Journal of Hydrology: Regional Studies 52 (2024) 101744

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Reconstruction of missing streamflow series in human-regulated catchments using a data integration LSTM model

Arken Tursun [a], Xianhong Xie [a,*], Yibing Wang [a], Yao Liu [a], Dawei Peng [a], Yusufujiang Rusuli [b,c], Buyun Zheng [a]

[a] *State Key Laboratory of Remote Sensing Science, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China*
[b] *Institute of Geographical Science and Tourism, Laboratory of Information Integration and Eco-Security, Xinjiang Normal University, Urumqi 830054, China*
[c] *Xinjiang Key Laboratory of Lake Environment and Resources in Arid Zone, Xinjiang Normal University, Urumqi 830054, China*

ARTICLE INFO

Keywords:
Data integration
LSTM
Multiscale static attributes
Streamflow reconstruction
Human activity
Yellow River Basin

ABSTRACT

Study region: Yellow River Basin in China, where streamflow dynamics were significantly impacted by human activities.
Study focus: We introduced a deep learning-based method, i.e., Data Integration (DI) with Long Short-Term Memory (LSTM), which leverages Global Flood Awareness System (GloFAS) streamflow data. Multiscale (Catchment, River) attributes were incorporated into the DI LSTM to represent human disturbances on land surface. We employed this method to reconstruct daily streamflow series in 60 human-regulated catchments across the Yellow River Basin, and identified the sensitivity of the DI LSTM model to the multiscale attributes.
New hydrological Insights for the Region: Our findings revealed that the DI LSTM model achieved favourable performance in streamflow estimation, with the highest Kling-Gupta efficiency (KGE) reaching up to 0.9, outperforming the Regular LSTM model, which was forced by meteorological variables. Multiscale attributes can enhance the DI model performance, particularly in large catchments with significant human activities. A two-step validation demonstrated the high accuracy of the reconstructed streamflow data across the Yellow River Basin, as the KGEs for streamflow estimation in 40 catchments are over 0.6. In summary, the DI LSTM model shows great potential for reconstructing streamflow in human-regulated catchments in arid regions. The reconstructed daily streamflow data contribute valuable insights for monitoring changing hydrological conditions, especially in regions lacking extensive streamflow monitoring networks.

1. Introduction

Streamflow data is of paramount importance in water resources management. It serves as the cornerstone for effective management in critical areas including agricultural irrigation, industrial water utilization, domestic water supply, hydropower generation, and flood control (Abbott et al., 2019; Hall and Perdigão, 2021; Marvel et al., 2019). Streamflow observations in some countries are not publicly available due to the policy reasons (Feng et al., 2021; Kratzert et al., 2023). Indeed, the absence of long-term daily streamflow data poses a major challenge in providing support for early warning and effective monitoring of hydrological extremes such as

* Corresponding author.
E-mail address: xianhong@bnu.edu.cn (X. Xie).

https://doi.org/10.1016/j.ejrh.2024.101744
Received 28 October 2023; Received in revised form 8 February 2024; Accepted 13 March 2024

droughts and floods (Wei et al., 2023; Yokoo et al., 2022). Therefore, reconstruction of long-term daily streamflow series across data-sparse regions is urgently needed (Gholizadeh et al., 2023; Ibrahim Demir, 2022).

Over the past several decades, the most common approach for producing daily streamflow datasets involves a combination of in situ and satellite observations, along with advanced hydrological models (Elsaadani et al., 2021; Li et al., 2022; Seibert, 2021). For example, Lin et al. (2019) reconstructed the naturalized streamflow at the global scale by employing the variable infiltration capacity (VIC) model. Ghimire et al. (2023) assimilated historical streamflow data from US Geological Survey (USGS) streamflow monitoring sites employing the VIC model. Classical land surface models have the capability to simulate spatiotemporally continuous river discharge at regional or global scales (Wang et al., 2022; Zhu et al., 2020). However, the reconstructed streamflow datasets from these global hydrological models may not guarantee reliability due to their coarse spatial resolution and the potential influences of anthropogenic activities (Althoff et al., 2021; Li et al., 2022). Global simulations of streamflow data can be useful in understanding historical changes in hydrological cycles in natural catchments (Aerts et al., 2022). However, providing reliable daily streamflow data for strongly human-regulated catchments remains a challenging task (Ouyang et al., 2021; Thomas Lees, 2021).

In recent studies, deep learning methods, e.g., long short-term memory (LSTM)-based models, have been widely and successfully applied to streamflow estimation (Hashemi et al., 2022; Hunt et al., 2022). In LSTM-based streamflow modelling, previous studies have shown that LSTM could be successfully used in streamflow regionalization and significantly outperform previous traditional hydrological models (Chen et al., 2022; Frame et al., 2022). Arsenault et al. (2023) investigated the potential of LSTM for streamflow prediction in ungauged basins (PUB) across North America, and demonstrating their ability to simulate streamflow more accurately in ungauged catchments. Moreover, some studies also tried to integrate machine learning (ML) models with data assimilation (DA) methods to improve model performance (Feng et al., 2021; Li et al., 2022). Boucher et al. (2020) proposed a DA method integrated with ML and the GR4J model, and their results showed that ML could be successfully applied for hydrological forecasting. Tang et al. (2023) also indicated that a data integration (DI) LSTM model forced by Global Hydrological Models (GHMs) had great potential in streamflow prediction and outperformed traditional LSTM modelling across 1897 global catchments.

While substantial progress has been achieved in recent studies, the ML or DL methods were mostly concentrated on predicting hydrological variables in data-rich and relatively small and natural catchments (Kratzert et al., 2019; Shen, 2018). Moreover, most studies only employed catchment attributes to represent hydrological processes in the natural catchments (Botterill and Mcmillan, 2023; Gao et al., 2022). To the best of our knowledge, few studies explored the potential of LSTM in reconstructing daily streamflow across Yellow River Basin with significant human interventions. There are still gaps in combining multiscale (Catchment, River) attributes with reanalysis hydrological data to reconstruct credible streamflow data. This integration aims to facilitate the reconstruction

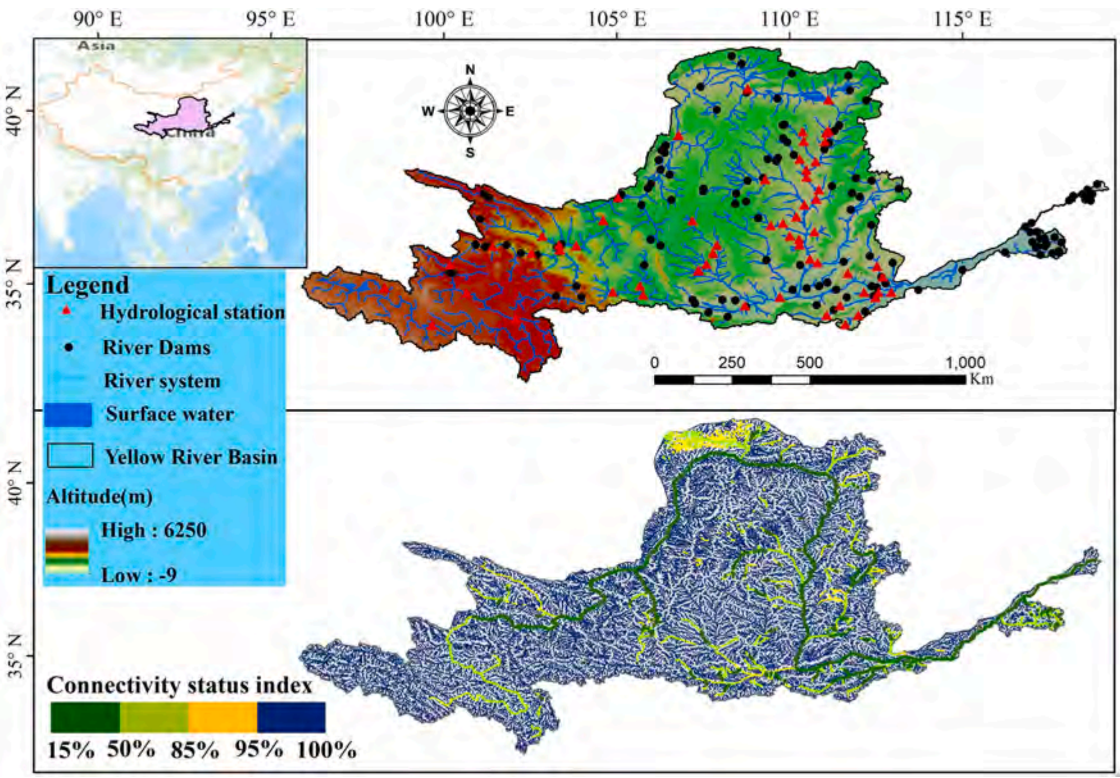

Fig. 1. The locations of stream gauges and river dams (above) and the river connectivity status index (CSI). The CSI value ranges from 0% to 100% for individual river reaches (below). If a value is at or above the threshold (95%), the river reach is considered to have good connectivity; otherwise, it is considered impacted.

of credible streamflow data, especially in catchments where data availability is limited.

This study aims to apply and evaluate our proposed DI LSTM and test the impacts of multiscale attributes on streamflow reconstruction in the Yellow River Basin, China. The specific objectives of this study are as follows: (1) to demonstrate the effectiveness of reanalysis streamflow forced-LSTM (DI LSTM) for reconstructing streamflow in human-regulated catchments; (2) to investigate the influence of different selections of static attributes (Base, River, Catchment, and Multiscale) on model performance; (3) to compare and validate the accuracy of reconstructed streamflow data using a two-step validation method.

2. Materials and methods

2.1. Study area

This study focused on the Yellow River Basin in northern China. This basin covers an area of approximately 750,000 km^2 and supports the livelihoods of 190 million people. The average annual precipitation in the basin is approximately 480 mm. Rainfall typically decreases from southeast to northeast and occurs mainly between June and September (Fig. 1). There are numerous dams and reservoirs across the Yellow River Basin, which have significantly disrupted the streamflow process, resulting in a loss of connectivity in stretches of the Yellow River (Liu et al., 2023; Zhang et al., 2019).

2.2. Data

2.2.1. Meteorological data

Meteorological data were obtained from the China Meteorological Forcing Dataset (CMFD) (He et al., 2020), which was produced by merging a variety of data sources, including China weather station observation data, TRMM satellite precipitation analysis data, GEWEX-SRB data, Modern Era-Retrospective Analysis for Research and Applications (MERRA) data, and the Global Data Assimilation System (GLDAS) data. The CMFD dataset is currently available for the period of 1979–2018 with a spatial resolution of 0.1 arc degrees (∼ 10 km) and a 3 hr temporal resolution. We computed the area-weighted spatial average for each variable in each catchment at daily time-step.

2.2.2. Streamflow observations

Daily streamflow observations (in cubic metres per second; hereafter m^3/s) for 60 catchments in the Yellow River Basin were obtained from the Loess Plateau Subcentre, National Earth System Science Data centre, National Science & Technology Infrastructure of China. The areas of the catchments range from 1298 to 508,878 km^2 (median area 11,401 km^2). Please note that consecutive streamflow data are available from 1979 to 1997. However, the data are completely missing for the period from 1998 to 2002 in 60 catchments. For the period from 2003 to 2018, the streamflow data are only available for 17 catchments (Fig. 2). During the period of missing data, streamflow data may have been recorded but might not be publicly-available.

