Journal of Environmental Monitoring

Cite this: *J. Environ. Monit.*, 2011, **13**, 2755

www.rsc.org/jem

PAPER

Optimizing water quality monitoring networks using continuous longitudinal monitoring data: a case study of Wen-Rui Tang River, Wenzhou, China

Kun Mei,[a] Yuanli Zhu,[a] Lingling Liao,[a] Randy Dahlgren,[ab] Xu Shang[a] and Minghua Zhang*[ab]

Received 23rd April 2011, Accepted 15th August 2011
DOI: 10.1039/c1em10352k

Identification of representative sampling sites is a critical issue in establishing an effective water quality monitoring program. This is especially important at the urban–agriculture interface where water quality conditions can change rapidly over short distances. The objective of this research was to optimize the spatial allocation of discrete monitoring sites for synoptic water quality monitoring through analysis of continuous longitudinal monitoring data collected by attaching a water quality sonde and GPS to a boat. Sampling was conducted six times from March to October 2009 along a 6.5 km segment of the Wen-Rui Tang River in eastern China that represented an urban–agricultural interface. When travelling at a velocity of ~2.4 km h^{-1}, this resulted in water quality measurements at ~20 m interval. Ammonia nitrogen (NH_4^+–N), electrical conductivity (EC), dissolved oxygen (DO), and turbidity data were collected and analyzed using Cluster Analysis (CA) to identify optimal locations for establishment of long-term monitoring sites. The analysis identified two distinct water quality segments for NH_4^+–N and EC and three distinct segments for DO and turbidity. According to our research results, the current fixed-location sampling sites should be adjusted to more effectively capture the distinct differences in the spatial distribution of water quality conditions. In addition, this methodology identified river reaches that require more comprehensive study of the factors leading to the changes in water quality within the identified river segment. The study demonstrates that continuous longitudinal monitoring can be a highly effective method for optimizing monitoring site locations for water quality studies.

1. Introduction

Water quality monitoring at the regional scale is an important component of water resource management, conservation, protection, and remediation.[1,2] A well-designed monitoring program can help reduce the cost of data collection and optimize information return on the monitoring investment. Selection of sample sites for many early water quality monitoring programs (during the 1960s) was often determined by arbitrary approaches (*e.g.*, bridge access) or personal knowledge of watershed conditions. Once monitoring sites were set, there was commonly no assessment of the monitoring network design effectiveness.[3–7]

An optimal design of water quality monitoring networks and efficiency improvements have been the subject of research since the 1970s,[8] and beginning in 1980s several approaches were evaluated to improve the design criteria and monitoring efficiency.[9] Following these preliminary studies, research focused on identifying the location of water quality monitoring stations for contrasting objectives, such as trend detection and remediation effectiveness monitoring.[1] More recent studies using multiple

[a]*The Environmental Geographic Information System Laboratory, School of Environmental Science and Public Health, Wenzhou Medical College, Wenzhou, 325035, China. E-mail: mhz.gis@gmail.com; Tel: +81 0577 8669 9595*
[b]*Department of Land, Air and Water Resources, University of California, Davis, California, 95616, USA*

Environmental impact

With the rapid economic development in China, water quality has been compromised. Our study showed that over 90% rivers in the watershed have been polluted, of which Wen-Rui Tang River is one of the major polluted rivers in the city. In order to benefit water resource management, protection and remediation, an efficient monitoring system should be developed. In this study, we used a longitudinal water quality monitoring approach combined with statistical method to optimize the locations of discrete monitoring sites. The analysis results showed that our method is both scientifically and economically valuable for locating representative monitoring sites for water quality assessment in Wen-Rui Tang River watershed.

statistical methods, genetic algorithms, and GIS tools have been used to identify monitoring sites.[10–21] Khalil provided a comprehensive review of statistical approaches for assessing and designing water quality monitoring networks.[22]

An important goal of water quality monitoring at the regional scale is to adequately characterize source area contributions with the fewest number of sample locations, therefore minimizing the monitoring cost. A source-search protocol has historically been employed as an efficient initial step to help identify spatial patterns in water quality parameters at discrete (fixed) sites within a given watershed.[23] The source-search technique involves periodic grab sampling (synoptic sampling) of representative sub-basins to determine concentrations/fluxes from the range of physical environments contained within the region. The search-source methodology becomes challenging at land-use interfaces as water quality conditions can change rapidly in both spatial and temporal dimensions.[24–27] Thus, a dense monitoring network would be required in areas with rapidly changing water quality conditions, such as highly populated regions, highly industrialized areas, or rapidly changing land-use conditions.[8] On the contrary, a single monitoring site may be sufficient in areas with stable water quality conditions. Therefore, methodologies for establishing effective monitoring networks remain a critical issue in water quality monitoring.

To address the multiple challenges associated with water quality monitoring across land-use interfaces (urban, agricultural, wetland), we examined the use of continuous longitudinal monitoring as a technique for optimizing spatial allocation of discrete monitoring sites within a regional water quality monitoring program. Longitudinal water quality acquisition is easily accomplished by coupling modern water quality sonde/sensor instrumentation and precision global positioning systems (GPS). Water quality sondes with multiple sensors can be attached directly to a boat for evaluation of the upper water column or submersible pumps can be attached to a boom at one or more depths to pump water through a flow-through cell containing a water quality sonde that is housed on the boat. Thus, multiple depths can be profiled in a single pass to determine if the water column is stratified. Water quality sondes can be programmed to collect data at a rapid frequency allowing detailed spatial analysis; for example ~20 m resolution for a travel velocity of 7 km h^{-1} using a 10 second data acquisition time. Location data can be synchronized with either an internal or an external GPS unit. At the end of a monitoring run, water quality and GPS data can be rapidly downloaded into a GIS for spatial analysis.

