

Global patterns in observed hydrologic processes

Received: 22 June 2024

Accepted: 11 February 2025

Published online: 31 March 2025



Hilary McMillan¹✉, Ryoko Araki¹, Lauren Bolotin¹, Dong-Hyun Kim¹, Gemma Coxon², Martyn Clark³ & Jan Seibert⁴

To manage water resources and forecast river flows, hydrologists seek to understand how water moves from precipitation, through watersheds, into river channels. However, we lack fundamental information on the spatial distribution and physical controls on global hydrologic processes. This information is needed to provide theoretical support for large-domain model simulations. Here, to address this issue, we present a global, searchable database of 400 research watersheds with published descriptions of dominant hydrologic flow pathways. This knowledge synthesis approach leverages decades of grant funding, fieldwork effort and local expertise. We use the database to test longstanding hypotheses about the roles of climate, biomes and landforms in controlling hydrologic processes. We show that aridity predicts the depth of water flow pathways and that terrain and biomes predict the prevalence of lateral flow pathways. These new data and search capabilities support efficient hypothesis testing to investigate emergent patterns that relate landscape organization to hydrologic function.

A fundamental challenge in hydrology is to explain and predict water movement through the terrestrial portion of the hydrologic cycle and into the network of streams and lakes¹. This knowledge is essential for applications ranging from water resource management to flood forecasting. Every watershed has a unique climatic and physiographic setting; therefore, understanding its hydrologic processes provides the opportunity for a step forward in hydrologic knowledge². This Analysis builds on such watershed-scale hydrologic knowledge, by synthesizing the behaviour of hundreds of catchments at the global scale, towards a generalization of hydrologic processes³. Synthesis of hydrologic process knowledge helps to explain where and when different processes occur, and how they are controlled by the climate and landscape^{4,5}. Answering this question has the potential to transform our ability to extrapolate process knowledge to data-scarce regions, to simplify predictive models by unifying perspectives from subdisciplines such as engineering hydrology and agricultural hydrology, and to improve water management based on knowledge of water storage and flow^{6,7}.

Geoscientists have recently proposed the concept of Earth's digital twin that moves towards a solution for comprehensive predictions

required to support decisions on water sustainability and reduce the impacts of water-related hazards. A digital twin is a dynamic simulation that aims to identically replicate the physical earth and its environments^{8–10}. It is a virtual representation of the physical system that combines big data, remote sensing and artificial intelligence with process understanding and numerical modelling, providing scientists and policymakers with perspectives on our changing environment from birds-eye overviews to deep dives¹¹. However, in hydrology, we lack the basic information needed to populate a digital twin, such as the dominant pathways and depths of water flow through watersheds, described here as hydrologic processes¹². This includes information to address fundamental questions such as whether most water moves overland, through shallow or deep soil layers, or as groundwater. Hydrologic theories on the global distribution and physical controls on processes are relatively rare; such syntheses oppose the view that emphasizes the uniqueness of each watershed¹³. In search of unifying hydrologic theories, Sivapalan⁶ proposes self-organization in landscape properties as the strongest clue to the drivers of hydrologic process, resulting in theories such

¹Department of Geography, San Diego State University, San Diego, CA, USA. ²School of Geographical Sciences, University of Bristol, Bristol, UK.

³Department of Civil Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada. ⁴Department of Geography, University of Zurich, Zurich, Switzerland. ✉e-mail: hmcmillan@sdsu.edu

as the control of topographic convergence on the dynamic spatial pattern of saturated areas.

Synthesis efforts across the field of hydrology have demonstrated the value of large-sample approaches towards scaling up our hydrologic understanding. So far, these efforts have focused on hydrological data, notably national to global databases of meteorology, streamflow, water chemistry, soil moisture and landscape attributes^{14–16}. In the realm of groundwater, studies have analysed permeability and porosity¹⁷ and depth to groundwater¹⁸ on continental to global scales. Alongside data-intensive research, there have been efforts to describe hydrographic features (for example, water bodies, river networks and observations) and their relationships, such as the Open Geospatial Consortium WaterML 2.0 standard for online water data^{19,20}. There have been only limited synthesis efforts that use field-based process knowledge across large samples of watersheds to explain large-scale patterns of hydrologic behaviour. Two recent papers synthesize data from multiple research watersheds to evaluate the drivers of hydrologic response^{21,22}. Several pioneering studies have analysed a single process on a global scale, merging many datasets to identify new patterns. These include groundwater depth data used to explain the global distribution of wetlands and hydrologic regulation of rooting depth^{23,24} as well as global stable isotope data used to show that ‘young’ water accounts for more than a third of global streamflow²⁵. However, no study has previously attempted to integrate global knowledge on watershed runoff generation processes.

Although there are hundreds of intensively studied watersheds around the globe where field measurements have revealed the respective dominant hydrologic processes²⁶, this knowledge is fragmented and challenging to search and integrate across regions and continents²⁷. Searches by location or watershed name in citation databases or search engines may return many articles without clear runoff process descriptions and searches by process type are complicated by ambiguity in process naming²⁸. Journal articles providing the clearest explanation of hydrologic process knowledge typically follow best practices in two steps. First, the authors use their expert knowledge of the watershed to translate field measurements of stores or fluxes into derived information about hydrologic processes. Second, the authors communicate the derived information to the reader in a consolidated and concise section that summarizes process knowledge. Such a summary is here called a hydrologist’s ‘perceptual model’ of the watershed functioning and can be expressed both as text and a schematic figure describing the main stores and fluxes that transport water within the watershed²⁹. These perceptual models are a valuable tool to share hydrologic process knowledge and identify knowledge gaps, test hypotheses to upscale process knowledge, and complement other forms of knowledge sharing such as datasets of climate, streamflow and physical features of watersheds^{30,31}. Aggregating process knowledge on a large scale is needed to satisfy new demands from the hydrologic modelling community, where continental-domain modelling frameworks have the potential to tailor the model structure to suit the dominant processes in individual watersheds^{32,33}.

This Analysis describes an effort to compile and synthesize process knowledge from perceptual models of research watersheds across the globe. Our synthesis enables us to test hypotheses from the literature that propose overarching drivers of patterns of hydrologic processes. We test the classic theory that aridity controls the balance of dominant flow processes between surface and subsurface flow^{34,35}, and a recent theory that lateral flow processes dominate in most global regions³⁶. Underlying our synthesis is a new open-source spatial database that brings together descriptions of hydrologic processes from intensively studied watersheds worldwide. The database is searchable using a hierarchical, taxonomic classification of process groups. The database enables advances in hydrologic theory to understand how climate and landscape control dominant processes, advances in hydrologic models that simulate spatially variable processes across different geographical domains, and advances in managing environmental change through tools such as digital twin environments.

A database of process knowledge

Our search has so far returned 400 published descriptions of hydrologic processes in watersheds with a median area of 6.9 km². Among these process descriptions, 131 are figures (including 63 identified in an earlier study of perceptual model figures³⁷) and 269 are text descriptions. Hydrologists began to publish process descriptions in the 1970s, and the number of such descriptions continues to grow, showing the potential for future expansion of synthesis efforts as process descriptions for an increasing number of locations are available each year (Fig. 1b). Figure 1a shows the locations of the watersheds, superimposed on global biome boundaries. Watershed locations include every continent, and all biomes, with the exception of mangroves and flooded grasslands (Fig. 1c). The spatial reach of these 400 process descriptions makes possible global analyses of watershed function, and the database provides a valuable public resource for the hydrologic community. The following sections demonstrate key hypothesis-testing research made possible by synthesizing this process information.

Clusters of watersheds occur in regions with a rich history of hydrologic process studies, that is, the USA, Europe, Brazil and Japan. Such clusters demonstrate the impact of scientific funding priorities on regional hydrologic knowledge and the benefits of formal or informal networks of research watersheds that encourage data sharing, collate watershed metadata and improve discoverability^{38,39}. The bias towards research watersheds in the Global North leads to a bias towards humid, temperate regions, similar to biases in the global river gauge network⁴⁰. Arid regions and grassland biomes are less well represented. These regions may not have dominant hydrologic processes that conform to the stereotypical hydrological cycle: in arid regions and grasslands, green (transpirable) water may be more important than blue (surface or ground) water, and relationships such as impacts of afforestation on streamflow may differ from temperate regions⁴¹. The deficit of research watersheds to study these regions has environmental justice implications⁴², as hydrological theories and tools in these regions have weaker linkages to process knowledge and, therefore, would be expected to produce poorer forecasts.

