



## Use of statistical models for the prediction of an indicator parameter of the groundwaters quality: Case study of fissured aquifers groundwaters of Ivory Coast

A. M. Kouassi<sup>1\*</sup>, K. E. Ahoussi<sup>2</sup>, K. E. Kouakou<sup>1</sup>, A. Mamadou<sup>3</sup>, J. Biemi<sup>2</sup>

<sup>[1]</sup> Institut National Polytechnique Félix Houphouët-Boigny (INP-HB) ; Département des Sciences de la Terre et des Ressources Minières (STeRMi) ; Laboratoire du Génie Civil, des Géosciences et des Sciences Géographiques ; BP 1093 Yamoussoukro (Côte d'Ivoire), Tél. (+225) 30 64 67 15 (Côte d'Ivoire).

<sup>[2]</sup> Université Félix Houphouët-Boigny d'Abidjan-Cocody, Unité de Formation et de Recherche des Sciences de la Terre et des Ressources Minières (UFR-STRM) ; Laboratoire des Sciences et Techniques de l'Eau et de l'Environnement (LSTEE) ; 22 BP 582 Abidjan 22 ; Tel : (+225) 22 48 38 03 (Côte d'Ivoire).

<sup>[3]</sup> Institut National Polytechnique Félix Houphouët-Boigny (INP-HB) de Yamoussoukro ; Ecole Supérieure des Mines et de Géologie (ESMG) ; Bp: 1093 Yamoussoukro (Côte d'Ivoire) ; Tél. : (+225) 30 64 36 07.

\* Corresponding Author: e-mail : [michel.a\\_kouassi@yahoo.fr](mailto:michel.a_kouassi@yahoo.fr) ou [michel.kouassi@inphb.edu.ci](mailto:michel.kouassi@inphb.edu.ci)

The objective of this study, carried out in Côte d'Ivoire, aims to evaluate and compare the predictive power of two statistical models with hydrochemical variables in order to simulate the electrical conductivity of groundwaters. An analysis of the performance and robustness of the different models using numerical criteria (correlation coefficient and Nash-Sutcliffe) and graphs, showed that both models are powerful and robust but the model 2 is more efficient and robust than the model 1. The criterion most suitable for the study of the reliability of the physical and chemical data was the Nash-Sutcliffe criterion.

Received: 18 October 2016  
Accepted: 8 November 2016  
Available online: 11 December 2016

### Keywords:

Statistical Models  
Hydrochemical variables  
Performance  
Robustness  
Ivory Coast

## Introduction

L'accroissement rapide de la démographie ainsi que l'amélioration du niveau de vie des populations engendre un développement économique et social et nécessitent proportionnellement un accroissement des besoins en eau. Au-delà de sa quantité, la qualité de l'eau est un critère essentiel d'évaluation de la ressource en eau. Ainsi, la prédiction des paramètres indicateurs de la qualité des eaux souterraines est actuellement au cœur de la recherche scientifique. De ce fait, la connaissance d'un certain nombre d'indicateurs tels que la conductivité électrique est primordiale. Dans la littérature, des méthodes statistiques

telles que les régressions linéaires et les réseaux de neurones artificiels ont été utilisées par plusieurs auteurs pour la simulation de variables relatives à la qualité de l'eau en général et la simulation de la conductivité électrique en particulier [1-11]. Une analyse physico-chimique peut être définie comme une suite d'opérations élémentaires, statistiquement indépendantes les unes des autres, qui commencent au moment de la prise d'essai (prélèvement d'un échantillon) et aboutissent à l'expression d'un résultat d'analyse de laboratoire qu'il faudra valider pour pouvoir disposer enfin d'une donnée analytique. Après validation des résultats par les laboratoires, il y a nécessité de contrôler la validité des résultats obtenus (contrôles physico-chimiques) dans l'objectif d'assurer la sécurité hydrique des populations,

nécessitant notamment l'obtention de résultats d'analyse fiables. La fiabilité des résultats implique l'utilisation de méthodes validées, mises en œuvre par un laboratoire compétent. Pour vérifier cette fiabilité, un certain nombre de critères sont utilisés (balance ionique, rapport conductivité électrique *in situ* et conductivité électrique mesurée au laboratoire, etc.). Kouassi et al. [6] ont utilisé les approches basées sur la balance ionique et la simulation de la conductivité électrique s'appuyant sur l'adéquation entre les conductivités électriques mesurées sur le terrain et celles simulées. Au regard de la divergence obtenue par Kouassi et al. [6] sur l'échantillon de Bondoukou dont la balance ionique a été déclarée non équilibrée alors qu'il a présenté de bonnes performances, deux modèles statistiques ont été développés par Kouassi et al. [7].

Les résultats obtenus par Kouassi et al. [7] ont montré que les variables les plus pertinentes et expressives pour la modélisation de la conductivité électrique des eaux souterraines sont la somme des cations majeurs et la somme des anions majeurs. Ainsi, ces deux paramètres ont été utilisés comme variables explicatives pour la conception d'un premier modèle. Aussi, pour mettre en évidence les impacts des activités anthropiques sur la qualité des eaux souterraines, un deuxième modèle a été envisagé par intégration de la température et du pH dans le premier modèle. Les erreurs-types associées aux coefficients de régression des différents modèles fluctuent entre 5 et 10 pour le premier modèle et oscillent en général entre 4 et 12 pour le modèle 2. Ces différents résultats traduisent des relations étroites entre la variable expliquée (conductivité électrique) et les variables explicatives. Cependant, une question essentielle reste posée. Ces modèles sont-ils capables de simuler des valeurs de conductivité électrique proches de celles observées ? En d'autres termes, ces modèles sont-ils performants ?

