Title
Water quality trend and change-point analyses using integration of locally weighted polynomial regression and segmented regression

Journal
Environmental Science and Pollution Research, 24(18)

ISSN
0944-1344

Authors
Huang, Hong
Wang, Zhenfeng
Xia, Fang
et al.

Publication Date
2017-06-01

DOI
10.1007/s11356-017-9188-x

Peer reviewed

Environ Sci Pollut Res
DOI 10.1007/s11356-017-9188-x

RESEARCH ARTICLE

Water quality trend and change-point analyses using integration of locally weighted polynomial regression and segmented regression

Hong Huang[1,2] · Zhenfeng Wang[1,2] · Fang Xia[1,2] · Xu Shang[1,2] · YuanYuan Liu[3] · Minghua Zhang[1,2] · Randy A. Dahlgren[1,2,4] · Kun Mei[1,2]

Received: 11 January 2017 / Accepted: 2 May 2017

Abstract Trend and change-point analyses of water quality time series data have important implications for pollution control and environmental decision-making. This paper developed a new approach to assess trends and change-points of water quality parameters by integrating locally weighted polynomial regression (LWPR) and segmented regression (SegReg). Firstly, LWPR was used to pretreat the original water quality data into a smoothed time series to represent the long-term trend of water quality. Then, SegReg was used to identify the long-term trends and change-points of the smoothed time series. Finally, statistical tests were applied to determine the significance of the long-term trends and change-points. The efficacy of this approach was validated using a 10-year record of total nitrogen (TN) and chemical oxygen demand (COD_{Mn}) from Shanxi Reservoir watershed in eastern China. Results showed that this approach was straightforward and reliable for assessment of long-term trends and change-points on irregular water quality datasets. The reliability was verified by statistical tests and practical considerations for Shanxi Reservoir watershed. The newly developed integrated LWPR-SegReg approach is not only limited to the assessment of trends and change-points of water quality parameters but also has a broad application to other fields with long-term time series records.

Keywords Water quality · Long-term trend assessment · Change-point analysis · Locally weighted polynomial regression · Segmented regression

Responsible editor: Marcus Schulz

✉ Kun Mei
meikun@iwaterlab.com

Hong Huang
Huanghongpanda@163.com

1 Southern Zhejiang Water Research Institute, Wenzhou Medical University, Wenzhou 325035, People's Republic of China

2 Key Laboratory of Watershed Environmental Science and Health of Zhejiang Province, Wenzhou 325035, People's Republic of China

3 School of Public Health and Management, Wenzhou Medical University, Wenzhou 325035, People's Republic of China

4 Department of Land, Air, and Water Resources, University of California, Davis, CA 95616, USA

Introduction

Many countries and regions in the world suffer from chronic water shortages, often from a scarcity of clean drinking water (Mukheibir 2010; Tiwari and Joshi 2012; Wang and Yu 2014). Therefore, water quality protection and remediation are very important aspects of sustainable social-economic development (Zhou et al. 2014). As a result, the protection and environmental remediation of drinking water sources have received considerable attention in research (Basu et al. 2014), since drinking water quality is closely related to human health (Corlin et al. 2016). In general, these research efforts include environmental policy and legislation planning (Syme and Nancarrow 2013; Xu et al. 2016), water quality standard development (Goncharuk 2013), and health risk assessment (Chiang et al. 2010; Houtman et al. 2014; Sun et al. 2015), water quality monitoring and modeling (Tropea et al. 2007; Sokolova et al. 2013; Chen et al. 2015), and protection and remediation technology (Zhang et al. 2011; Basu et al. 2014). An important aspect of water quality modeling for drinking water source protection is identifying changes in water quality trends and specific change-point timing, which are important information for environmental protection performance evaluation and

Published online: 23 May 2017

environmental decision-making (Bodo 1989; Chowdhury and Al-Zahrani 2014).

Water quality trend and change-point analyses commonly employ traditional statistical methods including parametric (i.e., linear regression, polynomial regression) and nonparametric (e.g., variants of the Mann-Kendall test) statistical methods (Donohue et al. 2001; Chang 2008; Kisi and Ky 2014). However, water quality is strongly influenced by both natural and human factors, and data records are often challenging to analyze from a statistical perspective as the data is often non-normally distributed, nonlinear fashion, non-monotonic trend, uneven time spacing, and with large seasonal variations. Linear regression may not be appropriate in cases of the data in nonlinear and non-normal distribution, while Mann-Kendall test may not be applicable in cases of non-monotonic trends and uneven sampling time spacing. Polynomial/logistic regression is based on global estimation, whereas data with these nonlinear patterns usually display local characteristics. Meanwhile, the water quality monitoring frequency is often at monthly and/or bimonthly which resulting in discrete and irregular time series data records at various sampling dates and sampling intervals. In general, trend assessment of statistically compromised time series datasets such as those that do not fully meet required statistical assumptions should not rely solely on abstract test statistics (Bodo 1989).

Appropriate graphing techniques may be a powerful data evaluation tool. As early as 1983, Chambers et al. (1983) noted that There is no single statistical tool as powerful as a well-chosen graph." The most classical graphical method is Cleveland's (1979, 1981) locally weighted scatterplot smoother (LOWESS), which was further developed by Cleveland and Devlin (1988) as the locally weighted polynomial regression (LWPR) procedure. LWPR can be directly applied to graphical analysis of 2D scatterplots for a series x_i versus corresponding times t_i. Soon afterwards, LOWESS was developed for seasonal-trend decomposition using loess (STL) (Cleveland et al. 1990). These graphical methods are based on various local smoothing techniques (Harding et al. 2016) and have been widely applied in environmental quality trend analysis including air quality (Li et al. 2014; Gong et al. 2015), water quality (Lee et al. 2010; Stow et al. 2015), etc. In general, graphical methods can play an important role in trend analysis in typical time series data, both as a diagnostic tool and as visual corroborative evidence when required assumptions for formal statistical tests are not met (Bodo 1989). However, graphical methods are not able to produce regression functions that can be mathematical described to determine the significance levels and define change-points (Liang 2014).

Segmented regression, also known as "piecewise regression" or "break-point regression", is a regression method applied to cases where independent variables are clustered into different intervals where the relationships between the variables are different. Nowadays, segmented regression (SegReg) has been widely used in trend and change-point analyses in research fields including medicine (Kazemnejad et al. 2014), hydrology (Shao and Campbell 2002), economics (Wu and Chang 2012), society (Mathews and Hamilton 2005), etc. In comparison with graphical methods, major advantages of SegReg are that the regression function is defined by a mathematical formula to describe the relationship, the significance levels of the trend can be determined statistically (Kazemnejad et al. 2014), and the change-point can be quantitatively defined (Taljaard et al. 2014). Since water quality time series are often statistically compromised, the direct application of SegReg might produce invalid results. However, SegReg would be an effective approach for trend and change-point analyses of water quality if the data could be appropriately pretreated.

