Monitoring and Modeling of Chlorophyll-a Dynamics in a Eutrophic Lake: M'koa Lake (Jacqueville, Ivory Coast)

Yapo Habib Kpidi[1], Ossey Bernard Yapo[1,3], Mamadou Guy-Richard Koné[2,*], Gabaze André Gadji[1], Agness Essoh Jean Eudes Yves Gnagne[1], Jean Stéphane N'dri[2], Nahossé Ziao[2]

[1]Laboratoire des Sciences de l'Environnement (LSE), (UFR-SGE), Université Nangui Abrogoua, 02 BP 801 Abidjan 02 - Côte-d'Ivoire
[2]Laboratoire de Thermodynamique et Physico-chimie du Milieu (LTPCM), (UFR-SFA), Université Nangui Abrogoua, 02 BP 801 Abidjan 02 - Côte-d'Ivoire
[3]Laboratoire Central de l'Environnement du Centre Ivoirien Anti-pollution (LCE-CIAPOL)
*Corresponding author: guyrichardkone@gmail.com

Abstract   In-situ measurements and physico-chemical analyzes of thirty (30) samples taken bimonthly from August 2015 to December 2016 on six (6) stations of Lake M'koa were carried out. A modeling study was made in order to determine a quantitative and qualitative relationship between *chlorophyll-a* and five physico-chemical descriptors (temperature, turbidity, oxidative power (R_H), nitrate ions (NO_3^-) and nitrite (NO_2^-)). These descriptors constituted the explanatory and predictive parameters of *chlorophyll-a* of samples taken from Lake M'koa. This study was carried out by using Principal Component Analysis (PCA), Ascending Hierarchical Classification (AHC), Multiple Linear Regression (RML) and Nonlinear (RMNL) methods. Two quantitative and qualitative linear and nonlinear models (RML and RMNL) have been proposed. These accredited models as good statistical indicators have been validated according to the rules established by the Organization of Economic Cooperation and Development (OECD). Statistical indicators of RMNL reveal more efficient predictions with R^2 = 0.942, **RMSE** = 0.049 and **F** = 291.986. The obtained results suggest that the combination of these five descriptors could be useful in predicting the property of *chlorophyll-a*. In addition, turbidity is the first most important descriptor for the prediction of *chlorophyll-a* at the M'koa Lake different stations.

Keywords: Eutrophication, chlorophyll-a, Quantitative Structure-Property Relationship (QSPR), physico-chemical descriptors

Cite This Article: Yapo Habib Kpidi, Ossey Bernard Yapo, Mamadou Guy-Richard Koné, Gabaze André Gadji, Agness Essoh Jean Eudes Yves Gnagne, Jean Stéphane N'dri, and Nahossé Ziao, "Monitoring and Modeling of Chlorophyll-a Dynamics in a Eutrophic Lake: M'koa Lake (Jacqueville, Ivory Coast)." *American Journal of Environmental Protection*, vol. 6, no. 1 (2018): 1-9. doi: 10.12691/env-6-1-1.

1. Introduction

The quality of surface water is one of the harmful environmental difficulties facing humanity [1,2] The undermining of these waters of surface quality can be due to agricultural activities, industrial and domestic waste ejected into the receiving environment without any treatment. Significant efforts have been made in recent decades to develop water management strategies in order to ensure good quality of water [3]. The quality of water is also one of the main cause of most public health problems in developing countries [4,5]. The classical approach to assess the trophic state of a given area consists of physico-chemical and biological measurements of a group of environmental variables, which is depending on several constraints. Nowadays, there is a growing interest concerning the development of alternative methods for the monitoring and rational management of water resources. Eutrophication is defined as a process of natural aging of aquatic environments, which is accelerated by human beings activities. These

human beings activities can be industrial effluents and/or domestic wastewater discharged without treatment or partially treated [6] in aquatic systems. The rejections are all riche in nitrogen and phosphorus and will enriched the aquatic area. These nutritive salts considered as trophic indicators [7] are likely to favor the appearance of algae and macrophytes whose growth is often linked to seasonal variations and various environmental factors [8,9]. Many environmental factors such as temperature, pH and dissolved oxygen content must also be considered. Indeed, the solubility of oxygen in water is inversely proportional to its temperature. The temperature (25 to 30°C average in our tropics), the concentration of dissolved oxygen and organic matter are physico-chemical factors that govern a lot of chemical and biological processes taking place in water [10]. The existence of various activities on the watershed can contribute to the degradation of the quality of its aquatic system [11]. To this end, a continuous supervision of its physico-chemical and biological quality, a deep knowledge of its hydro dynamism and different factors of disturbance of the equilibrium of this lacustrine ecosystem must be well-known [12]. Physico-chemical

and biological approaches to eutrophication allowed to quantify the trophic state of a given area by measuring environmental variables and compared them to established norms. This approach seems to be expensive in terms of equipment and reagents. to overcome the considerable logistical, chemical and investments requirement ᵈ for the occasional or continuous sampling in order to the characterize ᵒᶠ the aquatic environment [13], the development of alternative methods for the monitoring and the rational management of resources in waters constituted a real godsend., For the fulfillment of this aim, the use of modeling remains a way of approach. Indeed, dissolved oxygen have been increasingly modeled in the assessment of the health status of an aquatic area by using a statistical approach [14]. It appears necessary to model *chlorophyll-a* because it seems to express the environmental response to ecological disturbances and to circumscribe them. Thus, a Quantitative Structure Property Study (QSPR) is used to determine elaborated models from physico-chemical descriptors. This statistical approach can combine methods such as linear regression, nonlinear regression, and so on.

The general objective of this work is to design statistical models by using multiple regressions (linear and nonlinear) which will be able to predict the chlorophyllian property of M'koa Lake's water from environmental variables. Specifically, this involves identifying pertinent and expressive *chlorophyll-a* explanatory variables in order to develop simulation tools that will be useful for integrated and sustainable management of water resources.