Search results enable us to visualize global patterns of processes recorded in research watersheds (Fig. 2). The figure includes widely reported processes (subsurface stormflow and overland flow) and less common processes (organic layer subsurface stormflow and non-perennial flow). In Fig. 2a–c, process occurrence aligns with our expectations. Subsurface stormflow is widely reported, with exceptions in locations such as the central USA, parts of South America and Asia, and Spain. In these more arid locations, conditions that generate subsurface stormflow may be less likely to occur, that is, highly weathered soils with restrictive layers, rising water tables or wetting front propagation⁴³. Organic layer subsurface stormflow is found in northern permafrost regions where flow is restricted to the thawed active layer⁴⁴, and as part of stormflow in some humid or tropical forests⁴⁵. Overland flow is widely reported, with exceptions in temperate or boreal forested areas of the Pacific Northwest, Appalachians, Northern Europe and Japan. Forest soils typically have high macroporosity, promoting rapid drainage of water to deep soils and reducing overland flow⁴⁶. In Fig. 2d, process occurrence does not align with our expectations, as more non-perennial flow processes were recorded in the wetter eastern USA than the drier western USA. Although non-perennial flow regimes are most commonly associated with dryland areas, our results support recent work reporting their occurrence in diverse climate and geographic settings⁴⁷.

Global synthesis of observed hydrologic processes

We use the database to analyse and reveal overarching patterns in runoff generation mechanisms. In our first analysis, we visualize differences in hydrologic processes over space (Fig. 3). The figure shows the dominance of process types in each biome, with the ‘biome triangle’

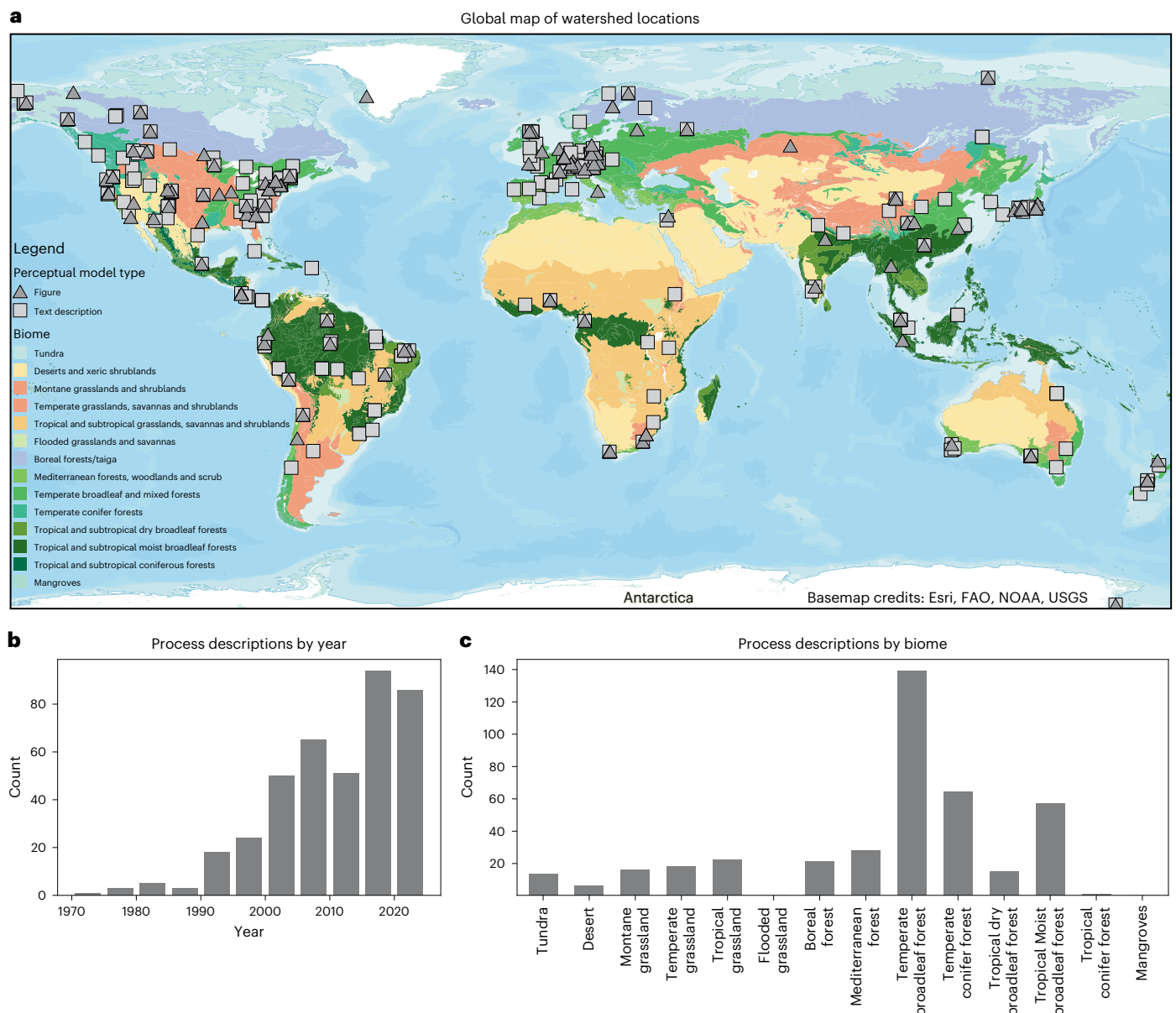


Fig. 1 | Research watershed locations, publication years and biomes.

a, Global map of research watershed locations with descriptions of hydrologic processes (symbols signify figure or text descriptions) superimposed on a global

classification⁷¹ of 14 biomes. FAO, Food and Agriculture Organization; USGS, United States Geological Survey. **b**, **c**, Counts of watershed process descriptions by publication year (**b**) and biome (**c**).

organized by temperature and precipitation⁴⁸. The size of each slice is proportional to the number of different processes regularly recorded in that category (processes that are recorded in at least one in eight watersheds in the biome). Some patterns shown in the figure follow our expectations, such as greater numbers of snow and ice processes recorded in cold regions, forests dominated by subsurface processes (as previously discussed) and fewer subsurface stormflow processes in hot and dry climates where shallow water tables are rarer. Other patterns are unexpected, but may be explained by examining the studies included in the database. For example, the surprising numbers of snow and ice processes recorded in deserts are due to research watersheds in the American West that are in the desert biome yet receive winter snowfall. In another unexpected result, the tropical grasslands biome has the highest proportion of subsurface and groundwater processes. We find that many studies from this biome describe soils with a high infiltration that promotes recharge to deep groundwater and minimizes infiltration excess flow, thereby differentiating this biome from

desert environments where biological soil crusts reduce infiltration⁴⁹. This finding might also reflect the emphases of studies in developing regions, which sometimes discuss broader hydrogeological features related to agricultural water use. Overall, the overview of spatial variability in processes shown in Fig. 3 provides guidance for global hydrologic and Earth system modelling efforts to ensure that models are capable of representing the common processes in each biome.

We build on this first analysis by investigating factors that control the difference in processes over space. We focus on testing two hydrologic theories that predict global patterns in which processes occur. The first are Dunne's^{34,35} hypotheses on the effect of aridity, slope and soil depth on the type and depth of dominant flow processes. Dunne predicts that, in arid-to-subhumid climates with thin vegetation, most runoff is derived from infiltration excess overland flow, with subsurface stormflow being less important. In humid climates with dense vegetation, under conditions of thinner soils and/or gentler, wider valleys, most runoff is derived from saturation excess (that is, precipitation

