Gleeson et al. “The global volume and distribution of modern groundwater”

1) Global data synthesis and calculation of total groundwater volume
   - Porosity → Total groundwater volume
   - Spatial data
   - Tritium
   - Watershed assignment

2) Numerical modeling
   - Model development
   - Groundwater age metrics

3) Analysis of modern groundwater storage for 30 aquifers
   - Model results + Tritium results
     - \( V_{\text{storage}} \) comparison
     - \( d_{\text{effective}} \) comparison

4) Global volume of modern and young groundwater
   - Model estimate
   - Tritium estimate
   - Model results

5) Groundwater recharge analysis

Figure S1. Schematic summary of methods.
1. Previous estimates of global groundwater volume

In our analysis of modern groundwater, we calculate a new estimate of the global volume of groundwater by analyzing tritium measurements from around the world and running thousands of numerical simulations of groundwater flow and age transport (Figure S1). Early estimates of the total volume of groundwater were 15 – 1175 km³, which is “absurdly small because a single major aquifer may contain hundreds of cubic kilometers of water”¹. Nace¹ also wrote that it was “not possible to estimate accurately the total amount of fresh groundwater” but roughly estimated 1 - 7 million km³ assuming 1-5% porosity for the upper 1 km. A widely referenced previous estimate² of the volume of global groundwater (Figure S2) used rough approximations of topography and assumed porosity. Total global groundwater stores of 23.4 million km³ were arbitrarily divided into 3.6 million km³ of active groundwater (elevations above local river elevation in a watershed assumed to have 15% porosity), 6.2 million km³ of less active groundwater (river elevations to sea level assumed to have 12% porosity), and 13.6 million km³ of stagnant groundwater (sea level to 2 km below sea level assumed to have 5% porosity).

Of the estimates that exist, only Garmonov² offered a quantitative, data-driven framework for the estimate of total groundwater resources on Earth (Figure S2 and Table S1). Other estimates have order-of-magnitude uncertainties and do not claim to accurately calculate groundwater storage¹,³-⁵. Any quantification of fluxes within the larger hydrologic cycle depends on an accurate estimate of the reservoir through which these fluxes occur, particularly for groundwater. Moreover, groundwater depletion can be better understood when the total available groundwater is known. While we do not attempt to determine the volume of total groundwater spatially, our new global estimate of total groundwater in the upper 2 km of the Earth’s crust incorporates 40 years of continued research and data collection.

Figure S2. Comparison of our results to the most recent estimate of global groundwater storage². The size of the water drop scales with the global volume of water stored in each groundwater compartment. See Figure 5 for comparison with other components of the water cycle.
Table S1. Summary of previous groundwater storage, groundwater recharge, and surface water runoff estimates. For the Global Land Data Assimilation System (GLDAS) estimates, it is assumed that groundwater recharge equals the remainder of precipitation after surface runoff and evapotranspiration as done in the work by Fan et al.7.

<table>
<thead>
<tr>
<th></th>
<th>Fresh storage</th>
<th>Recharge</th>
<th>Surface runoff</th>
<th>Recharge/Runoff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10⁶ km³</td>
<td>10⁶ km³/yr</td>
<td>10⁶ km³/yr</td>
<td>%</td>
</tr>
<tr>
<td>Nace (1969)¹</td>
<td>1 to 7</td>
<td>1.5</td>
<td>29.5</td>
<td>5.1</td>
</tr>
<tr>
<td>Nace (1971)⁴</td>
<td>4 to 60</td>
<td>6.0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Garmonov (1974)²</td>
<td>23.4 (3.6 active)</td>
<td>13.3</td>
<td>43.8</td>
<td>30.4</td>
</tr>
<tr>
<td>L'Vovich (1974)³</td>
<td>60 (4 active)</td>
<td>12.0</td>
<td>38.8</td>
<td>30.9</td>
</tr>
<tr>
<td>Döll et al. (2002)⁸</td>
<td>-</td>
<td>13.8</td>
<td>38.3</td>
<td>36.0</td>
</tr>
<tr>
<td>NRC (1986)⁵</td>
<td>15.3</td>
<td>-</td>
<td>40.0</td>
<td>-</td>
</tr>
<tr>
<td>Döll and Fiedler (2008)⁹</td>
<td>-</td>
<td>12.7</td>
<td>39.4</td>
<td>32.2</td>
</tr>
<tr>
<td>Wada et al. (2010)¹⁰</td>
<td>-</td>
<td>15.2</td>
<td>36.2</td>
<td>42.0</td>
</tr>
<tr>
<td>GLDAS⁷ CLM annual mean</td>
<td>-</td>
<td>17.7</td>
<td>21.6</td>
<td>81.9</td>
</tr>
<tr>
<td>GLDAS⁷ MOS annual mean</td>
<td>-</td>
<td>15.3</td>
<td>5.59</td>
<td>273</td>
</tr>
<tr>
<td>GLDAS⁷ NOAH annual mean</td>
<td>-</td>
<td>24.8</td>
<td>5.80</td>
<td>427</td>
</tr>
<tr>
<td>Gleeson et al.*</td>
<td>22.6 (0.35 young)</td>
<td>48.0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*current analysis, average permeability case only

2. Global data synthesis and calculation of the total groundwater volume

2.1. Porosity-depth models

For carbonate and siliciclastic rocks we fit Eq. 1 with global porosity-depth compilations of siliciclastic (30,122 values) and carbonate (10,481 values) petroleum reservoirs¹¹ (Figure 1a-b). Ehrenberg and Nadeau¹¹ report the percentiles P10, P50 and P90 of the distribution of porosity for 250 m depth intervals. We calculated P25-P75 uncertainty estimates from the reported P10-P90 range by assuming that porosity at each 250 m depth interval follows a Gaussian distribution and that the mean is represented by the P50 value. Taking into account that \( n \) at a given depth represents a distribution of values we fit Eq. 1 to obtain values of \( n_0 \) and \( \beta \) and their P25-P75 uncertainty range (Table S2).

The porosity of volcanic rock is highly variable both at the surface and with depth, since volcanic rock includes a wide variety of textures derived from different processes. Even the porosity of a single ash flow layer can be highly variable from 0.05 to 0.50 depending on the location within the ash layer¹². Based on a compilation of literature values, Gleeson et al.¹³ suggested a reasonable near-surface porosity for volcanic deposits is 0.09. Burns et al.¹⁴ suggested 0.025 is a reasonable porosity for continental flood basalts in the Columbia River basin, whereas Keller et al.¹⁵ measured porosities of ~0.10-0.30 for the basalts of Kilauea volcano in Hawaii. Laboratory experiments¹⁶ on fractured rock in hydrothermal systems indicate a porosity varying from 0.01 to 0.2. Tectonically deformed volcanic rocks in Patagonia¹⁷ have a measured porosity of 0.15 to 0.30. Since the porosity is highly variable both at the surface and at
depth and no discernible depth-porosity trends are evident in the literature, we applied a depth invariant and uncertain porosity of 0.09 ± 0.09 for volcanic rocks (Table S2).

