2022-03

High-quality reconstruction of China's natural streamflow

Miao, C

10.1016/j.scib.2021.09.022
Science Bulletin
Elsevier BV

High-quality reconstruction of China's natural streamflow

Chiyuan Miao*[1], Jiaojiao Gou[1], Bojie Fu[1,2], Qiuhong Tang[3], Qingyun Duan[1], Zhongsheng Chen[4], Huimin Lei[5], Jie Chen[6], Jiali Guo[7,8], Alistair G.L. Borthwick[9], Wenfeng Ding [10], Xingwu Duan [11], Yungang Li [11], Dongxian Kong[1], Xiaoying Guo[1], Jingwen Wu[1]

[1] State Key Laboratory of Earth Surface Processes and Resource Ecology, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China

[2] State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

[3] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

[4] College of Land and Resources, China West Normal University, Nanchong Sichuan 637009, China

[5] State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, 100084, China

[6] State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

[7] College of Hydraulic and Environmental Engineering, China Three Gorges University, Yichang 443002, China

[8] Engineering Research Center of Eco-environment in Three Gorges Reservoir Region, Ministry of Education, China Three Gorges University, Yichang, 443002, China

[9] School of Engineering, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3JL, UK

[10] Changjiang River Scientific Research Institute, Wuhan 430010, China

[11] Institute of International Rivers and Eco-security, Yunnan University, Kunming 650091, China

ABSTRACT

Reconstruction of natural streamflow is fundamental to the sustainable management of water resources. In China, previous reconstructions from sparse and poor-quality gauge measurements have led to large biases in simulation of the interannual and seasonal variability of natural flows. Here we use a well-trained and tested land surface model coupled to a routing model with flow direction correction to reconstruct the first high-quality gauge-based natural streamflow dataset for China, covering all its 330 catchments during the period from 1961 to 2018. A stronger positive linear relationship holds between upstream routing cells and drainage areas, after flow direction correction to 330 catchments. We also introduce a parameter-uncertainty analysis framework including sensitivity analysis, optimization, and regionalization, which further minimizes biases between modeled and inferred natural streamflow from natural or near-natural gauges. The resulting behavior of the natural hydrological system is represented properly by the model which achieves high skill metric values of the monthly streamflow, with about 83% of the 330 catchments having Nash–Sutcliffe efficiency coefficient (NSE) > 0.7, and about 56% of the 330 catchments having Kling-Gupta Efficiency coefficient (KGE) > 0.7. The proposed construction scheme has important implications for similar simulation studies in other regions, and the developed low bias long-term national datasets by statistical postprocessing should be useful in supporting river management activities in China.

Keywords: Natural streamflow; Reconstruction; Land surface model; Parameter uncertainty analysis; China

1. Introduction

Freshwater provides a fundamental basis for life on Earth [1], and the provision of sustainable water plays a pivotal role in ecosystems [2], economy [3], and politics [4]. A 2030 vision of sustainable freshwater withdrawal and supply is provided by Target 6.4 of the UN Sustainable Development Goals which aims to "substantially reduce the number of people suffering from water scarcity" [5]. Many studies worldwide have reported water scarcity driven by steadily increasing water demand, such as in southern Europe [1], northern China [6], the border between India and Pakistan [7], and the western United States [8]. Balancing water supply and demand for sustainable water management across different regions is highly dependent on accurate, reliable surface flow records. Collection of long-term observational streamflow datasets worldwide is necessary to provide an evidence-based foundation for hydrological studies [9]. Since the 1980s, numerous runoff compilation products, such as the Global Runoff Data Base (GRDB) [10], the Global Streamflow Indices and Metadata Archive (GSIM) [11], and the Dai and Trenberth Global River Flow and Continental Discharge Dataset [12] have been developed in order to monitor water deficits around the world. However, these global datasets pose challenges due to their unequal distribution of available resources to users; the datasets include information from sparsely distributed gauge stations and contain notable data gaps in certain regions, especially in China (Fig. S1, GRDB and GSIM provide two examples).

China has serious water scarcity problems given that it has only one-quarter of the world average of per capita freshwater resources [13, 14]. The uneven distribution of water resources driven by a monsoon climate and complicated topography has exacerbated water scarcity across China. In 2019, average surface water resources was reported to be about 78.32×10^4 m^3/km^2 in southern China, but only 9.43×10^4 m^3/km^2 in northern China [15]. To address potential solutions for China's water crisis, much government planning has been devoted to the determination of the extent and severity of China's water problems [16, 17]. Since the 1980s, the Nationwide Water Resources Investigation and Assessment Project has made comprehensive assessments of the volume, quality, spatiotemporal distribution, and development potential of water resources in China [18]. In such water assessment activities, natural flow records are required that contain no discernible abrupt changes or shifts compared with previous decades.

In fact, observed flow regimes are greatly affected by human activities, including water diversion, water withdrawal, and reservoir operation [19, 20]. By 2019, more than 7,000 global large dams (storage capacity > 0.1 km^3) had been constructed worldwide [21], with the overall area of irrigated agriculture constituting 40% of the total area used for agricultural production [22]. During the past three decades, China has experienced unprecedented economic growth involving accelerated industrialization and urbanization [23]. To accommodate food and water demands in China, people engineered rivers [24], drilled groundwater wells [25], and developed vast tracts of irrigated land [6]. For example, China has developed 58 million hectares of

88 irrigated land since the 1950s, which generate 70% of the country's total grain production [17].
89 Owing to these strong anthropogenic influences, traditional gauge measurements are unable to
90 represent long-term natural hydrological variations. Thus natural streamflow estimation is
91 needed for comparison with previous streamflow records [26] and to help interpret climate
92 information (e.g. precipitation and temperature) in hydrometeorology studies [27].

93 Reconstruction of natural flows is very complex because it involves several non-linear
94 hydrological processes and a variety of spatiotemporal scales, ranging from watershed to global
95 and from months to millennia. This complexity is the reason why understanding and
96 representing natural hydrological regimes is such a challenge. Traditional reconstruction
97 methods, such as the water balance method and regression method, encounter problems when
98 describing the natural flow in a basin with strong anthropogenic disturbances, scarce
99 observations, and dynamic surrounding environment [26, 28]. Hydrological models that
100 incorporate both the physical and climatic characteristics of basins appear effective tools to
101 reconstruct the flow of a regulated river system [29, 30]. Although several global or local natural
102 flows datasets have been developed, including the Global Runoff Reconstruction dataset [31],
103 the Global Reach-Level A Priori Discharge Estimates [30], and the Long-Term Land Surface
104 Hydrologic Fluxes and States Dataset [32], each involves reconstruction using information
105 from sparse training stations when applied to China. Noting the urgent need for long-term
106 reliable natural flow data, the present study enables hydrological models to be tuned accurately
107 to high quality data from natural or near-natural training stations. Flow direction correction and

parameter uncertainty analysis act as quality controls within the model, reducing potential errors between reconstructed and inferred natural streamflow. Our objective is to develop an effective high-quality scheme to reconstruct China's natural streamflow, leading to long-term continuous monthly streamflow datasets from 1961 to 2018 for 330 catchments. These long-term national datasets are useful in the assessment and allocation of natural water resources, and in studying the impact of climate change on the terrestrial water cycle in China. Such developed streamflow datasets without anthropogenic influences are vitally needed in the worldwide effort to tackle the water crisis and strengthen resilience to climate change.

2. Materials and methods

2.1. Model tools and Datasets

The Variable Infiltration Capacity (VIC) macroscale land-surface model [33] and coupled Lohmann routing model [34] were adopted to reconstruct natural streamflow records of 330 hydrological stations for the period 1961–2018. We ran the VIC model version 4.2 in water balance mode for daily time step increments. The VIC model involved three soil layers; the upper two layers generated surface flow according to the variable soil moisture capacity curve, and the lowest layer generated slow response runoff (baseflow) predicted by the nonlinear ARNO model [14]. The Lohmann routing model (hereinafter routing model) described the water's horizontal transport processes and transformed field runoff to streamflow at the outlet grid cell. More information about the two models is given in [14].

