ISSN : 0974 - 7435 Volume 9 Issue 5

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 9(5), 2014 [181-187]

The information mining of biological monitoring data in a heavily polluted river: The comparison between shannon-wiener diversity index and the multivariate analysis

Feng Li[1]*, Xiang-Yun Zeng[1], Xiao-Lin Long[1], Yan-Yan Lang[1], Yan-Mao Wen[2]

[1]School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510630, (CHINA)

[2]The Institute of Environmental Science, Sun Yat-sen University, Guangzhou, 510275, (CHINA)

ABSTRACT

The information mining of biological monitoring data is important for environment monitoring and accessment. Although Shannon-Wiener diversity index (SWI) has been widely used to explain the results of aquatic biological monitoring previously, it runs into problems in heavily polluted rivers. In this paper, a representative heavily polluted river has been selected, and the samples of sediment, pore water, phytoplankton, zooplankton, and zoobenthos were collected and analyzed, with a view to providing theoretical basis for biological data analysis in heavy polluted area. SWI, the multivariate analysis (combined by two multivariate analysis methods: cluster and Non-matric Multi-dimentional Scaling analysis) were used to analyze biological data of phytoplankton, zooplankton, and zoobenthos, with the results of the physical and chemical monitoring and assessment as reference. The results show that the results of SWI cannot effectively reflect the difference of pollution status of various stations in the heavily polluted river; despite the presence of some problems, multivariate analysis method is more suitable than SWI as far as information mining of biological monitoring in the heavily polluted river is concerned.

KEYWORDS

Information mining;
Biological monitoring;
Heavily polluted river;
Shannon-wiener diversity index (SWI);
Non-matric multi-dimentional scaling (MDS);
Cluster analysis.

INTRODUCTION

The biological monitoring is one comprehensive technology developed with biological method to monitor environmental quality, which can both save funding and illustrate problems easily[1,2]. Aquatic community monitoring is an important part of biological monitoring[3-5], which plays an important role in evaluation of water environment quality[6-8]. The keys and difficulties of biological monitoring are the information mining of biological monitoring data we gain, and analyzing the valuable information among them[9]. Shannon-Wiener diversity index (SWI) has been widely used to explain the results of aquatic biological monitoring previously[10]. However, SWI suffering the problem in a heavily polluted river, which it cannot identify the water polluted degree because of the species and quantities of aquatic animals are scarce and SWI tends to homogenization.

Recently, multivariate analysis method has been applied in some researches[11,12], and certain results have been achieved[13,14], but these studies are not involved in heavy pollution area[15,16]. So far, the application of multivariate analysis to information mining of biological monitoring data in heavily polluted river has been rarely reported. However, this is important especially for those developing countries where many rivers and lakes have become seriously polluted.

In this study, Foshan Waterway, a representative heavily polluted river in Pearl River Delta, has been selected. The samples of sediment, pore water, phytoplankton, zooplankton, and zoobenthos were collected and analyzed. The biological monitoring data were analyzed by SWI, the multivariate analysis (combining two methods of MDS and Cluster analysis)[10,17-19], and the analysis results were compared, with physicochemical monitoring and assessment results of sediment and pore water as the reference, so as to explore the method suitable for biological data analysis in heavily polluted area.

MATERIALS AND METHODS

Study area

Foshan waterway is a municipal river and the major drainage channel in Foshan City, flowing through urban area and major industrial district (Figure 1). With the wastewater discharged into the waterway for many years it has been a heavily polluted river and the main pollution factor was heavy metal.

Sample collection and analysis

According to pollution status of waterway in Foshan[20], 6 representative stations were selected in the river (the specific location of various sampling stations was shown in Figure1). The samples of phytoplankton, zooplankton, zoobenthos and pore water, sediment were collected.

The collection and processing of phytoplankton, zooplankton and zoobenthos were carried out according to Chinese national standards[21]. Porewater were prepared by centrifugation method, NH_3-N of pore water was determined by Nessler's reagent spectrophotometric method (GB7479-87); COD_{Cr} of interstitial water was determined by dichromate method (GB11914-89). The sediment samples were dried, ground into powder, and passed by 100 mesh (150 μm) sieve. About 0.5 g powder was taken to determine the content of heavy metals according to national standards of China[22].

