EVALUATION OF THE WATER QUALITY INDEX IN BILLINGS DAM THROUGH THE USE OF GEOPROCESSING TECHNIQUES

Giselle Schimidt Ragazzo[1], giselle.ragazzo@gmail.com
Joyce da Silva Bevilacqua[1,2], joyce@ime.usp.br
[1] Instituto de Matemática e Estatística da Universidade de São Paulo. Rua do Matão, 1010, 05508-900, São Paulo, SP, Brasil.
Centro Universitário Senac. Av. Eng. Eusébio Stevaux, 823, 04696-000, São Paulo, SP, Brasil.

Diana Sarita Hamburger, diana.sarita@ufabc.edu.br
Universidade Federal do ABC – CECF –Rua Santa Adélia, 166, 09210-170, Santo André, SP, Brasil.

Abstract. *The aim of this project is to present a proposal for a methodology to obtain information on Billings dam water quality through satellite images from Landsat 5, using the geoprocessing. The Water Quality Index - WQI is an indicator used by CETESB (Company of Environmental Sanitation Technology) since 1975 for water quality classification, composed of physical, chemical and biological parameters, which are: dissolved oxygen, biochemical oxygen demand, pH, turbidity, fecal coliform, total phosphorus , total nitrogen, total solids and temperature. The methodology is based on these nine limnological data, collected by CETESB bimonthly, at three collection points in Billings Reservoir for nearly 10 years. The data collected with the occurrence of rain within 24 hours preceding the time of collection were excluded. The Linear and Cubic Spline Interpolation was used to complete these historical series. Through the geographic coordinates of the points of collection and georeferencing of satellite images from Landsat 5, obtained in the same period of historica limnological data collection by INPE (National Institute for Space Research, Brazil), their locations were determined in the images. Temporal series were constructed with the values of digital images at the point of collection studied and a neighborhood of the point, for bands 1, 2 and 3 which have higher reflectance of the water. The identification of correlations between the limnological variables that compose the WQI and the digital values were made by the construction of linear trend for the digital values and the corresponding limnological data obtained by interpolation. For various parameters were correlated*
above 0.85 in the visible spectral range, despite the relatively low number of images and lack of synchronization between the dates of acquisition of data from CETESB and satellite images, demonstrating the feasibility of the technique.

Keywords*: geoprocessing, WQI, Billings, approximation methods, correlatio*n

1. INTRODUCTION

The aim of this work is to study the possibility to predict water quality of Billings reservoir from LANDSAT 5 orbital remote sensing images (orbit point 219/76) from 1995 to 2004 years. Comparing data from different sources is a main challenge and this case deals with digital values from remote sensing images and field data water quality index enhanced by different data periodicity and lack of information. During this ten years period, the water quality index was obtained bimonthly and images at each 16 days. Therefore, 60 data sets limnology parameters and 240 images were expected. However, different events as cloud cover or error in location of water collect, reduce significantly the available data set. The methodology applied to each type of data is schematized below.

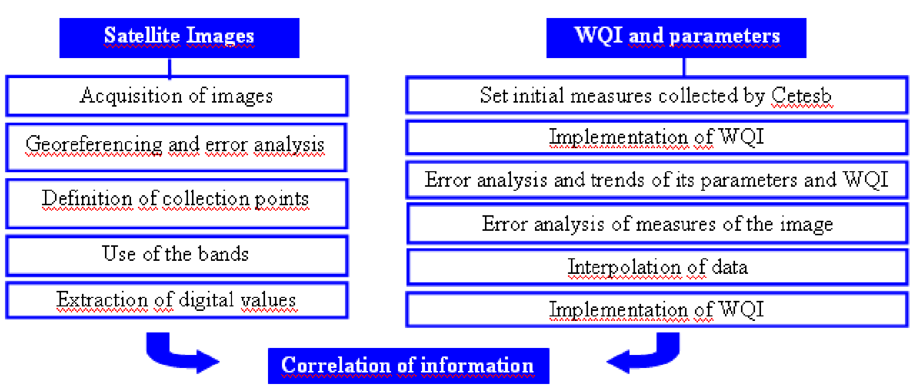

Figure 1. Methodology diagram

Comparison between limnology data and digital values showed in many cases a correlation above 0.9, suggesting the adequacy of the method. Another important benefit of this technique would be the possibility to spread the point collected analysis to the whole reservoir.

