Article

Distinct hydrologic response patterns and trends worldwide revealed by physics-embedded learning

Haoyu Ji[1], Yalan Song[1], Tadd Bindas[1], Chaopeng Shen [1] ✉, Yuan Yang[2], Ming Pan[2], Jiangtao Liu [1], Farshid Rahmani [1], Ather Abbas[3], Hylke Beck [3], Kathryn Lawson [1] & Yoshihide Wada [4]

To track rapid changes within our water sector, Global Water Models (GWMs) need to realistically represent hydrologic systems' response patterns – such as the baseflow fraction of streamflow – but are hindered by their limited ability to learn from data. Here we introduce a high-resolution, physics-embedded, big-data-trained model to reliably capture characteristic hydrologic response patterns (signatures) and their shifts. By realistically representing the long-term water balance, the model revealed widespread shifts – in some cases, more than 20% over 20 years – in fundamental green-blue-water partitioning and baseflow ratios worldwide. Shifts in these previously-assumed-static response patterns contributed to increasing flood risks in northern mid-latitudes, heightening water supply stresses in southern subtropical regions, and declining freshwater inputs to many European estuaries, all with ecological implications. With substantially more accurate simulations at monthly and daily scales than current operational systems, this next-generation model resolves large, nonlinear, seasonal runoff responses to rainfall (elasticity) and streamflow flashiness in semi-arid and arid regions. Our results highlight regions with management challenges due to large water supply variability and high climate sensitivity, and demonstrate an advanced tool to forecast seasonal water availability. This capability enables global-scale models to deliver reliable and locally-relevant insights for water management.

While extremes in precipitation and temperature are becoming more frequent[1], the terrestrial hydrologic system does not respond uniformly. Instead, the landscape modulates the impacts and feedbacks of these changes through complex and highly heterogeneous processes across space and time[2]. The landscape partitions precipitated water into evapotranspiration (ET) and runoff via surface and groundwater pathways, and releases these fluxes at basin-specific rates. Because of strong storage thresholds, memory, and nonlinear effects[3–5], the landscape can translate the same amount of warming or changes in precipitation into either muted or disproportionately large changes in floods and droughts with high spatial heterogeneity[6,7]. To account for such nonlinear effects, Global Water Models (GWMs) such as WaterGAP[8], DBH[9], H08[10], LPJml[11], MATSIRO[12], and PCR-GLOBWB[13] were developed with integrated rainfall-runoff, river routing, and human water use processes to describe the terrestrial water cycle. GWMs are important tools referred to by organizations like the

[1]Department of Civil and Environmental Engineering, The Pennsylvania State University, University Park, PA, USA. [2]Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA. [3]Physical Science and Engineering Department, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. [4]Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia. ✉e-mail: cshen@engr.psu.edu

Intergovernmental Panel on Climate Change (IPCC) to project future changes in flood hazards, water availability, and societal resilience[14]. In a series of high-impact analyses, GWMs have been employed to improve our understanding of climate change effects[15] and provide assessments of water resources[16].

The characteristic response patterns of a hydrologic system (often called hydrologic signatures) are interpretable, actionable, yet challenging-to-model summaries that help stakeholders anticipate changes, identify challenges, and manage risks[17,18]. Some signatures, including evaporation-precipitation ratio, baseflow-streamflow ratio, autocorrelation, runoff sensitivity to rainfall, and flashiness of the flow duration curve, have strong implications for water management and aquatic ecosystem health[18]. These signatures control the quantity, timing, variability, temperature, and quality of freshwater exported downstream, which, in turn, exert first-order controls on the ecosystem composition of the subsequent water bodies. For example, drought-induced reductions in freshwater inputs to estuaries, controlled by precipitation-streamflow elasticity (discussed more below), can increase salinity for open estuaries and reduce salinity for intermittent ones[19]. The signatures can be highly valuable for decision makers. For example, understanding the ET-runoff partitioning of a basin enables rapid assessment of water availability, while quantifying seasonal runoff sensitivity to rainfall readily allows estimation of how forecasted multi-month droughts will impact summertime streamflows. Similarly, understanding the baseflow ratio improves predictions of water quality and temperature[20]. However, a model that estimates these metrics needs to account for the hydrologic processes in all the (sometimes very large) upstream catchments, and must resolve systems' responses to inputs and recession behaviors at small timescales with high spatiotemporal precision. As we will show later, established GWMs can be challenged in these regards due to inadequate resolution and parameterization strategies.

Stakeholders worldwide have substantial urgent and unmet water prediction needs that current GWMs were not designed to address, due to the inherent tradeoffs between global coverage and high-quality local predictions, e.g., computational requirements. GWMs have so far mainly been tasked with describing large-scale, long-term change trends for entire climate zones, e.g., long-term-average runoff on continental-scale large rivers[21,22]. It is not clear if these tasks have been fulfilled to the extent permitted by data, as GWMs have shown divergent behaviors in projecting future extremes[15] and precipitation-recharge responses[23]. Moreover, coarse-resolution and loosely-calibrated GWMs are not intended to be practical water management tools at local scales, due to known complications from location-specific, scale-dependent hydrologic processes[24,25]. For short-term tasks like flood and drought forecasting, GWM accuracy is often below operational requirements for daily or sub-seasonal forecasts, e.g., having daily Nash-Sutcliffe model efficiency coefficient (NSE) values lower than 0.5[26]. Until recently, stakeholders have either relied on detailed and costly area-specific models developed and calibrated for local applications, or their needs have been left unfulfilled, especially in developing nations. Thus, advancing global hydrological modeling to provide communities around the world with locally-accurate predictions should be a core concern of the scientific community.

With traditional modeling approaches, large observational datasets such as streamflow and soil moisture cannot readily benefit global predictions and communities. There are relatively few avenues for such data to inform GWMs besides basic parameter calibration (e.g., with GloFAS[27]) and post-simulation bias correction (e.g., with WaterGAP[8] or GRFR[28]), which hinders the closure of the water balance. Even a rough calibration is hindered by the large computational demands of modeling at the global scale, and also suffers from the infamous issue of parameter nonuniqueness (equifinality)[29]. Parameter regionalization, or generalizing parameters in space, remains a large and persistent challenge despite myriad proposed schemes[30]. Such

methods cannot take advantage of the synergistic effects of large datasets, in which observations from diverse sites can jointly inform one model to increase its robustness[31].

Recently, deep neural networks (NNs) have shown a formidable ability to learn from data and generate hydrological predictions, but have not yet benefited terrestrial water cycle assessment tasks. Purely data-driven algorithms like long short-term memory (LSTM)[32], transformer[33], and diffusion[34] networks have demonstrated success at simulating different hydrologic variables[35,36] and especially streamflow[37–40] at lumped, small- to meso- basin scales. Nevertheless, their interpretability remains unsatisfactory: purely data-driven NNs do not have physical concepts like ET and baseflow when trained only on streamflow data and may not support the calculation of hydrologic signatures. Furthermore, it is uncertain whether NN models optimized for daily NSE or Kling-Gupta Efficiency (KGE) can satisfactorily reproduce long-term trends in large rivers or fill the gaps in data-scarce regions.

An opportunity has emerged to realistically describe the terrestrial water cycle, leveraging the advantages of both process-based and deep learning models. Physics-embedded machine learning (especially "differentiable") models[41] contain connected NNs and process-based equations which learn parameters or missing processes from data and are trained in a single step, enabling the complete tracing of inputs to outputs (see Supplementary Fig. S1). Such models have shown comparable performance to LSTM while maintaining process interpretability and diagnostic capability for untrained variables such as snowmelt, groundwater recharge, baseflow, and ET[42–46]. Due to the process-based components of the model, differentiable models generalize better than LSTM in data-scarce regions[47] and in representing extremes[48]. However, applying them for terrestrial water cycle assessment requires a performant and efficient river routing scheme to simulate major rivers, which necessitated improvements to our existing differentiable routing model[49]. It remained uncertain whether the newer learnable models or established GWMs not trained on large data could accurately represent the spatial variability of hydrologic signatures and their temporal changes for continental-scale rivers.

Here we demonstrate that by effectively learning from data at different scales, a physics-embedded differentiable hydrologic model can advance the representation of global hydrologic response patterns, revealing previously-unrecognized hydrologic shifts occurring over the last 20 years. We investigated multiple questions:

(1) How much has basic hydrologic partitioning (including evapotranspiration-to-precipitation and baseflow-to-runoff ratios) shifted worldwide in the past two decades according to data-trained models, and what are the implications for some important estuaries?

(2) How does winter and summer runoff worldwide respond to accumulated precipitation in the previous months, as characterized by seasonal runoff elasticity?

(3) Can big-data model training lead to the long-sought step change in GWM performance—improving reliability for continental-scale impact assessments while also increasing relevance to local stakeholders?

Differentiable HBV with Muskingum-Cunge routing (full version name δHBV2δMC2-Globe2-hydroDL, referred to as δHBV2 for brevity) is a hybrid, multiscale model that can learn from thousands of sites and output hydrologic fluxes and states at high spatial resolution. As it was only recently introduced in Song et al.[50], it has not yet been used to generate global hydrologic insights, nor has it been compared with GWMs. In short, a neural network generates physical parameters for a differentiable implementation of the conceptual hydrological model Hydrologiska Byråns Vattenbalansavdelning (HBV)[51]. Rainfall-runoff processes are simulated with HBV's equations at small unit basins (MERIT network, median catchment size ~37 km²) and then routed

downstream by a differentiable Muskingum-Cunge (MC) model (Methods). A total of 4746 basins with catchment areas less than 50,000 km^2 are used for model training and testing. Evaluations are carried out on (1) 33 large global rivers with mixed anthropogenic influences (called mixed-anthropogenic-impact rivers or mixed-impact rivers for short) previously used in intercomparisons[52] of six models (GWM0-GWM5) in ISIMIP2a[53]; (2) 28 other large rivers with less human impacts and a relatively minimal catchment area of 500,000 km^2 (natural rivers); and (3) >5000 smaller basins with relatively continuous streamflow records.

Results and discussion

In the following, we examine how basic hydrologic partitioning has changed between 2001 and 2020 and give practical examples of the implications for a number of US and European estuaries. To support the analysis in each case, we provide a benchmark against established GWMs in terms of matching various aspects of observations on large global rivers as well as smaller rivers. Using simulation benchmarks from both large and small global rivers, we explain how the differentiable model can better capture changes that are challenging for established GWMs.