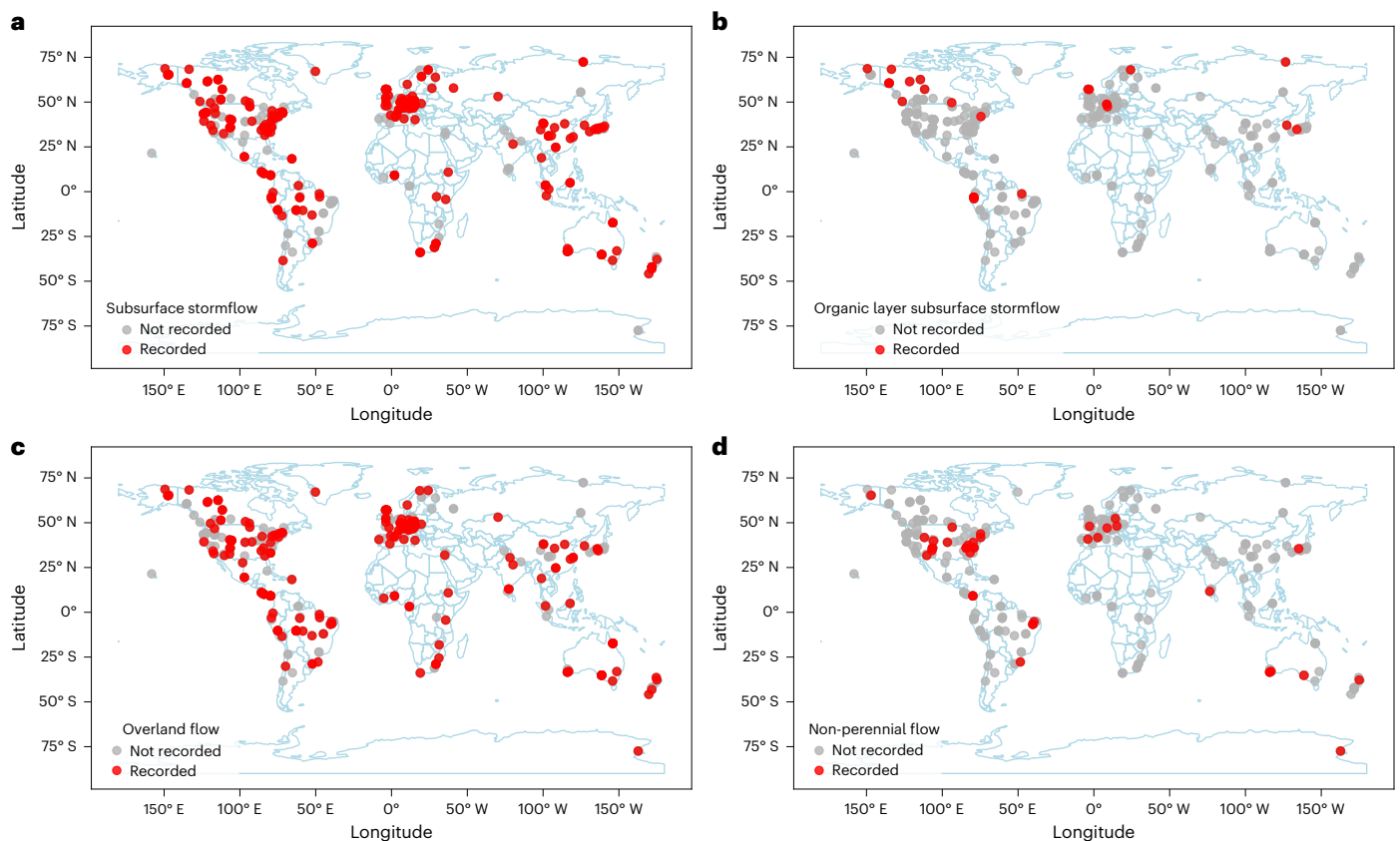


Fig. 2 | Locations of research watersheds where selected processes were recorded. a–d, Maps show recorded locations of subsurface stormflow (a), organic layer subsurface stormflow (b), overland flow (c) and non-perennial flow (ephemeral or intermittent channel flow) (d).

falling onto saturated areas) and the return flow of groundwater discharging to the surface. In humid climates, conditions of deeper soils and/or steeper, narrower valleys favour runoff derived from subsurface stormflow, with peaks produced by saturation excess and return flow.

We use the hydrologic process database to test whether Dunne's hypothesis is supported by the global body of watershed field studies. Figure 4 shows the total number of times processes are reported by vertically organized categories and Dunne's categories against global classes of aridity (aridity metric from Willmott and Feddes⁵⁰).

Our results agree with some parts of Dunne's hypothesis. There are four key points. First, arid watersheds tend to favour the shallowest (surface) or deepest (groundwater) processes, whereas humid watersheds favour mid-depth processes (soil drainage and subsurface stormflow) (Fig. 4a). The lesser importance of soil drainage and subsurface stormflow in arid and subhumid watersheds probably occurs, in part, because of deeper water tables, as opposed to humid watersheds where the water table is closer to the ground surface²³. Second, in accordance with Dunne's model, arid climates have a greater fraction of infiltration excess processes and a lesser fraction of saturation excess processes compared with humid climates (Fig. 4b). Third, the prevalence of subsurface stormflow varies little with aridity class, except for the most humid watersheds where a higher incidence of subsurface stormflow process is recorded. Fourth, and contrary to Dunne's hypothesis, saturation excess processes are more frequently described than infiltration excess processes for all aridity classes of research watersheds. These results could occur because of a myriad of controls on saturated areas—for example, saturated areas can develop because of ponding in surface depressions and topographic convergence of water, even when the water table is well below the ground surface⁵¹. Articles in the database report saturation excess overland flow in arid watersheds (aridity index <-0.5) in thin top soil layers or pockets^{52,53},

in areas of topographic convergence⁵⁴ and in alluvial near-stream areas⁵⁵. In humid watersheds, Dunne hypothesizes that thinner soils and/or gentler slopes favour surface flow, while deeper soils and/or steeper slopes favour subsurface flow. In our database, slope and soil depth show a negative correlation for individual watersheds, and we therefore analysed these variables individually. However, our results showed weak or no trend of subsurface stormflow prevalence with soil depth or slope (Extended Data Fig. 1).

As well as the occurrence of surface or subsurface processes, a second key knowledge gap in hydrology is how lateral and vertical flow processes drive the hydrologic response. This is important for global Earth system, hydrological and ecohydrological modelling capabilities, as such models often neglect lateral processes⁷ despite some recent efforts to include them^{56,57}. For example, a continental-scale integrated hydrology model that excluded lateral groundwater flow was shown to underestimate transpiration as a fraction of evapotranspiration by 15 percentage points⁵⁸. Therefore, the second hypothesis that we test is the Fan et al.³⁶ proposition that local terrain imposes a major hydrologic control by organizing down-valley drainage and convergence of surface and subsurface flow, creating drier hills and wetter valleys. They propose that lateral drainage must be accounted for in Earth system models to represent soil moisture, shallow groundwater, evapotranspiration and vegetation distribution and, therefore, energy, water and carbon fluxes. They hypothesize that the controls of lateral drainage on evapotranspiration and vegetation distribution are important everywhere except for in low-relief, arid areas (which have no surplus water to drain) and high-relief, ever-wet areas (which have plentiful water and limited water controls on evapotranspiration). Such areas are very limited, confined to Northern Africa and small areas in Western and Central Asia (low-relief, arid) and South-East Asia and small areas in Western Amazonia and Central Africa (high-relief, ever-wet).

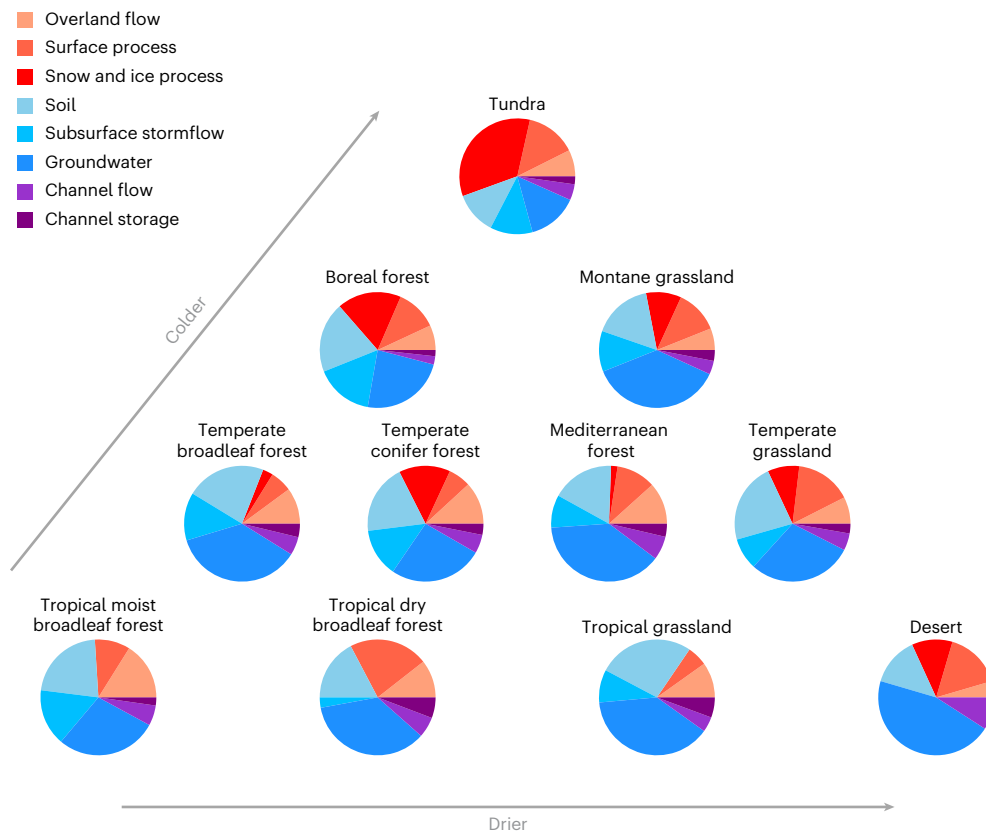


Fig. 3 | Process types recorded in each biome. Number of different processes recorded in each biome, organized by temperature and precipitation. Processes are grouped by type, where red colours indicate surface processes, blue colours indicate subsurface processes and purple colours indicate channel processes.

