Modelling Colored Dissolved Organic Matter (CDOM) using Neuro Fuzzy Technique: a Comparative Study.

HEDDAM Salim (1), LADLANI Ibtissem (2), HOUICHI Larbi (3), DJEMILI Lakhdar (4)

(1)Maître de conférences (MCB), Faculté des Science, Département des Sciences Agronomiques, Université 20 Août 1955, SKIKDA, *E-mail: heddamsalim@yahoo.fr*.
(2)Doctorant(e) en Sciences, Université Badji Mokhtar Annaba, Faculté des Sciences de l'Ingénieur, Département d'Hydraulique. *E-mail:* ladlaniibtissem@yahoo.fr
(3)Maître de conférences (MCA). Institut d'Architecture de Génie Civil et d'Hydraulique, Université Hadj Lakhdar - Batna. *E-mail:* houichilarbi@yahoo.fr
(4)Maître de conférences (MCA), Université Badji Mokhtar Annaba, Faculté des Sciences de l'Ingénieur, Département d'Hydraulique. *E-mail:* l_djemili@hotmail.com

Abstract— Colored dissolved organic matter (CDOM) is part of the dissolved organic matter (DOM), which can be mainly divided into two groups-natural organic matter (NOM) and anthropogenic organic matter. With two other components, chlorophyll and non-algal particles (NAP), CDOM plays an important role in determining photochemical characteristics of water in nature. The prediction of colored dissolved organic matter (CDOM) using artificial intelligence techniques (AI) has received little attention in the past few decades. In this study, colored dissolved organic matter (CDOM) was modelled using Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and multiple linear regression (MLR) models, as a function of Water temperature (TE), pH, specific conductance (SC) and turbidity (TU). Evaluation of the prediction accuracy of the models is based on the root mean square error (RMSE), mean absolute error (MAE), coefficient of correlation (CC) and Willmott's index of agreement (d).The results indicated that ANFIS can be applied successfully for prediction of colored dissolved organic matter (CDOM). In both models, 60 % of the data set was randomly assigned to the training set, 20 % to the validation set, and 20 % to the test set. The system proposed in this paper shows a novelty approach with regard to the usage of ANFIS models for colored dissolved organic matter (CDOM) concentration modelling.

Key-Words— Colored Dissolved Organic Matter, CDOM, ANFIS, MLR, modelling

I. INTRODUCTION

Dissolved organic matter, DOM, in natural waters is one of the largest pools of organic carbon in the biosphere. The fraction absorbing light from 300 to 800 nm, Colored dissolved organic matter (CDOM), historically referred to as Gelbstoff, yellow substances or humic material is the primary absorber of sunlight **[1]**. Colored dissolved organic matter (CDOM) is defined as the light absorbing component of total dissolved organic matter (DOM) that absorbs light in the ultraviolet and visible range of the electromagnetic spectrum **[2]**. Thus, CDOM is a major determinant of the optical properties of natural waters and it directly affects both the availability and spectral quality of light **[1]**. Increased supply of CDOM by rivers will reduce the photic depth in the shelf regions in particular, resulting in continued light limitation even after sea ice retreat **[3]**. Studying the concentration and distribution of CDOM in aquatic ecosystems, particularly the estuarine and coastal regions, will greatly improve the understanding of the dynamics of dissolved organic carbon (DOC), terrestrial-oceanic carbon cycle, and the impact of anthropogenic activities on water quality **[4]**. Knowledge of CDOM distributions and dynamics, the processes controlling CDOM, and its influence on optical properties are limited by the methods currently used for measurement **[5]**. To date, routine methods used for determination of Colored Dissolved Organic Matter (CDOM) in natural waters include rapidly quantifying CDOM via remote sensing based method.

The objectives of this study were to predict river CDOM concentration as a function of water quality variables by using Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and multiple linear regression (MLR) models.

II. METHODS

II.1. Adaptive Neuro-Fuzzy Inference Systems (ANFIS)

Adaptive neuro-fuzzy inference system (ANFIS) which was put forward in **[6]** is a fuzzy inference system structure that combined fuzzy logic with neural networks. It is mainly based on the fuzzy "if-then" rules from the Takagi and Sugeno type **[7]**. It involves a premise part and consequent part. To explain the procedure of the ANFIS simply, we consider two inputs x and y and one output f in the fuzzy inference system. Hence, the rule base will contain two fuzzy "if-then" rules as follows:

$$Rule1 = \text{if } (x \text{ is } A_1) \text{ and } (y \text{ is } B_1) \text{ then } (f_1 = p_1 x + q_1 y + r_1) \quad (01)$$

$$Rule2 = \text{if } (x \text{ is } A_2) \text{and } (y \text{ is } B_2) \text{ then } (f_2 = p_2 x + q_2 y + r_2) \quad (02)$$

Where x and y are the inputs, A_i and B_i are the fuzzy sets, f_i are the outputs within the fuzzy region specified by the fuzzy rule, p_i, q_i and r_i are the design parameters that are determined during the training process.