The primary objective of this paper is to demonstrate the use of continuous longitudinal water quality monitoring to optimize the location of discrete water quality monitoring sites in a study area located at an urban–agricultural–wetland interface in eastern China. The continuous longitudinal water quality data were incorporated into a GIS platform to facilitate a hierarchical cluster analysis to identify locations for implementation of a traditional synoptic water quality monitoring program. This approach provides a statistically defensible method to optimize discrete monitoring sites (reduce redundancy and eliminate data gaps) without losing key statistical information.

2. Methods

2.1 Study area

The Wen-Rui Tang River watershed is located at Wenzhou, Zhejiang Province, on the east coast of China (Fig. 1). The mainstem of the Wen-Rui Tang River has a length of 20.4 km within the urban area and connects to a network of interconnecting urban waterways with a total length of 1178 km. The average width of the urban portion of the Wen-Rui Tang River is about 50 m, with a range of 13 to 150 m. More than 75% of the watershed consists of a flat alluvial plain with elevations ranging from 3.0 to 4.2 m. The area has a subtropical oceanic climate

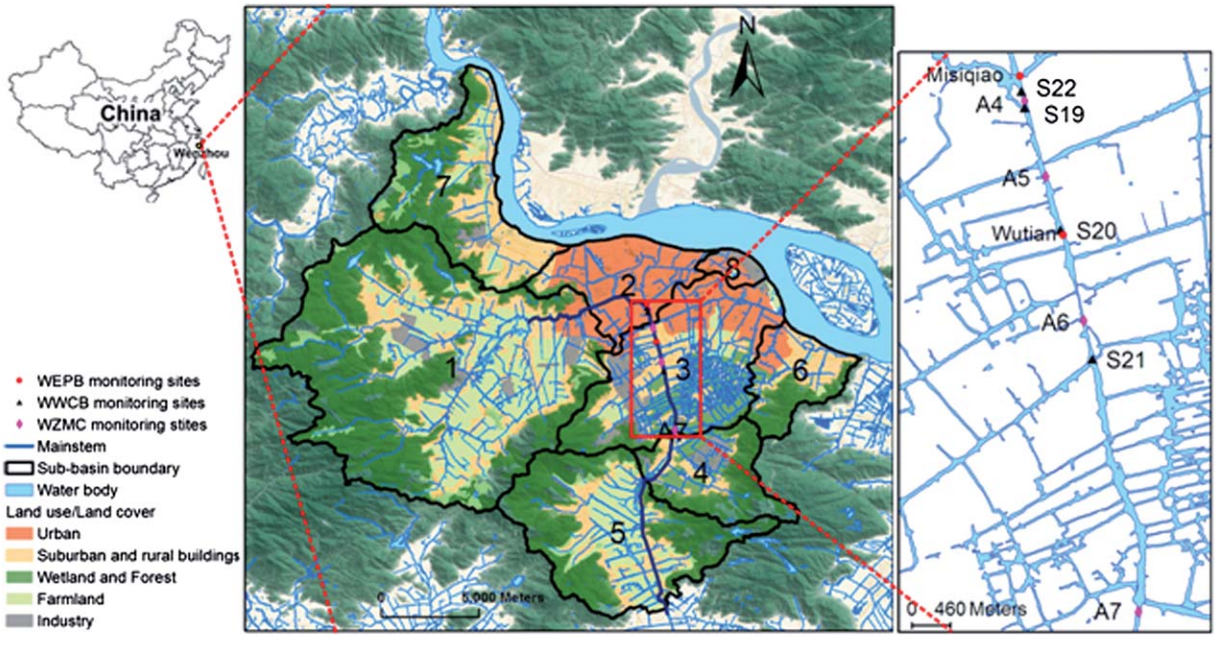

Fig. 1 Land-use map of Wen-Rui Tang River watershed and water quality monitoring sites located in the studied reach.

with an average annual rainfall of 1685 mm and annual runoff of 0.91 billion m^3. This study occurred along a 6.5 km reach on the mainstem of the river that transects urban land, agricultural land, and wetland uses.

Previous fixed-site, water quality monitoring sites were established along the studied reach by the Wenzhou Environmental Protection Bureau (WEPB) since 2000, the Wenzhou Water Conservation Bureau (WWCB) since 2006, and Wenzhou Medical College (WZMC) since 2008 (Table 1). The monitoring sites Misiqiao, S22, A4, and S19 are located in the urban area, stations A5, Wutian, S20, A6, S21 are in the rural–agricultural area, and station A7 is in an area surrounded by wetlands.

2.2 Continuous longitudinal sampling technique for data collection

A continuous, longitudinal water quality sampling technique was used to examine spatial patterns in water quality parameters in the 6.5 km river reach containing the discrete water quality collection sites from previous monitoring programs. A YSI 6920 multi-parameter water quality sonde (YSI Inc., OH, USA) was attached to a boom at a 1 m depth from a motor boat and the location was determined with a Trimble GEO-XH2005 GPS (Trimble Navigation Limited, Sunnyvale, CA). The YSI sonde was used to collect water quality data in real time for pH, water temperature, dissolved oxygen (DO), total ammonia nitrogen (NH_4^+–N), electrical conductivity (EC) and turbidity. Each water quality parameter was calibrated prior to data acquisition and the calibration was verified at the end of the sampling period (~3 h). Sampling of the 6.5 km river reach was conducted six times from March to October 2009 and included representative hydrologic conditions across both wet and dry seasons: 10 March, 28 April, 28 May, 20 July, 29 August and 3 October. The YSI sonde was programmed to take measurements at a 30 second frequency resulting in measurements at about a 20 m interval when the boat was travelling at a velocity of 2.4 km h^{-1}. This sampling protocol resulted in 325 water quality data points along the 6.5 km river reach. Water quality data were continuously examined by the boat operator using a YSI 650 Multiparameter Display System which allowed for immediate re-examination of questionable data (river section was re-measured) to ensure data quality. During selected sampling events, we also pumped water from a 2 m depth, using a submersible pump, through a flow-through cell housing the YSI sonde to determine whether stratification in water quality conditions existed within the water column. No differences were found between the 1 and 2 m depths for this river reach, therefore only the data for the 1 m depth are used in this analysis.