Green-blue-water and baseflow-surface runoff partitioning

Our assessment of terrestrial water partitioning relies on high-resolution simulations that can accurately capture water balances and their change trends. δHBV2 offers minimal bias for the long-term mean annual runoff (MAR)—it has a mean absolute bias of 32.4 mm/yr for large natural rivers and 42.0 mm/yr for large mixed-impact rivers (Fig. 1a, b, detailed values in Supplementary Tables S1 and S2). These values are more than 20% lower than the biases of GWM4 (62.4 and 52.2 mm/yr for natural and mixed-impact rivers, respectively) and 55% lower than those of GWM5 (75.4 and 93.1 mm/yr). δHBV2 has only 3 out of the 61 major-river basins (Congo, Xingu, Paraguai) with absolute biases over 100 mm/yr, which is much fewer than GWM4 (13), GWM5 (17), and GWM1 (42). δHBV2's high spatial R^2 values between MAR simulations and observations for mixed-impact (0.92) and natural (0.97) basins means most of the spatial variability in large-river MAR is explained by the model (Fig. 1d, e). The interannual variability in streamflow is also captured well by δHBV2, which achieves the lowest median annual-scale root-mean-square error (RMSE) values of 34.5 mm/yr and 39.8 mm/yr for mixed-impact and natural basins—around 40% lower than GWM4 (Fig. 1a, b). Importantly, δHBV2 is also the only model that achieved $R^2 > 0.4$ in describing the spatial variability of the temporal trends of the streamflow-to-precipitation ratio (Q/P, Fig. 1c, with detailed performance information in Supplementary Fig. S2) for the natural rivers. The most challenging trends are those in Africa, e.g., the Niger (GRDC 1834101) and Congo (GRDC 1147010) rivers, where the model could not capture the rising trend in Q/P due to a paucity of training data. Established GWMs tend to have R^2 values lower than 0.13 and exhibit large scattering in the estimated Q/P trends around the observed value, since such spatial heterogeneity in response patterns is challenging to grasp. Because these 61 large river stations were not used in training the model, δHBV2's high performance is due to the collective knowledge gained from the numerous small basins used for big-data training.

With refreshingly high accuracy and resolution, δHBV2 reveals significant trends in annual green-blue-water partitioning for many regions over the last 20 years (trends in Fig. 1d; long-term averages in Supplementary Fig. S3). Here we demonstrate the partitioning using evapotranspiration-to-precipitation ratio (ET/P) with local runoff-to-precipitation ratio Q'/P provided in Supplementary Fig. S4. It should be noted that local Q'/P and ET/P may not sum to one each year due to storage effects. Blue water returns to the ocean in rivers, while green water is tied to plant water use and carbon/energy/nutrient cycles and exits as ET. Thus, ET/P reflects the most fundamental partitioning of

the terrestrial water cycle. In general, midlatitudes in North America and Asia and tropical areas like Central America and Papua New Guinea have seen decreasing green water fluxes while Central Europe and subtropical and midlatitude regions in South America have seen them increasing. Some of these shifts are substantial—with a 1% change in this ratio per year, some regions have thus shifted 20% over the course of 20 years. These changes are correlated with trends in precipitation (Supplementary Fig. S5): where precipitation increases, blue water tends to increase, and vice versa, although the patterns do not fully match. This suggests that large-scale climate shifts affect water partitioning, and increasing rainfall can overflow storage thresholds to increase blue water[4].

The shifts in local baseflow-to-runoff ratio (baseflow/Q'; trends in Fig. 1e, long-term averages in Supplementary Fig. S6, basin-scale data in Supplementary Fig. S7) have overlaps with the blue-green-water shifts, but are significantly more widespread. Due to the important role of groundwater discussed previously, these shifts imply pervasive changes in stream temperature and water quality characteristics at the decadal scale. Thus, the baseflow ratio should not be treated as static, which is currently the standard practice[54]. This ratio also exhibits regional clustering that has not been noted before—basins within a large region tend to have similar shifts, presumably reflecting decadal-scale trends in the regional climate. The two ratios move correspondingly because both reflect increases in runoff and decreases in infiltrated or land-retained water. The processes of groundwater recharge and ET then compete for infiltrated water, as water exceeding the soil's water-holding capacity moves to replenish deeper moisture and groundwater, which later becomes baseflow. ET/P changes are more muted than those of baseflow/Q', e.g., regions like India show noticeably rising baseflow/Q' but little change in ET/P. It can be explained that changes in precipitation and infiltration leave large imprints on recharge and thus baseflow, while the magnitude of ET responses may be limited by the soil's ability to hold water. On a side note, the baseflow ratios in Supplementary Fig. S6 are noticeably higher than those from gage-based separation methods[55] due to conceptual and scale differences, and our simulated baseflow behavior is very similar to the observations if the same analysis method is applied (Supplementary Text S1 and Supplementary Fig. S7).

These shifts contribute to water excess and scarcity. Where blue-water fraction increases and baseflow ratio decreases in tandem (Fig. 1d, e, also see Supplementary Fig. S8), e.g., northeastern China, mid-latitude North America, and Papua New Guinea, there is a higher flooding potential, which has been documented in some studies[56,57]. Where model-diagnosed MERIT-basin-scale ET/P increases substantially, it often results from precipitation declines. In fact, we find a negative correlation between changes in ET/P and precipitation in space and time (Supplementary Fig. S8). When annual precipitation declines, there is less excess water that can overcome the storage thresholds to become runoff, so more precipitated water exits as green-water fluxes, and blue water drops disproportionately. The prominent shifts toward green-water fluxes which occur in Germany, central Siberia, southern Brazil, central Chile, the Congo basin, and northern Australia are due to declining precipitation as discussed above (Supplementary Fig. S5) and suggest a disproportionate decrease in streamflow available for human use in these regions. These depicted shifts are mainly climate-driven, because while land use and human water use could have contributed to the shifts in water balance, these two processes are not explicitly simulated by the current version of the model (see "Limitations"). Because the baseflow process and hydrologic processes are scale-dependent, these fine-grained insights about baseflow and runoff need to be obtained from a data-trained high-resolution model like this one.

As a direct consequence of the shifts, we witness statistically significant trends in freshwater inputs to estuaries over the last two decades, with significant implications for ecosystems. The data-trained
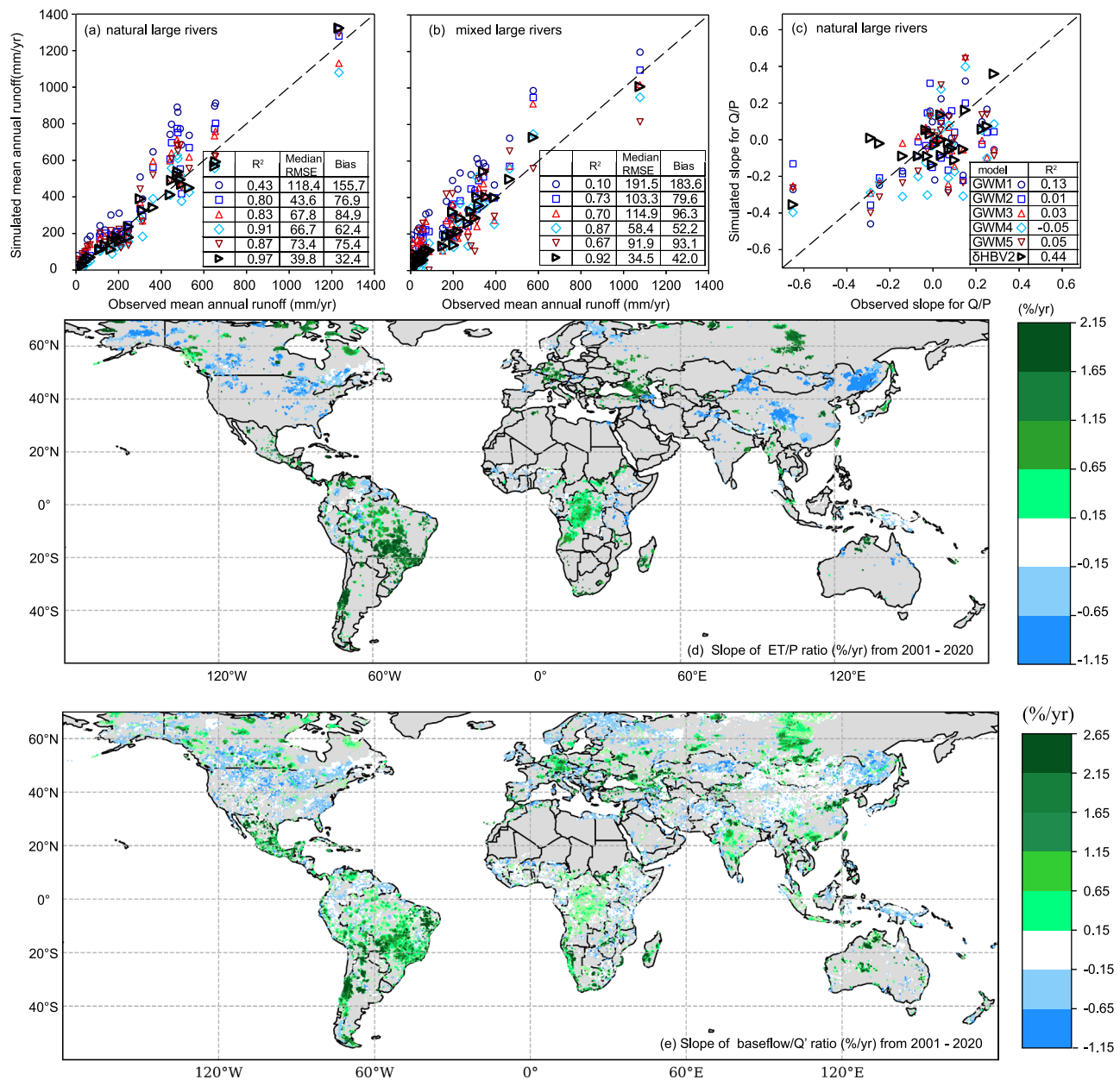
Fig. 1 | Global trends in water partitioning. Simulated versus observed mean annual runoff (MAR) for large rivers with natural (**a**) and mixed (**b**) anthropogenic influences from 1981–2000. Insets report R^2, bias, and root-mean-square error (RMSE, calculated by taking the median of each river's interannual RMSE) for each model. All models conserve mass and apply no post-simulation bias correction. **c** Simulated vs. observed trends in annual runoff-to-precipitation ratio (Q/P) for natural rivers, with each symbol representing one river simulation from one model. **d**, **e** Spatial trends in annual evapotranspiration-to-precipitation (ET/P) and local baseflow-to-runoff (baseflow/Q') ratios from 2001–2020, shown as percent change per year in basins with statistically significant trends (Mann–Kendall test, $p < 0.05$ colored). Some well-simulated basins, e.g., Ganges and Orinoco, were not represented in (**c**) as their records are too short (<10 years) to estimate the trend, and they are expected to elevate the R^2 if data were available. While δHBV2 was trained and validated against observations up to 2015, it was used to simulate water components through 2020. We removed parts of Siberia from the simulation due to having no training data there.

δHBV2 model identifies 10 stations flowing into estuaries out of the 55 analyzed as having significant declining trends in mean annual freshwater inflows from 2001 to 2020, mostly along the North Sea coast of Germany and France (Fig. 2b), which overlap with the increasing ET/P and baseflow/Q' ratios. The declining trends are substantial—a few German sites decline more than 1.5% per year, amounting to a 30% decline in 20 years. The changes in ET/P ratio (Fig. 1d) in Europe contributed to this decline. In contrast, US Mid-Atlantic estuaries have an increasing trend from 2001 to 2020. The trends for some of the mildly increasing/decreasing stations in the USA and France are not statistically significant (shown with thin marker borders in Fig. 2b), but the regional clustering of such trends is clear. Freshwater declines can

increase salinity and turbidity, which can drastically alter macrofaunal communities including fish as well as invertebrates like crustaceans, insects, and bivalves[58,59]. Combined with global sea level rise[60], large changes in estuarine salinity and macrofauna are expected. Consistent with this expectation, it was reported that macrofauna abundance in German estuaries decreased by around 31% and the total biomass decreased by around 45% for approximately this same period[61]. In that study, the changes were attributed entirely to sea level rise, but the decreasing freshwater inputs revealed here could have also played an important yet unacknowledged role.

