

Supplementary Information for:

# Twenty-five years of groundwater depletion and recharge across North America

Mohammad Haghiri<sup>1,\*</sup>

mhaghi2@uic.edu

Kerry Callaghan<sup>1</sup>

kerryrc@uic.edu

Andrew Wickert<sup>2</sup>

awickert@umn.edu

Roger Creel<sup>3</sup>

Roger.creel@tamu.edu

<sup>1</sup>Department of Earth and Environmental Sciences, University of Illinois Chicago, 845 West Taylor St, Chicago, 60607, Illinois, USA.

<sup>2</sup>Department of Earth & Environmental Sciences, University of Minnesota, 16 Church St., Minneapolis, 55455, Minnesota, USA.

<sup>3</sup>Department of Geology and Geophysics, Texas A&M University, 3115 TAMU, College Station, 77843, Texas, USA.

**\*Corresponding author**

E-mail: mhaghi2@uic.edu

## Contents of This File:

- Figures S1–S18
- Tables S1–S4



## Supplementary Overview

720 This Supplementary Information provides Figures S1–S18 and Tables S1–S4 supporting the main text: the model spin-up curve and equilibration timescale under constant 0 CE forcing (Fig. S1); standardized precipitation (SPI) and precipitation–evapotranspiration (SPEI) indices for North America (Fig. S2) and for individual watersheds (Fig. S3); the coupling between SPEI and simulated water-table depth, including lag-correlation and regression analyses (Fig. S4), with the associated variable and drought-index definitions (Table S1); the comparison of simulated water-table anomalies with snow-removed GRACE-  
725 derived groundwater storage, spatially (Fig. S5) and as watershed time series (Fig. S6); long-term trends in mean groundwater (Fig. S7) and surface-water (Fig. S8) depth, with their annual distributions (Figs. S9 and S10) and summary trend statistics (Tables S2 and S3); monthly wet–dry dominance (Fig. S11) and the interquartile range of the standardized water balance (Fig. S12) and of water-table depth (Fig. S13); validation of simulated versus observed water levels, including monthly  $R^2$  and NSE values (Table S4), the spatial residual map (Fig. S14), and lake-extent classification metrics (Fig. S15); and monthly water-  
730 table-depth differences (Fig. S16), monthly P–ET differences (Fig. S17), and watershed-scale climate–groundwater coupling (Fig. S18) for 2024.

### 1. Model Initialisation and Spin-up

### 2. Water-Table Depth and Hydroclimatic Forcing

### 3. Changing Water-Table Depth and Human and Natural Systems

### 735 4. Water Table Depletion Rates

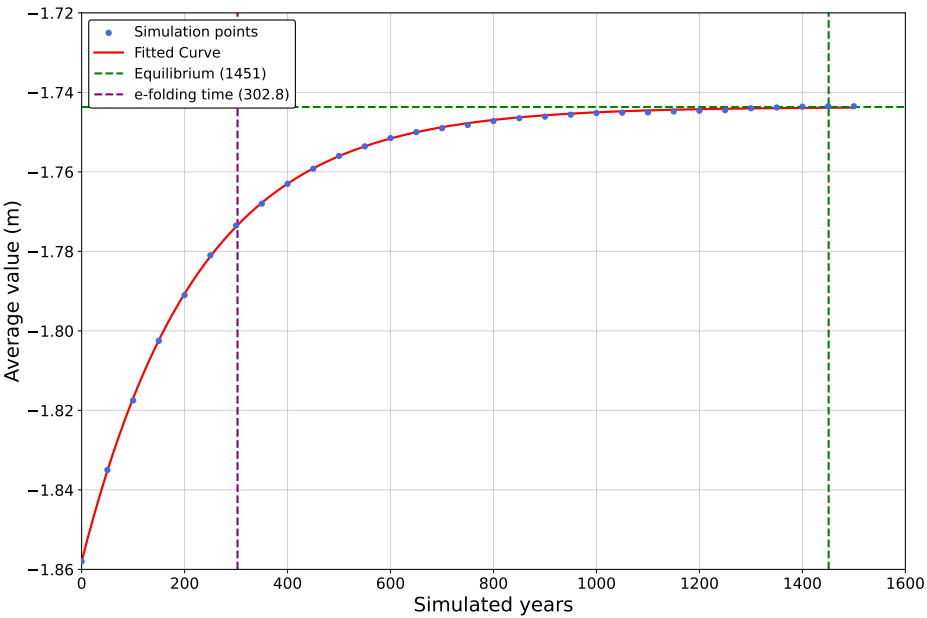
### 5. Validation of Water-Table Depth

### 6. Seasonal and Long-Term Sea-Level Equivalent (SLE) Dynamics

## Model Initialisation and Spin-up

To determine the length of time required for model spin-up, we initialised the WTM using the 0 CE input water table from  
740 previous work (Callaghan et al., 2025), and then ran the simulation to equilibrium under a constant 0–10 CE decadal climatic forcing solely to determine the required spin-up duration. Fig. S1 shows the changing average depth to water table over simulated time. The gradual leveling of the curve indicates the long memory and delayed response of continental-scale groundwater systems, emphasizing the importance of long-term transient simulations to understand changing water tables. Achieving a stable average water-table value confirms that subsurface hydraulic gradients and storage conditions have balanced under the  
745 imposed boundary conditions. To quantify the adjustment toward equilibrium, we fitted an exponential curve to the simulated trajectory and estimated the e-folding time, which represents the characteristic timescale of system response. The e-folding time is defined as the time required for the system to reduce its initial deviation from equilibrium to about 37%. The fitted

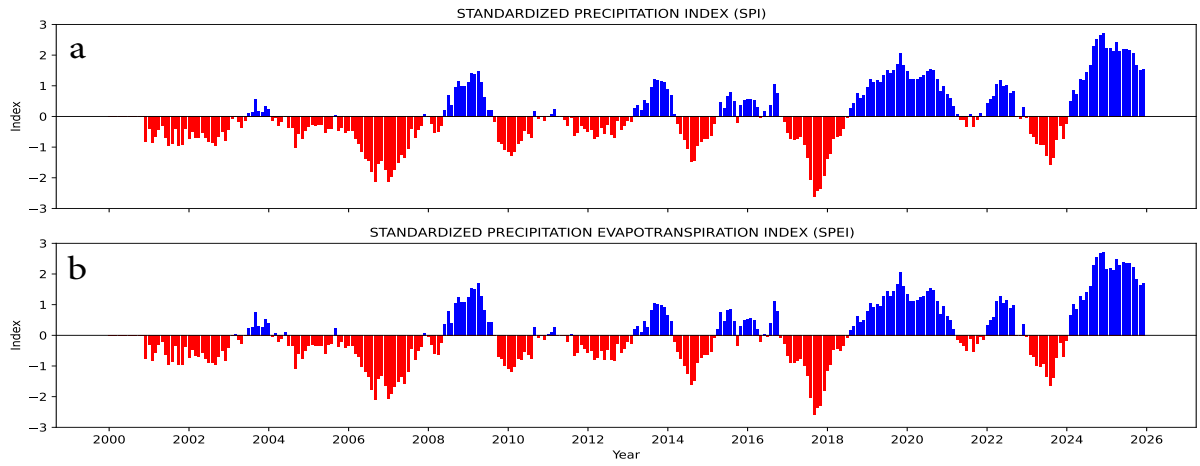
e-folding time is 302.8 years (Fig. S1), indicating slow continental-scale groundwater adjustment. The system reaches practical equilibrium after approximately 1451 years, corresponding to nearly five e-folding times. Accordingly, the 1451-year equilibrium time highlights the necessity of extended spin-ups for this large-scale hydrologic model to avoid inherited biases from initial states, ensuring that subsequent transient simulations of climate-driven groundwater variability (e.g., 2000–2025) represent real hydrologic dynamics rather than artifacts of initialization.



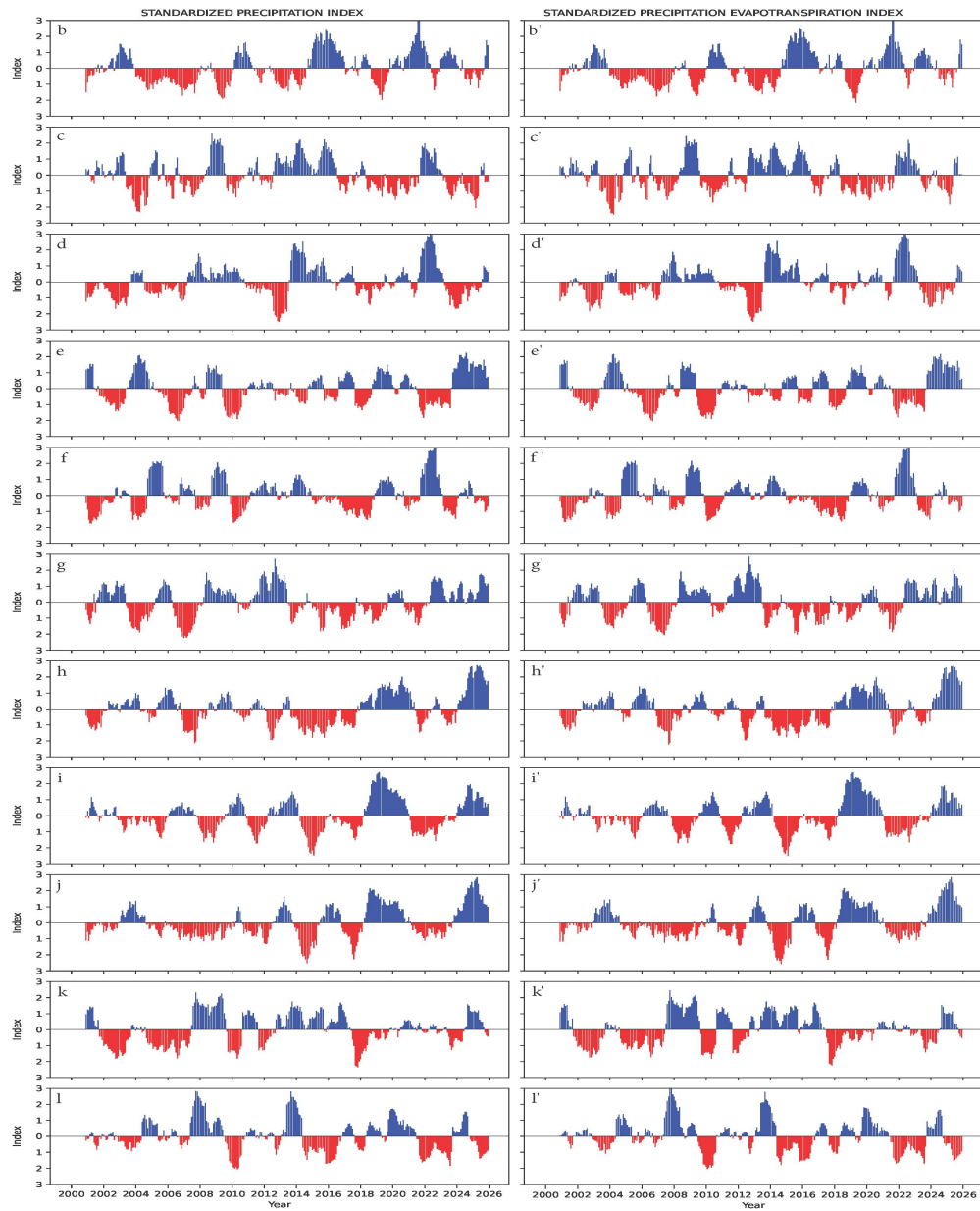
**Figure S1. Spin-up curve of the Water Table Model (WTM) under constant 0 CE climatic forcing.** Blue dots represent the simulated mean continental water-table depth through time, and the red curve shows the fitted exponential trend approaching equilibrium. The model achieves an e-folding time of 302.8 years and reaches equilibrium after approximately 1451 years, indicated by the dashed vertical lines.

