Spatiotemporal Features of Pollutant Loads in the Yan River Basin, a Typical Loess hilly and Gully Watershed in the Chinese Loess Plateau

Yanni Song
Xi'an Jiaotong University
Yiping Wu (rocky.ypwu@gmail.com)
Xi'an Jiaotong University https://orcid.org/0000-0002-5163-0884
Changshun Sun
Shaanxi Provincial Academy of Environmental Sciences
Fubo Zhao
Xi'an Jiaotong University
Jingyi Hu
Xi'an Jiaotong University
Ji Chen
University of Hong Kong
Linjing Qiu
Xi'an Jiaotong University

Research Letter

Keywords: COD, Nutrient loads, LOADEST, NPS pollution, SWAT

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-502118/v1

Version of Record: A version of this preprint was published at Geoscience Letters on February 19th, 2022.
See the published version at https://doi.org/10.1186/s40562-022-00220-3.

Spatiotemporal features of pollutant loads in the Yan River Basin, a typical loess hilly and gully watershed in the Chinese Loess Plateau

Yanni Song[a], Yiping Wu[a,*], Changshun Sun[b], Fubo Zhao[a], Jingyi Hu[a], Ji Chen[c], Linjing Qiu[a]

[a] Department of Earth & Environmental Science, Xi'an Jiaotong University, Xi'an, Shaanxi Province, 710049, China

[b] Shaanxi Provincial Academy of Environmental Sciences, Xi'an, 710061, China

[c] Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China

*Email: rocky.ypwu@gmail.com

Abstract

Water quality is the restrictive factor for both ecosystem health and social development in the Chinese Loess Plateau, a unique area with most severe soil erosion, fragile ecology, and water shortage. Understanding the characteristics of the pollutant loads is of vital importance for the sustainability of eco-environment in the Loess Plateau. This study investigated the spatiotemporal changes of chemical oxygen demand (COD), total nitrogen (TN), and total phosphorus (TP) loads by combining the Soil and Water Assessment Tool (SWAT) and regression model Load Estimator (LOADEST) in a typical loess hilly and gully watershed—the Yan River Basin (YanRB). Results showed that the model simulations of monthly streamflow and pollutant loads were in good agreement with those derived from the in-situ observations. The temporal variation analysis suggested that the pollutant loads were generally rising in the study period (2001–2018) at four of the five stations and reached the maximum in 2014, and the multi-year (i.e., 2001–2018 with 2013 being excluded due to extreme rainfall) average loads of COD, TN, and TP at the Tanjiahe station, which is close to the outlet of the basin, were 15,021 kg/d, 3,835 kg/d, and 168 kg/d, respectively. The spatial distribution of the TN and TP loads along the river seemed to be quite unique because the TP level were obviously higher at the midstream (e.g., Zhujiagou and Ganguyi) than the downstream (e.g., Tanjiahe), and the TN level decreased when the river flowed from Zhujiagou to Ganguyi. Further, the seasonal analysis indicated that the nutrient loads were the highest in summer, followed by autumn, and the loads in these two wet seasons contributed the most of the annual pollution loads—about 76%

and 84% for TN and TP, respectively, indicating the higher flow, the higher pollution load, a similar point based on the inter-annual analysis. In addition, the contribution analysis of point source and non-point source pollutions demonstrated that NPS led to most of the pollutant loads at the whole watershed—87%, 85%, and 84% of the COD, TN, and TP loads, respectively. Overall, this study provided spatiotemporal distributions of the key pollutant loads in the YanRB and can be valuable for water quality protection and pollution control in this area.

Keywords: COD; Nutrient loads; LOADEST; NPS pollution; SWAT

1 Introduction

Pollutant discharge causes widespread organic pollution and eutrophication, along with severe ecological destruction, and further threats human health and agricultural production (Dai et al., 2017; Lang et al., 2013; Schwarzenbach et al., 2010; Wang and Yang, 2016). The Chinese government has realized the importance of controlling water pollution, and published Action Plan for Prevention and Control of Water Pollution to protect the surface water and groundwater in April 2015 (Shi et al., 2015). However, making efficient water quality management measures is difficult because of multiple anthropogenic and natural influencing factors. According to the statistical data, the increase rate of wastewater and sewage discharge is 1.8 billion m^3 per year, and the daily discharge of industrial and domestic wastewater in China is about 164 million m^3 (Zhou and Li, 2018). Besides, Zou et al. (2020) used an inventory analysis to estimate the agricultural non-point source (NPS) pollution loads from 1978 to 2017 in China, and found that the pollution loads of chemical oxygen demand (COD), total nitrogen (TN) and total phosphorus (TP) increased by 91.0%, 196.2%, and 244.1%, respectively. Therefore, identifying the current pollution situation and understanding its relationships with human activities are important for water resources protection and pollution control.

Hydrometeorological factors (e.g., precipitation, runoff, and temperature), anthropogenic factors (e.g., land use, agricultural practices), and geographical factors (e.g., soil, vegetation, slope gradient) can largely affect the formation and transformation of pollutants. There were a number of studies assessing the water quality issues in recent years, especially about nutrients (e.g., TN, TP) and sediment loads,

75 urbanization, and reservoir operation (Oelsner and Stets, 2019; Tong et al., 2015; Xia
76 et al., 2020). Numerical water quality models are useful tools for evaluating pollution
77 loads at the watershed scale, and these models can be divided into classical empirical
78 models (e.g., RUSLE, SEDD, and PLOAD), statistical models (e.g., SPARROW,
79 WRTDS, and GAMMKS), physically-based models (e.g., SWAT, ANSWERS, and
80 AnnAGNPS), and intelligent data analysis methods such as Hybrid Double
81 Feedforward Neural Network and Fuzzy Binary Comparison methods (Lin et al., 2016;
82 Lin et al., 2018; Zou et al., 2020). The Soil and Water Assessment Tool (SWAT) model
83 has been widely used to simulate the sediment and nutrient load transport through
84 surface runoff, lateral flow, and groundwater flow (Chen et al., 2012). However, the
85 applications of SWAT in some regions are often disabled due to the data shortage. Jiang
86 et al. (2019) proposed the Hydro-Informatic Modelling System−pollution load (HIMS-
87 PL) model to simulate the processes of pollutant generation and transport, and
88 quantified the pollution load in semiarid and semi-humid areas. However, this model
89 needs intensive input data including the spatial attribute information of study area,
90 meteorology, streamflow, point source (PS) pollutant loads, and cross-section pollutant
91 concentrations. The Load Estimator (LOADEST) model, which has been widely used
92 for quantitative assessment of pollution loads in those ungauged watersheds or those
93 with limited gaging watershed, has been proven to be an efficient tool in water quality
94 assessment (Chen et al., 2015a; Chen et al., 2017; Duan et al., 2013).
95 Pollutants can be classified as PS and NPS pollution. PS pollution is relatively easy
96 to manage and control because it is easy to be monitored, but it is quite difficult to

quantify NPS pollution load and determine its pollutions sources because NPS is inseparable from precipitation, snowmelt, runoff, and soil erosion processes. Some studies have proposed methods for estimating NPS pollution loads (Hao et al., 2006; Li 2000), mainly including multivariate statistical analysis, dual stable isotopes, hydrological and water quality watershed models, geographical information systems (GIS) technology, and hydro-chemical characteristic analysis (Wei et al., 2020). In these methods, the statistical method has been widely used to analyze the relationships between flow and concentration (Li et al., 2011; Wang et al., 2015a).

The organic pollution, excessive nutrients (e.g., nitrogen and phosphorus), and heavy metal pollution are important components of water pollution (Zeng et al., 2021). In particular, the rapid economic development in Northwest China has led to substantial imbalances between the natural environment and industrial development. The Loess Plateau, mainly located in the middle reach of the Yellow River, is considered as one of the regions with the most severely eroded areas in the world, and the soil loss has significantly impacted nutrient losses (Zhao et al., 2020). Further, the Yellow River is facing the problem of water quality deterioration. It was reported that the water quality in almost 34.8% of main tributaries (e.g., the Qingjian, the Fen, the Shiwang, and the Yan Rivers) were generally below Class III standard (GB 3838-2002), and the main pollutants were COD, ammonia nitrogen, and TP. The Yan River, located in the Loess Plateau, is a main tributary in the middle Yellow River. Most existing studies focused on the land use, ecological evaluation, soil erosion, and hydrological processes (Chen et al., 2020; Su et al., 2011b; Su et al., 2012; Wang and Zhao, 2016; Zhou and Li, 2015),

but few studies were conducted to evaluate the spatiotemporal characteristics of pollutant loads of the YanRB due to its difficulty and data unavailability, though water quality degradation induced by organic matter and nutrients have become a main threat to the local eco-environment.

The primary target of this study was to investigate the spatiotemporal patterns of pollutants and identify highly polluted areas, and the specific objectives were to: 1) estimate the loads of the key pollutants (COD, TN and TP) at the main cross-sections of the river, 2) quantify the contributions of PS and NPS pollution to the total pollution load, 3) investigate the transport of pollutants along river under different hydrological conditions.

2 Materials and Method

2.1 Study area

The Yan River originates from Baiyu Mountain and flows from northwest to southeast through Zhidan, Ansai, Yan'an, and Yanchang counties before entering the Yellow River. The length of the river is 286.9 km, and the drainage area is 7,725 km^2 (Wu et al., 2020) (Fig. 1). The annual average air temperature of the basin ranged from 7.9°C in the west to 10.6°C in the east. The annual average precipitation is between 340 and 660 mm across the basin. Besides, the short-duration and high-intensity precipitation led to about 60.5% of the runoff occurring in the wet season, especially in July through September, and caused intensive soil erosion (Miao, 2018). In the YanRB, there are three dominate land use types—grassland, cropland, and woodland, accounting for 55.1%, 17.64%, and 26.16% of the basin, respectively.

Energy (e.g., oil and coal) is the dominant industry of the YanRB and has exerted a profound influence on the economic development. Major agricultural products in this area include apple, greenhouse vegetables, Oryza sativa, and strawberries (Su et al., 2011a), and the fertilization is quite intensive, leading to relatively higher nutrient loss to the river. We chose five cross-sections—Ansai in the upstream, Shiyaocun, Zhujiagou, and Ganguyi in the middle stream, and Tanjiahe in the downstream (Fig. 1) —to analyze the features of pollutant loads along the main river.

2.2 Model description

2.2.1 SWAT

The Soil and Water Assessment Tool (SWAT) model, developed by the Agricultural Research Service of the United States Department of Agriculture (USDA-ARS), is a comprehensive hydrological and water quality model (Arnold et al., 1998; Neitsch et al., 2011). SWAT takes spatially variable soil and land cover conditions into consideration to delineate the watershed. The Hydrological Response Units (HRUs) are the basic and least computing units that possess unique land use, soil property, and slope. The main outputs of SWAT are surface runoff, lateral flow, baseflow, evapotranspiration, water yield, sediment load, and nutrient loads, etc.