This paper is organized as follows: sections 2 and 3 describe, respectively, limnology and remote sensing data; approximations techniques applied to the data can be seen in section 4; results are discussed in section 5; conclusions in section 6.

2. WATER QUALITY INDEX

The Water Quality Index – *WQI* defines a measurement for public water provision considering the parameters: dissolved oxygen (DO), biochemical oxygen demand (BOD), pH, turbidity (TB), fecal coliform (FC), total phosphorus (TP), total nitrogen (TN) and total solids (TS). The weighted index is calculated by Eq. (1).

$$WQI : \Re^8 \rightarrow \Re, \quad WQI(x_1, x_2, \ldots x_8) = \prod_{i=1}^{8} q_i^{w_i}, \tag{1}$$

being vector $\boldsymbol{x}$ the concentration in the water sample, $\boldsymbol{q}$ the quality percentual and $\boldsymbol{w}$ the weight associated with variable $i=1,\ldots,8$. Each component in $\boldsymbol{x}$ was normalized to a [0, 100] percentual scale of quality, considering 0 as the worst condition. Each parameter $\boldsymbol{w}$ is defined by its importance to the index. Table 1 presents the $\boldsymbol{w}$ values defined by CETESB *(*Company of Environmental Sanitation Technology) study. The water temperature is a *WQI* variable that CETESB assumes as constant and equal to maximum quality, 92.5%.

Table 1. Weight distribution used in *WQI*

variable	weight	value
DO	w_1	0.17
BOD	w_2	0.1
pH	w_3	0.12
TB	w_4	0.08
FC	w_5	0.15
TP	w_6	0.1
TN	w_7	0.1
TS	w_8	0.08

This index was adapted by CETESB, based on a 1970's study of National Sanitation Foundation (United States) that includes the most relevant variables for public water provision. CETESB classifies the *WQI* in 5 categories according to Tab. 2, according Contract DAEE/CETESB, Term 49/79, 2004.

Table 2. CETESB Classification of Water Quality Index

Classification of *WQI*	Value of *WQI*
Very good	(79, 100]
Good	(51, 79]
Regular	(36, 51]
Bad	(19, 36]
Very bad	[0, 19]

A normalizing function for each variable is defined by $\boldsymbol{F(x) = q}$, composed by many least square approximation, according to variable characteristics which differs for different sub-intervals of the domain of $\boldsymbol{x}$. Figure 2 shows the CETESB model for the pH variable and Tab. 3 the approximation curves for each sub-interval. Similar models are developed for the other variables.

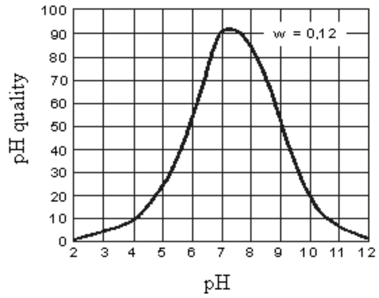

Figure 2. Curve of the pH quality

Table 3. Approximations by the method of least squares for pH

Interval	Approximate function	Parameter *a*	Parameter *b*	Parameter *c*
[2, 2]	constant and equal to 2	-	-	-
(2, 4]	$a + bx + cx^2$	13.6	-10.64	2.4364
(4, 6.2]	$a + bx + cx^2$	155.5	-77.36	10.2481
(6.2, 7]	$a + bx + cx^2$	-657.2	197.38	-12.9167
(7, 8]	$a + bx + cx^2$	-427.8	142.05	-9.695
(8, 8.5]	$a + bx$	216	-16	-
(8.5, 9]	ae^{bx}	1415823	-1.1507	-
(9, 10]	$a + bx$	288	-27	-
(10, 12]	$a + bx + cx^2$	633	-106.5	4.5
more than 12	constant and equal to 3	-	-	-

Many CETESB data sets were excluded because of imprecise spatial location, mistaken surveys and inconsistent values. The data obtained after rainy days were also excluded, remaining a 39 limnology data sets. Figure 3 shows point data collection location and Tab. 4 the coordinates of those points, according to Alonso et al. (1995) and Whately (2003).

