Water Quality Characterization and Index Optimization

-The Case of Agricultural Rivers in Southern Ehime, Japan-

Edward Lapong¹, Masayuki Fujihara², Tomoki Izumi³, Kunihiro Hamagami⁴,
Noriyuki Kobayashi⁵ and Toshiko Kakihara⁶

Abstract: Water quality characterization is important, and often, the first step in watershed water quality management. The characterization activity centers on the quantification of water quality parameters relevant to the monitoring objective and the analyses of the impact of the parameters on the state of water body. This study characterizes the general water quality of a network of agricultural rivers in Southern Ehime, Japan and conducted Factor Analysis (FA) and Cluster Analysis (CA) to determine the causal factors and spatial variation of pollution. The Water Quality Index (WQI), based on nine water quality parameters, was computed and optimized—determining the parameters that could best predict the overall water quality. Results show that the river sampling sites have good overall water quality and that physico-chemical parameters pH, DO, and BOD₅ best correlated and predicted WQI. Based on the FA, the primary factors that influence pollution in the sites are the solids or sediment transport (turbidity and sediment), inorganic pollution (NO₃-N, PO₄-P) and trophicity (total N, total P, fecal bacteria). CA, on the other hand, somehow provides information on the effect of the tributaries to the receiving rivers.

Keywords: Water quality; Water quality index (WQI); Optimization, Factor Analysis; Cluster Analysis

1 Introduction

The monitoring and analysis of water quality are the key elements of water quality characterization and are primarily important in the evaluation of the environmental status of water bodies. Information on water quality and pollution sources is important for the implementation of sustainable water-use management (Bu *et al.*, 2010; Crosa *et al.*, 2006; Sarkar *et al.*, 2007).

In rural areas, agricultural activities often influence the river water quality, whether it is pollutants from point and non-point sources. Some of the identified effects of agricultural production processes on any river body involve general water quality deterioration due to nutrient enrichment, destruction of spawning grounds for aquatic life, general fish kill and deterioration of its aesthetic value.

The evaluation of water quality has become a critical issue recently and, although water monitoring for different purposes is well-defined, the overall water quality is sometimes difficult to evaluate due to large number of samples and varying concentrations for many parameters (Chapman, 1992; Rauch *et al.*, 1998).

The use of Water Quality Index (WQI) is a simple

method that overcomes problems in water quality evaluation and could provide on-hand information about the changes and trends of water quality. WQI is a value computed from a set of parameters that represents the overall water quality and pollution level. Unlike mathematical and computational modeling that requires adequate knowledge of hydraulics and hydrodynamics and requires extensive validation (Chapman, 1992; Shanahan *et al.*, 1998; Somlyódy *et al.*, 1998), WQI is a simple and understandable tool that attempts to provide a mechanism for presenting a cumulatively derived numerical expression defining water quality (Miller, 1986).

The problems of data reduction and interpretation, characteristics change in water quality parameters and parameter indicator identification can be approached through the use of some statistical tools. Factor Analysis (FA) is a multivariate statistical technique used to identify important factors that explains most of the variances of a system. It is designed to reduce the number of variables into a small number of indices while preserving the original relationship of the variables (Davis, 1986; Manly, 1986; Wackernagel, 1995).

Another statistical tool found useful in environmental monitoring is the Cluster Analysis (CA). It is a method that uses squared Euclidian distances to segregate or desegregate factors in a given system.

Cognizant to the above premises, this study aims to provide insights on the applications of the WQI and statistical techniques, *i.e.* FA and CA, to characterize and optimize river water quality and to provide baseline information of the present state of the rivers' water quality. Specifically, WQI is used to determine the general water quality and its optimum correlation with other water quality parameters. FA is employed to determine the factors that primarily

¹Ph.D Student, The United Graduate School of Agricultural Sciences, Ehime University, 3-5-7 Tarumi, Matsuyama, Ehime 790-8566, Japan

²Professor, Faculty of Agriculture, Ehime University, 3-5-7 Tarumi, Matsuyama, Ehime 790-8566, Japan (Corresponding Author) Email: fuji-hara@agr.ehime-u.ac.jp

³Assistant Professor, Faculty of Agriculture, Ehime University, 3-5-7 Tarumi, Matsuyama, Ehime 790-8566, Japan

⁴Associate Professor, Faculty of Agriculture, Iwate University, 3-18-8 Ueda, Morioka, Iwate 020-8500, Japan

⁵Associate Professor, Faculty of Agriculture, Ehime University, 3-5-7 Tarumi, Matsuyama, Ehime 790-8566, Japan

⁶Lecturer, Institute of Education and Student Support, Ehime University, 3 Bunkyo-cho, Matsuyama, Ehime 790-8577, Japan

Figure 1: Location of monitoring and sampling sites

caused the pollution, and CA to show the sampling sites clusters in terms of pollution.

