

---

# Prediction of the Dynamics of Total Chlorophyll-a in the Waters of the Ebrie Lagoon, Abidjan (Ivory Coast)

Yapo Habib Kpidi<sup>1</sup>, Georges Stephane Dembele<sup>2,5,\*</sup>, Tiama Guy Ballet<sup>1</sup>, Esmel Amari<sup>3</sup>,  
Ossey Bernard Yapo<sup>1,4</sup>, Affian Kouadio<sup>3</sup>

<sup>1</sup>Laboratory of Environmental Sciences (LSE), UFR of Environmental Sciences and Management (UFR SGE), Nangui Abrogoua University, Abidjan, Ivory Coast

<sup>2</sup>Laboratory of Thermodynamics and Environmental Physico-Chemistry, Basic and Applied Sciences Department (UFR SFA), Nangui Abrogoua University, Abidjan, Ivory Coast

<sup>3</sup>Laboratory of Geosciences of Sedimentary Environments and Energies, UFR of Earth Sciences and Mining Resources, Université Félix Houphouët-Boigny, Abidjan, Ivory Coast

<sup>4</sup>Central Environmental Laboratory, Ivorian Anti-Pollution Centre (LCE-CIAPOL), Abidjan, Ivory Coast

<sup>5</sup>Ivorian Group for Research in Disease Modeling (GIR2M), Abidjan, Ivory Coast

## Email address:

1997sageme@gmail.com (G. S. Dembele)

\*Corresponding author

## To cite this article:

Yapo Habib Kpidi, Georges Stephane Dembele, Tiama Guy Ballet, Esmel Amari, Ossey Bernard Yapo, Affian Kouadio. Prediction of the Dynamics of Total Chlorophyll-a in the Waters of the Ebrie Lagoon, Abidjan (Ivory Coast). *International Journal of Environmental Monitoring and Analysis*. Vol. 9, No. 5, 2021, pp. 114-121. doi: 10.11648/j.ijema.20210905.12

**Received:** July 6, 2021; **Accepted:** July 19, 2021; **Published:** September 3, 2021

---

**Abstract:** The aquatic environments which represent a vital resource for man are threatened by various pollution of anthropogenic origin. Chlorophyll-a being the consequence of the nutrient enrichment of water and linked to primary production, participates in the degradation of water quality by the proliferation of phytoplankton. Algal development depends on physical, chemical and biological factors in the aquatic environment. The objective of the study is to determine a better model for predicting chlorophyll-a concentrations from the environmental variables of the waters of the Ebrié lagoon. The predictive model by multilinear regression of total chlorophyll-a allowed to generate excellent statistical indicators and final descriptors of influence ( $R^2 = 0.9076$ ; RMCE = 0.1296;  $F = 216.009$ ). The external validation tests of Tropsha et al and Roy et al are all verified and the model obtained has a good predictive power. Thus, suspended solids (SS) (33%) and pH (23%) were the most influential descriptors of the water model of the Ebrié lagoon. At the level of the applicability domain of the model the value of the threshold lever is  $h^*=0.875$ . The observations of the training set are all within the domain, so the prediction of chlorophyll-a concentrations is reliable.

**Keywords:** Physico-chemical, Biological, Total Chlorophyll-a, Multilinear Regression, Ebrié Lagoon Waters

---

## 1. Introduction

The issue of surface water quality is one of the major environmental challenges facing humanity [1]. The degradation of the quality of this water comes from agricultural activities, industrial and domestic effluents discharged into the receiving environment without prior treatment. Water quality is also at the heart of most public health problems in developing countries [1]. Nutrients from these discharges can cause a series of symptomatic changes including the development of algae and macrophytes that appear undesirable and harmful to the various

uses of water. Indeed, the concentration of chlorophyll-a, as an estimate of the phytoplankton biomass is a good indicator of the state of eutrophication of the environment studied. It is at the base of the pelagic trophic chain and is therefore responsible for an essential part of primary production in aquatic environments. Fresh water and particularly surface water which represents a vital resource for humans, is threatened by various types of pollution of human origin. Phytoplankton reacts to these alterations and can be considered as an indicator of the degradation of the quality of inland waters. Its proliferation has a direct impact on aquatic ecosystems, leading to changes in the

diversity and dynamics of species. The development of algal populations is governed by chemical, biological and physical factors. The advent of new tools for rational management of water resources is a boon. To this end, an approach based on so-called statistical models makes it possible to predict one or more descriptors of the process highlighted as a function of causal variables measured in the field [1, 2]. The objective of this study is to determine a better model for predicting total chlorophyll-a concentrations from the environmental variables measured in the waters of the Ebrié lagoon.

## 2. Materials and Methods

### 2.1. Presentation of the Study Area and Sampling Stations

The Ebrié lagoon, with an area of 566 km<sup>2</sup> [3], Stretches

125 km along the coast of Cote d'Ivoire, between 3 ° 40' and 4 ° 50' West, at the latitude of 5 ° 50' North. The volume of this lagoon is approximately 2,5x10<sup>9</sup> m<sup>3</sup>, the average depth is 4.8 m and a few pits near Abidjan exceed 20 m. The Ebrié lagoon is separated from the Atlantic Ocean by a sandy coastline of 1 to 8 km, crossed in its central part by the Vridi canal since 1951. The study focused on the typical estuarine part in front of Abidjan, due to the increase in pollution in various closed bays in this area. The Ebrié lagoon is fed by ocean waters, through the Vridi Canal, by freshwater streams, the main one, the Comoé River and finally precipitation [3]. Samples were taken at eleven stations in the estuarine zone of the Ebrié lagoon (Figure 1). Several criteria guided our choices, notably urbanization, industrial zones and proximity to the Vridi canal.