Data analysis and analytical quality control

Shannon-Wiener diversity index were calculated by Excel 2003[10]. The two multivariate statistical analysis, Cluster analysis and MDS analysis were conducted by SPSS19.0[23,24].

All glassware and plasticware were cleaned by soaking in 10% HNO_3 (v/v) for 24h, followed by soaking and rinsing with deionized water (Milli-Q). All reagents used in the experiment were at least of analytical grade. The blank, replicate, and spiked samples were analyzed in each batch of sampling.

RESULTS AND ANALYSIS

The environmental conditions of the river based on physicochemical monitoring and assessment

Although physicochemical monitoring has long been a practical routine in environmental monitoring and assessment, it has been recognized that biological monitoring may provide a more comprehensive situation of environmental contamination[7,14,17]. However, the physicochemical monitoring and assessment results could still offer the reference for biological monitoring. Indeed, the results combined with biological monitoring and physicochemical monitoring can provide more reliable results than the results considering only either one[25].

Thinking that most of the heavy metals may deposit and accumulate in sediments in the riverbed, contaminated sediments and pore water in sediments have selected as the main object of the study[26,27]. In order to effectively reflect the environmental conditions, a total of 10 pollution indexes with great impact on aquatic, Pb, Cd, Hg, Cr, As, Cu, Ni, Zn of sediment, NH_3-N and COD_{Cr} of pore water, were selected as parameters to characterize ecological risk, and Hakanson potential ecological risk index method was used to evaluate the physicochemical monitoring data which were shown in TABLE1 and TABLE2. Physical and chemical monitoring results (TABLE2) showed that the heavy metal content of A4 station was the highest among various stations, 6 kinds of heavy metals were in the

