Stochastic Environmental Research and Risk Assessment
https://doi.org/10.1007/s00477-018-1532-2

ORIGINAL PAPER

Influential factors detection for surface water quality with geographical detectors in China

Jiaxin Wang[1,2] · Maogui Hu[1,4] · Fengsong Zhang[1] · Bingbo Gao[3]

Abstract
Surface water quality is a matter of serious concern in China. This study quantitatively analyzes the spatial–temporal characteristics of surface water quality among 100 monitoring stations in China during 2015. A geographical detector was used to detect the influential annual and seasonal factors. Surface water quality is primarily controlled by the content of nutrient pollutants and organic pollutants. Natural factors (precipitation, temperature, soil erosion, and terrain) and anthropogenic factors [land use type, population density, and per capita gross domestic product (GDP-per-capita)] were selected as geographical proxies to be tested for their explanatory power for surface water quality. Results indicated that the top three factors influencing the annual mean of nutrient pollutants were the population density, terrain, and precipitation, the explanatory power of which was 0.82, 0.35, and 0.24, respectively. The interactive explanatory power for population density and terrain was 0.88 and for population density and precipitation was 0.87, both exhibiting enhanced interaction relationships. The top three factors influencing the annual mean of organic pollutants were population density, temperature, and basin, the explanatory power of which was 0.46, 0.29, and 0.27, respectively. The interactive explanatory power for population density and basin was 0.80 and for terrain and precipitation was 0.82, both demonstrating a nonlinear enhanced interaction relationship. For seasonal changes, the nutrient pollutants and organic pollutants were both affected by agricultural runoff due to seasonal farming. This study revealed that anthropogenic factors influenced surface water quality two to three times more than natural factors.

Keywords Surface water quality · Spatial–temporal analysis · Geographical Detector · CCME-WQI

1 Introduction

Surface water is an important resource for agricultural, drinking, environmental, and industrial uses. However, surface water quality is a matter of serious concern in many countries, including China (Varol et al. 2012). According to the 2015 China Environmental Status Bulletin, 35.5 percent of surface water quality were polluted to different degrees (Ministry of Environmental Protection 2016). Both anthropogenic influences and natural processes degrade surface waters (Phung et al. 2015; Zhou et al. 2017) and impair their use for drinking, industrial, agricultural, recreation, or other purposes (Varol et al. 2012; Simeonov et al. 2003). To effectively protect water resources, prevent water pollution, and improve water quality, it is urgent to explore the factors influencing water quality in China quantitatively.

✉ Maogui Hu
humg@lreis.ac.cn

1 State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

2 University of Chinese Academy of Sciences, Beijing 100049, China

3 Beijing Research Center for Information Technology in Agriculture, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

4 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

Published online: 13 March 2018

Many parameters are surrogates for water quality (Susanna and Wenli 2002; Nazeer et al. 2014; Seeboonruang 2012; Terrado et al. 2010; Zhao et al. 2015). Ha and Stenstrom (2003) selected 42 candidate variables among 90 water quality variables, which were detected in more than 25 percent of monitoring samples. Later, machine learning was used to assess the classification of the water quality by training the water quality parameters (Li et al. 2013). Discriminant analysis was used to reduce the number of sampling parameters, which allowed a reduction in the dimensionality of the large data set and indicated a few significant parameters were responsible for large variations in water quality (Huang et al. 2011; Varol et al. 2012; Phung et al. 2015; Shrestha and Kazama 2007). Recently, researchers combined a series of water quality parameters into a single index with a simple formula to evaluate water quality (Akkoyunlu and Akiner 2012; Misaghi et al. 2017). The Canadian Council of Ministers of the Environment Water Quality Index (CCME WQI) is widely used because of the advantage of flexibility (Terrado et al. 2010; Liu and Wu 2016).

The factors affecting surface water quality include natural and anthropogenic sources (Simeonov et al. 2003). In nature, changes in temperature and precipitation can affect the flow of water and thus affect the movement and diffusion of pollutants; the increase of temperature can influence chemical reactions in the water, affecting water quality and aquatic ecology (Wang et al. 2013; Dyer et al. 2014; Whitehead et al. 2009). Moreover, water quality varies significantly with the seasons (Varol et al. 2012; Ouyang et al. 2006; Sundaray et al. 2006; Nienie et al. 2017), soil erosion (Phung et al. 2015), terrain (Sun et al. 2013) and so on. Anthropogenic factors include pollutant emissions from agriculture, industry and urban wastewater (Zhou et al. 2017; Ha and Stenstrom 2003; Seeboonruang 2012; Zhao et al. 2015; Chen et al. 2016; Shen et al. 2014; Wilson 2015). It was found that in the Fuji river basin of Japan, the effects of meteorological factors on water quality are minor, agricultural factors are secondary, and industrial pollution effects are the most serious (Shrestha and Kazama 2007). However, the dominant factor might vary in different countries and regions. In summary, the primary factors influencing water quality include the natural factors of temperature, precipitation, soil erosion and terrain, and the anthropogenic factors of land use type, landscape pattern, agricultural runoff, agricultural activities, domestic sewage, and industrial effluents. Of course, natural and anthropogenic factors cannot be separated completely. Some factor, like soil erosion, belongs to both natural and anthropogenic factors.

Recent researches focused mainly on the micro-scale of small watersheds and rarely involved the macro-scale distribution of surface water at the national level. Most studies statistically analyzed the correlation between water parameters, and lack direct quantitative exploration of the relationships between parameters and influential factors. The effect of interactions between factors was also not considered in most researches. In this study, real-time monitoring data of 100 stations around China in 2015 was used as published by the surface water quality automatic monitoring network (http://online.watertest.com.cn/), including pH, dissolved oxygen (DO), chemical oxygen demand (COD), and ammonium (NH_4^+-N), to analyze the spatial and temporal distribution and variations of surface water quality from season to season. Several proxy variables were selected to represent the influential factors, including natural (temperature, precipitation, soil erosion and terrain) and anthropogenic factors (land use type, population density, and per capita gross domestic product (GDP-per-capita) and quantitate seasonal variations. The geographical detector method (Wang et al. 2010) was used to quantitatively analyze relationships between surface water quality and environmental risk factors in China. The results would be helpful to find the specific contribution of each important factor and their interactions in China, which can guide measures to be taken according to different conditions.

2 Materials and methods

2.1 Determinants of surface water quality and their proxies

Surface water quality is interactively determined by the content of organic pollutants (such as carbohydrates and proteins), nutrient pollutants (such as NH_4^+-N and P), activities of aquatic organisms, and pH levels, indicated by COD, NH_4^+-N, DO and pH values, respectively (Simeonov et al. 2003; Zheng 2012). Due to the mobility of water, the surface water quality is similar within a basin. Basin was chosen as a proxy variable to reflect the water mobility. Activities of aquatic organisms and pH levels are related to the content of organic pollutants and nutrient pollutants. The main sources of organic pollutants and nutrient pollutants are the surface runoff, domestic sewage, and industrial effluents. The surface runoff brings humus, pesticides, and chemical fertilizers into the surface water; it is the driving force for soil erosion, and its direction is related to the terrain, and its size is related to the precipitation. Soil erosion, terrain and precipitation were selected as proxy variables. Land use type, population density, and GDP-per-capita were chosen as proxy variables to represent domestic sewage and industrial effluents. In addition, temperature affects activities of the aquatic organisms. The

