

# Prediction of some physico-chemical parameters of water in Alton Reservoir, Suffolk, England

HIEDA A. NASCIMENTO SILVA  
UFPA, Federal University of Pará, Belém  
Department of Applied Mathematics  
Pará 66075-110  
BRAZIL

PAOLA G. VINUEZA NARANJO  
Universidad Nacional de Chimborazo  
Faculty of Engineering  
Avenida Antonio José de Sucre, Riobamba  
ECUADOR

LENA P. SOUZA RODRIGUES  
UFRA, Federal Rural University of Amazonia  
Institute of Mechanical Engineering  
Estr. Principal da Ufra, 2150 - Curió Utinga  
BRAZIL

DIOVANNI MORAES DE ARAUJO  
UFPA, Federal University of Pará, Belém  
Department of Applied Mathematics  
Pará 66075-110  
BRAZIL

YOMARA PINHEIRO PIRES  
UFPA, Federal University of Pará, Belém  
Department of Applied Mathematics  
Pará 66075-110  
BRAZIL

*Abstract:* Predict water quality variables such as Chlorophyll-a (CHL), Algae, Turbidity and Total Suspended Solids (TSS) are important for the analysis of freshwater ecosystems, that are significant not only for human populations but also essential for plant and animal diversity. However, monitoring all these variables from space is a very challenging task, which becomes particularly difficult when dealing with cyanobacteria blooms, because in high concentrations, they form scum on the water surface, which is a concern for public health due to the production of toxins. This article describes empirical algorithms to estimate these variables using Landsat-8 and Sentinel-2 images, multi-spectral instrument data, the Landsat spatial resolution (30 m) and imagery from the Sentinel-2 sensor, with a resampled 10 m spatial resolution can be used for environmental monitoring. These images, analyzed by Wavelets Neural Networks can be very useful to estimate physico-chemical and biological parameters of water. This approach is applied in Alton water reservoir, Suffolk, UK using spatial and temporal scales. The Alton Reservoir is the largest in Suffolk, with a perimeter of over 8 miles (13 km). This article presents techniques based on wavelets neural networks and fuzzy neural networks, namely the radial basis function, the adaptive neuro-fuzzy inference system and Least Square Estimat, which are well suited to predict data sequences stemming from real-world applications techniques. The prediction behavior shows good forecasts as (NMSE = 0.00004; MARE = 0.00078, MSE =0.00013) for test data, results showed that the predicted values have good accurate. This article contributes to improving efficiency to monitor water quality parameters and confirm the reliability and accuracy of the approaches proposed for monitoring water reservoirs.

*Key-Words:* Wavelet transform, water reservoir, time series analysis, chemico physical parameters, deep learning, remote sensing.

Received August 4, 2020; Accepted January 25, 2020  
Revised February 6, 2020

## 1 Introduction

Alton Water reservoir is a pumped storage reservoir lying 6 miles south of Ipswich, Suffolk, UK, is a linear reservoir, with input water pumped into the north-west end and drawoff tower in the main basin at the southeast end.

River water as the Alton reservoir is the most important resource for drinking water, ground water and it also affects the climate of that area. It is filled by pumped abstraction from the River Gipping over winter and used for potable water supply over summer. Inputs are dosed with ferric sulphate and are pumped

into the northwest end of the reservoir while water is removed through the draw off tower in the main basin [4].

The northwest end is shallow with a mean depth inside the bund of 1 m, increasing gradually along the reservoir with a maximum depth at full capacity of 18 m in the main basin, which is artificially mixed using six pneumatic helices to prevent stratification. The mean depth is 9 m [5].

Environmental monitoring is fundamental for reservoir management, and several studies around the world have been performed in order to evaluate the water quality of these ecosystems. There are many different methods of mathematical modelling, which have been developed to forecast physico-chemical and biological parameters of freshwater [6–8].

In these methods, a statistical approach, such as time series modelling, is a conventional method that has been widely used [9–11].

Statistical modelling has many advantages over mathematical models. But the shortcomings of the statistical approach include handling nonlinear characteristics of data because the statistical models are usually based on the linear correlations of the data can be expressed with a correlation coefficient [12].

To overcome the shortcomings of the statistical methods we propose one models that address the non-linearity, including Wavelet Artificial Neural Network (WANN) and An adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS). Routine sampling was used at sites along the reservoir to determine spatial and temporal patterns in algal biomass, turbidity and water chemistry.

The results shown in some researches using wavelet transform showed more accuracy when compared with the calculated data that do not use wavelets [13–17]. Wavelet analysis is a better tool than the Fourier transform. Wavelets used for time-scale demonstration of the time series and their interaction to resolve time series with non-stationarity [12, 18].

This study presents an alternative method for predicting chlorophyll-a levels, solids suspended and the turbidity of the water reservoir applying Wavelet Neural Network (WNN) and remote sensing techniques, satellite image, to assess water quality, proposing one model for the monitoring of multiple variables related to the ecological systems.

There exists a plethora of applications where WNN's have been successfully applied. These include nonlinear system identification, function learning and time series forecasting [19–21]. Some researchers have applied wavelet based models for analysis of irradiance time series and employ the analysis for prediction [23–25], used discrete wavelet transform to decompose the time series into various scales

and apply feedforward neural network at each scale, as described in [22].

In this study, neuro fuzzy and wavelet coupled model is introduced to predict the monthly Chlorophyll-a (CHL) levels, Algae, Turbidity and Total Suspended Solids (TSS) of Alton Water Reservoir considering water collection from 6 (six) years (2014, 2015, 2016, 2017, 2018 and 2019).

The study has an importance as the quality of water impacts many people that use water from Alton reservoir, people depend on river water for their daily usage. This paper reviews historical data for the site between 2014 and 2019, and discusses the combined artificial intelligence to predict water quality parameters.

The rest of this paper is organized as follows. Section II shows the study area and sampling stations for water quality monitoring. Section III discusses the wavelet neural network modeling approach, provides an introduction to artificial neural networks. The training and validation of wavelet neural network for time series forecasting are described in section IV and section V. Finally, conclusions are given in section VI, which is followed by the references.



Figure 1: Alton Water Reservoir Map, Suffolk, UK.

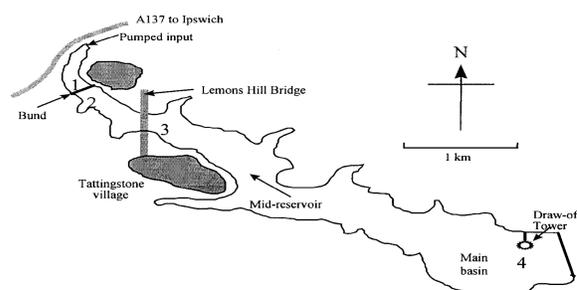


Figure 2: Alton Water reservoir, Suffolk, UK, showing sample sites used in this study, inside the bund (site 1), outside bund (site 2), open water site (site 3) main basin (site 4)

## 2 Remote Sensing Data Processing

The present paper aims to analyze the acquired between February 2014 and December 2019 by the Sentinel-2 and Landsat 8. The Sentinel-2 mission is a land monitoring constellation of two satellites that provide high resolution optical imagery and provide continuity for the current SPOT and Landsat missions. The mission provides a global coverage of the Earth's land surface every 10 days with one satellite and 5 days with 2 satellites, making the data of great use in on-going studies [3].