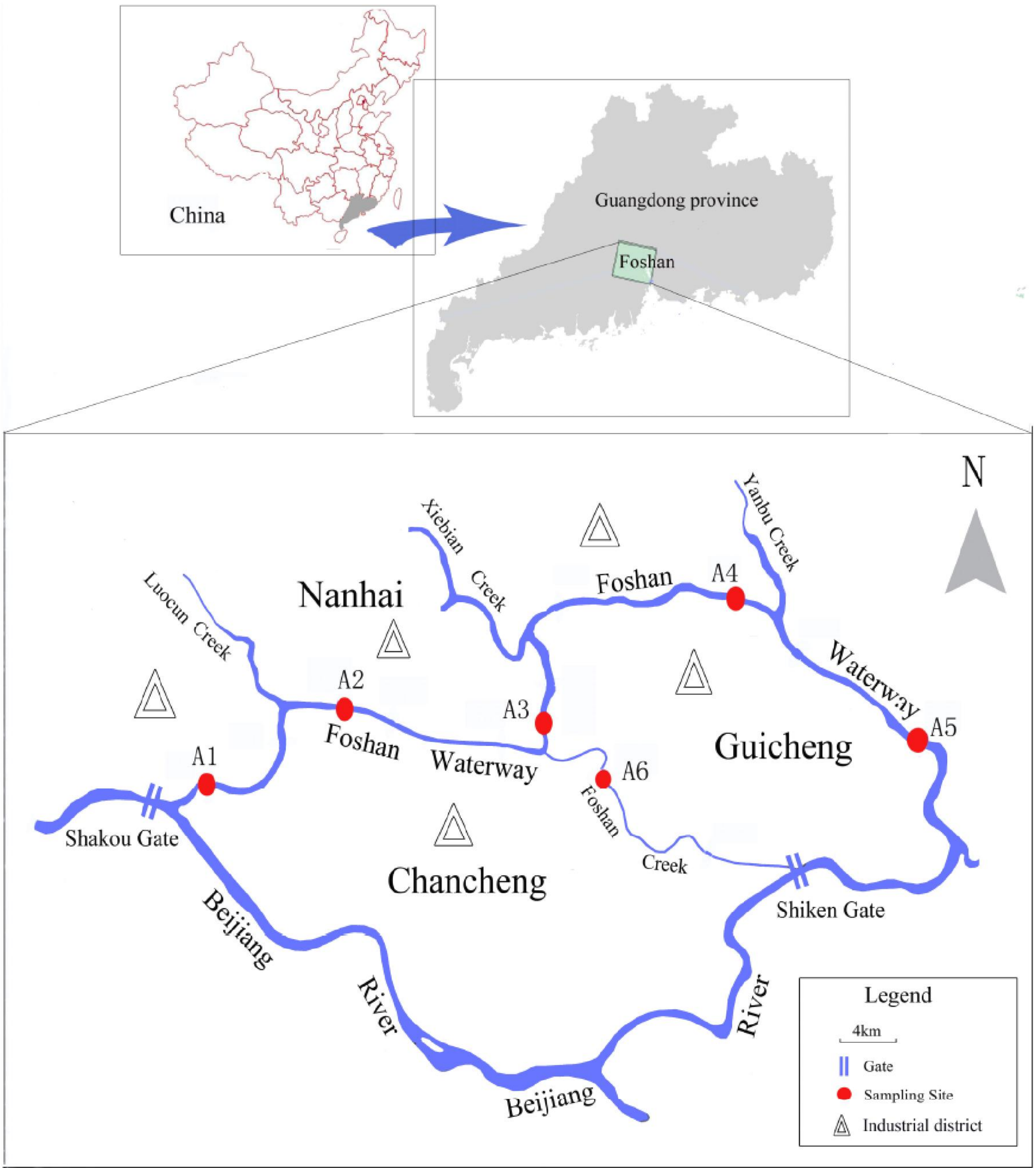

Figure 1 : Locations of the sampling stations and the distribution of the main industrical districts along with Foshan waterway

first place among 8 heavy metals, Hg was in third place, Cu was in second place, COD_{Cr} and NH_3-N in sediment pore water were also high; various pollution indexes in A1 station were the lowest among various stations.

Hakanson potential ecological risk index method uses enrichment degree of sediment heavy metals and its biological toxicity coefficient for weighted sum to get ecological risk index (RI), and RI is adopted as the evaluation basis[20,26-28]. It has been reported that Hakanson potential ecological risk index method not only can reflect the ecological effects of various heavy metals in the research area, but also can distinguish different harm degree of each sampling area and quantitatively divide the degree of potential ecological risk, being a widely used method[20,26,27]. As shown in TABLE 3, RI value of the rest sampling stations already exceeded 600 except A1, so its ecological risk had reached great degree; viewed from a variety of heavy metals, Hg and Cd showed the largest harm degree, RI

FULL PAPER

TABLE 1 : Abundance of zoobenthos, phytoplankton and zooplankton in various sampling stations

Stations	Phytoplankton (10^4 number/L)				Zooplankton (number/L)			
	Cyanophyta	Bacillariophyta	Chlorophyta	Cryptophyta, Euglenophyta	Protozoa	Rotifera	Cladocera	Copepoda
A1	3.12	16.20	13.12	6.19	0.93×10^4	0.50	0.69	1.08
A2	6.86	28.62	8.00	8.27	0.82×10^4	2.17	1.07	1.11
A3	19.01	37.44	61.44	12.53	1.31×10^4	3.80	2.23	2.22
A4	12.83	38.16	109.12	16.22	1.17×10^4	8.61	3.08	4.03
A5	5.232	7.20	13.12	3.12	0.97×10^4	1.94	0.77	0.61
A6	16.63	34.38	50.56	18.10	1.59×10^4	19.79	9.90	10.80

Stations	Zoobenthos (number/m^2)										
	Nemathelminthes, Nematoda sp.	Annelida	Tubifex sp.	Limnodrilus sp.	Branchiura sp.	Naididae sp.	Mollusca	Corbicula fluminea	Cipangopaludina	Oncomelania	Limnoperna lacustris
A1	1802	0	7043	819	246	0	0	737	164	82	82
A2	328	0	737	491	0	0	0	0	0	0	0
A3	655	0	2948	901	0	0	0	164	491	0	0
A4	737	0	2211	1065	410	0	0	328	0	0	0
A5	2744	0	71540	12695	2334	860	0	778	0	369	246
A6	1474	0	2539	819	0	0	0	0	0	0	0

