



Water quality index using multivariate factorial analysis

Christiane Coletti¹, Roberto Testezlaf¹, Túlio A. P. Ribeiro¹, Renata T. G. de Souza¹ & Daniela de A. Pereira²

ABSTRACT

The evaluation of environmental effects generated by agricultural production on water quality became essential in Brazil after the creation of policies for the use and conservation of water resources. For such, water quality indices have been considered with the purpose of showing the spatial and temporal variation of water quality in a watershed. The objective of this study was to develop a water quality index (WQI) applying the Multivariate Factorial Analysis (MFA) statistical technique, which could indicate the influence of agricultural activities in the quality of water resources. Water in a predominantly farm watershed was monitored from Sept. 2003 to Sept. 2004. Monthly water collections were carried out at six sample points, and eight parameters were analyzed: nitrate, ammoniacal nitrogen, ammonia, total phosphorus, electrical conductivity, pH, suspended solids and turbidity, which were considered important due to the agricultural management adopted in the region. Results indicated a contamination of agricultural origin along the basin. Factorial analysis showed that ammonia, ammoniacal nitrogen and nitrate parameters were the ones that most contributed in determining the WQI.

Key words: agricultural contamination, water resources, watershed

Índice de qualidade de água usando análise fatorial multivariada

RESUMO

A avaliação dos impactos gerados pela agricultura sobre a qualidade da água tornou-se primordial no Brasil após o estabelecimento das políticas de uso e conservação dos recursos hídricos, razão por que índices de qualidade de água são propostos com o intuito de se mostrar a evolução da qualidade da água no tempo e no espaço. O objetivo primeiro neste estudo, foi desenvolver um índice de qualidade de água (IQA), utilizando-se a técnica estatística da Análise Fatorial Multivariada, avaliando-se a influência das atividades agrícolas na qualidade dos recursos hídricos; realizou-se, para se atingir o proposto, um monitoramento da qualidade da água em uma bacia hidrográfica de uso predominantemente agrícola, no período de set/03 a set/04. Realizaram-se, também, coletas mensais de água em seis pontos, analisando-se oito variáveis (nitrito, nitrogênio amoniacal, amônia, fósforo total, condutividade elétrica, pH, turbidez e sólidos suspensos), julgadas importantes pelo manejo agrícola adotado na região. Os resultados obtidos indicaram uma contaminação de origem agrícola ao longo da bacia, sendo que, pela análise fatorial, amônia, nitrogênio amoniacal e nitrito foram as variáveis que mais contribuíram na determinação do IQA.

Palavras-chave: contaminação agrícola, recursos hídricos, bacia hidrográfica

¹ FEAGRI/UNICAMP, CP 6011, CEP 13083-970, Campinas, SP. Fone(s): (82) 3377-5544; (19) 3521-1024. E-mails: chriscoletti@yahoo.com.br; bob@agr.unicamp.br; tulior@agr.unicamp.br; renata.torres@gmail.com

² IMECC/UNICAMP, CP 6065, CEP 13083-859, Campinas, SP. Fone: (19) 3521-5925. E-mail: dani.araujo@gmail.com

INTRODUCTION

The reduction in the quantity and quality of water resources has become a worldwide concern, including countries with great water potential such as Brazil, since the availability of water is one of the main factors that limits socioeconomic development.

Merten & Minella (2002) observed that the water quality concept is not necessarily a state of purity of the water, but refers to chemical, physical and biological characteristics that determine its different uses. According to Sperling (1996) since the process of qualifying a water resource becomes complex due to the number of parameters involved, water quality indices are proposed with the intent of summarizing the analyzed variables and expressing them in a single number with the objective of showing the temporal and spatial evolution of water quality.

The search for an indicator that best characterizes a water source under study requires the use of statistical techniques. According to Haase et al. (1989), one of the methods used in the formulation of water quality indices is based on the multivariate factorial analysis technique, which was used by Carvalho et al. (2000) to evaluate a watershed water quality. The main objective of this analysis is to study the correlation structure of an initial set of "p" variables ($X_1, X_2 \dots X_p$), replacing it with a smaller set of hypothetical variables, which, lower in number and with a simpler structure, explain most of the variation in the original variables.