Perhaps partly attributable to model limitations, the above-mentioned declines in freshwater inputs to European estuaries
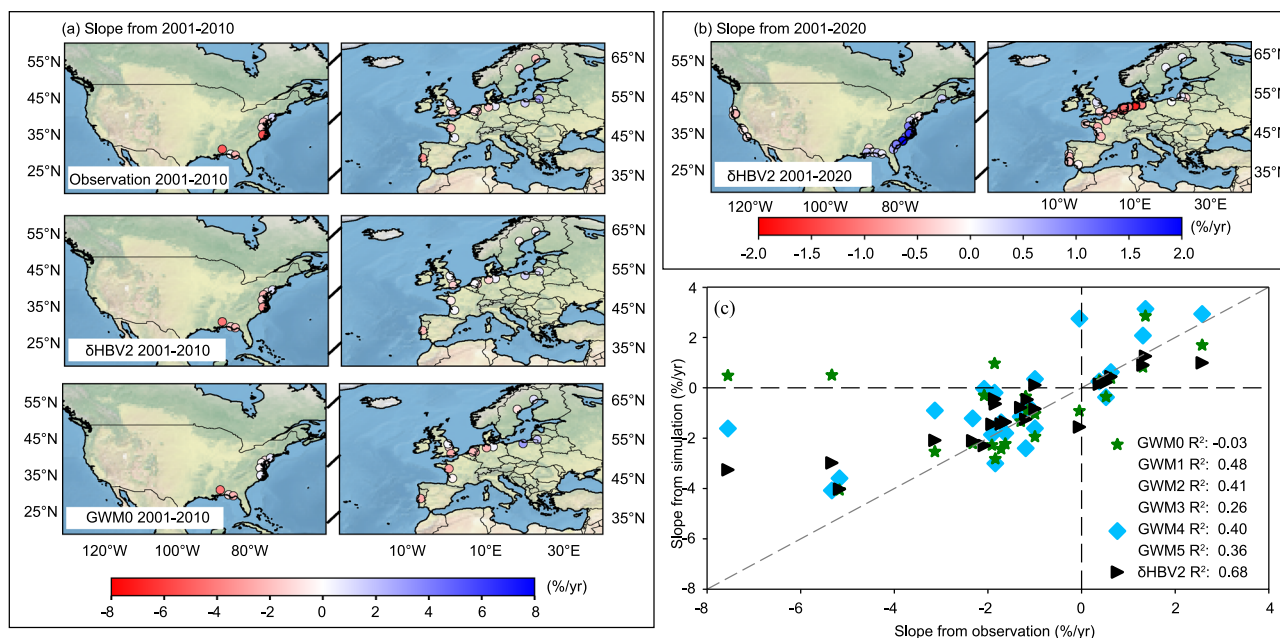
Fig. 2 | Estuary inflow trends and simulations by global water models (GWMs). **a** Simulated estuary inflow trends from δHBV2 and GWM0 compared to observed decadal-scale trends over the period of 2001 to 2010. **b** Simulated estuary inflow trends for δHBV2 from 2001 to 2020. **c** Scatter maps of observed estuary inflow trends and simulated trends from δHBV2 and GWMs from 2001 to 2010. Each point represents a station flowing into one estuary, with the symbol shape and color indicating the corresponding model (we only show GWM0 and GWM4 as examples). The R^2 value for each model is indicated in the legend.

(Fig. 2c) have not previously been reported. δHBV2 describes noticeably different trends than established GWMs for these estuaries, as shown at sites where a comparison is possible (Fig. 2a). We found a stronger match of δHBV2 predictions with observed trends (reaching R^2 values around 0.68) compared to the GWMs (GWM0: −0.05, GWM4: 0.46, with other models ranging between them). GWM4, for example, overestimated the freshwater declines of European sites in 2001–2010 while not showing any trend for the US sites. Freshwater inputs show clear decadal-scale oscillations, as US Atlantic estuaries have notable declines from 2001 to 2010 and overall rising trends from 2001 to 2020 (Fig. 2a, top row), but GWM0 largely misses the downward swing (Fig. 2a, bottom row). Since δHBV2's conceptual model backbone is not fundamentally different from other models, the parameterization through big-data training appears to have greatly improved the sensitivity to decadal-scale changes, allowing us to identify and predict these trends.

Seasonal streamflow response patterns

To support assessing models' seasonal rainfall-runoff responses (called elasticity), we evaluate GWMs' abilities to capture monthly flow fluctuations (Fig. 3a), monthly runoff autocorrelation (Fig. 3b) which relates to the recession behavior, and elasticity itself (Fig. 3c). For the natural rivers, δHBV2 ranks first among all models with a median correlation of 0.89 at the monthly scale (Fig. 3a), showing success at capturing the hydrological seasonality, although GWM0, which is enhanced by post-processing, achieves a slightly higher NSE. Unlike GMW0, δHBV2 does not impose a post-simulation bias correction and can thus ensure consistency for the internal hydrologic fluxes like ET, baseflow, and recharge, meaning these variables are also indirectly informed by the learning process and consistent with alternative estimates. The strong representation for autocorrelation (Fig. 3b for large natural rivers) suggests the model releases storage-dependent baseflow with the correct timing after conditioning by data, which is challenging for established GWMs. The established GWMs show substantial scattering in the arid regions for major rivers (high range of elasticity in Fig. 3c), while the data-trained model stays closer to the 1-to-1 line. Overall, in all categories of the evaluated hydrologic

signatures (one value per basin) related to rainfall responses, δHBV2 scores the highest except in ACF(1) (autocorrelation function with 1 month lag), where it places third with a margin of 0.01 (Supplementary Table S3). The difference in winter elasticity is smaller than in summer, presumably because the response in the wetter regime is faster and more linear.

When examined regionally, δHBV2 tends to perform well in boreal or northern midlatitude rivers but tends to be challenged in river basins where human water use is significant, e.g., Rio Grande de Santiago in the "northern dry" category, Cooper Creek in "southern dry" and the Columbia River, for which a large fraction of the catchment is arid (Fig. 4). Other rivers with significant reservoirs, e.g., Danube and Missouri, can also be challenging. Among the 33 mixed-impact rivers, 19 rivers have at least one GWM (excluding GWM0—see discussion above) with a monthly NSE of 0.2 or above, and are regarded as meaningful for benchmarking (Fig. 4a). Excluding GWM0, δHBV2 has the highest monthly NSE (median -0.77) for 15 of these 19 rivers (Fig. 4a).

The high resolution of δHBV2 enables the diagnosis of seasonal local runoff sensitivity to precipitation inputs (elasticity, ε), showing contrasting summer and winter ε values that mostly complement each other (Fig. 3d, e), with large summer ε values in arid and semi-arid regions[62]. The precipitation that generates the highest seasonal ε has an aggregation length of 6 months for summer and 3 months for winter. With the exception of central-western North America where there is some overlap, the high values for summer and winter ε tend to be staggered (clustered in different regions), identifying these regions as being "summer responsive" or "winter responsive". Summer ε is the largest in arid and semi-arid regions, e.g., Sahel, central and south Africa, central and southern South America, central Asia, and northern China and Siberia (Fig. 3d). In these regions, summer ε is high due to a low runoff baseline caused by high evapotranspiration and the dominance of storage-dependent groundwater releases driven by precipitation accumulated in previous months. Winter ε has a smaller range in values than summer ε, likely due to more linear hydrological responses once key thresholds—such as land abstraction—are exceeded, resulting in more proportional runoff responses to precipitation.
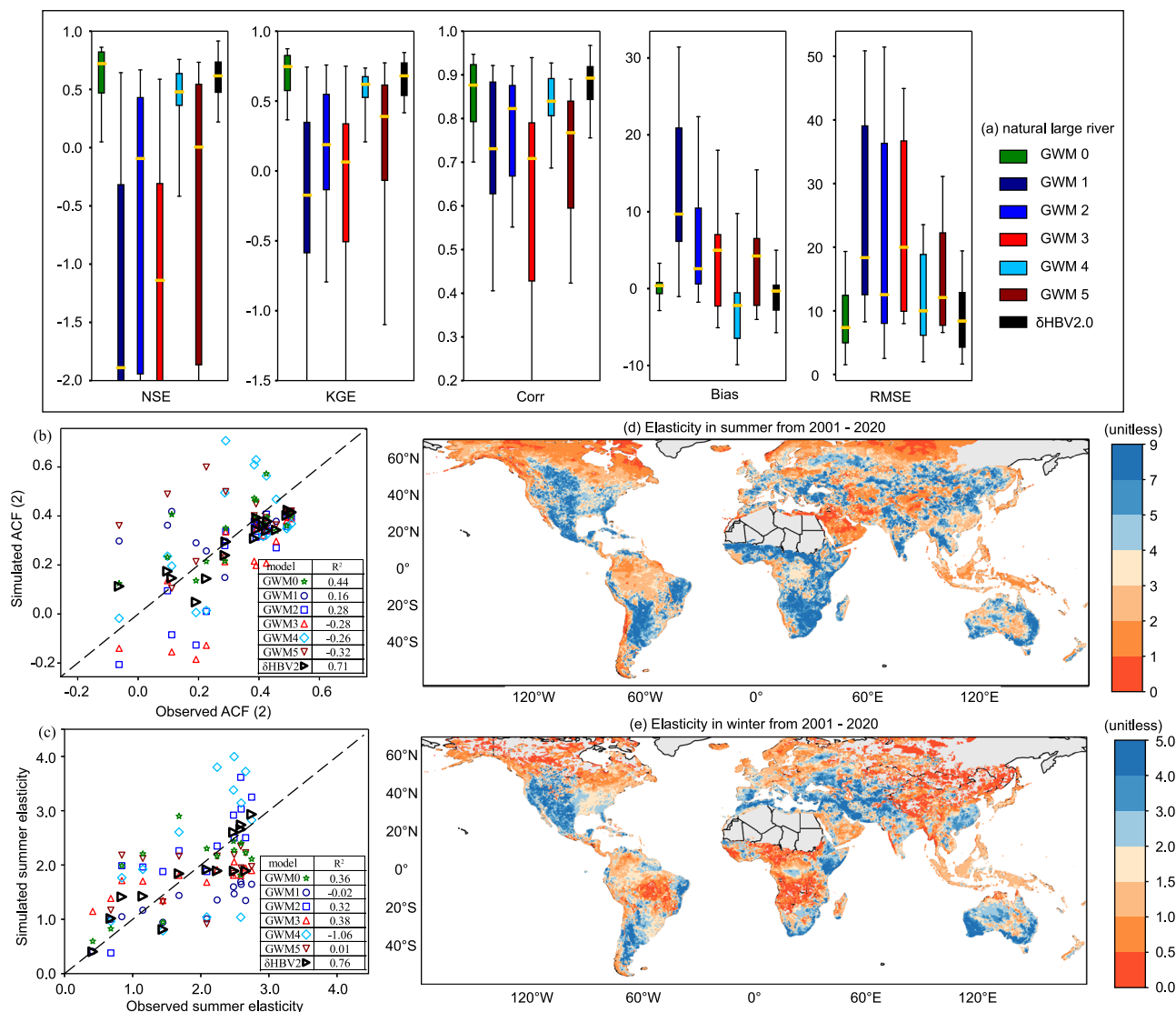
Fig. 3 | Monthly evaluations of GWMs and δHBV2-inferred seasonal elasticity. **a** Monthly performance metrics−Nash-Sutcliffe Efficiency (NSE), Kling-Gupta Efficiency (KGE), correlation (corr), bias, and root-mean-square error (RMSE)−for GWMs over large natural rivers. The top, center line, and bottom of each box plot indicate the 75th percentile (Q3), the median, and the 25th percentile (Q1), while the top and bottom whiskers indicate maximum and minimum values within 1.5 times the interquartile range (Q3−Q1) from the upper and lower quartiles, respectively. Boxes are arranged from left to right in the same order as the legend. **b** Simulated vs. observed autocorrelation at a 2-month lag (ACF(2)) from 1981−2000. **c** Simulated vs. observed summer streamflow elasticity to 6-month precipitation (1981−2010). Each point in **b**, **c** represents a model−observation pair for one river. **d**, **e** Spatial patterns of local runoff elasticity to precipitation in summer and winter (2001−2020). Only regions with available data are shown.