TABLE 2 : The content of heavy metals in sediment and COD_{Cr} and NH_3-N concentration in pore water

Stations	sediment mg/kg								pore water mg/L	
	Pb	Cd	Hg	Cr	As	Cu	Ni	Zn	COD_{Cr}	NH_3-N
A1	61.5	1.26	0.02	89.6	5.77	42.9	52.2	129.0	27.2	8.7
A2	172.6	3.70	3.2	359.8	38.76	186.1	61.4	445.7	220.3	29.8
A3	258.6	2.79	6.12	385.1	20.51	371.8	114.3	562.8	537.0	73.6
A4	299.9	8.5	5.2	450.9	47.9	360.4	121.1	1364.4	269.9	18.4
A5	88.0	1.8	3.5	93.8	32.5	73.7	57.9	225.1	54.0	11.7
A6	166.7	2.52	6.00	124.8	32.97	116.2	64.1	720.9	318.4	40.2

of Hg in various stations was greater than 320 except A1, belonging to extremely strong ecological harm. This result was consistent with the actual observation. Viewed from field situation, the water in most rivers of Foshan waterway generally has black smelly phenomenon, aquatic species are scarce, and fishes and shrimps are disappeared. According to the evaluation results of Hakanson potential ecological risk index method, A4 station had the largest potential ecological risk, and A1 station had the smallest. The pollution degree of each sampling station from heavy to light was as follows: A4 > A3 > A6 > A2 > A5 > A1, and the results were consistent with field observation and physicochemical monitoring results. Taking A4 and A1 as examples, A4 station was affected by sewage outfalls such as Yuelishayong, Libian sluice, Lubian sluice and Xinglong sluice, and electroplating industrial zone was once nearby, with high pollution level; A1 station was located in the upstream of Foshan waterway, which was near Shakou sluice and not influenced by the pollution, and the ecological environmental quality was the best in Foshan waterway.

The assessment results of SWI in the river

TABLE 3 : The potential ecological risk index of heavy metals in sediments

Stations	Pb	Cd	Hg	Cr	As	Cu	Ni	Zn	Risk Index
A1	12.30	75.60	3.20	2.99	3.85	7.15	8.42	1.61	115.12
A2	34.52	222.00	512.00	11.99	25.84	31.02	9.90	5.57	852.84
A3	51.72	167.40	979.20	12.84	13.67	61.97	18.44	7.04	1312.28
A4	59.98	510.00	832.00	15.03	31.93	60.07	19.53	17.06	1545.60
A5	17.60	108.00	560.00	3.13	21.67	12.28	9.34	2.81	734.83
A6	33.34	151.20	960.00	4.16	21.98	19.37	10.34	9.01	1209.40
Risk Index	209.46	1234.20	3846.40	50.14	118.94	91.86	75.97	43.10	

According to the evaluation results of zoobenthos by SWI (TABLE 4), the pollution degree of various stations from heavy to light was A5 > A6 > A2 > A1 > A3 > A4. This result varied greatly with physical and chemical monitoring results and evaluation results of Hakanson potential ecological risk index. With A4 and A5 stations as examples, the actual observation and physicochemical monitoring results showed that A4 station was the most seriously polluted station; A5 station was located in the end of Foshan waterway and communicated with Guangzhou section of Pearl River, which had relatively few pollution sources in the surrounding, with light pollution degree. As shown in TABLE 2, various pollution indexes of A5 station were only higher than A1 station, which was the second to last among various stations, and its pollution degree was significantly lower than that in A4. As shown in TABLE 4, the evaluation results of zoobenthos, phytoplankton and zooplankton by SWI had many contradictions. According to evaluation results of zooplankton by Shannon-Wiener diversity index, A6 station had the most serious pollution degree, which A1 station showed the lightest pollution degree; the pollution degree of various stations from heavy to light was A6 > A4 > A5 > A2 > A3 > A1. According to evaluation results of phytoplankton by SWI, A4 station had the most serious pollution degree, which A6 station showed the lightest pollution degree; the pollution degree of various stations from heavy to light was A4 > A2 > A3 > A1 > A5 > A6. The results showed great difference. A6 station had the most serious pollution according to zooplankton evaluation results, which became the cleanest according to phytoplankton evaluation results. Zooplankton and phytoplankton had great relevance, and the appearance of so big difference was not normal.