Change-point detection is the identification of an abrupt variation in process behavior due to distributional or structural change, whereas a trend can be defined as the estimation of a gradual departure from past norms (Sharma et al. 2016). For water quality time series analysis, identifying changes in long-term trends is important, yet identifying specific change-points is also important. The objective of this paper is to provide an integrated LWPR-SegReg approach to analyze water quality trends and change-points. The efficacy of this approach was demonstrated by trend and change-point analyses for a 10-year record of key water quality parameters (TN and COD_{Mn}) in Shanxi Reservoir watershed of Zhejiang Province, China. Innovative and importance aspects of the integrated LWPR-SegReg approach include its ability to define change-points and trends visually and to quantitatively define the change-points and trends with statistical rigor.

Materials and methods

Locally weighted polynomial regression(LWPR) approach

Trend analysis determines whether the measured values of a water quality variable increase or decrease during a given time period (Naddafi et al. 2007). For a water quality series WQ_i, a basic linear trend analysis model is:

$$WQ_i = \alpha t_i + \beta + \varepsilon_i, \tag{1}$$

where t_i is time, a is the regression coefficient indicating the slope of the line, β is the regression constant, and ε_i is an irregular noise term.

LWPR usually employs a local linear polynomial regression model, but a local nonlinear regression model can also be used in some circumstances (Bodo 1989). For a water quality series WQ_i, LWPR is:

$$WQ_i = f(t_i) + \varepsilon_i, \tag{2}$$

where $f(t_i)$ is a smoothed function and ε_i is an irregular noise term.

The local polynomial fits are typically first (linear) or second (quadratic) order using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away (Cleveland and Devlin 1988). The traditional weight function is the tricube weight function; however, any other function that satisfies the properties can be used. More details concerning the LWPR approach can be found in Rajagopalan and Lall (1998) and Proietti and Luati (2011). In this research, LWPR models were fitted in R using the "loess" function available in the "stats" package.

Segmented regression (SegReg) approach

Segmented regression with segments separated by breakpoints (i.e., change-points) is useful for quantifying abrupt changes in water quality over time (Shao and Campbell 2002, Kazemnejad et al. 2014). The least squares method is applied separately to each segment; each regression line is optimized to minimize the sum of squares of the differences (SSD). For a water quality series WQ_i with m change-points (CPs), a segmented linear regression with $m + 1$ segments is depicted as:

$$\begin{cases} WQ_i = \alpha_1 t_i + \beta_1 \ (t_i \le CP_1) \\ WQ_i = \alpha_2 t_i + \beta_2 \ (CP_1 < t_i \le CP_2) \\ \vdots \\ WQ_i = \alpha_{m+1} t_i + \beta_{m+1} \ (t_i \le CP_m) \end{cases}, \tag{3}$$

where t_i is time, α_m is the regression coefficient indicating the slope of each line segment, and β_m is the regression constant; $\alpha > 0$ and $\alpha < 0$ indicate increasing and decreasing trends, respectively.

Statistical tests are then performed to ensure that the trend is significant. The commonly used indexes are the correlation coefficient squared (R^2) and P value. If no significant change-point is detected, a single regression without a change-point should be used. More details concerning SegReg can be found in Mathews and Hamilton (2005) and Wu and Chang (2012). In this research, SegReg models were developed in R using the "segmented" package.

Integration of LWPR and SegReg

Since water quality time series are often statistically compromised, neither LWPR nor SegReg is able to accomplish long-term trend and change-point analyses independently (Liang 2014). However, with the integration of LWPR and SegReg, it is possible to deal with many of the problems associated with statistical assumptions. For a water quality series WQ_i, the integration of LWPR and SegReg is performed in three major steps:

1) LWPR (Eq. 2) is used to pretreat the original data series (WQ_i) versus corresponding times t_i, into a new smoothed series (SM_i) representing the long-term trend of water quality versus corresponding times t_i
2) SegReg (Eq. 3) is used to quantify the relationships between the smoothed water quality time series SM_i versus corresponding times t_i
3) Statistical tests are used to estimate the significance of the long-term trends and change-points

The original water quality time series are often statistically compromised (i.e., contain long-term trends, non-monotonic trends and storm event/seasonal variations) and are sometimes irregular and discrete (i.e., uneven time spacing and/or low frequency in sampling), and the direct application of SegReg might produce invalid results, since the least squares method is applied separately to each segment, which requires a normal distribution of the data. In order to extract the long-term trend information, the pretreatment step is necessary to remove the compounded noise. By means of the pretreatment approach (i.e., LWPR), the non-monotonic trends and seasonal variations are removed, and the long-term trend information is preserved and provided for subsequent SegReg analysis, which is able to quantitatively identify the change-points and trends in water quality data with statistical rigor. The integration of LWPR and SegReg methods is necessary and innovative for achieving these goals.

Model testing

Study area and data collection

Shanxi Reservoir is a multi-annual regulating reservoir located in the uplands of the Feiyun River watershed in Zhejiang Province, China (Fig. 1). This reservoir is the major drinking water source of seven million people in the local region, with a total watershed area of 1529 km^2 and total storage capacity of 1.8×10^9 m^3. The region has a subtropical monsoon climate with mean annual precipitation of 1870 mm and temperature of 17 °C. Forest and cultivated lands account for 75 and 15% of the watershed land area, respectively (Mei et al. 2016). The population density within the watershed is 236 per/km^2, which is about 1.6 times the national average (143 per/km^2). Since polluting industries have been moved or closed, the remaining major pollution originates from non-point sources of