Of the selected scenes, only the data of those presenting clear skies and cloud-free conditions over sample sites were considered. The free availability of data from the United States Geological Survey (USGS) (<http://earthexplorer.usgs.gov/>), as well as the temporal resolution of the sensor, provide certain advantages over other sensors. Orthorectified and terrain corrected MSI imagery was obtained. Imagery of the USGS website is provided in GeoTIFF.

The fieldwork data were collected the same days before and/or after the overpass of the sensor for all scenes. The main characteristics of the Sentinel - 2 sensor are summarized in Tab. 1.

Processing of the Sentinel-2 images was undertaken using the pixel extraction and map layouts using the latest version of SeaDAS (version 7.4, NASA, Washington, DC, USA) available on <https://seadas.gsfc.nasa.gov/> and ArcMap (version 10.5.1, ESRI, Redlands, CA, USA), respectively. Image data for optical band 8, band 4, band 3 and band 2 were extracted for further processing.

Then the data were radiometrically calibrated from relative to absolute radiance values using a linear model. The imagery was then geometrically corrected and atmospheric correction was applied as described in [17].

The atmospheric correction of satellite measurements in aquatic ecosystems is very important for the reason that a large part of radiation detected by the sensor is backscatter from the atmosphere. Thus, to properly identify the pixel content in an image in terms of water quality, the atmospheric correction presents a critical step and very important in data processing of satellite images.

The Sentinel-2 satellites have a Multi-Spectral Instrument (MSI) with 13 spectral bands that range from the visible range to the shortwave infrared (SWIR). Bands come in variable resolutions from 10 to 60 meter and their wavelength is determined based on specific purposes. The spatial resolution of Sentinel-2 is dependent on the particular spectral band. The list of band with their central wavelengths and resolutions are shown below in the Table 1:

Table 1: Wavelengths and Bandwidths of the 3 Spatial Resolutions of the MSI instruments for Sentinel. S.R.:Spatial Resolution; C. W.:Central Wavelength

S.R. (m)	Bands	C.W. (nm)
	B02 -Blue	490
	B03 - Green	560
	B04 - Red	665
	B08 - NIR	842
	B05-Vegetation Red Edge	705
	B06-Vegetation Red Edge	740
	B07-Vegetation Red Edge	783
	B08A-Narrow NIR	865
	B11-SWIR	1610
	B12-SWIR	2190
	B01-Coastal aerosol	443
	B09-Water vapour	940
	B10-SWIR - Cirrus	1375

The landsat-8 satellite was also used in this research, this satellite was launched in 2013, and its Operational Land Imager (OLI) provides high quality multispectral images at the resolution of 30 metres (15 for panchromatic) and a revisiting time of 16 days [1]. It aims to provide data continuity to the Landsat Earth observation programme, started in the 1970s.

The primary characteristics of the OLI instrument, from landsat-8, relevant to this research are presented in table 2, reporting characteristics of the analyzed bands: band 2, band 3 and band 4. Throughout this document, the spectral bands will be referred to by their band number, name or center wavelength. The list of band with their central wavelengths are shown below in the Table 2:

Table 2: Landsat-8 Operational Land Imager (OLI).

Landsat-8			
Band	Band	Wavelength(nm)	Bandwidth(nm)
2	Blue	482	60
3	Green	561	57
4	Red	655	38

## 3 Inference Models

Work on artificial neural network has been motivated right from its inception by the recognition that the human brain computes in an entirely different way from the conventional digital computer. The brain is a highly complex, nonlinear and parallel information processing system. It has the capability to organize its

structural constituents, known as neurons, so as to perform certain computations many times faster than the fastest digital computer in existence today. The brain routinely accomplishes perceptual recognition tasks, e.g. recognizing a familiar face embedded in an unfamiliar scene, in approximately 100-200 ms, whereas tasks of much lesser complexity may take days on a conventional computer.

A neural network is a machine that is designed to model the way in which the brain performs a particular task. The network is implemented by using electronic components or is simulated in software on a digital computer. A neural network is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experimental knowledge and making it available for use.

Neural networks and fuzzy logic are very suited to solving a prediction problem dealing with real-world sequences. In this paper, three models of this kind are proposed for the study of the considered physico-chemical parameters and quality assessment of Alton river water, UK.

They are introduced in the following, considering as a baseline for benchmarking in the successive tests: the linear predictor reported in Equation, where  $S(t)$  can be predicted by considering a function approximation  $Y = f(\underline{x})$ ,  $f : \mathbb{R}^N \rightarrow \mathbb{R}$ , by using linear models each input vector  $\underline{x}_t$  is made of  $N$  subsequent samples of  $S(t)$ , and the output  $y_t$  is the sample to be predicted.

In this sense, CHL, Algae, Turbidity and TSS were estimated through a common least-squares estimator (LSE); the well-known ARIMA model, which handles non-stationarity and seasonality and whose parameters are determined by a maximum likelihood approach. We also use RBF, ANFIS and Wavelets Neural Networks to predict these parameters and compare the best results presented by the analyzed estimators.

All of these models used lagged samples at the input, as determined by the embedding procedure, similarly to the adopted neural networks.

The collected values from water sampling station, previously evaluated in the laboratory, were taken from the Alton Reservoir database and compared to the proposed wavelet transformation of remote sensing images and analysis by an ANN method, described below:

### 3.1 Wavelet Neural Networks

The neural network represents a nonlinear function,  $g(x)$ , which is further defined as a composition of other nonlinear functions  $f_i(x)$ . This is conveniently

represented as a network structure, with arrows depicting the dependencies between variables [22], see Figure “1”. Such an architecture composition is the nonlinear weighted sum, where:

$$y = g(x) = \left( \sum_i w_j f_i \left( \sum w_{ji} x_j \right) \right), \quad (1)$$

Here  $x$  is the input vector comprising of elements  $x = 1, 2, 3, \dots, n$ , or  $x = [x_1, x_2, \dots, x_n]$   $f$  is the nonlinear function commonly referred to as the activation function.

$W_{ij}$  represent the element of the weight matrix connecting input layer to the hidden layer, from  $j^{th}$  neuron in the input layer to  $i^{th}$  neuron in hidden layer. Similarly  $w_i$  is the element of the weight vector connecting hidden layer to the output layer.  $g(x)$  is the overall nonlinear function corresponding to ANN. In this case, we obtain a single scalar output  $y = g(x)$ . A linear sum of individual elements of input vector (in the input layer) is fed to the hidden layer nonlinear functions. Then the output of each activation function is linearly combined to provide the final network output [22].

In the Wavelet neural network (WNN), a three layer ANN with  $f$  as the activation function of the hidden neurons and with a linear neuron in the output layer can be constructed and it can be referred as wavelet neural network (WNN).