In addition, snow storage and gradual snowmelt could also reduce the runoff response to precipitation. Summer ε is highest in central Asia and the northern Middle East, western Australia, the northern and eastern African coasts, eastern Brazil, and southern Patagonia (Fig. 3e). In these regions, there is a more direct and immediate runoff response to winter rainfall. Such a refined understanding, to our knowledge, has not been offered before by a GWM.

For low-flow-dependent aquatic or riparian ecosystems, local seasonal ε is arguably more ecologically impactful than annual ε. High summer ε regions are vulnerable to precipitation changes, as reduction of seasonal precipitation there could have an outsized impact on summer low flows. Hydrologically, this is because the reduced rainfall would not be sufficient to exceed some storage thresholds, resulting in disproportionate reductions in streamflows (blue water). Low-flow-dependent aquatic ecosystems in these regions could thus be sensitive to long-term changes in the precipitation regime or season-long droughts. However, high summer ε also means stakeholders can use

seasonal outlooks of precipitation to reliably predict streamflow (and inflow to the downstream water bodies) in the coming summer and prepare to intervene if possible. The ε patterns appear noticeably different from some previous work that employed the Budyko curve for crude analysis[63], partly because here we analyze seasonal ε rather than annual ε, and partly because a data-trained, high-resolution hydrologic model is now available.

Daily streamflow variability and trends

To further validate the relevance of high-resolution δHBV2 for local water management, we compare its daily streamflow simulations in small-to-medium basins (<50,000 km²) with LSTM, lumped δHBV, and a widely-used operational global-scale product, GloFAS, as previous GWMs lack sufficient spatial resolution for this scale. We use several datasets (ds0-ds2, see "Methods") where comparisons are reasonable between different models. In a test spanning both training and testing basins, δHBV2 performs well, with a median daily NSE of 0.63 for all
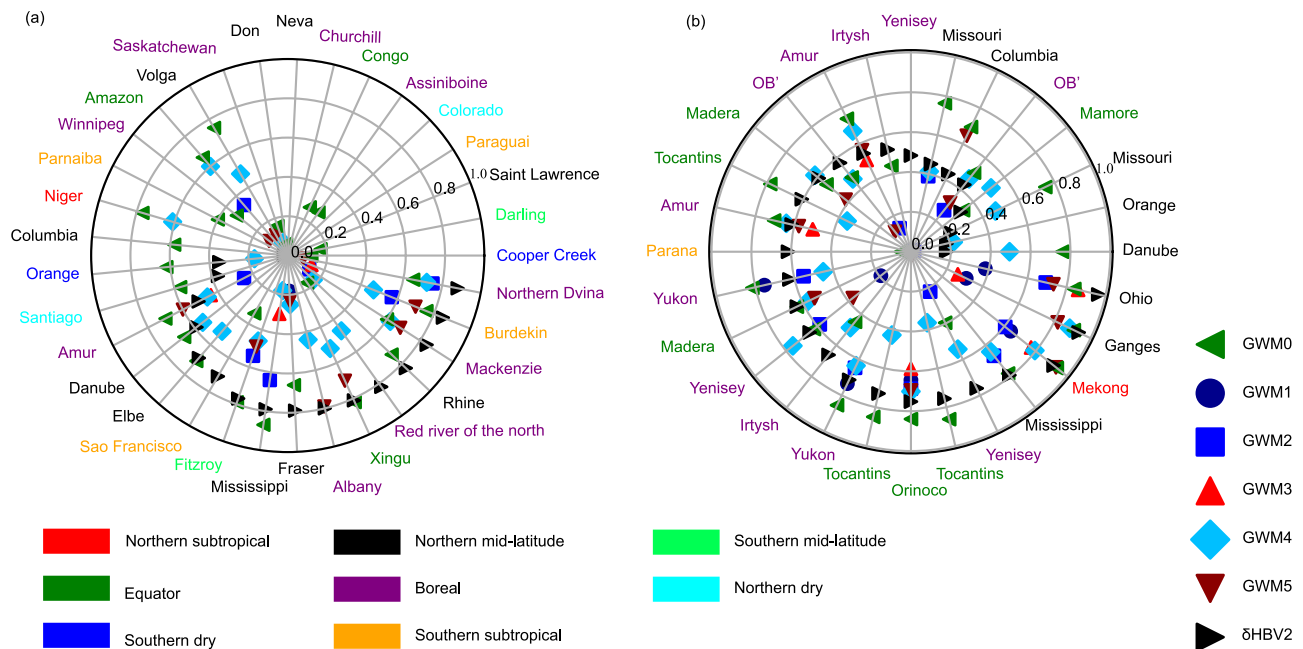
Fig. 4 | GWM performance across large global rivers. Monthly Nash-Sutcliffe Efficiency (NSE) scores (1981–2000) for global water models (GWMs) over **a** mixed-anthropogenic-impact and **b** natural basins. Rivers are arranged clockwise by δHBV2 performance (highest NSE on the outer ring, lowest at center) to show the best model over each basin. Marker shape and color represent different GWMs; river name color denotes biome type. The figure format and basin selection follow conventions from previous studies. GWM0 includes post-simulation bias correction; others do not. Amur, Columbia, Danube, Irtysh, Missouri, Madera, OB', Orange, Tocantins, and Yenisey are repeated, as the points in the right panel are taken from more upstream, natural gages of these rivers.

basins and 0.53 for ungauged basins, which is more than double Glo-FAS's median NSE of -0.26 (Fig. 5a). On another set of basins where comparison with GWMs is possible, δHBV2 noticeably leads GloFAS which, in turn, leads GWM0 (Supplementary Fig. S9). The strong performance of δHBV2 stems from its big-data training for para-meterization and generalization, which allows it to leverage data synergy[31]. Distributed model δHBV2 shows a clear improvement compared to the original lumped model δHBV1.0 with its ability to resolve high-resolution heterogeneity; it even edges out LSTM in terms of KGE, with a median of 0.74 vs. 0.73 for the higher-performing basins (Fig. 5b). This advantage against δHBV1.0 is more pronounced in data-rich regions, e.g., North America, where the subbasin-scale simulations can be better constrained by small training basins (Supplementary Table S4). There is still a slight gap between LSTM and δHBV2 in the lower half for daily KGE values (bottom left-hand side of Fig. 5b), likely due to systematic forcing biases and unrepresented heavy water uses in some basins[48]. Both δHBV2 and LSTM thus represent the state of the art for smaller basins, allowing them to accurately capture the flow variability. However, LSTM cannot simulate diagnostic variables like ET, baseflow, recharge, soil moisture, and snowmelt for providing narratives to stakeholders, and can suffer more when extrapolating to data-scarce regions and unseen extreme events[48]. Regardless, these results suggest that models with global coverage can finally be locally relevant for tasks such as flood forecasting and short-term water management.

The data-trained δHBV2 model showed large high-flow flashiness for arid and semi-arid regions and rather low flashiness for tropical rainforests with latitudes above 45 degrees (Fig. 5c). High-flow flashi-ness is quantified by the slopes of the flow duration curve between the 1st and 33rd exceedance percentiles (S_{FDC}) that are below −5 on the log scale, which corresponds to more than a 50 times difference between the 33rd and 1st percentile flows. Arid regions have the most negative slopes due to very low baseflows and prevalent quick, heavy storms. The eastern US, western and central Europe, and southern China have moderate S_{FDC} values of −3 to −1, as baseflow can be substantial. The

tropical rainforests, including the Amazon, Congo, and those throughout Pacific Asia, have the smallest S_{FDC}, since even during the dry season, streamflow in these regions can be significant. While these results are not surprising, they had not been shown at the global scale using a high-resolution model.

Examining the changes in flashiness over time, we find widespread and statistically-significant but spatially-mixed trends that do not easily match other spatial patterns studied here (Fig. 5d). S_{FDC} promi-nently increases and becomes less negative in Mexico, western South America, and northern India, indicating a less variable distribution of streamflows in the last 20 years. However, central-western USA, southern Africa, the south fringe of the Sahara Desert, central Asia, and Ethiopia have seen large declining and thus more negative S_{FDC} values, highlighting increasing streamflow variability. These changes are regional but can be substantial; we believe they may be caused by changes in precipitation intensities. Increasing flashiness poses the need for a greater storage capacity to ensure water supply resilience as well as control floods, although ecological consequences must also be considered. High resolution GWMs consistent with daily data are crucial for these applications.

