Global Daily Discharge Estimation Based on Grid-Scale Long Short-Term Memory (LSTM) Model and River Routing

Yuan Yang[1], Dapeng Feng[2], Hylke E. Beck[3], Weiming Hu[4], Agniv Sengupta[4], Luca Delle Monache[1], Robert Hartman[5], Peirong Lin[6], Chaopeng Shen[7], and Ming Pan[1]

[1]University of California San Diego
[2]Stanford University
[3]King Abdullah University of Science and Technology
[4]University of California, San Diego
[5]Robert K. Hartman Consulting Services
[6]Peking University
[7]Pennsylvania State University

October 14, 2023

Abstract

Accurate global river discharge estimation is crucial for advancing our scientific understanding of the global water cycle and supporting various downstream applications. In recent years, data-driven machine learning models, particularly the Long Short-Term Memory (LSTM) model, have shown significant promise in estimating discharge. Despite this, the applicability of LSTM models for global river discharge estimation remains largely unexplored. In this study, we diverge from the conventional basin-lumped LSTM modeling in limited basins. For the first time, we apply an LSTM on a global 0.25° grid, coupling it with a river routing model to estimate river discharge for every river reach worldwide. We rigorously evaluate the performance over 5332 evaluation gauges globally for the period 2000-2020, separate from the training basins and period. The grid-scale LSTM model effectively captures the rainfall-runoff behavior, reproducing global river discharge with high accuracy and achieving a median Kling-Gupta Efficiency (KGE) of 0.563. It outperforms an extensively bias-corrected and calibrated benchmark simulation based on the Variable Infiltration Capacity (VIC) model, which achieved a median KGE of 0.466. Using the global grid-scale LSTM model, we develop an improved global reach-level daily discharge dataset spanning 1980 to 2020, named GRADES-hydroDL. This dataset is anticipated to be useful for a myriad of applications, including providing prior information for the Surface Water and Ocean Topography (SWOT) satellite mission. The dataset is openly available via Globus.

Hosted file

975488_0_art_file_11460659_s24dmg.docx available at https://authorea.com/users/673109/articles/671985-global-daily-discharge-estimation-based-on-grid-scale-long-short-term-memory-lstm-model-and-river-routing

Hosted file

975488_0_supp_11456081_s22fyr.docx available at https://authorea.com/users/673109/articles/671985-global-daily-discharge-estimation-based-on-grid-scale-long-short-term-memory-lstm-model-and-river-routing

Global Daily Discharge Estimation Based on Grid-Scale Long Short-Term Memory (LSTM) Model and River Routing

Yuan Yang[1*], Dapeng Feng[2,†], Hylke E. Beck[3], Weiming Hu[1,‡], Agniv Sengupta[1], Luca Delle Monache[1], Robert Hartman[4], Peirong Lin[5], Chaopeng Shen[2], and Ming Pan[1]

[1] Center for Western Weather and Water Extremes, Scripps Institution of Oceanography, University of California San Diego, CA, USA

[2] Civil and Environmental Engineering, Pennsylvania State University, PA, USA

[3] Physical Science and Engineering Division, King Abdullah University of Science and Technology, Kingdom of Saudi Arabia

[4] Robert K. Hartman Consulting Services, Roseville, CA, USA

[5] Institute of Remote Sensing and GIS, Peking University, China

[†] Now at: Department of Earth System Science, Stanford University, Stanford, CA, USA, Institute for Human-Centered Artificial Intelligence, Stanford University, Stanford, CA, USA

[‡] Now at: School of Integrated Sciences, James Madison University, VA, USA

*Corresponding author: Yuan Yang (yuanyangthu@gmail.com)

Key Points:

- For the first time, the global applicability of LSTM model at 0.25° grid-scale and river routing for discharge estimation is explored.
- Globally, the grid-scale LSTM model outperforms a calibrated and bias-corrected benchmark simulation based on the process-based VIC model.
- Using the LSTM model, we create a highly accurate daily global reach-level discharge dataset covering 1980 to 2020.

Abstract

Accurate global river discharge estimation is crucial for advancing our scientific understanding of the global water cycle and supporting various downstream applications. In recent years, data-driven machine learning models, particularly the Long Short-Term Memory (LSTM) model, have shown significant promise in estimating discharge. Despite this, the applicability of LSTM models for global river discharge estimation remains largely unexplored. In this study, we diverge from the conventional basin-lumped LSTM modeling in limited basins. For the first time, we apply an LSTM on a global 0.25° grid, coupling it with a river routing model to estimate river discharge for every river reach worldwide. We rigorously evaluate the performance over 5332 evaluation gauges globally for the period 2000-2020, separate from the training basins and period. The grid-scale LSTM model effectively captures the rainfall-runoff behavior, reproducing global river discharge with high accuracy and achieving a median Kling-Gupta Efficiency (KGE) of 0.563. It outperforms an extensively bias-corrected and calibrated benchmark simulation based on the Variable Infiltration Capacity (VIC) model, which achieved a median KGE of 0.466. Using the global grid-scale LSTM model, we develop an improved global reach-level daily discharge dataset spanning 1980 to 2020, named GRADES-hydroDL. This dataset is anticipated to be useful for a myriad of applications, including providing prior information for the Surface Water and Ocean Topography (SWOT) satellite mission. The dataset is openly available via Globus.

1 Introduction

River discharge, presenting the accumulation of surface water flowing into rivers and ultimately reaching the ocean or other water bodies, plays a crucial role in the global water cycle (Tuozzolo et al., 2019; Wada et al., 2017). Accurate global river discharge estimation is of vital importance across various fields, including water resources (S. Liu et al., 2020; Oki & Kanae, 2006), climate change (Gerten et al., 2008; Milly & Dunne, 2020; Trabucco et al., 2008), natural hazards (Coughlan de Perez et al., 2016; Yang et al., 2021), biodiversity (Ficke et al., 2007; Vörösmarty et al., 2010) and energy production (Chen et al., 2016; Xu et al., 2023). Normally, gauging stations are deemed the most reliable data source for measuring river discharge (Fekete et al., 2002; Zaitchik et al., 2010). However, a significant proportion of the world's rivers remain ungauged due to a combination of technical, economic, and political constraints (Gleason & Smith, 2014; Hannah et al., 2011; Riggs et al., 2023). Encouragingly, recent advancements in remote sensing (RS), exemplified by the Surface Water and Ocean Topography (SWOT) mission, the first satellite mission dedicated to discharge estimation, have opened new avenues for global river discharge monitoring, even in ungauged basins (Biancamaria et al., 2016; Bjerklie et al., 2018; Gleason & Durand, 2020; Yang et al., 2019). Nonetheless, the temporal coverage of satellite observations is largely limited to the recent two decades and RS-based discharge estimations hinge upon prior knowledge of river discharge to reduce uncertainties and improve accuracy (Durand et al., 2023; Tuozzolo et al., 2019).

As a result, considerable efforts have been made by the modeling community to estimate river discharge based on various rainfall-runoff models, which use meteorological data, such as precipitation and temperature, as inputs to predict the runoff or discharge. The existing rainfall-runoff modeling approaches, depending on the extent to which physical process knowledge is imposed in the simulation, can be categorized into fully data-driven and process-based approaches, and the latter further range from conceptual to physically based approaches.

Historically, thanks to the continuously improved understanding of hydrological processes, process-based models served as the preferred choice for discharge estimation. A way forward pioneered in the field of large-scale hydrology has been to utilize advanced process-based models, together with the optimal combination of in situ and satellite observations, as well as reanalysis, to reconstruct spatiotemporal seamlessly river discharge globally (Alfieri et al., 2020; Harrigan et al., 2020; Hersbach et al., 2020). For example, the European Commission's Copernicus Emergency Management Service (CEMS) Global Flood Awareness System (GloFAS; http://www.globalfloods.eu/) employs a coupled land surface model (the Hydrology Tiled ECMWF Scheme for Surface Exchanges over Land - HTESSEL) and flow routing model (LISFLOOD) to generate long-term river discharge at daily time steps and 0.1° grid resolution (Alfieri et al., 2020; Harrigan et al., 2020). GloFAS has undergone significant upgrades, with the latest version (GloFAS v4.0) featuring an enhanced resolution of 0.05° resolution, approximately four times higher than its predecessor. Using the Variable Infiltration Capacity (VIC) model (Liang et al., 1994, 1996) that is well calibrated and bias-corrected and the Routing Application for Parallel computatIon of Discharge (RAPID) river routing model (David et al., 2011), Lin et al. (2019) produced the first reach-level naturalized daily river discharge, the Global Reach-Level A Prior Discharge Estimates for SWOT (GRADES), over 2.94 million river reaches globally for 1979-2014. Building upon the GRADES legacy, Yang et al. (2021) made significant enhancements to spatial and temporal resolutions, coverage, as well as input data, and developed the global 3-hourly river discharge data record during the 40-year period of 1980-2019. This data record, referred to as Global Reach-Level Flood Reanalysis (GRFR), exhibits improved simulation capabilities, particularly for high extremes, and serves as a valuable resource for flood analysis.

Recently, the availability of extensive datasets and advancements in computing technologies have facilitated the development of numerous modern data-driven techniques, predominantly based on machine learning (ML). These ML models directly learn intricate non-linear response patterns from massive amounts of data, without requiring explicit knowledge of the underlying physical processes and strong structural assumptions (Feng et al., 2022; LeCun et al., 2015; Prasad et al., 2017; Schmidhuber, 2015; Shen, 2018; Shen et al., 2023). These advances have also drawn the attention of the hydrological community, inspiring new efforts to apply ML models to rainfall-runoff modeling. Among many efforts, a popular model is the Long Short-Term Memory (LSTM) neural network, a specifically designed version of recurrent neural network (RNN) for long-term sequential datasets (Greff et al., 2016; Hochreiter & Schmidhuber, 1997), which has garnered significant attention from hydrologists. With long-term memory, LSTM excels in capturing both periodic and chaotic behaviors within time-series data, as well as learning their long-range dependencies with higher accuracy (Fang et al., 2017; Hu et al., 2019; Mouatadid et al., 2019). This makes LSTM particularly suitable for hydrologic modeling. The LSTM model can outperform a baseline process-based model in simulating rainfall-runoff relations and demonstrated the feasibility of employing LSTM for this task, which then sparked a proliferation of research on LSTM-based rainfall-runoff modeling. Multiple researchers have demonstrated LSTM's seemingly incomparable performance in simulating runoff (Feng et al., 2020, 2021; Frame et al., 2022; Gauch et al., 2021; Konapala et al., 2020; Kratzert et al., 2021; Kratzert, Klotz, Shalev, et al., 2019; Lees et al., 2021; J. Liu, Bian, et al., 2023; Nearing et al., 2021; Reichstein et al., 2019; Sun et al., 2021). However, it is noteworthy that most of these applications focus on data-rich regions such as CONUS, and Great Britain, with regionally trained networks. Limited pilot studies attempted to explore the transferability of LSTM to other basins beyond the CONUS. For instance, Ma et al. (2021) demonstrated that transferring the LSTM model weights trained over the CONUS to other regions, e.g., China, Chile, and Great Britain, and moderately retraining the models using local data can greatly enhance the accuracy as compared to locally trained models. Recently, two studies implemented LSTM across the globe (Koya & Roy, 2023; Tang et al., 2023). However, they tested the performance only over training basins and the performance in other regions was not explored yet. Overall, these basin-lumped studies represent only limited investigations of specific regions, and are not directly applicable to global application. As discussed earlier, a globally trained LSTM model ready to be applied to any land surface on earth could serve as a powerful tool for global river discharge estimation. However, to the best of our knowledge, no such research endeavor has been conducted thus far, and the applicability of LSTM in global river discharge estimation remains unclear.

Therefore, diverging from the conventional basin-lumped LSTM modeling, we stepped forward to apply an LSTM on the grid scale and coupled it with a river routing model to estimate river discharge for every river reach worldwide. We aim to evaluate the effectiveness and performance of LSTM in comparison to process-based models for global river discharge estimations. The rest of this study is organized as follows. Section 2 provides a description of the methodology, including details of experiments, dataset, training basin selection, and model evaluation. Section 3 shows the global performance of LSTM and its strengths and weaknesses compared to the baseline process-based model. Based on the optimized LSTM experiment, a new global river discharge data record is introduced in Section 4. The study concludes in Section 5 with a summary of the main findings.

