Use of a Multiple Regression Model in the Estimation of Water Borehole Flows in the Middle of Cracked Basement in Côte d'Ivoire

Assoué Kouakou Sylvestre Kouadio[1], Ouedraogo Moussa[1,2], Ismaïla Ouattara[1,2], Issiaka Savane[1]

[1]Laboratoire Géosciences et Environnement (LGE), UFR des Sciences et Gestion de l'Environnement (SGE), University of Nangui Abrogoua (UNA), Abidjan, Côte d'Ivoire
[2]Laboratoire Géosciences et Environnement (LGE), UFR Sciences Géologique et Minière (SGM), University of Man (UM), Man, Côte d'Ivoire
Email: sylvestreassoue@gmail.com

Abstract

The objective of this study was to propose a mathematical regression model to estimate the exploitation flow rate of a water borehole from geophysical parameters in the midst of a fissured basement in the central-eastern part of Côte d'Ivoire. The data of the study are both hydrogeological and geophysical parameters from one hundred and eleven (111) data sheets of (111) water and geophysical boreholes. Two methods were used. The Normal Principal Component Analysis (NPCA) method applied to the data made it possible to select the explanatory variables (geophysical parameters) for borehole productivity. The multiple linear regression method subsequently made it possible to propose a mathematical model capable of estimating the exploitation rate from the geophysical parameters. The results indicate a very strong correlation (0.87) between longitudinal conductivity and flow rate, with flow rate and apparent resistivity negatively correlated. The multiple linear regression method highlighted two relevant explanatory variables, longitudinal conductivity and apparent resistivity. These two geophysical parameters contributed to a mathematical model in the form $Q = C_1 X_1 + C_2 X_2 + \cdots + C_n X_n + C_0$. the real model obtained in this work is $Q = 0.82 Cl - 0.12 Rho.app + 2.5$. The resulting model is efficient with a correlation of 86% in calibration and 95% in validation. A bias of 0.37 in calibration and 0.82 in validation is observed. Finally, the square root mean square error (RMSE) is 3.10 to 3.38 respectively in calibration and validation.

Keywords

Productivity, Drilling, Apparent Resistivity, Longitudinal Conductivity,

Linear Regression, Fissured Aquifers

1. Introduction

Access to water and sanitation is one of the major challenges of the 21st century. According to the WHO [1], more than 2.1 billion people do not have access to safe drinking water and more than 4.5 billion people do not have access to sanitation facilities. As a result, these disastrous living conditions cause the death of nearly 10,000 people a day, the vast majority of whom are children. In the face of this uncertain future for drinking water in Africa and particularly in the West African sub-region, various national and sub-regional programme's are being put in place [2]. In Côte d'Ivoire, basement aquifers are the main sources of drinking water supply for the population. According to the Directorate for Water Resources Management and Protection (DGPRE) of the Ministry of Water and Forestry, the quantity of water contained in the basement is estimated at about 113 billion m^3, of which 35.5 billion m^3 are renewed each year. Despite the availability of this resource, drinking water supply remains a concern. According to the National Drinking Water Office (ONEP) in 2018, the drinking water coverage rate was 61 percent in urban areas, 76 percent in rural areas and 13 percent in semi-urban areas. This disparity is said to be linked to a lack of a serious and in-depth study on the productivity of water drilling in areas of cracked basement. In general, the drinking water supply of populations in rural and semi-urban areas is provided by basement aquifers.

Indeed, basement aquifers are formed from fractures created in the bedrock by tectonic constraints and weathering [3]. Crystalline basement regions are known for their structural complexity. The associated water resources reflect this complexity: highly heterogeneous, constraining identifier and to be characterized [4]. [5] describe the characterization of water resources in fractured environments as one of the greatest challenges for modern hydrogeologists. This is why the knowledge of the potentiality of these fractured aquifers associated with crystalline and crystallophyllized basement requires the implementation of an appropriate prospecting methodology [6]. The methods commonly used by these eminent researchers and other hydrogeologists are based on boreholes or piezometers [7]. However, the information obtained from these works has a low resolution due to their punctual nature. Thus, the location of these structures according to the fractures they capture plays a major role in their productivity. In addition, the high cost of construction and their destructive nature makes it impossible to build a large number of these structures.

It is therefore necessary to use complementary methods, in particular non-destructive geophysical methods [4]. The combination of these geophysical and hydrogeological methods could help to gain a deeper understanding of the productivity of water drilling in basement environments.

The general objective of our research is to propose a mathematical model capable of estimating the flow rate of a water borehole based on geophysical parameters in a cracked basement environment.

2. Presentation of the Study Area

The study area is located in the central-eastern part of Côte d'Ivoire between longitudes 3˚40' and 4˚55' West and latitudes 6˚20' and 8˚10' North and includes nine departments. It includes the N'zi, Iffou and Moronou regions (Figure 1).