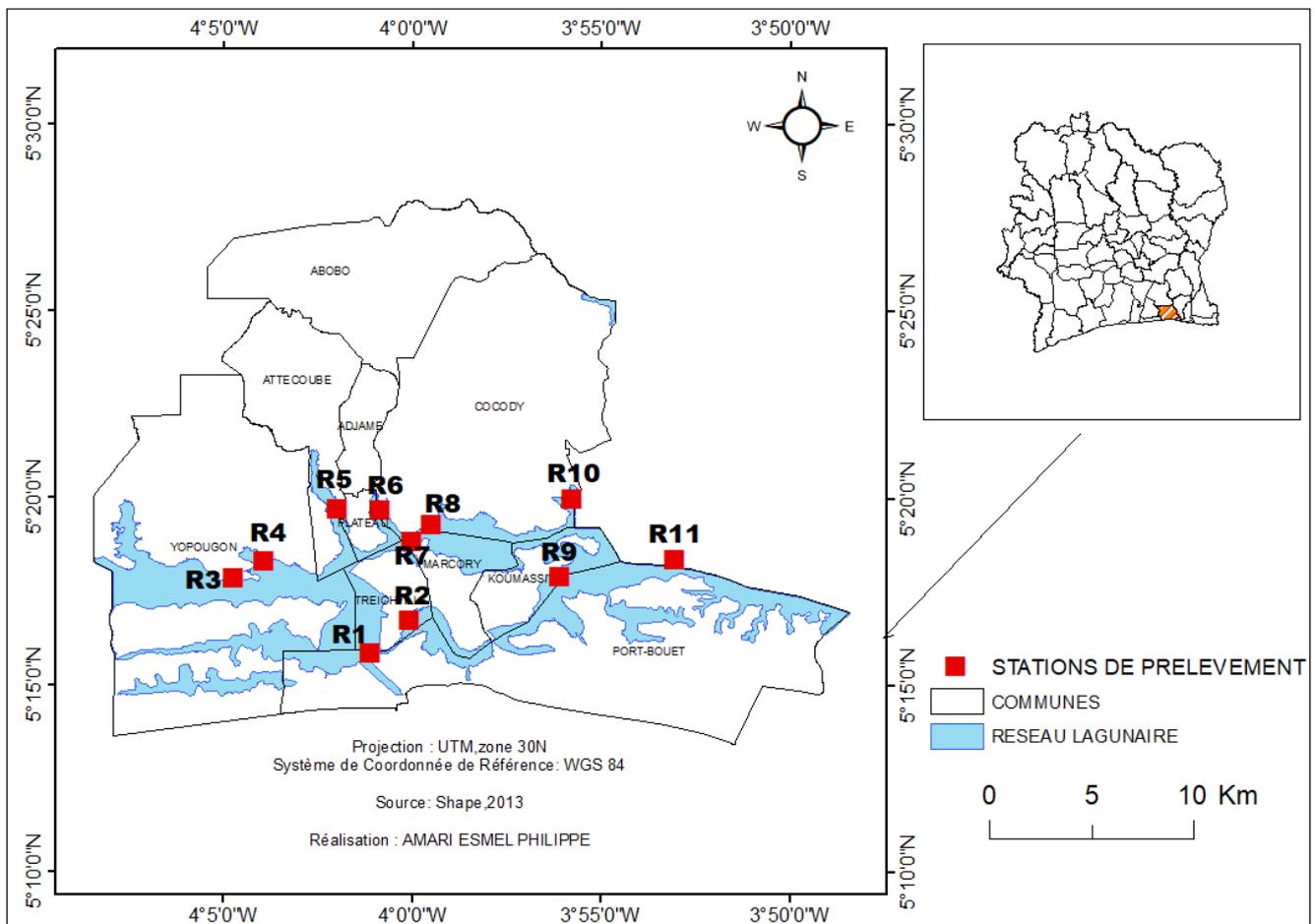


Figure 1. Location of the study area and sampling stations.

### 2.2. Sampling and Analysis

The study data used to carry out this work was provided by the Ivorian Anti-Pollution Center (CIAPOL). They were collected monthly over the period 1993-1998 with an annual average size of 10 samples, along the estuary in the urban area of the Ebrié lagoon, particularly the RNO monitoring stations (Figure 1). One hundred and forty-four (144) bimonthly measurement campaigns were carried out to better

estimate the data. In fact, samples were taken during tides of greater amplitude in the spring water period and of lower amplitude during the dead water sequences.