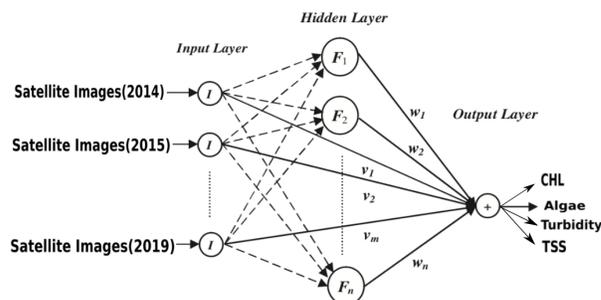


Figure 3: Proposed Architecture: A three layer Wavelet Artificial Neural Network

In this sense,  $[x_1, x_2, x_3, \dots, x_n]$  represents satellite images pixels, imagens were collected in 2014, 2015, 2016, 2017, 2018 and 2019, and the water samples were collected in the same period and analyzed in the laboratory by specialists.

In order to predict CHL, Algae, Turbidity and TSS the neural networks were trained using water collections and satellite imagery. All of the experiments had been performed using MATLAB\_R2018a, running on a 2.9-GHz Intel Core i7 platform equipped with 16 GB of memory.

In addition to the three proposed neural and fuzzy neural models, the linear (LSE) predictor was adopted for benchmarking. LSE did not have parameters to be set in advance, while for RBF and ANFIS, we had adopted the default options provided by the software platform for training and model regularization (RBF and ANFIS models were trained by using the supported functions in the econometrics toolbox, neural network toolbox and fuzzy logic toolbox of MATLAB, respectively). The data set contains information about surface water sample from Alton Water reservoir, Suffolk, UK.

### 3.2 ANFIS

This approach is based on wavelet neural networks and fuzzy logic, treating the inference system as a function approximation problem. An ANFIS neural network implements a fuzzy inference system to approximate the function  $y = f(\underline{x})$ ,  $f : \mathbb{R}^N \rightarrow \mathbb{R}$ . It is composed of  $M$  rules of Sugeno first-order type, where the  $k$ th rule,  $k = 1 \dots M$ , is:

If  $x_1$  is  $B_1^{(k)}$ , ..., and  $x_N$  is  $B_N^{(k)}$  then

$$y^{(k)} = \sum_{j=1}^N a_j^{(k)} x_j + a_0^{(k)}, \quad (2)$$

where  $\underline{x} = [x_1 \ x_2 \ \dots \ x_N]$  is the input to the network and  $y^{(k)}$  is the output of the rule. The antecedent part of the rule depends on the membership functions (MFs)  $\mu_{B_j^{(k)}}(x_j)$  of the fuzzy input variables  $B_j^{(k)}$ ,  $j = 1 \dots N$ , the consequent part is determined by the coefficients  $a_j^{(k)}$ ,  $j = 0 \dots N$ , of the crisp output  $y^{(k)}$ . By using standard options for composing the input MFs and combining the rule outputs [26], the output of the ANFIS network is represented by:

$$\tilde{y} = \frac{\sum_{k=1}^M \mu_{\underline{B}^{(k)}}(\underline{x}) y^{(k)}}{\sum_{k=1}^M \mu_{\underline{B}^{(k)}}(\underline{x})}, \quad (3)$$

where  $\tilde{y}$  is the estimation of  $y$  and  $\mu_{\underline{B}^{(k)}}(\underline{x})$  is the composed MF of the  $k$ th rule.

Several clustering algorithms, followed by a suited classification procedure, can be applied to associate each ANFIS rule with the right input pattern. Using a single sample output we used ANFIS for estimation of the physico-chemical parameters of water lake.

### 3.3 RBF

An RBF neural network is used to build up a function approximation model having the following structure:

$$lf(\underline{x}) = \sum_{i=1}^M \lambda_i \phi(\|\underline{x} - \underline{c}_i\|),$$

where  $\underline{x} \in \mathbb{R}^N$  is the input vector,  $\phi(\cdot)$  is a radial basis function centered in  $\underline{c}_i$  and weighted by an appropriate coefficient  $\lambda_i$ . The choice of  $\phi(\cdot)$  and  $\underline{c}_i$  must be considered for the ability of the network in its approximation capability. Commonly used types of radial basis functions include:

- *Gaussian*

$$l\phi(r) = e^{-(\epsilon r)^2};$$

- *Multiquadric*

$$l\phi(r) = \sqrt{1 + (\epsilon r)^2};$$

- *Inverse Quadratic*

$$l\phi(r) = \frac{1}{1 + (\epsilon r)^2}.$$

Several methods can be used to minimize the error between desired output and model output and hence, to identify the parameters  $\underline{c}_i$  and  $\lambda_i$ . The NN built in this way can be used to estimate a single parameter.

### 3.4 LSE

The basic Least Square Estimator is used to estimate these parameters using values of the decomposed images at each sampling station. The three wavelet output can be viewed as a sequence  $S$  of values. The real value of the parameter in analysis is attached at the end of the sequence, resulting as the last sample of the sequence itself. This is the sample to be estimated. The function estimator has this form:

$$\begin{aligned} l\underline{x}_t &= [S(t) \ S(t-1) \ \dots \ S(t-N+1)], \\ y_t &= S(t+m), \\ f_{lin}(\underline{x}_t) &= \sum_{j=1}^N \lambda_j x_{tj}, \end{aligned} \quad (4)$$

We obtain:

$$\tilde{S}(t+m) = \sum_{j=1}^N \lambda_j S(t-j+1), \quad (5)$$

considering the statistical properties of  $S(t)$ , as the autocorrelation function, it is possible to determinate the parameters  $\lambda_j, j = 1 \dots N$ , of function  $f_{lin}$ .

## 4 Experimental Results

The prediction performances of the considered models were investigated by means of several simulation tests carried out using data from Alton Reservoir, Suffolk, UK. Water from the Alton is currently abstracted and treated for use as drinking water by Anglian Water.

Sample sites were positioned inside the bund (site 1, annual mean depth 1 m), outside the bund (site 2, 3 m), in open water (site 3, 4 m) and in the main basin (site 4, 14 m). Sampling was carried out at two weekly intervals from March to October from 2014 to 2019. As previously stated, we propose remote monitoring of the reservoir using Landsat 8 and Sentinel-2 images to predict parameters of the water in four (4) points as can be seen in Figure 4. The following sampling stations 1, 2, 3, 4 were selected for analysis for five (5) years, (Figure 4).