L'objectif général de cette étude vise l'évaluation du pouvoir prédictif de deux modèles statistiques à variables hydrochimiques à simuler la conductivité électrique des eaux souterraines dans le socle cristallin et cristallophyllien. L'hypothèse de travail a consisté à supposer qu'il existe une concordance entre la fiabilité au sens de la balance ionique et la fiabilité au sens de la simulation de la conductivité électrique.

### Présentation de la zone d'étude

La Côte d'Ivoire, espace de notre travail, fait partie des pays du Golfe de Guinée. Elle s'étend sur une superficie de 322 462 Km<sup>2</sup>, environ 1% du continent africain. Ses frontières dessinent approximativement un carré s'inscrivant entre les coordonnées de 2°30' et 8°30' de longitude Ouest, 4°30' et 10°30' de latitude Nord avec, au Sud, une façade littorale de 550 Km. Elle est limitée au Nord par le Mali et le Burkina Faso, à l'Est par le Ghana, à l'Ouest par la Guinée et le Libéria et au Sud par l'océan atlantique. La figure 1 montre la situation de la Côte d'Ivoire en Afrique de l'Ouest et les zones de

prélèvement. Ces zones tests concernent les régions de l'ex-N'zi-Comoé (Dimbokro, Daoukro, Bongouanou et M'bahiakro), Guiglo-Duekoué, Tiébissou-Didiévi, Bondoukou et de Tiassalé. Deux grandes zones climatiques se côtoient: le climat équatorial et le climat tropical. Le climat subéquatorial dans le Sud est caractérisé par des précipitations abondantes, qui atteignent en moyenne 1 800 mm. Le climat tropical humide de savane couvre le Centre du pays. Les précipitations annuelles moyennes sont proches de 1 200 mm.

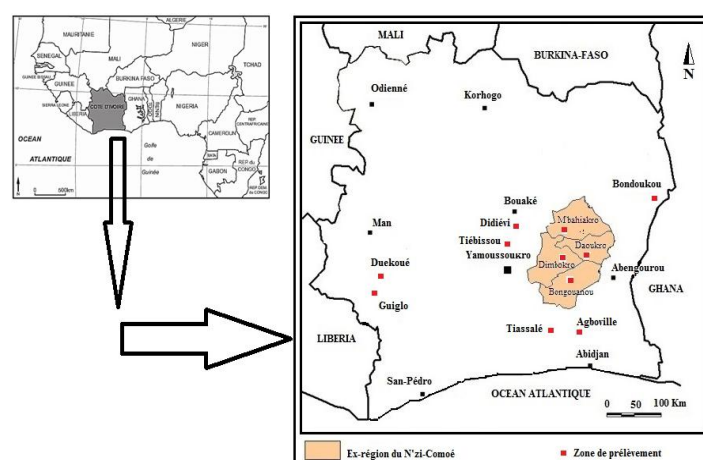


Fig. 1: Carte de la zone d'étude et sites d'échantillonnage des eaux

Le climat tropical atténué concerne principalement la partie nord du pays et les précipitations moyennes annuelles enregistrées sont généralement inférieures à 1000 mm. Le régime climatique régnant au niveau de l'Ouest de la Côte d'Ivoire est un climat de montagne avec des hauteurs pluviométriques annuelles dépassant parfois les 2 000 mm. Deux grands types de paysages végétaux se partagent le territoire ivoirien : un paysage forestier et un paysage de savane. Le premier correspond à la moitié sud du pays et appartient au domaine guinéen. Le second occupe la moitié nord de la Côte d'Ivoire et se rattache au domaine soudanais [12]. Le socle ivoirien est composé des domaines archéen, paléoprotérozoïque et SASCA. Le domaine archéen a été affecté par des déformations tectoniques très importantes attribuées à l'orogénèse léonienne (3500-2900 Ma) et à l'orogénèse libérienne (2900-2500 Ma). Il est caractérisé par les reliefs les plus accidentés de la Côte d'Ivoire. Le domaine archéen est dominé par des granites, des migmatites, des granito-gneiss, des amphibolo-pyroxénites, des charnockites. Le domaine paléoprotérozoïque est subdivisé en deux grands sous-ensembles (ensemble des granitoïdes et formations volcano-sédimentaires). Il a été structuré au cours du mégacycle éburnéen. Ce domaine est constitué d'un socle granito-gneissique dominé par les granites et de formations volcano-sédimentaires (métasédiments et roches vertes). Les métasédiments sont constitués en majeure partie de grès et de schistes. Les roches vertes sont constituées de metabasalte,



L'équation du modèle 2 est la suivante (équation 2):

$$CE(\mu S/cm) = 60,337 \times \sum \text{Cations} + 58,345 \times \sum \text{Anions} + 0,953 \times T - 51,866 \times \text{pH} + 298,950 \quad (2)$$

- CE : conductivité électrique simulée ( $\mu S/cm$ ) ;
- $\sum \text{cations}$  : somme des cations majeurs (méq/l) ;
- $\sum \text{anions}$  : somme des anions majeurs (méq/l) ;
- n : nombre de cations majeurs ;
- m : nombre d'anions majeurs ;
- T : température de l'eau ( $^{\circ}C$ ) ;
- pH : pH de l'eau ;