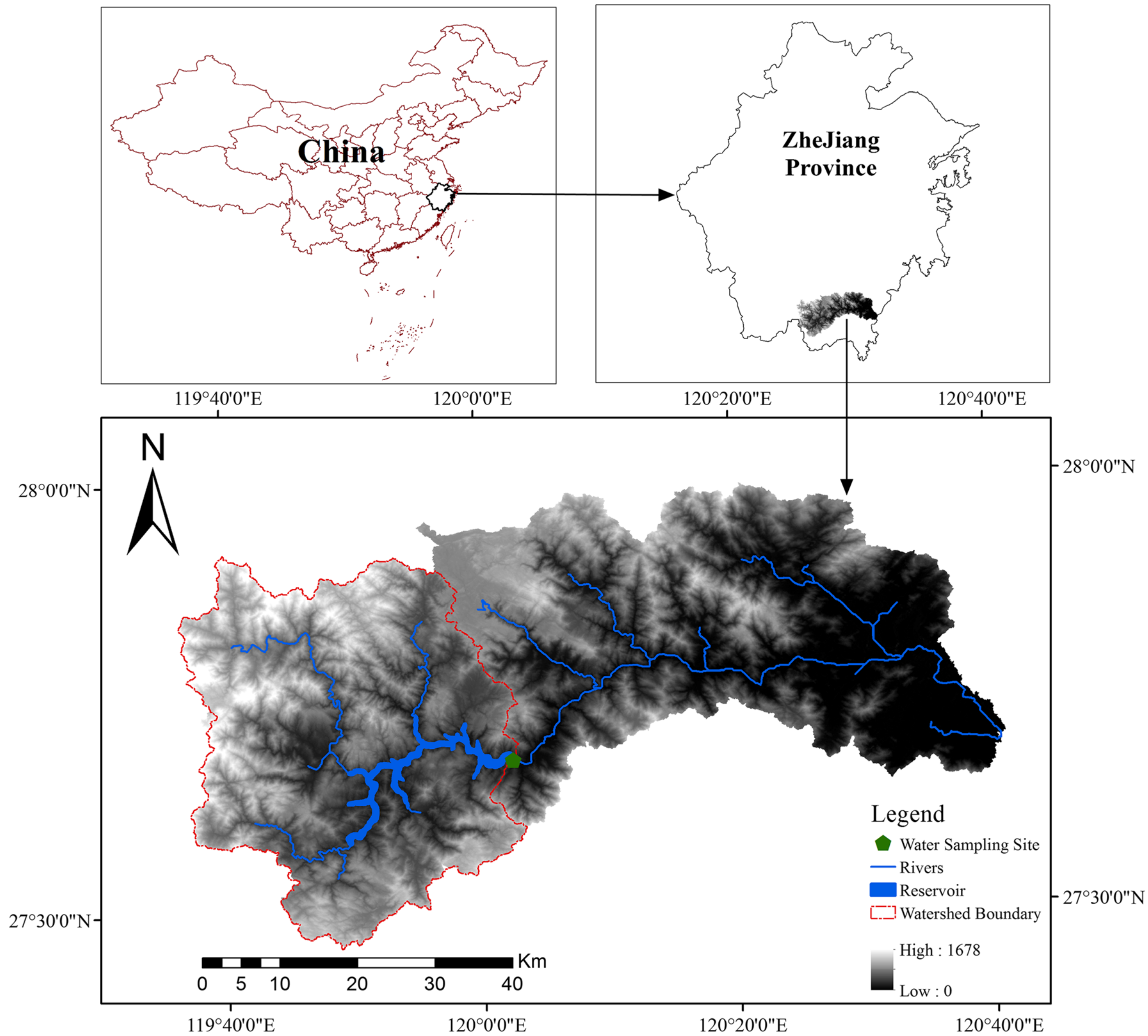

Fig. 1 Geographic location of study area and sampling site

atmospheric deposition, agricultural fertilizer, livestock, and domestic waste (Dong et al. 2016).

Total nitrogen (TN) and chemical oxygen demand (COD_{Mn}) are key water quality concerns in Shanxi Reservoir and were therefore selected as the focus of this study (Dong et al. 2016; Mei et al. 2016). Reservoir water quality was monitored by the local Water Resources Bureau on an approximately monthly basis during 2005–2014 (Fig. 1); however, the sampling date was often different in each month. COD_{Mn} was measured by the acid permanganate method. TN was measured following alkaline potassium persulfate digestion using an UV spectrophotometer. These water quality parameters displayed both long-term non-monotonic trends and seasonal variations, since water quality was influenced by both natural and human factors (Fig. 2). Major pollutants for Shanxi Reservoir originate from non-point sources, and the seasonal variation of water quality may be influenced by runoff fluctuation, since the study area has a subtropical monsoon climate with distinct wet and dry seasons (Huang et al. 2014a). However, there was no significant relationship between monthly average inflow runoff rate and monthly discrete TN and COD_{Mn} concentrations (Fig. 3).

In general, the transport and transformation of nitrogen in this large reservoir is very complex, and the seasonal variation in TN concentration is impacted by many factors requiring further research. The long-term trends, non-monotonic trends, and seasonal variations in the original water quality time series data for both TN and COD_{Mn} concentrations highlight the importance of smoothing the time series prior to further analysis. The efficacy of the integrated LWPR-SegReg approach was assessed by trend and change-point analyses of TN and COD_{Mn} from the Shanxi Reservoir water quality dataset.

Efficiency of the LWPR-SegReg approach

The distributional or structural changes of a time series result from changes in the data distribution patterns or changes in the distribution of parameters, including changes in mean value,

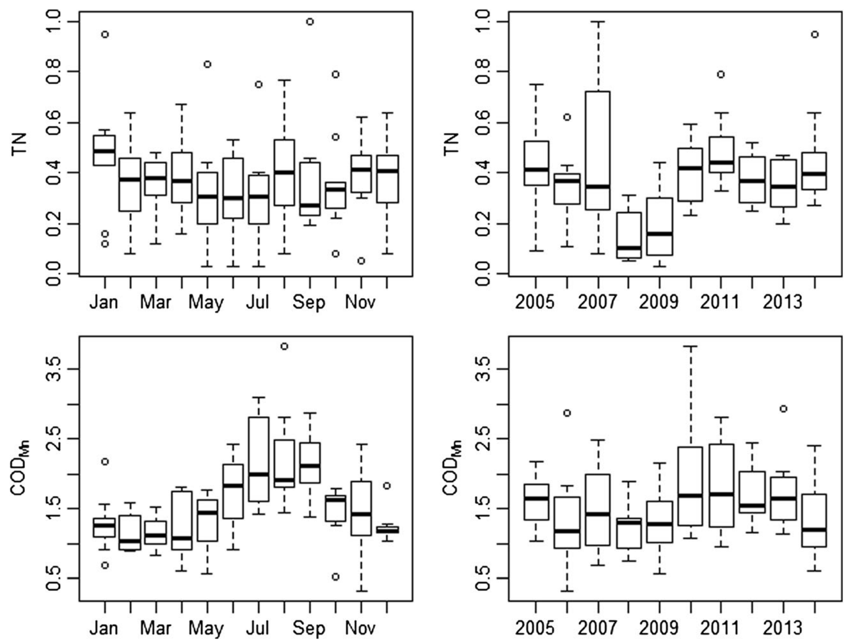

Fig. 2 Annual and monthly variations of TN and COD_{Mn} (mg L^{-1}). *Circles* in box plots are outliers, and *lines* in the graphs indicated values are at minimum, 25% percentile, median, 75% percentile, and maximum values

variance, and trends. Using the LWPR approach, the original monthly irregular and discrete TN and COD_{Mn} time series were smoothed by LWPR and then regressed by SegReg. The LWPR-SegReg results for the TN and COD_{Mn} time series are shown in Fig. 4 and Fig. 5, respectively. Using LWPR, the irregular and discrete original TN and COD_{Mn} time series (blue triangles) were smooth into a new time series (green circles), and then the smoothed time series were used for SegReg resulting in the segment lines (red lines) (Fig. 4 and Fig. 5; model parameters are listed in Table 1). The smoothed new time series generated from LWPR revealed that the TN and COD_{Mn} series showed non-monotonic trends within the 2005–2014 record, yet the change-points cannot be exactly defined from the graphs (Fig. 4 and Fig. 5).

