Journal of Water Resources and Ocean Science
2014; 3(2): 22-29
Published online May 30, 2014 (http://www.sciencepublishinggroup.com/j/wros)
doi: 10.11648/j.wros.20140302.12

Monthly predicted flow values of the Sanaga River in Cameroon using neural networks applied to GLDAS, MERRA and GPCP data

SIDDI Tengeleng[1, 2, 4], NZEUKOU Armand[1, *], KAPTUE Armel[5], TCHAKOUTIO SANDJON Alain[1, 3], SIMO Théophile[6], Djiongo Cedrigue[6]

[1]Laboratoire d'Ingénierie des Systèmes Industriels et de l'Environnement (LISIE), Institut Universitaire de Technologie Fotso Victor, Université de Dschang, P.O Box 134 Bandjoun, Cameroun
[2]Higher Institute of the Sahel, University of Maroua, P.O. Box 46 Maroua, Cameroon
[3]Laboratory for Environmental Modeling and Atmospheric Physics (LAMEPA) Department of Physics University of Yaounde 1 P.O Box 812 Yaounde, Cameroon
[4]Laboratory of Mechanics and Modeling of Physical Systems (L2MSP) - University of Dschang, Cameroon
[5]Geographic Information Science Center of Excellence, South Dakota State University, Brookling, SD 57007, USA
[6]Laboratory of Automatic and Applied Informatics (LAIA), IUT-FV, University of Dschang

Email address:
siddit2000@yahoo.fr (SIDDI T.), armand.nzeukou@gmail.com (NZEUKOU A.), armel.kaptue@sdstae.edu (KAPTUE A.), stchakoutio@yahoo.com (TCHAKOUTIO S. A.), sitheo1@yahoo.fr (SIMO T.) cedrigueboris.djiongokenfack@gmail.com (Djiongo C.)

To cite this article:
SIDDI Tengeleng, NZEUKOU Armand, KAPTUE Armel, TCHAKOUTIO SANDJON Alain, SIMO Théophile, Djiongo Cedrigue. Monthly Predicted Flow Values of the Sanaga River in Cameroon Using Neural Networks Applied to GLDAS, MERRA and GPCP Data. *Journal of Water Resources and Ocean Science.* Vol. 3, No. 2, 2014, pp. 22-29. doi: 10.11648/j.wros.20140302.12

Abstract: The aim of our study is to predict the discharge rate of the river Sanaga using neural network techniques. Our investigations have taken place in the Sanaga watershed area in Cameroon. The measurement station is situated in the locality of Edea-Song-Mbengue (04°04'15"N, 10°27'50"E) where we have obtained monthly values of the river Sanaga discharge rates that have been measured in situ from January 1989 to December 2004. We have trained neural networks (NN), each with data of parameters such as the surface albedo, the total cloud fraction, the evaporation, the outgoing longwave radiation, the air temperature, the specific humidity, the surface runoff and the precipitation height. The precipitation values have been obtained from GPCP (Global Precipitation Climatology Project) and those of the other parameters from the data assimilation systems GLDAS (Global Land Data Assimilation System) and MERRA (Modern Era-Retrospective analysis for Research and Application). As desired outputs of the NN during the learning process, we have used the measured river runoff values. After introducing temporal delays of 01 and 02 months in the learning-process, we could observe the presence of the memory effect of the parameters used on the temporal evolution of the river discharge rate. After analysis of the performance's criteria of the NN with the help of the calculated Root Means Square Errors (RMSE) and determination coefficients between predicted values and in situ observed ones, we have perceived that the NN which takes into account the two-month delay can predict the river discharge rate with a strong correlation.

Keywords: River Runoff, GLDAS, GPCP, MERRA, Neural Network, Sanaga Watershed area

1. Introduction

In sub-Saharan Africa in general and in Cameroon, in particular, electrical energy is mostly produced by hydroelectric dams. The mastering of the temporal evolution of the discharge rate of the rivers concerned allows better adapting the generators used. The river discharge rate is a hydrological parameter that can be nowadays evaluated by modern measurement instruments [1-3]. However, the acquirement, the maintenance and the operability of these logistics lead to significant costs that constitute a handicap in evaluating this parameter in sub-Saharan Africa. Collecting direct measurement of stream flow (discharge) on a continuous basis is challenging, especially during large flood events [4]. A common practice is to convert records of water-stages into discharges by using a pre-established stage-discharge relationship. Such relationships are often referred to as a rating curve. Unfortunately, the stage-discharge relationship is not always a simple unique relationship. The rating-curve has to be actualized more times in accordance with the number of the possible runoff values. Furthermore, some factors are able to modify temporarily or definitely the water flow and consequently the rating-curve [5]. Thus, it is important to regularly measure the river runoff in order to determine the relationship stage-discharge and to control its variation [6]. It is also important to emphasize that the forecast of river discharge in a given region allows better managing the problems relied to water such as floods and droughts [7].