## 2.2. Evaluation de la performance des modèles

L'évaluation de la qualité des modèles au niveau du calage et de la validation a été réalisée sur les données de l'ex-région du N'zi-Comoé (Centre-Est de la Côte d'Ivoire). L'échantillon de calage est de 128 forages. La validation est effectuée sur un jeu de 65 forages. La transposition des modèles a été réalisée sur plusieurs zones tests de Côte d'Ivoire. Il s'agit des régions de Guiglo-Duékoué, Bondoukou, Tiébissou-Didiévi, Tiassalé et Agboville. Afin d'évaluer la performance et de comparer les différents modèles entre eux, deux indices de performance ont été calculés parmi tant d'autres : le coefficient de corrélation R et le critère de Nash-Sutcliffe. En effet, plusieurs auteurs [6, 7, 23, 24] ont utilisé ces indices pour rendre compte de la performance des modèles. Des représentations graphiques ont été également réalisées pour compléter les indices numériques.

### 2.2.1. Coefficient de corrélation

Le coefficient de corrélation est utilisé pour estimer la corrélation linéaire entre la conductivité électrique mesurée sur les échantillons d'eau et celle simulée avec le modèle empirique. Son équation est la suivante (équation 3):

$$R = \frac{\sum_i (X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum_i (X_i - \bar{X})^2 \times \sum_i (Y_i - \bar{Y})^2}} \quad (3)$$

- $X_i$  : conductivité électrique mesurée ( $\mu S/cm$ ) ;
- $Y_i$  : conductivité électrique simulée ( $\mu S/cm$ ) ;
- $\bar{X}$  : conductivité électrique moyenne mesurée ( $\mu S/cm$ ) ;
- $\bar{Y}$  : conductivité électrique moyenne simulée ( $\mu S/cm$ ).

Dans le cadre de cette étude, deux variables sont considérées comme corrélées si leur coefficient de corrélation est supérieur ou égal à 0,60 (60%).

### 2.2.2. Nash-Sutcliffe

Le critère d'optimisation de Nash-Sutcliffe est généralement utilisé pour évaluer la performance des modèles. Dans le cas de notre étude, le critère est le rapport de la somme du carré des écarts entre la conductivité mesurée et celle simulée par les modèles et la somme des écarts entre la conductivité électrique mesurée et la moyenne des conductivités mesurées. Il est régi par l'équation suivante (équation 4):

$$\text{Nash - Sutcliffe} = 100 \times \left[ 1 - \frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right] \quad (4)$$

- $X_i$  : conductivité électrique mesurée ( $\mu S/cm$ ) ;
- $Y_i$  : conductivité électrique simulée ( $\mu S/cm$ ) ;
- $\bar{X}$  : conductivité électrique moyenne mesurée ( $\mu S/cm$ ).

Le critère de Nash-Sutcliffe mesure la performance du modèle par rapport à un modèle qui utiliserait la valeur moyenne comme valeur simulée. Le critère de Nash-Sutcliffe varie entre  $-\infty$  et 100%. Un critère de Nash-Sutcliffe compris entre 0 et 100% signifie que le modèle est meilleur que si la moyenne avait été utilisée comme valeur simulée. Un critère de Nash-Sutcliffe négatif signifie que la valeur moyenne est plus performante que le modèle testé. Le modèle est considéré comme performant quand les conductivités estimées se rapprochent de celles mesurées sur le terrain, c'est-à-dire quand la valeur du critère de Nash-Sutcliffe est proche de 100%. Le critère de Nash-Sutcliffe varie entre  $-\infty$  et 100% et l'échelle suivante peut être utilisée pour apprécier la qualité de la simulation:

- Nash-Sutcliffe  $\geq 90\%$  , le modèle est excellent ;
- Nash-Sutcliffe compris entre 80% et 90%, le modèle est très bon ;
- Nash-Sutcliffe compris entre 60% et 80%, le modèle est bon ;
- Nash-Sutcliffe  $\leq 60\%$  , le modèle est mauvais.

En effet, une valeur du critère de moins de 60% ne donne pas une concordance satisfaisante entre les grandeurs observées et simulées par les modèles. En effet, les performances en terme de critère de Nash-Sutcliffe sont l'image de l'adéquation du modèle et du jeu de paramètres calés de l'échantillon étudié.

Le seuil admis qui est de  $S=0,6$  ou 60% a été testé. Pour cela, le test de nullité a été appliqué. La démarche méthodologique suivie se présente comme suit:

- i) déterminer la valeur du critère t ou fonction discriminante qui s'exprime comme suit (équation 5):

$$t = \sqrt{n-2} \frac{r}{\sqrt{1-r^2}} \quad (5)$$

avec :

- n : taille de l'échantillon ;
- r : coefficient de corrélation.

ii) déterminer  $t'$  la valeur tabulée de  $t$  à partir de la table de Student. On admet une probabilité au dépassement ( $P$ ). La fonction discriminante  $t$  suit une loi de Student à  $n-2$  degrés de liberté ( $v=n-2$ ) ;

iii) comparer  $t$  et  $t'$  ;

iv) conclure : si  $t$  ( $t$  observé) <  $t'$  ( $t$  théorique) alors acceptation de l'hypothèse nulle sinon refus ou rejet de l'hypothèse nulle ( $H_0$ : les deux variables sont indépendantes ;  $H_1$ : les deux variables sont dépendantes ou corrélées).