The smoothed time series for TN and COD_{Mn} were divided into three segments (red lines) with two change-points (Fig. 4 and Fig. 5); adjusted R^2 were >0.99 and P values were <0.001 demonstrating excellent model fit. The first change-points for the TN and COD_{Mn} time series were 2008.567 and 2008.336, and the slopes of the first segments were −0.047 and −0.040, respectively (Table 1). This indicates that TN and COD_{Mn} showed small decreasing trends before the first half of 2008. The second change-points for the TN and COD_{Mn} time series were 2011.738 and 2011.631, and the slopes of the second segments were 0.094 and 0.130, respectively (Table 1). This indicates that TN and COD_{Mn} showed small increasing trends from the second half of 2008 to the first half of 2011. The slopes of the third segments for the TN and COD_{Mn} time series were −0.052 and −0.133, respectively (Table 1), which indicate decreasing trends since the second half of 2011. Based on statistical tests (Table 1), we conclude that the long-term trends and change-points were significant (Shao and Campbell 2002; Mathews and Hamilton 2005; Wu and Chang 2012; Kazemnejad et al. 2014).

In general, the LWPR approach is able to extract long-term trend information from statistically compromised water quality time series (WQ_i), providing a smoothed data series (SM_i) for subsequent analysis (Bodo 1989; Stow et al. 2015). Further, the subsequent SegReg is able to quantify the relationships between the smoothed water quality time series (SM_i) versus the corresponding times (t_i). That is the powerful advantage of the LWPR-SegReg approach. The practical

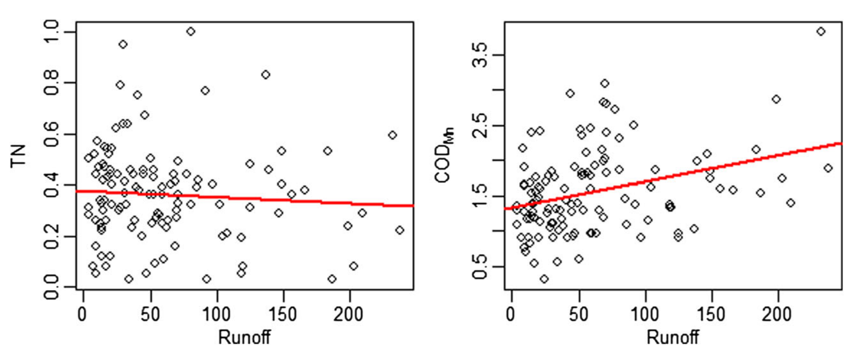

Fig. 3 Monthly average inflow rate (m^3 s^{-1}) to Shanxi Reservoir vs TN and COD_{Mn} concentrations (mg L^{-1})

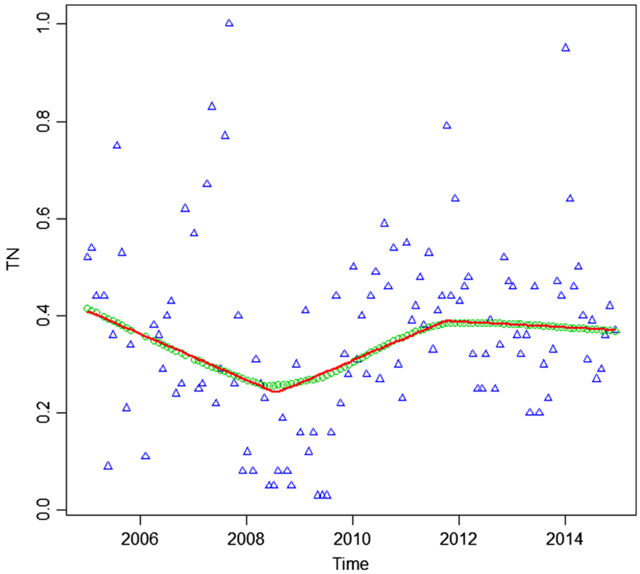

Fig. 4 LWPR-SegReg of monthly irregular and discrete TN concentrations. *Blue triangles* are original TN data, *green circles* are pretreated TN time series derived from LWPR, and *red line* is the SegReg results of the pretreated TN time series

value of this research was to examine the efficacy of past pollution control practices in the watershed, as well as to predict future trends. The existence of change-points indicates that the change in water quality trends within the time series was non-monotonic, and the corresponding specific times of the change-points were identified. Whether the water quality was improved or deteriorated can be identified according to the slopes of the segments derived from the LWPR-SegReg approach, and future trends can be predicted from the slope of the latest segment. Having demonstrated the efficacy of the

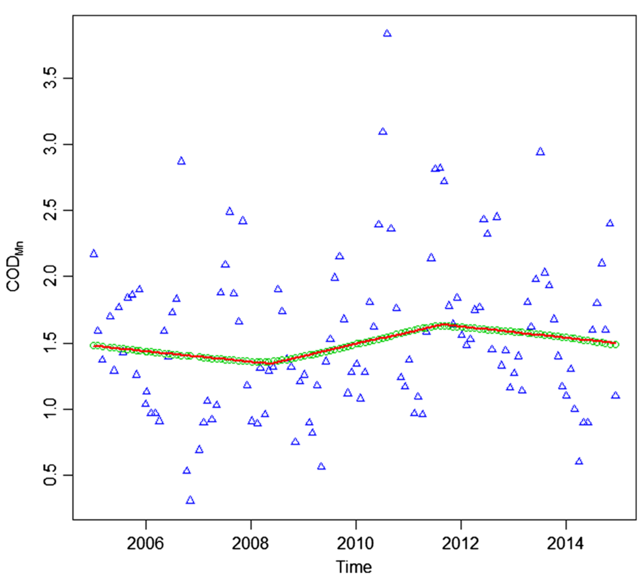

Fig. 5 LWPR-SegReg of monthly irregular and discrete COD_{Mn} concentrations. *Blue triangles* are original COD_{Mn} data, *green circles* are pretreated COD_{Mn} time series derived from LWPR, and *red line* is the SegReg results of the pretreated COD_{Mn} time series

Table 1 SegReg model parameters for pretreated TN and COD_{Mn} time series data

Model	Parameter	Estimated	Std. err	*P* value
TN	α1	−0.047	0.0006	<0.001
	α2	0.094	0.0009	
	α3	−0.052	0.001	
	CP 1	2008.567	0.019	
	CP 2	2011.738	0.036	
	Adjusted R^2	0.993		
COD_{Mn}	α1	−0.040	0.0007	<0.001
	α2	0.130	0.001	
	α3	−0.133	0.001	
	CP 1	2008.336	0.016	
	CP 2	2011.631	0.015	
	Adjusted R^2	0.997		

LWPR-SegReg approach, the reliability is analyzed in the following section.

