

Modélisation de la relation pluie-débit à l'aide des réseaux de neurones artificiels

Rainfall-Runoff Modelling Using Artificial Neural Networks

S. Lek, I. Dimopoulos, M. Derraz and Y. El Ghachtoul

Volume 9, Number 3, 1996

URI: <https://id.erudit.org/iderudit/705255ar>

DOI: <https://doi.org/10.7202/705255ar>

[See table of contents](#)

Publisher(s)

Université du Québec - INRS-Eau, Terre et Environnement (INRS-ETE)

ISSN

0992-7158 (print)

1718-8598 (digital)

[Explore this journal](#)

Cite this article

Lek, S., Dimopoulos, I., Derraz, M. & El Ghachtoul, Y. (1996). Modélisation de la relation pluie-débit à l'aide des réseaux de neurones artificiels. *Revue des sciences de l'eau / Journal of Water Science*, 9(3), 319–331.
<https://doi.org/10.7202/705255ar>

Article abstract

Identification of the elementary processes of the hydrological cycle in a drainage basin, and the comprehensive description of each of them, lead to hydrological models with a complex structure including a high number of relatively inaccessible parameters. Moreover these processes, even when simplified, are generally non-linear. Using models with a smaller number of parameters, in order to cope with non-linearity, is therefore necessary.

In this perspective, we propose an artificial neural network for rainfall-runoff modeling. Performances of this method in non-linear modeling have been already demonstrated in several scientific fields (biology, geology, chemistry, physics). In the present work, we use the error back-propagation algorithm with a three-layer neural network. The transfer functions belong to the sigmoidal type at each layer. To predict the runoff at a given moment, the input variables are the rainfall and the runoff values observed for the previous time period. The structure of the network (number of hidden nodes, learning coefficient and momentum values) is optimized to guarantee a good prediction of the runoff, using a set of test data (validation set) not used in the training phase.

Data compiled in our model are a ten year set of rainfall-runoff values collected by the Rabat hydraulic administration (September 1983 to April 1993) in the Beth Wadi catchment. In this study, we develop two types of models according to two different time steps (daily and weekly). The data are subdivided into two sets: a first set to train the model (training set) and a second set to test the model (validation set). For the daily timestep model, we used data of the last two years: April 1991 to April 1993. The initial 365 data (April 1991 - April 1992) constitute the training set and the 365 remaining data constitute the validation set. For the weekly data (Monday to Sunday averages), we have 502 pairs of values. We worked by preserving the last 120 values as the validation set and trained the neural network with the remaining data, i.e. 382 pairs of values of weekly rainfall-runoff.

Three types of estimation have been carried out:

1. *at instant* prediction: prediction of runoff at time t taking into account rainfall values at time t , as well as runoff and rainfall values at preceding times (until $t-1$);
2. *one step ahead* prediction: prediction of runoff at time t from rainfall and runoff values at the preceding times (until $t-1$);
3. *multistep* prediction: prediction of runoff values for a period from $t-jh$ until t , given that values of the runoff for the period 1 to $t-jh-1$ and values of the rainfall at times 1 to t are available (h is the timestep).

The step time is daily for the *at instant* prediction and weekly for *one step ahead* and *multistep* predictions. The choice of input variables is determined by autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses on runoff values, and cross-correlation function (CCF) analysis between rainfall and runoff values. For the *at instant* prediction, the input vector is composed by runoff values of the four days preceding day t , and rainfall values for the three last preceding days as well as its value on day t . For the *one step ahead* prediction, the input vector is composed of runoff values of the five weeks preceding week t , and rainfall values for the three preceding weeks (without considering the rainfall at time t). Finally, for the *multistep* prediction, the input vector is the same as for the *one step ahead* prediction but rainfall values include time t . The runoff values for the week $t-jh+1$, as well as for the following weeks, are computed by feed backing to the input vector the runoff value predicted for the preceding week.