The VIC model was established on a 0.25°×0.25° grid, with meteorological forcing including daily precipitation, maximum and minimum temperatures, and wind speed variables taken from ~2,400 weather stations [35, 36]. Land surface characteristics were provided by 5-arcmin soil texture datasets from the Food and Agriculture Organization of the United Nations, 1 km land cover data from the University of Maryland (http://glcfapp.glcf.umd.edu:8080/esdi/index.jsp), and information on leaf area index (LAI) from the 0.25° monthly database compiled by Myneni and colleagues [37]. Routing-related information, including eight-direction (D8) flow model and flow fraction, for each land cell was extracted from 1 km elevation data (http://westdc.westgis.ac.cn). High-resolution river flowlines determined by Lin et al. [30] were adopted to correct the D8 values, with flowlines extracted from 3-arcsec (~90 m) resolution MERIT DEM under a 25 km^2 channelization threshold.

To control the quality of the reconstructed streamflow data, a large dataset of monthly inferred natural/near-natural streamflow records from 1961 to 1979 for 330 gauge stations was used to train and test the models (Fig. 1). Besides the selected near-natural research period (noting the boom in socio-economic activity in China after 1980), we also adopted inferred gauge data from naturalized flows in the absence of water management effects. Nearly half of all gauge records are naturalized by the Ministry of Water Resources of China based on the water balance principle (Fig. S2 and Supplementary Text1 online). The remaining stations are all near natural with no upstream dams present or a low level of dam influence (Fig. S2). In

148 summary, data from these 330 gauge stations are of sufficient quality to reconstruct natural
149 streamflow because of their high gauge density, low regulation influence (naturalized records),
150 long record length (at least 10 years' cover), and suitable catchment area (larger than 1000 km^2)
151 (Fig. S2). More information about the models and datasets is given in [14, 38].
152
153 **< Fig. 1 >**
154
155 *2.2 Model training and test*
156 Both the land-surface model and routing model were trained. First, we corrected routing
157 networks in the coupled model because streamflow is a spatial integrator of field runoff [39],
158 and gauge locations, drainage area, and upstream flow direction must be correctly ascertained
159 before the land surface model is calibrated. China's hydrological gauge network suffers from a
160 number of negative influence factors, e.g., gauge relocation, station renaming, and missing data,
161 especially before China's 'Reform and Opening-up' period [40]. For example, 'Xiaodong' in
162 the Yellow River basin was renamed 'Wuzhi' after June 1968. We compiled basic information,
163 including location and drainage area, on 330 gauge stations derived from '*Hydrological*
164 *Yearbook of China*', local water resources information releases, and previous hydrological
165 research works. Then we manually adjusted the D8 flow direction based on the high-resolution
166 river flowlines according to the correct gauge location and catchment area. For the routing
167 model test, we checked that the number of upstream routing cells (0.25°×0.25°) matched the

observed drainage area for each gauge station.

Parameter tuning enables model predictions to fit corresponding observations [41]. In the present study, we considered 13 tunable streamflow-related parameters (Table S1). A parameter-uncertainty analysis framework, which includes sensitivity analysis, optimization, and regionalization, was used to train the VIC land-surface model [38]. Parameter sensitivity analysis was undertaken a priori to reduce parameter dimensionality, thus lightening the computational burden for parameter optimization. Parameter regionalization transferred optimized parameters from gauged to ungauged catchments. We used four global sensitivity analysis methods (sum-of-trees, multivariate adaptive regression splines, delta test, and metamodel-based Sobol′) to screen important parameters for the 10 water resources regions of China [14]. Then an adaptive surrogate modeling-based optimization (ASMO) was used to find the optimal parameter solution. The multiscale parameter regionalization (MPR) technique estimated values for the optimal parameter in ungauged areas [38] (Fig. S3). Supplementary Text3 (online) provides further details of the parameter uncertainty analysis methodology.

As a benchmark reference to natural records, streamflow data from a subset of 230 stations were used to train the coupled land-surface and routing model, thus reconstructing China's natural streamflow (Fig. S4). We tuned the sensitive parameters of modeled streamflow for each training station in succession from upstream to downstream during the calibration period (1961–1969) and estimated the corresponding parameters for ungauged areas. Next, the combined model was run for the validation period (1970–1979). The Nash–Sutcliffe efficiency

188 coefficient (NSE) was set as the objective function during the model training stage. To check
189 whether the model met the expectations of the training stage, inferred natural data from the
190 remaining 100 stations were used to test model performance without any manipulation of the
191 tuned parameter or settled river network (Fig. S4). Four skill metrics were chosen to evaluate
192 model performance: Pearson's Correlation Coefficient (CC), NSE, Percent bias (Pbias, unit: %),
193 and Kling-Gupta Efficiency coefficient (KGE). For NSE and KGE metrics, a higher score
194 indicates better reconstruction, with a perfect score being unity. Supplementary Text4 (online)
195 lists the equations used to compute four skill metrics.
196
197 *2.3 Model performance exploration*
198 Three catchment descriptors (drought index, area, and runoff efficiency index) were used
199 to explore the influence of watershed characteristics on model performance, and identify
200 possible obstacles to natural streamflow estimation. The drought index, defined as the ratio of
201 potential evapotranspiration to precipitation, reflects the dry-wet hydroclimatic regimes in a
202 given basin. Runoff efficiency, defined as the fraction of precipitation that becomes runoff,
203 represents land surface partitioning of precipitation, runoff, and evapotranspiration. Multi-year
204 averages of the drought index and runoff efficiency were calculated for 330 gauge stations
205 during the period from 1961 to 1979. Potential evapotranspiration was calculated by the method
206 of Hargreaves and Samani [42]. We examined the spatial pattern of the model performance
207 metrics with respect to the selected catchment descriptors, and considered the relationship of

208 the catchment descriptors to uncertainty in the hydrological modeling process.

209 *2.4 Model statistical post-processing*

210 To further improve the credibility of the reconstructed natural streamflow dataset, we
211 adopted the Scaled Distribution Mapping (SDM) statistical post-processing method [43] to
212 reduce systematic bias in estimated natural streamflow from a calibrated hydrological model
213 (e.g., bias from meteorological forcing and model structure) [30, 44, 45]. The SDM method
214 scales the referred natural streamflow distribution by raw model simulated natural streamflow
215 changes in magnitude and the likelihood of events under nonstationary conditions [43]. We
216 fitted gamma distribution parameters using maximum likelihood to the monthly referred and
217 simulated natural streamflow (lower limit set to 0.1 m^3/s) for the baseline period (1961–1979),
218 and to the monthly simulated natural streamflow for the whole simulation period (1961–2018).
219 Then the bias-corrected streamflow was calculated using the scaled fitted gamma distribution
220 for all cumulative density function values corresponding to the streamflow time series for each
221 gauge station.

222

223 **3. RESULTS AND DISCUSSION**

224 *3.1. Reconstruction performance after flow direction corrections*

225 Flow direction is inherently important in hydrological modeling because a basin's outflow
226 results from the spatial integration of upstream drainage. Fig.2a shows the extracted 0.25°×0.25°
227 D8 flow direction based on the digital elevation model of China. The extracted flow directions

agree with general flow patterns in China: most runoff discharges eastwards into the Pacific Ocean after passing over all or part of the three-step staircase topography of western China, whereas runoff in the Northwest River drainage system vanishes within dry land (Fig. 2a-b). However, the automatically extracted flow direction becomes unreliable when describing rivers passing through flat lands, narrow gorge ('throat') areas, or sharply curved river bends [46]. Flow direction correction is therefore a prerequisite to reconstructing natural streamflow.

< Fig. 2>

We manually corrected the D8 flow direction based on the 330 gauge locations, corresponding gauge areas, and high-resolution river flowlines [30]. Fig.3a shows the spatial distribution of reconstructed natural river discharge based on corrected flow direction. The routing model properly represents the shape of China's large rivers, including the Yangtze River, the Yellow River, and the Songhua River. Moreover, the upstream routing cells maintain a positive linear relationship with the observed drainage areas after flow direction correction was applied to data from the 330 gauge stations (Fig. 3b), indicating the corrected flow networks are satisfactory within the routing model.