The relief of the study area is dominated by a series of hills that rise to 300, 400 and even 500 m [8]. Located at the border between savannah and forest, it is covered by three types of vegetation. We have the dense forest in the Centre to the South and East, the shrubby forest in the North and North-west. And shrubby savannah from the Northeast to the West. The forest that covered most of the region is gradually retreating from north to south and extending eastward mainly due to seasonal bush fires. This forest is increasingly being destroyed in favour of the large rubber plantations cultivated in the region.
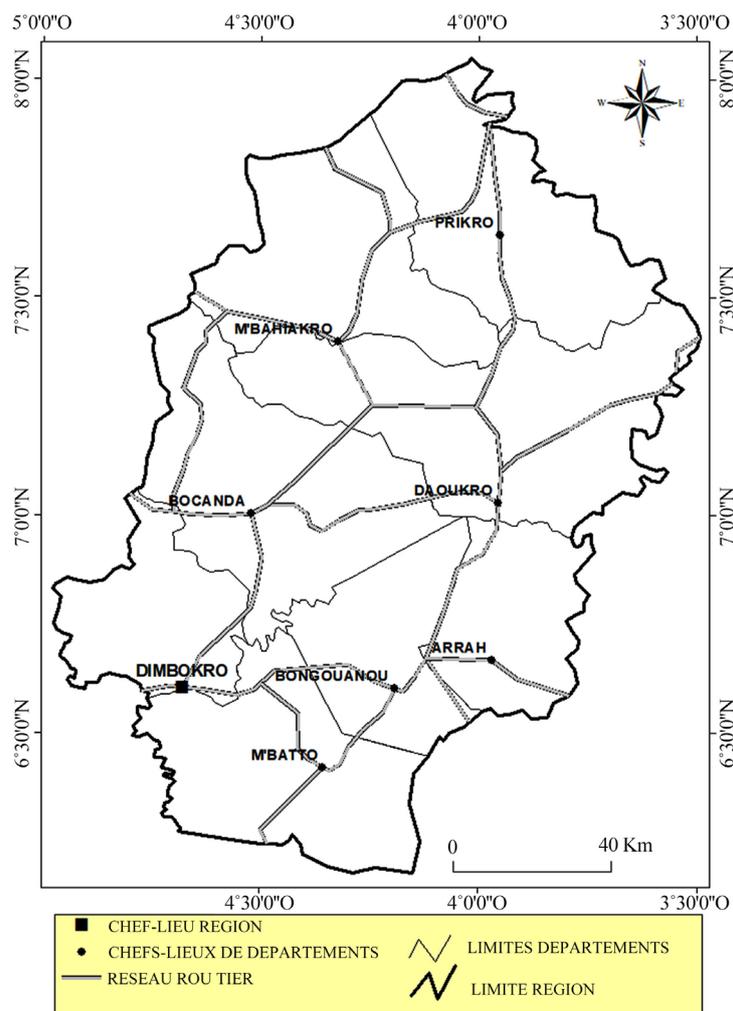
Figure 1. Geographic location of the study area.

The study area has a transitional equatorial climate. It is particularly hot and relatively dry with two rainy seasons from April to July and October to November and two dry seasons from December to March and August to September. Winds are generally light. Only the harmattan is important especially in the northern part of the region, but its influence decreases rapidly towards the south. However, with climate change, this regime is now disrupted.

The search for water in areas where the basement is crystalline or crystallo-phyllized is limited to areas where the rock has undergone the phenomenon of fracturing. The fractured environment is heterogeneous and its hydraulic characteristics are determined by the geometry and density of the cracks. The hydraulic continuity of this hydrogeological system depends on the interconnection of the cracks [9]. Crack aquifers are much larger but irregularly shaped reservoirs and are located at depths of 20 to 60 m or more, where the fractures are decompressed by lithostatic stress [10]. According to the same author, these aquifers develop a higher permeability than alteration and are characterized by low useful porosity. In outcrops, distension fractures with or without quartz filling are often observed and constitute very good drains for both recharge and groundwater exploitation (**Figure 2**) [11].

3. Materials and Methods

3.1. Data

The data used in this research was obtained from the technical files of 111 boreholes drilled in 2010 within the framework of the (West African Monetary Economic Union) UEMOA projects in the N'Zi, Iffou and Moronou Regions and within the framework of the 2013 Presidential Emergency Programme (PPU) in the Iffou Region (**Figure 3**). These technical files present, among others, the hydrogeological sounding data sheets, test pumping data sheets, and data sheets for electrical dragging and sounding. These data were processed using computer tools such as STATISTICA 7.1, IPI2WIN and ARCGIS 10.

3.2. Methods

3.2.1. Expressions of Some Parameters

Two hydrogeological parameters, *i.e.* the operating flow (Q) and the specific flow (Qs), and four geophysical parameters, such as longitudinal conductance (Cl), transverse resistance (Rt), apparent resistivity ($Rho.app$) and simulated alteration thickness (Eas), were analysed. The method used to determine the specific flow rate (Qs) is based on the following simple Equation (1).

$$Qs = \frac{Qd}{s} \tag{1}$$

where Qs is the Specific flow rate in m³/h/m, Qd is the flow rate of the last pumping stage in m³/h and s, the drawdown at the end of the 4 hours of pumping.
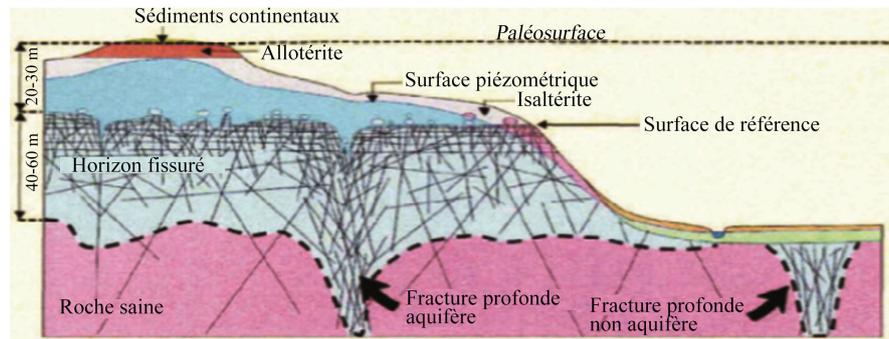
Figure 2. Typical weathering and fracturing profile of crystalline basement aquifers from [12].
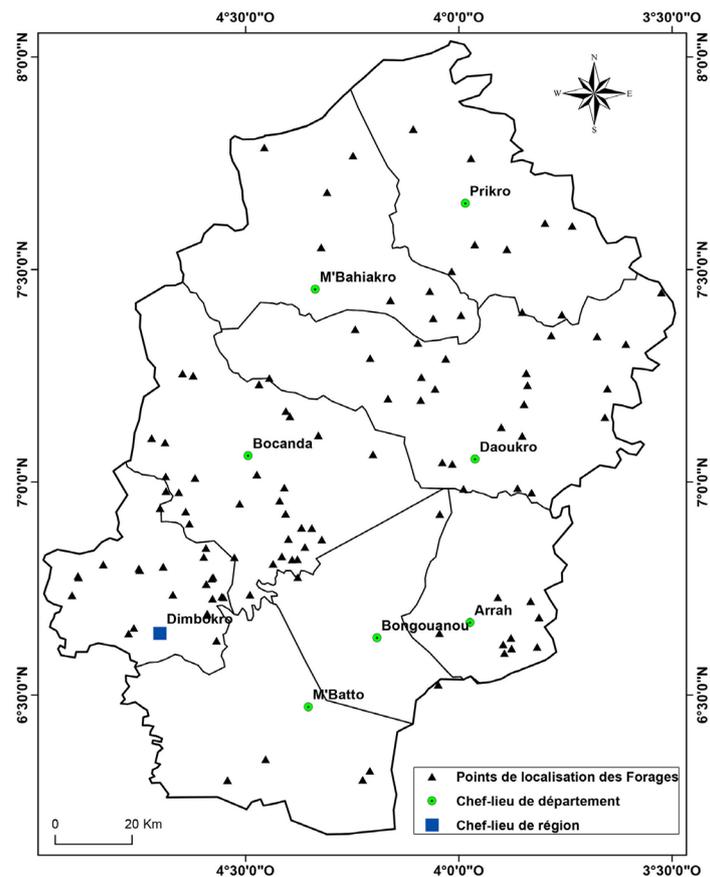
Figure 3. Location of water point data.