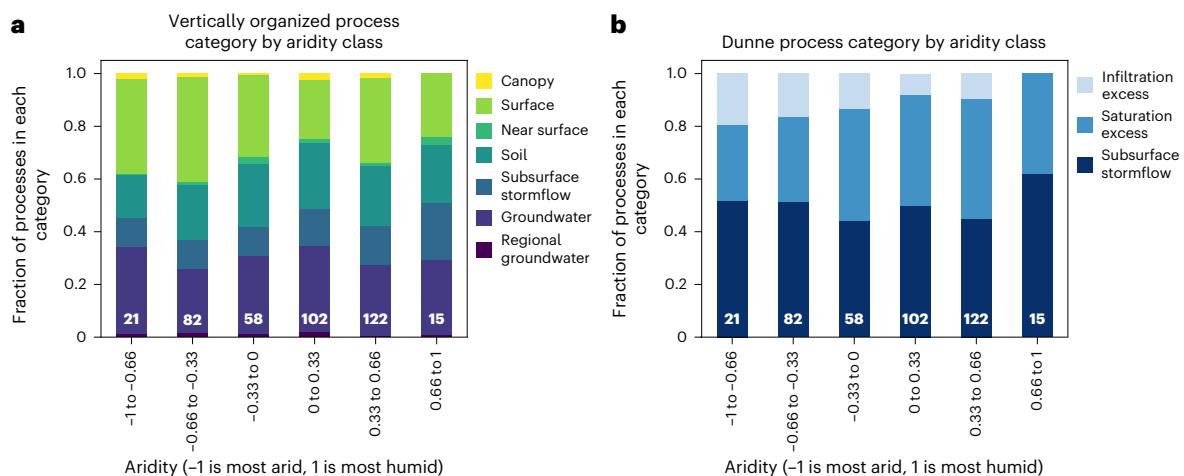


Fig. 4 | Process types recorded by aridity class. a, b, Stacked bar charts showing fractions of process types recorded as occurring in research watersheds by aridity class for vertically organized process categories from canopy to groundwater (a) and Dunne process categories (infiltration excess, saturation

excess or subsurface stormflow) (b). Positive aridity indices indicate humid climates, and negative aridity indices indicate arid climates. White numbers on bars indicate the number of process descriptions in each aridity class.

The process descriptions in our database do not have sufficiently described details on controls on energy and vegetation to test the full nuances of Fan et al.'s hypothesis, but we can map locations where lateral flow processes dominate over vertical flow processes, and vice versa. To achieve this, for each watershed in the database, we extracted all processes included in the figure or text description that describe water movement in the soil, counting whether each describes lateral or vertical flow. Watersheds where lateral processes outnumber vertical processes are defined as lateral-dominated (Fig. 5).

We found that process descriptions for research watersheds record a greater or equal number of lateral soil water and subsurface stormflow processes compared with vertical processes in 84% of watersheds (Fig. 5, red and grey dots). The spatial distribution in lateral flow processes suggests control by climate or landform. The bar graphs show that lateral flow processes are more often dominant in mountains, humid forests and cold regions. Vertical processes are dominant in fewer regions, typically in drier biomes, including temperate grassland in the interior plains of the USA, Mediterranean scrub in

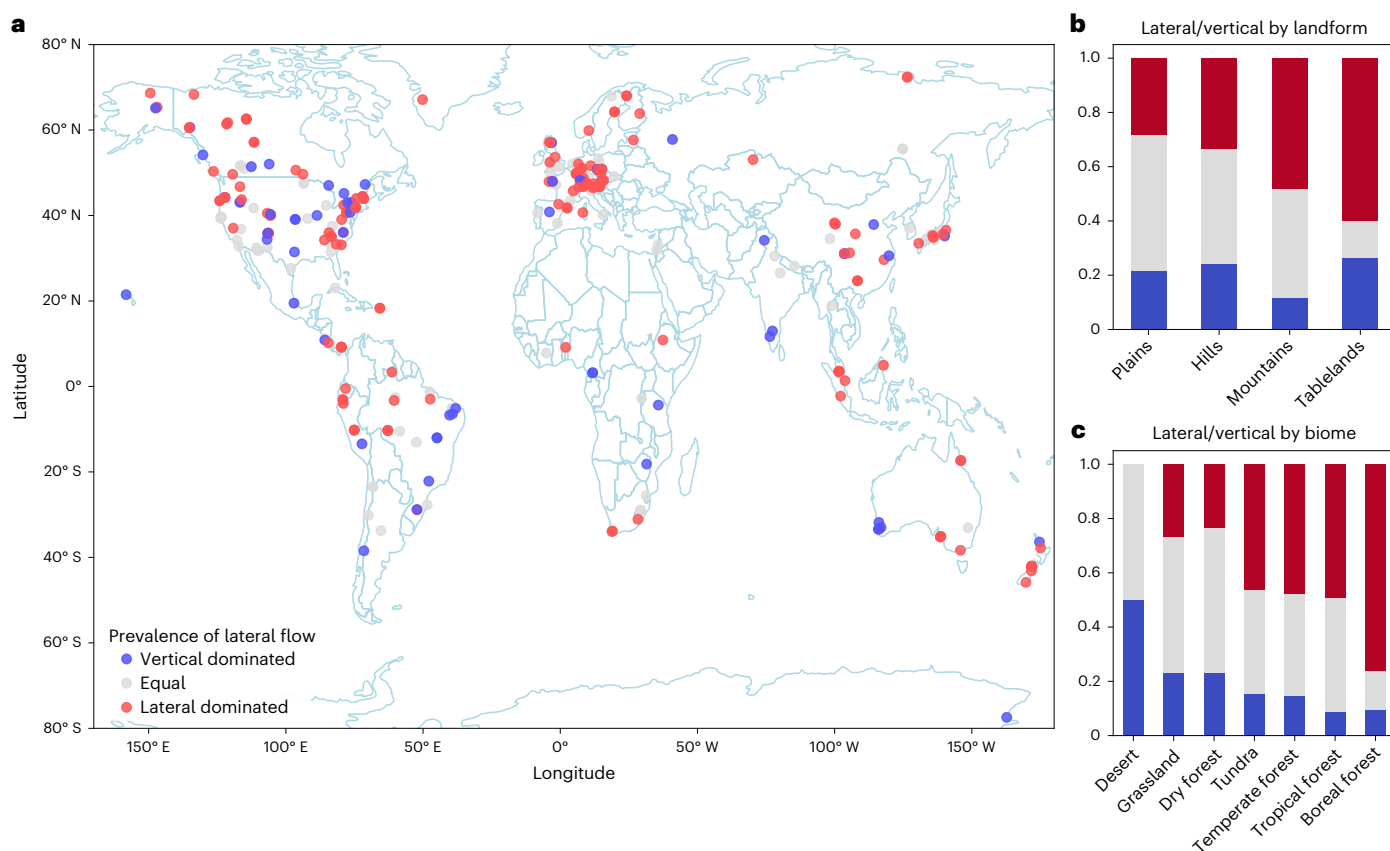


Fig. 5 | Lateral and vertical process dominance by location, landform and biome. **a–c**, Global map showing watersheds in which investigators recorded a greater number of lateral or vertical soil water flow processes or equal numbers of each, shown by location (**a**) as well as by landform type (**b**) and biome type (**c**) (bar graphs).

Southwestern Australia and savannas in the east of Africa and South America. These areas often lie in cratons with subdued topography and less well-developed river networks. Therefore, Earth system, hydrological or ecohydrological models that neglect lateral movement of water in the soil are missing a process that has been found to be important across most of the globe⁷.

