

## ARTICLE OPEN



# Changes in China's river water quality since 1980: management implications from sustainable development

Hanxiao Zhang<sup>1,6</sup>, Xianghui Cao<sup>2,6</sup>, Shouliang Huo<sup>1✉</sup>, Chunzi Ma<sup>1</sup>, Wenpan Li<sup>3</sup>, Yong Liu<sup>4</sup>, Yingdong Tong<sup>5</sup> and Fengchang Wu<sup>1</sup>

Human activities and climate change threaten water quality in China's rivers. We simulated the monthly concentrations of riverine total nitrogen (TN), ammonia-nitrogen (NH<sub>3</sub>-N), total phosphorus (TP), and chemical oxygen demand (COD<sub>Mn</sub>) in 613 sub-watersheds of the nation's 10 major river basins during the 1980–2050 period based on a 16-year (2003–2018) monitoring dataset using the stacking machine-learning models. The results showed that water quality improved markedly, except for the TN concentration, which was probably due to the lack of a TN control target and assessment system. Quantitative analysis indicated that anthropogenic factors were the primary controls compared with climatic drivers and geographical drivers for TN, TP, and NH<sub>3</sub>-N concentrations. On the basis of all 17 sustainable development goals (SDGs) relevant to water quality in China, the water resources, water environment, aquatic ecology and water security should be considered collectively to achieve improvements in the ecological status of China's rivers.

*npj Clean Water* (2023)6:45 | <https://doi.org/10.1038/s41545-023-00260-y>

## INTRODUCTION

Rivers are important channels for matter migration and transport between land and lakes or oceans, providing abundant fresh water resources for potable water, irrigation, aquaculture, navigation, and power generation<sup>1,2</sup>. However, river ecosystems are experiencing widespread deterioration and are globally threatened by anthropogenic activities and climate change<sup>3,4</sup>. A global study revealed that nearly 80% (4.8 billion) of the world's population (for 2000) lives in areas with a high incidence of threat (>75%) to human water security<sup>5</sup>. Even worse, one-third of the world's population lacks access to safe potable water<sup>6</sup>. In the face of current challenges, it is urgent to diagnose the threats to river water quality over a broad range of time and space scales, remedy their underlying causes, and limit the threats from the source to protect river freshwater resources<sup>5</sup>.

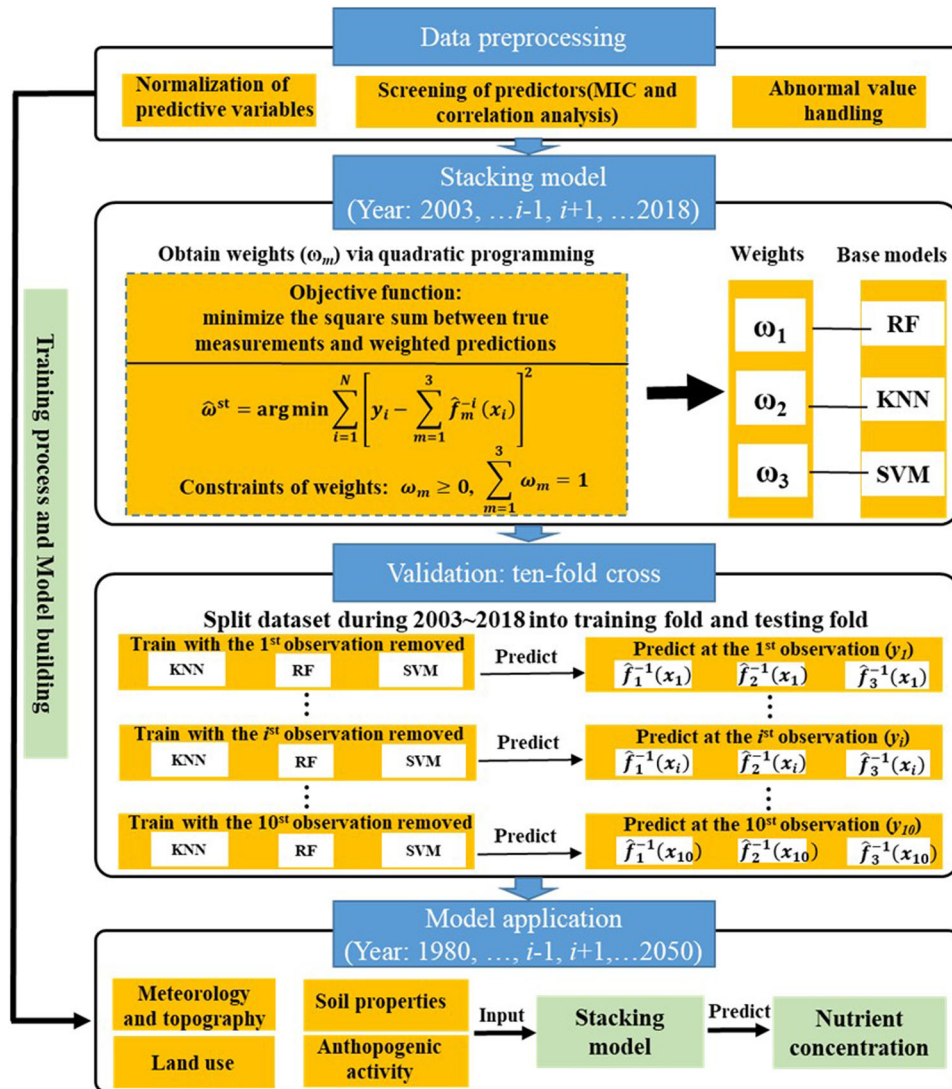
China's rivers have suffered profound water quality impairments due to the undeniable pressure of economic development on the environment since China's Reform and Opening-up in 1978<sup>7</sup>. Water pollution in China has been confirmed to be a major cause of 40 billion cubic meters of water shortage in China per year<sup>8</sup>. The elevated input of anthropogenic nutrients is a critical cause of reduced water quality in Chinese rivers. According to estimates from multi-scale models, the total dissolved nitrogen (TDN) and total dissolved phosphorus (TDP) input to rivers in China in 2012 were 28 Tg and 3 Tg, respectively<sup>9</sup>. Furthermore, excess nutrients from rivers were transported to lakes and the ocean, resulting in frequent episodes of blooms and red tide, endangering human and aquatic health and ecosystem services<sup>10</sup>. Fortunately, inland water quality across China displayed marked improvement or was maintained at favorable levels nationwide from 2003 to 2017, which is attributed to reductions in nutrient discharge<sup>11,12</sup>. In 2022, a national investigation of 3641 sampling sites in rivers, lakes, and reservoirs across China showed that 12.1% of sampling

sites had water quality lower than Class III according to the China Surface Water Environmental Quality Standard (GB3838-2002), while 0.7% of sites had a more severe condition at worse than Class V<sup>13</sup>. Over the past four decades since the Reform and Opening-up, under the background of balancing economic development and environmental protection, it is essential to identify the water quality patterns and underlying mechanisms in China's rivers to provide references and information for river water quality protection in developing countries<sup>7</sup>.

