{obs}} < 1.0 \text{ m}$) means 80 % of the monthly and 71 % of the annual scores are skill-

ful, an improvement of 5 and 4 percentage points in the monthly and annual validations, respectively (Fig. D2).

The spatial variation in $\text{KGE-NP}_{\text{skill}}$ score corroborates some of the conclusions derived from the cumulative distribution functions (Fig. 5a). For instance, lower scores are evident along regions where water table depths are expected to be deeper, such as arid regions and mountainous terrain (Fig. 5a). Furthermore, depth-dependency becomes evident when comparing relatively lower scores in the higher elevations of the Appalachian, Pyrenees, and Great Divide mountains to higher scores in the adjacent low lying regions (Fig. 5a). Similarly, the higher regions of the Indo-Gangetic Plain display lower $\text{KGE-NP}_{\text{skill}}$ scores when compared to

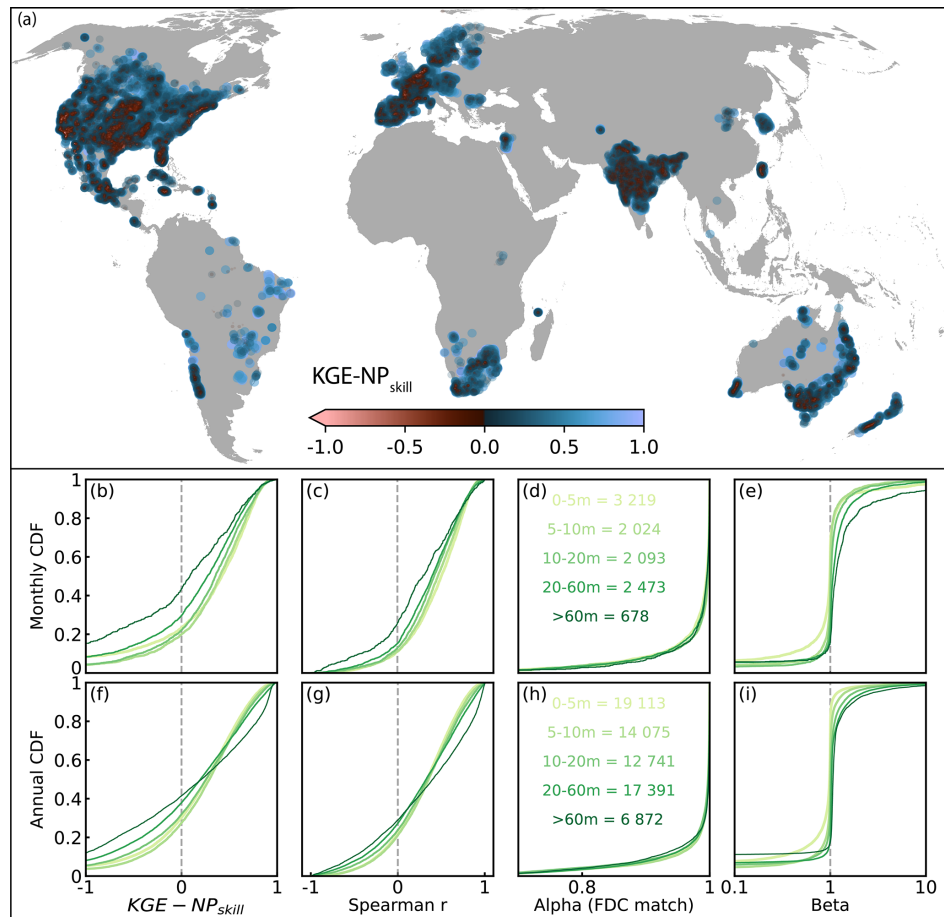


Figure 5. $KGE-NP_{skill}$ spatial distribution (a) and CDF (b, f). Constituent components (c, g) correlation, (d, h) alpha, and (e, i) beta of $KGE-NP$ for monthly and annual groundwater heads simulated by GLOBGM from 1965–2019, grouped by water table depth. Gray lines in panels (b) and (f) symbolize a $KGE-NP_{skill}$ of 0 which is the threshold at which greater values indicate predictions that improve upon the mean flow benchmark. Gray lines in panels (c) and (g) indicate the point at which correlations are positive to the right and negative to the left. Gray lines in panels (e) and (i) separate over-estimation to the right and under-estimation to the left.

coastal regions (Fig. 5a). Figure 5a and the analyses shown in Fig. D3 show that the regions of poor performance do not line up well with the regions in Fig. 4 where GLOBGM is expected to provide less accurate results due to limitations in process representation. In addition, there are no significant differences in performance between locations outside or inside major abstraction areas (Fig. D3).

The spatial distribution of TWSA trends shows broad-scale agreement between GLOBGM and GRACE/GRACE-FO (Fig. E1a). GLOBGM matches GRACE/GRACE-FO in both trend direction and magnitude for 45.7 % of the world (Fig. E1a). The two agree in direction but disagree in magnitude over 16.6 %; GLOBGM is too wet over 9.8 % and too dry over 6.8 % (Fig. E1a). GRACE/GRACE-FO disagree in direction for only 7.1 % (Fig. E1a). Over 17.8 % GRACE/GRACE-FO indicates a significant trend where GLOBGM simulates no trend. Conversely, GLOBGM simulates a trend when GRACE/GRACE-FO indicates no trend

over 12.8 % of the world. GLOBGM simulates the direction of TWSA trend in known regions of water storage decline, such as the south-western United States, northern India, and Pakistan, and the Middle East (Fig. E1). Lack of agreement aligns with karst, permafrost, and mountainous regions (Fig. E1); GLOBGM is expected to show reduced accuracy in these regions (Sect. 3.2.1). Notable examples include the Arctic, sub-Arctic regions North America and eastern Russia, where permafrost is present. Furthermore, disagreement is evident across prominent mountain ranges like the Andes and the Himalayas are evident (Fig. E1). Model disagreement is also evident in major karst landscapes. This is particularly evident across Mexico, Northern Africa, the Gulf of Guinea, and the Tibetan Plateau (Fig. E1). In addition, disagreement is prevalent in islands and around the world's major water bodies, such as the Great Lakes, Rift Valley Lakes, and Caspian Sea (Fig. E1).

3.2.3 Machine Learning Bias Correction

The machine learning based bias correction demonstrated adequate performance on the holdout set. The binary sign classifier was capable of correctly predicting bias direction in 80 % of the cases, which was comparable to the validation (80 %) and train (90 %; Table F1) sets. Similarly, the magnitude regressor achieved an RMSE of 37.4 in the train set, which was comparable to the validation (25.3) and test (36.7; Table F1) sets. The steady-state depth of the water table emerges as the most important feature (Fig. F1a, b). The ML-corrected GLOBGM estimates show a predominantly negative bias for central North America and Europe (Fig. 6a); whereas, the Indian subcontinent and semi-arid regions of southern Africa display a positive bias (Fig. 6a).

When grouped by depth classes, the original model showed a consistent pattern of over-estimation for shallower wells (0–10 m) and under-estimation for deeper wells (> 20 m; Fig. 6d, e). The ML correction reduced the median bias across all depth classes and improved the bias for the majority of the locations (Figs. 6d, e & F1c, d). After ML correction, the median bias shifted closer to zero in all categories and the interquartile range contracted (Fig. 6d, e). This reduction was particularly notable for the class > 60 m, where the original model showed the largest biases (Fig. 6d, e). The bias correction was more effective on the annual time scale compared to the monthly time scale, with a more pronounced reduction in bias (Fig. 6d, e). Post correction bias reductions translated into improved KGE-NP scores. For both the annual and monthly time scales, the ML-corrected simulations displayed higher median KGE-NP scores and reduced interquartile ranges across all depth classes (Fig. 6b, c). These improvements led to a $\sim 0.3\%$ – 1.7% increase in the proportion of observations with skillful predictions ($\text{KGE-NP}_{\text{skill}} > 0$; Fig. F1e).

3.3 GLOBGM v1.1 in comparison to previous modeling efforts

The raw steady-state GLOBGM v1.1 estimates are comparable to previous global 30 arcsec resolution efforts (GLOBGM v1.0 & Fan et al., 2017; Fig. 7a, b). Although Fan et al. (2017) achieves the lowest mean absolute bias overall (18.11 m), the CDFs intersect at approximately 20 m, indicating a lack of overall dominance. GLOBGM v1.0 and GLOBGM v1.1 exhibit a higher proportion of sites with absolute bias below 20 m, yet above this threshold Fan et al. (2017) provide more accurate estimations. This crossover suggests that Fan et al. (2017) achieves its lower mean absolute bias primarily through superior performance at locations with moderate-to-large errors, while GLOBGM v1.1 more effectively captures water table depth at the majority of locations (approximately 75 %) but at the cost of a heavier error tail (Fig. 7a, b). Applying the optional machine-learning bias

correction (ML-GLOBGM v1.1) reduces the absolute bias further relative to observations (Fig. 7a, b).

Compared to the 5 arcmin counterparts (de Graaf et al., 2017) the 30 arcsec simulations show marked improvements (Fig. 7c, d). Collectively, different versions of GLOBGM display an approximately 19 and 15 percentage points increase in the number of locations that improve upon the mean benchmark ($\text{KGE-NP}_{\text{skill}} > 0$), at the monthly and annual resolution respectively (Fig. 7c, d). The raw uncorrected GLOBGM v1.1 simulation represents an improvement over v1.0, particularly at the monthly resolution, 8 and 1 percentage points increase at the monthly and annual resolution, respectively (Fig. 7c, d). The ML-correction provides further improvement (Fig. 7c, d).

3.4 Historical Reference Simulation

The hyper-resolution groundwater model reveals the added spatial detail in mean annual water table depth (Fig. 8a). Finer scale interactions between river systems and groundwater are evident at this hyper-resolution. For example, the shallow reaches of the Amazon and Mississippi rivers, with their respective deltas, are visible through shallow water tables (Fig. 8e, b). Other major deltas, like the Indus and Ganges-Brahmaputra (Fig. 8d), are also well resolved at this resolution and shallower water table depths along the coast compared to more interior regions are present (Fig. 8b–g). Average water table depths largely resemble our current broad understanding of groundwater dynamics and reflect the influences of precipitation, topography and geology on driving the distribution of groundwater reserves.

Water table depths are seen to be deeper in the arid regions and shallower in humid areas, reflecting the difference in groundwater recharge rates which are driven by large-scale precipitation patterns (Reinecke et al., 2024). More mountainous regions exhibit deeper water table depths which is expected given the relatively lower infiltration rates, lower permeability, and tendency for water to move down slope by streams towards lower lying ground under the influence of gravity (Somers and McKenzie, 2020). GLOBGM reproduces deepening water table depths in regions where groundwater depletion is known to occur, such as the southern and central High Plains in the United States (Figs. 9 & G1), Iran (Figs. 9 & G1), Indo-Gangetic Plain (Figs. 9 & G1; Jasechko et al., 2024). Interestingly, GLOBGM identifies a number of regions where water table depths have risen, the majority of which are located in the northern latitudes and Arctic Circle, such as Northern Scandinavia and the Russian Far North. This could conceivably be due to climate change enhancing precipitation and groundwater recharge dynamics; however these trends align with permafrost regions where GLOBGM has been shown to be less reliable (McKenzie et al., 2021). In addition, simulated rising water tables match regions that have been shown to display increases in groundwater reserves such as the Guarani aquifer in South America,

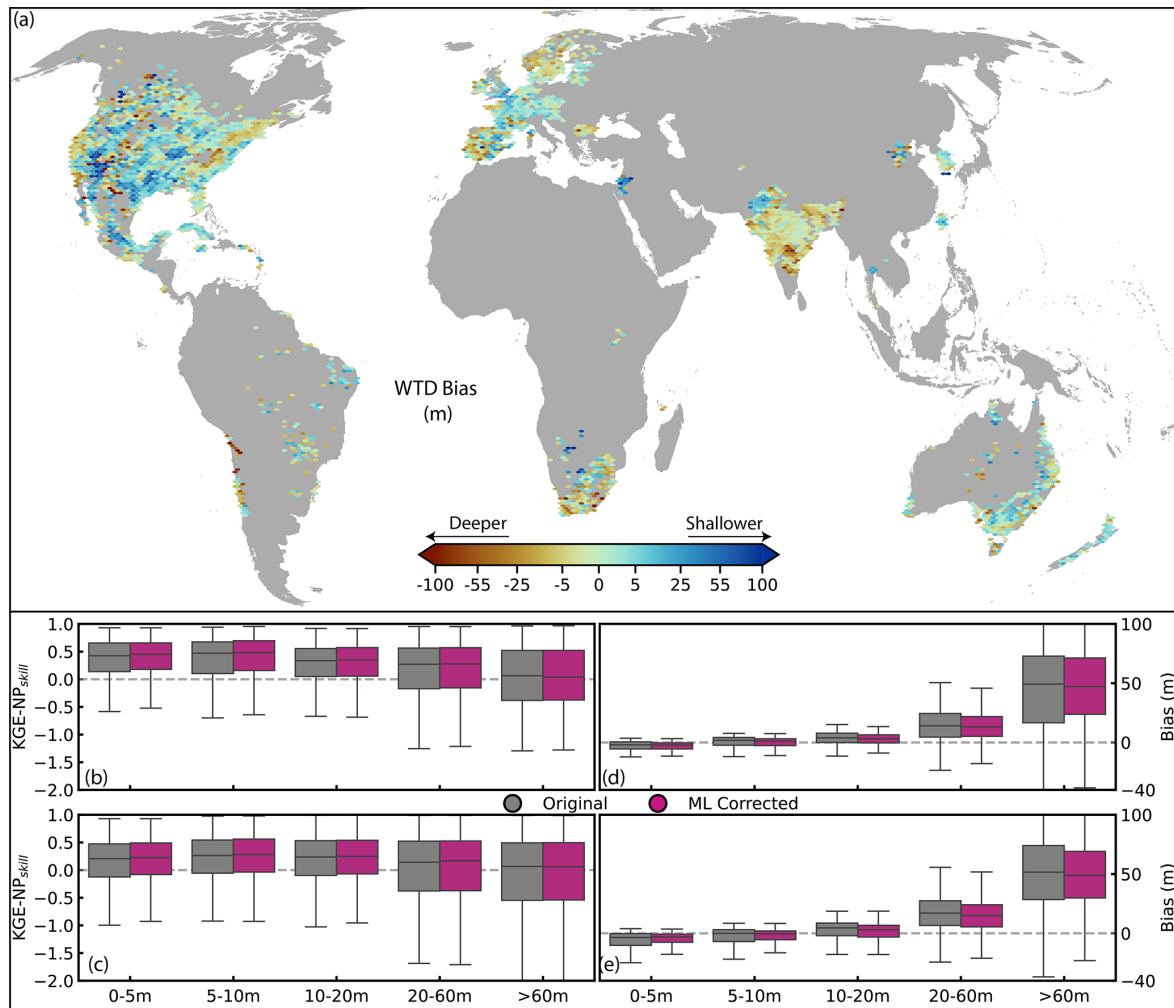


Figure 6. (a) Spatial distribution of ML-corrected mean water table depth bias (m). Original and ML-corrected (b) monthly and (c) annual KGE-NP_{skill} scores, calculated from heads, grouped by water table depth classes. Original and ML-corrected (d) monthly and (e) annual mean water table depth bias (m) across the same water table depth classes.

the northern Great Plains in North America, and the Karoo Basin in South Africa (Figs. 9 & G1; Jasechko et al., 2024; Taylor et al., 2013).