The relationships between hydrogeological and geophysical parameters in the study of aquifers are established through the work [13] which proposed the following formulas.

$$F = \frac{\rho}{\rho_w} \tag{2}$$

$$F = \phi^{-m} \tag{3}$$

where, *F* is the formation factor, ρ is the rock resistivity in ohm·m, ρ_w is the resistivity of formation water in ohm·m, ϕ is the total porosity of the forma-

tion, m is the cementing coefficient of the formation. The cementing factor depends on the permeability of the rock and therefore on the fracturing density for magmatic rocks. The knowledge of these hydrogeophysical relationships has led to the establishment of the following expressions [14].

$$Cl = \sum_{i=1}^{n} \frac{b_i}{\rho_i} \tag{4}$$

$$Rt = \sum_{i=1}^{n} b_i \cdot \rho_i \tag{5}$$

where Cl is the longitudinal conductance, b_i are the layers thicknesses, ρ_i are the layers resistivity and Rt the transverse resistance.

3.2.2. Statistical Analysis of the Data: NPCA and Linear Regression Model

Hydrogeological and geophysical parameters from the analysis of drilling data and geophysical studies were subjected to a standardized principal component analysis (SCA). The NPAC, which is a statistical tool, defines the main factors whose correlation with the variables allows an explanation of the phenomena involved. In order to assume that the phenomenon involved is sufficiently expressed, the cumulative sum of the contributions of the main factors must be about 70% [6]. This method made it possible to study the links and the grouping of the parameters studied.

In addition, multiple linear regression was used to predict the values of a dependent variable from explanatory or independent variables [15]. Indeed, multiple linear regressions are used to determine the most satisfactory linear relationship for predicting the dependent value that produces the smallest standard error. In such a model, each independent variable is weighted so that the value of the regression coefficients maximizes the influence of each variable in the final equation. It is possible to manipulate multiple independent variables from multiple linear regressions, but only one dependent variable. Multiple linear regression has already been used by several authors such as [16] [17]. Also, [18] has used it to study the productivity of water drilling in basement environments. In these different works, the models designed have demonstrated their capacity to simulate the productivity of water drilling in basement environments in general and particularly in Côte d'Ivoire and Chad. In this study, the characteristics explained (dependent variables) are the limit drilling depth and the thickness of the drilled basement. As regards the explanatory variables, these are drilling parameters whose relationship with productivity would be proven by multivariate statistical analysis. These are the most relevant and influential parameters revealed by the statistical analysis. In a multiple linear regression, the equation is in the following form Equation (6).

$$Y = C_1 X_1 + C_2 X_2 + \cdots + C_n X_n + C_0 \tag{6}$$

where Y is the variable explained, X_i is the explanatory variable, C_0 is the constant, $C_i (1 \le i \le N)$ the weighting coefficients of the explanatory variable X_i.

Indeed, Y is a vector of observed values of water drilling flow, X_i is a matrix of independent or explanatory variables, C_i is a vector of parameters or regression

coefficients to be estimated, and C_0 is a vector of residuals or random disturbances. Linear regression estimates the vector C_i as a least squares solution [19] [20].

$$C_i = \left(X_i^{\mathrm{T}} X_i \right)^{-1} \left(X_i^{\mathrm{T}} Y \right) \quad (7)$$

with X^T the transpose of X.

Multiple regression is a variant of the simple regression method that can help deal with collinearity by iteratively choosing the variables with the highest explanatory value. An ascending multiple regression starts with no variable, or a subset of the available variables, and adds the most significant variable (the one with the lowest p-value, combined with estimated F-statistics) at each step of the model. A stepwise downward regression starts with all available variables and removes the least significant variable at each step. This is the second form of multiple regression that was chosen for this study.

One of the most relevant steps in the development of a model is the estimation of its parameters [21]. According to this author, there are multiple estimation methods; this is due to the fact that the nature of the parameters is varied and that there is no globally satisfactory estimation method. Estimation involves making choices to establish a calibration strategy. The strategies are of course a function of the objective sought by the model, taking into account its relevance and performance. The calibration of the models was based on the principle of the "split-sample test", which consists of calibrating two-thirds (2/3) of the sample of available data and validating the remaining third (1/3). Thus, the calibration of the models was carried out on a sample of 74 boreholes and the validation on a sample of 37 boreholes. The estimation of the weighting coefficients of the selected variables was carried out by calibration. It consisted in adjusting the numerical values assigned to the model parameters to best reproduce the observed response. In presenting the calibration results, it is important to associate each regression coefficient with the associated standard error, which is an indicator that can be assimilated to the standard deviation. In fact, the standard error is to the regression coefficient what the standard deviation is to the mean of a variable. It therefore consists of a measure of the variability of the regression coefficient. Thus, if several regressions were performed on as many sub-samples drawn from the same main sample, the value of the regression parameters thus obtained would differ from one time to the next. The "robustness" of a given coefficient will be greater the smaller the variation around the most likely value, the coefficient itself. This is precisely what the standard error of the coefficient measures. The entire methodology has been implemented in the STATISTICA 7.1 software.