Several studies have been devoted to the patterns of water quality in China's rivers and their associated drivers, such as quantifying inputs of N and P to Chinese rivers from different sources at multiple scales<sup>9</sup>, the cycle of nutrients in river systems including sources, transformation, and flux<sup>14,15</sup>, and the spatial water quality patterns and critical covariates of river impairment<sup>12,16–18</sup>. However, there are still gaps in understanding the spatiotemporal variation and underlying mechanism of water quality of China's rivers over the past four decades. First, the lack of long-term regular frequency nationwide monitoring data is the major bottleneck in the study of impacts on river water quality, because the traceable and available monitoring data only extends from 2003<sup>12,19</sup>. Second, the identification of the driving mechanisms of river water quality variation are subject to the temporal and spatial scale resolution of factorial models and explanatory variables (including natural geographic characteristics, socio-economic indicators, land use data, and meteorological factors)<sup>12,17</sup>. Finally, it is challenging to bridge the barriers to scientific research and management applications, and apply the understanding of historical river water quality variation and driving mechanisms to future water quality management and the achievement of the sustainable development goals (SDGs)<sup>7,10</sup>.

The study assembled a 16-year (2003–2018) monthly data from 613 riverine water-quality monitoring sites as well as watershed

<sup>1</sup>State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China. <sup>2</sup>China Institute of Geological Environmental Monitoring, Beijing 100081, China. <sup>3</sup>Water Office, China National Environmental Monitoring Centre, Beijing 100012, China. <sup>4</sup>Key Laboratory of Water and Sediment Sciences Ministry of Education, College of Environmental Science and Engineering, Peking University, Beijing 100871, China. <sup>5</sup>College of Environmental Science and Engineering, Tianjin University, Tianjin 300072, China. <sup>6</sup>These authors contributed equally: Hanxiao Zhang, Xianghui Cao. ✉email: huoshouliang@126.com



**Fig. 1 Framework used for simulating nutrient concentrations in Chinese rivers.** Details of the model stacking process including data processing, stacking model, ten-fold cross validation, and model application. Monitor data of TN, TP,  $\text{NH}_3\text{-N}$ , and  $\text{COD}_{\text{Mn}}$  concentration from 2003 to 2018 are used to stimulate the monthly data of 1980 to 2018, and the interdecadal data 2020–2050 under two future scenarios (SSP2-RCP4.5 and SSP5-RCP8.5).

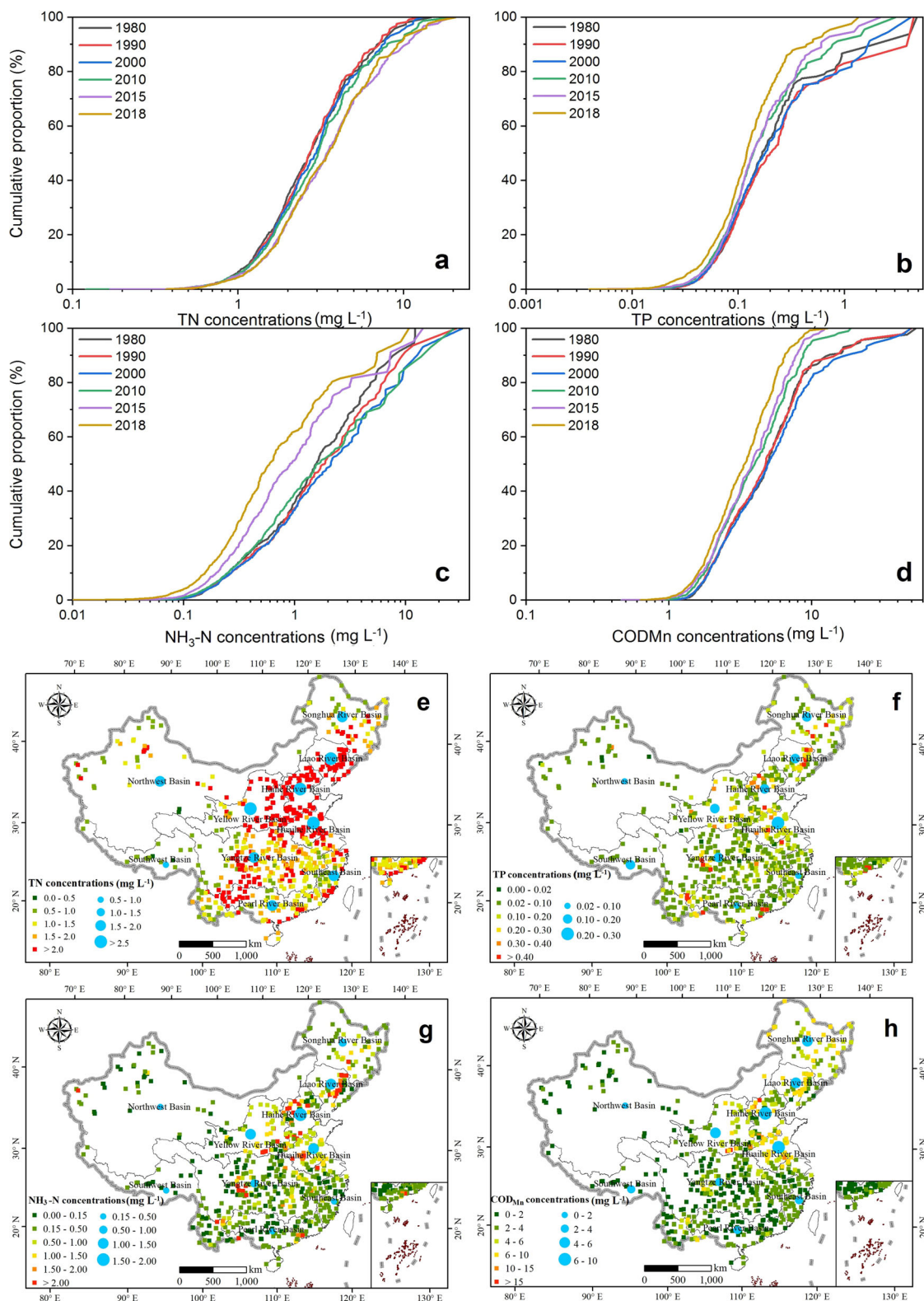
characteristics (e.g., longitude, latitude, land-use patterns, net anthropogenic N/P inputs, and soil properties) and climate conditions at a national scale to build a set of stacking machine-learning models. The stacking machine-learning models integrated the different base models' results, which could reduce variance and improve the stability of the final model<sup>20</sup>. Three base models were selected for their high popularity and performance based on previous studies, including random forest (RF), support vector machine (SVM), and k-nearest neighbors (KNN)<sup>19,20</sup>. The stacking model was used to simulate and predict the annual and monthly variations in river water quality during the period of 1980 to 2018 (Fig. 1). We then use two future scenarios (SSP2-RCP4.5 and SSP5-RCP8.5) to predict the decadal trends in water quality between 2020–2050. Multiple linear regression (MLR) models and correlation analyses were employed to quantify the relative contributions of anthropogenic, climatic and geographical factors to changes in riverine TN,  $\text{NH}_3\text{-N}$ , and TP as well as the  $\text{COD}_{\text{Mn}}$ . On the basis of the relations of SDGs to water quality, sustainable water quality management policies were proposed to achieve a better aquatic environment for China's rivers, as well as other developing countries.

## RESULTS AND DISCUSSION

### Spatiotemporal trends in TN, TP, $\text{NH}_3\text{-N}$ , and $\text{COD}_{\text{Mn}}$ concentrations

Comparison between measured and predicted values in the 10 large river basins showed that our machine-learning models were generally able to recreate TN, TP,  $\text{NH}_3\text{-N}$ , and  $\text{COD}_{\text{Mn}}$  concentrations at a significance level of  $p < 0.01$  and with a low predictive bias estimated using the  $R^2$ , root mean square error (RMSE), Nash-Sutcliffe efficiency (NSE), and the mean absolute error (MAE) (Supplementary Fig. 1, Supplementary Table 1). The accuracy statistics of 10-fold cross validation on both test and validation datasets also indicated the goodness of fit and predictive power of the stacking model for TN, TP,  $\text{NH}_3\text{-N}$ , and  $\text{COD}_{\text{Mn}}$  concentrations (Supplementary Tables 2–5). Although this data-driven machine learning stacking model had limited interpretability and deductibility, it showed promise with its robustness and stability in the induction capability of the educated ensemble machine learning<sup>19,20</sup>.

