A Water Quality Index Model using Stepwise Regression and Neural Networks Models for the Piabanha River Basin in Rio de Janeiro, Brazil.

H23A-1229

Mariana Villas Boas, UFRJ/COPPE – marianadvb@globo.com
Francisco Olivera, Texas A&M University
José Paulo S. de Azevedo, UFRJ/COPPE

AGU FALL MEETING
San Francisco | 9-13 December 2013

INTRODUCTION

The evaluation of water quality through "indexes" is widely used in environmental sciences. There are a number of methods available for calculating water quality indexes (WQI), usually based on site-specific parameters.

In Brazil, WQI were initially used in the 1970s and were adapted from the methodology developed in association with the National Science Foundation (Brown et al, 1970). Specifically, the WQI "IQA/SCQA", developed by the Institute of Water Management of Minas Gerais (IGAM), is estimated based on nine parameters: Temperature Range, Biochemical Oxygen Demand, Fecal Coliforms, Nitrate, Phosphate, Turbidity, Dissolved Oxygen, pH and Electrical Conductivity.

The goal of this study was to develop a model for calculating the IQA/SCQA, for the Piabanha River basin in the State of Rio de Janeiro (Brazil), using only the parameters measurable by a Multiparameter Water Quality Sonde (MWQS) available in the study area. These parameters are: Dissolved Oxygen, pH and Electrical Conductivity. The water quality data used in the study were obtained by the Geological Survey of Brazil in partnership with other public institutions (i.e. universities and environmental institutes) as part of the project "Integrated Studies in Experimental and Representative Watersheds".

The use of this model will allow to further the water quality monitoring network in the basin, without requiring significant increases of resources, because the water quality measurement with MWQS is less expensive than the laboratory analysis required for the other parameters.

STUDY AREA AND DATA

The Piabanha watershed is located in southeastern Brazil and it covers an area of approximately 2050 km² mostly inserted in the mountainous region of the state of Rio de Janeiro.

Geological Survey of Brazil (CPRM), in partnership with other institutions (i.e. universities and environmental institutes) , develops the project entitled "Integrated Studies in Experimental and Representative Watersheds - EIBEX" in the watershed. For this purpose, it was defined a representative basin, inserted in Piabanha watershed, with 400km². The representative basin has an equivalent social, economic, physical, and environmental reality of the Piabanha watershed. And, also, the characteristics of land use and vegetation cover are the same which allow the extrapolation of the results for the entire watershed. Then, three experimental watersheds were defined inside of the representative basin related to the major uses and soil occupation; such as: urban occupation at 47 km², agricultural use at 30 km², and preserved forest at 13km² where the studies are developed (MCT/FINEP, 2010).

A water quality monitoring network was implemented in the representative basin to understand the relationships and processes involved in the area. The network consists of nine monitoring locations (gauges) with a sampling frequency of five times per year since 2009. The dataset used in this study has three years of data. Figure 1 illustrates the Piabanha representative and experimental watersheds as well as the water quality monitoring network. Some parameters are measured in loco using a multiparametric probe (i.e. dissolved oxygen, electrical conductivity, temperature and pH) and others are analyzed in laboratory. (Villas-Boas *et al*, 2011)

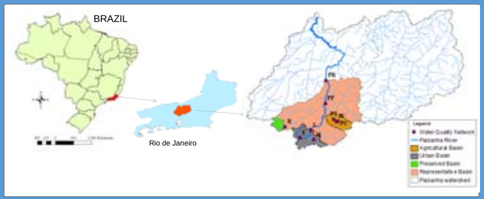

Figure 1 – Localization of Piabanha representative and experimental watersheds and the water quality monitoring network.

MATERIAL AND METHODS

IQA/SCQA

National Sanitation Foundation (NSF) in the United States has initiated the research on the water quality index in 1970. The initial idea was to develop an index (WQI/NSF) to represent the water quality based on the main water quality parameters. For this purpose, 142 water quality specialists were surveyed to select the main water quality parameters and their respective weights, considering the improvement and protection of water for consumption. Therefore, nine water quality parameters were selected to compose the index as presented in Table 1 (Brown et al, 1970).

The System of Calculation of Water Quality - IQA / SCQA, published in June 2005, was developed under the National Environment Program - PNMA as a partnership between the Brazilian Environment Ministry and the government of state of Minas Gerais (SEMAD (2005)) .The IQA/SCQA uses the same parameters established by NSF representing the physic- chemical and biological characteristics of water adapting the weights for Brazilian reality and always considers the temperature change equal zero. The Table 1 presents the parameters and weights used to perform IQA/SCQA compared to WQI/NSF.

Table 1 – Parameters and weights used to perform WQI/NSF and IQA/SCQA.