3.2.3. Criteria for Evaluating the Quality of the Model Developed
1) Evaluation of model performance
The performance of the models and their robustness were evaluated using numerical and graphical criteria. The analysis of the simulation results focuses on

the performance of the models in the calibration and validation phases. Indeed, according to [22], the calibration performances are less revealing of the real simulation capacities of the models. These are better expressed by the validation. To evaluate the performance of the models developed in this study, the correlation coefficient, the square root of the root mean square error and the biased mean were used as evaluation criteria. The correlation coefficient (R) is used to measure the linear correlation between the real or measured variable and the calculated variable. Its formula is Equation (8).

$$R = \frac{\sum_i (Q_i - \bar{Q})^2 \times (Q_i' - \bar{Q}')}{\sqrt{\sum_i (Q_i - \bar{Q})^2 \times \sum_i (Q_i' - \bar{Q}')^2}} \tag{8}$$

with, Q_i are the measured flow rate, Q_i' are the simulated flow rate, $\bar{Q}_i$ are the average of measured flows and $\bar{Q}_i'$ the average of simulated flows.

The relation is said to be perfect if $R = 1$, very strong if $0.8 \le R < 1$, strong if R is between 0.5 and 0.8, medium if R is between 0.2 and 0.5, weak if R is between 0 and 0.2 and nil if $R = 0$. The square root mean square error (RMSE) is used as the measure of the overall performance of the model. The model is well optimized if the RMSE value is close to zero, which tends towards a perfect forecast. Its mathematical formulation is given by the following relationship Equation (9).

$$\sqrt{\frac{\sum_i^n (Q_i - Q_i')^2}{n}} \tag{9}$$

with Q_i the Observed flow, Q_i' are the simulated flow, n is the sample size.

Bias is a criterion for highlighting the difference between two quantities. It must be minimized (the optimum is the null value). It then gives the relative error between the observed values and those simulated during the analyses. When the biased mean "B" tends towards zero, the model results are unbiased, *i.e.* the two values are close and therefore the model is efficient. This parameter is defined by the following relationship Equation (10).

$$B = \frac{1}{n} \sum_{n=1}^{i} |Q_i - Q_i'| \tag{10}$$

where Q_i are the observed flow rate, Q_i' are the simulated flow.

In addition to the numerical evaluation, a graphical analysis comparing the observed flows to the simulated flows was carried out in order to assess the quality of the modelling carried out. In practice, if the simulation was perfect, *i.e.*, if each of the values simulated by the model was equal to the observed value, the resulting scatterplot would be aligned and merged with the line of equation $y = x$. However, since the modeling is not perfect, the qualitative assessment of the performance of the different models consisted in assessing the dispersion of the scatter of the scatterplot around the first diagonal.

2) Assessment of model robustness

One of the most widely used techniques to assess the robustness of a model is the double-sampling technique [23]. This technique makes it possible to test the

adaptability of the model regardless of its complexity. The robustness criterion (R') of the model was assessed by the difference in the correlation coefficient in validation (R validation) and calibration (R calibration). The model is said to be robust if the robustness criterion is less than or equal to 10%. The expression of the robustness criterion is Equation (11).

$$R' = 100 \times \left| R_{\text{validation}} - R_{\text{calage}} \right| \tag{11}$$

4. Results and Discussion

4.1. Statistics of the Parameters Studied

Six parameters have been presented in the parameter characteristics Table 1 below. Apart from flow rate and apparent resistivity, which are parameters for which data are experimental, the other five (longitudinal conductance, transverse resistance, simulated weathering thickness and specific flow rate) are parameters for which data have been calculated. The coefficient of variation (Cv) varies from 51.97% to 731.69%. It is greater than 100% except for the weathering thickness, which is 51.97%. These values show that the parameters studied are not very variable.

4.2. Multiple Regression Model of Water Drilling Productivity

4.2.1. Identification of Relevant Explanatory Variables by the NPCA Method

1) Analysis and Interpretation of Own Values

The analysis of Table 2 below shows three determining factors with eigenvalues of 2.93, 1.17 and 0.89 and cumulative percentages of variances of 37.02%, 67.62% and 85.28% respectively. The cumulation of these three first factors being higher than 70% is enough determining factor to express the relations between the hydrogeological and geophysical parameters. In view of the percentages expressed, the factor analysis will focus on two factorial levels. The factorial planes F1 - F2 and F1 - F3.

Table 1. Table of study parameters.

Paramètres	Number	Minimum	Maximum	average	Ecart-type	Cv (%)
Apparent resistivity (*rho.app*)	111	42	2050	435.80	459.7	105.48
Longitudinal conductance (*Cl*)	111	0.000806	3.51	0.53	0.80	150.94
Transversal resistance (*Rt*)	111	762.35	199,694,050	2,861,528	20,937,715	731.69
Calculated weathering thickness (*Ep.cal*)	111	4.22	100	38.73	20.13	51.97
Flow rate (*Q*)	111	0.38	43.20	5.43	6.67	122.83
Specific flow (*Q/s*)	111	0.02	4.09	0.45	0.72	160

Table 2. Eigenvalues of hydrogeological and geophysical parameters.

Factors	Eigenvalue %	Variance Cumulative %	Eigenvalue Cumulative	Cumulative Variance
Factors 1	2.93	48.83	2.93	48.83
Factors 2	1.17	14.85	4.10	68.37
Factors 3	0.89	11.56	4.99	83.22

2) Factor Plan F1 - F2 and F1 - F3

Figure 4 below shows the projection of the variables on the factorial plane F1 - F2. Flow rate, specific flow rate and longitudinal conductivity are the variables at the negative end of the F1 factorial axis. At the positive end of the same factorial axis we have the apparent resistivity. The F1 factorial axis, which groups together the parameters flow rate, specific flow rate, longitudinal conductivity and apparent resistivity, expresses the productivity of water drilling. However, the apparent resistivity, because of its position opposite to that of the flow, specific flow and longitudinal conductance, explains why a high resistivity could reduce the productivity of a borehole. On factorial axis 2, we observe a clustering of the simulated weathering thickness and the transverse resistance on the lower side. This information shows that the factorial axis F2 would express the thickness of alteration.