This technique permits learning the behavior of data from the reduction of the parameters' original space dimension, thus permitting the selection of the most representative variables for the water resource being analyzed, as explained by Andrade et al. (2007). Factorial analysis thus permits the definition of more sensitive indicators, which can facilitate a monitoring program as well as the evaluation of changes that have taken place in the water resources. Toledo & Nicoletta (2002) demonstrated that the use of the index based on the factorial analysis technique is very useful when one intends to evaluate changes that have taken place in the watershed axis. However, indices based on statistical techniques cannot be generalized for all water sources since the water system has its own peculiar characteristic (Haase et al., 1989).

Irrigated agriculture can modify the quantity and/or the quality of water resources, affecting availability for several purposes. Irrigation is an agricultural technique that is being used more and more to increase crop productivity. According to Paz et al. (2000) the development of irrigated agriculture requires technological and economical procedures to optimize the water use to improve system application efficiency and yield gains based on the crop response to water application and other supplies without compromising the availability and quality of the resource.

The objective of this study is to develop a water quality index (WQI) applying the Multivariate Factorial Analysis (MFA) statistical technique, which could indicate the effects of agricultural activities in the quality of a watershed's water resources.

MATERIAL AND METHODS

The Water Quality Index (WQI) was established using multivariate factorial analysis applied on the water quality data of a watershed, monitored during a 13-month period.

Characterization of the watershed

The Rio das Pedras watershed, located in the cities of Mogi Guaçu and Estiva Gerbi, SP, and which belongs to the Mogi Guaçu River watershed, was chosen as the study area. This watershed was selected due to the significant activities of tomato crop production in the area, with constant applications of agrochemicals associated with furrow irrigation, known for its low application efficiency and water losses. This combination is considered a potential generator of diffused loads that are harmful to the availability of water resources. Thus, six water collection points were chosen, taking into consideration urban nuclei. Care was taken not to gather samples from regions near city districts since the objective was to evaluate only the agricultural impacts on water quality in the Rio das Pedras Watershed. In order to facilitate the location of the selected points, since the distances between points was relatively large, and to permit future researches to be conducted to expand the database, GPS equipment was used to georeference the sample collection points. Figure 1 shows the hydrological map for the Rio das Pedras Watershed in 2004 and the location of the water sample collection points.

Monitoring the climate

An automated Campbell meteorological station with a CR10X datalogger model was used to gather weather conditions. Its installation in the watershed area was carried out in bare ground conditions and followed the conditions pre-determined by technical standards (EPA, 2000), which recommend a minimum distance from buildings, fences, trees and power lines. Only atmospheric precipitation values (mm h^{-1} or mm d^{-1}) were used in this study despite the potential to record other weather data.

Parameters for monitored water quality and methods of analysis

Water sample collections were carried out monthly in the field using new plastic bottles with 500 mL of total volume, and three samples were always gathered from each point. The samples were kept cold from collection to analysis, as indicated in Standard Methods (APHA, 1995).

The analyzed parameters were: Electrical Conductivity (CE), pH, Ammoniacal Nitrogen (NIA), Ammonia (AMO), Nitrate (NIT), Total Phosphorus (FOT), Suspended Solids (SS) and Turbidity (TU). The laboratory analyses were always performed the day after the collection. Same day analyses were only performed in cases of turbidity and suspended solids; pH and electrical conductivity tests were carried out in the field. For the purpose of assessing the organic contamination of water sources, the dissolved oxygen (DO) was also included as a monitored parameter in the experiment. However, due to technical problems in the equipment, the collected data showed no reliability, determining a discontinuity in its measurement.