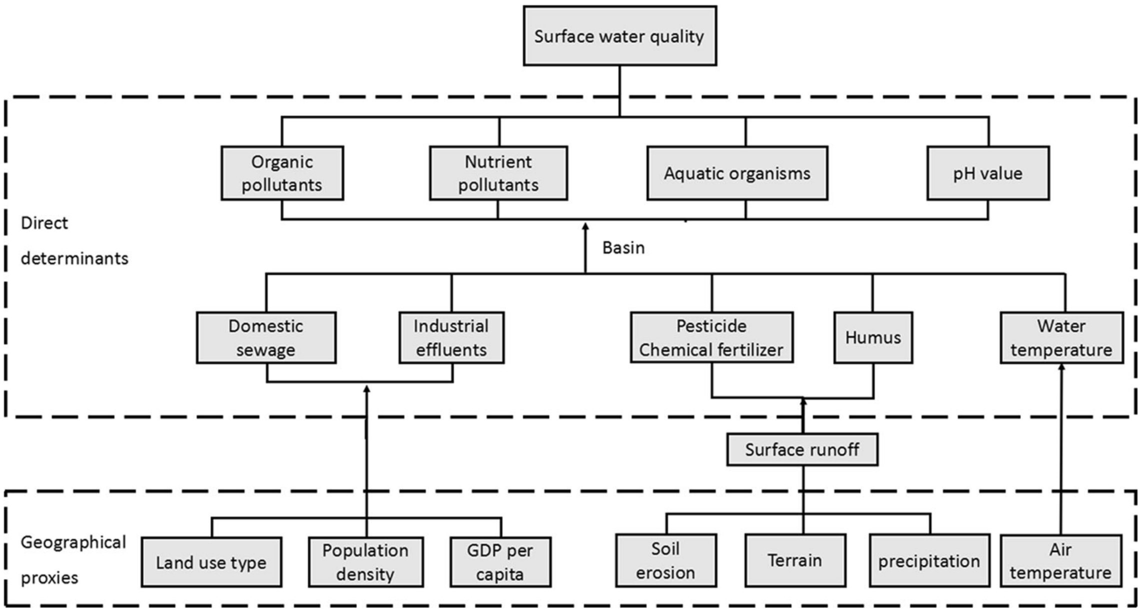


Fig. 1 Determinants and their proxies

proxy variable associations of potential factors that could affect the surface water quality are illustrated in Fig. 1.

2.2 Data

2.2.1 Surface water quality indicators

The Ministry of Environmental Protection in China has built 100 automatic water quality monitoring stations on 63 rivers and 13 large lakes, which are belong to the top seven river basins in China (Haihe River, Liaohe River, Huaihe River, Huanghe River, Songhua River, Yangtze River, and Pearl River) (Fig. 2). The stations on Amur River, Tumen River were merged into Songhua River, and the stations on Yalu River were merged into Liaohe River. The monitoring indicators include pH, DO, total organic carbon (TOC), COD, and NH_4^+-N. The water quality categories of stations were identified according to the National Standard of Environmental Quality for Surface Water (GB3838—2002) (Table 1),[1] with the exception that TOC has no established standard.

The monitoring indicators data for pH, DO, TOC, COD, and NH_4^+-N were released every 4 h in 2015. The monthly average for each indicator was calculated for every station and was used for the spatial and seasonal analysis of surface water quality and detection of geographical factors in China.

[1] http://online.watertest.com.cn/help.aspx.

2.2.2 Geographical proxies

The data sets of geographical proxies were provided by the Data Center for Resources and Environmental Sciences, Chinese Academy of Sciences (RESDC) (http://www.resdc.cn). The temperature and precipitation data were interpolated using the inverse distance weighted method based on the annual mean data from 1915 weather stations in China (Fig. 3a, b). Data for the terrain were derived from the Landscape Atlas of the People's Republic of China (1:1,000,000) (Fig. 3c). Soil erosion was categorized according to the general requirements of the Classification and Grading Standards for Soil Erosion (SL 190-96), the industry standard of the People Republic of China (Fig. 3d). Land use types were derived from the remote sensing monitoring database of land use in 2010 (Fig. 3e). Seeboonruang (2012) determined that nonpoint source pollution resulting from land use is defined as a diffuse source of contamination from a wide area, and it is often difficult to attribute this contamination to a single location. Cao and Sun (2012) discussed the influence of the spatial pattern of the dominant types of landscape on water quality. The dominant types of landscapes in the subbasins were calculated as the basin land use type to evaluate the influence on surface water quality (Fig. 3f). For the basin, it was generated using the digital elevation model (DEM) for watershed analysis in ArcGIS 10.2. The DEM was created from the Shuttle Radar Topography Mission data collected in 2000 (Fig. 3g). For the population density and GDP-per-capita information, the raw raster data in 2010 was transferred into vector data at the province level (Fig. 3h, i).

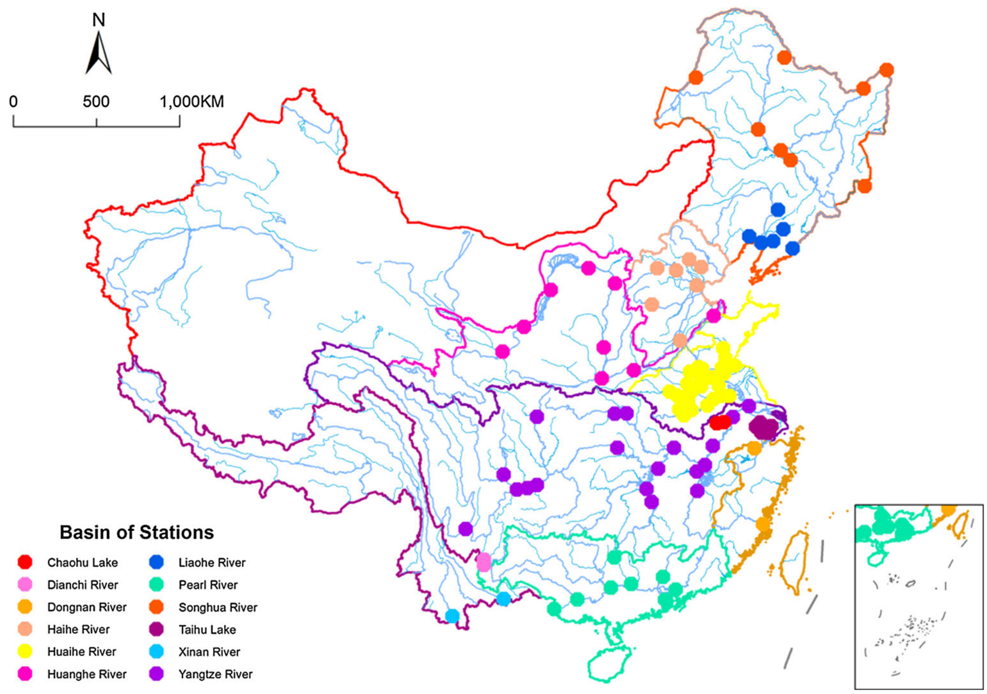


Fig. 2 The distribution of 100 automatic water quality monitoring stations

Table 1 Standard limits of indicators in the National Standard (GB3838-2002)

Indicators	Categories				
	Level I	Level II	Level III	Level IV	Level V
pH (unitless)	6–9				
DO (mg/L)≥	Saturation 90% (or 7.5)	6	5	3	2
COD (mg/L) ≤	2	4	6	10	15
NH_4^+-N (mg/L)≤	0.15	0.5	1	1.5	2

2.3 Method and Model

2.3.1 CCME WQI

The CCME WQI method was used to integrate the parameters into a single index to evaluate the surface water quality. The index yields a number between 0 (worst water quality) and 100 (best water quality), divided into five descriptive categories. The range of categories can be modified for every particular case of study (Terrado et al. 2010). The mathematical formulation is shown in Eq. (1) (Akkoyunlu and Akiner 2012; Terrado et al. 2010):

$$\mathrm{WQI} = 100 - \frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732} \quad (1)$$

where F_1 is the percentage of indicators that depart from their standard limits at least once in the monitoring samples, named as failed indicators, relative to the total number of indicators measured (Eq. 2). P is the number of failed indicators. N represents the total number of indicators.

