Application of multivariate statistical methods in the assessment of water quality in selected locations in Jialing River basin in Guangyuan, China

Yiming Zhang, Yunxiang Li, Youping Li, Wenyi Da, Maolei Yu and Qiumei Quan

ABSTRACT

This study evaluated temporal and spatial variations in water quality to understand the characteristics of the Jialing River watershed in Guangyuan City, China. Data on 17 parameters obtained from seven sites from 2012 to 2015, with a total of 329 samples for each parameter, were analyzed using multivariate statistical techniques. Observation months were grouped into two periods (Period A, May–November; Period B, December–April) according to similarities in water quality characteristics through time analysis and cluster analysis (CA). Water temperature (TEMP), flow rate (Q), dissolved oxygen (DO), oils, fluoride (F) and cadmium (Cd) were the most significant parameters for discriminating between the two periods. Through a spatial analysis, the sites were classified into two groups (Groups 1 and 2). Q, total phosphorus (TP), oils, F and fecal coliform bacteria (F. *coli*) were the most significant parameters for discriminating between the two groups. Results suggested that TEMP, DO and Cd as functions of time, TP and F. *coli* as functions of space, and Q, oils and F as functions of both time and space should be monitored closely. The main sources of water pollutant were surface runoff and industrial wastewater, of time, and wastewater from agricultural irrigation, industrial wastewater, and municipal sewage, of space.

Key words | Jialing River watershed, multivariate statistical techniques, sources of water pollutants, temporal and spatial variations, water quality

Yiming Zhang
Yunxiang Li
Youping Li
Wenyi Da
Maolei Yu
Qiumei Quan (corresponding author)
College of Environmental Science and Engineering,
China West Normal University,
Nanchong, Sichuan, 637000,
China
E-mail: *meimeiq@163.com*

INTRODUCTION

The Jialing River watershed is a major tributary of the Yangtze River (Wu *et al.* 2012; Yan *et al.* 2015). It flows from north to south through major cities, including Baoji, Hanzhong, Guangyuan, Nanchong, and Chongqing (Wu *et al.* 2012). The main economic source of the Jialing River watershed is agriculture, and the growth of crops depends on irrigation (Li 2015). Jialing River covers a wide area and is an important source of drinking water. Therefore, the quality of water in the river significantly affects the life and ecological environment of residents (Yan *et al.* 2015).

The water quality of Jialing River has become a popular research subject due to recent industrial and urban construction developments in the area (Li 2015). River water quality assessment is necessary, especially in areas where river water serves as a drinking water source and is threatened by pollution caused by various human activities along the river course (Amadi 2011; Batayneh & Zumlot 2012). The temporal and spatial variation characteristics of river water quality can provide dynamic information for the effective management of the water environment and have

doi: 10.2166/ws.2018.058

become an important means of water environment management (Shrestha & Kazama 2007; Shrestha & Muangthong 2014; Muangthong & Shrestha 2015).

In recent decades, the quality of water has been heavily affected by human agricultural and industrial activities. Ali (2013) used bioassessments to evaluate the biological conditions and identify the degree of water quality degradation in the Suez Irrigation Canal and the results showed the significant impact of human activities along the canal banks on the canal ecosystem. Many researchers have also combined the water quality index (WQI) and GIS technology to reveal the patterns of different heavy metals in aquatic environments and analyze their spatial distributions through maps. Multivariate statistical techniques, such as cluster analysis (CA), discriminant analysis (DA), and principal component analysis, are used to characterize and evaluate surface water quality because they can verify temporal and spatial variations caused by natural and anthropogenic factors linked to seasonality (Helena *et al.* 2000; Singh *et al.* 2004, 2005; Muangthong & Shrestha 2015). In the present study, data sets obtained from the Guangyuan section of Jialing River were subjected to different multivariate statistical techniques to explore temporal and spatial variations in water quality. The results provide a scientific basis for ecological system management, environmental protection, and water pollution control in the Jialing River basin.

EXPERIMENTAL SITE, MATERIALS AND METHODS

Study site and monitored parameters

Guangyuan City (31° 31′ N–32° 56′ N, 104° 36′ E–106° 45′ E) is in the northern region of Sichuan, China (Figure 1). The Guangyuan section of Jialing River is located in the middle of the basin, and water quality in the river is affected by upstream water quality and significantly influences the water quality downstream (Li 2015). The population in Guangyuan is 3.05 million. The entire basin has a total area of approximately 16,319 km^2, the forest cover rate of the land is 55.3%, and the cultivable land is 1,688.98 km^2. The basin experiences a humid subtropical monsoon climate and annual average temperature, sunshine, and rainfall of