The YSI water quality sonde and the GPS unit were time-synchronized to provide location data for each water quality measurement (precision of ~30 cm). At the end of a sampling event, water quality and GPS data were downloaded into a GIS desktop for spatial analysis. Some location data were lost under bridges where the satellite signal was not available. In these cases, a linear interpolation for the missing sampling sites was calculated from adjacent known locations.

2.3 Data preprocessing

Water quality data were required to be pre-processed before data analysis because differences in the boat speed, travel routes and sampling time interval resulted in differences in the total numbers of sampling measurements of each sampling event: March = 313, April = 381, May = 554, July = 659, August = 494, and October = 187. Since differences in absolute sampling site locations

Table 1 Description of WEPB, WWCB and WZMC sampling stations at the study area[a]

Station	Coordinates	Sampling interval	Water quality parameters
WEPB		Once per two months from Apr. 2000 to Dec. 2010	Water temperature (WT), pH, TSS, EC, DO, COD_{Mn}, BOD, NH_4^+–N, NO_3^-–N, NO_2^-–N, volatile phenol, CN compounds, arsenic (As), mercury (Hg), chromium (Cr), lead (Pb), cadmium (Cd), petroleum, TP.
Misiqiao	120°40′03″E 27°59′47″N		
Wutian	120°40′30″E 27°58′14″N		
WWCB		Once a month from Apr. 2006 to Mar. 2007	DO, BOD, COD_{Mn}, pH, NH_4^+–N, TN.
S22	120°40′00″E 27°59′35″N		
S19	120°40′01″E 27°59′35″N		
S20	120°40′27″E 27°58′35″N		
S21	120°40′44″E 27°57′43″N		
WZMC		Once a month from Jan. 2009 to Dec. 2010	WT, pH, DO, EC, turbidity, NH_4^+–N, NO_3^-–N, NO_2^-–N, PO_4^{3-}–P, TOC, TN, COD_{Cr}, total bacterial, coliform, *E. coli*, *Salmonella*.
A4	120°40′13″E 27°59′26″N		
A5	120°40′20″E 27°59′01″N		
A6	120°40′36″E 27°58′07″N		
A7	120°41′02″E 27°56′11″N		

[a] *Note*: TSS = total suspended solids, EC = electrical conductivity, DO = dissolved oxygen, COD = chemical oxygen demand, BOD = biochemical oxygen demand, TN = total nitrogen, PO_4^{3-}–P = phosphate, TOC = total organic carbon.

hinder statistical analyses, the data were re-sampled by the nearest point method. The middle line of the river was divided into 129 sections with 130 link points. Those link points were taken as the re-sampling sites and the water quality data from nearby sampling sites were assigned to those points. Thus, the final dataset consisted of 130 sample locations for 6 water quality parameters and 6 sample times (data matrix = $130 \times 6 \times 6$). In terms of land use, Sites 1–22 represent the urban area, Sites 23–90 represent the rural–agricultural area, and Sites 91–130 represent the wetland area. The Kolmogorov–Smirnov (K–S) statistic was used to test the goodness-of-fit of the data to a normal distribution using the Tinn-R tool.[28] Data that were not normally distributed were transformed to normality using log transformation for further analysis.

2.4 Statistical analysis

Cluster analysis (CA) is a method of unsupervised learning used to assign a set of objects into clusters so that objects in the same cluster are similar in some aspects.[29] An important step in clustering is to select a distance measure that will determine how the similarity of two objects is calculated. This will influence the formation of the clusters, as some objects may be close to one another according to one distance and farther away according to another.[30] The distance measure chosen in this study was the Euclidean distance which should only be used for expression data that are suitably normalized.

Hierarchical cluster analysis (HCA) was employed to the water quality data matrix to investigate the grouping of continuous longitudinal sampling sites within the 6.5 km study area. It is a general approach to cluster analysis, which provides intuitive similarity relationships between any one object and the entire dataset. A key component of the analysis is the repeated calculation of distance measures between objects, and between clusters once objects begin to be grouped into clusters. The outcome is represented graphically as a dendrogram.[31] The initial data for the hierarchical cluster analysis of N objects are a set of $N \times (N-1)/2$ object-to-object distances and a linkage function for computation of the cluster-to-cluster distances. The two main categories of methods for HCA are divisive methods and agglomerative methods. In practice, the agglomerative methods are of wider use. For each step, the pair of clusters with the smallest cluster-to-cluster distance is fused into a single cluster. The method used in this paper is Ward's Linkage.[32] The linkage function specifying the distance between two clusters is computed as the increase in the "error sum of squares" (ESS) after grouping two clusters into a single cluster. Ward's method seeks to choose the successive clustering steps so as to minimize the increase in ESS at each step. The ESS of a set X of N_X values is the sum of squares of the deviations from the mean value or the mean vector (centroid). For a set X, the ESS is described by the following expression:[33]

$$\mathrm{ESS}(X) = \sum_{i=1}^{N_X} \left| x_i - \frac{1}{N_X} \sum_{j=1}^{N_X} x_j \right|^2$$

where, $|\cdot|$ is the absolute value of a scalar value or the norm (the "length") of a vector. The distance between clusters X and Y is described by the following expression:

$$D(X,Y) = \mathrm{ESS}(XY) - [\mathrm{ESS}(X) + \mathrm{ESS}(Y)]$$

where, XY is the combined cluster resulting from fusion of clusters X and Y; $\mathrm{ESS}(\cdot)$ is the error sum of squares described above. SYSTAT 12.0 (Systat Software, Inc., San Jose, CA) was used for cluster analysis.

