Wavelet Transform and Deep Learning approach to predict physico chemical parameters of water

Hieda Adriana Nascimento Silva¹, Paola G. Vinueza Naranjo², Lena Patricia Souza Rodrigues³, Diovanni Moraes de Araujo ¹ and Yomara Pinheiro Pires ¹

¹ UFPA, Federal University of Pará, Belém, Pará 66075-110, Brazil; hieda@ufpa.br; diovannini@ufpa.br; yomara@ufpa.br
² Universidad Nacional de Chimborazo, Avenida Antonio Jose de Sucre, Riobamba, Ecuador; paolag.vinueza@unach.edu.ec
³ UFRA, Federal Rural University of Amazonia, Estr. Principal da Ufра, 2150 - Curίо Utinga, Belém, Pará, Brazil; patricia.souza@ufra.edu.br
E-mail: hieda@ufpa.br

Abstract. Alton water reservoir located within the stunning countryside of Suffolk in UK has a history of cyanobacterial bloom, that are single-celled organisms that live in fresh, brackish, and marine water. The traditional approaches to monitoring water reservoirs are often limited by the need for data collection which often is time-consuming and expensive. In addition, Chlorophyll-a, algae and turbidity are important variable for the analysis of water quality, that are significant not only for human populations but also essential for plant and animal diversity. The main objective of this study is to predict these chemico physical parameters from 2014 to 2019 using time series analysis, satellite imagery, wavelet transform and deep learning. The satellites images were used to predict these parameters in Alton reservoir, manually selected samples were employed to validate estimated parameters using Wavelet Neural Networks. The results predicted by the neural network show good results, and good approximation to laboratory results, suggesting that the proposed model is suitable for environmental monitoring, contributing to monitor water quality parameters.

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 northwest 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 [1].

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 [2].
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 [3, 4, 5].

In these methods, a statistical approach, such as time series modelling, is a conventional method that has been widely used [6, 7, 8].

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 [9].

To overcome the shortcomings of the statistical methods we propose one models that address the nonlinearity, including an artificial neural network (ANN), fuzzy and wavelet tranform from 2014 to 2019 routine sampling was used at sites along the reservoir to determine spatial and temporal patterns in algal biomass (Chl a concentrations), 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 [10, 11, 12, 13, 14]. 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-stationaritie [15, 9].

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 [16, 17, 18]. Some researchers have applied wavelet based models for analysis of irradiance time series and employ the analysis for prediction [20, 21, 22], used discrete wavelet transform to decompose the time series into various scales and apply feedforward neural network at each scale, as described in [19].

In this study, neuro fuzzy and wavelet coupled model is introduced to predict the monthly Chlorophyll-a, Algae Total, and Turbidity levels of Alton Water Reservoir considering water collection from 5 years.

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.

2. Methodology

2.1. 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 [19], see Figure “1”. Such an architecture composition is the nonlinear weighted sum, where:

\[
y = g(x) = \left( \sum_i w_j f_i \left( \sum w_j x_j \right) \right),
\]  

(1)
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)

Here \( x \) is the input vector comprising of elements \( x = 1, 2, 3, ..., n \), or \( x = [x_1, x_2, ..., x_n] \) \( f \) 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 [19].

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 neural network
In this sense, \([x_1, x_2, x_3, \ldots, x_n]\) represents satellite images pixels, images 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 algae, Chlorophyll-a concentration, and Turbidity 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.