2.2.3. GloFAS river discharge reanalysis data

The European Centre for Medium-Range Weather Forecasts (ECMWF) generates Global Flood Awareness System (GloFAS) reanalysis streamflow data by applying the global LISFLOOD model. It provides a valuable global dataset of daily streamflow values at a spatial resolution of 0.05°. An evaluation of GloFAS against a global network of 1801 in-situ river discharge observation stations was conducted (Harrigan et al., 2020; Swain et al., 2023), and the result revealed that the GloFAS data exhibits skilful in 86% of catchments, as determined by the modified Kling–Gupta Efficiency Skill Score (KGESS) against a mean flow benchmark. One of the innovative features of these data is their operational production, which enables them to be accessible within 2–5 days after collection, nearly in real-time. The GloFAS-ERA5 river discharge reanalysis product is available at https://cds.climate.copernicus.eu/cdsapp#!/dataset/cems-glofas.

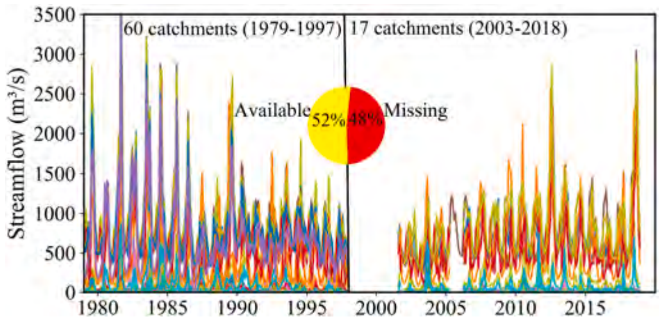

Fig. 2. Consecutive streamflow years from 1979 to 1997 for 60 catchments, missing data from 1998 to 2002 for the all 60 catchments, and from 2003 to 2018 for 17 catchments.

2.2.4. Multiscale attributes

The catchment attributes can be divided into two categories: (i) those obtained from HydroATLAS (Linke et al., 2019) and (ii) others obtained from the daily CMFD time series. The catchment attributes derived from HydroATLAS were from the available shapefile in the dataset (level 12). To derive the catchment attributes, we first computed the spatial join of the HydroATLAS polygons and the catchment boundaries and then calculated the catchment attributes as an area-weighted aggregate (Kratzert et al., 2023).

The Yellow River Basin has undergone significant human regulation. Therefore, it is important to include the river and catchment attributes for LSTM-based streamflow modelling. We obtained river reach scale attributes (Table 1) from the Global Free Flowing River data (FFR) (Grill et al., 2019). The development of this dataset is grounded in the HydroROUT global river routing model based on the HydroSHEDS database (Linke et al., 2019). The FFR dataset contains attribute details for each river reach, encompassing dominant pressure indicators such as degree of regulation (DOR), degree of fragmentation (DOF), consumptive water use (USE), sediment trapping (SED), road density (RDD), and urban areas (URB). Each individual river reach is assigned an integrated connectivity status index (CSI), which measures connectivity levels ranging from 0% to 100%. The geometric dataset provides global river attribute information for every river reach, including values for pressure indicators (DOF, DOR, SED, USE, RDD, and URB) as well as CSI values, are available at https://doi.org/10.6084/m9.figshare.7688801.

It should be noted that the method of calculating river-reach attributes significantly differs from that of catchment attributes. We considered only the river-reach attributes at the outlet of each catchment and did not calculate the average value of all river-reaches within the catchment. This deliberate choice allowed us to represent anthropogenic signatures at the river-reach scale. We combine river-reach attributes with catchment attributes, to form a new type of attribute collection, i.e., multiscale attributes, which represent natural and human related conditions on land surfaces and in river systems. As outlined in Table 1, the multiscale attributes have been derived from various sources, including hydrological, physiographic, climatological, soil and geological, land cover characteristic, and anthropogenic influence factors.

2.3. An overview of DI LSTM

LSTM networks, originally proposed by Hochreiter and Schmidhuber (1997), belong to the family of recurrent neural networks (RNNs) and effectively tackle the challenges posed by vanishing and exploding gradients (Anderson and Radić, 2022; Klotz et al., 2022; Li et al., 2022; Li et al., 2022). The input, forget, and output gates in LSTM directly determine what information to accept, forget, and output, respectively (Fig. 3). These gates are all trained automatically and simultaneously and use input data to predict the target variable (Frame et al., 2022; Wunsch et al., 2022; Yin et al., 2022). The memory cell enables LSTM to learn long-term dependencies such as snow and subsurface water storage, which are needed for streamflow predictions. The base model without DI is referred to as the Regular model, as it does not use GloFAS data.

In this study, the DI LSTM model was employed to enhance streamflow simulations by integrating the GloFAS reanalysis streamflow data alongside the Regular LSTM model inputs. The addition of GloFAS data aimed to address missing data issues and improve the accuracy of streamflow prediction. The hyperparameters governing the DI LSTM model's architecture and training

Table 1
Multiscale (catchment, river) attributes used to train the LSTM.

Static variables	Static variable description	Median	Range
Catchment attributes area	catchment area (km^2)	11401	[1298,508878]
ele_mt	Elevation (m)	1675	[985,4800]
p_mean	Mean daily precipitation (mm d^{-1})	1.9	[0.3,2.7]
ari_ix	Global aridity index (%)	58.6	[37,75]
pnv_pc	Potential extent of mixed forest (%)	0.54	[0,81]
pet_mean	Mean daily potential evaporation (mm d^{-1})	3.7	[2.3,4.3]
aet_mm	Monthly mean actual evapotranspiration (mm)	45	[34,65
high_prec_dur	Average duration of high precipitation events (-)	1.3	[1.2,1.4]
high_prec_freq	Frequency of high precipitation days (-)	0.05	[0,0.8]
low_prec_freq	Frequency of low precipitation days (-)	0.57	[0.50,0.77]
low_prec_dur	Average duration of low precipitation events (-)	4.9	[3.5,8.3]
swc_pc_	Soil water content (%)	58	[30,73]
glc_pc	Forest cover extent (%)	1.45	[0.01,12.6]
cly_pc	Clay fraction in soil (%)	18.2	[10,22]
frac_snow	Fraction of precipitation falling as snow (mm d^{-1})	0.06	[0,0.35]
moisture_index	Mean annual moisture index (-)	-0.5	[-0.75, 0.3]
pst_pc	Pasture cover extent (%)	36	[9,58]
ire_pc	Irrigated area extent (%)	10	[0,31]
River attributes CSI	River connectivity status index (%)	68	[21,99]
DOR	Degree of regulation (%)	46	[0,100]
DOF	Degree of fragmentation (%)	38	[0,100]
URB	Night light intensity in urban areas (%)	8	[0,97]
USE	Water use for irrigation, industry, municipal (%)	13	[0,38]
Num_dams	Number of dams on the river (-)	18	[0,133]

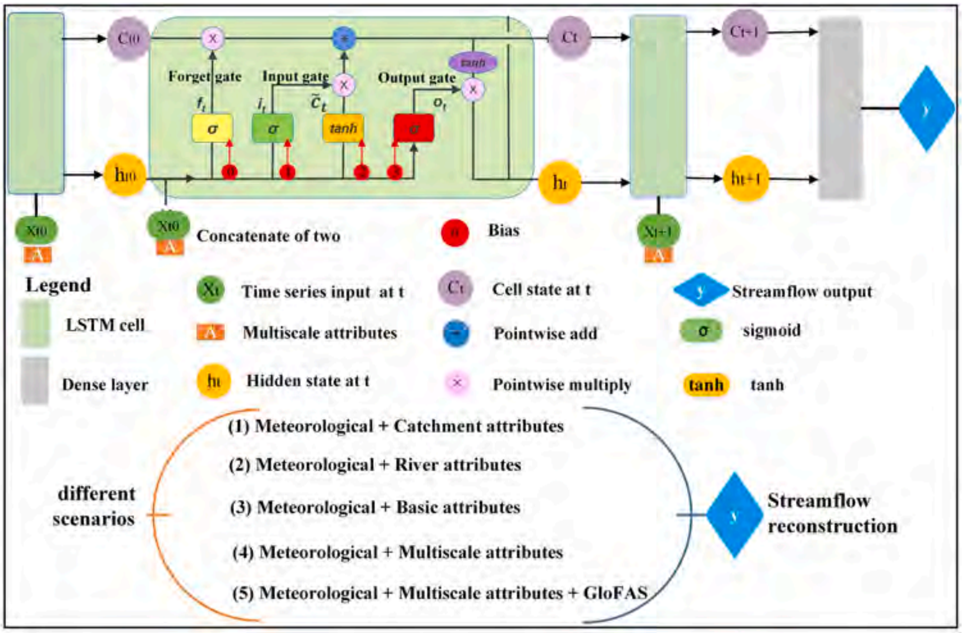

Fig. 3. Structural graph for the DI LSTM neural networks, including memory cells, forgetting gates, input gates, and output gates. The model's input data consists of three components: meteorological data and Multiscale attribute data, and global reanalysis streamflow data which comprise the different integration strategies.

process were consistent with those of the Regular LSTM model, as outlined in Table 2. These parameters typically include the number of LSTM layers, units per layer, learning rate, and dropout rate, among others. The GloFAS reanalysis streamflow data were seamlessly incorporated into the LSTM model as additional input features. This integration allows the model to leverage the rich information provided by GloFAS to refine its prediction and capture a more comprehensive understanding of streamflow dynamics. For both the Regular LSTM and DI models, we utilized the mean square error (MSE) loss function as the objective function. In this study, we tried other loss functions including root mean square error (RMSE), Nash-Sutcliffe Efficiency (NSE). Among these functions, MSE consistently yielded the best performance in optimizing the models' parameters and enhancing their predictive accuracy.

To prepare individual features that describe the topographic, land cover, and climatic properties of each catchment as static inputs, we also considered static attributes at the river reach scale, representing the pressure level on the river reach from human activities. These river and catchment attributes were selected to provide valuable hydrological signatures that enabled the LSTM model to differentiate between streamflow patterns across various catchments. All of these river reaches and catchment attributes are presented in Table 1. To predict the final time step of a specific daily discharge, we provided the LSTM model with 365 days of dynamic data and passed the final hidden output through a fully connected (linear) layer.

2.4. Experimental design

Fig. 4 presents a graphical overview of the proposed DI LSTM, which was supplemented with GloFAS reanalysis streamflow data as one of its inputs for streamflow estimation. The DI LSTM model requires several inputs, including daily meteorological data and daily reanalysis discharge data from GloFAS, as well as multiscale static attributes such as catchment and river scale characteristics. We conducted two-step evaluation to evaluate the performance of the model. In the first step, we reconstructed the daily streamflow data

Table 2
List of DI LSTM model's hyper parameters.