The porosity of crystalline rock is generally low and controlled by the density and aperture of the secondary fracture network, rather than the primary porosity, which can be assumed to be negligible (on the order of 10^3 to 10^6 according to Almén et al. 18). Based on a compilation of literature values, Gleeson et al. 13 suggested a reasonable near surface porosity for crystalline rocks is 0.01. Data from a variety of geologic settings 18-20 indicate secondary porosity in crystalline rock rarely exceeds 0.01. Since the porosity of crystalline rock is variable but very low compared to other lithologies, estimates of total groundwater volume do not depend significantly on the trend of crystalline porosity with depth. We therefore used a depth invariant and uncertain porosity of 0.01 ± 0.01 for crystalline rocks (Table S2).

### Table S2. Porosity and compressibility of the considered rock types.

<table>
<thead>
<tr>
<th>Lithology</th>
<th>Porosity at surface n₀</th>
<th>Porosity-depth relationship</th>
<th>Compressibility β (m⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbonate sediments</td>
<td>0.19 (0.12-0.24)</td>
<td>exponential function</td>
<td>3.3×10⁻⁴ (2.0×10⁻⁴-2.7×10⁻⁴)</td>
</tr>
<tr>
<td>Siliciclastic sediments</td>
<td>0.23 (0.18-0.28)</td>
<td>exponential function</td>
<td>9.8×10⁻⁴ (8.2×10⁻⁶-8.5×10⁻⁵)</td>
</tr>
<tr>
<td>Volcanic</td>
<td>0.09 (0 – 0.18)</td>
<td>constant</td>
<td>n/a</td>
</tr>
<tr>
<td>Crystalline</td>
<td>0.01 (0 – 0.02)</td>
<td>constant</td>
<td>n/a</td>
</tr>
</tbody>
</table>

### Table S3. The volumetric fraction of each lithology and the equivalent groundwater height of pore water stored in each lithology.

<table>
<thead>
<tr>
<th>Lithology</th>
<th>Volumetric fraction</th>
<th>Groundwater equivalent (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbonate sediments</td>
<td>0.10</td>
<td>29 (21-38)</td>
</tr>
<tr>
<td>Siliciclastic sediments</td>
<td>0.30</td>
<td>131 (105-158)</td>
</tr>
<tr>
<td>Volcanic</td>
<td>0.04</td>
<td>8 (0-15)</td>
</tr>
<tr>
<td>Crystalline rocks</td>
<td>0.56</td>
<td>11 (0-22)</td>
</tr>
<tr>
<td>Global sum</td>
<td>1.0</td>
<td>179 (125-233)</td>
</tr>
</tbody>
</table>

Note: Groundwater equivalent is weighted by volumetric fraction of each lithology in the upper 2 km of the crust.

2.2. Calculation of the total groundwater volume using porosity-depth models

The upper crust is volumetrically dominated by sedimentary and crystalline rock, with a lesser volume of volcanics 21. For each of the four lithologies, the area of each lithology was calculated from Hartmann and Moosdorf 22, assuming the area presently overlain by ice and water has the same lithologic distribution as the area not overlain by ice and water. The distribution of rock types with depth was based on global data on the thickness of sedimentary basins 23. If the thicknesses of sedimentary basins were less than 2 km, the depths below the sedimentary basin were assigned the non-sedimentary porosity (crystalline or volcanic porosity depending on the fraction of each lithology at the surface). The fraction of the total sediment
volume that consists of carbonates or siliciclastic sediments was assumed to be equal to the fractional
distribution at the surface\cite{22}. Volcanic rocks were assumed to form 9.7% of the total fraction of sediments
in the upper 2.0 km of the crust, following estimates of the surface coverage of volcanic rocks by
Hartmann and Moosdorf\cite{22}. Volcanic rocks were included in the sediment category, because these have
predominantly mapped as sediments in the Crust2.0 dataset\cite{23}. Results of the compilation and analysis are
presented in Table S3 and Figures 1a-b and S3.

For each lithology, a groundwater equivalent (the height of water if removed from the ground and stored
at the land surface) was calculated by integrating the lithology-specific porosity decay and weighing each
lithology by their volumetric fraction in the upper 2 km (Table S3).

The global volume of groundwater is the product of the global sum of the groundwater equivalent and the
considered land area. The global volume of groundwater is 22.6 (15.8 – 29.5) million km\(^3\) assuming a
total land area in our analysis of 126.3 million km\(^2\). Using the total land area of the previous analysis\cite{2}
(135 million km\(^2\)), the global volume of groundwater is 24.2 million km\(^3\) which is consistent (~3% more)
with the previous estimate of 23.4 million km\(^3\).

![Figure S3. Porosity-depth data of (a) 30,122 siliciclastic and (b) 10,481 carbonate reservoirs\cite{3} and best-fit exponential porosity-depth curves; (c) global distribution of lithology with depth and (d) global average porosity.](image)
2.3. Additional information on the global tritium ($^3$H) data synthesis and calculation of modern groundwater

A global data base of $^3$H concentrations in groundwater was compiled for 3,769 globally-distributed samples from 160 publications. The $^3$H data were unevenly distributed within 55 countries spanning Africa (count = 499), Asia (count = 1340), Europe (count = 617), the Malay Archipelago and Oceania (count = 76), North America (count = 1096) and South America (count = 141). Samples have been collected from confined (12% of dataset), partially confined (23% of dataset) and unconfined aquifers (65% of dataset). Since $^3$H sample is the measured quantity, implementing the mixing model requires calculation of $^3$H modern and $^3$H old. However, both values have varied through time – from “infinitely” long ago to 50 years prior to sampling for $^3$H old, and the 50 years preceding the time of sampling for $^3$H modern. For $^3$H modern, $^3$H in precipitation was determined at an annual time step for individual sample locations using a global model of tritium in precipitation \cite{Doney2005} developed for 1960-2005 (Figure S4). For $^3$H old, $^3$H in precipitation prior to 1960 was set to a range of 1 to 10 T.U. (before accounting for radioactive decay) as supported by pre-bomb $^3$H records from ice cores, wine and lakes \cite{Zhang2011} (Figure S5). And finally, the radioactive decay of all $^3$H pools was accounted for prior to calculation with the mixing model. These steps are elaborated further below.