To sum up, the above mentioned evaluation results of SWI had a larger gap with the results of the results of physicochemical monitoring and assessment. In addition, from the analysis results of SWI, the diversity index of various stations basically had no difference, which was very weak to judge the pollution degree difference. Consequently, the results in the river showed that SWI may not be suitable for analysis of biological monitoring results in heavily polluted area.

TABLE 4 : The calculation results of Shannon-Wiener diversity index

Stations	Phytoplankton	Zooplankton	Zoobenthos
A1	1.77	1.94	1.70
A2	1.70	1.89	1.51
A3	1.76	1.90	1.76
A4	1.50	1.70	1.99
A5	1.81	1.85	1.14
A6	1.85	1.69	1.44

The assessment results of multivariate analysis in the river

The difficulty of biological monitoring in heavily polluted rivers lies in the rare aquatic organisms[1,2]. The situation often occur that half of biological data in matrix data is zero. According to this characteristic, the choice of method to measure the similarity of the biological samples shall not be subject to the situation[25]. The MDS analysis based on the Bray-Curtis similarity coefficient, one of multivariate analysis, has been regarded as is a good method to deal with the problem[17]. When combined with Cluster analysis, anther multivariate analysis method, the MDS analysis can better explore biological monitoring data in heavily polluted rivers[10-12]. The steps are as follows. First of all, based on Bray-Curtis similarity measurement, hierarchica1 cluster of group average clustering method was used to draw the clustering tree diagram (Figure 2), and MDS was used to analyze. To make the results have more practical significance, MDS analysis results should compare with hierarchical clustering analysis results, and the corresponding cluster group should be sketched in MDS graph (Figure 3).

Through analysis, the evaluation results of plankton by multivariate analysis (Figure 3a and Figure 3b) were basically consistent with Hakanson potential ecological risk index method (Table 3). The stations ranked the first third places in Hakanson potential ecological risk index method were divided into the group with heavy pollution, and the stations ranked the last third places were divided into the group with light pollution, the grouping condition of phytoplankton and zooplankton was basically consistent (Figure 2 a, b and Figure 3 a, b). But the grouping condition of zoobenthos had difference with plankton in A2 and A3 stations (Figure 2c, Figure 3c): according to survey results of plankton, A2 station belonged to the group with light pollution, and A3 station belonged to the group with heavy pollution, but this was just opposite to the survey results of zoobenthos. From field observation, A3 station was greatly affected by Foshan sewage, which has a soy sauce factory nearby, and the river was often in black and odorous condition, so its pollution status was severe than A2 station. From chemical monitoring results (TABLE2), Cd and As indexes of A3 station was slightly lower than A2 station, and the other pollution indexes in A3 station were all higher than A2 station. From analysis results of Hakanson potential ecological risk index method (TABLE 3), the pollution degree of A3 station was higher than that of A2. Therefore, the pollution degree of A3 station was higher than A2 station from various aspects, and the survey results of zoobenthos might have error. The reason might be that the appearance of flood in June washed down the species such as *Corbicula fluminea* and *Cipangopaludina chinensis* from relatively clean water in upstream. Thus, multivariate statistical analysis method avoided the mistake by SWI, and its result was more intuitive to express by diagram.

From theoretic analysis, multivariate analysis method is based on sample similarity, which judges the difference of pollution degree in various stations through similarity determination of community structure, so it has wide application range, and can reflect the water environment quality from clean water to heavy polluted river (contamination resistant species are difficult to grow).

Although multivariate analysis method is not perfect, the test results show that the method is superior to the Shannon-Wiener diversity index method in biological monitoring results analysis in heavy polluted area.