The rainfall-runoff models allow a good estimation for one or several timesteps, daily as well as weekly. In the validation set, correlation coefficients between observed and estimated values are high. In the *at instant* prediction, we obtain the Pearson correlation coefficient $R=0.772$ and the Spearman correlation coefficient $CR=0.958$. The weak value of R as compared to CR is explained by a few extremely high values of error of prediction. In the *one step ahead* prediction ($R=0.887$ and $CR=0.782$) and *multistep* prediction ($R=0.908$ and $CR=0.727$), the R coefficients are higher than CR . This confirms that predicted values are in good agreement with the peaks of observed values (absence of large exceptional errors). In all cases, the results obtained are better than those obtained with linear methods. The neural network models can thus be recommended for time series studies in environmental sciences.

Modélisation de la relation pluie-débit à l'aide des réseaux de neurones artificiels

Rainfall-runoff modelling using artificial neural networks

S. LEK¹, I. DIMOPOULOS¹, M. DERRAZ², Y. EL GHACHTOUL³

Reçu le 19 juin 1995, accepté le 27 novembre 1995.

SUMMARY

Identification of the elementary processes of the hydrological cycle in a drainage basin, and comprehensive description of each of them, lead to models with a complex structure including a high number of hardly accessible parameters. Moreover these processes, even when simplified, are generally non-linear. Using models with a smaller number of parameters, in order to cope with non-linearity, is therefore necessary.

In this perspective, we propose an artificial neural network for rainfall-runoff modelling. Performances of this method in the non linear modelling have been already demonstrated in several scientific fields (biology, geology, chemistry, physics). In this work, we use the error backpropagation algorithm with a three-layer neural network. The transfer functions belong to the sigmoid type at each layer. To predict the runoff at a given moment, the input variables are the rainfall and the runoff values observed on the previous time period. The structure of the network (number of hidden nodes, learning coefficient and momentum values) is optimized to guarantee a good prediction of the runoff, using a set of test data (validation set) not used in the training phase.

Data compiled in our model are a ten years set of rainfall-runoff values collected by the Rabat hydraulic administration (September 1983 to April 1993) in the Beth Wadi basin. In this study, we develop two types of models according to two different time steps (day and week). The data are subdivided into two sets: a first set to train the model (training set) and a second set to test the model (validation set). For the daily time step model, we used data of the last two years: April 1991 to April 1993. The 365 first data constitute the training set and the 365 remaining data constitute the validation set. For the weekly data (Monday to Sunday averages), we have 502 couples of values. We worked by preserving the 120 last values as validation set and trained the neural network with the other data, i.e. 382 couples of values of weekly rainfall-runoff.

1. Labo Biologie Quantitative, U.P.S., 4R3, 118 route de Narbonne, 31062 Toulouse Cedex, France.

2. Faculté des Sciences, Département de Biologie, B.P. 4010, Meknes, Maroc.

3. Administration de l'Hydraulique, Division de la qualité de l'eau, Rabat, Maroc.

** Les commentaires seront reçus jusqu'au 15 juin 1997.

Three types of estimation have been realized:

- (i) "at instant" prediction: prediction of runoff at time t taking into account that rainfall values at time t , as well as runoff and rainfall values at preceding times (untill $t-1$);
- (ii) "one step ahead" prediction: prediction of runoff at the time t from rainfall and runoff values at the preceding times (untill $t-1$);
- (iii) "multistep" prediction: prediction of runoff values for a period from $t-jh$ until t , given that values of the runoff for the period 1 to $t-jh-1$ and values of the rainfall at time 1 to t are available (h is the time step).