The runoff values of a river can be estimated after using some remote sensed information [8-9]. These values could be influenced by some atmospheric parameters and soil properties like the surface albedo, the total cloud fraction, the evaporation, the outgoing long wave radiation, the air temperature, the specific humidity, the surface runoff and the precipitation height.

The estimation of atmospherical and hydrological parameters with the help of neural networks has been in the past and is still now an interesting subject of scientific research. Scientists like Deming et al. [10] have used them to validate atmospherical profiles of temperature; Moreau et al. [11] have developed a neural network named « Gated Expert » (GE) in order to retrieve liquid water content of the atmosphere over the ocean, using radiometric data; Lek et al. [12] have used the neural model to establish the relationship river runoff-rain; Tesch and Randeu [13] and Laurence et al. [14] have used them to predict the river discharge rate for short-term.

The aim of this work is to use the neural network techniques to predict the discharge rate of the river Sanaga in Cameroon, precisely at the measurement point Song-Mbengue, through assimilated data issued from data bases MERRA, GLDAS and GPCP. In the following sections, we will briefly give the functioning principle of the neural networks and present thereafter the study zone which is the watershed of Sanaga basin, the data used, the methodology required by the developed neural networks and will then end by presenting the obtained results.

2. Brief Description of Functioning Principle of Neural Networks

The neural network (NN) is formed by simple computation elements, the neurons that are connected to others by weights (synapses). During the learning phase, it learns the statistical relationship between input- and output data and determines the weights that have to be attributed to each connection between the neurons in order to approximate with accuracy this relationship [15].

In opposition to traditional methods, the neural network does need neither a detailed comprehension of the physics behind the problem that has to be solved nor a step by step conception of a detailed model describing this physical phenomenon. Only a general comprehension of the observation and measure technique is required in order to determine the most appropriate input parameters that statistically permit a large coverage of possible cases. Once trained, the algorithms possess a higher execution speed as the physical methods that need fastidious computation steps for each inversion [16]. The neural network method has been developed and implemented by many authors and can be considered nowadays as a powerful tool for estimation in several scientific research domains [17-18].

The structure of a neural network can be defined as an oriented graph whose nodes represent the neurons and the bows the different connexions between them. The nodes are gathered in input, hidden, and output layers (see figure 1).

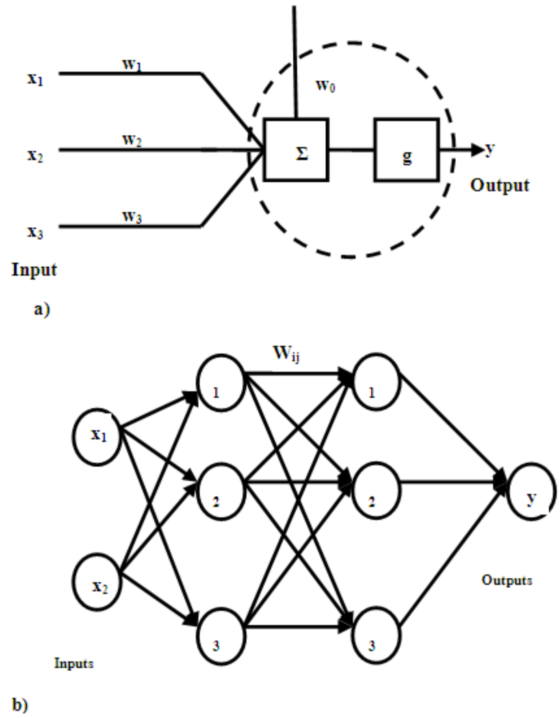

Figure 1. *a) a neuron model; b) an example of a neural network with one input layer (02 inputs), two hidden layers (03 neurons in each) and one output layer (01 neuron)*

In a given neural network, the neuron receives and adds n inputs x_j, each possessing a connection weight w_i. To the result a bias w_0 is added. The final obtained result is applied to a transfer function (or activation function) g in order to obtain the output y. If X is the input vector, the output can be calculated as follows:

$$g : IR^n \rightarrow IR$$
$$y(X) = g\left(\sum_{i=1}^{n} w_i \cdot x_i + w_0\right) = g\left(\sum_{i=0}^{n} w_i \cdot x_i\right) = g\left(W^T \cdot \tilde{X}\right) \quad (1)$$