Limitations

Besides water withdrawals and very cold or very dry climates, certain downstream riverine and hydrological processes including large nat-ural inland lakes, wetlands, and major reservoirs pose challenges for δHBV2. Large inland lakes and wetlands can store water, attenuate peak flows, sustain base flows, and induce floodplain water losses. They are not well simulated by the Muskingum-Cunge formulation. Examples include the Neva (Lake Ladoga), Paraguai (the Pantanal wetlands), Amazon (floodplains), Winnipeg (Lake Winnipeg), and Saint Lawrence (the Great Lakes) rivers. Due to the sizes of such storages, their impacts can even be noticeable at the monthly time scale. Addi-tionally, flood-control dams, hydroelectric dams, and dams for other purposes such as irrigation and recreation each operate with distinct objectives, further contributing to the complexity of their impacts[64].
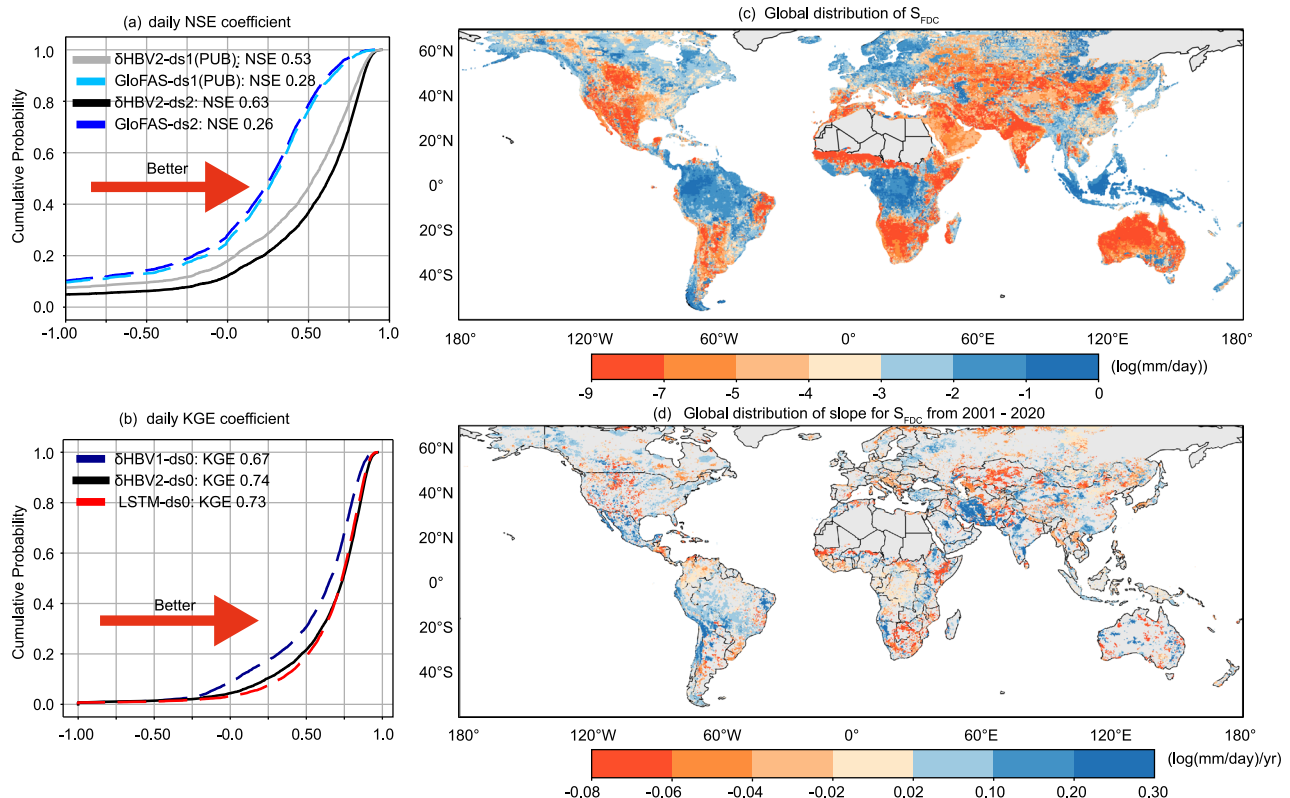
Fig. 5 | Flow duration curve (FDC) slopes and GWM performance comparisons. **a** Cumulative distribution function (CDF) plots comparing δHBV2 with a widely-used operational global-scale product, GloFAS, on 5558 basins (*ds2*) as well as for prediction in ungauged basins (PUB) on 2509 basins not used in model training (*ds1*). **b** CDF plots for temporal test results from δHBV1.0, δHBV2, and long short-term memory (LSTM, a purely data-driven model) on 4746 gages (ds0). *ds0-ds2* are different sets of basins for comparing models, as explained in the "Global datasets" section of the "Methods". **c** Global distribution map of Flow Duration Curve slope between 1% and 33% exceedance flow (S_{FDC}), used to indicate flashiness or the prevalence of sudden, heavy streamflows—a more negative slope indicates a high flashiness. **d** Global distribution map of the temporal change rate (slope) of S_{FDC}.

Combined with significant water withdrawals, it is challenging for the model to accurately capture such flow behaviors, which is only exacerbated in places with relatively few gages like Africa, e.g., the Niger and Congo rivers. Overall, due to structural limitations in both the HBV and Muskingum-Cunge models, δHBV2 has a limited capacity to represent anthropogenic impacts such as reservoirs, land use changes, and human water uses, although physical parameters learned from observational data can partially compensate for some of the resulting errors.

It is noteworthy that in our training and forward simulations, land-use inputs are regarded as static, so the model's long-term output shifts mainly reflect changes in the climate inputs. It is possible that the dynamic parameters produced by the NNs jointly trained on stream-flow data can capture some interannual covariation of vegetation-related characteristics due to climate shifts, but we do not expect this to be a major effect. The impacts of land use change, e.g., in central Asia and Ethiopia, are not explicitly simulated, which should be considered in future efforts.

While we do not have an ensemble of models to produce a formal uncertainty estimate, we present maps of large basin and upstream-catchment evaluations (Supplementary Fig. S2) as a gauge of model reliability in different regions at different scales. Some large basins, e.g., Ganges, Siberian, and southern Brazilian ones, are well simulated in terms of both NSE and long-term trends, despite not having any training basins. This suggests the model does possess some ability to generalize in space. Nevertheless, some poor-performing basins still tend to be due to a lack of training stations, especially African and north-central Asian ones, where hydrologic dynamics may also be systematically different from the training basins. These regions are expected to improve when

we learn from additional data, including expanded training stations, new streamflow estimates from the Surface Water and Ocean Topography (SWOT) mission[65], and non-streamflow observations such as soil moisture. Future efforts can also leverage an ensemble based on different training data to assess uncertainty.

Summary
Driven by climatic shifts, the terrestrial water cycle is undergoing significant changes in quantity, timing, and hydrologic response patterns. Using our high-resolution, high-accuracy, physics-embedded model, we identified coherent and widespread shifts between 2001 and 2020 in fundamental water partitioning—between evapotranspiration and runoff as well as between surface and subsurface runoff. These changes are primarily nonlinear responses to precipitation variations. For example, North America and parts of Asia have seen increased blue-water and surface runoff fractions leading to higher flood risks, while the Southern Hemisphere, tropical rainforests, and parts of Europe have experienced increases in green-water fluxes (i.e., evapotranspiration) and baseflow fractions leading to reduced river flows. Consequently, freshwater inputs into some European estuaries have declined markedly, with associated ecological impacts. Arid and semi-arid regions already show high flow variability and runoff elasticity, making them especially vulnerable to future shifts in seasonal precipitation patterns. Some arid regions also see precipitation pattern changes leading to even higher flow variability. We provide a high-resolution map of the response patterns and changes to identify future challenges in water supplies and aquatic ecosystems.

While understanding hydrologic response patterns is helpful for water management, analyzing them and their changes requires a

model that can accurately diagnose hydrologic fluxes with high spatiotemporal resolution. Traditional models often struggle to extract information from large datasets to characterize the landscape's hydrologic responses, manifesting in the difficulty to describe the spatial distributions of hydrologic signatures and their trends. Meanwhile, purely data-driven methods do not provide diagnostic variables or respect physical laws like mass conservation. Our approach—differentiable, physics-embedded learning—addresses both limitations. It offers a globally-consistent, high-resolution, physically-coherent picture of how hydrologic response patterns are shifting in response to the climate, enabling local-scale fine-grained analyses and decision-making while revealing many previously unrecognized changes described above. This modeling capability is essential for understanding and managing future water availability, aquatic ecosystem risks, and hydrologic resilience in a changing climate.

Methods
Global datasets
Differentiable models in this work were trained on daily streamflow observations from a recently-compiled global dataset by Abbas et al.[66], composed of various published resources as well as global and national databases indicated in Supplementary Table S5. The initial dataset, composed of approximately 34,000 catchments, was narrowed down to 4746 catchments under 50,000 km² with at least 95% of the observations available from 1980 to 2020. The data availability criterion was relaxed for data-sparse regions such as Africa, however. The catchment area selection was to ensure effective model training and was also partly due to limited computational resources. The meteorological forcings dataset includes the daily precipitation from Multi-Source Weighted-Ensemble Precipitation (MSWEP) V2.8[67] and maximum and minimum daily temperatures from Multi-Source Weather (MSWX) V1[68]; details are provided in Supplementary Table S6. Static attributes including topography, climate patterns, land cover, and soil and geological characteristics were derived from diverse sources and are also listed in Supplementary Table S6.

We used MERIT-Basins[69] as our hydrological simulation unit to build the distributed model and discretize the global river flow. This product delineates global flowlines into discrete reaches and associates each reach with its predefined drainage area based on the 90-m MERIT-Hydro[70] digital elevation model (DEM) dataset. The resolution of MERIT basins is much finer than the ~0.5 degree grid resolution of GWMs provided in the Inter-Sectoral Impact Model Intercomparison Project (ISIMIP). As a result, there is a discrepancy in the catchment area represented in the models, with MERIT having better topological correctness. Due to the relatively lower spatial resolution, GWMs were only benchmarked on large rivers at monthly or annual temporal scales to ensure a fair comparison. The first set of comparisons thus focus on annual and monthly evaluations on 33 large rivers benchmarked in the literature[52], which have major human influences including reservoirs and water withdrawals. To evaluate the models under more natural conditions, we further added 28 large rivers (sometimes using upstream gages of rivers among the first 33) as a second comparison set.

To benchmark δHBV2 against δHBV1.0, LSTM, and GloFAS, we compiled a third comparison set of thousands of smaller gages. GloFAS may still have catchment-area discrepancy with our data-driven models, so we restricted the comparison with GloFAS to gages where the catchment-area discrepancy between all models and the gage-registered area was less than 20%. To separately demonstrate the models' temporal, spatial, and overall generalizability, we used three small-gage datasets (*ds0-ds2*). *ds0* contains 4746 training basins which were also used for the temporal test. *ds1* contains 2509 basins used to test prediction in ungauged basins (PUB); none of these basins were used in the training of δHBV2 or the calibration of GloFAS. *ds2* contains all 5558 small basins from *ds0* and *ds1*, excluding those with a large

area discrepancy to enable a fair comparison between δHBV2 and GloFAS.

Models
As an overview, we used two differentiable models for global streamflow prediction in this work. Both models used neural networks (NNs) to generate physical parameters for the process-based Hydrologiska Byråns Vattenbalansavdelning (HBV) models[51] for hydrologic simulation. The first differentiable model is a basin-lumped model (δHBV1.0) with inputs and parameters defined for the entire catchment upstream of each prediction gage[42,47]. The second model is a high-resolution model defined on MERIT unit basins, with a differentiable implementation of Muskingum-Cunge (δMC) as the routing method. This model is called δHBV2δMC2-Globe2-hydroDL, or δHBV2 for short. The lumped δHBV1.0 can only provide a point prediction at a basin's outlet, while the high-resolution δHBV2 can provide hydrologic predictions across the entire seamless MERIT river network. δHBV2 is compared to δHBV1.0, a pure machine learning model (LSTM), and other GWMs.

Basin-lumped differentiable model (δHBV1.0)
The differentiable model, δHBV1.0, uses basin-averaged (lumped) inputs to predict streamflow at the catchment outlet. An LSTM network is used to support regionalized parameterization of the HBV model. δHBV1.0 can be described succinctly as:

$$\theta_b = \text{LSTM}(x_b^{1:t}, A_b) \tag{1}$$

$$q_b^{1:t} = \text{HBV}(\theta_b, x_b^{1:t}) \tag{2}$$

where θ_b represents the physical parameters of HBV. A_b represents the static attributes for each basin, while x_b represents the meteorological forcings used to drive the HBV model, including precipitation, mean temperature, and potential evaporation. Components of A_b and x_b are listed in Supplementary Table S6. q_b is the simulated streamflow in units of mm/day and can be converted to Q_b in units of m³/s using the basin area. Subscript b denotes the lumped inputs and outputs for the particular basin, b.