A final component of this study was to determine whether water quality parameters measured by the YSI sonde could serve as proxies for more difficult to measure constituents (*e.g.* turbidity for total phosphorus (TP), NH_4^+–N for chemical oxygen demand (COD)). The existence of strong relationships between constituents measured by continuous, longitudinal monitoring would allow for other constituents to be potentially assessed from the continuous, longitudinal monitoring as well. Spearman's rank correlation coefficients were calculated for all possible correlations among measured constituents.

3. Results

3.1 Data collected by longitudinal continuous sampling

The continuous, longitudinal sampling, which was conducted from March to October 2009, reflected the impact of seasonal variation on water quality. The atmospheric temperature during the study period ranged from 6.7 °C in March to 32.9 °C in July. The total precipitation from March to October was 1377 mm, while total precipitation of 2009 was 1611 mm. The highest monthly precipitation occurred in August (404 mm) resulting from a typhoon and approximately 60% of the total annual precipitation occurred between June and September (Fig. 2).

The spatial trends for NH_4^+–N, EC, DO and turbidity are shown in Fig. 3. Water temperature and pH were not evaluated as they displayed no direct relationship with water pollutants in the study area.

Ammonia nitrogen. NH_4^+–N in the study area originates primarily from untreated human and animal wastes entering the waterways. The spatial and temporal patterns of NH_4^+–N reveal serious contamination of surface water, with NH_4^+–N values generally above 2 mg L^{-1} (Fig. 3a). NH_4^+–N concentrations above 2 mg L^{-1} exceed Level V—the maximum level for surface water quality established by Chinese Law.[34] The October 2009 data for NH_4^+–N are not shown because the NH_4^+–N sensor did not meet our quality control requirements. NH_4^+–N concentrations showed a general increase in the downstream direction indicating increased loading during transport through the study area. The lowest NH_4^+–N concentrations were in August following the typhoon, which caused dilution and appreciable flushing of pollutants from the river system. In contrast, the highest NH_4^+–N concentrations were found in late April following an extended dry period.

Electrical conductivity. The cations and anions contributing to EC originate primarily from human (*e.g.*, cooking with salt) and industrial activities. The spatial pattern for EC was similar for each month (Fig. 3b) with increasing EC values to near Site 85 followed by relatively stable values to the end of the study area. The lowest EC values (<250 μS cm^{-1}) occurred in October following the rainy season while the highest EC values (>550 μS

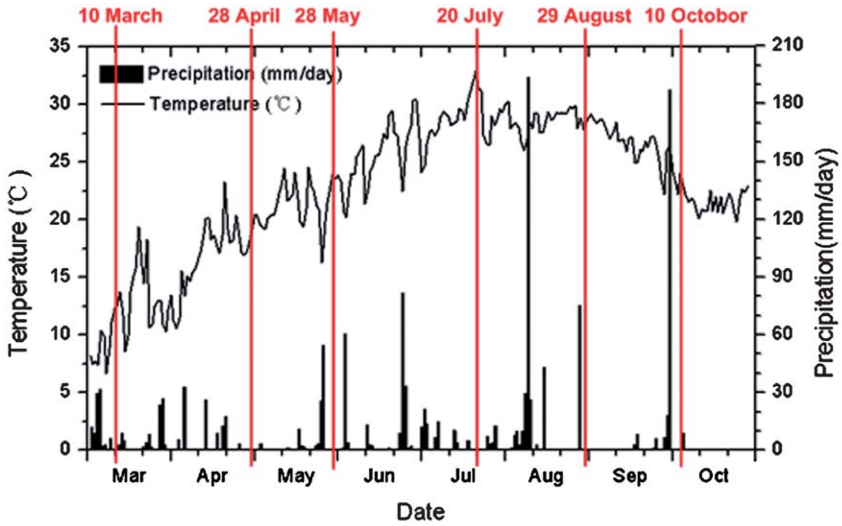

Fig. 2 Time series plot of precipitation and temperature during the study period. Six continuous, longitudinal sampling events conducted on the following dates (red vertical bars): 10 March, 28 April, 28 May, 20 July, 29 August and 3 October in 2009.

cm^{-1}) were in April following the dry season. While the August typhoon appeared to flush and dilute the NH_4^+–N concentrations in the waterways, the typhoon appeared to flush salts into the urban waterways resulting in higher EC values than might be expected.

Dissolved oxygen. DO was consistently below water saturation values for a given water temperature (100% saturation is water in equilibrium with atmospheric oxygen concentration) due to the high biological oxygen demand (BOD) resulting from organic matter and ammonia inputs from human and industrial wastes (Fig. 3c). Most waters were strongly impaired with respect to DO values, falling below the minimum Level V water quality standard for surface waters of less than 20% saturation (less than about 2 mg L^{-1}). Spatially, the lowest DO values (<5% DO saturation) were consistently found around Site 61. Temporally, the highest DO values were in April and appeared to be associated with an active algal bloom in which algae contributed oxygen to the water column through photosynthesis.

