



Prédiction des concentrations des *Escherichia coli* des eaux de puits de M'pody / Prediction of *Escherichia coli* concentrations in the well water of M'pody

Meless Djedjro Franck Renaud^{1,3*}, Gbagbo Tchape Aubin¹, Kouassi-Agbessi Therese Brah¹, Kpaibe Sawa Andre Philippe^{1,2}, Amin N'cho Christophe^{1,2}

Reçu pour publication le 15 octobre 2022 - Avis favorable de publication en novembre 2022

Résumé

Contexte/objectif. Dans le village M'pody de la circonscription d'Anyama, une épidémie de diarrhée a été détectée en janvier 2020 et a affecté 69 personnes. Selon la population touchée, ces cas de diarrhée étaient liés à la consommation d'eau du système hydraulique villageois amélioré, qui n'était plus entretenu depuis près de trois ans. L'objectif de ce travail était de développer un modèle de caractérisation bactériologique des nappes phréatiques du village M'pody à partir des paramètres physicochimiques et de la météorologie pour estimer la concentration des *Escherichia coli* par puits.

Méthodes. La méthode a consisté à réaliser quatre campagnes d'échantillonnage d'eau pour chacun des 72 puits que comptaient cette région, soit un total de 288 visites. Le développement du modèle de prédiction de l'indicateur *E. coli* a été effectué selon un modèle linéaire mixte. L'évaluation des performances de notre modèle a été réalisée par des techniques de ré-échantillonnages que sont le bootstrap et la validation croisée k-fold.

Résultats. Le modèle prédictif explique 30,24 % de la variance des concentrations en *E. coli*. Il est basé sur 9 variables. La validation des performances du modèle par bootstrap a donné un très faible biais relatif < 5%, des erreurs de prédiction moyenne et des erreurs de prédictions absolues par K-fold inférieur à 2,5.

Conclusion. La validation des performances prédictives par K-fold et bootstrap a montré que les prédictions du modèle sont précises pour une utilisation future.

Mots clés :

-Modèle linéaire mixte;
- Bootstrap;
-Analyse en composantes Principale

ABSTRACT

Background/Purpose: In the village of M'pody in the Anyama district, an outbreak of diarrhoea was detected in January 2020 and affected 69 people. According to the affected population, these cases of diarrhoea were linked to the consumption of water from the improved village water system, which had not been maintained for almost three years. The objective of this work was to develop a bacteriological characterisation model for the water tables of the village of M'pody based on physicochemical parameters and meteorology to estimate the concentration of *Escherichia coli* per well.

Methods: The method consisted of four water sampling campaigns for each of the 72 wells in the area, for a total of 288 visits. The development of the *E. coli* indicator

1-Institut National d'Hygiène Publique, Abidjan, Côte d'Ivoire

2- Département de Chimie Analytique - Bromatologie, Chimie générale, Chimie Minérale, UFR des Sciences Pharmaceutiques et Biologiques, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

3- Département Physique, Biophysique, Mathématiques, Statistiques Et Informatique, UFR des Sciences Pharmaceutiques et Biologiques, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

*Correspondant : Meless Djedjro Franck Renaud- melessrenaud@gmail.com

prediction model was carried out using a linear mixed model. The performance of our model was evaluated using bootstrap and k-fold cross-validation techniques. Results: The predictive model explained 30.24% of the variance in *E. coli* concentrations. It is based on 9 variables. Validation of the model's performance by bootstrapping gave a very low relative bias < 5%, mean prediction errors and absolute prediction errors by K-fold less than 2.5.

Conclusion: Validation of the predictive performance by K-fold and bootstrap showed that the model predictions are accurate for future use

Keywords:

-Linear Mixed Model;
-Bootstrap;
-Principal Component Analysis

INTRODUCTION

L'eau est l'un des éléments essentiels de la vie. Pour sa consommation humaine, elle doit être exempte de tout agent pathogène qui présente un danger potentiel pour la santé humaine. L'accès à l'eau potable constitue un problème majeur dans les pays en développement comme la Côte d'Ivoire et d'autres pays d'Afrique. En effet, en 2010, 783 millions de personnes n'avaient pas accès à l'eau potable provenant de sources améliorées^[1]. Néanmoins, parmi les sources d'approvisionnement en eau potable, les eaux souterraines sont traditionnellement les ressources en eau privilégiées pour l'eau potable, car elles sont plus à l'abri des pollutions que les eaux de surface^[2]. En Côte d'Ivoire dans les zones rurales comme urbaines certaines populations utilisent de l'eau de consommation provenant généralement de la nappe phréatique notamment de puits traditionnels. Cependant, ces aquifères peuvent être pollués par les activités anthropiques et être responsables de maladies hydriques^[3,4]. La pollution des eaux souterraines représente l'un des aspects les plus inquiétants et l'utilisation de ces eaux à des fins alimentaires représente un danger pour la santé^[5]. La consommation d'une eau contaminée par les microorganismes est à l'origine des épidémies^[6,7].