< Fig. 3 >

Three regions of interest, the Tarim basin in the Northwest River drainage system (Region A), the Three Parallel Rivers in the Southwest River drainage system (Region B), and the Huai River basin (Region C) (Fig. 2A-C, Fig. 3c-e), were chosen to test model reconstruction performance before and after flow direction correction. Region A is located in an arid endorheic basin with average annual precipitation below 50 mm [47]; Region B is located in a typical valley basin where three rivers run parallel through deep gorges; and Region C is located in a flat basin crisscrossed by a river of average elevation 87 m. For Region A, routing cells were removed where there was no water or the flow was insufficient to reach the next cell (Fig. 2A-A'). The parallel river flow direction was separated out for Region B and the intricate tributary network combined for Region C (Fig. 2B-B', Fig. 2C-C'). Selecting a single demonstration catchment for each region of interest, we found that the modeled streamflow provided a satisfactory match to the inferred natural streamflow after modification for insufficient or excess upstream water flow (Fig. 3c-e). Although the reconstruction performance performed well after flow direction correction, the modeled streamflow still exhibited large biases with the inferred natural streamflow (Fig. 3c-e), and so additional land-surface runoff model parameter optimization is needed before model application.

3.2. Reconstruction performance after parameter uncertainty analysis

Effective model training enables the land surface model to match closely the behavior of the natural hydrological system, a prerequisite for using the model to reconstruct the natural

268 streamflow. A parameter uncertainty analysis framework was developed that included
269 sensitivity analysis, optimization, and regionalization. Based on previous sensitivity analysis
270 [14], a list of VIC tunable parameters selected from 13 tunable parameters for each water
271 resources region was identified as sensitive in terms of streamflow simulation (Table S1-S2,
272 online). We explored the transferability of the hydrological model to ungauged catchments, and
273 estimated parameters based on physiographical predictors for the basins. The parameter
274 regionalization method provided estimates of spatial continuous parameter sets (Fig. S5, online)
275 [38].

276 Fig. 4 shows model performance expressed using CC, NSE, Pbias (%), and KGE skill
277 metrics. Overall, the model succeeds well in representing the natural streamflow at both training
278 and test stations (Fig. 4 and Table. S3), confirming that it provides a reliable reconstruction of
279 the natural streamflow in China. CC ranges from 0.69 to 0.99 with a mean of 0.93 at training
280 stations and 0.49 to 0.99 with a mean of 0.89 at test stations (Table S3, online). These high CC
281 values indicate that the model properly captures trends in the natural streamflow dynamics.
282 Pbias measures the absolute bias of modelled and referred natural streamflow. About 65% of
283 stations (216 of 330 stations) exhibit Pbias within ±20%. Given the high values obtained for
284 both CC and Pbias, most stations exhibit a certain degree of systematic bias in the simulated
285 streamflow across China. This systematic bias is likely to have resulted from
286 over/underestimates of precipitation forcing, inaccurate referred natural streamflow records,
287 and inherent model structure. About 83% and 56 % of the 330 hydrological stations have

288 respective NSE and KGE skill scores larger than 0.7 (Fig. 4); this result is generally better than
289 obtained by other local [32, 48, 49] or global [30, 31, 50] streamflow estimation studies. The
290 Yangtze River basin exhibits best model performance, with highest skill metric values at
291 training and test stations (Table S3). The training/test station with lowest CC and NSE values
292 (0.35/0.22 and 0.69/0.49, Table S3) was located in the Northwest River drainage system (Table
293 S3). Uncertainty arising from inferred streamflow, model structure, and metrological forcing
294 may cause this spatial heterogeneity in model performance, given that the hydrological
295 processes in the Northwest River drainage system are more difficult to capture than in the
296 Yangtze River basin.

297 **< Fig. 4 >**

298

299 *3.3. Possible obstacles to natural streamflow reconstruction*

300 Despite the overall good performance attained after step-by-step model training and testing,
301 it still proved necessary to identify 'bad' gauges and explore the reasons for such reconstruction
302 anomalies. Fig. 5 shows the spatial distribution of model performance indicated by four skill
303 metrics at the 330 gauge stations. For CC, all gauges exhibit strong relations between modeled
304 natural streamflow and inferred natural streamflow, especially for the Yangtze River basin,
305 Pearl River basin, and Southwest River drainage system. Compared to CC, spatial discrepancies
306 in Pbias, NSE, and KGE skill scores are more pronounced across the 330 stations (Fig. 5). Low
307 bias (±20%) is exhibited between modeled and inferred natural streamflow at most gauges. We

found that natural streamflow is generally underestimated in the Songhua River basin and Liao River basin. NSE values across 330 stations are generally high in the majority of southern basins and low in northern basins, particularly in the northwest (Fig. 5), highlighting a possible relationship between model performance and the dynamic environment of individual basins.

< Fig. 5>

Three catchment descriptors (drought index, area, and runoff efficiency index) were used to investigate whether the NSE pattern in streamflow simulation can be explained by watershed characteristics (Fig. 6). We found a negative correlation between model performance and drought index ($r = -0.40$; $P < 0.001$), and positive correlations between NSE and basin area ($r = 0.37$; $P < 0.001$) and runoff efficiency ($r = 0.22$; $P < 0.001$). The same significant spatial pattern was found for the CC metric to that of the NSE metric for drought index, catchment area and runoff efficiency, whereas KGE and Pbias are partially correlated with the catchment descriptors (Fig. S6). In other words, the model performed better for a basin with wetter hydroclimatic conditions, larger drainage area, and higher runoff efficiency (Fig. 5 and Fig. 6).

< Fig. 6 >

Beyond inherent uncertainty in the inferred natural streamflow data, meteorological forcing

328 uncertainty and model structure uncertainty could also cause poor streamflow reconstruction in
329 arid or semiarid basins. Meteorological forcing, especially precipitation, is an important
330 atmospheric upper boundary condition for streamflow generation, and the accuracy of such data
331 directly affects parameter values and model performance [51]. Uncertainty arising from a
332 combination of inadequate meteorological stations [14] and inaccurate measurements [52] in
333 harsh environments may cause deterioration in performance of streamflow simulation for dry
334 hydro-regions. For example, the precipitation pattern exhibits considerable discrepancies within
335 the different precipitation products in Northwestern China (Fig. S7-S8). Thus the
336 meteorological forcing still has certain uncertainties in this study, at least for the Northwestern
337 regions. In terms of model structure uncertainty, the VIC land surface model is better able to
338 simulate streamflow in humid regions than in arid regions where the infiltration excess runoff
339 generation mechanism is important [53-55].

340 Higher skill scores are obtained for basins of larger drainage area, partly because the
341 climatic and physiographic data are usually less reliable for smaller basins [56, 57]. Previous
342 studies simulated runoff in 269 catchments ranging in area from 10 to 130,000 km^2, and found
343 that NSE increased with catchment area, demonstrating that scale effects influence hydrological
344 model performance [58]. One would expect more reliable runoff simulation for large basins
345 because improved rainfall estimates can be made for such basins where there is usually a greater
346 density of rain gauges [58, 59]. Like the drought index, runoff efficiency is a predictor of
347 watershed partitioning, which expresses the relative fraction of precipitation that exits in a

watershed as runoff [60]. A possible reason for poor model performance when applied to a basin with low runoff efficiency is that snowmelt could account for a certain fraction of runoff in such a basin, e.g., the Songhua River basin, the Liao River basin, and the Northwest River drainage system. The snowmelt-related physical parameterization scheme is not fully represented in the VIC model [61], and other key hydrological processes (e.g., glacier dynamics) are also missing.

In summary, we found that it is difficult to reconstruct the natural streamflow in a basin with dry hydrological climate, small drainage area, or low runoff efficiency, such as in the northwest regions of China. This problem arises from a dearth of gauged streamflow data in the northwest regions where the environment is harsh and rivers are smaller, particularly in the Tibetan Plateau. Future hydrological modeling work should focus on decreasing this source of uncertainty (along with that from input data, observation, and model structure) in order to improve the accurate simulation of runoff in arid and semi-arid regions across China. It is also worth exploring the use of multi-model ensembles to reduce uncertainty in streamflow simulation.