$$\mathrm{F1} = \left(\frac{P}{N}\right) \times 100 \quad (2)$$

F_2 represents the percentage of monitoring samples that depart from the standard limits, named as failed samples (Eq. 3). Q is the number of failed samples. M represents the total number of samples.

$$\mathrm{F2} = \left(\frac{Q}{M}\right) \times 100 \quad (3)$$

F_3 represents the times by which failed samples values exceed their standard limits, using asymptotic function in order to yield a range between 0 and 100 (Eq. 4).

$$F3 = \left(\frac{q}{0.01 \times q + 0.01}\right) \quad (4)$$

The value of q is calculated by the formulation in Eq. (5). S_i is the number of times by which an sample concentration

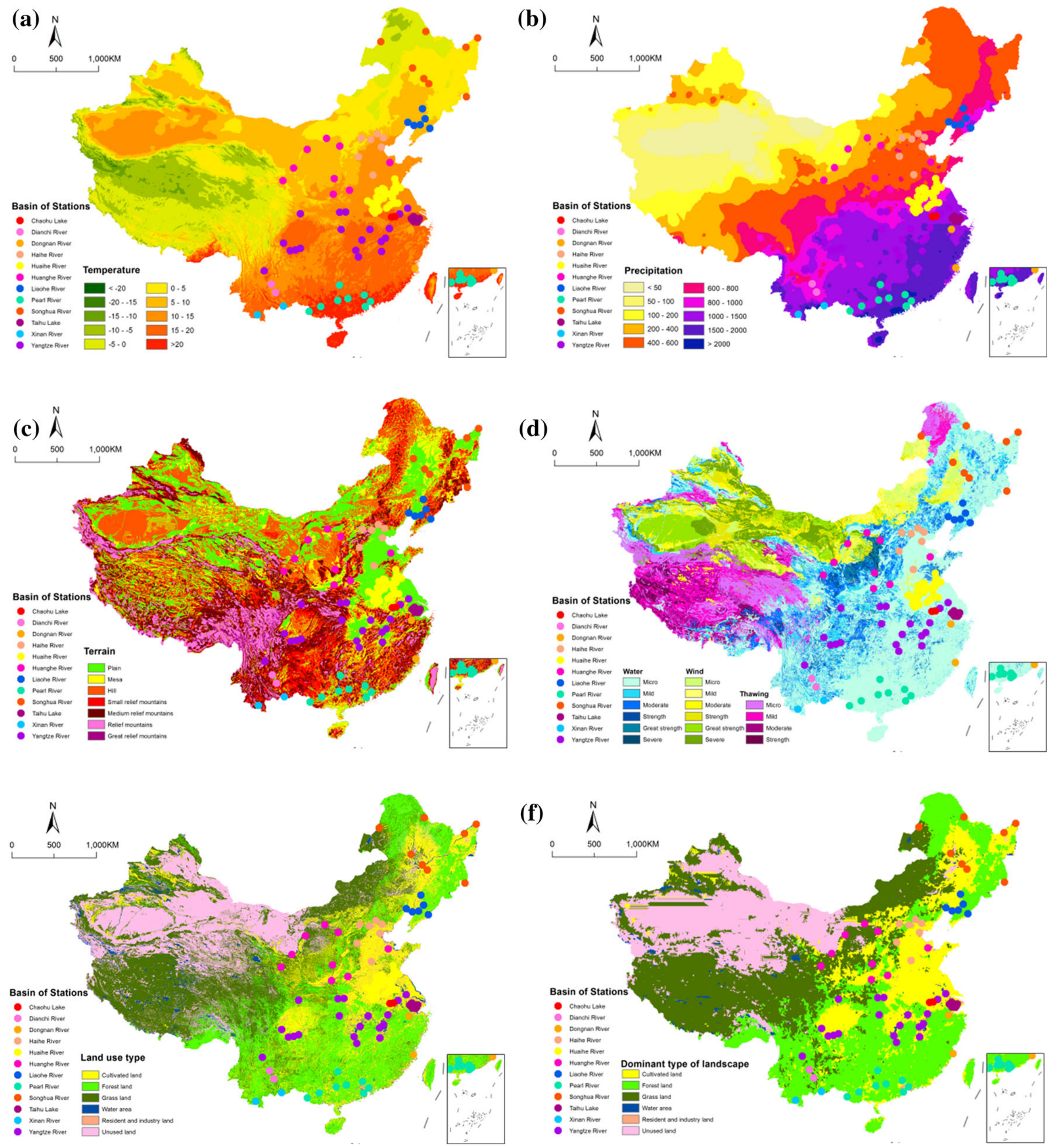


Fig. 3 The spatial distribution of geographic proxies. **a** Annual mean temperature, **b** annual mean precipitation, **c** terrain, **d** soil erosion, **e** land use type in 2010, **f** dominant type in sub-basin, **g** basin based on DEM, **h** population density in 2010, **i** GDP-per-capita in 2010

is greater than (or less than, when the standard limits are minimum).

$$q = \frac{\sum_{i=1}^{n} S_i}{M} \quad (5)$$

When the samples value must not exceed the standard limits (such as for COD and NH_4^+-N), S_i is calculated by the formulation in Eq. (6), where c_i is the failed samples value and c_s is the standard limits.

$$S_i = \frac{c_i}{c_s} - 1 \quad (6)$$

When the samples value must not fall below the standard limits (such as DO), S_i is calculated by the formulation in Eq. (7).

$$S_i = \frac{c_s}{c_i} - 1 \quad (7)$$

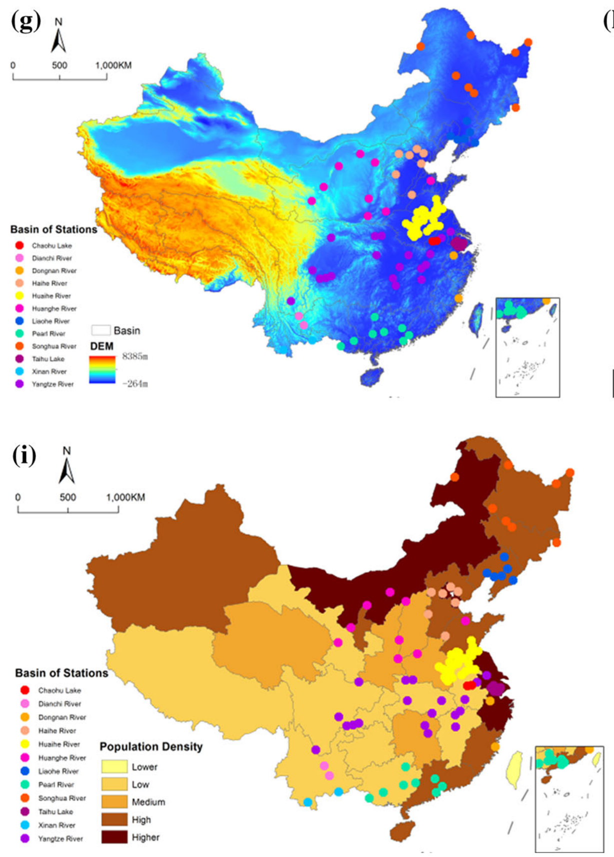


Fig. 3 continued

2.3.2 Geographical detector

The geographical detector method proposed by Wang et al. (2010) was used to compare the spatial consistency of surface water quality versus the geographical layers (e.g., temperature, precipitation, terrain, land use, land type, etc.) in which potential influence factors exist. Each geographical factor was divided into different strata; different strata have different attribute values. If one factor dominates the cause of surface water quality, the surface water quality will exhibit a spatial distribution similar to that of the geographical factor (Wang et al. 2010) and the variance of surface water quality within the strata of the geographical factor is less than that between the strata (Wang et al. 2016), that is, spatially stratified heterogeneity exists.

