

Use of ultraviolet–visible spectrophotometry associated with artificial neural networks as an alternative for determining the water quality index

Edson Marcelino Alves  · Ramon Juliano Rodrigues · Caroline dos Santos Corrêa · Tiago Fidemann · José Celso Rocha · José Leonel Lemos Buzzo · Pedro de Oliva Neto · Eutímio Gustavo Fernández Núñez

Received: 4 November 2017 / Accepted: 24 April 2018 / Published online: 2 May 2018
© Springer International Publishing AG, part of Springer Nature 2018

Abstract The water quality index (WQI) is an important tool for water resource management and planning. However, it has major disadvantages: the generation of chemical waste, is costly, and time-consuming. In order to overcome these drawbacks, we propose to simplify this index determination by replacing traditional analytical methods with ultraviolet-visible (UV–Vis) spectrophotometry associated with artificial neural network (ANN). A total of 100 water samples were collected from two rivers located in Assis, SP, Brazil and calculated the WQI by the conventional method. UV–Vis spectral analyses between 190 and 800 nm were also performed for each sample followed by principal component analysis (PCA) aiming to reduce the number of variables. The scores of the principal components were used as input to calibrate a three-layer feed-forward neural network. Output layer was defined by the WQI values. The modeling efforts showed that the optimal ANN architecture was 19-16-1, trainlm as training function, root–mean–square error (RMSE) 0.5813, determination coefficient between observed and predicted

values (R^2) of 0.9857 ($p < 0.0001$), and mean absolute percentage error (MAPE) of $0.57\% \pm 0.51\%$. The implications of this work's results open up the possibility to use a portable UV–Vis spectrophotometer connected to a computer to predict the WQI in places where there is no required infrastructure to determine the WQI by the conventional method as well as to monitor water body's in real time.

Keywords Water quality index · UV–Vis spectrophotometry · Artificial neural networks · Principal component analysis · Water pollution · Water analysis

Abbreviations

ANN	Artificial neural network
BOD ₅	Biochemical oxygen demand
C _{CM}	Cost of the conventional methodology
C _H	Hardware cost
C _{PM}	Proposed methodology cost
C _{uv}	Unit variable cost
DO	Dissolved oxygen
FCI	Fixed capital investment
HLS	Humic-like substances
MAPE	Mean absolute percentage error
N _{minimum}	Minimum number of samples
NSFWQI	National Sanitation Foundation Water Quality Index
PCA	Principal component analysis
PLS	Partial least square
R ²	Determination coefficient
RMSE	Root–mean–square error

E. M. Alves (✉) · R. J. Rodrigues · C. dos Santos Corrêa · T. Fidemann · J. C. Rocha · J. L. L. Buzzo · P. de Oliva Neto
Departamento de Ciências Biológicas, Universidade Estadual Paulista “Júlio de Mesquita Filho” Campus – Assis, Avenida Dom Antônio, 2100, Assis, SP 19806-900, Brazil
e-mail: edsonmaralves@hotmail.com

E. G. F. Núñez
Centro de Ciências Naturais e Humanas (CCNH), Universidade Federal do ABC, Avenida dos Estados, 5001, Santo Andre, SP 09210-580, Brazil

UV–Vis Ultraviolet–visible
WQI Water quality index

Introduction

Water quality indexes (WQI) are dimensionless single numbers (0–100), resulting from mathematical equations, which are based on physical, chemical, and microbiological parameters. Thus, these indexes provide an indication of overall water quality, allowing timeline comparison of points distributed in the same waterbody or among different water sources (Brown et al. 1972). There are different water quality indexes developed worldwide since different national and international agencies involved in water quality assessment and pollution control define water quality criteria for different water uses, considering different indicator parameters (Bharti and Katyal 2011). In the early 1970s, the National Sanitation Foundation (NSF) supported the creation of a water quality index that became known as NSFQI, which comprises a comprehensive study that has been discussed in various papers (Bharti and Katyal 2011). Brown et al. (1972) developed this index (NSFWQI) by selecting nine parameters rigorously (biochemical oxygen demand BOD₅, dissolved oxygen (DO), fecal coliforms, pH, nitrate, phosphate, temperature, suspended solids, and turbidity), developing a common scale and assigning weights to the parameters. Based on experts' opinions, rating curves were also developed to attribute values for the variation of the water quality level caused by different levels of each of the selected parameters (Poonam et al. 2013). The assessment of the index aims to summarize a great number of information in a form that allows quick recognition and interpretation of trends over time and space. The WQI results are useful in the operational management to easily identify points requiring priority action and helps in the modification of the policies, which are formulated by various environmental monitoring agencies (Tyagi et al. 2013). Besides this, the index facilitates the communication with the laypeople, while maintaining the initial precision of measurement. On the other hand, one of the major disadvantages of NSFQI is that some analyses have a significant cost, are time-consuming, and generate chemical waste requiring subsequent treatment. Moreover, any lack of data in one of the nine parameters limits its calculation.

Due to these disadvantages, new approaches using spectrophotometric techniques have been proposed in order to make the determination of some WQI parameters by indirect estimation (Van den Broeke et al. 2006).

The spectrophotometry in the ultraviolet and visible ranges (UV–Vis) is a fast and simple technique that allows the obtainment of extremely relevant data able to identify and quantify compounds in water (Vanrolleghem and Lee 2003). Several scientific communications support spectral information as an alternative to water analysis, since the characteristics of the water UV–Vis spectrum can be used for qualitative detection of changes in its composition (Rieger et al. 2006; Torres and Bertrand-Krajewski 2008). According to Van den Broeke et al. (2006), the principles used in submersible UV–Vis spectrophotometer probes allow the determination of multiple parameters in the UV–Vis range with just a sample. This approach can efficiently replace conventional analysis methods including chemical oxygen demand, biochemical oxygen demand, total organic carbon, dissolved organic carbon, nitrates, nitrites, turbidity, and suspended solids. Many other studies have been found in the literature dealing with the use of absorbance as surrogate of different contaminants in water (Anumol et al. 2015; Gerrity et al. 2012; Roccaro et al. 2015; Yan et al. 2014). Although being promising, only the use of spectrophotometric techniques is not enough to determine the WQI, since it only replaces some of the parameters that are taken into account in the calculations of this index.

The artificial neural networks (ANNs) are versatile and have been applied to predict water salinity (Maier and Dandy 1996), total dissolved solids, electrical conductivity turbidity (Najah et al. 2009), chemical oxygen demand, biochemical oxygen demand (Rene and Saidutta 2008), and even other parameter such as eutrophication (Kuo et al. 2007), rainfall (Wu et al. 2010), and suspended sediment load estimation (Chen and Chau 2016; Olyaei et al. 2015). Besides the use of spectrophotometric techniques, other approaches, such as artificial intelligence, is being increasingly used to predict and forecast water quality parameters (Alizadeh and Kavianpour 2015; Khuan et al. 2002). The ANNs are computational techniques based on biological neurons, capable of gradual learning over time as well as to recognize extremely complex patterns (Farmaki et al. 2010). The ANNs are composed by several computational units called artificial neurons, which perform a number of simple operations and transmit the results to

neighboring neurons through connections characterized by weights. The presence of weights between connections is critical, since they define the signal propagation along the network. The architecture of an ANN restricts the type of problem on which the network can be used and is defined by the number of layers, the number of neurons in each layer, and the type of connection between neurons (Ferneda 2006).