The structure and hyperparameters of LSTM can be found in Supplementary Text S2 and S3. It is a sequence-to-sequence neural network capable of processing input time series and capturing both long-term and short-term tendencies through its memory cell states[32], then producing the target time series (e.g., physical parameters of HBV). While many other architectures have been attempted, so far it is still challenging to surpass LSTM for deterministic prediction tasks in hydrology[71]. It can generate either static or dynamic parameters—i.e., the whole time series can be used as daily-varying parameters or one day's specific value can be selected as a static parameter. In our work, we trained our model with a sequence of 365 days following a 365-day warmup period, and static parameters were derived from the last day of LSTM outputs.

The structure and physical parameters of the HBV model are provided in Supplementary Table S7. HBV can simulate snow accumulation and melt, soil moisture, evapotranspiration, and runoff generation. Both NNs and physical components of δHBV1.0 and δHBV2 are implemented on PyTorch[72], a Python library that supports automatic differentiation and highly parallel execution on Graphical Processing Units. The NN for parameterization and the re-implemented HBV equations are trained within a single pipeline to allow gradient calculation throughout the entire workflow as well as optimization via gradient descent (Eq. (3)). The overall framework can be trained in parallel over many basins to obtain a regionalized and generalized mapping from inputs to HBV parameters. Once trained, it can easily generate parameters for untrained basins. The goal of the training process is to minimize the loss function, which is based on the root-

mean-square error calculated between streamflow simulations and observations:

$$\text{Loss}_{\text{RMSE}} = (1.0 - \alpha)\sqrt{\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}(q_b^t - q_b^{*t})^2}{B * T}} + \alpha_l \sqrt{\frac{\sum_{b=1}^{B}\sum_{t=1}^{T}(\log(q_b) - \log(q_b^*))^2}{B * T}} \tag{3}$$

where B (100) is the total number of gage basins in a training batch, and T (365) is the number of the evaluated simulation days for the batch. The mean square error of the log-transformed streamflow is included in the loss function to better consider the low flows. α is a weight used to balance the performance trade-off between high and low flows, where a large α will emphasize low flows (Eq. (3)). We adopted the same weight as used in Feng et al.[42], which was 0.25.

High-resolution, multiscale differentiable model (δHBV2)

δHBV2 is a high-resolution, multiscale differentiable model which has been developed[50] to better capture the heterogeneity of large basins and address hydrologic scale discrepancies. The connected NNs take the inputs at the resolution of small MERIT unit basins (median basin area is 37 km²) to estimate HBV parameters and HBV runs hydrologic simulations at the same resolution. The unit-basin-level runoff is aggregated and routed to the gage basin outlet for comparison with observations. To summarize, an LSTM network and a multilayer perceptron (MLP) are used to generate dynamic (daily time-variant) and static HBV parameters, respectively:

$$\theta_{d,m}^{1:t} = \text{LSTM}(x_m^{1:t}, A_m) \tag{4}$$

$$\theta_{s,m} = \text{MLP}(A_m) \tag{5}$$

where x_m and A_m are the same meteorological forcings and geographic attributes used as inputs in δHBV1.0, where subscript m denotes the basin-averaged variable for MERIT unit basins, and subscripts s or d respectively denote static or daily-varying parameters. The daily-varying parameters, $\theta_{d,m}^t$, are to compensate for missing processes like vegetation dynamics, deep water storage, and return flow. To suppress overfitting, we only chose three parameters from HBV as daily-varying parameters: the shape coefficient of effective flow (β^t), the shape coefficient evapotranspiration (η^t), and a dynamic recession coefficient of fast flow (k_0^t). The definitions of these variables can be found in Supplementary Table S7. All other parameters, $\theta_{s,m}$, in HBV are assumed to be static in time. An MLP is employed to generate static parameters separately to reduce the computational utilization of the model, and its structure can be found in Supplementary Text S4. Because MERIT basins already have high spatial resolution, only 4 parallel HBV components were used in δHBV2—that is, each MERIT basin has 4 subbasin-scale components.

The HBV model is used to predict runoff for MERIT unit basins using dynamic and static parameters along with forcing inputs:

$$q_m^{1:t} = \text{HBV}\left(\theta_{d,m}^{1:t}, \theta_{s,m}, x_m^{1:t}\right) \tag{6}$$

where q_m^t represents the runoff of MERIT unit basin m at time step t.

The runoff of all MERIT unit basins upstream of the target training gage are summed to obtain the total amount of runoff, $q_b'^t$, generated in the gage basin (Eq. (7)), which is further routed to the gage basin outlet by an intrinsic unit hydrograph formula (Eqs. (8) and (9)):

$$q_b'^t = \sum_{m=1}^{M} q_m^t \tag{7}$$

$$q_b^t = \int_0^t \xi(s) q_b'^s(t-s)ds \tag{8}$$

$$\xi(s) = \frac{1}{\Gamma(\theta_{ra})\theta_{rb}^{\theta_{ra}}} s^{\theta_{ra}-1} e^{-\frac{1}{\theta_{rb}}} \tag{9}$$

where M is the number of MERIT unit basins within the drainage area of the gage. $\xi(s)$ is the gamma distribution-based unit hydrograph. θ_{ra} and θ_{rb} are the static routing parameters that describe the shape of the hydrograph, which is also predicted by the MLP network.

Following Song et al.[50], we employ a loss function based on normalized mean square error:

$$\text{Loss}_{\text{NMSE}} = \frac{1}{B}\sum_{b=1}^{B}\frac{\sum_{t=1}^{T}(q_b^t - q_b^{*t})^2}{(\sigma(q_b^{*t}) + \epsilon)} \tag{10}$$

where $\sigma(q_b^{*t})$ is the standard deviation of the observed runoff of basin b in the whole training time span, which can avoid overweighting large and/or wet basins in the training. ϵ is a small value used to avoid a zero denominator.

This model is multiscale, trained at a finer resolution (MERIT unit basin) and constrained at a coarser resolution (gage basin outlet). This design allows δHBV2 to resolve both spatial heterogeneity and rainfall-runoff nonlinearity in the forcings and attributes at the MERIT-basin resolution. Such a scale also allows nonlinear and threshold behaviors to better manifest. For example, concentrated rainfall in a small mountainous region can lead to substantial saturation and runoff, but if the same rainfall is spread across a large basin, then very little runoff would occur. Once trained, δHBV2 can simulate runoff for around 2.94 million MERIT unit basins worldwide and then routes the runoff using a differentiable Muskingum-Cunge model (δMC) described in the next section. Our highly efficient system completes a 10-year global rainfall−runoff simulation and routing in 7 hours and 10 days separately on a single NVIDIA A100 GPU, which are reduced to 2 hours and 2.5 days when using 4 GPUs.

Since the NNs in δHBV2 only provide the parameters, a mass balance between all hydrologic fluxes and states is strictly enforced throughout the model at the MERIT-level basins by the HBV modules and the routing scheme. Training the aggregated model's behavior on streamflow data at the gages will indirectly condition the model's behavior at the smaller MERIT basins. In addition, since there are gages of varying catchment sizes, with some basins <100 km², but a median area of around 600 km² and a maximum area of around 50,000 km², we can constrain the model's behavior at different spatial scales. We also emphasize that we neither need nor have ground truth data to directly supervise the neural network's outputs, i.e., the physical parameters. Gradient information is propagated backward from the loss function defined between simulated and observed streamflow, through the process-based model, to update the neural networks in this end-to-end framework.

Based on our earlier benchmarks, a model with the simple HBV structure and NN-based differentiable parameter learning can roughly equal the performance of locally-calibrated operational hydrologic models[42], achieving around a median NSE of 0.64 on the CAMELS dataset. However, its regionalized parameterization offers benefits in applications to ungauged basins and lowers parameter equifinality[73]. Adding multiple hydrologic response units to implicitly represent heterogeneity within each basin largely elevates the performance to around 0.71. Setting two or three parameters as dynamic and adding a capillary flux further elevates it to 0.75[48], which is nearly equivalent to state-of-the-art LSTM on small basins while performing better for extremes. The recent addition of multiscale learning and differentiable

Muskingum-Cunge routing further elevates performance in arid regions and major rivers.

Differentiable Muskingum-Cunge routing (δMC)

Both lumped- and high-resolution models mentioned above inherently use a unit hydrograph formula for channel routing during the training process. However, this approach can be cumbersome for producing streamflow at every river reach, as it requires basin delineation at each prediction point, which also introduces inaccuracies for large global rivers. To provide a streamflow product that seamlessly covers all river reaches in the MERIT network, we use an explicit routing model, differentiable Muskingum-Cunge (δMC), applied to the MERIT flowlines[49]. The Muskingum-Cunge (MC) scheme solves a continuity equation for a mass balance and a simplified form of the momentum equation by assuming a prismatic channel shape, and conveys the flow from upstream to downstream in the river network.

Similar to δHBV, δMC incorporates a Kolmogorov-Arnold Network (KAN)[74] to learn hydraulic parameters and channel characteristics used in MC. The model can be summarized as:

$$n_m, p_m, q_m = KAN_h(A_r) \quad (11)$$

where n_m is Manning's n, and p_m and q_m are channel geometric parameters of the river flowline corresponding to the mth MERIT basin. A_r are static attribute inputs of KAN_h listed in Supplementary Table S8. The KAN's structure can be found in Supplementary Text S5.

With the parameters from Eq. (11), we solve a discretized MC equation with a finite difference scheme:

$$Q_{t+1} = c_1 I_{t+1} + c_2 I_t + c_3 Q_t + c_4 q' \quad (12)$$

where I and Q respectively denote the inflow and outflow of a flowline with the units of m³/s. q' is the incremental inflow from the MERIT unit basin of the flowline, converted from q_m^{t+1} from the δHBV2 simulation. c_1, c_2, c_3, and c_4 are the Muskingum-Cunge coefficients calculated from the hydraulic parameters[49,75]. The Q_{t+1} for all flowlines in the river network is computed using a matrix composed of their Muskingum-Cunge coefficients. The training objective is to minimize the same loss function as in Eq. (10). Thus the parameters from the trained NN should be interpreted as effective values to maximize routing accuracy and are not necessarily the ground truth, although future work can incorporate additional datasets and constraints to improve physical realism, as in Chang et al.[76] and Al Mehedi et al.[77].

Pure machine learning model, LSTM

An LSTM network can be trained alongside process-based models to generate physical parameters in differentiable models, or trained independently to directly predict streamflow, in which case it is considered a pure machine learning model. The structure and hyperparameters of the LSTM for streamflow are the same as those of the LSTM used for parameter generation in Eqs. (1) and (4) (Supplementary Text S2 and S3). However, without the physical components between the LSTM and the loss function, the network is trained solely for streamflow prediction:

$$Q_b^{1:t} = LSTM(x_b^{1:t}, A_b) \quad (13)$$

where the inputs x_b and A_b are the same lumped inputs used in δHBV1.0. Q_b is the streamflow simulation at the gage, b.

The training of the LSTM is directly guided by the requirement to minimize errors between its outputs and streamflow observations, calculated using the loss function in Eq. (10). This model has demonstrated superior performance over traditional hydrological models in many studies[35,38,78] but still lacks interpretability of the simulated

processes and cannot provide internal variables. Here we used it as a high-performance benchmark model to test against differentiable models.