3.4. Reconstruction performance after statistical post-processing

Fig. S9 (online) illustrates the probability density function distribution of the Pbias metric before and after bias correction (BC) via the statistical post-processing procedure for monthly streamflow at 330 hydrological stations across China. Fig. S9 clearly shows that Pbias values cluster much closer to zero after BC, and most stations are in the ±10% acceptable range. For regions with underestimated streamflow such as the Songhua River basin and Liao River basin,

368 the estimated natural streamflow time series have been corrected to match the referred natural
369 streamflow (Fig. 5, and Fig. S10 online). Among all stations, the mean value of the Pbias metric
370 decreased from 17.13% to 2.27% after BC, and the other three skill metrics, especially NSE
371 and KGE, are largely improved after BC (Table S4, online).

372 In further exploration of the performance of the reconstructed streamflow, we chose three
373 typical stations to compare the difference between the reconstructed streamflow and the gauged
374 streamflow. Huangjiagang, Luanxian, and Yingluoxia stations are associated respectively with
375 reservoir-impacted, irrigation-impacted, and natural flow regimes. Fig. S11 shows that the
376 reconstructed natural streamflow (Nat) provides a good match with the referred natural
377 streamflow (Nat') at all three stations but exhibits large discrepancy with gauged streamflow
378 (Obs) at Huangjiagang and Luanxian stations during the period 1980–2000. However, the flow
379 conditions at Huangjiagang station were modified by the upstream Danjiangkou Reservoir
380 (completed in 1968) through attenuation of floods peaks and increased in low water levels [62],
381 thus progressively narrowing the gap in streamflow seasonality between flood and dry seasons
382 (Fig. S11a). At Luanxian station in the Hai River basin, peak flow attenuation and bimodal
383 structure were triggered because of heavy irrigation (of an area that constitutes 24% of the total
384 basin) and extremely low available water resources (3.30×10^4 m^3/km^2 in 2019 [15], Fig. 1 and
385 Fig. S11b). At Yingluoxia station, the effect of human activities is less obvious, with the natural
386 streamflow providing a close match to the observed streamflow during the period 1980–2000
387 (Fig. S11c). These results demonstrate that the influences of human activities have been

effectively removed from the reconstructed natural streamflow series, which therefore have the potential to be applied in large-scale hydrological studies for water resources management and climate change assessment of terrestrial water availability.

CONCLUSION

Long-term continuous natural streamflow at near-continental scale can be reconstructed from a coupled physically-based land surface model and routing model. High-quality monthly natural streamflow datasets from 1961 to 2018 are built for 330 catchments across China. The total drainage area of the 330 catchments is 45.2×10^5 km^2, or about 54.7% of the non-ice, non-desert land area of China. The resulting developed natural gauge datasets are reliable because of the multi-level quality controls used herein, including flow direction correction within the routing model, parameter uncertainty analysis in the land surface model, and statistical post-processing, which minimize biases between reconstructed and inferred natural streamflow.

Flow direction corrections ensured that field runoff followed the correct accumulation route in each catchment, especially those with catchments with flat land, narrow gorges, and/or sharply curved river bends. After application of flow direction correction to 330 catchments, a stronger positive linear relationship was achieved between the streamflow in upstream routing cells and that in the drainage areas. Parameter uncertainty analysis (comprising parameter sensitivity analysis, parameter optimization, and parameter regionalization) was then used to

train the land surface model so that its results closely matched the behavior of the natural hydrological system. Supported by a well-trained model system, about 83% of the 330 catchments exhibited NSE > 0.7, and about 56% of the 330 catchments exhibited KGE > 0.7. The systematic bias of estimated natural streamflow from a calibrated hydrological model was reduced by the statistical post-processing technique with Pbias metric decreased from 17.13% to 2.27%. The reconstructed natural streamflow dataset provides a reliable representation of natural hydrological process in regions affected by intensive human activity.

Our high-quality, long-term, natural streamflow reconstruction modelling framework has the potential to support water resources management and allocation at near-continental scale. The model may also be used to identify possible physical mechanisms and processes of natural hydrological regime analysis, thus providing additional decision-making support for water managers. Although the model performed well overall in terms of reconstructed streamflow, specific natural streamflow reconstruction may be required for basins with dry hydrological climate, small drainage area, or low runoff efficiency. Future hydrological modeling work should focus on decreasing uncertainty arising from incomplete input data, observations, and model structure, perhaps through conducting multi-model ensembles and adopting higher resolution climate forcing. The gauge-based natural streamflow dataset developed in this study is available by means of a reasonable academic research request to the corresponding author (and is partially publicly available at xxx). Due to underlying confidentiality agreements, the data cannot be utilized for commercialization of research findings.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This research was supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (No.2019QZKK0405) and the National Natural Science Foundation of China (No. 41877155).

Author contributions

Chiyuan Miao designed research and wrote the paper with inputs from all coauthors. All authors collected and analyzed data.

References

[1] Gudmundsson L, Seneviratne SI, Zhang XB. Anthropogenic climate change detected in European renewable freshwater resources. Nat Clim Chang, 2017; 7: 813-8.

[2] Ellison D, Morris CE, Locatelli B, et al. Trees, forests and water: Cool insights for a hot world. Glob Environ Change, 2017; 43: 51-61.

[3] Flörke M, Schneider C, McDonald RI. Water competition between cities and agriculture driven by climate change and urban growth. Nat Sustain, 2018; 1: 51-8.

448 [4] D'Odorico P, Davis KF, Rosa L, et al. The global food-energy-water nexus. Rev Geophys,
449 2018; 56: 456-531.
450 [5] United Nations. Transforming our world: The 2030 agenda for sustainable development.
451 New York: United Nations, 2015.
452 [6] Piao SL, Ciais P, Huang Y, et al. The impacts of climate change on water resources and
453 agriculture in China. Nature, 2010; 467: 43-51.
454 [7] Oki T, Kanae S. Global hydrological cycles and world water resources. Science, 2006; 313:
455 1068-72.
456 [8] Barnett TP, Pierce DW, Hidalgo HG, et al. Human-induced changes in the hydrology of
457 the western United States. Science (New York, NY), 2008; 319: 1080-3.
458 [9] Tetzlaff D, Carey SK, McNamara JP, et al. The essential value of long-term experimental
459 data for hydrology and water management. Water Resour Res, 2017; 53: 2598-604.
460 [10] The Global Runoff Data Centre. The world-wide repository of river discharge data and
461 associated metadata. Koblenz, Germany, 1988, http://grdc.bafg.de.
462 [11] Do HX, Gudmundsson L, Leonard M, et al. The Global Streamflow Indices and Metadata
463 Archive (GSIM) – part 1: The production of a daily streamflow archive and metadata.
464 Earth Syst Sci Data, 2018; 10: 765-85.
465 [12] Dai AG. Dai and trenberth global river flow and continental discharge dataset. 2017,
466 https://doi.org/10.5065/D6V69H1T.
467 [13] Ge LQ, Xie GD, Zhang CX, et al. An evaluation of China's water footprint. Water Resour

468 Manag, 2011; 25: 2633-47.
469 [14] Gou JJ, Miao CY, Duan QY, et al. Sensitivity Analysis-Based Automatic Parameter
470 Calibration of the VIC Model for Streamflow Simulations Over China. Water Resour Res,
471 2020; 56: 1-19.
472 [15] Ministry of Water Resources of China. China's Water Resources Bulletin 2019. Beijing:
473 Ministry of Water Resources of China, 2020.
474 [16] World Bank Group. Watershed: A New Era of Water Governance in China: Policy brief
475 (English). 2018.
476 [17] Yu CQ. China's water crisis needs more than words. Nature, 2011; 470: 307-307.
477 [18] Wang GS, Dai N, Yang JQ, et al. Water resources assessment and prediction in China.
478 Proceedings of the International Association of Hydrological Sciences, 2016; 374: 79-84.
479 [19] Haddeland I, Heinke J, Biemans H, et al. Global water resources affected by human
480 interventions and climate change. Proc Natl Acad Sci USA, 2014; 111: 3251-6.
481 [20] Tang QH. Global change hydrology: Terrestrial water cycle and global change. Sci China
482 Earth Sci, 2019; 63: 459-62.
483 [21] Lehner B, Liermann CR, Revenga C, et al. Global reservoir and dam database (GRanD)
484 v1.3. 2019, http://globaldamwatch.org/data/.
485 [22] Meier J, Zabel F, Mauser W. A global approach to estimate irrigated areas – a comparison
486 between different data and statistics. Hydrol Earth Syst Sc, 2018; 22: 1119-33.
487 [23] Fang JY, Yu GR, Liu LL, et al. Climate change, human impacts, and carbon sequestration