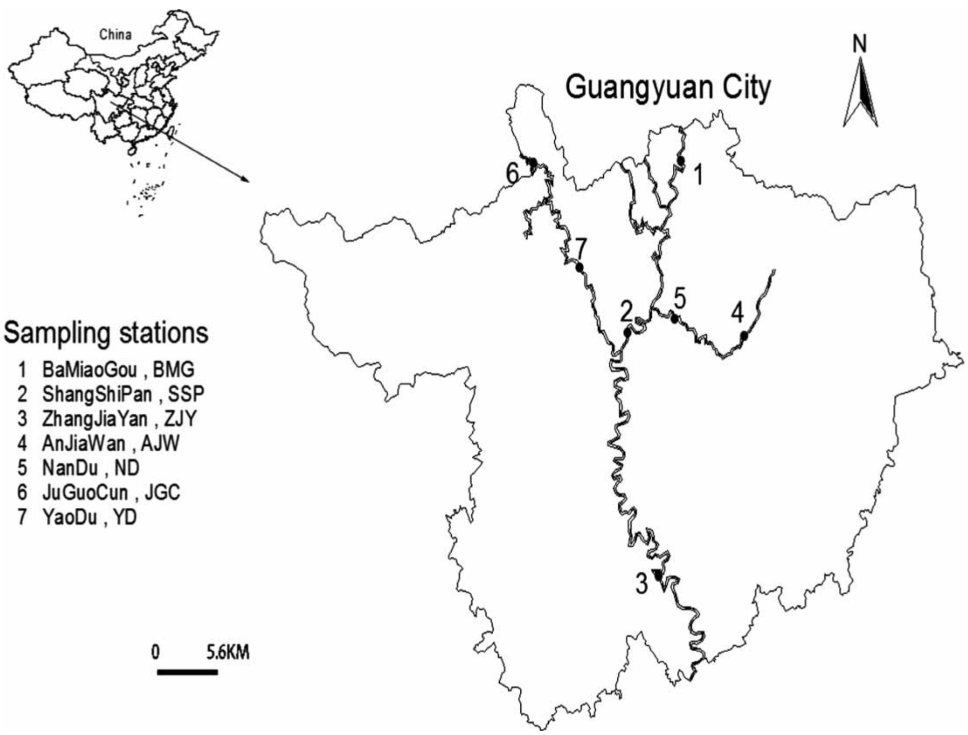

Figure 1 | Map of the study area and surface water quality monitoring stations in Jialing River watershed (Guangyuan).

17 °C, 1,398 h and 1,000 mm, respectively (Guangyuan Statistics Intranet 2010). The precipitation of the Jialing River basin is unevenly distributed, with 70%–90% rainfall distributed in June–September, and the runoff mainly originates from precipitation (Long *et al.* 2008). Water samples were collected from seven sites, namely, Bamiaogou (BMG), Shangshipan (SSP), Zhangjiayan (ZJY), Anjiawan (AJW), Nandu (ND), Juguocun (JGC), and Yaodu (YD), at monthly intervals between January 2012 and November 2015 from the Guangyuan Environmental Monitoring Center (GEMC). BMG, SSP, and ZJY are the monitoring sites for Jialing River. ND and AJW are the monitoring sites for Nan River, which is one of the branches of Jialing River. YD and JGC are the monitoring sites for Bailong River, which also is one of the branches of Jialing River. Both Nan River and Bailong River flow into Jialing River. SSP, ZJY, ND, and AJW are located in the urban area. BMG, YD, JGC, and ZJY are located in the country, but BMG is polluted by industrial wastewater and material.

Seventeen parameters were selected based on their sampling continuity at all the selected monitoring sites. These parameters included water temperature (TEMP), flow rate (Q), pH, electrical conductivity (EC), dissolved oxygen (DO), 5-day biochemical oxygen demand (BOD_5), chemical oxygen demand (COD_{Mn}), fecal coliform bacteria (F. *coli*), total nitrogen (TN), total phosphorus (TP), oils, ammonia nitrogen (NH_3–N), fluoride (F), copper (Cu), lead (Pb), cadmium (Cd), and zinc (Zn). All of these parameters, except for Q (m^3/s), pH, EC (μS/cm), TEMP (°C), and F. *coli* (cfu/L), are expressed in mg/L. GEMC had sampled, preserved, and analyzed these water quality parameters according to the Technical Specifications Requirements for Monitoring of Surface Water and Waste Water (HJ/T 91-2002, SEPA 2002). Environmental quality standards for surface water (GB3838-2002, in China) are shown in Table 1.

Analytical methods

Most multivariate statistical methods require variables to obey the normal distribution. Before multivariate statistical analysis, the normality of the distribution of each variable was checked by analyzing kurtosis and skewness statistical tests (Johnson & Wichern 1992; Zhou *et al.* 2007a). The original data demonstrated values of kurtosis ranging from −1.188 to 173.975 and skewness ranging from −0.907 to 11.942, indicating that distributions were far from normal with 95% confidence. Since most of the values of kurtosis or skewness were greater than zero, the original data were transformed in the form $x' = \log_{10}(x)$ (Kowalkowski *et al.* 2006; Papatheodorou *et al.* 2006). After log-transformation, the kurtosis and skewness values ranged from −1.176 to 1.801 and −0.588 to 1.406, respectively. But the kurtosis and skewness values of TN were 2.612 and −0.455, respectively, indicating that the distribution of the log-transformed TN was also non-normal, therefore it was not considered in the following study.