Le seuil de probabilité au dépassement retenu est  $P_1=5\%$ . Le seuil retenu pour les critères de performance est de  $S=0,60$  (ou 60%). La taille de l'échantillon retenue est la taille minimale qui est de 42.

### 2.2.3. Evaluation de la qualité des modèles au moyen de représentations graphiques

Outre l'évaluation à l'aide d'un critère numérique, une analyse comparative des profondeurs observées aux profondeurs simulées a été réalisée. En effet, des tracés de diagrammes de dispersion des profondeurs simulées en fonction des profondeurs observées a aussi permis d'apprécier la qualité de la modélisation effectuée. En pratique, si la simulation était parfaite, c'est-à-dire, si chacune des valeurs simulées par les modèles était égale à la valeur observée, le nuage de points constitué serait aligné et confondu à la droite d'équation  $y = x$ . Cependant, comme la modélisation n'est pas parfaite, l'évaluation du point de vue qualitative de la performance des différents modèles a donc consisté à analyser et à apprécier la dispersion du nuage de points autour de la première diagonale.

### **2.3. Critère de robustesse des modèles**

L'une des techniques les plus utilisées, pour évaluer la robustesse d'un modèle dans les situations les plus proches possibles de celles dans lesquelles il est censé fonctionner, est la technique du double échantillon. Cette technique permet de tester l'adaptabilité du modèle quelle que soit sa complexité. La robustesse implique que les résultats pouvant être obtenus par simulation avec le modèle ne divergent pas lorsque la valeur des variables d'entrée est légèrement modifiée du fait des erreurs et des incertitudes qui leur sont associées, mais qu'ils sont sensibles aux variations des facteurs dont on cherche à estimer les effets. Le critère de robustesse des différents modèles est évalué par la différence du critère de Nash-Sutcliffe en calage et en validation. Lorsque la valeur du critère de robustesse est comprise entre -10% et +∞, le

modèle est dit robuste. La robustesse se définit par la formule suivante (équation 6):

$$R' = 100 \times (\text{Nash-Sutcliffe}_{\text{Validation}} - \text{Nash-Sutcliffe}_{\text{Calage}}) \quad (6)$$

## **Résultats et Discussion**

### **1. Résultats du test de nullité du seuil choisi**

Les résultats du test de validation du seuil choisi ( $S=0,60$  ou 60%) montrent que pour une probabilité  $P=0,05$  ; la valeur de  $t$  ( $t$  expérimental: 1,98) est supérieure à celle de  $t'$  ( $t$  théorique: 1,96), pour toute valeur du seuil supérieure ou égale à 0,3 ou 30%. On peut déduire que l'hypothèse nulle est rejetée et donc le seuil de  $S=0,6$  ou 60%, est valable du point de vue statistique.

### **2. Performance des modèles en calage et validation**

Le calage a permis de reproduire d'une façon très satisfaisante la conductivité électrique observée sur le terrain avec des coefficients de corrélations ( $P=5\%$ ) de  $R=0,985$  et de  $R=0,986$  correspondant respectivement au modèle 1 et au modèle 2 (Tableau 1). En outre, les valeurs du critère de Nash-Sutcliffe ( $P=5\%$ ) pour ces deux modèles sont de 96,851% pour le modèle 1 et de 97,104% pour le modèle 2. Ces valeurs qui sont toutes supérieures à 90% traduisent donc une excellente simulation de la conductivité électrique des eaux souterraines. Il ressort de l'analyse du tableau I que les deux modèles présentent en calage des performances presque identiques. Les résultats de la validation des différents modèles sont consignés dans le tableau I. Il est constaté de fortes corrélations entre les conductivités observées et celles simulées par les différents modèles. En effet, le modèle 1 a enregistré le niveau de corrélation le plus élevé (0,960) suivi du modèle 2 avec un coefficient de corrélation de 0,900. Le critère de Nash-Sutcliffe quant à lui, affiche une forte valeur de 87,61 % pour le modèle 1 et une valeur non satisfaisante de 19,18% pour le modèle 2.

Ces résultats expriment une bonne performance en validation du modèle 1 tandis que le modèle 2 apparaît non performant en validation.

**Tableau 1:** Performances des modèles en calage et validation

Critères	Calage		Validation	
	M-1	M-2	M-1	M-2
<b>Coefficient de corrélation</b>	0,985	0,986	0,960	0,900
<b>Nash-Sutcliffe</b>	96,851	97,104	87,610	19,180
<b>NB: M : modèle</b>				

Les représentations graphiques des résultats obtenus à partir des deux modèles ont permis d'apprécier la qualité des valeurs simulées (Figure 3).

Les résultats de l'analyse graphique confirment les résultats numériques. En effet, le graphe du modèle 1 (Figure 3)



présente des points qui se regroupent généralement autour de la première bissectrice pour des valeurs de conductivité inférieures à 600  $\mu\text{S}/\text{cm}$ . Au-delà de cette valeur, les points s'éloignent de la première bissectrice et passe en dessous de celle-ci, ce qui traduit une surestimation de la conductivité électrique. Cela s'explique par les difficultés qu'éprouve le modèle à simuler des valeurs élevées de la conductivité électrique. Vu que la majorité (83%) des valeurs des données de validation demeurent inférieure à 600  $\mu\text{S}/\text{cm}$ . La qualité du graphe de la figure 3 laisse voir une mauvaise simulation de la conductivité électrique par le modèle 2 en validation. En effet, la quasi-totalité des points du graphique restent en dessous de la première bissectrice. Cela explique une surestimation des valeurs de la conductivité par le modèle 2 surtout pour les valeurs fortes.