2. Materials and Methods

2.1. Presentation of the Study Area

Situated in the region of large bridges (*Grands Ponts*), 62 km west of Abidjan, Lake M'koa is an endorheic lake without surface outfall, in Jacqueville (Côte d'Ivoire), constitutes the framework of this study (Figure 1). A shallow natural lake with a maximum depth of 3.9 m and an area of about 0.180 km^2, Lake M'koa, in the sub-equatorial Attiean climate, is maintained by infiltration [10].
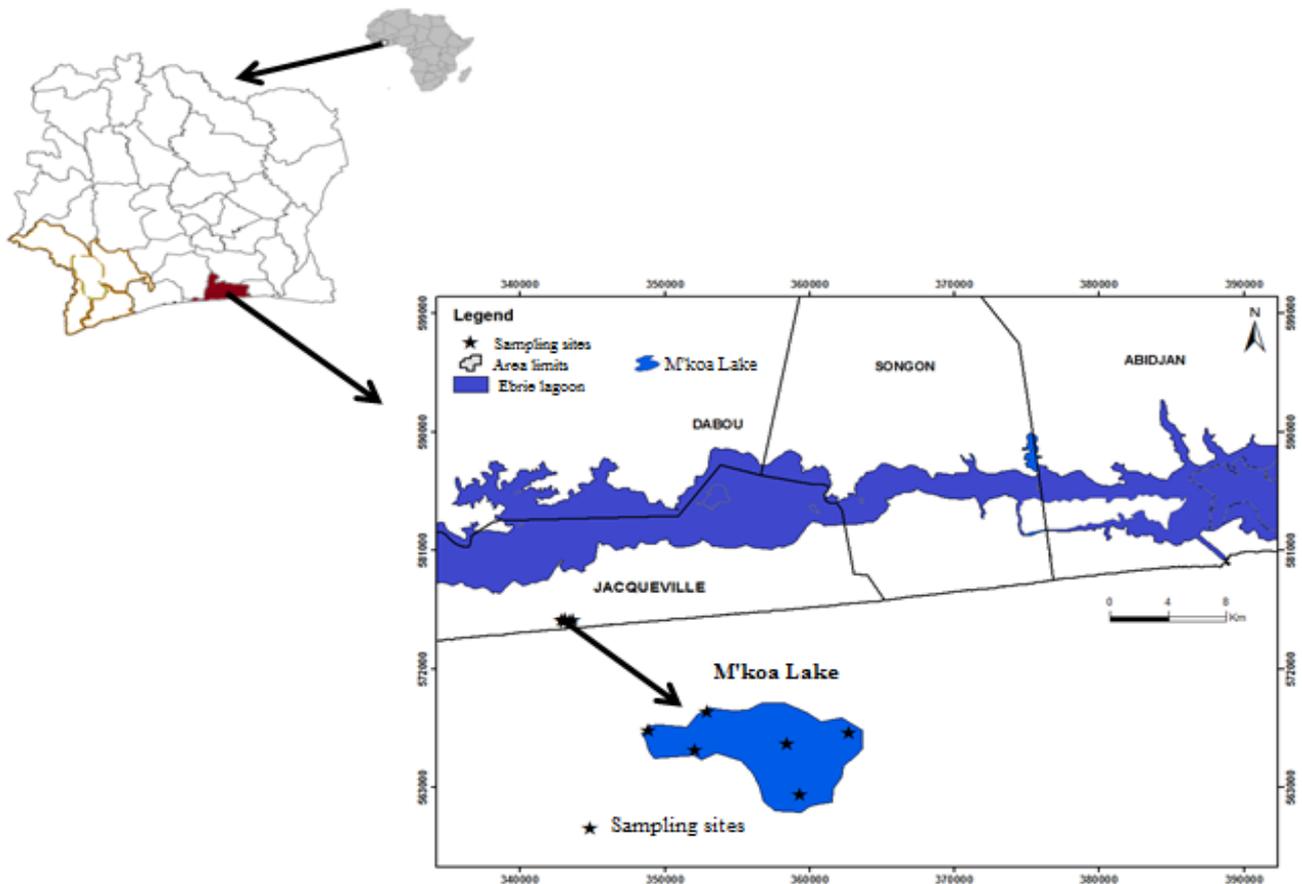

Figure 1. Mapping of the study area

2.2. Physico-Chemical Descriptors

Chlorophyll-a is the main form of chlorophyll present in organisms that carry out photosynthesis. It is the pigment present in almost all plants and captures the light necessary for photosynthesis. This process is very important because it allows plants to transform carbon dioxide (CO_2) into organic matter. Turbidity is a strong indicator for wetland monitoring and planning. Water

quality supervising using modeling is essential for mapping turbidity in water [15]. This indicator is a useful one which allows to predict the state of health of the wetland in order to maintain its ecosystem healthy [16]. Ammonia (NH_4^+), nitrite (NO_2^-) and nitrate (NO_3^-), which are the inorganic nitrogenous forms, are common pollutants in water of surface and groundwater. These inorganic forms are causing enormous health problems on humans and animals [17,18]. The oxidizing power (R_H)

permits to determine the oxidizing or reducing character of the sampled water evaluated as follows:

$$R_H = \frac{E_h}{0.0992 * T} + 2pH \qquad (1)$$

Eh: average redox potential; T: temperature in kelvin degree $(273 + T \,°C)$.

The value of the oxidizing power allows to classify waters in the following categories according to Rejset [19]:

- $R_H \geq 23$, the medium is called oxidizing;
- $15 < RH < 23$, the medium is described as anoxic;
- $R_H < 15$, the medium is said to be reducing.

These different physico-chemical descriptors were determined by *kpidi et al.* [10]. The modeling was done by using the multilinear and nonlinear regression method implemented in Excel spreadsheets [20] and XLSTAT version 2014 [21].

2.3. Estimation of the Predictive Capacity of a Model

Chlorophyll-a from thirty (30) studied samples showed significant variation in concentrations (17.16 to 72.8 μg/L). For the fulfilment of our aim, a log-based ten-fold transformation of the activity was applied to obtain higher mathematical values when the structures are biologically very efficient [22,23]. Biological properties are generally estimated by the opposite of the logarithm. *Chlorophyll-a* expressed by its potential (*pChl-a*) which is defined by equation (2):

$$pChl_a = -\log_{10}(Chl_a). \qquad (2)$$

Thus, this range of reduced variation of concentrations permit to define a better quantitative relationship between the *chlorophyll-a* and the physico-chemical descriptors of the taken samples. The quality of a model is determined by taking care with various statistical analysis indicators including the coefficient of determination R^2, the standard error (RMSE), the correlation coefficients of the cross validation Q_{CV}^2 and Fischer F. R^2, RMSE and F relate to the adjustment of calculated and experimental values. They describe the predictive power within the model limits and allow to express the accuracy of the calculated values on the test set [24,25]. As for the cross-validation coefficient Q_{CV}^2, it gives information on the predictive power of the model, which can be "internal" because it is calculated from the structures used to construct this model. The coefficient of determination R^2 gives an evaluation of the dispersion of the calculated values around the experimental ones. The quality of the modeling is better when the points evaluated by the coefficient R^2 are close to the line of adjustment [26]. R^2 is expressed as follows:

$$R^2 = 1 - \frac{\sum \left(y_{i,\exp} - \hat{y}_{i,theo}\right)^2}{\sum \left(y_{i,\exp} - \overline{y}_{i,\exp}\right)^2} \qquad (3)$$