The concentration trends of TN, TP,  $\text{NH}_3\text{-N}$  and  $\text{COD}_{\text{Mn}}$  were different throughout the country from 1980 to 2018 (Fig. 2a–d).



**Fig. 2** Nutrient concentrations in Chinese rivers between 1980 and 2018. Panels a–d display the cumulative proportions of the stimulated annual average concentration of TN, TP, NH<sub>3</sub>-N, and CODMn in China's rivers. e–h Average nutrient concentrations in 613 rivers of China from 1980 to 2018 (mg L<sup>-1</sup>). The blue circle size represents the average nutrient concentrations for the 10 major basins during 1980 to 2018.

The proportion of sampling sites with simulated TN concentrations lower than 1.5 mg L<sup>-1</sup> in 1980, 1990, 2000, 2010, 2015, and 2018 was 19.83%, 20.25%, 19.60%, 18.65%, 17.41%, and 18.12%, respectively. These trends suggest that TN pollution increased

during the observational period. At present, although there is a threshold of river TN concentrations in the China Surface Water Environmental Quality Standard (that is, 1 mg L<sup>-1</sup> for Class III), it has not been included in the assessment system by water

environment management<sup>21</sup>. NH<sub>3</sub>-N concentrations increased from 1980 to 2010, and then decreased from 2010 to 2018. COD<sub>Mn</sub> increased from 1980 to 2000, and then decreased from 2000 to 2018 (Fig. 2). This result is generally consistent with the findings of a previous study, which found that NH<sub>3</sub>-N and COD concentrations overall decreased from 2003 to 2017 based on monthly monitoring data obtained from inland water bodies (including rivers and lakes) in China<sup>12</sup>. TP concentrations increased from 1980 to 2015, and then decreased from 2015 to 2018 (Fig. 2). In 2000, the Chinese government proposed the Total Amount of Pollutants Control Plan, taking COD as one of the control indexes of 12 major pollutants, and remarkable results have been achieved<sup>7,22</sup>. The Law of the People's Republic of China on the Prevention and Control of Water Pollution enacted in 2008 has strictly tightened the regulation of water environmental protection<sup>23</sup>. Then, the construction of an ecological civilization was elevated to a national strategy in 2012, and a series of discharge reduction targets for TP, COD, and NH<sub>3</sub>-N were set during the 13th 5-Year Comprehensive Work Plan from 2016 to 2020<sup>24</sup>. However, nutrient concentrations are not expected to decrease under the influence of future human activities and climate change (Supplementary Fig. 2). Thus, the smooth implementation of these policies and action plans is still needed to gradually decouple economic growth from its environmental impacts<sup>7,10</sup>.

Due to differences in geographical conditions among the watersheds, changes in nutrient concentrations differed between basins during the period 1980–2018 (Supplementary Figs. 3–6). Except for the Huaihe River ( $n = 48$ ,  $p < 0.05$ ,  $r = -0.21$ ), average TN concentrations increased, especially within the Southeast River ( $n = 41$ ,  $p < 0.05$ ,  $r = 0.86$ ) and the Pear River ( $n = 66$ ,  $p < 0.01$ ,  $r = 0.97$ ), where the temporal trends were statistically significant (Supplementary Fig. 3). Average TP concentrations decreased in most basins of China, especially within the Songhua River ( $n = 45$ ,  $p < 0.05$ ,  $r = -0.75$ ) and the Yellow River ( $n = 67$ ,  $p < 0.01$ ,  $r = -0.76$ ) (Supplementary Fig. 4). However, NH<sub>3</sub>-N concentrations and COD<sub>Mn</sub> displayed several phased changes in some basins. NH<sub>3</sub>-N and COD<sub>Mn</sub> concentrations increased during the early part of the observational period, whereas NH<sub>3</sub>-N and COD<sub>Mn</sub> concentrations decreased in most regions of China in recent years (Supplementary Figs. 5, 6).

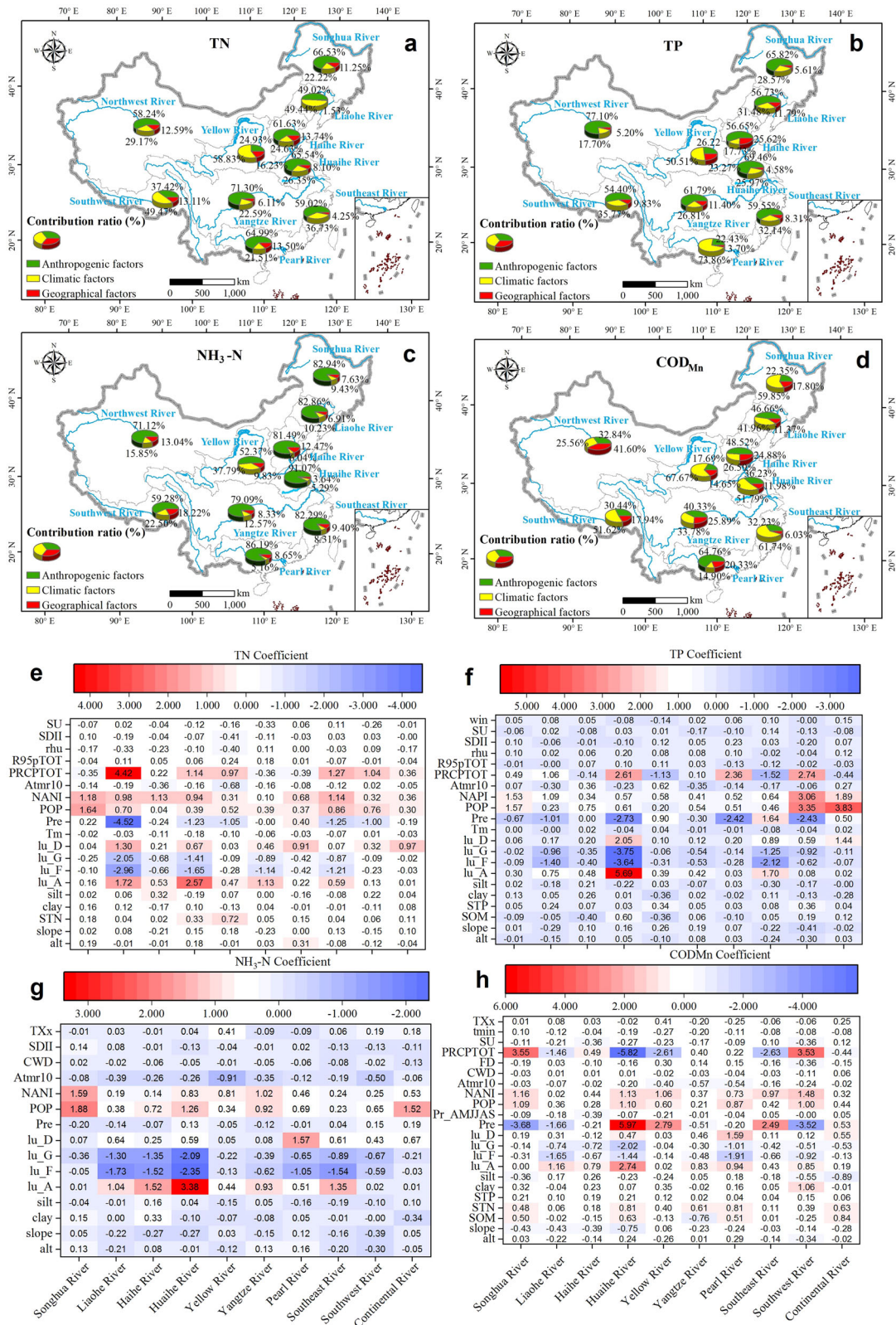
Our results show that TN concentrations exhibited significant spatial variations with poorer water quality in eastern China between 1980–2018 ( $p < 0.01$ ; Fig. 2e). TP, NH<sub>3</sub>-N and COD<sub>Mn</sub> concentrations did not vary significantly between basins (Fig. 2f–h). The proportions of sampling sites with TN concentration greater than 1.5 mg L<sup>-1</sup> was 62.3%, which suggests that TN represents a relatively serious pollution problem in most regions of China under the current water quality standards. The proportion of sampling sites with simulated observations where TP concentration was greater than 0.2 mg L<sup>-1</sup>, NH<sub>3</sub>-N concentration was greater than 1.0 mg L<sup>-1</sup> and the proportion of COD<sub>Mn</sub> concentration was greater than 6.0 mg L<sup>-1</sup> was 13.3%, 16.3%, and 13.7%. Compared with the historical distribution of nutrients, the proportion with higher concentrations increased between the period of 2020–2050 (Supplementary Fig. 7). Under the SSP2-RCP4.5 scenario, the proportions of simulated observations where TN concentration is greater than 1.5 mg L<sup>-1</sup>, TP concentration is greater than 0.2 mg L<sup>-1</sup>, NH<sub>3</sub>-N concentration is greater than 1.0 mg L<sup>-1</sup> and the proportion of COD<sub>Mn</sub> concentration is greater than 6.0 mg L<sup>-1</sup> are 75.8%, 18.2%, 40.2%, and 17.6%, respectively. Under SSP5-RCP8.5 scenario, the proportion of total simulated observations where TN, TP, NH<sub>3</sub>-N and COD<sub>Mn</sub> concentration were greater than 1.5 mg L<sup>-1</sup>, 0.2 mg L<sup>-1</sup>, 1.0 mg L<sup>-1</sup>, and 6.0 mg L<sup>-1</sup> are 75.2%, 19.6%, 43.2%, and 17.5%, respectively. The results make it clear that human activities and climate change will significantly influence riverine nutrient concentrations, especially within the Yellow, Huaihe and Haihe rivers ( $p < 0.01$ ).