### 3. Performances des modèles en transposition

La transposition des modèles dans plusieurs zones tests de la Côte d'Ivoire a donné des résultats consignés dans le tableau 2. Au niveau du modèle 1, les conductivités électriques simulées et celles mesurées présentent de très fortes corrélations allant de 0,90 (Guiglo-Duekoué) à 0,99 (Bondoukou). Quant au modèle 2, les conductivités électriques simulées et celles mesurées présentent de très fortes corrélations allant de 0,93 (Tiébissou-Didiévi) à 0,99 (Bondoukou) sauf la région de Guiglo-Duekoué qui affiche une faible valeur de 0,19. Les deux modèles présentent des performances presque similaires au regard du coefficient de corrélation.

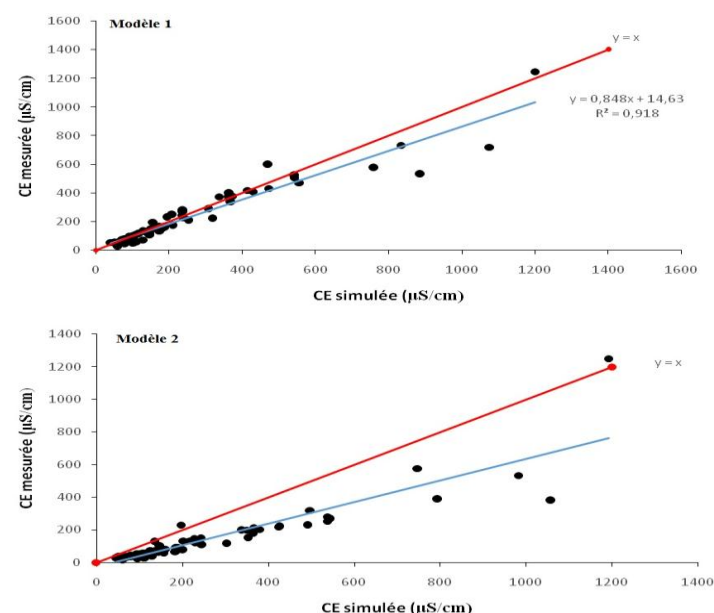


Fig. 3: Conductivités mesurées et simulées par les modèles 1 et 2 en validation

Des valeurs satisfaisantes voire très satisfaisantes du critère de Nash-Sutcliffe (78,23%-96,81%) ont été obtenues par les deux

modèles. Cependant, des valeurs non satisfaisantes ont été obtenues par les deux modèles au niveau de Guiglo-Duekoué. En effet, le modèle 2 a enregistré une performance négative (-826,84%) quand le modèle 1, lui a obtenu une performance de 29,48%.

Les valeurs de Nash-Sutcliffe sont généralement plus élevées au niveau du modèle 2. Ces résultats traduisent une simulation satisfaisante voire très satisfaisante de la conductivité électrique dans les différentes zones tests en dehors de Guiglo-Duekoué par les deux modèles au regard du critère de Nash-Sutcliffe à P=5%.

Tableau 2: Performances des modèles en transposition

	Corrélation		Nash-Sutcliffe	
	M-1	M-2	M-1	M-2
Tiébissou-Didiévi	0,94	0,93	80,77	81,21
Agboville	0,92	0,94	78,23	94,65
Tiassalé	0,97	0,96	92,94	96,81
Bondoukou	0,99	0,99	91,82	95,07
Guiglo-Duekoué	0,90	0,19	29,48	-826,84

NB: M : modèle

Les représentations graphiques des cas de Tiassalé et Bondoukou qui ont enregistré les valeurs les plus élevées du coefficient de corrélation et du critère de Nash-Sutcliffe sont illustrées aux figures 4 et 5.

Les graphes présentent des points regroupés autour des courbes de tendance et de la première bissectrice ; ce qui montre une bonne simulation des conductivités électriques en transposition.

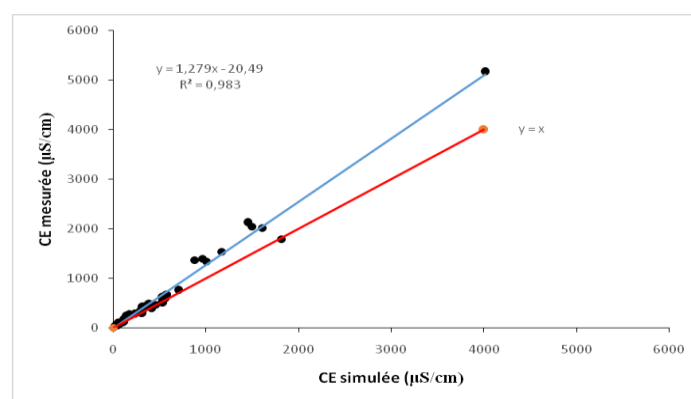
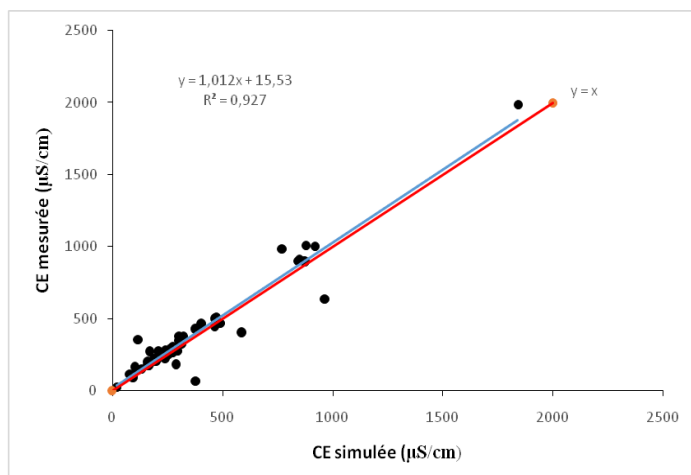