Particularly, this study aims to assess and optimize the prediction of overall water quality in a network of agricultural rivers in Southern Ehime Prefecture, Japan which flows into a relatively clean Shimanto River (referred to as the '*last clear stream of Japan*'). The Shimanto river is located in the adjacent Kochi Prefecture and its water quality has been observed to be deteriorating due to the polluted waters of its tributaries.

2 Materials and Methods

2.1 Study site

The study site is a network of agricultural river located in Southern Ehime Prefecture, Japan. It is located between latitudes $33^{\circ}8' \text{ N}$ and $33^{\circ}20' \text{ N}$, and between longitudes $132^{\circ}34' \text{ E}$ and $132^{\circ}53' \text{ E}$. The rivers are classified as small (N, M, H) to large rivers (Shimanto), according to United Nations Environment Programme (UNEP) and World Health Organization (WHO) classification. In fact, N river flows into M river, M river flows into H river, and H river flows into the Shimanto river. (**Figure 1**)

N, M and H rivers have a watershed area of 25 km^2 , 76 km^2 , and 186 km^2 , respectively; main stream length of 8 km, 17 km, and 37 km, respectively. N and M rivers are in the 3rd stream order and H river in the 4th stream order, classifying them as small rivers. The Shimanto river, on the other hand, has a watershed area of $2,270 \text{ km}^2$ and a stream length of 196 km.

H river is one of the biggest tributaries of the Shimanto river, located in the latter's middle course and apparently discharges sediment-laden and polluted waters, especially during the months of April to September which corresponds to the agricultural production season [particularly rice planting] in the areas. Its waters flow into the Shimanto river that apparently deteriorates the latter's water quality, tarnishing its reputation as the '*last clear stream of Japan*'.

The rivers' watersheds are basically agricultural with considerable area devoted to rice production. Rice produc-

tion is the biggest agricultural activity from April to September, with land preparation activities starting as early as the end of March. Rice production season corresponds to the spring and summer seasons which are also the period of high rainfall and typhoons, aggravating the high-sediment effluents from the paddy rice fields.

In terms of rice paddy area, N river has 1.1 km^2 , M has 6.9 km^2 , and H has 3.4 km^2 , representing 80%, 88% and 70% of their arable area, respectively.

2.2 Data monitoring and collection

The river discharge and water quality were monitored periodically from April 2010 to September 2011: biweekly during the agricultural production season (April-September) and once a month during October-March.

Table 1: Water quality parameters

Parameters	Monitoring/Analysis Methods
Discharge, Q	data logger
Turbidity, T	bench turbidimeter
Total Suspended Solids, TSS	filtration-oven
Temperature, ΔT	<i>in-situ</i>
pH	<i>in-situ</i>
Conductivity, Cond	<i>in-situ</i>
Dissolved oxygen, DO	<i>in-situ</i>
Biochemical Oxy. Demand, BOD ₅	residual DO
Total Fecal Coliform Bacteria, FCB	cfu plate count
Total Nitrogen, TN	spectrophotometric
Nitrate, NO ₃ -N	spectrophotometric
Total Phosphorus, TP	spectrophotometric
Phosphate, PO ₄ -P	spectrophotometric

Note: Laboratory analyses are based on APHA (1995) procedures

The parameters, its acronyms as used in this study, and its monitoring/analysis method are summarized in **Table 1**.

Discharge is determined by an established discharge rating curve and monitoring the hourly water level using a data logger. Physico-chemical water quality parameters (T,

pH, Cond, DO) are monitored *in-situ* using a multi-parameter sensor probe (W-22XD Horiba) and water samples are manually collected for laboratory analysis of the other parameters. Temperature for every sampling site is measured twice: in the water collection site and in a similarly-conditioned site 1 mile upstream. The difference represents ΔT and is used in the computation of water quality index.

There were a total of 200 data sets for every parameter during the monitoring period—25 for each sampling sites.