2 Methodology

2.1 Experimental design

To comprehensively explore the potential of LSTM models and evaluate their strengths and weaknesses compared to process-based hydrological models, we conducted a series of four experiments (Table 1). Note that instead of using the LSTM at the basin scale to do the rainfall-discharge modeling in the previous studies, here we applied the LSTM at 0.25° grid-scale to estimate global daily 0.25° runoff, which can then be routed to generate the global river discharge. We explain the four experiments as follows.

1. *VIC*. We employed the Variable Infiltration Capacity (VIC; Liang et al., 1994, 1996) land surface model for runoff modeling. To reduce model biases, we performed grid-level parameter calibration and bias correction (postprocessing) against ML-derived, global runoff characteristic maps from the Global Streamflow Characteristics Dataset (GSCD; Beck et al., 2015). More details about the VIC setup can be found in Yang et al. (2021). We aggregated the original 0.05°, 3-hourly runoff to 0.25°, daily. To our knowledge, this is one of the best global simulations achieved based on process-based models.

2. *LSTM(VIC)*. A single LSTM model was trained against 0.25° grid-scale VIC-modeled runoff derived from the above experiment. Then, this trained LSTM model was applied to all global 0.25° land surface grids to generate global runoff. To reduce the training burden as well as assess the generalization capability of LSTM in untrained regions, we implemented a $1/8^2$ sampling density, where $1/8^2$ represents sampling one grid-cell from each 8x8 patch, and resulted in 4153 grids for training. This experiment serves as a surrogate model to evaluate the ability of

LSTM models to reproduce the underlying processes exhibited by hydrological models (Shen et al., 2023; Tsai et al., 2021).

3. *LSTM(obs)*. We first trained a single LSTM model against discharge observations from selected training basins (see Section 2.3 for the details) and applied this trained LSTM model to global 0.25°grids. This is a classic ML strategy using observations as the training target. The evaluation of this experiment allows us to gain insights into the applicability of the LSTM model for global runoff estimation.

4. *LSTM(VIC+obs)*. A single LSTM model was first trained against VIC-modeled runoff (same as *LSTM(VIC)*), and then re-trained against discharge observations. The LSTM model trained twice was later applied to the global grid scale. This experiment aimed to investigate the potential benefits of incorporating hydrologic simulations generated by process-based models in improving the performance of the LSTM model.

Table 1. Experiments conducted in this study.

Experiment Name	Model	Training Data	Purpose
VIC	VIC	-	Benchmark
LSTM(VIC)	LSTM	VIC-modeled runoff from 4153 0.25° grids	Surrogate model
LSTM(obs)	LSTM	Discharge observations from 4144 basins	Classic ML
LSTM(VIC+obs)	LSTM	VIC-modeled runoff from 4153 0.25° grids + discharge observations from 4144 basins	Testing the added value of VIC

For each LSTM experiment, a single global LSTM model was trained using training period data from all training basins or grids so that the network can learn a more general understanding of the rainfall-runoff process. The LSTM models were trained using 20 years' worth of data from 1 January 1980 to 31 December 1999, and evaluated using another 21 years' worth of data from 1 January 2000 to 31 December 2020. The LSTM networks were trained on sequences of 365 days of six meteorological features and 10 static basin attributes (detailed in Table 2) to simulate the discharge at each time step. The objective function was the Root-Mean-Squared Error (RMSE), calculated on the transformed discharge (see Section 2.2 for more details about data pre-processing), aimed at improving low flow representation. The Adadelta algorithm (Zeiler, 2012) was used as the optimization method. Hyperparameter combinations (Table S1) from Feng et al. (2020) were utilized, and through a simple validation process, it was determined that these hyperparameters remained optimal for discharge estimation in the current study. A fast and flexible LSTM code that was capable of leveraging the optimized NVIDIA CUDA Deep Neural Network (cuDNN) library from the PyTorch Deep Learning platform was implemented. It took about 15 hours of computational time on an NVIDIA P100 Graphical Processing Unit (GPU) to train *LSTM(VIC)* and *LSTM(obs)* to convergence (300 epochs).

188 **Table 2.** Summary of the forcing and attribute variables used as the input to the LSTM model.

	Variable	Data Source	Units
Forcing	Daily mean precipitation	MSWEP V2.80 (Beck et al., 2019) (https://www.gloh2o.org/mswep/)	mm/d
	Daily maximum temperature	ERA5 (Hersbach et al., 2018) (https://cds.climate.copernicus.eu/cds app#!/dataset/reanalysis-era5-complete?tab=overview)	℃
	Daily minimum temperature		℃
	Daily mean surface downwelling shortwave		W/m^2
	Daily mean 10m wind		m/s
	Monthly LAI climatology	PROBAV VITO LAI (https://land.copernicus.eu/global/products/lai)	-
Attributes	Mean daily precipitation	MSWEP V2.80	mm/d
	High precipitation duration - the average duration of high precipitation events (number of consecutive days ≥ 5 times mean daily precipitation		days
	Fraction of precipitation falling as snow (i.e., on days colder than 0 ℃)	MSWEP V2.80 and ERA5	-
	Aridity - P/PET, where PET is estimated by the Hargreaves (1994) method		-
	Frozen days - days colder than 0 ℃	ERA5	days
	Area	basin boundary file	km^2
	Mean elevation	GMTED (Amatulli et al., 2018) (https://doi.pangaea.de/10.1594/PANGAEA.867115)	m above sea level
	Mean slope		°
	Geological permeability	GLHYMPS V2 (Huscroft et al., 2018) (https://borealisdata.ca/dataset.xhtml?persistentId=doi%3A10.5683/SP2/TTJNIU)	m^2
	Soil sand content	SoilGrids (Hengl et al., 2017) (https://soilgrids.org/)	%

2.2 Datasets

For the VIC model, we used the precipitation from Multi-Source Weighted-Ensemble Prediction (MSWEP) version 2.80 (Beck et al., 2019) and other meteorological fields (surface air temperature, pressure, incoming shortwave and longwave radiation, humidity, and wind speed) from ERA5 (Hersbach et al., 2018) as forcing inputs. MSWEP is a global dataset with high quality, which ingests a wide range of data sources (in situ gauges, satellite products, and reanalysis products), makes distributional bias corrections, as well as corrections of systematic terrestrial biases using river discharge observations. ERA5 is the latest climate reanalysis dataset produced by the European Centre for Medium-Range Weather Forecasts (ECWMF) and has been widely used in meteorological and hydrological applications.

As input features to the LSTM models, we adopted six meteorological variables, including precipitation from MSWEP V2.80, daily maximum and minimum temperatures, downwelling shortwave radiation, mean 10-m wind from ERA5, as well as PROBAV VITO's monthly leaf area index (LAI) climatology (Table 2). Kratzert, Klotz, Herrnegger, et al. (2019) and Kratzert, Klotz, Shalev, et al. (2019) have shown that including basin attributes can improve overall model performance since they contain information that helps to distinguish different basin-specific rainfall-runoff behaviors. Therefore, we calcualted the top 10 sensitive attributes according to Kratzert, Klotz, Shalev, et al. (2019), including climate, topography, and soil attributes (Table 2) as additional inputs to train the LSTM model. These attributes remained constant in time throughout the simulation (training and evaluation) and were directly concatenated with the forcings and provided as inputs.

Globally, we compiled daily discharge records for 19999 river gauges from multiple sources, including the United States Geological Survey (USGS) National Water Information System (NWIS), Global Runoff Data Centre (GRDC), European Water Archive (EWA) of EURO-FRIEND-Water, Water Survey of Canada Hydrometric Data, etc. (Beck et al., 2020). Figure 1a shows that North America and Europe have much higher gauge densities. It is worth mentioning that discharge records in Asia and Africa are only from GRDC gauges, most of which are located in large basins or provide only monthly data or data before 1995. This poses a challenge in terms of limited gauge availability for training and evaluation processes in these regions.

To reduce the differences between basins of varying sizes and wetness levels during the calculation of the loss function, we adopted the pre-processing procedures following Feng et al. (2020). First, we normalized the daily discharge by basin area and mean daily precipitation to get a dimensionless discharge value as the target variable. For *LSTM(VIC)*, the runoff was directly normalized by mean daily precipitation. Then we transformed the distributions of daily discharge and precipitation from Gamma to as close to Gaussian as possible by $v^* = log_{10}(\sqrt{v} + 0.1)$ (v and v^* are the variables before and after transformation, respectively). Finally, for efficient learning, all input features (meteorological variables and static basin attributes), as well as the output (discharge/runoff), were standardized to have zero mean and unit variance over all 4144 training basins (*LSTM(obs)*) /4153 training grids (*LSTM(VIC)*) collectively.

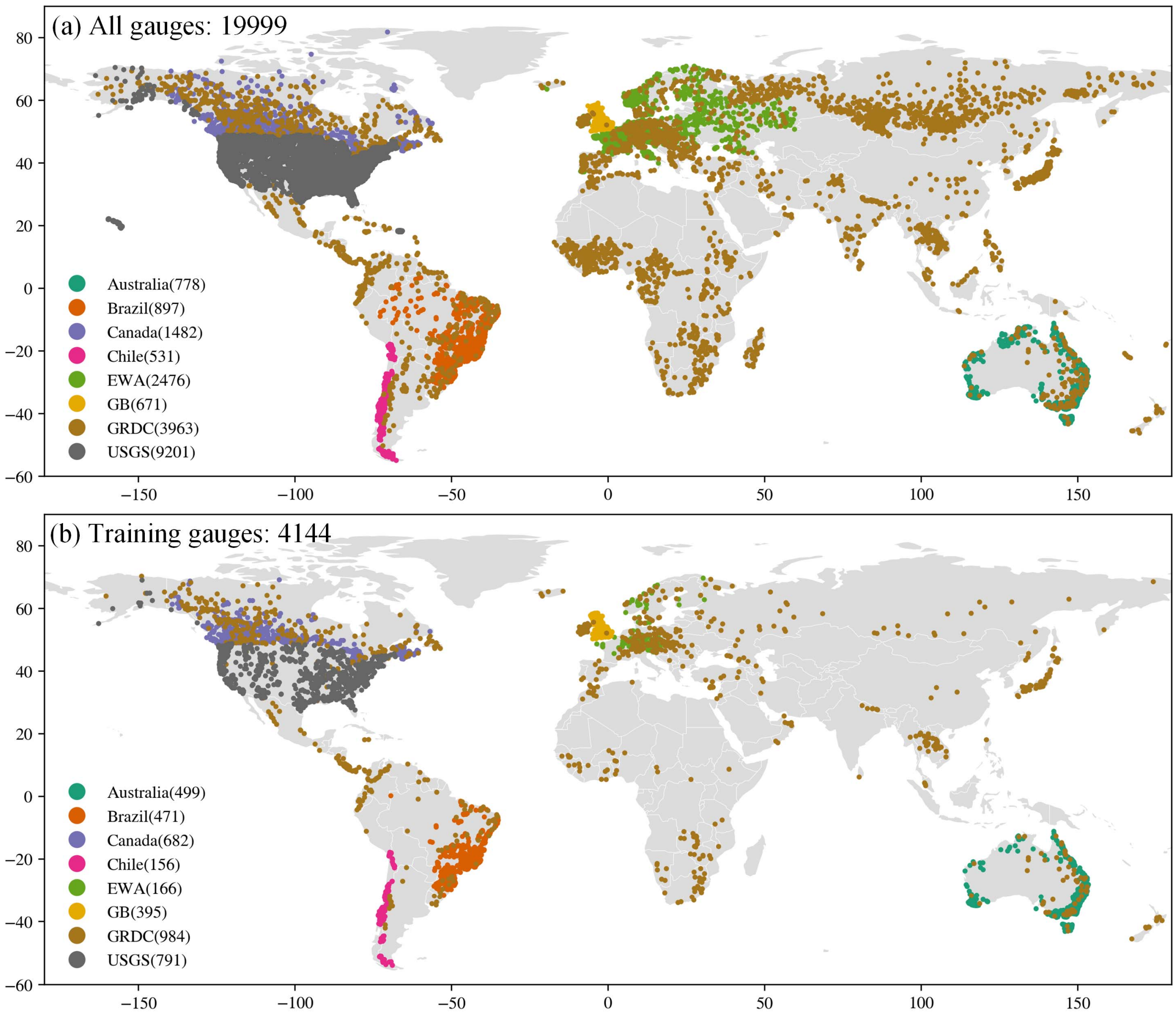

Figure 1. Spatial distribution of (a) all basins and (b) selected training basins. The legend shows the different data sources and corresponding gauge numbers. Australia: the Australian Bureau of Meteorology(BoM), Brazil: the HidroWeb portal of the Brazilian Agência Nacional de Águas, Canada: the Water Survey of Canada Hydrometric Data (HYDAT), Chile: the Chilean Center for Climate and Resilience Research (CR2), EWA: the European Water Archive of EURO-FRIEND-Water, GB: Great Britain, GRDC: the Global Runoff Data Centre, USGS: the United States Geological Survey National Water Information System. More details about the data sources can be found in Beck et al. (2020).