3.5 Ensemble-mean Scenario Simulations

The trends in water table depths between 2015 and 2100 show that the different SSP-RCP scenarios do indeed suggest alternative futures for groundwater reserves (Figs. 10 & G2). At the global scale, more regions exhibit rising than declining water table depths (Figs. 10 & G2). The intensity of change and the proportion of areas projected to undergo change differ among the three SSP-RCP scenarios. More specifically, SSP1-RCP2.6 shows the lowest projected rate of change in water table depths, followed by SSP3-RCP7.0, whilst SSP5-RCP8.5 displays the largest projected rate of change (Figs. 10 & G2).

Many regions show consistent directions of change across all three scenarios; however, some differences are evident. The regions in which water table depths are projected to rise are the Sahel, Horn of Africa, North America's Pacific Coast, Tibetan Plateau, Northern Polar regions and Arabian Peninsula (Figs. 10 & G2). In contrast, water table depths are projected to decline in the Mediterranean, south eastern Australia, and the African and South American tropics (Figs. 10 & G2). The Alps and the Amazon are some of the few regions that show divergence in the direction of change. A declining water table depth is projected for the Amazon under SSP3-RCP7.0 and SSP5-RCP8.5, in contrast to a projected stable water table depth under SSP1-RCP2.6. Water table depth in the Alps is projected to rise under SSP1-RCP2.6 and decline under SSP3-RCP7.0 and SSP5-RCP8.5 scenarios (Figs. 10 & G2). Projected trends indicate that water tables will rise across the majority of continents in the com-

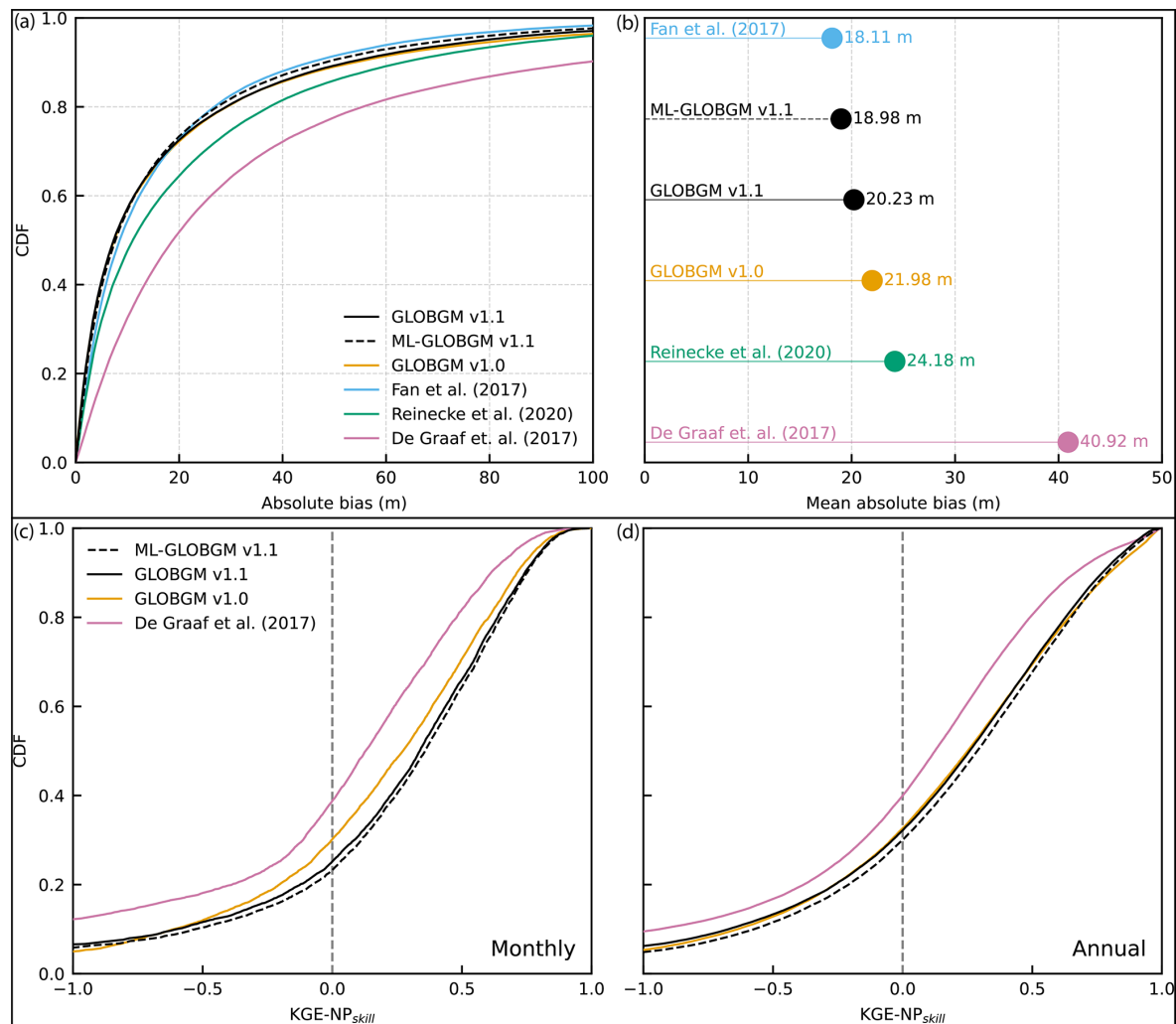


Figure 7. (a) CDF and (b) mean absolute bias in simulated steady-state water table depth validated against observations. (c) Monthly and (d) annual KGE-NP_{skill} scores for transient simulated hydraulic heads validated against observations.

ing century (Fig. 11). This rise is persistent in Africa and North America, accelerating under higher emission scenarios (Fig. 11b, d). In Asia, the water table remains relatively stable until the 2010s before beginning to rise, again intensifying with higher emissions (Fig. 11c). Australasia follows a similar upward trajectory before stabilizing around 2050, with minimal variation between scenarios (Fig. 11g). Conversely, Central and South America exhibit early increases followed by a divergence: SSP1-RCP2.6 leads to stability, while the other scenarios result in deepening water tables (Fig. 11e). Small islands show an initial rise until 2010 followed by a decline, with no discernible difference between the scenarios (Fig. 11f). Europe is the distinct outlier; it is the only region where the water table markedly deepens, particularly under higher emission scenarios, while SSP1-RCP2.6 maintains relative stability (Fig. 11a).

4 Discussion

4.1 Model Development and Methodological Advances

Compared to its predecessor, GLOBGM v1.1 incorporates four methodological changes: calibration, recharge downscaling, a dynamic elevation routine and an improved spin-up procedure. In addition to the raw model outputs, we also provided a dataset which was bias corrected using a machine learning based method.

The contributions of the dynamic elevation routine and the recharge downscaling routine yielded increased accuracy when compared to the uncalibrated GLOBGM v1.0 and v1.1. Nevertheless, accurate estimation of groundwater recharge remains a challenge in global hydrological modeling, with its estimates varying widely across models due to limited observational data and high spatial heterogeneity of recharge processes (Reinecke et al., 2024; Berghuijs et al., 2024).

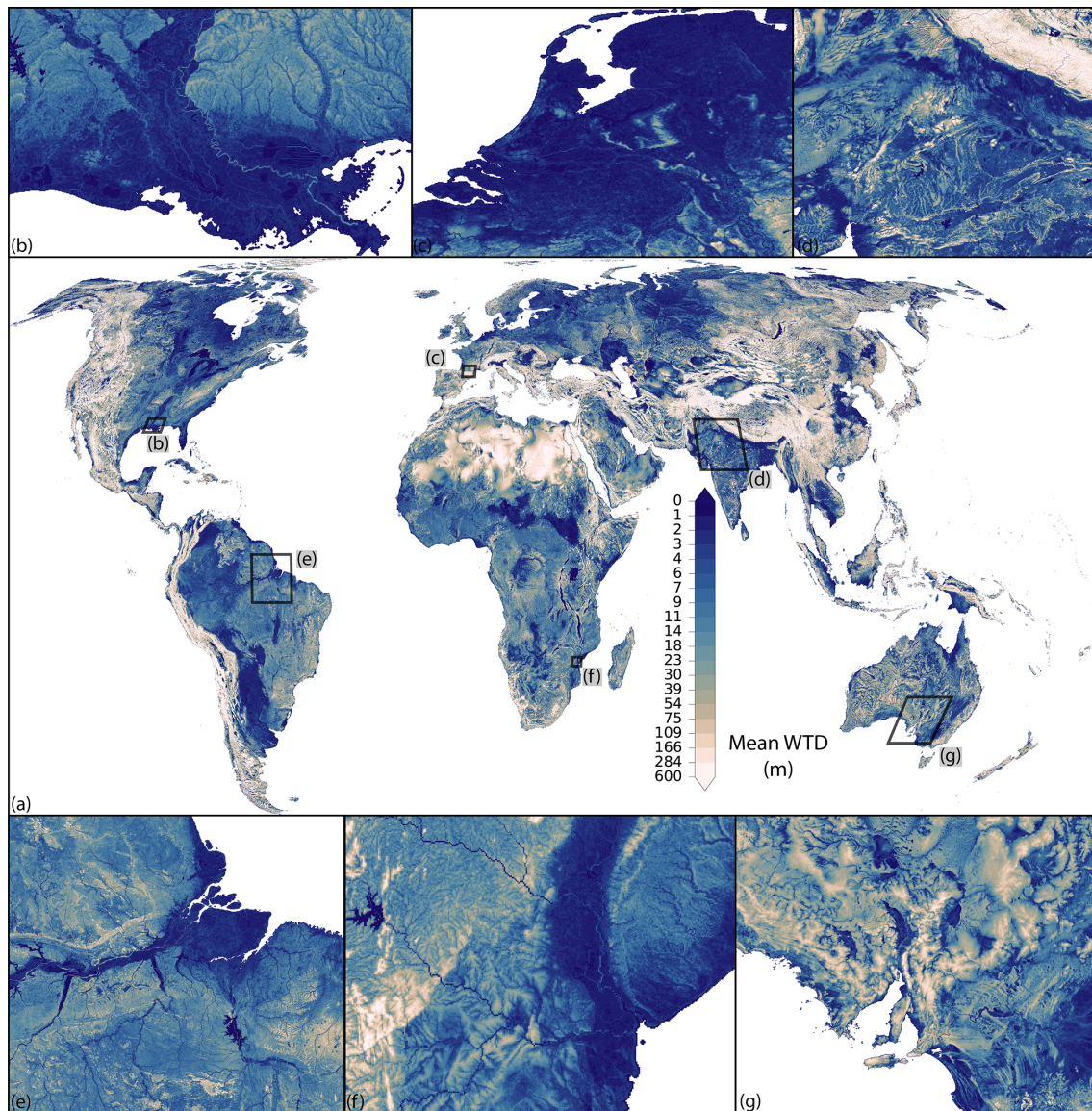


Figure 8. (a) Mean annual water table depth (1965–2019) simulated by GLOBGM. With zoomed insets providing more detail over (b) the Mississippi delta, (c) Netherlands, (d) Indo-Gangetic Plain, and (e) the Amazon basin, (f) Zambezi delta and (g) Southern Australia. In areas where a confining layer is present, only the uppermost model layer is depicted, thus providing a perspective from the ground surface downward.

Working within these limitations, we demonstrate that introducing a relatively simple statistical estimate of recharge at the hyper-resolution yields improved groundwater head estimates. In agreement, recent work has demonstrated that machine learning approaches are capable of predicting recharge fields at the global scale (Jung et al., 2024).

GLOBGM v1.1, calibrated through lithology-specific parameter adjustments, produces simulations with reduced bias and improved spatial agreement with observed water table depths, consistent with findings from regional and global studies reporting performance gains from calibrations (Fan et al., 2013; Zell and Sanford, 2020). However, applying

prefactors at the global-scale, or even based on lithological class as done here, shifts the entire distribution. Ideally, calibration would cause the upper and lower tails of the bias distribution to converge toward zero. Using global objective functions and calibration factors that influence large regions with very different hydrological characteristics inherently prevents incorporation of needed regional sensitivity; this is compounded by uncertainties in the datasets used to define lithological groupings and the uneven distribution of observations. Efforts focused on the collection and dissemination of observed groundwater data to broaden the scope and coverage of the current geographically constrained (Fig. C1)

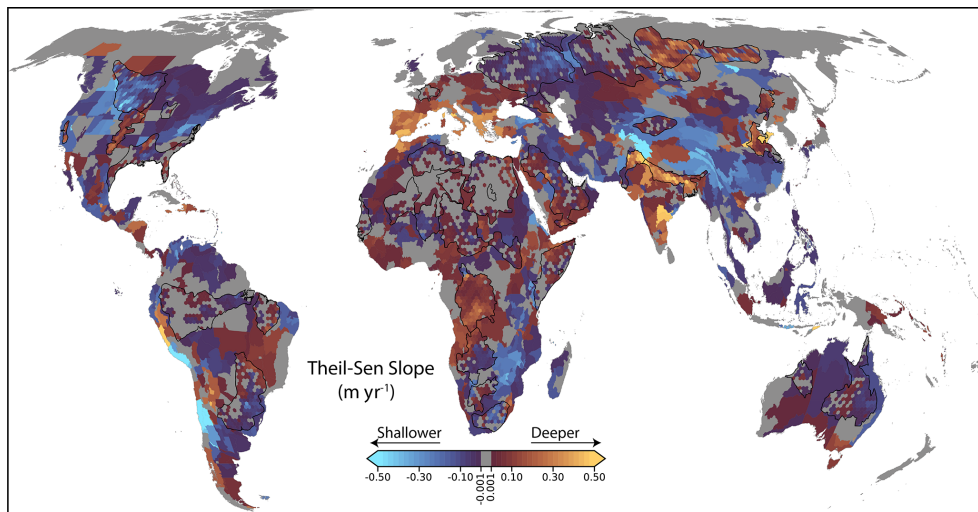


Figure 9. Theil-Sen slope of annual water table depth (1965–2019) simulated by GLOBGM. Gray pixels indicate regions where slopes are either insignificant, less than 0.001 m yr^{-1} , or both. Note that this figure represents the lower model layer which is representative of unconfined and confined aquifers.