2.3 Data analysis

The overall water quality of the rivers is represented by Water Quality Index (WQI). WQI was computed using the National Sanitation Foundation Water Quality Index, NSF WQI (Ott, 1978; Dunnette, 1979; Miller et al., 1986). NSF WQI is one of the most popularly-used indices and considered 'highly acceptable' for general water quality evaluation as it is derived from a combination of physical, chemical, and bacteriological indicators (Karami, *et al.*, 2009). It has been widely tested and applied to data from a number of geographical areas all over the world (Samantary, *et al.*, 2009).

NSF WQI is computed using 9 parameters: T, TSS, ΔT , pH, DO, BOD₅, FCB, NO₃-N and PO₄-P. The mathematical expression and equivalent rating for WQI is given by,

$$WQI = \sum_{i=1}^n I_i W_i \quad (1)$$

where I_i is the index value for i th water quality parameter, W_i is the weight of each parameter in terms of importance, and n is the number of parameters (1, 2, ..., 9). The index value is a standardized rating value for each parameter with range 0–100, 0 for worst and 100 for best quality. It is based on established 'Q-curves' with actual parameter values as abscissa and 0–100 rating as ordinate. The weight of the parameters are given as: T (0.08), TSS (0.07), ΔT (0.10), pH (0.11), DO (0.17), BOD₅ (0.11), FCB (0.16), NO₃-N (0.10) and PO₄-P (0.10).

A correlation analysis, using the non-parametric Spearman rank correlation (r_s), was conducted using actual parameter and index values to determine the statistically significant correlated parameters and the parameters which are highly correlated with the WQI. The result of the correlation was also used to determine the optimum model equations for WQI. Correlation analysis was done (a) using aggregate values of all sampling sites and (b) using separate values for every sampling site—to determine location-specificity of the parameters.

Factor Analysis (FA), a multivariate statistical technique, was conducted to determine the water quality parameters that highly influence the pollution in the rivers. FA identifies the parameters that explain most of the variances of the system. In mathematical terms, FA involves the following steps: (1) coding the variables $x_1, x_2, \dots, x_n$, to have zero means and unit variance, (2) calculate the covariance matrix C , (3) find the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ and the corresponding eigenvectors $a_1, a_2, \dots, a_n$, (4) discard components that

only account for a small proportion of the variation of the dataset, and (5) develop the factor loading matrix to infer principal parameters (Ouyang, *et al.*, 2006).

In this study, all variables except pH, were logarithmically transformed to normalized the data distribution. Only the principal factors with eigenvalues greater than 1 were considered for interpretation. Also, the factors were rotated twice using Varimax method to reduce the factors into 2 and get a 2-dimensional representation for the vectors of the parameter factor loading or coefficient.

Moreover, an Agglomerative Hierarchical Cluster Analysis using complete linkage method was conducted to assess the pollution spatial variation, that is, similarity of the pollution level of the sampling sites.

All data analyses—correlation, equation model optimization, FA, CA, statistical—were conducted with the aid of XLStat program software version 2011.

3. Results and Discussion

3.1 Water quality parameters and WQI correlation

The correlation matrix in **Table 2**, showing Spearman rank correlation between parameter index and WQI, showed that WQI correlates best with physico-chemical parameters, *i.e.* pH, DO, and BOD₅. It is less correlated with chemical and bacteriological parameters NO₃-N, PO₄-P, FCB; least with physical parameters T and TSS; and not at all with temperature. Among parameters, highly significant correlation ($r_s > 0.4$, $\alpha = 0.01$) exists between T:TSS, T:FCB, TSS:FCB, BOD₅:FCB, FCB:PO₄-P. From these, it could be inferred that using parameter indices, the parameters that correlate best with other parameters are T, TSS, and FCB.

Using actual parameter values, the correlation analysis (**Table 3**) showed that WQI correlates best with physico-chemical parameters, *i.e.* BOD₅, DO and conductivity. It is less correlated with TSS, FCB, TP, PO₄-P; least with Q and T; and not at all with temperature. Among parameters, the highly and significantly correlated parameters ($r_s > 0.4$, $\alpha = 0.01$) are Q:Cond, Q:NO₃-N, T:TSS, T:FCB, T:TN, T:TP, T:PO₄-P, TSS:Cond, TSS:TP, TSS:PO₄-P, pH:FCB, DO:BOD₅, DO:FCB, BOD₅:FCB, FCB:TP, FCB:PO₄-P, TN:NO₃-N, TN:TP, TN:PO₄-P, and TP:PO₄-P. From these, it is apparent that the parameters which correlate best with other parameters are T, TSS, FCB and PO₄-P.