Depuis ces dernières décennies, les études réalisées sur les ressources en eau de la Côte d'Ivoire par Ahoussi et al. (2013), Traoré et al. (2015), Ouattara et al. (2016), Soro et al. (2010)^[8-11] avaient signalé de nombreux foyers de pollution aussi bien au niveau des eaux de surface que des eaux souterraines,

dans les localités d'Abidjan que dans les zones rurales. En janvier 2020, à M'pody, village situé à une soixantaine de kilomètres de la ville d'Anyama, une épidémie de diarrhée est survenue faisant environ 69 victimes, en majorité des enfants de 0 à 5 ans. Cette épidémie serait due à la consommation de l'eau provenant de l'hydraulique améliorée villageoise^[12].

En Côte d'Ivoire, l'Institut National d'Hygiène Publique est la structure habilitée à contrôler la qualité de l'eau de consommation à travers le laboratoire d'hygiène. Ce contrôle passe par la caractérisation microbiologique et physicochimique. Les résultats de l'analyse de la majorité des paramètres physicochimiques sont disponibles dans un très court laps de temps alors que ceux de la microbiologie nécessitent un temps plus long souvent plus de 24 heures^[13]. Compte tenu de ces délais, l'eau aura déjà été consommée par la population lorsque les résultats des analyses deviennent disponibles. Par ailleurs, la plupart des normes sur la qualité microbiologique de l'eau s'appuie sur les coliformes fécaux en principalement sur *Escherichia coli*. Il nous a paru intéressant de développer des outils de détection du risque sanitaire lié à ce paramètre avant la confirmation microbiologique 24 à 48 heures. Cet outil pourrait servir pour une utilisation en cas de déclaration de maladie hydrique dans cette localité et permettra de lancer une intervention précoce telle que la fermeture des puits ou la désinfection de l'eau par l'hypochlorite de sodium avant la disponibilité des résultats de laboratoires.

METHODES

Collecte et mise en forme des données disponibles

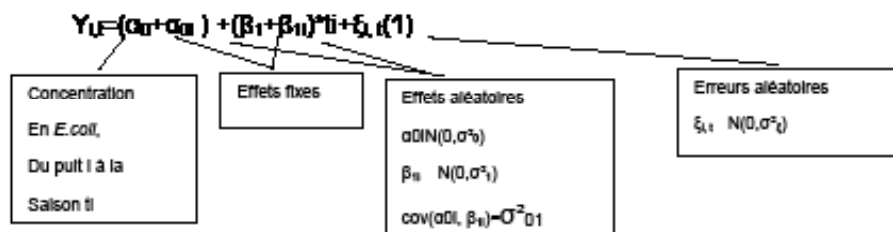
Il s'agissait d'une étude écologique de type cohorte rétrospective réalisée sur l'ensemble des 72 puits que comptait la localité de M'pody au cours des quatre saisons de l'année 2020

(grande saison sèche (février) et pluvieuse (juin), petite saison sèche (août) et pluvieuse (octobre). Quatre prélèvements ont été réalisés sur chacun des 72 puits. Ces prélèvements ont été effectués pour chaque puits à une date bien précise pour chaque saison. Les échantillons ont été rangés dans une glacière contenant des accumulateurs

de froid, à l'abri de la lumière, à une température comprise entre 4° C et 8° C et transportés au laboratoire en respectant la chaîne de froid. Les paramètres physico-chimiques classiques ont été déterminés à l'aide de méthodes électrochimiques et spectrophotométriques. L'analyse microbiologique a été réalisée par la technique de filtration sur membrane puis ensemencement sur milieu spécifique. Concernant l'inspection visuelle un questionnaire a été administré aux chefs de ménage par les inspecteurs pour l'analyse des sources de pollution potentielle. La mise en forme des données a consisté à ramener toutes ces variables sur une base comparable. Nous avons également exploité les données météorologiques de la ville d'Anyama pour l'année 2020 enregistrées par la Société d'exploitation et de développement aéroportuaire, aéronautique et météorologique^[14]. Les variables indépendantes disponibles pour expliquer la variable dépendante concentration d'indicateurs fécaux incluaient 21 paramètres chimiques, 3 paramètres physiques, 9 paramètres météorologiques et 30 variables d'inspection sanitaire explicatives.