2.3 Training basin selection

We performed a screening to identify suitable training basins globally. As previously mentioned in Section 2.2, the sole data source for Asia and Africa is GRDC, which primarily consists of large basins with monthly data or data before 1995. This poses a challenge as it restricts the data availability for effective training and evaluation processes. Therefore, different thresholds were applied to non-GRDC and GRDC basins for criteria 2 and 3, allowing for the inclusion of selected basins in Asia, Siberia, and Africa. The following procedure was thus implemented:

1) The absolute relative difference between the area derived from the basin boundary file and reported by each data provider had to be smaller than 10%.

2) Only basins within the range of 50-5,000 km^2 for non-GRDC basins or 50-10,000 km^2 for GRDC basins were selected. This criterion was applied since channel routing effects become apparent at the daily scale in larger basins (Gericke & Smithers, 2014).

3) At least 25 years of observed daily data for non-GRDC basins, or 5 years for GRDC basins had to be available (not necessarily continuous) during 1980-2020 to ensure there was sufficient data.

4) To minimize artifacts and anthropogenic influence, basins were specifically chosen based on three criteria: a) reservoir influence ≤ 0.1, b) urban area fraction ≤ 0.1, c) irrigated fraction ≤ 5%, and d) a "reference" flagged if the data source is USGS.

The selection procedure described above finally produced a list of 4144 basins for LSTM training (Fig 1b), most of which were located in North America, Europe, Australia, and Brazil.

2.4 Model evaluation

Taking the runoff generated by the above four experiments as inputs, we implemented the Routing Application for Parallel Computation of Discharge (RAPID; David et al., 2011), a river routing model that uses a matrix-based version of the Muskingum method to calculate the flow and volume of water for each reach on a river network. The detailed setup of the RAPID model can be found in Yang et al. (2021). Notably, to our knowledge, no such global grid-scale application of LSTM models coupled with a routing model has been conducted in previous research.

We did the evaluation during the period of 2000-2020, different from the training period, to test the temporal generalization ability of the developed global LSTM models. We selected the gauges meeting the following criteria: 1) < 500 m from the closest reach, 2) a small (≤ ±10%) discrepancy between the area derived from the basin boundary file and reported by each data provider, 3) a small (≤ ±10%) discrepancy between gauge area derived from the basin boundary file and upstream area of a river reach, 4) ≥ 3 years of valid data during the evaluation period 2000-2020, 5) the gauges with the smallest area difference in cases where multiple gauges matched a single river reach or vice versa, 6) training gauges were excluded from the evaluation to assess the spatial generalization ability of the developed LSTM models. 5332 (daily)/ 5331(monthly) gauges were selected by these six criteria. Note that our evaluation conducted over different time periods and different gauges from the training, involves not only temporal generalization, but also spatial generalization, which inevitably poses much greater challenges to the LSTM networks.

Additionally, we conducted an evaluation on a subset of gauges with little anthropogenic influence. This subset of gauges satisfies not only the above-mentioned six criteria but also fulfills the criterion 4 in the training basin selection. Finally, 1123 gauges (daily) / 1128 gauges (monthly) were chosen for the evaluation, focusing on those with little anthropogenic influence.

Metrics adopted to evaluate model performance include the modified Kling-Gupta Efficiency (KGE, Kling et al., 2012), Correlation Coefficient (CC), Relative variability (RV), and Relative Bias (RB). CC measures the dynamic errors (temporal coherence). RV describes the bias in variability, and RB is widely used to indicate the magnitude of over- or under-estimations compared to the observations. KGE adds together CC, RV and RB, and is considered a more balanced metric. While all these metrics evaluated the performance over the entire time series, we also used the percent bias of the top 2% peak flow range (FHV) and the percent bias of the bottom 30% low flow range (FLV) (Yilmaz et al., 2008), to highlight the performance of the model for peak flows and baseflow, respectively. To compare with previous LSTM studies, the Nash-Sutcliffe Efficiency coefficient (NSE; Nash & Sutcliffe, 1970) was also calculated. All metrics were calculated for each evaluation basin for the period 2000-2020.

3 Results and Discussions

3.1 Model performance among the four experiments

Figure 2 shows the performance comparison among the four experiments over the global 5332 gauges for the period of 2000-2020 at the daily scale. Notably, all LSTM experiments exhibit comparable or superior results versus to the benchmark VIC model. The benchmark VIC model was already calibrated and bias-corrected against observation-based runoff characteristics, and was considered a significant advance compared to existing modeling literature (Yang et al., 2021). This comparison highlights the ability of a global grid-scale LSTM model to capture and learn hydrologic behaviors across diverse basin.

LSTM(obs) stands out among the four experiments for having the best overall performance, with the highest median KGE (0.563) and CC (0.811), the lowest bias (1.25%) and the closest variance (0.992). The median FHV and FLV of *LSTM(obs)* are 4.833%, and -4.023%, respectively, indicating its good ability to reproduce both high and low flow. *LSTM(obs)* performs much better than *LSTM(VIC)*, whose median KGE is 0.471. This is expected since *LSTM(obs)* uses discharge observations as the training target while *LSTM(VIC)* uses the modeled runoff, which may already include some bias. *LSTM(VIC+obs)* also exhibits much better performance than *LSTM(VIC)*, showing that imperfection and/or bias in the *LSTM(VIC)* are reduced through actual observations-based training and indicating the importance of using observations during the training process. However, *LSTM(VIC+obs)* does not produce any benefits compared to *LSTM(obs)* in this case, indicating that simulated data from VIC cannot provide additional added value to enhance global discharge estimation. This is partially because *LSTM(obs)* is already very strong, and the observations may already provide sufficient information for LSTM networks to construct gradient-like features. However, we envision future scenarios where *LSTM(VIC+obs)* could have value when there are not sufficient observations available for training the LSTM network (see Section 3.4).

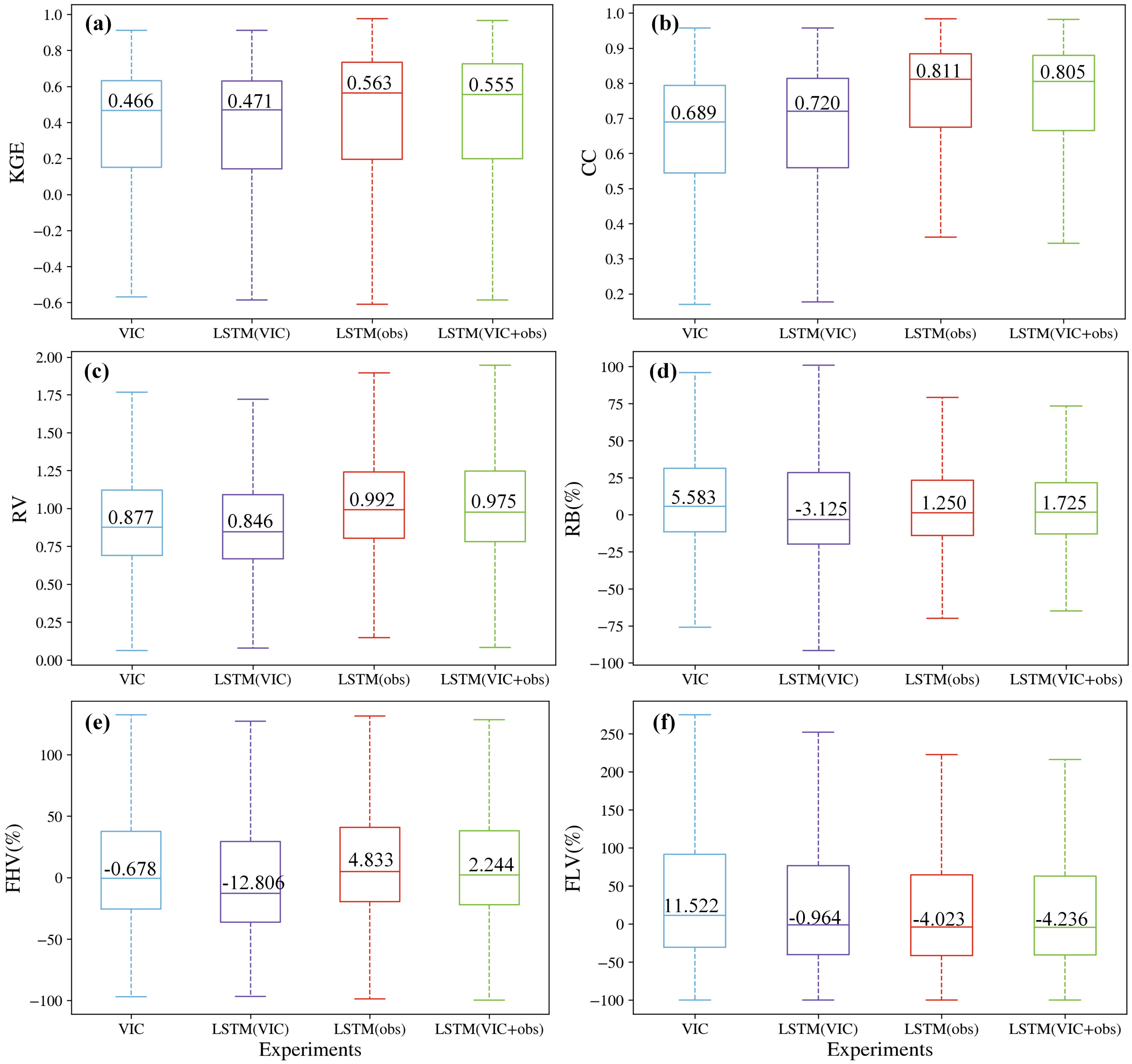

Figure 2. Performance of the four experiments at evaluation gauges for the evaluation period of 2000-2020: (a) KGE, (b) CC, (c) RV, (d) RB, (e) FHV and (f) FLV. All metrics are calculated for 5332 global gauges at the daily scale.

Additionally, an evaluation was conducted on a subset of gauges with little anthropogenic influence. The median KGE of the four experiments are 0.454, 0.445, 0.599, and 0.589, respectively. *LSTM(obs)* exhibits superior performance again. Compared to the results of all 5332 evaluation gauges, *LSTM(obs)* shows better performance on gauges with little anthropogenic influence (0.599 vs 0.563).

Upon initial examination, the performance of *LSTM(obs)* across the 5332 evaluation gauges in this study (with a median KGE of 0.563 and a median NSE of 0.476) appears comparatively lower than previous global basin-lumped LSTM studies. For example, Koya & Roy (2023) achived a median KGE of 0.647 for 2610 basins, and Tang et al. (2023) reported a median NSE of 0.59 for 1897 basins with area large than 9,000 km^2. However, it's important to note that these two studies applied LSTM at the basin scale, that is, they trained and tested LSTM at the same basins, thus limiting discharge generation to training basins. In contrast, this study pushed forward the application of LSTM at the grid scale and then coupled it with the RAPID river routing model to estimate the discharge everywhere globally. The evaluation conducted in this study is a comprehensive evaluation of the entire framework (grid-scale LSTM + routing), which is inevitably more challenging than previous studies. When we conducted the similar basin-scale LSTM application globally, we achived a median KGE of 0.733 and a median NSE of 0.689 over 4144 basins for the period 2000-2020, surpassing the performance of the two aforementioned global studies. Overall, the grid-scale LSTM coupled with a river routing model trade a certain amount of performance metrics for the discharge of every river reach worldwide.