The time step is daily for the "at instant" prediction and weekly for "one step ahead" and "multistep" predictions. The choice of input variables is determined by autocorrelation function (ACF), partial autocorrelation function (PACF) analysis on runoff values and cross-correlation function (CCF) analysis between rainfall and runoff values. For the "at instant" prediction, the input vector is composed by runoff values of the four days preceding the t day, and rainfall values of the three last preceding days as well as its value on day t . For the "one step ahead" prediction, the input vector is composed by runoff values of the five weeks preceding the week t , and rainfall values of the three preceding weeks (without considering the rainfall at time t). Finally, for the "multistep" prediction, the input vector is the same as for the "one step ahead" prediction but rainfall values include time t . The runoff values for the week $t-jh + 1$, as well as for the following weeks are computed by feed backing to the input vector the runoff value predicted for the preceding week.

The rainfall-runoff models allow a good estimation for one or several time steps, daily as well as weekly. In the validation set, correlation coefficients between observed and estimated values are high. In the "at instant" prediction, we obtain the Pearson correlation coefficient $R = 0.772$ and the Spearman correlation coefficient $CR = 0.958$. The weak value of R as compared to CR is explained by a few extremely high values of error of prediction. In the "one step ahead" prediction ($R = 0.887$ and $CR = 0.782$) and "multistep" prediction ($R = 0.908$ and $CR = 0.727$), the R coefficients is higher that CR . This confirms that predicted values are in good agreement with the peaks of observed values (absence of large exceptional errors). In all cases, the results obtained are better than those obtained with linear methods. The neural network models can thus be recommended for the time series studies in environmental sciences.

Key words : modelling, neural networks, retropropagation, rainfall-runoff, time series.

RÉSUMÉ

Identifier tous les processus physiques élémentaires du cycle hydrologique qui peuvent avoir lieu dans un bassin versant et attribuer à chacun d'eux une description analytique permettant la prévision conduisent à des structures complexes employant un nombre élevé de paramètres difficilement accessibles. En outre ces processus, même simplifiés, sont généralement non linéaires. Le recours à des modèles à faible nombre de variables, capables de traiter la non-linéarité, s'avère nécessaire.

C'est dans cette optique que nous proposons une méthode de modélisation de la relation pluie et débit basée sur l'utilisation de réseaux neuronaux. Les performances de ces derniers dans la modélisation non linéaire ont été déjà prouvées dans plusieurs domaines scientifiques (biologie, géologie, chimie, physique). Dans ce travail, nous utilisons l'algorithme de la rétropropagation des erreurs avec un réseau à 3 couches de neurones. La fonction de transfert appliquée est de type sigmoïde. Pour prédire le débit à un moment donné, on présente à l'entrée du réseau des valeurs de pluies et de débits observés à des instants précédents. La structure du réseau est optimisée pour obtenir une bonne capacité prévisionnelle sur des données n'ayant pas participé au calage.

L'application du réseau à des données pluviométriques et débitométriques du bassin de l'oued Beth permet d'obtenir de bonnes prévisions d'un ou plusieurs pas de temps, aussi bien journalières qu'hebdomadaires. Pour les données n'ayant pas participé au calage, les coefficients de corrélation entre les valeurs observées et les valeurs estimées par les différents modèles sont élevés. Ils varient de 0,72 à 0,91 pour les coefficients de corrélation de Pearson et de 0,73 à 0,95 pour les coefficients de Spearman.

Mots-clés : modélisation, réseaux neuronaux, rétropropagation, relation pluie et débit, série chronologique.

1 – INTRODUCTION

Au Maroc, pays à climat aride ou semi-aride, l'approvisionnement en eau potable et industrielle en plus de l'irrigation est actuellement assuré à plus de 60 % par les eaux de surface. Pour une bonne gestion hydraulique, une prévision des débits joue un rôle très important, mais elle est souvent difficile à réaliser.

Dans le bassin de Sebou, l'un des plus importants bassins marocains, se situe l'Oued Beth (34°00' N, -5°55' W) qui draine un sous bassin de 4 500 km² (fig. 1) à couverture végétale très réduite. Les précipitations moyennes sont de 620 mm par an, concentrées à de courtes périodes (DERRAZ, 1995). Pour évaluer les flux de nutriments responsables de la dégradation de la qualité de l'eau du barrage El Kansera situé en aval, il nous a paru important de pouvoir disposer d'un modèle prédictif de débit.