The correlation analyses suggest that WQI correlates better with parameter indices, as it gives higher and readily interpretable coefficients. This may be due to the normalization of data which gives all the parameter a direct positive relationship with WQI. This is not the case when using actual parameter values, as the parameters have different trend relationship with WQI: *i.e.* T, TSS, BOD, FCB, N, and TP have negative relationship with WQI; DO has positive relationship with WQI; while pH has positive relationship with WQI at pH < 7.5 and negative at pH > 7.5.

On the other hand, correlation among individual water quality parameters is better when using actual parameter values. This is due to the special relationship existing between water quality parameters which diminish when they are transformed into parameter indices.

Table 2: Spearman rank correlation matrix of parameter index values and WQI considering aggregate data set

Variables	T	TSS	ΔT	pH	DO	BOD ₅	FCB	NO ₃ -N	PO ₄ -P	WQI
T	1									<i>0.155</i>
TSS	-0.595	1								-0.154
ΔT	0.053	-0.060	1							0.132
pH	-0.213	-0.005	0.071	1						0.634
DO	0.053	-0.001	0.086	<i>0.143</i>	1					0.464
BOD ₅	-0.127	0.064	-0.106	0.065	0.057	1				0.437
FCB	0.556	-0.409	0.132	0.025	0.115	-0.436	1			0.228
NO ₃ -N	0.314	-0.306	-0.020	0.197	0.013	-0.002	0.325	1		0.284
PO ₄ -P	0.379	-0.337	<i>0.174</i>	0.007	0.125	-0.205	0.479	<i>0.158</i>	1	0.192
WQI	<i>0.155</i>	-0.154	0.132	0.634	0.464	0.437	0.228	0.284	0.192	1

Values in boldface are significant correlation at $\alpha=0.01$; boldface-italicized at $\alpha=0.05$

Table 3: Spearman rank correlation matrix of actual parameter values and WQI considering aggregate data set

Variables	Q	T	TSS	Temp	pH	Cond	DO	BOD ₅	FCB	T-N	NO ₃ -N	T-P	PO ₄ -P	WQI
Q	1													<i>0.181</i>
T	0.111	1												-0.160
TSS	-0.166	0.585	1											-0.241
Temp	-0.114	-0.265	-0.282	1										-0.068
pH	0.226	-0.344	-0.055	0.124	1									-0.132
Cond	-0.402	<i>0.191</i>	0.607	-0.204	0.269	1								-0.262
DO	-0.160	-0.067	0.126	0.015	0.163	0.266	1							-0.359
BOD ₅	-0.193	-0.127	0.085	-0.012	0.270	0.307	0.746	1						-0.433
FCB	0.049	0.544	0.282	-0.118	-0.432	-0.076	-0.458	-0.438	1					-0.234
T-N	-0.245	0.417	0.252	-0.027	-0.227	0.088	-0.194	-0.080	0.335	1				-0.152
NO ₃ -N	-0.621	<i>0.196</i>	0.253	0.036	-0.351	0.270	0.143	<i>0.174</i>	0.075	0.488	1			-0.145
T-P	-0.229	0.554	0.593	-0.155	-0.072	0.346	-0.366	-0.209	0.554	0.651	0.369	1		-0.207
PO ₄ -P	-0.133	0.494	0.483	-0.073	-0.090	0.182	-0.381	-0.209	0.562	0.548	0.301	0.905	1	-0.208
WQI	<i>0.181</i>	-0.160	-0.241	-0.068	-0.132	-0.262	-0.359	-0.433	-0.234	-0.152	-0.145	-0.207	-0.208	1

Values in boldface are significant correlation at $\alpha=0.01$; boldface-italicized at $\alpha=0.05$

Table 4: Highly-significant correlated parameters ($r_s \geq 0.6$, $\alpha \leq 0.01$)