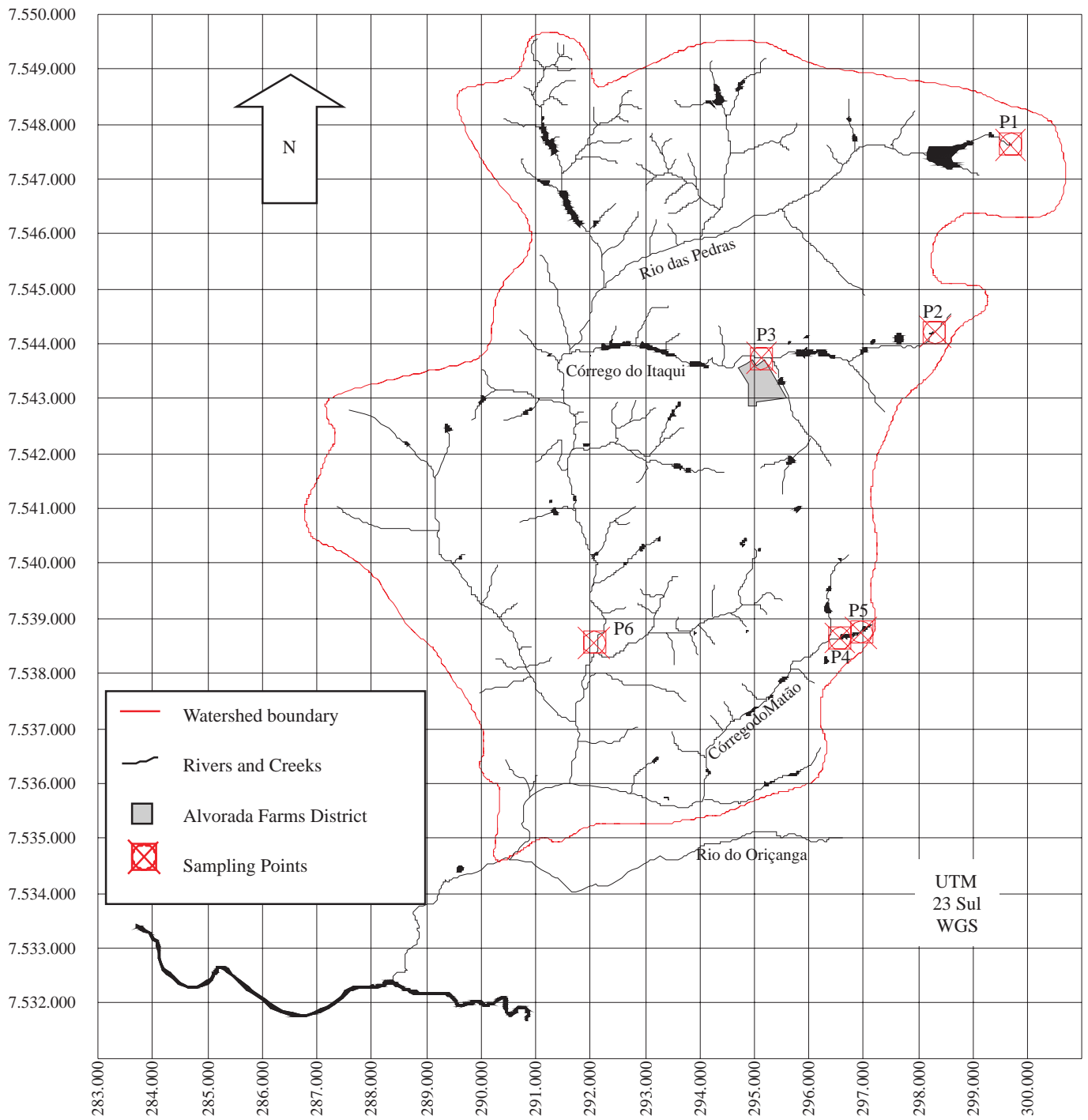


Figure 1. Hydrological map of the Rio das Pedras Watershed and the water sample collection points

It is important to emphasize that the estimation of this parameter is essential in water quality analysis and its inclusion is required in this type of research.

Analyses of suspended solids, turbidity, pH and electrical conductivity parameters were performed according to Standard Methods (APHA, 1995). Colorimetric determinations were used for analyses of nitrate, ammoniacal nitrogen, ammonia and total phosphorus in water using these reagents: nitra-ver 5, nessler reagent, phos-ver 3, potassium persulfate; a digester and a spectrophotometer; the latter being a method accepted by USEPA (United States Environmental Protection

Agency) to monitor receptor bodies and Wastewater Treatment Systems launches.