Prédire le débit du futur nécessite les connaissances d'informations de débits et/ou de pluies du passé et du présent. La disponibilité en outils informatiques de plus en plus puissants permet d'étudier ce problème plus facilement et surtout

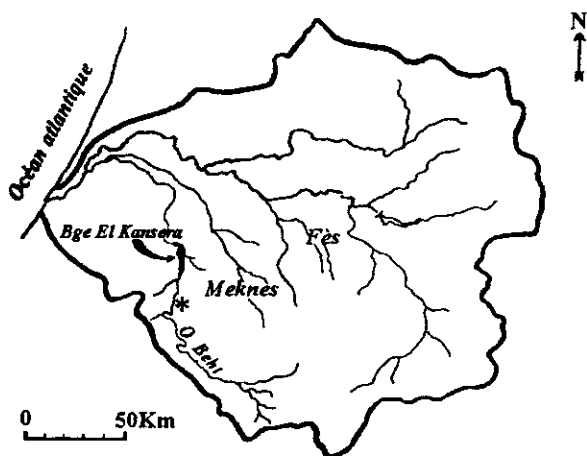


Figure 1

Localisation de l'oued Beth dans le bassin du Sebou. L'astérisque (*) indique l'emplacement de mesure.

Location of the Beth wadi in the Sebou basin. The asterisk (*) indicates the measurement site.

avec plus de rapidité et de précision. Avec cette nouvelle possibilité, nous nous proposons de modéliser la relation pluie et débit avec une nouvelle approche, les réseaux de neurones artificiels. Cette technique a été déjà utilisée par plusieurs auteurs. Dans le domaine de l'hydrologie, on peut notamment citer CUBERO (1991) pour des études de la demande en eau ; BEAUDEAU *et al.* (1994) pour des prévisions des crues turbides ; DIMOPOULOS *et al.* (1996) pour une utilisation simultanée du filtre de Kalman et des réseaux neuronaux pour la prévision du débit ; RAMAN et SUNILKUMAR (1995) qui comparent un modèle linéaire (de type ARMA) et les réseaux de neurones pour la prévision de débit. D'autres auteurs ont appliqué des méthodes neuronales pour étudier des séries chronologiques dans différents domaines, principalement en économie. On peut notamment citer CHEN *et al.* (1990), CHAKRABORTY *et al.* (1992) et HOPTROFF (1993) qui appliquent avec succès la méthode pour la prévision des séries chronologiques.

Dans ce travail, en nous appuyant sur les données de l'oued Beth, nous proposons différents types de prévisions : « à l'instant », à « un pas » et « multipas ».

2 – MÉTHODOLOGIE

La plupart des méthodes statistiques conventionnelles de prévision sont basées sur des modèles linéaires. Le succès pratique de ces approches est limité par leur linéarité. Nous proposons ici une méthode de modélisation basée sur l'un des types de réseaux neuronaux, réseau multicouche sans rétroaction entraîné par l'algorithme de rétropropagation (RUMELHART *et al.*, 1986). Il s'agit d'un principe inspiré par le mode de fonctionnement des neurones humains avec une fonction qui transforme l'activation en réponse de type non linéaire.

Un réseau comporte typiquement trois types de couches de neurones : une couche d'entrée, une ou plusieurs couche(s) cachée(s) et une couche de sortie. Tous les neurones d'une couche donnée, sauf ceux de la dernière couche, émettent une connexion vers chaque neurone de la couche en aval. Dans la majorité des cas, pour limiter le temps de calcul et surtout lorsque les résultats sont en général satisfaisants, c'est un réseau à une seule couche cachée qui est utilisé (*fig. 2*).

La couche d'entrée comporte n neurones codant les n informations ($X_1 \dots X_n$) à l'entrée du réseau. Le nombre de neurones de la couche cachée est choisi par l'utilisateur de manière que les résultats obtenus soient optima. Enfin, la couche de sortie comporte un seul neurone responsable du calcul de la valeur à prédire. A chaque connexion entre les neurones de deux couches successives est associée un poids modifiable au cours de l'apprentissage en fonction des jeux de données en entrée et en sortie.