Multivariate analysis of the water quality data set was performed with two techniques: CA (Jarvie *et al.* 1998; Singh *et al.* 2005; Zhou *et al.* 2007a; Varol & Şen 2009) and DA (Johnson & Wichern 1992; Singh *et al.* 2005; Zhou *et al.* 2007a). To avoid misclassifications, CA was applied to experimental data standardized through z-scale transformation using Ward's method of linkage with squared Euclidean distance as the measure of similarity. Raw data were processed through DA and the stepwise method (Liu *et al.* 2003; Simeonov *et al.* 2003). Non-parametric tests were often used to evaluate the correlation structure between water quality parameters with non-normal distributions (Wunderlin *et al.* 2001; Singh *et al.* 2004; Shrestha & Kazama 2007). All mathematical and statistical computations were performed with IBM SPSS 23.0.

Table 1 | The standard of surface water environment quality (GB3838-2002, China)

	DO	COD_{Mn}	BOD_5	NH_3–N	TP	TN	Cu	Zn	F	Cd	Pb	oils	F. *coli*
I	7.5	2	3	0.15	0.02	0.2	0.02	0.05	1	0.001	0.01	0.05	200
II	6	4	3	0.5	0.1	0.5	1	1	1	0.005	0.05	0.05	2,000
III	5	6	4	1	0.2	1	1	1	1	0.005	0.05	0.05	10,000

CA is an exploratory analysis and classifies objects so that each of them can be similar to others in the cluster with respect to a predetermined selection criterion. The resulting clusters of objects should exhibit high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity (Varol & Şen 2009). In the evaluation of water quality, CA is often based on sampling points and time to analyze the spatial and temporal variations in water quality or based on the evaluation index to analyze the similarity between the indexes (Singh *et al.* 2005; Zhou *et al.* 2007a, 2007b). In this study, CA was performed on the standardized data sets by using Ward's method with squared Euclidean distances as a measure of similarity (Jarvie *et al.* 1998; Singh *et al.* 2005).

DA is a multivariate statistical method that generally uses a stepwise model to construct the discriminant functions (DFs) that present important information for each group (Johnson & Wichern 1992; Singh *et al.* 2005). In this study, DA operates on the original data using a stepwise mode to confirm the groups found by CA and evaluate spatiotemporal variations in terms of discriminant variables. The monitoring periods (temporal) and sites (spatial) were regarded as grouping variables, and the measured parameters were regarded as independent variables (Zhou *et al.* 2007a). Further details of the method can be found in the study of Johnson & Wichern (1992).

RESULTS AND DISCUSSION

Temporal similarity and period grouping

Temporal CA was used to generate a dendrogram (Figure 2) that grouped the months into two clusters with similar physiochemical water quality characteristics. Period A covered May–November, which included the high-flow period (July–September) and the mean-flow period (April–May and October–November) of Jialing River. Period B covered December–April, which included the low-flow period (December–March) of the river.

As shown in Table 2, the value of Wilks' lambda for the discriminant function was small (0.305), the chi-square value was high (384.064) and the p level (0.000) was below 0.05. These figures suggested that the temporal DA

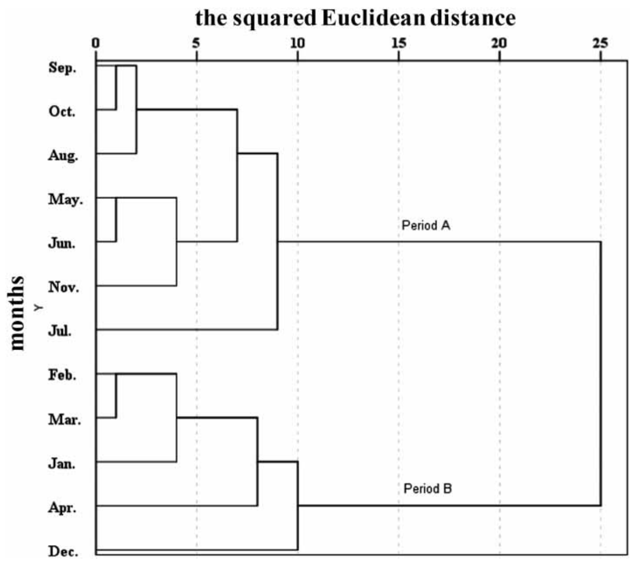

Figure 2 | Dendrogram showing the clustering of monitoring periods in Jialing River watershed (Guangyuan).