Water Quality Index (WQI)

The water quality index (WQI) was determined using the statistical technique called Factorial Analysis (MINITAB, 1998). The statistical model that formalizes Factorial Analysis is expressed by Eq. 1.

$$z_{ij} = \sum_{p=1}^m a_{jp} F_{pi} + u_j Y_{ji} \quad (i = 1, 2, K, N; j = 1, 2, K, n) \quad (1)$$

where:

$a_{jp}F_{pi}$ – contribution of the common p factor to the linear combination;

$u_j Y_{ji}$ – residual error in the representation of the observed Z_{ij} measurement.

The model was adjusted using the average of the results from the three simple samples collected from the six points distributed over the watershed, for the eight monitored parameters over the thirteen months of the experiment.

The elaboration of the WQI using this technique required three phases: a) preparation of the correlation matrix; b) extraction of the common factors and the possible reduction of space, and c) the rotation of axes related to the common factors, aiming for a simple and easily interpreted solution.

First, a basic matrix was built without any gaps, a requirement demanded by the factorial technique being employed. Using the matrix from the original data, a correlation matrix was obtained using the Spearman Correlation Coefficient, which revealed the linear dependency among those variables being studied (Steel & Torrie; 1980; Bhattacharyya & Johnson, 1977).

Common factors were then found using the procedure called Factor Analysis. Their numbers were determined based on the percent of total variance in the variables. This is explained by the set of factors associated with their representativeness to the actual study situation.

With the intent of obtaining a simpler factorial load matrix, the axes were rotated using the Varimax Procedure. Since the obtained results were not efficient, the axes without rotation were used.

RESULTS AND DISCUSSION

Table 1 shows the basic descriptive statistics for the variables being studied at the six sampling points during the 13 months of collection, between September 2003 and September 2004.

Table 1. Basic descriptive statistics of the eight variables being studied at the six sampling points from September/2003 to September/2004

Variable	Mean	Median	Standard deviation	Coefficient of variation (%)
Ammoniacal Nitrogen (NIA)	0.07	0.05	0.09	128.57
Ammonia (AMO)	0.09	0.06	0.10	111.11
Nitrate Nitrogen (NIT)	0.13	0.10	0.11	84.61
Total Phosphorus (FOT)	0.41	0.39	0.20	48.78
Electrical Conductivity (CE)	0.02	0.01	0.01	50.00
pH	6.62	6.19	1.70	25.68
Suspended Solids (SS)	13.63	8.33	12.88	94.50
Turbidity (TU)	7.22	6.66	5.85	81.02

In the Table 1, it is possible to observe the high values for the coefficient of variation shown by the variables, demonstrating the variability during the sampling period. This can be explained by the anthropic interventions that occur in the area along the watershed and by the variation in precipitation and flow during the study period. Table 1 also

shows the proximity between the mean and the median values for the parameters FOT, pH e TU.

The Spearman correlation matrix for the monitored variables is found in Table 2. The observation of Table 2 shows the high direct correlation for the following pairs of variables: (NIA-AMO), (NIA-NIT) and (NIT-AMO), at a significant level of 5%. This reveals dependence among variables caused by natural processes. This interaction among variables can be explained since they are all derived from nitrogenated compounds. According to Medeiros et al. (2002) and Piedras et al. (2006), nitrogen undergoes several changes in form and states of oxidation in the aquatic world. Those of greatest interest and in descending order of state of oxidation are: nitrate (NO_3^-), nitrite (NO_2^-), ammoniacal nitrogen ($\text{NH}_4\text{-N}$) or ammonia (NH_3 or NH_4^+) and organic nitrogen (dissolved or in suspension).

Table 2. Correlation matrix for the eight variables studied for water quality

Variable	NIA	AMO	NIT	FOT	CE	pH	SS
AMO	0.000*						
NIT	0.000*	0.000*					
FOT	0.634	0.587	0.721				
CE	0.001*	0.000*	0.007*	0.970			
pH	0.002*	0.000*	0.000*	0.021*	0.008*		
SS	0.069	0.048*	0.024*	0.666	0.499	0.127	
TU	0.000*	0.000*	0.003*	0.395	0.000*	0.055	0.365

* reject the null hypothesis, at significant level of 5%, that the variables are independent

It is possible to infer that the addition of ammoniacal nitrogen determined the increase in ammonia and nitrate concentrations. With the increases in nitrate, the quantity of organic matter increases, which leads to the consequent increase in turbidity and suspended solids, as explained by Andrade et al. (2007) in the Rio Acaraú basin study, when the turbidity and suspended solids variables had high correlation, as well as the pairs of variables “color and turbidity” and “color and ammonia”. This author explained that the correlations between color, turbidity and suspended solids can be understood by the fact that the water color is defined by the reflection and refraction of light on dissolved or suspended materials, linking the water color intensification to the contribution of municipal and industrial sewage and agricultural activities.