Reliability of the LWPR-SegReg approach

To further confirm the reliability of the integrated LWPR-SegReg approach, the commonly utilized STL method was applied for trend analysis of the TN and COD_{Mn} time series from Shanxi Reservoir. More details concerning the theory and application of STL can be found in Cleveland et al. (1990) and Liang (2014), respectively. STL analysis was performed in R using the "forecast" package. The time interval between water quality data points was assumed to be consistent for STL analysis purposes. Using STL, the original data series were decomposed into seasonal, trend, and residual patterns, respectively. STL analysis indicated that TN concentration trends in Shanxi Reservoir were generally decreasing before 2009, increasing during the 2009 and 2010 period, and then decreasing after 2010 (Fig. 6). COD_{Mn} concentration trends followed an identical pattern to that of TN during the study period (Fig. 7).

In comparison to the LWPR-SegReg approach, the STL method was unable to specifically identify the timing of the change-points, and the trends derived from STL fluctuated within each time period. This results from STL holding the seasonal patterns the same throughout the whole time period (Cleveland et al. 1990); however, the seasonal pattern is somewhat variable from year-to-year. For LWPR, the seasonal patterns were designed to be functions of time and were removed locally according to their corresponding times (Bodo 1989; Stow et al. 2015) resulting in the smoothed data trends (green circles in Fig. 4 and Fig. 5) for subsequent SegReg analysis. However, the non-monotonic change trends for the TN and COD_{Mn} time series determined by LWPR-SegReg and STL

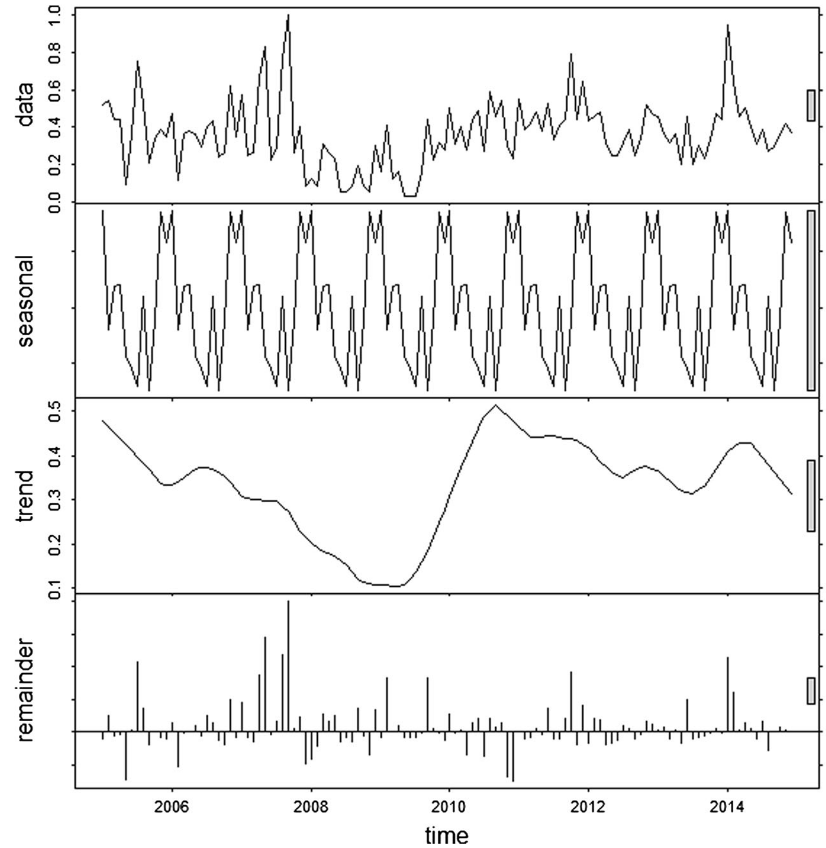

Fig. 6 STL analysis of monthly irregular and discrete TN concentrations (mg L^{-1})

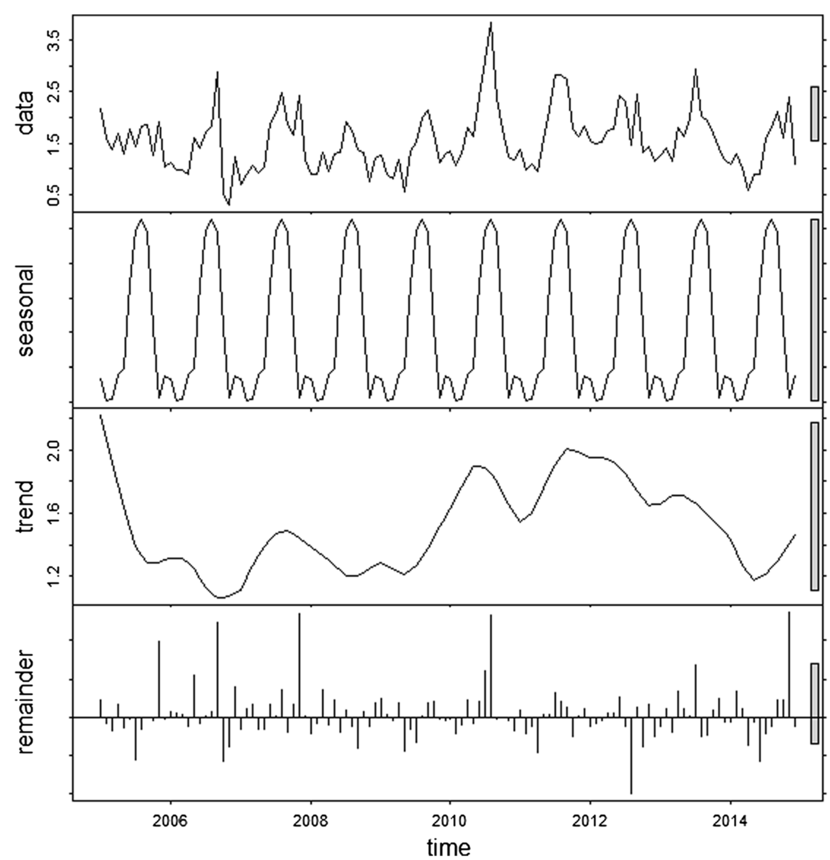

Fig. 7 STL analysis of monthly irregular and discrete COD_{Mn} concentrations (mg L^{-1})

were generally consistent, which further supports the reliability of the new LWPR-SegReg approach.