All Sites	Sampling Sites							
	1	2	3	4	5	6	7	8
TP:PO ₄ -P ^a	DO:BOD ^c	TP:PO ₄ -P ^a	TP:PO ₄ -P ^b	TP:PO ₄ -P ^b	TP:PO ₄ -P ^b	TP:PO ₄ -P ^b	TP:PO ₄ -P ^b	TP:PO ₄ -P ^a
DO:BOD ^c	DO:TP ^c	Q:T ^d	TN:TP ^c	DO:BOD ^c	DO:BOD ^c	Q:TP ^c	Q:T ^b	Q:T ^b
TN:TP ^d	DO:PO ₄ -P ^c	TN:TP ^d	DO:BOD ^c	Q:PO ₄ -P ^d	Q:PO ₄ -P ^c	Q:FCB ^c	DO:BOD ^c	Q:FCB ^c
Q:NO ₃ -N ^d	FCB:TP ^c	DO:BOD ^d	DO:PO ₄ -P ^c	Q:TP ^d	Q:FCB ^c	FCB:PO ₄ -P ^c	T:FCB ^d	DO:BOD ^d
	FCB:PO ₄ -P ^c	T:FCB ^d	Q:pH ^c	TN:TP ^d	Q:TP ^d	DO:BOD ^c	DO:TP ^d	TN:PO ₄ -P ^d
	TP:PO ₄ -P ^c		T:pH ^c		FCB:PO ₄ -P ^d	Q:PO ₄ -P ^d	Q:FCB ^d	
	Q:T ^d		Q:FCB ^c			FCB:TP ^d	TSS:FCB ^d	
	Q:FCB ^d		pH:FCB ^d			DO:PO ₄ -P ^d	T:TSS ^d	
			DO:TP ^d			TN:TP ^d		
			FCB:PO ₄ -P ^d			DO:TP ^d		
						Q:DO ^d		

Note: ^a ($r_s > 0.9$), ^b ($0.9 \leq r_s > 0.8$), ^c ($0.8 \leq r_s > 0.7$), ^d ($0.7 \leq r_s > 0.6$)

Separate analysis in the respective river sampling sites gives significantly higher correlation coefficients and more correlated parameters, as compared to aggregated data. In this case, the parameters best correlated with WQI are the physico-chemical parameters pH, DO, and BOD. And among parameters, TP:PO₄-P, DO:BOD and Q:T were found to be the best correlated parameters (Table 4). A high TP:PO₄-P and DO:BOD correlation could be expected as these parameters as naturally closely related—PO₄-P

being a component of TP and BOD₅ being based upon DO. On the other hand, a high Q:T correlation suggests that Q would be a good predictor for turbidity.

Also from Table 4, by individual parameter: Q is best and highly correlated with T, FCB, and PO₄-P; T with Q and FCB; DO with BOD, TP, and PO₄-P; BOD with DO; FCB with Q, TP, and PO₄-P; TN with TP; TP with PO₄-P, Q, DO, and FCB; and PO₄-P with TP, FCB, and DO.

Table 5: Average WQI of the sampling sites from April 2010 to September 2011

Month/Period	Sampling Sites							
	1	2	3	4	5	6	7	8
All Period	75.9	76.7	76.0	75.7	77.7	75.1	78.5	78.2
April-September	75.1	76.1	75.8	75.3	77.3	75.3	78.3	78.0
October-March	77.1	77.9	76.3	76.5	78.3	74.7	78.7	78.6

Table 6: NSF WQI rating

WQI	Rating
0 ~ 25	Very Bad
25 ~ 50	Bad
50 ~ 70	Medium
70 ~ 90	Good
90 ~ 100	Excellent

3.2 WQI of river-sampling sites

The average WQI of the sampling sites considering all period and two seasons in a year is shown in **Table 5**. Based on the NSF WQI rating criterion (**Table 6**), the sampling sites have 'Good' water quality.

The average WQI among the river-sampling sites shows no apparent significant difference, although the Shimanto river (Sites 7, 8) consistently has the highest values, suggesting that it is the cleanest and the least polluted.

However, during April-September which is the agricultural production season and the period of rice planting, the sampling sites have relatively lower WQI values, implying that they are less clean or more polluted during this period

3.3 WQI optimum prediction models

The optimization procedure involves the selection of equation model [$WQI = f(\text{water quality parameter})$] that yields the highest correlation, highest model efficiency, and least error. Fundamental functions, *i.e.* linear, log-linear, power, logarithmic, exponential, polynomial, were tested using

parameter indices and actual parameter values. Single factor regression analysis was done, treating the parameters as single independent variable affecting WQI.

The optimization analysis conducted revealed that equation models developed using parameter indices as independent variables are better. This coincides with the correlation analysis results in *Section 3.1*. Hence, when predicting WQI, the actual parameter values should first be transformed to equivalent parameter indices.