3.2 Spatial pattern of *LSTM(obs)* performance

Figure 3 maps the spatial pattern of *LSTM(obs)* performance over the global 5332 (daily) / 5331(monthly) gauges for 2000-2020. The temporal dynamics are effectively simulated over most parts of the world except for complex terrains like the central CONUS. Approximately 82.4% (91.3%) of the gauges exhibit CC higher than 0.6 at daily (monthly) scale. The model reproduces flow variability well, albeit with a tendency to underestimate in the central CONUS and overestimate in northern Canada. The total flow volume is also well captured by *LSTM(obs)*, with about 54.5% (74.2%) gauges having RB within ±20% (±40%). Large overestimations (e.g., RB >100%) and large underestimations (e.g., RB<-60%) are mostly located in the central CONUS, eastern Brazil and Africa, which are mainly arid regions where a small absolute error (e.g., 0.1 m^3s^{-1}) leads to a large relative error. Overall, about 55.8% (daily) / 65.0% (monthly) gauges exhibit KGE values larger than 0.6, hence indicating that *LSTM(obs)* shows good performance globally. Consistent with previous hydrological model-based results (Alfieri et al., 2020; Lin et al., 2019; Yang et al., 2021), LSTM also struggles with arid basins. For example, arid regions like the central CONUS, Africa, and eastern Brazil show negative KGE values. Several factors could contribute to the poor performance in arid regions, including: 1) highly non-linear response due to substantial transmission losses, 2) low quality of precipitation data due to the prevalence of short-duration, localized convective events, and 3) very low discharge volumes that cannot provide effective training samples to the LSTM network.

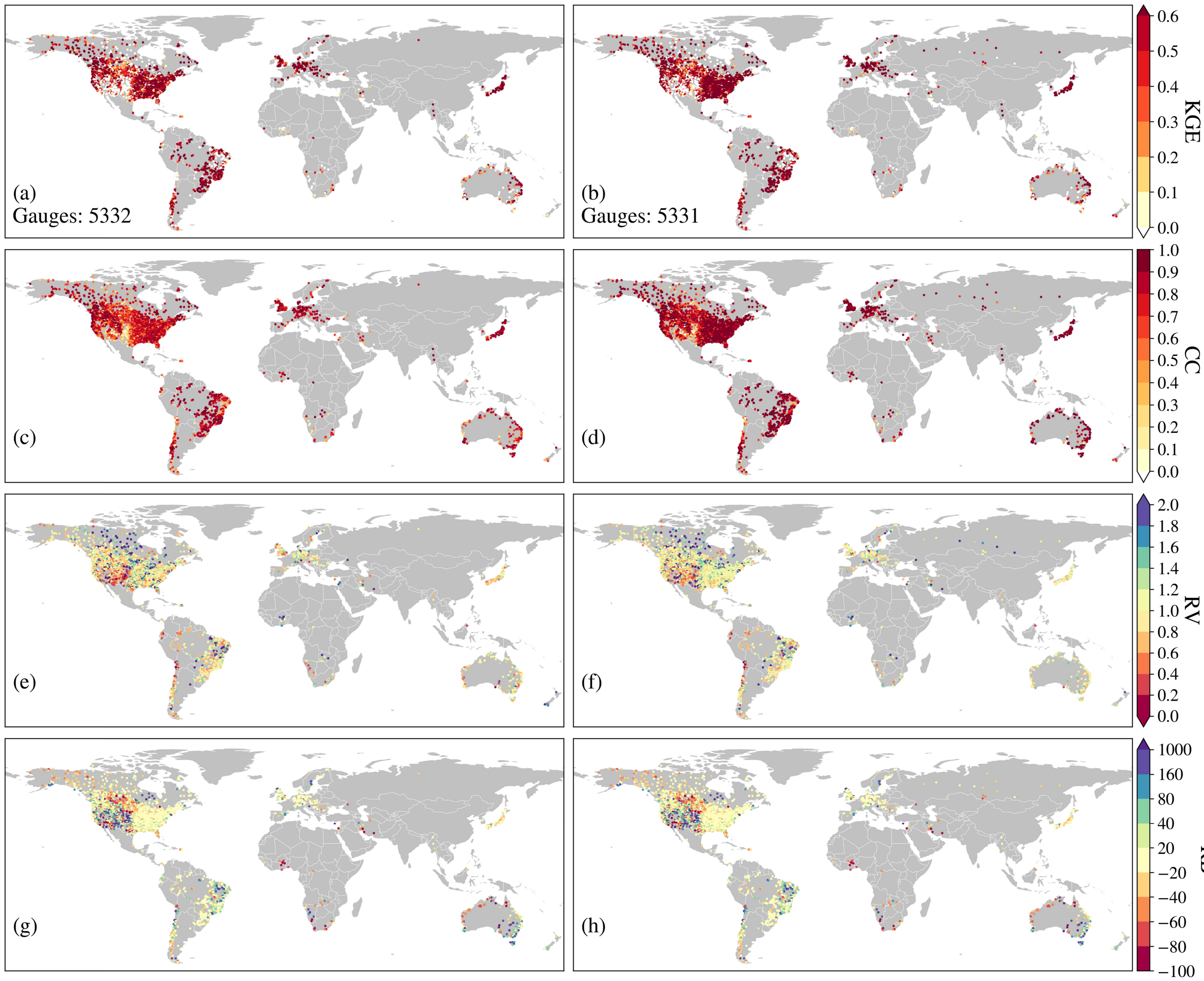

Figure 3. Skill metrics for simulated discharge from *LSTM(obs)*. (a), (b) KGE, (c), (d) CC, (e), (f) RV, and (g), (h) RB for (left) daily and (right) monthly scales.

3.3 Performance of *LSTM(obs)* compared to *VIC*

The performance of *LSTM(obs)*, the best-performed LSTM experiment, is compared to that of the benchmark *VIC* model to assess the effectiveness of LSTM in global river discharge estimates. Figure 4 shows the histogram comparison between *LSTM(obs)* and *VIC* for the KGE, CC, RV, and RB values for the 5332 evaluation gauges at the daily scale. Tables in Figure 4 show the statistics of Kolmogorov-Smirnov test (K-S test; Eghbali, 1979; Smirnov, 1948), which is used to validate whether the improvements are significant. Significant improvements can be seen in the overall KGE as evidenced by a rightward shift of the red bars in Figure 4(a). The percentages of KGE falling in ranges <0.7, [0.7,0.8], [0.8,0.9], and [0.9,1] are 69.3%, 17.2%, 11.8%, and 1.7%, improved from 85.6%, 10.7%, 3.7%, and 0.0%, respectively. The number of gauges which have negative KGE values is also reduced. The improvements in the three components of KGE are also significant. In terms of CC, about 52.7% of gauges show values larger than 0.8 for *LSTM(obs)*, while only about 23.2% of gauges for *VIC*, indicating that the temporal dynamics are obviously and significantly improved. The substantial improvement in CC provides valuable insights into flow timing, which would be critical for flood prevention and water management. Compared with *VIC*, *LSTM(obs)* tends to overestimate the flow variability to a greater degree which, at the same time, also reduces the severe underestimation of the flow variability. Therefore, a more balanced distribution of RV can be obtained, and the percentages of gauges falling in ranges <0.9, and >1.1 are 37.3% and 37.0% for *LSTM(obs)*, respectively, while 26.5% and 53.2% for *VIC*. More gauges fall in ranges around 1.0, for example, such as the range [0.9, 1.1]. The *VIC* runoff biases have already been corrected against nine runoff percentiles, further improvements in runoff biases can be difficult. However, significant improvements can be seen in RB as indicated by the higher percentage falling within ±10%.

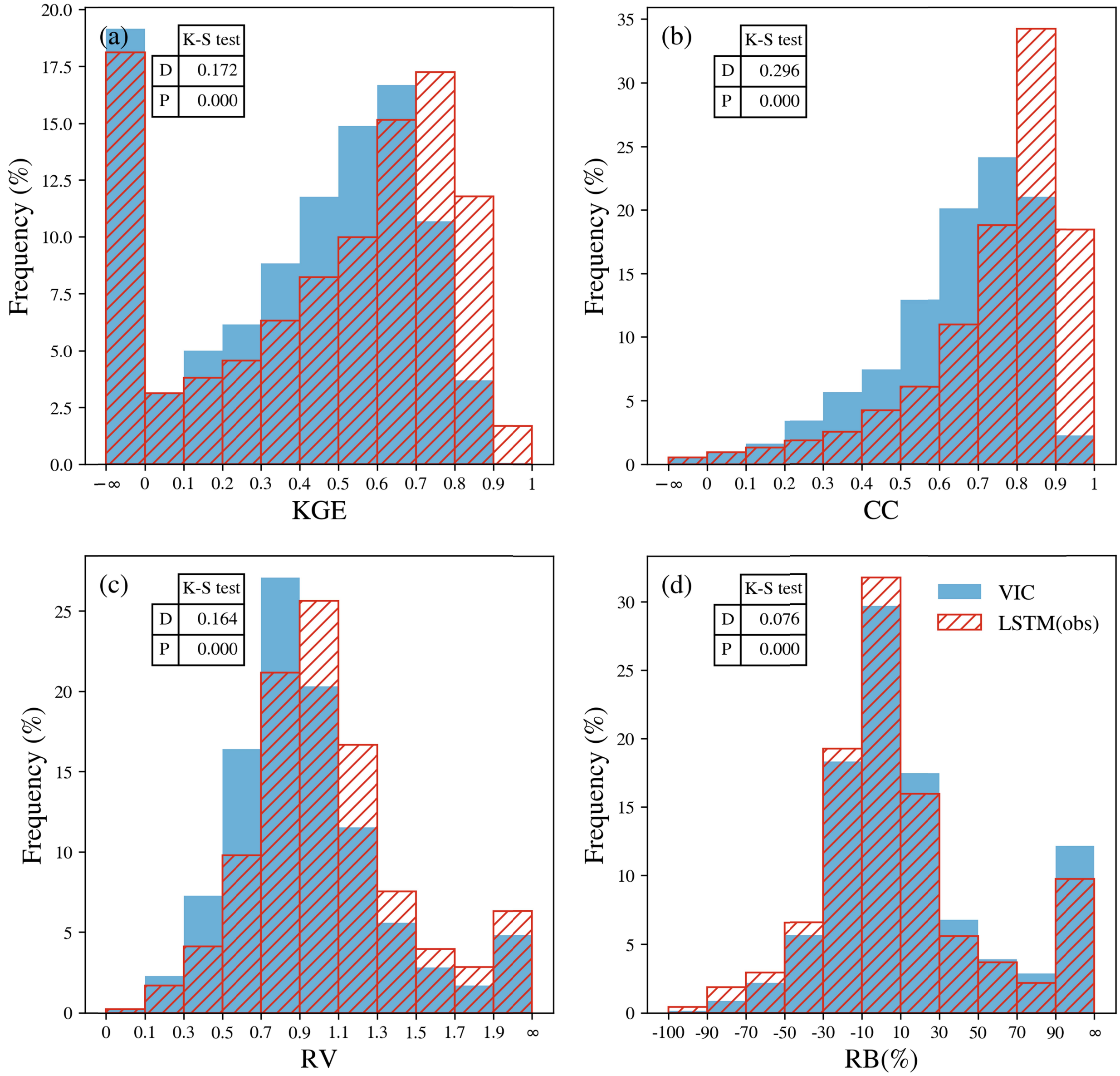

Figure 4. The histogram of (a) KGE, (b) CC, (c) RV and (d) RB for *LSTM(obs)* and *VIC*. D and P are K-S test statistics.