Whereby

$$x_0 = 1,\ W = (w_0, w_1, \cdots, w_n)^T \text{ and } \tilde{X} = (1, x_1, \cdots, x_n)^T \quad (2)$$

3. Study Zone and Used Data

3.1. Study Zone

Cameroon is situated in central Africa, between the 1st and the 13th degree north latitude, and between the 8th and the 17th east longitude (see figure 2a). The country is characterized by a dense hydrographical network. During the whole year, there is water in the Sanaga, but the discharge rate of this river varies from one year to another. As study zone, we have focused on the watershed area of Sanaga that is approximately delimited by the geographical coordinates (3,51°N, 10.10°E) and (6,33°N, 14,42°E) (see figure 2b).

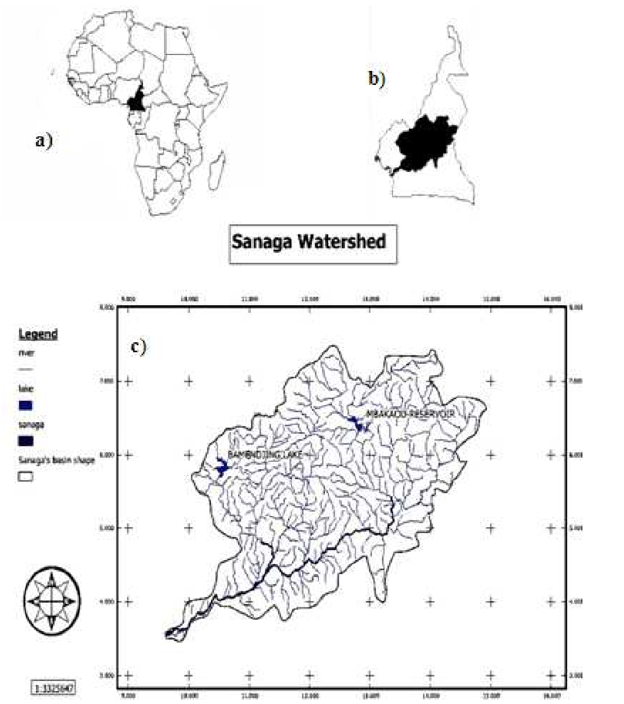

Figure 2. *Sanaga Watershed area; a) localization of Cameroon in Africa; b) localization of watershed area of Sanaga in Cameroun ((3,51°N, 10.10°E); (6,33°N, 14,42°E)); c) Illustration of the watershed area of Sanaga (measure point in Song-Mbengue : 04°04'15"N, 10°27'50"E).*

3.1. Used Data

At disposal, we have monthly values of the discharge rate of Sanaga (m^3/s) obtained in situ in the measurement station of Song-Mbengue, between January 1989 and December 2004. We have also used the data assimilation systems MERRA (Modern Era-Retrospective analysis for Research and Application), GPCP (Global Precipitation Climatology Project) and GLDAS (Global Land Data Assimilation System) [19-22].

From MERRA, we have obtained the values of the surface albedo, total cloud fraction, evaporation, outgoing longwave radiation, air temperature, specific humidity and surface runoff. From GPCP are issued the monthly accumulated rain; and from the GLDAS system, we have obtained the surface runoff.

The general characteristics of the monthly parameter means are represented in the figure 3. They concern the standardized monthly mean values from January 1989 to December 2004. Table 1 presents the statistical values of the mean, standard deviation and variation coefficients.

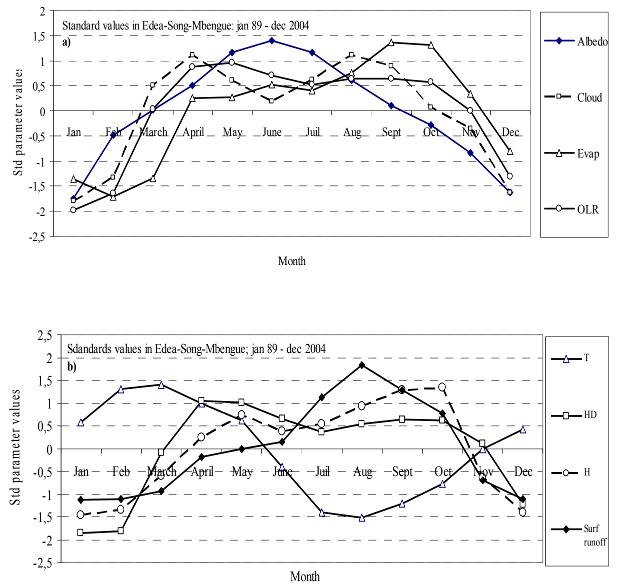

Figure 3. *Evolution of standardized monthly parameter values in the watershed area of the river Sanaga: a) albedo; cloud fraction; evaporation and OLR (Outgoing Long Radiation); b) temperature; humidity; accumulated precipitation and surface runoff.*