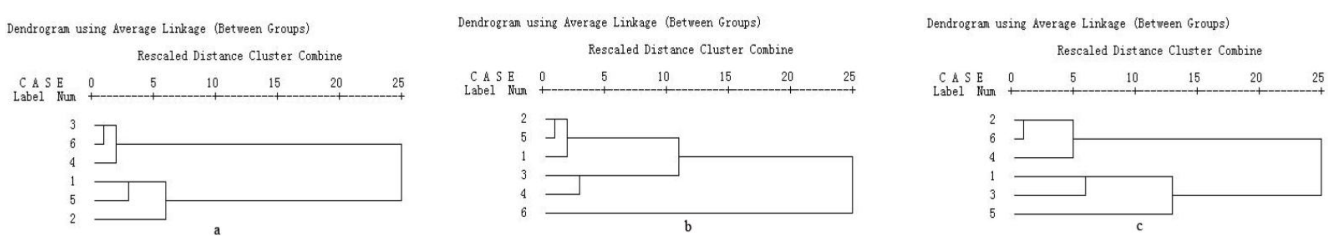

Figure 2 : Cluster dendrogram phytoplankton (a), zooplankton(b) and zoobenthos(c) were used hierarchical cluster of group average clustering method based on Bray-Curtis similarity mesurement

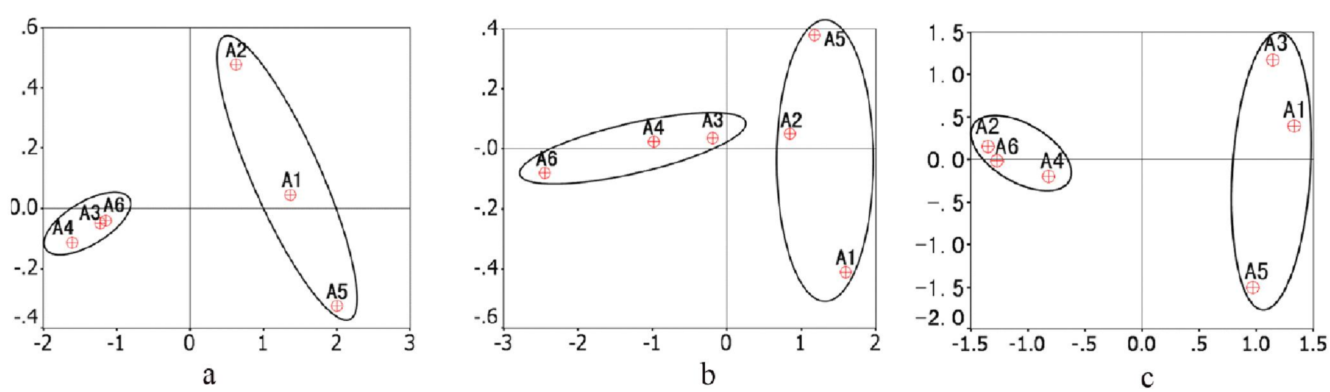

Figure 3 : Non-metric multivariate scaling (MDS) analysis of phytoplankton (a), zooplankton(b) and zoobenthos(c)

CONCLUSIONS

SWI analysis results cannot effectively reflect the difference of pollution status of various stations in the heavily polluted river judging from the fact that the analysis results have larger gap with actual pollution status. Despite the presence of some problems, multivariate analysis method is more suitable than SWI as far as information mining of biological monitoring in the heavily polluted river is concerned.

ACKNOWLEDGEMENTS

Thank for the help of Hui-juan Tang, Jin-ying Wu, Shu-you He, Guang-hui Yu, Ge Mai, and Xian-ge Wei. This work was partly founded by National Natural Science Foundation of China (No.41001341), Nature Science Foundation of Guangdong Province, China (No. 9151401501000015), Basic Scientific Research Project of China's Ministry of Education (No. 2012ZM0082), and Open Foundation of State Key Laboratory of Subtropical Science (No. 2011KB12).