### 2.3. Physico-chemical Descriptors

Chlorophyll-a is a chlorophyll generally found in organisms that perform photosynthesis. It is a pigment of plants; its role is to capture the light necessary for photosynthesis. Photosynthesis is essential because it allows plants to modify

carbon dioxide (CO<sub>2</sub>) into organic matter. Temperature is an important parameter for the study of water masses. It plays a determining role in biological cycles: growth and reproduction of organisms, planktonic development, microbial activity. It conditions evaporation, the solubility of gases and, in particular, dissolved oxygen, the rates of chemical and biochemical reactions, the density of water [4]. The pH is a physicochemical parameter characterizing the acidity or basicity of a medium, also plays an important role in biological processes in living environments.

Dissolved oxygen is an important parameter of the environment which governs the majority of chemical and biological processes in aquatic ecosystems. The dissolved oxygen concentration (O<sub>2</sub> (%)) is the result of physical (temperature, air-water exchange and salinity), chemical (chemical oxidation) and biological (photosynthesis, respiration and regeneration) factors [2].

The measurement of suspended solids (SS) is necessary in the study of aquatic environments. On the one hand, the particles reduce the transparency of the water and, as a result, the primary photosynthetic production and, on the other hand, they have a large contact surface for physicochemical, chemical or biological exchanges with the water seawater or lagoon [4]. Depending on their nature, they are also a significant source of nutrients for wildlife. Nitrate ion (NO<sub>3</sub><sup>-</sup>) is the stable oxidized form of nitrogen. It enters the nitrogen cycle as the main medium for phytoplankton growth. It is then regenerated from organic forms by bacteria. It results from the nitrification reaction [4].



Formerly "faecal streptococci", enterococci are spherical bacteria, in pairs or chains, Gram positive, catalase negative, facultative anaerobes that hydrolyze esculin in the presence of bile [5]. They are also found on plants. They are nevertheless considered to be indicators of faecal pollution, and their main advantage lies in the fact that they are resistant to desiccation and persist longer in water [4]. They therefore provide additional information on the origin of wastewater.

#### 2.4. Estimation of the Predictive Capacity of a Model

The prediction of total chlorophyll-a was made using a linear regression model. Indeed, given the significant variation of total chlorophyll-a in the waters of the Ebrié lagoon, a base 10 logarithmic transformation was applied in order to bring this variation to a lower level for adequate modeling. Total chlorophyll -a was chosen as the variable to be explained (Y) for their modeling because it seems to relate the effect of ecological disturbances on the environment. The whole modeling process was done using Excel and Xlstat 2016 software.

$$pChl_{a\ tot} = -\log_{10}(Chl_{a\ tot}) \quad (1)$$

The reliability of a model is based on the determination of certain statistical analysis parameters such as the coefficient

of determination R<sup>2</sup>, the standard deviation (S) or the root mean square error (RMSE). R<sup>2</sup>, S and F reflect the difference between calculated and experimental values. These parameters show the predictive ability of the model, and allow to estimate the accuracy of the values calculated on the training set [6, 7]. This forecasting power is said to be "internal" because it is calculated from the structures used to build this model. The coefficient of determination R<sup>2</sup> gives an evaluation of the dispersion of the theoretical values around the experimental values. The quality of the modeling is characterized by the closeness of the points around the fitting line [8]. The fit of the points to this line can be assessed by the coefficient of determination.

$$R^2 = 1 - \frac{\sum(y_{i,exp} - \hat{y}_{i,theo})^2}{\sum(y_{i,exp} - \bar{y}_{i,exp})^2} \quad (2)$$

Or:

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

$\hat{y}_{i,theo}$ : Theoretical value of total chlorophyll-a

$\bar{y}_{i,exp}$ : Average value of the experimental values of the total chlorophyll a.

When the R<sup>2</sup> value is close to 1, the theoretical and experimental values are correlated.

Another statistical indicator used is the Root of the Mean Squares of RMCE Errors. This parameter evaluates the reliability and accuracy of the model:

$$\text{RMCE} = \sqrt{\frac{\sum(y_{i,exp} - y_{i,theo})^2}{n-k-1}} \quad (3)$$

The Fisher F test is used to evaluate the level of statistical significance of the model, i.e. the quality of the choice of descriptors that constitute the model.

$$F = \frac{\sum(y_{i,theo} - y_{i,exp})^2}{\sum(y_{i,exp} - y_{i,theo})^2} * \frac{n-k-1}{k} \quad (4)$$

The coefficient of determination of the cross-validation Q<sup>2</sup><sub>cv</sub> allows to evaluate the accuracy of the predicted values of the training set. It is determined using the following formula:

$$Q_{cv}^2 = \frac{\sum(y_{i,theo} - \bar{y}_{i,exp})^2 - \sum(y_{i,theo} - y_{i,exp})^2}{\sum(y_{i,theo} - \bar{y}_{i,exp})^2} \quad (5)$$

#### 2.5. Acceptance Criteria for a Model

According to Eriksson *et al.* [9], the performance of a mathematical model is described by the value Q<sup>2</sup><sub>cv</sub>, when Q<sup>2</sup><sub>cv</sub> > 0.5 the model is satisfactory and the model excellent when Q<sup>2</sup><sub>cv</sub> > 0.9. According to these authors, given a training set, a model will perform well if the acceptance criterion R<sup>2</sup> - Q<sup>2</sup><sub>cv</sub> < 0.3 is met.