To study the spatial pattern of the improvement of model performance, the differences in skill metrics between *LSTM(obs)* and *VIC* are shown in Figure 5, and the skill metrics of *VIC* are shown in Figure S1. The gauges with small differences (±0.1 for KGE, CC and |RV-1|, ±10% for |RB|) are not counted and shown in Figure 5. *LSTM(obs)* shows overwhelming improvements in CC. 2172 gauges that have higher CC values in *LSTM(obs)*, while only 275 gauges in *VIC* with small difference magnitudes. The most obvious improvement in CC is located in the Rocky Mountains across Canada and CONUS, with CC values improved from less than 0.6 (Figure S1(c) and S1(d)) to more than 0.8 over most gauges in that region. The underestimations in flow variability have been largely reduced, especially in the central CONUS, eastern Brazil, and Australia. Overall, more gauges (1890 vs 1527) witness the improvements in RV. In terms of RB, *LSTM(obs)* reduces the overestimation in northern Chile and eastern Brazil and the underestimation in southern Chile. The overestimation in the northern part of the central CONUS and Africa has turned into underestimation. Also, more underestimations occur in Alaska and northeastern Canada. Overall, *LSTM(obs)* shows better performance over more gauges than the *VIC* model. 2500 (about 48%) gauges experience a boost of larger than 0.1 in KGE, but there are some regions with stronger improvements, for example in the western CONUS, eastern Brazil, Chile (Figure 5(a) and 5(b)). However, it is important to acknowledge that LSTM is not a silver bullet. For example, the negative KGE values still exist in Texas, New Mexico, and Arizona, perhaps because the short time scale of runoff generation in these basins (e.g., flash flood) are not easily handled by LSTM (Ma et al., 2021). The benefits of *LSTM(obs)* over *VIC* don't exhibit a discernible spatial pattern, nor show any correlations with topography, climate, and gauge basin area (scatter plots are omitted for brevity). The comparison between *LSTM(obs)* and *VIC* clearly shows that *LSTM(obs)* finds rainfall-runoff relationships in some basins that *VIC* cannot emulate, thus highlighting that there is substantial room to improve *VIC* overall. At the same time, the fact that *VIC* performs better in certain basins (Figure 5(b)) indicates the potential value of having physical constraints in a hydrological model.

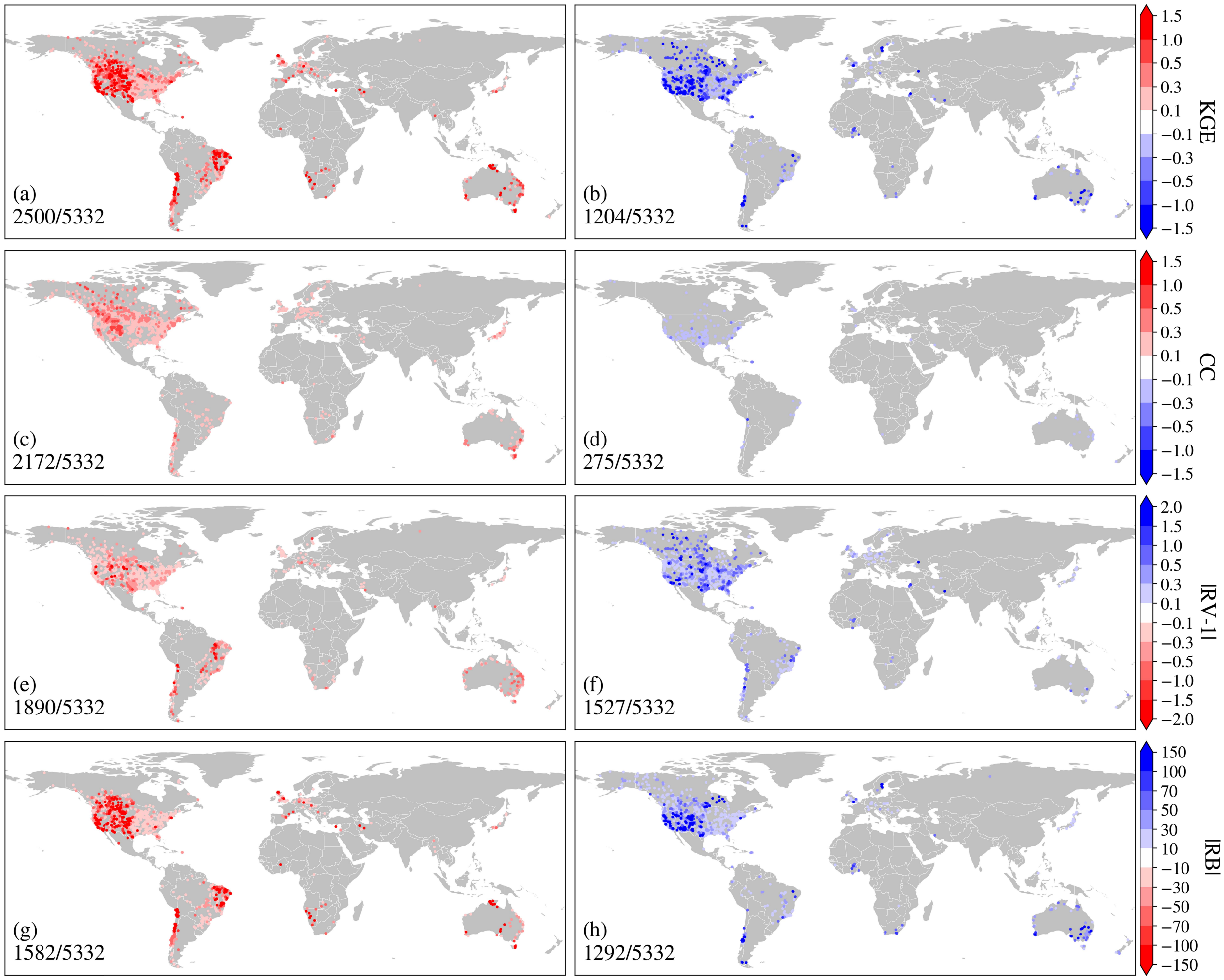

Figure 5. The skill metrics differences between *LSTM(obs)* and the benchmark *VIC* model. For RV and RB, we use the the absolute value of (RV-1) and RB to show the metrics difference. The red dots indicate that *LSTM(obs)* performs better than *VIC*, blue dots the other way around. #1 of #1/5332 stands for the gauge number with better performance (left: *LSTM(obs)* has better performance, right: *VIC* has better performance). The gauges with metric differences lying in the white range are not shown and counted in this figure.

3.4 The effect of the number of training gauges

To have a better understanding of the effect of initialization through pre-training ML models with simulated data from process-based models, we conducted a sensitivity analysis in the *LSTM(VIC+obs)* training process (the same setups except for the number of training gauges). Different numbers (N) of training gauges (N=100, 500, 1000…) were randomly sampled. To reduce the sampling uncertainty, each N (except 4144, the total number of training gauges) was sampled three times. Figure 6 shows that the more training data used to train LSTM networks, the better performance obtained for both *LSTM(obs)* and *LSTM(VIC+obs)*. Notably, utilizing a small number of training gauges (100 for the whole globe) would introduce large uncertainties in RB, thus resulting in worse performance than *VIC* and *LSTM(VIC)*, which don't incorporate any observations from training gauges. It is inspiring that when there are not enough training gauges, pre-training LSTM networks by the modeled runoff from *VIC* can improve the model performance as evidenced by slightly higher KGE values in *LSTM(VIC+obs)*. This result is consistent with previous research on lake temperature (Jia et al., 2018, 2021; Read et al., 2019), which showed that pretraining ML models with synthetic data can be helpful when limited observed data were available. However, when there are sufficient training data, for example, 3000 or 4144 basins in this case, *LSTM(VIC+obs)* doesn't surpass the *LSTM(obs)*, which counters our intuition that model initialization using the simulated data from a process-based hydrological model could improve model performance (Jia et al., 2018; Ma et al., 2021; Read et al., 2019). Several potential explanations may be considered to account for this. First, previous studies used the same training target (e.g. lake temperature or discharge) in the same basins for both the pre-training and retraining process. In this study, although we normalized the discharge data (the training target in the retraining process) by basin area, it is still different from *VIC* 0.25° runoff (the training target in the pretraining process), especially for larger basins where the routing process plays a more substantial role. Additionally, the training runoff grids in the pretraining process are different from the training basins in the retraining process. Second, previous studies focused on training and testing LSTM models on a single lake or basin, while this study trained LSTM over thousands of basins, then applied the trained LSTM at grid-scale, and tested over non-training basins, which introduced large uncertainties to the performance. Third, this may be related to the limited transfer learning ability of LSTM. Other advanced ML algorithms, such as Transformer (J. Liu, Bian, et al., 2023; Vaswani et al., 2017) and its variants or physics-informed differentiable models (Feng et al., 2022, 2023; Shen et al., 2023), could be explored in the future to see whether any gains can be obtained by pre-training ML models using process-based model's simulated data at the global scale.

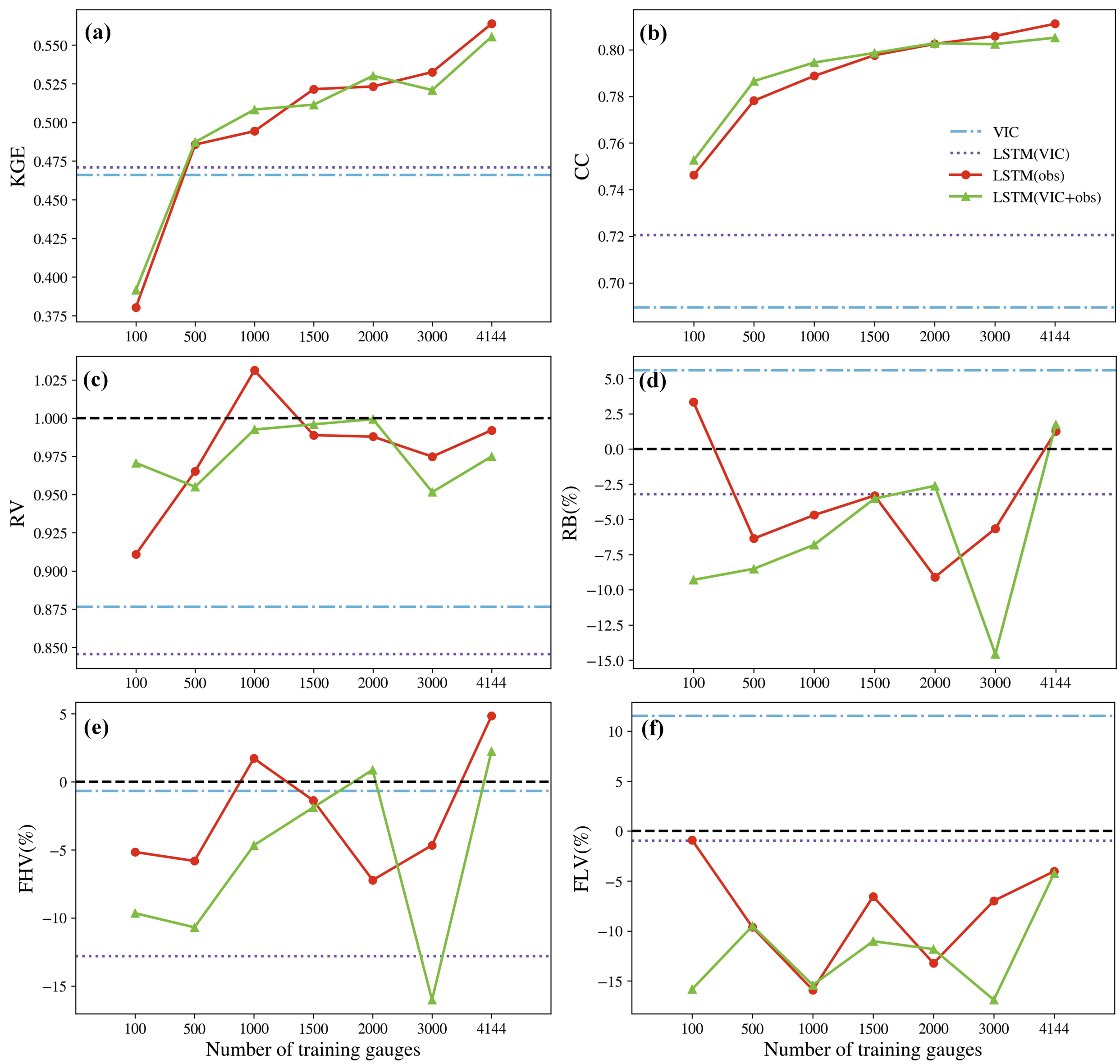

Figure 6. The median values of skill metrics from the four experiments with different numbers of training gauges. The blue, purple, red, and green lines are for *VIC*, *LSTM(VIC)*, *LSTM(obs)*, and *LSTM(VIC+obs)*. The black lines stand for the perfect values of skill metrics.