Analyses statistiques

Le développement de ce modèle a été réalisé en trois étapes :



Validation des performances des modèles

Pour évaluer les performances de nos modèles nous avons utilisé des techniques de ré-échantillonnages que sont le bootstrap et la validation croisée K-fold.

Validation Bootstrap

Les étapes spécifiques de la validation Bootstrap de nos modèles sont les suivantes [15] :

- Générer un échantillon bootstrap en rééchantillonnant à partir des données et/ou du modèle estimé
- Obtenir les estimations pour tous les paramètres du modèle pour l'échantillon bootstrap
- Répétez les étapes 1 et 2 B (1000) fois pour obtenir la distribution bootstrap des estimations des paramètres, puis calculez

- Choix des variables explicatives pour alimenter le modèle
- Mise en place du modèle linéaire mixte
- Validation des performances du modèle

Sélection des variables

Une analyse bivariée sur l'ensemble des variables a été réalisée en fixant un seuil de p-valeur de 20 % dans le choix des variables explicatives. Par la suite le critère d'Akaike dans la procédure pas à pas descendante nous a aidé dans le choix définitif des variables explicatives. La revue de la littérature a complété le choix des variables explicatives.

Modèles linéaires à effets mixtes

Nous avons réalisé un spaghetti plot pour choisir le modèle soit à pente aléatoire ou à intercept aléatoire ou à pente et intercept aléatoire. Transformer la variable concentration en *E. coli* en logarithme si possible. Le modèle pourrait s'écrire s'il est à intercept et pente aléatoire sous cette forme :

la moyenne, l'écart-type et l'IC à 95 % de cette distribution.

Soit le paramètre estimé pour le ième échantillon bootstrap. Étant donné un ensemble de données, la valeur attendue de l'estimateur bootstrap sur la distribution bootstrap est calculée comme la moyenne des estimations de paramètres provenant des B (1000) échantillons bootstrap

$$\bar{\theta}_b^* = \frac{1}{B} \sum_{b=1}^B (\theta_b^*) \quad (2)$$

L'écart type bootstrap est obtenu comme l'écart-type de l'échantillon du paramètre θ_b^*

$$SE_{\theta_b^*} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\theta_b^* - \bar{\theta}_b^*)^2} \quad (3)$$

Une autre approche consiste à utiliser une normale pour construire un IC bootstrap, en utilisant l'estimation du SE_b

$$\widehat{\theta}_b - \widehat{SE}_b * Z_{1-\alpha/2} < \theta < \widehat{\theta}_b + \widehat{SE}_b * Z_{1-\alpha/2} \quad (4)$$

$Z_{1-\alpha/2}$ désigne le quantile de la distribution normale standard

Le biais relatif de l'estimation asymptotique sera obtenu en comparant l'estimation asymptotique et la valeur réelle θ_0 comme

$$RBiais(\theta_0) = \frac{1}{n} \sum_{b=1}^n \left(\frac{\widehat{\theta}_b - \theta_0}{\theta_0} * 100 \right) \quad (5)$$

Validation K fold

La validation k-fold consiste à diviser l'échantillon original en k (5) échantillons (ou « blocs »), puis on sélectionne un des 5 échantillons comme ensemble de validation pendant que les k-1 autres échantillons constituent l'ensemble d'apprentissage. Après apprentissage, on peut calculer une performance de validation. Puis on répète l'opération en sélectionnant un autre échantillon de validation parmi les blocs prédéfinis. À l'issue de la procédure nous obtenons ainsi 5 scores de performances, un par bloc. Un certain nombre de métriques d'ajustement du modèle seront calculé comme le R^2 , RMSE, MAE.

RÉSULTATS

Sélection des variables

La sélection des variables en analyse bivariée a permis de retenir un ensemble de variable explicatives que sont la turbidité, pH, la vitesse du vent de la veille et du jour du prélèvement, la température minimale et maximale de la veille et du jour du prélèvement, la conductivité, la température de l'eau, les concentrations en nitrates, en chlore, en calcium, en sodium, en fer, en manganèse, en phosphate, en Titre Alcalimétrique Complet, en bicarbonate, la précipitation de la veille et du jours de prélèvement. Ce sont les 12 variables citées plus haut au niveau du clustering. Le critère d'Akaike a permis de retenir 9 variables associées à la variable log *E.coli*

Modèles linéaires à effets mixtes

Un spaghetti plot présenté à la Figure 1 et le test de rapport de vraisemblance avec les estimations basées sur l'approche du maximum de vraisemblance restreinte ($p > 0,05$) permet de choisir un modèle linéaire à intercept aléatoire.