in this study was significant (Zhou *et al.* 2007a). The DFs and classification matrices (CMs) obtained from DA are shown in Tables 3 and 4, respectively. DA produced a CM with almost 91% correct assignments, a percentage that is higher than most of the correct assignments produced by backward stepwise DA (Shrestha & Kazama 2007; Zhou *et al.* 2007b). The temporal DA results suggested that TEMP, Q, DO, oils, F and Cd were the most significant

Table 2 | Wilk's lambda and chi-square values from DA of temporal variations in water quality

	Fun. (s)	R	Wilks' lambda	Chi-square	Sig.
Temporal	1	0.834	0.305	384.064	0.000

Table 3 | Classification function coefficients for DA of temporal variation

Parameters	Period A	Period B
TEMP	1.678	1.086
Q	−0.018	−0.025
DO	11.965	13.118
oils	−29.830	23.768
F	16.089	22.348
Cd	2,101.702	263.842
Constant	−67.492	−70.701

Table 4 | Classification matrix for DA of temporal variation

		Period assigned by DA[a]	
	Percent correct	**Period A**	**Period B**
Period A	92.9	182	13
Period B	90.2	14	120
Total	90.8	196	133

[a]Checked by cross-validation method.

parameters for discriminating between Periods A and B and accounted for most of the expected temporal variation in water quality.

The results showed that TEMP was significantly higher ($p = 0.000$) in Period A (21.49 °C) than in Period B (9.68 °C) and revealed a clear-cut temporal effect (Figure 3(a)) because Period A included all the warm months. A clear inverse relationship was observed between TEMP and DO. This relationship is natural because warm water saturates easily with oxygen and contains minimal DO (Varol & Şen 2009). DO has been used as an indicator of the ecological health of aquatic ecosystems (Qadir & Malik 2009). The DO content during most of the period was higher than the second national standard limit (6 mg/L, GB3838-2002) required for the protection of aquatic life (DeZuane 1997). The average Q and Cd in Period A (98.33 m^3/s and 0.00019 mg/L) were higher than those in Period B (83.06 m^3/s and 0.00008 mg/L, $p_{(Q)} = 0.028$, $p_{(Cd)} = 0.000$). Rainfall was also higher in Period A than in Period B because Period A covered the rainy months. The main sources of the water pollutant Cd were surface runoff and industrial wastewater (Li *et al.* 2013). The high Cd content in Period A can be related to urban