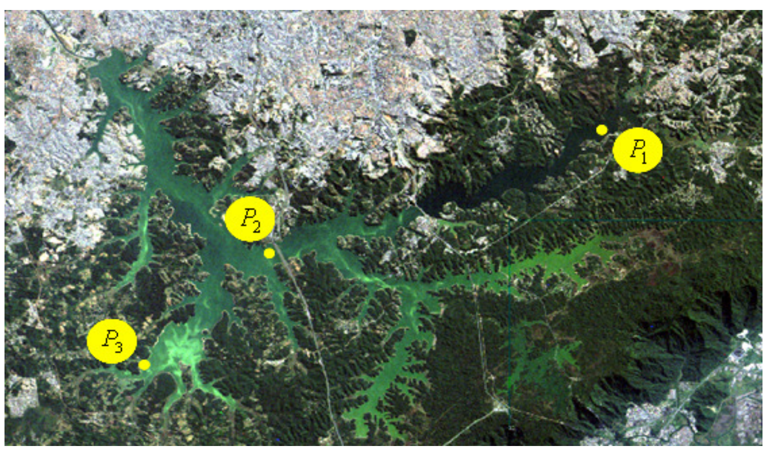

Figure 3. Points data collection location in Billings dam

Table 4. Location and period of *WQI* data sets

Point	Location	Measurement period
*P*1	23°44’2"*S* 46°26’50"*W*	01/1995 to 09/2004
*P*2	23°46’54"*S* 46°34’48"*W*	03/1995 to 09/2004
*P*3	23°50’41"*S* 46°39’20"*W*	03/1999 to 09/2004

The consistency of CETESB data was analyzed considering the error propagation in *WQI* if an error occurs in original values of each variable. In the sensitivity study variation in the errors magnitude from 1 to 50 percent were applied to each variable considering constant all the other variables. The result of this analysis is in Tab. 5, being the most sensible variables are dissolved oxygen (DO) and pH. Even in those cases that the variation is above 10% the impact in *WQI* does not sufficient to alter the water category.

Table 5. The sensitivity analysis of the *WQI* and its parameters

Parameter	Variations	1%	5%	10%	50%
DO	rate of change	4.55	4.73	4.97	4.7
	percentage change	0.31%	1.61%	3.39%	16.03%
FC	rate of change	0	0	0	0
	percentage change	0.05%	0.24%	0.46%	2%
BOD	rate of change	-1.15	-0.97	-0.94	-0.86
	percentage change	0.07%	0.3%	0.59%	2.69%
TN	rate of change	-0.66	-0.66	-0.66	-0.66
	percentage change	0.01%	0.04%	0.08%	0.41%
TF	rate of change	-7.39	-7.39	-7.39	-7.39
	percentage change	0%	0.01%	0.03%	0.14%
TB	rate of change	-6.54	-1.31	-0.65	-0.13
	percentage change	9.24%	9.24%	9.24%	9.24%
TS	rate of change	0	0	0	0
	percentage change	0.02%	0.02%	0.02%	0.02%
pH = 9	rate of change	-5.5	-6.12	-7.34	-4.58
	percentage change	0.67%	3.71%	8.9%	27.75%
pH = 6	rate of change	7.03	7.21	7.45	7.35
	percentage change	0.55%	2.81%	5.81%	28.67%

3. REMOTE SENSING

LANDSAT 5 remote sensing images are collected each 16 days generating a 7 band image and with a 30 meters spatial resolution (Maguire. *et al.* (1991); Novo., 2008; Engesat, 2008). In the period from 1995 to 2004 data set should be composed by 240 images (acquired at INPE, National Institute for Space Research, Brazil, 2008). Limitations in the collection, related to days after rainfalls, cloud cover, lack of data and inadequate signal, compromised the data. After excluding those images only 21 were left.