The integrated LWPR-SegReg approach indicated that TN and COD_{Mn} concentrations in Shanxi Reservoir had three different trends that decreased before the first half of 2008, increased during the second half of 2008 to the first half of 2011, and then decreased after the second half of 2011 (Fig. 4 and Fig. 5). Firstly, the initial increase in TN and COD_{Mn} concentrations in Shanxi Reservoir (2005 to first half of 2008) was due to processes associated with the first filling of the reservoir that began in 2000 and reached normal operating water levels in 2005 (Shi 2010). Once the runoff was captured within the reservoir, various pollutants were retained, recycled, or lost due to processes within the reservoir (e.g., sedimentation, biological uptake/transformation). Being a large multi-annual regulating reservoir, the initial self-cleaning (assimilative) capacity was substantial (Liu et al. 2013), and therefore, TN and COD_{Mn} concentrations decreased slowly in the first several years. Secondly, TN and COD_{Mn} concentrations increased slowly in the second period of the record (the second half of 2008 to the first half in 2011). Once the reservoir's assimilation capacity was exceeded (Liu et al. 2013), the water quality deteriorated as the pollution loads exceeded the reservoir's assimilation capacity, such as due to recycling of nutrients from the sediments to the water column (Wei et al. 2009). A previous study showed a deterioration of water quality due to cyanobacteria blooms in some areas of Shanxi Reservoir after 2008, particularly in 2010, which caused serious concerns for drinking water operations (Shi 2010). These algal blooms were consistent with the increasing trends in TN and COD_{Mn} concentrations during the second half of 2008 to the first half in 2011. Thirdly, TN and COD_{Mn} concentrations decreased after the second half of 2011. The algal blooms in 2010 resulted in adoption of comprehensive environmental regulations for the entire Shanxi Reservoir watershed. These efforts included implementation of five watershed remediation strategies including domestic sewage treatment, residential garbage treatment, livestock pollution treatment, conservation and restoration of major tributaries, and development of an online real-time water quality monitoring and forecasting system (Dong et al. 2016; Mei et al. 2016). As a result, water quality in Shanxi Reservoir demonstrated slow improvements during the latest period. In general, the trends and change-points identified by the integrated LWPR-SegReg approach were confirmed by the hydrologic and biogeochemical conditions of the reservoir, as well as watershed-scale implementation of environmental regulation and remediation actions.

The practical application of the LWPR-SegReg approach was realized in examining the efficiency of past pollution control strategies and to predict future water quality trends. The existence of change-points indicates that the changes in trends within the time series are non-monotonic, and the corresponding change-points reveal when the trends changed. Whether the water quality improved or deteriorated was identified by the slopes of the segments derived from the LWPR-SegReg approach, and trends for the near future were predicted from the trend of the latest segment.

Advantage of the LWPR-SegReg approach

STL, LOWESS and LWPR are graphical methods that have been improved to assess long-term trends in environmental quality time series records (Bodo 1989; Lee et al. 2010; Stow et al. 2015). SegReg has also been widely used in trend and change-point analyses in many research fields (Mathews and Hamilton 2005; Wu and Chang 2012; Kazemnejad et al. 2014), including hydrology (Shao and Campbell 2002). For water quality trend and change-point analyses, graphical methodsare limited as they do not produce the regression functions necessary to determine significance levels and change-point detection, while SegReg is limited by statistically compromised data issues common in water quality time series records.

SegReg was performed directly on the original monthly irregular and discrete TN and COD_{Mn} datasets; the fitting results for the SegReg models are shown in Figs. 8 and 9, and model parameters are listed in Table 2. SegReg on the original time series data for TN and COD_{Mn} identified three segments with two change-points, indicating that the trends were both decreasing and increasing within the 2005–2014 record (Figs. 8 and 9). The first change-points for TN and COD_{Mn} were 2009.516 and 2010.177 with corresponding

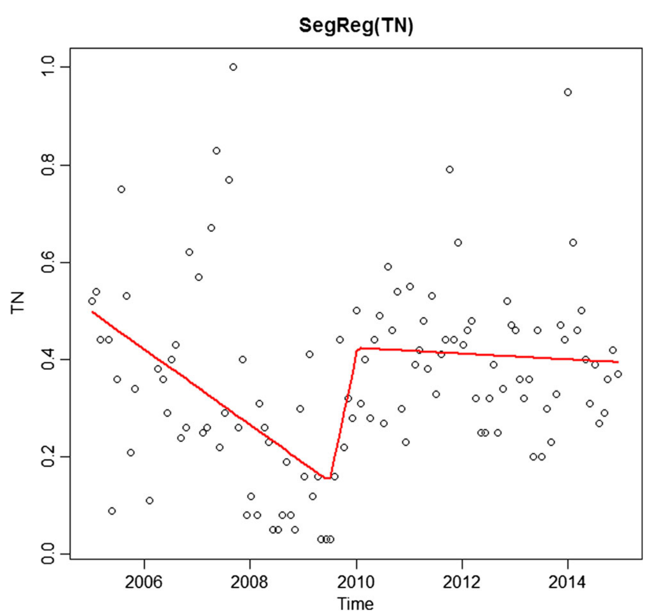

Fig. 8 Direct SegReg of monthly irregular and discrete TN concentrations (mg L^{-1})

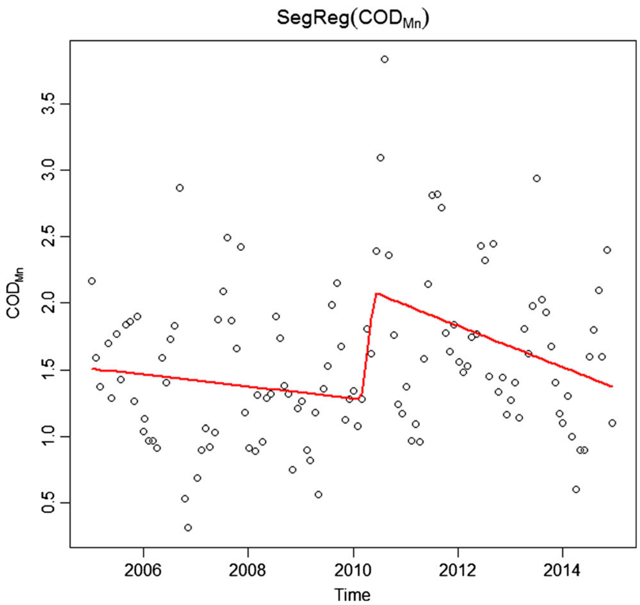

Fig. 9 Direct SegReg of monthly irregular and discrete COD_{Mn} concentrations (mg L^{-1})