Figure 5 shows a grouping of flow, specific flow and longitudinal conductivity on the F1 factor axis as in Figure 4. The difference with respect to the factorial plane F1 - F2 is on the factorial axis F3 which shows an opposite position between the calculated weathering thickness and the transverse resistance. We can retain that the longitudinal conductance and the apparent resistivity are geophysical parameters that also express the productivity of a water well.

3) Analysis of the correlation matrix between the variables

Analysis of the correlation matrix in Table 3 shows that there is a good correlation between hydrogeological and geophysical parameters. This is the case between flow and longitudinal conductance. The correlation between these two parameters is strong (0.87). These results confirm the grouping of these parameters on the factorial axes illustrated above. This correlation is also remarkable with the apparent resistivity where the correlation is average and negative (−0.45). This explains an inverse evolution between these two parameters. These correlations are very weak between the transverse resistance, which is a geophysical parameter, and the flow rate. These values are (−0.04), (−0.04) and (−0.05) respectively. Between the geophysical parameters, the only significant correlation is between apparent resistivity and longitudinal conductivity (−0.41). It is also negative and could mean that low apparent resistivity leads to high longitudinal conductivity. These results from this part of the study show that the geophysical parameters that best express productivity are longitudinal conductivity and apparent resistivity.
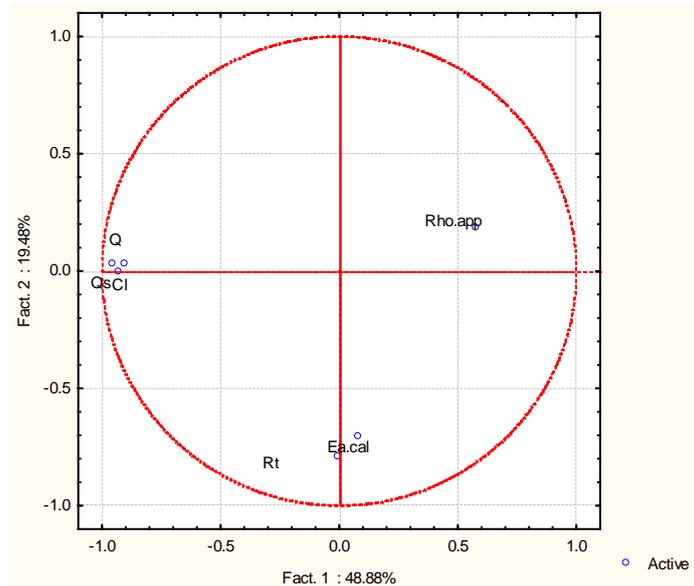
Figure 4. Projection of the variables on the factorial plane F1 - F2.
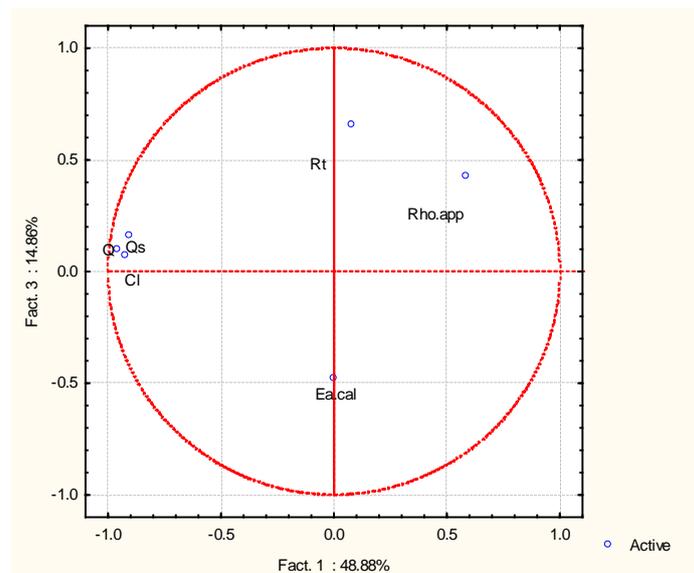
Figure 5. Projection of the variables on the factorial plane F1 - F3.

Table 3. Correlation table between variables.

Variables	*Qal*	*Q/s*	*T*	*Rho.app*	*Ea.cal*	*Cl*	*Rt*
Qal	1.00	**0.85**	**0.70**	**−0.45**	−0.04	**0.87**	−0.05
Q/s	**0.85**	1.00	**0.66**	**−0.34**	−0.03	**0.79**	−0.04
T	**0.70**	**0.66**	1.00	**−0.30**	0.01	**0.53**	−0.04
Rho.app	**−0.45**	**−0.34**	**−0.30**	1.00	−0.11	**−0.41**	0.05
Ea.cal	−0.04	−0.03	0.01	−0.11	1.00	0.02	0.16
Cl	**0.87**	**0.79**	**0.53**	**−0.41**	0.02	1.00	−0.05
Rt	−0.05	−0.04	−0.04	0.05	0.16	−0.05	1.00

4.2.2. Statistical Models of Linear Regression

1) Design of multiple linear regression models

The models were designed using the backward or backward elimination regression method. The results are presented in Table 4. First, step 0 includes the 4 explanatory variables which are the geophysical parameters. These parameters are: longitudinal conductance, apparent resistivity, transverse resistance and calculated alteration thickness. At this step, the multiple R is 88% and the set of variables in the model of the equation explains nearly 78% of the variance ($R^2 = 0.78$). This step shows us that longitudinal conductivity has an important weight (82%) in the prediction of borehole flow rates. Next comes the apparent resistivity, which influences the productivity estimate by 12%. However, transverse resistance and calculated alteration thickness have no significant influence (0% and 4% respectively). Next, step 1 gives us 3 parameters, namely longitudinal conductance, apparent resistivity and calculated thickness of alteration. The transverse resistivity being very insignificant (0.0004) in the prediction of productivity was rejected in step 2. The weights of the variables remain the same in this step as well as the multiple R and the R. The standard error varies slightly at the C_0 constant from step 0 to step 1. Finally, step 2 includes two explanatory variables which are the apparent resistivity with an influence weight of (−0.12) or (12%) and the longitudinal conductivity with a much larger influence weight of (0.83) or (83%). These values show that the longitudinal conductivity can help predict the flow rate of a borehole. Also, the R^2 of 78% means that the apparent resistivity and the longitudinal conductivity are likely to predict the productivity of a borehole. The set of standard errors is less than 1%. These results express a good relationship between the flow rate (explained variable) and the explanatory variables (apparent resistivity and longitudinal conductance).