The pure water spectral reflectance curve has a high pick at the blue wavelengths, diminishing and lowering until the infrared wavelength (Girard and Girard, 2003). Mineral particles, planktons on the water can change those spectral characteristics.

Figure 4 shows that the most appropriate bands of LANDSAT 5 for this study were: the visible blue (0.45 – 0.52μm), green (0.52 – 0,6 μm) and red (0.63 – 0.69 μm), respectively, identified in this work by band 1, band 2 and band 3. All 21 images were resampled to the UTM Projection (Zone 23, South Hemisphere, Corrego Alegre Datum). The georreferencing had a spatial error of less than 0.3 pixel, less than the spatial imprecision of the *WQI* CETESB point data.

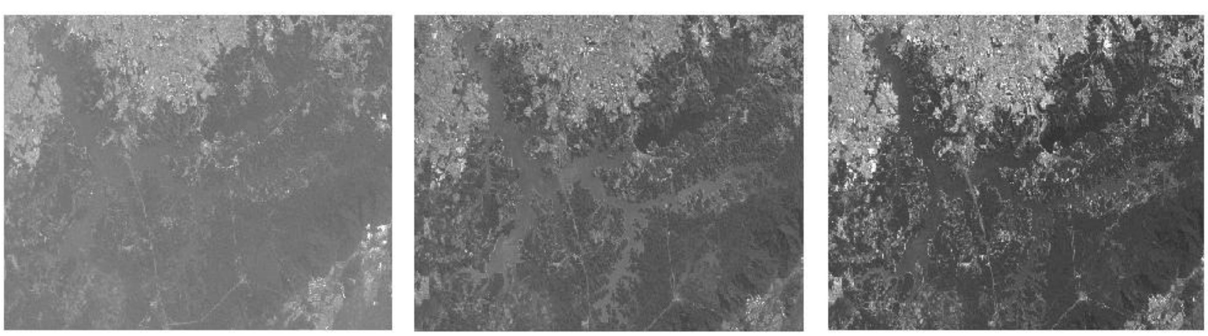

Figure 4. From left to right band 1, band 2 and band 3.

The location of the 3 *WQI* points in the images was done. For each point P_i, and in each image band B_i , i=1,2,3 (blue, green and red), digital values DV were defined by Eq. (2) as the mean of the first order neighborhood including the diagonal pixels.

$$DV = \frac{\sum_{i,j=1}^{3} x_{ij}}{9} \qquad (2)$$

Figure 5 shows the DV relative to P_1. The consecutive data can be identified by straight lines. It can be seen the data deficit to execute a temporal analysis.

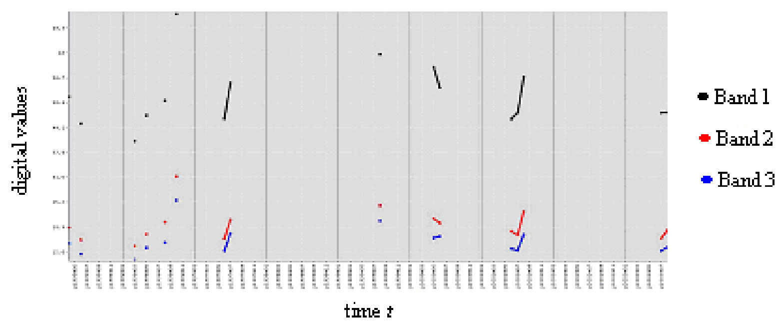

Figure 5. Digital Values of images to the point of collection P_1

4. DATA TREATMENT

The data from both sets (digital values and water samples) available to use in the correlation analysis can be together seen in Fig. 6, were the red and black lines indicate, respectively, turbidity quality and remote sensing data for the P_1 point. Similar figures were constructed for each variable and each point, to visualize the total data and, consequently, the lack of information for each case.

To treat those data means to interpolate water variables in order to make possible a temporal analysis. The best period for comparison between these two sets is identified as 12 months, beginning at January, as is shown by the vertical lines in Fig. 6.