The SWAT model was designed for long-term continuous simulations at daily time step for watershed management decisions (Sudheer et al., 2007). In this study, we used the Sequential Uncertainty Fitting version 2 (SUFI-2) algorithm to optimize the parameters, and the calibrated/validated SWAT was used to obtain the long-term streamflow, which were then used for subsequent pollutant load estimation with

LOADEST (see the following section).

2.2.2 LOADEST

The Load Estimator (LOADEST) is an empirical statistical model, developed by the United States Geological Survey (USGS) and can estimate the long-term pollutant loads using the streamflow and pollutant concentration regression relationships (Sharifi et al., 2017). The setting of the model requires at least 12 continuous flows and concentration data (Park and Engel, 2016). Considering the different diffusion modes of pollutants and the adaptability of pollutant loads to river dynamics, 11 regression equations are available to users in the LOADEST. The model uses Tobit regression to process the discrete water quality and streamflow data at the monitoring point and automatically selects the predefined model through the lowest Akaike Information Criterion (AIC) and the Schwarz Posterior Probability Criterion (SPPC) (Runkel et al., 2004).

The model parameter estimation is established based on three statistical algorithms, including the adjusted maximum likelihood estimation (AMLE), the minimum variance unbiased estimate (MVUE), and the least absolute deviation (LAD) method. This study selected the AMLE estimation method, which can eliminate the influence of censored data as much as possible and has been proven to have a high simulation accuracy (Aulenbach, 2013; Duan et al., 2014). The specific principle of the AMLE is described as follows (Cohn et al., 1992):

$$\hat{L}_{AMLE} = exp\left(a_0 + \sum_{j=1}^{M} a_j X_j \right) H(a, b, s^2, \alpha, \kappa)$$

where $\hat{L}_{AMLE}$ is the AMLE estimate of instantaneous load, a and b are functions of the explanatory variables, κ and α are parameters of the gamma distribution, s^2 is the

185 residual variance, a_0 and a_j are maximum likelihood estimates corrected for first-order
186 bias, the bias correction factor [$H(a,b,s2,\alpha,\kappa)$] is an approximation of the infinite series.
187 2.3 Data
188 The main input data for SWAT include digital elevation model (DEM), soil, land use,
189 and meteorological data. The DEM data with a 30m resolution was obtained from the
190 National Geomatics Center (http://www.ngcc.cn/ngcc/). Land use data in 2010 with
191 30m resolution were provided by the Institute of Remote Sensing and Digital Earth,
192 Chinese Academy of Sciences. Meteorological data (1951–2018) was collected from
193 the Data Center of the China Meteorological Administration (http://data.cma.cn/),
194 including precipitation, relative humidity, wind speed, maximum, and minimum air
195 temperature. The solar radiation was estimated based on sunshine duration data, and the
196 calculation details are available in our previous study (Zhang et al., 2019b). The
197 observed monthly runoff data (2001–2008) at the Ganguyi flow gaging station, with a
198 drainage area of 5,891 km^2, were obtained from the Yellow River Water Resources
199 Commission.
200 In this study, the in-situ observed runoff data were only available for Ansai, Yan'an,
201 Ganguyi flow gaging stations, covering the period of 2015-2018, and the runoff data
202 for the other periods (2001–2014) were from SWAT simulations. Water quality data
203 (2017–2018) including chemical demand oxygen (COD), total nitrogen (TN), and total
204 phosphorus (TP) concentration at five monitoring stations were obtained from the water
205 quality monitoring station, with TN missing from January through May in 2017.
206 2.4 Model setup and verification
207 The ArcSWAT (version 2012) was used to prepare the model input for driving SWAT

208 to simulate the streamflow of the YanRB. Based on the combination of land use, soil,
209 and slope, the YanRB was discretized into 99 subbasins and 686 HRUs. We used the
210 SWAT-CUP (SWAT Calibration and Uncertainty Programs) program to optimize the
211 model parameters. The correlation coefficient (R^2), Nash efficiency coefficient (NSE),
212 and percent bias (PBIAS) were selected to quantify the fitness between the
213 measurements and simulations. Theoretically, model simulation is rated satisfactory
214 when $R^2 \geq 0.5$, NSE$\geq$0.5, -25%$\leq$PBIAS$\leq$25% according to Moriasi et al. (2007) (see
215 Appendix).

216 In this study, we used flow and water quality data (i.e., COD, TN, and TP) as input
217 for the LOADEST model. Then, we set up a few key variables of interest (e.g., observed
218 date, time, streamflow, and concentration) for LOADEST to generate the long-term
219 series of each key variable based on the optimal equations. Finally, the fitness of the
220 best candidate equation was validated by comparing observed and simulated loads. A
221 few statistical terms, including R^2, NSE, PBIAS, and the probability plot correlation
222 coefficient (PPCC), were used to evaluate the model performance when selecting the
223 final equation for estimating constituent loads.

224 2.5 Quantification of PS and NPS pollutant loads

225 2.5.1 Digital filtering method

226 Baseflow is an integral portion of streamflow that comes from groundwater or other
227 delayed sources (Hall, 1968; Zhang et al., 2017; Zhu et al., 2019) and recognized as an
228 important hydrological pathway for nutrient export (Drewry et al., 2006; He and Lu,
229 2016; Wang et al., 2015b), especially in arid and semi-arid areas (Yang et al., 2020).

Many methods have been used to estimate baseflow, such as graphical methods, statistical methods, and chemical methods (Miller et al., 2015; Zhang et al., 2019a). There were a variety of factors, such as climate and geological conditions, that may affect baseflow (Segura et al., 2019). Dou and Huang (2010) found that the digital filtering method was suitable for the baseflow segmentation of rivers in the Loess Plateau. For our study area, the surface runoff and the pollutant concentration are greatly influenced by external factors such as rainfall intensity and human activities. Thus, the signal of pollutants carried by surface runoff is strong and can be regarded as a high-frequency signal. In comparison, due to the regulation and storage of the precipitation, the baseflow and the concentration are relatively stable in the dry season, which can be regarded as a low-frequency signal (Eckhardt, 2005). Therefore, the baseflow calculation in this study can be expressed as (Lyne and Hollick, 1979):

$$q_i = \beta q_{i-1} + \frac{1+\beta}{2}[Q_i - Q_{i-1}] \quad (1)$$

where q_i is the filtering flow on the i^{th} day, Q_i is the total flow on the i^{th} day, and β is the filter parameters with a reference value of 0.925.

2.5.2 Pollutant separation method

As a statistical method, the hydrological estimate has received much attention, especially in areas where the annual rainfall distribution varies greatly. This method is based on the assumptions that the NPS pollution load mainly comes from the rainfall-runoff process, and the pollutant load in the dry season is predominantly contributed by the PS. Based on the segmentation of baseflow, we assumed that the discharge of PS pollutant is relatively stable and can be derived from the baseflow in dry season

252 (December to March). The calculation for the total pollution load is shown in Eq. (2)
253 (Xin et al., 2017):

254 $$W_t = \sum_{i=1}^{n} C_p Q_p \Delta t + \sum_{i=1}^{n} C_{np} Q_{np} \Delta t \qquad (2)$$

255 where W_t is the total pollution loads in the river channel; C_p is the PS pollutant
256 concentration in the i^{th} month, represented by the average concentration of the water
257 quality data during the dry season; Q_p is the baseflow; C_{np} is the NPS pollutant
258 concentration in the i^{th} month; Q_{np} is the surface flow, n is the number of months, and
259 Δt is the time in the i^{th} month.

260 $$W_t = \sum_{i=1}^{n} CQ\Delta t \qquad (3)$$

261 where C is observed pollutant concentration in the i^{th} month, and Q is the average flow
262 in the i^{th} month. The total pollution load minus the point source pollution load equals
263 the NPS pollution load.

264 **3. Results**

265 3.1 Model performance evaluation

266 3.1.1 Hydrological modeling

267 Based on the sensitivity analysis, we identified eight parameters and derived the
268 optimal values using SUFI2 (Table 1). Fig. 2 showed the visual comparison of
269 simulated monthly streamflow against the observations at the Ganguyi flow gaging
270 station. Although there was an underestimation of streamflow during the wet season,
271 the overall performance was satisfactory during both calibration (2001–2004) and
272 validation (2005–2008) periods, with R^2 and NSE ranging from 0.50 to 0.67 and PBIAS
273 being less than 15%. According to the evaluation standard by Moriasi et al. (2007), the

model performance in streamflow simulation can be rated as satisfactory.

3.1.2 Water quality model evaluation

The LOADEST simulated and observed COD, TN, and TP loads with the 95% confidence interval at Ganguyi water quality monitoring station during 2017–2018 were shown in Fig. 3a, c, and e. Although the model underestimated the peak values in July, it can still capture the temporal patterns of the three pollutants. To further evaluate the model performance, the scatter plots (Fig. 3b, d, and f) showed that the R^2 was above 0.69, NSE was above 0.56, and |PBIAS| was lower than 25% for the COD, TN and TP loads. Besides, we found that the estimation of TN and TP loads had a better performance than COD. For five water quality monitoring stations, the statistical evaluation terms showed that the modeling loads of COD, TN, and TP had good performances, with the R^2 ranging from 0.52 to 0.94, the NSE ranging from 0.49 to 0.90, the PBIAS ranging from -6.9 to -24.1%, and the PPCC ranging from 0.88 to 0.99 (Fig. 1). Overall, the model can accurately simulate the pollutant loads in the YanRB.