### Attribution of riverine water quality variability

The covariates related to nutrient concentration were divided into three categories including anthropogenic, climatic and geographical factors. The contribution ratios of these anthropogenic, climatic and geographical factors for nutrient concentrations are quantified by inputting data from each of the 10 river basins into the MLR model separately (Fig. 3a–d). Among the covariates considered, the anthropogenic predictors had a larger contribution (24.93–71.29% for TN, 22.43–77.10% for TP, and 52.37–91.06% for NH<sub>3</sub>-N) compared with climatic drivers (21.51–58.83% for TN, 17.73–73.86% for TP, and 5.16–37.79% for NH<sub>3</sub>-N) and geographical drivers (4.25–16.23% for TN, 3.70–25.61% for TP, and 6.90–18.21% for NH<sub>3</sub>-N) when inputting data into the MLR model using the data for each of the 10 river basins (Fig. 3a–d). However, the pattern of contribution ratios for the anthropogenic, climatic, and geographical factors for COD<sub>Mn</sub> differed between river basins. In general, the geographical drivers had a larger contribution (41.60%) than the climatic drivers (25.55%) and anthropogenic predictors (32.83%) in the Northwest Inland River. The climatic drivers had a larger contribution to the COD<sub>Mn</sub> concentration compared with geographical drivers and anthropogenic predictors in the Songhua River, Yellow River, Huaihe River, Southwest River and Southeast River. This finding may be due to the fact that river nitrogen and phosphorus are mainly derived from anthropogenic sources and are therefore highly correlated with human activities<sup>25–27</sup>. A previous study found that point sources accounted for 75% of the TDP input and agricultural non-point sources accounted for 72% of TDN input in Chinese rivers in 2012<sup>9</sup>. However, the possible sources of COD may be more related to natural sources, including endogenous sources from the degradation of algae and aquatic plants and the release of sediments, and exogenous sources from atmospheric sedimentation and the import of terrestrial vegetation and soil organic matter<sup>28,29</sup>. Climate change-induced rising temperature, hydrological intensification, and extreme weather affect the timing and magnitude of dissolved organic matter delivery from terrestrial ecosystems to surface water<sup>2,29</sup>.

Moreover, the regression coefficients between TN, TP, and NH<sub>3</sub>-N concentrations and anthropogenic drivers were higher than those for the natural drivers and geographical drivers (Fig. 3e–h). Among the selected geographical drivers, elevation, slope, soil total nitrogen (STN), soil total phosphorus (STP), and soil organic matter (SOM) displayed a discernible contribution to nutrient variability in most river basins. Precipitation factors (Pre and PRCPTOT) were strong predictors for the concentrations of TN, TP, and COD<sub>Mn</sub> across the 10 studied basins, with the exception of NH<sub>3</sub>-N. In contrast, air temperature (Atmr10) was a weak predictor of nutrient concentrations across the 10 studied basins (Supplementary Table 6).

Regarding the selected anthropogenic drivers, our analysis shows that the percentage of farmland, forestland, grassland, urban area, population and anthropogenic N and P inputs were critical covariates for nutrient riverine levels (Fig. 3e–h). The extent of urban area and farmland within any given watershed showed a consistently positive relationship with nutrient concentrations. By contrast, forestland and grassland displayed a negative relationship with nutrient concentrations. Population had a distinct signature on nutrients levels in the Songhua (NH<sub>3</sub>-N, TN, TP and COD<sub>Mn</sub>), Huaihe (NH<sub>3</sub>-N and COD<sub>Mn</sub>), Southwest (TP), and Northwest Inland (NH<sub>3</sub>-N and TP) rivers. A relevant, previously conducted study used nighttime light intensity to characterize population and concluded that the nighttime light intensity had a distinct signature (contribution > 35%) on the two nutrient levels (TP and NH<sub>3</sub>-N) in the Yellow and Pearl River Basins<sup>17</sup>, which is different from our study. Anthropogenic N and P inputs had higher contributions to the variability of both nutrients in the Songhua, Haihe, Huaihe, Yangtze, Southwest, and Northwest Inland rivers, where a somewhat higher contribution was



**Fig. 3** Evaluation of the percentage contribution and effect of anthropogenic, climatic and geographical factors to nutrient variability in ten major basins between 1980 and 2018. **a–d** Percentage contributions of anthropogenic, climatic, and geographical factors used multiple linear regression. **e–h** Regression coefficients for every predictor.

registered with the regression coefficient of  $>1$  (Fig. 3e–h). The results suggested that with socio-economic growth, the rivers in western and inland regions of China (except eastern regions) have registered severe water quality impairments, which should be given more attention in the future.