The *q*-statistic in the geographical detector can explore the spatially stratified heterogeneity of surface water quality in the stratum of a geographical factor and detect the extent to which the geographical factor explains the spatially stratified heterogeneity of surface water quality (Wang and Xu 2017). The mathematical formulation is shown in Eq. (8) (Wang et al. 2016; Wang and Xu 2017).

$$q = 1 - \frac{\sum_{h=1}^{L} N_h \sigma_h^2}{N\sigma^2} \tag{8}$$

The surface water quality was composed of N units and was stratified into h = 1, 2, …, L strata; stratum h is composed of N_h units; σ^2 and σ_h^2 express the variance of the population and the stratum, respectively.

The value of the *q*-statistic is within [0, 1]. When the *q* value approaches 1, the value of σ_h^2 is close to 0, which means that this factor has the same distribution as the surface water quality (Huang et al. 2014).

The interactive detector in the geographical detector software can be used to analyze the effect of the interaction of two or multiple factors on surface water quality. Table 2 shows the interactive results of two factors.[2]

The value of $q(X1 \cap X2)$ represents the explanatory power of the interaction of the two factors, X1 and X2, on surface water quality. The interactions between two factors are categorized as nonlinear weaken, weaken, binary enhance, independent, and nonlinear enhance, which depends on the relationship between $q(X1 \cap X2)$ and $q(X1), q(X2)$.

[2] http://www.geodetector.org/.

Table 2 Redefined interaction relationships

Description	Interaction
$\mathbf{q}(\mathbf{X1} \cap \mathbf{X2}) < Min(\mathbf{q}(X1), \mathbf{q}(\mathbf{X2}))$	Weaken, nonlinear
$\mathbf{Min}(\mathbf{q}(X1), \mathbf{q}(\mathbf{X2})) < \mathbf{q}(\mathbf{X1} \cap \mathbf{X2}) < Max(\mathbf{q}(X1), \mathbf{q}(\mathbf{X2}))$	Weaken, uni-
$\mathbf{q}(\mathbf{X1} \cap \mathbf{X2}) > \mathbf{Max}(\mathbf{q}(X1), \mathbf{q}(\mathbf{X2}))$	Enhance, bi-
$\mathbf{q}(\mathbf{X1} \cap \mathbf{X2}) = \mathbf{q}(X1) + \mathbf{q}(\mathbf{X2})$	Independent
$\mathbf{q}(\mathbf{X1} \cap \mathbf{X2}) > \mathbf{q}(X1) + \mathbf{q}(\mathbf{X2})$	Enhance, nonlinear

3 Results

3.1 Spatial–temporal variations in surface water quality

Surface water quality varies in different basins. Table 3 shows the variations of different indicators in different basins, including the maximum, minimum, and mean values of the indicators and water quality level in the basin which had the highest ratio in the monitoring stations. The surface water quality varies among basins, the water quality of Haihe River was serious, which was worse than level V; that of the Dianchi Lake was heavy, reaching level V; that of Huai River, Liaohe River, Songhua River and Taihu Lake was middle, reaching level IV; that of Chaohu Lake was mild, reaching level III; that of Huanghe River, South-east Rivers, South-west Rivers, Yangtze River, and Pearl River was good, ranked at level II. Figure 4 shows the annual spatial water quality level for stations and basins in this study.

Figure 5 is a histogram of the mean of the CCME-WQI grouped by watershed changing across seasons. Obviously, the value varies with season. The surface water quality was determined as follows: for the Chaohu Lake, Yangtze River, and Southeast Rivers, it was best in winter and worst in summer; for the Haihe River and Southwest Rivers, it was best in summer and worst in autumn; for the Yellow River and Liao River, it was better in summer and autumn and worse in winter and spring; for the Taihu Lake and Pearl River, it was better in spring and autumn and worst in summer; for the Dianchi Lake, Huai River, and Songhua River, it was best in spring and varied with season.

The annual mean values for COD, NH_4^+-N, DO, and pH in each station were contrasted with the standard limits of indicators in Table 1. The surface water quality was divided by different indicators into categories. Statistical analysis signified that the number of stations in which the surface water quality was determined by NH_4^+-N, COD, DO, and pH was 67, 73, 14, and 4, respectively. In addition, in 46 stations, the NH_4^+-N and COD categories were the same. Clearly, NH_4^+-N and COD were the key indicators determining the surface water quality.

3.2 *q*-statistic of geographical factors

The risk detector was used to analyze the effects of 8 different influential factors by season, including temperature, precipitation, terrain, soil erosion, land use type, GDP—per-capita, population density, and basin, on NH_4^+-N, COD, and CCME-WQI. The value of the *q*-statistic—reflecting the proportion of each geographical factor explained by the three indicators—is listed in Table 4.

Evaluation of the risk detector disclosed that the primary geographical factors are ranked by the value of the *q*-statistic for the annual means of the indicators as NH_4^+-N: population density (0.82) > precipitation (0.35) > terrain (0.24); COD: population density (0.46) > temperature (0.29) > basin (0.27); and the CCME-WQI: basin (0.49) > terrain (0.47) > dominant land use type (0.24). Season had variable influence on the geographical factors explaining the three indicators.

3.3 Interactive *q*-statistic of geographical factors

The interactive detector was used to disclose the interactive influence of geographical factors on the three indicators. The results for NH_4^+-N, COD, and the CCME-WQI, respectively, are designated in Tables 5, 6, and 7.

For the annual mean of the NH_4^+-N indicator, the *q* values for terrain, precipitation, and population density were 0.24, 0.08, and 0.82 respectively; however, the interactive value of population density and terrain was 0.88, and that of population density and precipitation was 0.87. For the annual mean of the COD indicator, *q* the values for terrain, precipitation, population density, and basin were 0.25, 0.21, 0.46, and 0.27, respectively; however, the interactive value of precipitation and terrain was 0.82, and that of population density and the basin was 0.80. For the annual mean of the CCME-WQI indicator, the *q* values for terrain, precipitation, and basin were 0.51, 0.14, and 0.49, respectively; however, the interactive value of precipitation and terrain was 0.73, and that of terrain and the basin was 0.75. These interactive values of the *q*-statistic appeared to be higher than any value for the *q*-statistic of solo factors. Most of the interactive results belonged to the types of binary enhancement and nonlinear enhancement.

Table 3 Descriptive statistics of indicators in 12 basins (unit: mg/L)