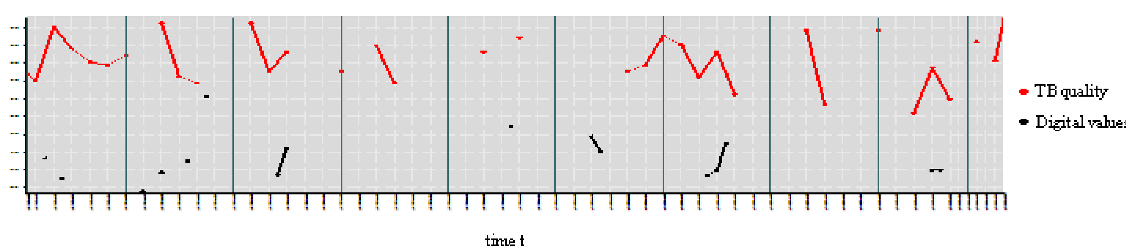

Figure 6. Digital values and TB quality for P_I

Three different approximations were applied to water variables sets: linear interpolation of two consecutive points; cubic spline interpolation to a year set points; linear regression applied to one year data set. Spline interpolations provides non expected negative values, for variables which have values near zero or present a significant variation of standard deviation for some time intervals. This phenomenon is observed in, pH, fecal coliform, total phosphorus and total nitrogen. This inconsistency was treated replacing those negative values by zero, meaning null concentration of these variables in the water (Fig. 7). The consistency of this procedure is based on the prior sensitivity analysis. The linear interpolation inserts trends in data sets especially when the lack of consecutive data is large. Linear regression applied to a period of one year data identify the main tendency of this year.

Variation rate of digital values were calculated by two different ways: angular coefficient of the linear regression line for one year data (annual adjustment); linear interpolation of consecutive data (consecutive adjustments). Figure 8 shows the adjusted straight method of least squares through digital values available in each year and Fig.9 straight lines through the DV adjusted in consecutive dates.

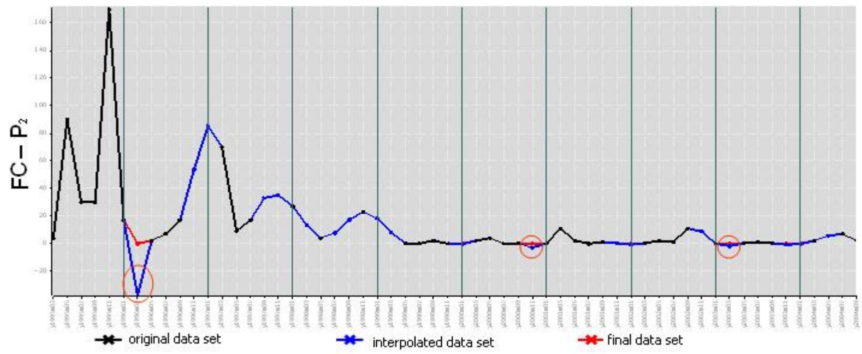

Figure 7: Original and final data set for one limnology variable (FC).