488 in China. Proc Natl Acad Sci USA, 2018; 115: 4015-20.
489 [24] Long D, Yang W, Scanlon BR, et al. South-to-North water diversion stabilizing Beijing's
490 groundwater levels. Nat Commun, 2020; 11: 3665-74.
491 [25] de Graaf IEM, Gleeson T, Rens van Beek LPH, et al. Environmental flow limits to global
492 groundwater pumping. Nature, 2019; 574: 90-4.
493 [26] Wang H, Wang JH, Qin DY, et al. Theory and methodology of water resources assessment
494 based on dualistic water cycle model. J Hydraul Eng, 2006; 12: 1496-502.
495 [27] Lehner F, Wood AW, Vano JA, et al. The potential to reduce uncertainty in regional runoff
496 projections from climate models. Nat Clim Chang, 2019; 9: 926-33.
497 [28] Hernández-Henríquez MA, Mlynowski TJ, Déry SJ. Reconstructing the natural
498 streamflow of a regulated river: A case study of la grande rivière, Québec, Canada. Can
499 Water Resour J, 2010; 35: 301-16.
500 [29] Piniewski M. Natural streamflow simulation for two largest river basins in Poland: a
501 baseline for identification of flow alterations. Proceedings of the International Association
502 of Hydrological Sciences, 2016; 373: 101-7.
503 [30] Lin PR, Pan M, Beck HE, et al. Global reconstruction of naturalized river flows at 2.94
504 million reaches. Water Resour Res, 2019; 55: 6499–516.
505 [31] Ghiggi G, Humphrey V, Seneviratne SI, et al. GRUN: An observation-based global gridded
506 runoff dataset from 1902 to 2014. Earth Syst Sci Data, 2019; 11: 1655-74.
507 [32] Zhang XJ, Tang Qh, Pan M, et al. A long-term land surface hydrologic fluxes and states

508 dataset for China. J Hydrometeor, 2014; 15: 2067-84.
509 [33] Liang X, Lettenmaier DP, Wood EF, et al. A simple hydrologically based model of land
510 surface water and energy fluxes for general circulation models. J Geophys Res Atmos,
511 1994; 99: 14415-28.
512 [34] Lohmann D, Nolte-Holube R, Raschke E. A large-scale horizontal routing model to be
513 coupled to land surface parametrization schemes. Tellus A, 1996; 48: 708-21.
514 [35] Shen Y, Xiong AY, Wang Y, et al. Performance of high-resolution satellite precipitation
515 products over China. J Geophys Res Atmos, 2010; 115: 1-17.
516 [36] He J, Yang K, Tang WJ, et al. The first high-resolution meteorological forcing dataset for
517 land process studies over China. Sci Data, 2020; 7: 1-11.
518 [37] Myneni RB, Ramakrishna R, Nemani R, et al. Estimation of global leaf area index and
519 absorbed par using radiative transfer models. IEEE Trans Geosci Remote Sens, 1997; 35:
520 1380-93.
521 [38] Gou JJ, Miao CY, Samaniego L, et al. CNRDv1.0: The china natural runoff dataset version
522 1.0. Bull Amer Meteor Soc, 2021; 10.1175/BAMS-D-20-0094.1.
523 [39] Lohmann D, Raschke E, Nijssen B, et al. Regional scale hydrology: I. Formulation of the
524 VIC-2L model coupled to a routing model. Hydrol Sci J, 1998; 43: 131-41.
525 [40] He H. China gauging station network. Advanced in Water Science, 2010; 21: 460-5 (in
526 Chinese).
527 [41] Gupta HV, Sorooshian S, Yapo Patrice O. Status of automatic calibration for hydrologic

528 models: Comparison with multilevel expert calibration. J Hydrol Eng, 1999; 4: 135-43.
529 [42] Hargreaves GH, Samani ZA. Reference crop evapotranspiration from temperature. Appl
530 Eng Agric, 1985; 1: 96-9.
531 [43] Switanek MB, Troch PA, Castro CL, et al. Scaled distribution mapping: a bias correction
532 method that preserves raw climate model projected changes. Hydrol Earth Syst Sc, 2017;
533 21: 2649–66.
534 [44] Bum Kim K, Kwon H-H, Han D. Bias-correction schemes for calibrated flow in a
535 conceptual hydrological model. Hydrol Res, 2021; 52: 196–211.
536 [45] Chen C, Haerter JO, Hagemann S, et al. On the contribution of statistical bias correction
537 to the uncertainty in the projected hydrological cycle. Geophys Res Lett, 2011; 38: L20403.
538 [46] Yan D, Wang K, Qin T, et al. A data set of global river networks and corresponding water
539 resources zones divisions. Sci Data, 2019; 6: 219-29.
540 [47] Tao H, Gemmer M, Bai Y, et al. Trends of streamflow in the tarim river basin during the
541 past 50 years: Human impact or climate change? J Hydrol, 2011; 400: 1–9.
542 [48] Wang GQ, Zhang JY, Jin JL, et al. Assessing water resources in China using PRECIS
543 projections and a VIC model. Hydrol Earth Syst Sc, 2012; 16: 231-40.
544 [49] Xie ZH, Yuan F, Duan QY, et al. Regional parameter estimation of the VIC land surface
545 model: methodology and application to river basins in China. J Hydrometeorol, 2007; 8:
546 447-68.
547 [50] Beck HE, van Dijk AIJM, de Roo A, et al. Global evaluation of runoff from 10 state-of-

548 the-art hydrological models. Hydrol Earth Syst Sc, 2017; 21: 2881-903.
549 [51] Xu HL, Xu C-Y, Chen H, et al. Assessing the influence of rain gauge density and
550 distribution on hydrological model performance in a humid region of China. J Hydrol,
551 2013; 505: 1-12.
552 [52] Tang GQ, Clark MP, Papalexiou SM, et al. Have satellite precipitation products improved
553 over last two decades? A comprehensive comparison of GPM IMERG with nine satellite
554 and reanalysis datasets. Remote Sens Environ, 2020; 240: 111697-715.
555 [53] Xie ZH, Su FG, Xu L, et al. Applications of a surface runoff model with horton and dunne
556 runoff for VIC. Adv Atmos Sci, 2003; 20: 165-72.
557 [54] Liang X, Xie ZH. Important factors in land–atmosphere interactions: surface runoff
558 generations and interactions between surface and groundwater. Glob Planet Change, 2003;
559 38: 101-14.
560 [55] Atkinson SE, Woods RA, Sivapalan M. Climate and landscape controls on water balance
561 model complexity over changing timescales. Water Resour Res, 2002; 38: 50-66.
562 [56] Beck HE, van Dijk AIJM, de Roo A, et al. Global-scale regionalization of hydrologic
563 model parameters. Water Resour Res, 2016; 52: 3599-3622.
564 [57] Nester T, Kirnbauer R, Gutknecht D, et al. Climate and catchment controls on the
565 performance of regional flood simulations. J Hydrol, 2011; 402: 340–56.
566 [58] Merz R, Parajka J, Blöschl G. Scale effects in conceptual hydrological modeling. Water
567 Resour Res, 2009; 45: 1-15.

568 [59] Zhuo L, Dai Q, Han D. Meta-analysis of flow modeling performances-to build a matching
569 system between catchment complexity and model types. 2015; 29: 2463–77.
570 [60] Condon LE, Atchley AL, Maxwell RM. Evapotranspiration depletes groundwater under
571 warming over the contiguous United States. Nat Commun, 2020; 11: 873-80.
572 [61] Pan M, Sheffield J, Wood EF, et al. Snow process modeling in the North American Land
573 Data Assimilation System (NLDAS): 2. Evaluation of model simulated snow water
574 equivalent. J Geophys Res Atmos, 2003; 108: 1-11.
575 [62] Xu J. Underlying gravel layers in a large sand bed river and their influence on downstream-
576 dam channel adjustment. 1996; 17: 351–9.
577

578 **Figure Captions**

579 **Fig. 1.** Distribution of surface water resources over China, indicating locations of the 330
580 natural/near-natural gauge stations used in this study. Surface water volume was derived from
581 *China's Water Resources Bulletin* in 2019
582 (http://www.mwr.gov.cn/sj/tjgb/szygb/202008/t20200803_1430726.html). Gray boundaries
583 delineate 10 water resources regions: I, Songhua River; II, Liao River; III, Hai River; IV, Yellow
584 River; V, Huai River; VI, Yangtze River; VII, Southeast River drainage system; VIII, Pearl
585 River; IX, Southwest River drainage system; and X, Northwest River drainage system.

