African Journal of Environmental Science and Technology

Full Length Research Paper

Integration of hydro-climatic data and land use in neural networks for modeling river flows: Case of Lobo river in the southwest of Cote d'ivoire

Yao Blaise KOFFI[1]*, Kouassi Ernest AHOUSSI[1], Amani Michel KOUASSI[2], Ouattara KOUASSI[1], Loukou Christophe KPANGUI[1] and et Jean BIEMI[1]

[1]Université de Cocody, Unité de Formation et de Recherche (UFR) des Sciences de la Terre et des Ressources Minières (STRM), 22 BP 582 Abidjan 22, Abidjan, Côte d'Ivoire.
[2]Institut National Polytechnique Félix Houphouët-Boigny (INP-HB), Département des Sciences de la Terre et des Ressources Minières (STeRMi), BP 1093 Yamoussoukro, Côte d'Ivoire.

In a context of climate disruption due to uncontrolled human activities, the classical models of rainfall-runoff modeling are almost unusable. In addition on the Lobo River (Southwest of Côte d'Ivoire), no simulation study has been carried out yet despite that, this river has flooded fields and villages causing huge losses in September 2007. Neural networks appear in this case as a solution for simulating flows in the context of non-linearity between rainfall and flow of this river. Climatic data (rainfall, temperature and PET) and land use will be phased in neural models to simulate monthly flows of the river Lobo. Four (4) neuronal model variants were constructed from three (3) hydro-climatic parameters (rainfall, potential evapotranspiration and flow) and the land acquired from Landsat ETM + 1990 and ETM + 2000. Two types of models have been created: the unguided model and the guided model. The simulation with the unguided model did not provide a satisfactory result. In effect, the value of Nash is only 22.90%. However, the NASH value of the guided model is much better than the previous one (85.01, 83.38 and 84.05%). These results help to highlight the importance of land use on the performance of neural networks. This study also demonstrated the ability of artificial neural networks to simulate the nongauged river flows in the context of climate disruption.

Key words: Côte d'Ivoire, Lobo, multilayer perception, remote sensing, simulation, hydrology, model, neuron, flow.

INTRODUCTION

The South-western region of Côte d'Ivoire has become over the last two decades, the new place where coffee and cocoa flourishes. This has fostered many economic activities in this area. These agricultural activities have led to an increased use of water resources. According to the works of Tahoux (1993), the uncontrolled use of land in this region of Côte d'Ivoire has caused the damage of the forest which led to a breakdown of the rainfall balance and severely disrupted the flow of rivers. In the context of climate change due to uncontrolled human activities, the classical models of rainfall-runoff modeling are almost unusable. It should also be noted that the Lobo river has not been a subject of study despite of the catastrophic events of September 2007 where field and villages were flooded causing huge losses. Neural networks appear in this case as a solution for the flow