3.2 Spatiotemporal distribution of pollutant loads

3.2.1 Annual pollution loads in the YanRB

Fig. 4 demonstrated the monthly average loads of COD, TN, and TP at five stations from 2001 to 2018, excluding 2013. The pollutant loads fluctuated greatly within a year because precipitation dominated the pollutant delivery. For example, the COD load on January 2018 was 5,241 kg/d and 27,336 kg/d for July in 2018. We also found both TN and TP load at Ganguyi, a station in the middle reach, exceeded the final cross-section (i.e., Tanjiahe) during 2017–2018. From Fig. S1, we found that the annual average loads

296 of COD, TN and TP were relatively stable before 2009, and there was a slight upward
297 trend for COD and TP loads at four of the five stations (i.e., Ansai, Shiyaocun, Ganguyi,
298 and Tanjiahe) during the 17-year study period, though the trend was not statistically
299 significant. Further, there was a decreasing trend in annual average pollutant loads at
300 the Zhujiagou station. Specifically, the annual average COD load changed from 9,815
301 kg/d in 2001 to 5,213 kg/d in 2018 (about 47% decrease), annual average TN load
302 decreased from 3,002 kg/d in 2001 to 1,852 kg/d in 2018 (about 38% decrease), and
303 annual average TP load declined from 88 kg/d in 2001 to 50 kg/d in 2018 (about 43%
304 decrease). In addition, the temporal variation showed that the pollutant loads were
305 generally rising and reached the maximum in 2014, and the trend was statistically
306 significant at four of the five stations (i.e., Ansai, Shiyaocun, Ganguyi, and Tanjiahe
307 station).
308 3.2.2 Monthly pollution loads in the YanRB
309 Fig. 5 showed the multi-year average loads of COD, TN, and TP in each calendar
310 month at five stations from 2001 to 2018, excluding 2013. It can be seen that the
311 monthly distribution of pollutant loads is very uneven, and the maximum value
312 occurred in July. The spatial distribution of the COD loads increased from upstream to
313 downstream of the river, reflecting that the transportation and confluence of COD
314 pollutant with streamflow. The TN loads decreased when river flowed from Zhujiagou
315 to Ganguyi station, and this was mainly caused by relatively more intensive agricultural
316 practices such as fertilization and flooding irrigation. The spatial feature of the TP loads
317 along the river seemed to be quite unique because the TP level was obviously higher at

318 the Zhujiagou and Ganguyi station in the midstream than that at the Tanjiahe in the
319 downstream.
320 3.2.3 Seasonal pollution loads in the YanRB
321 Estimated seasonal loads of TN and TP at five water quality monitoring stations were
322 highly variable in the YanRB. The result showed that the greatest loads occurred in
323 summer due to high streamflow, followed by autumn, and the loads in these two wet
324 seasons contributed the most of the annual pollution load—about 76% and 84% for TN
325 and TP, respectively, reflecting the effect of seasonal runoff patterns and agricultural
326 practices on pollution loads (Fig. 6). The multi-year average proportion of TP loads in
327 summer was 52%, a little higher than that of TN loads (49%). Compared to Ganguyi,
328 the proportion of the TN loads in summer at Zhujiagou were significantly higher, which
329 might be attributed to relatively larger farmland areas around Zhujiagou (see Fig. 1).
330 Therefore, it should be noted that the YanRB might face the challenge of eutrophication
331 in summer and the nutrient losses due to the intensive agricultural practices in autumn.
332 3.3 Contributions of PS and NPS pollution loads
333 The different water quality data sources, hydrological separation methods, and
334 baseflow period assumptions caused largely uncertain for the investigation of NPS
335 pathways (Zhu et al., 2019). The separation of baseflow using digital filtering method
336 and simulated daily streamflow data derived from SWAT model at five stations showed
337 that the baseflow accounted for 35–45% of streamflow. Fig. 7 showed the contributions
338 of NPS pollution load to the total pollution loads. The percentage of NPS contributions
339 ranged from 67% to 94%, indicating that the NPS pollution was the dominant factor of

the total pollution loads in the YanRB. Our spatial analysis showed that the percentage of the NPS pollution for the Zhujiagou station was lower than other stations. We also found that the NPS contribution percentage decreased gradually for pollutant loads, especially for TP loads, which decreased by 19% and 14% at the Zhujiagou and Tanjiahe station, respectively, in 2018. In contrast to other pollutants, the contribution of PS to TP was higher, and this is probably because NPS phosphorus was removed in some degree due to adsorption with sediment and accompanied deposition in lands or channels. Besides, PS phosphorus load may become higher due to increased domestic wastewater caused by population growth and urbanization.

4 Discussion

4.1 Water pollution features

In the YanRB, the PS pollution mainly comes from industrial and domestic wastewater, whereas NPS pollution mainly comes from agriculture, livestock, and poultry breeding, and large-scale oil and coal mining. From Fig. 4, we found that this watershed was more seriously polluted by the organic pollutants, which probably mainly come from oil and coal mining, livestock, and poultry breeding. The spatial distribution of TP loads along the river showed that the maximum was at Ganguyi, a station in the middle reach, followed by a sharp decrease at Tanjiahe (a station closes to the estuary). One reason is that the slope of the river channel in the downstream area (after Ganguyi) became less, and the river velocity became lower with reduced hydraulic condition, causing the deposition of phosphorus with sediment. Another reason can be the check dams in gullies/channels and few tributaries in the downstream,

362 leading to the intercept and adsorption of phosphorus during its transport in channels
363 (Guo et al., 2016; Sun et al., 2019). Overall, the spatial distribution of nutrients
364 exhibited a non-linear change from upstream to downstream, and this phenomenon can
365 be attributed to various cumulative effects (USEPA, 2015) such as different
366 hydrodynamic conditions, population density, topography, and the intensive cultivation
367 in some specific areas. Moreover, the meteorological, hydrological, and underlying
368 surface conditions are also important influencing factors in the spatiotemporal
369 characteristics of pollution in the loess hilly and gully. Wu et al. (2016) found that the
370 distributions of NPS pollution load and sediment yield are closely related to land use
371 types and soil erosion of the Majiagou watershed, one subbasin of the YanRB. In our
372 study area, the farming practices such as fertilization and tillage exerted great influences
373 on TN and TP loads in middle areas.

374 With respect to management of this region, our results showed that the three stations
375 (i.e., Shiyaocun, Zhujiagou, and Ganguyi) locating in the Baota district contributed
376 most to the pollutant loads, which was in agreement with the study by Wu et al. (2015).
377 Therefore, it is significant to pay much attention to this area for pollution control. In
378 addition, we found that the pollutant loads at Ganguyi were the highest (i.e., higher than
379 that of Tanjiahe which was close to the estuary), and thus the area around Ganguyi has
380 become a highly polluted area in recent years. Therefore, application of potential
381 measures (e.g., riparian filter zone, drip irrigation technology, hyperaccumulator plants
382 in farmland zone) in this area would be cost-effective. Furthermore, effective treatment
383 technologies and management of industrial and municipal wastewater are also

384 important to alleviate the deterioration of the river water quality (Wu and Chen, 2013).

385 4.2 Implications of the flood–pollution loads relationship

386 In previous studies on pollution of the Yan River, experimental approaches were used
387 to investigate sediment and nutrient losses due to erosion at the slope scale, and little
388 attention was paid to the changes of COD, TN, and TP under different hydrological
389 conditions, though these pollutants caused adverse effects on the aquatic ecosystem.
390 During our study period, an extremely-heavy rainstorm occurred in 2013, which caused
391 a rapid surface flow, creating and reshaping the dispersal pollutant patterns in a river
392 system (Grygar et al., 2014). It is noted that an extreme rainfall event can affect the
393 statistical trend analysis of the pollution load at the annual or monthly scale. Therefore,
394 we excluded an extremely–high water year (2013) when investigating annual/seasonal
395 average characteristics of pollutant loads, as shown in Section 3.2. To illustrate the
396 difference of the pollutant loads in extreme and normal hydrological condition, Fig. 8
397 compared the pollutant loads during 2013 and averaged loads across other years (i.e.,
398 2001–2012 and 2014–2018) at four stations (i.e., Shiyaocun, Zhujiagou, Ganguyi, and
399 Tanjiahe). The spatial analysis indicated that the impact of the high-intensive rainfall
400 on the downstream (from Ganguyi to Tanjiahe) was great due to the confluence of the
401 river tributaries in 2013, suggesting the severe threat of flooding events to the aquatic
402 environment. It has been recognized that pollutant fluxes were very large during an
403 extreme year because the pollutants temporarily stored in channels or slope lands are
404 quickly entrained and transferred during floods (Chen et al., 2015b). For example, the
405 average value of COD was 155,683 kg/d at Ganguyi during 2013, which is nearly 9

406 times the average load (18,211 kg/d) across other years. Our study demonstrated that
407 more attention should be paid to the impacts of the extreme hydrological year on water
408 quality for developing cost-effective prevention and controlling measures. This also
409 tells that a finer scale (e.g., day and hour) study of pollutant changes in an extreme
410 hydrological event (i.e., before and after the flooding) deserves further investigation.

411 4.3 Uncertainty analysis of the LOADEST modeling

412 The performance of the load estimation method (LOADEST) is dependent on many
413 factors including constituent type, streamflow characteristics, sampling frequency, and
414 water-quality record consistency (Lee et al., 2016). In terms of regression models, the
415 uncertainty may be higher because the LOADEST model assumes that all explanatory
416 variables have zero error (Pellerin et al., 2014). In this study, the uncertainty of the
417 LOADEST modeling deserves attention because the discrete and low sampling
418 frequency of TN could affect the simulation accuracy, and the hydrological change
419 could cause poor relations between constituent concentration and discharge. Further,
420 the streamflow fluctuated greatly, resulting in greater residual errors at high streamflow,
421 mainly because the regression model assumed a stationary concentration–flow
422 relationship. For example, positive hysteresis occurs when constituent concentrations
423 on the rising limb of a storm hydrograph are higher than those measured at equivalent
424 flows on the falling limb (Amirreza et al., 2017). Therefore, it would be significant but
425 challenging to improve the accuracy of load estimation methods. In addition,
426 topography plays an important role in pollutant diffuse, and the combination of
427 statistical models and watershed delineation would help identify contaminated areas

and achieve load reduction targets based on water environmental capacity (Park et al., 2015). In the future study, the uncertainties in land surface conditions, hydrological processes and extreme hydrological events are deserved to be investigated for better estimation of pollutant fluxes.

5. Conclusions

In this study, we investigated the spatiotemporal changes of pollutant loads by combining the hydrological model SWAT and the statistical water quality model LOADEST in the YanRB. We also quantified PS and NPS load using the digital filtering and hydrological statistical methods. The temporal analysis indicated that the pollutant loads showed a slight upward trend at four of five stations (i.e., Ansai, Shiyaocun, Ganguyi, Tanjiahe), and reached the maximum in 2014 due to the high precipitation under the normal water year conditions. The seasonal proportions of TN and TP loads indicated that there was a relatively higher level of pollutant loads in summer, followed by autumn, due to the high flow and frequent agricultural practices in these two wet seasons. Further, we found that the pollutant loads were concentrated in July, with relatively higher rainfall, suggesting much attention should be given to this month and extreme rainfall. Our spatial result showed that the most polluted area was in the midstream area (around Ganguyi) where more attention and measures should be considered. The high-water year was 2013, and the loads in this typical year were almost 9 times that of other years, indicating the potential risk in wet years. In addition, we found that NPS pollution contributed substantially to loads of COD, TN, and TP, with a contribution percentage of 87%, 85%, and 84%, respectively. Overall, our study

can provide decision makers with identification of pollution-prone areas/channels and the potential causes, contributing to the sustainable water quality management at the watershed scale on the Loess Plateau.