Or:

$y_{i,exp}$: Experimental value of *chlorophyll-a*,

$\hat{y}_{i,theo}$: Calculated value of *chlorophyll-a* and

$\overline{y}_{i,exp}$: Experimental average value of *chlorophyll-a*

The closer the value of R^2 is to 1, the better the computed and experimental values are correlated. Moreover, the variance σ^2 is determined by the relation (4)

$$\sigma^2 = (RMSE)^2 = \frac{\sum \left(y_{i,\exp} - y_{i,theo}\right)^2}{n - k - 1} \qquad (4)$$

Where k is the number of independent variables (descriptors), n is the observation number of the test or training set and n-k-1 is the degree of freedom.

The standard error or standard deviation **RMSE** is another statistical used indicator. It allows to evaluate the reliability and the accuracy of a model:

$$RMSE = \sqrt{\frac{\sum \left(y_{i,\exp} - y_{i,theo}\right)^2}{n - k - 1}}. \qquad (5)$$

The Fisher **F** test is also used to measure the level of statistical significance of the model, in other words the quality of the choice of descriptors constituting the model.

$$\mathbf{F} = \frac{\sum \left(y_{i,theo} - y_{i,\exp}\right)^2}{\sum \left(y_{i,\exp} - y_{i,theo}\right)^2} * \frac{n - k - 1}{k}. \qquad (6)$$

The coefficient of determination of the cross-validation Q_{CV}^2 which makes it possible to evaluate the accuracy of the prediction on the test set is determined by means of the following relation:

$$Q_{cv}^2 = \frac{\sum \left(y_{i,theo} - \overline{y}_{i,\exp}\right)^2 - \sum \left(y_{i,theo} - y_{i,\exp}\right)^2}{\sum \left(y_{i,theo} - \overline{y}_{i,\exp}\right)^2}. \qquad (7)$$

2.4. Statistical Analyzes

Principal Component Analysis (PCA) is a data analysis tool that can be used to explain the structure of correlations or covariances using linear combinations of original data. Its use permit to interpret the data in a reduced space [27]. It was used to appreciate, certainly, the relations between the different measured variables, above all to access their structuring in order to be able to gather them by zone. The fact of grouping them per zone thus meets the objective of the slighted approach, which is to correlate the classes of *physico-chemical* descriptors obtained during the sampling.

The Ascending Hierarchical Classification (AHC) aims to partition into homogeneous classes a set of individuals (an individual is an observation and, in our case, they are samples) [28]. It organizes individuals, defined by a number of variables and modalities, by grouping them hierarchically on a dendrogram. It aggregates those that are most similar by using measures of dissimilarity or distance between individuals to form classes. It is made from the data of individuals and variables. AHC allowed to establish a typology of the samples according to the temperature, turbidity, R_H, NO_3^- and NO_2^-.

The Multiple Linear Regression (RML) statistical technique is used to study the relationship between a dependent variable (Property) and several independent variables (descriptors). This statistical method minimizes

the differences between the actual and predicted values. It also allowed to select the descriptors used as input parameters in multiple nonlinear regression (RMNL). As for the analysis of nonlinear multiple regression (RMNL), it is also used for betterment of the structure-property relationship in order to evaluate quantitatively the property. It is the most common tool for studying multidimensional data. It is based on the following XLSTAT preprogrammed functions:

$$y = a + (bx_1 + cx_2 + dx_3 + ex_4)$$
$$+ (fx_{12} + gx_{22} + hx_{32} + ix_{42}) \quad (8)$$

Where a, b, c, d, ... represent the parameters and, x_1, x_2, x_3, x_4,... represent the variables. RML and RMNL were generated using the XLSTAT software version 2014 [21], to predict *chlorophyll-a*. The equations of the different models were evaluated by the coefficient, the mean squared error (RMSE), the Fischer test (F) and the cross correlation coefficient (Q^2_{CV}) [29].

2.5. Criterion of Acceptance of a Model

The performance of a mathematical model, for *Eriksson et al.* [30], is characterized by a value of $Q^2_{cv} > 0.5$ for a satisfactory model and for an excellent model, $Q^2_{cv} > 0.9$. According to *Tropsha et al.* [31,32,33], for the external validation set, the predictive power of a model can be obtained from the following five criteria

1) $R^2_{Test} > 0.7$

2) $Q^2_{CvTest} > 0.6$

3) $\left| R^2_{Test} - R^2_0 \right| \leq 0.3$,

4) $\dfrac{\left| R^2_{Test} - R^2_0 \right|}{R^2_{Test}} < 0.1$ and $0.85 \leq k \leq 1.15$,

5) $\dfrac{\left| R^2_{Test} - R'^2_0 \right|}{R^2_{Test}} < 0.1$ and $0.85 \leq k' \leq 1.15$.

Roy and Roy [34] have further refined the predictive skill of a model. They developed quantities such as r^2_m and Δr^2_m, called metric values. r^2_m determines the proximity between the observed activities and the predicted ones. The metric values r^2_m and Δr^2_m are calculated from the observed and predicted activities. Currently, these two different variants r^2_m and Δr^2_m can be calculated for the test set (internal validation) or for the test set (external validation). A model is acceptable for these authors, if both of these criteria are met.

$$\overline{r^2_m} = \frac{\left(r^2_m + r'^2_m \right)}{2} > 0.5, \quad \Delta r^2_m = \left| r^2_m - r'^2_m \right| < 0.2$$

Where: $r^2_m = r^2 * \left(1 - \sqrt{(r^2 - r^2_0)} \right)$ and $r'^2_m = r^2 * \left(1 - \sqrt{(r^2 - r'^2_0)} \right)$.