However, the SS variable has the opposite effect, according to Table 2. Since the nitrate has a saline nature, its increase also influences the increased electrical conductivity of the water.

On the other hand, increased FOT implied a growth in the quantity of algae, which caused an increase in TU and SS. This fact characterizes a process of water eutrophication, explained Mansor et al. (2006), which quantifies rural diffuse contributions of N and P in the surface water. Thus, the correlation matrix revealed independent relation among the SS and NIA, FOT, CE and pH parameters.

Although a good correlation between the TU and SS variables was expected, the statistical analysis showed that the variables were independent in this study and the option was made to carry out the water quality index without the suspended solids parameter (SS).

Table 3 shows the breakdown of the correlation matrix, which, when conditioned upon the existing interrelations among original data, resulted in a reduction of space for observed variables translated by the common factors. After a few attempts, it was concluded that three factors would be adequate for the case study. After that, the value for each factor was calculated for each sampling station. This value was called a factorial score and it comprises the Water Quality Index (WQI).

Table 3. Matrix for factorial loads, observed variance and the commonality of water quality variables

Variable	Factorial loads			Estimated commonality
	F1	F2	F3	
NIA	0.915	-0.193	0.073	0.880
AMO	0.932	-0.202	0.036	0.910
NIT	0.859	-0.001	-0.081	0.745
FOT	0.195	0.175	0.833	0.763
CE	0.287	-0.673	-0.224	0.586
pH	0.535	-0.035	-0.580	0.624
TU	0.193	-0.776	-0.159	0.665
Variance for each factor	3.06	1.51	1.19	
Variance (%)	38.3	18.9	14.9	
Accumulated Variance (%)	38.3	57.2	72.1	

Using the main components method, three factors associated with self-values that were greater than the unit were selected, which explained 72.1% of the total variance of original variables, indicating a good degree of information conservation.

Since the first factor (F1) explains the greatest variability of data (38.3%), it is adopted as a water quality index (WQI).

Observing the values of estimated commonalities, notice that 91.0% of the AMO parameter is explained by three common factors (F1, F2 and F3), while 58.6% of the CE variance is explained by these factors. Thus, upon analyzing the oneness (a complement of commonality), it is observed that the smallest failure to explain unitary total variance occurs in the AMO variable (9.0%) and the largest failure occurs in the CE variable (41.4%).

Using the Barlett criterion, which is also used by Toledo & Nicolella (2002), the Water Quality Index (WQI) was represented by the following expression (Eq. 2).

$$WQI = \hat{F} = 0.306Z_1 + 0.308Z_2 + 0.292Z_3 + 0.147Z_4 + 0.022Z_5 + 0.139Z_6 - 0.014Z_7 \quad (2)$$

where: Z_i – standardized variables in the following order NIA, AMO, NIT, FOT, CE, pH, TU with $i = 1, \dots, 7$; where: 0.306; 0.308; 0.292; 0.147; 0.022; 0.139; and 0.014 = the weight of each variable in the WQI.

The WQI calculated in this study is in the interval between 0 and 2, and the lower it is the better the quality of the water. According to the Z_i standardized variable coefficients shown in the WQI, AMO was the variable that most influenced the index, followed by NIA and NIT, and the TU variable was the least influent.

According to Sperling (1996) and Vidal et al. (2000), in water sources and rivers the determination of the predominant portion of nitrogen may provide information about the pollution status. The nitrogen compounds in organic form or ammonia refer to the recent pollution, while nitrite and nitrate pollution more remote contamination. Since ammoniacal nitrogen is one of the first steps in the organic matter decomposition, its presence can be related to the precarious construction of wells and the lack of aquifer protection (Alaburda & Nishihara, 1998).

According to Silva & Araújo (2003), elevated levels of nitrates indicate contamination by the inappropriate disposal of human or industrial waste, or the use of nitrogenated fertilizers in agriculture.