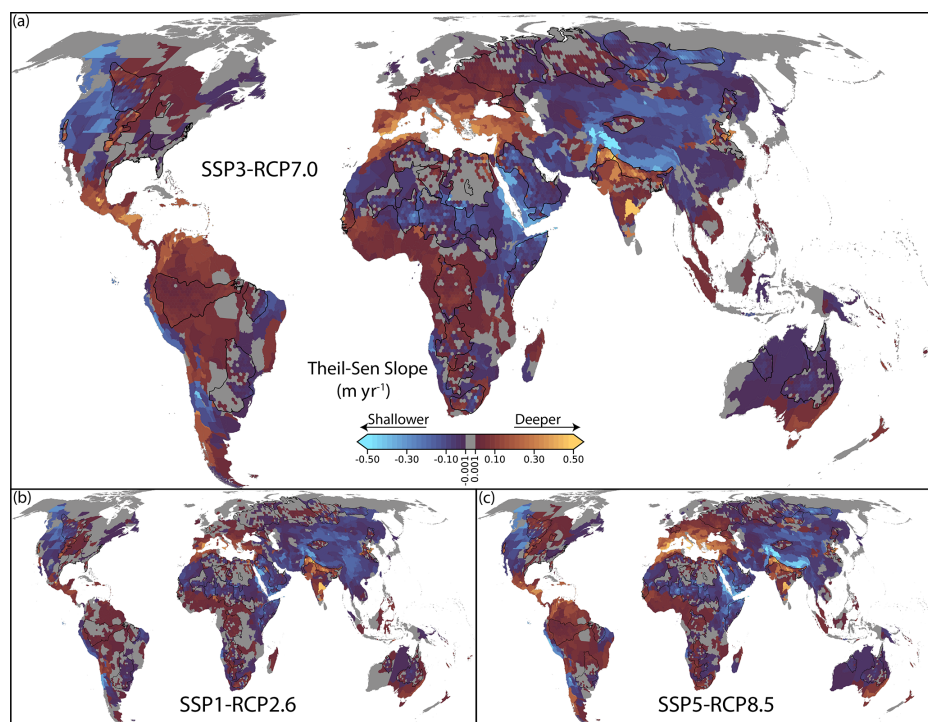


Figure 10. Theil-Sen slope of annual water table depth (2015–2100) simulated by GLOBGM under three SSP-RCP scenarios. Gray pixels indicate regions where slopes are either insignificant, less than 0.001 m yr^{-1} , or both. Note that this figure represents the lower model layer which is representative of unconfined and confined aquifers.

large sample datasets are central to achieving realistic parameter estimates through calibration (Jasechko and Perrone, 2021; Bätthge et al., 2026). Future studies should investigate whether alternative calibration techniques such as sequential calibration (Béraud et al., 2024) or data assimilation approaches that incorporate observed data (Béraud et al.,

2024; White, 2018) may aid in realizing locally specific calibration sensitivities. Similarly, regionalization and regularization methods have been shown to improve calibration of global hydrological models (Kupzig et al., 2024; Doherty, 2003).

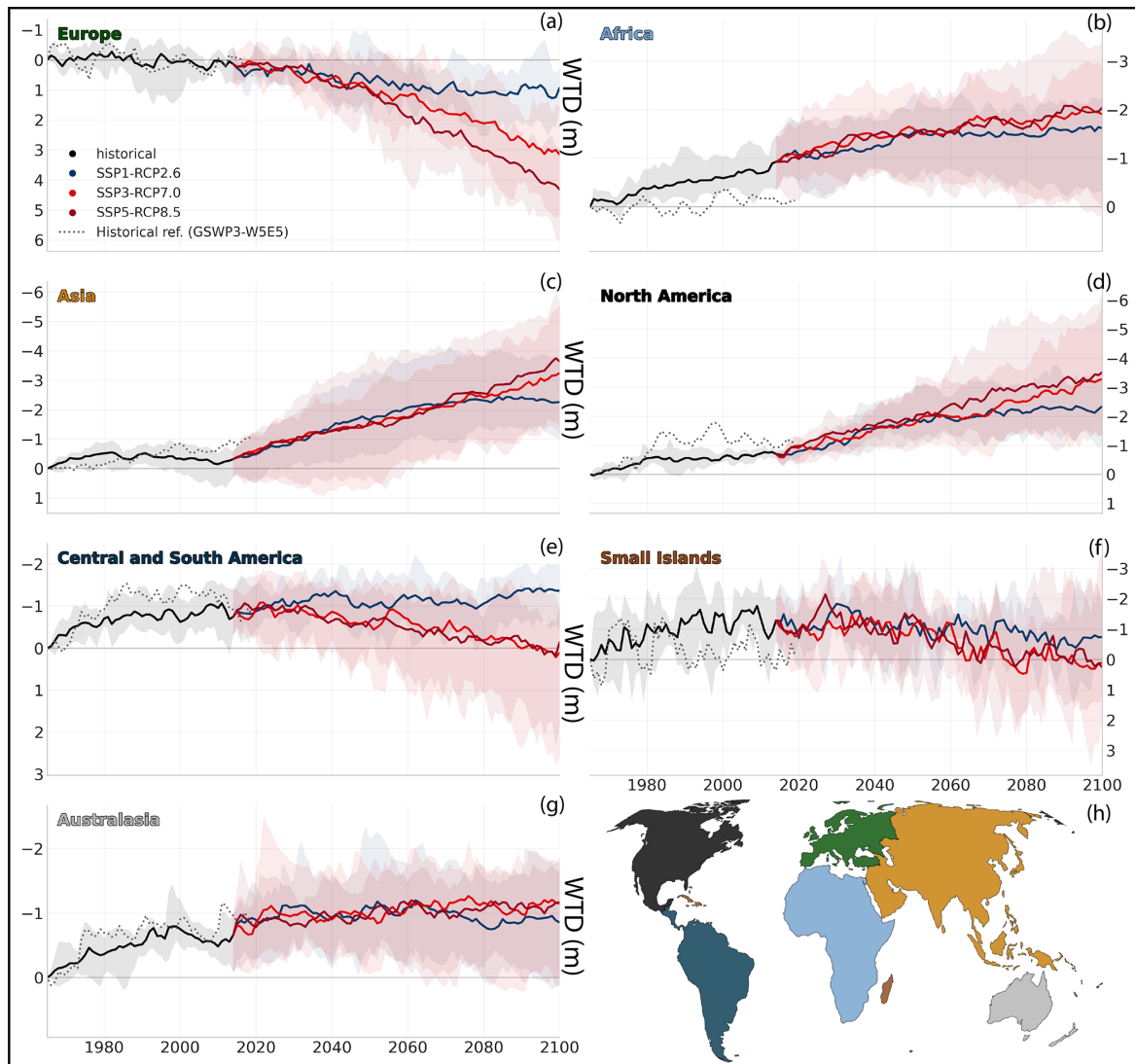


Figure 11. Time series of regional mean WTD (m) anomalies relative to 1965 for the historical-reference, historical and three Shared Socioeconomic Pathways: SSP1-RCP2.6, SSP3-RCP7.0, and SSP5-RCP8.5. Shaded bands depict the spread of the models (min–max). The map in panel (h) defines the aggregation regions.

The approach used to estimate initial groundwater heads resulted in a more accurate estimation of initial states than what would have been accomplished had the traditional approach been used (i.e., steady state simulation followed by transient sub-model without abstraction for 20 iterations). Nonetheless, results also suggest room for improvement. Ideally, the model would display strict dynamic equilibrium for all locations; however, this would require a disproportionate amount of computational power to achieve relatively small changes in initial groundwater head estimates. The estimation of accurate and representative initial states is an often cited problem for global groundwater models (Reinecke et al., 2024; de Graaf et al., 2017) and Earth system models in general (Werndl, 2019). Opportunities exist to improve this by using advances in initial condition estimation algo-

rithms, such as those based on sequence acceleration (Khatala, 2024) or by making use of artificial intelligence surrogate models where repetitive iterations can be simulated at a faster rate (Asher et al., 2015; Bennett et al., 2024; Pawusch et al., 2025; Triplett et al., 2025). Similar to the importance of regional sensitivities during calibration, future studies should aim to develop initial condition estimation frameworks that explicitly incorporate regional responses.

The two-stage bias correction method demonstrates that systematic errors in GLOBGM can be reduced using machine-learning-based methods, analogous to recent physics-guided ML groundwater digital twins (Ma et al., 2026b, a). The high importance of water table depth suggests that errors are linked to limitations in representing groundwater at different depths. This correction does not enforce mass

conservation; the corrected water table depths nonetheless benefit studies of groundwater-dependent ecosystems (Otoo et al., 2025).

4.2 GLOBGM v1.1 Validation, Performance, Historical Reference Simulation

The majority of the point-based observations (75 %–67 %) can be considered skillful, and when accounting for locations where poor performance is a result of metric artifacts, this increases to 80 % and 71 % of locations for the monthly and annual scales. Monthly validation reveals depth-dependent performance, with ~ 80 % skillful predictions for shallow-intermediate wells (0–20 m) versus 56 % for deep wells (> 60 m). In contrast, depth-dependent performance is less apparent at the annual resolution (~ 80 % skillful). This, combined with the overall lower accuracy at annual versus monthly timescales, suggests that the model captures seasonal variability more reliably than interannual variation. Reduced accuracy for deeper wells may partly be attributed to the delays that occur in deeper unsaturated zones that are not modeled in our approach.

GLOBGM and GRACE/GRACE-FO TWSA trends agree in direction for 62.3 % of the world and agree on direction and magnitude for 45.7 %. Sign disagreement is present over 7.1 % of land surface area. This is in line with other global hydrological models (Scanlon et al., 2018; Tiwari et al., 2025). Disagreement is prevalent in the higher latitudes where permafrost is present, and in mountainous and karst dominated regions (Sect. 3.2.1). Apart from limitations in representation of hydrogeology, the disagreement in high latitude and mountainous areas may also be due to trends in glacier and permafrost melt not included in our models. Disagreement over the Sahara and Sahel is consistent with poor meteorological forcing (Liu et al., 2024). Besides these regions, disagreement is prevalent in regions where recharge is governed by indirect, focused, and threshold driven processes (Cuthbert et al., 2019). A related limitation applies along the equatorial band, where models do not capture recharge generated by high-intensity precipitation (Jasechko and Taylor, 2015); this is especially relevant for East Africa (Taylor et al., 2013). Disagreement around large water bodies can be attributed to improper surface water dynamics within PCR-GLOBWB, which are especially sensitive to small errors in rainfall and evaporation in endorheic basins, and as such is not indicative of GLOBGM performance. Sign disagreement is also disproportionately common across small islands and archipelagos, including Indonesia, the Philippines, and the Caribbean. GRACE/GRACE-FO trends in these settings are subject to land-ocean leakage, in which gravity signals from ocean mass change contaminate the retrieval over adjacent narrow or fragmented land areas (Pfeffer et al., 2023). The comparison in these regions is then unreliable regardless of how GLOBGM performs.

GLOBGM v1.1 achieves sufficient skill for global-scale groundwater analysis and demonstrates clear progress against the current state of the art. At 30 arcsec resolution, GLOBGM v1.1 and its bias-corrected counterpart (ML-GLOBGM v1.1) perform comparably to Fan et al. (2017) and improve on GLOBGM v1.0. The two models trade off differently: Fan et al. (2017) achieves a lower mean absolute bias overall, but GLOBGM v1.1 captures water table depth more accurately at the majority of locations (~ 75 %), while Fan et al. (2017)'s lower mean error comes from stronger performance specifically at sites with moderate-to-large biases. Against the coarser-resolution (de Graaf et al., 2017) (5 arcmin), the 30 arcsec GLOBGM simulations improve on the mean benchmark by approximately 19 and 15 percentage points at monthly and annual resolutions, respectively. Within the GLOBGM lineage, v1.1 improves on v1.0 most at the monthly resolution, with a more modest gain annually, consistent with the model's stronger seasonal than interannual performance noted above.

Results are broadly consistent with the outputs of other global and regional groundwater models, though important model differences must be acknowledged. A key distinction is that GLOBGM explicitly incorporates human groundwater abstraction, a feature shared by some models but absent from others. Despite this structural difference, agreement across models is strong for well-documented depletion hotspots (de Graaf et al., 2017; Döll et al., 2014; Yang et al., 2025) such as declining trends in the High Plains Aquifer, California's Central Valley, Indo-Gangetic Plain, and North China Plain. Spatial agreement is strongest in sedimentary basins, which is where the majority of global groundwater use occurs (de Graaf et al., 2017). The relatively homogeneous subsurface structure is well suited to the parameterization approaches common across these models. Overall, confidence in GLOBGM simulations is highest in sedimentary basins (Verkaik et al., 2024; de Graaf et al., 2017). By contrast, performance in topographically complex terrain and mountainous regions remains uncertain in GLOBGM v1.1, which is the case for most global groundwater models. This limits the use of the results for assessing biodiversity, ecosystem functioning, and other non-consumptive uses of groundwater that depend on baseflow and shallow water table dynamics in these regions. Addressing this uncertainty represents an important frontier for the next generation of global groundwater models (Reinecke et al., 2024).

4.3 Scenario Based Simulations

When considering the future of global groundwater resources as simulated by GLOBGM v1.1, there is a tendency toward intensification of change in groundwater reserves in the coming century. Overall, more regions are projected to experience an increase in groundwater reserves. This finding aligns well with CMIP6 based studies showing an intensification of the hydrological cycle and is in congruence with previous re-

gional studies (Eccles et al., 2025; Berghuijs et al., 2024). For example, there is broad agreement between scenarios in terms of direction in the Sahel, Arabian Peninsula, Indus, and the Mediterranean. Stable water tables are projected for the Amazon, western Europe, and central Africa under SSP1-RCP2.6, yet under SSP3-RCP7.0 and SSP5-RCP8.5 water tables are projected to decline. These regional differences in direction and magnitude show the importance of uncertainty-based climate mitigation efforts as they can lead to divergent groundwater futures in certain regions.

GLOBGM differs from previous assessments of groundwater reserves by incorporating transient CMIP6 precipitation and temperature forcing, and its inclusion of groundwater abstraction to meet human water demand. Where recharge dominates the groundwater balance, GLOBGM broadly agrees with earlier projections; where abstraction is intensive, it diverges. GLOBGM projects a deepening water table over southern Europe, the Amazon, the California Central Valley, the central and southern High Plains Aquifer, and the Senegalo-Mauritanian Basin, and rising water tables over the Taoudeni-Tanezrouft Basin, the Lake Chad Basin, the Ogaden-Juba Basins, Iullemeden-Irhazer Aquifer System and Guarani Aquifer System. These patterns of rising and deepening are corroborated by a previous CMIP6-based water table depth study (Costantini et al., 2023) and recharge projections. Recharge is expected to decrease over the coming century in southern Europe, the Amazon, the California Central Valley, the central and southern High Plains Aquifer, and the Senegalo-Mauritanian Basin (Meixner et al., 2016; Berghuijs et al., 2024; Kumar et al., 2025; Hirata et al., 2025). In contrast, the Taoudeni-Tanezrouft Basin, the Lake Chad Basin, the Ogaden-Juba Basins, Iullemeden-Irhazer Aquifer System and Guarani Aquifer System are expected to receive increased recharge (Berghuijs et al., 2024). Over the United Kingdom, GLOBGM shows no consistent signal, in line with the current consensus that climate-change effects on UK groundwater will be modest (Jackson et al., 2015; Parry et al., 2024).

In several heavily pumped regions projections diverge. Over the Indus, Costantini et al. (2023) project a rising water table, in line with projected recharge increases (Berghuijs et al., 2024) and total water storage anomalies (Wu et al., 2020). Likewise, Costantini et al. (2023) projected rising water tables over Germany, where recharge is expected to remain stable (Kumar et al., 2025; Berghuijs et al., 2024), while GLOBGM projects a deepening water table depth. Because neither Costantini et al. (2023) nor Wu et al. (2020) represent groundwater abstraction, they likely underestimate depletion in these regions. Our simulated decline is consistent with a national scale, observation based, study which projects declining groundwater levels across most of Germany (Wunsch et al., 2022). Condon et al. (2020) isolated the effect of warming on water table depths over the western US, by holding precipitation constant and omitting anthropogenic water use and projected deepening water table depths in regions where

groundwater is shallow and connected to the root zone, while in already water limited arid regions little change is projected. GLOBGM simulates a more heterogeneous response with deepening over the California Central Valley and High Plains, where groundwater abstraction for irrigation is common, and rising water tables over the rest of the western US.

Taken together, this comparison shows that GLOBGM largely reproduces the sign of change projected by earlier climate-driven studies, whereas in heavily abstracted regions it projected deeper water tables. This reiterates that anthropogenic alteration of the water cycle is an important process to include in groundwater simulations (Galelli et al., 2025).