Table 3: Test set results

NMSE error			
PARAMETERS	ANFIS	RBF	LSE
Chlorophyll-a	0.00232	0.00332	0.00179
Algae	0.37451	0.34596	0.47188
Turbidity	0.01413	0.08589	0.01404
Total Suspended Solids	0.00216	0.00114	<b>0.00004</b>

Table 4: MSE

MSE error			
PARAMETERS	ANFIS	RBF	LSE
Chlorophyll-a	0.00201	0.00287	0.00155
Algae	0.33906	0.31321	0.42722
Turbidity	0.00891	0.05412	0.00885
Total Suspended Solids	0.01725	0.00355	<b>0.00013</b>

Table 5: MARE

MARE error			
PARAMETERS	ANFIS	RBF	LSE
Chlorophyll-a	0.00504	0.00602	0.00366
Algae	0.04794	0.04121	0.06800
Turbidity	0.01492	0.02594	0.01483
Total Suspended Solids	0.00567	0.00558	<b>0.00078</b>

Table 6: MAPE

MAPE error			
PARAMETERS	ANFIS	RBF	LSE
Chlorophyll-a	14.43727%	33.47396%	3.64851%
Algae	19.22293%	17.69944%	25.73034%
Turbidity	2.34159%	3.15459%	2.50602%

This study proposes a method for predicting Chlorophyll-a, Algae, Turbidity and Total Suspended Solids, case study in Alton reservoir using Neural Networks and remote sensing techniques to assess water quality offering a operative method that can be available for the monitoring of variables related to the ecological systems with precise and less expensive sampling than the methods currently used for analysis in water reservoirs. Waveletes Neural Network demonstrated good results between observed and estimated

Multidisciplinary approach to assess the feasibility for the Sentinel-2 and Landsat 8 sensors to monitor this reservoir. An understanding of the ecological characteristics of reservoirs, including bio-physical and chemical features is important for their water management. Biological studies are important to assess uses of water in reservoirs due to their close relation to the effects of algal blooms. Enhanced phytoplankton growth is a major concern for policy and management particularly when the reservoir is used for recreation, aquaculture or potable supplies.

The images from Landsat 8 satellite (launched, 2013), Sentinel-2A (launched 2015) and Sentinel-2B (launched 2017) were used for this research, offer 10 m to 30 m multi-spectral global coverage, were obtained during the same timeframe, for example, if a water sampling was conducted in March 2014, a satellite image was retrieved in March 2014. Together, these satellite advance the virtual constellation paradigm for mid-resolution land imaging [27].

As stated before, we used a set of satellite images from reservoir, corresponding to the sampling stations. The satellites images were cropped, 32x32 pixels, using Matlab functions.

The digital values of the pixels of the images cut

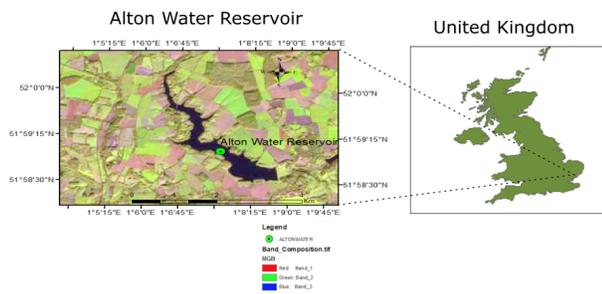


Figure 4: Study Area, Alton Water Reservoir Map, Suffolk, UK.

in the vicinity of the collection stations of water samples were used as input for the ANN. The digital pixel value is an average of radiance values, emittance or backscatter of the different targets that can be contained in the pixel. Thus, the possible differences between the digital values of the images of the different hydrological cycles used in the study were related to the output data of analyzed parameters, forming the input/output pairs for the WANN training. The Figure 5 displays a pixel matrix and its corresponding digital values to exemplify the processing of the proposed method used here in this study.

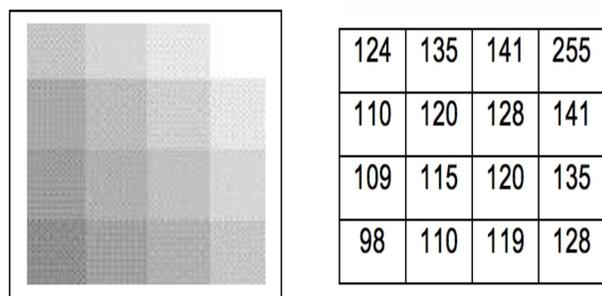


Figure 5: Digital values of the pixels

One sampling station was initially chosen for analysis and an image of the water sampling point 32x32 pixels was cropped, corresponding to an array containing 1024 pixels. Subsequently, the wavelet transform was applied, with only one level of decomposition, resulting in a matrix array of 16x16 pixels for each of the following three components: Horizontal (H), vertical (V) and diagonal (D).

The conversion of the arrays to the H, V and D components to their respective column-matrices was performed, and subsequently a concatenation of the three arrays (each containing 256 pixels) was executed, generating a vector with 768 column size (256 x 3). The images of the water sampling collection points were decomposed by wavelet transform into its three wavelet components, and were used as input to

the WANN. The tests were conducted considering the image representations isolated for each wavelet component, with satisfactory results.

However, when the input data of the three wavelet components was considered, the approximations were even better, which motivated the choice of this arrangement in the proposed solution.

Transformed images by applying wavelet transform to train and test a WANN were used with the three algorithms explained (ANFIS, RBF and LSE). The output parameters of the inference system are physico-chemical values. Performance is compared to the actual, real, values collected from the sampling stations to forecast the values of the successive year, 2019.

Results showed the application of a method to estimate CHL, Algae, Turbidity and TSS using Artificial Neural Network, satellite images, band 2, band 3, band 4 from Sentinel-2 and Landsat 8 which demonstrated good accuracy in the predictions for the years 2015 to 2019 considered here in this study.

Figures 6, 7, 8, 9, 10, 11 displays the validation results for 2019 for several of the water sampling stations, with the laboratory results being the “observed values” and those obtained by wavelet transformation of the remote sensing images and subsequent analysis by WANN, proposed herein, being the “estimated values”, regarding CHL, Algae, Turbidity and TSS.