The criteria of Tropsha *et al.* [10, 11, 12] are used for the external validation set to determine the predictive power of a model. These criteria are as follows:

- 1) R<sup>2</sup><sub>Test</sub> > 0.7, 2) Q<sup>2</sup><sub>cv Test</sub> > 0.6, 3) |R<sup>2</sup><sub>Test</sub> - R<sup>2</sup><sub>0</sub>| ≤ 0.3,
- 4)  $\frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} < 0.1$  et  $0.85 \leq k \leq 1.15$ , 5)  $\frac{|R_{Test}^2 - R_0^2|}{R_{Test}^2} < 0.1$  et

$0.85 \leq k' \leq 1.15$ .

On the other hand, Roy and Roy [13], proposed more accurate methods for the predictive ability of a QSAR model. They proposed parameters  $r_m^2$  et  $\Delta r_m^2$ , called metric values.  $r_m^2$  determines the difference between the observed activity and the prediction. Metric values  $r_m^2$  et  $\Delta r_m^2$  are determined from the observed and predicted chlorophyll-a. These two quantities  $r_m^2$  et  $\Delta r_m^2$ , can be determined for the learning set (internal validation) or for the validation set (external validation). According to these authors, a model is acceptable when these criteria are met.

$$\overline{r_m^2} = \frac{(r_m^2 + r'^2_m)}{2} > 0.5$$

$$\Delta r_m^2 = |r_m^2 - r'^2_m| < 0.2$$

$$\text{Or } r_m^2 = r^2 * \left(1 - \sqrt{(r^2 - r_0^2)}\right) \text{ et } r'^2_m = r^2 * \left(1 - \sqrt{(r^2 - r'^2_0)}\right)$$

## 2.6. Statistical Analyzes

The statistical technique of multiple linear regression (MLR) is generally used to study the relationship between a dependent variable (Property) and several independent variables (Descriptors). This statistical tool reduces the differences between experimental and predicted values. It is based on the following preprogrammed XLSTAT functions:

$$y = a + (bx_1 + cx_2 + dx_3 + ex_4) + (fx_{12} + gx_{22} + hx_{32} + ix_{42}) \quad (6)$$

Where a, b, c, d,... represent the parameters and,  $x_1, x_2, x_3, x_4, \dots$  represent the variables.

## 2.7. Scope of Applicability

The domain of applicability of a model is the physico-chemical, structural or biological space, in which the model equation is applicable to make predictions for new observations [14]. It corresponds to the area of the chemical space including the observations of the test set and similar

observations, which are close in the same space [15]. It is therefore necessary to establish the DA of any QSAR model. This is what the Organization for Economic Cooperation and Development (OECD) recommends in developing a QSAR model [16]. There are different methods for determining the applicability domain of a model [15]. The method used in this work is that of leverage. This method is based on the variation of the standardized residuals of the dependent variable with the distance between the descriptor values and their mean, called leverage [17]. The  $h_{ii}$  are the diagonal elements of an H matrix called the hat matrix, H is the projection matrix of the real values of the explained variable  $Y_{exp}$  in the space of the predicted explained variable values  $Y_{pred}$  such as:

$$Y_{pred} = HY_{exp} \quad (7)$$

H is defined by the expression (8):

$$H = X(X^t X)^{-1} X^t \quad (8)$$

The range of applicability is delimited by a limit value of the lever noted  $h^*$ . This value is determined by  $3 \frac{p+1}{n}$ , where n is the number of observations in the training set, and p is the number of physicochemical parameters in the model [18, 19]. At the level of standardized residuals, the threshold values used are  $\pm 3\sigma$ ,  $\sigma$  being the standard deviation of the real values of the quantity to be explained [20]: this is the "three sigma rule" [21].

## 3. Results and Discussion

This modeling study was carried out on 36 observations. The observations were divided into 2 groups, twenty-four (24) were used for the learning game and twelve (12) for the validation game. The objective of this work is to model total chlorophyll a from certain physicochemical descriptors. Table 2 shows the physicochemical and biological descriptors taken into account by the established pChl-a total model.

Table 1. Characteristics of the physicochemical and biological descriptors of the model.

Observation	Temp (°C)	pH	O <sub>2</sub> (%)	Strep.F (N/100)	NO <sub>3</sub> (mg/l)	SS (mg/l)	pChl-a total
Training Set							
1	25.5	7.4	97	192	0.0420	6.4	1.2504
2	27	6.79	116	37550	0.1170	1.2	0.4624
3	28	7.68	108	693	0.1750	9.6	0.9294
4	28.5	7.4	125	1950	0.1670	9.6	0.9294
5	28	6.72	108	2398	0.1170	7.4	0.7160
6	30.1	9.03	137	585	0.0870	0.4	1.8859
7	29.6	7.2	65	10175	0.1310	4.8	0.6721
8	29.3	7.38	77	34	0.0680	6.2	1.0212
9	30.4	8	107	40	0.0180	7.8	1.0212
10	31.1	8.43	175	20	0.0010	10.8	1.2201
11	27.7	7.7	56	9734	0.0220	2.8	1.1173
12	28.6	7.48	134	1335	0.0620	17.2	0.8451
13	27.2	7.7	98	600	0.0390	20	0.7404
14	29.3	7.45	132	3000	0.2030	7	0.7404
15	29.2	7.54	147	500	0.2940	29.2	0.2304
16	30.3	8.12	139	7427	0.0870	1.2	1.2355