### Analysis of natural and anthropogenic drivers

Our attempt to characterize the signature of anthropogenic activities required an assessment of the role of natural factors, such as meteorological conditions and geographical characteristics, in shaping riverine water quality. When considering all 613 sub-watersheds across China, our analysis was able to discern a weak negative relationship between elevation and/or slope and nutrient concentrations (Fig. 3e–h), suggesting that sites at higher elevations and/or that possessed steeper slopes exhibited lower nutrient levels. In general, anthropogenic activities increase with the decrease of altitude, and plains and lowland areas are susceptible to intensive agricultural cultivation, livestock and poultry farming, urban development, and population aggregation<sup>30</sup>. The slope of the watershed determines the speed of water flow and the severity of soil erosion. Compared with lowland rivers, mountainous rivers with higher slopes are expected to experience faster flow rates and more severe erosion, resulting in shorter retention times for water and a weaker river self-purification capacity<sup>2</sup>. Our results confirmed the negative correlation between the altitude and nutrient concentration of China's rivers, and found that the theoretical positive effect of slope on nutrient concentration was offset by other human activities, such as impoundment to regulate hydraulic and nutrient retention<sup>17,31</sup>.

The timing and magnitude of exogenous nutrient inputs and the factors promoting internal nutrient migration and transformation in rivers may be affected by long-term meteorological forcing<sup>32</sup>. Increasing air temperature will influence the riverine thermal regimes, as well as the physical and chemical properties of water (i.e., water pH, salinity, solubility, viscosity, and diffusion rates), further affecting biochemical processes such as nitrification, denitrification, sediment mineralization and re-release<sup>14,33,34</sup>. The role of precipitation or extreme precipitation is also a dominant factor affecting hydrological regimes, including hydraulic characteristics, water level, flow rate, inundation pattern, and water cycles<sup>3,35</sup>. Moreover, changes in the amount, frequency, and intensity of precipitation will mobilize nutrients on land through surface and subsurface processes for gathering non-point pollution, and release higher concentrations of sediment through erosion and resuspension<sup>36,37</sup>. A previous study suggested that precipitation dominated the interannual variability of riverine N loading across the continental United States during the period of 1987–2007<sup>3</sup>. Our results show that the relationship between nutrient concentrations in rivers and mean and extreme meteorological factors varies according to geographical region and water quality indicators (Fig. 3), possibly due to the covariance between meteorological factors and other natural and human activity variables<sup>14,33</sup>.

Compared to the relatively minor effects of climate and geographical factors considered, the population, NANI/NAPI, and the percentage of specific land-use types were found to be the stronger predictors of riverine nutrients levels (except for  $\text{COD}_{\text{Mn}}$ ) and explained most of the nutrient variation collectively (Fig. 3). It has been confirmed that continuous urbanization and intensive agricultural development have had a profound impact on nutrient inputs from land to rivers<sup>25,26</sup>. It is noteworthy that the average explanatory rates of the predicted variables used after screening for the variations in  $\text{COD}_{\text{Mn}}$ ,  $\text{NH}_3\text{-N}$ , TN, and TP concentrations in the 10 river basins were 52.92%, 35.74%, 72.15%, and 31.97%, respectively (Supplementary Table 7). Pollution control measures, as indicated by the proportion of land with drainage systems and the capacity of sewage treatment plants, and the construction of

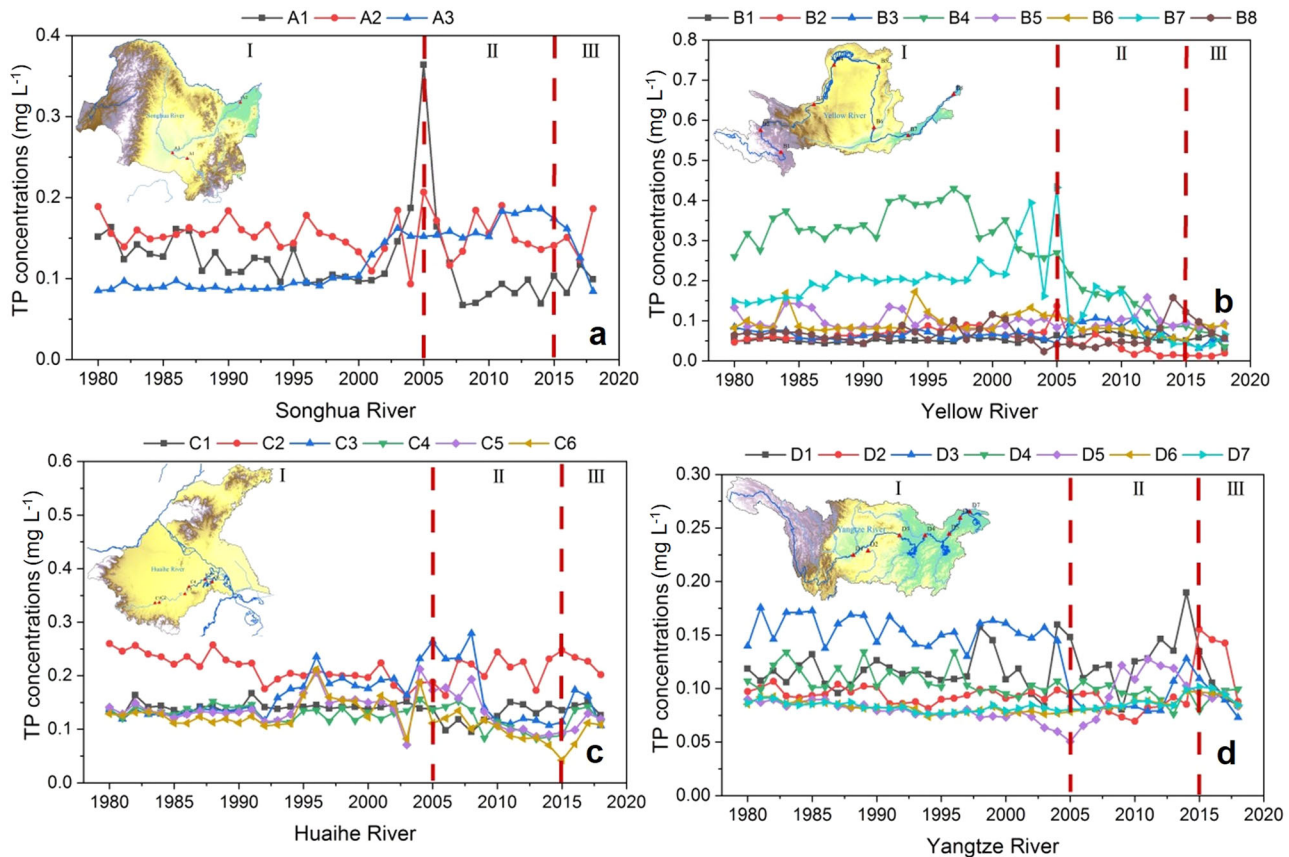
water conservancy facilities should be taken into account in further studies due to their impacts on river nutrient input and migration, although it is difficult to collect these data with high resolution and accuracy in China since 1980<sup>17</sup>.

### Management implications for future environmental efforts

Our analysis suggested that anthropogenic activities and natural factors have a significant impact on riverine nutrients levels. In addition, changes in water environmental management policies have played an important role in water quality improvement. As the discharge control standard gradually received attention, the water environmental management policy was transferred to target the discharge of COD since 2000. Therefore,  $\text{COD}_{\text{Mn}}$  concentrations have decreased since then, especially in the Yellow River and the Huaihe River (Supplementary Fig. 8). The Action Plan for Prevention and Control of Water Pollution in China was formulated in 2015 to strengthen the prevention, and control of water pollution, and all water functional areas are required to meet the water quality requirements. During the period of 2007–2017, the N and P loads exported from agriculture have significantly decreased from  $1.598 \times 10^9$  to  $7.195 \times 10^8$  kg and  $1.087 \times 10^8$  to  $7.62 \times 10^7$  kg, respectively<sup>38</sup>. TP and  $\text{NH}_3\text{-N}$  concentrations have also decreased with the change in water policy in recent years (Fig. 4, Supplementary Fig. 9). However, TN was not included in China's surface water quality control targets over the past years, and thus TN concentrations have not significantly decreased (Supplementary Figs. 3 and 10), which will result in potential detriment to aquatic ecosystems. Therefore, some mitigation measures should be taken to manage N to restore water quality in China<sup>21</sup>.