Table 1. *Statistical parameter values: mean, standard deviation and variation coefficient concerning the period from January 1989 to December 2004 in the region of the Sanaga watershed area (Cameroon)*

Statistical parameters	**Albedo [fraction]**	**CLDTOT [fraction]**	**EVAP [$kg/m^2/s$]**	**OLR [W/m^2]**
μ (mean)	0.123	0.826	4.60E-05	385.02
σ (std dev)	0.0058	0.0572	4.00E-06	13.57
CV (var coef)	0.047	0.0692	0.087	0.035
Statistical parameters	**T [K]**	**HD [kg/kg]**	**H [mm]**	**Surf runoff [$kg/m^2/s$]**
μ (mean)	289.30	0.01519	140.993	6.14E-07
σ (std dev)	1.162	0.001641	88.1979	5.16E-07
CV (var coef)	0.00402	0.10808	0.6255	0.840

4. Methodology

Training of a neural network is an iterative process that ends if the NN has sufficiently learnt the relationship existing between the inputs, and the outputs presented to him. Many stop-criteria exist, but the main important ones are the iteration number, and the error calculated between the desired values and those evaluated by the NN.

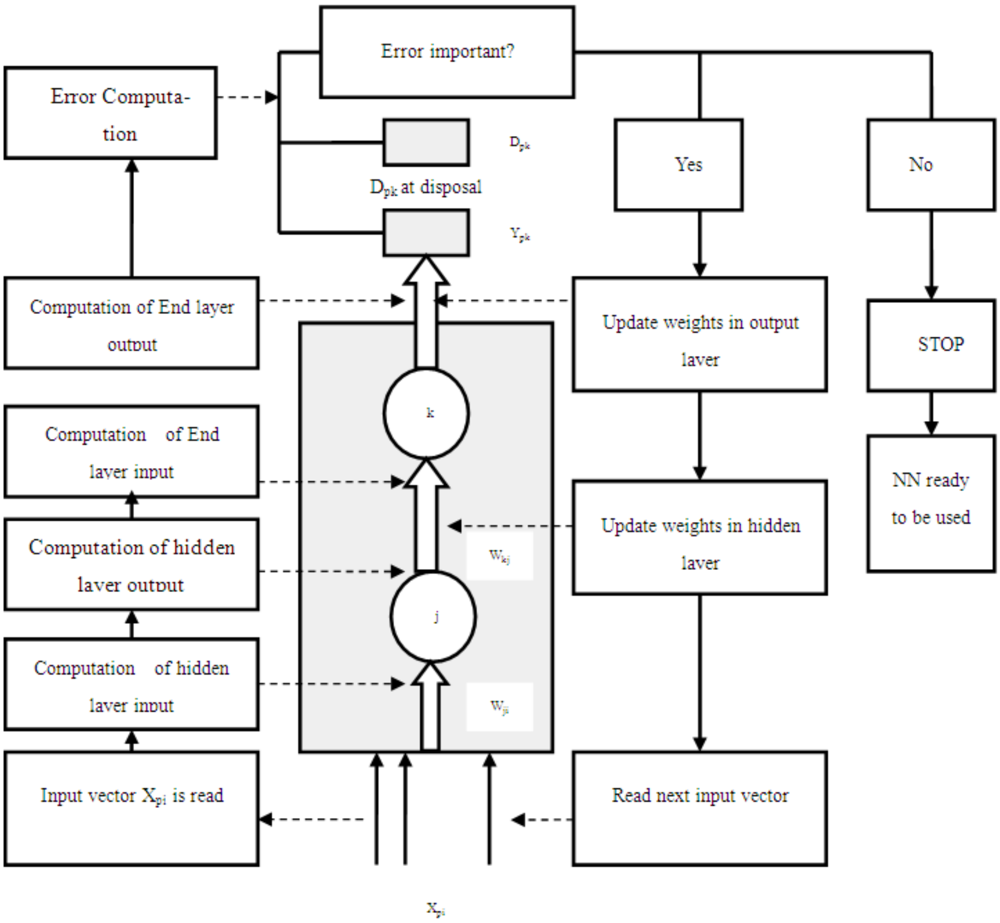

Figure 4. *Learning process of a neural network*

For a convenient prediction of the outputs corresponding to given inputs of a trained NN, it is necessary, during the learning process, that all possibilities of inputs / outputs combinations exist. In the figure 4, we have presented a diagram describing the learning process of a NN which can be condensed in 05 main points:

1. From the input matrix, X_{pi} is an input vector x_{pi} read (sample);
2. Input- and output values of existing hidden layers are calculated;
3. Input- and output values of the last layer (output layer) are computed;
4. The error between desired available value and NN output is evaluated;
5. If the calculated error is not significant, the learning process is stopped, and the NN can be considered as trained and used as a prediction tool for other outputs corresponding to inputs that are, up to now, unknown by the NN. If the evaluated error is important, the next input vector (sample) is read from the available input matrix, and a new iteration begins.