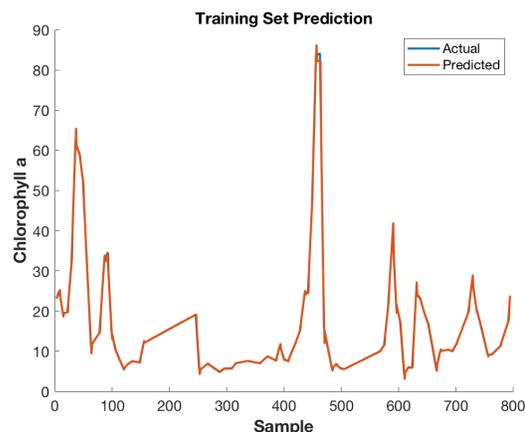


Figure 6: ANFIS: Training Set Prediction for Chlorophyll-a

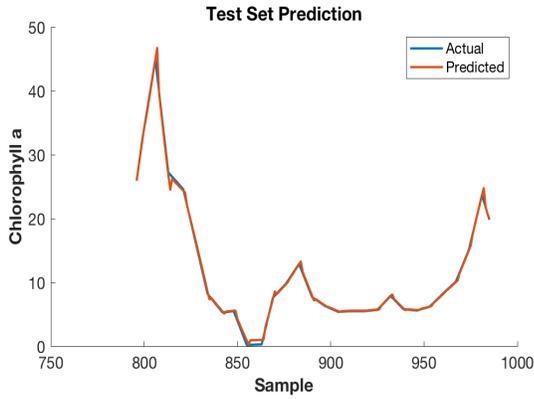


Figure 7: ANFIS: Test Set Prediction for Chlorophyll-a

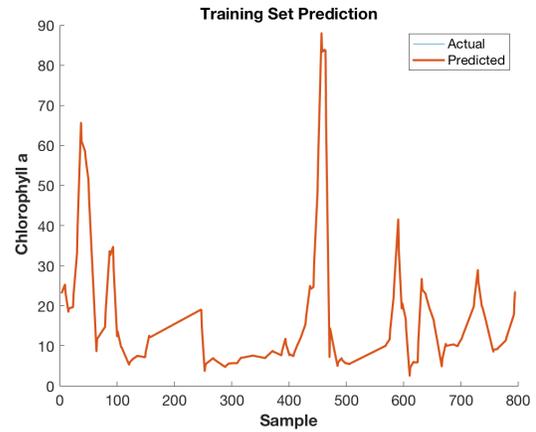


Figure 10: LSE: Training Set Prediction for Chlorophyll-a

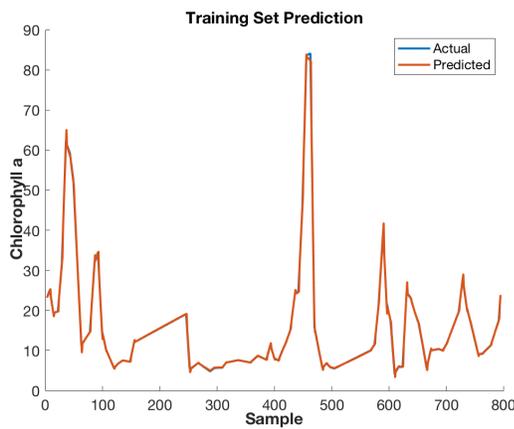


Figure 8: RBF: Training Set Prediction for Chlorophyll-a

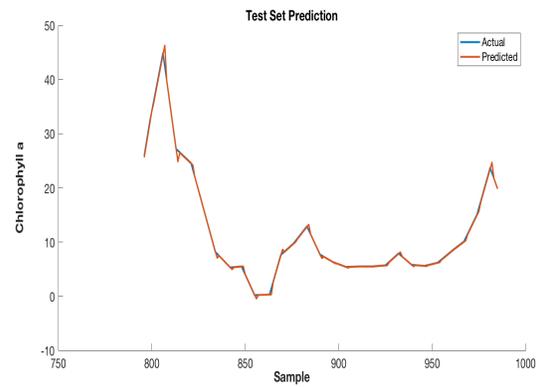


Figure 11: LSE: Test Set Prediction for Chlorophyll-a

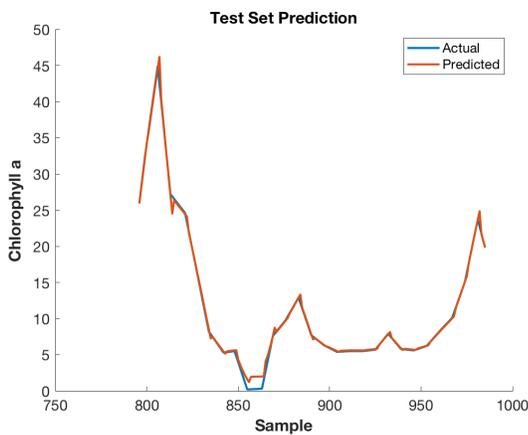


Figure 9: RBF: Test Set Prediction for Chlorophyll-a

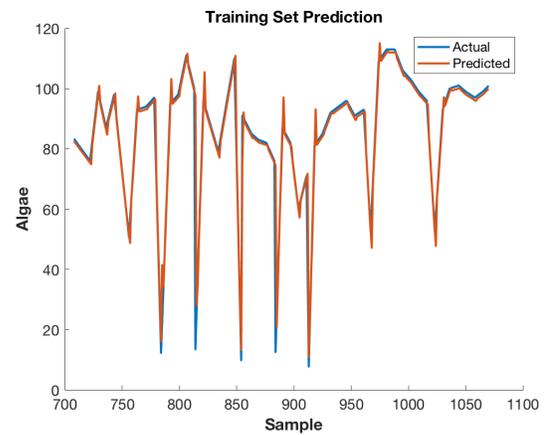


Figure 12: ANFIS: Training Set Prediction for Algae

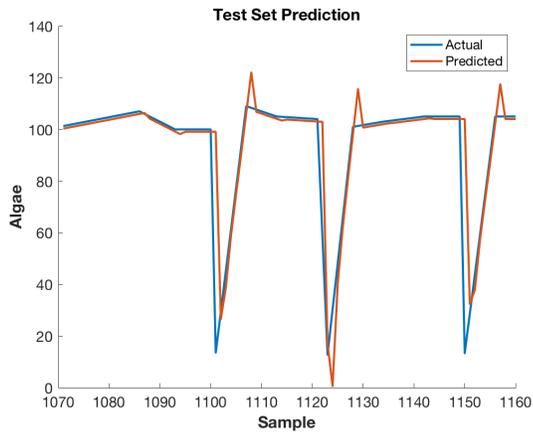


Figure 13: ANFIS: Test Set Prediction for Algae

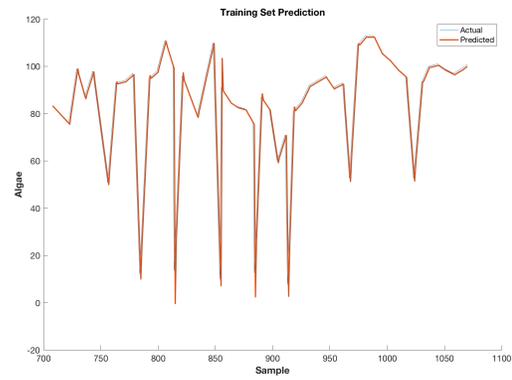


Figure 16: LSE: Training Set Prediction for Algae

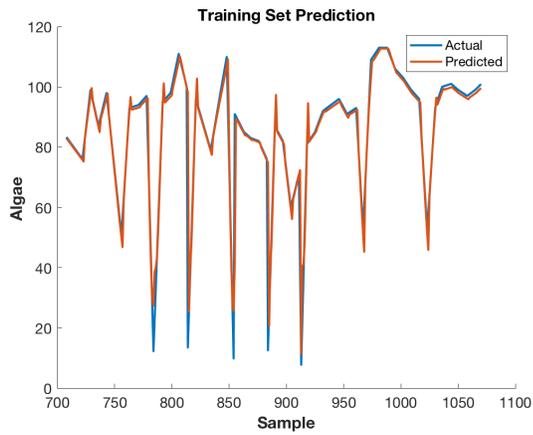


Figure 14: RBF: Training Set Prediction for Algae

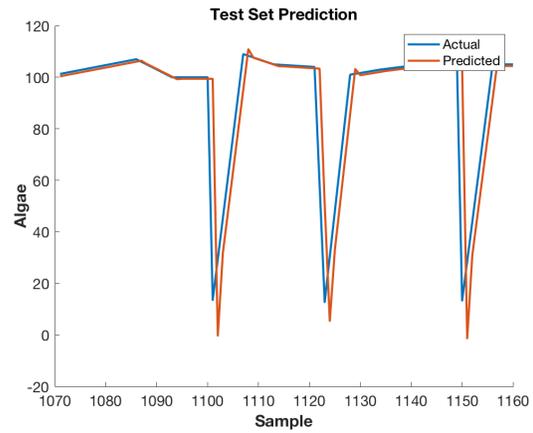


Figure 17: LSE: Test Set Prediction for Algae

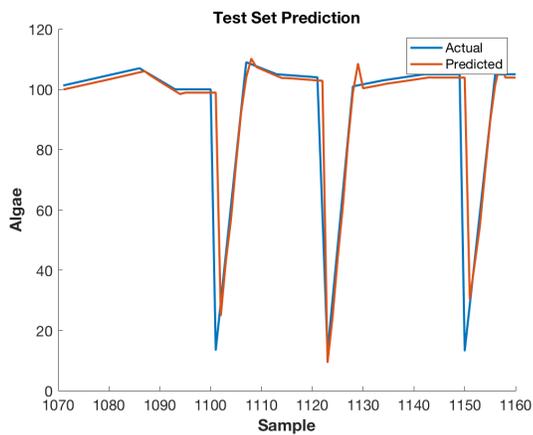


Figure 15: RBF: Test Set Prediction for Algae

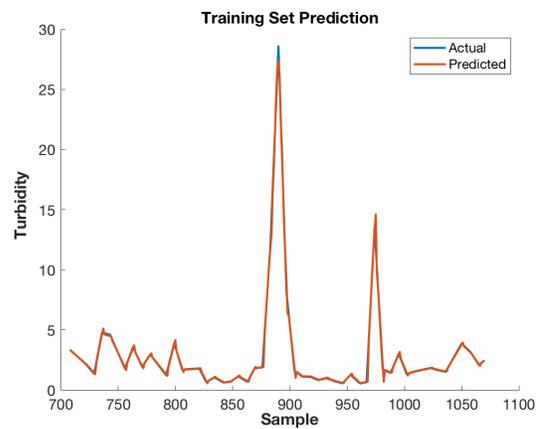


Figure 18: ANFIS: Training Set Prediction for Turbidity

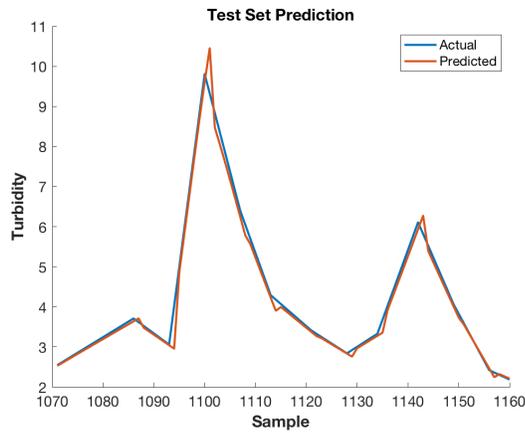


Figure 19: ANFIS: Test Set Prediction for Turbidity