Although the recent decreases in TP,  $\text{NH}_3\text{-N}$ , and  $\text{COD}_{\text{Mn}}$  concentrations in most rivers indicates that China's nutrient control measures have been effective, it has to be noted that reaching a good ecological status will required significant time (Supplementary Fig. 2). Currently, problems exist within China's water environmental management system, such as unified water environmental quality standards for all regions, separation of water quantity management and water quality management, and coordination between different management organizations. The findings from this study indicate that China now needs more flexible regional water strategies to cope with the different regional trends and sources of nutrient loadings to freshwaters. The ongoing revision of China's Water Pollution Prevention and Control Law should propose significant changes to the current water governance structure and reflect its flexibility between regions to further reduce pollutant loadings and nutrient concentrations in rivers.

A sustainable pathway is essential for achieving a reduction in riverine TN, TP, and  $\text{NH}_3\text{-N}$  concentrations as well as  $\text{COD}_{\text{Mn}}$  in the near future. All 17 SDGs have targets that are relevant to water quality in rivers in China<sup>10</sup>. Two SDGs, namely SDG 6 "Clean Water and sanitation" and SDG14 "Life under Water", are particularly relevant to water quality. For example, reducing nutrient pollution in shallow groundwater and surface water resources could help to achieve universal and equitable access to safe and affordable drinking water by 2030 (SDG 6.1)<sup>8</sup>. Achieving a substantial increase water-use efficiency across all sectors and ensuring sustainable withdrawals and a sustainable supply of freshwater to address water scarcity by 2023 (SDG 6.4) could reduce nutrient pollution in water systems via improving water-use efficiency in agriculture and upgrading the domestic wastewater pipe networks to reduce nutrient leaching and runoff to waters. Reducing the river export of nutrients could help to achieve the target of the prevention and significant reduction of marine pollution of all kinds (in particular pollution due to land-based activities), and including marine debris and nutrient pollution by 2025 (SDG 14.1)<sup>10</sup>. On the basis of sustainable development goals, the water resources, water environment, aquatic ecology and water risk



**Fig. 4** Temporal trends in TP concentrations with a change in water environmental management policies in several typical watersheds between 1980 and 2018. I represents the standard discharge control used between 1980 and 2005; II represents the target total amount control approach used between 2005 and 2015; and III represents the water environmental quality improvement between 2015 and 2018. **a** Songhua River; **b** Yellow River; **c** Huaihe River; and **d** Yangtze River.

should be considered together to achieve good ecological status in China's rivers. In formulating future policies, special attention needs to be paid to pollution discharge, sewage systems and climate change regarding their economic, societal, institutional and technical feasibility in ensuring the effectiveness of the policies in pollution control.

## METHODS

### Machine-learning models for simulating concentrations of water quality parameters

The collected variable data were screened using a Maximum Information Coefficient (MIC). Actually, the process was based on three steps: (1)  $MIC > 0.25$ ; (2) deletion of the prediction variables with collinearity (Spearman correlation analysis,  $R > 0.8$ ); (3) retainment of the prediction index that has a high correlation with a response index (Spearman correlation analysis,  $R > 0.4$ ). The general processes inherent in the models are depicted in Fig. 1. We utilized the model stacking method, which provided a composite prediction based on the results of multiple base models (that is, RF, SVM, and KNN)<sup>20</sup>. The model stacking algorithm uses a two-layered learning framework where the outputs generated by individual base models are input into another model to generate final predictions<sup>39,40</sup>. The learning process of the stacking model is categorized into three steps: stacking generation, stacking pruning, and stacking integration. The phase of stacking generation mainly refers to the generation of base models, whereas the last two steps optimally combine the base model predictions to form a final set of predictions using a second-level algorithm.

Different methods can be used to combine the base models, among which linear combination is the most widely used<sup>41–43</sup>. A linear stacking model has a prediction function expressed as:

$$y = \omega_1 f_1 + \omega_2 f_2 + \dots + \omega_M f_M \quad (1)$$

where  $y$  represents the stacking target;  $f_1, f_2, \dots, f_M$  denote the base model predictions from  $M$  individual algorithms ( $M=3$  in this study); and  $w_m$  ( $m=1, \dots, M$ ) is the weight assigned for each base model. The key problem with this approach lies in how to obtain the optimal set of weights (Fig. 1). A quadratic programming-based algorithm was adopted to estimate the set of weights using the R software package. We then assumed that the dataset which we intend to estimate the weights for  $N$  observations. First, a base model  $m_i$  is trained using the dataset with the  $i^{\text{th}}$  observation removed.  $\hat{f}_m^{-i}(x_i)$  represents the prediction of the model  $m$  for the  $i^{\text{th}}$  observation. The estimation of the weights is obtained from the least square linear regression of  $y_i$  (observed value of the  $i^{\text{th}}$  observation) on the linear combination of  $\hat{f}_m^{-i}(x_i)$ ,  $m=1, \dots, M$ . The optimal set of stacking weights are estimated by minimizing the following objective function under two constraints:

$$\hat{\omega}^{st} = \underset{\omega}{\operatorname{argmin}} \sum_{i=1}^N \left[ y_i - \sum_{m=1}^M \omega_m \hat{f}_m^{-i}(x_i) \right]^2 \quad (2)$$

$$\omega_m \geq 0, m = 1, 2, 3, \dots, M$$

$$\sum_{m=1}^M \omega_m = 1, m = 1, 2, 3, \dots, M$$

where  $\hat{\omega}^{st}$  is the objective function, and  $x_i$  refers to the  $i^{th}$  observation composed of all environmental variables. The two above constraints are reasonable if we interpret the weights as posterior model probabilities. It is worth noting that the  $i^{th}$  observation is removed from the training data when training model  $m$  to avoid assigning unfairly high weights to models with higher complexity<sup>44</sup>. More detailed information on the approach can be seen in the Supporting Information.

The predicting performance of the training and testing datasets provided complementary information for model validation. Training primarily exhibited model robustness, i.e., stability and balance of model predictability in the presence of data shuffling. Testing measures the model's performance on the unseen data and addresses the model fitness. In this context, we used the Pearson correlation coefficient ( $R^2$ ) as the statistical metric to quantify the predictive performance of the models (Supplementary Tables 1–5). To supplement the Pearson correlation coefficient and provide an in-depth assessment of model accuracy, we calculated the RMSE, NSE, and MAE. NSE estimates the correspondence between observed and predicted values<sup>45</sup>.