Knowing the water quality of a region gives a measure of environmental equilibrium and is a useful tool for the rational and sustainable management of available resources. Nonetheless, the major challenge in this context is to guarantee that every municipality is able to monitor and supervise the quality of their water sources as quickly and efficiently as possible. In order to overcome this issue, we propose in this work to simplify the determination of WQI by replacing all the traditional analytical methods with UV–Vis spectrophotometry associated with ANN.

UV–Vis spectral changes of water in defined collection points could be well correlated with the WQI, so the use of an appropriate ANN would estimate the value of WQI from spectrophotometric patterns. The use of water UV–Vis spectrophotometric patterns associated with ANN to the prediction of the WQI contributes to the innovative nature of this study simplifying the determination of the index, reducing costs, analysis time, and generation of chemical waste.

Thus, the main stages for establishment of this methodology were (a) to determine the WQI of two rivers located in Assis, SP, Brazil, (b) to obtain the UV–Vis spectrum of each sample collected, and (c) to obtain an ANN to determine the WQI value from the water UV–

Vis spectrophotometric patterns. The methodology presented here could be applied with certain adjustments to any other river and places with lack of infrastructure as a useful alternative to WQI prediction. The adjustments for reproducibility of this methodology are related with the calibration of an ANN for a specific site before its establishment. A data set containing results of wavelength scan and WQI from water samples of a specific site could be used to calibrate an ANN according to the approach disclosed here.

Materials and methods

Study area

The chosen study area is located in the Médio Paranapanema watershed and essentially corresponds to the two headwaters of Assis, SP, the Água da Porca stream and Jacu stream. Five sampling points were chosen in each stream from their respective headwaters (Table 1).

We highlight that any other river could be a feasible alternative as a study case to apply the proposed methodology. The implication of choosing this area is that it has population of approximately 100,000 inhabitants. According to a census conducted in 2017 by the Brazilian Institute of Geography and Statistics (IBGE 2017), 94.6% of the Brazilian municipalities have population with less than 100,000 inhabitants; thus, the chosen area is representative regarding to the urbanization impact on water body’s caused by the major Brazilian cities.

Table 1 Sampling point locations

Sampling points	Latitude S	Longitude W	Sampling site description
AP-P1	22 38 34.85	50 25 6.05	Headwater next to Assis’ bus station
AP-P2	22 38 35.34	50 25 5.06	Lake formed by damming of the headwater
AP-P3	22 38 7.61	50 25 10.84	Point 2 km downstream from the headwater
AP-P4	22 36 32.05	50 26 8.49	Before Assis’ supply reservoir
AP-P5	22 36 32.51	50 26 58.90	After Assis’ supply reservoir
JS-P1	22 39 56.10	50 24 36.49	Headwater next to Perimetral Avenue
JS-P2	22 40 2.76	50 24 30.33	Headwater behind the town hall of Assis
JS-P3	22 40 51.20	50 24 17.02	Point 1.7 km downstream from JS-P2
JS-P4	22 41 36.86	50 24 15.73	Before Assis’ wastewater treatment station
JS-P5	22 42 27.34	50 23 56.24	After Assis’ wastewater treatment station

AP Água da Porca stream, JS Jacu stream

WQI determination

Water samples were collected once a month from March to December 2015, in the five collection points of each river, totaling 10 replicates of each collection point and a total set of 100 samples. The samples were collected in 1-L Duran Schott bottles (Duran, USA), previously autoclaved and identified. The volume collected at each point was 2 L. The measurement of pH and temperature was done in situ using a portable multiparameter meter (Orion 5-Star Plus, Thermo Scientific, Beverly, USA). After sample collection (between 8:00 and 10:00 a.m.), they were transported to the Laboratory of Process Engineering of Sao Paulo State University—Campus/ Assis for the accomplishment of physical, chemical, and microbiological analyses. The holding times for these samples were maximal 3 h.

The water quality index was determined according to the NSFQWI described by Brown et al. (1972). The equations used in this index are as follows:

$$WQI = \sum_{i=1}^n Wi \cdot Qi \quad (1)$$

$$\sum_{i=1}^n Wi = 1 \quad (2)$$

where Wi is the weighting factor associated with i th water quality parameter, Qi is the rating value for i th parameter, and n is the number of parameters.

The physical, chemical, and biological parameters analyzed were as follows: temperature change, pH, total nitrogen, DO, BOD₅, total phosphorus, thermotolerant coliform, turbidity, and suspended solids. Analyses of water samples for each parameter followed the methodology presented by American Public Health Association (APHA 2005).

UV–Vis spectrophotometry

A wavelength scan between 190 and 800 nm, at intervals of 1.0 nm, was conducted for each collected sample using a spectrophotometer (UV-M51, Bel Engineering, Piracicaba, Brazil). The analyses were carried out in quartz cuvettes, with 1.0 cm path length, after 3 h of sampling, as a time standardization. This procedure was performed in order to occur sedimentation of particles and thus prevent maximum absorbance values. Deionized water was used to set the spectrophotometer to zero. Afterwards, the graph of absorbance versus

wavelength was plotted for each sample using Microsoft Excel 2013 (Microsoft Corporation, Redmond, WA, USA), saving the data for subsequent analyses.

Data preprocessing

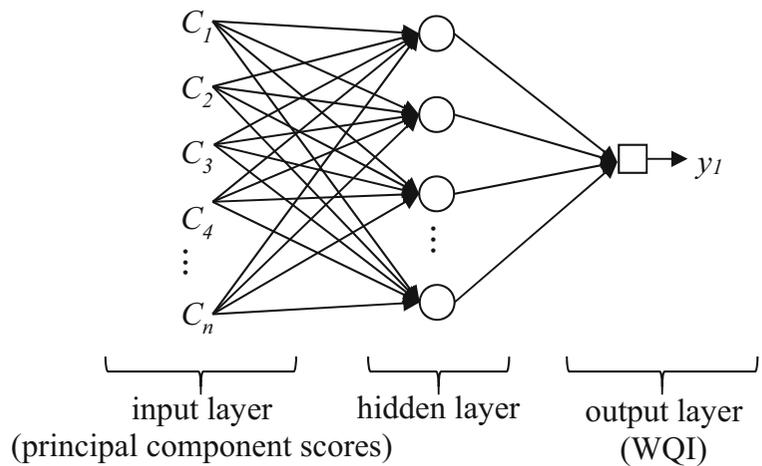
In order to ensure convergence of the ANN modeling, the absorbance values between 190 and 800 nm (611 variables) were preprocessed by normalization to the range from 0 to 1. These 611 variables correspond to the absorbance values obtained by the wavelength scan of each single sample from 190 to 800 nm using the cited spectrophotometer. The normalization process adjusts values measured on different scales to a common scale. Then, principal component analysis (PCA) was performed in the data set as described by Wu et al. (2010) for the purpose of dimension reduction with a minimum loss of information. The PCA approach uses all of the original variables to obtain a smaller set of principal components (PCs) which can be used to approximate the original variables (Sumi et al. 2012). As a result of the transformation, the first principal component has the largest possible variance; each succeeding component has the highest possible variance under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components.