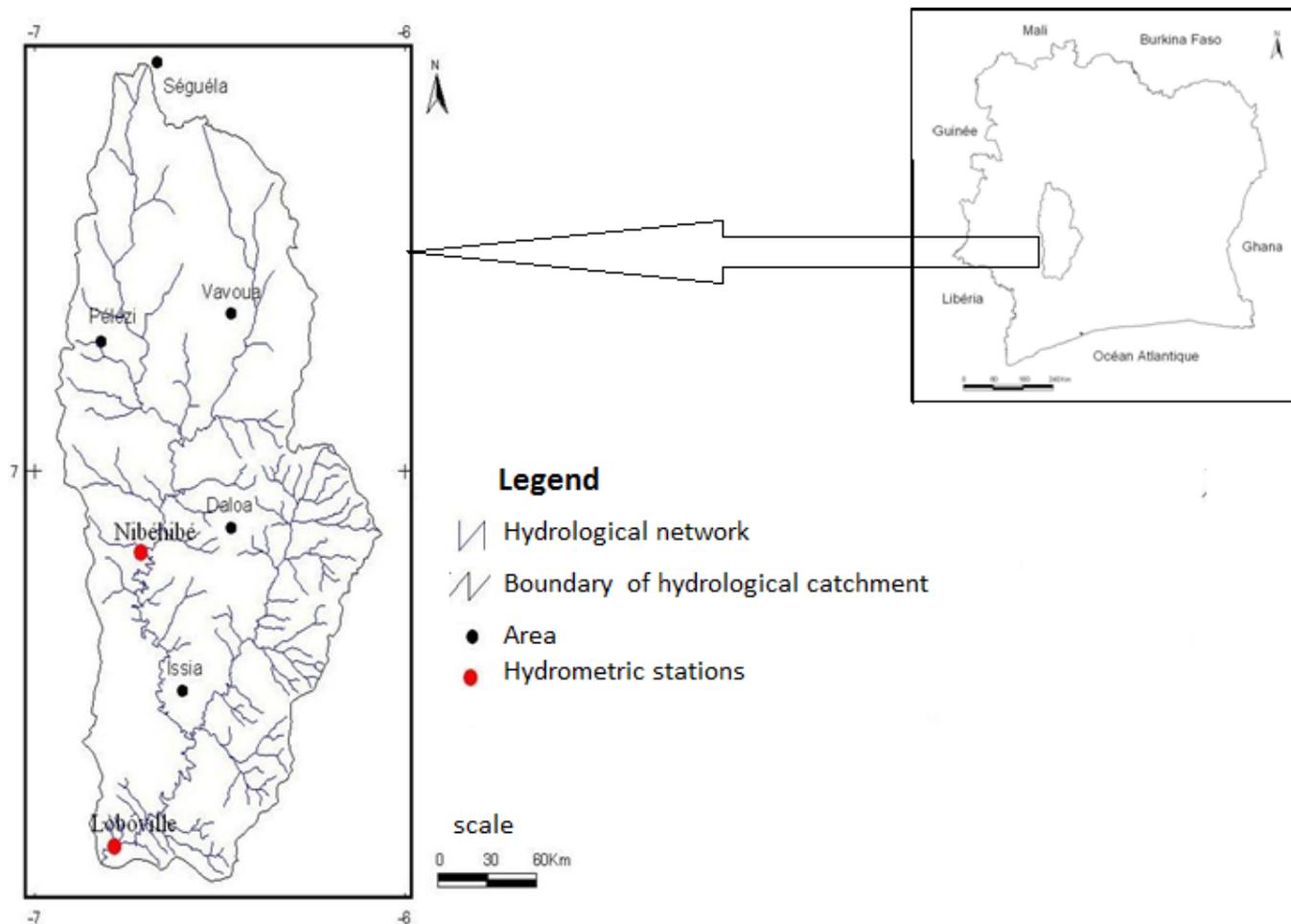*Corresponding author. E-mail: yaomonie@yahoo.fr. Tel: (+225) 22 48 38 03.

Figure 1. Localisation of the hydrological catchment.

simulation in the context of non-linearity between rainfall and river flow. Other methods are linear ones but here we are in a non-linear context. Climatic data (rainfall, temperature and evapotranspiration) and land use will be gradually integrated in artificial neural models to simulate monthly runoff of the Lobo river.

MATERIALS AND METHODS

Study area

The river Lobo is one of the main tributaries of the river Sassandra (Figure 1). The hydrological catchment of this river has two hydrometric stations: the station of Nibéhibé in the center and the station of Loboville in the south. This hydrological catchment covered an area of 12,724 km^2. It is the rainy region with lateritic soils (Adjanohoun et al., 1971). This hydrological catchment is subjected to a transition equatorial attenuated regime.

Data

The monthly rainfall (from 1991 to 2011), the monthly Lobo river flows (from 1991 to 2011) and the land use of Lobo river catchment

has been use in this study. These hydro-climatics data were provided by the Department of Human Hydraulics (DHH) and the Operating Company and Airport Development and Meteorology (SODEXAM). Regarding the land use of Lobo river catchment, it is obtained from the Landsat ETM + 1990 and Landsat ETM + 2000 images. In this study, the input of artificial neural networks like the potential evapotranspiration (PET) was calculated by the method of Thornthwaite. This method although simple, gives very good results (Koffi, 2007; Kouassi, 2007). The general principle of the processing of Landsat ETM + 1990 and Landsat ETM + 2000 is based on the contrasts between the observed object and its environment (Kouassi, 2009). Thus, from all landscape units or objects identified by visual interpretation, we defined four classes of land in case, the habitats, water, crops and forests. Given the inevitable case of confusion, other landscape units were ignored as individual classes because they were part of other more important. This is the case in urban areas (roads, habitats), bare soil (rock outcrops, bare surfaces).