Basin	Indicators	Max	Min	Mean	Level	Basin	Indicators	Max	Min	Mean	Level
Chaohu Lake	DO	8.86	7.32	8.09	III	Liaohe River	DO	10.04	6.92	9.14	IV
	pH	7.83	7.60	7.71			pH	7.59	7.00	7.41	
	NH_4^+-N	0.53	0.32	0.42			NH_4^+-N	1.01	0.01	0.43	
	COD	4.56	4.27	4.41			COD	6.13	1.33	3.28	
Dianchi Lake	DO	8.57	6.13	7.35	V	Songhua River	DO	9.30	7.32	8.38	IV
	pH	8.76	7.95	8.35			pH	7.50	6.55	7.00	
	NH_4^+-N	0.52	0.42	0.47			NH_4^+-N	0.45	0.17	0.32	
	COD	10.48	9.09	9.78			COD	7.04	3.79	5.44	
South-east Rivers	DO	8.58	7.63	8.10	II	Taihu Lake	DO	8.83	4.05	7.39	IV
	pH	7.60	7.07	7.33			pH	7.88	7.08	7.57	
	NH_4^+-N	0.17	0.11	0.14			NH_4^+-N	1.29	0.20	0.54	
	COD	2.46	2.13	2.29			COD	7.30	3.19	4.72	
Haihe River	DO	10.95	2.66	7.91	Worse than V	South-west Rivers	DO	7.59	6.24	6.91	II
	pH	8.37	7.16	7.92			pH	8.01	7.67	7.84	
	NH_4^+-N	3.06	0.05	0.79			NH_4^+-N	0.33	0.18	0.25	
	COD	6.51	1.97	4.04			COD	2.71	1.64	2.17	
Huai River	DO	11.07	3.59	8.15	IV	Yangtze River	DO	10.50	6.04	8.30	II
	pH	8.33	7.16	7.70			pH	8.14	6.62	7.58	
	NH_4^+-N	1.37	0.19	0.57			NH_4^+-N	0.48	0.11	0.23	
	COD	8.86	0.95	4.78			COD	3.51	1.19	2.38	
Huanghe River	DO	10.34	3.17	8.20	II	Pearl River	DO	8.83	2.73	7.32	II
	pH	8.40	6.67	7.67			pH	7.79	7.00	7.41	
	NH_4^+-N	5.78	0.20	0.98			NH_4^+-N	1.20	0.14	0.34	
	COD	16.56	2.03	4.84			COD	3.17	1.23	1.73	

4 Discussion

The main purpose of this study was to exhibit the spatial and temporal distribution and variations of surface water quality, and to quantitatively analyze the influence factors of the surface water quality in China. The NH_4^+-N, COD and CCME-WQI are chosen as the commonly used indices in evaluating the surface water quality (Bouza-Deano et al. 2008; Jin et al. 2014; Alexakis et al. 2016; Zhao et al. 2016). Geographical detectors-based risk assessment and its application in the water quality study could be used to explore the power of influence factors to the water quality (Wang et al. 2016). The geographical detector method was used to identified that precipitation, fertilizer and elevation was the significant and powerful variables controlling the groundwater nitrate contamination (Gao et al. 2015; Nolan et al. 2015; Shrestha and Luo 2017).

The NH_4^+-N indicator reflects the contents of nutritional pollutants (Simeonov et al. 2003). The top three factors for NH_4^+-N concentration are population density, terrain, and precipitation. The explanatory power of them are 0.82, 0.35, and 0.24, respectively (Table 4). A research on the water quality in the Great Lakes coastal wetlands also reported that with human population related stress, these wetlands had higher ammonium concentration (Morrice et al. 2008). Population density reflects the discharge of domestic sewage and industrial effluents (Samal et al. 2012). Besides, agricultural and urban expansion were related to NH_4^+-N concentration because of leachate from intensive animal agriculture (Burkholder et al. 2006; Rothenberger et al. 2006). Terrain and precipitation represent surface runoff, and surface runoff affects the simultaneous diffusion of agricultural pesticides and chemical fertilizers (Wang et al. 2013). Moreover, soil erosion processes have a powerful relationship with nutrient losses in catchments (Panagopoulos et al. 2011). Linear correlation between the nutrient load of Miyun reservoir in Beijing and the density of people of the basin was also found to be statistically significant (Jiao et al. 2015). Precipitation can also speed up atmospheric deposition and affect surface water quality (Simeonov et al. 2003). The main source of NH_4^+ in the precipitation, which is influenced by climate and human activities, is NH_3 volatilized from agricultural fertilizer, animal and human excreta.

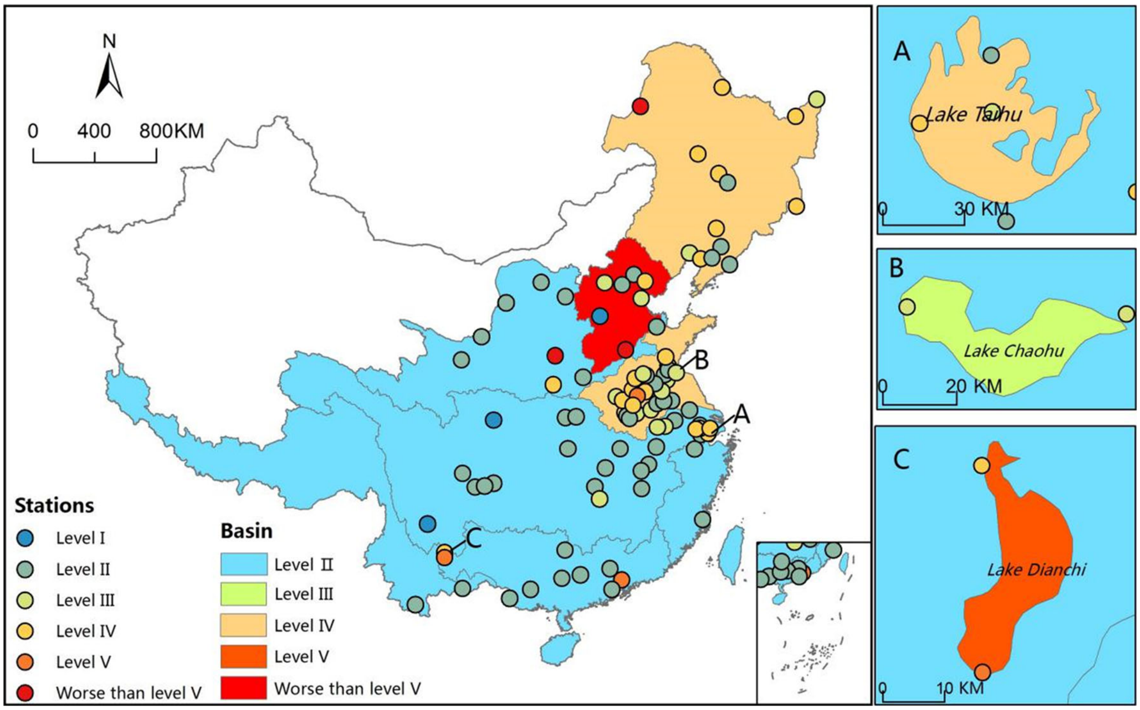
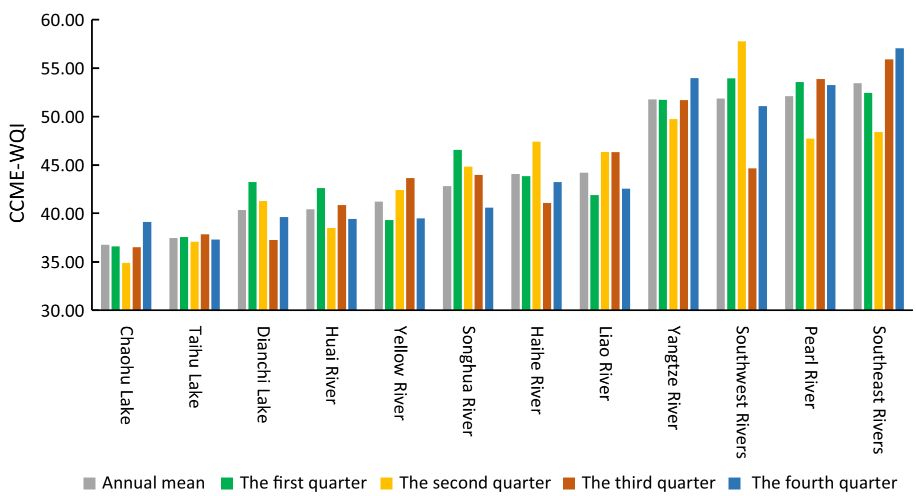


Fig. 4 Annual water quality level in 2015 for stations and basins in this study

Fig. 5 Seasonal means of the CCME-WQI



Pina-Ochoa and Alvarez-Cobelas (2009) demonstrated that NH_4^+-N might be desorbed with the infiltration, leak into the soil, enter the river and affect the surface water quality. In conclusion, domestic sewage, industrial effluents, agricultural pesticides, and chemical fertilizers contribute to the nutritional pollutants. The anthropogenic factors play more important roles in the surface water quality than natural factors.