3. Results and Discussion

The set of descriptor values for twenty (20) samples of the test set and ten (10) other samples of the validation set are presented in Table 1.

Table 1. Experimental physico-chemical and *pChl-a* descriptors of test and validation test

	Temperature (°C)	R_H	Turbidity (NTU)	NO_2^- (mg.L^{-1} N-NO$_2^-$)	NO_3^- (mg.L^{-1} N-NO$_3^-$)	Chl-a (µg/mL)	pChla
Training Set							
1	30.550	12.102	11.385	0.002	0.250	17.160	4.765
2	30.150	12.174	37.700	0.001	0.380	23.660	4.626
3	30.400	12.301	35.700	0.001	0.250	19.760	4.704
4	30.450	12.203	31.500	0.001	0.350	17.160	4.765
5	28.900	13.906	36.700	0.002	0.200	50.700	4.295
6	29.200	14.041	40.600	0.003	0.450	37.750	4.447
7	29.400	13.843	36.000	0.004	0.500	47.670	4.322
8	29.600	13.756	46.200	0.004	0.450	60.230	4.220
9	29.300	14.101	23.200	0.001	0.300	34.780	4.459
10	29.100	14.072	42.100	0.002	0.260	59.800	4.223
11	29.100	14.352	35.300	0.003	0.400	45.180	4.345
12	29.100	14.982	29.700	0.003	0.400	49.730	4.303
13	29.100	14.275	27.500	0.002	0.280	34.130	4.467
14	32.000	14.085	49.900	0.006	0.300	72.800	4.138
15	31.100	16.069	51.100	0.001	0.120	64.350	4.191
16	31.000	14.236	47.100	0.004	0.200	57.530	4.240
17	31.400	14.480	50.000	0.002	0.500	36.400	4.439
18	31.300	13.958	80.900	0.003	0.400	55.900	4.253
19	31.400	14.000	70.500	0.004	0.300	57.200	4.243
20	31.300	14.134	81.400	0.001	0.300	66.300	4.178
Validation Set							
21	31.300	14.370	49.900	0.001	0.200	36.400	4.439
22	29.700	13.950	46.100	0.000	0.400	60.450	4.219
23	29.900	13.810	31.300	0.004	0.400	41.930	4.378
24	31.300	13.958	80.900	0.003	0.400	55.900	4.253
25	31.400	14.000	70.500	0.004	0.300	57.200	4.243
26	31.300	14.134	81.400	0.001	0.300	66.300	4.178
27	28.400	38.098	11.100	0.004	0.200	60.450	4.219
28	30.600	12.288	50.400	0.012	0.800	20.150	4.696
29	29.700	13.776	30.700	0.004	0.500	34.230	4.466
30	29.800	13.652	28.200	0.001	0.500	35.750	4.447

3.1. Typology of the Waters of Lake M'koa

The Pearson (n) correlation matrix (Table 2), the correlation circle (Figure 2), the Cartesian diagrams according to F1 and F2 (Figure 3 and Figure 4) and the dendrograms of the stations are shown below.

The PCA data matrix gathers the average values of five (5) variables representing physico-chemical descriptors, and thirty (30) individuals representing the original samples. The resulting matrix provides information on the negative or positive correlation between the variables. Turbidity is positively correlated with *pChl-a* (r = 0.7094 and p <0.05) at a significant level. The examination of the community circle of Figure 2 associated with the analysis

of the PCA factor structure (Table 1), indicate two (02) main components represented by F1 and F2 which corresponds respectively to 33.33% and 29.36%, all explained variance. These two factors totalizing 62.69% of the total variance are sufficient to interpret the entire PCA data. In fact, high temperature and turbidity are positively correlated with factor 1. Conversely, *pChl-a* is negatively and moderately correlated. Factor 1 reflects a pollution gradient. Concerning factor 2, NO_2^- and NO_3^-, which participate to the synthesis of *chlorophyll-a*, are strongly and positively correlated with the opposite of R_H, which is negatively correlated with factor 2. Factor 2 seems to indicate an increasing gradient of oxidation.

Table 2. Correlation matrix (Pearson (n)) between the different physico-chemical descriptors

Variables	Temperature (°C)	R_H	Turbidity (NTU)	NO_2^- (mg.L^{-1} N-NO$_2^-$)	NO_3^- (mg.L^{-1} N-NO$_3^-$)	*pChla*
Temperature (°C)	**1**					
R_H	-0.3384	**1**				
Turbidity (NTU)	**0.7094**	-0.2692	**1**			
NO_2^- (mg.L^{-1} N-NO$_2^-$)	0.0816	0.0653	0.0541	**1**		
NO_3^- (mg.L^{-1} N-NO$_3^-$)	-0.0728	-0.2710	0.0348	**0.5703**	**1**	
pChla	-0.1127	-0.2977	**-0.4923**	0.0351	0.2932	**1**

Bold values are different from 0 to a significant level for p <0.05. Very significant for p <0.01. Very highly significant for p <0.001.
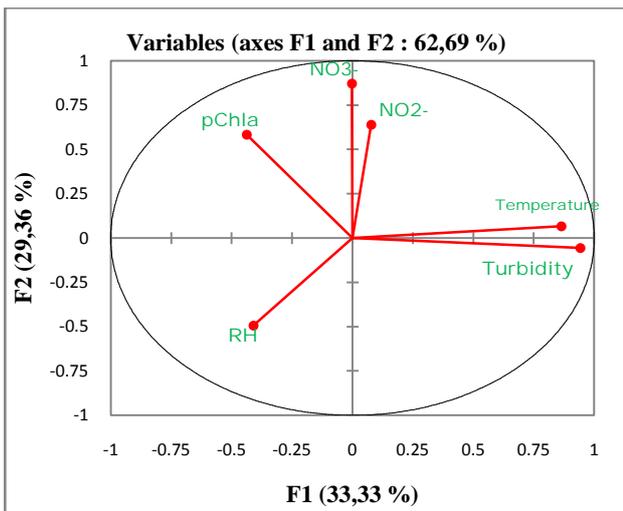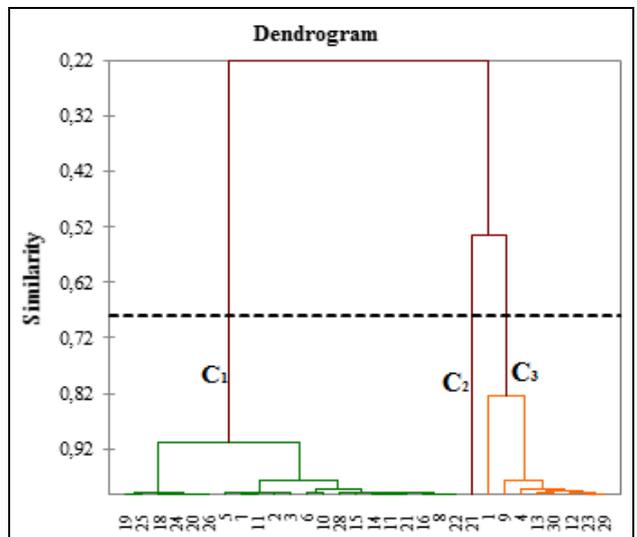