Table 4. Table of coefficients of explanatory variables.

Step	Explanatory variable	Coefficient (C_i)	R	R^2	Error-type
	Constante (C_0)	2.88			0.70
	Rho.app	−0.120			0.0007
0	*Ep.cal*	−0.040	0.88	0.78	0.009
	Cl	0.82			0.43
	Rt	0.0004			0.000
	Constante (C_0)	2.88			0.69
	Rho.app	−0.120			0.0007
1	*Ep.cal*	−0.040	0.88	0.78	0.009
	Cl	0.82			0.42
	Constante (C_0)	2.50			0.55
2	*Rho.app*	−0.12	0.88	0.78	0.0007
	Cl	0.83			0.42

2) Results of multiple linear regression models

Following the analysis of Table 4 above, we retain two models of equations.

- Model 1 with the parameters longitudinal conductance (*Cl*), apparent resistivity (*Rho.app*) and calculated alteration thickness (*Ep.cal*). It is expressed in the form of the following equation.

$$Q = 0.82\,Cl - 0.12\,Rho.app - 0.04\,Ep.cal + 2.88 \qquad (12)$$

- Model 2 with the parameters longitudinal conductance (*Cl*) and apparent resistivity (*Rho.app*). It is expressed in the form of the following equation

$$Q = 0.82\,Cl - 0.12\,Rho.app + 2.50 \qquad (13)$$

4.2.3. Evaluation Criteria for the Proposed Equations

1) Results of the quality of the models developed through performance

Three evaluation criteria were selected to verify the performance of the model. Only Model 2 was subjected to the evaluation criterion after the top-down elimination regression test. Table 5 shows a very high correlation R (86%) in calibration and 95% in validation. This very high correlation expresses the good relationship between the observed drilling rates and the drilling rates simulated from equation 12 above. The bias B is 0.37 in calibration and 0.82 in validation. These values tend towards zero. They show that the relative error between the observed and simulated flows is small. However, this relative error is higher in validation than in calibration. The values of the square root mean square error (RMSE) vary from 3.10 to 3.38 respectively in calibration and validation. The RMSE best expresses the overall performance of the equation when they tend towards zero. Overall, the results obtained show that the model of the equation developed performs well.

2) Results of the quality of the models developed using robustness

The robustness of the model was tested by the double sample technique. This technique allowed us to determine the robustness criterion R', which is 9%. This value allowed us to test the robustness of the developed model because R' is less than 10%. These results confirm those obtained in Table 5.

3) Graphical validation of the developed model

The developed model was tested graphically by a correlation study betweenthe observed and simulated flow in calibration and validation. Figure 6 and Figure 7 below show that there is a strong correlation (0.86 in calibration and 0.95 in validation) between the observed and simulated flows. This correlation is stronger and reflects a good performance of the model in validation. The graphs also show that the points are well grouped around the straight line of the equation for flows below 12 m³/h. Above 12 m³/h, we observe a dispersion of values around the equation line. However, this dispersion is important in Figure 6 (calibration). However, these graphical results confirm the numerical results obtained above.

5. Discussion

The study of the prediction of drilling productivity from geophysical parameters was carried out by analyzing six parameters, including four geophysical parameters

Table 5. Performance of the developed model.

Evaluation Criteria	Calibration performance	Validation performance
	Model 2	Model 2
R	0.86	0.95
Biais (*B*)	0.37	0.82
RMSE	3.10	3.38
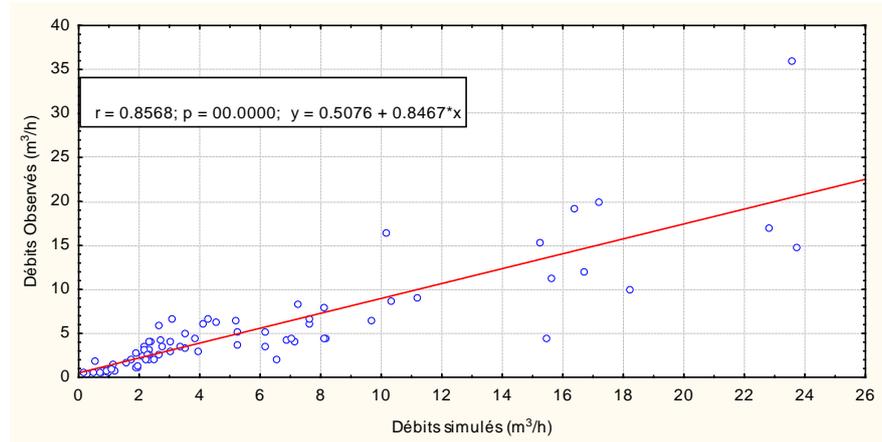
Figure 6. Flows simulated and observed by the calibration model.
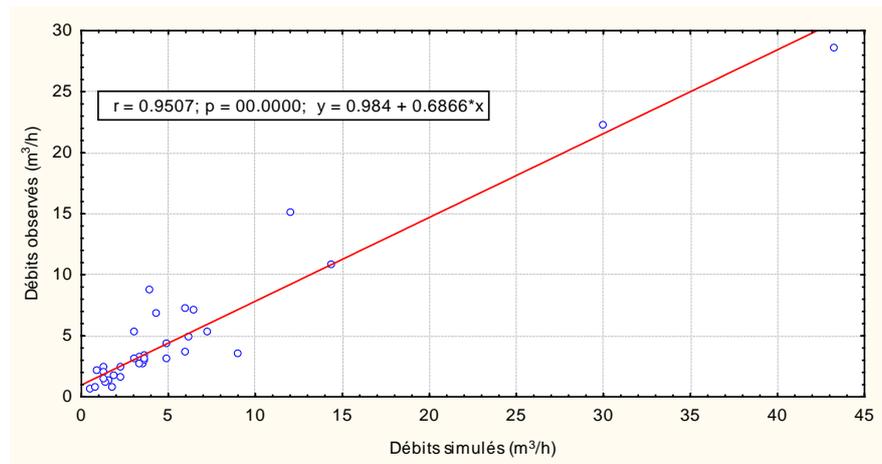
Figure 7. Flows simulated and observed by the model under validation.