No	LSTM parameters	Value
1	Number of LSTM layers	2
2	Hidden states	124
3	Initial forget bias	3
4	Dropout rate	0.4
5	Learning rate	0.005
6	Batch size	256
7	Optimizer	Adam
8	Number of training epochs	50
9	Sequence length	365
10	Loss function	MSE

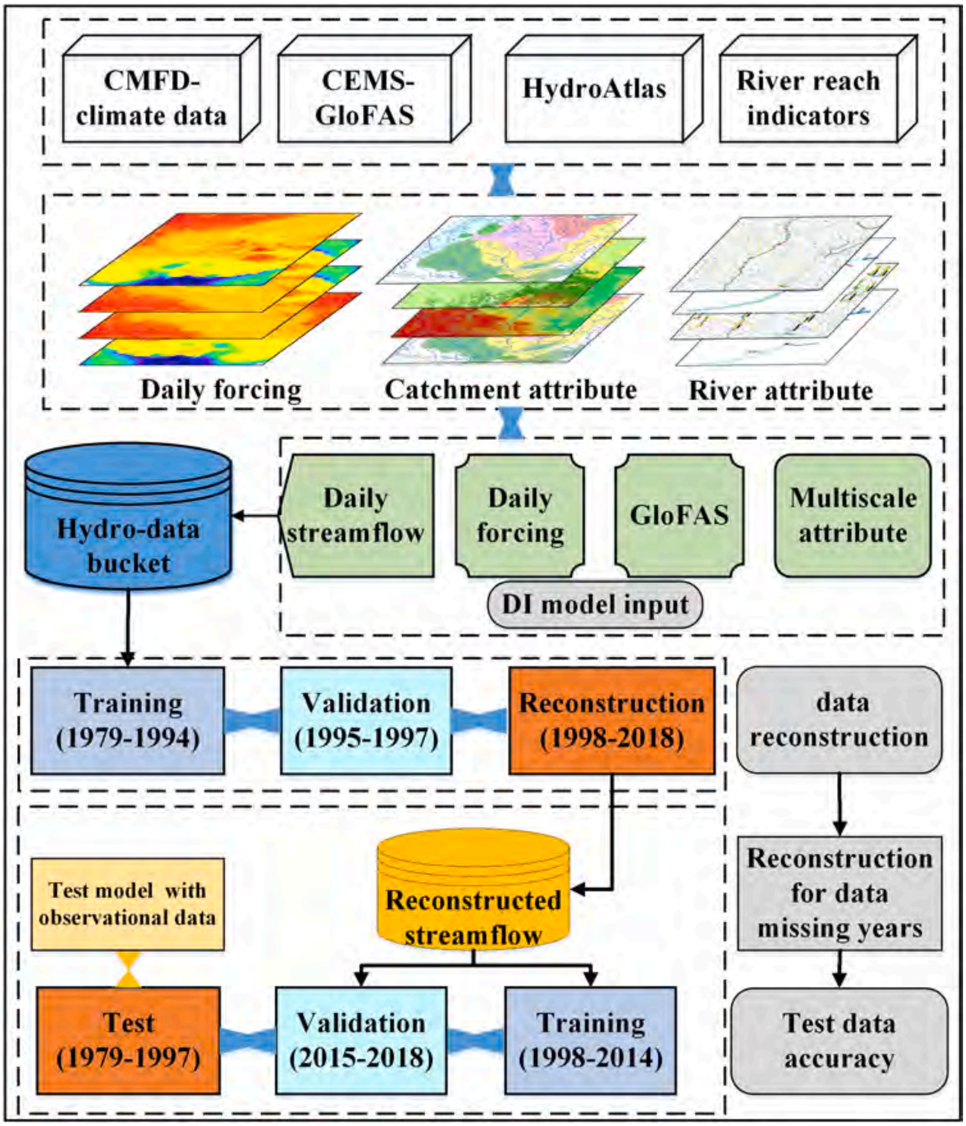

Fig. 4. Sketch of the DI LSTM model framework.

for the years with missing data (1998–2018) for all 60 catchments. However, only 17 catchments had available data to test the model performance. Therefore, further evaluation of the reconstructed data accuracy was necessary. In the second step, we trained and validated the LSTM model using the reconstructed streamflow data and assessed its performance using daily observational data during the test period (1979–1997).

2.4.1. Setup with the GloFAS data

We trained two different LSTM models. The first was referred to as the Regular LSTM model, which was forced by a set of meteorological variables, including temperature, precipitation, wind speed, and net solar radiation, along with basic attributes (catchment area, cropland fraction, forest fraction, soil water content, river connectivity status index, and number of dams). The second model was subjected to the same meteorological forcing and attributes as the first model, but we integrated the GloFAS reanalysis discharge data and named this model as the DI LSTM model. Both the Regular and the DI LSTM models were trained and tested using the same length of training data. These experiments allowed us to further examine whether the GloFAS reanalysis discharge data could provide valuable information for streamflow modelling across both human-regulated and data-limited catchments. We divided streamflow data into three periods (training: 1979–1994, validation: 1995–1997, reconstruction: 1998–2018), while only 17 catchments were available to test model performance during the model reconstruction period.

2.4.2. Setup with the multiscale attributes

The LSTM model can learn hydrological similarity from different groups of static attributes. Thus, we trained the model with different selections of multiscale attribute subsets as inputs, and designed four different scenarios: (1) the basic attribute scenario, in which the LSTM utilized only six basic attributes (catchment area, cropland fraction, forest fraction, soil water content, river connectivity status index, and number of dams); (2) the catchment attribute scenario, in which the LSTM model was fed with all available catchment attributes, as shown in Table 2 (Catchment); (3) the river attribute scenario, in which the model was fed with river attributes, as shown in Table 2 (River); and (4) the multiscale attributes scenario, in which all available catchment and river attributes were used as static attributes in the LSTM model. Please note that in these scenarios, we used the same meteorological variables from the CMFD and the same reanalysis discharge data from GloFAS to force the LSTM models. The four scenarios can identify the contribution of each type of attribute in LSTM-based streamflow modelling. By averaging predictions from these models (four input-selection models × five random seeds = 20 total simulations), we obtained an ensemble result called the input-selection ensemble (Feng et al., 2021). We evaluate the accuracy of the reconstructed streamflow data using the available data from 17 catchments. However, recognizing that the limited dataset may not provide a comprehensive assessment of data reliability, and it could introduce uncertainties for the remaining 43 catchments within the total of 60.

Two-step validation for reconstructed streamflow data

In this experiment, we employed a two-step validation strategy to thoroughly assess the accuracy of our reconstructed streamflow data. In the initial step, we conducted the evaluation with a holistic approach: the Regular LSTM model was trained using the reconstructed streamflow data and meteorological data as inputs. We did not incorporate the GloFAS data in this particular step. Then meticulously evaluated its performance against the observational streamflow data. In the second step, we further compared accuracy of our reconstructed streamflow data with GloFAS streamflow data.

This experiment was interesting and worthwhile. We trained and validated the model using the reconstructed streamflow data (training set: 1998–2014, validation set: 2015–2018) and tested its performance against streamflow observational data (test set: 1979–1997). In addition, we further evaluated the potential of the LSTM model for streamflow data estimation.

2.5. Evaluation metrics

In this study, the simulation accuracy of each model was evaluated using statistical error measurements and the Kling-Gupta efficiency (KGE) criterion (Yilmaz et al., 2008). The KGE criterion is widely employed to evaluate the effectiveness of hydrological models. It is expressed as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \quad (1)$$

where r represents the correlation coefficient between the observed and model-simulated streamflow; a denotes the ratio of the standard deviation of model-simulated flows to the standard deviation of observed flows, and b signifies the ratio of the mean of the

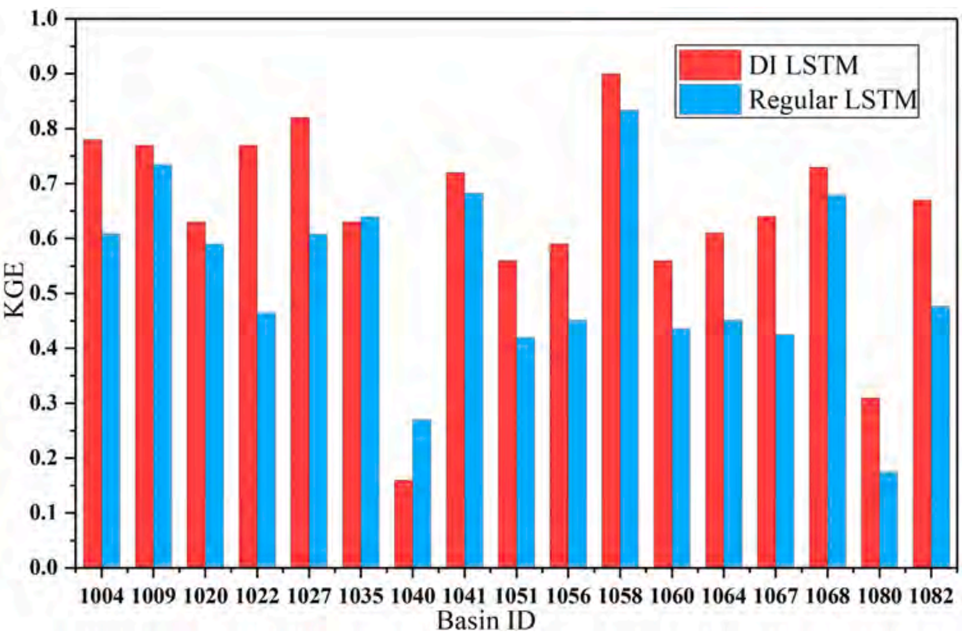

Fig. 5. Performance of the DI LSTM and the Regular LSTM models in the estimation period (1998–2018).

simulated flows to the mean of the observed flows.

Additionally, we included the percent biases of the top 2% high flow values (FHV) and the bottom 30% low flow range (FLV) (Yilmaz et al., 2008), as indicators of the model performance for peak flow and low flow, respectively.

3. Results

3.1. Benefit of integrating the GloFAS reanalysis discharge data

Fig. 5 shows the difference in model performance between the DI LSTM and the Regular LSTM models. The DI LSTM model demonstrates better performance, with a maximum KGE value of 0.9 and a minimum of 0.16. In contrast, the Regular LSTM model exhibits a range of KGE values, reaching a high of 0.83 and a low of 0.17. The mean and median KGE values for the DI LSTM model are 0.638 and 0.621, respectively, but the Regular LSTM model yields mean and median KGE values of 0.521 and 0.477, respectively. The difference in KGE values (ΔKGE) between the DI and Regular LSTM models is substantial, with Δmean KGE and Δmedian KGE values of 0.11 and 0.16, respectively. In summary, the DI LSTM model significantly outperforms the Regular LSTM model, particularly evident in the context of its median KGE value. It is important to note that nearly all catchments, except for Catchment IDs 1035 and 1040, exhibit significant improvements due to the inclusion of GloFAS reanalysis streamflow data. These improvements are particularly prominent in arid catchments and those characterized by lower annual discharge.

In Fig. 6, we presented two catchment hydrographs as examples to illustrate the performance differences between the DI and Regular LSTM models. Notably, substantial improvements were found in Catchment A (ID 1004), where the DI model achieved a KGE value of 0.78, surpassing the Regular model's KGE value of 0.60 (Fig. 6 A). It is evident that the DI LSTM model adeptly simulated both peak flows and baseflow, while the Regular LSTM model underestimated the peaks. Moving to Catchment B (Fig. 6B), the DI and regular LSTM models delivered KGE values of 0.90 and 0.83, respectively. Both models were proficient in generating reliable streamflow simulations and effectively replicated the observed magnitudes of peak flows and baseflows at Catchment ID 1058. These results unequivocally demonstrate the superior performance of the DI LSTM model in comparison to the Regular LSTM model.