4.4 Synthesis

Because the data follow the ISIMIP framework, the same framework guides how they can be used. The historical reference simulation supports model evaluation against observations, as done here, and uncertainty studies that compare models (Reinecke et al., 2024, 2026). Future projections can be used to support studies that aim to investigate the climate-related risks to groundwater reserves under varying levels of climate change and socioeconomic conditions. Future users are especially encouraged to use the existing Earth systems datasets to relate groundwater dynamics to other Earth system components, as suggested in the review by Huggins et al. (2023). Besides the advantages provided by the high spatial and temporal resolution estimates of groundwater reserves presented here, the inclusion of multiple climate models across three combined climate and socioeconomic projections allows for the incorporation of uncertainty and enhances the robustness of results derived from future studies. For example, water table depth is a critical parameter when assessing the ecological impacts of groundwater depletion (Otoo et al., 2025). By analyzing these impacts using multiple models, it becomes possible to quantify uncertainty and improve confidence in projections. A pertinent case is southern Africa, a region identified here as potentially facing divergent groundwater futures depending on the trajectory of climate change (Sect. 3.5), highlighting the value of multi-GCM assessments in regions with high climatic and hydrological variability. As stated by Reinecke et al. (2026) in the proposed ISIMIP groundwater sector framework, multi-model groundwater heads and water table depth data may facilitate cross-sectoral research on the impact of changing groundwater reserves on social, economic, and ecological systems. Ultimately, the results presented here may serve as a baseline for future studies aimed at exploring the role of adaptation and mitigation strategies.

5 Conclusions

This study provides the first hyper-resolution (~ 1 km) monthly estimates of global groundwater heads and water table depths for past and future periods. Using the global

groundwater model GLOBGM in combination with CMIP6 climate projections for baseline and future periods, the simulated groundwater levels offer a valuable resource to enhance understanding of global groundwater at unexplored spatial and temporal scales. By incorporating dynamic drainage elevation, recharge downscaling, calibration, a novel spin-up procedure, and bias correction, simulations yield predictions that are skillful for 71 %–80 % of observations. Moreover, this work identifies opportunities for the research community to improve future modeling efforts, including the development of more efficient spin-up procedures and refined initial condition estimations and multi-groundwater model estimates and projections. By integrating anthropogenic and climatic influences on groundwater, this study provides a basis for assessing how global change affects groundwater availability and for evaluating groundwater dynamics within the broader Earth system.

Appendix A: Performance of the GAM in predicting groundwater recharge

The predictive skill of the Generalized Additive Model used for groundwater recharge was assessed by leave-one-out cross validation (Sect. 2.1.2). For each of the 5188 observations, the model was refit on the remaining 5187 and subsequently used to predict the single recharge point which was held out. These predictions were pooled and model performance was evaluated using the coefficient of determination R^2 , root-mean-square-error (RMSE), and the mean absolute error (MAE).

The fitted GAM demonstrated satisfactory predictive skill across the leave-one-out cross validation, achieving an R^2 of 0.7, RMSE of 0.2 m yr^{-1} and a MAE of 0.1 m yr^{-1} . The spatial distribution of the leave-one-out cross validation shows that for the majority of regions residuals are centered around zero (Fig. A1).

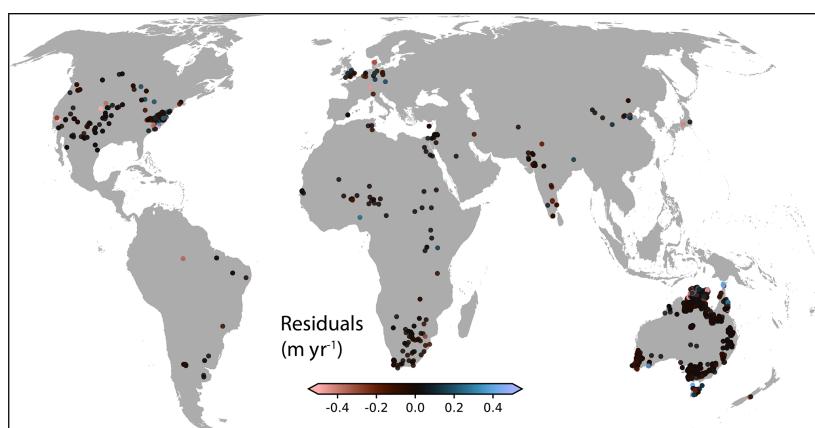


Figure A1. Global distribution of leave-one-out cross-validation residuals for the Generalized Additive Model (GAM) of groundwater recharge.

Appendix B: Derivation of the KGE-NP Mean Flow Benchmark Threshold and Skill Score Normalization

Knoben et al. (2019) define a benchmark as the mean flow simulation, defined as $\text{Sim}(t) = \overline{\text{Obs}}$ for all time steps t . Under this benchmark, $\beta = 1$: since $\mu_{\text{sim}} = \overline{\text{Obs}} = \mu_{\text{obs}}$ by definition, $\alpha = 0$: since the simulated time series is constant, $\sigma_{\text{sim}} = 0$ and $r = 0$: since a constant series has no temporal variation, the linear correlation with observations is formally undefined; following Knoben et al. (2019), it is assigned a value of zero, as there is no relationship between the fluctuations of the two series. Substituting these values into the standard KGE equation:

$$\begin{aligned} \text{KGE}_{\text{bench}} &= 1 - \sqrt{(0 - 1)^2 + (0 - 1)^2 + (1 - 1)^2} \\ &= 1 - \sqrt{2} \approx -0.41 \end{aligned} \quad (\text{B1})$$

Critically, the threshold of -0.41 holds for any observation time series, making it a universal benchmark for the standard KGE. The non-parametric KGE (KGE-NP; Pool et al., 2018) replaces the parametric components of the standard KGE with rank-based and distribution-free analogues:

$$\text{KGE-NP} = 1 - \sqrt{(r_s - 1)^2 + (\alpha_{\text{NP}} - 1)^2 + (\beta - 1)^2} \quad (\text{B2})$$

where r_s is the Spearman rank correlation between simulated and observed, $\alpha_{\text{NP}} = 1 - \text{MAE}(\text{FDC}_{\text{sim}}, \text{FDC}_{\text{obs}})$ is a non-parametric variability ratio derived from the mean absolute error between the normalized flow duration curves (FDCs) of simulated and observed, and $\beta = \mu_{\text{sim}}/\mu_{\text{obs}}$ is the bias ratio, unchanged from the standard KGE.

Applying the same mean flow benchmark ($\text{Sim}(t) = \overline{\text{Obs}}$ for all t) to the KGE-NP formulation: $\beta = 1$, $r = 0$. However for α_{NP} the following applies:

$$\alpha_{\text{NP}} = 1 - \text{MAE}(\text{FDC}_{\text{sim}}, \text{FDC}_{\text{obs}}) \quad (\text{B3})$$

$$= 1 - \frac{1}{n \overline{\text{Obs}}} \sum_{i=1}^n |\text{Obs}_i^{\text{sorted}} - \overline{\text{Obs}}_{\text{obs}}| \quad (\text{B4})$$

$$= 1 - \frac{\text{MAD}_{\text{obs}}}{\mu_{\text{obs}}} \quad (\text{B5})$$

Substituting all three terms of the mean flow benchmark into the KGE-NP equation:

$$\begin{aligned} \text{KGE-NP}_{\text{bench}} &= 1 - \sqrt{(0 - 1)^2 + \left(-\frac{\text{MAD}_{\text{obs}}}{\mu_{\text{obs}}}\right)^2 + (1 - 1)^2} \\ &= 1 - \sqrt{1 + \left(\frac{\text{MAD}_{\text{obs}}}{\mu_{\text{obs}}}\right)^2} \end{aligned} \quad (\text{B6})$$

Therefore, unlike the standard KGE, the KGE-NP benchmark threshold is not universal since it depends on $\text{MAD}_{\text{obs}}/\mu_{\text{obs}}$, the coefficient of mean absolute deviation of the observed record. As a result, a single fixed KGE-NP threshold analogous to -0.41 cannot be defined. To recover a universal metric suitable for global-scale evaluation and multi-model comparison, KGE-NP scores were normalized into a skill score following Knoben et al. (2019):

$$\text{KGE-NP}_{\text{skill}} = \frac{\text{KGE-NP}_{\text{model}} - \text{KGE-NP}_{\text{bench}}}{1 - \text{KGE-NP}_{\text{bench}}} \quad (\text{B7})$$

A $\text{KGE-NP}_{\text{skill}} < 0$ indicates that a model is worse than the mean flow benchmark, $\text{KGE-NP}_{\text{skill}} = 0$ indicates that a model is equal to the mean flow benchmark, and $\text{KGE-NP}_{\text{skill}} > 0$ indicates that a model is better than the mean flow benchmark. The benchmark $\text{KGE-NP}_{\text{bench}}$ is computed individually for each well from its observed record and places all wells on a common scale, where zero universally denotes the boundary between a model that adds value over the mean flow benchmark and one that does not.

Appendix C: Spatio-temporal Distribution in observations of groundwater head

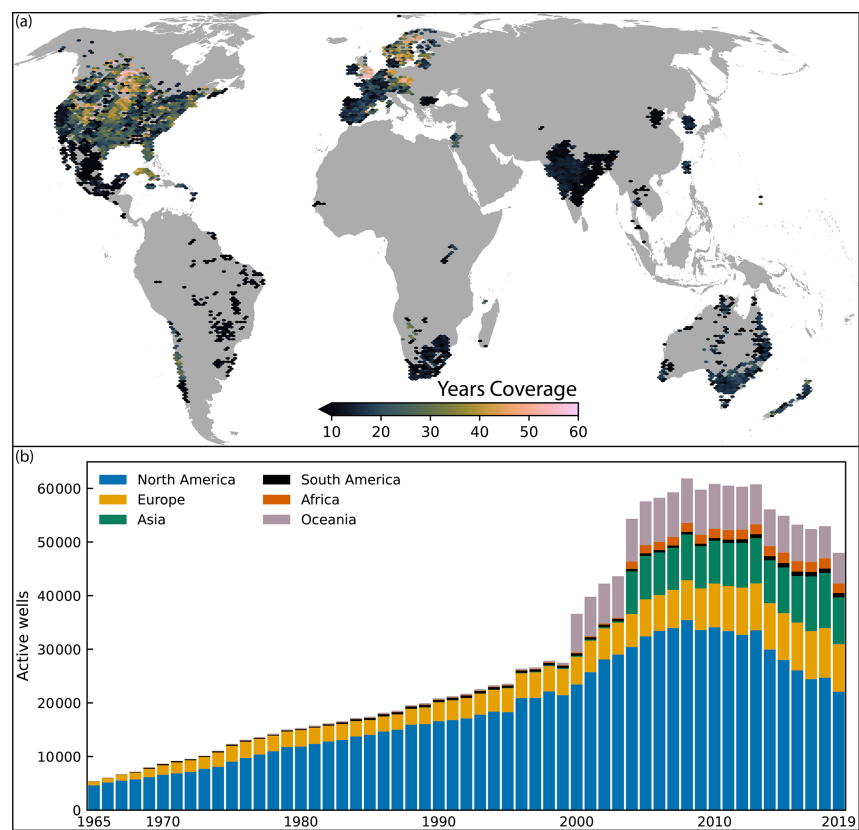


Figure C1. Global distribution and temporal coverage of groundwater observations. **(a)** Spatial distribution of active groundwater monitoring wells worldwide, colored by the number of years of available data. **(b)** Number of active groundwater monitoring wells by continent from 1965 to 2019.

Appendix D: KGE- NP_{skill} Post-hoc Analysis

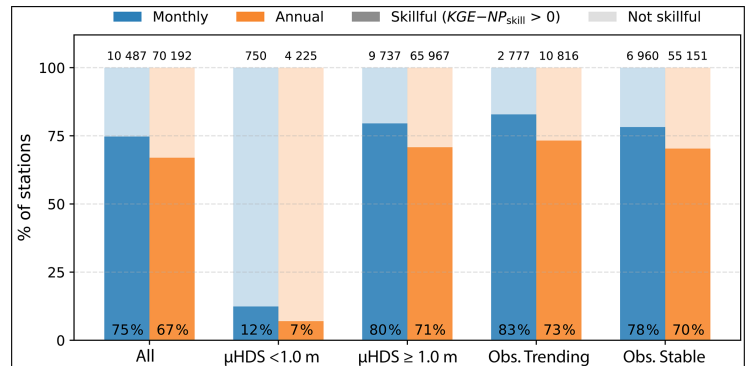


Figure D1. Grouped stacked bar charts showing the percentage of skillful validations, defined as $KGE-NP_{skill} > 0$. Annotations at the top of each bar indicate the total number of observation wells. The percentages at the bottom of show the proportion of observation wells classified as skillful. Wells are split into trending/stable by their absolute Theil-Sen slope using a threshold of 0.01 m yr^{-1} . We choose this threshold instead of the 0.1 m yr^{-1} of Jasechko et al. (2024) because this is more conservative in defining a trends within the stable class.

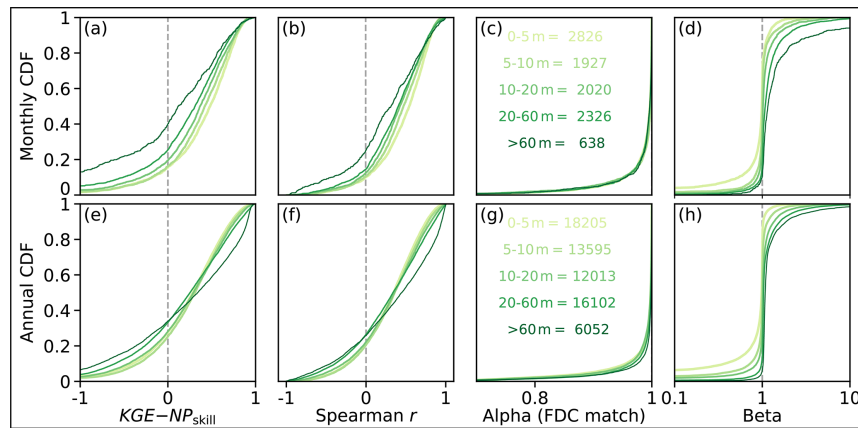


Figure D2. (a, e) KGE-NP_{skill} CDF filtered for locations where statistical artifacts result in poor performance. Constituent components (b, f) correlation, (c, g) alpha, and (d, h) beta of KGE-NP for monthly and annual groundwater heads simulated by GLOBGM from 1965–2019, grouped by water table depth. Gray lines in panels (a) and (e) symbolize a KGE-NP_{skill} of 0 which is the threshold at which greater values indicate predictions that improve upon the mean flow benchmark. Gray lines in panels (b) and (f) indicate the point at which correlations are positive to the right and negative to the left. Gray lines in panels (d) and (h) separate over-estimation to the right and underestimation to the left.