Figure 2. Correlation circle of *physico-chemical* descriptors according to F1 x F2
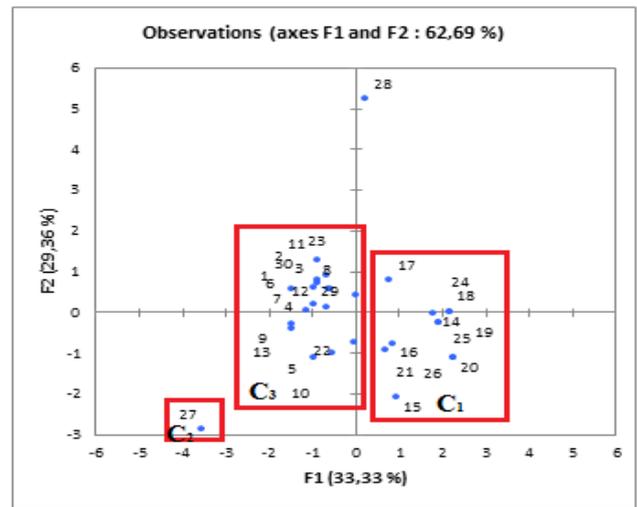
Figure 4. Cartesian diagram according to F1 and F2
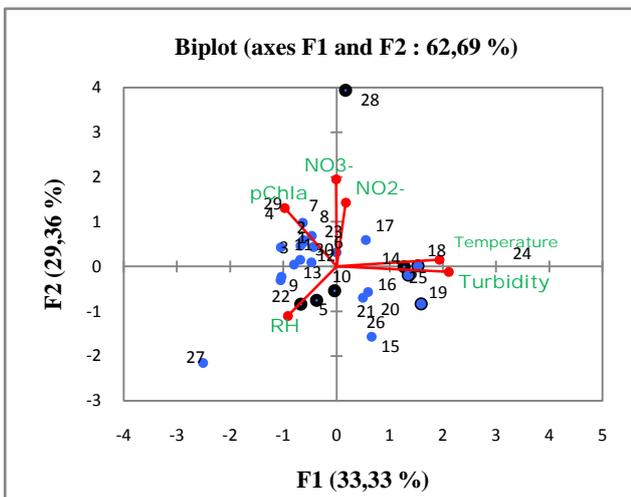
Figure 3. Cartesian diagram according to F1 and F2: correlation between the descriptors used and *chlorophyll-a*

Figure 5. Dendrograms of the samples

The Cartesian diagram of Figure 3 shows that the sample **28** is rich in inorganic nitrogen (NO_2^- and NO_3^-) when the samples **5**, **10** and **22** with a high oxidizing power are inversely proportional to the first one. At the same time, samples **14**, **18**, **19** and **25** have high turbidity and temperature.

Figure 4 illustrates a sample distribution in three classes (**C_1**, **C_2** and **C_3**) characterizing the physico-chemical descriptors. Thus, the class **C_1** characterizes the pollution marked by the turbidity and the temperature while the class **C_2** is marked by the inorganic nitrogen compounds (NO_2^- and NO_3^-) and finally the class **C_3** marked by R_H. This classification is corroborated by the Ascending Hierarchical Classification which also highlights these three groups of samples.

3.2. Multiple Linear Regression (RML)

The equation of the RML model with these statistical indicators is presented below. Figure 6 shows the fitting right of the experimental and theoretical data of *pChl-a* test sets (blue dots) and validation (red dots) of the model. This obtained model relates the *Chlorophyll-a* property and the physico-chemical descriptors of the samples used. The negative or positive sign of the coefficient of a descriptor of model reflects the effect of proportionality in the evolution of *chlorophyll-a* and this parameter of the regression equation. The negative sign indicates an inversely proportional evolution between *Chlorophyll-a*

and this descriptor. The positive sign reflects the opposite effect. The equation obtained is presented below:

$$pChl-a = 4.18721 + 0.05935 * Temperature$$
$$- 0.09422 * R_H - 0.00637 * Turbidity \quad (9)$$
$$- 60.55103 * NO_2^- + 0.40608 * NO_3^-$$

$$N = 20, R^2 = 0.914, Q_{CV}^2 = \mathbf{0.914},$$
$$\mathbf{RMSE} = 0.060, \mathbf{F} = 190.503.$$

The negative signs of the R_H coefficients, turbidity and NO_2^- indicate that these descriptors contribute more to the synthesis of *chlorophyll-a* in the aquatic environment. Temperature and concentration of NO_3^- which are positive descriptors of the regression equation appear to contribute to the increasing production of *chlorophyll-a*. The significance of the model is shown by the high value of the Fischer coefficient which is F = 190.503. In addition, the robustness of this model is translated by the correlation coefficient of the cross validation $Q_{CV}^2 = 0.914 > 0.9$. The regression line established by the experimental and theoretical data of *pChl-a* test and validation sets is shown in Figure 6.