2.2. 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 \(y = f(x), f: \mathbb{R}^N \rightarrow \mathbb{R}\). It is composed of \(M\) rules of Sugeno first-order type, where the \(k\)th rule, \(k = 1 \ldots M\), is:

\[
    \text{If } x_1 \text{ is } B_1^{(k)}, \ldots, \text{ and } x_N \text{ is } B_N^{(k)} \text{ then }
    \]

\[
    y^{(k)} = \sum_{j=1}^{N} a_j^{(k)} x_j + a_0^{(k)},
\]

where \(x = [x_1 x_2 \cdots x_N]\) 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 \(B_j^{(k)}, j = 1 \ldots N\), the consequent part is determined by the coefficients \(a_j^{(k)}, j = 0 \ldots N\), of the crisp output \(y^{(k)}\). By using standard options for composing the input MFs and combining the rule outputs [23], the output of the ANFIS network is represented by:

\[
    \tilde{y} = \frac{\sum_{k=1}^{M} \mu_{B_j^{(k)}}(x) y^{(k)}}{\sum_{k=1}^{M} \mu_{B_j^{(k)}}(x)},
\]

where \(\tilde{y}\) is the estimation of \(y\) and \(\mu_{B_j^{(k)}}(x)\) 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 to estimate Algae values.

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

\[
    lf(x) = \sum_{i=1}^{M} \lambda_i \phi(\|x - \varsigma_i\|),
\]
where \( x \in \mathbb{R}^N \) is the input vector, \( \phi(\cdot) \) is a radial basis function centered in \( c_i \) and weighted by an appropriate coefficient \( \lambda_i \). The choice of \( \phi(\cdot) \) and \( c_i \) 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 \( c_i \) and \( \lambda_i \). The NN built in this way can be used to estimate a single parameter.

### 3. Experimental Results

#### 4. Sampling and Experimental Results

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 LandSat8 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 1. Test set results | \( \text{NMSE error} \) |
|---------------------------|----------------------|
| **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 |

Map of the sample location is shown in the "Figure 1" and "Figure 2".

This study proposes a method for predicting Chlorophyll-a concentrations, Algae Total and Turbidity, 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...
Table 2. MSE

| 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 |

Table 3. MARE

| 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 |

Table 4. MAPE

| 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% |

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 [24].

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 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.

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 image of the water sampling collection point, decomposed via wavelet into its three wavelet components, was used as the ANN input. 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 Algae values. Performance is compared to the actual Algae values taken from the sampling stations to forecast the values of the successive year, 2019.

Results showed the application of a method to estimate Chlorophyll a concentration, algae total, and turbidity using Artificial Neural Network, satellite imagens, band 2, band 3, band 4 from Sentinel and Landsat 8 which demonstrated good acurrace 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 samplings 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 chlorophyll a, algae and turbidity.
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.
Figure 16. LSE: Training Set Prediction for Algae

Figure 17. LSE: Test Set Prediction for Algae

Figure 18. ANFIS: Training Set Prediction for Turbidity

Figure 19. ANFIS: Test Set Prediction for Turbidity

Figure 20. RBF: Training Set Prediction for Turbidity

Figure 21. RBF: Test Set Prediction for Turbidity

prediction. A WNN 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, $NMSE = 0.00179$, $MSE = 0.00155$ 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.

References

[1] 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.
[2] 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.
[3] 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.
[4] 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
[5] Hameed, Mohammed and Sharqi, Saadi Shartooh 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.
[6] 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.
[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] Silva, Hieda Adriana Nascimento and Rosato, Antonello and Altílio, 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

[14] 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.

[15] Adamowski, Jan and Chan, Hiu Fung. A wavelet neural network conjunction model for groundwater level forecasting. *Journal of Hydrology* - Elsevier 2011, 407, 1-4, 28–40.

[16] Iyengar, S Sitharama and Phoha, Vir V. Foundations of wavelet networks and applications. *Chapman and Hall/CRC - Book 2002*

[17] 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.

[18] 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.

[19] 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.

[20] Cao, Shuanghua and Cao, Jiacong. Forecast of solar irradiance using recurrent neural networks combined with wavelet analysis. *Applied Thermal Engineering* Elsevier 2005, 25, 2-3, 161–172.

[21] 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.

[22] 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 - Elsevier2008*, 21, 8, 1255–1263.

[23] 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.

[24] 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.