### Water-Table Depth and Hydroclimatic Forcing

The SPI and SPEI time series (Fig. S2-S3) show patterns of climatic variability across the watersheds examined between 2000 and 2025. Monthly precipitation (P) and precipitation minus evapotranspiration (P–ET) were aggregated over each watershed using area-weighted spatial averages derived from the same input data used for WTM. Accumulated series were constructed at the selected temporal scale using rolling monthly sums. SPI was calculated by standardizing the accumulated precipitation time series at each cell (used to create graph for each watershed), and SPEI was calculated by standardizing the accumulated P–ET time series. Standardization was performed using the mean and standard deviation of the respective time series over the analysis period. Positive values represent wetter-than-average conditions and negative values represent drier-than-average conditions. These figures provide the temporal evolution of watershed-scale climatic variability and complement the groundwater responses shown in the main text.



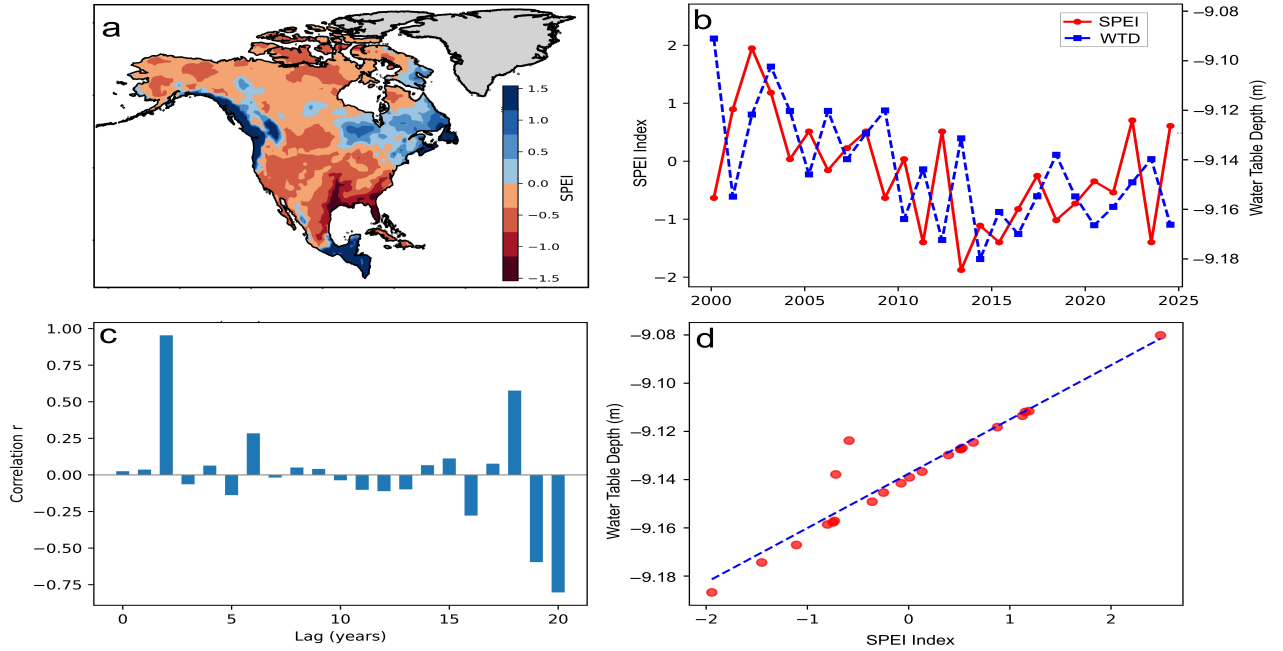
**Figure S2. Standardized monthly precipitation (SPI) and precipitation–evapotranspiration (SPEI) indices for North America, 2000–2025.** (a) SPI computed from domain mean precipitation using a 12 month moving sum and z-score standardization over the full record. (b) SPEI computed from the domain mean climatic balance ( $P - E$ ) using the same 12 month window and standardization. Blue bars indicate wetter than normal conditions (index  $> 0$ ) and red bars indicate drier than normal conditions (index  $< 0$ ).



**Figure S3. Standardized monthly precipitation (SPI) and precipitation–evapotranspiration (SPEI) indices for different watersheds, 2000–2025.** In this figure, two columns are shown and indicated as (b–l) and (b'–l'), which represent each watershed (panel labels (b–l) correspond to major North American watersheds as labeled in Fig. 2). The panels (b–l) show SPI, and (b'–l') show SPEI. Blue bars denote wet anomalies (index > 0), and red bars denote dry anomalies (index < 0). SPI reflects precipitation-only variability, whereas SPEI integrates both precipitation and evaporative demand, resulting in greater drought sensitivity under periods of high atmospheric dryness. Each horizontal panel corresponds to an individual watershed, allowing comparison of climatic wetness and dryness patterns across regions. Some values are as high as 3.16; we here limit the y axis to a range from -3 to +3 for viewability.

## Relationship between SPEI and simulated groundwater variability

Fig. S4 illustrates the link between climatic water-balance anomalies and simulated water table dynamics. Spatial patterns of the SPEI (Fig. S4a) identify regions experiencing anomalously wet and dry conditions based on precipitation minus evapotranspiration (P–ET). The time series comparison (Fig. S4b) shows that interannual fluctuations in water-table depth closely track variations in the SPEI, with wetter conditions generally associated with shallower water tables and drier conditions corresponding to deeper levels. Lagged correlation analysis (Fig. S4c) reveals a peak positive correlation at a multi-year lag, indicating a delayed groundwater response to climatic forcing consistent with groundwater storage and slow recharge processes. To create this analysis, we systematically shifted the SPEI time series by 0–20 years relative to WTD and recalculated the Pearson correlation at each lag to quantify the temporal response of groundwater to climatic variability. The maximum correlation occurs at an optimal lag of 2 years, suggesting that groundwater systems integrate climatic anomalies over multiple years before fully responding. Such delayed responses reflect the concept of hydrologic memory, whereby subsurface storage and slow recharge processes buffer short-term climate variability and transmit signals to groundwater systems over multi-year timescales. Similar multi-year lagged responses between precipitation anomalies and groundwater levels have been documented in previous studies of regional aquifer systems Cuthbert et al. (2019); Fan et al. (2013b). Fig. S4d shows the linear regression between watershed-averaged WTD and SPEI at the lag corresponding to the maximum correlation identified in Fig. S4c. SPEI values were first temporally shifted by the optimal lag (2 years), and only overlapping years between the shifted SPEI series and WTD were retained. Annual mean values were then paired and a least-squares linear regression was fitted to quantify the relationship between SPEI (independent variable) and WTD (dependent variable). This behavior is consistent with previous studies showing that groundwater responses to climatic variability can occur over multi-year timescales due to the buffering effect of aquifer storage and delayed recharge processes.



**Figure S4. Climate-groundwater coupling across North America** (a) Per pixel SPEI for as a 25-year average, computed from annual precipitation minus evapotranspiration ( $P-ET$ ) and standardized over all land pixels so that positive values indicate anomalously wet conditions and negative values indicate moisture deficits. (b) Time series of the SPEI (red line, left axis) and mean simulated water-table depth (WTD, blue dashed line, right axis) for the selected area over 2000–2025. (c) Pearson correlation coefficient ( $r$ ) between the SPEI and WTD as a function of lag (0–20 years), where a positive lag denotes a delayed groundwater response to the climate index. (d) Scatterplot of WTD versus SPEI at the best lag identified in panel c, with a least squares regression line.

Table S1 provides detailed definitions of all variables, indices, and parameters used in the construction of the drought metrics, spatial anomalies, water-table depth, lag-correlation analysis, and groundwater-sensitivity estimates shown in Fig. S4.

785 Annual precipitation is  $P$ , annual evapotranspiration  $ET$ , and their difference  $(P - ET)$  represents the climatic water balance, where positive values indicate anomalously wet conditions and negative values indicate dry conditions. The function  $F(\cdot)$  is the cumulative distribution function (CDF) of the fitted probability model (Gamma for SPI and log-logistic for SPEI), and  $\Phi^{-1}(\cdot)$  is the inverse standard normal transform used to convert probabilities into standardized z-scores. The standardized precipitation index (SPI) is defined as  $SPI = \Phi^{-1}(F(P))$ , and the standardized precipitation–evapotranspiration index (SPEI) as  $SPEI = \Phi^{-1}(F(P - ET))$ . The spatial SPEI anomaly map (Fig. S4a) is computed as  $Z(x, y) = ((P - ET)(x, y) - \mu) / \sigma$ , where  $\mu$  and  $\sigma$  are the spatial mean and standard deviation of  $(P - ET)$  across all land pixels for the selected year.

For negative SPEI anomalies (Fig. S4b), the drought index for year  $t$  is computed as  $I(t) = \frac{1}{N} \sum_{(x,y) \in \Omega} Z(x, y, t)$ , where  $\Omega$  is the polygon domain (North America) and  $N$  is the number of valid pixels. The simulated water-table depth is denoted by  $WTD(x, y, t)$  at each grid cell and aggregated to the polygon mean as  $WTD(t) = \frac{1}{N} \sum_{(x,y) \in \Omega} WTD(x, y, t)$ . Lagged  
795 climate-groundwater coherence (Fig. S4c) is quantified using the Pearson correlation between the drought index and water-

table depth at lag  $k$ , defined as  $r(k) = \text{corr}(I(t), \text{WTD}(t+k))$ . The optimal groundwater response time is defined as  $k^* = \arg \max_k |r(k)|$ . Fig.S4d shows the linear regression between the lag-adjusted drought index and water-table depth, expressed as  $\text{WTD}(t) \approx a + bI(t - k^*)$ , where  $a$  is the intercept and  $b$  is the regression slope representing the change in water-table depth per unit change in climate drought index.