Le coefficient de détermination (R^2) est le pourcentage de la variation totale de la variable de réponse expliquée par la droite de régression. Sa formule est :

Récupération	réf	
Recyclage	0,004	0,25
Rénovation	0.002	0.30

est la somme des différences au carré entre la valeur

prédite et observée, (8) est la somme des différences au carré entre la valeur moyenne et observée,

L'erreur quadratique moyenne (RMSE) est l'erreur de prédiction moyenne (racine carrée de l'erreur quadratique moyenne).

$$489 \quad 0,71 \quad 0,28 - 1,8$$

(9)

L'erreur absolue moyenne (MAE) est l'erreur de prédiction absolue moyenne. Il est moins sensible aux valeurs aberrantes.

réf

(10)

Les logiciels utilisés pour l'analyse de nos données sont le logiciel R version 4.1.1.1. Le seuil de significativité est de 5 %.

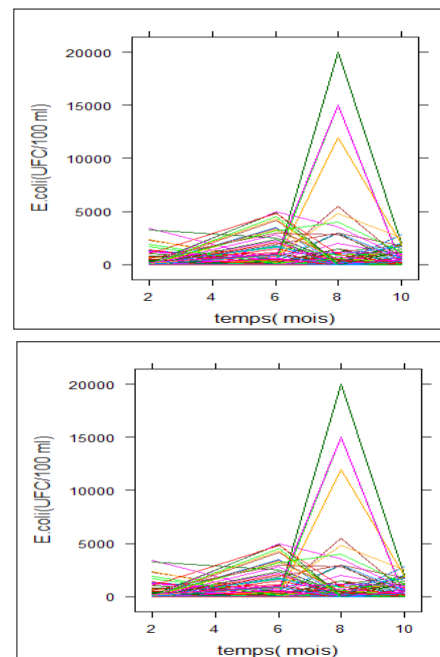


Fig. 1 : Evolution de la concentration en E. coli puis logE. coli pour chaque puits au cours du temps / Evolution of E. coli concentration and then logE. coli for each well over time

Le modèle prédictif explique 30,24 % de la variance des concentrations en *E. coli* (transformation logarithmique). Il est basé sur la turbidité, le pH, le fer, le chlore, le sodium, la température minimale et maximale du jour de prélèvement, les précipitations précédant la journée de prélèvement, la présence de latrines à moins de 15 m du puit. La variabilité inter-individuelle de la concentration en *E.coli* au temps 0 est $\sigma^2 = 0,85$. Le modèle retenu s'écrit sous la forme suivante :

$$\text{Log}(E. coli+1)_{ij} = (0.65 + \alpha_{0i}) + 0.02 * (\text{Turbidité}_{ij}) + 0.90 * (\text{pH}_{ij}) + 0.07 * (\text{Fer}_{ij}) - 0.04 * (\text{Cl}_{ij}) + 0.04 * (\text{Na}_{ij}) - 0.12 * (\text{Température min J}_{ij}) + 0.05 * (\text{Température max J}_{ij}) + 0.6 * (\text{Précipitations du jour 1}_{ij}) + 0.25 * (\text{Latrine à moins de 15 m}_{ij}) + \xi_{ij}$$

$$\forall i \in [[1 ; 72]], \forall tij \in [[2 ; 6 ; 8 ; 10]] \quad (11)$$

Validation des performances des modèles

La validation des performances passe par la vérification des conditions d'application d'un modèle linéaire mixte que sont la normalité des erreurs aléatoires et résiduelles et

l'homoscédasticité. Le log des concentrations en *E. coli* a l'air d'évoluer linéairement avec le temps (Figure 2). Il y a une homoscédasticité marquée car la variance du log de la concentration semble ne pas trop augmenter avec le temps. Le plot des résidus en fonction des valeurs prédites par le modèle est utile pour détecter une hétéroscédasticité. On peut conclure à l'homoscédasticité. Le qqplot montre des points qui ondulent autour d'une droite : l'hypothèse de normalité des résidus semble ici raisonnable.