6. Appendix

In order to assess model performance compared to observations, the following popular criteria were used in this study:

(I) The percentage bias (PBIAS) measures the average difference of the observed and simulated value. The optimal value of PBIAS is 0.0, with low-magnitude values indicating accurate model simulation, while positive or negative values indicate overestimation or underestimation bias, respectively (Patrice et al., 1996).

$$PBIAS = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_{i,sim} - Y_{i,obs}}{Y_{i,obs}} \times 100\right) \quad (1)$$

(II) The Nash-Sutcliffe Efficiency (NSE) (Nash and Sutcliffe, 1970)) is a normalized statistic that measures the goodness of fit. If the efficiency becomes negative, model predictions are worse than a prediction performed using the average of all observations. while the efficiency is closer to 1, corresponding to a perfect match of modelled to the observed data

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Y_{i,sim} - Y_{i,obs}\right)^2}{\sum_{i=1}^{n}\left(Y_{i,obs} - \bar{Y}_{obs}\right)^2} \quad (2)$$

(III) The R^2 reveals the strength and direction of a linear relationship between the simulated and observed value. The difference between the NSE and the R^2 is that the NSE can interpret model performance in replicating individually observed values, while

the R^2 does not (Green and van Griensven, 2008).

$$R^2 = \frac{\left(\sum_{i=1}^{n}(Y_{i,obs} - \bar{Y}_{obs})(Y_{i,sim} - \bar{Y}_{sim})\right)^2}{\sum_{i=1}^{n}(Y_{i,obs} - \bar{Y}_{obs})^2 \sum_{i=1}^{n}(Y_{i,sim} - \bar{Y}_{sim})^2} \quad (3)$$

(IV) The PPCC test by using the correlation coefficient r between the ordered observation X_i and the corresponding fitted quantiles M_i was provided by Filliben (1975) for normality test. The corresponding fitted quantiles of this test are determined by plotting position for each observation. It is assumed that the observations could have been drawn from the fitted distribution if the value of r is close to 1.0.

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(M_i - \bar{M})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(M_i - \bar{M})^2}} \quad (4)$$

X and M denote the mean values of the observations X_i and the fitted quantiles M_i, respectively, and n is the sample size. The estimate of the order statistic median for M_i is shown as follows:

$$M_i = \Phi^{-1}(P_i) \quad (5)$$

$$P_i = 1 - (0.5)^{\frac{1}{n}}, \quad i=1 \quad (6)$$

$$P_i = \frac{i - 0.3175}{n + 0.365}, \quad i=2, 3, ..., n-1 \quad (7)$$

$$P_i = (0.5)^{\frac{1}{n}}, \quad i=n \quad (8)$$

Φ^{-1} is the inverse of cumulative distribution function.

7. Availability of data and materials

The data supporting SWAT analyses can be accessed publicly from the link provided in this paper, and the simulation data are available from the corresponding author on reasonable request.

8. Competing interests

The authors declared that they have no conflict of interests.

9. Funding

National Natural Science Foundation of China (31961143011), the Innovation Team of Shaanxi Province (2021TD-52), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB40020205), the Shaanxi Major Theoretical and Practical Program (20ST-106), the Shaanxi Key Research and Development Program of China (2018ZDXM-GY-030).

10. Authors' contributions

Yanni Song collected hydrological and water quality data, and prepared for the manuscript. Yiping Wu provided supervision, methodology, and financial support for this study, and checked/corrected all the versions of the manuscript. Changshun Sun provided water quality data. Fubo Zhao commented on earlier versions of the manuscript. Jingyi Hu commented on earlier versions of the manuscript. Ji Chen provided resources. Linjing Qiu provided hydrological data. All authors read and approved the final manuscript.

11. Acknowledgments

This study was funded by the National Natural Science Foundation of China (31961143011), the Innovation Team of Shaanxi Province (2021TD-52), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB40020205), the Shaanxi Major Theoretical and Practical Program (20ST-106), the Shaanxi Key Research and Development Program of China (2018ZDXM-GY-030), and the National

Thousand Youth Talent Program of China. We also thank the HPCC Platform in Xi'an Jiaotong University for computing equipment and computer maintenance.

12. Author information

Department of Earth & Environmental Science, Xi'an Jiaotong University, Xi'an, Shaanxi Province, 710049, China

Yanni Song, Yiping Wu, Fubo Zhao, Jingyi Hu, Linjing Qiu

Shaanxi Provincial Academy of Environmental Sciences, Xi'an, 710061, China

Changshun Sun

Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China

Ji Chen

13. Reference

Amirreza, S.;Haw, Y.;Carlington, W.;Gregory, M.C.;Wade, C.;Bahram, M.;Megan, L.;Ali, S.;Sangchul, L. and Judith, D. (2017) Effect of Water Quality Sampling Approaches on Nitrate Load Predictions of a Prominent Regression-Based Model. Water 9(11), 895-.

Arnold, J.G.;Srinivasan, R.;Muttiah, R.S. and Williams, J.R. (1998) Large Area Hydrologic Modeling and Assessment Part I: Model Development. JAWRA Journal of the American Water Resources Association 34(1), 1-17.

Aulenbach, B.T. (2013) Improving regression-model-based streamwater constituent load estimates derived from serially correlated data. Journal of Hydrology 503, 55-66.

Chen, D.;Hu, M.;Guo, Y. and Dahlgren, R.A. (2015a) Reconstructing historical changes in phosphorus inputs to rivers from point and nonpoint sources in a rapidly developing watershed in eastern China, 1980–2010. Science of the Total Environment 533, 196-204.

Chen, D.;Li, J.;Yang, X.;Zhou, Z.;Pan, Y. and Li, M. (2020) Quantifying water provision service supply, demand and spatial flow for land use optimization: A case study in the YanHe watershed. Ecosystem Services 43, 101117.

Chen, L.;Liu, R.;Huang, Q.;Chen, Y.;Gao, S.;Sun, C.;Shen, Z.;Ou, S. and Chen, S. (2012) An integrated simulation-monitoring framework for nitrogen assessment: a case study in the Baixi watershed, China. Procedia Environmental Sciences 13, 1076-1090.

Chen, L.;Sun, C.;Wang, G.;Xie, H. and Shen, Z. (2017) Event-based nonpoint source pollution prediction in a scarce data catchment. Journal of Hydrology 552, 13-27.

Chen, Y.;Song, X.;Zhang, Z.;Shi, P. and Tao, F. (2015b) Simulating the impact of flooding events on non-point source pollution and the effects of filter strips in an intensive agricultural watershed in China. Limnology 16(2), 91-101.

Cohn, T.A.;Gilroy, E.J. and Baier, W.G. (1992) Estimating fluvial transport of trace constituents using a regression model with data subject to censoring, The Survey.

Dai, X.;Zhou, Y.;Ma, W. and Zhou, L. (2017) Influence of spatial variation in land-use patterns and topography on water quality of the rivers inflowing to Fuxian Lake, a large deep lake in the plateau of southwestern China. Ecological Engineering 99, 417-428.

Dou, L. and Huang, M.-B. (2010) Applied Study of Baseflow Separation Methods in Watersheds of Loess Plateau. Bulletin of Soil and Water Conservation.

Drewry, J.;Newham, L.;Greene, R.;Jakeman, A.J. and Croke, B. (2006) A review of nitrogen and phosphorus export to waterways: Context for catchment modelling. Marine and Freshwater Research 57, 757-774.

Duan, S.;Powell, R.T. and Bianchi, T.S. (2014) High frequency measurement of nitrate concentration in the Lower Mississippi River, USA. Journal of Hydrology 519, 376-386.

Duan, W.;Takara, K.;He, B.;Luo, P.;Nover, D. and Yamashiki, Y. (2013) Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River, Japan, 1985 to 2010. Science of the Total Environment 461, 499-508.

Eckhardt, K. (2005) How to construct recursive digital filters for baseflow separation. Hydrological Processes: An International Journal 19(2), 507-515.

Filliben, J.J. (1975) The Probability Plot Correlation Coefficient Test for Normality. Technometrics 17(1), 111-117.

Green, C.H. and van Griensven, A. (2008) Autocalibration in hydrologic modeling: Using SWAT2005 in small-scale watersheds. Environmental Modelling & Software 23(4), 422-434.

Grygar, T.M.;Elznicová, J.;Bábek, O.;Hošek, M.;Engel, Z. and Kiss, T. (2014) Obtaining isochrones

560 from pollution signals in a fluvial sediment record: a case study in a uranium-polluted
561 floodplain of the Ploučnice River, Czech Republic. Applied Geochemistry 48, 1-15.
562 Guo, Y.;Chang Huang, C.;Pang, J.;Zhou, Y.;Zha, X. and Mao, P. (2016) Reconstruction palaeoflood
563 hydrology using slackwater flow depth method in the Yanhe River valley, middle Yellow
564 River basin, China. Journal of Hydrology, S0022169416307235.
565 Hall, F.R. (1968) Base-Flow Recessions—A Review. Water Resources Research 4(5).
566 Hao, F.;Yang, S.;Cheng, H.;B. U., Q. and Zheng, L. (2006) A method for estimation of non-point
567 source pollution load in the large-scale basins of China. Acta Scientiae Circumstantiae 26(3),
568 375-383.
569 He, S. and Lu, J. (2016) Contribution of baseflow nitrate export to non-point source pollution. Science
570 China Earth Sciences 59.
571 Jiang, Y.;Liu, C.;Hao, S.;Zhao, H. and Li, X. (2019) A framework to develop a watershed pollution load
572 model for semiarid and semihumid areas. Journal of Hydrology 579, 124179.
573 Lang, M.;Li, P. and Yan, X. (2013) Runoff concentration and load of nitrogen and phosphorus from a
574 residential area in an intensive agricultural watershed. Science of the Total Environment 458,
575 238-245.
576 Lee, C.J.;Hirsch, R.M.;Schwarz, G.E.;Holtschlag, D.J.;Preston, S.D.;Crawford, C.G. and Vecchia, A.V.
577 (2016) An evaluation of methods for estimating decadal stream loads. Journal of Hydrology
578 542, 185-203.
579 Li;Jiake~*;LI;Huaien~;SHEN;Bing~;and;LI;Yajiao~ and Xi'an (2011) Effect of non-point source
580 pollution on water quality of the Weihe River. International Journal of Sediment Research 01,
581 53-64.
582 Li , H. (2000) Mean concentration method for estimation of nonpoint source load and its application.
583 Actaentiae Circumstantiae.
584 Lin, C.;Wu, Z.;Ma, R. and Su, Z. (2016) Detection of sensitive soil properties related to non-point
585 phosphorus pollution by integrated models of SEDD and PLOAD. Ecological Indicators 60,
586 483-494.
587 Lin, C.;Ma, R. and Xiong, J. (2018) Can the watershed non-point phosphorus pollution be interpreted
588 by critical soil properties? A new insight of different soil P states. The ence of the Total
589 Environment 628-629(JUL.1), 870-881.