For the learning of the different parameters used, we have employed the neural network toolbox available in MATLAB [23-24].

Sometimes, in accordance with different stages of the parameter values used, it is necessary to work with normalized input values situated between -1 and +1. For this purpose, we have used the normalized NN input values of parameters (equation 3):

$$X_{norm} = 2 \cdot \frac{X - X_{min}}{X_{max} - X_{min}} - 1 \qquad (3)$$

Whereby X, X_{min} and X_{max} represent respectively the real value of the parameter, its minimal value and its maximal value.

As inputs parameters of the NN we have: the albedo (%), the cloud coverage (%) (CLDTOT), the evaporation (EVAP) (kg/m^2/s), the outgoing long radiation (OLR) (W/m^2), the air temperature T (°K); the specific humidity (HD) (kg/kg), the accumulated rain (H) (mm) and the surface runoff (SURF RUNOFF) (kg/m^2/s). We have obtained the values from the NASA website and through the data assimilation systems MERRA, GPCP and GLDAS. All these NN input values correspond to the period from January 1989 to December 1996. As outputs of the three NN, we have used as desired values, during the learning process, the in situ measured river discharge rates respectively from January 1989, from February 1989 and from March 1989, to December 1997. With the first NN, we have predicted the river discharge rates as from January1997 to November 2004, without any time delay accordingly to the training period. The second and the third NN predict these values respectively with a temporal delay of 01 (as from February 1997 to October 2004) and 02 months (as from March 1997 to September 2004). The temporal delays that we have introduced allow perceiving the memory effect of the used input parameters on the evolution of the temporal river discharge rate. Table 2 gives the different periods corresponding to the training and to the prediction phases.

Table 2. *Training and prediction periods for the three NN*

NN	Training period	Prediction period
1	January 19889 – December 1996	January 1997 – November 2004
2	January 19889 – December 1996	February 1997 – October 2004
3	January 19889 – December 1996	March 1997 – September 2004

For better understanding the efficiency of the different neural networks during the prediction of the river Sanaga discharge rate in Song-Mbengue, we have evaluated the performance criteria through computation of the Root Mean Square Error (RMSE) (equation 4) and determination coefficient (r) (equation 5):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(x_i^n - x_i^s\right)^2}{N-1}} \tag{4}$$

$$r = \frac{\sum_{i=1}^{N}\left(x_i^n - \overline{x}^n\right)\cdot\left(x_i^s - \overline{x}^s\right)}{\sqrt{\sum_{i=1}^{N}\left(x_i^n - \overline{x}^n\right)^2}\cdot\sqrt{\sum_{i=1}^{N}\left(x_i^s - \overline{x}^s\right)^2}} \tag{5}$$

Whereby x_i^n and x_i^s are respectively the NN simulated values, the in situ measured ones and N the total number of samples.

5. Results and Discussions

We have trained for each temporal delay three NN with 02 hidden layers (8-10-10-1; 8-10-20-1; 8-20-10-1). In this designation the first number represents the number of inputs, the second the number of neurons in the first layer, the second the number of neurons in the third layer and the fourth the number of outputs of the NN. We have observed that the structure 8-10-10-1 is better trained as the other structures. With the 02 months' time delay, we have compared in figure 5 the predicted and the measured river discharge rate values. We have found a strong correlation of 0.95 orders.

5.1. Training Phase of the Neural Networks

We have trained for each temporal delay three NN with 02 hidden layers (8-10-10-1; 8-10-20-1; 8-20-10-1). In this designation the first number represents the number of inputs, the second the number of neurons in the first layer, the second the number of neurons in the third layer and the fourth the number of outputs of the NN. We have observed that the structure 8-10-10-1 is better trained than other structures. With the 02 months' time delay, we have compared in figure 5 the predicted and the measured river discharge rate values. We have found a strong correlation of 0.95 orders.