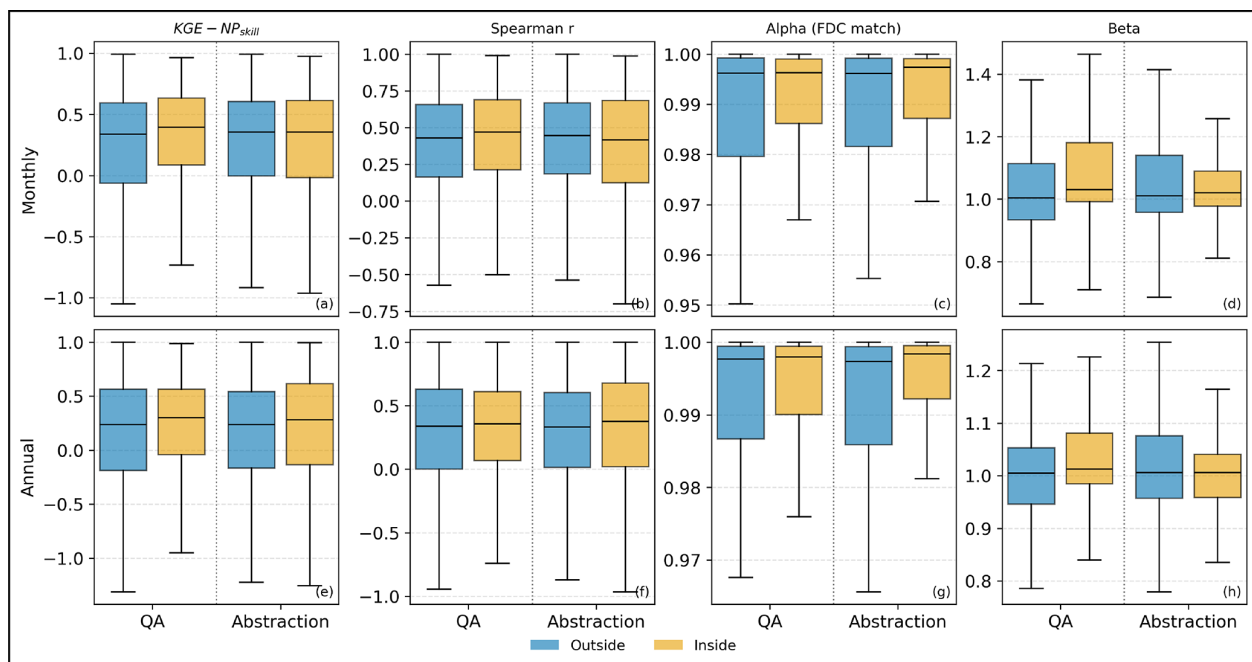


Figure D3. (a, e) Boxplots comparing KGE-NP_{skill} inside and outside quality assurance regions (QA, Sect. 3.2.1) and abstraction regions, for monthly (a–d) and annual (e–h) and its constituent components (b, f) correlation, (c, g) alpha, and (d, h) beta.

Appendix E: GRACE/GRACE-FO Comparison

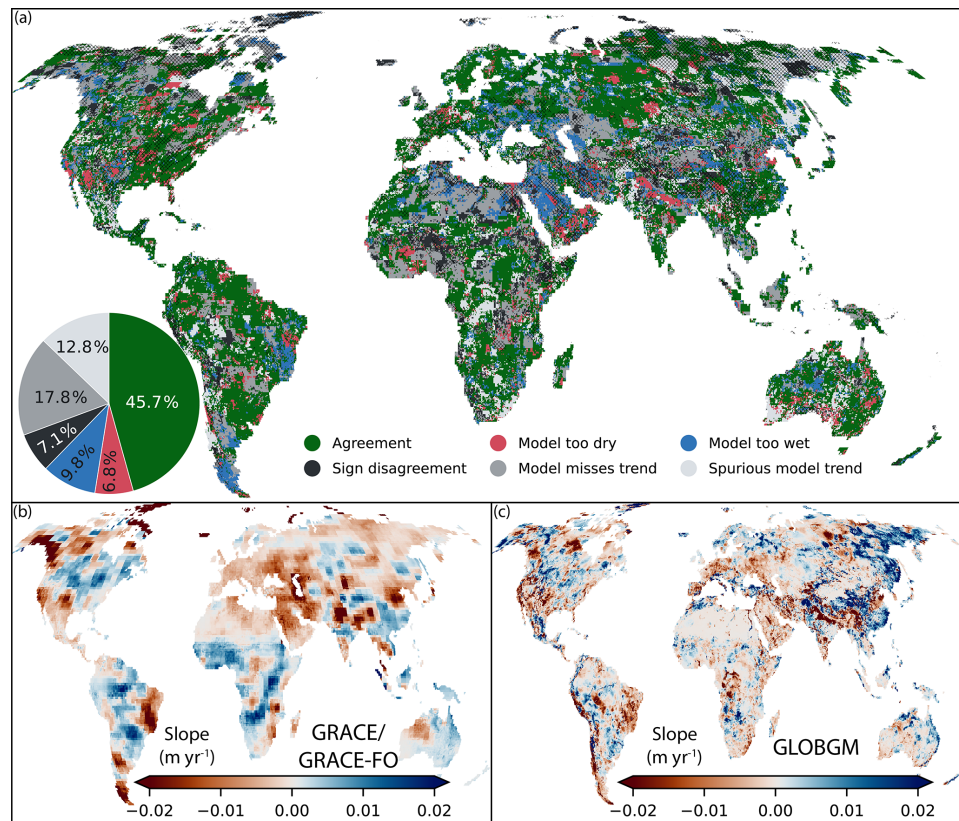


Figure E1. (a) Agreement between GRACE/GRACE-FO and GLOBGM Total Water Storage Anomalies with the inset showing the proportion of land area per agreement class and hatching indicates cells where more than half the underlying 30 arcsec area carries a model quality flag (mountainous terrain, karst, or permafrost). Theil-Sen trends in deseasonalised Total Water Storage Anomalies (2002–2019) from (b) GRACE/GRACE-FO and (c) GLOBGM.

Appendix F: Machine Learning Bias Correction

Table F1. Performance metrics of the two-stage correction ML model for the training, validation, and holdout test sets.

Stage	Metric	Train	Validation	Test (Holdout)
Stage 1: Binary Classifier	Log-loss	0.32	0.47	0.50
	Sign Accuracy	0.9	0.8	0.8
Stage 2: Magnitude Regressor	RMSE	37.4	25.3	36.7

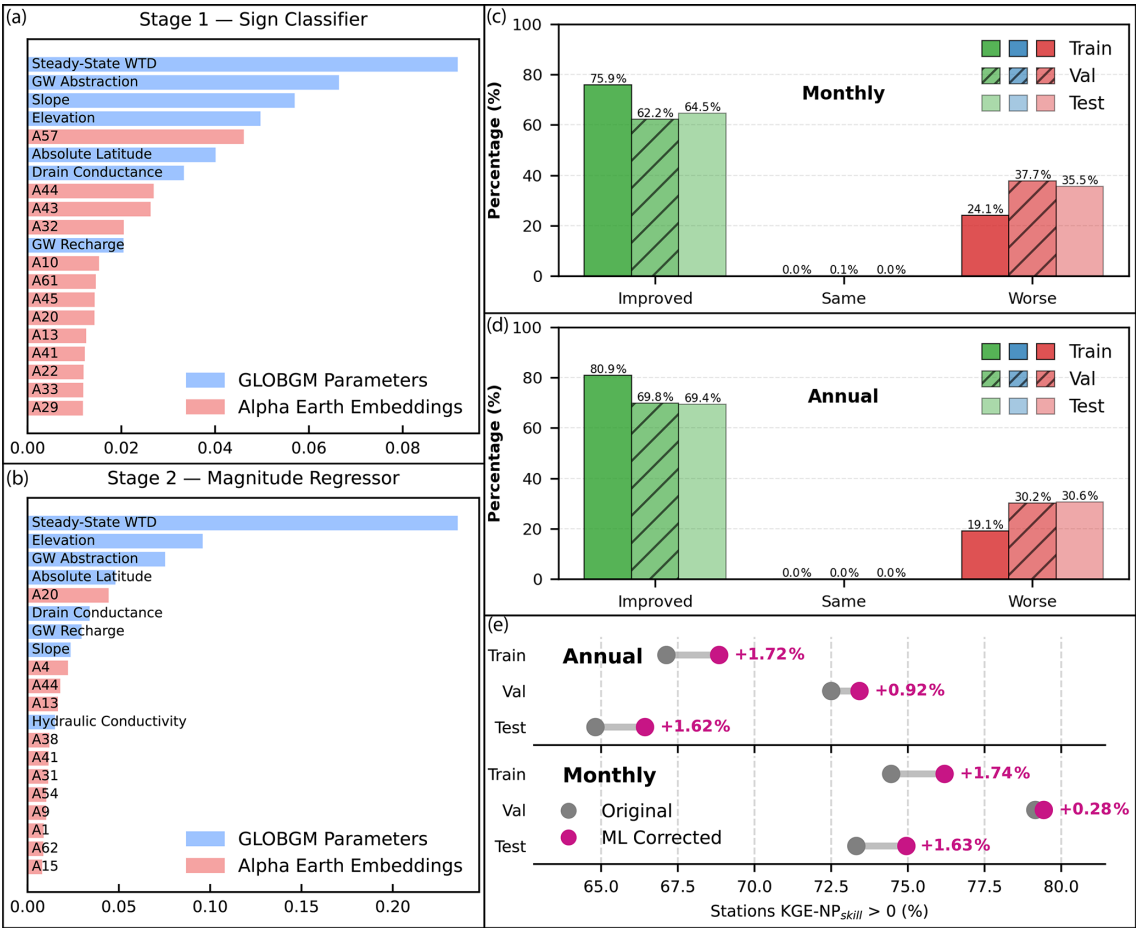


Figure F1. Feature importance of predictors of the (a) binary classifier and (b) magnitude regressor LightGBM models. Percentage of observation wells where bias improved, stayed the same, or worsened at the monthly (c) and annual (d) resolution, separated by train, validation, and test sets. (e) Percentage of observations where $KGE-NP_{skill} > 0$ before and after bias correction.

Appendix G: Unaggregated Annual Theil-Sen Trends

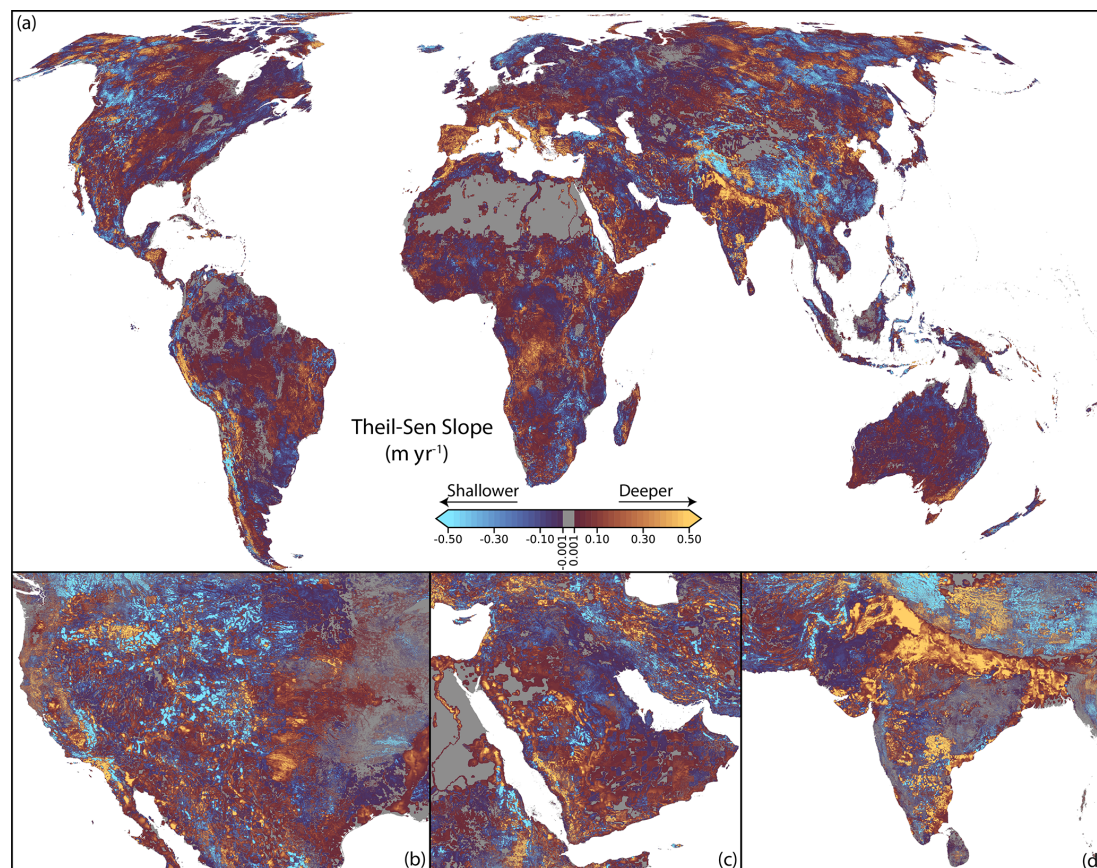


Figure G1. (a) Theil-Sen slope of annual water table depth (1965–2019) simulated by GLOBGM. Gray pixels indicate regions where slopes are either insignificant, less than 0.001 m yr⁻¹, or both. Note that this figure represents the lower model layer which is representative of unconfined and confined aquifers. Zoomed insets provide more detail over (b) United States, (c) the Arabian Peninsula, and (d) the Indo-Gangetic Plain.

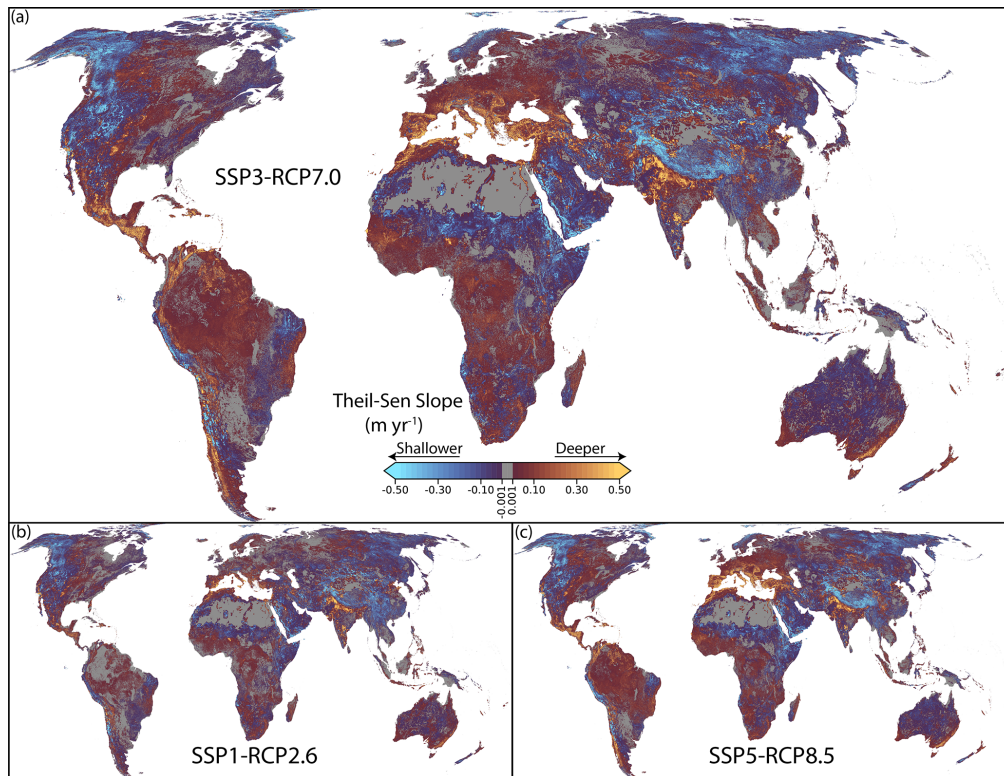


Figure G2. Theil-Sen slope of annual water table depth (2015–2100) simulated by GLOBGM under three SSP-RCP scenarios. Gray pixels indicate regions where slopes are either insignificant, less than 0.001 m yr^{-1} , or both. Note that this figure represents the lower model layer which is representative of unconfined and confined aquifers.