The classification method used in this study is the maximum likelihood which is the subject of learning on a subset of the data for which the class is known.

Modeling by artificial neural networks

The neural network is defined as an assembly of identical structural

Table 1. Artificial neural networks models.

Neural network model	Architecture	Input	Output
Unguided model with input rain	[1 10 1]	Rain (t)	Flow rate (t)
Guided model with input rain an Flow rate	[2 3 1]	Rain (t), Q(t-1),	Flow rate (t)
Guided model with input rain, flow rate and land use	[3 7 1]	Rain (t), Q(t-1) Land use	Flow rate (t)
Guided model with rain, flow rate, PET and land use	[4 4 1]	Rain (t), ETP, Q(t-1), Land use	Flow rate (t)

elements called cells (or neurons) interconnected like the cells of the vertebrate nervous system (Coulibaly et al., 1999; Schmitt et al., 2001; Vazken et al., 2004). It comes from the early work by McCulloch and Pitts (1943) in the field of artificial intelligence for modeling the functioning of the human brain based primarily on the concept of neurons. The data used as input to build different neural networks are composed of: rain (t), the flow rate Q (t-1), potential evapotranspiration (PET) and the land use. These variants of models are reported in Table 1.

Process for model calibration (model calibration and validation)

Before the calibration process, the ANN database has been normalized. It is a process of normalized amplitude values accepted by the network. This allows the model to quickly converge to the desired values. The values of the input variables in the software are between (0 and 1). This database was divided into three groups: calibration, testing and validation with respectively the proportions 50, 25 and 25%. Thus, the data from 1991 to 1998 were used to stall, very important phase of the modeling process by networks of neurons. During this process the parameters of neural models are adjusted using supervised learning.

The values of the connection weights are first created at random and the system searches through successive iterations to obtain data modeling. This alteration of connections is obtained by Levenberg algorithm Marquartd (LM) (Dreyfus et al., 2004). The period 1999 to 2001 was used for the test to see if the network responds are good as desired and the period 2002 to 2004 was used for validation.

Performance criteria

There are several criteria to appreciate the performance of the models in hydrological sciences. Two criteria to measure the numerical error are often used. In this work the Nash Sutcliffe Efficiency (NSE) and the correlation coefficient (R^2) are used. The Nash criterion (Nash and Sutcliffe, 1970) is used to measure the degree of fit between observed and simulated values for the calibration and validation.

$$Nash = \left(1 - \frac{\sum_i (X_i - X_{sim})^2}{\sum_i (X_i - \bar{X})^2}\right) \times 100$$

With: X_i, flow observed; X_{sim}, simulated flow, $\bar{X}$, average flow observed; Nash: Nash criterion.

The correlation coefficient or Pearson coefficient, R^2, is simply the square root of the coefficient of determination, and its sign (±) gives the direction of the relationship. It allows studying the existence of relationship between two variables.

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

With: X_i, flow observed; Y_i, simulate flow, $\bar{X}$, average flow observed; $\bar{Y}$, average rates simulated.

By convention, we say that the relationship between observed flows (Xi) and simulated flow (Yi) is: perfect, if R = 1; very strong, if R> 0.8; strong, if R is between 0.5 and 0.8; average, when R is between 0.2 and 0.5; and low, if R is between 0 and 0.2 (Kachroo, 1986; Koffi et al., 2007).

RESULTS

Land use in the catchment of Lobo river

The statistics supervised classification performed with satellite imagery shows that forests occupy the largest area of the basin. However, results from the analysis show that area occupied by forests decreased by 7.74% from 1990 to 2000. However, water has the lowest percentage and cultivation areas have increased significantly from 30.27 to 32.28% for a period of 10 years. Overall, the proportions of each element decreased in 2000. Table 2 summarizes the proportions of each element used in this study.