The COD indicator reflects the contents of organic pollutants (Rahaman et al. 2015; Barakat et al. 2016). Hernandez-Romero et al. (2004) reported that high COD values in a tropical coastal wetland in southern Mexico were associated with mangrove- enriched organic matter. In this study, the top three factors influencing the COD are population density, temperature and basin. The explanatory power of them are 0.46, 0.29, and 0.27, respectively (Table 4). Population density is, to some degree, an indicator of integrative human activity influence, and is associated with urban areas, nutritional and organic pollutant loads. However, there was study finding that compared to population, water discharge is relatively correlated with water quality indices in the Adige basin (Diamantini et al. 2017). The reason might be that the high COD value is mostly attributed to the point pollution sources, such as,

Table 4 The q-statistic of geographical factors by season

Indicators	Season	A	B	C	D	E	F	G	H
NH_4^+-N	Spring	0.05	0.08	0.23	0.38	0.19	0.10	0.84	0.15
	Summer	0.11	0.07	0.21	0.14	0.14	0.03	0.43	0.19
	Autumn	0.11	0.08	0.23	0.13	0.10	0.01	0.43	0.21
	Winter	0.08	0.09	0.23	0.39	0.20	0.10	0.87	0.15
	Annual mean	0.08	0.35	0.24	0.08	0.19	0.08	0.82	0.16
COD	Spring	0.19	0.14	0.18	0.26	0.13	0.04	0.67	0.19
	Summer	0.35	0.26	0.27	0.08	0.06	0.03	0.24	0.38
	Autumn	0.32	0.21	0.23	0.06	0.03	0.04	0.28	0.33
	Winter	0.22	0.17	0.28	0.19	0.10	0.01	0.55	0.20
	Annual mean	0.29	0.14	0.25	0.21	0.07	0.02	0.46	0.27
CCME-WQI	Spring	0.10	0.10	0.40	0.05	0.17	0.03	0.14	0.41
	Summer	0.18	0.11	0.47	0.07	0.25	0.02	0.24	0.42
	Autumn	0.12	0.15	0.34	0.08	0.15	0.04	0.24	0.44
	Winter	0.20	0.14	0.51	0.13	0.25	0.13	0.21	0.49
	Annual mean	0.16	0.08	0.47	0.14	0.24	0.05	0.23	0.49

A temperature, *B* precipitation, *C* terrain, *D* soil erosion, *E* dominant land use type, *F* GDP-per-capita, *G* population density, *H* basin, ∩ = interaction

Table 5 Interactive seasonal q-statistic values of geographical factors for NH_4 + -N

Season	A	B	C	A ∩ C	B ∩ C
Spring	0.23	0.08	0.84	0.88	0.87
Summer	0.21	0.07	0.43	0.57	0.55
Autumn	0.23	0.08	0.43	0.59	0.58
Winter	0.23	0.09	0.87	0.90	0.91
Annual mean	0.24	0.08	0.82	0.88	0.87

A terrain, *B* precipitation, *C* population density, ∩ interaction

Table 7 Interactive seasonal q-statistic values of geographical factors for the CCME-WQI

Season	A	B	D	A ∩ B	A ∩ D
Spring	0.40	0.10	0.41	0.66	0.69
Summer	0.47	0.11	0.42	0.70	0.79
Autumn	0.34	0.15	0.44	0.61	0.73
Winter	0.51	0.14	0.49	0.79	0.76
Annual mean	0.51	0.14	0.49	0.73	0.75

A terrain, *B* precipitation, *D* Basin, ∩ = interaction

Table 6 Interactive seasonal q-statistic values of geographical factors for COD

Season	A	B	A ∩ B	C	D	C ∩ D
Spring	0.18	0.14	0.86	0.67	0.19	0.87
Winter	0.28	0.17	0.83	0.55	0.20	0.81
Annual mean	0.25	0.21	0.82	0.46	0.27	0.80

	A	B	A ∩ B	D	A ∩ D
Summer	0.27	0.26	0.74	0.38	0.80
Autumn	0.23	0.21	0.72	0.33	0.73

A terrain, *B* precipitation, *C* population density, *D* basin, ∩ = interaction

industrial waste and sewage discharge (Fucik et al. 2014), which is related to the basin area. In general, the larger the basin area was, the higher the discharge was. Additionally, urban land use was the important factor influencing COD in highly urbanized areas (Kannel et al. 2007). The urban land use is a factor associating with population. Additionally, it is noticed that temperature has impact on COD, because physio-chemical and microbial processes in surface water play essential roles in organic matter and nutrient distribution and transport (Zhai et al. 2014; Rahaman et al. 2015; Wu et al. 2018).

The result showed that the explanatory power of population density to COD is lower than that to NH_4^+-N, because of the greater content of NH_4^+-N in domestic sewage coming from human and animal excreta. Besides, this study demonstrated that the q-statistic value of the basin for COD is higher than that for NH_4^+-N. It means that the distribution of organic pollutants in the same sub-basin is more homogeneous than nutritional pollutants, which may be related to the activities of aquatic organisms. The content of nutritional pollutants in algal regions is different from that having no algae, while the organic pollutants have no constraint because bacteria are everywhere.

For seasonal variations, the population density and soil erosion have more explanatory power in winter and spring than in summer and autumn for the NH_4^+-N and COD indicators (Table 4). In winter and spring, relatively less precipitation lead to lower water discharge or runoff, which enhancing the role of human activity play in regulating pollutant (i.e. NH_4^+-N and COD). The pollutant source mainly derived from point source and low surface water discharge (McKee et al. 2001). Therefore, population density is more powerful for NH_4^+-N and COD in winter and spring. On one hand, the pollutant concentration may be diluted due to precipitation. On the other hand, precipitation causes the serious soil erosion and the increase of pollutant content. The soil erosion also reflects the state of surface runoff, which can carry agricultural pesticides and chemical fertilizers into the surface water. There are seasonal differences in the agricultural pesticide and chemical fertilizer use due to seasonal farming. In the summer and autumn, the entire country is engaged in crop management and harvesting activities, which means large amounts of pesticides and chemical fertilizers are added to farmland, diffusing along the surface runoff. Similar to our study, the research by Shiddamallayya and Pratima (2008) also present that the maximum concentrations of chemical oxygen demand were found during summer season when the nutrient load was high due to precipitation. In the winter and spring, farmland is left fallow in the north, causing surface runoff to carry fewer nutritional pollutants and organic pollutants, which is similar in the south during the summer and autumn. This is in accordance with the work done by Mckee et al. (2001), which also indicated that nutritional pollutant concentrations were significantly related to population density and agricultural activities.

The differences in the distribution of pesticides and chemical fertilizers used in the winter and spring were more similar to the distribution of soil erosion than in summer and autumn according to the q-statistic results (Table 4). Therefore, the seasonal variations of pesticides and chemical fertilizers caused the seasonal distribution of NH_4^+-N and COD, thereby influencing the surface water quality. These results reflect the same seasonal variation pattern of agricultural activities and the seasonal pattern of the natural and human factors influence on the pollutant.

For the CCME-WQI, the annual and seasonal changes were both smaller than that of other indicators, because it integrated the variation of other parameters, and elucidate the overall water quality (CCME 2001, 2003; Hong et al. 2008; Alexakis et al. 2016; Sethy et al. 2017). The top two influencing factors on the season were the basin and terrain. The data indicated that the distribution of the CCME-WQI was more homogeneous in the basin than that of other indicators. The CCME-WQI can be used to compare the differences in water quality at the macro-scale, while the single indicator destroys the limit of basin and reflects the difference at the micro-scale. Moreover, recent reports suggest that the CCME-WQI with modified categorization scheme is believed to support water managers to integrate and know well the picture of overall water quality (Boyacioglu 2010).