590 Lyne, V.D. and Hollick, M. (1979) Stochastic Time-Variable Rainfall-Runoff Modeling.
591 Miao, Y. (2018) Hydrological Elements Variation Feature Analysis of the Yanhe River Basin over the
592 Last 50 Years. Geospatial Information.
593 Miller, M.P.;Johnson, H.M.;Susong, D.D. and Wolock, D.M. (2015) A new approach for continuous
594 estimation of baseflow using discrete water quality data: Method description and comparison
595 with baseflow estimates from two existing approaches. Journal of Hydrology 522, 203-210.
596 Moriasi, D.;Arnold, J.;Van Liew, M.;Bingner, R.;Harmel, R.D. and Veith, T. (2007) Model Evaluation
597 Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions
598 of the ASABE 50.
599 Nash, J.E. and Sutcliffe, J.V. (1970) River flow forecasting through conceptual models. Part I a
600 discussion of principles. Journal of Hydrology 10, 282-290.
601 Neitsch, S.;Arnold, J.;Kiniry, J. and Williams, J.R. (2011) Soil and water asessment tool theoritical
602 documentation: version 2009, Texas Water Resources Institute technical report No. 406. Texas
603 Water Resources Institute, Texas A&M University, Texas. USA.
604 Oelsner, G.P. and Stets, E.G. (2019) Recent trends in nutrient and sediment loading to coastal areas of
605 the conterminous US: Insights and global context. Science of the Total Environment 654,
606 1225-1240.
607 Park, Y.;Engel, B.;Frankenberger, J. and Hwang, H. (2015) A Web-Based Tool to Estimate Pollutant
608 Loading Using LOADEST. Water 7(9), 4858-4868.
609 Park, Y.S. and Engel, B.A. (2016) Identifying the correlation between water quality data and
610 LOADEST model behavior in annual sediment load estimations. Water 8(9), 368.
611 Patrice;O.;Yapo;and;Hoshin;Vijai;Gupta;and;Soroosh and Sorooshian (1996) Automatic calibration of
612 conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology.
613 Pellerin, B.A.;Bergamaschi, B.A.;Gilliom, R.J.;Crawford, C.G.;Saraceno, J.F.;Frederick,
614 C.P.;Downing, B.D. and Murphy, J.C. (2014) Mississippi River Nitrate Loads from High
615 Frequency Sensor Measurements and Regression-Based Load Estimation. Environmental ence
616 & Technology 48(21), 12612-12619.
617 Runkel, R.L.;Crawford, C.G. and Cohn, T.A. (2004) Load Estimator (LOADEST): A FORTRAN
618 program for estimating constituent loads in streams and rivers.
619 Schwarzenbach, R.P.;Egli, T.;Hofstetter, T.B.;Gunten, U.v. and Wehrli, B. (2010) Global Water

620 Pollution and Human Health. Annual Review of Environment and Resources 35(1), 109-136.
621 Segura, C.;Noone, D.;Warren, D.;Jones, J.A.;Tenny, J. and Ganio, L.M. (2019) Climate, Landforms,
622 and Geology Affect Baseflow Sources in a Mountain Catchment. Water Resources Research
623 55.
624 Sharifi, A.;Yen, H.;Wallace, C.W.;McCarty, G.;Crow, W.;Momen, B.;Lang, M.W.;Sadeghi, A.;Lee, S.
625 and Denver, J. (2017) Effect of water quality sampling approaches on nitrate load predictions
626 of a prominent regression-based model. Water 9(11), 895.
627 Shi, X.;Li, L. and Zhang, T. (2015) Water Pollution Control Action Plan, a Realistic and Pragmatic
628 Plan—— An Interpretation of Water Pollution Control Action Plan. Environmental Protection
629 ence.
630 Su, C.;Fu, B.;Lu, Y.;Lu, N.;Zeng, Y.;He, A. and Lamparski, H. (2011a) Land use change and
631 anthropogenic driving forces: A case study in Yanhe River Basin. Chinese Geographical ence
632 21(5), 587.
633 Su, C.;Fu, B.;Lu, Y.;Lu, N.;Zeng, Y.;He, A. and Lamparski, H. (2011b) Land use change and
634 anthropogenic driving forces: A case study in Yanhe River Basin. Chinese Geographical
635 Science 21(5), 587.
636 Su, C.;Fu, B.;Wei, Y.;Lü, Y.;Liu, G.;Wang, D.;Mao, K. and Feng, X. (2012) Ecosystem management
637 based on ecosystem services and human activities: a case study in the Yanhe watershed.
638 Sustainability Science 7(1), 17-32.
639 Sudheer, K.;Chaubey, I.;Garg, V. and Migliaccio, K. (2007) Impact of time-scale of the calibration
640 objective function on the performance of watershed models. Hydrological Processes 21, 3409-
641 3419.
642 Sun, P.;Wu, Y.;Yang, Z.;Sivakumar, B.;Qiu, L.;Liu, S. and Cai, Y. (2019) Can the Grain-for-Green
643 Program Really Ensure a Low Sediment Load on the Chinese Loess Plateau? Engineering
644 5(5), 855-864.
645 Tong, Y.;Zhao, Y.;Zhen, G.;Chi, J.;Liu, X.;Lu, Y.;Wang, X.;Yao, R.;Chen, J. and Zhang, W. (2015)
646 Nutrient loads flowing into coastal waters from the main rivers of China (2006–2012).
647 Scientific reports 5, 16678.
648 USEPA (2015) Connectivity of streams and wetlands to downstream waters: A review and synthesis of
649 the scientific evidence, US Environmental Protection Agency Washington, DC.

650 Wang, C. and Zhao, H. (2016) The assessment of urban ecological environment in watershed scale.
651 Procedia Environmental Sciences 36, 169-175.
652 Wang, H.;Zening, W.;Caihong, H. and Xinzhong, D. (2015a) Water and nonpoint source pollution
653 estimation in the watershed with limited data availability based on hydrological simulation
654 and regression model. Environmental Science and Pollution Research 22(18), 14095-14103.
655 Wang, Q. and Yang, Z. (2016) Industrial water pollution, water environment treatment, and health risks
656 in China. Environmental Pollution 218, 358-365.
657 Wang, Y.;Liu, X.;Li, Y.;Liu, F.;Shen, J.;Ma, Q.;Yin, J. and Wu, J. (2015b) Rice agriculture increases
658 base flow contribution to catchment nitrate loading in subtropical central China. Agriculture,
659 Ecosystems and Environment 214, 86-95.
660 Wei, M.;Duan, P.;Gao, P.;Guo, S. and Li, M. (2020) Exploration and application of hydrochemical
661 characteristics method for quantification of pollution sources in the Danjiangkou Reservoir
662 area. Journal of Hydrology, 125291.
663 Wu, L.;Gao, J.-e.;Ma, X.-y. and Li, D. (2015) Application of modified export coefficient method on the
664 load estimation of non-point source nitrogen and phosphorus pollution of soil and water loss
665 in semiarid regions. Environmental Science and Pollution Research 22(14), 10647-10660.
666 Wu, L.;Liu, X. and Ma, X.Y. (2016) Spatio-temporal variation of erosion-type non-point source
667 pollution in a small watershed of hilly and gully region, Chinese Loess Plateau.
668 Environmental Science & Pollution Research 23(11), 10957-10967.
669 Wu, L.;He, Y. and Ma, X. (2020) Using five long time series hydrometeorological data to calibrate a
670 dynamic sediment delivery ratio algorithm for multi-scale sediment yield predictions.
671 Environmental Science and Pollution Research, 1-16.
672 Wu, Y. and Chen, J. (2013) Investigating the effects of point source and nonpoint source pollution on
673 the water quality of the East River (Dongjiang) in South China. Ecological Indicators 32(sep.),
674 294-304.
675 Xia, Y.;Zhang, M.;Tsang, D.C.;Geng, N.;Lu, D.;Zhu, L.;Igalavithana, A.D.;Dissanayake, P.D.;Rinklebe,
676 J. and Yang, X. (2020) Recent advances in control technologies for non-point source pollution
677 with nitrogen and phosphorous from agricultural runoff: current practices and future
678 prospects. Applied Biological Chemistry 63(1), 1-13.
679 Xin, X.K.;Yin, W. and Li, K.F. (2017) Estimation of non-point source pollution loads with flux method

in Danjiangkou Reservoir area, China. Water ence and Engineering 10(2), 134-142.

Yang, Q.;Li, Z.;Han, Y. and Gao, H. (2020) Responses of Baseflow to Ecological Construction and Climate Change in Different Geomorphological Types in The Middle Yellow River, China. Water 12, 304.

Zeng, Z.;Luo, W.-G.;Wang, Z. and Yi, F.-C. (2021) Water Pollution and Its Causes in the Tuojiang River Basin, China: An Artificial Neural Network Analysis. Sustainability 13(2).

Zhang, J.;Zhang, Y.;Song, J. and Cheng, L. (2017) Evaluating relative merits of four baseflow separation methods in Eastern Australia. Journal of Hydrology 549, 252-263.

Zhang, J.;Song, J.;Cheng, L.;Zheng, H.;Wang, Y.;Huai, B.;Sun, W.;Qi, S.;Zhao, P.;Wang, Y. and Li, Q. (2019a) Baseflow estimation for catchments in the Loess Plateau, China. Journal of Environmental Management 233, 264-270.

Zhang, S.;Wu, Y.;Sivakumar, B.;Mu, X.;Zhao, F.;Sun, P.;Sun, Y.;Qiu, L.;Chen, J. and Meng, X. (2019b) Climate change-induced drought evolution over the past 50 years in the southern Chinese Loess Plateau. Environmental Modelling and Software 122, 104519.

Zhao, J.;Feng, X.;Deng, L.;Yang, Y.;Zhao, Z.;Peng, C. and Fu, B. (2020) Quantifying the Effects of Vegetation Restorations on the Soil Erosion Export and Nutrient Loss on the Loess Plateau. Frontiers in Plant Science 11.

Zhou, L. and Li, Y. (2018) A study on China's water pollution status and water environment management strategy. Environment & Development.

Zhou, Z. and Li, J. (2015) The correlation analysis on the landscape pattern index and hydrological processes in the Yanhe watershed, China. Journal of Hydrology 524, 417-426.

Zhu, Y.;Chen, L.;Wei, G.;Li, S. and Shen, Z. (2019) Uncertainty assessment in baseflow nonpoint source pollution prediction: The impacts of hydrographic separation methods, data sources and baseflow period assumptions. Journal of Hydrology 574, 915-925.

Zou, L.;Liu, Y.;Wang, Y. and Hu, X. (2020) Assessment and analysis of agricultural non-point source pollution loads in China: 1978–2017. Journal of Environmental Management 263, 110400.