Discussion and conclusions

In hydrology, overarching hypotheses about the roles of climate, biome and landform in controlling runoff generation processes have been rare and have been difficult to test experimentally. However, such hypotheses can be tested and brought to life by interrogating process descriptions from hundreds of research watersheds. This knowledge synthesis approach leverages decades of grant funding, fieldwork effort and local expertise created through field hydrology programmes around the globe. In this study, we analysed published descriptions of hydrologic processes from 400 watersheds around the globe. We found that patterns of process occurrence for subsurface flow, organic layer subsurface flow and overland flow could be related to expected differences in process with climate and land cover. By contrast, non-perennial flow occurred in locations not traditionally associated with that process. Process occurrence showed distinct differences by biome, with forested biomes being dominated by subsurface processes, while grassland, desert and tundra showed greater numbers of surface processes.

To investigate underlying controls on these spatial patterns in processes, we tested how the occurrence of surface or subsurface processes is related to watershed aridity. We found a global tendency for arid watersheds to favour shallow (surface) or deep (groundwater) processes, whereas humid watersheds favour mid-depth processes

(soil drainage and subsurface stormflow). Such analysis can be adapted to relate any process categorization to any climate or landscape metric. These data and search capabilities open the door to an efficient cycle of hypothesis generation and testing to investigate emergent patterns that relate landscape organization and hydrologic function. For modelling applications, the database can be used to determine where different processes are important in controlling river flows. This is critical in Earth system, hydrological or ecohydrological model development, where missing flow components such as lateral flow may compromise predictions, and in streamflow forecasting models such as the National Oceanic and Atmospheric Administration (NOAA)'s proposed Next Generation Water Resources Modeling Framework, which will allow spatial variability in model structure to suit regional characteristics including the dominant local processes⁵⁹.

Underlying our synthesis is a global, searchable database of research watersheds with 400 published descriptions of dominant hydrologic flow pathways. We used a hierarchical taxonomy to label process descriptions, enabling flexible aggregation, analysis and visualization of process patterns. The database offers many future opportunities for deeper investigation into how hydrologists generate knowledge about runoff generation processes. By examining the underlying journal articles in the database, we can assess what types of field evidence are most valuable to deduce dominant flow pathways and evaluate the strength of evidence in individual watersheds. This information will be valuable for the design of future hydrologic observatories or research watersheds that seek to analyse dominant runoff generation processes for applications such as flood mitigation. The database adds value to existing research watershed studies, by putting them in the context of neighbouring watersheds and those from similar landscapes. By highlighting outlying watersheds compared with regional patterns,

we can assess landscape characteristics that contribute to anomalies in hydrologic response. In the future, artificial intelligence methods might be leveraged to evaluate very large numbers of articles to extract process knowledge. Early efforts show the potential for the hydrologic sciences to treat ‘text as data’ and extract topic and location data from journal articles, although further improvements in process identification and geolocation accuracy will be required⁶⁰.

We conclude this Analysis with a call to action to preserve hydrologic process knowledge. Today’s focus on artificial intelligence and machine learning has brought about a new appreciation of the value of data across many scientific fields. Despite this, in developing the database, we became keenly aware of the risk of loss of hydrologic data in the form of historical, watershed-specific knowledge. We found journal articles that included process descriptions but in which pinpointing the location of the watershed was difficult. Older articles (for example, pre-1990) commonly included a high-resolution map of a small watershed but indicated the exact watershed location only roughly on a large-domain map. Process descriptions in such studies cannot be linked to spatial data such as land use or topography and cannot be included in a spatial database. Across all research watersheds, we appeal to authors to expose missing or obscured process descriptions. Many articles provided field data inventories or statistics but did not include quantitative information on storages and fluxes, nor interpret datasets to provide expert descriptions of dominant hydrologic processes. Our informal discussions with field scientists hint at a common worry that there is not yet enough information for an authoritative description of dominant processes, which may, therefore, never be published. In other examples, process information was included in the article, but was spread across multiple sections and was not easily accessible. Instead, we found that alternative or interdisciplinary sources, such as a collection on forest hydrology⁶¹, sometimes provided more complete process descriptions than watershed-specific journal articles. In summary, developing a global understanding of hydrologic processes is critical to underpin accurate forecasting models and digital twin simulations of Earth’s water resources. Uncovering and preserving process descriptions enables us to conserve decades of field-based domain knowledge for future hydrologic generations.

Methods

Process description source criteria

Process descriptions included in the hydrologic process database were gathered from published journal articles and reports from recognized organizations such as US Geological Survey, US Forest Service or South Africa’s Water Research Commission. Descriptions were required to include a substantially complete description of runoff generation processes, that is, movement of water from the land surface to stream channel, in either figure or text form. We required that process descriptions resulted from field investigation of the watershed, whereby local hydrologic measurements were used to determine active stores and fluxes of water. On a 4-point scale for the strength of hydrologic field evidence, such evidence would be classed as stage 3 or stage 4 (where 4 is the strongest)^{62,63}. This requirement was applied to the field investigation as a whole, not to individual processes. Descriptions of meteorology, land surface or floodplain processes were outside the study scope. We included descriptions at hillslope to watershed scales, the scales at which research watersheds are commonly instrumented and investigated to understand runoff generation. For process description figures, we followed the criteria from McMillan et al.³⁷, that the figure should relate to one specific watershed, that it should label stores and fluxes and that it should relate to the physical watershed and not to a computer model. Text descriptions were additionally required to be presented as a coherent section and not scattered throughout the article. While some sources might contain sufficient scattered information to describe runoff generation in the watershed, it was impractical to disentangle such process knowledge from the remainder of the text.

We searched a wide variety of sources to maximize the number of process descriptions found. We used a systematic search from Google Scholar using the search terms ‘runoff generation’ and ‘(perceptual model’ or ‘conceptual model’ or ‘conceptual diagram’). We used the first 500 results from this search (ordered by relevance). We used reference lists from previous publications cataloguing processes in research watersheds^{28,46,63–65}, from the *Hydrological Processes* special issue ‘Research and Observatory Catchments: the Legacy and the Future’ and from the experimental hydrology wiki (<https://experimental-hydrology.net>). We used lists of research watersheds from the Long Term Ecological Research, NorthWatch and Critical Zone Observatory networks and searched Google Scholar using search terms of the watershed name and ‘runoff generation’. We reached out directly to field and catchment hydrologists involved in the Consortium of Universities for the Advancement of Hydrologic Science, Inc. interest group on ‘Research and Observatory Catchments’ and made a public call on social media for relevant papers. All of these sources were screened for eligible process descriptions as above.

Process description analysis

For each eligible process description, we collected information about the source document, the watershed and the process description. For the source document, we collected the citation, the DOI and licence information. For the watershed, we collected the name, area, and latitude and longitude of the basin outlet. These attributes were sometimes described in the text, but in other cases had to be manually derived from maps included in the article or from other articles describing the same watershed. We urge future authors to include latitude and longitude values with more than three-decimal-place precision in all articles describing research watersheds.

For the process description in a figure or in the text, we collected a list of named processes, and the equivalent standard process names from a taxonomic classification of hydrologic processes²⁸. If processes were described as very rarely or never occurring, we did not include them in the list. Other than that criterion, we did not label processes as dominant or otherwise in the watershed as the decision was found to be too subjective to produce repeatable results. Using standard process names enables us to search for equivalent but differently named processes in different watersheds, and to search hierarchically for groups of processes. We describe processes as being ‘recorded’ in the watershed if the article names the process in the figure or text description. Our criteria specify that the article must include a substantially complete description of runoff generation processes. Therefore, we assume that non-recorded processes have a low probability of occurring and, thus, that over multiple watersheds the number of recordings and importance of the process scale with each other. We recorded whether the process description used subpanels (figure description) or subsections (text description) to describe spatiotemporal variability by specifying processes at different time periods or spatial zones (for example, by season, or by hillslope position). We recorded the number of divisions and the type of classification. For the purposes of our analysis, processes that occur in any time period or spatial zone were associated with the watershed location as specified by the coordinates of the watershed outlet. We collected additional information conveyed by the figure or text such as types of vegetation, soil and geology and whether any information on uncertainty was presented.

Process database specifications

The process descriptions are stored in a PostgreSQL relational database. Each figure or text description is linked to metadata, including the article citation and watershed location. Further, each description is linked to a list of processes that occur in the watershed, with standardized process names from a hydrologic process taxonomy²⁸. This taxonomy allows hierarchical searching for watersheds exhibiting any process, including its subprocesses, such as overland flow, including

infiltration excess and saturation excess flow processes. A schema for the database is shown in Extended Data Fig. 2. A public version of the database is available as described in the 'Data availability' section; this version redacts original text/figures for articles without an open-access licence.