First, we calculated the spatio-temporal maps of meteoric tritium for the years 1965 to 2005 using a published model for $^3$H in precipitation \cite{Zhang2011}. $^3$H has been monitored at globally-distributed locations for more than 60 years (e.g., \textit{Begemann and Libby} \cite{Begemann1964}. \textit{Zhang et al.} \cite{Zhang2011} have developed an empirically-based model for 1960 to 2005, building upon work of \textit{Doney et al.} \cite{Doney2005}). The model, named the Modified Global Model for Tritium in Precipitation (MGMTP) \cite{Zhang2011}, uses factor analysis based upon long-term monitoring data for $^3$H in precipitation to develop a globally-distributed estimate of annual meteoric $^3$H activities for each year from 1960 to 2005. We extracted annual estimates of meteoric $^3$H for the years spanning 1960 to 2005 from the MGMTP for a 5°×5° grid of the land surfaces (Figure S4). These annual $^3$H values were used to estimate the activity of $^3$H at each groundwater well in our compiled database (3,769 samples) by matching each well location to the nearest 5°×5° grid location in the MGMTP derived map. That is, for each sample $^3$H modern is a time-variant input function of the $^3$H in precipitation within 50 years of the time of sample collection.

Second, $^3$H old which is $^3$H in precipitation falling prior to the onset of the MGMTP model (in year 1960) was assumed to have an initial $^3$H concentration that varied between 1 to 10 T.U. (where 1 T.U. is equivalent to 1 $^3$H atom in $10^{18}$ hydrogen (protium and deuterium) atoms, equaling 3.24 p Ci L$^{-1}$, or 0.12 Bq L$^{-1}$) as supported by wine samples from temperate latitudes prior to the onset of widespread atmospheric thermonuclear testing (Figure S5).

Third, we accounted for the radioactive decay of $^3$H from the approximate time of precipitation to the time that each sample was collected. The half-life ($T_{1/2}$) of $^3$H is 12.32±0.02 years \cite{Doney2005}. We calculated the decay-corrected $^3$H activity of precipitation that fell any year within the 50 years prior to the collection of any given water sample within our database using:

$$A = A_0 e^{(-0.693t)/(T_{1/2})}$$

(S1)

where $A$ represents the radioactive-decay-corrected $^3$H activity, $A_0$ represents the annual $^3$H-activity as measured when the precipitation first fell to the land surface \cite{Zhang2011}, $t$ is the time elapsed from the time that the precipitation fell to the sampling date, and $T_{1/2}$ represents the half-life of $^3$H (12.32 years \cite{Doney2005}).

Finally, we used the decay-corrected \cite{Doney2005} and model-based \cite{Zhang2011} estimates of meteoric $^3$H concentrations in the mixing model (Eq. 2). Thus, the mixing model assigns a range of decay-corrected meteoric $^3$H activities
spanning the 50 years prior to the collection of each sample to envelope possible $^3$H activities for recharge within 50 years of sampling ($^3$H$_\text{young}$). The $^3$H activity of precipitation falling prior to 50 years of sampling ($^3$H$_\text{old}$) are near to zero (i.e., less than 1 T. U.) in nearly all cases. Each sample therefore has a range of values for $R_{\text{modern},3H}$ since the calculation accounts for the spread in the input parameters in the mixing model. The ensemble of $R_{\text{modern},3H}$ calculations with depth comprises Figure 2a.

We acknowledge that hydrodynamic dispersion and heterogeneous groundwater flow path velocities likely allow for groundwater mixing with different age distributions. Our analysis includes this mixture of all groundwater because we split the total groundwater age distribution using a threshold age (i.e., older than 50 years and younger than 50 years).

Figure S4. Tritium concentration in precipitation for 1960-2005 calculated for a 5° by 5° grid using the modified global model for tritium in precipitation$^6$. 


Figure S5. Tritium variations in precipitation (Ottawa, Canada shown) and ice cores\textsuperscript{9} (squares), large lakes\textsuperscript{7} (circles) and wine\textsuperscript{8} (diamonds). \textsuperscript{3}H concentrations in meteoric waters span three orders of magnitude due to atmospheric thermonuclear tests completed prior to the signing of the Partial Nuclear Test Ban Treaty in 1963.

3. Additional information on the numerical modeling of groundwater age transport

We did not simulate the distribution of solutes in our modeling framework, as the goal of this work is oriented towards groundwater age and not groundwater salinity. In many groundwater systems, solute concentrations increase with groundwater residence times through weathering reactions\textsuperscript{30} and can lead to changes in water density that affect groundwater flow\textsuperscript{31}. These significant changes in solute concentration primarily occur over very large (i.e., intermediate to regional scale) groundwater flowpaths that are not likely to have young (< 100 year residence time) groundwater, depending on the chemical processes and associated reaction rates\textsuperscript{32,33}. However, some groundwater systems have shallow saline groundwater resulting from a variety of hydrologic and geologic processes\textsuperscript{34-38}.

Since modeling the groundwater flow and age transport field for all 933,639 watersheds is prohibitive even with 2D models, we reduced the number of models by binning watershed properties. The frequency distributions of water table gradients and watershed half-widths guided the discretization. The roughly log-normally distributed water table gradients were divided into 33 bins based on the logarithms of water table gradients increasing in width away from the median value (Figure S6). The watershed half-widths, which exhibited a long-tailed normal distribution, were also split into 33 bins (Figure S6). To better represent the larger (wider) watersheds, 16 evenly spaced bins for every 5 km were added to the original 33 bins starting with 10 km up to 100 km.
Using this binning procedure, the groundwater age distributions for Earth’s watersheds could have been modeled using 14,553 combinations of the hydrologic input parameters, but we ran an additional 29,106 models changing the original combinations by one order of magnitude in $k_0$ to provide an estimate of the uncertainty in each modeled groundwater age distribution. Uncertainty of any of the datasets used in the numerical modeling could alter the resulting groundwater age distribution for a watershed. All of the hydrologic inputs contributed to the modeled groundwater age distributions sometimes with complex interactions between parameters. Initial results suggest permeability and hydraulic gradient were the most important controls on age distributions. Hydraulic gradient is derived from the water table data which is more easily measured and less uncertain than permeability. Therefore, permeability was considered the most important parameter controlling uncertainty. Additional, detailed analysis of how each hydrologic parameter contributes to the model is the topic of ongoing research that we hope to address in future work.

We used an unstructured finite-element mesh with triangular elements to solve Eq. (5) and Eq. (6). An adaptive meshing algorithm automatically refined the mesh when needed to more accurately solve and stabilize the age transport problem. When refinement was necessary, the mesh was refined for 60% of the elements with the highest solution errors using the a posteriori estimate of the $L_2$ norm\textsuperscript{39}. If the models did not converge on a solution after two such refinements, a universal mesh refinement was applied and then the model was re-run. This process was repeated up to three times until all of the model runs successfully converged. The number of triangular mesh elements ranged from 3,000-656,000 with an average of 90,000 elements. Due to the automated nature of running thousands of models, a detailed analysis of mesh convergence was not possible, but the combined benefits of the adaptive mesh and universal mesh refinement led to consistent results when compared to finer meshes during the early model development and vetting. The model runs took a total several months of wall-clock computer time with an average model solution time of 3 minutes with some taking several hours to solve. We then analyzed the results following Eq. 9 and Eq. 10 (Figure S7).
4. Modern groundwater storage for 30 specific aquifers

Tritium concentrations are only available for a minority of aquifers globally. Therefore, we developed two approaches for testing the agreement between the simulation and \(^3\text{H}\)-derived estimates of modern groundwater storage for 30 aquifers with the most tritium samples. The first method compared the volume of modern groundwater storage between the two methods for estimating modern groundwater volumes. \(^3\text{H}\) estimates of young groundwater storage volumes were calculated using well samples from a particular aquifer with aquifer-average porosity values. To calculate the modeling estimate of the modern groundwater storage volume for an aquifer, all of the watersheds within the areal extent of that aquifer were summed, using the well sample locations to define the aquifer extent. In the second method, the \(d_{\text{effective,}3\text{H}}\) calculated from integrating \(R_{\text{modern,}3\text{H}}\) was compared with the \(d_{\text{effective}}\) (using \(T=50\) years) based on the watershed-scale modeling.