Layer 1: This layer is composed of a number of computing nodes whose activation functions are fuzzy logic membership functions. In the study, bell-shaped membership functions are employed.

$$«_{A_i}(x) = \frac{1}{1+\left[(x-c_i)/a_i\right]^{2b_i}} \quad (03)$$

Where $\{a_i, b_i, c_i\}$ represents the parameter set.

Layer 2: Every node in this layer is a fixed node, marked by a circle node, labeled∏, which multiplies the incoming signals and outputs the product.

$$O_i^2 = w_i = «_{A_i} \times «_{B_i}, \quad i=1, 2, \quad (04)$$

The output signal w_i denotes the firing strength of a rule.

Layer 3: Every node i in this layer is a fixed node, marked by a circle node, labelled N. The ith node calculates the ratio of the ith rule's firing strength to the sum of all rules firing strengths:

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2} \quad i=1, 2, \quad (05)$$

Layer 4: Every node in this layer is an adjustable node, marked by a square node, with the node function as:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i x + r_i) \quad i=1, 2, \quad (06)$$

Where $\overline{w}_i$ is the output of Layer 3 and ($\{p_i, q_i, r_i\}$) is the parameter set of this node.

Layer 5: The single node in this layer is a fixed node labelled Σ, which computes the overall output as the summation of all incoming signals.

II.2. Multiple Linear Regression Model (MLR)

Multiple linear regression (MLR) is a well-known method of mathematically modeling the relationship between a dependent variable and one or more independent variables. In general, response variable Y may be related to *n regressor* variables. The following model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n \quad (7)$$

Is called a multiple linear regression model with n regressor variables. Where β_0 is a constant and β_i, i = 1… n are regression coefficients.

(2)

III. STUDY AREA AND DATA COLLECTION

To assess the potential of the models we used continuous and discrete data collected from the Connecticut River at Middle Haddam station , operated by the United States Geological Survey (USGS Station No: 01193050, Latitude 41°32'30", Longitude 72°33'13") from January 2012-September 2012. Water temperature (TE deg.C), pH (std.unit), specific conductance (SC uS/cm), turbidity (TU FNU) and Colored Dissolved Organic Matter (CDOM ppb QSE) data were used to create both the input and output variables of the models presented in this study. The USGS maintain long-term data monitoring networks for these data in the study area. 999 records of 4 input variables (TE, pH, SC and TU) and one output variable (CDOM) were used for the models. The Middle Haddam dataset (999 records) was divided into three groups. In both the models, 60% of the data set was randomly assigned as the training set, which corresponds to 601 input–output pairs, 20% was used for validation, which corresponds to 199 input–output pairs, and 20% was used for testing, which corresponds to 199 input–output pairs. Note that both the ANFIS and MLR models employ the same training, validation and test datasets for an appropriate performance comparison. The data sets were mathematically evaluated by calculating the following evaluation criteria: root mean square error (RMSE), mean absolute error (MAE), coefficient of correlation (CC) and Willmott's index of agreement (d).

IV. RESULTS AND DISCUSSIONS

The comparison of the models was based on both graphical plots and statistical indices (Table 1, Figs. 1 and 2). The ANFIS model developed herein was found to yield better agreement with experimental observations for the training, validation, and testing data set compared to data predicted by the multiple linear regression (MLR) model. **Table 1** shows a statistical analysis of the ANFIS and MLR models for training, validation and testing performances. According to **Table 1**, for the ANFIS model in the training phase, the values of CC, RMSE, MAE, and d, are 0.995, 0.348, 0.252, and 0.997, respectively. In addition, in the validation phase, the values of CC, RMSE, MAE, and d, are 0.995, 0.649, 0.539, and 0.993, respectively. Finally, in the testing phase, the values of CC, RMSE, MAE, and d, are 0.990, 0.641, 0.493, and 0.993, respectively. The results of the MLR models for CDOM prediction are presented in **Table 1**. It may be seen from **Table 1,** the RMSE, MAE, CC and d values for all models in the case of multiple linear regression (MLR) were found to be lower than those for the ANFIS models, thereby establishing the superiority of the ANFIS models. The prediction accuracy for the regression models was lower when compared to MLR models for all the three phases. The scatterplots of the observed versus calculated values of the colored dissolved organic matter CDOM for the ANFIS and MLR analyzed herein are shown in **Fig. 1** and **2** for the training, validation and testing phases, respectively. It may be seen from these figures, for the ANFIS model, the correlation coefficients between observed and forecasted values is very high and that scatter points are predominantly scattered closely along the straight line, indicating that the forecasting values closely approximate the observed values. The ANFIS model is consistently superior to the MLR in all phases.