3.2. Impacts of multiscale attributes on model performance

The influence of different attributes on the model's performance is evident in Fig. 7. Among the various attribute-selected models, the classical catchment attribute model exhibited poor performance compared to the river attribute model. The integration of both river and catchment attributes in the multiscale attribute model significantly enhanced model performance relative to other individual models. Surprisingly, even the base-attribute model, with only six attributes, outperformed the catchment attribute model. Furthermore, the ensemble model (an average of results from the Catchment, River, Basic, and Multiscale models) improved streamflow

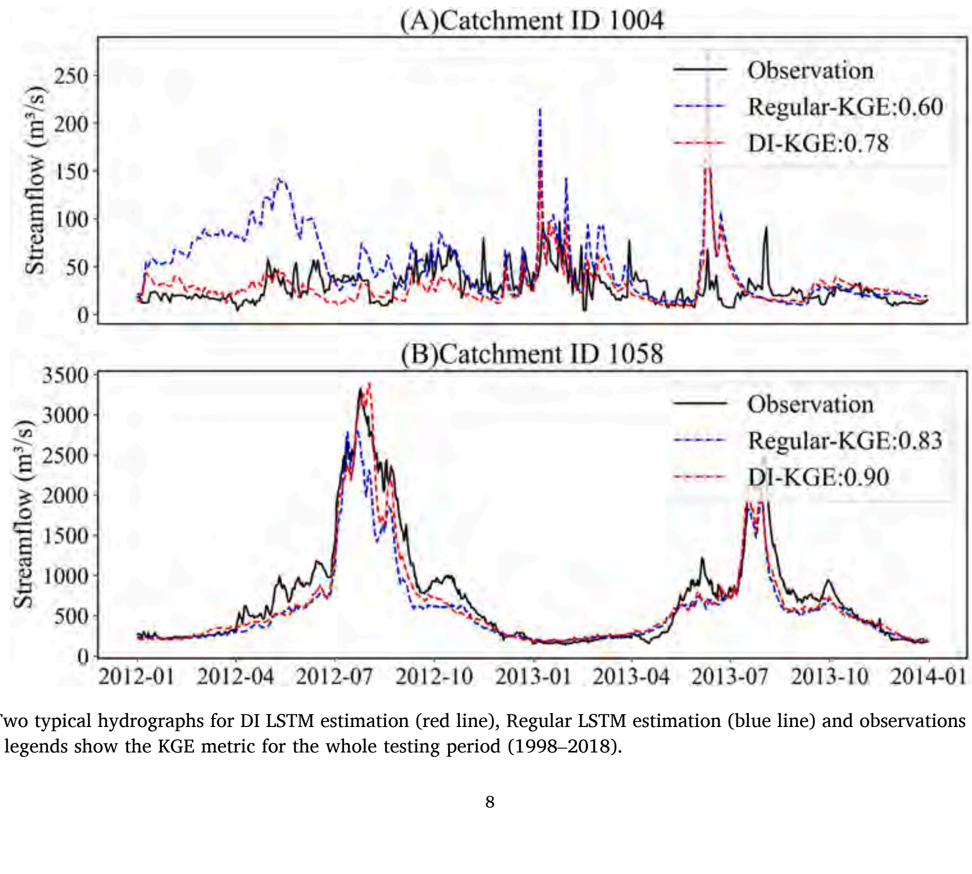

Fig. 6. (A-B). Two typical hydrographs for DI LSTM estimation (red line), Regular LSTM estimation (blue line) and observations (black line). The numbers in the legends show the KGE metric for the whole testing period (1998–2018).

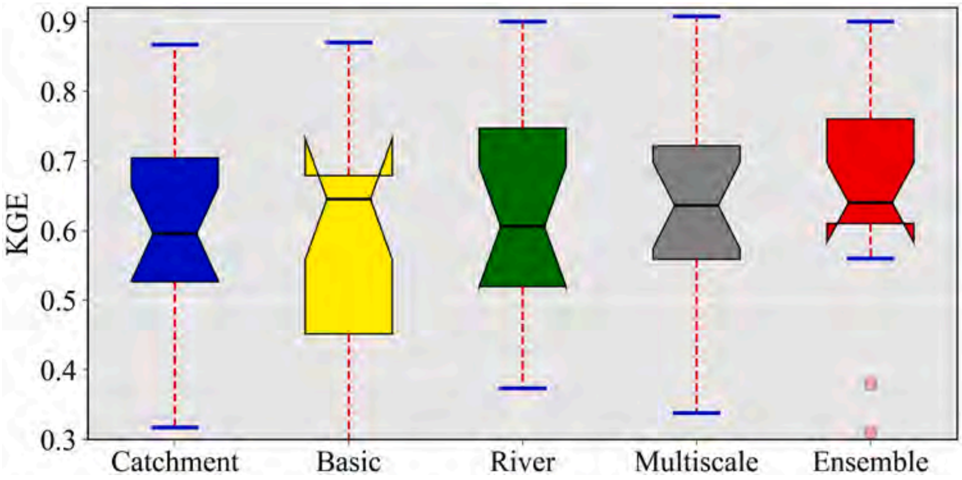

Fig. 7. Performance of including different attributes in LSTM for streamflow reconstruction in data missing years (1998–2018).

prediction in comparison to the individual models. The ensemble model achieved a mean KGE value of 0.65, whereas the other five individual models had a mean KGE value of 0.6, indicating an average improvement in KGE of 0.05 across all basins. The box plots revealed that the lower and upper whiskers of the ensemble model tended to be higher than those of the individual models included in the ensemble. However, it is essential to note that specific catchments still displayed low KGE values even after forming the ensemble. Two catchments notably exhibited lower KGE values in the other five individual models as well, suggesting that there were no significant improvements after combining the individual models.

Fig. 8 provides a comprehensive overview of the DI LSTM model's performance in estimating streamflow across the 17 catchments. The model exhibited notable superiority in larger catchments, consistently achieving KGE scores that exceeded 0.5. Nevertheless, a gradual decline in model performance became evident as we moved towards the middle and lower stream regions. This decline can be attributed primarily to the significant human interventions that exerted a substantial impact on streamflow dynamics, particularly in large and dry catchments. Conversely, the model displayed improved performance in the humid southern catchments characterized by more pronounced human interventions, with KGE scores surpassing 0.60.

3.3. Two-step validation of the reconstructed streamflow

While our initial evaluation in subsection 3.2 provided valuable insights into the model's streamflow reconstruction capabilities, it was limited to a subset of 17 catchments. Therefore, we pursued a comprehensive assessment of the reconstructed streamflow data in two distinct phases. In the first phase, we trained and validated a regional LSTM model using reconstructed streamflow data from 1998 to 2018. We then rigorously assessed the model's performance by comparing its predictions with daily observational data recorded

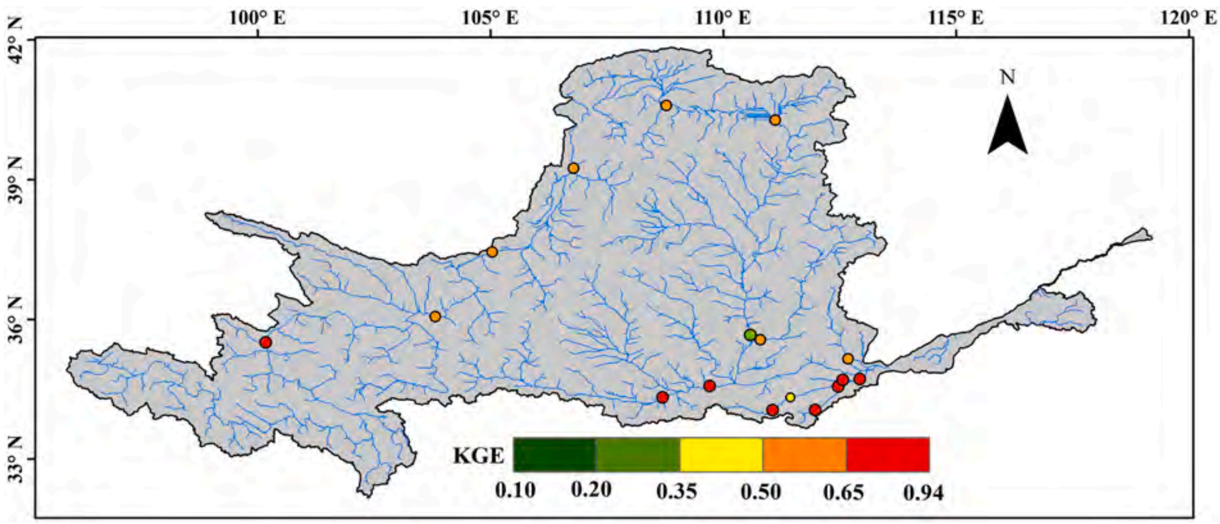

Fig. 8. KGE spatial patterns of the ensemble results for the DI LSTM model.

between 1979 and 1997. In the second phase, we conducted a meticulous comparative analysis by comparing the reconstructed streamflow data with the GloFAS reanalysis data.

Fig. 9 shows the spatial KGE distribution of the LSTM during the test period. The notably high KGE values observed during the evaluation period were somewhat unexpected, as our model had been trained using the reconstructed streamflow data and subsequently assessed against observational data. Initially, we held the assumption that the model's performance might be compromised in human-regulated catchments. However, our findings revealed the model's proficiency across both upstream and downstream catchments. Even within large and heavily human-regulated catchments, the model continued to deliver commendable performance, achieving a KGE value of 0.5. Among the 60 catchments, a total of 40 exhibited KGE values exceeding 0.6, while 50 others surpassed the 0.3 threshold, leaving just 4 catchments with KGE values falling below 0.2. These empirical findings provide compelling evidence of the LSTM's proficiency in daily streamflow estimation, particularly within catchments subject to substantial human regulation. Predominantly, those catchments with KGE values below 0.4 were concentrated within the middle stream regions, characterized by both notably low annual streamflow and extensive human intervention. As a consequence, within these challenging catchments, the LSTM model consistently exhibited a proclivity to overestimate baseflow while concurrently underestimating the magnitudes of flow peaks.

Fig. 10 presents the spatial distribution of model performance for two distinct flow regimes observed during the test period. Regarding FHV, which represents the top 2% of peak flow ranges, it is evident that most catchments exhibit FHV values lower than −40, indicating that the LSTM tends to underestimate peak flow in certain catchments. Conversely, catchments with FHV values higher than 0 are primarily situated in the upper and lower reaches of the river, where the DI LSTM model tends to overestimate the peak flow. As for FLV, representing the bottom 30% of low flows, it is noticeable that in the majority of arid and human-regulated catchments in the middle reaches, the LSTM model tends to overestimate baseflow, particularly in catchments with FLV values greater than 80. While those with lower FLV values are primarily found in the lower reaches. These findings highlight the challenges the LSTM model faced in accurately simulating baseflow within catchments characterized by significant aridity and human regulation.

The GloFAS data cover the period from 1979 to present, but we only have streamflow observations for 17 catchments from 1979 to 2018. To ensure a consistent comparison, we focused on the overlapping period of 1979–1997. It is necessary to assess the accuracy of different types of models for streamflow reconstruction. The △KGE ($KGE_{reconstructed}$ -KGE_{GloFAS}) clearly indicated the remarkable performance of the LSTM model (Fig. 11). We observed a significant improvement in performance for the LSTM model, with a mean △KGE of 0.64 and a median △KGE of 0.46, respectively. Among the 60 catchments, △KGE values were above zero for 59 of them. Which implies that LSTM model notably improved the accuracy across Yellow River Basin. However, the daily streamflow obtained by the GloFAS is considered unreliable due to substantial human intervention across the Yellow River Basin. Consequently, a noticeable difference in KGE values exists between our reconstructed streamflow and the GloFAS data, particularly in the middle and downstream regions. These areas are predominantly arid and subject to significant human intervention.