586 **Fig. 2.** Flow direction extraction and correction. The top two panels display the automatically
587 extracted flow direction information and its general flow patterns throughout China. The bottom
588 six panels show flow directions obtained for three regions of interest before (A, B, and C) and
589 after (A', B', and C') modification of the flow directions of 'land' cells. Arrows and grid colors
590 indicate the D8 flow direction for each grid cell. The red circle denotes the outlet cell of the
591 example catchment within each region of interest, and corrected cells are delineated by red
592 rectangles.

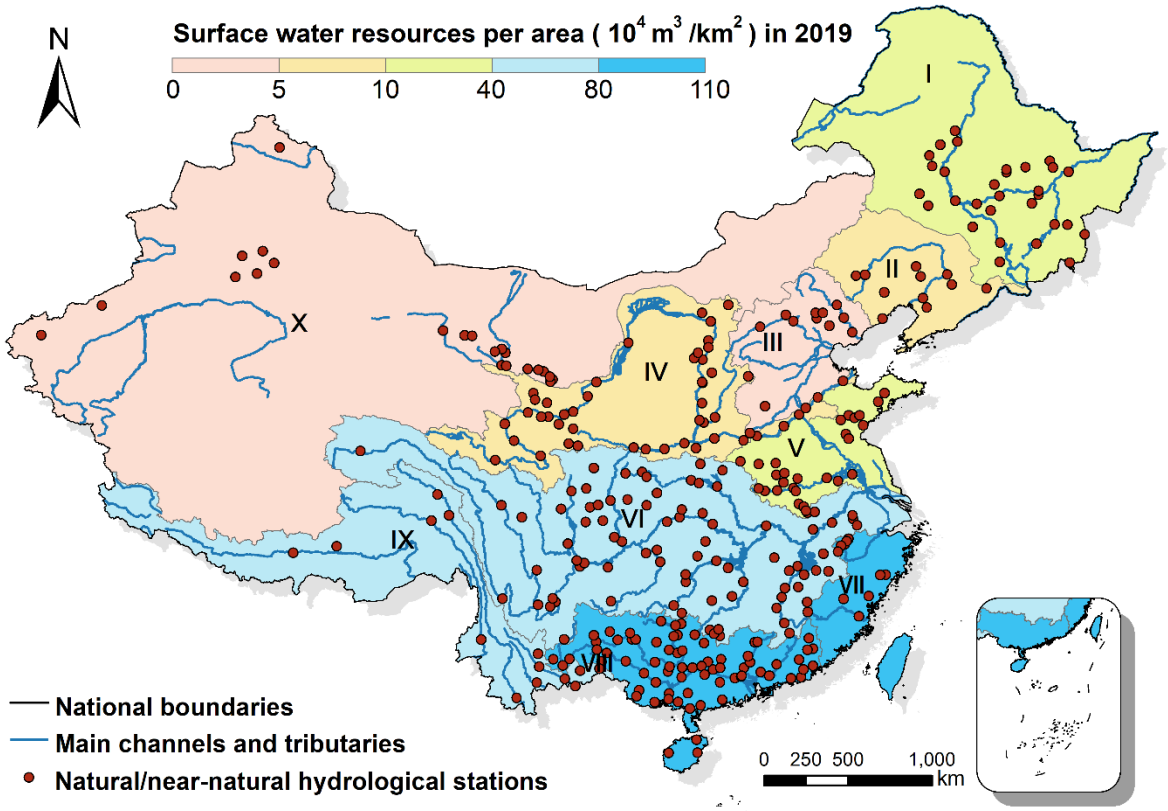
593 **Fig. 3.** Model performance before and after river flow direction correction. (a) Spatial
594 distribution of reconstructed natural river discharge based on the correct flow direction; each
595 grid cell represents the mean upstream natural river discharge in the period from 1961 to 2018.
596 (b) Relationship between drainage area and number of routing cells before and after flow
597 direction correction for data from 330 gauge stations. (c–e) Three examples showing the time-

series of reconstructed streamflow without and with flow direction correction. Red and blue dots/lines indicate the original and modified flow direction. Inferred natural streamflow indicated by orange dots in (c–e) is taken as the benchmark for flow direction evaluation.

Fig. 4. Model performance evaluation by four metrics: CC (correlation coefficient), NSE (Nash Sutcliffe efficiency coefficient), Pbias (percent bias, %), and KGE (Kling-Gupta Efficiency). The left column shows the model training results for 230 training gauge stations during calibration (red dots, 1961–1969) and validation (blue dots, 1970–1979), and the right column shows model results for 100 test gauge stations during the period from 1961 to 1979.

Fig. 5. Spatial pattern of four model performance metrics during the period 1961 to 1979. Gray boundaries indicate the 10 water resources regions of China: I, Songhua River; II, Liao River; III, Hai River; IV, Yellow River; V, Huai River; VI, Yangtze River; VII, Southeast River drainage system; VIII, Pearl River; IX, Southwest River drainage system; and X, Northwest River drainage system.

Fig. 6. Possible impact factors on model performance. a, c, and e are spatial distributions of three catchment descriptors (drought index, area, and runoff efficiency) at 330 gauge stations across China; b, d, and f display relationships between model performance (expressed as NSE) and drought index, area, and runoff efficiency.



617

618 **Fig. 1.** Distribution of surface water resources over China, indicating locations of the 330

619 natural/near-natural gauge stations used in this study. Surface water volume was derived from

620 *China's Water Resources Bulletin* in 2019

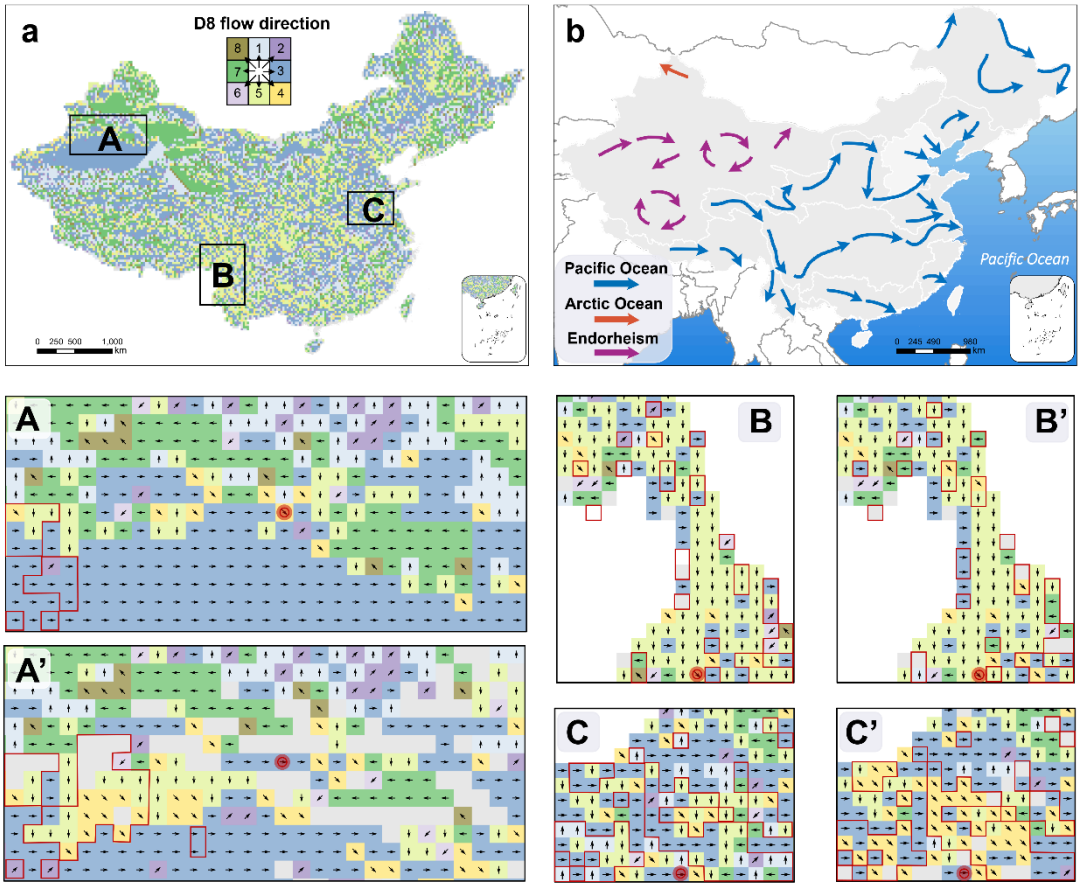
621 (http://www.mwr.gov.cn/sj/tjgb/szygb/202008/t20200803_1430726.html). Gray boundaries