Model training and evaluation

We used 4746 small-gage basins (*dsO*), having areas ranging from 21 km² to 49,821 km² with a median area of 583 km², to train both versions of the differentiable models from 1980 to 2000. The distribution of the global training basins is shown in Supplementary Fig. S10. The hyperparameters and optimization configurations of the embedded NNs in the differentiable models are provided along with their structures in Supplementary Text S2–S5. The high-resolution δHBV2 is comprehensively evaluated to demonstrate its performance in major rivers and smaller basins through a fair comparison with GWMs, δHBV1.0, and LSTM. We first evaluated the performance of δHBV2 in capturing the water balance for both long-term and seasonal periods over 61 global large rivers (first and second comparison sets) from 1981 to 2000, in comparison with previous GWMs from phase 2a of ISIMIP2a[53]. The locations of the large rivers are provided in Supplementary Fig. S10. We further evaluated δHBV2 performance with small gages using *dsO-ds2* from the third comparison set, by comparing it with δHBV1.0, LSTM, and GloFAS[27] (a GWM from ECMWF's operational system). The evaluation on small gages includes temporal, spatial, and overall tests using *dsO*, *ds1*, and *ds2*, respectively. In the temporal test, δHBV2 is evaluated on its training basins, *dsO*, but tested over a different time span (from 2001 to 2015). In the spatial test, it is evaluated on basins it did not see in training, *ds1*, from 1980 to 2010. In the overall test, it is evaluated on all the basins in *ds2*, from 1980 to 2015.

Metrics like NSE and KGE that are commonly used by the community are used to evaluate model performance. Metric definitions can be found in Supplementary Table S9. To assess the impacts of past changes in precipitation, temperature, and water partitioning, we also identified estuaries in North America and Europe from a global estuary database developed by the Sea Around Us project[79]. The identified estuaries have at least one streamflow station upstream. We then evaluated the changes in freshwater inflow through the main stem river into these estuaries.

Global water models

Global water models (GWMs)—including global hydrological models, land surface models, and dynamic global vegetation models—simulate the terrestrial water cycle at the global scale. In this study, we compared our results with six established GWMs: WaterGAP2, DBH, H08, LPJmL, MATSIRO, and PCR-GLOBWB. The 0.5° simulations, forced by GSWP3 atmospheric data, were obtained from the ISMIP2a protocol. Readers are referred to Müller Schmied et al.[80] for more information about these models. Due to the resolution discrepancy, we only compared GWMs at major river outlets whose catchments are distinctly larger than the grid size of GWMs, and at monthly, annual, or decadal scales.

Among these models, GWM0 is unique in its calibration and two-stage bias-correction strategies. First, an area-based correction coefficient is uniformly applied to all grid cells within a basin to adjust runoff, aiming to match observed streamflow at the annual scale within a 1% margin. If this adjustment is insufficient, a second-stage correction is applied directly to the simulated streamflow at target gages, without further modifying the runoff. A more detailed description can be found in Supplementary Text S6.

Hydrologic signatures

Hydrologic signatures are important metrics or indices to describe the statistical and dynamical properties of hydrologic data, e.g., streamflow. In our work, we analyzed models' abilities to capture different hydrologic signatures including elasticity, slope of the flow duration curve, and auto-correlation function.

Elasticity

Following Zhang et al.[62], we suppose that the variability of streamflow at each time interval i is mainly influenced by the variability of precipitation, and thus the streamflow variability can be estimated using the equation:

$$dQ_i = \frac{\partial Q}{\partial P} dP_i \tag{14}$$

and can be further written as:

$$\frac{dQ_i}{Q} = \frac{\partial Q}{\partial P} \frac{P}{Q} \frac{dP_i}{P} = \frac{\partial Q/Q}{\partial P/P} \frac{dP_i}{P} = \varepsilon_{P_i} \frac{dP_i}{P} \tag{15}$$

The above equation assumes that precipitation influences runoff generation at its corresponding period, which is more appropriate for small-to-medium-size basins and does not account for the significant lag effect observed in large rivers.

In this work, we define the interval for precipitation as i and interval for streamflow as j, so the equation can be written as:

$$\frac{dQ_j}{Q} = \frac{\partial Q}{\partial P} \frac{P}{Q} \frac{dP_i}{P} = \frac{\partial Q/Q}{\partial P/P} \frac{dP_i}{P} = \varepsilon_{P_i} \frac{dP_i}{P} \tag{16}$$

where P_i and Q_j are the summed precipitation and streamflow data at the predefined aggregation intervals i and j. For winter elasticity, we consider December, January, and February as the precipitation and streamflow intervals for the Northern Hemisphere and June, July, and August for rivers located in the Southern Hemisphere. As the summer season retains more water compared to winter, we consider the lag effect of precipitation, and define March–August (6 months) as the precipitation interval and June–August as the streamflow response interval for the Northern Hemisphere, and use the opposite months for the Southern Hemisphere. We tested different window lengths, and 6 months gave the highest elasticity for summertime discharge. P and Q in the equation are the mean values for the cumulative precipitation and streamflow over all time intervals. ε_{P_i} can be interpreted as the sensitivity of streamflow to the variability of precipitation, and can be regarded as the slope of a linear regression line. In this regression, the x-axis represents the precipitation deviations from the mean (dP_i), normalized by the mean precipitation (P), while the y-axis represents the corresponding calculation for streamflow. To ensure the reliability and fairness of the elasticity calculation, we uniformly used GSWP3 as the precipitation dataset and applied the F-test, retaining only results with p values below 0.1 as in Zhang et al.[62]. To better understand the physical meaning of ε_{P_i}, we can assume that if the ε_{P_i} of one river is equal to 2, it means that a 1% increase in precipitation would lead to approximately a 2% increase in streamflow.

Slope of FDC

Flow duration curves (FDCs) are widely used to characterize streamflow variability. They are constructed by ranking the streamflow data in descending order and calculating the exceedance probability for each value. The slope of an FDC is normally computed over a specific percentile range (e.g., 1%–33%) to quantify the flow variability. In our work, the slope for daily simulation is defined as:

$$slope = \frac{Log_{10}(Q_{33}) - Log_{10}(Q_1)}{P_{33} - P_1} \tag{17}$$

where $P_{33} - P_1$ indicates the difference between exceedance probabilities at two points on the FDC, and is roughly equal to 0.32 when there is enough data. P represents the exceedance probability and can

be calculated as:

$$P_i = \frac{r}{N+1} \tag{18}$$

where r is the rank of streamflow value (starting from 1 for the highest flow) and N is the total number of observations. For the monthly FDC comparison (in Supplementary Table S3), the log-transformation was not adopted due to the more evenly-distributed monthly time series.

Auto-correlation function

The auto-correlation function (ACF) is normally used to describe the correlation between a time series and its lagged values across multiple timescales, and is expressed as:

$$ACF(k) = \frac{\sum_{t=1}^{n_t-k}(X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^{n_t-k}(X_t - \bar{X})^2} \tag{19}$$

where X_t is the streamflow value at time t, $\bar{X}$ is the mean streamflow value, and k is the lag time.

Data availability

A model simulation data repository is available at https://doi.org/10.5281/zenodo.17042358. All input data used in this work are publicly available. The estuary datasets were from Global Estuary Database and can be downloaded at https://www.pigma.org. The MERIT-Basins datasets can be downloaded at https://www.reachhydro.org/home/params/merit-basins. Simulations for other GWMs are available at https://www.isimip.org. Static geographic attributes were derived from diverse sources and are listed in Supplementary Table S5. MSWEP and MSWX forcing can be downloaded from www.gloh2o.org. The streamflow dataset was compiled and provided by Ather et al.[66], with the GRDC data being described by Abbas et al.[66] but downloaded directly from the GRDC at https://grdc.bafg.de/data/data_portal, as they do not permit redistribution of their data.

Code availability

The codes of δHBV1.0 and the multiscale δHBV2 (δHBV2δMC2-Globe2-hydroDL) are available at the repository https://doi.org/10.5281/zenodo.14827983.

References

1. Hirabayashi, Y. et al. Global flood risk under climate change. *Nat. Clim. Change* **3**, 816–821 (2013).
2. Kumar, S. et al. Terrestrial contribution to the heterogeneity in hydrological changes under global warming. *Water Resour. Res.* **52**, 3127–3142 (2016).
3. Zhang, X., Tang, Q., Liu, X., Leng, G. & Di, C. Nonlinearity of runoff response to global mean temperature change over major global river basins. *Geophys. Res. Lett.* **45**, 6109–6116 (2018).
4. McDonnell, J. J., Spence, C., Karran, D. J., van Meerveld, H. J. (Ilja) & Harman, C. J. Fill and-spill: a process description of runoff generation at the scale of the beholder. *Water Resour. Res.* **57**, e2020WR027514 (2021).
5. Van Loon, A. F. et al. Review article: Drought as a continuum—memory effects in interlinked hydrological, ecological, and social systems. *Nat. Hazards Earth Syst. Sci.* **24**, 3173–3205 (2024).
6. Sharma, A., Wasko, C. & Lettenmaier, D. P. If precipitation extremes are increasing, why aren't floods? *Water Resour. Res.* **54**, 8545–8551 (2018).
7. Chiang, F., Mazdiyasni, O. & AghaKouchak, A. Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nat. Commun.* **12**, 2754 (2021).

8. Alcamo, J. et al. Development and testing of the WaterGAP 2 global model of water use and availability. *Hydrol. Sci. J.* **48**, 317–337 (2003).

9. Tang, Q., Oki, T., Kanae, S. & Hu, H. Hydrological cycles change in the Yellow River Basin during the last half of the twentieth century. *J. Clim.* **21**, 1790–1806 (2008).

10. Hanasaki, N., Inuzuka, T., Kanae, S. & Oki, T. An estimation of global virtual water flow and sources of water withdrawal for major crops and livestock products using a global hydrological model. *J. Hydrol.* **384**, 232–244 (2010).

11. Bondeau, A. et al. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Glob. Change Biol.* **13**, 679–706 (2007).

12. Takata, K., Emori, S. & Watanabe, T. Development of the minimal advanced treatments of surface interaction and runoff. *Glob. Planet. Change* **38**, 209–222 (2003).

13. Van Beek, L. P. H., Wada, Y. & Bierkens, M. F. P. Global monthly water stress: 1. Water balance and water availability. *Water Resour. Res.* **47**, 2010WR009791 (2011).

14. Pörtner, H.-O. et al. *Climate Change 2022—Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, 2023).

15. Gudmundsson, L. et al. Globally observed trends in mean and extreme river flow attributed to climate change. *Science* **371**, 1159–1162 (2021).

16. Wada, Y., de Graaf, I. E. M. & van Beek, L. P. H. High-resolution modeling of human and climate impacts on global water resources. *J. Adv. Model. Earth Syst.* **8**, 735–763 (2016).

17. McMillan, H. K. A review of hydrologic signatures and their applications. *WIREs Water* **8**, e1499 (2021).

18. McMillan, H. K., Gnann, S. J. & Araki, R. Large scale evaluation of relationships between hydrologic signatures and processes. *Water Resour. Res.* **58**, e2021WR031751 (2022).

19. Scanes, E., Scanes, P. R. & Ross, P. M. Climate change rapidly warms and acidifies Australian estuaries. *Nat. Commun.* **11**, 1803 (2020).

20. Rodrigue, M., Magnan, M. & Boulianne, E. Stakeholders' influence on environmental strategy and performance indicators: a managerial perspective. *Manag. Account. Res.* **24**, 301–316 (2013).