Turbidity. Turbidity originates from particulate matter (organic and inorganic) suspended in the water column and is therefore a good proxy for total suspended solids (TSS) in the water column. Since several other water quality parameters are associated with suspended sediments (*e.g.*, pesticides, heavy metals, phosphorus), turbidity may serve as a proxy for several other pollutants. Spatially, turbidity values were relatively similar throughout the study area in March, April, and July, while they showed a general decline in the downstream direction during May and October (Fig. 3d). The large spike in turbidity values in August is related to the typhoon that resulted in large inputs of sediments to the rural–agricultural river segment.

3.2 Classification of sampling sites

Considering sampling sites as cases and NH_4^+–N, EC, DO or turbidity values for each of the six sampling times as variables, the analysis resulted in the grouping of sampling sites into two or three clusters based on the different water quality parameters (Fig. 4). Visual assessment in GIS revealed that the sampling sites in these clusters share similar characteristics (*e.g.* land use, location of point and non-point pollution sources, *etc.*) (Fig. 5). These results indicate that for synoptic sampling of water quality, one site in each cluster is needed to represent a reasonably accurate spatial assessment of each water quality parameter.

Ammonia nitrogen. According to NH_4^+–N values, sampling sites were classified into two clusters (Fig. 4a). Sites 1 to 68 and Sites 84 to 87 were grouped as Cluster 1, while Sites 69 to 83 and Sites 88 to 130 were grouped as Cluster 2. A *t*-test showed that the Cluster 1 mean value was lower than that of Cluster 2 ($p < 0.001$). The *t*-test result confirms that classification of the two sampling clusters was reasonable.

Electrical conductivity. Sampling sites for EC values were also classified into two distinct clusters (Fig. 4b). The EC values in Cluster 1 (Site 1 to 67) were lower than in Cluster 2 (Sites 68 to 130). A *t*-test confirmed a significant difference ($p < 0.001$) in monthly EC values and average EC values between Cluster 1 and Cluster 2.

Dissolved oxygen. DO values segregated into three clusters (Fig. 4c): Sites 1 to 53 identified as Cluster 1, Sites 54 to 89 identified as Cluster 2, and remaining sites identified as Cluster 3. Analysis of variance (ANOVA) was used to test the difference among means for these clusters. The results showed that there were significant differences ($p < 0.001$) between Cluster 1 and Cluster 2 and between Cluster 2 and Cluster 3 ($p < 0.001$), however, no significant difference was identified between Cluster 1 and Cluster 3 ($p > 0.10$). The DO values in Cluster 2 were lower than in the other two clusters.

Turbidity. In contrast to dissolved oxygen, turbidity values displayed a more complex clustering with Sites 1 to 61 classified into Cluster 1, Sites 86 to 102 classified into Cluster 3, and other sites classified into Cluster 2 (Fig. 4d). The results of the ANOVA

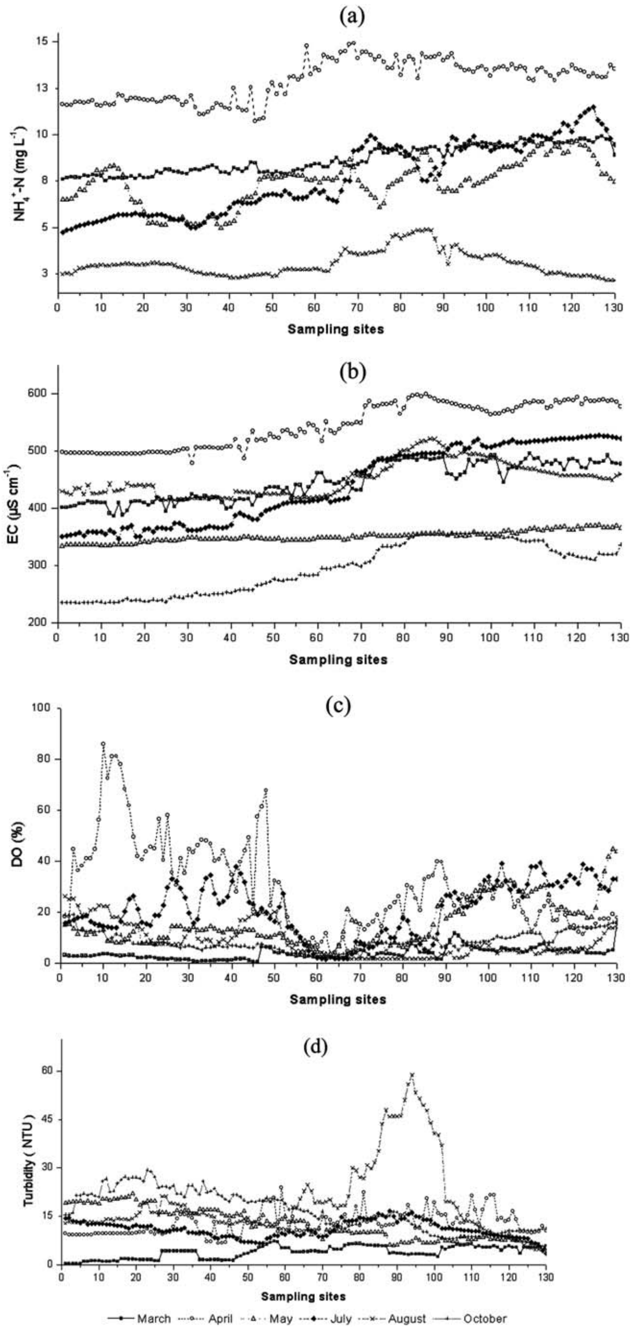

Fig. 3 The spatial and temporal trends of water quality parameter concentration: (a) NH_4^+–N, (b) EC, (c) DO and (d) turbidity.