Table Captions

Table 1. Sensitive parameters and optimized values of SWAT using SUFI-2.

Table 2. Evaluation of LOADEST performance in simulating COD, TN, and TP loads.

709

710 **Figure Captions**

711 Fig. 1 Location, elevation, and land use of the Yan River Basin (YanRB).

712 Fig. 2 Evaluation of the model performance in streamflow simulation at the Ganguyi
713 gaging station during the four-year (2001-2004) calibration and four-year (2005–2008)
714 validation periods.

715 Fig. 3 Comparison of the simulated versus the observed COD (a and b), TN (c and d),
716 and TP (e and f) loads. The scatterplots on the right panel represent comparisons
717 between observed and simulated loads.

718 Fig. 4 Estimated monthly average loads of COD, TN, and TP at five stations along the
719 mainstream of the Yan River.

720 Fig. 5 Estimated multi-year average loads of COD, TN, and TP in each calendar
721 month at five stations in the Yan River Basin.

722 Fig. 6 The multi-year average proportions of TN and TP loads in each season at five
723 water quality monitoring stations. The big pie charts represent the multi-year average
724 proportions of TN and TP, respectively.

725 Fig. 7 The contributions of PS and NPS pollutant loads at five water quality monitoring
726 stations.

727 Fig. 8 Boxplots of COD, TN, and TP loads during flood and non-flood periods at four
728 water quality monitoring stations: (a) Shiyaocun, (b) Zhujiagou, (c) Ganguyi, and (d)
729 Tanjiahe. The horizontal lines represent (from the top) the maximum, the third quartile,
730 the median, the first quartile and the minimum. The red circles represent average values,

and the blue squares represent outliers.

Figures

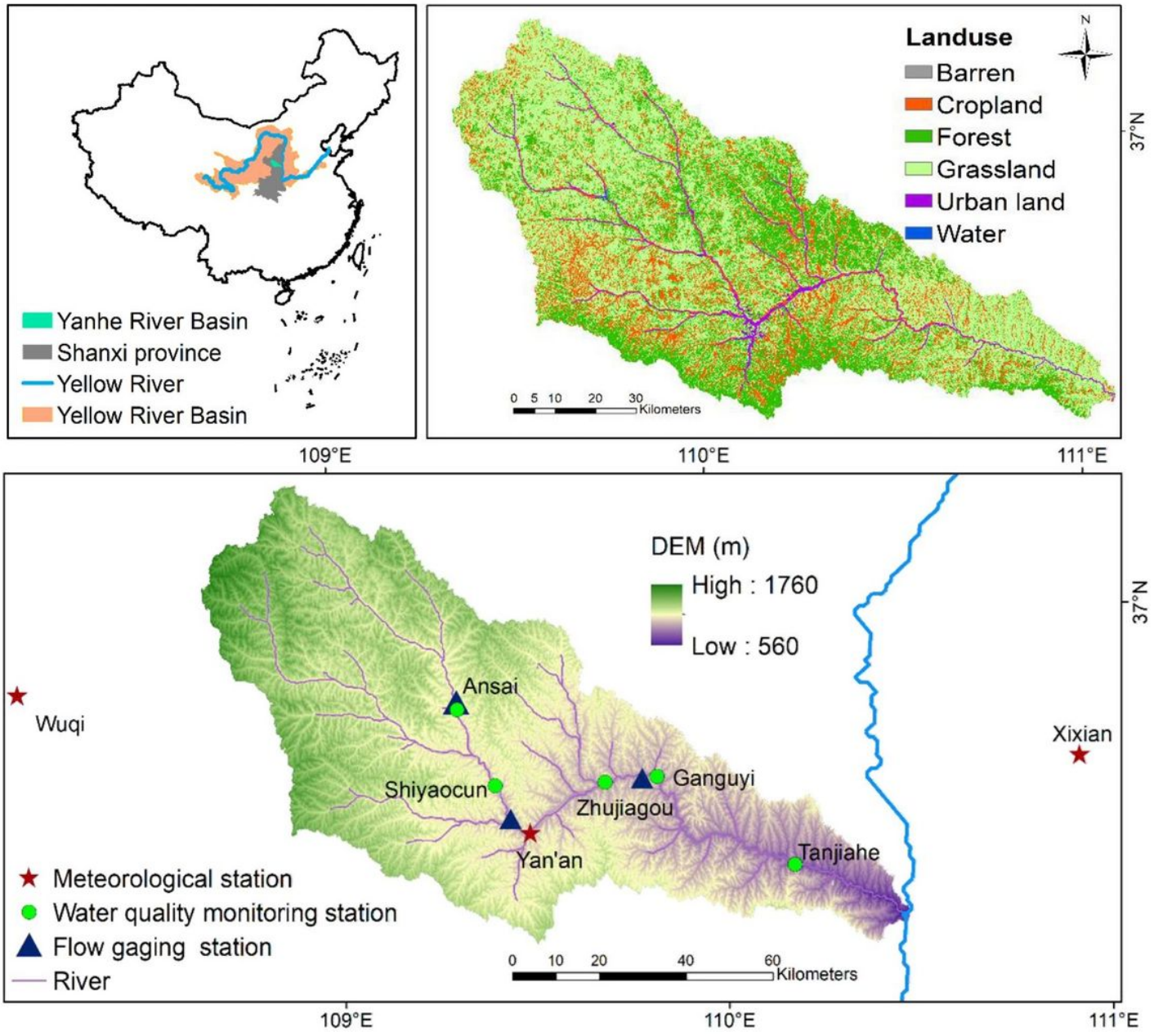

Figure 1

Location, elevation, and land use of the Yan River Basin (YanRB). Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

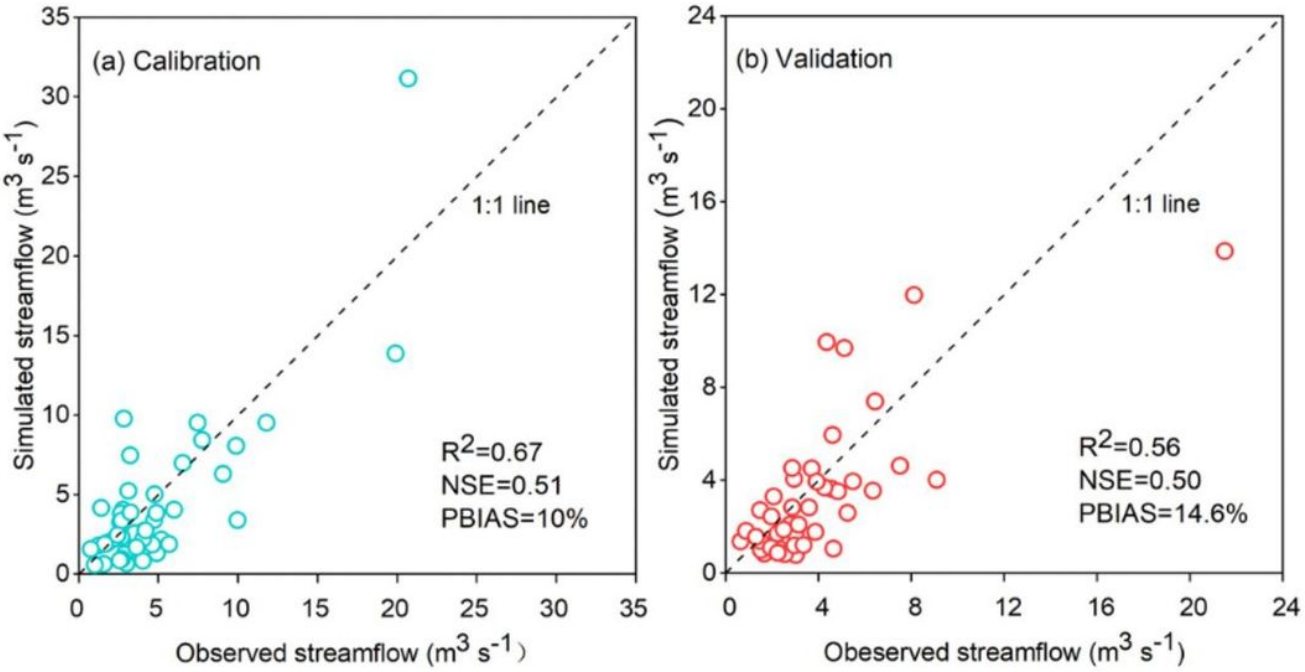

Figure 2

Evaluation of the model performance in streamflow simulation at the Ganguyi gaging station during the four-year (2001–2004) calibration and four-year (2005–2008) validation periods.