Artificial neural network calibration

A simple three-layer ANN, equipped with the Levenberg–Marquardt training algorithm and hyperbolic tangent sigmoid transfer functions, was used as the benchmark model to obtain the WQI value based on the samples UV–Vis spectrophotometric patterns (Fig. 1). ANNs with one hidden layer can fit multidimensional mapping problems arbitrarily well and are considered to provide enough complexity to accurately simulate the nonlinear properties of this kind of problem by given consistent data and enough neurons in its hidden layer. The feed-forward back-propagation algorithm was employed to speed up the condition of convergence.

The obtained data, pretreated spectra, and WQI values were randomly distributed into three groups (70, 15, and 15% of the total data) to train, to test, and to validate, respectively, the ANN model. The implementation of the ANN was performed by using an algorithm developed in MATLAB platform that selects the minimum number of principal components used as inputs, and the number of neurons in the hidden layer

Fig. 1 ANN architecture and information pathway



that gives the best ANN model. The root-mean-square error (RMSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE) between the ANN predicted data and the real data were used for evaluation of the quality of the developed model. The RMSE, MAPE, and R^2 are frequently used as criteria to evaluate the performance of a network by comparing the error obtained from converged neural network runs and the measured data. For more details about equations and the advantages of adopting these particular evaluation criteria, see Olyaei et al. (2015) and Nabavi-Pelesaraei et al. (2017).

Economic analysis

An economic analysis was done in order to define the minimum number of samples (N_{minimum}), from which it would be financially viable to introduce a device that incorporates the methodology developed in this work. The operational variable costs, such as reagents, electric power, and labor, were considered; thereby, the unit variable cost (C_{uv}) for WQI single determination was defined. For fixed capital investment (FCI), all equipment and glassware utilized for physical, chemical, and microbiological assays for WQI quantification were taken into account. The cost of the conventional methodology (C_{CM}) can be represented mathematically as the following equation:

$$C_{CM} = FCI + C_{uv} \cdot N \tag{3}$$

On the other hand, the overall cost of the possible device derived from proposed methodology (C_{PM})

includes the cost with hardware (C_H) (a portable spectrophotometer and a computer) and the cost for ANN standardization using 100 samples. The C_{PM} can be represented mathematically as the following equation:

$$C_{PM} = C_H + C_{uv} \cdot 100 \tag{4}$$

The costs for sample collection and laboratory building were not included in the economic analysis, since this simplification does not interfere in N_{minimum} calculations. Brazilian currency (Brazilian Real, BRL) was used as reference currency.

Results

WQI

The average WQI values for each sampling point over the 10-month study are shown in Fig. 2 (see Appendix Table 5 for a table showing the results of all physical, chemical, and biological parameters analyzed in each sample). As observed, only the points AP-P1, AP-P4, AP-P5, and JS-P1 have, on average, good rating (Fig. 2a, b). On the other hand, JS-P5 point stands out with the lower WQI on average. For the other points, a precise classification cannot be inferred, because their standard error values cover both good and medium classes.

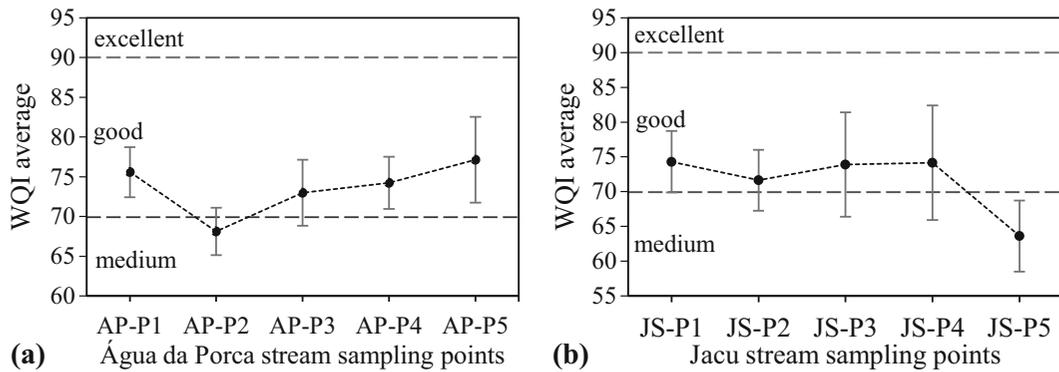


Fig. 2 WQI average value along the Água da Porca stream (a) and Jacu stream (b) between March and December 2015. Error bars are the standard deviations

Water UV–Vis spectrum

Figure 3 shows the average absorbance between 190 and 800 nm for each sampling point. The UV–Vis spectra were characterized by peaks with maximum absorbance around 190 and 210 nm (Fig. 3a, b). Looking closely at the JS-P5 spectrum, a subtle decreasing concave shape on a wide range of wavelengths can be observed, with a slight shoulder at around 260–280 nm. A decreasing profile at this same wavelength range was observed for points AP-P2, AP-P3, AP-P4, and AP-P5, however without a marked shoulder. For the major part of the spectra for both streams from 500 nm on, the absorbance values were found tending to zero.

Artificial neural network

The best ANN architecture was reached by using the scores of the first 19 principal components as inputs and

16 neurons in the hidden layer (Table 2). As shown in Table 3, the first 19 significant components explained the variability of original spectral data in 99.99% (cumulative coefficient of determination, $R^2 = 0.9999$, corresponds to the fraction of the variance in original data that is explained by the significant principal components), which comprised 611 descriptor variables (absorbance values related to 611 wavelengths, 190–800 nm). The UV–Vis spectral data were well correlated with WQI values (Fig. 4). WQI modeling by optimal ANN showed satisfactory correlations: 0.9933 for training, 0.9940 for validation, and 0.9880 for testing. The optimal ANN performance was assessed by the R^2 (0.9857), which measures the proportion of the variance in the dependent variable that is predictable from the independent variable and by the RMSE (0.5813), which indicates the discrepancy between the observed and forecasted values. Briefly, the predictions of the model are optimum if R^2 and RMSE are found to be close to 1