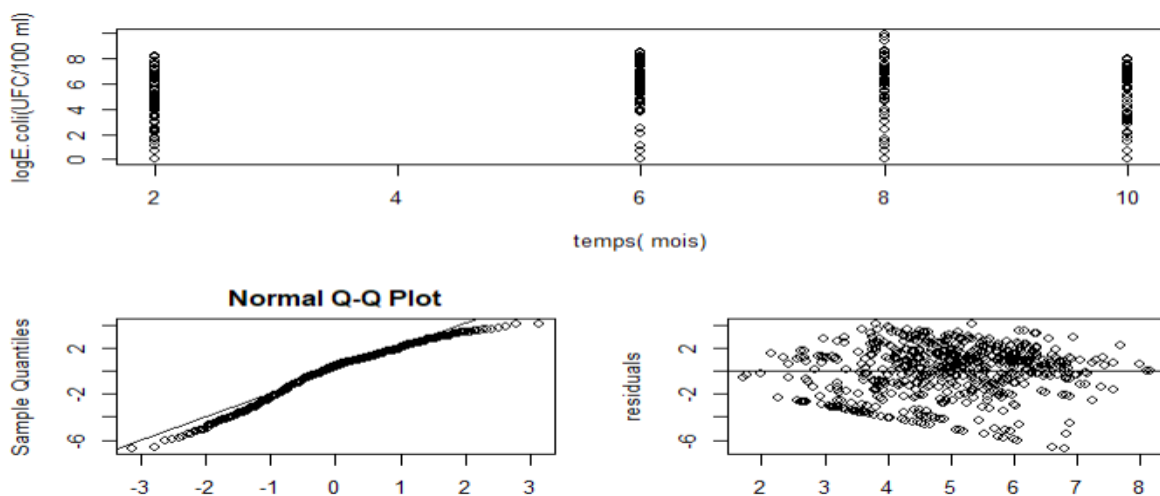


Fig. 2 : Vérification des hypothèses des résidus validité du modèle logE .coli / Residual hypothesis testing validity of the logE .coli model

Validation bootstrap

Le tableau I suivant indique les paramètres du modèle sur la donnée empirique, les moyennes de réplication, les erreurs standards puis les biais

relatifs pour chaque paramètre des variables explicatives.

Tableau I : Estimation des performances des modèles par validation bootstrap/ Estimation of model performance by bootstrap validation

Paramètres	Observées	Moyennes des réplifications	Erreurs standards	Biais relatifs
Intercept	0,6549	0,6870	1,4928	4,9
Turbidité	0,0208	0,0209	0,0089	0,3384
pH	0,9022	0,8940	0,2763	0,9146
Fer	0,0705	0,0700	0,0301	0,6582
Cl	-0,0460	-0,0462	0,0113	0,3998
Na	0,0436	0,0438	0,0239	0,4506
Température min jj	-0,1193	-0,1188	0,0504	0,4247
Température max jj	0,0480	0,0479	0,0411	0,0659
Précipitation du j-1	0,6033	0,6042	0,2493	0,1356
Latrine à moins de 15 m	0,2517	0,2519	0,3251	0,0901

Validation kfold

le R^2 conditionnel moyen est de 0,31, l'erreur de prédiction moyenne (RMSE) est de 1,95 et l'erreur de prédiction absolue moyenne(MAE) est de 1,6 (Fig. 3)