slopes of −0.077 and −0.044, the second change-points were 2010.006 and 2010.384 with slopes of 0.637 and 3.927, and the slopes for the third segment were −0.565 and −4.038, respectively (Table 2). Based on practical considerations, the long-term trends and change-points determined from direct SegReg analysis of the original time series were not fully reliable. Firstly, TN and COD_{Mn} should not be decreasing as early as the beginning of 2010, since the implementation of watershed pollution control measures only began after 2010, and the improvement of water quality would be expected to lag behind the implementation of watershed management practices. Secondly, the slopes for the second and third segments (particularly for COD_{Mn}) would not be expected to be so big due to the buffering capacity of water quality constituents by the large volume of water stored in the reservoir. For instance, due to the long-term accumulation of nutrients in soil and groundwater within watersheds (Chen et al. 2014a; Huang et al. 2014b; Van Mete and Basu 2015), a lag effect in nutrient transport from legacy nutrient sources has been shown to increase riverine nutrient concentrations in many regions even after implementation of extensive pollution control efforts (Stålnacke et al. 2003; Onderka and Mrafková 2012; Chen et al. 2014b; Dong et al. 2016). Therefore, water quality improvement in large watersheds and reservoirs requires long time periods following implementation of pollution control measures, and the response is generally not very rapid. In general, the accuracy and objectivity of LWPR-SegReg results were much better than the use of STL, LWPR, or SegReg alone.

Table 2 SegReg model parameters of original TN and COD_{Mn} time series data

Model	Parameter	Estimated	Std. err	*P* value
TN	α1	−0.077	0.018	<0.001
	α2	0.637	0.671	
	α3	−0.565	0.670	
	CP 1	2009.516	0.295	
	CP 2	2010.006	0.328	
	Adjusted R^2	0.169		
COD_{Mn}	α1	−0.044	0.046	0.349
	α2	3.927	10.029	
	α3	−4.038	10.029	
	CP 1	2010.177	0.338	
	CP 2	2010.384	0.227	
	Adjusted R^2	0.105		

Neither LWPR nor SegReg was able to accomplish long-term trend and change-point analysis independently, whereas their integration creates a powerful new method to analyze water quality time series data. An important advantage of the integrated LWPR-SegReg approach is the ability of LWPR to extract long-term trend information from water quality time series (Cleveland 1979; Bodo 1989; Stow et al. 2015) and subsequently providing an appropriate dataset for SegReg analysis. Another notable advantage is the ability of SegReg to quantitatively detect change-points and trends with statistical rigor (Shao and Campbell 2002; Kazemnejad et al. 2014). Furthermore, this approach is easily performed in R packages, and all results can be graphed to provide an effective visualization of the time series dynamics.

Conclusions

The integrated LWPR-SegReg approach was demonstrated to be straightforward and effective for determining long-term trends and change-points in irregular water quality time series. The practical value of the integrated LWPR-SegReg approach is the ability to successfully evaluate the efficacy of pollution control strategies, as well as to predict future water quality trends. While this approach was developed for use with water quality data, it has applications for use with many types of time series records.

The study revealed that the TN and COD_{Mn} concentrations in Shanxi Reservoir watershed decreased before the first half of 2008, increased during the second half of 2008 to the first half of 2011, and then decreased gradually in response to pollution control actions after the second half of 2011. Given the considerable lag effect resulting from legacy nutrient retention in watershed and reservoir waters/soils/sediments, improving water quality conditions in the reservoir will require a long-term effort.

Acknowledgements We gratefully acknowledge the useful comments from the editor and anonymous reviewers. This work was supported by the National Natural Science Foundation of China (No. 40161554), Natural Science Foundation of Zhejiang (No. LQ16C030004), Public Welfare Science and Technology Project Plan of Wenzhou (No. S20140014), and Science Research Funding of Wenzhou Medical University (No. QTJ14045).

Compliance with ethical standards

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

References

Basu A, Saha D, Saha R et al (2014) A review on sources, toxicity and remediation technologies for removing arsenic from drinking water. Res Chem Intermed 40(2):447–485

Bodo BA (1989) Robust graphical methods for diagnosing trend in irregularly spaced water quality time series. Environ Monit Assess 13(2): 407–428

Chambers JM, Cleveland WS, Kleiner B (1983) Graphical methods for data analysis. Duxbury Press, Boston

Chang H (2008) Spatial analysis of water quality trends in the Han River basin, South Korea. Water Res 42(13):285–304

Chen DJ, Hu MP, Guo Y et al (2014a) Influence of legacy phosphorus, land use, and climate change on anthropogenic phosphorus inputs and riverine export dynamics. Biogeochemistry 123(1–2):99–116

Chen DJ, Huang H, Hu MP et al (2014b) Influence of lag effect, soil release, and climate change on watershed anthropogenic nitrogen inputs and riverine export dynamics. Environ Sci Technol 48(10): 5683–5690

Chen Z, Zhu Z, Yin L et al (2015) The changing water quality characteristics from urban drinking water sources in Guangdong, China. Water Resour Manag 29(4):987–1002

Chiang CT, Lian L, Su CC et al (2010) Spatiotemporal trends in oral cancer mortality and potential risks associated with heavy metal content in Taiwan soil. Int J Environ Res Public Health 7(11): 3916–3928

Chowdhury S, Al-Zahrani M (2014) Trend analysis in water quality of Al-Wehda Dam, north of Jordan. Environ Monit Assess 186(10): 6127–6143

Cleveland WS (1979) Robust locally weighted regression and smoothing scatterplots. J Am Stat Assoc 74(368):829–836

Cleveland WS (1981) LOWESS: a program for smoothing scatterplots by robust locally weighted regression. Am Stat 35(1):54–54

Cleveland WS, Devlin SJ (1988) Locally–weighted regression: an approach to regression analysis by local fitting. J Am Stat Assoc 83(403):596–610

Cleveland RB, Cleveland WS, McRae JE et al (1990) STL: a seasonal–trend decomposition procedure based on loess. J Off Stat 6(1):3–33

Corlin L, Rock T, Cordova J et al (2016) Health effects and environmental justice concerns of exposure to uranium in drinking water. Curr Environ Health Rep 3(4):434–442

Dong X, Mei K, Shang X et al (2016) Analysis of variation trend of water quality based on Mann–Kendall test and rescaled range analysis. J Ecol Rural Environ 32(02):277–282 (in Chinese with English Abstract)

Donohue R, Davidson WA, Peters NE (2001) Trends in total phosphorus and total nitrogen concentrations of tributaries to the Swan–Canning Estuary, 1987 to 1998. Hydrol Process 15(15):2411–2434