Process depth analysis

To create Fig. 4, showing the fraction of processes in vertically organized categories, we assigned each process in each watershed to a depth category, as follows. Categories were specified via the hierarchical taxonomy²⁸ so that each category included all subprocesses related to a parent process. Canopy processes included canopy evaporation, interception, canopy snow processes and canopy sublimation processes. Surface processes included all surface processes in the taxonomy with the exception of canopy processes (thereby including evapotranspiration, snow, glacier and frozen ground processes, overland flow and infiltration). Near-surface processes included subsurface stormflow in the organic horizon. Soil processes included all water fluxes in the unsaturated zone below the organic horizon. Subsurface stormflow processes included lateral matrix and macropore flow at soil horizons and at the soil–bedrock interface, and included variable source area subsurface stormflow. Groundwater processes included all groundwater fluxes within the watershed and included groundwater flow, perched water table processes, displacement of groundwater and return flow. Regional groundwater processes included regional groundwater flux and deep groundwater losses.

For each watershed in the database, we calculated an aridity index using the Thornthwaite moisture index that is based on the ratio of precipitation and potential evapotranspiration⁵⁰, using monthly average climate values from the Climatic Research Unit Time Series (CRU TS) v3.23 dataset⁶⁶. For each aridity class, we selected all watersheds in that class, and pooled all processes that were recorded in those watersheds with their number of occurrences. The number of process occurrences in each depth category was divided by the total number of process occurrences in the aridity class to give the fraction of processes in each depth category shown in Fig. 4.

Soil depth and slope analysis

Extended Data Fig. 1 shows the fraction of overland flow and subsurface stormflow process by landscape slope class and soil thickness class. Slope values were derived from a 1-km-pixel-size classification of median slope (among a variety of global topographic variables), based on a 250-m Global Multi-resolution Terrain Elevation Data 2010 dataset⁶⁷. Larger pixel sizes from 5 km to 100 km were available, but we selected 1 km based on the small size of most watersheds in the database (median watershed size 6.9 km²). Slope classes were selected on a log scale based on the distribution of slope values in our dataset. Soil thickness values were derived from a global 1-km dataset of thickness of soil, regolith and sedimentary deposit layers that gives an average value of the thickness of unconsolidated material (including soil on upland hillslopes and sedimentary deposits in valley bottoms)⁶⁸.

Lateral process analysis

To create Fig. 5, showing whether each watershed location is dominated by lateral or vertical processes, we considered only processes in the soil and subsurface stormflow categories. We used this subset of processes to balance the intention of the article by Fan et al.³⁶, which guided the analysis, with the level of detail available in the database. Fan et al.³⁶ describe lateral flow as the lateral ridge-to-valley convergence that routes water from hillslopes towards the nearest stream channel, creating drier hills and wetter valleys. In process descriptions in the database, soil water flux is typically differentiated into vertical drainage in the unsaturated zone, and lateral flow that occurs only at impeding layers including soil horizons and at the bedrock interface. Vertical drainage is also known as percolation, recharge and vertical

matrix or macropore flow. Lateral flow at impeding layers is also known as subsurface stormflow, throughflow and interflow. In the saturated (groundwater) zone, lateral and vertical movement of water are not as clearly distinguished, with articles often referring to groundwater flow or groundwater displacement but not specifying the degree to which lateral or vertical water movement is involved. Therefore, for this Analysis, we did not include groundwater processes. For soil and subsurface stormflow categories, we assigned each process recorded in each watershed to a lateral or vertical category. As for the previous analyses, our method assumes that the number of recordings and importance of the lateral or vertical category scale with each other. Categories were specified via the hierarchical taxonomy²⁸ so that each category included all subprocesses related to a parent process. Lateral processes included all subsurface stormflow processes and included lateral matrix flow, lateral macropore flow, lateral unsaturated flow and topographic convergence of catchment waters. Vertical processes included vertical matrix flow, vertical macropore flow and vertical drainage to groundwater. For each watershed in the database, we counted the number of recorded soil and subsurface stormflow processes according to the categories above. Watersheds where the recorded number of lateral processes was greater than, equal to or less than the number of vertical processes were respectively specified as lateral-dominated, equal or vertical-dominated.

The bar graphs in Fig. 5 show the fraction of lateral-dominated, equal or vertical-dominated watersheds by landform type and by biome type. Landform types are derived from a 250-m resolution World Terrestrial Ecosystems raster classification with four classes (plains, hills, mountains and tablelands), developed by the US Geological Survey, Esri and The Nature Conservancy⁶⁹. These classes are a simplified reclassification of the World Ecological Facets Landform Classes derived using the Hammond Landform Classification Algorithm⁷⁰. Biomes are labelled according to a global classification of 14 biomes by a team convened for the World Wildlife Fund⁷¹. We used the ArcGIS layer of these biomes from the RESOLVE Ecoregions and Biomes dataset⁷². For visual clarity, the bar graph combines the following biomes: temperate broadleaf and conifer forests combined into temperate forest; tropical dry broadleaf forest and mediterranean forest combined into dry forest; and tropical, temperate and montane grassland combined into grassland.

Data availability

The perceptual model database in the current study, including access via a GIS dashboard, is available via Zenodo at <https://doi.org/10.5281/zenodo.14829661> (ref. 73) and as a continuously updated version via GitHub at <https://github.com/Ry4GIT/perceptual-models>. The RESOLVE Ecoregions and Biomes data are available at https://hub.arcgis.com/datasets/37ea320eebb647c6838c23f72abae5ef_0/about. The CRU TS v3.23 data are available at https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_3.23/. The Global Multi-resolution Terrain Elevation Data 2010 data are available at <https://doi.org/10.1594/PANGAEA.867115>. The thickness of soil, regolith and sedimentary deposit data are available at <https://doi.org/10.3334/ORNLDAAAC/1304>. The World Terrestrial Ecosystems data are available at <https://doi.org/10.5066/P9DO61LP>.

Code availability

The code used to analyse data and produce the figures in this Analysis is available via Zenodo at <https://doi.org/10.5281/zenodo.14829661> (ref. 73) and as a continuously updated version via GitHub at <https://github.com/Ry4GIT/perceptual-models>.

References

- Brooks, P. D. et al. Hydrological partitioning in the critical zone: recent advances and opportunities for developing transferable understanding of water cycle dynamics. *Water Resour. Res.* **51**, 6973–6987 (2015).