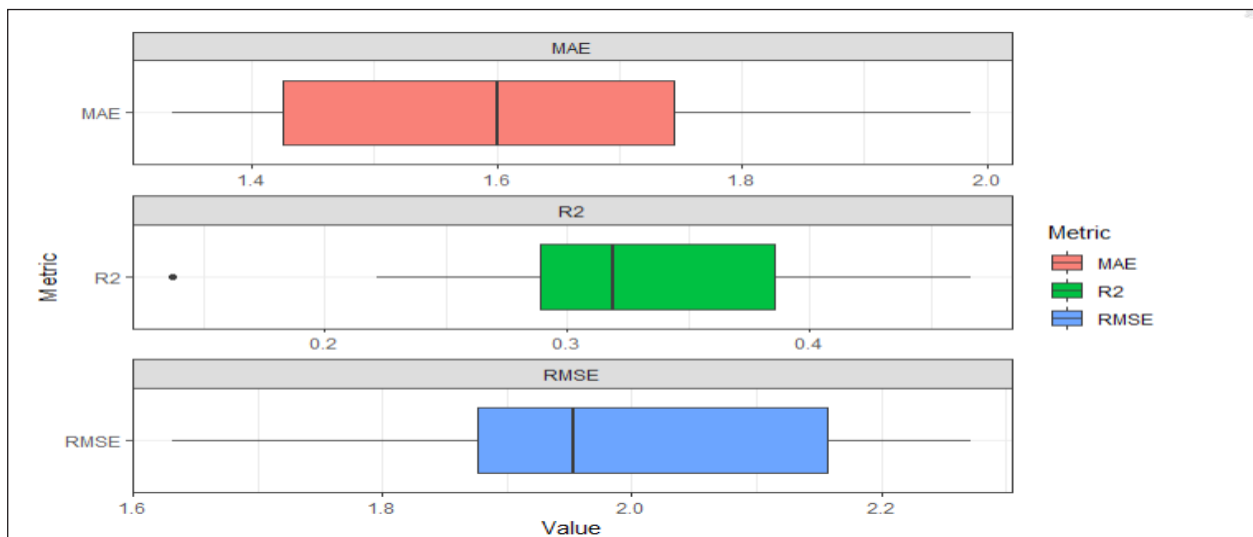


Figure 3 : Validation des performances du modèle logE.coli par validation k-fold/ Performance validation of the logE.coli model by k-fold validation

DISCUSSION

A M'pody, village de la circonscription d'Anyama, une épidémie de diarrhée est survenue en janvier 2020 suite à la consommation de l'eau provenant de l'hydraulique améliorée villageoise. Pour mettre en place un système d'intervention précoce en cas de déclaration de maladie hydrique dans cette localité nous avons pensé à mettre en place un modèle de caractérisation bactériologique des nappes phréatiques du village M'pody à partir des paramètres physicochimiques et de la météorologie pour estimer la concentration des germes indicateurs de pollution fécale par puits.

Dans le cadre de cette étude, Le choix final des variables explicatives s'est porté sur 9 variables. La variance expliquée par les effets fixes et les effets de groupe (R^2 conditionnel) était de 30,25%. Les variables turbidité et précipitation étaient corrélées positivement à la concentration en *E. coli*. Ces variables sont plus susceptibles de se retrouver dans les modèles prédictifs des coliformes fécaux. Ce fut le cas dans une étude de Hebert et al., 2004 sur la modélisation de la qualité bactériologique d'un site potentiel de baignade à la baie de Beauport (Québec), 2003^[16]. Les précipitations enregistrées durant les 24 heures précédant l'échantillonnage ont eu un

impact important sur la qualité microbiologique de l'eau étant donné plus de 50 % des puits de cette zone disposaient d'installations sanitaires en amont des puits. Ces fosses septiques en amont pouvaient être une source de transport bactérien lors des pluies.

L'utilisation de la turbidité comme prédicteur était due au fait que la turbidité aurait pour origine la présence de matières en suspension. L'étude de N'Diaye et *al.* a montré une corrélation positive et fortement significative entre la présence de coliformes fécaux et la turbidité^[17].

Le fer est un nutriment indispensable aux bactéries. Il intervient à la fois comme macroélément et comme oligo-élément cationique. Il joue un rôle important dans le métabolisme bactérien en tant que donneur d'électrons et est impliqué dans de nombreuses réactions enzymatiques, notamment au niveau de la superoxyde dismutase, qui est impliquée dans la production d'oxygène. Le fer entre également dans la synthèse des acides nucléiques des bactéries, virus et parasites^[18]. La température et le pH sont deux éléments importants dans la croissance bactérienne.

E. coli se multiplie à des températures comprises entre 7 °C et 50 °C, la température optimale étant de 37 °C. Certaines souches se développent dans des aliments acides, jusqu'à un pH de 4,4^[19]. La variable explicative chlore a une corrélation négative avec les concentrations de coliformes fécaux et d'entérocoques. La présence de chlore dans le puits indique une non-conformité. Ne pas utiliser cette variable dans ce modèle biaiserait l'estimation de la concentration bactérienne étant donné que la population l'utilise.

La meilleure façon de mesurer la puissance prédictive d'un modèle est de tester le modèle sur un ensemble de données indépendantes qui n'est pas utilisé pour l'estimation des paramètres. Cependant, les ensembles de données indépendantes sont souvent indisponibles, difficiles à collecter et coûteux^[20]. Une façon de résoudre le problème est la validation croisée (CV). CV est une méthode de ré-échantillonnage des données en divisant un ensemble de

données en deux : un ensemble de données d'apprentissage et un ensemble de données de test. L'ensemble de données d'apprentissage est utilisé pour ajuster un modèle, et l'ensemble de données de test est utilisé pour évaluer les performances prédictives du modèle ajusté grâce aux erreurs de prédiction. Ce processus est répété plusieurs fois et l'estimation CV de l'erreur est l'erreur de prédiction moyenne sur les ensembles de données de test^[21].