Observation	Temp (°C)	pH	O <sub>2</sub> (%)	Strep.F (N/100)	NO <sub>3</sub> (mg/l)	SS (mg/l)	pChl-a total
17	30.1	7.62	49.3	205	0.0000	8.8	0.8976
18	32	7.8	72	100	0.0100	7	1.1875
19	28.4	8.19	50.3	480	0.1130	2.6	1.5172
20	29.1	7.51	45.8	225	0.1040	2.2	0.7404
21	25.7	7.87	110	1370	0.0460	4.5	1.3464
22	26.5	7.42	64.8	685	0.0430	25.6	0.3617
23	27.5	7.59	113	503	0.0940	12.4	0.8808
24	30.4	8.06	128	1000	0.0290	3.6	1.3997
Validation Set							
25	27	7.07	95	575	0.1430	17.4	0.6128
26	27	7.65	127	7400	0.1490	10.8	0.6435
27	29.2	7.6	91	21500	0.0060	2.6	0.8633
28	29.1	6.66	164	4680	0.1430	9.2	0.7243
29	27.9	7.55	149	10625	0.0920	5.8	1.1173
30	28.2	7.08	140	100	0.0850	0.4	1.1461
31	30.2	7.75	147	850	0.2420	3.6	1.1271
32	30.4	7.36	146	420	0.08	4.6	0.9956
33	27	7.71	50.7	95	0.0210	3.6	1.2967
34	30.1	7.11	56.5	2050	0.08	25.5	0.1761
35	30	7.30	88	30	0.1240	4.2	0.9395
36	30.2	8.08	107	1400	0.0230	10	1.0334

### 3.1. Total Chlorophyll-a Model of the Waters of the Ebrié Lagoon

The total chlorophyll-a potential (pChl-a tot) from the logarithmic transformation to base 10 of the total chlorophyll-a concentrations was used for the data modeling.

The equation of this model is presented below:

$$pChl_{a\ tot} = -0.05334 * Temp(^{\circ}C) + 0.35836 * pH + 0.00174 * O_2(\%) - 0.00002 * Strep.F(N/100) - 0.87403 * NO_3(mg/l) - 0.03359 * SS(mg/l)$$

The positive sign of the coefficient of the potential in H<sup>+</sup> ion (pH) and of the percentage of dissolved oxygen (O<sub>2</sub> (%)) indicates large values of these descriptors contributing to photosynthesis in the aquatic environment. However, the negative sign of the coefficient of temperature (Temp (°C)), faecal streptococci (Strep.F), nitrate (NO<sub>3</sub><sup>-</sup>) and suspended solids (SS) content indicates that the synthesis of chlorophyll -a total will be improved for low values of these descriptors.

Table 2 presents the statistical indicators of the pChl-a tot model of the waters of the Ebrié lagoon.

Table 2. Performance of the lagoon water test sets.

R <sup>2</sup>	R <sup>2</sup> ajusté:	Standard Error: RMCE	Fisher	Q <sup>2</sup> <sub>cv</sub>
0.9076	0.8767	0.1296	216.009	0.9076

Modeling by multilinear regression of the total chlorophyll-a contents of the waters of the Ebrié lagoon allowed the establishment of a model for the prediction of data. Since the robustness of a model is determined by statistical parameters, the best known of which is the coefficient of determination (R<sup>2</sup>), it evaluates the part of variance explained by the established model. In addition, 90.76% of the variances of pChl-a tot were explained by the model obtained from the waters of the Ebrié lagoon. The established multilinear regression model was accepted and validated. First, the coefficient of determination recorded was high and very close to 1. Indeed, the appropriate coefficient of determination to qualify a multilinear model as excellent is

at least 0.90 and therefore close to 1. This result is in agreement with that of Duidéa [22], Kpidi *et al.*, [23]. According to these authors, the greater the value of R<sup>2</sup> ≥ 0.90 and therefore close to 1, the better the established model. Therefore, the predicted and experimental values are strongly correlated and demonstrate the stability of the established model. This can be effectively used for the prediction of total chlorophyll-a contents.

Another criterion to support the stability of the model is the standard error or RMCE. Knowing that the standard error is to the multilinear regression coefficient what the standard deviation is to the mean of a variable, its estimate could be a measure of the dispersion of the observed values of the dependent variable on the regression line. This seems weak for the established model. Indeed, the estimated standard error of the pChl-a tot model is 0.1296. According to Guendouzi [24], the low standard error value would reflect a good statistical fit of the model and a high reliability of the prediction as obtained by Kpidi *et al.* [24].

Also, the value of the Fisher test (F = 216.009) is relatively high compared to the critical value from the Fisher table, F<sub>cr</sub> = 2.36. This value 216.009 from Fisher's test, greater than the critical value shows that the error made is less than what the model explains. This model is acceptable because it agrees with the acceptance criteria of Erickson *et al* [25]. R<sup>2</sup> - Q<sup>2</sup><sub>cv</sub> = 0.901 - 0.901 = 0.000 < 0.3.