The first step for these comparisons was to determine the hydrologic properties of the aquifers. To do this, each well used for \(^3\text{H}\) samples was assigned hydrologic data from the global spatial data synthesis using its location. Then, these hydrologic data were used to assign outputs from the numerical simulation. The same binning procedure was used to assign the hydrologic inputs to the simulation parameter space. Additional models were run with the permeability at the surface for the \(^3\text{H}\) data changed by one order of
Figure S7. Examples of groundwater age distributions within the original flow domain (a-c), the same age distributions referenced from the surface of the domain (d-f), the resulting depth-specific age probability distribution for each model (g-i), and the calculated $R_{young}$ for each flow system (j-l).

These example age distributions are for 10 km long groundwater flow domains with a porosity, $n_0=0.2$, $\beta \alpha =0.001$ m$^{-1}$. Continuous zero values in the bottom row figures are removed for plotting purposes. Each row of data in g-i sum to unity and the vertical line marks $\tau = 50$ yr. White lines in a-c are groundwater streamlines that delineate streamtubes with 10% of the flow.
magnitude from the base value. Only the young groundwater metrics for a timescale of 50 years were used for this analysis.

Once the model results were assigned to the aquifer-related watersheds, the two comparison methodologies diverged. For the volume-based comparison, the results from all watersheds within 10 km of the well locations for a given aquifer were collected. Young groundwater volumes for each watershed were summed to give the aquifer-wide estimate of young storage ($V_{storage}$), and the areas of these watersheds were also summed. The most common lithology (i.e., for porosity and permeability decay values) in these watersheds was then used with $R_{modern,3H}$ in Eq. 11 to calculate the $d_{equivalent,3H}$ over each aquifer. Next, $d_{equivalent,3H}$ was multiplied by the total area of the enveloped watersheds to give an aquifer-wide $^3$H estimate of the volume of young groundwater storage ($V_{storage,3H}$). This comparison of aquifer young groundwater volumes provided a consistent spatial framework for comparing the simulation and $^3$H results. However, the comparison relies upon an unbiased sampling of the aquifer (i.e., well locations are evenly distributed throughout the aquifer) and the accuracy of enforcing lithologic homogeneity in the calculation of $d_{equivalent,3H}$.

In the comparison of $d_{effective}$ values, the $R_{young,3H}$ was integrated in depth following Eq. 9 with the well data for each aquifer, yielding a single value for each aquifer (Figure S8). However, the simulation-based analysis operated at the watershed-scale, whereby each well sample had a corresponding watershed in which it was located. Thus, for this analysis, the $d_{effective}$ was used as the terminal indicator of young groundwater storage so as not to require additional assumptions of hydrogeologic properties and distributions across the aquifer. Since different scales were considered (i.e., aquifer vs. watershed), the flow simulation-based estimates should be lower to much lower than the $^3$H estimates.

The results of the two geomatic comparisons between the modeling and tritium-based approaches for estimating young groundwater storage indicated that approximately two-thirds of the aquifers have matching storage volumes or $d_{effective}$ values within the uncertainty of the two methods (Figure S8-10, Table S4).

The comparison of young groundwater storage volumes for the selected aquifers showed that the tritium estimates were always larger than the simulation estimates for all but six aquifers, and the model-derived volumes were on average one order of magnitude less than the tritium estimate (Figure S9a, Table S4).

Second, we used the ratio of $d_{effective}$/ $d_{effective,3H}$ as a metric of the agreement between the geochemical and simulated young groundwater estimates. Considering only the median $d_{effective}$ from the models with the median $d_{effective,3H}$ resulted in mainly one order of magnitude difference between the methods (Table S4, black dots in Figure S9b), but the disparity reached four orders of magnitude for some aquifers, such as the Najd Aquifer. Median simulation-derived $d_{effective}$ results primarily under-predicted the $d_{effective,3H}$. Uncertainty in the $d_{effective,3H}$ arises from the uncertainty in $R_{modern,3H}$, whereas the uncertainty in the $d_{effective}$ from the simulations arises from both the heterogeneity in the hydrologic datasets sampled by the distribution of wells in each aquifer and the uncertainty within the hydrologic datasets themselves. The colored uncertainty range in the $d_{effective}$ ratio in Figure S8b considered only the 25th–75th percentiles of $d_{effective,3H}$ with published permeability values to guide the models and taking the median of the simulated $d_{effective}$. Considering the uncertainty of $d_{effective,3H}$ and model $d_{effective}$ together defined the range of the grey boxes in Figure S9b. Combining all of the geochemical and modeling uncertainties led to favorable (i.e., overlapping) comparisons of $d_{effective}$ and $d_{effective,3H}$ for a majority of the aquifers tested (Figure S10).
Aquifers that did not share any overlap between the geochemical and model analyses did not follow any consistent geographic, spatial, or hydrologic trends.

In addition to the two geomatic assessments of the similarity between the numerical model and $^3$H estimates of modern groundwater, we also performed the $V_{storage}$ analysis using the $d_{equivalent}$ values assigned using recharge along with either the water table gradient or the porosity (Figure S9c-d). Since this method of assigning models to watersheds did not include alternative parameter combinations (i.e., unlike the geomatic assignment that considered models with $k$ changed an order of magnitude), the only uncertainty that was considered in comparing $V_{storage}$ arose from $^3$H. Again, the numerical model results predicted less storage of modern groundwater for the majority of the aquifers.

The modeling and tritium analyses provided separate methods for estimating young groundwater storage, but both approaches incorporate different simplifying assumptions that makes calibration of one with the other not always meaningful. Firstly, the modeled $d_{effective}$ and $d_{equivalent}$ values were based on a depth-integrated probability of the occurrence of young groundwater across an entire flow system, but the $d_{effective,3H}$ and $d_{equivalent,3H}$ were based on the proportion of young groundwater in well samples that integrate multiple scales of flow and potentially significant heterogeneity. However, the models restricted these storage metrics to include only the shallowest topographically-driven flow system bounded by perennial streams. In regions with small surface watersheds, the models would incorrectly limit the extent of young groundwater to shallower flow systems that in reality could mix with larger scale, more regional flowpaths. The models also only considered a single lithology per domain, whereas the $^3$H samples integrate the complexity of their hydrologic and hydrogeologic histories. Secondly, the simulations modeled the age of the groundwater, not the transport of $^3$H, whereby specifying a $^3$H molecular diffusivity value would be required. Additionally, the processes that led to the measured $^3$H concentrations occurred in three-dimensional space and through time, whereas the models were steady-state and two-dimensional. Within the $^3$H analysis, uncertainty arose predominantly from the estimated $^3$H concentration in groundwater recharge and the stochastic groundwater mixing model, which did not include groundwater dynamics. Thus, the purpose of comparing the model and $^3$H results was to understand how the uncertainty in each analysis biased the estimates of young groundwater storage for each aquifer.