Dans notre étude, la performance prédictive de notre modèle par validation k-fold a été donnée comme suite, l'erreur de prédiction moyenne (RMSE) variait de 1,9 à 2,15 avec une moyenne de 1,95. L'erreur de prédiction absolue variait de 1,45 à 1,75 avec une moyenne (MAE) de 1,6 et le R² moyen était de 0,31. L'idée de base lors de la mesure de ces paramètres de performances était de voir à quel point les prédictions du modèle étaient mauvaises ou erronées par rapport aux valeurs réelles observées. Ainsi, un RMSE élevé est «mauvais» et un RMSE faible est «bon». 4 cas de figure peuvent se présenter

- RMSE faible, R² élevé (le meilleur des cas)
- Faible RMSE, faible R²
- RMSE élevé, R² élevé
- RMSE élevé, R² faible (le pire des cas)^[22]

Le modèle se trouvait dans la deuxième situation avec des RMSE et R² faibles. La prédiction du modèle a été précise. En d'autres termes, les résidus étaient proches de zéro. Les estimations bootstrap des paramètres ont été considérées comme non biaisées lorsque le biais relatif était inférieur à 5%, modérément biaisées lorsque le biais relatif était compris entre 5 et 10 % et fortement biaisées s'il était supérieur à 10 %^[23]. Dans notre étude, le biais de chaque paramètre estimé était inférieur à 5 % donc non biaisé. Notre étude a rencontré des limites avec des R² conditionnel moyen faible qui pourraient signifier la non prise en compte de certains prédicteurs, l'absence de données indépendantes pour la validation des performances prédictives.

a montré que les prédictions de nos modèles sont précises et les estimations bootstrap des paramètres sont non biaisés. Les modèles mis en place pourraient servir en cas de déclaration de maladies hydriques dans cette localité avant le retour des résultats de l'analyse microbiologique.

CONCLUSION

L'existence des bases de données fiables a permis le développement d'un modèle statistique de prédiction de la concentration des germes indicateurs de pollutions fécale de cette localité. Ce modèle a pu être expliqué (*E.coli*) par neuf variables explicatives. La validation des performances prédictives par K fold et bootstrap