(longitudinal conductance, transverse resistance, apparent resistivity and calculated alteration thickness) and two hydrogeological parameters (flow, specific flow) that influence productivity. Indeed, authors such as [24] in the Baya basin; [25] at Aboisso, [10] at Man-Danané have shown that the hydrogeological parameters influencing productivity are flow, transmissivity and specific flow. The study of the coefficient of variation of these parameters shows that it is greater than 100% for all parameters except the calculated alteration thickness, which is 51.97%. This confirms the high variability of the data for the parameters studied. Indeed, according to [13], the water saturation rate of a rock influences its apparent re-

sistivity. In other words, a water-saturated rock has a high productivity. With regard to the results obtained after the identification of the relevant explanatory variables, parameters such as flow rate, specific flow rate and longitudinal conductivity are governed by the same phenomenon which is productivity [11] [25] [26]. Also, the calculated weathering thickness and the apparent resistivity have little influence on productivity. The correlation between these hydrogeological and geophysical parameters was studied using the NPCA method and the multiple linear regression method to establish a model. This model showed a perfect correlation between the flows obtained after drilling and the simulated flows obtained from the developed model. This explains why surface geophysical parameters such as longitudinal conductance and apparent resistivity can be used to predict drilling productivity. Authors such as [16] [17] in Côte d'Ivoire [27] in Morocco, [18] in Chad and [28] in Central African Republic have conducted studies in which they used surface geophysical parameters to determine a model of borehole productivity. This model was validated by two main criteria, the performance criterion through calculations of the correlation R between the observed and simulated flows in calibration and validation. This correlation R is 0.86 in calibration and 0.95 in validation. Also the bias calculation which gave values of 0.37 and 0.82 only confirmed the performance of the equation. In addition, another criterion which is the robustness criterion gave values that confirm the reliability of the equation. These criteria were also used by [17] in the N'zi-comoé region to validate equation models.

6. Conclusions

Prediction of water drilling productivity based on surface geophysical parameters was done in a methodical manner. First, the choice of the hydrogeological parameters governing borehole productivity was made. Thus, two parameters, namely the flow rate and the specific flow rate, were selected as explanatory variables. But the flow rate being the parameter observable after a hydrogeological drilling was retained as the explained parameter. Also, four geophysical parameters such as longitudinal conductance, transverse resistance, calculated alteration thickness and apparent resistivity were analyzed and retained as explanatory parameters. Similarly, the study of coefficients of variation gave values greater than 100% except for the non-calculated alteration thickness, which is 51.97%. In view of these results, we can say that the parameters studied are highly variable in the research area.

Finally, a model for estimating drilling productivity was established based on surface geophysical parameters. This model allowed us to understand that the longitudinal conductance with a weighting coefficient of 0.86 and the apparent resistivity with a weighting coefficient of (−0.12) can be used to simulate a drilling rate after a geophysical study. To validate this model, we performed the performance test by determining the correlation R to show the link between the observed and simulated flows. The values of R = 0.86 in calibration and 0.95 in va-

lidation obtained show that there is a very strong correlation between the simulated and observed flows. The values of Bias B (0.37 in calibration and 0.82 in validation) tend towards zero, they also confirm the performance of the developed model. Another confirmation, that of the robustness test gave an R' value of 9%. This also confirms the efficiency of the developed model. After this study, we retain that the developed model can predict the productivity of a water well after a geophysical study.

Authors' Contributions

A.K.S.K., O.M., O.I., and I.S., designed the study, developed the methodology and wrote the manuscript. A.K.S.K., collected the data and performed the analysis; O.M., O.I., and I.S., supervised the data analysis; O.M., contributed to paper write up.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] OMS-UNICEF (2017) Progrès en matière d'eau, d'assainissement et d'hygiène, mise à jour 2017 et évaluation des ODD. Rapport commun OMS/UNICEF.

[2] Youan, T.M. (2008) Contribution De La Télédétection Et Des Systèmes D'informations Géographiques A La Prospection Hydrogéologique Du Socle Précambrien d'Afrique De L'ouest: Cas De La Région De Bondoukou (Nord-Est De La Côte d'Ivoire). Thèse De Doctorat, Université De Cocody, Abidjan (Côte d'Ivoire), 237 p.

[3] Lachassagne, P., Wyns, R. and Dewandel, B. (2011) The Fracture Permeability of Hard Rock Aquifer Is Due Neither to Tectonics, Nor to Unloading, But to Weathering Processes. *Terra Nova*, **23**, 145-161.
https://doi.org/10.1111/j.1365-3121.2011.00998.x

[4] Soro, D. (2017) Caractérisation et modélisation hydrogéologique d'un aquifère en milieu de socle fracturé: Cas du site expérimental de Sanon, région du plateau central au Burkina Faso). Thèse de Doctorat en cotutelle, Université Pierre et Marie Curie-Paris 6 (UPMC), Institut International d'Ingénierie de l'Eau et de l'Environnement (2iE), 303 p.

[5] Neuman, S.P. (1974) Effect of Partial Penetration on Flow in Unconfined Aquifers Considering Delayed Gravity Response. *Water Resources Research*, **10**, 303-312.
https://doi.org/10.1029/WR010i002p00303

[6] Kouadio, K.E. (2005) Stratégie de prospection des nappes de fissures par analyse

spatiale du potentiel de productivité et optimisation de la profondeur des forages. Cas du Denguélé (Nord-Ouest de la Côte d'Ivoire). Thèse unique de doctorat, Université de Cocody, 181 p.

[7] Guérin, R. (2005) Borehole and Surface-Based Hydrogeophysics. *Hydrogeology Journal*, **13**, 251-254. https://doi.org/10.1007/s10040-004-0415-4

[8] Yao, D., Delor, C., Gadou, G., Kohou, P., Okou, A., Konate, S. and Diaby, I. (1995) Notice explicative de la carte géologique feuille de Dimbokro. 21 p.