The result of the optimization process is presented in **Table 7**. It shows that the physico-chemical parameters pH, DO, and BOD₅ are the best predictors of WQI. This is in concordance with the non-parametric Spearman correlation analysis in which these are the parameters that best correlates to WQI. On the other hand, it was also shown that DO is the best predictor-parameter when considering aggregate data, while it is pH considering respective sampling sites, with the exception of Sampling Sites 4, 6 and 8 (DO) and Sampling Site 5 (BOD₅).

The optimum model equation types are either linear or quadratic, with Nash-Sutcliffe model efficiency coefficients ranging from 0.370 to 0.717. Correlation coefficients (*r*) and error (CV-RMSE) have a closer range at 0.632~0.846 and 0.01~0.028, respectively.

3.4 Factor Analysis

The Factor Analysis (FA) was carried out on XLStat program software using actual parameters. Actual parameter values (and not parameter indices) were used in order to preserve the original and exact relationship of the parameters which would somehow change or diminish when

Table 7: Optimum correlated parameter with prediction equation, correlation coefficient *r*, Nash-Sutcliffe model efficiency coefficient *E_f*, and coefficient of variation-root mean squared error *CV (RMSE)*

River/Site	Parameter	Equation	<i>r</i>	<i>E_f</i>	CV(RMSE)
All	DO	$WQI = 0.222DO + 56.06$	0.705	0.497	0.026
1	pH	$WQI = 0.137pH + 64.69$	0.608	0.370	0.024
2	pH	$WQI = -0.010pH^2 + 1.546pH + 19.04$	0.764	0.584	0.028
3	pH	$WQI = 0.001pH^2 - 0.032pH + 72.43$	0.654	0.428	0.021
4	DO	$WQI = 0.219DO + 55.94$	0.828	0.687	0.025
5	BOD ₅	$WQI = 0.005BOD^2 - 0.717BOD + 102.56$	0.649	0.419	0.017
6	DO	$WQI = 0.001DO^2 + 0.039DO + 61.93$	0.846	0.717	0.025
7	pH	$WQI = 0.126pH + 67.99$	0.632	0.399	0.018
8	DO	$WQI = 0.232DO + 56.26$	0.666	0.443	0.018

Table 8: Principal factor eigenvalues and variances

Descriptor	F1	F2	F3	F4	F5	F6	F7
Eigenvalue	3.98	2.32	1.41	1.06	0.45	0.17	0.06
Variability (%)	30.6	17.8	10.9	8.2	3.5	1.3	0.5
Cumulative %	30.6	48.5	59.3	67.5	71.0	72.3	72.8

Table 9: Principal factor loading (F) and derived loading after Varimax rotation (D) showing causal parameters (*values in boldface*)

Parameter	F1	F2	F3	F4	D1	D2
Q	-0.287	-0.577	0.676	-0.251	-0.358	-0.536
T	0.678	-0.007	0.300	-0.422	0.672	-0.094
TSS	0.593	0.385	0.349	-0.106	0.637	0.306
Temp	0.101	-0.096	-0.343	0.341	0.088	-0.108
pH	-0.371	0.228	0.556	0.637	-0.339	0.273
Cond	0.268	0.619	0.216	0.173	0.345	0.580
DO	-0.368	0.737	0.105	-0.342	-0.270	0.778
BOD ₅	-0.275	0.701	0.116	-0.130	-0.183	0.731
FCB	0.658	-0.388	0.006	-0.129	0.603	-0.469
T-N	0.647	0.065	-0.108	-0.020	0.651	-0.018
NO ₃ -N	0.456	0.445	-0.424	-0.078	0.509	0.383
T-P	0.947	0.061	0.192	0.238	0.947	-0.061
PO ₄ -P	0.835	-0.053	0.138	0.215	0.821	-0.160

Figure 2: Scree plot of eigenvalues and variances

transformed into parameter indices.

The variables, except pH, were transformed into log-values for normalization. Only the principal factors with eigenvalues closed to or greater than 1 were considered for interpretation.

The table and scree plot (Table 8 and Figure 2) of eigenvalues and variances showed that the first 4 factors have eigenvalues greater than 1 and accounts for 68% of the total variances of the dataset. F1 and F2 have considerably higher eigenvalues and explain 48% of the variances.