Additional uncertainty in the comparisons arose from the uncertainty in the model input parameters. A number of assumptions are inherent in the calculation and use of the watershed half-widths (e.g., what is the characteristic length scale for a watershed and how that translates to groundwater flow) and hydraulic gradients (e.g., soil permeability and only horizontal groundwater flow were used in the global water table depth dataset, while vertical flow and near-surface permeability were used in our analysis) in the numerical models. However, quantifying the uncertainty in these parameters is not straightforward and could require the monumental task of reproducing hundreds of previous studies. Instead, the uncertainty in the models is addressed using the near-surface permeability, which has constrained uncertainty ranges$^{13}$. 


Figure S8. Aquifer profiles of $R_{\text{modern},3H}$ and resulting estimates in the effective depth to modern groundwater calculated from the tritium profiles ($d_{\text{effective},3H}$, horizontal dashed line and grey area). Table S5 reports the median values (dashed line) of $d_{\text{effective},3H}$ for each aquifer.
Figure S9. Comparison of modeled and $^3$H-based estimates of a) the geomatic results for the volume of modern (50 yr) storage ($V_{\text{storage}}$), b) the geomatic results for the effective depth of modern groundwater ($d_{\text{effective}}$), and the storage of modern groundwater using recharge to assign groundwater models with matching c) water table gradients and d) porosity for 30 aquifers with the most $^3$H samples. In a) and b) the colored bar shows the uncertainty range only considering the $^3$H analysis, and the grey bars show the combined uncertainty from the $^3$H calculation and permeability in the numerical simulations. Black dots in a) show the ratio of storage volumes calculated from the average permeability models and the median $d_{\text{equivalent,}^3\text{H}}$ and in b) show the ratio of the median $d_{\text{effective}}$ from the models and the median $d_{\text{effective,}^3\text{H}}$ for each aquifer. In c) and d), the colored range reflects the uncertainty in the $^3$H estimate, and the dot shows the median value.
Table S4. Aquifers used for the comparison of $V_{\text{storage,}3H}$ and $d_{\text{effective,}3H}$ derived from $3H$ samples with $V_{\text{storage}}$ and $d_{\text{effective}}$ calculated from numerical models using hydrologic parameters extracted to the well locations from the spatial datasets. The model results in this table do not include the uncertainty introduced by permeability nor from the $3H$ data, whereas Figure S9 does.

<table>
<thead>
<tr>
<th>Continent</th>
<th>Aquifer</th>
<th>Wells</th>
<th>$V_{\text{storage,}3H}$ [km$^3$]</th>
<th>$V_{\text{storage}}$ [km$^3$]</th>
<th>$V_{\text{storage}}$ ratio</th>
<th>$d_{\text{effective,}3H}$ [m]</th>
<th>$d_{\text{effective}}$ [m]</th>
<th>$d_{\text{effective}}$ ratio</th>
<th>Log10(Models/$3H$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>Sidi Bouzid plain</td>
<td>48</td>
<td>2.5 0.12</td>
<td>-1.32</td>
<td>56.7</td>
<td>14.7</td>
<td>-0.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Karoo</td>
<td>20</td>
<td>18.4 0.10</td>
<td>-2.25</td>
<td>28.1</td>
<td>1.5</td>
<td>-1.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chad basin</td>
<td>28</td>
<td>1212.9 3.96</td>
<td>-2.49</td>
<td>376.2</td>
<td>1.7</td>
<td>-2.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tadla basin</td>
<td>74</td>
<td>115.7 121.46</td>
<td>0.02</td>
<td>135.2</td>
<td>15.6</td>
<td>-0.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kalahari</td>
<td>48</td>
<td>4242.4 20.74</td>
<td>-2.31</td>
<td>101.9</td>
<td>1.5</td>
<td>-1.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>Canterbury Plains</td>
<td>29</td>
<td>3.5 0.09</td>
<td>-1.60</td>
<td>43.4</td>
<td>23.7</td>
<td>-0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Perth basin</td>
<td>34</td>
<td>3.6 0.01</td>
<td>-2.69</td>
<td>28.5</td>
<td>0.2</td>
<td>-2.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Krishna-Godavari basin</td>
<td>60</td>
<td>31.5 0.59</td>
<td>-1.72</td>
<td>32.2</td>
<td>1.8</td>
<td>-1.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Taoyuan Chungli</td>
<td>8</td>
<td>4.7 0.02</td>
<td>-2.47</td>
<td>41.7</td>
<td>0.3</td>
<td>-2.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tableland</td>
<td>95</td>
<td>2089.2 52.72</td>
<td>-1.60</td>
<td>77.7</td>
<td>4.7</td>
<td>-1.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paleogene limestone</td>
<td>37</td>
<td>23.2 0.84</td>
<td>-1.44</td>
<td>32.8</td>
<td>6.1</td>
<td>-0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Batinah coast</td>
<td>19</td>
<td>0.1 0.51</td>
<td>0.61</td>
<td>1.7</td>
<td>12.2</td>
<td>0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Najd</td>
<td>48</td>
<td>457.6 0.97</td>
<td>-2.67</td>
<td>155.3</td>
<td>0.0</td>
<td>-4.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>North China plain</td>
<td>151</td>
<td>2018.0 10.56</td>
<td>-2.28</td>
<td>268.6</td>
<td>1.8</td>
<td>-2.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Yinchuan plain</td>
<td>37</td>
<td>486.7 12.34</td>
<td>-1.60</td>
<td>236.5</td>
<td>10.0</td>
<td>-1.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bengal basin</td>
<td>109</td>
<td>466.5 5.83</td>
<td>-1.90</td>
<td>42.6</td>
<td>2.6</td>
<td>-1.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>Pisa coast</td>
<td>59</td>
<td>0.6 1.19</td>
<td>0.30</td>
<td>17.6</td>
<td>0.1</td>
<td>-2.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Paris basin</td>
<td>31</td>
<td>1.0 5.67</td>
<td>0.73</td>
<td>47.4</td>
<td>442.6</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>S. Turkey coast</td>
<td>24</td>
<td>209.5 84.40</td>
<td>-0.39</td>
<td>76.1</td>
<td>27.7</td>
<td>-0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molasse basin</td>
<td>63</td>
<td>5.0 0.18</td>
<td>-1.43</td>
<td>39.0</td>
<td>7.9</td>
<td>-0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aquitaine basin</td>
<td>49</td>
<td>183.9 47.81</td>
<td>-0.58</td>
<td>185.5</td>
<td>132.4</td>
<td>-0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Americas</td>
<td>Santiago basin</td>
<td>13</td>
<td>0.9 0.08</td>
<td>-1.02</td>
<td>86.3</td>
<td>0.2</td>
<td>-2.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central Oklahoma</td>
<td>59</td>
<td>43.8 0.43</td>
<td>-2.01</td>
<td>69.3</td>
<td>0.9</td>
<td>-1.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>E. Snake River plain</td>
<td>39</td>
<td>146.7 379.37</td>
<td>0.41</td>
<td>122.0</td>
<td>361.8</td>
<td>0.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Villa de Reyes basin</td>
<td>14</td>
<td>15.8 7.50</td>
<td>-0.32</td>
<td>73.7</td>
<td>151.9</td>
<td>0.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pleistocene till plain</td>
<td>76</td>
<td>3.6 24.72</td>
<td>0.84</td>
<td>22.0</td>
<td>0.1</td>
<td>-2.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Miss. Embayment:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Memphis</td>
<td>40</td>
<td>0.6 0.12</td>
<td>-0.71</td>
<td>15.9</td>
<td>12.0</td>
<td>-0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Middle Rio Grande</td>
<td>168</td>
<td>185.8 91.99</td>
<td>-0.31</td>
<td>99.3</td>
<td>356.6</td>
<td>0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central High Plains</td>
<td>24</td>
<td>353.1 15.91</td>
<td>-1.35</td>
<td>57.3</td>
<td>7.6</td>
<td>-0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>California basin</td>
<td>60</td>
<td>239.2 2.67</td>
<td>-1.95</td>
<td>123.6</td>
<td>3.6</td>
<td>-1.53</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wells with $3H$ samples for comparison = 1564
Figure S10. Comparison of the effective depth to modern groundwater calculated from the tritium profiles (horizontal dashed line and grey area) and numerical simulations (box and whisker) with all quantified sources of uncertainty for each data-rich aquifer. Figure S9b is a summary of these data.
5. Assignment of groundwater age transport model results to global watershed distribution