622 delineate 10 water resources regions: I, Songhua River; II, Liao River; III, Hai River; IV, Yellow

623 River; V, Huai River; VI, Yangtze River; VII, Southeast River drainage system; VIII, Pearl

624 River; IX, Southwest River drainage system; and X, Northwest River drainage system.

625



626

627 **Fig. 2.** Flow direction extraction and correction. The top two panels display the automatically

628 extracted flow direction information and its general flow patterns throughout China. The bottom

629 six panels show flow directions obtained for three regions of interest before (A, B, and C) and

630 after (A', B', and C') modification of the flow directions of 'land' cells. Arrows and grid colors

631 indicate the D8 flow direction for each grid cell. The red circle denotes the outlet cell of the

632 example catchment within each region of interest, and corrected cells are delineated by red

633 rectangles.

634

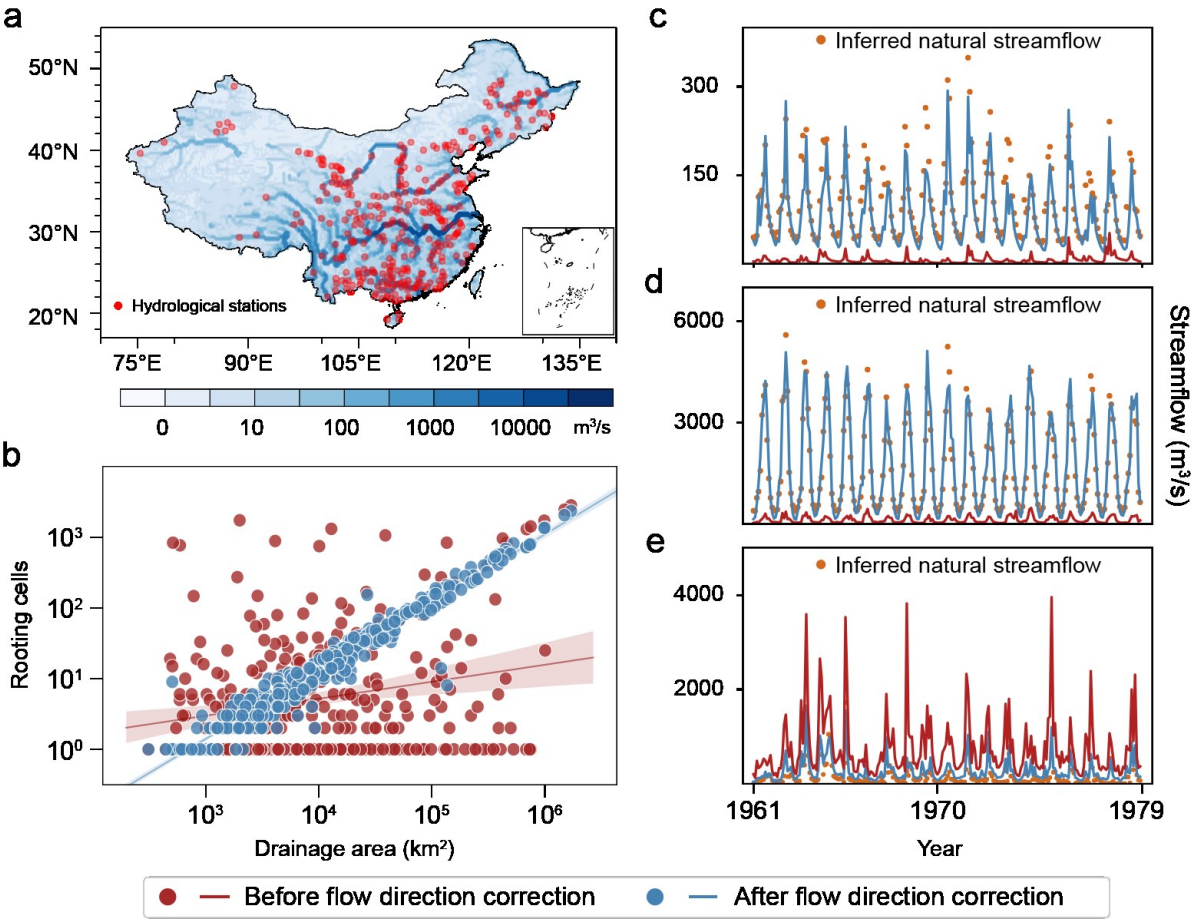
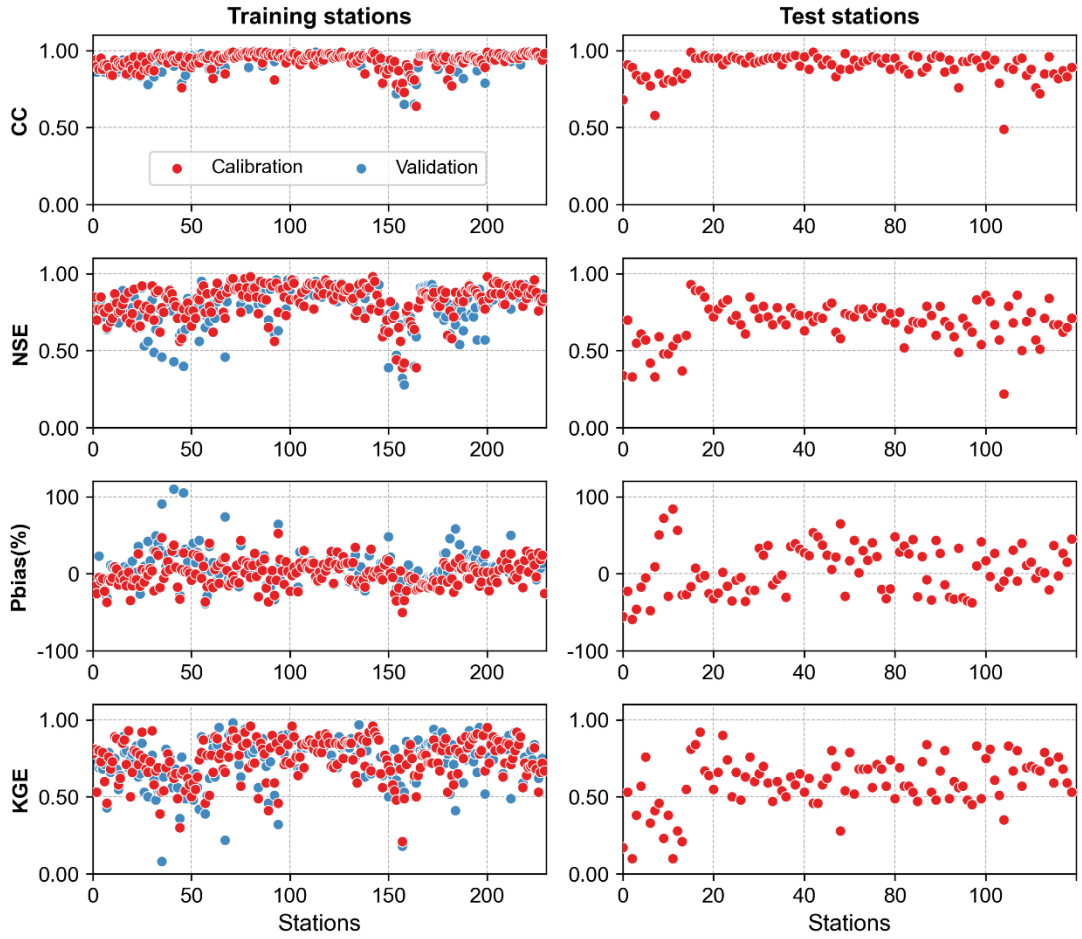


Fig. 3. Model performance before and after river flow direction correction. (a) Spatial distribution of reconstructed natural river discharge based on the correct flow direction; each grid cell represents the mean upstream natural river discharge in the period from 1961 to 2018. (b) Relationship between drainage area and number of routing cells before and after flow direction correction for data from 330 gauge stations. (c–e) Three examples showing the time-series of reconstructed streamflow without and with flow direction correction. Red and blue dots/lines indicate the original and modified flow direction. Inferred natural streamflow indicated by orange dots in (c–e) is taken as the benchmark for flow direction evaluation.