Code and data availability. The model code used to run the simulations has been archived on Zenodo and is accessible via <https://doi.org/10.5281/zenodo.17065147> (van Jaarsveld et al., 2025). Simulated groundwater heads and water table depths are available on YODA, the research data management service of Utrecht University. For the historical reference simulation (ISIMIP3a), the dataset includes long-term, annual, and monthly groundwater head and water table depth averages. For the future projections (ISIMIP3b), long-term and annual averages are provided for each GCM under the three SSP-RCP scenarios. Ensemble means for long-term, annual, and monthly estimates of groundwater head and water table depth are provided across the three scenarios. The raw model outputs can be accessed directly: historical-reference-gswp3-w5e5: <https://doi.org/10.24416/UU01-AKSHOX> (van Jaarsveld, 2025a), globgm-cmip6-monthly: <https://doi.org/10.24416/UU01-1BXLDP> (van Jaarsveld, 2025b), globgm-cmip6-annual: <https://doi.org/10.24416/UU01-V6B9YS> (van Jaarsveld, 2025c), globgm-cmip6-average: <https://doi.org/10.24416/UU01-SLRFI7> (van Jaarsveld, 2025d). The spin-up information, quality assurance, GRACE/GRACE-FO agreement, and the machine learning bias correction datasets are available at <https://doi.org/10.24416/UU01-16EJ3Y> (van Jaarsveld, 2025e). Instructions and code for accessing the data are available at https://vanjaarsveldbarry.github.io/globgm_cmip6/ (last access: 23 August 2026).

Author contributions. BvJ: Data curation, Formal analysis, Software, Validation, Visualization, Writing – original draft preparation. NW: Conceptualization, Project administration, Supervision, Writing – review & editing. NGO: Conceptualization, Data curation, Software, Writing – review & editing. EHS: Conceptualization, Data curation, Software, Writing – review & editing. JV: Data curation, Software, Writing – review & editing. DZ: Conceptualization, Data curation, Writing – review & editing. MFPB: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative (grant no. EINF-11758, EINF-9757 & NWO-2024.031). Huub Stoffers and Stefan Wolfsheimer are thanked for their support regarding the use of Snellius. MB acknowledges support from the ERC Advanced Grant scheme (Grant no. 101019185). Myrthe Leijnse is thanked for guidance in accessing and using the hotspot polygons. Maisam Mohammadi Dadkan and Vincent Brunst are thanked for their guidance and assistance with hosting data on YODA.

Financial support. This research has been supported by the European Research Council, EU HORIZON EUROPE European Research Council (grant no. 101019185).

Review statement. This paper was edited by Gabriele Messori and reviewed by Richard Taylor and one anonymous referee.

References

- Asher, M. J., Croke, B. F. W., Jakeman, A. J., and Peeters, L. J. M.: A Review of Surrogate Models and Their Application to Groundwater Modeling, *Water Resour. Res.*, 51, 5957–5973, <https://doi.org/10.1002/2015WR016967>, 2015.
- Bäthge, A., Ruz Vargas, C., Lischeid, G., Collenteur, R., Cuthbert, M., Fleckenstein, J., Flörke, M., de Graaf, I., Gnann, S., Hartmann, A., Huggins, X., Moosdorf, N., Wada, Y., Wagener, T., and Reinecke, R.: A Global-Scale Time Series Dataset for Groundwater Studies within the Earth System, *Scientific Data*, 13, 401, <https://doi.org/10.1038/s41597-026-06966-1>, 2026.
- Bennett, A., Tran, H., De La Fuente, L., Triplett, A., Ma, Y., Melchior, P., Maxwell, R. M., and Condon, L. E.: Spatio-Temporal Machine Learning for Regional to Continental Scale Terrestrial Hydrology, *J. Adv. Model. Earth Sy.*, 16, e2023MS004095, <https://doi.org/10.1029/2023MS004095>, 2024.
- Béraud, T., Claprood, M., and Gloaguen, E.: A Sequential Ensemble Smoother for Multiple Data Assimilation in Hydrogeological Modeling, *Front. Water*, 6, <https://doi.org/10.3389/frwa.2024.1462914>, 2024.
- Berghuijs, W. R., Collenteur, R. A., Jasechko, S., Jaramillo, F., Luijendijk, E., Moeck, C., van der Velde, Y., and Allen, S. T.: Groundwater Recharge Is Sensitive to Changing Long-Term Aridity, *Nat. Clim. Change*, 14, 357–363, <https://doi.org/10.1038/s41558-024-01953-z>, 2024.
- Bierkens, M. F. P.: Global Hydrology 2015: State, Trends, and Directions, *Water Resour. Res.*, 51, 4923–4947, <https://doi.org/10.1002/2015WR017173>, 2015.
- Bierkens, M. F. P. and Wada, Y.: Non-Renewable Groundwater Use and Groundwater Depletion: A Review, *Environ. Res. Lett.*, 14, 063002, <https://doi.org/10.1088/1748-9326/ab1a5f>, 2019.
- Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., De Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., Van De Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-Resolution Global Hydrological Modelling: What Is next?: “Everywhere and Locally Relevant”, *Hydrol. Process.*, 29, 310–320, <https://doi.org/10.1002/hyp.10391>, 2015.
- Bierkens, M. F. P., Sutanudjaja, E. H., and Wanders, N.: Large-scale sensitivities of groundwater and surface water to groundwater withdrawal, *Hydrol. Earth Syst. Sci.*, 25, 5859–5878, <https://doi.org/10.5194/hess-25-5859-2021>, 2021.
- Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D’Andrea, F., Davini, P., De Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J.-L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M.-A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J.-Y., Guenet, B., Guez, E., L., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J.-B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the IPSL-CM6A-LR Climate Model, *J. Adv. Model. Earth Sy.*, 12, e2019MS002010, <https://doi.org/10.1029/2019MS002010>, 2020.
- Brown, C. F., Kazmierski, M. R., Pasquarella, V. J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., and Kohli, P.: AlphaEarth Foundations: An Embedding Field Model for Accurate and Efficient Global Mapping from Sparse Label Data, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2507.22291>, 2025.
- Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L., and Karger, D. N.: CHELSA-BIOCLIM+ A Novel Set of Global Climate-Related Predictors at Kilometre-Resolution, *EnviDat*, <https://doi.org/10.16904/ENVIDAT.332>, 2022a.
- Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L., and Karger, D. N.: Global climate-related predictors at kilometer resolution for the past and future, *Earth Syst. Sci. Data*, 14, 5573–5603, <https://doi.org/10.5194/essd-14-5573-2022>, 2022b.
- Chen, Z., Goldscheider, N., Auler, A., Bakalowicz, M., Broda, S., Drew, D., Hartmann, J., Jiang, G., Moosdorf, N., Richts, A., Stevanovic, Z., Veni, G., Dumont, A., Aureli, A., Clos, P., and Krombholz, M.: World Karst Aquifer Map (WHYMAP WOKAM), BGR, IAH, KIT, UNESCO, https://doi.org/10.25928/B2.21_SFKQ-R406, 2017.
- Cochand, M., Molson, J., and Lemieux, J.-M.: Groundwater Hydrogeochemistry in Permafrost Regions, *Permafrost Periglac.*, 30, 90–103, <https://doi.org/10.1002/ppp.1998>, 2019.
- Condon, L. E., Atchley, A. L., and Maxwell, R. M.: Evapotranspiration Depletes Groundwater under Warming over the Contiguous United States, *Nat. Commun.*, 11, 873, <https://doi.org/10.1038/s41467-020-14688-0>, 2020.
- Condon, L. E., Kollet, S., Bierkens, M. F. P., Fogg, G. E., Maxwell, R. M., Hill, M. C., Fransen, H.-J. H., Verhoef, A., Van Loon, A. F., Sulis, M., and Abesser, C.: Global Groundwater Modeling and Monitoring: Opportunities

- and Challenges, *Water Resour. Res.*, 57, e2020WR029500, <https://doi.org/10.1029/2020WR029500>, 2021.
- Costantini, M., Colin, J., and Decharme, B.: Projected Climate-Driven Changes of Water Table Depth in the World's Major Groundwater Basins, *Earths Future*, 11, e2022EF003068, <https://doi.org/10.1029/2022EF003068>, 2023.
- Custodio, E.: Intensive Groundwater Development: A Water Cycle Transformation, a Social Revolution, a Management Challenge, in: *Re-Thinking Water and Food Security*, CRC Press, ISBN 978-0-429-21299-4, 2010.
- Cuthbert, M. O., Taylor, R. G., Favreau, G., Todd, M. C., Shamsud-duha, M., Villholth, K. G., MacDonald, A. M., Scanlon, B. R., Kotchoni, D. O. V., Vouillamoz, J.-M., Lawson, F. M. A., Adjomayi, P. A., Kashaigili, J., Seddon, D., Sorensen, J. P. R., Ebrahim, G. Y., Owor, M., Nyenje, P. M., Nazoumou, Y., Goni, I., Ousmane, B. I., Sibanda, T., Ascott, M. J., Macdonald, D. M. J., Agyekum, W., Koussoubé, Y., Wanke, H., Kim, H., Wada, Y., Lo, M.-H., Oki, T., and Kukuric, N.: Observed Controls on Resilience of Groundwater to Climate Variability in Sub-Saharan Africa, *Nature*, 572, 230–234, <https://doi.org/10.1038/s41586-019-1441-7>, 2019.
- de Graaf, I. E. M., van Beek, L. P. H., Gleeson, T., Moosdorf, N., Schmitz, O., Sutanudjaja, E. H., and Bierkens, M. F. P.: A Global-Scale Two-Layer Transient Groundwater Model: Development and Application to Groundwater Depletion, *Adv. Water Resour.*, 102, 53–67, <https://doi.org/10.1016/j.advwatres.2017.01.011>, 2017.
- de Graaf, I. E. M., Gleeson, T., van Beek, L. P. H., Sutanudjaja, E. H., and Bierkens, M. F. P.: Environmental Flow Limits to Global Groundwater Pumping, *Nature*, 574, 90–94, <https://doi.org/10.1038/s41586-019-1594-4>, 2019.
- de Graaf, I. E. M., Marinelli, B., and Liu, S.: Global Analysis of Groundwater Pumping from Increased River Capture, *Environ. Res. Lett.*, 19, 044064, <https://doi.org/10.1088/1748-9326/ad383d>, 2024.
- Diak, M., Böttcher, M. E., Ehlert Von Ahn, C. M., Hong, W.-L., Kędra, M., Kotwicki, L., Kozirowska-Makuch, K., Kuliński, K., Lepland, A., Makuch, P., Sen, A., Winogradow, A., Silberberger, M. J., and Szymczycha, B.: Permafrost and Groundwater Interaction: Current State and Future Perspective, *Front. Earth Sci.*, 11, 1254309, <https://doi.org/10.3389/feart.2023.1254309>, 2023.
- Doherty, J.: Ground Water Model Calibration Using Pilot Points and Regularization, *Groundwater*, 41, 170–177, <https://doi.org/10.1111/j.1745-6584.2003.tb02580.x>, 2003.
- Döll, P., Hoffmann-Dobrev, H., Portmann, F. T., Siebert, S., Eicker, A., Rodell, M., Strassberg, G., and Scanlon, B. R.: Impact of Water Withdrawals from Groundwater and Surface Water on Continental Water Storage Variations, *J. Geodyn.*, 59–60, 143–156, <https://doi.org/10.1016/j.jog.2011.05.001>, 2012.
- Döll, P., Fritsche, M., Eicker, A., and Müller Schmied, H.: Seasonal Water Storage Variations as Impacted by Water Abstractions: Comparing the Output of a Global Hydrological Model with GRACE and GPS Observations, *Surv. Geophys.*, 35, 1311–1331, <https://doi.org/10.1007/s10712-014-9282-2>, 2014.
- Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., De Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y., Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., Van Der Schalie, R., Seneviratne, S. I., Smolander, T., and Lecomte, P.: ESA CCI Soil Moisture for Improved Earth System Understanding: State-of-the-Art and Future Directions, *Remote Sens. Environ.*, 203, 185–215, <https://doi.org/10.1016/j.rse.2017.07.001>, 2017.
- Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., Stock, C. A., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P. G., Griffies, S. M., Guo, H., Hallberg, R. W., Harrison, M., He, J., Hurlin, W., McHugh, C., Menzel, R., Milly, P. C. D., Nikonov, S., Paynter, D. J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B. G., Robinson, T., Schwarzkopf, D. M., Sentman, L. T., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, Y., and Zhao, M.: The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall Coupled Model Description and Simulation Characteristics, *J. Adv. Model. Earth Sy.*, 12, e2019MS002015, <https://doi.org/10.1029/2019MS002015>, 2020.
- Eccles, R., Syktus, J., Trancoso, R., Chapman, S., Wasko, C., Evans, J. P., Thatcher, M., Di Virgilio, G., and Stassen, C.: Substantial Increases in Future Precipitation Extremes – Insights from a Large Ensemble of Downscaled CMIP6 Models, *npj Nat. Hazards*, 2, 60, <https://doi.org/10.1038/s44304-025-00107-1>, 2025.
- Fan, Y., Li, H., and Miguez-Macho, G.: Global Patterns of Groundwater Table Depth, *Science*, 339, 940–943, <https://doi.org/10.1126/science.1229881>, 2013.
- Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B., and Otero-Casal, C.: Hydrologic Regulation of Plant Rooting Depth, *P. Natl. Acad. Sci. USA*, 114, 10572–10577, <https://doi.org/10.1073/pnas.1712381114>, 2017.
- Felfelani, F., Wada, Y., Longuevergne, L., and Pokhrel, Y. N.: Natural and Human-Induced Terrestrial Water Storage Change: A Global Analysis Using Hydrological Models and GRACE, *J. Hydrol.*, 553, 105–118, <https://doi.org/10.1016/j.jhydrol.2017.07.048>, 2017.
- Ford, T. W. and Frauenfeld, O. W.: Surface–Atmosphere Moisture Interactions in the Frozen Ground Regions of Eurasia, *Sci. Rep.*, 6, 19163, <https://doi.org/10.1038/srep19163>, 2016.
- Gädeke, A., Krysanova, V., Aryal, A., Chang, J., Grillakis, M., Hanasaki, N., Koutroulis, A., Pokhrel, Y., Satoh, Y., Schaphoff, S., Müller Schmied, H., Stacke, T., Tang, Q., Wada, Y., and Thonicke, K.: Performance Evaluation of Global Hydrological Models in Six Large Pan-Arctic Watersheds, *Climatic Change*, 163, 1329–1351, <https://doi.org/10.1007/s10584-020-02892-2>, 2020.
- Galelli, S., Turner, S. W. D., Pokhrel, Y., Yi Ng, J., Castelletti, A., Bierkens, M. F. P., Pianosi, F., and Biemans, H.: Advancing the Representation of Human Actions in Large-Scale Hydrological Models: Challenges and Future Research Directions, *Water Resour. Res.*, 61, e2024WR039486, <https://doi.org/10.1029/2024WR039486>, 2025.
- Garrick, D. E., Hall, J. W., Dobson, A., Damania, R., Grafton, R. Q., Hope, R., Hepburn, C., Bark, R., Boltz, F., De Stefano, L., O'Donnell, E., Matthews, N., and Money, A.: Valuing Water for Sustainable Development, *Science*, 358, 1003–1005, <https://doi.org/10.1126/science.aao4942>, 2017.
- Gelman, A. and Stern, H.: The Difference Between “Significant” and “Not Significant” Is Not Itself Statistically Significant, *Am. Stat.*, 60, 328–331, <https://doi.org/10.1198/000313006X152649>, 2006.