Si les états des neurones de la couche d'entrée sont déterminés par les variables à l'entrée du réseau, les autres neurones (de la couche cachée et de sortie) doivent évaluer l'intensité de la stimulation en provenance des neurones de la couche précédente par la relation suivante :

$$a_j = \sum_{i=1}^I X_i W_{ij} \quad (1)$$

avec a_j : activation du $j^{\text{ème}}$ neurone de la couche courante ; X_i : valeur de sortie de la $i^{\text{ème}}$ neurone de la couche précédente ; W_{ij} : poids synaptique de la connexion entre le $j^{\text{ème}}$ neurone de la couche précédente et le $i^{\text{ème}}$ neurone de la couche courante.

La réponse des neurones est une fonction non linéaire de type sigmoïde :

$$f(a_j) = \frac{1}{1 + \exp^{-a_j}} \quad (2)$$

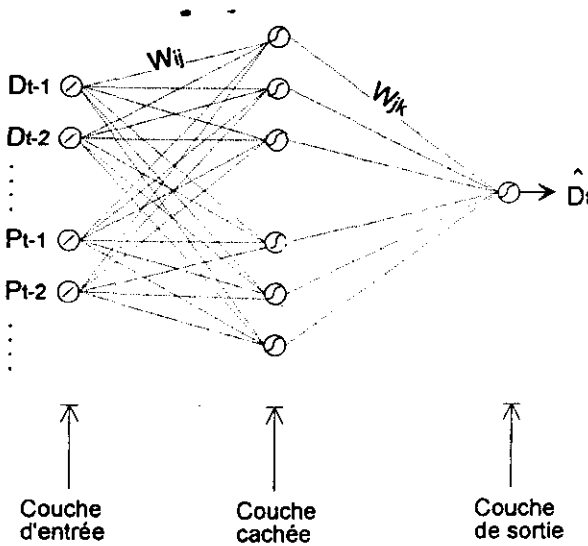


Figure 2 Réseau de neurones à 3 couches utilisé dans cette étude.
Three-layered neural network structure, as used in this work.

La technique de rétropropagation s'apparente à un apprentissage supervisé (pour apprendre, le réseau doit connaître la réponse qu'il aurait dû donner). Elle modifie ensuite l'intensité de connexion de manière que sa réponse soit le plus proche possible de la réponse désirée. On trouvera dans de nombreux articles, notamment RUMELHART *et al.* (1986), CARPENTER (1989), WEIGEND *et al.* (1992), GALLANT (1993) le détail des algorithmes de rétropropagation des erreurs. Notons enfin, qu'on peut utiliser des paramètres comme η (coefficient d'apprentissage) et α (momentum), qui servent à accélérer l'apprentissage tout en permettant au réseau de ne pas tomber dans des minima locaux (références citées plus haut).

L'apprentissage du réseau est effectué de façon que la somme des carrés des erreurs (SSE) soit minimale. SSE est donnée par la relation :

$$SSE = \sum_{j=1}^N (Dt_j - \hat{Dt}_j)^2 \quad (3)$$

avec D_{t_j} : valeur attendue en sortie du réseau (valeur « théorique ») ; \hat{D}_{t_j} : valeur calculée par le réseau (neurone de la couche de sortie) ; N : nombre de données de l'ensemble de calage.

3 – DONNÉES ET EXPÉRIMENTATIONS

Nous disposons de dix années d'observations journalières effectuées par l'administration de l'hydraulique de Rabat (de septembre 1983 à avril 1993), ce qui représente 3 530 mesures de pluie et de débit.