### Factors controlling TN, TP, NH<sub>3</sub>-N, and COD<sub>Mn</sub> variability

In this study, an MLR model was employed to assess the factors controlling nutrient levels in the water bodies. MLR models provide insights into the relationships between the response variables and multiple explanatory variables. The influence of each explanatory variable on the response variable is determined based on the ratio of the standardization coefficients of different explanatory variables to the sum of the absolute values of the total standardization coefficients<sup>46</sup>. The considered response and explanatory variables are shown in Supplementary Table 6. This approach has been widely applied to simulate water quality and to identify the key driving factors<sup>3,32</sup>. The MLR model applied in this study used a connection function to establish the relationship between the response variables (nutrient concentrations) and explanatory variables (environmental factors). Considering a response variable  $Y$  and  $p$  explanatory variables  $X_1, \dots, X_p$ , and  $n$  observations for MLR, that is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p-1} x_{i,p-1} + \beta_p x_{ip} + \varepsilon_i \quad (3)$$

where  $\varepsilon_i \sim N(0, \delta^2)$  for  $i = 1, \dots, n$ . After testing the collinearity of the predictive variables, this study used the annual mean data of the 613 sub-watersheds from 1980 to 2018 to input the data of 10 major river basins into the model separately. The ordinary least square method obtains the best function by minimizing the sum of squares of errors to estimate the standardization coefficient. The significance of standardized coefficients and fitting equations were tested using  $t$  tests and  $F$  tests, respectively<sup>46</sup>. The standardized coefficient ( $r$ ) between response variables and explanatory variables was used to compare the influence of each variable on nutrient variability. Influence, in this case, is presented as a contribution percentage of each variable as follows:

$$C_i = \frac{r_i}{|r_1| + |r_2| + \dots + |r_{p-1}| + |r_p|} \times 100\% \quad (4)$$

where  $C_i$  represents the contribution percentage of variable  $i$ ,  $i = 1, 2, 3, \dots, p$ , and  $r_i$  represents the standardized coefficient between response variables and the explanatory variable  $i$ .

### Data sources

In this study, four selected water quality parameters, including COD<sub>Mn</sub>, TN, NH<sub>3</sub>-N, and TP were selected to describe water quality in China. Monthly data were collected between 2003–2018 from 613 river water quality monitoring sites in the nation's 10 major river basins from the China National Environmental Monitoring Center (<http://www.mee.gov.cn/hjzl/shj/dbszdczb/>). The COD<sub>Mn</sub>, TN, NH<sub>3</sub>-N, and TP concentrations were analyzed in a laboratory

using the standard testing procedures recommended by the Ministry of Environmental Protection of China<sup>47</sup>, which did not change over the reported time period. The 10 river basins included the Songhua, Liaohe, Haihe, Yellow, Huaihe, Yangtze, Southeast, Pearl, Southwest, and Northwest Inland rivers. Spatial data were also collected from the basins, including the geographical conditions, physicochemical soil properties, climatic conditions, land use, anthropogenic discharges, and socio-economic development (Supplementary Table 6). The elevations and slopes of each water monitoring station were determined on the basis of a digital elevation model (resolution of  $1 \times 1$  km) from the Resource and Environment Science and Data Center (<http://www.resdc.cn/Default.aspx>). The local meteorological conditions (e.g., temperature, precipitation, wind speed, and extreme climate index) over the spatial domain covered by the national network were obtained from the CN05.1 dataset, which was obtained from the China Meteorological Administration and constructed by the "anomaly approach". Data interpolation between sites was based on many station observations (–2400) in China<sup>48,49</sup>. The dataset has a spatial resolution of  $0.25^\circ \times 0.25^\circ$ . The soil properties of each sub-basin were extracted from a digital soil properties map obtained from the Institute of Soil Science, Chinese Academy of Sciences (<http://www.issas.cas.cn/>). The scale of the utilized digital soil properties map is 1:1,000,000. The land-use dataset from China (with a resolution of  $30 \times 30$  m) was obtained from the Institute of Geographic Sciences and Natural Resources Research (IGSNRR, Chinese Academy of Sciences) (<http://www.resdc.cn/Default.aspx>). The net anthropogenic N and phosphorus (P) inputs (NANI and NAPI) were estimated based on the reported discharge activity data and discharge coefficient<sup>50,51</sup>. The discharge activity data were obtained from the Chinese Statistical Yearbook (<https://data.cnki.net/Yearbook/Navit?type=type&code=A>). Gross domestic product (GDP) and population density (POP) represent two important social economic indicators that may affect pollution sources and the input of pollutants to water bodies<sup>52</sup>. The spatial distribution of social economic data (resolution of  $1 \times 1$  km) was obtained from the Resource and Environment Science and Data Center in the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (<https://www.resdc.cn/>).

### DATA AVAILABILITY

The historical GDP and POP data were obtained from the China Statistical Yearbooks (1980–2018) (<http://www.stats.gov.cn/english/Statisticaldata/>). Annual land use data for the COAC at a resolution of  $30 \times 30$  m was obtained from National Earth System Science Data Center for a period of 39-years (1980–2018), which integrate AVHRR, MODIS and Landsat data using the BFAST algorithm. Water quality data were obtained from the China National Environmental Monitoring Center (<https://szdjcnemc.cn:8070/GJZ/Business/Publish/Main.html>). All data are also immediately available from the corresponding author upon reasonable request. The sites of National Earth System Science Data Center of China and the China National Environmental Monitoring Center can be translated into English by the Google Translate plugin in Chrome or the Google Translate site for website translation (<https://translate.google.com/?sl=zh-CN&tl=en&op=websites>).

Received: 6 February 2023; Accepted: 23 May 2023;

Published online: 06 June 2023

### REFERENCES

1. Grill, G. et al. Mapping the world's free-flowing rivers. *Nature* **569**, 215–221 (2019).
2. Palmer, M. & Ruhí, A. Linkages between flow regime, biota, and ecosystem processes: implications for river restoration. *Science* **365**, eaaw2087 (2019).
3. Sinha, E., Michalak, A. & Balaji, V. Eutrophication will increase during the 21st century as a result of precipitation changes. *Science* **357**, 405–408 (2017).
4. Stokral, M. et al. Global multi-pollutant modelling of water quality: scientific challenges and future directions. *Curr. Opin. Environ. Sustain.* **36**, 116–125 (2019).