In Fig. 12, we present three catchment (from upstream to downstream) hydrographs to compare the accuracy of the reconstructed streamflow with that of GloFAS. In Catchment A (Fig. 12 A), the reconstructed streamflow had a KGE value of 0.86, while the GloFAS data yielded a KGE value of 0.66 in the source region. Both reconstructed streamflow and GloFAS data successfully captured the baseflow, but the reconstructed streamflow data closely represented the hydrological process and appeared similar to the real observational series. For Catchment B (Fig. 12B), the reconstructed streamflow and the GloFAS data exhibited KGE values of 0.48 and −0.37, respectively, resulting in a △KGE of 0.85. Notably, the reconstructed streamflow data underestimated the peak flow, but they effectively simulated the streamflow pattern in this human-regulated catchment. Conversely, the GloFAS streamflow data overestimated both the peak flow and baseflow, leading to noticeable differences in the △KGE in the middle reaches. Last, in Catchment C,

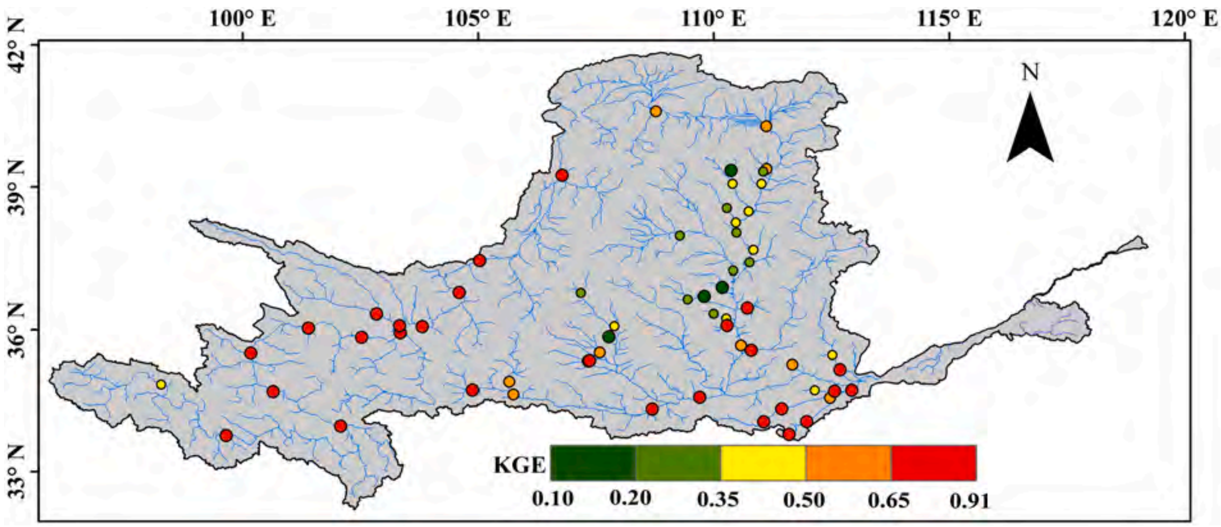

Fig. 9. Performance of the LSTM for streamflow estimation for 1979–1997.

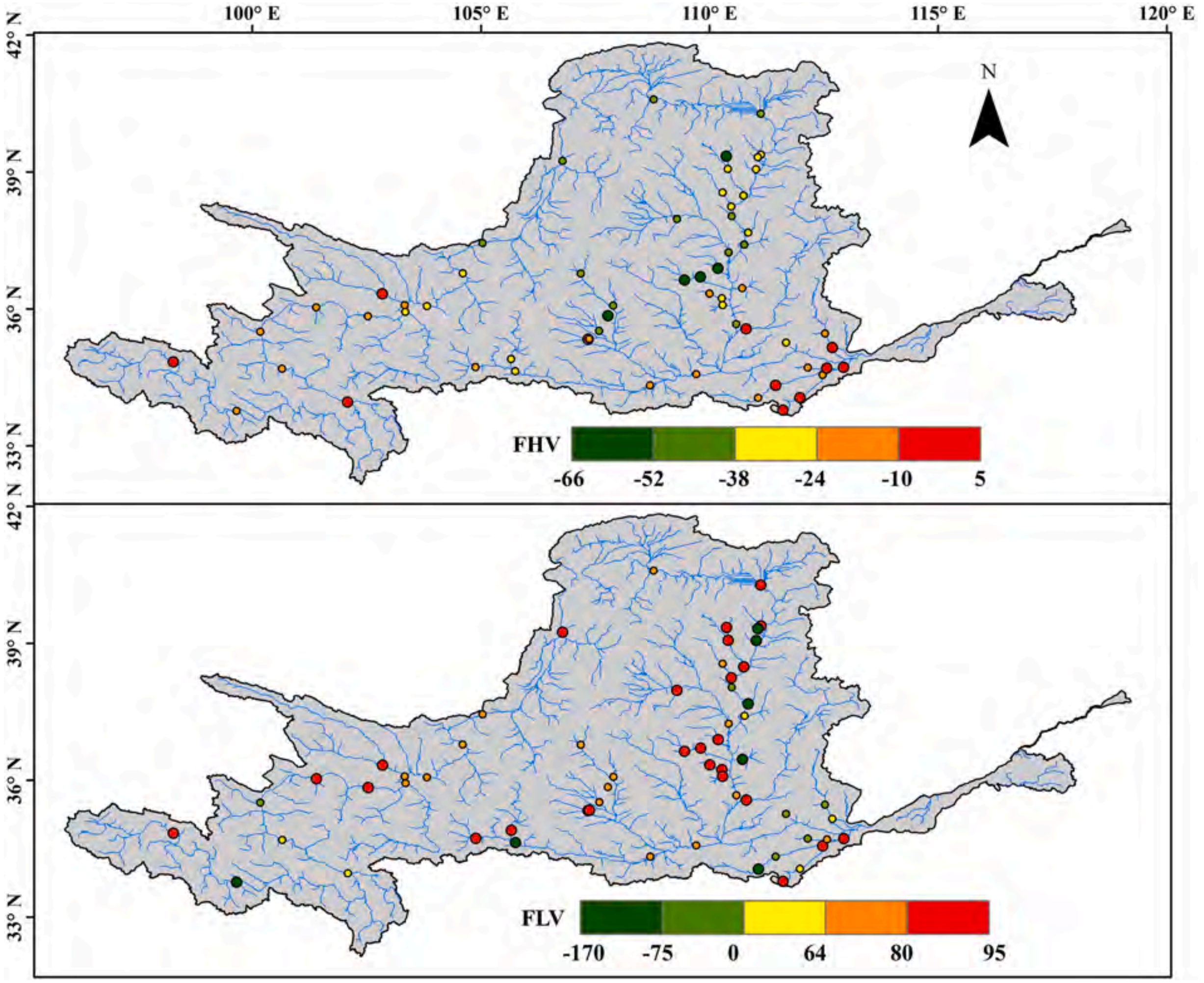

Fig. 10. Spatial patterns of the low flow percent bias (FLV) and high flow percent bias (FHV).

located downstream with more significant anthropogenic intervention compared to that in the upper regions (Fig. 12 C), the reconstructed streamflow had a KGE value of 0.78, while the GloFAS data had a KGE value of −0.90. The substantial difference of 1.68 in the △KGE between the two datasets was highly significant. Indeed, it is noteworthy that the reconstructed streamflow data closely approached the level of the real observational data in strongly human-regulated catchment, whereas the GloFAS streamflow data exhibited a lower accuracy. These examples vividly demonstrated the significant improvements achieved by the LSTM model compared to the GloFAS data.

period (1979–1997).

3.4. Variation in the reconstructed streamflow

Utilizing the reconstructed streamflow data, we conducted an in-depth analysis of seasonal and annual streamflow trends spanning the last four decades (1979–2018) across the 60 catchments within the Yellow River Basin. The results of this trend analysis unveiled distinctive regional variations in streamflow patterns, as depicted in Fig. 13. Notably, discernible surges in streamflow were observed across all four seasons and annually, denoted by the red and yellow colorations, respectively. These trends were predominantly witnessed in the upper and middle reaches of the Yellow River. Conversely, the lower reaches displayed a contrasting pattern with consistent declines in streamflow across all seasons, marked by the green and dark green colorations, respectively. Within catchments situated in the middle and lower reaches, a decreasing trend was observed in summer, while an increasing trend was noted during the winter season. Specifically, our analysis revealed that an increase in flow was observed in 80% of the catchments during winter (40 catchments), with 57% (34 catchments) and 52% (31 catchments) displaying rising trends in fall and summer, respectively. However, during the spring season, only 38% (23 catchments) of the catchments exhibited an increase in flow. Conversely, a declining trend in flow was observed in 20% (12 catchments) of the catchments during winter and in 43% (26 catchments) and 48% (29 catchments) of the watersheds in fall and summer, respectively. A notable 62% (37 catchments) of the catchments showed declining trends during the

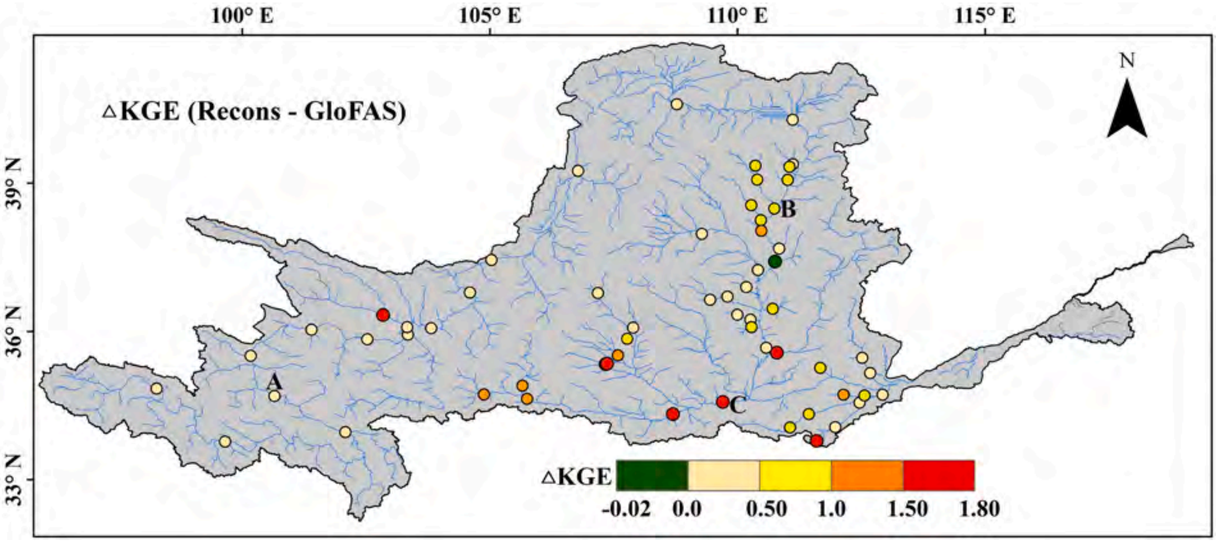

Fig. 11. Spatial distribution of the difference in KGE values between the reconstructed and the GloFAS streamflow across Yellow River Basin stream gauges. (A-C) Locations of hydrograph comparisons between the two data sets for the three catchments.

spring season. Additionally, we noted a decrease in flow within 42% of the catchments when assessing the annual trend (25 catchments), with the majority of these catchments distributed in the lower reaches of the basin.

Fig. 14 presents an overview of the annual peak flow trend across the Yellow River Basin. During the extensive period from 1979 to 2018, it is noteworthy that approximately 12% of all catchments, totalling 7 out of 60, exhibited an increasing trend. However, it is crucial to emphasize that the predominant nature of these trends was negative, with a median ranging between −10 and 0 $m^3.s^{-1}.\ a^{-1}$, indicating a consistent and clear declining trend in annual peak flow until 2018. In the upper reaches, certain catchments displayed increasing trends, and a majority of the positive trends were concentrated in the source region. Analysing the annual peak flow trend holistically, there is a distinct propensity towards declining peak flow overall, with more pronounced declines observed in the lower reaches. Simultaneously, with the exception of three catchments, the middle and lower reaches generally exhibited minimal declines, and in some instances, even displayed increasing trends. This observation suggests a potential impact of anthropogenic interventions in these regions on peak flow dynamics, particularly in the lower reaches.

4. Discussion

4.1. Reliability of the reconstructed streamflow

In this study, we employed a novel two-step validation strategy to thoroughly assess the accuracy of the reconstructed streamflow data. This approach significantly differs from previous studies (Aerts et al., 2022; Gholizadeh et al., 2023; Sadler et al., 2022). In the first step, we conducted an evaluation of the accuracy of the reconstructed streamflow data using the available data from 17 catchments. We trained our LSTM model using the reconstructed streamflow data and then meticulously evaluated its performance against observational streamflow data. We hypothesized that if the reconstructed data were indeed reliable, the model would have acquired valuable insights during the training period, resulting in exceptional performance during the evaluation phase when evaluated with real observational data. As anticipated, the model trained with the reconstructed data demonstrated excellent performance across all 60 catchments, confirming the reliability and accuracy of our reconstructed streamflow data across Yellow River Basin (Fig. 9).