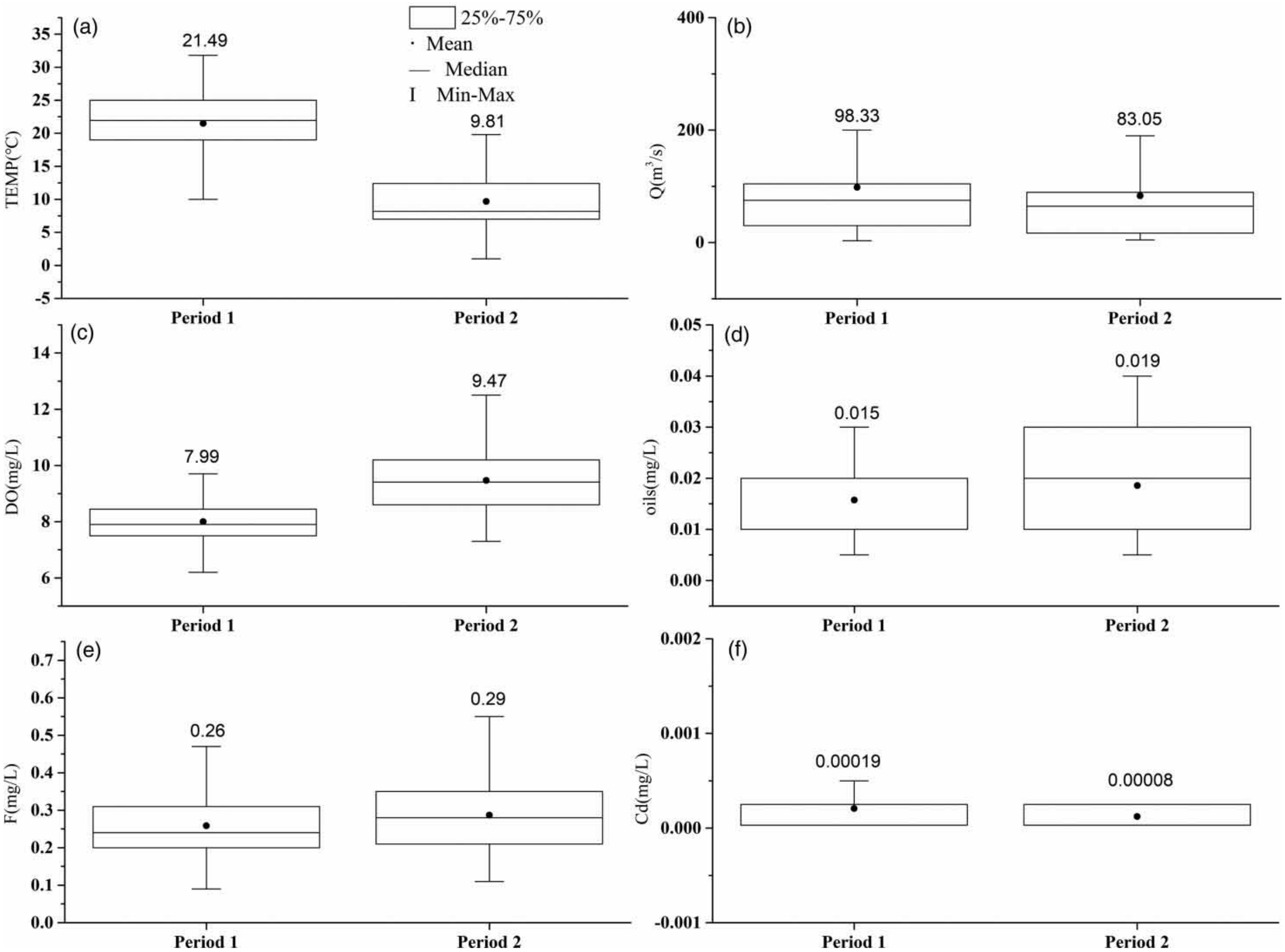

Figure 3 | Temporal variations: (a) temperature, (b) Q, (c) DO, (d) oils, (e) F, and (f) Cd.

runoff, traffic density, and automobile washing, which is common along stream banks. The maximum concentration of Cd (0.004 mg/L) was within the second national standard limit but higher than the WHO standard of 0.003 mg/L. Thus, Cd concentration should be controlled, and attention should be focused on waste disposal during Period A. The average concentrations of oils and F under the same pattern were lower in Period A (0.0157 mg/L and 0.0259 mg/L) than in Period B (0.0186 mg/L and 0.0287 mg/L, $p_{(oils)} = 0.017$, $p_{(F)} = 0.014$) and within the second national standard limit in both periods. The pollution source of oils and F may be from industrial wastewater. The government should strengthen monitoring of industrial wastewater and develop the waste disposal industries whose facilities should comply with state standards.

Spatial similarity and site grouping

The dendrogram showed that the monitoring locations may be grouped into two main clusters for the analysis of spatial variation (Figure 4). JGC, YD, and ZJY formed Group 1, and the remaining sites formed Group 2. The classification of these groups changed with the significance level. Group 2 is characterized by a larger squared Euclidean distance compared with Group 1. In Group 1, JGC, YD, and ZJY, which are in the country, exhibited lower levels of pollution than the other sites. Most of these stations receive pollution from agricultural and farm effluents. Group 2 included AJW, ND, BMG, and SSP, which have higher pollution levels than the other sites. Most of these stations receive pollution from domestic wastewater and industrial effluents located in urban areas.

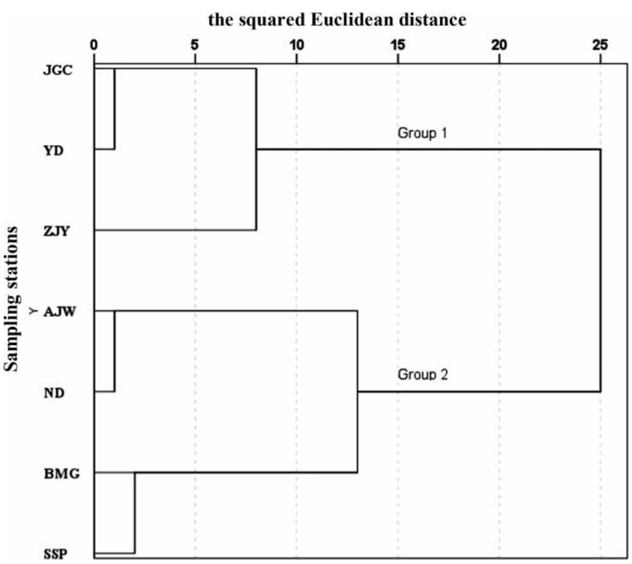

Figure 4 | Dendrogram showing the clustering of monitoring sites in Jialing River basin (Guangyuan).

As shown in Table 5, the values of Wilks' lambda and chi-square for the discriminant function were 0.560 and 187.976, respectively, and the p level (0.000) was below 0.05. These values indicated that the spatial DA in this study had a good discriminatory ability and was significant (Zhou *et al.* 2007b). Tables 6 and 7 respectively show the DFs and CMs obtained from DA. Spatial DA produced CMs with 83% correct assignments, which is close to those of other reports (Shrestha & Kazama 2007; Zhou *et al.* 2007b). DA showed that Q, oils, TP, F, and F. *coli* were the discriminating parameters in space.