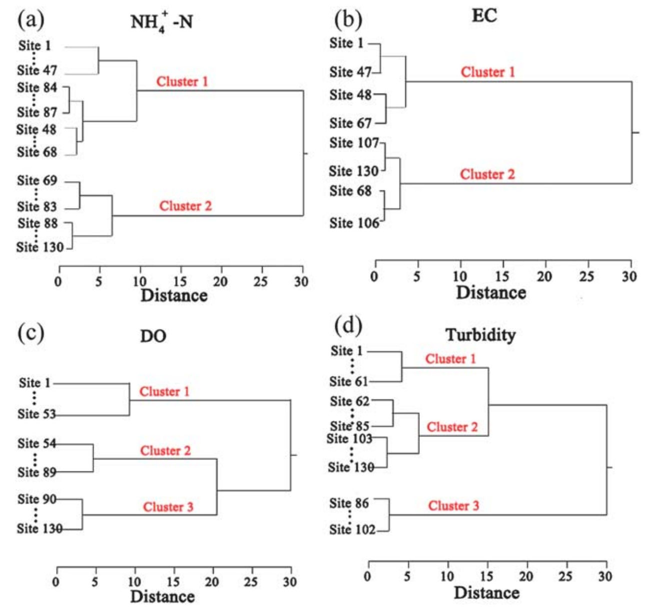

Fig. 4 Dendrogram showing different clusters of sampling sites based on different water quality parameters: (a) NH_4^+–N, (b) EC, (c) DO and (d) turbidity.

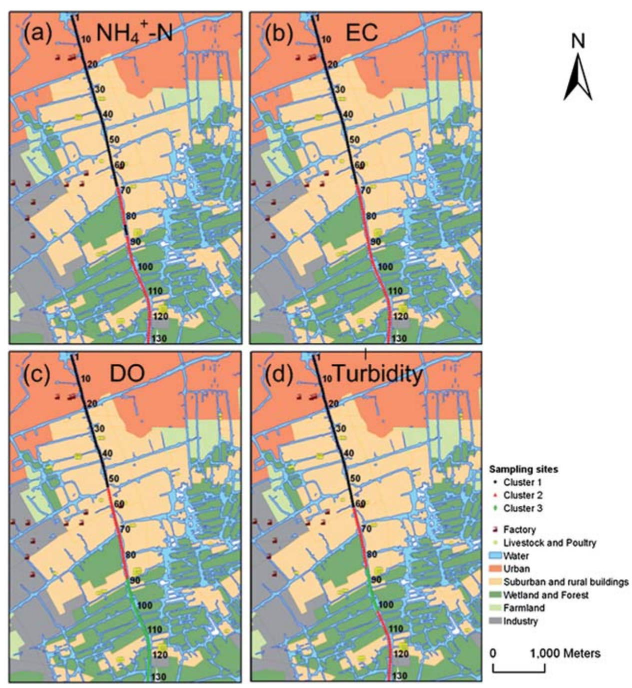

Fig. 5 Spatial distribution of sampling site clusters based on (a) NH_4^+–N, (b) EC, (c) DO and (d) turbidity values.

showed that there were significant differences among these clusters ($p < 0.001$).

3.3 Relationships of water quality parameters

In correlation analysis, longitudinal continuous sampling data and fixed-site sampling data from WZMC monitoring stations (A4, A5, A6 and A7) were used to determine whether the continuous data could be used as a proxy for other water quality constituents. Ten water quality parameters (EC, NH_4^+–N, DO, turbidity, nitrite (NO_2^-–N), nitrate (NO_3^-–N), total nitrogen (TN), phosphate (PO^{3-}–P), total organic carbon (TOC) and chemical oxygen demand (COD_{Cr})) were assessed by Spearman's rank correlation coefficient (Table 2). EC was positively correlated with NH_4^+–N ($r = 0.657$) and TN ($r = 0.607$), and negatively correlated with turbidity ($r = -0.638$). The strong correlation between NH_4^+–N and TN ($r = 0.905$) revealed that NH_4^+–N was the largest contributor to TN, as compared to NO_2^-–N ($r = 0.134$) and NO_3^-–N ($r = -0.171$). COD_{Cr} showed positive correlations with NH_4^+–N, EC, TN, PO_4^{3-}–P, TOC, and

Table 2 Spearman's rank correlation coefficients between the water quality parameters ($n = 39$)

r	NH_4^+–N	EC	DO	Turbidity	NO_2^-–N	NO_3^-–N	TN	PO_4^{3-}–P	TOC	COD_{Cr}
NH_4^+–N	1									
EC	0.657^a	1								
DO	-0.346^c	−0.124	1							
Turbidity	−0.314	-0.638^a	−0.159	1						
NO_2^-–N	0.095	0.082	0.143	−0.046	1					
NO_3^-–N	−0.240	−0.105	0.374^c	−0.203	0.098	1				
TN	0.905^a	0.607^a	−0.269	-0.356^c	0.134	−0.171	1			
PO_4^{3-}–P	0.245	0.510^a	0.172	−0.236	−0.115	−0.075	0.351^c	1		
TOC	0.391^c	0.488^b	0.112	-0.410^b	0.040	0.040	0.526^a	0.589^a	1	
COD_{Cr}	0.573^a	0.577^a	−0.256	0.162	-0.547^a	-0.495^b	0.548^a	0.462^b	0.531	1

[a] Correlation is statistically significant at the $p < 0.001$ level.[b] Correlation is statistically significant at the $p < 0.01$ level.[c] Correlation is statistically significant at the $p < 0.05$ level.

especially with NH_4^+–N ($r = 0.573$) and EC ($r = 0.577$). A negative correlation existed between COD_{Cr} and NO_3^-–N ($r = -0.485$) reflecting the loss of nitrate to denitrification under anoxic conditions imposed by high oxygen demand. There were no obvious correlations between DO, turbidity and other parameters.