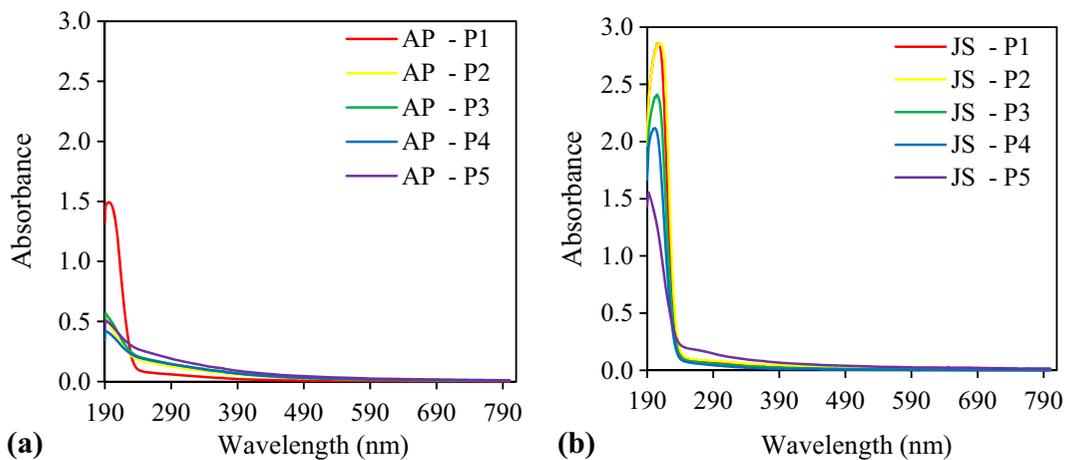


Fig. 3 Average UV–Vis spectra for samples of Água da Porca stream (a) and Jacu stream (b)

Table 2 The best fitting models of the applied ANN modeling

Topology	R^2	RMSE	MAPE (%)
16-18-1	0.9659	1.4194	0.8755
17-18-1	0.9757	0.9883	0.6603
18-16-1	0.9803	0.7995	0.6488
<i>19-16-1</i>	<i>0.9857</i>	<i>0.5813</i>	<i>0.5715</i>
20-14-1	0.9779	0.8947	0.5946
21-18-1	0.9809	0.7733	0.5896

The italicized numbers are relatively high values for R^2 and relatively low values for RMSE and MAPE. Scanning was performed with input neurons number ranging from 1 to 30. Topologies with input neuron numbers below 16 and over 21 presented no good performances and were not listed in the table

ANN artificial neural network, R^2 coefficient of determination, RMSE root-mean-square error, MAPE mean absolute percentage error

and 0, respectively. The accuracy of ANN modeling, evaluated by means of MAPE, showed a value of 0.57% with a standard deviation of 0.51%.

Economic analysis

The economic analysis was based on a rigorous calculation of operational costs and fixed capital investment for conventional WQI methodology, as well as the cost for the possible device that can be produced using the proposed approach (Table 4). The cost for ANN standardization using 100 samples (354,414.88 BRL) and the cost of a portable spectrophotometer and computer (7500 BRL) composed the overall cost of the proposed methodology (42,914.88 BRL). Then, the minimum number of samples (465) that justified financially the present methodology was calculated by the interception of lines, which represent costs of conventional and proposed methodologies as functions of sample number (Fig. 5).

Discussion

UV-Vis spectroscopy has been increasingly used in water quality monitoring, since many of the compounds that affect negatively the water have characteristic absorbance profile. Normally, if the UV spectrum of natural water is flat and close to zero, the pollution probability is very low (Thomas and Burgess 2007). In this way, several studies have shown the correlation between UV absorption and some pollutants. Rieger et al. (2004) developed a

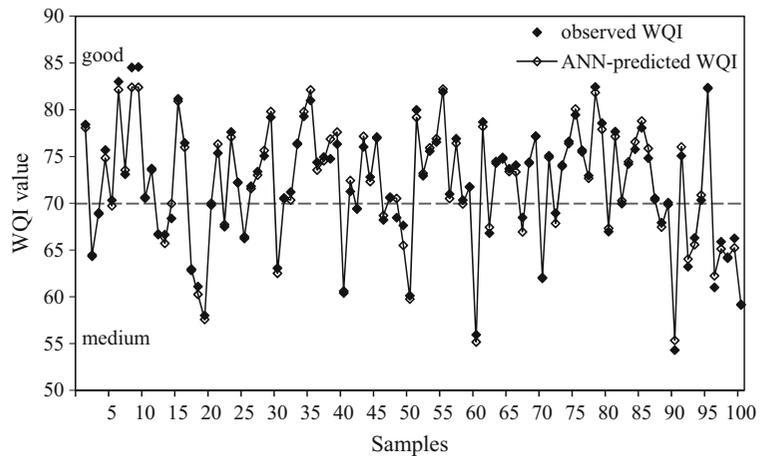
Table 3 Percentage of the total variance explained by each one of the first 19 principal components

<i>i</i> th principal component	Percentage (%)	Cumulative percentage (%)
1	87.3137	87.3137
2	8.0653	95.3790
3	2.0272	97.4062
4	1.0675	98.4737
5	0.9526	99.4263
6	0.2779	99.7042
7	0.1323	99.8365
8	0.0842	99.9207
9	0.0262	99.9469
10	0.0216	99.9685
11	0.0107	99.9792
12	0.0042	99.9834
13	0.0034	99.9867
14	0.0020	99.9887
15	0.0013	99.9900
16	0.0011	99.9911
17	0.0010	99.9922
18	0.0009	99.9930
19	0.0008	99.9938

methodology for the measurement of nitrite and nitrate, simultaneously, based on UV absorption followed by data analysis using partial least square (PLS). According to them, both nitrate and nitrite have their highest absorbance between 210 and 240 nm and only very weak absorption peak at 300 nm (nitrate) and at 360 nm (nitrite). Thus, the peaks with maximum absorbance around 190 and 210 nm observed in the present work could be explained by the presence of nitrate. As nitrate is the most stable form of nitrogen in water, resulting from the oxidation of all other dissolved N compounds (nitrite, ammonia or organic nitrogen), it is contained in the majority of surface water, specially wastewater (Thomas et al. 2017).

Brookman (1997) investigated the absorbance at 280 nm for farm slurry effluents and found a good exponential relationship between absorbance and BOD₅, showing that this approach could be used to predict this parameter. Dobbs et al. (1972) showed that UV absorption of total organic matter at 254 nm can be applied for its quantification in effluents and surface waters. The UV absorptions between 260 and 270 nm observed in this study, including the slight shoulder, is likely due to the presence of organic matter and phenolic-type chromophores,

Fig. 4 Comparison of observed and ANN-predicted WQI values for each individual observation



respectively. According to Thomas and Burgess (2007), humic substances or humic-like substances (HLS) composed of humic acids, fulvic acids, and related substances correspond to the major part of organic matter. These substances have chromophore groups susceptible to be significantly absorbed by the UV light, but no specific spectrum shape is usually associated with this organic matrix. However, higher percentage of phenolic-type chromophores associated with some HLS of ponds or wetlands was found to be responsible to produce marked shoulders.