The low value of the standard error RMSE = 0.060 attests to the good similarity between the predicted and experimental values (Figure 7). This curve shows a similar evolution of these data by the multilinear model *pChl-a* of Lake M'koa, despite some recorded differences.
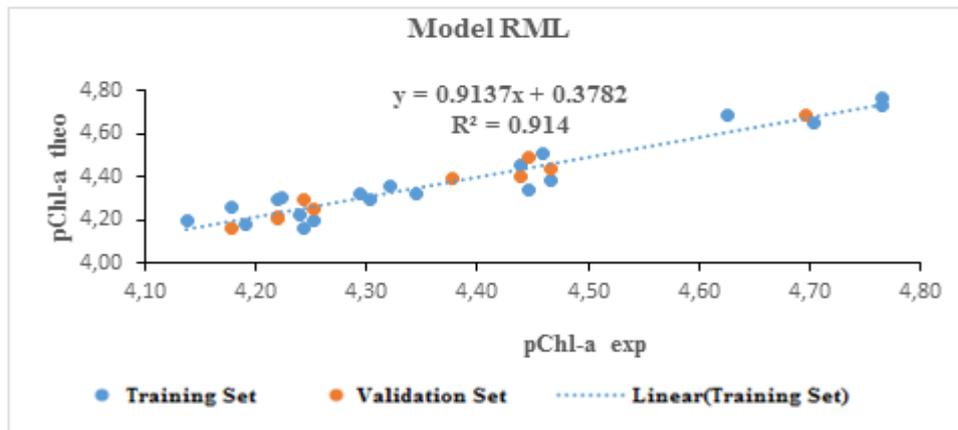

Figure 6. Regression line of the test and validation phases of the RML model
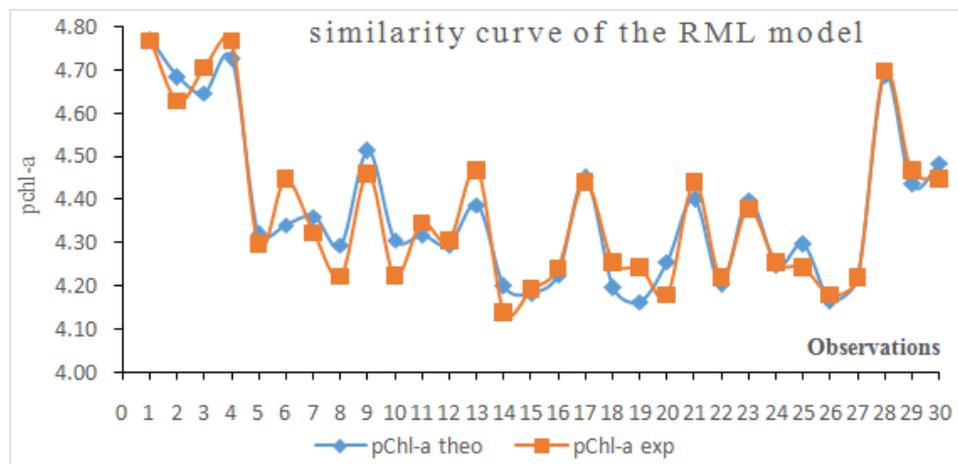

Figure 7. Similarity curve of the experimental and predicted values of the RML model

Verification of Tropsha's Criteria

$$R_{Test}^2 = 0.968 > 0.7, Q_{Cv\ Test}^2 = 0.968 > 0.6,$$

$$\left| R_{Test}^2 - R_0^2 \right| = 0.025 < 0.3, \frac{\left| R_{Test}^2 - R_0^2 \right|}{R_{Test}^2} = 0.025 < 0.1$$

and $0.85 \le k(=1.00) \le 1.15$;

$$\frac{\left| R_{Test}^2 - R_0^{'2} \right|}{R_{Test}^2} = 0.033 < 0.1 \text{ and } 0.85 \le k^{'}(=1.00) \le 1.15.$$

All values meet the Tropsha criteria, so the model is acceptable for the prediction of M'koa Lake *chlorophyll-a*.

3.3. Nonlinear Multiple Regression (RMNL)

The statistical nonlinear regression method was used to improve quantitatively the prediction of *pChl-a*. It takes into account the five selected descriptors (Temperature, R_H, Turbidity, NO_2^-, NO_3^-). It is the most common tool for studying multidimensional data. This statistical method is applied to the data in Table 1 for thirty (30) samples associated with the five (5) physico-chemical descriptors. The resulting equation is expressed as follows:

$$pChl-a = 19.06051 - 0.77605 * \text{Temperature}$$
$$-0.5375 * R_H - 0.00313 * \text{Turbidity}$$
$$+68.70479 * NO_2^- + 1.70536 * NO_3^-$$
$$+0.01424 * \text{Temperature}^2 + 0.01607 * R_H^2 \qquad (10)$$
$$-0.00004 * \text{Turbidity}^2 - 21705 * \left(NO_2^- \right)^2$$
$$-2.01122 * \left(NO_3^- \right)^2$$

$$N = 20, \mathbf{R}^2 = 0.942, \mathbf{Q}_{CV}^2 = 0.942,$$
$$\mathbf{RMSE} = 0.049 \text{ et, } \mathbf{F} = 291.986.$$

The significance of the model is defined by the Fischer coefficient F = 291.986. Concerning the robustness of this model, the correlation coefficient of the cross validation $Q_{CV}^2 = 0,942 > 0.9$, the model is said to be excellent. The line of this model is illustrated in Figure 8.

The very low value of the standard error, RMSE = 0.049, also demonstrates the good similarity between predicted and experimental values (Figure 9). This curve reflects a very good analogue evolution of the experimental values predicted by the Lake M'koa RMNL model, despite some discrepancies.

Verification of Roy's Criteria

$$\overline{r_m^2} = \frac{\left(r_m^2 + r_m^{'2} \right)}{2} = 1.000 > 0.5;$$

$$\Delta r_m^2 = \left| r_m^2 - r_m^{'2} \right| = 0.000 < 0.2$$

which $r_m^2 = 1.000$ *and* $r_m^{'2} = 1.000$.

The two metric criteria of Roy *et al.* are verified, the model is therefore acceptable for the prediction of *chlorophyll-a* at M'koa Lake. On both models, the model obtained by the RMNL statistical method has a much better predictive ability. However, since this model is a function of five physico-chemical descriptors, it is essential to determine the contribution of each of them in the prediction of the *chlorophyll-a* property. Indeed, the knowledge of this contribution permits to establish the order of priority of the various descriptors and to define the choice of the parameters to be optimized for the good prediction and understanding of the eutrophication of the M'koa lake.
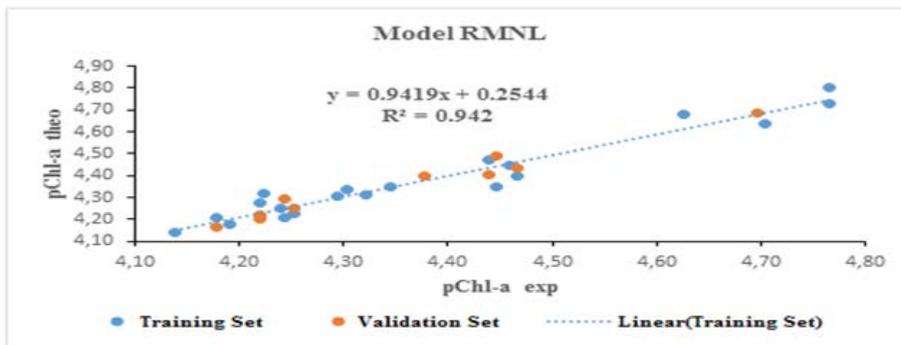