These relationships suggest that the continuous monitoring data (EC, NH_4^+–N, DO, turbidity) have limited value for predicting other water quality constituents (only COD_{Cr} and TN from either EC or NH_4^+–N) at this study site. However, many of these relationships will be site specific, therefore continuously monitored parameters could possibly serve as proxies for other water quality constituents in other systems.

4. Discussion

4.1 Spatial and temporal patterns of water quality as affected by land-use types

The continuous data collected by longitudinal sampling with the water quality sonde provide a rigorous evaluation of spatial and temporal patterns for the Wen-Rui Tang River water quality assessment. Based on these data, it was our objective to provide a statistically defensible approach to optimize discrete sampling sites for a synoptic monitoring program along the 6.5 km study area so as to reduce redundancy and eliminate spatial data gaps. The HCA results showed that the physical and chemical parameters evaluated had spatial patterns similar to those identified in a previous study.[35] The spatial distribution of pollutants appeared to be associated with different land-use types, such as urban, agricultural, and wetland. Fig. 3 indicates that each parameter shows an infection point (whether up or down) around Site 60. The main reasons for this pattern include: (1) this region is the transitional zone from urban/residential to agriculture land use, which means that human activities and pollution source types are different, and (2) the waterway in this region passes by a wetland district and the interaction between the river and wetland affects several water quality parameters.

4.2 Optimization of the synoptic monitoring network

The results of HCA effectively captured the spatial patterns for water quality constituents. EC and NH_4^+–N concentrations were correlated ($r = 0.657$) in this study and provided a similar set of two clusters: Cluster 1 = Sites 1–68, Cluster 2 = Sites 69–130. EC and NH_4^+–N concentrations were relatively stable until they entered the rural area around Site 60 (Fig. 5a and b). At this point, concentrations tended to increase reflecting the contribution from agricultural non-point source pollution and sewage discharged to the water body without sewage treatment. This causes the water quality deterioration for EC and NH_4^+–N in Cluster 2. The cluster analysis indicated that one site in each of the two clusters is enough to reflect the NH_4^+–N and EC characteristics of the whole 6.5 km study area. Due to the strong correlation between NH_4^+–N and EC with COD_{Cr}, monitoring for COD_{Cr} should also follow a similar spatial strategy.

Three clusters were identified based on DO% saturation values. Although DO is affected by many factors, the wetland was found to have a major effect on the spatial patterns of DO in the study area. DO values were lowest in Cluster 2 when the river flowed through the rural–agricultural area, and then water purification by the wetland caused the DO value to increase after Site 90 (Fig. 5c). This results in the three clusters being separated as before, in and after the rural–agricultural area, respectively. This result also implies that for effective synoptic monitoring, only one station in each cluster is needed to represent the distinct DO segments of the study area.

The sample site clusters based on turbidity were somewhat different from the results based on EC, NH_4^+–N and DO. The sampling sites before the river that flowed through the rural–agricultural area around Site 61 were classified as one cluster as identified for EC and NH_4^+–N; and points in the Sites 62–85 segment were a bit lower and formed a different cluster. Once the water entered into the rural–agricultural area, turbidity values increased and formed a separate cluster. After the water passed through the wetland area (after Site 103), turbidity values decreased and this segment was clustered into the same group as Sites 52–85 (Fig. 5d). However, Cluster 3 was separated from Cluster 2 because of the high values in August. The sampling date in August was the day after a heavy rain which would lead to surface erosion. The HCA result without turbidity values of August was two significantly different clusters ($p < 0.001$): Sites 1–64 in Cluster 1 and Sites 65–130 in Cluster 2. Therefore, a minimum of two synoptic monitoring sites is required to characterize the turbidity conditions within the 6.5 km study area.

Based on the integrated analysis for all four continuously monitored parameters, it appears that establishing synoptic sampling sites near Sites 30, 70 and 110 would be sufficient to spatially characterize these constituents. We suggest that the existing monitoring stations established by the three independent monitoring programs could be redesigned to reduce costs. The WEPB program has two monitoring stations in the range of Sites 1 to 60 that belong to Cluster 1. Thus, it would be more effective to move one of these two stations to the range of Sites 70 to 130. The WWCB program could remove one station from its four existing stations. One practical plan is that they can keep stations S20 and S21, then move one station to the range between Sites 110 and 130. The WZMC program could also eliminate one station to save the analysis cost. This could be achieved by keeping A5 and moving A6 down to the range of Sites 70 to 90, and moving A7 up to the range of Sites 110 to 130.

5. Conclusion

The use of continuous longitudinal monitoring using water quality sondes was demonstrated to be an effective method for acquiring data that can be used with HCA to identify discrete sites for synoptic water quality monitoring networks. The technique is especially useful along river reaches that experience rapid changes to water quality conditions due to changing land use (*e.g.*, urban–agricultural–wetland interfaces). This approach also provides a statistically defensible method for evaluating the effectiveness of current synoptic water quality networks and to optimize designs to reduce redundancy and eliminate data gaps. Significant correlations between water quality parameters measurable by water quality sondes and other water quality constituents that can be measured only in the laboratory could allow sonde-monitored parameters to serve as proxies for other constituents. New advances in water quality sonde technologies (*e.g.*, sensors for chlorophyll, blue-green algal, nitrate, hydrocarbons, oil, and colored organic matter (CDOM)) will continue to enhance the capabilities of continuous longitudinal water quality data acquisition.