Figure 2A shows the variation of the calculated WQI value for the six points monitored over the collection period, and the corresponding values for pluviometric precipitation. It was observed that, in general, the P1 collection point showed the best water quality indices, while P6 had the worst index of all the sampling points. This was expected, since P1 is the source of the Rio das Pedras (main river), corresponding to the highest point of the watershed, and P6 corresponds to the point of greatest contribution, which even receives urban waste.

March 2004 demonstrated the best overall WQI values, whereas May 2004 had the worst. It is believed that the low water quality index during this period is due to the application of fertilizers before transplanting the tomato to the watershed since most producers plant from February to April, applying pre-planting fertilizers from January to March.

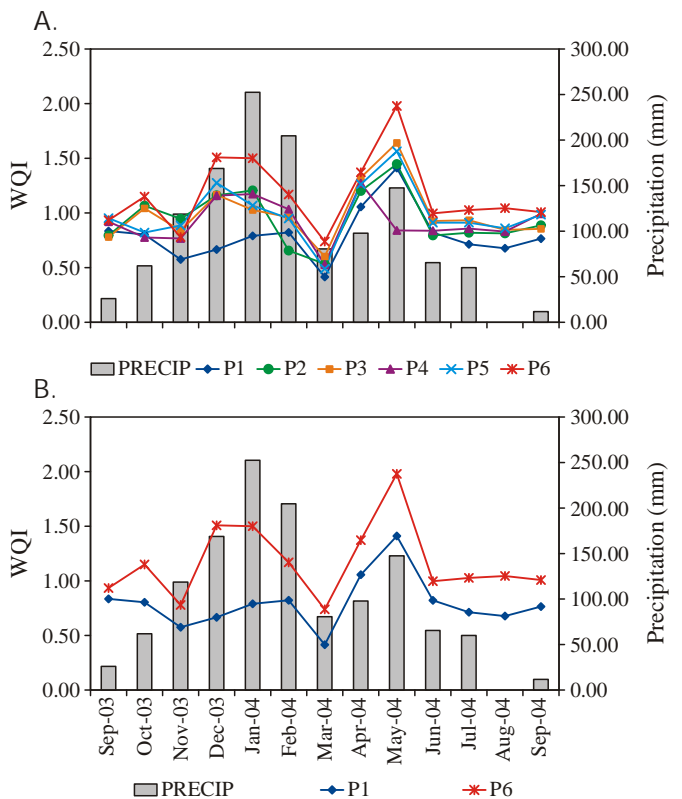


Figure 2. Variation of WQI and of precipitation for Rio das Pedras Watershed: (A) For all sampling points (B) For points P1 and P6

Thus, a worsening in water quality can be observed, especially at P6, which receives the contribution of almost the entire watershed.

Precipitation (the sum of daily rains that occurred each month) influenced the water quality index since its behavior caused alterations in the data in some of the monitored months, as occurred in the researches of Carvalho et al. (2000) and Zimmermann et al. (2008). It is important to take into consideration that the relationship between the days when there was greatest precipitation and the days when field sample collections were taken was not studied.

A comparison of the WQI calculated for the source of the Rio das Pedras (P1) and the point of greatest contribution (P6) is shown in Figure 2B. From Figure 2B it is possible to observe the existence of a similar behavior for the two sampled points despite the distance between them. Furthermore, the occurrence of contaminations along the watershed was confirmed since the water quality at P6 (downstream from the source of the Rio das Pedras) proved to be inferior to P1 over the thirteen months of the collection. However, this contamination is not caused only by agricultural activity, since P6 also receives urban effluents.

CONCLUSIONS

1. The developed water quality index demonstrated the deteriorating conditions of water quality in the Rio das Pedras Watershed due to agricultural activities, although the analysis was not comprehensive due the lack of monitoring other variables.

2. The elaboration of the WQI permitted the characterization of water quality along the watershed, demonstrating its potential as a tool to contribute to the monitoring of water resource availability in the region.

ACKNOWLEDGEMENTS

To FAPESP for the financial support for this project and for the scholarship that was granted (Processes: 01/00566-7).