Simulations with neural models

During calibration, the model [1 10 1] gives a poor correlation between simulated flows and observed flows (R = 0.70). Also in validation, the correlation is not high (R = 0.40). The poor correlation between the flows calculated by the model and flows observed is also reflected in the process of calibration and validation by the discrepancies between the different hydrographs (the value of Nash is equal to 22.90% (Table 3 and Figure 2A).

Regarding the model [1 3 1], it gives a good correlation between simulated flows and observed flows (R = 0.92) during calibration. This good correlation is also observed during the validation phase where R = 0.94, which is reflected in the evolution of hydrographs. Indeed, some overlap between these two hydrographs is observed (Table 3 and Figure 2B). But there is a difference between those hydrographs in November 2002 and

Table 2. Proportion of land use.

Theme	Image ETM+ 1990	Image ETM+ 2000
Percentage (%)		
% of spaced built	17.67	24.36
% of Water	6.64	5.68
% of crops	30.27	32.28
% of Forest	45.42	37.68
Total	100	100
Area (km^2)		
Spaced built	2.25	3.10
Water	0.84	0.72
Crops	3.85	4.11
Forest	5.78	4.79
Total	12.72	12.72

Table 3. Statistical performances of artificial neural networks models.

	Calibration			Validation		
Model	NNC	Nash	R	NNC	Nash	R
[1 10 1]$_P$	10	32.91	0.60	10	22.90	0.20
[2 3 1]	3	73.53	0.92	3	85.01	0.84
[3 7 1]	7	73.34	0.95	7	83.38	0.83
[4 4 1]	4	75.24	0.97	4	84.05	0.84

November 2003. The addition of a third element to the previous model slightly improves the performance of the model [3 7 1]. The results show a very good correlation between the observed flows and simulated flows during calibration (R = 0.95) and validation (R = 0.93). Simulated and observed hydrographs show a perfect harmony during the calibration, however, for validation, there is a slight discrepancy between the simulated and observed hydrographs (Table 3 and Figure 2C). For the last model [4 4 1] with four (4) hidden neurons, the correlation coefficients R during the calibration and validation move in the same order of magnitude as those of the model [3 7 1] with seven (7) hidden neurons. The correlation coefficients of 0.94 were obtained with this model for calibration and validation.

Hydrographs demonstrate the perfect correlation between the observed flows and simulated flows in both phases (calibration and validation). This model [4 4 1] gives a good difference between the simulated flow and observed flow (Table 3 and Figure 2D). Among the developed model, only the model [1 10 1] is an unguided model. Other models [2 3 1], [3 7 1] and [4 4 1] are guided models. Latter simulate better rates compared to unguided model. However, the model [4 4 1] simulates better compared to others. Value of the correlation

coefficient R obtained with this model is 0.86. This reflects a perfect relationship between the observed flows and simulated flows (Koffi, 2007). But, this study showed an overestimation of low flows and under-estimation of flood flows.