The discriminating parameters identified by spatial DA were used to evaluate different patterns associated with spatial variations in water quality (Figure 5). The average Q was higher in Group 1 (142.03 m^3/s) than in Group 2 (54.74 m^3/s, $p = 0.000$, Figure 5(a)). The average concentration of F. *coli* in Group 2 (3,085.11 cfu/L) was higher than that in Group 1 (2,470.92 cfu/L, $p = 0.000$), and both averages were above the second national standard limit (2,000, GB3838-2002). These results indicated that water quality was negatively affected by anthropogenic and

Table 5 | Wilk's lambda and chi-square values of DA of spatial variation in water quality

	Fun. (s)	R	Wilks' lambda	Chi-square	Sig.
Spatial	1	0.663	0.560	187.976	0.000

Table 6 | Classification function coefficients for DA of spatial variation

Parameters	Group 1	Group 2
Q	0.003	−0.015
oils	137.139	189.239
TP	27.843	56.590
F	32.516	42.080
F. *coli*	0.001	0.002
Constant	−17.913	−24.663

Table 7 | Classification matrix for DA of spatial variation

		Regions assigned by DA[a]	
	Percent correct	Group 1	Group 2
Group 1	75.2	106	21
Group 2	88.8	35	167
Total	83.0	141	188

[a]Checked by cross-validation method.

animal activities (Whitlock *et al.* 2002). The concentrations of oils, TP and F were within the second national standard limits and lower than those reported by other authors (Zhou *et al.* 2007a; Varol & Şen 2009; Muangthong & Shrestha 2015). The trends of these parameters indicated that the average concentrations in Group 2 were higher than those in Group 1 ($p_{(oils)} = 0.000$, $p_{(TP)} = 0.000$, $p_{(F)} = 0.045$, Figure 5(b)–5(d)). Therefore, the management of rational pesticides and the principles of rational fertilization should be strengthened at the sites in Group 2. The sites in Group 1 were relatively far from pollution sources, and the concentration of pollutants was diluted by river confluence. Sites in Group 2 were more affected by anthropogenic activities and characterized by relatively turbulent flow, washing activities along the stream banks, urban runoff, and municipal effluents, compared with the sites in Group 1. Group 2's discharge was easily influenced by wastewater from agricultural irrigation, industrial wastewater, and municipal sewage because most of its monitoring sites were near urban areas or 'unsewered' villages (Shrestha & Muangthong 2014; Muangthong & Shrestha 2015). Therefore, F. *coli* should be