LITERATURE CITED

- Alaburda J. E.; Nishihara L. Presença de compostos de nitrogênio em águas de poços. *Revista de Saúde Pública*, v.32, n.2, p.160-165, 1998.
- Andrade, E. M. de; Araújo, L. de F. P.; Rosa, M. F.; Disney, W.; Alves, A. B. Seleção dos indicadores da qualidade das águas superficiais pelo emprego da análise multivariada. *Engenharia Agrícola*, v.27, n.3, p.683-690, 2007.
- APHA – American Public Health Association: Standard methods for the examination of water and wastewater. 16.ed. Washington: APHA ; AWWA; APCF, 1995. 1268p.
- Bhattacharyya, G. K.; Johnson, R. A. Statistical concepts and methods. New York: John Wiley & Sons, 1977. 639p.
- Carvalho, A. R.; Schlittler, F. H. M.; Tornisiolo, V. L. Relações da atividade agropecuária com parâmetros físicos químicos da água. *Química Nova*, v.23, n.5, p.618-622, 2000.
- EPA – U.S. Environmental Protection Agency. Meteorological monitoring guidance for regulatory modeling applications, 2000. <http://www.epa.gov/scram001/guidance/met/mmgrma.pdf>. 02 May 2007.
- Haase, J.; Krieger, J. A. H.; Possoli, S. Estudo da viabilidade do uso da técnica de análise fatorial como um instrumento na interpretação da qualidade da água da bacia do Guaíba, RS, Brasil. *Ciência e Cultura*, v.41, n.6, p.576-582, 1989.
- Mansor, M. T. C.; Teixeira Filho, J.; Roston, D. M. Avaliação preliminar das cargas difusas de origem rural, em uma sub-bacia do Rio Jaguari, SP. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.10, n.3, p.715-723, 2006.
- Medeiros, M. A. C. de; Sobrinho, G. D.; Albuquerque, A. F. de; Oliveira, A. C. de; Vendemiatti, J. A. de S. Handout: Química sanitária e laboratório de saneamento II. Limeira: UNICAMP/CESET, 2002. 55p.
- Merten, G. H.; Minella, J. P. Qualidade da água em bacias hidrográficas rurais: um desafio atual para sobrevivência futura. *Agroecologia e Desenvolvimento Rural Sustentável*, v.3, n.4, p.33-38, 2002.
- MINITAB®, Statistical software. Release 12. Minitab Inc. State College: Minitab, 1998.
- Paz, V. P. da S.; Teodoro, R. E. F.; Mendonça, F. C. Recursos hídricos, agricultura irrigada e meio ambiente. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.4, n.3, p.465-473, 2000.
- Piedras, S. R. N.; Oliveira, J. L. R.; Moraes, P. R. R.; Bager, A. Toxicidade aguda da amônia não ionizada e do nitrito em alevinos de *Cichlasoma facetum* (Jenyns, 1842). *Ciência e Agrotecnologia*, v.30, n.5, p.1008-1012, 2006.
- Silva, R. C. A.; Araújo, T. M. Qualidade da água do manancial subterrâneo em áreas urbanas de Feira de Santana (BA). *Ciência & Saúde Coletiva*, v.8, n.4, p.1019-1028, 2003.
- Sperling, M. von. Introdução à qualidade de águas e ao tratamento de esgotos. 2 ed. Belo Horizonte: DESA/UFMG, 1996. 243p.
- Steel, R. G. D.; Torrie, J. H. Principles and procedures of statistics: A Biometrical Approach. New York: McGraw-Hill. 1980. 633p.
- Toledo, L. G. de; Nicoletta, G. Índice de qualidade de água em microbacia sob uso agrícola e urbano. *Scientia Agrícola*, v.59, n.1, p.181-186, 2002.
- Vidal, M.; López, A.; Santoalla, M. C.; Valles, V. Factor analysis for the study of water resources contamination due to the use of livestock slurries as fertilizer. *Agricultural Water Management*, v.45, n.1, p.1-15, 2000.
- Zimmermann, C. M.; Guimarães, O. M.; Peralta-Zamora, P. G. Avaliação da qualidade do corpo hídrico do rio Tibagi na região de Ponta Grossa utilizando análise de componentes principais (PCA). *Química Nova*, v.31, n.7, p.1727-1732, 2008.