**Table S1.** Summary of drought indices and variables used in Fig. S4, including SPEI spatial anomaly, water-table depth, lag correlation, and groundwater sensitivity.

| Index / Variable                                           | Input Data                          | Probability Distribution          | Definition / Equation                                                      |
|------------------------------------------------------------|-------------------------------------|-----------------------------------|----------------------------------------------------------------------------|
| SPI (Standardized Precipitation Index)                     | Precipitation $P$                   | Gamma (or Normal for annual data) | $\text{SPI} = \Phi^{-1}(F(P))$                                             |
| SPEI (Standardized Precipitation–Evapotranspiration Index) | Climatic water balance $(P - ET)$   | Log-logistic                      | $\text{SPEI} = \Phi^{-1}(F(P - ET))$                                       |
| SPEI spatial anomaly (Panel A)                             | Annual $(P - ET)$ at each grid cell | Normal (spatial standardization)  | $Z(x, y) = \frac{(P - ET)(x, y) - \mu}{\sigma}$                            |
| SPEI (Panel B)                                             | SPEI anomaly $Z(x, y, t)$           | —                                 | $I(t) = \frac{1}{N} \sum_{(x, y) \in \Omega} Z(x, y, t)$                   |
| Water-table depth (WTD; Panel B)                           | Simulated annual WTD                | —                                 | $\text{WTD}(t) = \frac{1}{N} \sum_{(x, y) \in \Omega} \text{WTD}(x, y, t)$ |
| Lagged correlation (Panel C)                               | Index $I(t)$ and $\text{WTD}(t)$    | Pearson correlation               | $r(k) = \text{corr}(I(t), \text{WTD}(t+k))$                                |
| Best-response lag                                          | Lag–correlation curve $r(k)$        | —                                 | $k^* = \arg \max_k  r(k) $                                                 |
| WTD–SPEI regression (Panel D)                              | Index at best lag $I(t - k^*)$      | —                                 | $\text{WTD}(t) \approx a + bI(t - k^*)$                                    |

**Land-cover classification and its correspondence with water table depth change**

Fig. 2 in the main text presents the classification of land-cover types and their spatial correspondence with simulated water-table-depth (WTD) change across North America between 2000 and 2025. Panel (a) shows the original land-cover dataset Commission for Environmental Cooperation (2023), which includes 19 vegetation and surface classes representing forests, shrublands, grasslands, wetlands, croplands, urban areas, barren land, water, and snow/ice. This dataset serves as the basis for identifying regions with varying degrees of human influence. Panel (b) shows a reclassification of the original land-cover map into three aggregated categories (Foley et al., 2005; Cuthbert et al., 2019): (1) strong human effect, including cropland and urban areas; (2) mixed human and climate effect, including grasslands, shrublands, and mixed forests; and (3) mostly climate-driven regions, including boreal forests, wetlands, tundra, barren land, water, and snow/ice. This grouping is used to simplify the interpretation of human versus climate controls on groundwater dynamics. Panel (c) shows the simulated difference in WTD between 2000 and 2025, expressed in centimeters, where positive values indicate a shallower water table (rise) and negative values indicate a deeper water table (decline). To assess spatial correspondence, a systematic grid of points ( $1.5^\circ$ ) is applied across the domain, where black dots indicate locations where human-influenced regions (classes 1 and 2) coincide with declining WTD, and white dots indicate climate-driven regions (class 3) coinciding with rising WTD. All maps are derived from the Water Table Model (WTM) outputs and processed at a consistent spatial resolution, with land-cover classification used as a static descriptor and WTD change calculated from transient model simulations.

**Spatial and temporal comparison of water table depth and terrestrial water storage anomalies**

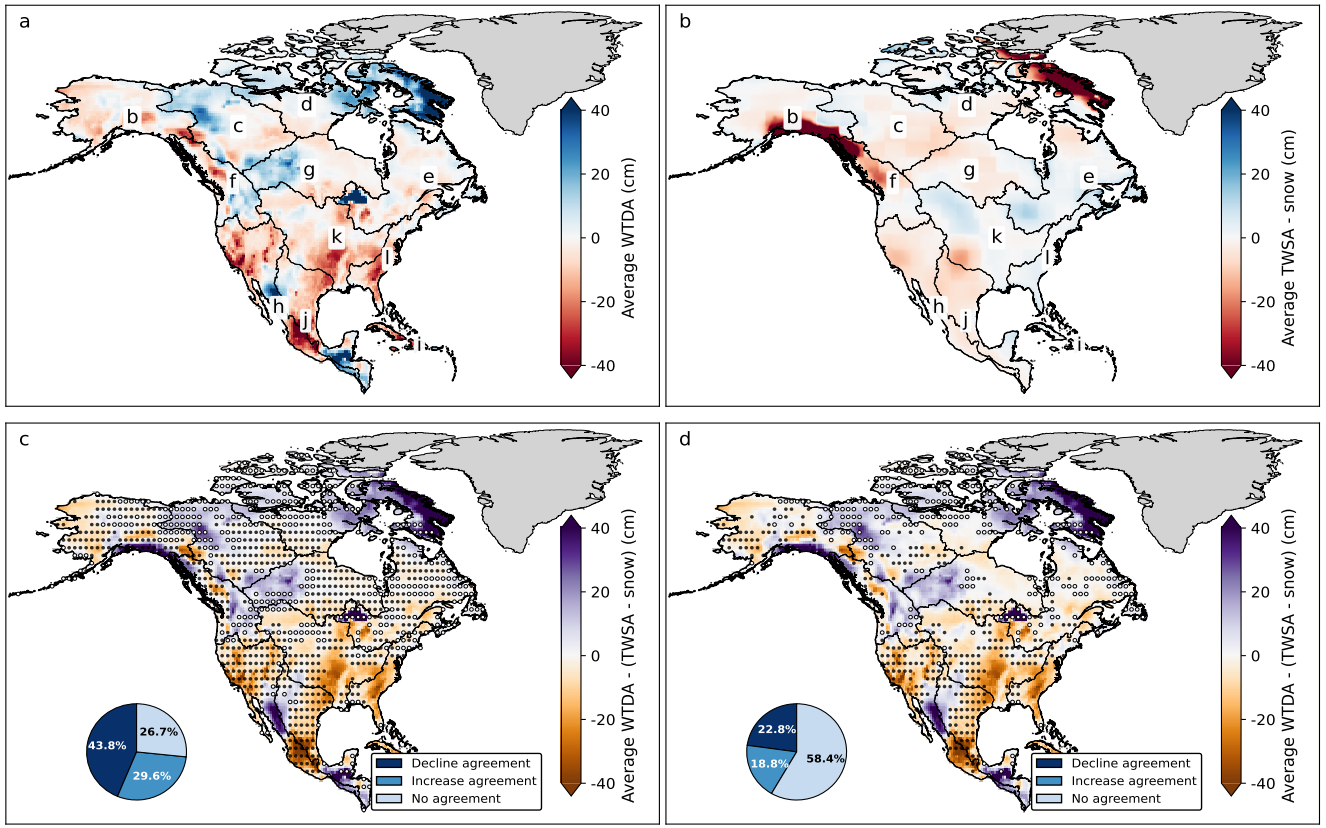
For comparison to GRACE data, we derived water table depth anomalies (WTDA) from monthly Water Table Model simulations and aggregated these to annual means. A 2004–2009 baseline was used to calculate annual anomalies, consistent with the baseline period of the GRACE and GRACE-FO mascon dataset used for comparison (GRACE-FO, D. N. W., 2024). Terrestrial water storage anomalies (TWSA) were obtained from the GRACE and GRACE-FO mascon dataset (JPL RL06.3 v04) using liquid water equivalent thickness GRACE-FO, D. N. W. (2024). The GRACE and GRACE-FO mascon dataset (GRACE-FO, D. N. W., 2024) uses a 2004–2009 baseline period, and so WTDA fields were similarly referenced to this baseline to ensure consistency in the anomaly comparison. Monthly GRACE data were aggregated to annual means. WTDA fields were reprojected to the GRACE grid using bilinear interpolation, and mean annual anomaly maps were calculated over the overlapping analysis period. In supplementary Fig. S5c, dots indicate locations where WTDA and TWSA shared the same anomaly sign, or where their absolute difference was within a tolerance of 20 cm (i.e.,  $|\text{WTDA} - \text{TWSA}| \leq 20\text{cm}$ ), so that pixels with opposite signs but near-zero magnitudes were not misclassified as disagreement. For pixels falling within the tolerance band, the sign of the mean of WTDA and TWSA determined whether they were assigned to decline or increase agreement. In panel (d), the same 20 cm tolerance criterion was applied, with agreement further constrained using grouped land-cover classes (human-influenced, mixed, and climate-driven): decline agreement was assigned to classes 1–2 and increase agreement was assigned



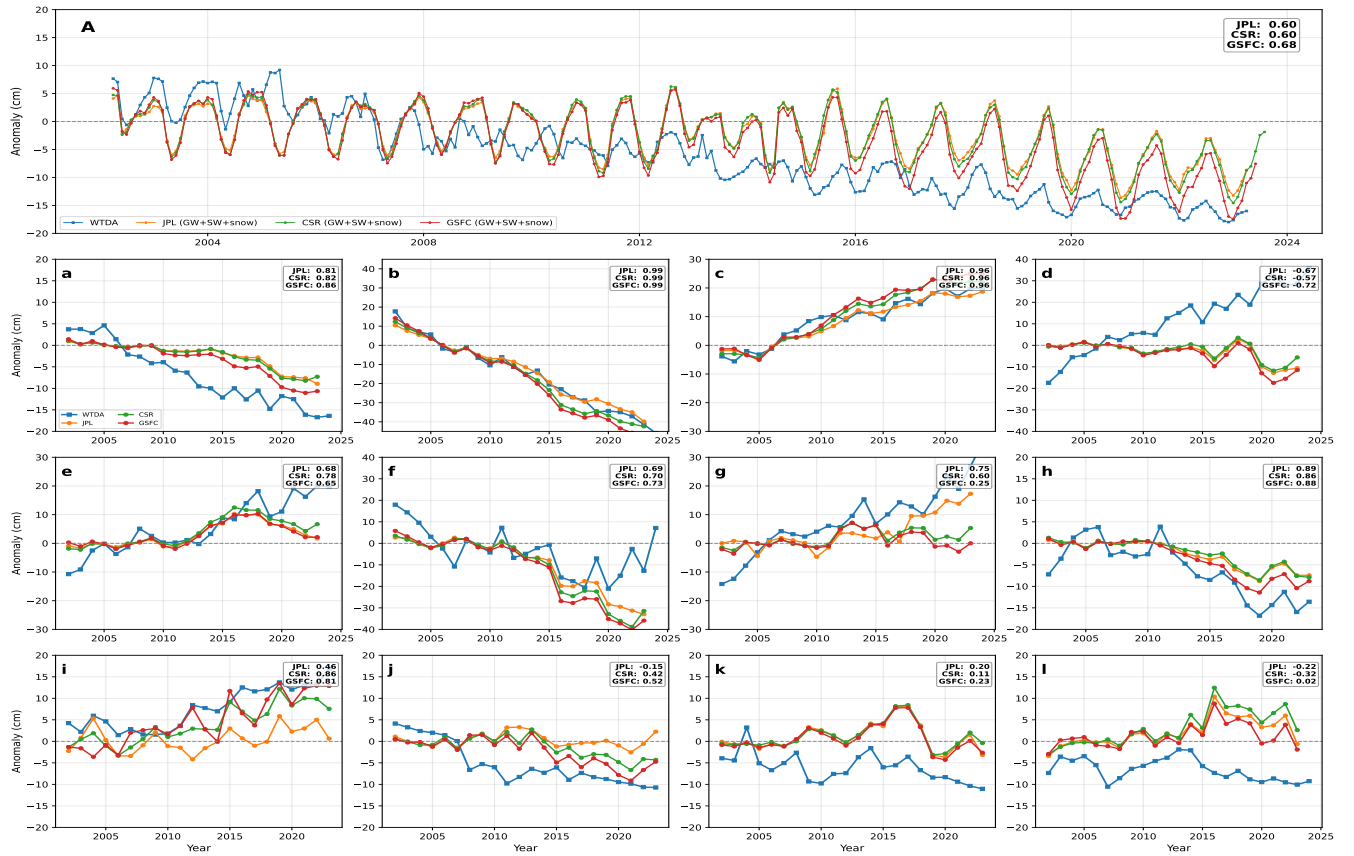
to class 3 as discussed in Section 5. Agreement percentages were calculated as the fraction of valid grid cells satisfying each condition across the North American domain.