Pour notre étude, nous avons décidé de travailler selon deux pas de temps différents avec pour chacun une subdivision des données en deux ensembles : un pour le calage du modèle et un autre pour la validation du modèle :

- pas de temps journalier (données journalières) : vu l'importance de la base de données, nous avons considéré uniquement les deux dernières années, c'est-à-dire d'avril 1991 à avril 1993. Nous utilisons les premières 365 données comme jeu d'apprentissage et les 365 autres comme jeu de test ;

- pas de temps hebdomadaire : les moyennes hebdomadaires (lundi à dimanche) de pluie et de débit ont été calculées. On obtient ainsi 502 couples de valeurs de pluie et débit hebdomadaires. Nous travaillons en conservant les 120 dernières valeurs comme jeu de validation après apprentissage effectué sur le reste des données, soit 382 couples de valeurs de pluie et débit hebdomadaires.

Trois types de prévision ont été réalisés :

- prédiction « à l'instant » : prédiction du débit au moment t en considérant que les valeurs pour la pluie jusqu'au moment t , ainsi que pour les débits aux moments précédents (jusqu'à $t-1$), sont disponibles ;

- prédiction « un pas » : on prédit le débit au moment t en connaissant les valeurs du débit et de la pluie jusqu'au moment $t-1$;

- prédiction « multipas » : prédiction du débit pour une période $t-jh$ jusqu'à t , étant donné que les valeurs du débit pour la période 1 à $t-jh-1$ et les valeurs de la pluie pour la période 1 à t sont disponibles (h est le pas de temps égal à la semaine).

Le pas de temps est journalier pour la prédiction « à l'instant » et hebdomadaire pour les prédictions à « un pas » et « multipas ». Le choix des variables en entrée est déterminé par l'analyse d'autocorrélation (ACF), et d'autocorrélation partielle (PACF) des valeurs du débit et des corrélations croisées (CCF) entre le débit et la pluie. Pour la prédiction « à l'instant », le vecteur des entrées est composé par les valeurs du débit aux quatre jours précédentes le jour t , et les valeurs de la pluie aux trois jours précédents ainsi que sa valeur au moment t . Pour la prédiction « un pas », le vecteur des entrées est constitué par les valeurs du débit aux cinq semaines précédentes la semaine t , et les valeurs de la pluie aux trois semaines précédentes (sans considérer la pluie au temps t). Enfin, pour la prédiction « multipas », le vecteur des entrées est le même que pour la prédiction à « un pas » augmenté de la valeur de pluie au temps t . La valeur du débit pour la semaine $t-jh + 1$, ainsi que les semaines suivantes est calculée en utilisant comme entrée le débit prédit pour la semaine précédente.

4 – CRITÈRES DE VALIDATION

La validation et la performance du modèle sont jugées à travers :

- de l'étude des erreurs (valeurs observées – valeurs estimées) : en s'assurant que leur moyenne est statistiquement nulle et qu'elles sont stationnaires par rapport à la variable estimée en sortie ;
- des valeurs du coefficient de corrélation linéaire R de Pearson entre le débit estimé et le débit observé ;
- des valeurs du coefficient de corrélation de rangs CR de Spearman.

Les deux coefficients de corrélation peuvent donner des résultats différents selon la position des valeurs extrêmes par rapport à la majorité des observations. En effet, R est sensible à des valeurs extrêmes (erreurs exceptionnelles, mais importantes). Même si l'erreur moyenne est faible, R peut être pénalisé par quelques fortes erreurs. Par contre CR (coefficient R calculé sur les rangs des observations) est moins influencé par des valeurs extrêmes (ces valeurs ne modifient pas l'ordre).

Les trois critères énumérés ci-dessus sont appliqués sur des ensembles de validation (jeu de test) qui ne participe pas à l'élaboration du modèle.