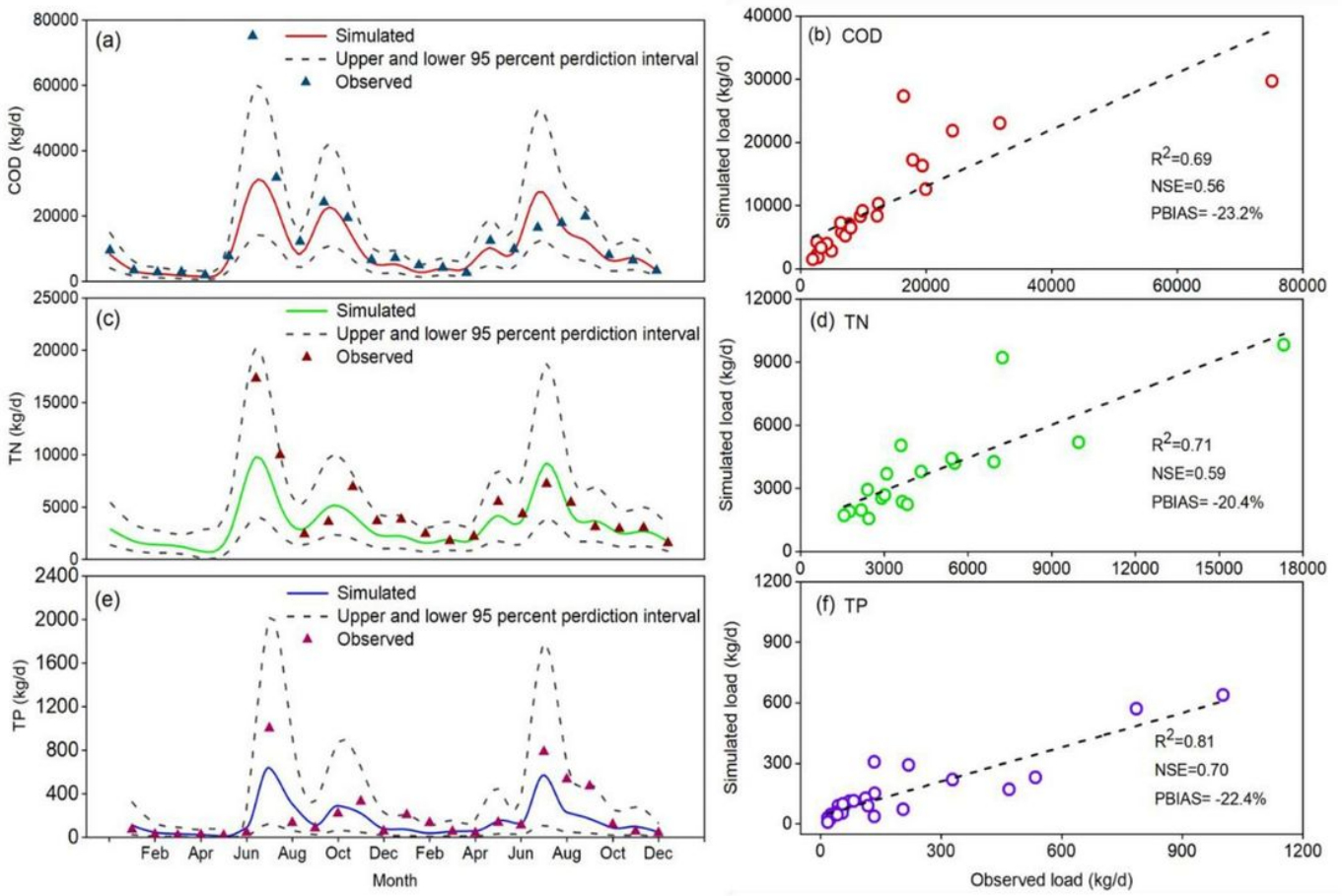

Figure 3

Comparison of the simulated versus the observed COD (a and b), TN (c and d), and TP (e and f) loads. The scatterplots on the right panel represent comparisons between observed and simulated loads.

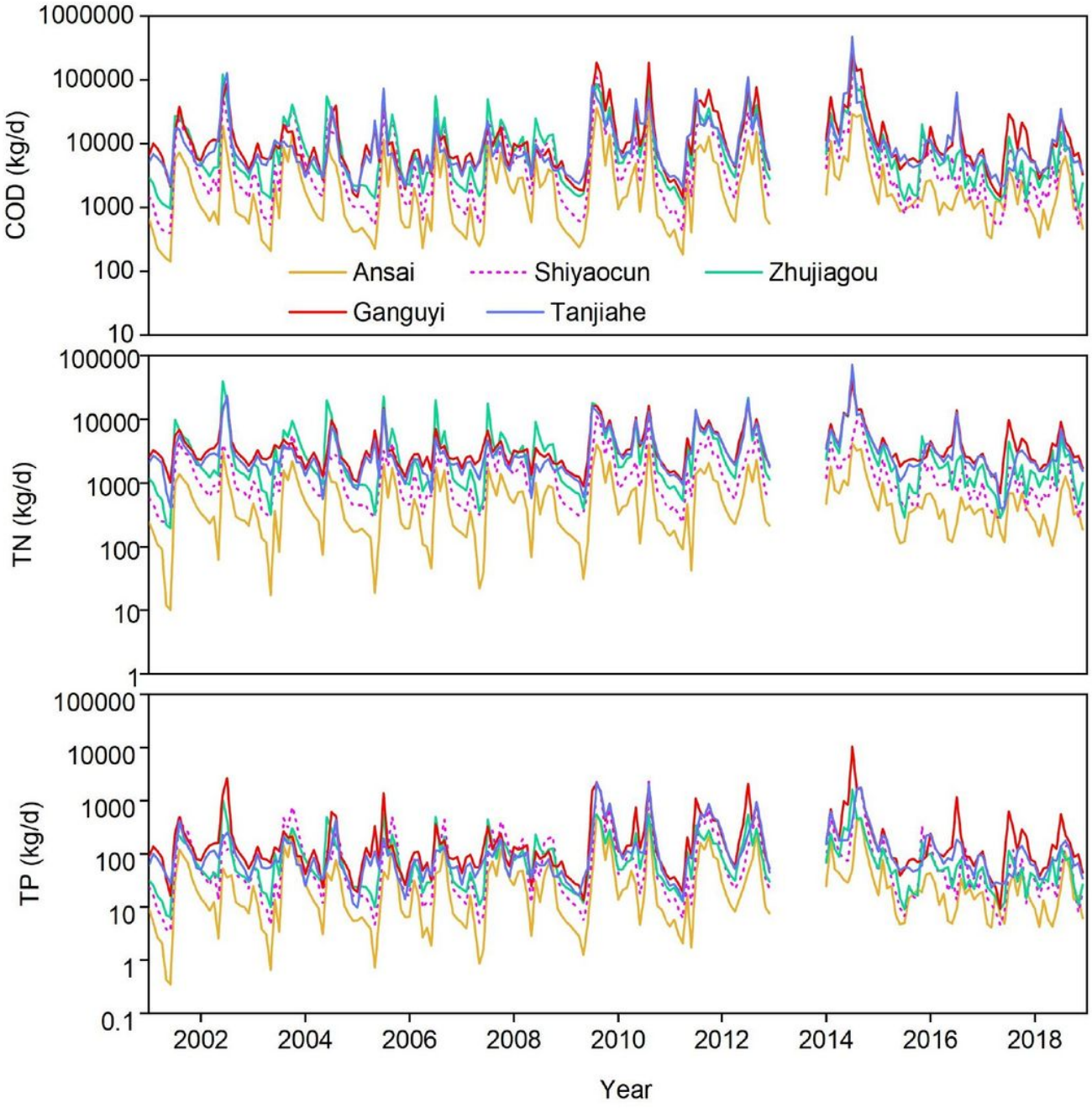

Figure 4

Estimated monthly average loads of COD, TN, and TP at five stations along the mainstream of the Yan River.

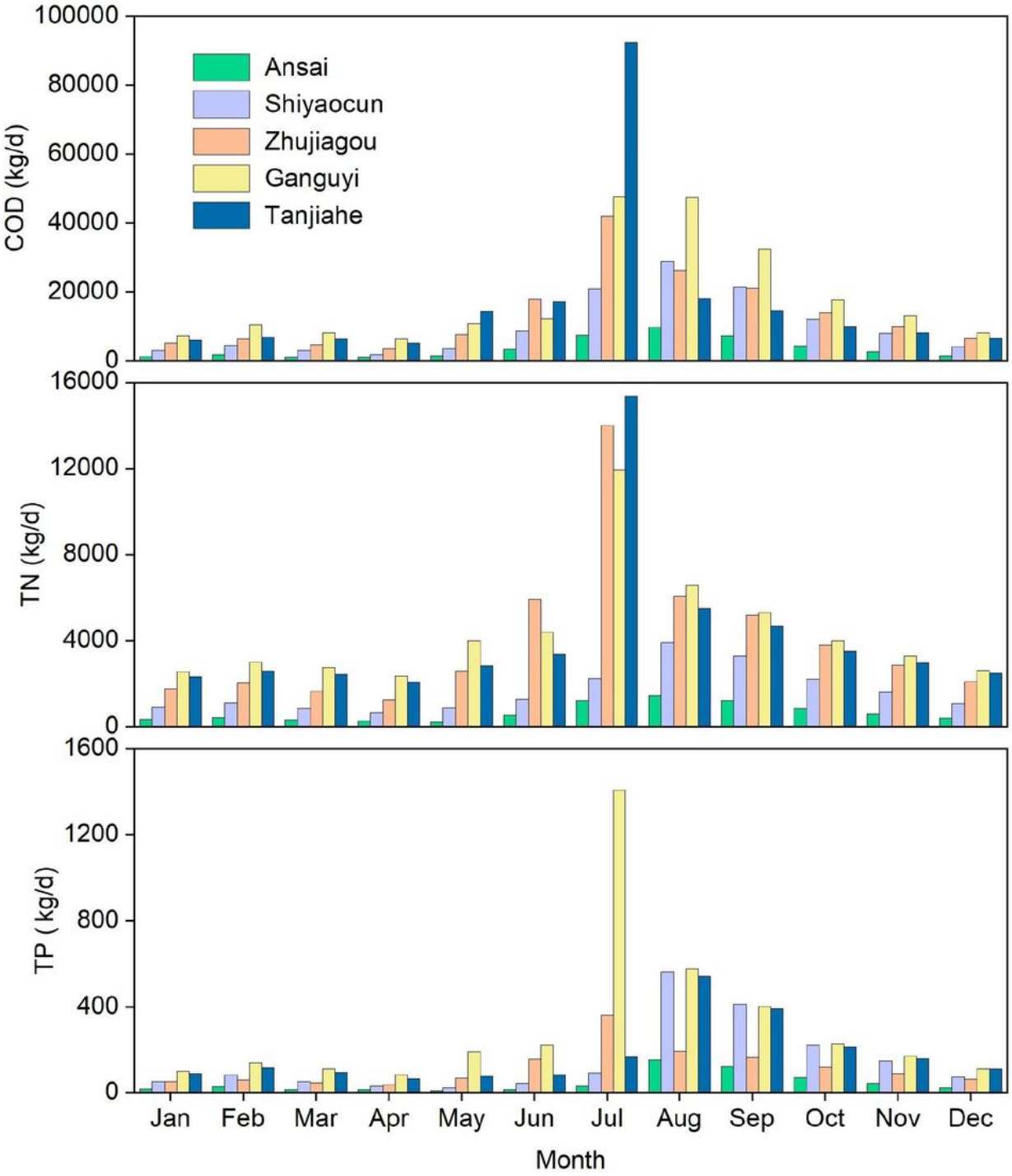

Figure 5

Estimated multi-year average loads of COD, TN, and TP in each calendar month at five stations in the Yan River Basin.