According to Huber and Frost (1998), analysis of the general shape of the spectrum or absorption at a specific wavelength can be also used to derive more specific parameters, such as turbidity and suspended solids by using algorithm for compensation of light scattering caused by suspended particles. Due to this physical

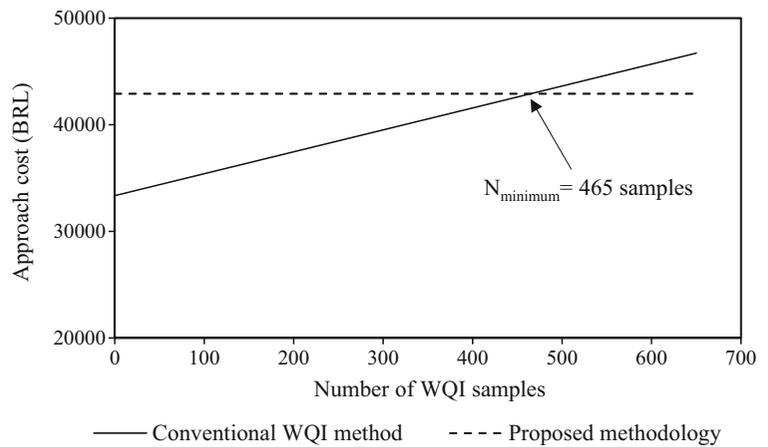
diffuse absorption, the absorbance value and the general mean slope of the spectrum in the 250–350 nm region can vary. Therefore, a large amount of valuable information can be obtained using the entire spectrum of UV–Vis absorption of the water, which supports its use as a consistent alternative to the WQI determination when associated with artificial intelligence techniques.

Many articles use ANN to predict some of the water quality parameters due to its capability to predict non-linear time series. Neto et al., (2014) demonstrated the effectiveness of virtual sensors in monitoring physico-chemical parameters and metal concentrations in water of reservoirs in the Amazon area by ANN and remote sensing images. Their results provide researchers an accurate and less expensive alternative resource in environmental monitoring processes. Gazzaz et al. (2012), in a previous study, developed an ANN model for the

Table 4 Primary data used for economic analysis

Cost type	Item	Unit variable cost (BRL/WQI sample)
Operational variable costs (conventional methodology)	Reagents	2.34
	Electric power	6.97
	Labor	11.25
Total unit variable cost (C_{uv})		20.56
Fixed capital investments (FCI) (conventional methodology)	Item	Value (BRL)
	Equipments	32,055.00
	Glassware	1304.00
Total		33,359.00
Device based on present methodology	Portable computer	1000.00
	Portable spectrophotometer	6500.00
Total		7500

Fig. 5 Determination of minimum number of samples, from which the proposed methodology is economically viable



prediction of water quality index for Kinta River, Malaysia from physical, chemical, and microbiological data. The results demonstrated that the proposed model was able to simplify all the mathematical calculations involved with the WQI, reducing the computation time, but not to reduce the analytical work. Alizadeh and Kavianpour (2015) examined the ability of ANN and wavelet-neural network (WNN) models to predict daily values of salinity, temperature and DO based on the existing measured data of chlorophyll, DO, salinity, turbidity, and water temperature in Hilo Bay, Pacific Ocean. The results showed that the model was able to monitor the ocean parameters even when data were missing. Apart from this applicability, the use of soft computing techniques such as ANN is related to many studies aiming to reduce the computation time and effort and the possibility of errors in the calculation. The approach here presented allows the prediction of the WQI and reduces, in addition to the mathematical calculations, all the analytical work involved in the quantification of the nine parameters that composes the WQI, when the method is standardized.

The ANN model developed in this work was created using pretreated absorbance values as input data, one hidden layer, and a single output layer defined by the WQI. This ANN benchmark architecture was similar to other previously reported for water quality index modeling, where one hidden layer were also suitable (Khuan et al. 2002). According to Sheremetov et al. (2014), one hidden layer is sufficient for the large majority of problems, and ANNs with more hidden layers are extremely hard to train.

The appropriate number of neurons in one hidden layer suggested by Fletcher and Goss (1993) should range from

$(2n^{1/2} + m)$ to $(2n + 1)$, where n is the number of input neurons and m is the number of output nodes. In this work, the best ANN model showed $n = 19$ and $m = 1$; therefore, the hidden layer neuron number (16 neurons) is in compliance with this law, once it ranges from 10 to 39. Increasing the number of neurons in the hidden layer increases the non-linear mapping capacity of the network. However, when this number is too large, the model may overadjust to the data in the presence of noise in the training samples. It is said that the network is subject to overfitting. On the other hand, a network with few neurons in the hidden layer may not be able to perform the desired mapping, which is called underfitting.

Regarding the network performance, Prajithkumar et al. (2014) obtained a correlation coefficient between the measured and predicted values by ANN close to our study ($R = 0.9900$, $p < 0.01$) when studied the prediction of the WQI for the Pavana River, India using modular ANN model. However, this study used as input data the following parameters: concentration of DO, nitrates, calcium, total suspended solids, total hardness, alkalinity, chlorides and values of pH, turbidity, and conductivity. Thus, this ANN modeling, despite the good results achieved, still has the disadvantage regarding the need of time-consuming analysis and the generation of pollutants to the prediction of WQI.

From the results of the present work, a portable UV-Vis spectrophotometer coupled to a computer could be used to predict the WQI in situ, at low cost, without chemical waste generation, and almost in real time, which would allow quickly decisions regarding the preservation of the water resources. Brazil is a privileged country with regard to the water, being the first placed in water availability in the world. Although

plentiful, this resource is unevenly distributed between the regions and the use of certain water sources has been restricted due to contamination. In addition, most of the rivers located in small towns do not have the quality of their waters monitored. Therefore, all advantages of the proposed approach would be very useful, as it makes possible to monitor the water quality in places where there is no needed infrastructure to determine the WQI by the conventional method.

The economic analysis demonstrated that the developed methodology can be used when intensive monitoring of water resources is demanded. Chovanec and Winkler (1994) described the national river monitoring program of Austria. The program was established, following a Federal Act on Water Law in 1990, with 140 sampling sites, and subsequently was extended to 250 sites aiming to allow more detailed interpretations. As samples were taken bimonthly, a total of 500 samples were analyzed per month. Therefore, the presented methodology could be, for example, a financially viable alternative to water quality monitoring programs such that cited. Moreover, it can be useful for geographical regions or cities with many aquifers and/or rigorous environmental policies to ensure watershed protection.

Conclusion

Through this work, it was possible to verify that UV–Vis spectroscopy in combination with ANN model was able to determine the WQI. The implications of this work’s results open up the possibility to use a portable UV–Vis spectrophotometer connected to a computer to predict the WQI in places where there is no required infrastructure to determine the WQI by the conventional method. Once this methodology is established, no sample preparation and no reagents are required, reducing the analysis cost, time, and the generation of chemical waste. Suggested improvements and recommendation for future work is to use samples from different aquifers in order to demonstrate the comprehensiveness of the presented approach.

Funding information The authors gratefully acknowledge São Paulo Research Foundation (FAPESP) for Scientific Initiation Scholarship (grant 2014/26025-2) and Ivan Esperança Rocha, director of the Faculdade de Ciências e Letras de Assis, for financial support.