All model results, including $d_{\text{effective}}$ and $d_{\text{equivalent}}$, were assigned to the 933,639 HydroSHEDS watersheds by mapping each watershed to the binned distributions of input hydrologic parameters driving the 43,659 modeled generic groundwater systems. Additional model results with one order of magnitude change in the $k_0$ above and below the published data were also assigned to each watershed to account for the uncertainty in the permeability data. Distributions of young groundwater storage were calculated for 25, 50, 75, and 100 years (Figure S11). Watersheds located where water table depths were >100 m were removed from the calculation of global young groundwater storage due to limitations in the original analysis. To calculate the global volume of young groundwater, we summed the volume of young groundwater from each watershed, calculated by multiplying the watershed area by $d_{\text{equivalent}}$.

![Figure S11. Global young groundwater equivalent, $d_{\text{equivalent}}$, maps for the geomatic model assignment.](Image)
6. Analysis of groundwater recharge in the numerical simulations

Groundwater recharge is the mechanism for the replenishment of groundwater resources and is critical for understanding the hydrodynamics in hydrogeologic systems. In our simulations, groundwater recharge was assigned to the domain by prescribing a Dirichlet head condition at the upper boundary that distributes the recharge and discharge fluxes as a function of the hydrologic properties in the modeling domain. Thus, recharge (and discharge) fluxes can be calculated as an output from our models, but was not an input parameter. However, the assigned head gradients are calculated from water table depths\(^7\) that were modeled using globally-distributed, average modern recharge estimates from Döll et al.\(^9\). Other global groundwater recharge estimates were calculated by Wada et al.\(^{10}\) and using land surface-climate models through the Global Land Data Assimilation System (GLDAS)\(^6\).

Few quantitative estimates of groundwater recharge exist at the continental scale that can be compared to the simulation results. Comparison between different recharge estimates on the shallow, unconfined water table position are relatively insensitive to significantly different recharge conditions\(^7\), making quantification of recharge based on shallow groundwater flow uncertain. Thus, the recent previous estimates represent the state-of-the-science but may not consider all of the hydrological processes nor fully incorporate hydrological heterogeneity commented on below.

Despite the potential shortcomings of these global estimates of groundwater recharge, they provide benchmarks for understanding biases in our numerical simulations. Our models transformed the lateral two-dimensional water table depth analysis into a simplified vertical cross-section domain that was meant to integrate both the hydrologic and hydrogeologic systems at the watershed scale. And whereas the previous recharge analyses are discretized based on a regular grid, our recharge analysis explained below used the HydroSHEDs watersheds as the spatial unit, which complicates how these estimates are compared with the current work. To test the validity of the simulations to quantify the young groundwater storage, the recharge calculated from the groundwater models should be close to previous estimates, of which Döll et al.\(^9\) was used as the input for the water table study\(^7\) that parameterized our models.

To calculate the watershed-integrated groundwater recharge that our models indirectly impose, a similar methodology to that used for the storage length-scales is adopted. Given that the simulations are steady-state, the groundwater recharge imposed by the prescribed head condition equals the groundwater discharge from the domain, which is required to conserve mass (Eq. 5). Thus, the up-gradient half of the domain is the recharge zone with a flux magnitude equal to the discharge flux but with the opposite sign. Total recharge and discharge fluxes \([\text{m}^2/\text{yr}]\) are calculated for each of the simulations by integrating the fluxes into and from the upper domain boundary. The recharge equivalent, \(r_{\text{equivalent}}\) \([\text{m/yr}]\), for a watershed is then calculated by dividing the total recharge flux by the watershed half-width \((L_{\text{half-width}})\):

\[
r_{\text{equivalent}} = \frac{\tau}{L_{\text{half-width}}} \int_{L_{\text{recharge}}} r(x) \, dx
\]

(S2)

with \(r(x)\) the recharge rate into the domain over the recharge zone, \(L_{\text{recharge}}\). Then, \(r_{\text{equivalent}}\) can be multiplied by a time-scale to give a length-scale of recharge that occurs over a given time period, \(\tau\). Using \(\tau = 1\) year, \(r_{\text{equivalent,1}}\) gives a watershed-wide integrated length-scale of recharge that can be multiplied by the watershed areas to give the volumetric rate of groundwater recharge in those watersheds. These
recharge rates can then be summed globally to compare with previous recharge estimates. Similarly, a
time-scale of 50 years can be chosen to compare the volume of groundwater storage younger than 50
years with the expected volume of recharge that would occur over 50 years. Since the storage calculation
accounts for groundwater of various ages mixing, the storage volume is expected to be less than the
recharged volume because older groundwater mixes with incoming younger groundwater, especially
where heterogeneity is significant40,41. Equivalent recharge volumes less than young groundwater storage
volumes could be expected in groundwater systems where very active young groundwater dispersively
mixes with older groundwater, yielding a mean age younger than the cut-off for the definition of
groundwater youth (e.g., 50 years).