Table 1. Performances of the ANFIS and MLR models in different phases

Model	Training				Validation				Testing			
	CC	RMSE	MAE	d	CC	RMSE	MAE	d	CC	RMSE	MAE	d
ANFIS	0.995	0.348	0.252	0.997	0.995	0.649	0.539	0.993	0.990	0.641	0.493	0.993
MLR	0.771	2.426	2.041	0.863	0.826	2.295	1.818	0.897	0.767	2.488	2.091	0.858

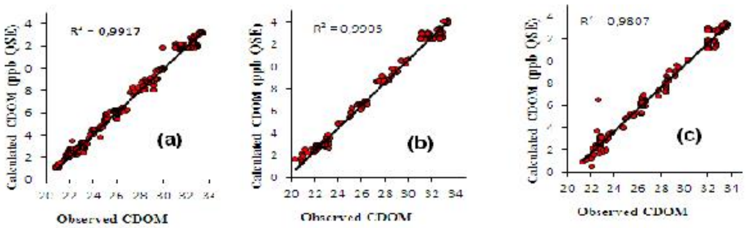

Figure 1. Scatterplots of calculated versus observed values for the ANFIS model for, (a) Training, (b) Validation and (c) Testing.

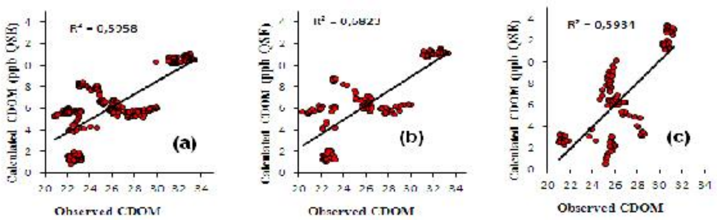

Figure 2. Scatterplots of calculated versus observed values for the MLR model for, (a) Training, (b) Validation and (c) Testing.

V. CONCLUSION

In this paper, we have proposed the Adaptive neuro-fuzzy inference system (ANFIS) for modeling colored dissolved organic matter (CDOM), as a function of Water temperature (TE), pH, specific conductance (SC) and turbidity (TU). The experimental results show that the proposed method outperforms the conventional multiple linear regression (MLR) method. CDOM was predicted with high accuracy of approximately 99%. It has been concluded that ANFIS could be used as a powerful and simple alternative technique for prediction of colored dissolved organic matter (CDOM).

REFERENCES

[1] Kowalczuk P., Stoń-Egiert J., Cooper W.J., Whitehead R.F., Durako M.J. (2005). Characterization of chromophoric dissolved organic (CDOM) in the Baltic Sea by excitation

emission matrix fluorescence spectroscopy. Marine Chemistry 96 273-292.

[2] Singh S., D’Sa E., Swenson E. (2010). Seasonal variability in CDOM absorption and fluorescence properties in the Barataria Basin, Louisiana, USA. Journal of Environmental Sciences, 22(10) 1481-1490.

[3] Stedmon C.A., Amon R.M.W., Rinehart A.J., Walker S.A. (2011).The supply and characteristics of colored dissolved organic matter (CDOM) in the Arctic Ocean: Pan Arctic trends and differences. Marine Chemistry 124 108-118.

[4] Zhu, W., Yu Q. (2011). Inversion of Chromophoric Dissolved Organic Matter (CDOM) from EO-1 Hyperion Imagery for Turbid Estuarine and Coastal Waters. IEEE Transactions on Geoscience and Remote Sensing.

[5] Shanmugam, P. (2011). New models for retrieving and partitioning the colored dissolved organic matter in the global ocean: Implications for remote sensing’’. Remote Sensing of Environment 115 1501-1521.

[6] Jang JSR. (1993) ANFIS: Adaptive network based fuzzy inference system. IEEE Trans.Syst. Man Cybernetics, 23, 665-683.

[7] Takagi T, Sugeno M. (1985). Fuzzy identification of systems and its applications to Modeling and control. IEEE Transaction on Systems, Man, and Cybernetics, 15, 116-132.11.