In previous studies, the GloFAS reanalysis data demonstrated proficiency in 86% of the 1801 case study catchments across the globe (Chen et al., 2019; Harrigan et al., 2020; Zhao et al., 2022). In the second step, we conducted a comprehensive assessment of the accuracy of the GloFAS discharge data compared to our reconstructed streamflow data. Our investigations unveiled the substantial potential of the reconstructed streamflow data generated by the DI LSTM model for effectively representing hydrological processes (Fig. 11). In contrast, the accuracy of GloFAS streamflow data within the Yellow River Basin was notably inferior. Notably, the reanalysis discharge data exhibited significant discrepancies, encompassing both overestimations and underestimations of streamflow within the Yellow River Basin. These discrepancies were particularly evident when compared to actual observational data, as illustrated in Fig. 12. This variance is attributable to the absence of static attributes modules within the LISFLOOD river channel routing module, a limitation that adversely impacted the model's performance throughout the Yellow River Basin. Drawing from these comparative analyses, we can confidently affirm the reliability and accuracy of the reconstructed streamflow data derived from the DI LSTM across Yellow River Basin.

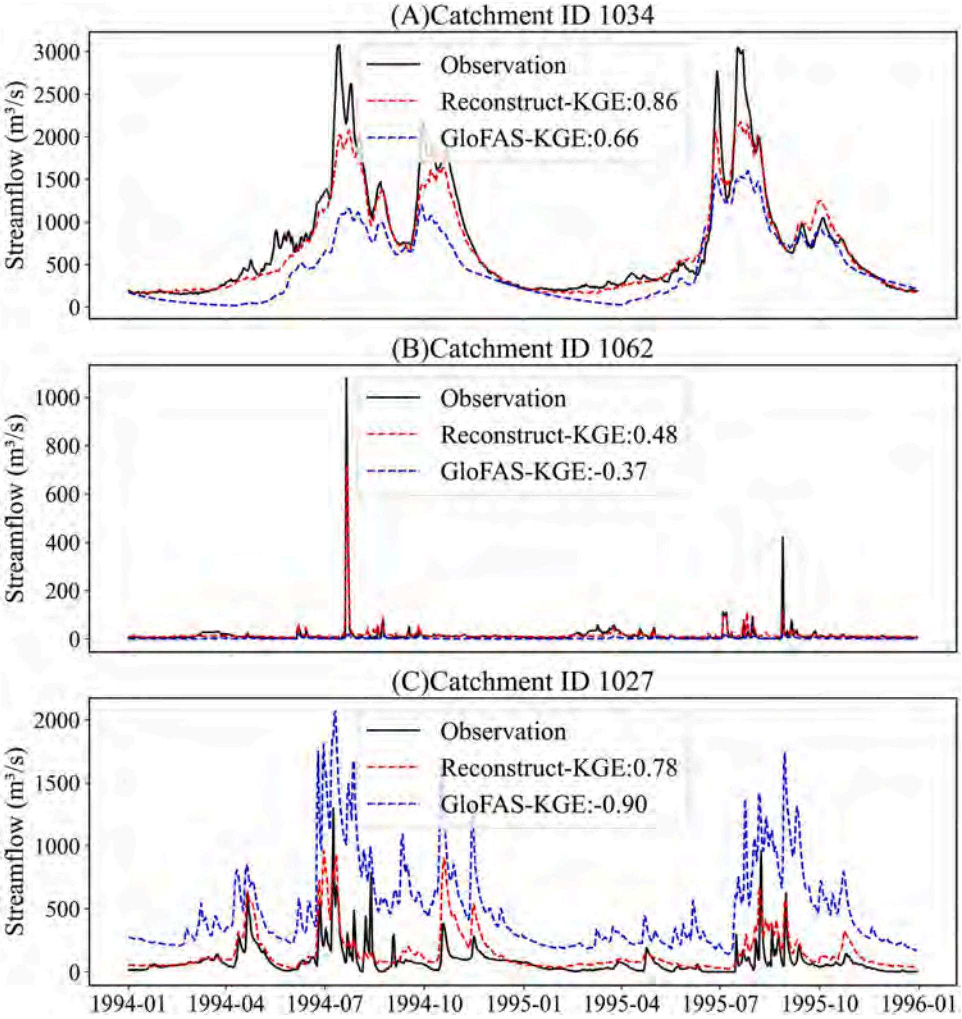

Fig. 12. (A–C) Hydrograph comparisons between different datasets for three typical basins. The numbers in the legends show the KGE metric for the whole testing.

4.2. Implications for streamflow reconstruction

Prior work has documented the success of modelling streamflow with LSTM in reference basins with minimal anthropogenic impacts (Feng et al., 2022; Gauch et al., 2021). Here we mainly focus on how to reconstruct missing streamflow series within exceedingly large and human-regulated catchments. We proposed to combine river-reach and multiscale attributes to create new sets of multiscale attributes, including the sufficient hydrological signatures about catchment characteristics and anthropogenic signals. Furthermore, we integrate the GloFAS data with meteorological forcing data to develop data integration (DI) LSTM model. Our results demonstrate that the Multiscale DI LSTM model yields superior simulations compared to the Regular LSTM model when applied to large and human-regulated catchments (Fig. 5). The high performance for the significantly regulated catchments was somewhat unexpected, because we had earlier thought that diverse human impacts would create immense challenges for DI LSTM and there may not be reliable mapping relationships that could be learned in the human-regulated catchments. The implication of our study is that incorporating GloFAS reanalysis streamflow data as inputs to the LSTM model improves its performance in catchments with sparse gauge data. Moreover, using multiscale attributes as static inputs to the LSTM in human-regulated catchments improves the model performance over equivalent models that only use classical catchment-averaged attributes. Therefore, the DI LSTM model with multiscale attributes has a great potential to learn intricated hydrological process and produce reliable streamflow data in large and human-regulated catchments.

4.3. Limitation and future work

While the DI LSTM model demonstrated great promise here, we still do not consider this method a silver bullet that solves all

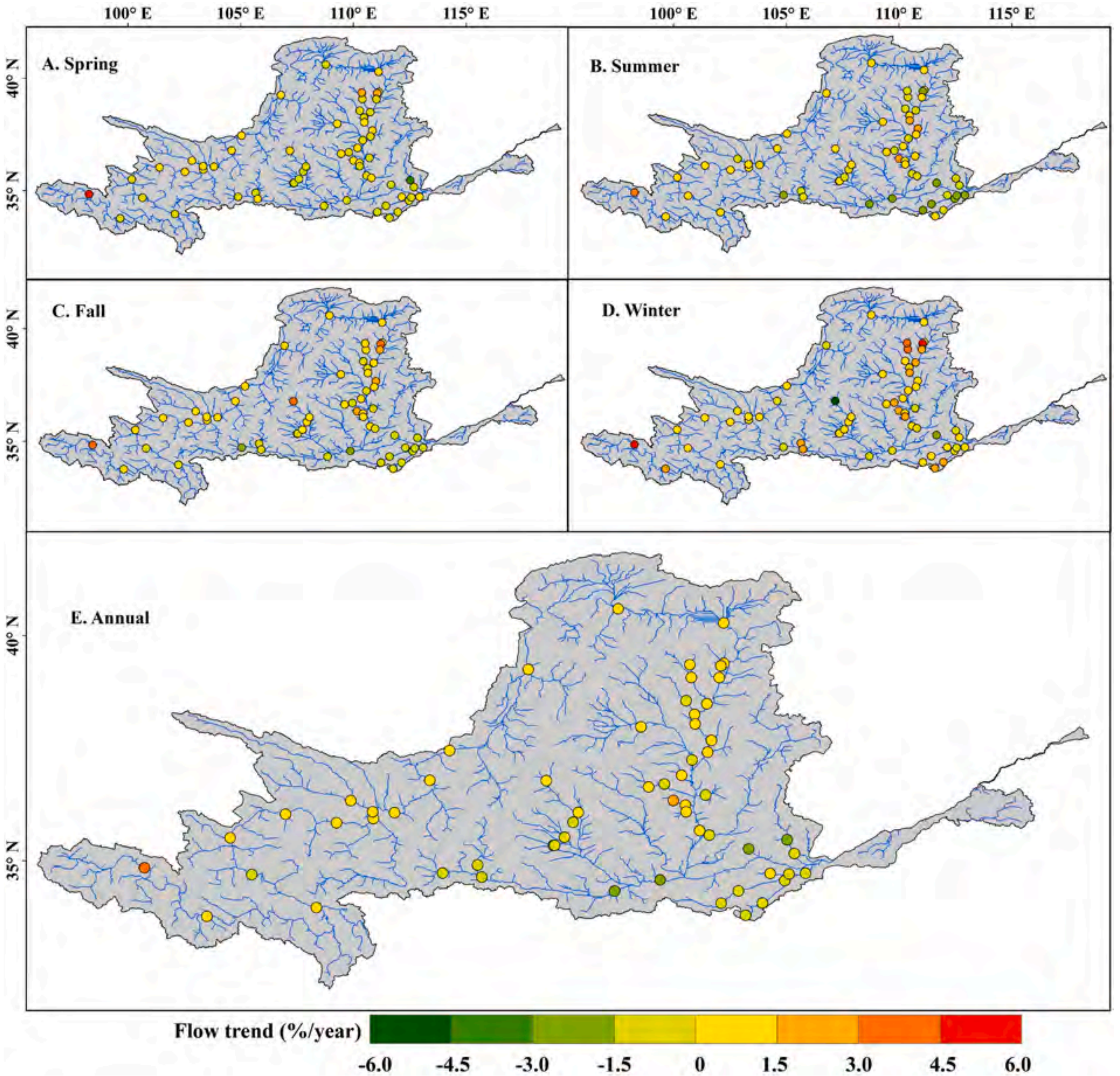

Fig. 13. (A–E) Seasonal and annual streamflow trends over 40 years (1979–2018). (A) Spring (March, April and May); (B) Summer (June, July and August); (C) Fall (September, October and November); (D) Winter (December, January and February). (E) Annual.

hydrological problems. There are several limitations to consider in this study. First, the DI model faces challenges in certain human-regulated catchments in the middle reaches. These catchments are characterized by 1-day flash flows due to reservoir regulation or other water use activities (Liu et al., 2023). Second, the complexity of the training strategy in the LSTM model introduces sources of uncertainty (Cho and Kim, 2022; Klotz et al., 2022; Moosavi et al., 2022). Factors such as model structure and hyperparameter tuning, including the number of hidden layers, units, and learning rates, can significantly influence model performance, and decisions regarding the training period and data splitting for validation are essential. Furthermore, while it is evident that multiscale attributes have improved model performance when compared to classical catchment attributes, it is crucial to acknowledge that the assumption of treating multiscale attributes as static may not align with the dynamic nature of these attributes in the real world, especially over extended time periods. For future studies, it is imperative to develop strategies that enable the integration of dynamic or evolving multiscale attributes or features into the modelling framework.