645

646 **Fig. 4.** Model performance evaluation by four metrics: CC (correlation coefficient), NSE (Nash

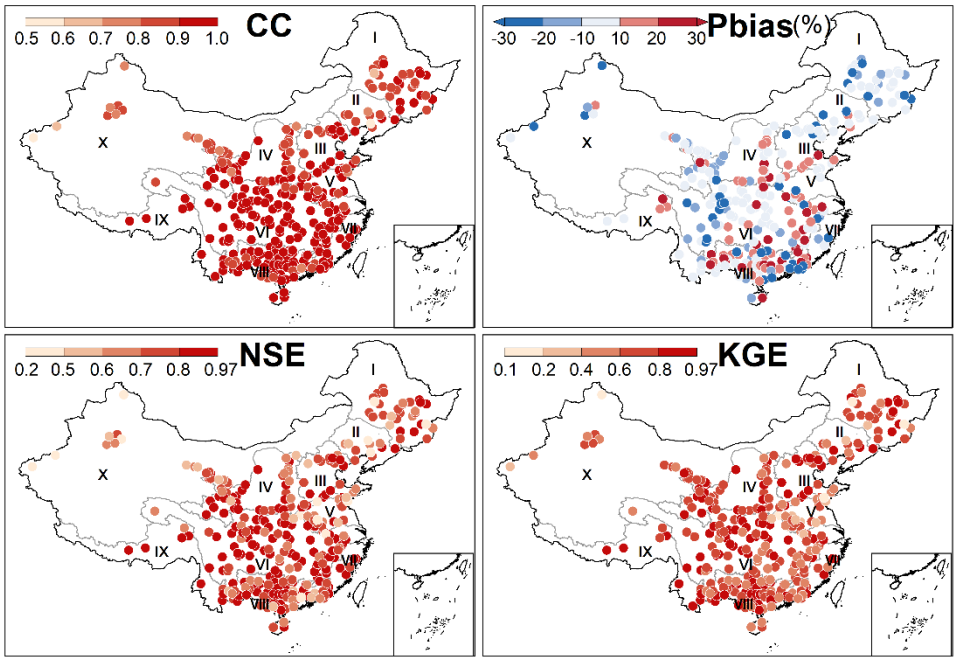
647 Sutcliffe efficiency coefficient), Pbias (percent bias, %), and KGE (Kling-Gupta Efficiency).

648 The left column shows the model training results for 230 training gauge stations during

649 calibration (red dots, 1961–1969) and validation (blue dots, 1970–1979), and the right column

650 shows model results for 100 test gauge stations during the period from 1961 to 1979.

651



652

653 **Fig. 5.** Spatial pattern of four model performance metrics during the period 1961 to 1979. Gray

654 boundaries indicate the 10 water resources regions of China: I, Songhua River; II, Liao River;

655 III, Hai River; IV, Yellow River; V, Huai River; VI, Yangtze River; VII, Southeast River

656 drainage system; VIII, Pearl River; IX, Southwest River drainage system; and X, Northwest

657 River drainage system.

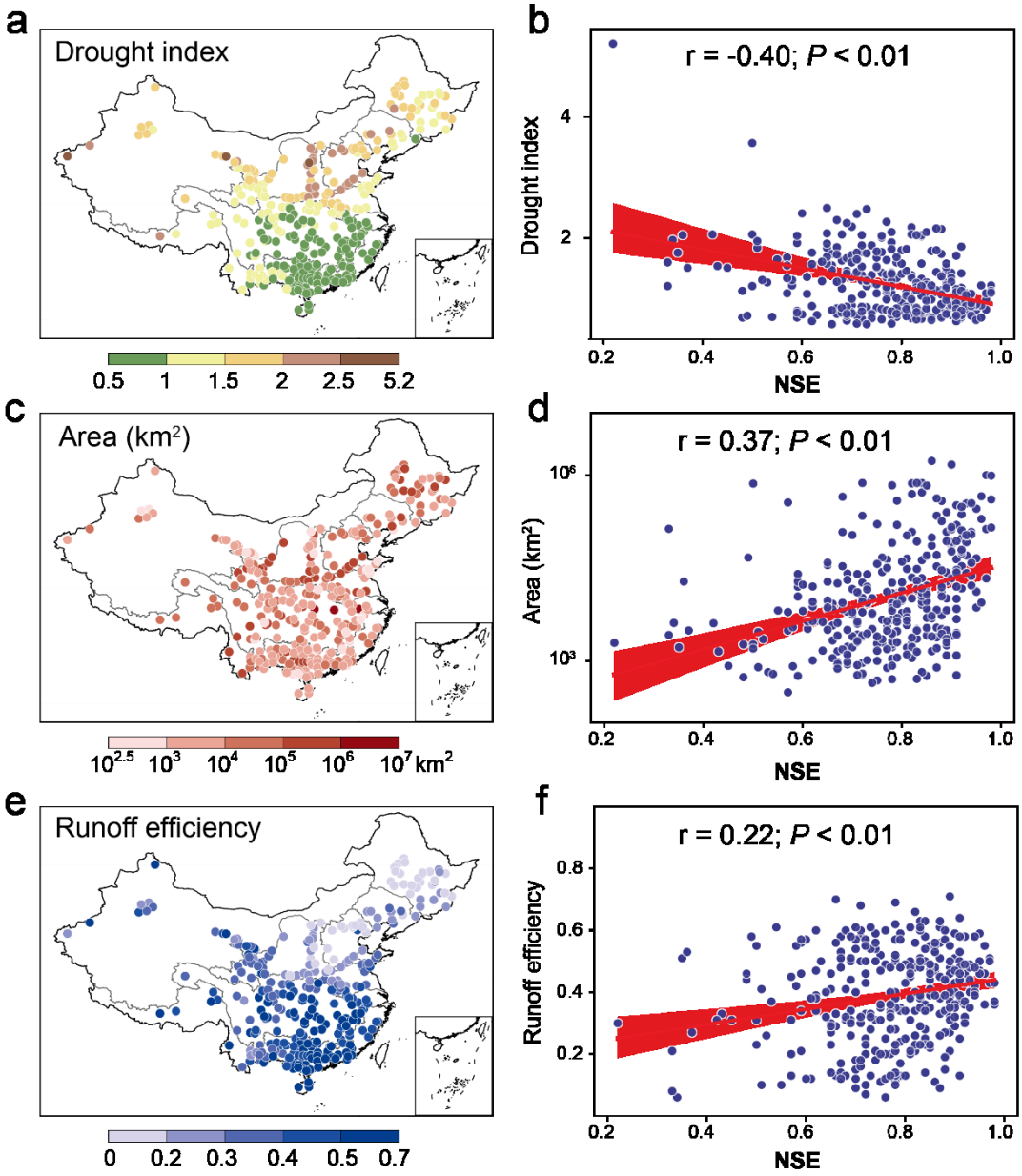


Fig. 6. Possible impact factors on model performance. a, c, and e are spatial distributions of three catchment descriptors (drought index, area, and runoff efficiency) at 330 gauge stations across China; b, d, and f display relationships between model performance (expressed as NSE) and drought index, area, and runoff efficiency.