Table 9 presents the factor loading of the different parameters and the derived factor loading values after Varimax rotation.

The first principal factor, F1, explains 31% of the total variances and is highly correlated with solids/sediment transport (T, TSS), inorganic pollution

(NO₃-N, PO₄-P) and trophicity (TN, TP, FCB) factors. These are the parameters that primarily affect the rivers' water quality and could be associated with the communities and agricultural activities along the river course. T and TSS are apparently contributed by the sediment-laden drainage from paddy fields, especially during rice planting season, and from surface and non-point erosion sources. NO₃-N and PO₄-P could be associated with the use of fertilizers and pesticides, whose excesses reach the rivers through runoff and sediment particulates during drainage and storm events. FCB has a strong linkage to the effluents from households and livestock farms.

The second principal factor, F2, which accounts 18% of the total variance is highly correlated to oxygen-related physico-chemical parameters, *i.e.* conductivity, DO and BOD₅. These parameters indicate organic pollution and are secondary factors that affect the river water quality.

The last two principal factors, F3 and F4, are highly correlated to Q and other physico-chemical parameters, *i.e.* temperature and pH. These parameters are associated with salient environmental factors such as precipitation and temperature. They are the tertiary factors that influence the river water quality and could provide insights on the effect of river flow and seasonal change.

After Varimax rotation, the principal factors were reduced to two, eliminating lesser-correlated factors and infusing its causal parameters to a more highly-correlated factor. The first factor (D1) is associated with the sediment/solids transport factors (T, TSS), organic and inorganic (FCB, TN, NO₃-N, TP, PO₄-P) factors, and pH. The second factor (D2) is correlated with discharge and physico-chemical parameters, *i.e.* temperature, conductivity, DO and BOD₅.

Figure 3: Vectors of correlations between factors and water quality parameters

The vectors of factor-parameter correlation coefficient (**Figure 3**) indicates that among F1 factors, TP has the highest influence in the water quality followed by $\text{NO}_3\text{-N}$ and $\text{PO}_4\text{-P}$, while T and TSS has the least influence. Among F2 factors, DO has the biggest influence in the water quality, closely followed by conductivity and BOD.

3.5 Cluster Analysis

The Agglomerative Hierarchical Cluster Analysis was done using the parameter indices and was carried out on the XLStat program software. The assessment of the pollution spatial variation is based on the similarity of the pollution level or rank of the sampling sites.

The cluster analysis is performed using an automatic truncation option, which is based on the entropy of the given set of data and tries to create homogenous group. The dendrogram (**Figure 4**) shows the sampling sites as grouped by the algorithm. It is composed of 3 groups or clusters: *Cluster A* (Sites 3, 4, 6), *Cluster B* (Sites 7, 8) and *Cluster C* (Sites 1, 2, 5). Dissecting Cluster A and with reference to

Table 5, it is opined that Sampling Site 3 is less polluted than Sampling Sites 4 and 6. For Cluster B, Sampling Sites 7 and 8 have the relatively same pollution level. And for Cluster C, Sampling Site 1 is more polluted than Sampling Sites 2 and 5.

The pollution index values (**Table 5**) in association with the clusters would indicate the effect of a more polluted to the lesser polluted river after combination of their flow (at a confluence). The Shimanto river (Sampling Site 7) was affected by the flow of H river, thus diminishing its water quality after their confluence at Sampling Site 8. On the other hand, the water quality of H river (Sampling Site 4) greatly diminished after the inflow of M river to its course at Sampling Site 5.

4. Conclusion

This study presents the use of Water Quality Index (WQI) in water quality characterization and prediction optimization, as well as the application of some statistical tools in the characterization of water quality in agricultural rivers.

The analysis results show that the physico-chemical parameters, *i.e.* pH, DO and BOD, correlates best with the overall water quality, as represented by the WQI. These are also the parameters that could best predict the overall water quality, thus, monitoring of these parameters would be useful to feasibly estimate the water quality.

Transforming the parameter values into indices was found necessary as a prerequisite in water quality prediction, as WQI correlates better with parameter indices than with actual values. This may due to the normalization of data which gives all the parameters a direct positive relationship with WQI. Conversely, in determining correlation among water quality parameters, actual parameter values were found to produce higher and more significant correlation, thus, would be more advantageous. This is due to the special relationship existing between water quality parameters which changes or diminishes when they are transformed into parameter indices.