- Gleeson, T., Cuthbert, M., Ferguson, G., and Perrone, D.: Global Groundwater Sustainability, Resources, and Systems in the Anthropocene, *Annu. Rev. Earth Pl. Sc.*, 48, 431–463, <https://doi.org/10.1146/annurev-earth-071719-055251>, 2020.
- 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, <https://doi.org/10.1038/s44221-023-00160-y>, 2023.
- Gruber, S.: Derivation and analysis of a high-resolution estimate of global permafrost zonation, *The Cryosphere*, 6, 221–233, <https://doi.org/10.5194/tc-6-221-2012>, 2012.
- Hagemann, S. and Gates, L. D.: Improving a Subgrid Runoff Parameterization Scheme for Climate Models by the Use of High Resolution Data Derived from Satellite Observations, *Clim. Dynam.*, 21, 349–359, <https://doi.org/10.1007/s00382-003-0349-x>, 2003.
- Hamed, K. H. and Ramachandra Rao, A.: A Modified Mann-Kendall Trend Test for Autocorrelated Data, *J. Hydrol.*, 204, 182–196, [https://doi.org/10.1016/S0022-1694\(97\)00125-X](https://doi.org/10.1016/S0022-1694(97)00125-X), 1998.
- Hartmann, J. and Moosdorf, N.: The New Global Lithological Map Database GLiM: A Representation of Rock Properties at the Earth Surface, *Geochem. Geophys. Geosy.*, 13, 2012GC004370, <https://doi.org/10.1029/2012GC004370>, 2012.
- Hirata, R., Goodarzi, L., Rörig, F. S., Alves, L. M., and Bertolo, R.: Climate Change Impacts on Groundwater: A Growing Challenge for Water Resources Sustainability in Brazil, *Environ. Monit. Assess.*, 197, 784, <https://doi.org/10.1007/s10661-025-14235-8>, 2025.
- Huggins, X., Gleeson, T., Castilla-Rho, J., Holley, C., Re, V., and Famiglietti, J. S.: Groundwater Connections and Sustainability in Social-Ecological Systems, *Groundwater*, 61, 463–478, <https://doi.org/10.1111/gwat.13305>, 2023.
- IGRAC: The Global Groundwater Monitoring Network (GGMN), <https://doi.org/10.58154/6Z0Y-DA34>, 2024.
- Ilstedt, U., Bangués Tobella, A., Bazié, H. R., Bayala, J., Verbeeten, E., Nyberg, G., Sanou, G., Benegas, L., Murdiyarso, D., Laudon, H., Sheil, D., and Malmer, A.: Intermediate Tree Cover Can Maximize Groundwater Recharge in the Seasonally Dry Tropics, *Sci. Rep.*, 6, 21930, <https://doi.org/10.1038/srep21930>, 2016.
- Jackson, C. R., Bloomfield, J. P., and Mackay, J. D.: Evidence for Changes in Historic and Future Groundwater Levels in the UK, *Progress in Physical Geography: Earth and Environment*, 39, 49–67, <https://doi.org/10.1177/0309133314550668>, 2015.
- van Jaarsveld, B.: GLOBGM_CMIP6: A Global Hyper-Resolution Groundwater Dataset for Assessing Historical and Future Groundwater Dynamics Under Climate and Socioeconomic Change – [globgm_historical_reference_gswp3-w5e5], YODA [data set], <https://doi.org/10.24416/UU01-AKSHOX>, 2025a.
- van Jaarsveld, B.: GLOBGM_CMIP6: a global hyper-resolution groundwater dataset for assessing historical and future groundwater dynamics under climate and socioeconomic change – [ssp_rcp_monthly], YODA [data set], <https://doi.org/10.24416/UU01-1BXLDP>, 2025b.
- van Jaarsveld, B.: GLOBGM_CMIP6: a global hyper-resolution groundwater dataset for assessing historical and future groundwater dynamics under climate and socioeconomic change – [ssp_rcp_annual], YODA [data set], <https://doi.org/10.24416/UU01-V6B9YS>, 2025c.
- van Jaarsveld, B.: GLOBGM_CMIP6: A Global Hyper-Resolution Groundwater Dataset for Assessing Historical and Future Groundwater Dynamics Under Climate and Socioeconomic Change – [ssp_rcp_average], YODA [data set], <https://doi.org/10.24416/UU01-SLRFI7>, 2025d.
- van Jaarsveld, B.: GLOBGM_CMIP6: a global hyper-resolution groundwater dataset for assessing historical and future groundwater dynamics under climate and socioeconomic change – [quality_assurance], YODA [data set], <https://doi.org/10.24416/UU01-16EJ3Y>, 2025e.
- Jasechko, S. and Perrone, D.: Global Groundwater Wells at Risk of Running Dry, *Science*, 372, 418–421, <https://doi.org/10.1126/science.abc2755>, 2021.
- Jasechko, S. and Taylor, R. G.: Intensive Rainfall Recharges Tropical Groundwaters, *Environ. Res. Lett.*, 10, 124015, <https://doi.org/10.1088/1748-9326/10/12/124015>, 2015.
- Jasechko, S., Seybold, H., Perrone, D., Fan, Y., Shamsudduha, M., Taylor, R. G., Fallatah, O., and Kirchner, J. W.: Rapid Groundwater Decline and Some Cases of Recovery in Aquifers Globally, *Nature*, 625, 715–721, <https://doi.org/10.1038/s41586-023-06879-8>, 2024.
- Jung, H., Saynisch-Wagner, J., and Schulz, S.: Can explainable AI Offer a New Perspective for Groundwater Recharge Estimation? – Global-Scale Modeling Using Neural Network, *Water Resour. Res.*, 60, e2023WR036360, <https://doi.org/10.1029/2023WR036360>, 2024.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., ISBN 978-1-5108-6096-4, 2017.
- Kendall, M. G.: *Rank Correlation Methods*, Griffin, London, 4th edn., 2nd impr. edn., ISBN 978-0-85264-199-6, 1975.
- Khatiwala, S.: Efficient Spin-up of Earth System Models Using Sequence Acceleration, *Sci. Adv.*, 10, eadn2839, <https://doi.org/10.1126/sciadv.adn2839>, 2024.
- Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*, 23, 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.
- Kollet, S., Belleflamme, A., Condon, L., Fahad, M., Goergen, K., Maxwell, R., and Naz, B.: Global Groundwater Modeling: Proof-of-concept of 3D Variably Saturated Flow Simulation at Kilometer Resolution, *J. Hydrol. X*, 30, 100213, <https://doi.org/10.1016/j.hydroa.2025.100213>, 2026.
- Kuang, X., Liu, J., Scanlon, B. R., Jiao, J. J., Jasechko, S., Lancia, M., Biskaborn, B. K., Wada, Y., Li, H., Zeng, Z., Guo, Z., Yao, Y., Gleeson, T., Nicot, J.-P., Luo, X., Zou, Y., and Zheng, C.: The Changing Nature of Groundwater in the Global Water Cycle, *Science*, 383, eadf0630, <https://doi.org/10.1126/science.adf0630>, 2024.
- Kumar, R., Samaniego, L., Thober, S., Rakovec, O., Marx, A., Wanders, N., Pan, M., Hesse, F., and Attinger, S.: Multi-Model Assessment of Groundwater Recharge Across Europe Under Warming Climate, *Earths Future*, 13, e2024EF005020, <https://doi.org/10.1029/2024EF005020>, 2025.

- Kupzig, J., Kupzig, N., and Flörke, M.: Regionalization in global hydrological models and its impact on runoff simulations: a case study using WaterGAP3 (v 1.0.0), *Geosci. Model Dev.*, 17, 6819–6846, <https://doi.org/10.5194/gmd-17-6819-2024>, 2024.
- Kuzma, S., Bierkens, M. F., Lakshman, S., Luo, T., Saccoccia, L., Sutanudjaja, E. H., and van Beek, L. P. H.: Aqueduct 4.0: Updated Decision-Relevant Global Water Risk Indicators, World Resources Institute, <https://doi.org/10.46830/writn.23.00061>, 2023.
- Lange, S. and Büchner, M.: ISIMIP3b Bias-Adjusted Atmospheric Climate Input Data, ISIMIP Repository [data set], <https://doi.org/10.48364/ISIMIP.842396.1>, 2021.
- Lange, S., Menz, C., Gleixner, S., Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Müller Schmied, H., Hersbach, H., Buontempo, C., and Cagnazzo, C.: WFDE5 over Land Merged with ERA5 over the Ocean (W5E5 v2.0), ISIMIP Repository [data set], <https://doi.org/10.48364/ISIMIP.342217>, 2021.
- Lannelongue, L., Grealey, J., and Inouye, M.: Green Algorithms: Quantifying the Carbon Footprint of Computation, *Adv. Sci.*, 8, 2100707, <https://doi.org/10.1002/adv.202100707>, 2021.
- Lehner, B., Verdin, K., and Jarvis, A.: New Global Hydrography Derived From Spaceborne Elevation Data, *Eos, Transactions American Geophysical Union*, 89, 93–94, <https://doi.org/10.1029/2008EO100001>, 2008.
- Leijnse, M., Bierkens, M. F. P., Gommans, K. H. M., Lin, D., Tait, A., and Wanders, N.: Key Drivers and Pressures of Global Water Scarcity Hotspots, *Environ. Res. Lett.*, 19, 054035, <https://doi.org/10.1088/1748-9326/ad3c54>, 2024.
- Li, F. and Kusche, J.: Reproducing GRACE Total Water Storage Change at Finer Spatial Scales, *Geophys. Res. Lett.*, 53, e2025GL119881, <https://doi.org/10.1029/2025GL119881>, 2026.
- Liu, R., Zhang, X., Wang, W., Wang, Y., Liu, H., Ma, M., and Tang, G.: Global-Scale ERA5 Product Precipitation and Temperature Evaluation, *Ecol. Indic.*, 166, 112481, <https://doi.org/10.1016/j.ecolind.2024.112481>, 2024.
- Ma, Y., Condon, L. E., Koch, J., Bennett, A., Defnet, A., Tijerina-Kreuzer, D., Melchior, P., and Maxwell, R. M.: High Resolution US Water Table Depth Estimates Reveal Quantity of Accessible Groundwater, *Commun. Earth Environ.*, 7, 45, <https://doi.org/10.1038/s43247-025-03094-3>, 2026a.
- Ma, Y., Tijerina-Kreuzer, D., Defnet, A., Artavanis, G., Condon, L. E., and Maxwell, R. M.: A Prototype Hyper-Resolution Groundwater Digital Twin for the Contiguous United States: Integrating Physics-Based Modeling, Machine Learning, and Observations, *J. Hydrol.*, 671, 135189, <https://doi.org/10.1016/j.jhydrol.2026.135189>, 2026b.
- Maples, S. R., Fogg, G. E., and Maxwell, R. M.: Modeling Managed Aquifer Recharge Processes in a Highly Heterogeneous, Semi-Confined Aquifer System, *Hydrogeol. J.*, 27, 2869–2888, <https://doi.org/10.1007/s10040-019-02033-9>, 2019.
- Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., Hohenegger, C., Ilyina, T., Jahns, T., Jimenez-de-la Cuesta, D., Jungclaus, J., Kleinen, T., Kloster, S., Kracher, D., Kinne, S., Kleberg, D., Lasslop, G., Kornbluh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Möbis, B., Müller, W. A., Nabel, J. E. M. S., Nam, C. C. W., Notz, D., Nyawira, S.-S., Paulsen, H., Peters, K., Pincus, R., Pohlmann, H., Pongratz, J., Popp, M., Raddatz, T. J., Rast, S., Redler, R., Reick, C. H., Rohrschneider, T., Schemann, V., Schmidt, H., Schnur, R., Schulzweida, U., Six, K. D., Stein, L., Stemmler, I., Stevens, B., Von Storch, J.-S., Tian, F., Voigt, A., Vrese, P., Wieners, K.-H., Wilkenskjeld, S., Winkler, A., and Roeckner, E.: Developments in the MPI-M Earth System Model Version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO₂, *J. Adv. Model. Earth Sy.*, 11, 998–1038, <https://doi.org/10.1029/2018MS001400>, 2019.
- McKenzie, J. M., Kurylyk, B. L., Walvoord, M. A., Bense, V. F., Fortier, D., Spence, C., and Grenier, C.: Invited perspective: What lies beneath a changing Arctic?, *The Cryosphere*, 15, 479–484, <https://doi.org/10.5194/tc-15-479-2021>, 2021.
- Meixner, T., Manning, A. H., Stonestrom, D. A., Allen, D. M., Ajami, H., Blasch, K. W., Brookfield, A. E., Castro, C. L., Clark, J. F., Gochis, D. J., Flint, A. L., Neff, K. L., Niraula, R., Rodell, M., Scanlon, B. R., Singha, K., and Walvoord, M. A.: Implications of Projected Climate Change for Groundwater Recharge in the Western United States, *J. Hydrol.*, 534, 124–138, <https://doi.org/10.1016/j.jhydrol.2015.12.027>, 2016.
- Moeck, C., Grech-Cumbo, N., Podgorski, J., Bretzler, A., Gurdak, J. J., Berg, M., and Schirmer, M.: Data for: A Global-Scale Dataset of Direct Natural Groundwater Recharge Rates: A Review of Variables, Processes and Relationships, Eawag: Swiss Federal Institute of Aquatic Science and Technology [data set], <https://doi.org/10.25678/0001NG>, 2020a.
- Moeck, C., Grech-Cumbo, N., Podgorski, J., Bretzler, A., Gurdak, J. J., Berg, M., and Schirmer, M.: A Global-Scale Dataset of Direct Natural Groundwater Recharge Rates: A Review of Variables, Processes and Relationships, *Sci. Total Environ.*, 717, 137042, <https://doi.org/10.1016/j.scitotenv.2020.137042>, 2020b.
- Mohan, C., Western, A. W., Wei, Y., and Saft, M.: Predicting groundwater recharge for varying land cover and climate conditions – a global meta-study, *Hydrol. Earth Syst. Sci.*, 22, 2689–2703, <https://doi.org/10.5194/hess-22-2689-2018>, 2018.
- Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., and Köster, J.: Sustainable Data Analysis with Snakemake, *F1000Research*, 10, 33, <https://doi.org/10.12688/f1000research.29032.2>, 2021.
- Niazi, H., Wild, T. B., Turner, S. W. D., Graham, N. T., Hejazi, M., Msangi, S., Kim, S., Lamontagne, J. R., and Zhao, M.: Global Peak Water Limit of Future Groundwater Withdrawals, *Nat. Sustain.*, 7, 413–422, <https://doi.org/10.1038/s41893-024-01306-w>, 2024.
- Otoo, N. G., Sutanudjaja, E. H., van Vliet, M. T. H., Schipper, A. M., and Bierkens, M. F. P.: Mapping groundwater-dependent ecosystems using a high-resolution global groundwater model, *Hydrol. Earth Syst. Sci.*, 29, 2153–2165, <https://doi.org/10.5194/hess-29-2153-2025>, 2025.
- Parry, S., Mackay, J. D., Chitson, T., Hannaford, J., Magee, E., Tanguy, M., Bell, V. A., Facer-Childs, K., Kay, A., Lane, R., Moore, R. J., Turner, S., and Wallbank, J.: Divergent future drought projections in UK river flows and groundwater levels, *Hydrol. Earth Syst. Sci.*, 28, 417–440, <https://doi.org/10.5194/hess-28-417-2024>, 2024.
- Pawusch, L., Scheurer, S., Nowak, W., and Maxwell, R. M.: HydroStartML: A Combined Machine Learning