2. Penna, D. A recipe for why and how to set up and sustain an experimental catchment. *Hydrol. Process.* **38**, e15163 (2024).
3. Blöschl, G. Hydrologic synthesis: across processes, places, and scales. *Water Resour. Res.* **42**, W03S02 (2006).
4. Kirchner, J. W. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* **42**, W03S04 (2006).
5. Paola, C. et al. Toward a unified science of the Earth's surface: opportunities for synthesis among hydrology, geomorphology, geochemistry, and ecology. *Water Resour. Res.* **42**, W03S10 (2006).
6. Sivapalan, M. in *Encyclopedia of Hydrological Sciences* Ch. 13 (Wiley 2006); <https://doi.org/10.1002/0470848944.hsa012>
7. Clark, M. P. et al. Improving the representation of hydrologic processes in Earth system models. *Water Resour. Res.* **51**, 5929–5956 (2015).
8. Bauer, P., Stevens, B. & Hazeleger, W. A digital twin of Earth for the green transition. *Nat. Clim. Chang.* **11**, 80–83 (2021).
9. Brocca, L. et al. A Digital Twin of the terrestrial water cycle: a glimpse into the future through high-resolution Earth observations. *Front. Sci.* **1**, 1190191 (2024).
10. Li, X. et al. Big Data in Earth system science and progress towards a digital twin. *Nat. Rev. Earth Environ.* **4**, 319–332 (2023).
11. Blair, G. S. Digital twins of the natural environment. *Patterns* **2**, 100359 (2021).
12. Blöschl, G. et al. Twenty-three unsolved problems in hydrology (UPH)—a community perspective. *Hydrol. Sci. J.* **64**, 1141–1158 (2019).
13. Beven, K. J. Uniqueness of place and process representations in hydrological modelling. *Hydrol. Earth Syst. Sci.* **4**, 203–213 (2000).
14. Dorigo, W. A. et al. The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.* **15**, 1675–1698 (2011).
15. Kratzert, F. et al. Caravan—a global community dataset for large-sample hydrology. *Sci. Data* **10**, 61 (2023).
16. Vlah, M. et al. MacroSheds: enabling continental-scale comparison of watershed biogeochemistry. *AGU Fall Meeting Abstracts* **2022**, H52J-0581 (2022).
17. Gleeson, T., Moosdorf, N., Hartmann, J. & van Beek, L. V. A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity. *Geophys. Res. Lett.* **41**, 3891–3898 (2014).
18. Jasechko, S., Seybold, H., Perrone, D., Fan, Y. & Kirchner, J. W. Widespread potential loss of streamflow into underlying aquifers across the USA. *Nature* **591**, 391–395 (2021).
19. Almoradie, A., Jonoski, A., Popescu, I. & Solomatin, D. Web based access to water related data using OGC WaterML 2.0. *Int. J. Adv. Comput. Sci. Appl.* **3**, 83–89 (2013).
20. Blodgett, D., Johnson, J. M., Sondheim, M., Wiczorek, M. & Frazier, N. Mainstems: a logical data model implementing mainstem and drainage basin feature types based on WaterML2 part 3: HY features concepts. *Environ. Model. Softw.* **135**, 104927 (2021).
21. Wlostowski, A. N. et al. Signatures of hydrologic function across the critical zone observatory network. *Water Resour. Res.* **57**, e2019WR026635 (2021).
22. Zhang, L. et al. CHOSEN: a synthesis of hydrometeorological data from intensively monitored catchments and comparative analysis of hydrologic extremes. *Hydrol. Process.* **35**, e14429 (2021).
23. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science* **339**, 940–943 (2013).
24. Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B. & Otero-Casal, C. Hydrologic regulation of plant rooting depth. *Proc. Natl Acad. Sci. USA* **114**, 10572–10577 (2017).
25. Jasechko, S., Kirchner, J. W., Welker, J. M. & McDonnell, J. J. Substantial proportion of global streamflow less than three months old. *Nat. Geosci.* **9**, 126–129 (2016).
26. Tetzlaff, D., Carey, S. K., McNamara, J. P., Laudon, H. & Soulsby, C. The essential value of long-term experimental data for hydrology and water management. *Water Resour. Res.* **53**, 2598–2604 (2017).
27. Arora, B. et al. Building cross-site and cross-network collaborations in critical zone science. *J. Hydrol.* **618**, 129248 (2023).
28. McMillan, H. A taxonomy of hydrological processes and watershed function. *Hydrol. Process.* **36**, e14537 (2022).
29. Beven, K. J. & Chappell, N. A. Perceptual perplexity and parameter parsimony. *Wiley Interdisc. Rev. Water* **8**, e1530 (2021).
30. Wagener, T. et al. Knowledge gaps in our perceptual model of Great Britain's hydrology. *Hydrol. Process.* **35**, e14288 (2021).
31. Wagener, T. et al. On doing hydrology with dragons: realizing the value of perceptual models and knowledge accumulation. *WIREs Water* **8**, e1550 (2021).
32. Clark, M. P. et al. A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resour. Res.* **51**, 2515–2542 (2015).
33. Johnson, J. M. et al. Comprehensive analysis of the NOAA National Water Model: a call for heterogeneous formulations and diagnostic model selection. *J. Geophys. Res.* **128**, e2023JD038534 (2023).
34. Dunne, T. Relation of field studies and modeling in the prediction of storm runoff. *J. Hydrol.* **65**, 25–48 (1983).
35. Dunne, T. in *Hillslope Hydrology* (ed. Kirkby, M. J.) 227–293 (Wiley, 1978).
36. Fan, Y. et al. Hillslope hydrology in global change research and Earth system modeling. *Water Resour. Res.* **55**, 1737–1772 (2019).
37. McMillan, H., Araki, R., Gnan, S., Woods, R. & Wagener, T. How do hydrologists perceive watersheds? A survey and analysis of perceptual model figures for experimental watersheds. *Hydrol. Process.* **37**, e14845 (2023).
38. Melo, D. C. D. et al. The big picture of field hydrology studies in Brazil. *Hydrol. Sci. J.* **65**, 1262–1280 (2020).
39. Coxon, G. et al. CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain. *Earth Syst. Sci. Data* **12**, 2459–2483 (2020).
40. Krabbenhoft, C. A. et al. Assessing placement bias of the global river gauge network. *Nat. Sustain.* **5**, 586–592 (2022).
41. Linton, J. Is the hydrologic cycle sustainable? A historical–geographical critique of a modern concept. *Ann. Assoc. Am. Geogr.* **98**, 630–649 (2008).
42. Mohai, P., Pellow, D. & Roberts, J. T. Environmental justice. *Annu. Rev. Environ. Resour.* **34**, 405–430 (2009).
43. McGuire, K. J., Klaus, J. & Jackson, C. R. James buttle review: interflow, subsurface stormflow and throughflow: a synthesis of field work and modelling. *Hydrol. Process.* **38**, e15263 (2024).
44. Bring, A. et al. Arctic terrestrial hydrology: a synthesis of processes, regional effects, and research challenges. *J. Geophys. Res. Biogeosci.* **121**, 621–649 (2016).
45. Bonell, M. in *Forests, Water and People in the Humid Tropics* (eds Bonell, M. & Bruijnzeel, L. A.) 314–406 (Cambridge Univ. Press, 2005); <https://doi.org/10.1017/cbo9780511535666.020>
46. Bonell, M. Progress in the understanding of runoff generation dynamics in forests. *J. Hydrol.* **150**, 217–275 (1993).
47. Shanafield, M., Bourke, S. A., Zimmer, M. A. & Costigan, K. H. An overview of the hydrology of non-perennial rivers and streams. *WIREs Water* **8**, e1504 (2021).
48. Whittaker, R. H. Classification of natural communities. *Bot. Rev.* **28**, 1–239 (1962).
49. Belnap, J. The world at your feet: desert biological soil crusts. *Front. Ecol. Environ.* **1**, 181–189 (2003).

50. Willmott, C. J. & Feddema, J. J. A more rational climatic moisture index. *Prof. Geogr.* **44**, 84–88 (1992).
51. Loague, K. R-5 revisited: 2. Reevaluation of a quasi-physically based rainfall–runoff model with supplemental information. *Water Resour. Res.* **26**, 973–987 (1990).
52. Martínez-Mena, M., Albaladejo, J. & Castillo, V. M. Factors influencing surface runoff generation in a Mediterranean semi-arid environment: Chicamo watershed, SE Spain. *Hydrol. Process.* **12**, 741–754 (1998).
53. Lange, J. et al. Runoff generation from successive simulated rainfalls on a rocky, semi-arid, Mediterranean hillslope. *Hydrol. Process.* **17**, 279–296 (2003).
54. George, R. J. & Conacher, A. J. Mechanisms responsible for streamflow generation on a small, salt-affected and deeply weathered hillslope. *Earth Surf. Process. Landforms* **18**, 291–309 (1993).
55. Liu, F., Parmenter, R., Brooks, P. D., Conklin, M. H. & Bales, R. C. Seasonal and interannual variation of streamflow pathways and biogeochemical implications in semi-arid, forested catchments in Valles Caldera, New Mexico. *Ecohydrology* **1**, 239–252 (2008).
56. Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M. & Perket, J. Representing intrahillslope lateral subsurface flow in the community land model. *J. Adv. Model. Earth Syst.* **11**, 4044–4065 (2019).
57. Chaney, N. W., Torres-Rojas, L., Vergopolan, N. & Fisher, C. K. HydroBlocks v0. 2: enabling a field-scale two-way coupling between the land surface and river networks in Earth system models. *Geosci. Model Dev.* **14**, 6813–6832 (2021).
58. Maxwell, R. M. & Condon, L. E. Connections between groundwater flow and transpiration partitioning. *Science* **353**, 377–380 (2016).
59. Cosgrove, B. et al. NOAA's National Water Model: advancing operational hydrology through continental-scale modeling. *J. Am. Water Resour. Assoc.* **60**, 247–272 (2024).
60. Stein, L. et al. Wealth over woe: global biases in hydro-hazard research. *Earth's Future* **12**, e2024EF004590 (2024).
61. Amatya, D. M. et al. in *Forest Hydrology: Processes, Management and Assessment* (eds Amatya, D. M. et al.) 219–239 (CABI, 2016); <https://doi.org/10.1079/9781780646602.0219>
62. Elsenbeer, H. & Vertessy, R. A. Stormflow generation and flowpath characteristics in an Amazonian rainforest catchment. *Hydrol. Process.* **14**, 2367–2381 (2000).
63. Barthold, F. K. & Woods, R. A. Stormflow generation: a meta-analysis of field evidence from small, forested catchments. *Water Resour. Res.* **51**, 3730–3753 (2015).
64. Klaus, J. & Jackson, C. R. Interflow is not binary: a continuous shallow perched layer does not imply continuous connectivity. *Water Resour. Res.* **54**, 5921–5932 (2018).
65. McMillan, H. K., Gnann, S. J. & Araki, R. Large scale evaluation of relationships between hydrologic signatures and processes. *Water Resour. Res.* **58**, e2021WR031751 (2022).
66. Harris, I. P. D. J., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 dataset. *Int. J. Climatol.* **34**, 623–642 (2014).
67. Amatulli, G. et al. A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Sci. Data* **5**, 180040 (2018).
68. Pelletier, J. D. et al. A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling. *J. Adv. Model. Earth Syst.* **8**, 41–65 (2016).
69. Sayre, R. et al. An assessment of the representation of ecosystems in global protected areas using new maps of World Climate Regions and World Ecosystems. *Glob. Ecol. Conserv.* **21**, e00860 (2020).
70. Karagulle, D. et al. Modeling global Hammond landform regions from 250-m elevation data. *Trans. GIS* **21**, 1040–1060 (2017).
71. Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on Earth: a new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience* **51**, 933–938 (2001).
72. Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67**, 534–545 (2017).
73. McMillan, H. et al. RY4GIT/perceptual-models: v1.2: McMillan et al., 2025 release (v1.2). *Zenodo* <https://doi.org/10.5281/zenodo.14829661> (2025).