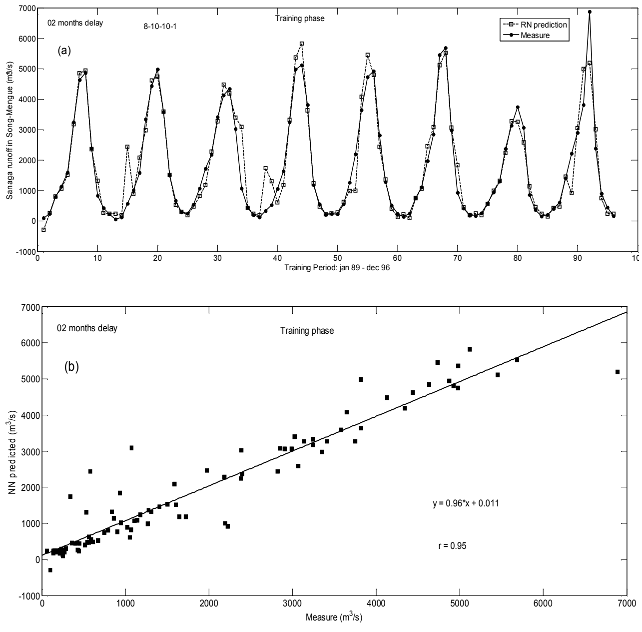

Figure 5. *Training of the neural network 8-10-10-1 with a time delay of 02 months (8 input parameters, 10 neurons in the 1st hidden, 10 neurons in the 2nd hidden layer (see the legend)*

5.2. Prediction Phase of the Neural Networks

After execution of the input values through the trained NN we can observe that the 02 months delay leads equally to better predictions. Thus, we obtain for the period from March 1997 to September 2004 the predicted river discharge values of Sanaga at the measurement point Song-Mbengue that we compare in figure 6 to the values observed in situ. We obtain equally the existence of a strong correlation in order of 0.89 between predicted and measured values.