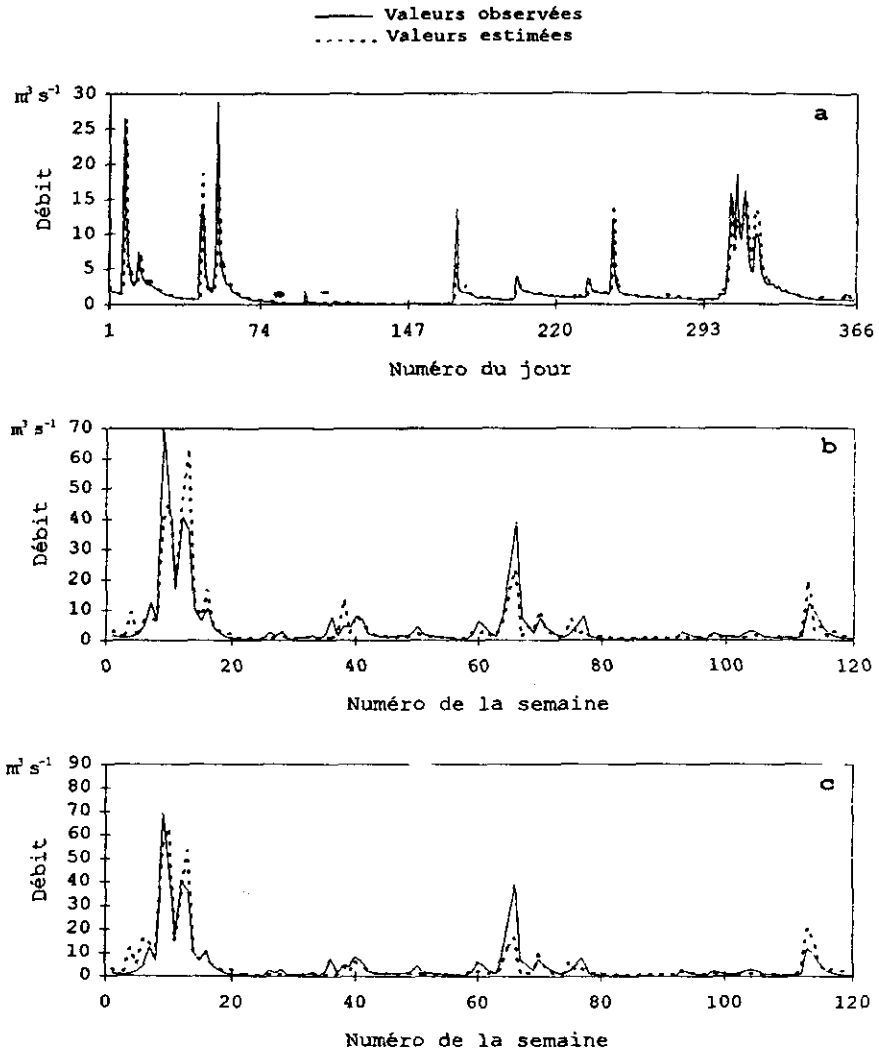
5 – RÉSULTATS

En modélisation par les réseaux de neurones, pour éviter un surapprentissage (modélisation du bruit), les résultats obtenus à partir des données réservées à l'apprentissage doivent être validés sur un nouveau jeu de données. Nous n'insisterons donc ici que sur les résultats obtenus avec les ensembles de validation. Les données qu'ils contiennent, n'ayant pas participé à l'établissement du modèle, permettent de vérifier dans quelle mesure la relation pluie et débit est correctement restituée.

5.1 Prévision « à l'instant »

Le test a été effectué à partir d'un ensemble de données journalières de mai 1992 à avril 1993, à partir d'un modèle établi avec un ensemble de 365 valeurs de pluie et de débit mesurés de mai 1991 à avril 1992. La figure 3a montre que les valeurs de débit estimées par le réseau suivent bien les valeurs observées. On note cependant quelques sous-estimations ou surestimations, ce qui explique un assez mauvais coefficient de corrélation de Pearson $R = 0,722$. Par contre le coefficient de corrélation de rangs de Spearman est excellent ($CR = 0,958$) et confirme la capacité du modèle de bien prédire la majorité des débits. Le faible coefficient R par rapport à CR est expliqué par sa sensibilité due à la présence de quelques fortes erreurs de prédiction de débit.