5 Conclusions

Quantitatively understanding the relationship between water quality and influential factors should be helpful for the management of water quality. Geographical Detector can be used to detect influential factors of surface water quality at national scale. In China, the top three factors influencing the nutrient pollutants are the population density, terrain, and precipitation, whose explanatory powers are 0.82, 0.35, and 0.24, respectively. They are of enhanced interaction relationships. The top three factors influencing the organic pollutants are population density, temperature, and basin, whose explanatory powers are 0.46, 0.29, and 0.27, respectively. The relationship among the factors is a nonlinear enhanced interaction. Taken together, anthropogenic factors influence surface water quality two to three times more than natural factors. Consequently, water quality management should predominantly focus on controlling human activity related pollution emission.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant Nos. 41771434, 41531179). Project of Cultivation and Development for Science and Technology Innovation Base of Beijing (Z161100005016110).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Akkoyunlu A, Akiner ME (2012) Pollution evaluation in streams using water quality indices: a case study from Turkey's Sapanca Lake Basin. Ecol Indic 18:501–511. https://doi.org/10.1016/j.ecolind.2011.12.018

Alexakis D, Tsihrintzis VA, Tsakiris G, Gikas GD (2016) Suitability of water quality indices for application in lakes in the mediterranean. Water Resour Manag 30:1621–1633

Barakat A, El Baghdadi M, Rais J, Aghezzaf B, Slassi M (2016) Assessment of spatial and seasonal water quality variation of Oum Er Rbia River (Morocco) using multivariate statistical techniques. Int Soil Water Conserv Res 4(4):284–292. https://doi.org/10.1016/j.iswcr.2016.11.002

Bouza-Deano R, Ternero-Rodriguez M, Fernandez-Espinosa AJ (2008) Trend study and assessment of surface water quality in the Ebro River (Spain). J Hydrol 316:227–239
Boyacioglu H (2010) Utilization of the water quality index method as a classification tool. Environ Monit Assess 167(1–4):115–124
Burkholder JM, Dickey DA, Kinder CA, Reed RE, Mallin MA, McIver MR, Cahoon LB, Melia G, Brownie C, Smith J, Deamer N, Springer J, Glasgow HB, Toms D (2006) Comprehensive trend analysis of nutrients and related variables in a large eutrophic estuary: a decadal study of anthropogenic and climatic influences. Limnol Oceanogr 51(1):463–487
Canadian Council of Ministers of the Environment (CCME) (2001) Canadian water quality guidelines for the protection of aquatic life. CCME water quality index 1.0, User's Manual, Winnipeg. http://www.ccme.ca/assets/pdf/wqi_usermanualfctsht_e.pdf. Accessed 17 Apr 2015
Canadian Council of Ministers of the Environment (CCME) (2003) CCME Water quality index workshop,November 24–25, Halifax, NS, Canada. Canadian Council of Ministers of the Environment Winnipeg, MB. http://www.ccme.ca/initiatives/water.html?categoryid=82. Accessed 17 Apr 2015
Cao X, Sun J (2012) The impact of landscape pattern on water quality in Lake Dianchi watershed. Ecol Environ Sci 21(2):364–369
Chen X, Zhou W, Pickett ST, Li W, Han L (2016) Spatial-temporal variations of water quality and its relationship to land use and land cover in Beijing, China. Int J Environ Res Public Health 13(5):pii: E449. https://doi.org/10.3390/ijerph13050449
Diamantini E, Lutz SR, Mallucci S, Majone B, Merz R, Bellin A (2017) Driver detection of water quality trends in three large European river basins. Sci Total Environ 612:49–62
Dyer F, Elsawah S, Croke B et al (2014) The effects of climate change on ecologically-relevant flow regime and water quality attributes. Stoch Environ Res Risk Assess 28(1):67–82
Fucik P, Novak P, Zizala D (2014) A combined statistical approach for evaluation of the effects of land use, agricultural and urban activities on stream water chemistry in small tile-drained catchments of south Bohemia, Czech Republic. Environ Earth Sci 72(6):2195–2216
Gao BB, Wang JF, Fan HM, Xu K, Hu MG, Chen ZY (2015) A stratified optimization method for a multivariate marine environmental monitoring network in the Yangtze River estuary and its adjacent sea. Int J Geogr Inf Sci 29(8):1332–1349
Ha H, Stenstrom MK (2003) Identification of land use with water quality data in stormwater using a neural network. Water Res 37(17):4222–4230. https://doi.org/10.1016/S0043-1354(03)00344-0
Hernandez-Romero AH, Tovilla-Hernandez C, Malo EA, Bello-Mendoza R (2004) Water quality and presence of pesticides in a tropical coastal wetland in southern Mexico. Mar Pollut Bull 48(11–12):1130–1141
Hong SH, Yim UH, Shim WJ, Oh JR, Viet PH, Park PS (2008) Persistent organochlorine residues in estuarine and marine sediments from Ha Long Bay, Hai Phong Bay, and Ba Lat Estuary. Vietnam. Chemosphere 72(8):1193–1202
Huang J, Ho M, Du P (2011) Assessment of temporal and spatial variation of coastal water quality and source identification along Macau peninsula. Stoch Environ Res Risk Assess 25(3):353–361
Huang J, Wang J, Bo Y, Xu C, Hu M, Huang D (2014) Identification of health risks of hand, foot and mouth disease in China using the geographical detector technique. Int J Environ Res Public Health 11(3):3407–3423. https://doi.org/10.3390/ijerph110303407
Jiao J, Du PF, Lang C (2015) Nutrient concentrations and fluxes in the upper catchment of the Miyun Reservoir, China, and potential nutrient reduction strategies. Environ Monit Assess 187(3):110
Jin ZF, Chen LX, Li FL, Pan ZY, Jin MT (2014) Effects of water transfer on water quality and estimation of the pollutant fluxes from different sources into West Lake, Hangzhou City, China. Environ Earth Sci 73:1091–1101
Kannel PR, Lee S, Kanel SR, Khan SP, Lee YS (2007) Spatial-temporal variation and comparative assessment of water qualities of urban river system: a case study of the River Bagmati (Nepal). Environ Monit Assess 129:433–459
Li W, Yang M, Liang Z et al (2013) Assessment for surface water quality in Lake Taihu Tiaoxi River Basin China based on support vector machine. Stoch Environ Res Risk Assess 27(8):1861–1870
Liu L, Wu L (2016) Review on surface water quality index assessment method in foreign countries. Water Resour Prot 32(1):86–90
McKee LJ, Eyre BD, Hossain S, Pepperell PR (2001) Influence of climate, geology and humans on spatial and temporal nutrient geochemistry in the subtropical Richmond River catchment, Australia. Mar Freshw Res 52(2):235–248
Ministry of Environmental Protection (2016) China environmental status bulletin 2015. Environ Prot 7:6
Misaghi F, Delgosha F, Razzaghmanesh M, Myers B (2017) Introducing a water quality index for assessing water for irrigation purposes: a case study of the Ghezel Ozan River. Sci Total Environ 589:107–116. https://doi.org/10.1016/j.scitotenv.2017.02.226
Morrice JA, Danz NP, Regal RR, Kelly JR, Niemi GJ, Reavie ED, Hollenhorst T, Axler RP, Trebitz AS, Cotter AM, Peterson GS (2008) Human influences on water quality in Great Lakes coastal wetlands. Environ Manag 41(3):347–357
Nazeer S, Hashmi MZ, Malik RN (2014) Heavy metals distribution, risk assessment and water quality characterization by water quality index of the River Soan, Pakistan. Ecol Ind 43:262–270
Nienie AB, Sivalingam P, Laffite A, Ngelinkoto P, Otamonga J-P, Matand A, Mulaji CK, Mubedi JI, Mpiana PT, Poté J (2017) Seasonal variability of water quality by physicochemical indexes and traceable metals in suburban area in Kikwit, Democratic Republic of the Congo. Int Soil Water Conserv Res 5(2):158–165. https://doi.org/10.1016/j.iswcr.2017.04.004
Nolan BT, Fienen MN, Lorenz DL (2015) A statistical learning framework for groundwater nitrate models of the Central Valley, California, USA. J Hydrol 531:902–911
Ouyang Y, Nkedi-Kizza P, Wu QT, Shinde D, Huang CH (2006) Assessment of seasonal variations in surface water quality. Water Res 40(20):3800–3810. https://doi.org/10.1016/j.watres.2006.08.030
Panagopoulos Y, Makropoulos C, Baltas E, Mimikou M (2011) SWAT parameterization for the identification of critical diffuse pollution source areas under data limitations. Ecol Model 222(19):3500–3512
Phung D, Huang C, Rutherford S, Dwirahmadi F, Chu C, Wang X, Nguyen M, Nguyen NH, Do CM, Nguyen TH, Dinh TA (2015) Temporal and spatial assessment of river surface water quality using multivariate statistical techniques: a study in Can Tho City, a Mekong Delta area, Vietnam. Environ Monit Assess 187(5):229. https://doi.org/10.1007/s10661-015-4474-x
Pina-Ochoa E, Alvarez-Cobelas M (2009) Seasonal nitrogen dynamics in a seepage lake receiving high nitrogen loads. Mar Freshw Res 60(5):335–445
Rahaman SMB, Rahaman MS, Ghosh AK, Gain D, Biswas SK (2015) A spatial and seasonal pattern of water quality in the Sundarbans River systems of Bangladesh. J Coastal Res 31(2):390–397
Rothenberger MB, Burkholder JM, Brownie C (2006) Long-term effects of changing land use practices on surface water quality in a coastal river and lagoonal estuary. Environ Manag 44(3):505–523
Samal NR, Saha P, Roy PK, Biswasroy M, Venkat Ramana R, Mazumdar A (2012) Limnological analysis of an urban polluted