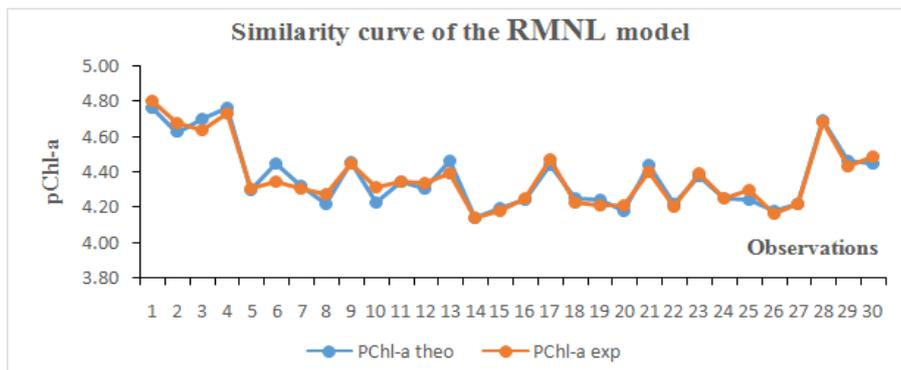**Figure 8**. Regression line of the RMNL test and validation phases

Figure 9. Similarity curve of the experimental and predicted values of the RMNL model
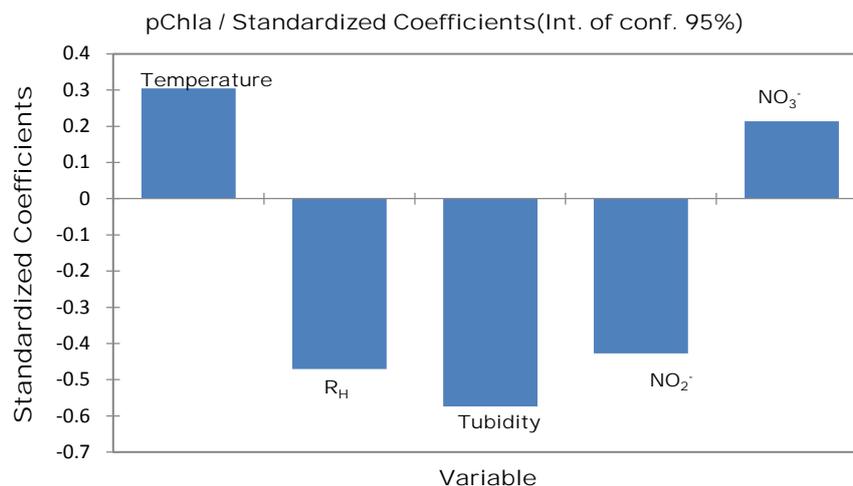
Figure 10. Contribution of physico-chemical descriptors in models

3.4. Analysis of the Contribution of the Descriptors

The contribution of the five physico-chemical descriptors in the prediction of *chlorophyll-a* series of samples was determined from the XLSTAT version 2014 software [21]. The different contributions are illustrated in Figure 10.

The order of priority of physico-chemical descriptors is classified according to the following sequence:

$$\textbf{Turbidity} > \textbf{R}_\textbf{H} > \textbf{NO}_2^- > \textbf{Temperature} > \textbf{NO}_3^-.$$

According to this sequence, turbidity displays a high proportion (-0.5743) and nitrate with 0.2140 the lowest proportion compared to other descriptors. It should be noted that turbidity is the most influential physico-chemical descriptor. Thus, to reduce the phytoplankton biomass of Lake M'koa, it is necessary to reduce the turbidity as much as possible. This means reducing or avoiding any activities that generate particular pollution in Lake M'koa.

4. Conclusion

This study allows to highlight relationships between *chlorophyll-a*, which is a fundamental property of phytoplankton manifestation, and the physico-chemical descriptors measured in the samples. The physico-chemical descriptors (Turbidity, R_H, NO_2^-, Temperature, NO_3^-) permit to explain and predict the property of *chlorophyll-a* because there are strong correlations between the calculated and experimental values. The multivariate analysis allowed us to detect three classes, namely class C1 characterizing the organic pollution expressed by turbidity and temperature. Class C2, which is characterized by inorganic nitrogen pollution (NO_2^- and NO_3^-) and finally class C3 expressed by R_H. The study of the robustness of the two models (RML and RMNL) presented a good stability and an excellent power of prediction. In addition, the RMNL model ($R^2 = 0.942$, RMSE = 0.049, F = 291.986) is better and is an effective tool for predicting *chlorophyll-a* at M'koa Lake. Moreover, the study of the contribution of physico-chemical descriptors has shown that turbidity is the priority descriptor in the prediction of *chlorophyll-a*.

References

[1] Zalewski M., Guidelines for the Integrated Management of the Watersheds. Phytotechnoloy and Ecohydrology, Newletter and Technical Publications. *Freshwater Management Series*, UNEP, (2002), 5, 237.

[2] Fu F., Dionysiou D. D., Liu H.. The use of zero-valent iron for groundwater remediation and wastewater treatment: *A review. J. Hazard. Mater*, (2014), 267, 194-205.

[3] Horn, A.L.; Rueda, F. J.; Hormann, G.; Fohrer, N., Implementing river water quality modelling issues in mesoscale watershed models for water policy demands-An overview on current concepts, deficits, and future tasks. *Phys. Chem. Earth*, (2004), *29*, 725-737.

[4] Montgomery M. A., Elimelech M., Water and sanitation in developing countries: Including health in the equation. *Environmental Science & Technology*, (2007), 41, 17-24.

[5] Organization, W.H. Water, sanitation and hygiene links to health facts and figures-updated. November 2004, (2010), Geneva.