Acknowledgements

Authors thank the following people for data collection and chemistry analysis: Xuyang Zhang, Ping Lu, Yujin Zhang and Bibo Long. This study was supported by the projects of Wenzhou City (No. 20082780125) and Science and Technology Department of Zhejiang Province (No. 2008C03009).

References

1 L. H. MacDonald, A. W. Smart and R. C. Wissmar, *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska*, U.S. Environmental Protection Agency, Region 10, Seattle, Washington, 1991.
2 K. H. Reckhow, A. S. Donigian, J. R. Karr, J. Mandrup-Poulsen, H. S. McDonald, V. Novotny, R. A. Smith and C. O. Yoder, *Assessing the TMDL Approach to Water Quality Management*, National Academy Press, Washington, DC, 2001.
3 F. S. Tirsch and J. W. Male, in *Proceedings of the Twentieth Annual Conference of American Water Resources Associations*, ed. T. M. Schad, AWRA Publications, 1984.
4 R. C. Ward, *J. Am. Water Resour. Assoc.*, 1996, **32**, 673–680.
5 R. O. Strobl and P. D. Robillard, *J. Environ. Manage.*, 2008, **87**, 639–648.
6 R. C. Ward, J. C. Loftis and G. B. McBride, *Environ. Manage.*, 1986, **10**, 291–297.
7 N. B. Harmancioglu, O. Fistikoglu, S. D. Ozkul, V. P. Singh and M. N. Alpaslan, *Water Quality Monitoring Network Design*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999.
8 S. K. Ning and N. B. Chang, *J. Environ. Monit.*, 2002, **4**, 121–126.
9 J. R. Skalski and D. H. McKenzie, *J. Environ. Manage.*, 1982, **14**, 237–251.
10 S. E. Cieniawski, J. W. Eheart and S. Ranjithan, *Water Resour. Res.*, 1995, **31**, 399–409.
11 D. Dutta, A. D. Gupta and V. Ramnarong, *Ground Water Monit. Rem.*, 1998, **18**, 139–147.
12 P. Reed, B. Minsker and A. J. Valocchi, *Water Resour. Res.*, 2000, **36**, 3731–3741.
13 S. K. Ning and N. B. Chang, *Environ. Monit. Assess.*, 2004, **91**, 145–170.
14 O. Ying, *Water Res.*, 2005, **39**, 2621–2635.
15 S. S. Shieh, J. Z. Chu and S. S. Jang, *Math. Geol.*, 2005, **37**, 29–48.
16 S. Y. Park, J. H. Choi, S. Wang and S. S. Park, *Ecol. Modell.*, 2006, **199**, 289–297.
17 R. Strobl, P. Robillard, R. Day, R. Shannon and A. McDonnell, *Environ. Monit. Assess.*, 2006, **122**, 319–334.
18 R. Strobl, P. Robillard, R. Shannon, R. Day and A. McDonnell, *Environ. Monit. Assess.*, 2006, **112**, 137–158.
19 A. Dhar and B. Datta, *Ground Water*, 2009, **47**, 806–815.
20 I. T. Telci, K. Nam, J. Guan and M. M. Aral, *J. Environ. Manage.*, 2009, **90**, 2987–2998.
21 F. Masoumi and R. Kerachian, *Environ. Monit. Assess.*, 2009, **161**, 247–257.
22 B. Khalil and T. B. M. J. Ouarda, *J. Environ. Monit.*, 2009, **11**, 1915–1929.
23 R. A. Dahlgren, K. W. Tate and D. S. Ahearn, in *Environmental Instrumentation and Analysis Handbook*, ed. R. D. Down and J. H. Lehr, John Wiley & Sons, Inc., NY, 2005, pp. 547–564.
24 D. S. Ahearn, R. W. Sheibley, R. A. Dahlgren, M. Anderson, J. Johnson and K. W. Tate, *J. Hydrol.*, 2005, **313**, 234–247.
25 B. R. Scanlon, R. C. Reedy, D. A. Stonestrom, D. E. Prudic and K. F. Dennehy, *Global Change Biol.*, 2005, **11**, 1577–1593.
26 V. S. Benson, J. A. VanLeeuwen, J. Sanchez, I. R. Dohoo and G. H. Somers, *J. Environ. Qual.*, 2006, **35**, 421–432.
27 C. E. Torgersen, R. E. Gresswell, D. S. Bateman and K. M. Burnett, *Spatial Identification of Tributary Impacts in River Networks*, John Wiley & Sons, Ltd, 2008.
28 Tinn-R Department Team, *Tinn-R: an Editor for R Language and Environment Statistical Computing*, 2004.
29 D. L. Massart and L. Kaufman, *The Interpretation of Analytical Data by the use of Cluster Analysis*, Wiley, New York, 1983.
30 . M. Otto, in *Analytical Chemistry*, Wiley-VCH, Weinheim, Germany, 1998, p. 916.
31 M. Forina, C. Armanino and V. Raggio, *Anal. Chim. Acta*, 2002, **454**, 13–19.
32 C. H. Romesburg, *Cluster Analysis for Researchers*, Krieger Publishing Company, Malabar, Florida, 1984.
33 J. H. Ward, *J. Am. Stat. Assoc.*, 1963, **58**, 236–244.
34 Sepa, China, 2002, vol. China SEPA GB3838-2002.
35 P. Lu, K. Mei, Y. Zhang, L. Liao, B. Long, R. Dahlgren and M. Zhang, *Environ. Monit. Assess.*, 2010, **180**, 501–520.