5. Conclusions

To reconstruct the missing streamflow data across the Yellow River Basin, this study proposes a DI method based on LSTM. In

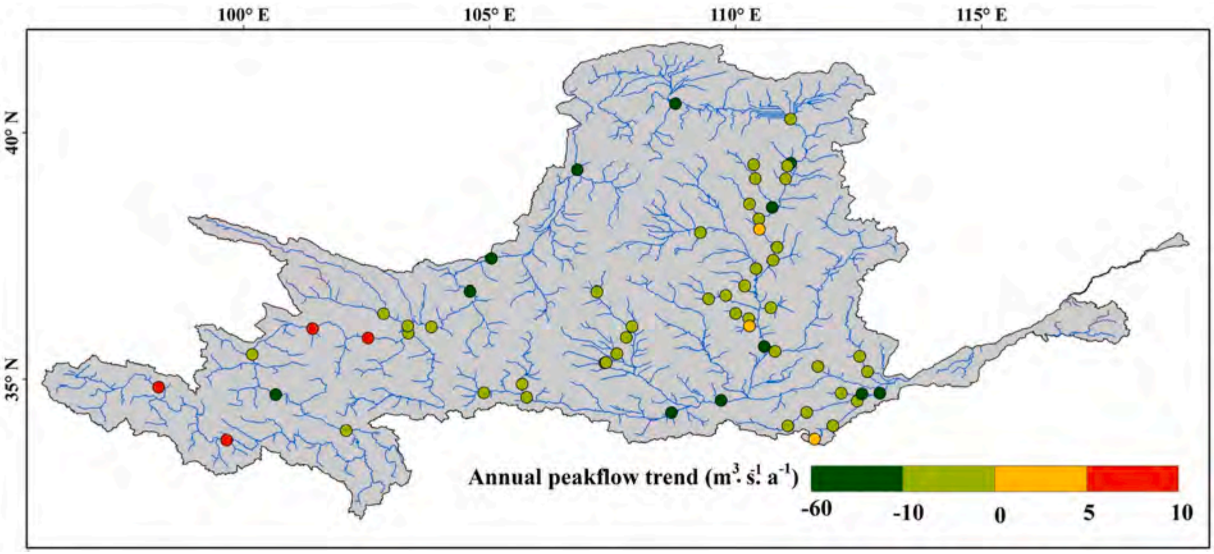

Fig. 14. Annual peak flow trends over 40 years from reconstructed data (1979–2018).

human-regulated catchments, we incorporated multiscale attributes within the DI LSTM to enhance the model performance. The validity of the multiscale DI LSTM and the reconstructed data in this study is verified through comparisons with global reanalysis streamflow data and benchmark models. The following key conclusions can be drawn from our research:

(1) The DI LSTM outperformed the Regular LSTM for streamflow estimation with a maximum KGE of 0.9 (min 0.16), while the KGE of the Regular LSTM ranged from 0.17 to 0.83. The DI LSTM can achieve satisfactory estimation for streamflow series across data sparse catchments like the Yellow River Basin, as indicated by the KGE values.
(2) The DI LSTM with multiscale attributes significantly outperformed other individual models in larger catchments which were disturbed by human activities, and the river attributes are more contributory than the catchment attributes. The input-selection ensemble (average of individual models) enhanced the model performance compared to other individual models.
(3) The two-step evaluation results indicate that the reconstructed streamflow data exhibits high accuracy in large human-regulated catchments across Yellow River Basin. We further suggest that our reconstructed streamflow data exhibit superior performance compared to the global reanalysis streamflow data (GloFAS). The long-term reconstructed data for the period 1979–2018 showed that streamflow declined in the mid-lower reaches of the Yellow River Basin and increased in the source region.

There are some limitations in this work. First, the complexity of the training strategy in the LSTM may introduce uncertainties, which makes the LSTM struggle to capture streamflow patterns in some catchments. Second, treating multiscale attributes as static may not well align with the dynamic nature of the land surface conditions, which implies the necessity of developing dynamic attributes in the LSTM framework. Despite these limitations, however, our study provides evidence that the proposed DI LSTM model effectively learns streamflow generation processes in human-regulated catchments. This suggests that when estimating streamflow in data-scarce and human-regulated catchments, it is an encouraging approach to take into account reanalysis streamflow and multiscale attributes as background information. Future research should explore the potential of including other reanalysis discharge data to improve the model performance across varying hydro-climatological conditions.

CRediT authorship contribution statement

Xianhong Xie: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Yibing Wang:** Investigation, Data curation. **Arken Tursun:** Writing – original draft, Visualization, Validation, Software, Methodology, Conceptualization. **Buyun Zheng:** Visualization. **Dawei Peng:** Software, Data curation. **Yusufujiang Rusuli:** Investigation, Formal analysis. **Yao Liu:** Validation, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Data will be made available on request.

Acknowledgements

This study was supported by a grant from the National Natural Science Foundation of China (No. 42271021) and Supported by Open Fund of State Key Laboratory of Remote Sensing Science and Beijing Engineering Research Center for Global Land Remote Sensing Products (No. OF202202). Acknowledgement for the data support from "Loess Plateau SubCenter, National Earth System Science Data Center, National Science & Technology Infrastructure of China. (http://loess.geodata.cn)". We would like to thank Frederik Kratzert for releasing"neuralhydrology" codebase.

Appendix A

Fig. A1 shows the eight attributes of all catchments, including potential evaporation, aridity, low and high precipitation duration days, cropland and forest fraction, elevation and area. The catchments located in the centre have the highest aridity index, low precipitation duration days and highest cropland fraction. The catchments range in size from 1298 to 508,878 km^2, and the elevation ranges from 568 to 4497 m.

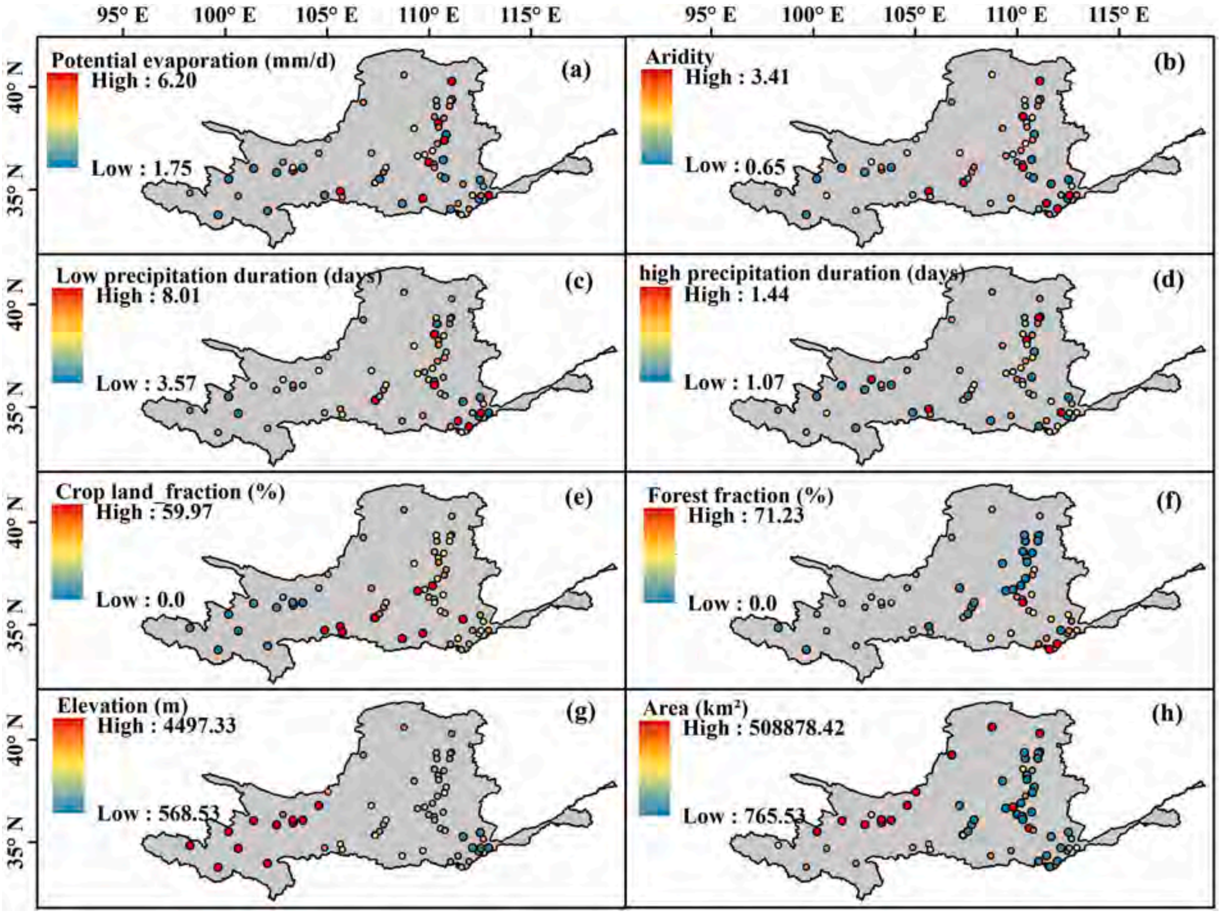

Fig. A1. Overview of eight catchment attributes: (a) mean daily potential evaporation, (b) aridity index, (c) low precipitation duration, (d) high precipitation duration, (e) cropland formation fraction, (f) forest formation fraction, (g) mean elevation, and (h) area of the catchments.

References

Abbott, B.W., Bishop, K., Zarnetske, J.P., Minaudo, C., Chapin, F.S., Krause, S., Hannah, D.M., Conner, L., Ellison, D., Godsey, S.E., Plont, S., Marçais, J., Kolbe, T., Huebner, A., Frei, R.J., Hampton, T., Gu, S., Buhman, M., Sara Sayedi, S., Ursache, O., Chapin, M., Henderson, K.D., Pinay, G., 2019. Human domination of the global water cycle absent from depictions and perceptions. Nat. Geosci. 12 (7), 533–540 https://coi.org/10.1038/s41561-019-0374-y.