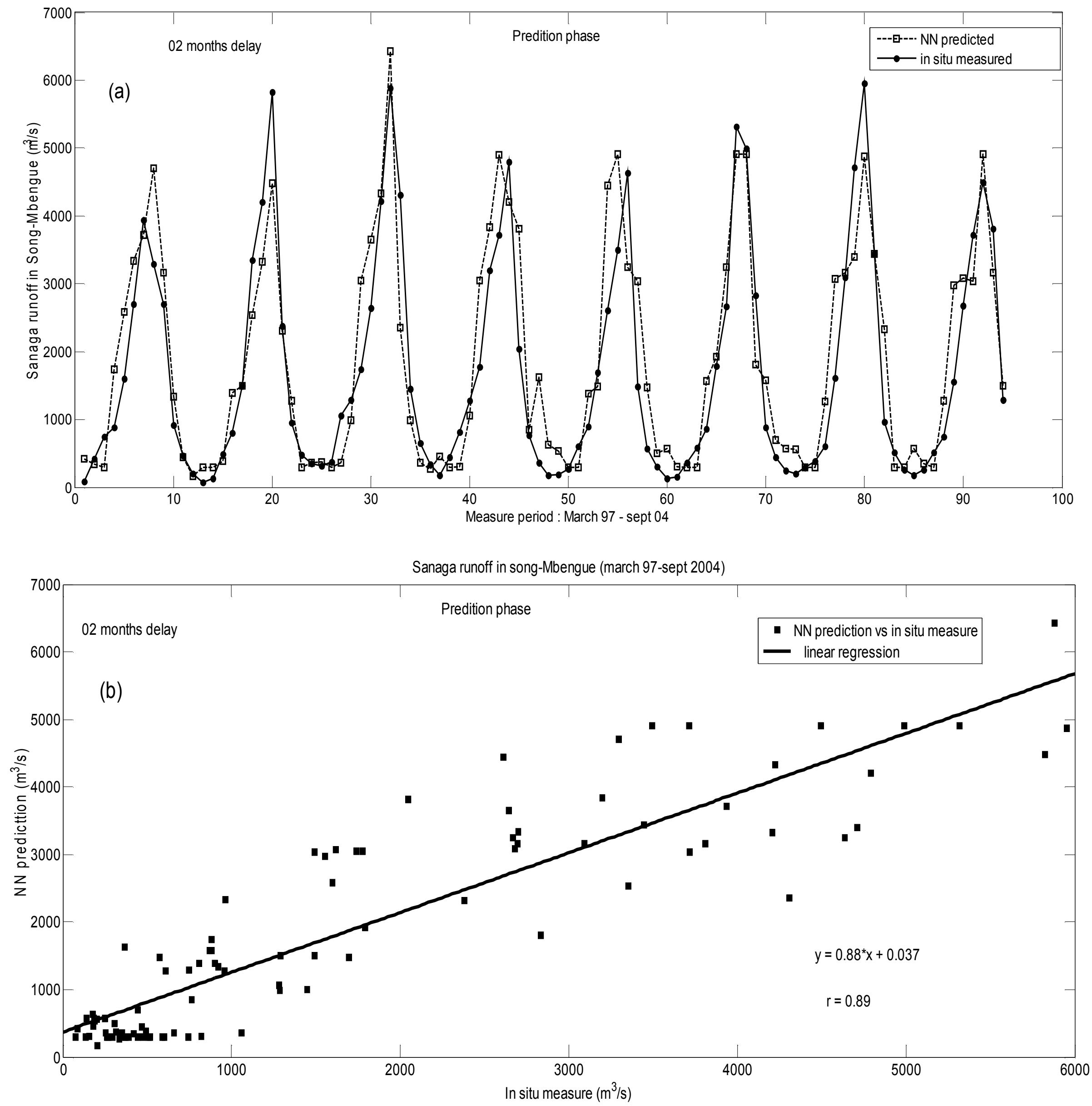

Figure 6. *a) Prediction with the 8-10-10-1 NN (squares) and measures (points) of the Sanaga discharge rate at Song-Mbengue in Cameroon during the period from March 97 to September 2004; b) Linear regression between measured and predicted values.*

The table 3 presents the different values of the performance criteria RMSE and r obtained

Table 3. *Root Mean Square Errors (RMSE) and Determination Coefficients (r) between the discharge rate values, calculated by the NN and measured in situ: training phase from January 1989 to December 1996; prediction phase from January 1997 to December 2004; use of the configuration 8-10-10-1 for all the three temporal delays (0 month, 01 months and 02 months delay)*

Delay time (months)	Training phase			Prediction phase	
	RMSE (m^3/s)	r	Iteration number needed	RMSE (m^3/s)	r
0	439.0	0.91	12	1056.41	0.80
1	434.0	0.92	17	882.00	0.83
2	430.9	0.95	18	735.50	0.89

6. Conclusion

In our study, we have investigated the prediction of the discharge rate of the river Sanaga in Cameroon through using assimilated data from GLDAS, MERRA and GPCP. A neural network technique has been used after the training of the models with values observed in the measurement station Song-Mbengue.

While developing the NN, several architectures have been tested, and the configuration (8-10-10-1) has been identified as the best one, independently on the temporal delay introduced. Furthermore, after analyzing the performance criteria (RMSE and r between predicted values and in situ measured ones), we have perceived that the NN which takes into account the temporal delay of 02 months produces better prediction of the river discharge rate with strong correlation.

Considering that the measures of river discharge rate values need enormous material and human resources, the use of NN models based on data from assimilation systems like GLDAS, MERRA and GPCP appears as an adequate, efficient and rapid mean with low cost that permits to predict the discharge rate values of rivers of Sanaga watershed area in Cameroon with a good spatio-temporal regularity.

Acknowledgments

We thank NASA through its service data (Goddard Earth Sciences Data and Information Services Center (GES DISC)).

Nomenclature

CLDTOT: Cloud coverage
CV: variation coefficient
EOS: Earth Observing System
EVAP: Evaporation
GES DISC: Goddard Earth Sciences Data and Information Services Center
GLDAS: Global Land Data Assimilation System
GOES: Geostationary Operational Environmental Satellite
GPCP: Global Precipitation Climatology Project
HD: Humidity
HDISC: Hydrology Data and Information System Center
μ: Mean value
MERRA: Modern Era-Retrospective analysis for Research and Application
NASA: National Atmospheric and Space Administration
NN: Neural Network
OLR: Outgoing Long Radiation
RMSE: Root Mean Square Error
σ: Standard deviation
Surf runoff: Surface runoff
T: Air temperature

References

[1] Chen, Y., C. "Flood discharge rate measurement of a mountain river - Nanshih River in Taiwan". Hydrol. Earth Syst. Sci., 2013, 17, 1951–1962.

[2] Negrel, P., Kosuth, P. and Bercher, N. "Estimating river discharge from earth observation measurements of river surface hydraulic variables". Hydrol. Earth Syst. Sci., 2011, 15, 2049–2058.

[3] EOS. "Space-Based Measurement of River Runoff Transactions", American Geophysical, Union. Eos, Vol. 86, No. 19, 10 May 2005.

[4] Emad, H. Habib and Ehab A. Meselhe. "Stage-discharge for Low-Gradient Tidal Streams Using Data-driven Models". Journal of Hydraulic Engineering, ASCE/May 2006.