DISCUSSION

This modeling involved two types of models (unguided model [1 10 1] and guided model [2 3 1], [3 7 1] and [4 4 1]). Unguided model [1 10 1] gives a correlation coefficient R below 50%, while guided models give correlation coefficients of 90%. Indeed, guided models are obtained by the use of the output flow (Q t-1) as an input, in contrast to unguided models where flow (Q t-1) is not used. The flow introduction as an input (Q t-1) allows the model to self-correct. According to Koffi (2007), this way of doing allows the network to be able to find the true results. Unguided model [1 10 1] gave a Nash coefficient of 32.91% in calibration and 22.90% in validation values that are well below 50%. This poor unguided model achievement can be explained by the non-linearity between the rainfall and the river flow study. This result supports the works of Dechemi et al. (2003)
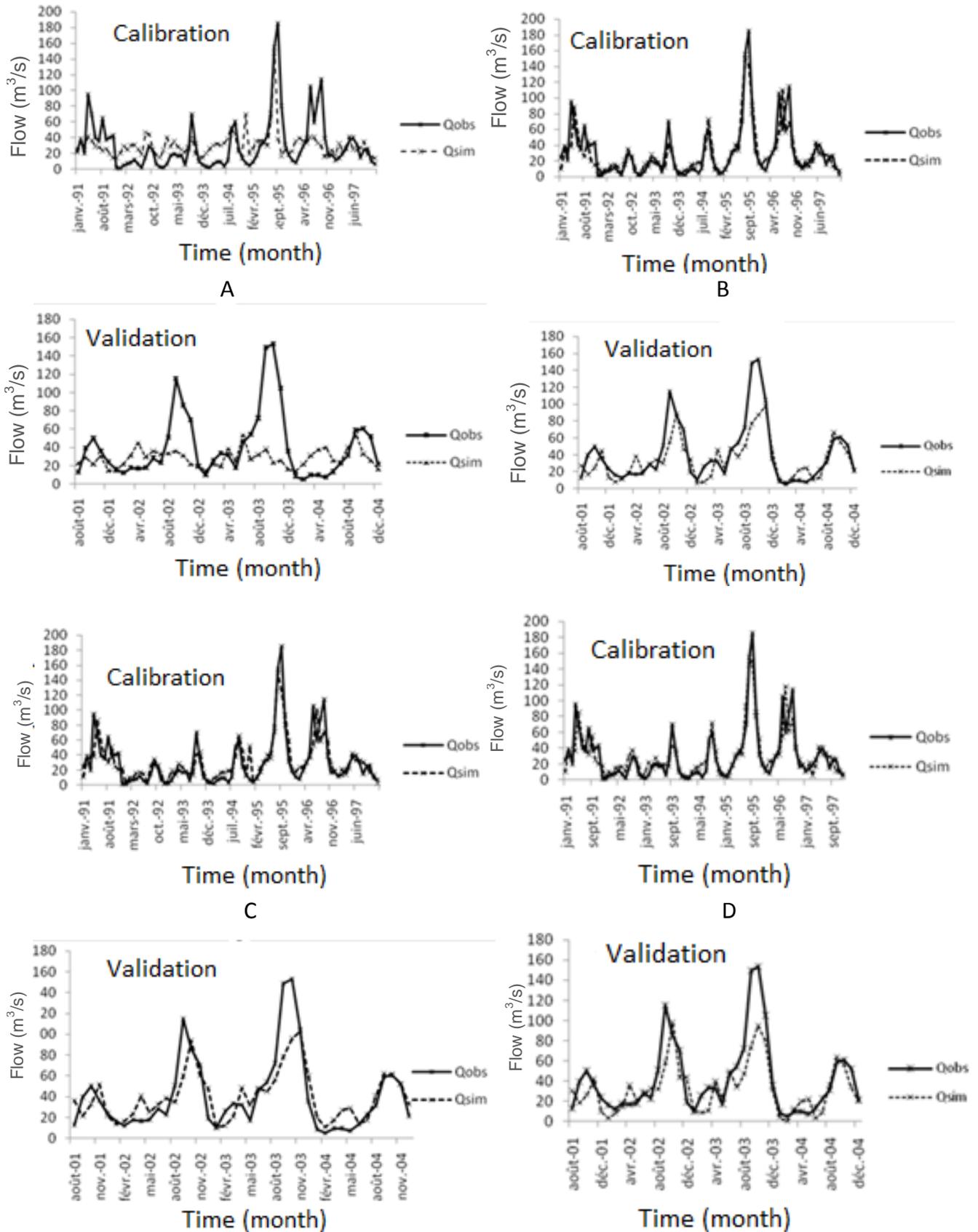
Figure 2. Hydrographs simulated and observed over the entire study period (A, B, C and D).

that show the complexity of the monthly rainfall-runoff relationship. Indeed, it is difficult to quantify some modeling parameters such as evapotranspiration. The performance of the model [4 4 1] show that land use influences the flow of water on a hydrology catchment. The vegetative activity and soil type are related and their combined actions greatly influence the flow surface. The cover holds a variable amount of atmospheric water (Musy, 2005). This idea is also supported by other authors such as Mouelhi (2003) and Brou et al. (1998) who say that human activities and land use also have an influence on the transformation of rainfall into runoff. However, the work of Tahoux (1993) and Brou (2005) have clearly shown that this area is strongly influenced by the migration of people in search of land.