4 Global daily reach-level river discharge dataset based on LSTM (GRADES-hydroDL)

Given the promising performance of *LSTM(obs)* observed in this study, we further develop an improved global reach-level daily database of discharge records spanning the period from 1980 to 2020. The setups are the same as *LSTM(obs)*, with the exception that the data from 1980-2020 are adopted in the training process to maximize the information learned. This updated dataset serves as an improved iteration of the previously known GRADES (Lin et al., 2019), and we now refer to it as GRADES-hydroDL. Here, hydroDL denotes a set of consistent hydrologic deep learning implementations using a specific library. The hydroDL library contains models for several versions of LSTM models for hydrologic and water quality variables, multiscale models, physics-informed differentiable models (ecosystem, routing - so far small scale), and in the future Transformer models. Besides streamflow (Feng et al., 2020, 2021), hydroDL has been employed to simulate soil moisture (Fang & Shen, 2020; J. Liu, Hughes, et al., 2023), stream temperature (Rahmani, Lawson, et al., 2021; Rahmani, Shen, et al., 2021), dissolved oxygen (Zhi et al., 2021, 2023), sediment (Chaemchuen et al., 2023), nitrate (Saha et al., 2023), phosphorous and snow water equivalent, etc. Furthermore, the neural networks in the library are utilized to support physics-informed differentiable modeling (Shen et al., 2023), where neural networks are integrated with physical descriptions to provide physical concepts, interpretability and intermediate fluxes (Aboelyazeed et al., 2023; Bindas et al., 2022; Feng et al., 2022, 2023; Shen et al., 2023). The sequence-to-sequence nature of the model makes it quite efficient, with a CONUS-scale training job to finish typically within 2 hours on a single 2080 Ti Graphical Processing Unit (GPU).

GRADES-hydroDL reproduces the global discharge very well (Figure 7). Figure 8(a) maps the mean daily river discharge from 1980 to 2020 for each river reach with a stream order larger than 5 and discharge value larger than 1 m^3/s, revealing the main river arteries of the world. Generally, river reaches situated further downstream exhibit a darker blue color, which indicates larger river discharge values. Discharge is concentrated in specific regions, with quantities ranging from nearly zero in the desert areas, such as the Sahara, Gobi, and Arabian deserts, to exceeding 50,000m^3/s near the river mouths of major rivers the Amazon, Mississippi, Yangtze, Congo, and Nile. The spatial pattern is similar to previous results (Harrigan et al., 2020; Lin et al., 2019). Figure 8(b) shows the inter-annual variability of global rivers, quantified by the coefficient of variation (CV) of annual flow calculated as the standard deviation divided by the mean annual discharge. The majority of rivers exhibit low inter-annual variability, with $CV <$ 0.4, indicating a steady flow pattern during the period from 1980 to 2020. However, certain regions, including the Great Plains, eastern Brazil, Argentina, Australia, southern Africa, area around 15°N in Africa, and eastern Indian Peninsula, experience substantial inter-annual variability. Matching with the Köppen-Geiger climate types (Beck et al., 2018), these least steady rivers are mostly located in arid regions that are driven by precipitation variability (Fielding et al., 2018; McMahon et al., 1987). Other rivers with large inter-annual variability can also be seen in the CONUS West Coast and northeastern China.

The analysis of flood seasonality holds significant importance in enhancing our understanding of the flood generation mechanisms and hence is critical in a number of applications, from flood risk estimation and water resources management to climate change investigations (Berghuijs et al., 2019; Blöschl et al., 2017; Collins, 2019; Dickinson et al., 2019; Hall & Blöschl, 2018; Villarini, 2016; Ye et al., 2017). Based on the annual maximum flows, two seasonality metrics were calculated globally:1) average timing of river flood $\overline{D}$, quantifying the time of the year in which flood events tend to occur and 2) concentration of floods R, quantifying how strong the seasonality is. Higher values denote a more concentrated flood season while lower values signify spread-out seasonal distributions. Detailed definitions are in Supporting Information Text S1. The mean seasonality of flood exhibits distinct regional features (Figure 8(c)). In low-latitude tropical regions, annual floods predominantly occur during the wet summer season of July-September in the Northern Hemisphere and December-February in the Southern Hemisphere. The Northern Hemisphere shows a more spatially heterogeneous pattern. High-latitudes and high-altitudes (e.g., Rockies, Alps) primarily experience late spring and summer floods due to the effect of snow storage and melt (Hall & Blöschl, 2018; Parajka et al., 2009; Villarini, 2016; Ye et al., 2017). The US West Coast is characterized by winter floods from December to February, which are often driven by atmospheric rivers that are long and narrow atmospheric features transporting moisture from the tropics to the midlatitudes and produce potentially significant runoff in the warm, heavy rainfall events (Leung & Qian, 2009; Neiman et al., 2011; Ralph et al., 2006). As we move northward, the flood seasonality in the eastern US and China transitions from late winter to spring, and from spring to summer, respectively. Complexity arises with factors such as land-sea interactions. For example, the flood season in Europe shifts from December-January in coastal areas to April-May in the interior because of increasing continentality (away from the Atlantic) (Blöschl et al., 2017; Hall & Blöschl, 2018). The seasonal concentration of flood R is shown in Figure 8(d). Globally, the seasonality is strong with $R > 0.9$ in most areas. Relatively weaker seasonality is found mainly in the southeastern CONUS, the Mediterranean, southern Brazil, and equatorial regions, with $R < 0.5$.

Due to its high accuracy, GRADES-hydroDL will be tremendously valuable for making better decisions on water-related issues such as flood control, integrated water resources management, and ecological environmental assessment. Additionally, as an updated version of GRADES, it can provide better prior information in support of the SWOT mission and other scientific applications requiring spatiotemporal continuous discharge estimates. In future studies, the updated GRADES-hydroDL shall supersede its predecessor GRADES. We highly recommend using this new data in relevant applications.

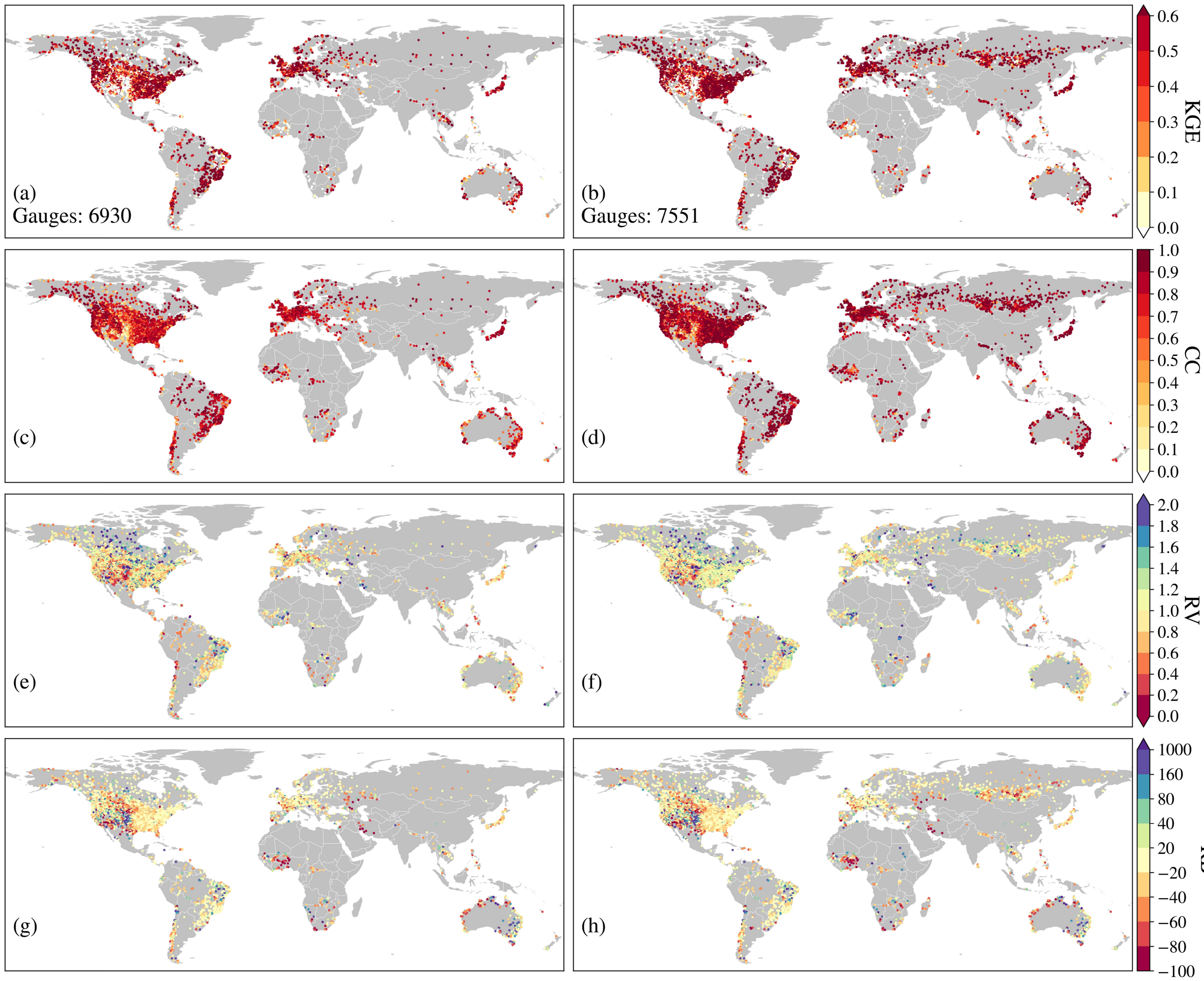

Figure 7. Same as Figure 3, but for GRADES-hydroDL during the period of 1980-2020.

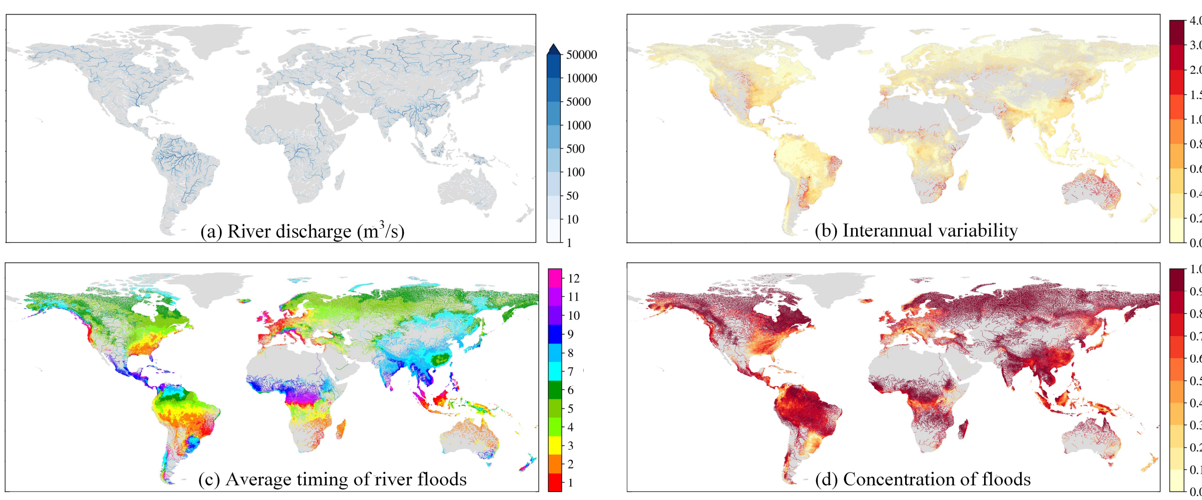

Figure 8. Characteristics for GRADES-hydroDL, 1980-2020. (a) Mean daily river discharge during 1980-2020 for river reaches of stream order ≥ 5 and discharge value ≥ 1 m^3s^{-1}, (b)

Interannual variability (CV) for river reaches of discharge value ≥ 1 m^{3}s^{-1}, (c) Average timing of river floods ($\overline{D}$) for river reaches of discharge value ≥ 1 m^{3}s^{-1}, (d) Concentration of floods within a year (R, if R=0, evenly distributed; if R =1, all floods occur on the same date) for river reaches of discharge value ≥ 1 m^{3}s^{-1}.