The regression graph of the RML model presenting experimental total chlorophyll-a as a function of predicted total chlorophyll-a is shown in Figure 2.

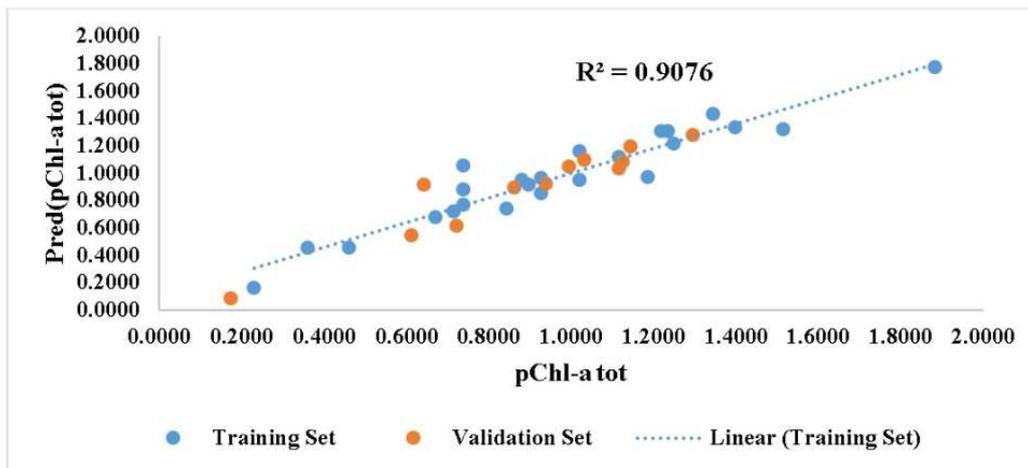


Figure 2. Regression curve for total chlorophyll-a test and validation sets.

Analysis of the regression curve of the RML model shows that all points are around the regression line. This result indicates that there is a small difference (RMCE = 0.1296) between the values of pChl-a tot exp and pChl-a th, therefore a good similarity at the level of these values. This similarity is illustrated in Figure 3.

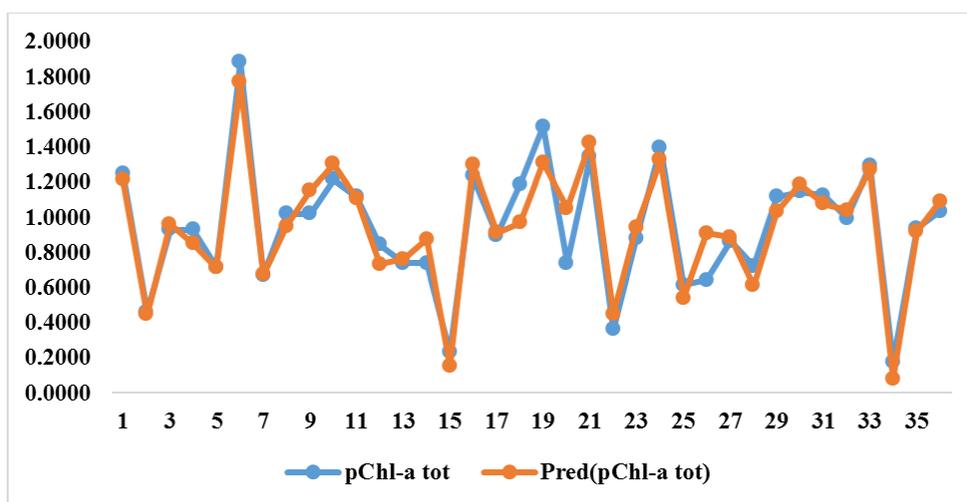


Figure 3. Similarity curve of experimental and predicted values of chl- a tot.

### 3.2. External Validation Criteria

The Tropsha and Roy criteria checks for the external validation set are shown in Tables 3 and 4.

Table 3. Verifications of the Tropsha criteria of the model's external validation set.

$R_{Test}^2$	$Q_{Cv Test}^2$	$ R_{Test}^2 - R_0^2 $	$\frac{ R_{Test}^2 - R_0^2 }{R_{Test}^2}$	$\frac{ R_{Test}^2 - R_0^2 }{R_{Test}^2}$	k	k'
0.9033	0.888	0.0004	0.0004	0.0154	0.9996	0.9894

Table 4. Checks of Roy's criteria of the model's external validation set.

$r_m^2$	$r'^2_m$	$\bar{r}_m^2 = \frac{(r_m^2 + r'^2_m)}{2}$	$\Delta r_m^2 =  r_m^2 - r'^2_m $
0.8852	0.7968	0.8410	0.0884

All the values meet the criteria of Tropsha and Roy, the model is therefore acceptable for the prediction of total chlorophyll-a in the waters of the Ebrié lagoon.

### 3.3. Analysis of the Contribution of Descriptors

Figure 4 provides information on the incremental impact of each descriptor in this model.