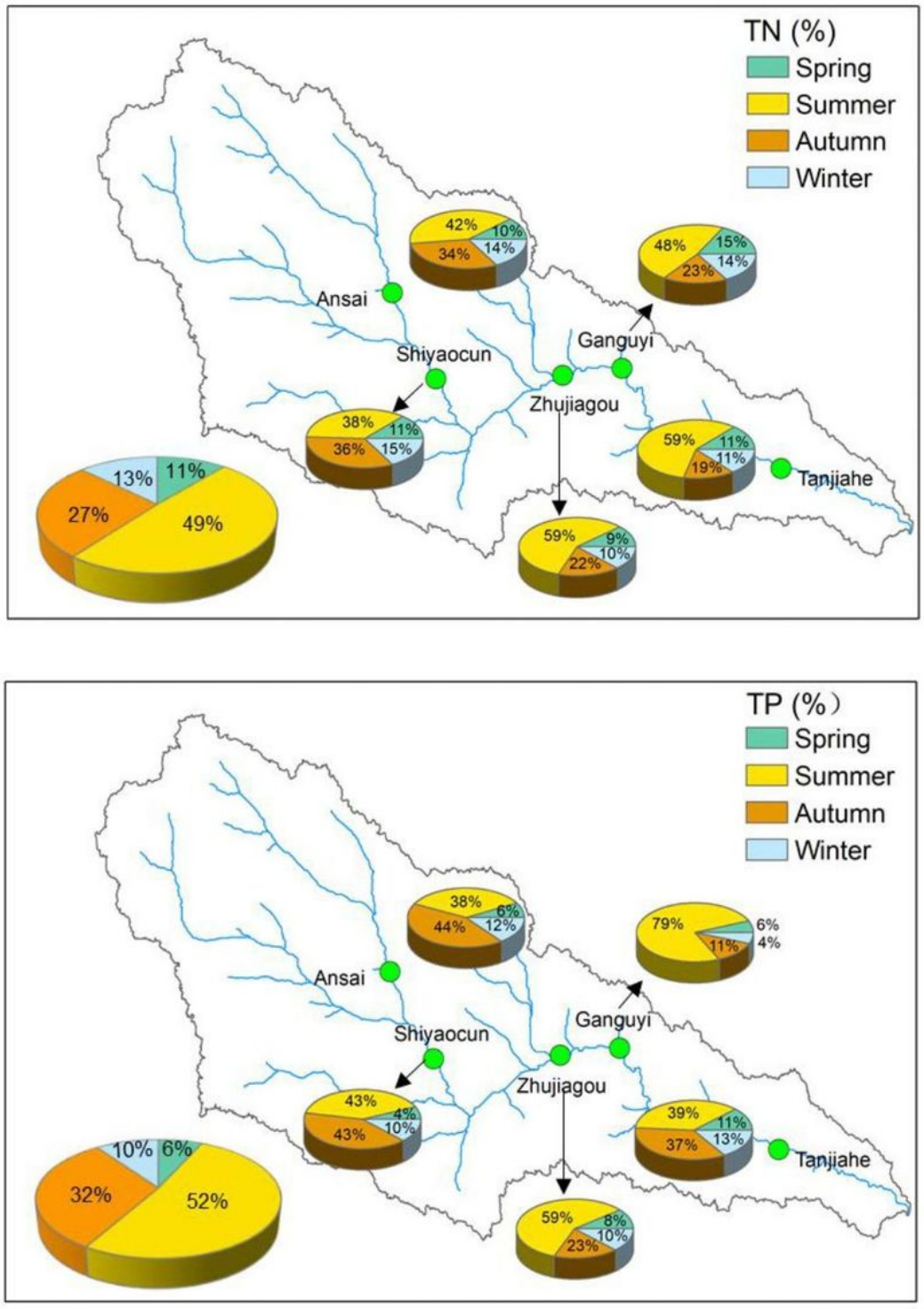

Figure 6

The multi-year average proportions of TN and TP loads in each season at five water quality monitoring stations. The big pie charts represent the multi-year average proportions of TN and TP, respectively. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country,

territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

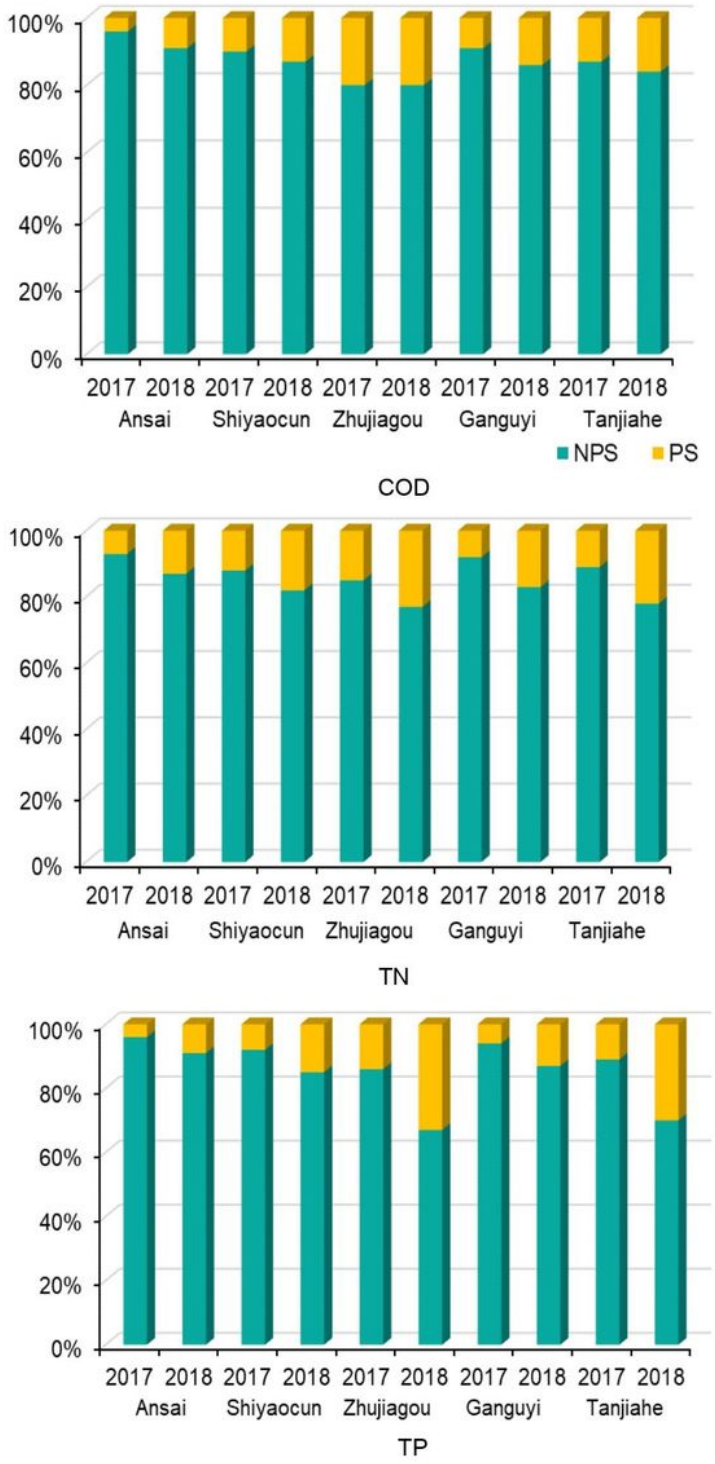

Figure 7

The contributions of PS and NPS pollutant loads at five water quality monitoring stations.

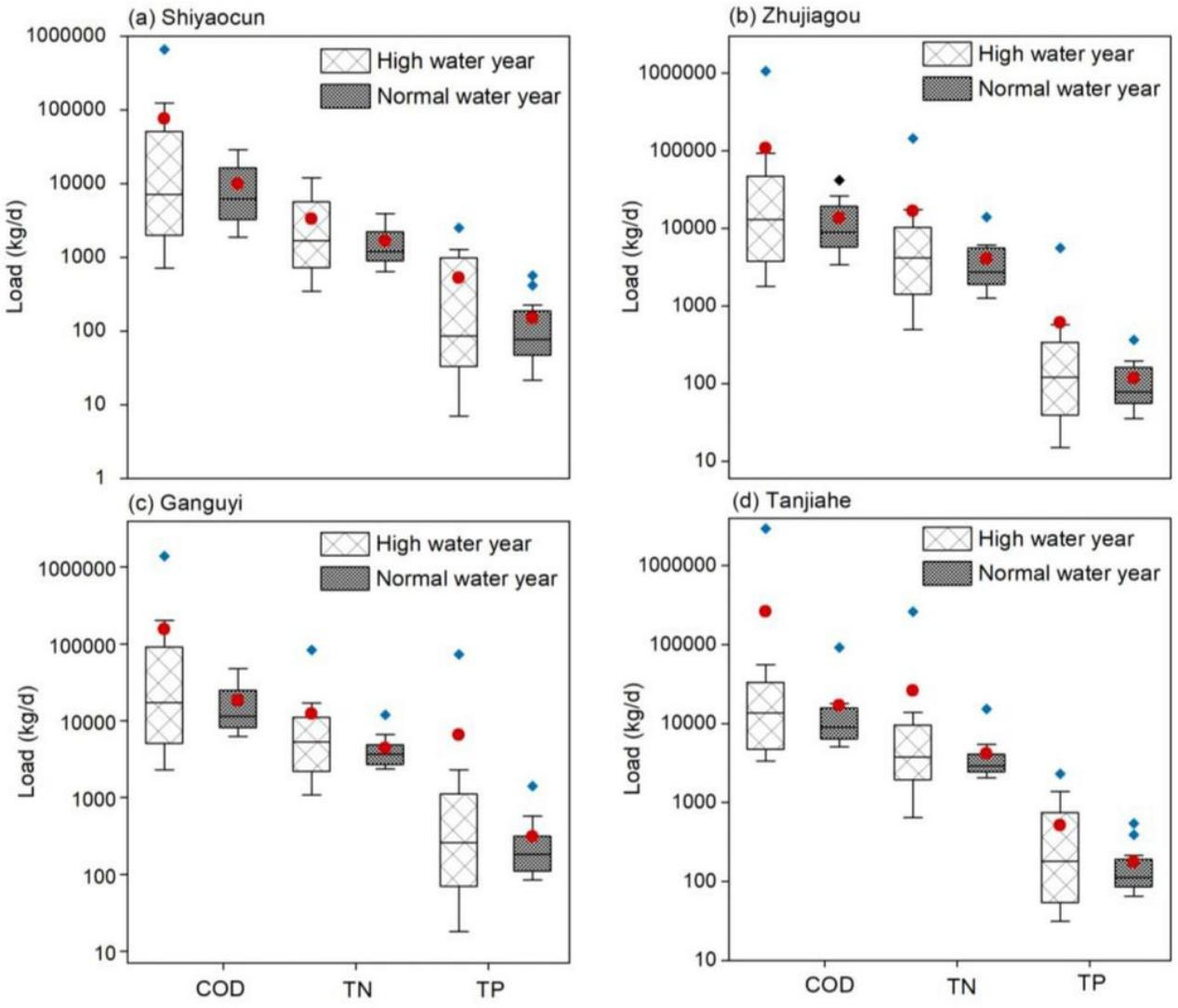

Figure 8

Boxplots of COD, TN, and TP loads during a high water year (2013) and the normal condition (averaged loads across 2001-2018 excluding 2013) at four water quality monitoring stations: (a) Shiyaocun, (b) Zhujiagou, (c) Ganguyi, and (d) Tanjiahe. For (a)-(d), max = top end of whisker, 3rd quantile = top edge of the rectangle, bold black line = median,1st quantile = bottom edge of the rectangle, min = bottom end of whisker. The red circles represent average values, and the blue diamonds represent outliers.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile.docx