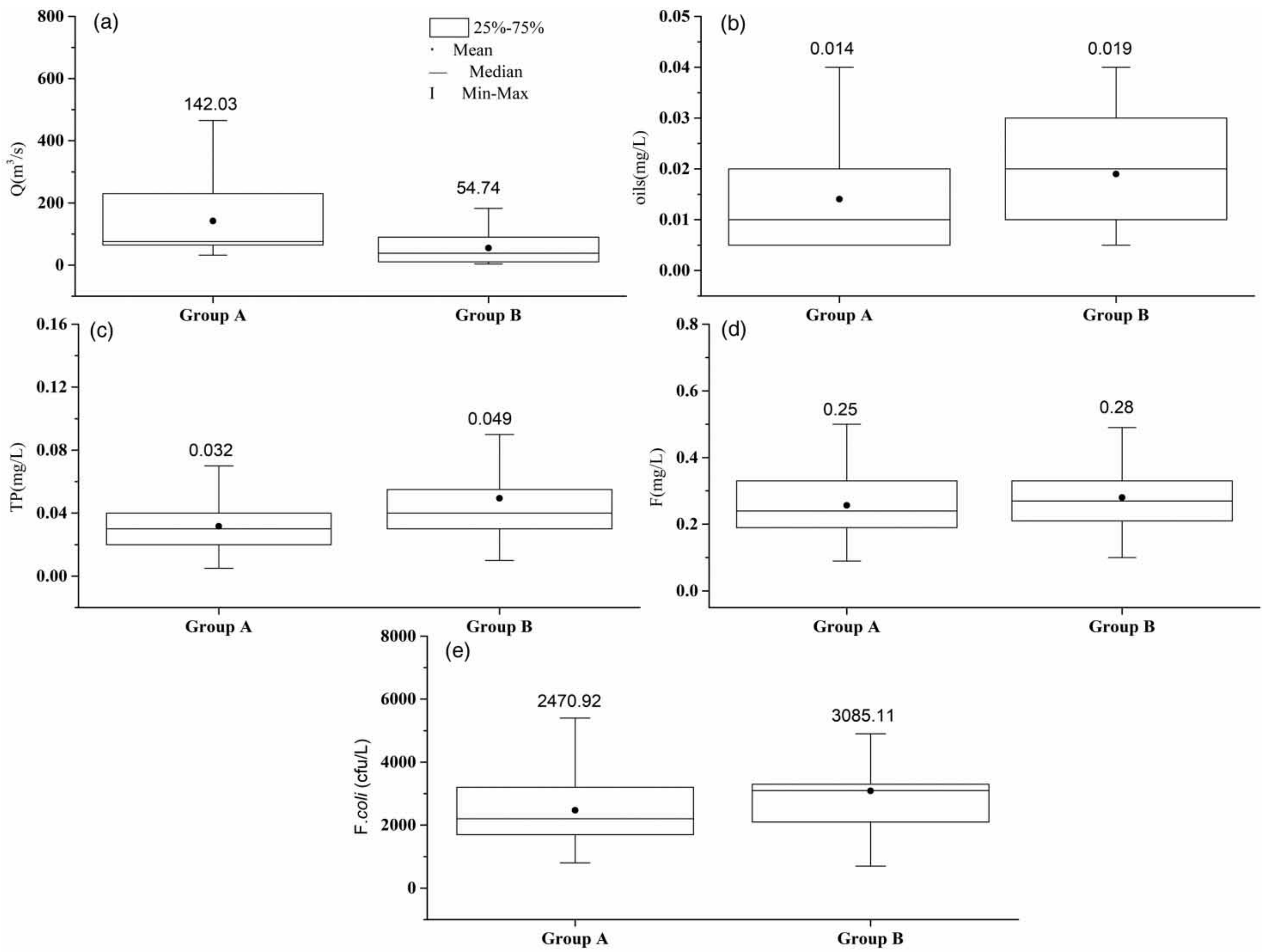

Figure 5 | Spatial variations: (a) Q, (b) oils, (c) TP, (d) F, and (e) F. *coli*.

monitored and controlled. Moreover, the construction and management of many municipal intercepting sewer pipes, wastewater, and garbage treatment plants should be supervised well. The wide application of agricultural machines was the source of oils and F.

CONCLUSIONS

The temporal similarity analysis of Jialing River basin showed that months can be divided into two periods: May–November (Period A) and December–April (Period B). TEMP, Q, DO, oils, F, and Cd were the discriminant variables. The temporal analysis assigned these groups with 90.8% accuracy. The average values of TEMP, Q, and Cd in Period A were higher than those in Period B. Therefore, the source of Cd should be controlled, and attention should be focused on waste disposal during Period A. The average values of DO, oils, and F in Period B were higher than those in Period A. During Period B, management should focus on managing oils and F.

The spatial similarity analysis showed that the sampling stations can be divided into two clusters. AJW, ND, BMG, and SSP, which were highly polluted regions, formed Group 2, and the remaining stations formed Group 1. Q, oils, TP, F, and F. *coli* were discriminant variables with 80.3% correct assignments. The average Q in Group 1 was higher than that in Group 2. The average values of oils, TP, F, and F. *coli* in Group 2 were higher than those in Group 1. These parameters, especially F. *coli*, should be controlled. Management can then establish a network of municipal intercepting sewers effectively.

In summary, management should concentrate on monitoring TEMP, Q, DO, oils, F, and Cd as functions of time and Q, oils, TP, F, and F. *coli* as functions of space. There were four main pollution sources including domestic sewage, industrial wastewater, agricultural fertilizer and pesticide inputs, and livestock and poultry-raising waste. The main sources of water pollutant were surface runoff and industrial wastewater as functions of time and wastewater from agricultural irrigation, industrial wastewater, and municipal sewage as functions of space. This study provides an improved understanding of the temporal and spatial variations in the Guangyuan section of the Jialing River watershed.

ACKNOWLEDGMENTS

The authors sincerely thank Guangyuan Environmental Protection Department for permission to use the data. This paper was supported by the Department of Science and Technology of Sichuan Province (CN) (2015JY0094). The opinions in this paper do not reflect the views or policies of GEMC.

REFERENCES

Ali, M. M. 2013 An analysis of the impact of human activities on water quality and ecological responses in the Suez irrigation canal. *Management of Environmental Quality* **19** (3), 377–401.

Amadi, A. N. 2011 Assessing the effects of aladimma dumpsite on soil and groundwater using water quality index and factor analysis. *Australian Journal of Basic & Applied Sciences* **5** (11), 763–770.

Batayneh, A. & Zumlot, T. 2012 Multivariate statistical approach to geochemical methods in water quality factor identification: application to the shallow aquifer system of the Yarmouk basin of north Jordan. *Research Journal of Environmental & Earth Sciences* **4** (7), 170–177.

DeZuane, J. 1997 *Handbook of Drinking Water Quality*. Wiley, New York, USA.

GB3838-2002 2002 Environmental quality standards for surface water. In: *Inspection and Quarantine of the People's Republic of China*.