Goncharuk V (2013) The new standard for drinking water. J Chem 2013:1–7

Gong J, Hu Y, Liu M et al (2015) Characterization of air pollution index and its affecting factors in industrial urban areas in northeastern China. Teratology 8(1):81–82

Harding LW Jr, Gallegos CL, Perry ES et al (2016) Long-term trends of nutrients and phytoplankton in Chesapeake Bay. Estuar Coasts 39(3):664–681

Houtman CJ, Kroesbergen J, Lekkerkerker-Teunissen K et al (2014) Human health risk assessment of the mixture of pharmaceuticals in Dutch drinking water and its sources based on frequent monitoring data. Sci Total Environ 496:54–62

Huang H, Chen DJ, Zhang BF et al (2014a) Modeling and forecasting riverine dissolved inorganic nitrogen export using anthropogenic nitrogen inputs, hydroclimate, and land-use change. J Hydrol 517: 95–104

Huang H, Zhang BF, Lu J (2014b) Quantitative identification of riverine nitrogen from point, direct runoff and base flow sources. Water Sci Technol 70(5):865–870

Kazemnejad A, Arsang Jang S, Amani F et al (2014) Global epidemic trend of tuberculosis during 1990–2010: using segmented regression model. J Res Health Sci 14(2):115–121

Kisi O, Ky M (2014) Comparison of Mann–Kendall and innovative trend method for water quality parameters of the Kizilirmak River. Turk J Hydrol 513(5):362–375

Lee HW, Bhang KJ, Parl SS (2010) Effective visualization for the spatio-temporal trend analysis of the water quality in the Nakdong River of Korea. Ecol Inform 5(4):281–292

Li L, Qian J, Qu CQ et al (2014) Spatial and temporal analysis of air pollution index and its timescale-dependent relationship with meteorological factors in Guangzhou, China, 2001–2011. Environ Pollut 190(7):75–81

Liang ZY, Liu Y, Sheng H et al (2014) Identification of water quality trends in Lake Dianchi using an integrated STL and RSI approach. Acta Sci Circumst 34(3):754–762 (in Chinese with English Abstract)

Liu RY, Li L, Li H et al (2013) Research on carrying capacity of water resource based on multi–reservoir regulation. Res J Appl Sci Eng Technol 6(23):4340–4343

Mathews TJ, Hamilton BE (2005) Trend analysis of the sex ratio at birth in the United States. Div Vital Stat 53(20):1–17

Mei K, Shang X, Wang ZF et al (2016) Influence of land use on memory effect of watershed nitrogen. Acta Sci Circumst 36(10):3856–3863 (in Chinese with English Abstract)

Mukheibir P (2010) Water access, water scarcity, and climate change. Environ Manag 45:1027–1039

Naddafi K, Hnari H, Ahmadi M (2007) Water quality trend analysis for the Karoon River in Iran. Environ Monit Assess 134:305–312

Onderka M, Mrafková L, Krein et al (2012) Long–term persistence of stream nitrate concentrations (memory effect). Water Air Soil Pollut 223:241–252

Proietti T, Luati A (2011) Rank prediction in graphs with locally weighted polynomial regression and EM of polynomial mixture models. J Stat Plann Inference 141(2):831–845

Rajagopalan B, Lall U (1998) Locally weighted polynomial estimation of spatial precipitation. J Geogr Inf Decis Anal 2(2):44–51

Shao Q, Campbell NA (2002) Modelling trends in groundwater levels by segmented regression with constraints. Aust N Z J Stat 44(2):129–141

Sharma S, Swayne DA, Obimbo C (2016) Trend analysis and change point techniques: a survey. Energy Ecol Environ. doi:10.1007/s40974-016-0011-1

Shi QW (2010) Cause of formation and countermeasures for cyanobacteria bloom in Shanxi Reservoir. J N Chin Inst Water Conserv Hydroelectr Power 12(9):142–145 (in Chinese with English Abstract)

Sokolova E, Pettersson TJR, Bergstedt O et al (2013) Hydrodynamic modelling of the microbial water quality in a drinking water source as input for risk reduction management. J Hydrol 497(7):15–23
Stålnacke P, Grimvall A, Libiseller C, Laznik M, Kokorite I (2003) Trends in nutrient concentrations in Latvian rivers and the response to the dramatic change in agriculture. J Hydrol 283(1–4):184–205
Stow CA, Cha YK, Johnson LT et al (2015) Long-term and seasonal trend decomposition of Maumee River nutrient inputs to western Lake Erie. Environ Sci Technol 49(6):3392–3400
Sun C, Zhang J, Ma Q et al (2015) Human health and ecological risk assessment of 16 polycyclic aromatic hydrocarbons in drinking source water from a large mixed-use reservoir. Int J Environ Res Public Health 12(11):13956–13969
Syme GJ, Nancarrow BE (2013) Incorporating community and multiple perspectives in the development of acceptable drinking water source protection policy in catchments facing recreation demands. J Environ Manag 129(18):112–123
Taljaard M, McKenzie JE, Ramsay CR et al (2014) The use of segmented regression in analyzing interrupted time series studies: an example inpre–hospital ambulance care. Implement Sci 9(1):1–4
Tiwari PC, Joshi B (2012) Environmental changes and sustainable development of water resources in the Himalayan headwaters of India. Water Resour Manag 26:883–907
Tropea AE, Ginn BK, Cumming BF et al (2007) Tracking long-term acidification trends in Pockwock Lake (Halifax, Nova Scotia), the water supply for a major eastern Canadian city. Lake Reservoir Manage 23(3):279–286
Van Mete KJ, Basu NB (2015) Catchment legacies and time lags: a parsimonious watershed model to predict the effects of legacy storage on nitrogen export. PLoS One 10(5):e0125971. doi:10.1371/journal.pone.0125971
Wang H, Yu X (2014) A review of the protection of sources of drinking water in China. Nat Res Forum 38(2):993–108
Wei GL, Yang ZF, Cui BS et al (2009) Impact of dam construction on water quality and water self–purification capacity of the Lancang River. China Water Resour Manag 23(9):1763–1780
Wu JL, Chang PC (2012) A trend–based segmentation method and the support vector regression for financial time series forecasting. Math Probl Eng 2012(11–12):243–253
Xu J, Feng P, Yang P (2016) Research of development strategy on China's rural drinking water supply based on SWOT-TOPSIS method combined with AHP-entropy: a case in Hebei Province. Environ Earth Sci 75(1):1–11
Zhang XJ, Chen C, Lin PF et al (2011) Emergency drinking water treatment during source water pollution accidents in China: origin analysis, framework and technologies. Environ Sci Technol 45(1):161–167
Zhou Y, Khu ST, Xi B et al (2014) Status and challenges of water pollution problems in China: learning from the European experience. Environ Earth Sci 72:1243–1254