Our initial estimate of the global recharge rate for all ice-free watersheds using the average permeability
values was $50.6 \times 10^3 \text{km}^3/\text{yr}$ ($5.21-523 \cdot 10^3 \text{km}^3/\text{yr}$ for permeability one order of magnitude below and
above the average values, respectively). Previous global groundwater recharge estimates are listed in
Table S1. Indeed, global estimates of precipitation available for runoff and potential groundwater
recharge based on measurements42 range from $36-40 \times 10^3 \text{km}^3/\text{yr}$ and others are listed in Table S1,
which set the upper limit of realistic potential groundwater recharge. Thus, we investigated the effect of
culling watersheds with extreme modeled recharge rates on the global groundwater recharge and the
modeled young groundwater storage. However, the uncertainty introduced by the permeability, which sets
the recharge rate, contained the range of previous recharge estimates.

The results of the culling are summarized in Table S5. First, we culled watersheds (n=48,365; area = 7.3
$\times 10^6 \text{km}^2$, 6.0% of land surface) where the water table from Fan et al.7 was outside the scope of their
analysis (i.e., > 100 m), reducing our global recharge estimate by $2.6 \times 10^3 \text{km}^3/\text{yr}$ (5.1%). In our analysis,
these areas would have hydraulic gradients set only by the topography since they were prescribed a
constant water table depth of 100 m. Additionally, these areas with deep water tables are primarily in arid
and semi-arid regions with significant topography that would likely have groundwater systems controlled
by recharge rates rather than topography43,44, where the latter is the assumption in our numerical
simulations. In addition, we culled all watersheds with annual groundwater recharge values that were
greater than mean annual precipitation using Climate Prediction Center Merged Analysis of Precipitation
(CMAP) data45, where 92,536 watersheds (10.0% of land surface) accounted for $42.1 \times 10^3 \text{km}^3/\text{yr}$
(83.2%) of our recharge estimate for the original permeability models. Thus, the culled estimate of the
global recharge rate as specified by the flow models was $5.9 \times 10^3 \text{km}^3/\text{yr}$ without areas where model
inputs were inconsistent with observational data or poorly constrained. This estimate may be within the
range of uncertainty in the previous recharge estimates, given the misfit between the recent models10 is
$2.5 \times 10^3 \text{km}^3/\text{yr}$. However, the true uncertainty in the model results is unknown, and large groundwater
fluxes (0.1-6.5 $\times 10^3 \text{km}^3/\text{yr}$ from submarine groundwater discharge46) and hydrological processes (e.g.,
groundwater recharge from surface waters) were not considered in these analyses that would require
additional recharge to balance the groundwater budget.

Watersheds with groundwater recharge equivalents greater than the CMAP precipitation rate were
primarily located in arid and mountainous regions (Figure S12). These areas of high modeled recharge
values highlight the potential disconnection and some inconsistency between surface and subsurface

provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at
http://www.esrl.noaa.gov/psd/
Our recharge analysis demonstrates the plausibility of magnitude changes of permeability contributing to uncertainty on the global scale (Figure S13). Part of the misfit of the global recharge rate for this study resulted from the assumptions built into the numerical models which restricted the analysis primarily to less arid regions of the world (i.e., primarily topography-limited flow). But, an order of magnitude change in permeability also resulted in an order of magnitude change in the global recharge rate. Thus, for the base permeability case, overestimating the recharge is a combined result of: (1) introducing too much recharge into arid and semi-arid regions, (2) discrepancies introduced in the dimensional simplification (complex 3D to simple 2D) and (3) homogenization of hydrologic properties within watersheds. Thus, the estimate of young groundwater storage prior to considering recharge may represent an upper limit of this storage, given that our average permeability case overestimates global groundwater recharge by a factor of at least two. Importantly, the alternative permeability cases (i.e., permeability changed an order of magnitude from the lithology-specified average) incorporate recharge rates that completely bound the expected global recharge rate from other studies (Table S5). Thus, while our estimates of young groundwater storage remain highly uncertain, the true storage of young groundwater on Earth is contained within the spread of these estimates.
Table S5. Summary of the final estimates for the modeled groundwater recharge and storage renewal rates using all ice-free watersheds, watersheds with a water table depth less than 100 m and watersheds with recharge equivalents less than the precipitation rate. Since recharge rates are calculated using the assigned permeability, different recharge rates were calculated for the same watersheds for the different permeability scenarios. The relative contributions and extent of the culls are shown in the subtractions section. Storage of 50 yr old groundwater (km³) divided by the land area and then divided by 50 years to give km³/yr.

<table>
<thead>
<tr>
<th>Watersheds</th>
<th>Area</th>
<th>Model recharge</th>
<th>Model recharge</th>
<th>Model recharge</th>
<th>Storage renewal</th>
<th>Storage renewal</th>
<th>Storage renewal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>10⁶ km³</td>
<td>%</td>
<td>10³ km³/yr</td>
<td>%</td>
<td>10³ km³/yr</td>
<td>%</td>
</tr>
<tr>
<td>All ice-free</td>
<td>933639</td>
<td>100</td>
<td>121.7</td>
<td>100</td>
<td>50.6</td>
<td>100</td>
<td>5.2</td>
</tr>
<tr>
<td>Water table &lt; 100 m</td>
<td>885275</td>
<td>94.8</td>
<td>114.4</td>
<td>94.0</td>
<td>48.0</td>
<td>94.9</td>
<td>5.0</td>
</tr>
<tr>
<td>Recharge equivalent &lt; precipitation (CMAP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base k</td>
<td>790982</td>
<td>84.7</td>
<td>102.3</td>
<td>84.0</td>
<td>5.9</td>
<td>11.7</td>
<td>- -</td>
</tr>
<tr>
<td>k · 10⁻¹</td>
<td>866577</td>
<td>92.8</td>
<td>112.3</td>
<td>92.2</td>
<td>- -</td>
<td>2.7</td>
<td>50.8</td>
</tr>
<tr>
<td>k · 10¹</td>
<td>633825</td>
<td>67.9</td>
<td>81.6</td>
<td>67.1</td>
<td>- -</td>
<td>- -</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Subtractions

<table>
<thead>
<tr>
<th>Watersheds</th>
<th>Area</th>
<th>Model recharge</th>
<th>Model recharge</th>
<th>Model recharge</th>
<th>Storage renewal</th>
<th>Storage renewal</th>
<th>Storage renewal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water table &gt; 100 m</td>
<td>48364</td>
<td>5.2</td>
<td>7.3</td>
<td>6.0</td>
<td>2.6</td>
<td>5.1</td>
<td>0.3</td>
</tr>
</tbody>
</table>

Recharge equivalent > precipitation (CMAP)

| permeability | 92540 | 9.9 | 12.2 | 10.0 | 42.1 | 83.2 | - - | - - | - - | 18.9 | 67.0 | - - | - - | - - | - - | - - |
| k · 10⁻¹     | 16877 | 1.8 | 2.2 | 1.8 | - - | 2.3 | 44.2 | - - | - - | - - | 2.0 | 38.4 | - - | - - | - - | - - | - - |
| k · 10¹      | 250776 | 26.9 | 32.8 | 26.9 | - - | - - | 485.7 | 92.9 | - - | - - | - - | 63.5 | 76.0 | - - | - - | - - | - - | - - |
Figure S12. Watershed locations that had a water table depth greater than 100 m (red) and a recharge equivalent greater than expected precipitation values for the average permeability case (blue). Increasing the permeability one order of magnitude increased the number of erroneous watersheds, whereas decreasing the permeability resulted fewer watersheds with groundwater recharge greater than precipitation.