835 In addition to the JPL solution, we processed two more mascon products to characterize observational uncertainty in the GRACE signal. The CSR GRACE/GRACE-FO RL06 mascon solution from the Center for Space Research at the University of Texas at Austin (Save et al., 2016), and the GSFC GRACE/GRACE-FO RL06 mascon solution from NASA Goddard Space Flight Center (Luthcke et al., 2013). All three products span the 2002–2025 period and were processed identically: monthly grids were aggregated to annual means, referenced to the common 2004–2009 baseline. Using three independent solutions rather than a single product allows us to quantify processing-related spread in the observed TWSA and to confirm that the  
840 agreement with simulated WTDA is not sensitive to the choice of GRACE mascon solution. Unless otherwise noted, the spatial comparisons in Fig. 3 and supplementary Fig. S5 panels (c) and (d) use the ensemble mean of the three solutions.

Annual and monthly time series were constructed by averaging monthly WTDA and TWSA values within each watershed as defined in fig 2, as well as across the continent. Again, WTDA outputs were converted to centimeters and referenced to the 2004–2009 baseline period. TWSA from GRACE mascon data (GRACE-FO, D. N. W., 2024) were processed using the same  
845 baseline period. Spatial aggregation was performed using Level-2 watershed polygons, with raster data masked and averaged within each watershed to produce representative regional time series.



**Figure S5. Comparison of simulated water table anomalies with isolated GRACE-derived groundwater storage.** (a) Mean water table depth anomaly (WTDA; cm) simulated by the Water Table Model, referenced to the 2004–2009 baseline period. Positive values (blue) indicate rising water tables and negative values (red) indicate water table decline. (b) Mean groundwater storage anomaly (GWSA; cm), obtained by removing the snow water equivalent components from the ensemble-mean terrestrial water storage anomaly of three GRACE and GRACE-FO mascon solutions (JPL, CSR, and GSFC). The snow water equivalent anomalies were computed from GLDAS-NOAH v2.1 over the same 2004–2009 baseline and subtracted pixel-by-pixel from TWSA. (c) Spatial difference between WTDA and GWSA (background map;  $WTDA - GWSA$ ). Dots mark locations where the two datasets share the same anomaly sign, or where their absolute difference falls within a 20 cm tolerance ( $|WTDA - GWSA| \leq 20$  cm), so that pixels with opposite signs but near-zero magnitudes are not misclassified as disagreement. Black dots denote co-occurring negative anomalies (decline agreement) and white dots denote co-occurring positive anomalies (increase agreement); for pixels within the tolerance band, the sign of the mean of WTDA and GWSA determines the dot category. (d) The same  $WTDA - GWSA$  difference map and 20 cm tolerance criterion as in panel (c), but with agreement further constrained by land-cover classification (Fig. 2b). Black dots denote co-occurring negative anomalies in human- and mixed-influenced regions (classes 1 and 2), and white dots denote co-occurring positive anomalies in climate-driven regions (class 3). Pie charts summarize the fractional area of decline agreement, increase agreement, and no agreement across the North American domain.



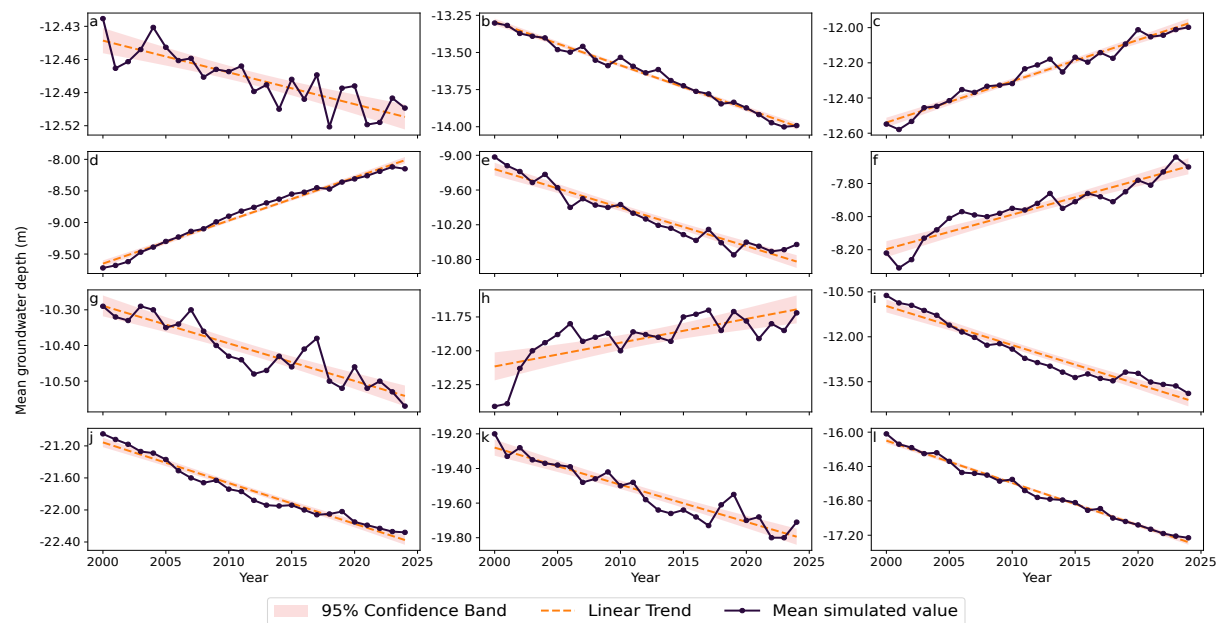
**Figure S6. Temporal comparison of simulated groundwater and surface water anomalies with three GRACE-derived, groundwater-plus-surface-water-isolated storage products across North America.** Panel A shows the monthly mean water-table depth anomaly (WTDA; blue) and the isolated groundwater-plus-surface-water storage anomaly derived from three independent GRACE and GRACE-FO mascon solutions – JPL (orange), CSR (green), and GSFC (red) – averaged across North America from 2002 to 2023. To isolate the groundwater and surface water signal, monthly snow water equivalent from the GLDAS Noah Land Surface Model (GLDAS\_NOAH025\_M v2.1) were subtracted from each GRACE terrestrial water storage anomaly ( $GW+SW = TWSA - snow$ ), making the satellite-derived signal directly comparable to the groundwater and surface water storage simulated by the Water Table Model. Panel a shows the corresponding annual mean anomalies for North America, and panels b–l show annual mean WTDA and isolated storage for individual Level-2 watershed regions. All anomalies are expressed in centimeters relative to the 2004–2009 baseline period. Pearson correlation coefficients between WTDA and each isolated GRACE product are annotated in each panel. The close correspondence between simulated groundwater and surface water variability and all three isolated satellite-derived products across most watersheds highlights the model’s ability to capture large-scale hydroclimatic fluctuations, while the spread among the three GRACE solutions provides a measure of observational uncertainty.

The differences between the simulated water table anomalies and GRACE-derived terrestrial water storage anomalies (Fig. 3) arise from a combination of factors, including differences in what the two anomalies represent (water tables alone for WTDA versus terrestrial water anomaly), structural limitations of the WTM, residual short- and long-term differences between bias-

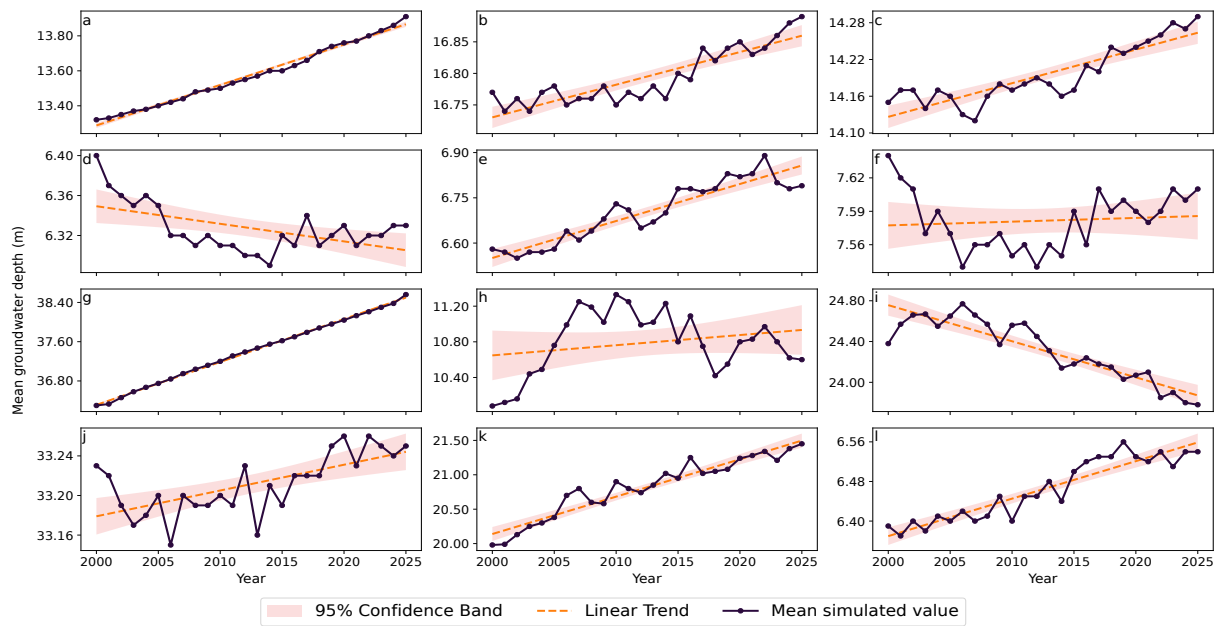
850 corrected and observed climate forcing, the exclusion of direct human water pumping and irrigation from WTM simulations, differences between WTM simulation resolution and GRACE spatial resolution, and limitations related to model resolution and input data availability. The pronounced storage decline in the southern watersheds (k and l) likely reflects in part the warm and dry bias that climate models exhibit over the central-southern Great Plains (Sun and Liang, 2023), a discrepancy expected to diminish as such biases are reduced.