5 Conclusions

For the first time, we implemented the LSTM model on the 0.25° grid scale to get seamless global runoff field and then coupled it with the RAPID river routing model to estimate reach-level daily discharge globally. The effectiveness and performance of LSTM in comparison to VIC for estimating global river discharge is investigated through four experiments.

The results demonstrate that the grid-scale LSTM model shows great potential in global river discharge estimation. When trained on the observed discharge data from 4144 basins globally, and then applied at 0.25° grid scale, LSTM can effectively capture and learn rainfall-runoff behaviors across diverse basins and accurately simulate global river discharge. Moreover, over the majority of regions, LSTM significantly outperforms the process-based VIC model, which is already calibrated and bias-corrected, highlighting the superiority of LSTM. However, LSTM is not a silver bullet that solves all problems. It still struggles with arid basins.

The performance of LSTM is affected by the number of training observations, which improves as the training data increases. Pre-training the global LSTM model with the simulated data from the process-based model is hard to provide additional added value to global river discharge estimation when sufficient observed discharge data is available, while it can enhance model performance in situations where observations are limited.

Based on the grid-scale LSTM model and the RAPID river routing model, an improved global reach-level daily database of discharge for the period 1980-2020 is developed. This data, referred to as GRADES-hydroDL, can serve as valuable resources for global hydrologic research. For example, this long-term data with high accuracy helps to better understand the global water resources and their variability at seasonal and long-term scales. Additionally, as an upgraded version of GRADES, GRADES-hydroDL can provide essential support to river-observing satellite missions, such as SWOT, facilitating the development of accurate discharge algorithms. The GRADES-hydroDL is available at https://app.globus.org/file-manager?origin_id=3380b1a4-698c-47c0-96db-e47e06b97295&origin_path=%2F.

Acknowledgments

This work was supported by NASA Grant NNH18ZDA001N-NEWS “Global and Regional Water and Energy Variations Under a Changing Climate”.

References

Aboelyazeed, D., Xu, C., Hoffman, F. M., Liu, J., Jones, A. W., Rackauckas, C., et al. (2023). A differentiable, physics-informed ecosystem modeling and learning framework for large-scale inverse problems: demonstration with photosynthesis simulations. *Biogeosciences*, *20*(13), 2671–2692. https://doi.org/10.5194/bg-20-2671-2023

Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., & Salamon, P. (2020). A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X*, *6*, 100049. https://doi.org/10.1016/j.hydroa.2019.100049

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., & Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific Data*, *5*(1), 180040. https://doi.org/10.1038/sdata.2018.40

Beck, H. E., Roo, A. de, & Dijk, A. I. J. M. van. (2015). Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments. *Journal of Hydrometeorology*, *16*(4), 1478–1501. https://doi.org/10.1175/JHM-D-14-0155.1

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Dijk, A. I. J. M. van, et al. (2019). MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1

Beck, H. E., Wood, E. F., McVicar, T. R., Zambrano-Bigiarini, M., Alvarez-Garreton, C., Baez-Villanueva, O. M., et al. (2020). Bias Correction of Global High-Resolution Precipitation Climatologies Using Streamflow Observations from 9372 Catchments. *Journal of Climate*, *33*(4), 1299–1315. https://doi.org/10.1175/JCLI-D-19-0332.1

Berghuijs, W. R., Harrigan, S., Molnar, P., Slater, L. J., & Kirchner, J. W. (2019). The Relative Importance of Different Flood-Generating Mechanisms Across Europe. *Water Resources Research*, *55*(6), 4582–4593. https://doi.org/10.1029/2019WR024841

Biancamaria, S., Lettenmaier, D. P., & Pavelsky, T. M. (2016). The SWOT mission and its capabilities for land hydrology. *Remote Sensing and Water Resources*, 117–147. https://doi.org/10.1007/978-3-319-32449-4_6

Bindas, T., Tsai, W.-P., Liu, J., Rahmani, F., Feng, D., Bian, Y., et al. (2022). *Improving large-basin streamflow simulation using a modular, differentiable, learnable graph model for routing* (preprint). Hydrology. https://doi.org/10.1002/essoar.10512512.1

Bjerklie, D. M., Birkett, C. M., Jones, J. W., Carabajal, C., Rover, J. A., Fulton, J. W., & Garambois, P.-A. (2018). Satellite remote sensing estimation of river discharge: Application to the Yukon River Alaska. *Journal of Hydrology*, *561*, 1000–1018. https://doi.org/10.1016/j.jhydrol.2018.04.005

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., et al. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351), 588–590. https://doi.org/10.1126/science.aan2506

Chaemchuen, P., Song, Y., Rahmani, F., Zhi, W., Li, L., Liu, X., et al. (2023, January 11). Deep Learning Insights into Suspended Sediment Concentrations Across the Conterminous United States: Strengths and Limitations. SSRN Scholarly Paper, Rochester, NY. https://doi.org/10.2139/ssrn.4322321

Chen, J., Shi, H., Sivakumar, B., & Peart, M. R. (2016). Population, water, food, energy and dams. *Renewable and Sustainable Energy Reviews*, *56*, 18–28. https://doi.org/10.1016/j.rser.2015.11.043

Collins, M. J. (2019). River flood seasonality in the Northeast United States: Characterization and trends. *Hydrological Processes*, *33*(5), 687–698. https://doi.org/10.1002/hyp.13355

Coughlan de Perez, E., Van den Hurk, B., Van Aalst, M. K., Amuron, I., Bamanya, D., Hauser, T., et al. (2016). Action-based flood forecasting for triggering humanitarian action. *Hydrology and Earth System Sciences*, *20*(9), 3549–3560. https://doi.org/10.5194/hess-20-3549-2016

David, C. H., Maidment, D. R., Niu, G.-Y., Yang, Z.-L., Habets, F., & Eijkhout, V. (2011). River Network Routing on the NHDPlus Dataset. *Journal of Hydrometeorology*, *12*(5), 913–934. https://doi.org/10.1175/2011JHM1345.1

Dickinson, J. E., Harden, T. M., & McCabe, G. J. (2019). Seasonality of climatic drivers of flood variability in the conterminous United States. *Scientific Reports*, *9*(1), 15321. https://doi.org/10.1038/s41598-019-51722-8

Durand, M., Gleason, C. J., Pavelsky, T. M., Prata de Moraes Frasson, R., Turmon, M., David, C. H., et al. (2023). A framework for estimating global river discharge from the Surface Water and Ocean Topography satellite mission. *Water Resources Research*, *59*(4), e2021WR031614. https://doi.org/10.1029/2021WR031614

Eghbali, H. J. (1979). K-S Test for Detecting Changes from Landsat Imagery Data. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 17–23. https://doi.org/10.1109/TSMC.1979.4310069

Fang, K., & Shen, C. (2020). Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel. *Journal of Hydrometeorology*, *21*(3), 399–413. https://doi.org/10.1175/JHM-D-19-0169.1

Fang, K., Shen, C., Kifer, D., & Yang, X. (2017). Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental U.S. Using a Deep Learning Neural Network. *Geophysical Research Letters*, *44*(21), 11,030-11,039. https://doi.org/10.1002/2017GL075619

Fekete, B. M., Vörösmarty, C. J., & Grabs, W. (2002). High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, *16*(3), 15–1. https://doi.org/10.1029/1999GB001254

Feng, D., Fang, K., & Shen, C. (2020). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *Water Resources Research*, *56*(9), e2019WR026793. https://doi.org/10.1029/2019WR026793

Feng, D., Lawson, K., & Shen, C. (2021). Mitigating Prediction Error of Deep Learning Streamflow Models in Large Data-Sparse Regions With Ensemble Modeling and Soft Data. *Geophysical Research Letters*, *48*(14), e2021GL092999. https://doi.org/10.1029/2021GL092999

Feng, D., Liu, J., Lawson, K., & Shen, C. (2022). Differentiable, learnable, regionalized process-based models with multiphysical outputs can approach state-of-the-art hydrologic prediction accuracy. *Water Resources Research*, *58*(10), e2022WR032404. https://doi.org/10.1029/2022WR032404

Feng, D., Beck, H., Lawson, K., & Shen, C. (2023). The suitability of differentiable, physics-informed machine learning hydrologic models for ungauged regions and climate change impact assessment. *Hydrology and Earth System Sciences*, *27*(12), 2357–2373. https://doi.org/10.5194/hess-27-2357-2023

Ficke, A. D., Myrick, C. A., & Hansen, L. J. (2007). Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries*, *17*, 581–613. https://doi.org/10.1007/s11160-007-9059-5

Fielding, C. R., Alexander, J., & Allen, J. P. (2018). The role of discharge variability in the formation and preservation of alluvial sediment bodies. *Sedimentary Geology*, *365*, 1–20. https://doi.org/10.1016/j.sedgeo.2017.12.022

Frame, J. M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., et al. (2022). Deep learning rainfall–runoff predictions of extreme events. *Hydrology and Earth System Sciences*, *26*(13), 3377–3392. https://doi.org/10.5194/hess-26-3377-2022

Gauch, M., Mai, J., & Lin, J. (2021). The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. *Environmental Modelling & Software*, *135*, 104926. https://doi.org/10.1016/j.envsoft.2020.104926

Gericke, O. J., & Smithers, J. C. (2014). Review of methods used to estimate catchment response time for the purpose of peak discharge estimation. *Hydrological Sciences Journal*, *59*(11), 1935–1971. https://doi.org/10.1080/02626667.2013.866712

Gerten, D., Rost, S., von Bloh, W., & Lucht, W. (2008). Causes of change in 20th century global river discharge. *Geophysical Research Letters*, *35*(20). https://doi.org/10.1029/2008GL035258

Gleason, C. J., & Durand, M. T. (2020). Remote sensing of river discharge: A review and a framing for the discipline. *Remote Sensing*, *12*(7), 1107. https://doi.org/10.3390/rs12071107

Gleason, C. J., & Smith, L. C. (2014). Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences*, *111*(13), 4788–4791. https://doi.org/10.1073/pnas.1317606111

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, *28*(10), 2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

Hall, J., & Blöschl, G. (2018). Spatial patterns and characteristics of flood seasonality in Europe. *Hydrology and Earth System Sciences*, *22*(7), 3883–3901. https://doi.org/10.5194/hess-22-3883-2018

Hannah, D. M., Demuth, S., van Lanen, H. A., Looser, U., Prudhomme, C., Rees, G., et al. (2011). Large-scale river flow archives: importance, current status and future needs. *Hydrological Processes*, *25*(7), 1191–1200. https://doi.org/10.1002/hyp.7794

Hargreaves, G. H. (1994). Defining and Using Reference Evapotranspiration. *Journal of Irrigation and Drainage Engineering*, *120*(6), 1132–1139. https://doi.org/10.1061/(ASCE)0733-9437(1994)120:6(1132)

Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., et al. (2020). GloFAS-ERA5 operational global river discharge reanalysis 1979–present. *Earth System Science Data*, *12*(3), 2043–2060. https://doi.org/10.5194/essd-12-2043-2020

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PloS One*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., et al. (2018). Operational global reanalysis: progress, future directions and synergies with NWP. https://doi.org/10.21957/TKIC6G3WM

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hu, R., Fang, F., Pain, C., & Navon, I. (2019). Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. *Journal of Hydrology*, *575*, 911–920. https://doi.org/10.1016/j.jhydrol.2019.05.087

Huscroft, J., Gleeson, T., Hartmann, J., & Börker, J. (2018). Compiling and Mapping Global Permeability of the Unconsolidated and Consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). *Geophysical Research Letters*, *45*(4), 1897–1904. https://doi.org/10.1002/2017GL075860

Jia, X., Willard, J., Karpatne, A., Read, J., Zwart, J., Steinbach, M., & Kumar, V. (2018). Physics Guided RNNs for Modeling Dynamical Systems: A Case Study in Simulating Lake Temperature Profiles (Version 2). https://doi.org/10.48550/ARXIV.1810.13075