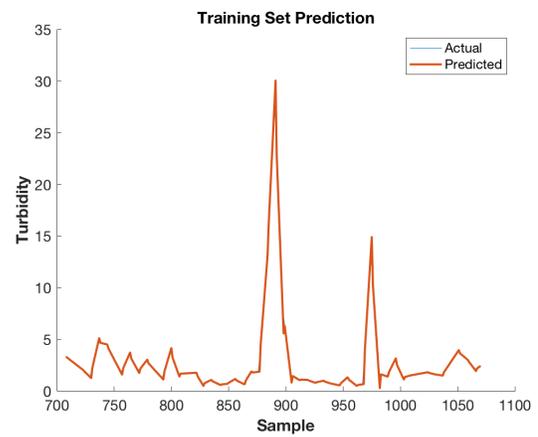


Figure 22: LSE: Training Set Prediction for Turbidity

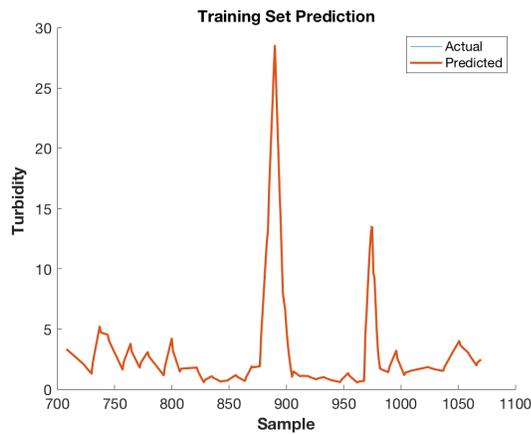


Figure 20: RBF: Training Set Prediction for Turbidity

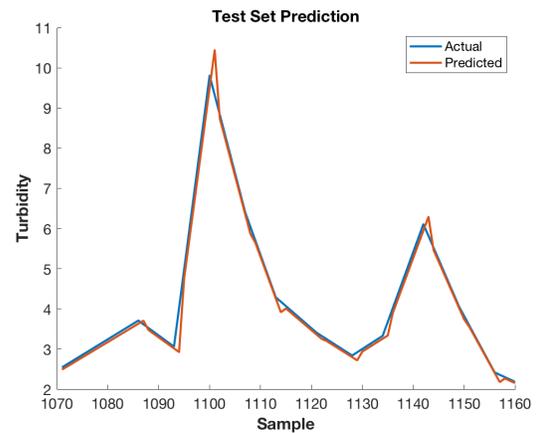


Figure 23: LSE: Test Set Prediction for Turbidity

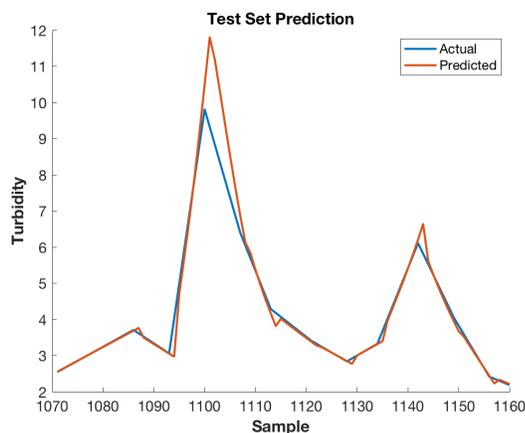


Figure 21: RBF: Test Set Prediction for Turbidity

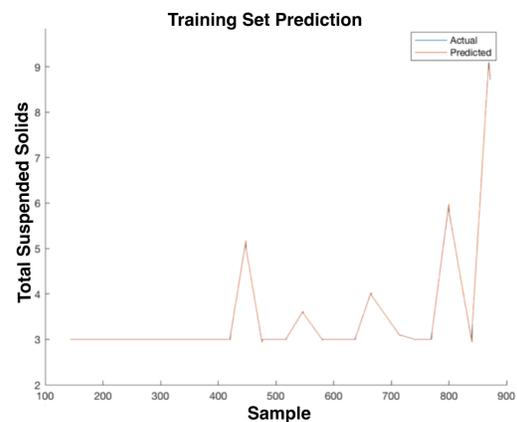


Figure 24: ANFIS: Training Set Prediction for Total Suspended Solids

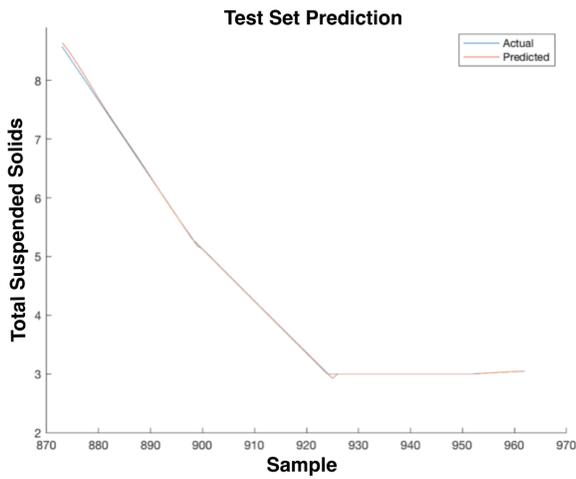


Figure 25: ANFIS: Test Set Prediction for Total Suspended Solids

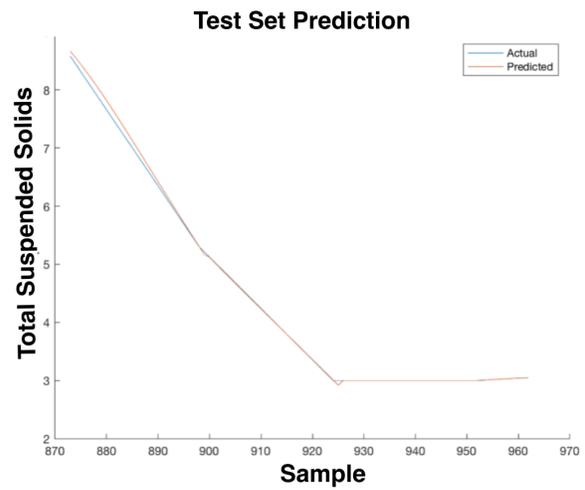


Figure 27: RBF: Test Set Prediction for Total Suspended Solids

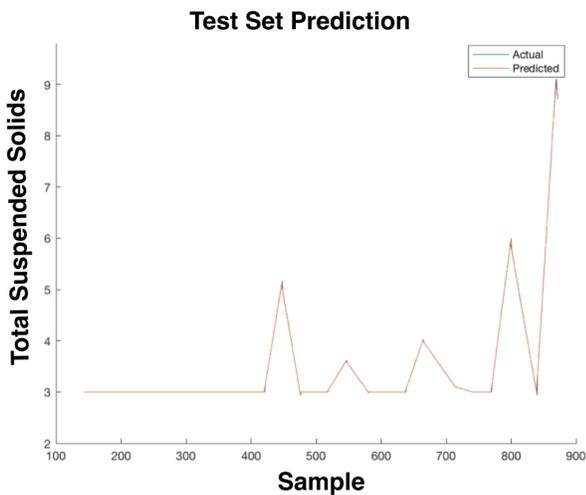


Figure 26: RBF: Training Set Prediction for Total Suspended Solids

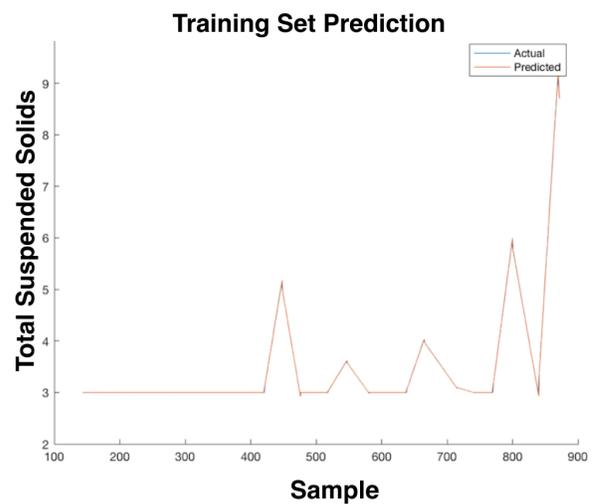


Figure 28: LSE: Training Set Prediction for Total Suspended Solids

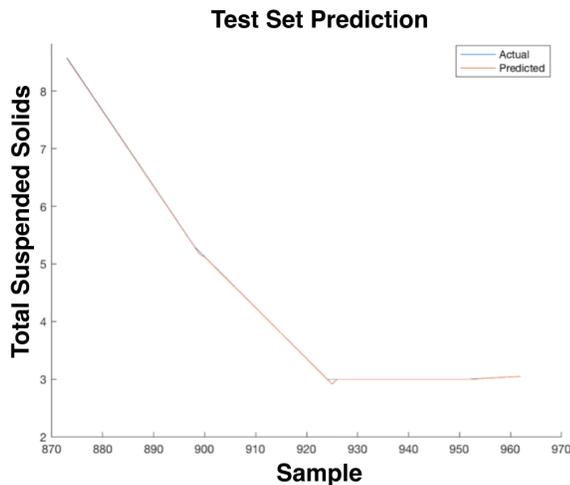


Figure 29: LSE: Test Set Prediction for Total Suspended Solids

## 5 Conclusions

In this paper, approaches based on wavelet neural network and fuzzy neural networks have been properly tailored to be efficiently applied to physico chemical parameters of water time series prediction. The WANN with multi-dimensional wavelet as the activation function of the hidden neurons is applied as the forecasting engine to implement the input/output mapping function of parameters prediction process. A technique is proposed and adapted as the training procedure to optimize the free parameters of the forecasting engine using satellite images from sample station.

Effectiveness of the proposed water quality parameters forecasting strategy as well as the effectiveness of its main components including the suggested WNN and training procedure is extensively evaluated by real-world data, water samples were collected periodically at the site. The models were trained using different training sets, datasets from 2014 to 2019.

The results obtained demonstrate that cloudless satellite images are adequate for the proposed objective, resulting in more reliable performances for neural network training. Wavelet Artificial Neural Networks demonstrated good results between observed and estimated,  $NMSE = 0.00004$ ,  $MSE = 0.00013$  when there are few clouds in the region, providing more efficient analysis of satellite imagery.

In particular, the approaches were very accurate when considering a 5-years training set to predict the successive year. The results showed that employment satellite images to estimate these parameters can be in a important tool for the systematic monitoring of water quality in reservoirs.