855 **Water Table Depletion Rates**

**Long-term trends in groundwater and surface water (2000–2025)**

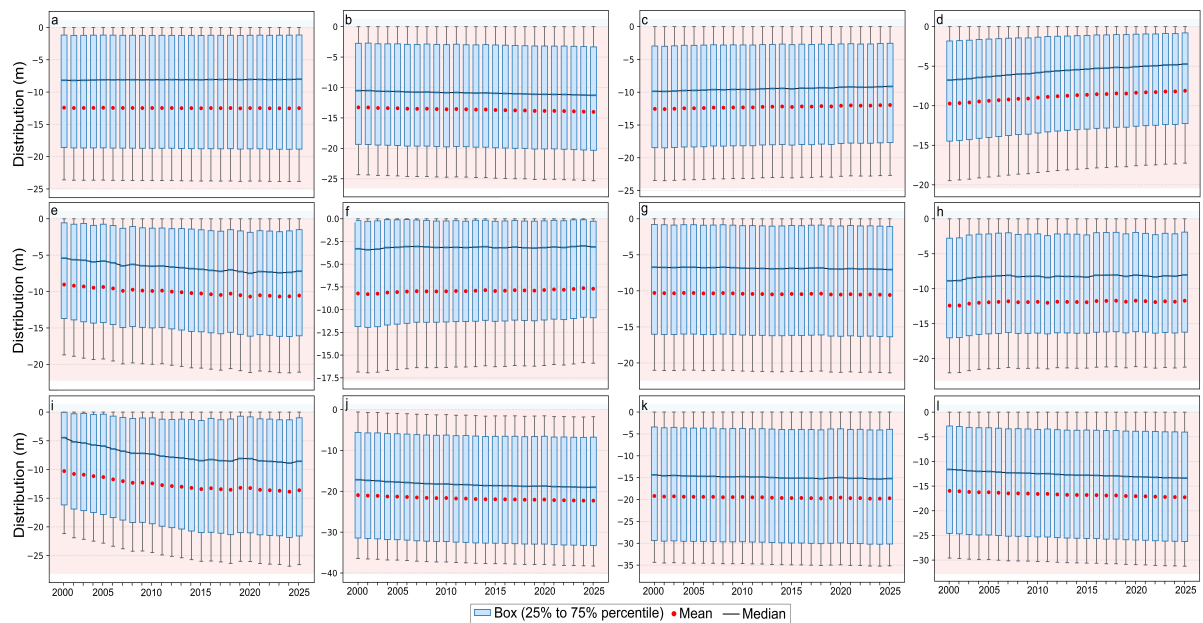


**Figure S7. Temporal trends of mean groundwater depth (m) from 2000 to 2025.** Panel (a) shows the trend for the entire North American domain, while panels (b–l) represent individual major watersheds. The purple solid lines depict observed annual mean groundwater depth, and the orange dashed lines indicate linear trends fitted to each time series. The shaded orange bands represent the 95 percent confidence intervals of the fitted regressions, illustrating the uncertainty around the estimated depth trends across regions.

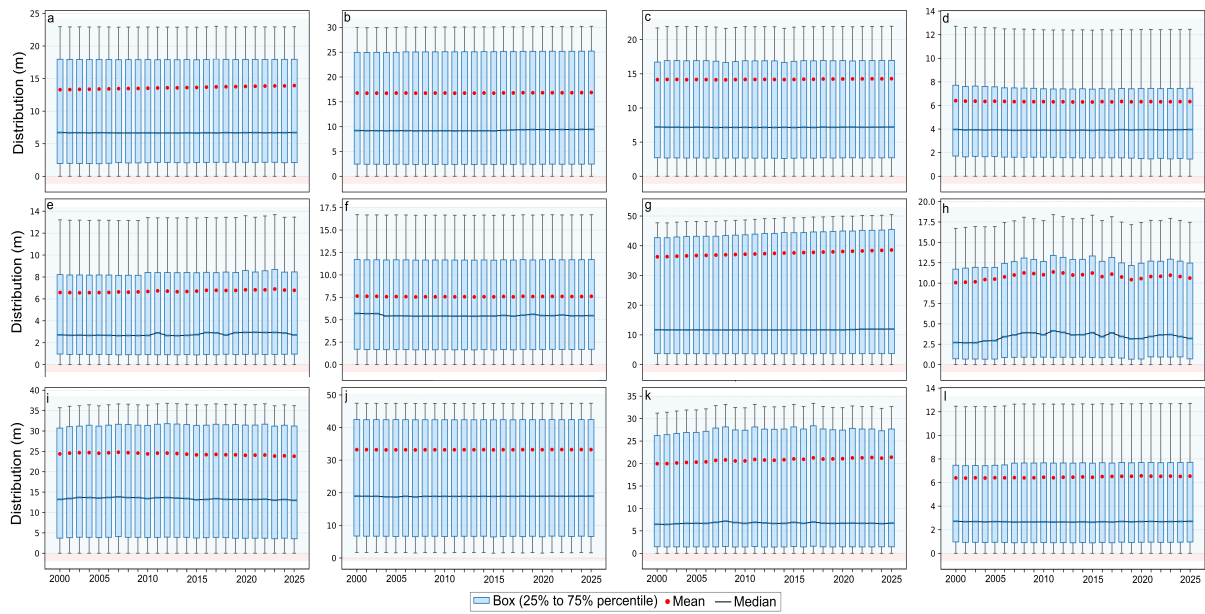


**Figure S8. Temporal trends of mean surface water depth (m) from 2000 to 2025.** Panel (a) shows the trend for the entire North American domain, while panels (b–l) display the trends for individual major watersheds. The purple solid lines represent the observed annual mean surface-water depth (m), and the orange dashed lines show the fitted linear trends for each watershed. The shaded orange envelopes denote the 95 percent confidence intervals, illustrating the statistical uncertainty associated with each trend estimate.

Figs. S9 and S10 show the annual statistical distributions of groundwater and surface-water depth across North America and major watersheds for the same period. For each year, boxplots were constructed using all grid-cell values within the selected spatial domain. Boxes represent the interquartile range (25th–75th percentile), whiskers denote the full range of values, the central line indicates the median, and the red markers denote the mean. The separation between mean and median indicates that extreme deep values skew the mean below the median, suggesting that localized but intense depletion events influence the continental average groundwater depth.



**Figure S9. Temporal distribution of groundwater depth (m) from 2000 to 2025.** Panel (a) shows the overall depth pattern for North America, while panels (b–l) represent individual major watersheds. Each box illustrates the interquartile range (25th–75th percentile) of depth, with whiskers extending to the full range. The red dashed line denotes the mean, and the black solid line inside the boxes indicates the median.



**Figure S10. Temporal distribution of surface water depth (m) from 2000 to 2025.** Panel (a) depicts the overall surface water depth (m) distribution for the entire North American domain, while panels (b–l) show the distributions for the major individual watersheds. Each yearly boxplot represents the interquartile range (25th–75th percentile) of depth, with whiskers extending to the minimum and maximum observed values. The red dots indicate the annual mean depth, and the black horizontal line inside each box represents the annual median.

The groundwater (Negative value in model output) results (Supplementary Fig. S7) quantitatively confirm the groundwater depletion across North America during 2000–2025. For the continental domain (panel a), the mean groundwater depth increased from approximately  $-12.42$  m in 2000 to  $-12.52$  m by 2025, yielding a negative slope of  $-0.004$  m/yr and reflecting a persistent acceleration in net loss. Among the watersheds, the Mackenzie (b) shows a steady decline from  $-13.28$  to  $-14.00$  m ( $-0.029$  m/yr). Hudson Bay (c) rises from  $-12.55$  to  $-12.00$  m ( $+0.022$  m/yr), consistent with recharge-dominated cold regions. The Arctic (d) shows a gradual rise in mean depth from  $-9.45$  to  $-8.10$  m, indicating sustained recharge. In contrast, subtropical and mid-latitude watersheds show pronounced depletion trajectories. The Caribbean (i) declines from  $-10.52$  to  $-13.55$  m ( $-0.121$  m/yr), Baja California (h) shows modest recharge from  $-12.25$  to  $-11.70$  m ( $+0.022$  m/yr), and the Pacific Coastal (f) trends upward from  $-8.21$  to  $-7.68$  m ( $+0.021$  m/yr), suggesting modest recharge linked to coastal precipitation anomalies. The Rocky Mountain–Prairie (g) watershed shows negative slope ( $-0.010$  m/yr). The Great Lakes–St. Lawrence (e) exhibits a steady decline from  $-9.00$  to  $-10.60$  m ( $-0.064$  m/yr), while the Southeastern U.S. (l) and Mississippi (k) watersheds experience strong depletion, with linear slopes near  $-0.048$  and  $-0.023$  m/yr, respectively. The Central America (j) watershed records depletion from  $-21.10$  to  $-22.38$  m ( $-0.051$  m/yr). These linear trends align with the widening interquartile ranges in the corresponding box plots (Supplementary Fig. S9). We provide a summary of groundwater changes across North America and its watersheds in Table S2.

The surface water (Positive value in model output) results (Supplementary Fig. S8) show watershed-dependent but measurable long-term changes in surface storage across North America from 2000 to 2025. For the continental domain (panel a), the mean surface-water distribution remains centered around 13.35–13.89 m, indicating slope +0.022 m/yr. The Mackenzie (b) watershed exhibits an increase in mean storage from 16.78 to 16.89 m (+0.004 m/yr), while Hudson Bay (c) rises from 14.15 to 14.29 m (+0.006 m/yr), consistent with enhanced high-latitude runoff. The Arctic basin (d) shows a slight decline from 6.40 to 6.34 m (−0.002 m/yr), reflecting modest reduction in annual precipitation. Great Lakes–St. Lawrence (e) increases from 6.58 to 6.85 m (+0.011 m/yr). The Pacific Coastal (f) region is nearly stationary, fluctuating between 7.64–7.61 m with a slope close to zero. Rocky Mountain–Prairie (g) shows one of the strongest increases, rising from 36.89 to 38.42 m (+0.061 m/yr), driven by precipitation-dominated hydrology. Baja California (h) increases slightly from 10.31 to 10.72 m (+0.016 m/yr). Caribbean (i) shows a stronger decrease, with mean values from 24.40 to 23.85 m (−0.022 m/yr). Central America (j) shows a small but consistent increase from 33.23 to 33.27 m (+0.002 m/yr). The Mississippi basin (k) exhibits a clear upward trend from 20.00 to 21.50 m (+0.060 m/yr), with widening variability after 2010, while the Southeastern U.S. (l) rises from 6.38 to 6.54 m (+0.006 m/yr). These linear trends are reflected in the boxplots (Supplementary Fig. S10), where the median–mean separation indicates greater variability in mid-latitude and subtropical watersheds compared to the narrower and more stable distributions in northern basins. We provide a summary of surface water changes across North America and its watersheds in Table S3.

**Table S2.** Mean groundwater depth (negative values in model output) and linear trends for North America and major watersheds, 2000–2025.

| Region                       | 2000 (m) | 2025 (m) | Slope (m/yr) | Trend            |
|------------------------------|----------|----------|--------------|------------------|
| North America (a)            | -12.42   | -12.52   | -0.004       | Depletion        |
| Mackenzie (b)                | -13.28   | -14.00   | -0.029       | Depletion        |
| Hudson Bay (c)               | -12.55   | -12.00   | 0.022        | Recharge         |
| Arctic (d)                   | -9.45    | -8.10    | 0.054        | Recharge         |
| Great Lakes–St. Lawrence (e) | -9.00    | -10.60   | -0.064       | Depletion        |
| Pacific Coastal (f)          | -8.21    | -7.68    | -0.021       | Recharge         |
| Rocky Mtn–Prairie (g)        | -10.30   | -10.55   | -0.010       | Depletion        |
| Baja California (h)          | -12.25   | -11.70   | 0.022        | Recharge         |
| Caribbean (i)                | -10.52   | -13.55   | -0.121       | Strong depletion |
| Central America (j)          | -21.10   | -22.38   | -0.051       | Depletion        |
| Mississippi (k)              | -19.20   | -19.78   | -0.023       | Depletion        |
| Southeastern U.S. (l)        | -16.02   | -17.23   | -0.048       | Depletion        |