Aerts, J.P.M., Hut, R.W., van de Giesen, N.C., Drost, N., van Verseveld, W.J., Weerts, A.H., Hazenberg, P., 2022. Large-sample assessment of varying spatial resolution on the streamflow estimates of the wflow_sbm hydrological model. Hydrol. Earth Syst. Sc. 26 (16), 4407–4430 https://coi.org/10.5194/hess-26-4407-2022.
Althoff, D., Rodrigues, L.N., Silva, D.D.D., 2021. Addressing hydrological modeling in watersheds under land cover change with deep learning. Adv. Water Resour. 154, 103965 https://coi.org/10.1016/j.advwatres.2021.103965.
Anderson, S., Radić, V., 2022. Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling. Hydrol. Earth Syst. Sc. 26 (3), 795–825 https://coi.org/10.5194/hess-26-795-2022.
Arsenault, R., Martel, J., Brunet, F., Brissette, F., Mai, J., 2023. Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models. Hydrol. Earth Syst. Sc. 27 (1), 139–157 https://coi.org/10.5194/hess-27-139-2023.
Botterill, T.E., Mcmillan, H.K., 2023. Using machine learning to identify hydrologic signatures with an encoder–decoder framework. Water Resour. Res 59 (3) https://coi.org/10.1029/2022WR033091.
Boucher, M.A., Quilty, J., Adamowski, J., 2020. Data assimilation for streamflow forecasting using extreme learning machines and multilayer perceptrons. Water Resour. Res 56 (6) https://coi.org/10.1029/2019WR026226.
Chen, K., Guo, S., Wang, J., Qin, P., He, S., Sun, S., Naeini, M.R., 2019. Evaluation of glofas-seasonal forecasts for cascade reservoir impoundment operation in the upper yangtze river. Water-Sui 11 (12), 2539 https://coi.org/10.3390/w11122539.
Chen, X., Wang, S., Gao, H., Huang, J., Shen, C., Li, Q., Qi, H., Zheng, L., Liu, M., 2022. Comparison of deep learning models and a typical process-based model in glacio-hydrology simulation. J. Hydrol. 615, 128562 https://coi.org/10.1016/j.jhydrol.2022.128562.
Cho, K., Kim, Y., 2022. Improving streamflow prediction in the wrf-hydro model with lstm networks. J. Hydrol. 605, 127297 https://coi.org/10.1016/j.jhydrol.2021.127297.
Elsaadani, M., Habib, E., Abdelhameed, A.M., Bayoumi, M., 2021. Assessment of a spatiotemporal deep learning approach for soil moisture prediction and filling the gaps in between soil moisture observations. Front. Artif. Intell. 4 https://coi.org/10.3389/frai.2021.636234.
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. Geophys Res Lett. 48 (14) https://coi.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 Resour. Res 58 (10) https://coi.org/10.1029/2022WR03240⟨4⟩.
Frame, J.M., Kratzert, F., Klotz, D., Gauch, M., Shalev, G., Gilon, O., Qualls, L.M., Gupta, H.V., Nearing, G.S., 2022. Deep learning rainfall–runoff predictions of extreme events. Hydrol. Earth Syst. Sc. 26 (13), 3377–3392 https://coi.org/10.5194/hess-26-3377-2022.
Gao, S., Zhang, S., Huang, Y., Han, J., Luo, H., Zhang, Y., Wang, G., 2022. A new seq2seq architecture for hourly runoff prediction using historical rainfall and runoff as input. J. Hydrol. 612, 128099 https://coi.org/10.1016/j.jhydrol.2022.128099.
Gauch, M., Mai, J., Lin, J., 2021. The proper care and feeding of camels: how limited training data affects streamflow prediction. Environ. Model. Softw. 135, 104926 https://coi.org/10.1016/j.envsoft.2020.104926.
Ghimire, G.R., Hansen, C., Gangrade, S., Kao, S.C., Thornton, P.E., Singh, D., 2023. Insights from dayflow: a historical streamflow reanalysis dataset for the conterminous united states. Water Resour. Res 59 (2) https://coi.org/10.1029/2022WR032312.
Gholizadeh, H., Zhang, Y., Frame, J., Gu, X., Green, C.T., 2023. Long short-term memory models to quantify long-term evolution of streamflow discharge and groundwater depth in alabama. Sci. Total Environ. 901, 165884 https://coi.org/10.1016/j.scitotenv.2023.165884.
Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P., Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., Mcclain, M.E., Meng, J., Mulligan, M., Nilsson, C., Olden, J.D., Opperman, J.J., Petry, P., Reidy Liermann, C., Sáenz, L., Salinas-Rodríguez, S., Schelle, P., Schmitt, R.J.P., Snider, J., Tan, F., Tockner, K., Valdujo, P.H., van Soesbergen, A., Zarfl, C., 2019. Mapping the world's free-flowing rivers. Nature 569 (7755), 215–221 https://coi.org/10.1038/s41586-019-1111-9.
Hall, J., Perdigão, R.A.P., 2021. Who is stirring the waters? Science 371 (6534), 1096–1097 https://coi.org/10.1126/science.abg6514.
Harrigan, S., Zsoter, E., Alfieri, L., Prudhomme, C., Salamon, P., Wetterhall, F., Barnard, C., Cloke, H., Pappenberger, F., 2020. Glofas-era5 operational global river discharge reanalysis 1979–present. Earth Syst. Sci. Data 12 (3), 2043–2060 https://coi.org/10.5194/essd-12-2043-2020.
Hashemi, R., Brigode, P., Garambois, P., Javelle, P., 2022. How can we benefit from regime information to make more effective use of long short-term memory (lstm) runoff models? Hydrol. Earth Syst. Sc. 26 (22), 5793–5816 https://coi.org/10.5194/hess-26-5793-2022.
He, J., Yang, K., Tang, W., Lu, H., Qin, J., Chen, Y., Li, X., 2020. The first high-resolution meteorological forcing dataset for land process studies over china. Sci. Data 7 (1) https://coi.org/10.1038/s41597-020-0369-y.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780 https://coi.org/10.1162/neco.1997.9.8.1735.
Hunt, K.M.R., Matthews, G.R., Pappenberger, F., Prudhomme, C., 2022. Using a long short-term memory (lstm) neural network to boost river streamflow forecasts over the western united states. Hydrol. Earth Syst. Sc. 26 (21), 5449–5472 https://coi.org/10.5194/hess-26-5449-2022.
Ibrahim Demir, Z.X.B.D., 2022. Waterbench: a large-scale benchmark dataset for data-driven streamflow forecasting. Earth Syst. Sci. data https://coi.org/https://doi.org/10.5194/essd-2022-52.
Klotz, D., Kratzert, F., Gauch, M., Keefe Sampson, A., Brandstetter, J., Klambauer, G., Hochreiter, S., Nearing, G., 2022. Uncertainty estimation with deep learning for rainfall–runoff modeling. Hydrol. Earth Syst. Sc. 26 (6), 1673–1693 https://coi.org/10.5194/hess-26-1673-2022.
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 Resour. Res 55 (12), 11344–11354 https://coi.org/10.1029/2019WR026065.
Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., Matias, Y., 2023. Caravan - a global community dataset for large-sample hydrology. Sci. Data 10 (1), 61 https://coi.org/10.1038/s41597-023-01975-w.
Li, H., Ke, C., Zhu, Q., Li, M., Shen, X., 2022. A deep learning approach to retrieve cold-season snow depth over arctic sea ice from amsr2 measurements. Remote Sens Environ. 269, 112840 https://coi.org/10.1016/j.rse.2021.112840.
Li, J., Miao, C., Zhang, G., Fang, Y.H., Shangguan, W., Niu, G.Y., 2022. Global evaluation of the noah-mp land surface model and suggestions for selecting parameterization schemes. J. Geophys. Res.: Atmospheres 127 (5) https://coi.org/10.1029/2021JD035753.
Li, Q., Zhu, Y., Shangguan, W., Wang, X., Li, L., Yu, F., 2022. An attention-aware lstm model for soil moisture and soil temperature prediction. Geoderma 409, 115651 https://coi.org/10.1016/j.geoderma.2021.115651.
Li, Y., Lv, C., Chang, J., Wang, Z., Kong, C., 2022. A bayesian data assimilation method to enhance the time sequence prediction ability of data-driven models. Aip Adv. 12 (10), 105021 https://coi.org/10.1063/5.0119688.
Lin, P., Pan, M., Beck, H.E., Yang, Y., Yamazaki, D., Frasson, R., David, C.H., Durand, M., Pavelsky, T.M., Allen, G.H., Gleason, C.J., Wood, E.F., 2019. Global reconstruction of naturalized river flows at 2.94 million reaches. Water Resour. Res 55 (8), 6499–6516 https://coi.org/10.1029/2019WR025287.
Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., Thieme, M., 2019. Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. Sci. Data 6 (1) https://coi.org/10.1038/s41597-019-0300-6.
Liu, Y., Xie, X., Tursun, A., Wang, Y., Jiang, F., Zheng, B., 2023. Surface water expansion due to increasing water demand on the loess plateau. J. Hydrol.: Reg. Stud. 49, 101485 https://coi.org/10.1016/j.ejrh.2023.101485.
Marvel, K., Cook, B.I., Bonfils, C.J.W., Durack, P.J., Smerdon, J.E., Williams, A.P., 2019. Twentieth-century hydroclimate changes consistent with human influence. Nature 569 (7754), 59–65 https://coi.org/10.1038/s41586-019-1149-8.
Moosavi, V., Gheisoori Fard, Z., Vafakhah, M., 2022. Which one is more important in daily runoff forecasting using data driven models: input data, model type, preprocessing or data length? J. Hydrol. 606, 127429 https://coi.org/10.1016/j.jhydrol.2022.127429.
Ouyang, W., Lawson, K., Feng, D., Ye, L., Zhang, C., Shen, C., 2021. Continental-scale streamflow modeling of basins with reservoirs: towards a coherent deep-learning-based strategy. J. Hydrol. 599, 126455 https://coi.org/10.1016/j.jhydrol.2021.126455.
Sadler, J.M., Appling, A.P., Read, J.S., Oliver, S.K., Jia, X., Zwart, J.A., Kumar, V., 2022. Multi-task deep learning of daily streamflow and water temperature. Water Resour. Res 58 (4) https://coi.org/10.1029/2021WR030138.

Seibert, J., 2021. A retrospective on hydrological modelling based on half a century with the hbv model. Hydrol Earth Syst Sc. https://coi.org/https://doi.org/10.5194/hess-2021-542..

Shen, C., 2018. Deep learning: a next-generation big-data approach for hydrology. Eos 99 https://coi.org/10.1029/2018EO095649.

Swain, S., Mishra, S.K., Pandey, A., Srivastava, P.K., Nandi, S., 2023. Characterization and assessment of hydrological droughts using glofas streamflow data for the narmada river basin, india. Environ. Sci. Pollut. R. https://coi.org/10.1007/s11356-023-27036-8.

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 Resour. Res 59 (7) https://coi.org/10.1029/2022WR034352.

Thomas Lees, S.R.F.K., 2021. Hydrological concept formation inside long short-term memory (lstm) networks. Hydrol. Earth Syst. Sci. https://coi.org/https://doi.org/10.5194/hess-2021-566.

Wang, Y., Xie, X., Shi, J., Zhu, B., Jiang, F., Chen, Y., Liu, Y., 2022. Accelerated hydrological cycle on the tibetan plateau evidenced by ensemble modeling of long-term water budgets. J. Hydrol. 615, 128710 https://coi.org/10.1016/j.jhydrol.2022.128710.

Wei, X., Wang, G., Schmalz, B., Hagan, D.F.T., Duan, Z., 2023. Evaluation of transformer model and self-attention mechanism in the yangtze river basin runoff prediction. J. Hydrol.: Reg. Stud. 47, 101438 https://coi.org/10.1016/j.ejrh.2023.101438.

Wunsch, A., Liesch, T., Cinkus, G., Ravbar, N., Chen, Z., Mazzilli, N., Jourde, H., Goldscheider, N., 2022. Karst spring discharge modeling based on deep learning using spatially distributed input data. Hydrol. Earth Syst. Sci. 26 (9), 2405–2430 https://coi.org/10.5194/hess-26-2405-2022.

Yin, H., Guo, Z., Zhang, X., Chen, J., Zhang, Y., 2022. Rr-former: rainfall-runoff modeling based on transformer. J. Hydrol. 609, 127781 https://coi.org/10.1016/j.jhydrol.2022.127781.

Yokoo, K., Ishida, K., Ercan, A., Tu, T., Nagasato, T., Kiyama, M., Amagasaki, M., 2022. Capabilities of deep learning models on learning physical relationships: case of rainfall-runoff modeling with lstm. Sci. Total Environ. 802, 149876 https://coi.org/10.1016/j.scitotenv.2021.149876.

Zhang, K., Xie, X., Zhu, B., Meng, S., Yao, Y., 2019. Unexpected groundwater recovery with decreasing agricultural irrigation in the yellow river basin. Agric. Water Manag. 213, 858–867 https://coi.org/10.1016/j.agwat.2018.12.009.

Zhao, T., Chen, Z., Tu, T., Yan, D., Chen, X., 2022. Unravelling the potential of global streamflow reanalysis in characterizing local flow regime. Sci. Total Environ. 838, 156125 https://coi.org/https://doi.org/10.1016/j.scitotenv.2022.156125.

Zhu, B., Huang, M., Cheng, Y., Xie, X., Liu, Y., Zhang, X., Bisht, G., Chen, X., Missik, J., Liu, H., 2020. Effects of irrigation on water, carbon, and nitrogen budgets in a semiarid watershed in the pacific northwest: a modeling study. J. Adv. Model Earth Syst. 12 (9) https://coi.org/10.1029/2019MS001953.