5. Vorosmarty, C. et al. Global threats to human water security and river biodiversity. *Nature* **467**, 555–561 (2010).
6. United Nations. Sustainable development goals. <https://www.un.org/sustainabledevelopment/water-and-sanitation/> (2020)
7. Lu, Y. et al. Forty years of reform and opening up: China's progress toward a sustainable path. *Sci. Adv.* **5**, eaau9413 (2019).
8. Tao, T. & Xin, K. A sustainable plan for China's drinking water: tackling pollution and using different grades of water for different tasks is more efficient than making all water potable. *Nature* **511**, 527–529 (2014).
9. Chen, X. et al. Multi-scale modeling of nutrient pollution in the rivers of China. *Environ. Sci. Technol.* **53**, 9614–9625 (2019).
10. Wang, M. et al. Accounting for interactions between Sustainable Development Goals is essential for water pollution control in China. *Nat. Commun.* **13**, 730 (2022).
11. Huang, J. et al. How successful are the restoration efforts of China's lakes and reservoirs? *Environ. Int.* **123**, 96–103 (2019).
12. Ma, T. et al. China's improving inland surface water quality since 2003. *Sci. Adv.* **6**, eaau3798 (2020).
13. The Ministry of Ecology and Environment, P.R.C. Annual Bulletin of China's Ecology and Environment. <https://www.mee.gov.cn/hjzl/shj/qgdbzslzk/> (2022).
14. Xia, X. et al. The cycle of nitrogen in river systems: sources, transformation, and flux. *Environ. Sci. -Proc. Imp.* **20**, 863–891 (2018).
15. Wang, S. et al. Reduced sediment transport in the yellow river due to anthropogenic changes. *Nat. Geosci.* **9**, 38–42 (2016).
16. Huang, F., Wang, X., Lou, L., Zhou, Z. & Wu, J. Spatial variation and source apportionment of water pollution in Qiantang River (China) using statistical techniques. *Water Res.* **44**, 1562–1572 (2010).
17. Huang, J. et al. Characterizing the river water quality in China: recent progress and on-going challenges. *Water Res.* **201**, 117309 (2021).
18. Zhou, Y. et al. Improving water quality in China: environmental investment pays dividends. *Water Res.* **118**, 152–159 (2017).
19. Ma, C. et al. Improving the estimation of nitrogen and phosphorus concentrations in lakes and reservoirs using a stacked approach. *Earths Future* **11**, e2022EF003013 (2023).
20. Chen, K. et al. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data. *Water Res.* **171**, 115454 (2020).
21. Yu, C. et al. Managing nitrogen to restore water quality in China. *Nature* **567**, 516–520 (2019).
22. Tong, Y. et al. Decline in Chinese lake phosphorus concentration accompanied by shift in sources since 2006. *Nat. Geosci.* **10**, 507–511 (2017).
23. He, G. Z., Zhang, L., Mol, A. P. J., Lu, Y. & Liu, J. Revising China's environmental law. *Science* **341**, 133 (2013).
24. Zhang, B. Five-year plan: supervise Chinese environment policy. *Nature* **534**, 179 (2015).
25. Ongley, E. D., Zhang, X. & Yu, T. Current status of agricultural and rural non-point source pollution assessment in China. *Environ. Pollut.* **158**, 1159–1168 (2010).
26. Shen, Z., Liao, Q., Hong, Q. & Gong, Y. An overview of research on agricultural nonpoint source pollution modelling in China. *Sep. Purif. Technol.* **84**, 104–111 (2012).
27. Huo, S. et al. Spatiotemporal differences in riverine nitrogen and phosphorus fluxes and associated drivers across China from 1980 to 2018. *Chemosphere* **310**, 136827 (2023).
28. Lapierre, J. F., Guillemette, F., Berggren, M. & Giorgio, P. D. Increases in terrestrially derived carbon stimulate organic carbon processing and CO<sub>2</sub> emissions in boreal aquatic ecosystems. *Nat. Commun.* **4**, 2972 (2013).
29. Creed, I. F. et al. Global change-driven effects on dissolved organic matter composition: implications for food webs of northern lakes. *Glob. Change Biol.* **24**, 3692–3714 (2018).
30. Zhou, J., Leavitt, P. R., Zhang, Y. & Qin, B. Anthropogenic eutrophication of shallow lakes: is it occasional. *Water Res.* **221**, 118728 (2022).
31. Maavara, T. et al. Global phosphorus retention by river damming. *Proc. Natl Acad. Sci.* **112**, 15603–15608 (2015).
32. Isles, P. D. F., Xu, Y., Stockwell, J. D. & Schroth, A. W. Climate-driven changes in energy and mass inputs systematically alter nutrient concentration and stoichiometry in deep and shallow regions of Lake Champlain. *Biogeochemistry* **133**, 201–217 (2017).
33. van Vliet, M. et al. Global river discharge and water temperature under climate change. *Glob. Environ. Change* **23**, 450–464 (2013).
34. Yang, C., Yang, P., Geng, J., Yin, H. & Chen, K. Sediment internal nutrient loading in the most polluted area of a shallow eutrophic lake (Lake Chaohu, China) and its contribution to lake eutrophication. *Environ. Pollut.* **262**, 114292 (2020).
35. Raymond, C. et al. Understanding and managing connected extreme events. *Nat. Clim. Chang* **10**, 611–621 (2020).
36. Green, M. B., Nieber, J. L., Johnson, G., Magner, J. & Schaefer, B. Flow path influence on an N:P ratio in two headwater streams: a paired watershed study. *J. Geophys. Res. Biogeosci.* **112**, G03015 (2007).
37. Stockwell, J. et al. Storm impacts on phytoplankton community dynamics in lakes. *Glob. Change Biol.* **26**, 2756–2784 (2020).
38. The Ministry of Ecology and Environment, P.R.C. Second National Pollutant Source Census Bulletin. [http://www.mee.gov.cn/xgk/xgk01/202006/t20200610\\_783547.html](http://www.mee.gov.cn/xgk/xgk01/202006/t20200610_783547.html). (2020)
39. Zhai, B. & Chen, J. Development of a stacked ensemble model for forecasting and analyzing daily average pm 2.5 concentrations in Beijing, China. *Sci. Total Environ.* **635**, 644–658 (2018).
40. Rice, J. S. & Emanuel, R. E. How are streamflow responses to the el nino southern oscillation affected by watershed characteristics? *Water Resour. Res.* **53**, 4393–4406 (2017).
41. Li, W. et al. Concentration estimation of dissolved oxygen in Pearl River Basin using input variable selection and machine learning techniques. *Sci. Total Environ.* **731**, 139099 (2020).
42. Shen, L. Q., Amatulli, G., Sethi, T., Raymond, P. & Domisch, S. Estimating nitrogen and phosphorus concentrations in streams and rivers, within a machine learning framework. *Sci. Data* **7**, 161 (2020).
43. Visser, H. et al. What drives the ecological quality of surface waters? A review of 11 predictive modeling tools. *Water Res* **208**, 117851 (2022).
44. Friedman, J., Hastie, T., Tibshirani, R. The Elements of Statistical Learning. vol. 1. Springer series in statistics, New York, NY, USA. <https://doi.org/10.1007/978-0-387-84858-7> (2001).
45. Jiang, S., Zheng, Y. & Solomatine, D. Improving AI system awareness of geoscience knowledge: symbiotic integration of physical approaches and deep learning. *Geophys. Res. Lett.* **47**, 1–11 (2020).
46. Korkmaz, M. A study over the general formula of regression sum of squares in multiple linear regression. *Numer. Methods Part. D. E.* **37**, 406–421 (2020).
47. PCR MEP. Water and wastewater monitoring and analysis methods. China Environmental Science Press, Beijing (2002).
48. Xu, Y. et al. A daily temperature dataset over China and its application in validating a RCM simulation. *Adv. Atmos. Sci.* **26**, 763–772 (2009).
49. Wu, J. & Gao, X. A gridded daily observation dataset over China region and comparison with the other datasets. *Chin. J. Geophys.* **56**, 1102–1111 (2013).
50. Goyette, J. O., Bennett, E. M., Howarth, R. W. & Maranger, R. Changes in anthropogenic nitrogen and phosphorus inputs to the St. Lawrence sub-basin over 110 years and impacts on riverine export. *Glob. Biogeochem. Cycles* **30**, 1000–1014 (2016).
51. Hu, M. et al. Long-term (1980–2015) changes in net anthropogenic phosphorus inputs and riverine phosphorus export in the Yangtze River basin. *Water Res.* **177**, 115779 (2020).
52. Xia, X. et al. Analyzing the contribution of climate change to long-term variations in sediment nitrogen sources for reservoirs/lakes. *Sci. Total Environ.* **523**, 64–73 (2015).

## ACKNOWLEDGEMENTS

The National Natural Science Foundation of China (No. 52225903, U2243209) and the National Key Research and Development Program of China (2022YFC3201900) supported this study.

## AUTHOR CONTRIBUTIONS

H.Z. and X.C. collected and analyzed data in addition to drafting the manuscript. S.H. conceived the idea and study design. C.M., Y.L., Y.T., and F.W. revised the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41545-023-00260-y>.

**Correspondence** and requests for materials should be addressed to Shouliang Huo.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



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

© The Author(s) 2023