**Wet-dry dominance and spatial variability metrics**

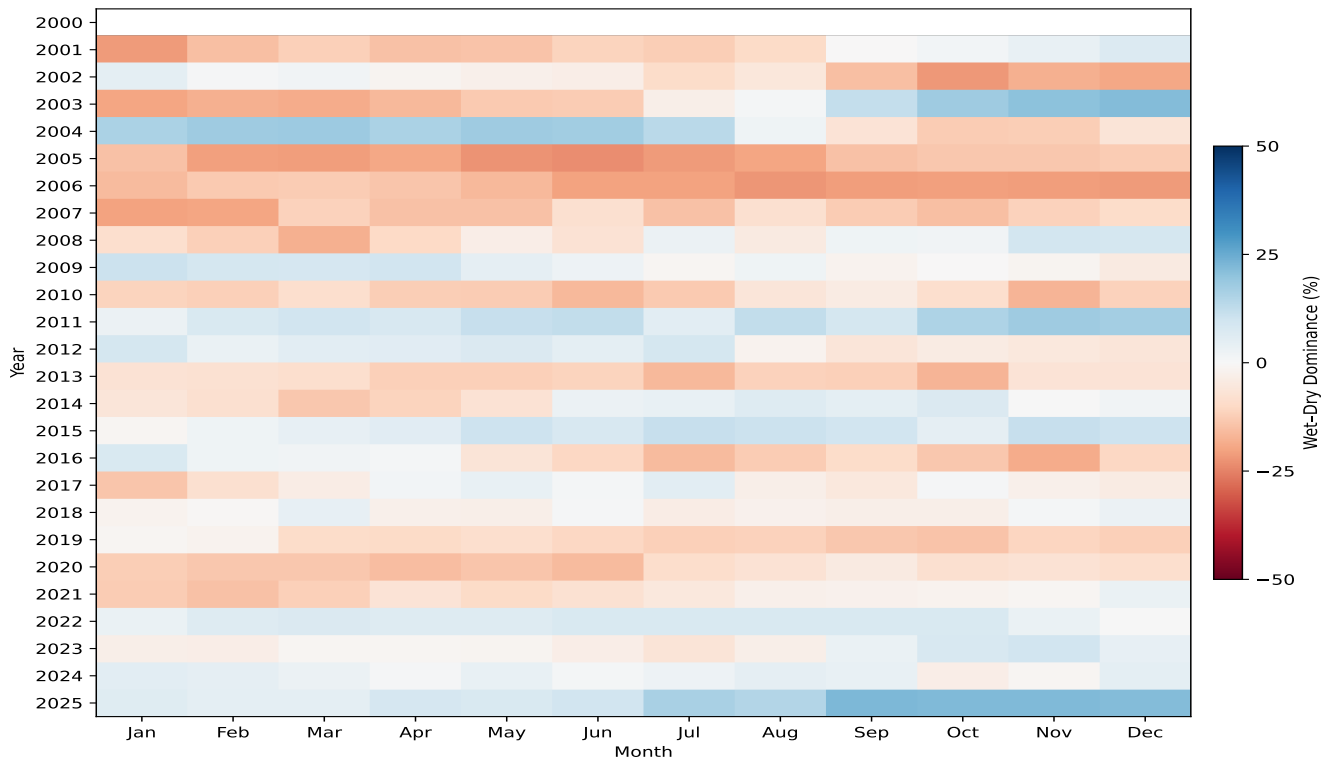
Fig. S11 presents a monthly heatmap of wet and dry dominance across North America from 2000 to 2025, based on the precipitation minus evaporation balance. Blue tones indicate wetter-than-average conditions, whereas red tones indicate drier-than-average conditions, with color intensity reflecting the magnitude of dominance. The figure reveals pronounced seasonal



**Table S3.** Mean surface-water storage (positive values in model output) and linear trends for North America and major watersheds, 2000–2025.

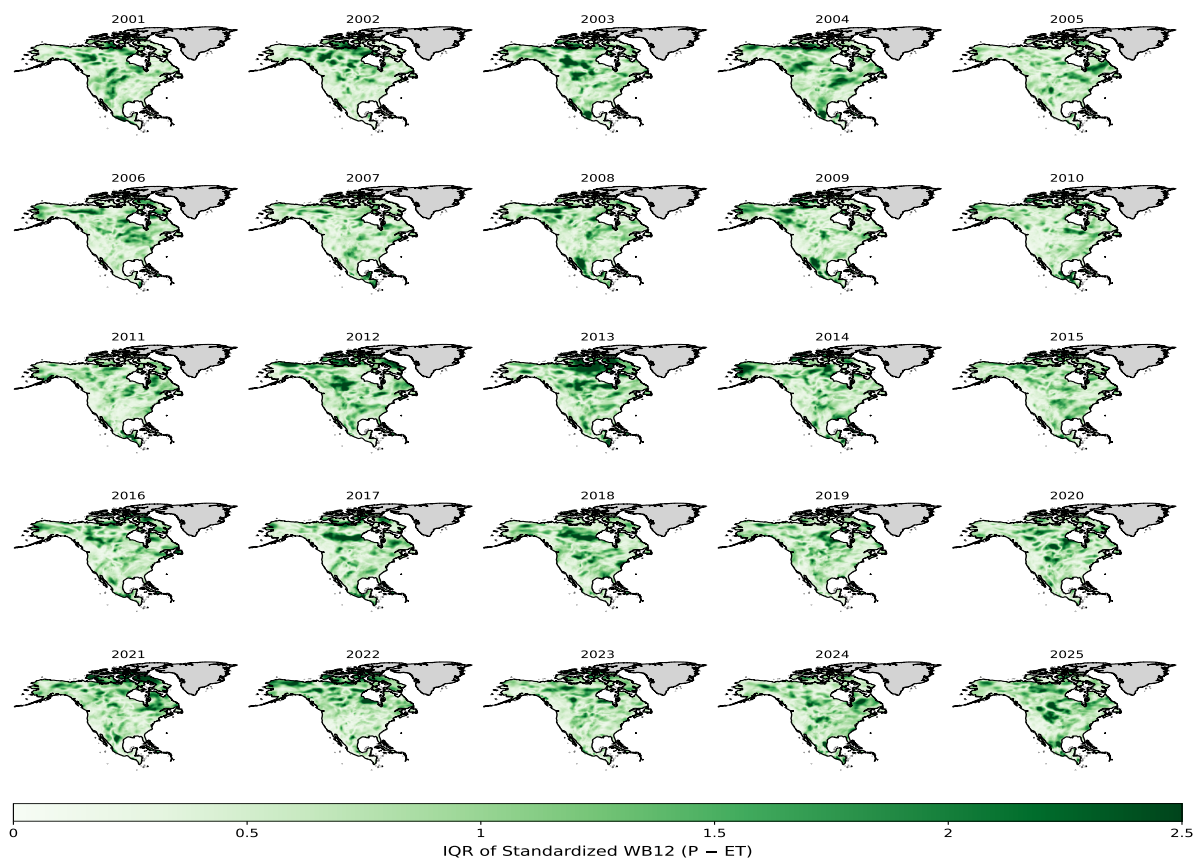
| Region                       | 2000 (m) | 2025 (m) | Slope (m/yr) | Trend    |
|------------------------------|----------|----------|--------------|----------|
| North America (a)            | 13.35    | 13.89    | 0.022        | Increase |
| Mackenzie (b)                | 16.78    | 16.89    | 0.004        | Increase |
| Hudson Bay (c)               | 14.15    | 14.29    | 0.006        | Increase |
| Arctic (d)                   | 6.40     | 6.34     | -0.002       | Stable   |
| Great Lakes–St. Lawrence (e) | 6.58     | 6.85     | 0.011        | Increase |
| Pacific Coastal (f)          | 7.64     | 7.61     | -0.001       | Stable   |
| Rocky Mtn–Prairie (g)        | 36.89    | 38.42    | 0.061        | Increase |
| Baja California (h)          | 10.31    | 10.72    | 0.016        | Increase |
| Caribbean (i)                | 24.40    | 23.85    | -0.022       | Decrease |
| Central America (j)          | 33.23    | 33.27    | 0.002        | Stable   |
| Mississippi (k)              | 20.00    | 21.50    | 0.060        | Increase |
| Southeastern U.S. (l)        | 6.38     | 6.54     | 0.006        | Increase |

and interannual variability, including periods of sustained dry or wet dominance and distinct transitions between hydroclimatic regimes over time.



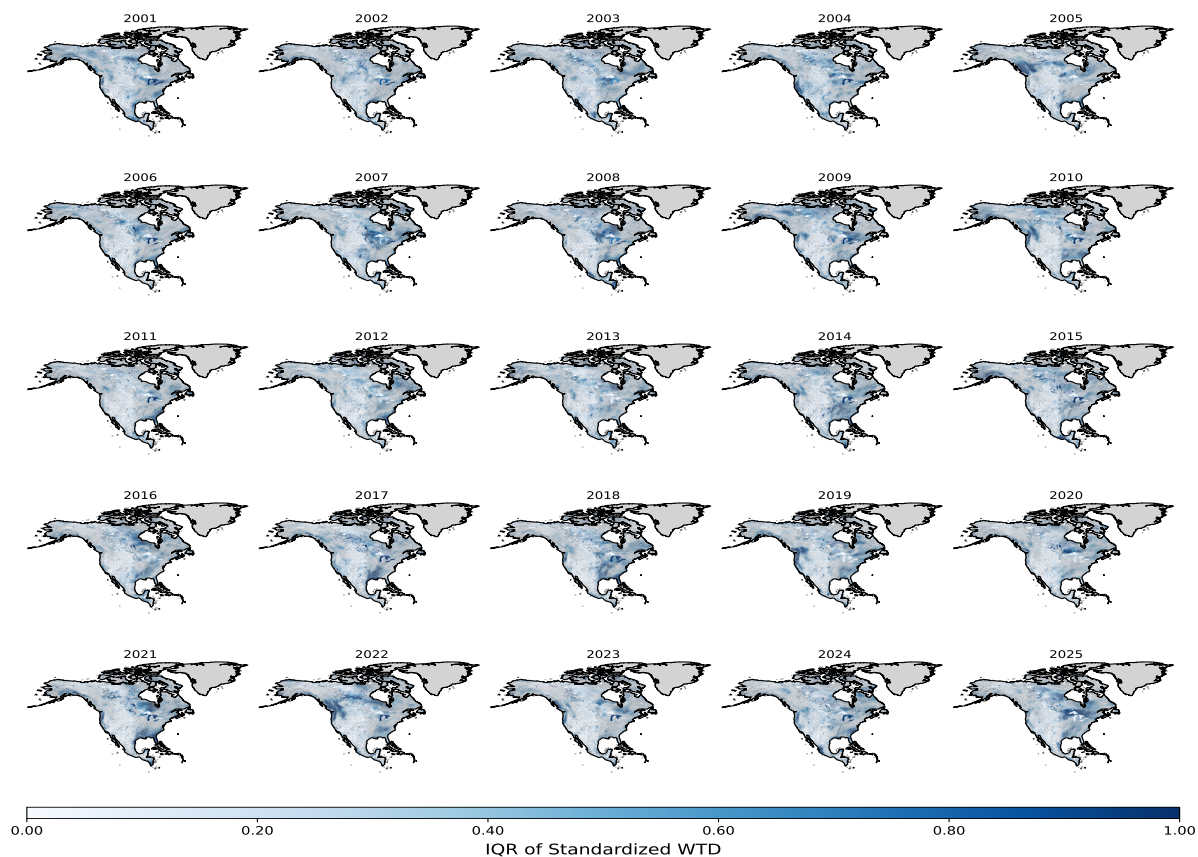
**Figure S11. Monthly wet–dry dominance across North America from 2000 to 2025.** Colors represent the percentage dominance of wet (blue, positive values) versus dry (red, negative values) conditions for each month and year, computed from the precipitation–evaporation balance. Values near zero indicate near-neutral conditions, while larger magnitudes indicate stronger wet or dry dominance.