REFERENCES

[1] N.Z.Arman, Salmiati, M.I.M.Said, S.Azman, S.A.M.Safeai;. IPCBEE, **35**, 52-56 **(2012)**.
[2] C.Colosio, F.M.Rubino, A.Alegakis, E.Ariano, G.Brambilla, S.Mandic-Rajcevic, F.Metruccio, C.Minoia, A.Moretto, C.Somaruga, A.Tsatsakis, R.Turci, F.Vellere; Toxicology Letters, **213**, 49- 56 **(2012)**.
[3] L.Stabili, C.Caroppo, R.A.Cavallo; Environmental Monitoring and Assessment, **121**, 301-323 **(2006)**.
[4] M.M.Ali; Management of Environmental Quality, **19**, 377-401**(2008)**.
[5] P.O.Raburu, F.O.Masese, C.A.Mulanda; Aquatic Ecosystem Health&Management, **12**, 197-205 **(2009)**.
[6] F.Ishaq, A.Khan; Pelagia Research Library, **3**, 18-31 **(2013)**.
[7] F.Ishaq, A.Khan; Biotechnology: An Indian Journal, **7**, **(2013)**.
[8] E.Jeppesen, P.Noges, T.A.Davidson, J.Haberman, T.Noges, K.Blank, T.L.Lauridsen, M.Sondergaard, C.Sayer, R.Laugaste, L.S.Johansson, R.Bjerring, S.L.Amsinck; Hydrobiologia, **676**, 279-297**(2011)**.
[9] M.A.Deyab, A.H.A.Khedr, M.A.EI-Naggar; Algological Studies, **112**, 123-140 **(2004)**.
[10] R.E.Kwiatkowski; Hydrobiologia, **68**, 247-256 **(1980)**.
[11] J.B.R.Agard, J.Gobin, R.M.Warwick; Marine Ecology Progress Series, **92**, 233-243**(1993)**.
[12] A.P.Muchaa, M.Teresa, S.D.Vasconcelos, A.A.Bordalo; Marine Environment Research, **60**, 531-550 **(2005)**.
[13] H.A.Hernaandez Arana, R.M.Warwick, M.J.Attrill, A.A.Rowden, Gerardo Gold-Bouchot; Marine Ecology Progress Series, **289**, 89-107 **(2005)**.
[14] J.Hyland, L.Balthis, I.Karakassis, P.Magni, A.Petrov, J.Shine, O.Vestergaard, R.Warwick; Marine Ecology Progress Series, **295**, 91-103 **(2005)**.
[15] M.G.Chapman; Marine Ecology Progress Series, **162**, 71-78 **(1998)**.
[16] D.Tagiiapietra, M.Paean, C.Wager; Estuarine, Coastal and Shelf Sciences, **47**, 217-226 **(1998)**.
[17] F.Li, Y.M.Wen, P.T.Zhu; Bull Environ Contam Toxicol, **81**, 90-94 **(2008)**.
[18] M.Bozzetti, U.H.Schulz; An index of biotic integrity based on fish assemblages for subtropical streams in southern Brazil. Hydrobiologia, **529**, 133-144 **(2004)**.
[19] K.R.Clarke, R.M.Warwick; Change in marine communities: An approach to statistical analysis and interpretation, Bourne Press, Bournemonth, **(1994)**.
[20] F.Li, X.G.Wei, Y.Y.Hang; China Water&Wastewater, **22**, 88-91 **(2006)**.
[21] Water and Wastewater Monitoring Method Editorial Committee, State Environmental Protection Administration of China, China Environmental Science Press, **(2002)**.
[22] China soil society; Soil agriculture chemical analysis method, China Agriculture Science and Technology Press, Beijing, **(2000)**.
[23] Y.M.Wu; Environmental statistics, China Environmental Science Press, Beijing, **(2002)**.
[24] Z.F.Cheng, F.C.Xu; Environmental data statistical analysis, Chemical Industry Press, Beijing, **(2006)**.
[25] A.Gomiero, A.Dagnino, C.Nasci, A.Viarengo; Science of the Total Environment, **442**, 534-544 **(2013)**.
[26] H.M.Fernandes; Environmental Pollution, **97**, 317-325 **(1997)**.
[27] D.M.D.Toro, J.A.Mcgrath, D.J.Hansen, W.J.Berry, P.R.Paquin, R.Mathew, K.B.Wu, R.C.Santore; Environmental Toxicology and Chemistry, **24**, 2410-2427 **(2005)**.
[28] L.H akanson; Water Research, **14**, 975-1001 **(1980)**.