Parameters	Units	WQI/NSF Weights	IQA/SCQA Weights
Dissolved Oxygen (DO)	%	0.17	0.17
Fecal Coliforms (FC)	NMP/100ml	0.16	0.15
pH	-	0.11	0.12
Biochemical Oxygen Demand (BOD)	mg/l	0.11	0.10
Nitrate (N)	mg/l	0.10	0.10
Total Phosphate (P)	mg/l	0.10	0.10
Temperature change	°C	0.10	0.10
Tubidity (T)	UNT	0.08	0.08
Total Solids (TDS)	mg/l	0.07	0.08

Stepwise linear regression

Stepwise regression is a systematic method for adding and removing terms from a multilinear model based on their statistical significance in a regression. At each step, the p value of an F-statistic is computed to test models with and without a potential term. If a term is not currently in the model, the null hypothesis is that the term would have a zero coefficient if added to the model. If there is sufficient evidence to reject the null hypothesis, the term is added to the model. Conversely, if a term is currently in the model, the null hypothesis is that the term has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the term is removed from the model.

Neural Networks

The Neural Network(NN) has the ability to fit arbitrary functions representing the nature of the contributing variables, mainly when the relationship between them is not known and susceptible to non-linearity, likewise water quality data.

The NN architecture used in this study comprised a Feedforward Neural Network with tan-sigmoid function in the hidden layer and a linear function in the output layer. Some different combinations were tested to find the number of neurons in each layer and the best one was using 4 neurons in the hidden layer and 1 in the output layer. The neural network training has been executed based on a backpropagation algorithm which includes performing computations backward through the network and was repeated several times to confirm that the global minimum had been found. Figure 2 shows the NN architecture.

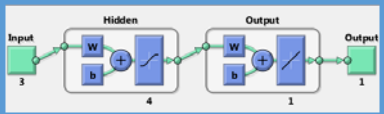

Figure 2 - NN architecture used in the application.

RESULTS AND DISCUSSION

The both models, stepwise regression and neural networks, were performed using the three MWQS water quality parameters (i.e. dissolved oxygen, electrical conductivity and pH) as input and the values of IQA/SCQA as output. The Figure 3 shows the correlation matrix between the 9 parameters used to calculate IQA/SCQA . It is possible to note that dissolved oxygen (DO) and electrical conductivity (EC) have a strong linear correlation with IQA/SCQA. The model results were evaluated according to the following validation statistics: coefficient of determination (R2), Root Mean Square Error (RMSE), Akaike information criterion (AIC) and Final Prediction Error (FPE). The results are presented in Table 2.

It is possible to observe that the neural network model produced a better fit than the regression model, having a greater R2 and smaller RMSE, AIC and FPE. The best performance of the second method can be attributed to the fact that the water quality parameters often exhibit nonlinear behaviors and neural networks are capable of representing nonlinear relationship efficiently, while the regression is limited to linear relationships.

.

Table 2- The results for model statistcs.

Statistcs	Stepwise Regression	Neural Networks
RMSE	6.19	4.86
R2	0.92	0.94
AIC	0.65	0.33
FPE	1.93	1.39

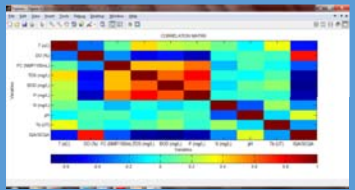

Figure 3 Correlation Matrix for IQA/SCQA water quality parameters.

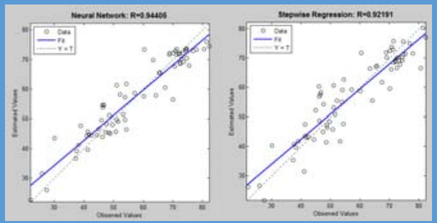

Figure 4 – Correlation between model estimated and observed values.

CONCLUSIONS AND RECOMMENDATIONS

This study has showed the good performance of two models, stepwise regression and neural networks, computing the water quality index, IQA/SCQA, using only three water quality parameters instead of the common nine. The neural network model had the best adjustment perharps because of the nonlinear behavior of the water quality variables. These results are important in order to reduce costs of future monitoring campaings, where it can be used only the MWQS parameters to calculate IQA/SCQA and evaluate the water quality in the watershed. It's recommended to repeat the application using a large dataset to increase the eficacy of the results.

REFERENCES

- Brown, R.M., McLelland, N.I., Deininger, R.A., Tozer, R.G. 1970. A Water Quality Index–Do we dare? Water & Sewage Works, October: 339-343.
- SEMAD; UCEMG / PNMA II (2005). "Sistema de cálculo da qualidade da água (SCQA) – Estabelecimento das equações do índice de qualidade das águas (IQA)". Relatório I.
- MCT/FINEP/CT-HIDRO, "Projeto-EIBEX-I – Estudos Integrados de Bacias Experimentais – Parametrização Hidrológica na Gestão de Recursos Hídricos das Bacias da Região Serrana do Rio de Janeiro" – Rel. Téc. Parcial 2, março de 2010.
- VILLAS-BOAS, M.D., CHAVES, C.F.B., SANTOS, F.J., SILVA, J.G.P., HENRIQUES, M.P., MONTEIRO, A.E.G.C, 2011. Avaliação preliminar da qualidade da água nas bacias experimentais e representativa do rio Piabanha a partir do cálculo do Índice de Qualidade de Água – IQA. In: XIX Simpósio Brasileiro de Recursos Hídricos; Novembro de 2011, Maceió, ABRH, 2011 Anais... CDROM.