Fig S12 shows annual maps of the interquartile range (IQR) of standardized 12-month water balance (WB12 = precipitation minus evapotranspiration) across North America for the period 2001–2025. The IQR quantifies the spread of monthly standardized WB12 values within each year and therefore represents the degree of hydroclimatic variability rather than mean wetness or dryness. Regions with higher IQR values exhibit larger fluctuations between wet and dry conditions over the course of a year, indicating stronger variability in the water balance, whereas lower IQR values reflect more temporally stable conditions.



**Figure S12. Annual interquartile range (IQR) of standardized 12-month water balance ( $WB12 = P - ET$ ) across North America, 2001–2025.** Each panel shows the spatial distribution of the IQR of monthly standardized WB12 values within a given year. Higher IQR values (darker green) indicate greater intra-annual variability in water-balance conditions, whereas lower values (lighter green) indicate more temporally stable conditions.

Fig. S13 summarizes how strongly groundwater conditions fluctuate within each year. At every grid cell, monthly WTD values were first standardized (z-scored) so variability is comparable across regions with different absolute water-table depths. Then, for each year, the IQR (75th minus 25th percentile) was calculated from the 12 monthly standardized values. Because IQR is robust to outliers, it captures the typical seasonal spread rather than being dominated by a few extreme months. Higher IQR (darker blue) highlights regions where WTD varies strongly within the year (stronger seasonality or higher hydroclimatic variability), whereas lower IQR (lighter blue) indicates relatively stable groundwater conditions through the year.



**Figure S13. Annual maps of intra-annual groundwater variability across North America (2001–2025), expressed as the interquartile range (IQR) of standardized water-table depth (WTD).** Each panel shows the spatial distribution of within-year groundwater variability for the corresponding year.

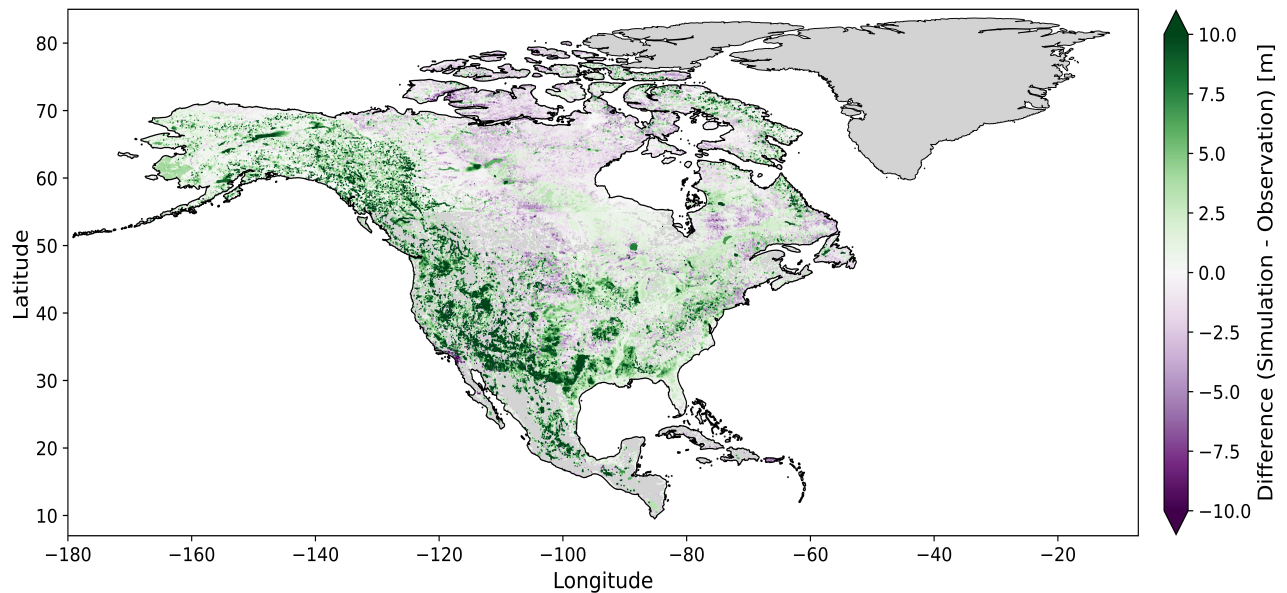
## 910 Validation of Water-Table Depth

Simulation accuracy was evaluated by comparing simulated groundwater and surface-water levels with observational datasets. Monthly validation against USGS well and surface-water records for 2024 yielded an average coefficient of determination ( $R^2$ ) of 0.9239 and Nash–Sutcliffe Efficiency of 0.9190. Monthly values of  $R^2$  and NSE are reported in Table S4. Comparison with FLake Choulga et al. (2014) lake depth observations, which are static and not temporally varying, was conducted to

915 assess model representation of surface-water dynamics. Supplementary Fig. S14 shows the relationship between simulated and observed water levels.

**Table S4.** Monthly R<sup>2</sup> and NSE values for 2024.

| Month    | R <sup>2</sup> | NSE    | Month     | R <sup>2</sup> | NSE    |
|----------|----------------|--------|-----------|----------------|--------|
| January  | 0.9539         | 0.9490 | July      | 0.9540         | 0.9491 |
| February | 0.9539         | 0.9490 | August    | 0.9540         | 0.9492 |
| March    | 0.9539         | 0.9490 | September | 0.9540         | 0.9492 |
| April    | 0.9539         | 0.9490 | October   | 0.9540         | 0.9493 |
| May      | 0.9539         | 0.9490 | November  | 0.9541         | 0.9493 |
| June     | 0.9539         | 0.9491 | December  | 0.9539         | 0.9490 |



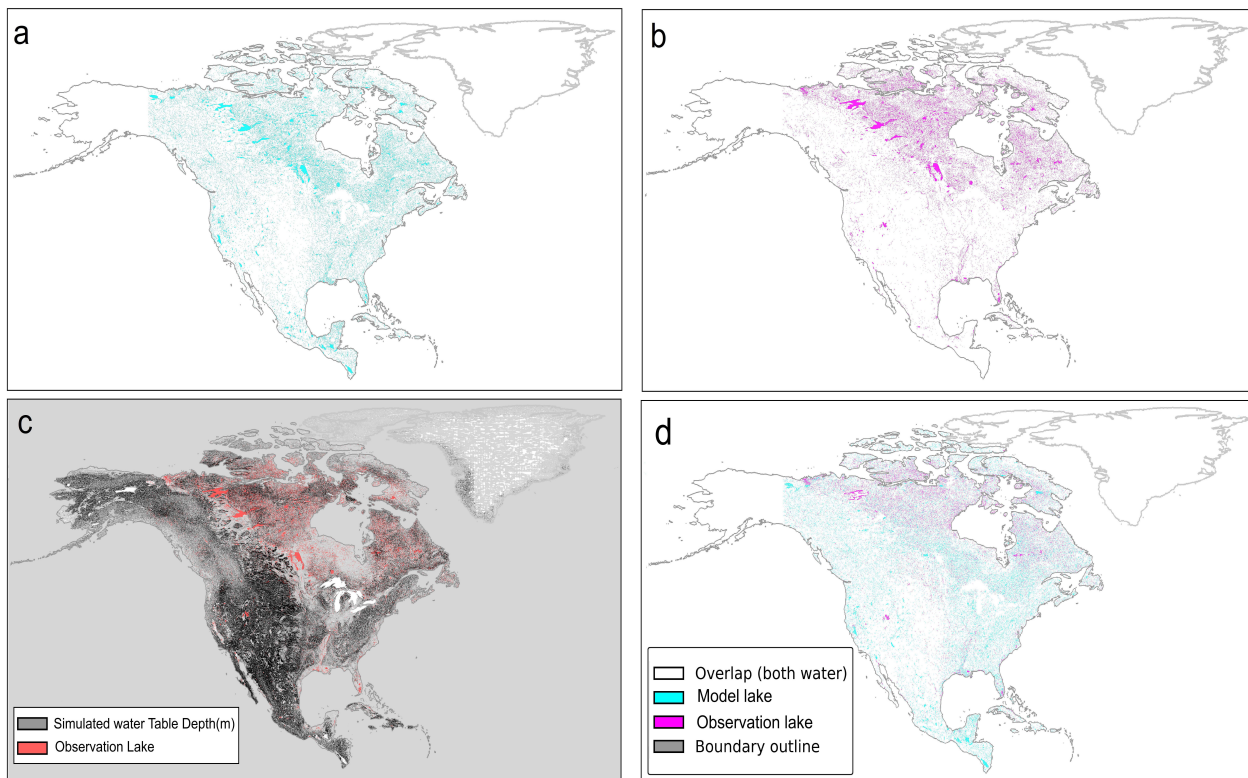
**Figure S14. Spatial comparison of simulated and observed water levels across North America (year 2024).** Difference between simulated and observed water levels (Simulation – Observation, in meters), aggregated to the model grid. Observations combine groundwater well records (Vector data with size 3 pt) with FLake lake levels, each matched to the simulation cell in which it falls. Green indicates that the simulated water level is higher (shallower water-table) than observed, purple indicates a lower (deeper) simulated level, and white indicates close agreement. Areas with no observational coverage are shown in light gray

Continental-scale validation against observed lake extent datasets is shown in Supplementary Fig. S15 by using these equations:

920     $\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN}$

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad \text{IoU} = \frac{TP}{TP + FP + FN}, \quad \text{Accuracy} = \frac{TP + TN}{N}$$

where  $TP$ ,  $FP$ ,  $FN$ , and  $TN$  denote true positives, false positives, false negatives, and true negatives, respectively, and  $N$  represents the total number of grid cells evaluated. These spatial metrics assess the model's ability to reproduce observed lake presence and quantify agreement between simulated and observed surface-water distributions (false negatives may reflect missing lake ice in the evaporation scheme; false positives may arise from small topographic depressions or small lakes not captured in the observational dataset, which only includes larger, permanent lakes). Fig. S15 presents a comparison between modeled surface-water extent inferred from simulated water-table depth and observed lake distributions across North America. Panel (a) shows the modeled standing-water extent, panel (b) shows the observed lake union dataset, and panel (c) displays their spatial overlap. Panel (d) presents the spatial classification of agreement and disagreement between modeled and observed surface-water areas. The dominance of lake-only water in northern latitudes likely reflects a limitation of the evaporation scheme used in the simulation, which does not explicitly account for seasonal lake ice. As a result, northern lakes may experience excessive simulated evaporation, leading to underestimation of lake extent in the model.