- and Physics-Based Approach to Reduce Hydrological Model Spin-up Time, *Adv. Water Resour.*, 206, 105124, <https://doi.org/10.1016/j.advwatres.2025.105124>, 2025.
- Perrone, D. and Jasechko, S.: Deeper Well Drilling an Unsustainable Stopgap to Groundwater Depletion, *Nat. Sustain.*, 2, 773–782, <https://doi.org/10.1038/s41893-019-0325-z>, 2019.
- Pfeffer, J., Cazenave, A., Blazquez, A., Decharme, B., Munier, S., and Barnoud, A.: Assessment of pluri-annual and decadal changes in terrestrial water storage predicted by global hydrological models in comparison with the GRACE satellite gravity mission, *Hydrol. Earth Syst. Sci.*, 27, 3743–3768, <https://doi.org/10.5194/hess-27-3743-2023>, 2023.
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, *SOIL*, 7, 217–240, <https://doi.org/10.5194/soil-7-217-2021>, 2021.
- Pool, S., Vis, M., and Seibert, J.: Evaluating Model Performance: Towards a Non-Parametric Variant of the Kling-Gupta Efficiency, *Hydrolog. Sci. J.*, 63, 1941–1953, <https://doi.org/10.1080/02626667.2018.1552002>, 2018.
- Reinecke, R., Foglia, L., Mehl, S., Trautmann, T., Cáceres, D., and Döll, P.: Challenges in developing a global gradient-based groundwater model (G³M v1.0) for the integration into a global hydrological model, *Geosci. Model Dev.*, 12, 2401–2418, <https://doi.org/10.5194/gmd-12-2401-2019>, 2019.
- Reinecke, R., Gnann, S., Stein, L., Bierkens, M., de Graaf, I., Gleeson, T., Essink, G. O., Sutanudjaja, E. H., Ruz Vargas, C., Verkaik, J., and Wagener, T.: Uncertainty in Model Estimates of Global Groundwater Depth, *Environ. Res. Lett.*, 19, 114066, <https://doi.org/10.1088/1748-9326/ad8587>, 2024.
- Reinecke, R., Akhter, T., Bähge, A., Dietrich, R., Gnann, S., Gosling, S. N., Grogan, D., Hartmann, A., Kollet, S., Kumar, R., Lammers, R., Liu, S., Liu, Y., Moosdorf, N., Naz, B., Nazari, S., Orazulike, C., Pokhrel, Y., Schewe, J., Smilovic, M., Stokal, M., Thiery, W., Wada, Y., Zuidema, S., and de Graaf, I.: The ISIMIP groundwater sector: a framework for ensemble modeling of global change impacts on groundwater, *Geosci. Model Dev.*, 19, 523–542, <https://doi.org/10.5194/gmd-19-523-2026>, 2026.
- Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J. T., Beaudoing, H. K., Landerer, F. W., and Lo, M.-H.: Emerging Trends in Global Freshwater Availability, *Nature*, 557, 651–659, <https://doi.org/10.1038/s41586-018-0123-1>, 2018.
- Saccò, M., Mammola, S., Altermatt, F., Alther, R., Bolpagni, R., Brancelj, A., Brankovits, D., Fišer, C., Gerovasileiou, V., Griebler, C., Guareschi, S., Hose, G. C., Korbel, K., Lictevout, E., Malard, F., Martínez, A., Niemiller, M. L., Robertson, A., Tanalgo, K. C., Bichuette, M. E., Borko, Š., Brad, T., Campbell, M. A., Cardoso, P., Celico, F., Cooper, S. J. B., Culver, D., Di Lorenzo, T., Galassi, D. M. P., Guzik, M. T., Hartland, A., Humphreys, W. F., Ferreira, R. L., Lunghi, E., Nizzoli, D., Perina, G., Raghavan, R., Richards, Z., Reboleira, A. S. P. S., Rohde, M. M., Fernández, D. S., Schmidt, S. I., Van Der Heyde, M., Weaver, L., White, N. E., Zagmajster, M., Hogg, I., Ruhi, A., Gagnon, M. M., Allentoft, M. E., and Reinecke, R.: Groundwater Is a Hidden Global Keystone Ecosystem, *Glob. Change Biol.*, 30, e17066, <https://doi.org/10.1111/gcb.17066>, 2024.
- Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., van Beek, L. P. H., Wiese, D. N., Wada, Y., Long, D., Reedy, R. C., Longuevergne, L., Döll, P., and Bierkens, M. F. P.: Global Models Underestimate Large Decadal Declining and Rising Water Storage Trends Relative to GRACE Satellite Data, *P. Natl. Acad. Sci. USA*, 115, E1080–E1089, <https://doi.org/10.1073/pnas.1704665115>, 2018.
- Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., Andrews, M. B., Andrews, T., Archibald, A. T., De Mora, L., Dyson, H., Elkington, M., Ellis, R., Florek, P., Good, P., Gohar, L., Haddad, S., Hardiman, S. C., Hogan, E., Iwi, A., Jones, C. D., Johnson, B., Kelley, D. I., Kettleborough, J., Knight, J. R., Köhler, M. O., Kuhlbrodt, T., Liddicoat, S., Linova-Pavlova, I., Mizielski, M. S., Morgenstern, O., Mulcahy, J., Neininger, E., O'Connor, F. M., Petrie, R., Ridley, J., Rioual, J.-C., Roberts, M., Robertson, E., Rumbold, S., Seddon, J., Shepherd, H., Shim, S., Stephens, A., Teixeira, J. C., Tang, Y., Williams, J., Wiltshire, A., and Griffiths, P. T.: Implementation of U.K. Earth System Models for CMIP6, *J. Adv. Model. Earth Sy.*, 12, e2019MS001946, <https://doi.org/10.1029/2019MS001946>, 2020.
- Sen, P. K.: Estimates of the Regression Coefficient Based on Kendall's Tau, *J. Am. Stat. Assoc.*, 63, 1379–1389, <https://doi.org/10.1080/01621459.1968.10480934>, 1968.
- Servén, D., Brummitt, C., and Abedi, H.: Dswah/pyGAM: V0.8.0, Zenodo [code], <https://doi.org/10.5281/zenodo.1476122>, 2018.
- Somers, L. D. and McKenzie, J. M.: A Review of Groundwater in High Mountain Environments, *WIREs Water*, 7, e1475, <https://doi.org/10.1002/wat2.1475>, 2020.
- Straatsma, M., Droogers, P., Hunink, J., Berendrecht, W., Buitink, J., Buytaert, W., Karssenbergh, D., Schmitz, O., Sutanudjaja, E. H., van Beek, L. P. H., Vitolo, C., and Bierkens, M. F.: Global to Regional Scale Evaluation of Adaptation Measures to Reduce the Future Water Gap, *Environ. Model. Softw.*, 124, 104578, <https://doi.org/10.1016/j.envsoft.2019.104578>, 2020.
- Sutanudjaja, E. H.: HYPFLOWSC16: HYdrological Projection of Future gLObal Water States with CMIP6, YODA [data set], <https://doi.org/10.24416/UU01-YM7A5H>, 2024.
- 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., Karssenbergh, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannamettee, E., Wisse, D., and Bierkens, M. F. P.: PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, *Geosci. Model Dev.*, 11, 2429–2453, <https://doi.org/10.5194/gmd-11-2429-2018>, 2018.
- Taylor, R. G., Scanlon, B., Döll, P., Rodell, M., van Beek, L. P. H., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J. S., Edmunds, M., Konikow, L., Green, T. R., Chen, J., Taniguchi, M., Bierkens, M. F. P., MacDonald, A., Fan, Y., Maxwell, R. M., Yechieli, Y., Gurdak, J. J., Allen, D. M., Shamsudduha, M., Hiscock, K., Yeh, P. J.-F., Holman, I., and Treidel, H.: Ground Water and Climate Change, *Nat. Clim. Change*, 3, 322–329, <https://doi.org/10.1038/nclimate1744>, 2013.
- Theil, H.: A Rank-Invariant Method of Linear and Polynomial Regression Analysis, *Indagat. Math.*, 12, 386–392, 1950.
- Tiwari, A. D., Pokhrel, Y., Boulange, J., Burek, P., Guillaumot, L., Gosling, S. N., Grillakis, M., Hanasaki, N., Koutroulis, A., Ostberg, S., Otta, K., Schmied, H. M., Satoh, Y., Scanlon, B., Stacke, T., and Yokohata, T.: Similarities and Divergent Patterns in Hydrologic Fluxes and Storages Simulated by Global Water Mod-

- els, *Nature Water*, 3, 550–560, <https://doi.org/10.1038/s44221-025-00435-6>, 2025.
- Todini, E.: The ARNO Rainfall – Runoff Model, *J. Hydrol.*, 175, 339–382, [https://doi.org/10.1016/S0022-1694\(96\)80016-3](https://doi.org/10.1016/S0022-1694(96)80016-3), 1996.
- Triplett, A., Bennett, A., Condon, L. E., Melchior, P., and Maxwell, R. M.: A Deep-Learning Based Parameter Inversion Framework for Large-Scale Groundwater Models, *Geophys. Res. Lett.*, 52, e2024GL114285, <https://doi.org/10.1029/2024GL114285>, 2025.
- UN Water: Groundwater Making the Invisible Visible, no. 2022 in The United Nations World Water Development Report, UNESCO, Paris, ISBN 978-92-3-100507-7, 2022.
- van Jaarsveld, B., Wanders, N., Otoo, N. G., Sutanudjaja, E. H., Verkaik, J., Zamrsky, D., and Bierkens, M. F. P.: GLOBGM_CMIP6: Software and Code, Zenodo [code], <https://doi.org/10.5281/zenodo.17065147>, 2025.
- Vega-Briones, J., Sutanudjaja, E. H., De Jong, S., and Wanders, N.: Modelling Groundwater Hydrological Drought and Its Recovery Given Natural and Anthropogenic Scenarios in South America, *Hydrol. Process.*, 38, e15340, <https://doi.org/10.1002/hyp.15340>, 2024.
- Verkaik, J., Sutanudjaja, E. H., Oude Essink, G. H. P., Lin, H. X., and Bierkens, M. F. P.: GLOBGM v1.0: a parallel implementation of a 30 arcsec PCR-GLOBWB-MODFLOW global-scale groundwater model, *Geosci. Model Dev.*, 17, 275–300, <https://doi.org/10.5194/gmd-17-275-2024>, 2024.
- Vörösmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B.: Global Water Resources: Vulnerability from Climate Change and Population Growth, *Science*, 289, 284–288, <https://doi.org/10.1126/science.289.5477.284>, 2000.
- Wada, Y. and Bierkens, M. F. P.: Sustainability of Global Water Use: Past Reconstruction and Future Projections, *Environ. Res. Lett.*, 9, 104003, <https://doi.org/10.1088/1748-9326/9/10/104003>, 2014.
- Wada, Y., van Beek, L. P. H., Wanders, N., and Bierkens, M. F. P.: Human Water Consumption Intensifies Hydrological Drought Worldwide, *Environ. Res. Lett.*, 8, 034036, <https://doi.org/10.1088/1748-9326/8/3/034036>, 2013.
- Werndl, C.: Initial-Condition Dependence and Initial-Condition Uncertainty in Climate Science, *Brit. J. Philos. Sci.*, 70, 953–976, <https://doi.org/10.1093/bjps/axy021>, 2019.
- West, C., Rosolem, R., MacDonald, A. M., Cuthbert, M. O., and Wagener, T.: Understanding Process Controls on Groundwater Recharge Variability across Africa through Recharge Landscapes, *J. Hydrol.*, 612, 127967, <https://doi.org/10.1016/j.jhydrol.2022.127967>, 2022.
- White, J. T.: A Model-Independent Iterative Ensemble Smoother for Efficient History-Matching and Uncertainty Quantification in Very High Dimensions, *Environ. Model. Softw.*, 109, 191–201, <https://doi.org/10.1016/j.envsoft.2018.06.009>, 2018.
- Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., De Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., Van De Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution Global Land Surface Modeling: Meeting a Grand Challenge for Monitoring Earth's Terrestrial Water, *Water Resour. Res.*, 47, 2010WR010090, <https://doi.org/10.1029/2010WR010090>, 2011.
- Wu, W.-Y., Lo, M.-H., Wada, Y., Famiglietti, J. S., Reager, J. T., Yeh, P. J.-F., Ducharme, A., and Yang, Z.-L.: Divergent Effects of Climate Change on Future Groundwater Availability in Key Mid-Latitude Aquifers, *Nat. Commun.*, 11, 3710, <https://doi.org/10.1038/s41467-020-17581-y>, 2020.
- Wunsch, A., Liesch, T., and Broda, S.: Deep Learning Shows Declining Groundwater Levels in Germany until 2100 Due to Climate Change, *Nat. Commun.*, 13, 1221, <https://doi.org/10.1038/s41467-022-28770-2>, 2022.
- Xie, J., Liu, X., Jasechko, S., Berghuijs, W. R., Wang, K., Liu, C., Reichstein, M., Jung, M., and Koirala, S.: Majority of Global River Flow Sustained by Groundwater, *Nat. Geosci.*, 17, 770–777, <https://doi.org/10.1038/s41561-024-01483-5>, 2024.
- Yang, C., Jia, Z., Xu, W., Wei, Z., Zhang, X., Zou, Y., McDonnell, J., Condon, L., Dai, Y., and Maxwell, R.: CONCN: a high-resolution, integrated surface water–groundwater ParFlow modeling platform of continental China, *Hydrol. Earth Syst. Sci.*, 29, 2201–2218, <https://doi.org/10.5194/hess-29-2201-2025>, 2025.
- Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deuschi, M., Tanaka, T., Hosaka, M., Yabu, S., Yoshimura, H., Shindo, E., Mizuta, R., Obata, A., Adachi, Y., and Ishii, M.: The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and Basic Evaluation of the Physical Component, *J. Meteorol. Soc. Jpn. Ser. II*, 97, 931–965, <https://doi.org/10.2151/jmsj.2019-051>, 2019.
- Zamrsky, D., Oude Essink, G. H. P., and Bierkens, M. F. P.: Global Impact of Sea Level Rise on Coastal Fresh Groundwater Resources, *Earths Future*, 12, e2023EF003581, <https://doi.org/10.1029/2023EF003581>, 2024.
- Zamrsky, D., Ruzzante, S., Compare, K., Kretschmer, D., Zipper, S., Befus, K. M., Reinecke, R., Pasner, Y., Gleeson, T., Jordan, K., Cuthbert, M., Castronova, A. M., Wagener, T., and Bierkens, M. F. P.: Current Trends and Biases in Groundwater Modelling Using the Community-Driven Groundwater Model Portal (GroMoPo), *Hydrogeol. J.*, 33, 355–366, <https://doi.org/10.1007/s10040-025-02882-7>, 2025.
- Zell, W. O. and Sanford, W. E.: Calibrated Simulation of the Long-Term Average Surficial Groundwater System and Derived Spatial Distributions of Its Characteristics for the Contiguous United States, *Water Resour. Res.*, 56, e2019WR026724, <https://doi.org/10.1029/2019WR026724>, 2020.