The model [4 4 1] is more representative of how the hydrological catchment of Lobo works. The difference of the architecture model, the quality and the quantity of data may also explain the differences in performance of the models studied.

Conclusion

The integration of physiographic and meteorological variables to the neural networks model was used to model the operation of Lobo hydrology catchment. Based on the ETM + satellite imagery geospatial database was created in order to represent the physiographic characteristics of the Lobo hydrology catchment. The calibration of the model parameters and validations performed on spatio-temporal Lobo catchment shows synchronization between observed and simulated Lobo river flow. The Nash of unguided model is around 85% with a correlation coefficient R of 94%. These values suggest good performance of neural networks and their applicability in hydrology modeling process in a changing climate.

ACKNOWLEDGEMENTS

REFERENCES

Adjanohoun E, Guillaumet JL. (1971). Le milieu naturel de Côte d'Ivoire. Mém. ORSTOM, Paris, n°50. p. 391.

Coulibaly P, Anctil F, Bobee B (1999). Prévision hydrologique par réseau de neurones artificiels: état de l'art. Can. J. Civ. Eng. 26:293-304.

Dechemi N, Benkaci T, Issolah A (2003). Modélisation des débits mensuels par les modèles conceptuels et les systèmes neuro-flous. Revue des sciences de l'eau, rev. Sci. Eau 16/3(2003):407-424.

Dreyfus G, Martinez J-M, Samuelides M, Gordon MB, Badran P, Thiria S, Hérault L (2004). Réseaux de neurones; Méthodologie et application $2^{ième}$ edition. p. 274.

Kachroo RK (1986). HOMS workshop on river flow forecasting, Nanjing, China. Unpublished internal report, Dep. Eng. Hydrol., University College Galway, Ireland. p. 149.

Koffi YB (2007). Modélisation pluie-débit en région tropicale humide : Application des réseaux de neurones sur quatre stations hydrométriques du Bandama Blanc (Bada, Marabadiassa, Tortiya, Bou), Nord de la Côte d'Ivoire). Thèse unique d'hydrogéologie, Université de Cocody Abidjan. p. 233.

McCulloch WS, Pitts WH (1943). A logical calculus ideas immanent in nervous activity. Bull. Math. Biophys. 5:115-133.

Mouelhi S (2003). Vers une chaîne cohérente de modèles pluie-débit conceptuels globaux aux pas de temps pluriannuel, annuel, mensuel et journalier. Thèse de Doctorat, Ecole Nationale du Génie Rural des Eaux et des Forêts de Paris. p. 323.

Musy A (2005). Hydrologie général, chapitre 2: Le basin versant et son complexe $4^{ème}$ semestre. p. 20.

Nash JE, Stucliffe JV (1970). River flow forecasting through conceptual models. Part1. A discussion of principles. J. Hydrol. 10:282-290.

Schmitt A, Le Blanc B, Corsini MM, Lafond C, Bruzek J (2001). Les réseaux de neurones artificiels : un outil de traitement de données prometteur pour l'anthropologie. Bull de Mém de la société d'Anthropologie de Paris, n.s., t. 13, 1-2, 143-150.

Tahoux née Touao KM (1993). Economie de plantation et organisation de l'espace du Sud Ouest ivoirien. Thèse de doctorat $3^{ème}$ cycle IGT, Université de Cocody. p. 558.

Vazken A, Jenni S, Oudin L, Perrir C, Mathevet C, Rosique JL, Tangara M, Anctil F, Lauzon N., Parent E, Penciolelli A, Gaume E, Moulin L, Brun J, Valdes D (2004). Prévision des crues sur le bassin amont de la Loire à partir de modèles pluie-débit globaux et de réseaux de neurones. Rapport final, Programme Risque inondation, Ministère de l'aménagement du territoire et l'environnement. p. 85.