Guangyuan Statistics Intranet 2010 Natural resources of Guangyuan. Guangyuan Statistics Intranet. http://www.scgytjj.gov.cn/info/23.htm.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M. & Fernandez, L. 2000 Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research* **34** (3), 807–816.

Jarvie, H. P., Whitton, B. A. & Neal, C. 1998 Nitrogen and phosphorus in east coast British rivers: speciation, sources and biological significance. *Science of the Total Environment* **210–211**, 79–109.

Johnson, R. A. & Wichern, D. W. 1992 *Applied Multivariate Statistical Analysis*, 5th edn. Prentice-Hall, Englewood Cliffs, NJ, USA.

Kowalkowski, T., Zbytniewski, R., Szpejna, J. & Buszewski, B. 2006 Application chemometrics in river water classification. *Water Research* **40**, 744–752.

Li, W. X. 2015 Investigation of water quality and pollution discharge in Jialing river basin. *Water Resources Planning and Design* **8**, 37–38.

Li, C. L., Hu, Y. M., Liu, M., Xu, Y. Y. & Sun, F. Y. 2013 Urban non-point source pollution: research progress. *Chinese Journal of Ecology* **32** (2), 492–500.

Liu, C. W., Lin, K. H. & Kuo, Y. M. 2003 Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the Total Environment* **313** (1–3), 77–89.

Long, T. Y., Li, J. C. & Liu, N. M. 2008 Adsorbed non-point source pollution load of Jialing River Basin. *Environmental Science* **29** (7), 1811–1817.

Muangthong, S. & Shrestha, S. 2015 Assessment of surface water quality using multivariate statistical techniques: case study of the Nampong River and Songkhram River, Thailand. *Environmental Monitoring & Assessment* **187** (9), 548.

Papatheodorou, G., Demopoulou, G. & Lambrakis, N. 2006 A long-term study of temporal hydrochemical data in a shallow lake using multivariate statistical techniques. *Ecological Modelling* **193**, 759–776.

Qadir, A. & Malik, R. N. 2009 Assessment of an index of biological integrity (IBI) to quantify the quality of two tributaries of river Chenab, Sialkot, Pakistan. *Hydrobiologia* **621** (1), 127–153.

SEPA 2012 *Technical Specifications Requirements for Monitoring of Surface Water and Wastewater HJ/T 91-2002*. China Environmental Science Press, Beijing, China.

Shrestha, S. & Kazama, F. 2007 Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environmental Modelling & Software* **22** (4), 464–475.

Shrestha, S. & Muangthong, S. 2014 Assessment of surface water quality of Songkhram River (Thailand) using environmetric techniques. *International Journal of River Basin Management* **12** (4), 341–356.

Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. & Kouimtzis, Th. 2003 Assessment of the surface water quality in northern Greece. *Water Research* **37** (17), 4119–4124.

Singh, K. P., Malik, A., Mohan, D. & Sinha, S. 2004 Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – a case study. *Water Research* **38** (18), 3980–3992.

Singh, K. P., Malik, A. & Sinha, S. 2005 Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques – a case study. *Analytica Chimica Acta* **538** (1–2), 355–374.

Varol, M. & Şen, B. 2009 Assessment of surface water quality using multivariate statistical techniques: a case study of Behrimaz Stream, Turkey. *Environmental Monitoring & Assessment* **159** (1–4), 543–553.

Whitlock, J. E., Jones, D. T. & Harwood, V. J. 2002 Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. *Water Research* **36** (17), 4273–4282.

Wu, L., Long, T. Y., Liu, X. & Mmereki, D. 2012 Simulation of soil loss processes based on rainfall runoff and the time factor of governance in the Jialing River Watershed, China. *Environmental Monitoring & Assessment* **184** (6), 3731–3748.

Wunderlin, D. A., Díaz, M. d. P., Amé, M. V., Pesce, S. F., Hued, A. C. & Bistoni, M. d. l. A. 2001 Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquía River Basin (Cordoba, Argentina). *Water Research* **35**, 2881–2894.

Yan, J. L., Jiang, T., Gao, J., Wei, S. Q., Lu, S. & Liu, J. 2015 Characteristics of absorption and fluorescence spectra of dissolved organic matter from confluence of rivers: case study of Qujiang River-Jialing River and Fujiang River-Jialing River. *Environmental Science* **36** (3), 869–878.

Zhou, F., Huang, G. H., Guo, H., Zhang, W. & Hao, Z. 2007a Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong. *Water Research* **41** (15), 3429–3439.

Zhou, F., Liu, Y. & Guo, H. 2007b Application of multivariate statistical methods to water quality assessment of the watercourses in northwestern new territories, Hong Kong. *Environmental Monitoring & Assessment* **132** (1–3), 1–13.

First received 14 June 2017; accepted in revised form 2 March 2018. Available online 19 March 2018