Fig. 4: Conductivités mesurées et simulées par le modèle 1 en transposition à Bondoukou



**Fig. 5:** Conductivités mesurées et simulées par le modèle 2 en transposition à Tiassalé

#### 4. Résultats de robustesse des modèles

Les résultats de robustesse des modèles sont consignés dans le tableau 3. En validation, la valeur du critère de robustesse est supérieure à -10% pour le modèle 1 (-9,24%) et largement inférieure au seuil d'acceptation de -10% pour le modèle 2 (-77,92%). En transposition, les valeurs du critère de robustesse vont de -67,37 à -3,91% pour le modèle 1 avec deux valeurs supérieures à -10% (Tiassalé et Bondoukou). Quant au modèle 2, les valeurs du critère de robustesse vont de -923,94 à -0,29% avec trois valeurs supérieures à -10% (Tiassalé, Bondoukou et Agboville). Ces résultats montrent qu'en validation, le modèle 1 a été plus robuste que le modèle 2, qui a manqué de robustesse. En transposition, les résultats montrent que le modèle 1 est plus robuste que le modèle 2. En somme, le modèle 1 est plus robuste que le modèle 2.

**Tableau 3:** Résultats de robustesse des modèles

Localités	Modèle 1 (%)	Modèle 2 (%)
N'zi-Comoé	-9,241	-77,924
Tiébissou-Didiévi	-16,081	-15,894
Tiassalé	-3,911	-0,294
Bondoukou	-5,031	-2,034
Guiglo-Duekoué	-67,371	-923,944

#### 5. Discussion

Pour les deux modèles, de fortes valeurs des critères de performances ont été généralement obtenues en calage, en validation interne et en validation par transposition spatiale. Cependant, en validation, une faible valeur de Nash-Sutcliffe ( $N=19,18\%$ ) a été obtenue par le modèle 2. En transposition, une faible valeur de coefficient de corrélation a été obtenue ( $R=0,19$ ) par le modèle 2 à partir des données de Guiglo-

Duekoué. De faibles valeurs de Nash-Sutcliffe ont été également obtenues par les deux modèles (modèle 1: 29,48% ; modèle 2: -826,84%) au niveau de Guiglo-Duekoué. Au regard du test de nullité, les mauvaises performances obtenues en termes de coefficient de corrélation et de Nash-Sutcliffe, ne sont pas significatives statistiquement.

En calage, il est constaté une performance presque similaire au niveau des deux modèles quel que soit le critère utilisé. En validation, le modèle 1 est plus performant que le modèle 2. En transposition, la performance est quasi-identique au niveau des deux modèles en s'appuyant sur le coefficient de corrélation en dehors de l'échantillon de Guiglo-Duekoué. En considérant le critère de Nash-Sutcliffe, le modèle 2 se trouve plus performant que le modèle 1. Au regard des valeurs des critères de performance, les modèles montrent une grande capacité à simuler la conductivité électrique. Le modèle 2 a démontré une grande capacité au plan de sa transposabilité. L'ensemble des résultats montrent que les mauvaises performances ont été obtenues au niveau de l'échantillon de Guiglo-Duekoué. L'analyse des résultats montre que le critère de Nash-Sutcliffe est plus habilité à analyser voir comparer les performances des différents modèles. Sur cette base, une analyse de la robustesse a été réalisée. Les résultats de robustesse des différents modèles montrent qu'en validation, le modèle 1 a été robuste tandis que le modèle 2 a manqué de robustesse. En transposition, le modèle 2 a été plus robuste que le modèle 1. La qualité d'un modèle est d'abord sa performance et vient ensuite la robustesse. Elle s'apprécie également à partir de la capacité du modèle à simuler des valeurs proches des valeurs observées dans des zones en dehors de celle sur laquelle il a été conçu. Ainsi, en s'appuyant sur les résultats de la transposition, on peut dire que le modèle 2 est plus performant et plus robuste que le modèle 1. Il ressort donc que l'ajout du pH et de la température aux deux premières variables a mis en évidence une amélioration du modèle. En effet, la conductivité électrique des eaux varie avec la température et le pH.

Manssouri et al. [9], ont construit un modèle de prédiction basé sur les réseaux de neurones artificiels de type perceptron multicouches pour prédire des indicateurs (conductivité électrique, etc.) de la qualité des eaux souterraines de la nappe de Souss-Massa-Maroc, à partir d'un certain nombre de variables optimisées par la méthode de la régression linéaire multiple, le coefficient de détermination obtenu du modèle RNA-MLP est de l'ordre de 98%.