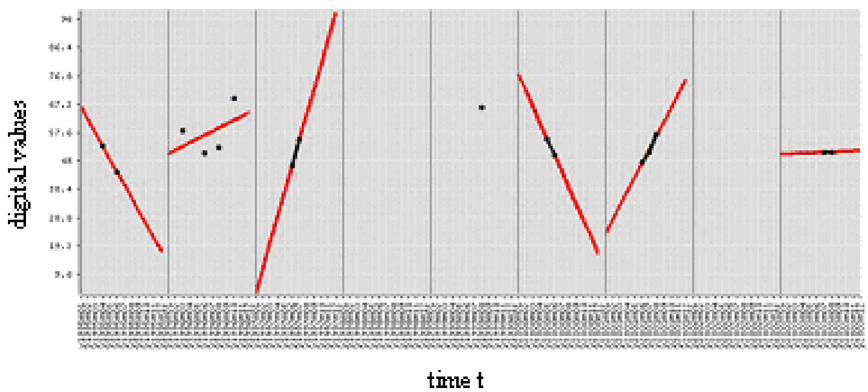

Figure 8. Annual adjustments for DV

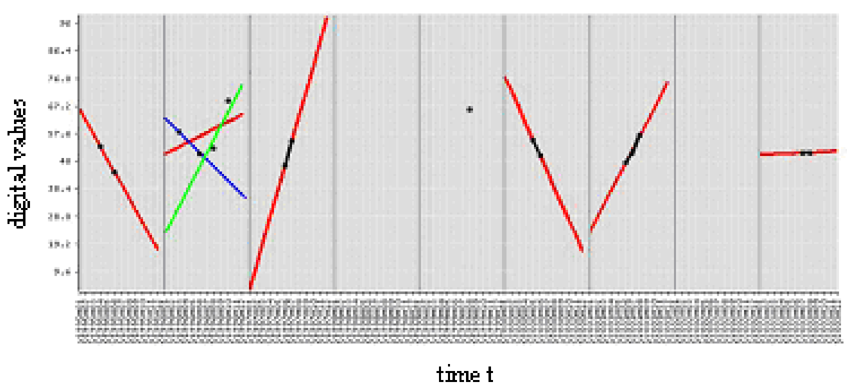

Figure 9. Consecutive adjustments for DV

5. RESULTS

Even with a sparse information, non simultaneous collect date, this technique shows some encouraging results. Classical correlation analysis Eq. (3) between variation rates of the two sets of data is done by each point and each band. The correlation considers also the techniques applied to the original data (Raw). Some variables from limnology data showed gross efficiency to obtain correlations, as can be seen in Tab. 6, for FC, TF and the index itself.

Comparing the results in three bands, band 1 is slightly better them the other two. However, it is not possible to guarantee that this band is always better than the others because, depending on the point location and local characteristics of the water, other bands contribution can be identified as shown in Tab. 8.

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}_{obs}\,\overline{y}_{obs}}{\sqrt{\left[\sum_{j=1}^{n} x_j^2 - n\overline{x}_{obs}^2\right]\left[\sum_{j=1}^{n} y_j^2 - n\overline{y}_{obs}^2\right]}} \tag{3}$$

The interpolation was the worst efficient technique because of a large lack of consecutive data, making trends in analysis. None correlation is found for point P_2 (Tab. 7), because the samples are collected under a viaduct, consequently, without light presence changing water characteristics.

Table 6. Correlations obtained by annual adjustments for P_1

P_1	Raw data			Linear interpolation			Natural cubic spline		
Limnological data	Band 1	Band 2	Band 3	Band 1	Band 2	Band 3	Band 1	Band 2	Band 3
FC	-0.889	-0.914	-0.872	-0.313	-0.514	-0.544	-0.474	-0.609	-0.672
pH	0.510	0.428	0.416	0.094	-0.069	0.211	0.156	0.003	0.28
TB	-0.312	-0.196	-0.104	0.010	-0.240	-0.085	0.207	-0.007	0.155
BOD	-0.233	-0.263	-0.099	-0.136	-0.364	-0.191	0.017	-0.182	0.026
DO	-0.684	-0.536	-0.542	-0.442	-0.315	-0.230	-0.599	-0.517	-0.511
TN	0.158	0.319	0.466	0.285	0.528	0.565	0.210	0.456	0.450
TF	0.859	0.892	0.676	0.578	0.631	0.780	0.442	0.504	0.681
TS	-0.514	-0.472	-0.187	-0.164	-0.043	0.224	-0.592	-0.478	-0.22
WQI	-0.885	-0.825	-0.806	-0.256	-0.265	0.032	-0.564	-0.575	-0.374