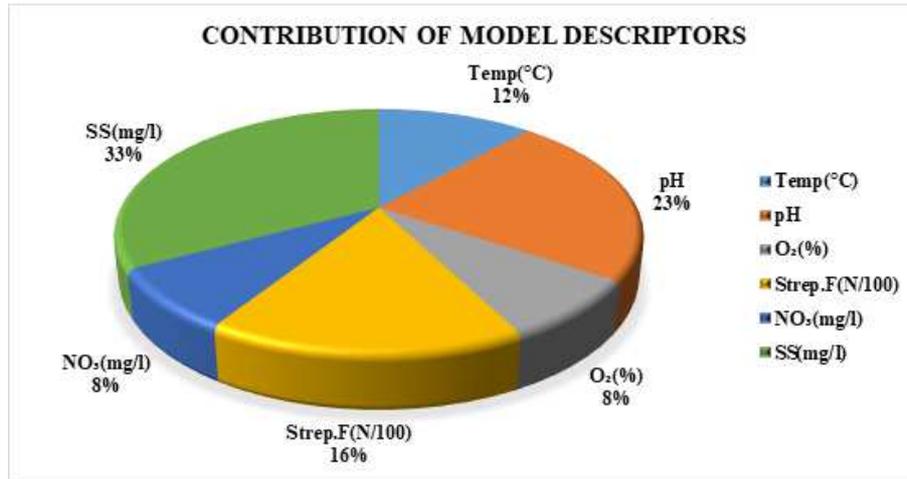


Figure 4. Contribution of model descriptors.

Each descriptor among those taken into account in the established model has a so-called incremental weight. This is measured by a normalized coefficient X for its part taken in the model. The remark is that in the waters of the Ebrié lagoon, the amount of suspended solids (SS) was recorded as being the most influential variable or having the strongest contribution to the establishment of said model with 33%.

### 3.4. Scope of the Model

In order to establish the domain of applicability of the model, we used the "Leverage" method. The values of the standardized residuals and of the levers  $h_{ii}$ , made it possible to draw the graph represented in figure.

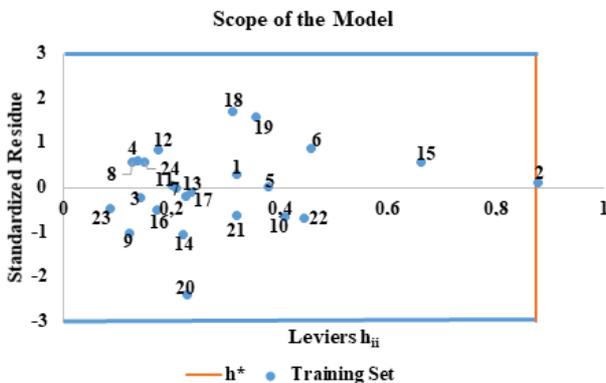


Figure 5. Graph of standardized residues of total chlorophyll-a according to the levers.

Figure 5 shows us that the 24 observations of the learning game and the 6 descriptors of the model, the threshold value of the levers  $h^*$  is 0.875. Also for the standardized residuals, the values are within  $\pm 3\sigma$  according to the "three sigma rule" [21]. These different elements define the domain of applicability of the model as shown in Figure 5.

## 4. Conclusion

In this work we have obtained a prediction model of total chlorophyll-a concentrations from environmental variables of

the waters of the Ebrié lagoon. pH, dissolved oxygen (O<sub>2</sub> (%)), temperature (Temp (°C)), faecal streptococci (Strep.F), nitrate content (NO<sub>3</sub><sup>-</sup>) and suspended matter (SS) are the physicochemical and biological parameters which best explain the total chlorophyll-a concentration. The statistical tool used for this work is multilinear regression (MLR). The statistical indicators of the model, namely  $R^2=0.9076$ ;  $RMCE= 0.1296$ ;  $F=216.009$ , the external validation test of Tropsha and Roy, show that the model is acceptable, robust and with good predictive power. In terms of the contribution of descriptors in the model, suspended solids (SS) with 33% of the incremental weight proved to be the most influential descriptor of the established model. The domain of applicability of the model, obtained from the lever method, showed that the prediction of total chlorophyll-a from observations is reliable.

## References

- [1] Y. H. Kpidi, O. B. Yapo, T. G. Ballet et M.-J. Ohou-Yao, "Daily variability of the physico-chemical quality of Lake M'koa of Jacqueville (Ivory Coast)", *International Journal of Biological and Chemical Sciences*, vol. 11, n°12, p. 901–910, 2017.
- [2] E. Gnagne Agness, «Characterization of the effluents drained by the wastewater network of the city of Abidjan and prediction of TSS and COD from turbidity measurement», PhD Thesis Abidjan, 2017.
- [3] K. M. Yao, B. S. Metongo, A. Trokourey et Y. Bokra, «Water pollution of the urban area of a tropical lagoon by oxidizable matter (Ebrié Lagoon, Ivory Coast)» *International Journal of Biological and Chemical Sciences*, vol. 3, n°14, 2009.
- [4] A. M. Kouassi, «Hydrochemistry and water quality of two tropical lagoons in Ivory Coast» PhD Thesis, Abidjan, 2005.
- [5] CEAEQ, «Research and enumeration of enterococci: membrane filtration method» 2006.
- [6] G. W. Snedecor et W. G. Cochran, «Methods, Statistical» *Oxford and IBH: New Delhi, India*; p. 381, 1967.