REFERENCES

- [1] **WATERAID**, 2013. Partout et pour tous. Une vision pour l'accès à l'eau potable, à l'hygiène et à l'assainissement après 2015. WaterAid, Londres, Royaume-Uni. n.d.
- [2] **Guergazi S, Achour S**. Caractéristiques physico-chimiques des eaux d'alimentation de la ville de Biskra. Pratique de la chloration. Larhyss J 2005;4:119–27.
- [3] **Festy B, Hartemann P, Ledrans M, Levallois P, Payment P, Tricard D**. Qualité de l'eau. Environ Santé Publique-Fond Prat 2003:333–68.
- [4] **Yapo O, Mambo V, Seka A, Ohou MJA, Konan F, Gouzile V, et al**. Evaluation de la qualité des eaux de puits à usage domestique dans les quartiers défavorisés de quatre communes d'Abidjan (Côte d'Ivoire): Koumassi, Marcory, Port-Bouet et Treichville. Int J Biol Chem Sci 2010;4.
- [5] **Laferrière M, Lavoie J, Payment P, Minville JJ**. L'industrie porcine: les risques liés à la santé humaine. Bull. Information Santé Environnem. Vol. 7. Québec: Unité de santé publique du Centre hospitalier régional du Grand-Portage; 1995. n.d.
- [6] **Angulo FJ, Tippen S, Sharp DJ, Payne BJ, Collier C, Hill JE, et al**. A community waterborne outbreak of salmonellosis and the effectiveness of a boil water order. Am J Public Health. 1 avr 1997;87(4):580–4. n.d.
- [7] **Balbus JM, Embrey MA**. Risk factors for waterborne enteric infections. Current opinion in gastroenterology. 2002;18(1):46–50. n.d.
- [8] **Ahoussi KE, Koffi YB, Kouassi AM, Soro G, Biémi J**. Étude hydrochimique et microbiologique des eaux de source de l'ouest montagneux de la Côte d'Ivoire : Cas du village de Mangouin-Yrongouin (sous-préfecture de Biankouman). Journal of Applied Biosciences 2013; 63: 4703– 471 n.d.
- [9] **Soro N, Ouattara L, Dongo K, Kouadio EK, Ahoussi EK, Soro G, et al**. Déchets municipaux dans le District d'Abidjan en Côte d'Ivoire: sources potentielles de pollution des eaux souterraines. Int J Biol Chem Sci 2010;4.
- [10] **Traoré A, Ahoussi KE, AKA N, Traoré A, Soro N.** Niveau de contamination par les pesticides des eaux des lagunes Aghien et Potou (Sud-Est de la Côte d'Ivoire). International Journal of Pure & Applied Bioscience 2015;3(4): 312–322 n.d.
- [11] **Ouattara A, Méité A, Dally T, Ouattara H, Kati-Coulibaly S**. Étude de la qualité des eaux de consommation dans la localité de N'Zianouan s/p de Tiassalé et des quartiers précaires de trois communes du District d'Abidjan (Koumassi, Treichville, Attécoubé). Journal of Applied Biosciences 2016, 102: 9708–9715. DOI: <http://dx.doi.org/10.4314/jab.v102i1.5> n.d.
- [12] **KOACI**. Côte d'Ivoire: 69 cas de diarrhée « détectés » à Anyama, aucune perte en vie humaine. KOACI n.d. https://www.koaci.com/article/2020/01/24/cote-divoire/sante/cote-divoire-69-cas-de-diarrhee-detectes-a-anyama-aucune-perde-en-vie-humaine_138802.html (accessed November 12, 2021).
- [13] **Rompré A, Servais P, Baudart J, De-Roubin M-R, Laurent P**. Detection and enumeration of coliforms in drinking water: current methods and emerging approaches. J Microbiol Methods. 2002;49(1):31–54. n.d.
- [14] **Nomades DC**. Météo à Anyama en 2020. Hist Météo n.d. <https://www.historique-meteo.net/afrique/cote-d-ivoire/anyama/2020/> (accessed May 19, 2022).
- [15] **Palm R**. Utilisation du bootstrap pour les problèmes statistiques liés à l'estimation des paramètres. Biotechnol Agron Société Environ 2002;6:143–53.
- [16] **????** modélisation de la qualité bactériologique d'un site potentiel de baignade à la baie de Beauport, QUEBEC, ETE, 2003 - (accessed May 16, 2022).
- [17] **N'Diaye AD, Sid MO, Kankou AO, Namr KI**. Evaluation des teneurs en coliformes fécaux par couplage avec des paramètres physicochimiques en ACP: cas des effluents de la STEP du périmètre maraicher de Sebka, Nouakchott. Cameroon J Exp Biol 2011;7:35–40.
- [18] **François Bricaire**. Fer et maladies infectieuses Paris - Recherche Google n.d. <https://www.google.com/search?q=Fer+et+maladies+infectieuses+Fran%C3%A7ois+Bricaire%2C+Paris&oq=Fer+et+maladies+infectieuses+Fran%C3%A7ois+Bricaire%2C+Paris&aqs=chrome..69i57.657j0j7&sourceid=chrome&ie=UTF-8> (accessed June 9, 2022).
- [19] **???** Escherichia coli (E. coli) n.d. <https://www.who.int/fr/news-room/fact-sheets/detail/e-coli> (accessed May 16, 2022).
- [20] **Snee RD**. Validation of regression models: methods and examples. Technometrics 1977;19:415–28.
- [21] **Yang Y, Huang S**. Suitability of five cross validation methods for performance evaluation of nonlinear mixed-effects forest models—a case study. For Int J For Res 2014;87:654–62.
- [22] **Wheeler W**. Evaluating linear regression models using RMSE and R². wwblog 2021. <https://medium.com/wwblog/evaluating-regression-models-using-rmse-and-r%C2%B2-42f77400efee> (accessed May 17, 2022).
- [23] **Thai H-T, Mentré F, Holford NH, Veyrat-Follet C, Comets E**. A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. Pharm Stat 2013;12:129–40

Déclaration de conflit d'intérêt :

Les auteurs déclarent aucun conflit d'intérêt

Contribution des auteurs :

- Meless Djedjro Franck Renaud, Gbagbo Tchape Aubin ont initié cette étude et ont écrit le premier draft de cet article.
- Meless Djedjro Franck Renaud a réalisé le traitement statistique
- Kpaibe SAP, Kouassi-ATB et Amin N'cho christophe ont supervisé la collecte, les analyses physiques et microbiologiques au laboratoire et la réalisation de cette étude.

- Amin N'cho christophe, Kpaibe SAP et Kouassi-ATB ont relu le manuscrit

Tous les auteurs ont contribué à la relecture et correction du manuscrit.

Tous les auteurs n'ont aucun conflit lié à ce travail et ont donné leur accord pour la publication de ce manuscrit

Remerciements

Nous voulons remercier spécialement nos encadreurs Mr Laurent Lehot et le Professeur Roch Giorgi de l'Université Aix Marseille.