**Acknowledgements:** The research was supported by the Federal University of Pará, Machine Learning Research Group, Federal Rural University of Amazonia and Universidad Nacional de Chimborazo, it was also supported by the Anglian Water, water industry company in England.

### References:

- [1] Roy, David P and Wulder, Michael A and Loveland, Thomas R and Woodcock, CE and Allen, Richard G and Anderson, Martha C and Helder, Dennis and Irons, James R and Johnson, David M and Kennedy, Robert and others. Landsat-8: Science and product vision for terrestrial global change research. *Remote sensing of Environment* **2014**, *145*, 154–172.
- [2] Silva, HA Nascimento and Laneve, G and Rosato, A and Panella, M. Retrieving Chlorophyll-a levels, transparency and TSS concentration from multispectral satellite data by using artificial neural networks. *2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL)* **2017**, 2876–2883.
- [3] Zhu, Zhe and Wang, Shixiong and Woodcock, Curtis E. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sensing of Environment* **2005**, *159*, 269–277.
- [4] Perkins, RG and Underwood, GJC. The potential for phosphorus release across the sediment–water interface in an eutrophic reservoir dosed with ferric sulphate. *Water Research - Elsevier* **2001**, *35*, 1399–1406.
- [5] Perkins, RG and Underwood, GJC. Gradients of chlorophyll a and water chemistry along an eutrophic reservoir with determination of the limiting nutrient by in situ nutrient addition. *Water Research - Elsevier* **2000**, *34*, 3, 713–724.
- [6] Ahmed, AA Masrur. Prediction of dissolved oxygen in Surma River by biochemical oxygen demand and chemical oxygen demand using the artificial neural networks (ANNs). *Journal of King Saud University-Engineering - Sciences - Elsevier* **2017**, *29*, 2, 151–158.
- [7] Chen, Wei-Bo and Liu, Wen-Cheng. Water quality modeling in reservoirs using multivariate linear regression and two neural network models *Advances in Artificial Neural Systems - Hindawi Publishing Corp.* **2015**,6
- [8] Hameed, Mohammed and Sharqi, Saadi Shar-tooh and Yaseen, Zaher Mundher and Afan,

Haitham Abdulmohsin and Hussain, Aini and Elshafie, Ahmed. Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Computing and Applications - Springer* **2017**, 28, 1, 893–905.

- [9] Diodato, Nazzareno and Guerriero, Luigi and Fiorillo, Francesco and Esposito, Libera and Revellino, Paola and Grelle, Gerardo and Guadagno, Francesco Maria. Predicting monthly spring discharges using a simple statistical model. *Water resources management - Springer* **2014**, 28, 4, 969–978.
- [10] Sachindra, DA and Huang, Fuchun and Barton, Andrew and Perera, BJC. Least square support vector and multi-linear regression for statistically downscaling general circulation model outputs to catchment streamflows. *International Journal of Climatology - Wiley Online Library* **2013**, 33, 5, 1087–1106.
- [11] Zhang, Qiang and Xu, Chong-Yu and Chen, Xiaohong and Zhang, Zengxin. Statistical behaviours of precipitation regimes in China and their links with atmospheric circulation 1960–2005. *International Journal of Climatology - Wiley Online Library* **2011**, 31, 11, 1665–1678.
- [12] Parmar, Kulwinder Singh and Bhardwaj, Rashmi. River water prediction modeling using neural networks, fuzzy and wavelet coupled model. *Water resources management - Springer* **2015**, 29, 1, 17–33.
- [13] Parmar, Kulwinder Singh and Bhardwaj, R. Water quality index and fractal dimension analysis of water parameters. *International Journal of Environmental Science and Technology - Springer* **2013**, 10, 1, 151–164.
- [14] Parmar, KS and Chugh, Priyanka and Minhas, Preeti and Bhardwaj, Rasmi and Sahota, HS. Seasonal variation of physico-chemical parameters and water quality indexing of Harike Lake. *Indian J. Environ. Protect* **2011**, 31, 482–6.
- [15] Chang, Fi-John and Chang, Ya-Ting. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in water resources - Elsevier* **2006**, 29, 1, 1–10.
- [16] Silva, Hieda Adriana Nascimento and Rosato, Antonello and Altilio, Rosa and Panella, Massimo. Water Quality Prediction Based on Wavelet Neural Networks and Remote Sensing. *2018 International Joint Conference on Neural Networks (IJCNN) - IEEE* **2018**, 1–6
- [17] Silva, HA Nascimento and Laneve, G and Rosato, A and Panella, M. Retrieving Chlorophyll-a levels, transparency and TSS concentration from multispectral satellite data by using artificial neural networks. *2017 Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL) - IEEE* **2017**, 2876–2883.
- [18] Adamowski, Jan and Chan, Hiu Fung. A wavelet neural network conjunction model for ground-water level forecasting. *Journal of Hydrology - Elsevier* **2011**, 407, 1-4, 28–40.
- [19] Iyengar, S Sitharama and Phoha, Vir V. Foundations of wavelet networks and applications. *Chapman and Hall/CRC - Book* **2002**
- [20] Zhang, Jun and Walter, Gilbert G and Miao, Yubo and Lee, Wan Ngai Wayne. Wavelet neural networks for function learning. *IEEE transactions on Signal Processing - IEEE* **1995**, 43, 6, 1485–1497.
- [21] Billings, Stephen A and Wei, Hua-Liang. A new class of wavelet networks for nonlinear system identification. *IEEE Transactions on neural networks - IEEE* **2005**, 16, 4, 862–874.
- [22] Sharma, Vishal and Yang, Dazhi and Walsh, Wilfred and Reindl, Thomas. Short term solar irradiance forecasting using a mixed wavelet neural network. *Renewable Energy - Elsevier* **2016**, 90, 481–492.
- [23] Cao, Shuanghua and Cao, Jiacong. Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. *Applied Thermal EngineeringElsevier* **2005**, 25, 2-3, 161–172.
- [24] Cao, Shuanghua and Weng, Wenbing and Chen, Jianbo and Liu, Weidong and Yu, Guoqing and Cao, Jiacong. Forecast of solar irradiance using chaos optimization neural networks. *2009 Asia-Pacific Power and Energy Engineering ConferenceBook* **2005**, 1-4.
- [25] Cao, Jiacong and Lin, Xingchun. Application of the diagonal recurrent wavelet neural network to solar irradiation forecast assisted with fuzzy technique. *Engineering Applications of Artificial Intelligence - Elsevier* **2008**, 21, 8, 1255–1263.
- [26] Jang, Jyh-Shing Roger and Sun, Chuen-Tsai and Mizutani, Eiji. Neuro-fuzzy and soft computing—a computational approach to learning and machine intelligence. *IEEE Transactions on automatic control[Book Review]* **1997**, 42, 10, 1482–1484.
- [27] Li, Jian and Roy, David P. A Global Analysis of Sentinel-2A, Sentinel-2B and Landsat-8 Data Revisit Intervals and Implications for Terrestrial Monitoring. *Remote Sensing - Multidisciplinary Digital Publishing Institute[Book Review]* **2017**, 9, 9, 902.