- [7] N. J.-B. Kangah, M. G.-R. Koné, C. G. Kodjo, B. R. N'guessan, A. L. C. Kablan, S. A. Yéo et N. Ziao, «Antibacterial Activity of Schiff Bases Derived from Ortho Diaminocyclohexane, Meta-Phenylenediamine and 1,6-Diaminohexane: Qsar Study with Quantum Descriptors» *International Journal of Pharmaceutical Science Invention*, vol. 6, n°13, pp. 38-43, 2017.
- [8] E. X. Esposito, A. J. Hopfinger et J. D. Madura, «Methods for Applying the Quantitative Structure-Activity Relationship Paradigm» *Methods in Molecular Biology*, vol. 275, pp. 131-213., 2004.
- [9] L. Eriksson, J. Jaworska, A. Worth, M. D. Cronin, R. M. Mc Dowell et P. Gramatica, «Methods for Reliability and Uncertainty Assessment and for Applicability Evaluations of Classification- and Regression-Based QSARs» *Environmental Health Perspectives*, vol. 111, n°110, pp. 1361-1375, 2003.
- [10] A. Golbraikh et A. Tropsha, «Beware of qsar» *J. Mol. Graph. Model*, vol. 20, pp. 269-276, 2002.
- [11] A. Tropsha, P. Gramatica et V. K. Gombar, «The importance of being earnest, validation is the absolute essential for successful application and interpretation of QSPR models» *QSAR Comb. Sci.*, vol. 22, pp. 69-77, 2003.
- [12] O. Ouattara, T. S. Affi, M. G.-R. Koné, K. Bamba et N. Ziao, «Can Empirical Descriptors Reliably Predict Molecular Lipophilicity ? A QSPR Study Investigation» *Int. Journal of Engineering Research and Application*, vol. 7, n°15, pp. 50-56., 2017.
- [13] Roy P. P. et K. Roy, «on some aspects of variable selection for partial least squares regression models» *QSAR Comb Sci*, vol. 27, pp. 302-313, 2008.
- [14] N. N.-Jeliazkova et J. Jaworska, « An Approach to Determining Applicability Domains for QSAR Group Contribution Models: An Analysis of SRC KOWWIN» *ATLA* 33, p. 461-470, 2005.
- [15] F. Sahigara, K. Mansouri, D. Ballabio, A. Mauri et V. C. a. R. Todeschini, «Comparison of Different Approaches to Define the Applicability Domain of QSAR Models» *Molecules*, vol. 17, pp. 4791-4810, 2012.
- [16] K. Roy et e. al, «A Primer on QSAR/QSPR Modeling Chapter 2 Statistical Methods in QSAR/QSPR» *Springer Briefs in Molecular Science*, pp. 37-59, 2015.
- [17] J. Jaworska, N. N. Jeliazkova et T. Aldenberg, «QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review» *ATLA* 33, p. 445-459, 2005.
- [18] M. Ghamali, S. Chtita, M. Bouachrine et T. Lakhlifi, «Méthodologie générale d'une étude RQSA/RQSP, Revue Interdisciplinaire,» vol. 1, n°11, 2016.
- [19] S. Chtita, M. Ghamali, R. Hmamouchi, B. Elidrissi, M. Bourass, M. Larif, M. Bouachrine et T. Lakhlifi, «Investigation of Antileishmanial Activities of Acridines Derivatives against Promastigotes and Amastigotes Form of Parasites Using QSAR Analysis» *Advances in Physical Chemistry*, pp. 1-16, 2016.
- [20] T. Asadollahi, S. Dadfarnia, A. Shabani, J. Ghasemi et M. Sarkhosh, «QSAR Models for CXCR2 Receptor Antagonists Based on the Genetic Algorithm for Data Preprocessing Prior to Application of the PLS Linear Regression Method and Design of the New Compounds Using In Silico Virtual Screening» *Molecules*, vol. 16, pp. 1928-1955, 2011.
- [21] S. Chtita, M. Larif, M. Ghamali, M. Bouachrine et T. Lakhlifi, «Quantitative structure-activity relationship studies of dibenzo [a,d]cycloalkenimine derivatives for non-competitive antagonists of N-methyl-D-aspartate based on density functional theory with electronic and topological descriptors,» *Journal of Taibah University for Science*, vol. 9, pp. 143-154, 2015.
- [22] M. Diudea, *QSPR/QSAR Studies for Molecular Descriptors*, New York. USA.: Nova Science: Huntington, 2000.
- [23] Y. H. Kpidi, O. B. Yapo, M. G.-R. Koné, G. A. Gadj, J.-E. Gnagne, J. S. N'dri et N. 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, n° 11, pp. 1-9, 2018.
- [24] A. Guendouzi, «Élaboration des modèles QSPR prédictifs des propriétés physico-chimiques à l'aide des descripteurs moléculaires» 2015.
- [25] Y. H. Kpidi, M. Koné, G. Ballet, L. Ouattara, O. B. Yapo et N. Ziao, «Modeling of the Dynamics and Prediction of Transpercy of a Mesotropic Tropical Lake: Bakré Lake (Abidjan, Côte d'Ivoire)» *Ressources and Environnement*, vol. 3, n°18, pp. 174-183, 2018.