21. Beck, H. E. et al. Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrol. Earth Syst. Sci.* **21**, 2881–2903 (2017).

22. Gädeke, A. et al. Performance evaluation of global hydrological models in six large Pan-Arctic watersheds. *Clim. Change* **163**, 1329–1351 (2020).

23. Gnann, S. et al. Functional relationships reveal differences in the water cycle representation of global water models. *Nat. Water* **1**, 1079–1090 (2023).

24. Blöschl, G. & Sivapalan, M. Scale issues in hydrological modelling: a review. *Hydrol. Process.* **9**, 251–290 (1995).

25. Song, Y. et al. Prominent impacts of hydrologic scaling laws on climate risks. Preprint at https://doi.org/10.21203/rs.3.rs-4584048/v1 (2024).

26. Moriasi, D. N. et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE.* https://doi.org/10.13031/2013.23153 (2007).

27. Hirpa, F. A. et al. Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *J. Hydrol.* **566**, 595–606 (2018).

28. Yang, Y. et al. Global reach-level 3-hourly river flood reanalysis (1980–2019). *Bull. Am. Meteorol. Soc.* **102**, E2086–E2105 (2021).

29. Beven, K. A manifesto for the equifinality thesis. *J. Hydrol.* **320**, 18–36 (2006).

30. Liu, Y. & Gupta, H. V. Uncertainty in hydrologic modeling: toward an integrated data assimilation framework. *Water Resour. Res.* **43**, 2006WR005756 (2007).

31. Fang, K., Kifer, D., Lawson, K., Feng, D. & Shen, C. The data synergy effects of time-series deep learning models in hydrology. *Water Resour. Res.* **58**, e2021WR029583 (2022).

32. Hochreiter, S. & Schmidhuber, J. Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).

33. Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems*, (eds Guyon, I. et al.) Vol. 30 (Curran Associates, Inc., 2017).

34. Chang, Z. et al. On the design fundamentals of diffusion models: a survey. *Pattern Recognit.* **157**, 109915 (2025).

35. Rahmani, F., Shen, C., Oliver, S., Lawson, K. & Appling, A. Deep learning approaches for improving prediction of daily stream temperature in data-scarce, unmonitored, and dammed basins. *Hydrol. Process.* **35**, e14400 (2021).

36. Liu, J., Hughes, D., Rahmani, F., Lawson, K. & Shen, C. Evaluating a global soil moisture dataset from a multitask model (GSM3 v1.0) with potential applications for crop threats. *Geosci. Model Dev.* **16**, 1553–1567 (2023).

37. Xiang, Z., Yan, J. & Demir, I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resour. Res.* **56**, e2019WR025326 (2020).

38. Feng, D., Fang, K. & Shen, C. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resour. Res.* **56**, e2019WR026793 (2020).

39. Zhong, L., Lei, H. & Yang, J. Development of a distributed physics-informed deep learning hydrological model for data-scarce regions. *Water Resour. Res.* **60**, e2023WR036333 (2024).

40. Wi, S. & Steinschneider, S. On the need for physical constraints in deep learning rainfall–runoff projections under climate change: a sensitivity analysis to warming and shifts in potential evapotranspiration. *Hydrol. Earth Syst. Sci.* **28**, 479–503 (2024).

41. Shen, C. et al. Differentiable modelling to unify machine learning and physical models for geosciences. *Nat. Rev. Earth Environ.* **4**, 552–567 (2023).

42. Feng, D., Liu, J., Lawson, K. & Shen, C. Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. *Water Resour. Res.* **58**, e2022WR032404 (2022).

43. Rahmani, F., Shen, C., Lawson, K., Feng, D. & Appling, A. Data Release: identifying structural priors in a hybrid differentiable model for stream water temperature modeling at 415 U.S. basin outlets, 2010-2016. U.S. Geological Survey data release. https://doi.org/10.5066/P9UDDHVD (2023).

44. Aboelyazeed, D. et al. A differentiable, physics-informed ecosystem modeling and learning framework for large-scale inverse problems: demonstration with photosynthesis simulations. *Biogeosciences* **20**, 2671–2692 (2023).

45. Song, Y. et al. When ancient numerical demons meet physics-informed machine learning: adjoint-based gradients for implicit differentiable modeling. *Hydrol. Earth Syst. Sci.* **28**, 3051–3077 (2024).

46. Wang, C. et al. Distributed hydrological modeling with physics-encoded deep learning: a general framework and its application in the Amazon. *Water Resour. Res.* **60**, e2023WR036170 (2024).

47. Feng, D. et al. Deep dive into hydrologic simulations at global scale: harnessing the power of deep learning and physics-informed differentiable models (δHBV-globe1.0-hydroDL). *Geosci. Model Dev.* **17**, 7181–7198 (2024).

48. Song, Y. et al. Physics-informed, differentiable hydrologic models for capturing unseen extreme events. Preprint at https://doi.org/10.22541/essoar.172304428.82707157/v2 (2025).

49. Bindas, T. et al. Improving river routing using a differentiable Muskingum-Cunge model and physics-informed machine learning. *Water Resour. Res.* **60**, e2023WR035337 (2024).

50. Song, Y., Bindas, T., Shen, C. & Ji, H. High-resolution national-scale water modeling is enhanced by multiscale differentiable physics-informed machine learning. *Water Resour. Res.* https://doi.org/10.1029/2024WR038928 (2025).

51. Seibert, J. Estimation of parameter uncertainty in the HBV model. *Hydrol. Res.* **28**, 247–262 (1997).

52. Zaherpour, J. et al. Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environ. Res. Lett.* **13**, 065015 (2018).

53. Gosling, S. N. et al. ISIMIP2a simulation data from the global water sector. ISIMIP Repository. https://doi.org/10.48364/ISIMIP.882536 (2023).

54. Gnann, S. J., Howden, N. J. K. & Woods, R. A. Hydrological signatures describing the translation of climate seasonality into streamflow seasonality. *Hydrol. Earth Syst. Sci.* **24**, 561–580 (2020).

55. Xie, J. et al. Majority of global river flow sustained by groundwater. *Nat. Geosci.* **17**, 770–777 (2024).

56. Mallakpour, I. & Villarini, G. The changing nature of flooding across the central United States. *Nat. Clim. Change* **5**, 250–254 (2015).

57. Wang, H., Wang, S., Shu, X., He, Y. & Huang, J. Increasing occurrence of sudden turns from drought to flood over China. *JGR Atmos.* **129**, e2023JD039974 (2024).

58. Prado, P., Caiola, N. & Ibáñez, C. Freshwater inflows and seasonal forcing strongly influence macrofaunal assemblages in Mediterranean coastal lagoons. *Estuar. Coast. Shelf Sci.* **147**, 68–77 (2014).

59. González-Ortegón, E. et al. Freshwater scarcity effects on the aquatic macrofauna of a European Mediterranean-climate estuary. *Sci. Total Environ.* **503–504**, 213–221 (2015).

60. The IMBIE Team. Mass balance of the Antarctic Ice Sheet from 1992 to 2017. *Nature* **558**, 219–222 (2018).

61. Singer, A. et al. Long-term response of coastal macrofauna communities to de-eutrophication and sea level rise mediated habitat changes (1980s versus 2018). *Front. Mar. Sci.* **9**, 963325 (2023).

62. Zhang, Y., Viglione, A. & Blöschl, G. Temporal scaling of streamflow elasticity to precipitation: a global analysis. *Water Resour. Res.* **58**, e2021WR030601 (2022).

63. Berghuijs, W. R., Larsen, J. R., Van Emmerik, T. H. M. & Woods, R. A. A global assessment of runoff sensitivity to changes in precipitation, potential evaporation, and other factors. *Water Resour. Res.* **53**, 8475–8486 (2017).

64. Ouyang, W. et al. Continental-scale streamflow modeling of basins with reservoirs: towards a coherent deep-learning-based strategy. *J. Hydrol.* **599**, 126455 (2021).

65. Andreadis, K. M. et al. A first look at river discharge estimation from SWOT satellite observations. *Geophys. Res. Lett.* **52**, e2024GL114185 (2025).

66. Abbas, A. et al. Comprehensive global assessment of 23 gridded precipitation datasets across 16,295 catchments using hydrological modeling. Preprint at https://doi.org/10.5194/egusphere-2024-4194 (2025).

67. Beck, H. E. et al. MSWEP V2 global 3-hourly 0.1° precipitation: methodology and quantitative assessment. *Bull. Am. Meteorol. Soc.* **100**, 473–500 (2019).

68. Beck, H. E. et al. MSWX: global 3-hourly 0.1° bias-corrected meteorological data including near real-time updates and forecast ensembles. *Bull. Am. Meteorol. Soc.* **103**, E710–E732 (2022).

69. Lin, P. et al. Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resour. Res.* **55**, 6499–6516 (2019).

70. Yamazaki, D. et al. MERIT hydro: a high-resolution global hydrography map based on latest topography dataset. *Water Resour. Res.* **55**, 5053–5073 (2019).

71. Liu, J., Bian, Y., Lawson, K. & Shen, C. Probing the limit of hydrologic predictability with the Transformer network. *J. Hydrol.* **637**, 131389 (2024).

72. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. in *Advances in Neural Information Processing Systems 32* (eds Wallach, H. et al.) 8024–8035 (Curran Associates, Inc., 2019).

73. Feng, D., Beck, H., Lawson, K. & Shen, C. The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. *Hydrol. Earth Syst. Sci.* **27**, 2357–2373 (2023).

74. Liu, Z. et al. KAN: Kolmogorov-Arnold networks. Preprint at https://doi.org/10.48550/ARXIV.2404.19756 (2024).

75. David, C. H. et al. River network routing on the NHDPlus dataset. *J. Hydrometeorol.* **12**, 913–934 (2011).

76. Chang, S. Y. et al. The geometry of flow: advancing predictions of river geometry with multi-model machine learning. *Water Resour. Res.* **60**, e2023WR036733 (2024).

77. Al Mehedi, M. A. et al. Spatiotemporal variability of channel roughness and its substantial impacts on flood modeling errors. *Earths Future* **12**, e2023EF004257 (2024).

78. Kratzert, F. et al. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrol. Earth Syst. Sci.* **23**, 5089–5110 (2019).

79. Zhang, J. et al. A global assessment of mangrove soil organic carbon sources and implications for blue carbon credit. *Nat. Commun.* **15**, 8994 (2024).

80. Müller Schmied, H. et al. Graphical representation of global water models. *Geosci. Model Dev.* **18**, 2409–2425 (2025).

Author contributions

C.S. conceived the study, wrote parts of the manuscript and edited the manuscript, H.J. performed experiments, ran the analysis and drafted the initial manuscript, Y.S. assisted with methodology, drafting the initial manuscript, and revision, T.B. and Y.W. assisted with methodology and editing the manuscript, J.L. and F.R. contributed to modeling and data processing methodology, A.A., M.P., Y.Y. and H.B. aided with dataset preparation, and K.L. and all other authors edited the manuscript.

Competing interests

C.S. and K.L. have financial interests in HydroSapient, Inc., a company which could potentially benefit from the results of this research. These interests have been reviewed by the University in accordance with its Individual Conflict of Interest policy, for the purpose of maintaining the objectivity and the integrity of research at The Pennsylvania State University. There are no other competing interests.

Additional information