Appendix

Table 5 Results of the parameters analyzed of each sample

Sample number	Stream	ΔT (°C)	pH	Turbidity NTU	DO (% sat)	BOD ₅ (mg/L)	Coliform NMP	Total P (mg/L)	Total N (mg/L)	SS (mg/L)	WQI
1	AP-P1	0.00	6.73	1.72	70.00	0.00	920.00	0.03	0.40	152.76	83.01
2	AP-P2	0.40	7.07	26.22	12.00	0.00	1600.00	0.04	0.93	603.00	73.11
3	AP-P3	- 0.40	6.89	40.27	51.00	1.23	920.00	0.04	0.58	134.00	84.52
4	AP-P4	- 0.40	6.94	34.64	72.00	2.11	240.00	0.06	0.26	120.60	84.57
5	AP-P5	0.70	6.79	90.22	56.00	1.23	240.00	0.41	0.45	138.69	70.64
6	JS-P1	0.00	7.41	6.22	97.00	1.58	350.00	0.03	0.36	152.76	83.01
7	JS-P2	0.70	11.47	8.71	80.00	5.96	0.00	0.04	0.39	603.00	73.11
8	JS-P3	2.10	7.51	1.81	92.00	0.97	170.00	0.04	0.15	134.00	84.52
9	JS-P4	1.50	7.42	0.65	91.00	0.88	240.00	0.06	0.22	120.60	84.57
10	JS-P5	1.10	7.27	21.11	71.00	5.27	240.00	0.41	0.98	138.69	70.64
11	AP-P1	0.00	7.08	4.40	56.20	2.88	240.00	0.04	0.52	103.45	76.45

Table 5 (continued)

Sample number	Stream	ΔT (°C)	pH	Turbidity NTU	DO (% sat)	BOD ₅ (mg/L)	Coliform NMP	Total P (mg/L)	Total N (mg/L)	SS (mg/L)	WQI
12	AP-P2	0.10	6.74	21.36	30.20	2.13	130.00	0.03	1.67	88.84	62.83
13	AP-P3	-0.90	6.77	35.95	58.50	3.81	1600.00	0.28	0.99	89.58	61.10
14	AP-P4	-1.00	6.44	42.69	71.60	6.09	540.00	0.09	0.82	160.47	58.01
15	AP-P5	0.20	6.72	12.75	99.00	4.19	49.00	0.69	0.32	163.15	69.81
16	JS-P1	0.00	7.10	9.40	90.00	5.20	920.00	0.04	1.30	103.45	76.45
17	JS-P2	-0.40	6.56	492.54	69.00	4.49	35,000.00	0.03	1.99	88.84	62.83
18	JS-P3	-0.70	7.48	150.60	62.00	3.96	24,000.00	0.28	2.77	89.58	61.10
19	JS-P4	-0.80	7.70	327.23	45.00	3.26	1600.00	0.09	4.17	160.47	58.01
20	JS-P5	-0.20	7.19	32.89	52.00	3.52	2.00	0.69	4.95	163.15	69.81
21	AP-P1	0.00	6.75	3.69	50.50	1.11	130.00	0.04	0.39	140.90	71.82
22	AP-P2	1.71	6.50	14.64	22.50	1.61	49.00	0.03	1.42	161.94	73.38
23	AP-P3	0.11	6.46	5.33	47.00	1.82	6.80	0.04	0.46	131.39	75.07
24	AP-P4	0.61	6.42	26.51	49.80	4.36	7.80	0.05	0.35	117.72	79.20
25	AP-P5	2.51	6.40	24.62	46.00	0.62	70.00	0.69	0.25	142.91	63.06
26	JS-P1	0.00	7.11	5.97	54.30	3.78	170.00	0.04	2.11	140.90	71.82
27	JS-P2	1.00	6.78	4.53	50.30	0.49	350.00	0.03	0.37	161.94	73.38
28	JS-P3	-0.20	7.09	2.53	60.00	0.67	350.00	0.04	4.67	131.39	75.07
29	JS-P4	1.20	7.16	1.36	57.60	0.54	79.00	0.05	0.48	117.72	79.20
30	JS-P5	2.30	7.26	24.84	46.20	3.32	79.00	0.69	4.59	142.91	63.06
31	AP-P1	0.00	6.82	12.45	50.20	1.90	540.00	0.03	0.59	99.29	74.38
32	AP-P2	1.20	6.86	16.31	62.20	4.13	240.00	0.02	0.79	151.49	74.95
33	AP-P3	1.20	6.77	8.70	53.90	0.65	79.00	0.03	0.28	126.50	74.76
34	AP-P4	0.30	6.58	16.47	65.10	1.36	23.00	0.04	0.21	102.38	76.33
35	AP-P5	2.20	6.74	13.17	60.40	0.50	7.80	0.56	0.19	126.09	60.41
36	JS-P1	0.00	7.02	4.41	63.30	1.92	1600.00	0.03	0.33	99.29	74.38
37	JS-P2	0.50	6.99	3.01	57.90	0.47	920.00	0.02	0.14	151.49	74.95
38	JS-P3	0.30	7.15	2.37	62.10	1.28	3500.00	0.03	0.16	126.50	74.76
39	JS-P4	-0.60	7.13	3.66	64.60	1.13	1600.00	0.04	0.21	102.38	76.33
40	JS-P5	0.10	7.15	19.37	46.50	3.57	3500.00	0.56	2.52	126.09	60.41
41	AP-P1	0.00	6.78	2.53	46.70	1.63	540.00	0.05	1.68	107.94	68.22
42	AP-P2	-0.80	6.86	10.48	40.30	2.10	170.00	0.02	1.19	150.75	70.57

Table 5 (continued)