Figure S13. Uncertainty estimates of the 50 year $d_{\text{equivalent}}$ using groundwater simulations with different near surface permeability ($k$) values. a) and b) the difference between the average permeability case of $d_{\text{equivalent}}$ and the $d_{\text{equivalent}}$ calculated with the permeability changed an order of magnitude below and above the average, respectively. c) and d) the normalized variation of the $d_{\text{equivalent}}$ results from the average permeability results to the other orders of magnitude.
7. Groundwater age fields of unsaturated flow systems

Our numerical models relied on the assumption that the groundwater systems were fully water saturated, using the gradient of the water table as the upper boundary condition. Thus, we used a linear water table with a constant slope derived from Fan et al.\textsuperscript{7} in our 2D models. However, if we modeled unsaturated groundwater flow using groundwater recharge as the upper boundary condition, the water table could become more parabolic\textsuperscript{44}, especially where groundwater discharges. The effect of this parabolic water table allows more of the 2D domain to be recharging groundwater, thus increasing the amount of young groundwater in the domain (Figure S14b-c). Therefore, we used the same method used to pair models with watersheds that had similar domain-wide recharge rates to compare the similarity of the age fields and $d_{equivalent}$ for saturated and variably flow models (Figure S14a).

The domain for the variably saturated flow model was constructed similarly to the saturated flow simulations, but the upper boundary was prescribed using a topographic gradient instead of using the water table. Variably saturated groundwater pressure ($p$) field was solved for in steady state using Richards equation\textsuperscript{50}:

$$\nabla \cdot \rho \left( - \frac{k_s}{\mu} k_r (\nabla p + \rho g \nabla z') \right) = R$$ \hspace{1cm} S3

where $\rho$ is fluid density, $k_s$ is the permeability, $k_r$ is the relative permeability and is a function of water content $\Theta$, $\mu$ is the fluid viscosity, $g$ is the gravitational constant, $z'$ is depth, and $R$ is a source term (kg m$^{-3}$ s$^{-1}$). We used the van Genuchten model\textsuperscript{51} for calculating $k_r$, $\Theta$, and the effective saturation, $Se$.

The liquid volume fraction is calculated with:

$$\Theta = \begin{cases} 
\Theta_r + Se(\Theta_s - \Theta_r) & \text{if } H_p < 0 \\
\Theta_s & \text{if } H_p \geq 0
\end{cases}$$ \hspace{1cm} S4

where $\Theta_r$ is the volumetric moisture content retained after draining and $\Theta_s$ is the moisture content when the subsurface is saturated. $H_p$ is the pressure head $p/(\rho g)$ and $Se$ is calculated using:

$$Se = \begin{cases} 
\frac{1}{[1 + (\alpha H_p)^n]^m} & \text{if } H_p < 0 \\
1 & \text{if } H_p \geq 0
\end{cases}$$ \hspace{1cm} S5

where $\alpha$ and $n$ are constants that depend on the subsurface material and $m = 1-1/n$. Finally, $k_r$ is calculated with:

$$k_r = \begin{cases} 
Se^l [1 - (1 - Se m)^m]^2 & \text{if } H_p < 0 \\
1 & \text{if } H_p \geq 0
\end{cases}$$ \hspace{1cm} S6

where $l$ is another constant that depends on the subsurface material. We chose $\alpha = 1$ m$^{-1}$, $n = 2$, and $l = 0.5$ for our variably saturated flow simulations.

We used a general mixed boundary condition (Cauchy) for the top of the domain that switched between a Neumann condition if $H_p < 0$ specifying groundwater recharge at the surface:
with \( r \) the recharge rate in L/T and a Robin boundary condition allowing the formation of a seepage face and groundwater discharge:

\[-n \cdot \rho u = -\rho r\]  \hspace{1cm} S7

\[-n \cdot \rho u = -\rho R_b H_p\]  \hspace{1cm} S8

with \( R_b \) the conductance of the interface and was set equal to the hydraulic conductivity \( k_g \mu^{-1} \). Since the location of where these boundary conditions should be applied is unknown a priori, the boundary conditions were free to move along the top of the domain during the solution of Eq S3 based on the calculated pressures. To stabilize this nonlinearity, the transition between the two types of boundary conditions was smoothed using a smoothed Heaviside step function centered on \( H_p = -0.5 \times 10^{-6} \) m with a transition zone width of \( 10^{-6} \) m. This mixed boundary condition required an extremely fine mesh along the top domain relative to the saturated deeper portions of the flow system, making domains with larger half-widths exponentially more computationally expensive to model. With the domain half-width of 1.5 km, 100,000-400,000 triangular mesh elements were used.

For the unsaturated models, we used a watershed half-width of 1500 m, a constant porosity of 0.19, and changed the permeability from \( 10^{-12} \) to \( 10^{-11} \), topographic gradient from \( 10^{-2.5} \) to \( 10^{-1.5} \), and recharge rate from 5-20 cm/yr. Larger domains, or domains with either lower permeability or lower topographic gradients, required prohibitively more mesh elements to accurately solve for the free surface. The resulting water table position was used to model saturated groundwater flow and groundwater age transport. Then, the groundwater age fields were used to calculate \( d_{\text{equivalent}} \).

Comparing saturated and unsaturated \( d_{\text{equivalent}} \) values for models with identical watershed half-widths and porosity, similar recharge rates and water table gradients, and unconstrained permeability values, the \( d_{\text{equivalent}} \) values for the unsaturated models are larger. However, these models represent a small subspace of the model parameters used for the global modeling analysis. As the recharge increases relative to the permeability or topographic gradient, the water table will approach the land surface and become increasingly similar to saturated groundwater models. Results suggest that unsaturated flow models may have up to \(~2\) times larger volume of modern groundwater, although it must be emphasized that these results from the small subset of this parameter space due to the computational costs of these simulations, and this relationship is the focus of ongoing research beyond the scope this current paper.
Figure S14. a) Comparison between $d_{equivalent}$ calculated using b)-c) unsaturated and d)-e) saturated groundwater flow models matched using the same procedure as the assignment of saturated groundwater models to watersheds based on recharge.
References