[9] Coulibaly, M. (2001) Hydrogéologie des milieux de socle: Hydrodynamique des réservoirs fissurés. Cas de la région de Niellé (Nord de la Côte d'Ivoire). Mémoire de DEA des Sciences de la Terre, option Hydrogéologie, Univ. Abidjan, 71 p.

[10] Lasm, T., Kouame, F., Oga, M., Jourda, J., Soro, N. and Kouadio, H. (2004) Etude de la productivité des réservoirs fracturé des zones de socle Cas du noyau archéen de Man-Danané (Ouest de la Côte de la d'Ivoire). *Revue Ivoirienne des Sciences & Technologie*, **5**, 97-115.

[11] Biemi, J. (1992) Contribution à l'étude géologique, hydrogéologique et par télédétection des bassins versants sub-sahéliens du socle précambrien de l'Afrique de l'Ouest: Hydrostructurale, hydrodynamique, hydrochimie et isotopie des aquifères discontinus de sillons et aires granitiques de la haute Marahoué (Côte d'Ivoire). Thèse de Doctorat d'état ès Sciences Naturelles, Univ. Abidjan, 493 p.

[12] Wyns, R., Baltassat, J.M., Lachassagne, P., Legchenko, A., Vairon, J. and Mathieu, F. (2004) Application of SNMR Soundings for Groundwater Reserves Mapping in Weathered Basement Rocks (Brittany, France). *Bulletin de la Société Géologique de France*, **175**, 21-34. https://doi.org/10.2113/175.1.21

[13] Archie, G. (1942) The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. *Transactions of the American Institute of Mining Metallurgical and Petroleum Engineers*, **146**, 54-62.
https://doi.org/10.2118/942054-G

[14] Sinan, M. and Razack, M. (2006) Estimation du champ de transmissivité d'un aquifère alluvial fortement hétérogène à partir de la résistance transversale. Application à la nappe du Haouz de Marrakech (Maroc). *Journal of Water Science*, **19**, 221-232. https://doi.org/10.7202/013540ar

[15] Belanger, M., El-Jabi, N., Caissié, D., Ashkar, F. and Ribi, J. (2005) 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*, **18**, 403-421.
https://doi.org/10.7202/705565ar

[16] Kouadio, E., Soro, N. and Savané, I. (2010) Stratégie d'optimisation de la profondeur des forages en contexte de socle: Application à la région du Denguélé, Nord-Ouest de la Côte d'Ivoire. *Revue des sciences de l'eau*, **23**, 1-15.
https://doi.org/10.7202/038921ar

[17] Kouassi, A., Kouakou, K., Koffi, Y. and Biemi, J. (2014) Développement de modèles statistiques pour la prévision du débit de forage d'eau en milieux de socle cristallin et cristallophyllien en Côte d'Ivoire. *International Journal of Innovation and Scientific Research*, **10**, 154-170.

[18] Gombert, P. (1997) Variabilité spatiale de la productivité aquifère du socle sahélien en hydraulique rurale. *Hard Rock Hydrosystems*, Vol. 241, 113-122.

[19] Benyahya, L., Caissie, D., Mertes, L.A.K., O'Daniel, S.J., Woessner, W.W. and Thomas, S.A. (2007) A Review of Statistical Water Temperature Models. *Canadian Water Resources Journal*, **32**, 179-193. https://doi.org/10.4296/cwrj3203179

[20] Daigle, A., St-hilaire, A., Peters, D. and Baird, D. (2010) Multivariate Modeling of

Water Temperature in the Okanagan Watershed. *Canadian Water Resources Journal*, **35**, 237-258. https://doi.org/10.4296/cwrj3503237

[21] Hingray, B., Picouet, C. and Musy, A. (2012) Hydrologie: Une Science pour l'Ingénieur. 2ème édition, Edition Presse Polytechniques et Universitaires, Romandes, 600 p.

[22] Nascimento, O. (1995) Appréciation à l'aide d'un modèle empirique des effets d'actions anthropiques sur la relation pluie débit à l'échelle d'un bassin versant. Thèse de Doctorat, Ecole Nationale des Ponts et Chaussées de Paris, Paris, 390 p.

[23] Kouassi, A.M., N'guessan, B.T.M., Kouamé, K.F., Kouamé, K.A., Okaingni, J.C. and Biémi, J. (2012) Application de la méthode des simulations croisées à l'analyse de tendances dans les relations pluie-débit à partir du modèle GR2M: Cas du bassin versant du N'Zi-Bandama (Côte d'Ivoire). *Comptes Rendus Géoscience*, **344**, 288-296. https://doi.org/10.1016/j.crte.2012.02.003

[24] Mangoua, O. (2013) L'évaluation des potentialités et de la vulnérabilité des ressources en eau souterraine des aquifères fissurés a été initié dans le bassin versant de la Baya (Est de la Côte d'Ivoire). Thèse unique, Univ. Nangui Abrogoua, 199 p.

[25] Dibi, B., Doumouya, I., Goula, B., Savane, I. and Biemi, J. (2004) Analyse statistique des paramètres influençant la productivité des forages d'eau en milieu cristallin et cristallophyllien dans la région d'Aboisso (Sud-Est de la Côte d' Ivoire). *Sud Sciences & Technologies*, **13**, 22-30.

[26] Savane, I. (1997) Contribution à l'étude géologique et hydrogéologique des aquifères discontinus du socle cristallin d'Odienné (Nord-Ouest de la Côte d'Ivoire). Apport de a télédétection et d'un système d'information hydrogéologique à référence spatiale. Thèse de doctorat d'Etat ès Sciences Naturelles, Univ. Cocody, Cocody, 322 p.

[27] Sinan, M. and Razack, M. (2008) An Extension to the DRASTIC Model for Assessing Groundwater Vulnerability to Pollution: Application to the Haouz Aquifer of Marrakech (Morocco). *Environmental Geology*, **57**, 349-363. https://doi.org/10.1007/s00254-008-1304-2

[28] Detay, M. and Doutambaye, C. (1989) Hydrogéologie statistique du socle précambrien de la République Centrafrique: Principaux résultats du programme d'hydraulique villageoise en zone cotonnière. Bullétin de liaison du CIEH, No. 76.