Cette performance se trouve dans l'ordre de celle obtenue au cours de ce travail.

L'une des applications de la modélisation de la conductivité électrique est l'étude de la fiabilité des données physico-chimiques. En effet, à l'instar de la balance ionique, l'étude de la corrélation entre les conductivités électriques observées sur le terrain et celles simulées est un moyen d'identification des erreurs commises lors de l'analyse des données physico-chimiques. Ainsi, les travaux de Kouassi et al. [6] sur l'étude de

la fiabilité des données des différentes zones tests a mis en évidence une balance ionique équilibrée (inférieure à 10%) pour les zones tests de Tiébissou-Didiévi, de l'ex-région du N'zi-Comoé, d'Agboville et de Tiassalé. Cependant, les échantillons des zones tests de Bondoukou et de Guiglo-Duekoué ont présenté des balances ioniques supérieures au seuil de 10% donc non équilibrées. Ainsi, tous les échantillons de balance équilibrée présentent de bonnes valeurs au niveau des critères de performance tandis que l'échantillon de la zone test de Guiglo-Duekoué qui selon Kouassi et *al.* [6] a présenté la plus mauvaise balance (22,13%), affiche également les plus mauvaises valeurs de tous les critères en calage, en validation interne et en transposition. Cependant, l'échantillon de la zone de Bondoukou ayant une balance non équilibrée présente des valeurs très satisfaisantes des critères de performance. Ce constat met en évidence une non concordance des résultats basés d'une part sur l'analyse de la balance ionique et d'autre part, sur la simulation de la conductivité électrique. Kouassi et *al.* [6] sont parvenus à la même conclusion au cours de leurs travaux à partir d'un modèle qui définit la conductivité électrique (CE) comme le rapport du résidu sec (mg/l) au coefficient 0,65 ( $CE = \text{résidu sec} / 0,65$ ). On peut dire donc que la mauvaise performance enregistrée au niveau de l'échantillon de Guiglo-Duekoué est due à la mauvaise qualité des données issues des analyses des laboratoires. Qu'en est-il de l'échantillon de Bondoukou ? Est-ce une limite de l'étude de la fiabilité des résultats d'analyses physico-chimiques à partir de la modélisation ? A quel seuil peut-on dire que l'échantillon est fiable ? L'analyse des critères de performance a déjà démontré la limite du coefficient de corrélation à travers une incapacité à mettre en relief une différence significative en termes de performance. En s'appuyant sur le critère de Nash-Sutcliffe, on pourrait retenir le seuil de 60% pour lequel une performance qui lui est supérieure traduirait une balance équilibrée. Le modèle 1, bien que performant, robuste et parcimonieux présente des limites quant à l'étude de la fiabilité des données d'analyses physico-chimiques. Ainsi, le modèle 2 se présente comme le modèle le plus adapté à l'étude de la fiabilité des données physico-chimiques. Cela permet d'affirmer que la pertinence de l'analyse de la fiabilité des résultats d'analyses physico-chimiques des eaux souterraines par la modélisation n'est plus pertinente que si le modèle tient compte des paramètres physiques influençant la conductivité électrique des eaux

## Conclusion

Deux modèles statistiques de simulation de la conductivité électrique des eaux souterraines ont été évalués en termes de performance et de robustesse en phases de calage, de validation interne et de validation par transposition spatiale. Plusieurs critères ont été utilisés (coefficient de corrélation, critère de Nash-Sutcliffe, critère de robustesse). Pour les deux modèles, de fortes valeurs des critères de performances ont

été généralement obtenues en calage, en validation interne et en validation par transposition spatiale traduisant une grande capacité pour les modèles à simuler la conductivité électrique des eaux souterraines. Les deux modèles sont donc performants et robustes. Mais, le modèle 2 est plus performant et plus robuste que le modèle 1. Il ressort donc que l'ajout du pH et de la température aux deux premières variables a mis en évidence une amélioration du modèle. En effet, la conductivité électrique des eaux varie avec la température et le pH. Les échantillons de balance équilibrée présentent de bonnes performances. L'ensemble des résultats montrent que les plus mauvaises performances ont été obtenues au niveau de l'échantillon de Guiglo-Duekoué qui a une balance non équilibrée. Cependant, l'échantillon de la zone de Bondoukou ayant une balance non équilibrée présente des valeurs très satisfaisantes des critères de performance. Ce constat présente une non concordance des résultats basés d'une part sur l'analyse de la balance ionique et d'autre part, sur la simulation de la conductivité électrique. Le modèle 2 se présente comme le modèle le plus adapté à l'étude de la fiabilité des données physico-chimiques avec le critère de Nash-Sutcliffe comme le critère le plus adapté. En s'appuyant sur le critère de Nash-Sutcliffe, on pourrait retenir le seuil de 60% pour lequel une performance qui lui est supérieure traduirait une balance équilibrée.