lake in Bangalore city in India. Desalin Water Treat 30(1–3):217–228. https://doi.org/10.5004/dwt.2011.2025

Seeboonruang U (2012) A statistical assessment of the impact of land uses on surface water quality indexes. J Environ Manag 101:134–142. https://doi.org/10.1016/j.jenvman.2011.10.019

Sethy SN, Syed TH, Kumar A (2017) Evaluation of groundwater quality in parts of the Southern Gangetic Plain using water quality indices. Environ Earth Sci 76(3):116

Shen Z, Hou X, Li W, Aini G (2014) Relating landscape characteristics to non-point source pollution in a typical urbanized watershed in the municipality of Beijing. Landsc Urban Plan 123:96–107. https://doi.org/10.1016/j.landurbplan.2013.12.007

Shiddamallayya N, Pratima M (2008) Impact of domestic sewage on fresh water body. J Enviorn Biol 29(3):303–308

Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. Environ Model Softw 22(4):464–475. https://doi.org/10.1016/j.envsoft.2006.02.001

Shrestha A, Luo W (2017) Analysis of groundwater nitrate contamination in the Central Valley: comparison of the geodetector method, principal component analysis and geographically weighted regression. Int J Geo Inf 6(10):297

Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsa D, Anthemidis A, Sofoniou M, Kouimtzis T (2003) Assessment of the surface water quality in Northern Greece. Water Res 37(17):4119–4124. https://doi.org/10.1016/S0043-1354(03)00398-1

Sun R, Wang ZZ, Chen L, Wang WW (2013) Assessment OF SURFACE WATER QUALITY at large watershed scale: land-use, anthropogenic, and administrative impacts. JAWRA J Am Water Resour Assoc 49(4):741–752. https://doi.org/10.1111/jawr.12033

Sundaray SK, Panda UC, Nayak BB, Bhatta D (2006) Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of the Mahanadi river-estuarine system (India)–a case study. Environ Geochem Health 28(4):317–330. https://doi.org/10.1007/s10653-005-9001-5

Susanna TYT, Wenli C (2002) Modeling the relationship between land use and surface water quality. J Environ Manag 66:377–393. https://doi.org/10.1006/jema.2002.0593

Terrado M, Barceló D, Tauler R, Borrell E, Campos SD, Barceló D (2010) Surface-water-quality indices for the analysis of data generated by automated sampling networks. TrAC Trends Anal Chem 29(1):40–52. https://doi.org/10.1016/j.trac.2009.10.001

Varol M, Gökot B, Bekleyen A, Şen B (2012) Spatial and temporal variations in surface water quality of the dam reservoirs in the Tigris River basin, Turkey. CATENA 92:11–21. https://doi.org/10.1016/j.catena.2011.11.013

Wang J, Xu C (2017) Geodetector: principle and prospective. Acta Geogr Sin 72(01):116–134

Wang JF, Li XH, Christakos G, Liao YL, Zhang T, Gu X, Zheng XY (2010) Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. Int J Geogr Inf Sci 24(1):107–127. https://doi.org/10.1080/13658810802443457

Wang Y, Wang P, Bai Y, Tian Z, Li J, Shao X, Mustavich LF, Li B-L (2013) Assessment of surface water quality via multivariate statistical techniques: a case study of the Songhua River Harbin region. China. Journal of Hydro-environment Research 7(1):30–40. https://doi.org/10.1016/j.jher.2012.10.003

Wang J-F, Zhang T-L, Fu B-J (2016) A measure of spatial stratified heterogeneity. Ecol Ind 67:250–256. https://doi.org/10.1016/j.ecolind.2016.02.052

Whitehead PG, Wilby RL, Battarbee RW, Kernan M, Wade AJ (2009) A review of the potential impacts of climate change on surface water quality. Hydrol Sci J 54(1):101–123. https://doi.org/10.1623/hysj.54.1.101

Wilson CO (2015) Land use/land cover water quality nexus: quantifying anthropogenic influences on surface water quality. Environ Monit Assess 187(7):424. https://doi.org/10.1007/s10661-015-4666-4

Wu ZS, Wang XL, Chen YW, Cai YJ, Deng JC (2018) Assessing river water quality using water quality index in Lake Taihu Basin. China. Sci Total Environ 5(612):914–922

Zhai XY, Xia J, Zhang YY (2014) Water quality variation in the highly disturbed Huai River Basin, China from 1994 to 2005 by multi-statistical analyses. Sci Total Environ 496:594–606

Zhao W, Zhu X, Sun X, Shu Y, Li Y (2015) Water quality changes in response to urban expansion: spatially varying relations and determinants. Environ Sci Pollut Res Int 22(21):16997–17011. https://doi.org/10.1007/s11356-015-4795-x

Zhao YM, Qin YW, Zhang L, Zhang BH, Ma YQ (2016) Water quality analysis for the Three Gorges Reservoir, China, from 2010 to 2013. Environ Earth Sciences 75(17):1255

Zheng X (2012) The influence analysis of total phosphor on algae growing and algae growing on pH and dissolved oxygen in surface water. Instrum Anal Monit 3:43–45

Zhou Y, Ma J, Zhang Y, Qin B, Jeppesen E, Shi K, Brookes JD, Spencer RGM, Zhu G, Gao G (2017) Improving water quality in China: environmental investment pays dividends. Water Res 118:152–159. https://doi.org/10.1016/j.watres.2017.04.035