Moreover, the prediction of water quality is better when conducted separately in each sampling sites, which proves the location-specificity of the water quality.

Factor analysis shows that the water quality parameters

Figure 4: Dendrogram of sampling sites clusters

associated to agricultural activities and domestic effluents, *i.e.* turbidity, suspended solids, and organic and inorganic pollutants, greatly influence water quality of the rivers. Cluster Analysis, on the other hand, somehow provides information on the effect of the tributaries to the receiving rivers, although additional monitoring is needed for better validation.

Acknowledgments

This study was kindly supported and partially funded by the Nanyo Regional Office of Ehime Prefecture government.

The authors would also like to thank the Kihoku Town office (Ehime Prefecture) and Nishitosa branch office of Shimanto City (Kochi prefecture) for the valuable support in the field observation and data gathering.

References

- [1] American Public Health Association-APHA (1995): *Standard methods for the examination of water and wastewater*.
- [2] Bu, H., Tan, X., Li, S., Zhang, Q. (2010): Temporal and spatial variations of water quality in the Junshui river of the South Qinling Mts., China, *Ecotoxicoology and Environmental Safety*, 73, pp. 907-913.
- [3] Chapman, D. (1992): *Water quality assessment*, Chapman and Hall (on behalf of UNESCO, WHO and UNEP), p. 585.
- [4] Crosa, G., Froebrich, J., Nikolayenko, V., Stefani, F., Gallid, P., Calamari, D. (2006): Spatial and seasonal variations in the water quality of the Amu Darya River (Central Asia), *Water Res.* 40, pp. 2237-2245.
- [5] Davis, J.C. (1986): *Statistical and data analysis in geology*, 2nd Edition, Wiley, New York, pp. 250-266.
- [6] Dunnette, D.A. (1979). A geographically variable water quality index used in Oregon. *J. Water Pollution Control Fed.*, 51, 53-61.
- [7] Karami, B., Dhumal, K.N., Golabi, M., and Jaafarzadeh, N. (2009): Optimizing the relationship between water quality index and physical and chemical parameters of water in Bamdezh Wetland, Iran, *J. Applied Science*, 9, pp. 3900-3905.
- [8] Manly, B.F.J. (1986): *Multivariate statistical methods: A primer*. Chapman and Hall, London, pp. 68-85.
- [9] Miller, M.W., Joung, H.M., Mahannah, C.N., and Garrett, J.R. (1986): Identification of water quality differences in Nevada through index application, *J. Environ. Quality*, 15, pp. 265-272.
- [10] Ott, W.R. (1978). *Environmental Indices, Theory and Practice*. Ann Arbor Science Publications, Ann Arbor, Michigan, 230 p.
- [11] Ouyang, Y., Nkedi-Kizza, P., Wu, Q.T., Shinde, D., and Huang, C.H. (2006): Assessment of seasonal variations in surface water quality, *Water Res.*, 40, pp. 3800-3810.
- [12] Rauch, W., Henze, M., Koncsos, L., Reichart, P., Shanahan, P., Somlyody, L., and Vanrolleghem, P. (1998): River water quality modeling I- State of the art, *Wat. Sci. Tech.*, 38, pp. 237-244.
- [13] Samantray, P., Mishra, B., Panda, C., and Rout, S. (2009): Assessment of water quality index in Mahanadi and Atharabanki rivers and Taldanda canal in Paradip Area, India, *J. Human Ecology*, 26(3), pp. 153-161.
- [14] Sarkar, S.K., Saha, M., Takada, H., Bhattacharya, A., Mishra, P., and Bhattacharya, B. (2007): Water quality management in the lower stretch of the river Ganges, East Coast of India: An approach through environmental education, *J. Clean Prod.*, 15, pp. 1559-1567.
- [15] Shanahan, P., Henze, M., Koncsos, L., Rauch, W., Reichart, P., Somlyody, L., and Vanrolleghem, P. (1998): River water quality modeling II- Problems of the art, *Wat. Sci. Tech.*, 38, pp. 245-252.
- [16] Somlyody, L., Henze, M., Koncsos, L., Rauch, W., Reichart, P., Shanahan, P., and Vanrolleghem, P. (1998): River water quality modeling III-Future of the art, *Wat. Sci. Tech.*, 38, pp.253-260.
- [17] Wackernagel, H. (1995): *Multivariate geostatistics: An introduction with applications*, Springer, New York, pp. 121-125.

Discussion open until June 30, 2013