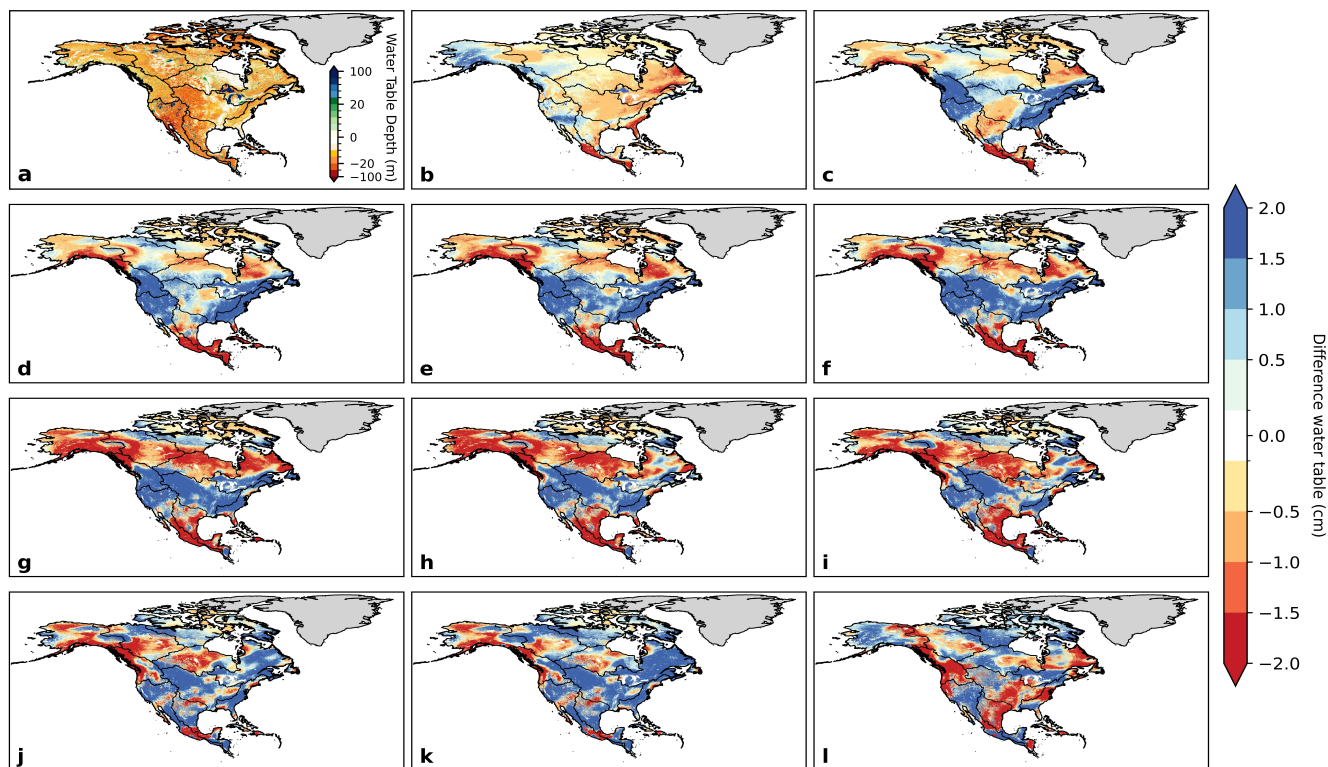
**Figure S15. Spatial comparison between modeled and observed lake extents across North America.** (a) Modeled surface-water areas, derived from grid cells where simulated water-table depth exceeds 0 m. (b) Observed lake extent from the merged observational lake dataset (union of available lake rasters). (c) Overlay of modeled and observed lake extents. (d) Spatial classification of agreement and disagreement between modeled and observed lake extents.

## Seasonal and Long-Term Sea-Level Equivalent (SLE) Dynamics

### 935 Seasonal groundwater variability

The simulations presented in this research were performed at monthly temporal resolution. Fig. S16 shows an example of the month-to-month change in WTD for the year 2024, capturing the monthly-scale dynamics of climate-driven water table fluctuation across the continent. These maps demonstrate the spatial and temporal evolution of groundwater storage throughout the year, highlighting seasonal recharge and depletion patterns across diverse climatic zones.

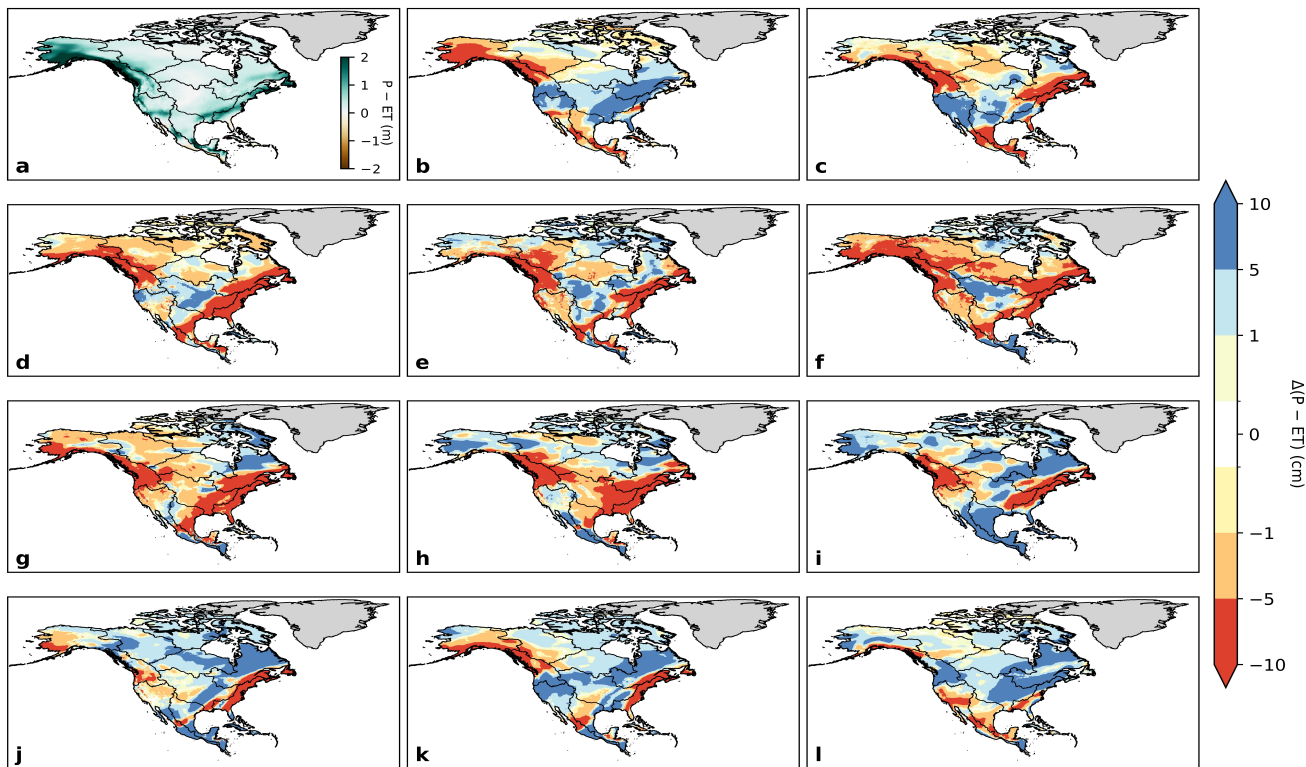




**Figure S16. Spatial distribution of monthly water-table depth differences across North America for the year 2024.** Panel (a) shows the modeled water-table depth for January 2024, serving as the baseline. Panels (b–l) illustrate the monthly (February to December) differences in water-table depth relative to January 2024, with blue shades indicating a water-table rise (positive anomaly) and red shades indicating a water-table decline (negative anomaly). The color bar represents the magnitude of change in centimeters.

940 Fig. S17 presents the spatial distribution of monthly differences in precipitation minus evapotranspiration (P–ET) across North America for 2024. Monthly P–ET values were calculated at each grid cell as the difference between precipitation and evapotranspiration at monthly temporal resolution. These maps were generated to document the monthly evolution of climatic water-balance conditions and to provide the corresponding climatic context for the groundwater variations shown in Supplementary Fig. S16 and Fig. 7 of the main text.



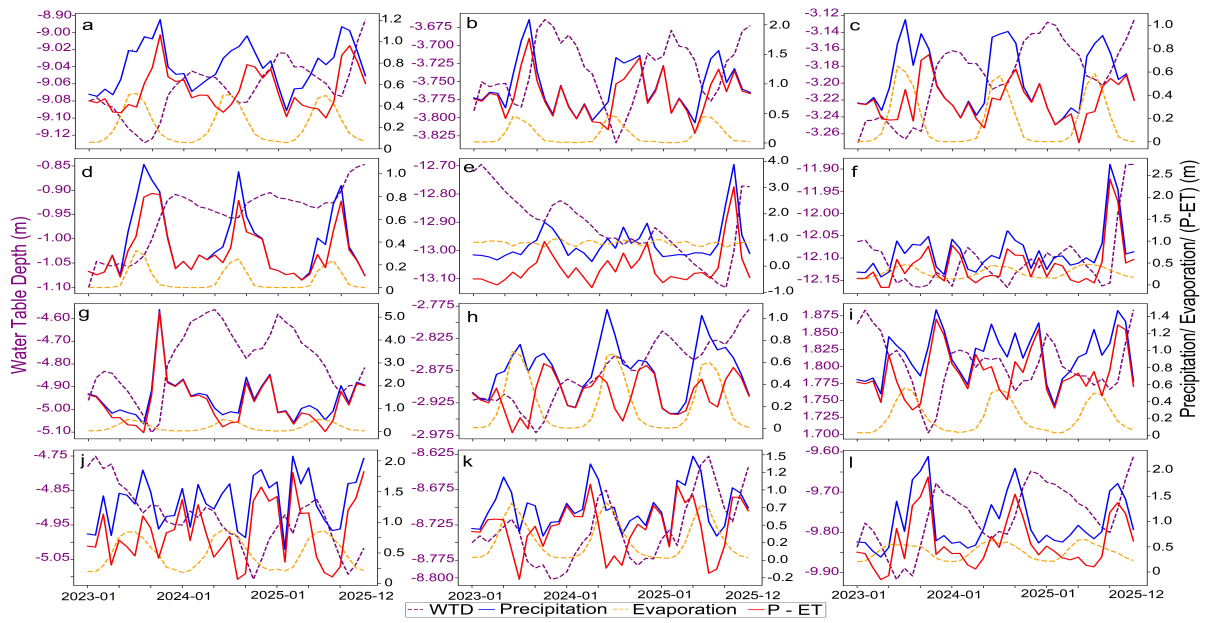


**Figure S17. Monthly differences in precipitation minus evapotranspiration (P-ET) across North America for the year 2024.** Panel (a) represents the January 2024 baseline, while panels (b-l) illustrate the monthly deviations in P-ET relative to January, with blue shades indicating positive anomalies (wetter) and red shades indicating negative anomalies (drier). The color bar quantifies the magnitude of P-ET change in centimeters.

#### 945 **Watershed-scale climate-groundwater coupling**

Continent-wide and major watershed-scale changing average WTD, P-ET, P, and ET from 2023 to 2025 are shown in Fig. S18. Spatial averages were computed for the entire North American domain and for each major watershed (Fig. 1) using area-weighted grid-cell means. Each panel shows the monthly time series of WTD together with the corresponding climatic variables for the selected region. All variables were aggregated from monthly model outputs at the 30 arc second resolution.

950 The close correspondence between precipitation–evaporation balance (P-ET) and water-table depth (WTD) emphasizes the direct influence of surface hydrologic inputs on subsurface storage. Overall, Fig. S18 demonstrates how the seasonal balance between precipitation and evaporation governs groundwater variability, with distinct spatial contrasts—humid regions showing dynamic but reversible fluctuations, while arid regions undergo sustained declines that contribute to the accelerating trend in groundwater depletion across the continent.



**Figure S18. Monthly variation of simulated water-table depth (WTD, purple dashed line), precipitation (blue solid line), evaporation (orange dashed line), and precipitation–evaporation balance ( $P - ET$ , red solid line) across major North American watersheds for the year 2023 to 2025. Each panel (a–l) corresponds to a Level-2 watershed, illustrating the monthly coupling between groundwater and climatic variables.**