## Références Bibliographiques

1. M.J Diamantopoulou, V.Z Antonopoulos, D.M Papamichail. *The use of a neural technique for the prediction of water quality parameters of Axios river in Northern Greece*. EWRA European water, **2005**, 11, 12, 55.
2. A.B Sebaiti. *Gestion optimisée des ressources en eau d'une nappe côtière : Application à la plaine d'Annaba (Nord-Est Algérie)*. Doctorat de l'Université de Lille, Sciences et Technologies, France, **2005**, 205.
3. A Abdallaoui, H El badaoui, L Lancelot. *Développement d'un modèle statistique neuronal pour la prédiction des concentrations des métaux lourds (Fe, Mn et Al) dans les eaux superficielles : cas de l'Oued Boufekrane, (Meknes, Maroc)*. Actes du 1<sup>er</sup> Colloque international sur « L'eau et le climat » à Rouen, du 25 au 26 septembre 2012, **2012**, 18.
4. N Zerki, R Bchitou, A Bouhaouss. 1<sup>er</sup> Colloque international REZAS12, *Ressources en eau dans les zones arides et semi-arides. Enjeux et perspectives. Cas du continent africain*, Béni-Mellal, Maroc, **2012a**, (14, 15 et 16 Novembre 2012).
5. N Zerki, A Bouhaouss, R Bchitou. *Modeling and Optimization the Effect of Some Anions on the Water Conductivity*. Australian Journal of Basic and Applied Sciences, **2012b**, 6, 10, 396.
6. A.M Kouassi, A Mamadou, K.E Ahoussi, J Biemi. *Simulation de la conductivité électrique des eaux souterraines en relation avec leurs propriétés géologiques: cas de la Côte*



- d'Ivoire. Revue Ivoirienne des Sciences et Technologie, **2013**, 21&22, 138.
7. A.M Kouassi, A Mamadou, K.E Ahoussi, J Biemi. *Conception de modèles statistiques à variables hydrochimiques pour la prédiction de la conductivité électrique des eaux souterraines*. Larhyss Journal, **2014**, 20, 189.
  8. T Manssouri, H Sahbi, I Manssouri. *Elaboration of stochastic mathematical models for the prediction of parameters indicative of groundwater quality Case of Souss Massa-Morocco-IJCER*. International Journal of Computational Engineering Research, **2014**, 4, 5, 31.
  9. T Manssouri, H Sahbi, I Manssouri, B Boudad. *Utilisation d'un modèle hybride base sur la RLMS et les RNA-PMC pour la prédiction des paramètres indicateurs de la qualité des eaux souterraines: cas de la nappe de SOUSS-MASSA-MAROC*. European Scientific Journal, **2015**, 11, 18, 35.
  10. N Zerki, K Achelhi, R Bchitou, A Bouhaouss. *Modeling and optimization the conductivity of wastewater discharged into the river Bouregreg (Rabat, Morocco) in function of ions: chlorides, sulphates, nitrates and bicarbonates*. Journal of Industrial Engineering Research, **2015a**, 1, 3, 35.
  11. N Zerki, K Achelhi, R Bchitou, A Bouhaouss. *Correlation study between Nitrate, Nitrite, Chloride and Sulfate in Natural Waters of Bouregreg River (Rabat, Morocco)*. Journal of Industrial Engineering Research, **2015b**, 1, 4, 129-135.
  12. T.Y Brou. Mémoire de Doctorat d'Etat, Université de Cocody, Abidjan, **2005**, 226p.
  13. A.N Kouamelan. Thèse Doc. Univ. Rennes I, Mémoires, Géosciences Rennes, **1996**, 73, 289.
  14. T Lasm, F Kouamé, M.S Oga, J.P.R Jourda, N Soro, H.B Kouadio. Revue Ivoirienne des Sciences et Technologie, **2004**, 5, 97, 1.
  15. J.P.R Jourda, E.V Djagoua, K. Kouamé, M.B Saley, C. Gronayes, J.J Achy, J Biemi, M Razack. Télédétection, **2006a**, 6, 2, 123.
  16. J.P.R Jourda, M.B Saley, E.V Djagoua, K.J Kouamé, J Biemi, M Razack. Télédétection, **2006b**, 5, 4, 339.
  17. T Lasm, M Youan-Ta, J.P.R Jourda, K.F Kouamé. European Journal of Scientific Research, **2008**, 21, 1, 196.
  18. M Youanta, T Lasm, J.P.R Jourda, K.F Kouame, M Razack. Télédétection, **2008**, 8, 2, 119.
  19. M Koita, H Jourde, D Ruelland, K Koffi, S Pistre, I Savané. Hydrological Sciences Journal, **2010**, 55, 5, 805.
  20. Y.A N'go, T Lasm, M Koita, I Savané. Télédétection, **2010**, 9, 1, 33.
  21. V.M Sorokoby, M.B Saley, K.F Kouamé, M.E.V Djagoua, M Bernier, K Affian, J Biemi. Télédétection, **2010**, 9, 3-4, 209.
  22. B Tagini. Thèse de Doctorat, Université de Lausanne, SODEMI, Abidjan-Côte d'Ivoire, **1971**, 302.
  23. M Belanger, N EL-Jabi, D Caissie, F Ashkar, J.M Ribí. *Estimation de la température de l'eau en rivière en utilisant les réseaux de neurones et la régression linéaire multiple*. Revue des Sciences de l'Eau, **2005**, 18, 3, 403.
  24. M Filion. *Modélisation de la qualité de l'eau et prévision des débits par la méthode des réseaux de neurones*. Mémoire présenté à la Faculté des études supérieures de l'Université Laval pour l'obtention du grade de Maître ès Sciences, **2007**, 109.