[5] Lillian, Oygarden. "Erosion and surface runoff in small agricultural catchments". IAHS Publ. 1996, no. 236, Global and Regional Perspectives Centre for Soil and Environmental Research

[6] Watanabe, Fumio, Kobayashi, Yukimitsu, Suzuki, Shinji, Hotta, Tomoki and Takahashi, Satoru. "Estimating the Volume of Surface Runoff from in Situ Measured Soil Sorptivity". Journal of Arid Land Studies, 2012, 22 (1), 95-98.

[7] Pekarova, P., Pavol, Miklanek and Jan Pekar. "Spatial and temporal runoff oscillation analysis of the main rivers of the world during the 19th – 20th centuries". Journal of Hydrology, 2003, 274, 62-79.

[8] Bjerkliea David, M., Delwyn, M., Laurence, C. Smith, Lawrence Dingman. "Estimating discharge rate in rivers using remotely sensed hydraulic information". Journal of Hydrology, 2005, 309, 191–209.

[9] Tarpanelli, Angelica, Barbetta, Silvia, Brocca, Luca and Moramarco, Tommaso. "Discharge Estimation by Using Altimetry Data and Simplified Flood Routing Modeling". Remote Sens., 2013, 5, 4145-4162; doi:10.3390/rs50941 45

[10] Deming J., Chaohua D. and Weison L. "Neural networks approach to high vertical resolution atmospheric temperature profile retrieval from space borne high spectral resolution infrared sounder measurements". Proc. SPIE 6064, Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning, 2006, 60641L; doi:10.1117/12.649743.

[11] Moreau, E., Mallet, C., Thiria, S., Mabbou, B., Badran, F. and Klapisz, C. "Atmospheric Liquid Water Retrieval Using a Gated Experts Neural Network". Journal of Atmospheric and Oceanic Technology, 2002, 19, 457-467.

[12] Lek, S., Dimopoulos, I., Ghachtoul, Y., El. « Rainfall-runoff modelling using artificial network". Revue des sciences de l'eau, rev. Sci. Eau, 1996, 3, 319-331.

[13] Tesch, R. and Randeu, W., L. "A neural network model for short term river flow prediction". Nat. Hazards Earth Syst. Sci., 2006, 6, 629-635.

[14] Laurence, C., Smith and Tamlin, M., Pavelsky. “Estimation of river discharge rate, propagation speed, and hydraulic geometry from space: Lena River, Siberia”. Water Ressources Research, 2006, Vol. 44, doi:10.1029/2007 WR006133.

[15] Shrivastava, G., Karmakar, S., Guhathakurta, P. and Manoj Kumar Kowar, M.K.). “Application of Artificial Neural Networks in Weather Forecasting: A Comprehensive Literature Review”. International Journal of Computer Applications, 2012, 51, (18), 0975-8887.

[16] Del Frate, F. and Giovanni S. “Neural Networks for the retrieval of water vapour and liquid water from radiometric data”. Radio Science, 1998, 33, (5), 1373-1386.

[17] Widrow, B., Lehr, M.A. “30 years of adaptive neural networks: Perceptron, Madaline, and Backpropagation”. Proceedings of the IEEE, 1990, (9), 78.

[18] Kou-Lin Shu, Xiaogang Gao, Sorooshi Sorooshia and Hoshin V. Gupta. “Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks”. Journal of applied meteorology, 1996, 36, 1176-1189.

[19] Huffman, G. J., Bolvin, D. T. “Version 1.2 GPCP One-Degree Daily Precipitation Data Set Documentation. Mesoscale”, Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, 2013.

[20] Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D. and Toll, D. “The Global Land Data Assimilation System”. Bulletin of American Meteorological Society, 2004, 381-394, doi: 10.1175/BAMS-85-3-381.

[21] Hongliang, F., Hrubiak, P. L., Hiroko, K., Rodell, M., Teng, W. L. and Vollmer, B.E. “Global Land Data Assimilation System (GLDAS) products from NASA, Hydrology Data and Information Services Center (HDISC)”. ASPRS Annual Conference, Portland Oregon, April 28 - May 2, 2008.

[22] Rui, H., Teng W. L., Vollmer B.E., Mocko, D.M., Beaudoing, H.K. and Rodell, M. “NASA Giovanni Portals for NLDAS/GLDAS Online Visualization, Analysis and Intercomparison”. American Geophysical Union, Fall Meeting, 2011.

[23] Hagan, M., Demuth, H., and Beate, M. “Neural Network design. Boston”: Pws publication, 1996.

[24] Demuth, H., Beate, M., and Hogan, M. “Neural Network toolbox user’s guide”. The Mathworks. Inc., Natrick, USA, 2009.