Acknowledgements

We thank T. Dunne and Y. Fan Reinfelder for advising us on comparisons with their work. H.M. was supported by NSF Hydrologic Sciences Program, NSF Division of Earth Sciences, grant/award 2322510. G.C. was supported by a United Kingdom Research and Innovation Future Leaders Fellowship (MR/V022857/1). M.C. was supported by the National Oceanic and Atmospheric Administration (NOAA), funding awarded to the Cooperative Institute for Research on Hydrology (CIROH) through the NOAA Cooperative Agreement with the University of Alabama, NA22NWS4320003.

Author contributions

H.M. conceived the study idea and design and acquired funding for the project. H.M., R.A., L.B. and D.-H.K. conducted the analyses. H.M. wrote the Analysis with important input from R.A., G.C. and M.C. All authors (H.M., R.A., L.B., D.-H.K., G.C., M.C. and J.S.) provided edits and feedback on the text and figures.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s44221-025-00407-w>.

Correspondence and requests for materials should be addressed to Hilary McMillan.

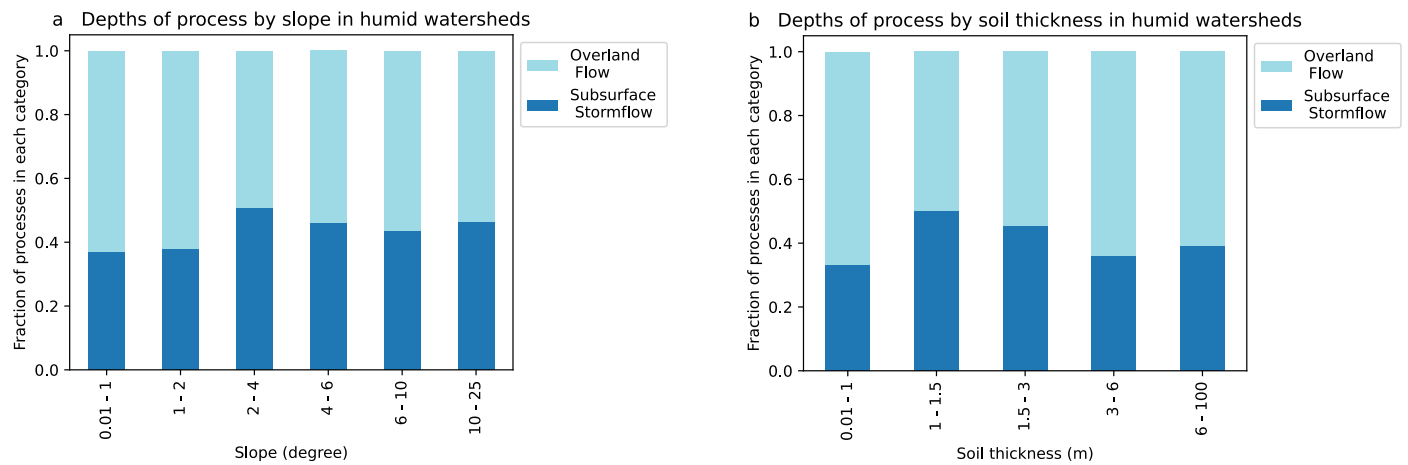
Peer review information *Nature Water* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

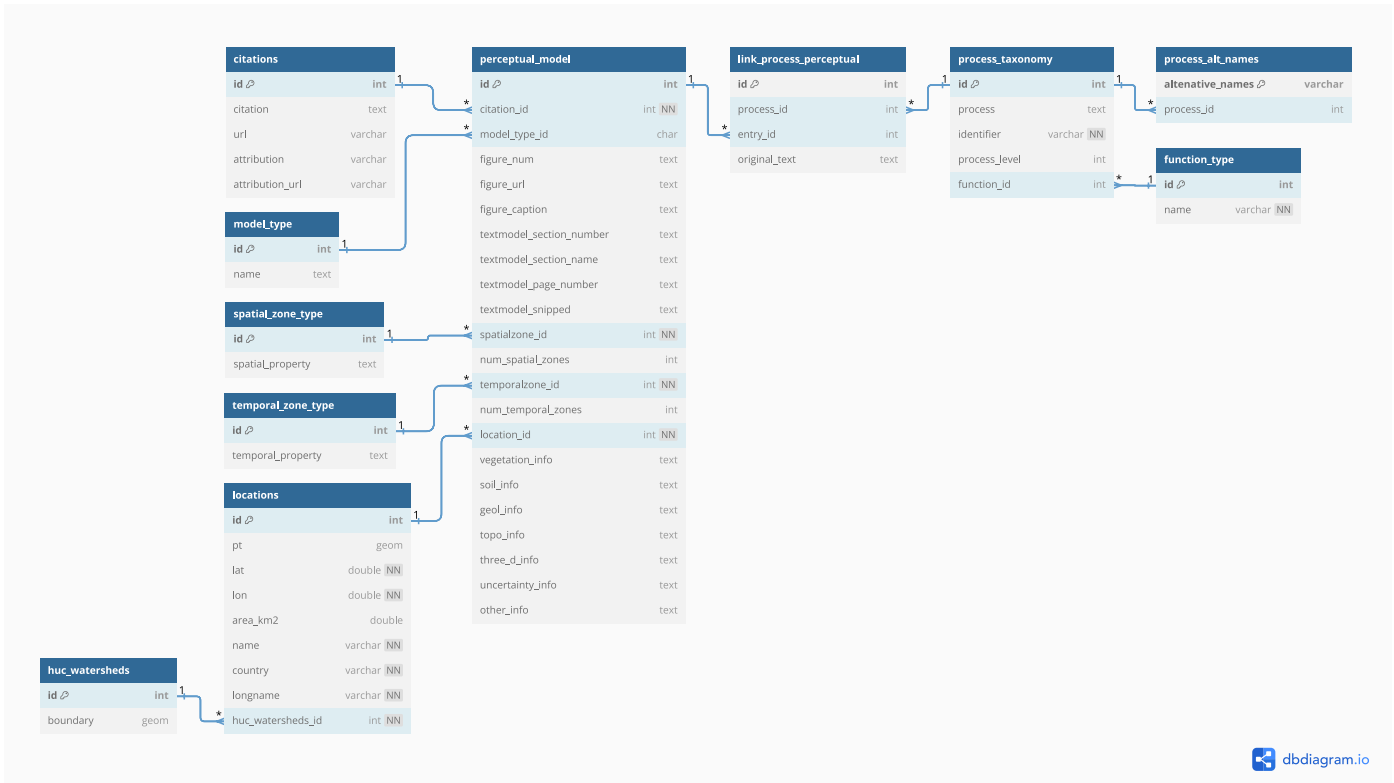
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025



Extended Data Fig. 1 | Process types recorded by slope and soil thickness class. Stacked bar charts showing fractions of overland flow vs. subsurface stormflow processes recorded in process descriptions of humid research watersheds by **a.** Slope class, and **b.** Soil Thickness class.



Extended Data Fig. 2 | Database schema for Hydrologic Process Database. Schema shows database tables with their names, fields, primary key and relationships between tables. Relationships are indicated as one (1) to many (*).