[6] Sheela A. M., Letha J. & Sabu J. Environmental status of a tropical lake system. *Environ Monit Assess*, (2011), 180, 427-449.

[7] Abell J. M., Ozkundakci D. & Hamilton D. P., Nitrogen and phosphorus limitation of phytoplankton growth in New Zealand Lakes: Implications for eutrophication control. *Ecosystems*, (2010), 13, 966-977.

[8] Coulibaly-Kalpy, J., Soumahoro, M-K., Niamien-Ebrotié, J. E., Yéo, K., Amon, L., Djaman, A. J. & Dosso, M., Déterminisme de la prolifération de la prolifération des cyanobactéries toxiques en Côte-d'Ivoire. *Int. J. Biol. Chem. Sci.*, (2017), 11, 266-274.

[9] Hadj, A. R., Quaranta, G., Gueddari, F., Million, D. & Clauer, N., The life cycle impact assessment applied to a coastal lagoon: the case of the Slimane lagoon (Tunisia) by the study of seasonal variations of aquatic eutrophication potential. *Environmental Geology*, (2008), 54, 1103-1110.

[10] Kpidi, Y. H., Yapo, O. B., Ohou-Yao, M.-J., Ballet, T. G., Variabilité journalière de la qualité physico-chimique du lac M'koa de Jacqueville (Côte d'Ivoire). *Int. J. Biol. Chem. Sci.*, (2017), 11, 901-910.

[11] N'Guessan, K. A., Konan, K. F., Kotchi, Y. B., Edia, O. E., Gnagne, T., Traoré, K. S. & Houenou, V. P., Prospects for rehabilitation of man-made lake system of Yamoussoukro (Ivory Coast). *Procedia Environmental Sciences*, (2011), 9, 140-147.

[12] Grizzetti, B., Bouraoui, F., Billen G., van Grinsven H. J. M., Cardoso A. C., Thieu V., Garnier J., Curtis C., Howarth R. W. & Jones P., Nitrogen as a threat to European water quality. In: "European nitrogen assessment" (Sutton M. A., Howard C. M., Erisman J. W., Billen G., Bleeker A., Grennfelt P., Van Grinsven H. J. M., & Grizzetti B., eds.). *Cambridge University Press*, (2011), 379-404.

[13] Neal, C., The potential for phosphorus pollution remediation by calcite precipitation in UK freshwaters. *Hydrology and Earth System Sciences*, (2001), 5, 119-131.

[14] Villeneuve V., Légaré S., Painchaud J. & Warwick V., Dynamique et modélisation de l'oxygène dissous en rivière. *Revue des Sciences de l'Eau*, (2006), 19, 259-274.

[15] Nechad B., Ruddick K. G. & Neukermans G., Calibration and Validation of a Generic Multisensor Algorithm for Mapping of Turbidity in Coastal Waters. *SPIE European International Symposium on Remote Sensing*, Berlin. (2009).

[16] Raihan, A., Mehebub, S., Haroon, S., Preparing turbidity and aquatic vegetation inventory for waterlogged wetlands in Lower Barpani sub-watersheds (Assam), India using geospatial technology, The Egyptian Journal of Remote Sensing and Space Sciences, Egypt. *J. Remote Sensing Space* Sci. (2016), pp (1-7).

[17] Wang, Q. H., Yu, L. J., Liu Y., Lin L., Lu R., Zhu J. p., He L. & Lu Z. L., Methods for the detection and determination of nitrite and nitrate: *A review. Talanta*, (2017), 165, 709-720.

[18] Zhang J., Yang C., Wang X. & Yang X., Colorimetric recognition and sensing of nitrite with unmodified gold nanoparticles based on a specific diazo reaction with phenylenediamin. *Analyst*, (2012), 137, 3286-3292.

[19] Rejset, F., Analyse des eaux: aspects réglementaires et techniques. *Centre Régional de documentation pédagogique d'Aquitaine*. (2002):

[20] Microsoft ® Excel ® 2013 (15.0.4420.1017) MSO (15.0.4420.1017) 64 Bits Partie de Microsoft Office Professionnel Plus. (2013).

[21] XLSTAT Version 2014.5.03 Copyright Addinsoft 1995-2014; XLSTAT and Addinsoft are Registered Trademarks of Addinsoft. (2014).

[22] Chaltterjee, S., Hadi, A. and Price, B., Regression Analysis by Examples. *Wiley VCH*, New York. (2000)

[23] Phuong, H.T.N., Synthèse et étude des relations structure/activité quantitatives (QSAR/2D) d'analogues Benzo [c] phénanthridiniques. Thèse de doctorat, Université d'Angers (2007).

[24] Snedecor, G. W. & Cochran, W. G., Statistical Methods. *Oxford and IBH, New Delhi, India*, (1967). 381.

[25] Diudea, M. V., QSPR/QSAR Studies for Molecular Descriptors. *Nova Science: Huntingdon, New York,* (2000).

[26] Esposito, E.X., Hopfinger, A.J. and Madura, J. D., Methods for Applying the Quantitative Structure-Activity Relationship Paradigm. *Methods in Molecular Biology*, (2004), 275,131-213.

[27] Clarke, K. R. & Ainsworth, M., A method of linking multivariate community structure to environmental variables. *Marine Ecology*, (1993), 92, 205-219.

[28] Escofier, B. & Pagès, J., Analyses factorielles simples et multiples: Objectifs, méthodes et interprétation. Dunod, Paris, (2008), 318.

[29] Rücker C. & Rücker G. J. (2007). Y-Randomization is a tool used in validation of QSPR/QSAR models. *Chem. Inf. Model*, 47, 2345-2357.

[30] Eriksson, L., Jaworska J., Worth A., Cronin M.T. D., Dowell R. M. Mc & Gramatica P., Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification and Regression-Based QSARs. *Environmental Health Perspectives*, (2003), 111, 1361-1375.

[31] Golbraikh A., Tropsha A., Beware of qsar. *J. Mol. Graph. Model*, (2002), 20, 269-276.

[32] Tropsha, A., Best practices for QSAR model development, validation, and exploitation. *Molecular Informatics*, (2010), 29, 476-488.

[33] Tropsha, A. Gramatica, P. & Gombar V. K., The importance of being earnest, validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.*, (2003), 22, 69-77.

[34] Roy, P. P. & Roy, K.. On some aspects of variable selection for partial least squares regression models. *QSAR Comb Sci.*, (2008), 27, 302-313.