Sample number	Stream	ΔT (°C)	pH	Turbidity NTU	DO (% sat)	BOD ₅ (mg/L)	Coliform NMP	Total P (mg/L)	Total N (mg/L)	SS (mg/L)	WQI
43	AP-P3	-1.10	6.91	5.70	51.60	0.55	130.00	0.30	0.45	117.18	68.46
44	AP-P4	-1.80	6.73	49.04	54.60	0.43	79.00	0.37	0.55	100.17	67.64
45	AP-P5	-0.90	6.80	22.24	57.90	0.36	49.00	0.79	0.41	105.19	60.12
46	JS-P1	0.00	6.92	4.78	56.90	3.45	2400.00	0.05	0.84	107.94	68.22
47	JS-P2	0.40	6.82	3.70	53.80	0.92	1600.00	0.02	0.31	150.75	70.57
48	JS-P3	-0.80	7.20	11.30	60.10	1.74	2400.00	0.30	0.55	117.18	68.46
49	JS-P4	-0.80	7.29	8.52	57.80	1.95	1600.00	0.37	0.40	100.17	67.64
50	JS-P5	-1.40	7.30	25.34	45.90	3.28	2400.00	0.79	3.92	105.19	60.12
51	AP-P1	0.00	6.84	10.69	51.90	4.08	0.40	0.03	1.44	115.44	70.99
52	AP-P2	0.20	6.83	15.54	62.60	5.04	70.00	0.03	1.38	163.68	76.91
53	AP-P3	-0.40	6.83	7.54	55.90	4.46	23.00	0.13	0.54	146.86	70.34
54	AP-P4	-0.80	6.62	9.54	65.20	4.42	23.00	0.14	0.30	129.98	71.77
55	AP-P5	1.50	6.95	10.82	71.60	4.86	4.50	0.95	0.34	175.41	55.94
56	JS-P1	0.00	7.41	4.55	62.30	4.39	3500.00	0.03	0.78	115.44	70.99
57	JS-P2	1.30	7.13	2.46	56.10	0.75	170.00	0.03	0.17	163.68	76.91
58	JS-P3	0.90	7.33	12.69	61.00	2.13	3500.00	0.13	0.64	146.86	70.34
59	JS-P4	0.50	7.32	17.82	62.80	1.93	1600.00	0.14	0.76	129.98	71.77
60	JS-P5	1.50	7.25	40.67	50.90	3.26	1600.00	0.95	6.11	175.41	55.94
61	AP-P1	0.00	6.93	7.72	65.90	0.97	130.00	0.08	2.64	127.64	74.08
62	AP-P2	-0.90	6.80	22.24	19.50	0.62	130.00	1.28	2.21	181.10	68.48
63	AP-P3	-0.90	6.80	5.58	47.50	2.07	46.00	0.22	0.91	131.99	74.40
64	AP-P4	-1.30	6.68	13.46	53.50	0.20	79.00	0.19	0.31	134.40	77.17
65	AP-P5	-0.20	6.44	15.77	48.10	0.15	46.00	1.08	0.37	159.86	61.99
66	JS-P1	0.00	7.10	5.40	62.20	3.27	240.00	0.08	0.62	127.64	74.08
67	JS-P2	-0.20	6.96	29.85	63.80	1.60	130.00	1.28	0.17	181.10	68.48
68	JS-P3	-1.20	7.22	31.14	61.10	0.40	170.00	0.22	0.62	131.99	74.40
69	JS-P4	0.50	7.04	5.66	60.20	0.81	130.00	0.19	0.21	134.40	77.17
70	JS-P5	0.50	7.27	30.93	49.00	3.66	46.00	1.08	4.60	159.86	61.99
71	AP-P1	0.00	6.90	12.96	84.00	6.14	240.00	0.05	3.17	135.88	75.49
72	AP-P2	1.70	6.62	6.47	49.10	3.26	240.00	0.04	1.89	180.63	72.96
73	AP-P3	-0.30	6.75	7.80	48.90	2.58	33.00	0.04	0.81	141.64	82.44

Table 5 (continued)

Sample number	Stream	ΔT (°C)	pH	Turbidity NTU	DO (% sat)	BOD ₅ (mg/L)	Coliform NMP	Total P (mg/L)	Total N (mg/L)	SS (mg/L)	WQI
74	AP-P4	-0.50	6.58	17.87	57.00	1.19	23.00	0.30	2.38	123.08	78.57
75	AP-P5	2.40	6.73	16.37	57.00	1.49	4.50	0.75	0.53	136.01	66.97
76	JS-P1	0.00	7.54	4.49	83.50	5.86	1600.00	0.05	0.88	135.88	75.49
77	JS-P2	-0.60	8.84	3.61	70.80	2.98	240.00	0.04	0.20	180.63	72.96
78	JS-P3	-1.40	7.33	1.85	82.40	3.12	46.00	0.04	0.53	141.64	82.44
79	JS-P4	0.00	7.38	2.30	62.30	1.93	33.00	0.30	0.35	123.08	78.57
80	JS-P5	0.50	7.36	18.41	54.90	4.23	17.00	0.75	4.52	136.01	66.97
81	AP-P1	0.00	7.20	4.52	67.30	4.21	33.00	0.03	1.66	139.70	74.83
82	AP-P2	2.00	6.79	11.27	33.30	2.19	33.00	0.07	1.17	155.44	70.54
83	AP-P3	0.60	7.01	12.50	53.00	2.67	70.00	0.07	0.55	102.11	67.93
84	AP-P4	0.00	6.72	18.54	53.10	1.33	33.00	0.09	0.36	88.98	70.06
85	AP-P5	2.00	6.78	24.46	61.10	1.78	13.00	0.41	0.41	93.06	54.31
86	JS-P1	0.00	7.56	3.66	76.60	3.55	920.00	0.03	1.39	139.70	74.83
87	JS-P2	0.20	8.30	5.38	52.50	1.84	350.00	0.07	1.41	155.44	70.54
88	JS-P3	-0.70	7.33	37.50	56.20	1.49	3500.00	0.07	1.52	102.11	67.93
89	JS-P4	0.20	7.29	36.37	62.30	1.40	9200.00	0.09	1.39	88.98	70.06
90	JS-P5	1.00	7.27	116.03	47.80	3.32	1600.00	0.41	4.22	93.06	54.31
91	AP-P1	0.00	7.13	7.44	54.00	1.45	130.00	0.14	0.64	139.70	61.00
92	AP-P2	0.70	6.82	26.46	33.00	2.30	920.00	0.12	1.45	155.44	65.90
93	AP-P3	0.30	6.93	34.40	50.50	3.15	920.00	0.16	1.11	102.11	64.13
94	AP-P4	-0.10	7.61	35.97	44.10	1.50	130.00	0.21	0.74	88.98	66.26
95	AP-P5	1.00	6.69	19.25	58.70	1.40	2.00	0.81	0.45	93.06	59.14
96	JS-P1	0.00	7.31	102.74	52.10	3.79	1600.00	0.14	1.11	139.70	61.00
97	JS-P2	-0.20	7.15	44.04	51.90	1.29	9200.00	0.12	0.75	155.44	65.90
98	JS-P3	-1.20	7.23	124.00	56.10	1.49	3500.00	0.16	0.89	102.11	64.13
99	JS-P4	-1.10	7.16	238.86	53.40	1.57	210.00	0.21	0.97	88.98	66.26
100	JS-P5	-0.90	7.23	552.08	55.10	4.35	350.00	0.81	0.97	93.06	59.14

AP Água da Porca stream, JS Jacu stream, ΔT temperature change, DO dissolved oxygen, BOD₅ biochemical oxygen demand, Total N total nitrogen, Total P total phosphorus, SS suspended solids, WQI water quality index