Table 7. Correlations obtained by consecutive adjustments for P_2

P_2	Raw data			Linear interpolation			Natural cubic spline		
Limnological data	Band 1	Band 2	Band 3	Band 1	Band 2	Band 3	Band 1	Band 2	Band 3
FC	-0.431	-0.515	-0.257	-0.374	-0.480	-0.279	-0.224	-0.328	-0.094
pH	0.003	0.092	-0.059	0.070	0.150	-0.039	0.005	0.082	-0.079
TB	0.229	0.278	0.133	-0.171	-0.127	-0.193	-0.099	-0.046	-0.201
BOD	-0.065	-0.079	-0.130	-0.317	-0.326	-0.372	-0.367	-0.406	-0.407
DO	0.299	0.287	0.283	0.083	-0.017	0.227	0.282	0.272	0.237
TN	-0.495	-0.584	-0.675	-0.019	-0.121	0.012	-0.318	-0.434	-0.232
TF	-0.280	-0.262	-0.175	-0.192	-0.189	-0.237	-0.239	-0.219	-0.108
TS	0.105	0.188	0.049	-0.202	-0.128	-0.246	-0.133	-0.055	-0.194
WQI	-0.392	-0.426	-0.608	-0.252	-0.310	-0.175	-0.055	-0.034	-0.065

Correlations for the point P_3 showed in Tab. 8 demonstrated that all variables can be used to identify the correlation in all bands indicating the feasibility of this technique. The annual adjustment is slightly better than the consecutive because the former contains one year information instead of the latter, which evaluation is done considering a sub set of the annual data.

Table 8. Correlations obtained for P_3

P_3	Annual adjustments			Consecutive adjustments		
Limnological data	Band 1	Band 2	Band 3	Band 1	Band 2	Band 3
pH	-0.467	-0.961	-0.564	-0.510	-0.974	-0.605
TB	-0.991	-0.592	-0.970	0.596	-0.165	0.501
BOD	0.844	0.199	0.778	0.847	0.205	0.781
DO	-0.303	-0.897	-0.410	-0.978	-0.829	-0.995
TN	-0.485	-0.966	-0.581	0.255	-0.520	0.144
TF	0.426	-0.356	0.321	0.541	-0.230	0.442
TS	-0.747	-0.997	-0.818	-0.231	-0.861	-0.340
WQI	-0.630	-0.996	-0.714	-0.531	-0.979	-0.624

6. CONCLUDING REMARKS

In this work we analyze the possibility of evaluating the quality of water using remote sensing images. Even without synchronism between water variables collect and satellite images acquisition the technique demonstrated to be feasible, considering the high correlations observed in some cases.

It indicates the possibility to collect relative information of Billings reservoir water quality through digital images, even though it is not possible to establish directly an *WQI* value from an image digital value. The proposed technique can be applied without an enormous data set and with low computational effort.

This study suggests that if limnology data should be collected in a more controlled way the results must be improved. The collect points must be chosen in order to guarantee light presence in the sample, maintenance of the geographic coordinates of each point during a long time period and synchronism between acquisition images and water sample collect.

7. ACKNOWLEDGEMENTS

The authors are in debt with Prof. Alexandre Saron for the Billings data collection, Prof. Benjamin Capellari for the enlightened discussions about geoprocessing questions and Carmen Lucia Midaglia for CETESB reports.

8. REFERENCES

Alonso.C, Perez.L, Toledo.A, 1995. "Reavaliação dos pontos de amostragem da rede básica de monitoramento da qualidade das águas interiores do Estado de São Paulo – UGRHI 06 – Alto do Tietê", CETESB.

Contrato DAEE/CETESB, Termo 49/79, 2004, "Modelo Matemático para Cálculo do Índice de Qualidade de Água (IQA)".

CETESB, "Relatórios anuais de qualidade das águas", http://www.cetesb.sp.gov.br.

ENGESAT, Satellite Informations, http://www.engesat.com.br.

Girard.M., Girard.C., 2003, "Processing of Remote Sensing Data", Editora Taylor & Francis, 1. ed.

INPE, Satellite Informations, http://www.inpe.br .

Maguire.D, Goodchild.M, Rhind.D, 1991. "GIS and Remote Sensing in Geographical Information Systems Longman & Technical"

Novo.E, 2008. "Sensoriamento Remoto: Princípios e Aplicações", Blucher, São Paulo.

Whately.M, 2003. "Avaliação e identificação de áreas e ações prioritárias para a conservação, recuperação e uso sustentável da Bacia Hidrográfica da Billings", São Paulo.

9. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.