Jia, X., Willard, J., Karpatne, A., Read, J. S., Zwart, J. A., Steinbach, M., & Kumar, V. (2021). Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. *ACM/IMS Transactions on Data Science*, *2*(3), 20:1-20:26. https://doi.org/10.1145/3447814

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, *424–425*, 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Konapala, G., Kao, S.-C., Painter, S. L., & Lu, D. (2020). Machine learning assisted hybrid models can improve streamflow simulation in diverse catchments across the conterminous US. *Environmental Research Letters*, *15*(10), 104022. https://doi.org/10.1088/1748-9326/aba927

Koya, S. R., & Roy, T. (2023, May 20). Temporal Fusion Transformers for Streamflow Prediction: Value of Combining Attention with Recurrence. arXiv. https://doi.org/10.48550/arXiv.2305.12335

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning. *Water Resources Research*, *55*(12), 11344–11354. https://doi.org/10.1029/2019WR026065

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kratzert, F., Klotz, D., Hochreiter, S., & Nearing, G. S. (2021). A note on leveraging synergy in multiple meteorological data sets with deep learning for rainfall–runoff modeling. *Hydrology and Earth System Sciences*, *25*(5), 2685–2703. https://doi.org/10.5194/hess-25-2685-2021

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lees, T., Buechel, M., Anderson, B., Slater, L., Reece, S., Coxon, G., & Dadson, S. J. (2021). Benchmarking data-driven rainfall–runoff models in Great Britain: a comparison of long short-term memory (LSTM)-based models with four lumped conceptual models. *Hydrology and Earth System Sciences*, *25*(10), 5517–5534. https://doi.org/10.5194/hess-25-5517-2021

Leung, L. R., & Qian, Y. (2009). Atmospheric rivers induced heavy precipitation and flooding in the western U.S. simulated by the WRF regional climate model. *Geophysical Research Letters*, *36*(3). https://doi.org/10.1029/2008GL036445

Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres*, *99*(D7), 14415–14428. https://doi.org/10.1029/94JD00483

Liang, X., Wood, E. F., & Lettenmaier, D. P. (1996). Surface soil moisture parameterization of the VIC-2L model: Evaluation and modification. *Global and Planetary Change*, *13*(1), 195–206. https://doi.org/10.1016/0921-8181(95)00046-1

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., et al. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water Resources Research*, *55*(8), 6499–6516. https://doi.org/10.1029/2019WR025287

Liu, J., Hughes, D., Rahmani, F., Lawson, K., & Shen, C. (2023). Evaluating a global soil moisture dataset from a multitask model (GSM3 v1.0) with potential applications for crop threats. *Geoscientific Model Development*, *16*(5), 1553–1567. https://doi.org/10.5194/gmd-16-1553-2023

Liu, J., Bian, Y., & Shen, C. (2023). Probing the limit of hydrologic predictability with the Transformer network (Version 1). https://doi.org/10.48550/ARXIV.2306.12384

Liu, S., Shi, H., & Sivakumar, B. (2020). Socioeconomic Drought Under Growing Population and Changing Climate: A New Index Considering the Resilience of a Regional Water Resources System. *Journal of Geophysical Research: Atmospheres*, *125*(15), e2020JD033005. https://doi.org/10.1029/2020JD033005

Ma, K., Feng, D., Lawson, K., Tsai, W., Liang, C., Huang, X., et al. (2021). Transferring Hydrologic Data Across Continents – Leveraging Data-Rich Regions to Improve Hydrologic Prediction in Data-Sparse Regions. *Water Resources Research*, *57*(5). https://doi.org/10.1029/2020WR028600

McMahon, T., Finlayson, B., Haines, A., & Srikanthan, R. (1987). Runoff variability: a global perspective. *IASH-AISH*, *168*(1987), 3–11.

Milly, P. C., & Dunne, K. A. (2020). Colorado River flow dwindles as warming-driven loss of reflective snow energizes evaporation. *Science*, *367*(6483), 1252–1255. https://doi.org/10.1126/science.aay9187

Mouatadid, S., Adamowski, J. F., Tiwari, M. K., & Quilty, J. M. (2019). Coupling the maximum overlap discrete wavelet transform and long short-term memory networks for irrigation flow forecasting. *Agricultural Water Management*, *219*, 72–85. https://doi.org/10.1016/j.agwat.2019.03.045

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., et al. (2021). What Role Does Hydrological Science Play in the Age of Machine Learning? *Water Resources Research*, *57*(3). https://doi.org/10.1029/2020WR028091

Neiman, P. J., Schick, L. J., Ralph, F. M., Hughes, M., & Wick, G. A. (2011). Flooding in Western Washington: The Connection to Atmospheric Rivers. *Journal of Hydrometeorology*, *12*(6), 1337–1358. https://doi.org/10.1175/2011JHM1358.1

Oki, T., & Kanae, S. (2006). Global Hydrological Cycles and World Water Resources. *Science*, *313*(5790), 1068–1072. https://doi.org/10.1126/science.1128845

Parajka, J., Kohnivá, S., Merz, R., Szolgay, J., Hlavčová, K., & Blöschl, G. (2009). Comparative analysis of the seasonality of hydrological characteristics in Slovakia and Austria / Analyse comparative de la saisonnalité de caractéristiques hydrologiques en Slovaquie et en Autriche. *Hydrological Sciences Journal*, *54*(3), 456–473. https://doi.org/10.1623/hysj.54.3.456

Prasad, R., Deo, R. C., Li, Y., & Maraseni, T. (2017). Input selection and performance optimization of ANN-based streamflow forecasts in the drought-prone Murray Darling Basin region using IIS and MODWT algorithm. *Atmospheric Research*, *197*, 42–63. https://doi.org/10.1016/j.atmosres.2017.06.014

Rahmani, F., Shen, C., Oliver, S., Lawson, K., & Appling, A. (2021). Deep learning approaches for improving prediction of daily stream temperature in data-scarce, unmonitored, and dammed basins. *Hydrological Processes*, *35*(11), e14400. https://doi.org/10.1002/hyp.14400

Rahmani, F., Lawson, K., Ouyang, W., Appling, A., Oliver, S., & Shen, C. (2021). Exploring the exceptional performance of a deep learning stream temperature model and the value of streamflow data. *Environmental Research Letters*, *16*(2), 024025. https://doi.org/10.1088/1748-9326/abd501

Ralph, F. M., Neiman, P. J., Wick, G. A., Gutman, S. I., Dettinger, M. D., Cayan, D. R., & White, A. B. (2006). Flooding on California's Russian River: Role of atmospheric rivers. *Geophysical Research Letters*, *33*(13). https://doi.org/10.1029/2006GL026689

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., et al. (2019). Process-Guided Deep Learning Predictions of Lake Water Temperature. *Water Resources Research*, *55*(11), 9173–9190. https://doi.org/10.1029/2019WR024922

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Riggs, R. M., Allen, G. H., Wang, J., Pavelsky, T. M., Gleason, C. J., David, C. H., & Durand, M. (2023). Extending global river gauge records using satellite observations. *Environmental Research Letters*, *18*(6), 064027. https://doi.org/10.1088/1748-9326/acd407

Saha, G. K., Rahmani, F., Shen, C., Li, L., & Cibin, R. (2023). A deep learning-based novel approach to generate continuous daily stream nitrate concentration for nitrate data-sparse watersheds. *Science of The Total Environment*, *878*, 162930. https://doi.org/10.1016/j.scitotenv.2023.162930

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research*, *54*(11), 8558–8593. https://doi.org/10.1029/2018WR022643

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., et al. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, 1–16. https://doi.org/10.1038/s43017-023-00450-9

Smirnov, N. (1948). Table for Estimating the Goodness of Fit of Empirical Distributions. *The Annals of Mathematical Statistics*, *19*(2), 279–281. https://doi.org/10.1214/aoms/1177730256

Sun, A. Y., Jiang, P., Mudunuru, M. K., & Chen, X. (2021). Explore Spatio-Temporal Learning of Large Sample Hydrology Using Graph Neural Networks. *Water Resources Research*, *57*(12), e2021WR030394. https://doi.org/10.1029/2021WR030394

Tang, S., Sun, F., Liu, W., Wang, H., Feng, Y., & Li, Z. (2023). Optimal Postprocessing Strategies With LSTM for Global Streamflow Prediction in Ungauged Basins. *Water Resources Research*, *59*(7), e2022WR034352. https://doi.org/10.1029/2022WR034352

Trabucco, A., Zomer, R. J., Bossio, D. A., van Straaten, O., & Verchot, L. V. (2008). Climate change mitigation through afforestation/reforestation: a global analysis of hydrologic impacts with four case studies. *Agriculture, Ecosystems & Environment*, *126*(1–2), 81–97. https://doi.org/10.1016/j.agee.2008.01.015

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., et al. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications*, *12*(1), 5988. https://doi.org/10.1038/s41467-021-26107-z

Tuozzolo, S., Lind, G., Overstreet, B., Mangano, J., Fonstad, M., Hagemann, M., et al. (2019). Estimating River Discharge With Swath Altimetry: A Proof of Concept Using AirSWOT Observations. *Geophysical Research Letters*, *46*(3), 1459–1466. https://doi.org/10.1029/2018GL080771

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention Is All You Need (Version 5). https://doi.org/10.48550/ARXIV.1706.03762

Villarini, G. (2016). On the seasonality of flooding across the continental United States. *Advances in Water Resources*, *87*, 80–91. https://doi.org/10.1016/j.advwatres.2015.11.009

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., et al. (2010). Global threats to human water security and river biodiversity. *Nature*, *467*(7315), 555–561. https://doi.org/10.1038/nature09440

Wada, Y., Reager, J. T., Chao, B. F., Wang, J., Lo, M.-H., Song, C., et al. (2017). Recent Changes in Land Water Storage and its Contribution to Sea Level Variations. *Surveys in Geophysics*, *38*(1), 131–152. https://doi.org/10.1007/s10712-016-9399-6

Wu, H., Adler, R. F., Tian, Y., Huffman, G. J., Li, H., & Wang, J. (2014). Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. *Water Resources Research*, *50*(3), 2693–2717. https://doi.org/10.1002/2013WR014710

Xu, R., Zeng, Z., Pan, M., Ziegler, A. D., Holden, J., Spracklen, D. V., et al. (2023). A global-scale framework for hydropower development incorporating strict environmental constraints. *Nature Water*, *1*(1), 113–122. https://doi.org/10.1038/s44221-022-00004-1

Yang, Y., Lin, P., Fisher, C. K., Turmon, M., Hobbs, J., Emery, C. M., et al. (2019). Enhancing SWOT discharge assimilation through spatiotemporal correlations. *Remote Sensing of Environment*, *234*, 111450. https://doi.org/10.1016/j.rse.2019.111450

Yang, Y., Pan, M., Lin, P., Beck, H. E., Zeng, Z., Yamazaki, D., et al. (2021). Global reach-level 3-hourly river flood reanalysis (1980–2019). *Bulletin of the American Meteorological Society*, *102*(11), E2086–E2105. https://doi.org/10.1175/BAMS-D-20-0057.1

Ye, S., Li, H.-Y., Leung, L. R., Guo, J., Ran, Q., Demissie, Y., & Sivapalan, M. (2017). Understanding Flood Seasonality and Its Temporal Shifts within the Contiguous United States. *Journal of Hydrometeorology*, *18*(7), 1997–2009. https://doi.org/10.1175/JHM-D-16-0207.1

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9). https://doi.org/10.1029/2007WR006716

Zaitchik, B. F., Rodell, M., & Olivera, F. (2010). Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resources Research*, *46*(6). https://doi.org/10.1029/2009WR007811

Zeiler, M. D. (2012, December 22). ADADELTA: An Adaptive Learning Rate Method. Retrieved July 20, 2023, from https://arxiv.org/abs/1212.5701v1

Zhi, W., Feng, D., Tsai, W.-P., Sterle, G., Harpold, A., Shen, C., & Li, L. (2021). From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale? *Environmental Science & Technology*, *55*(4), 2357–2368. https://doi.org/10.1021/acs.est.0c06783

Zhi, W., Ouyang, W., Shen, C., & Li, L. (2023). Temperature outweighs light and flow as the predominant driver of dissolved oxygen in US rivers. *Nature Water*, *1*(3), 249–260. https://doi.org/10.1038/s44221-023-00038-z