References

- Alizadeh, M. J., & Kavianpour, M. R. (2015). Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine Pollution Bulletin*, *98*, 171–178.
- Anumol, T., Sgroi, M., Park, M., Roccaro, P., & Snyder, S. A. (2015). Predicting trace organic compound breakthrough in granular activated carbon using fluorescence and UV absorbance as surrogates. *Water Research*, *76*, 76–87.
- APHA. (2005). *Standard methods for the examination of water and wastewater*. Washington: American Public Health Association.
- Bharti, N., & Katyal, D. (2011). Water quality indices used for surface water vulnerability assessment. *International Journal of Environmental Sciences*, *2*, 154–173.
- Brookman, S. K. E. (1997). Estimation of biochemical oxygen demand in slurry and effluents using ultra-violet spectrophotometry. *Water Research*, *31*, 372–374.
- Brown RM, McLelland NI, Deininger RA, O'Connor MF (1972) A water quality index—crashing the psychological barrier. In: Jenkins SH (ed) *Advances in water pollution research: Proceedings of the Sixth International Conference Held in Jerusalem*. Pergamon Press, New York, pp. 787–794.
- Chen, X. Y., & Chau, K. W. (2016). A hybrid double feedforward neural network for suspended sediment load estimation. *Water Resour Manag*, *30*(7), 2179–2194.
- Chovanec, A., & Winkler, G. (1994). River monitoring and water quality assessment in Austria. *Toxicological and Environmental Chemistry*, *44*, 161–176.
- Dobbs, R. A., Wise, R. H., & Dean, R. B. (1972). The use of ultra-violet absorbance for monitoring the total organic carbon content of water and wastewater. *Water Research*, *6*, 1173–1180.
- Farmaki, E. G., Thomaidis, N. S., & Efstathiou, C. E. (2010). Artificial neural networks in water analysis: theory and applications. *Int J Environ An Ch*, *90*, 85–105.
- Ferneda, E. (2006). Neural networks and its application in information retrieval systems (Redes neurais e sua aplicação em sistemas de recuperação de informação). *Ciência da Informação*, *35*, 25–30.
- Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: an application using bankruptcy data. *Inform. Manage*, *24*, 159–167.
- Gazzaz, N. M., Yusoff, M., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin*, *64*, 2409–2420.
- Gerrity, D., Gamage, S., Jones, D., Korshin, G. V., Lee, Y., Pisarenko, A., & Snyder, S. A. (2012). Development of surrogate correlation models to predict trace organic contaminant oxidation and microbial inactivation during ozonation. *Water Research*, *46*, 6257–6272.
- Huber, E., & Frost, M. (1998). Light scattering by small particles. *J Water Supply Res T*, *47*, 87–94.
- IBGE – Brazilian Institute of Geography and Statistics (Instituto Brasileiro de Geografia e Estatística) (2017). Retrieved from: https://ww2.ibge.gov.br/home/estatistica/populacao/estimativa2017/estimativa_dou.shtm. April 17, 2018.
- Khuan, L. Y., Hamzah, N., & Jailani, R. (2002) Prediction of water quality index (WQI) based on artificial neural network (ANN). In *Research and development, 2002. Student Conference on Research and Development Proceedings*, Shah Alam, pp. 157–161.
- Kuo, J. T., Hsieh, M. H., Lung, W. S., & She, N. (2007). Using artificial neural network for reservoir eutrophication prediction. *Ecological Modelling*, *200*(1), 171–177.
- Maier, H. R., & Dandy, G. C. (1996). The use of artificial neural networks for the prediction of water quality parameters. *Water Resources Research*, *32*(4), 1013–1022.
- Nabavi-Pelesaraei, A., Bayat, R., Hosseinzadeh-Bandbafha, H., Afrasyabi, H., & Chau, K. W. (2017). Modeling of energy consumption and environmental life cycle assessment for incineration and landfill systems of municipal solid waste management—a case study in Tehran Metropolis of Iran. *Journal of Cleaner Production*, *148*, 427–440.
- Najah, A., Elshafie, A., Karim, O. A., & Jaffar, O. (2009). Prediction of Johor River water quality parameters using artificial neural networks. *European Journal of Scientific Research*, *28*(3), 422–435.
- Neto, B. R., Hauser-Davis, R. A., Lobato, T. C., Saraiva, A. C. F., Brandão, I. L. S., Oliveira, T. F. O., & Silveira, A. M. (2014). Estimating physicochemical parameters and metal concentrations in hydroelectric reservoirs by virtual sensors: a case study in the Amazon Region. *Computing in Science & Engineering*, *4*(2), 43–53.
- Olyaie, E., Banejad, H., Chau, K. W., & Melesse, A. M. (2015). A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. *Environmental Monitoring and Assessment*, *187*(4), 189.
- Poonam, T., Tanushree, B., & Sukalyan, C. (2013). Water quality indices-important tools for water quality assessment: a review. *Int J Adv Chem*, *1*, 15–28.
- Prajithkumar, S., Verma, D. S., & Mane, S. J. (2014). Prediction of water quality index of Pavna River using ANN model. *Int J Eng Res Technol*, *3*, 121–125.
- Rene, E. R., & Saidutta, M. B. (2008). Prediction of water quality indices by regression analysis and artificial neural networks. *Int J Environ Res*, *2*, 183–188.
- Rieger, L., Langergraber, G., Thomann, M., Fleischmann, N., & Siegrist, H. (2004). Spectral in-situ analysis of NO₂, NO₃, COD, DOC and TSS in the effluent of a WWTP. *Water Science and Technology*, *50*, 143–152.
- Rieger, L., Langergraber, G., & Siegrist, H. (2006). Uncertainties of spectral in situ measurements in wastewater using different calibration approaches. *Water Science and Technology*, *53*, 187–197.
- Roccaro, P., Yan, M., & Korshin, G. V. (2015). Use of log-transformed absorbance spectra for online monitoring of the reactivity of natural organic matter. *Water Research*, *84*, 136–143.
- Sheremetov, L., Cosultchi, A., Martínez-Muñoz, J., Gonzalez-Sánchez, A., & Jiménez-Aguino, M. A. (2014). Data-driven forecasting of naturally fractured reservoirs based on nonlinear autoregressive neural networks with exogenous input. *Journal of Petroleum Science and Engineering*, *123*, 106–119.
- Sumi, S. M., Zaman, M. F., & Hirose, H. (2012). A rainfall forecasting method using machine learning models and its application to

- the Fukuoka city case. *International Journal of Applied Mathematics and Computer Science*, 22(4), 841–854.
- Thomas, O., & Burgess, C. (2007). *UV-visible spectrophotometry of water and wastewater*. Amsterdam: Elsevier.
- Thomas, O., Causse, J., Jung, A. V., & Thomas, M. F. (2017). Natural water. In: *UV-visible spectrophotometry of water and wastewater* (Second Edition). Elsevier, Amsterdam, pp. 225–259.
- Torres, A., & Bertrand-Krajewski, J. (2008). Partial least squares local calibration of a UV-visible spectrometer used for in situ measurements of COD and TSS concentrations in urban drainage systems. *Water Science and Technology*, 57, 581–588.
- Tyagi, S., Sharma, B., Singh, P., & Dobhal, R. (2013). Water quality assessment in terms of water quality index. *Am J Water Resou*, 1(3), 34–38.
- Van den Broeke, J., Langergraber, G., & Weingartner, A. (2006). On-line and in-situ UV/vis spectroscopy for multi-parameter measurements: a brief review. *Spectroscopy Europe*, 18, 15–18.
- Vanrolleghem, P. A., & Lee, D. S. (2003). On-line monitoring equipment for wastewater treatment processes: state of the art. *Water Science and Technology*, 47, 1–34.
- Wu, C. L., Chau, K. W., & Fan, C. (2010). Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*, 389(1), 146–167.
- Yan, M., Korshin, G. V., & Chang, H. S. (2014). Examination of disinfection by-product (DBP) formation in source waters: a study using log-transformed differential spectra. *Water Research*, 50, 179–188.