Les moyennes des valeurs de débits estimés et observés sont très proches, avec un écart type des valeurs estimées un peu plus faible (tabl. 1). La moyenne des erreurs est nulle avec un faible écart type. Bien que la figure 4a ne montre pas une distribution normale des erreurs, on remarque le regroupement d'une

**Figure 3**

Évolution des débits pour les trois types de prévision. Les lignes continues indiquent les valeurs de débit observé (en $\text{m}^3 \text{s}^{-1}$) dans l'ensemble de validation et les traits discontinus sont des valeurs estimées par les modèles. a : prédiction « à l'instant » ; b : prédiction à « à un pas » ; c : prédiction « multipas ».

Evolution of flows for the three prediction types. The solid line indicates observed values (in $\text{m}^3 \text{s}^{-1}$) in the validation set and the dashed line shows estimated values. a : "at instant" prediction ; b : "one-step ahead" prediction ; c : "multi-step" prediction.

très forte proportion de valeurs au niveau de la moyenne (valeur 0). C'est cette distribution des erreurs qui donne une forte valeur à CR. La figure 5a ne montre pas une dépendance marquée des résidus par rapport à des valeurs estimées

Tableau 1 Paramètres statistiques de la prédiction « à l'instant » sur les données journalières (ensemble de validation).**Table 1** Statistical parameters of the prediction "at instant" with daily data (validation set).

Valeur	Observée ($m^3 s^{-1}$)	Estimée ($m^3 s^{-1}$)	Résidus ($m^3 s^{-1}$)
Moyenne	1,870	1,834	0,036
Ecart type	3,331	3,008	2,384

par le modèle, à l'exception de quelques surestimations des valeurs de faibles débits et de quelques sous-estimations des valeurs de fortes débits (valeurs extrêmes responsables de la mauvaise valeur de R par rapport à CR). Notons enfin que presque la totalité des points sont situés sur la ligne d'ordonnée 0, valeur de la moyenne des erreurs.

5.2 Prédiction « un pas »

Le test a été effectué avec les 120 données hebdomadaires (de 1991 à 1993) après un apprentissage sur les 382 semaines (1983 à 1990). La figure 3b montre le profil des deux séries de valeurs observées et estimées par le modèle pour l'ensemble de validation. On note dans la globalité une assez bonne concordance qui justifie de bons coefficients de corrélation ($R = 0,887$, $CR = 0,782$). Comparé à la prédiction « à l'instant », les valeurs estimées par le modèle suivent mieux les pics. Les erreurs sur l'ensemble de données sont, par contre, plus élevées (par rapport à la prédiction « à l'instant »), ce qui explique le meilleur coefficient de corrélation R par rapport à CR.

Les paramètres statistiques des séries de valeurs observées et calculées sont pratiquement identiques (tabl. 2). La moyenne des résidus est d'environ 0,2, avec un faible écart type. La figure 4b montre une bonne répartition des résidus avec une majorité des valeurs au niveau du point zéro de la moyenne. On ne note pas de dépendance des résidus par rapport à des valeurs estimées par le modèle. A l'exception de 2 fortes valeurs, les autres points sont bien alignés sur la ligne d'ordonnée zéro des résidus (fig. 5b).

5.3 Prédiction « multipas »

Comme le cas précédent, les études ont été effectuées avec des données hebdomadaires. La prédiction multipas donne de bonnes estimations comme le

Tableau 2 Paramètres statistiques de la prédiction à « un pas » sur des données hebdomadaires (ensemble de validation).**Table 2** Statistical parameters of the prediction "at instant" with weekly data (validation set).

Valeur	Observée ($m^3 s^{-1}$)	Estimée ($m^3 s^{-1}$)	Résidus ($m^3 s^{-1}$)
Moyenne	4,582	4,388	0,194
Ecart-type	9,769	9,291	4,561

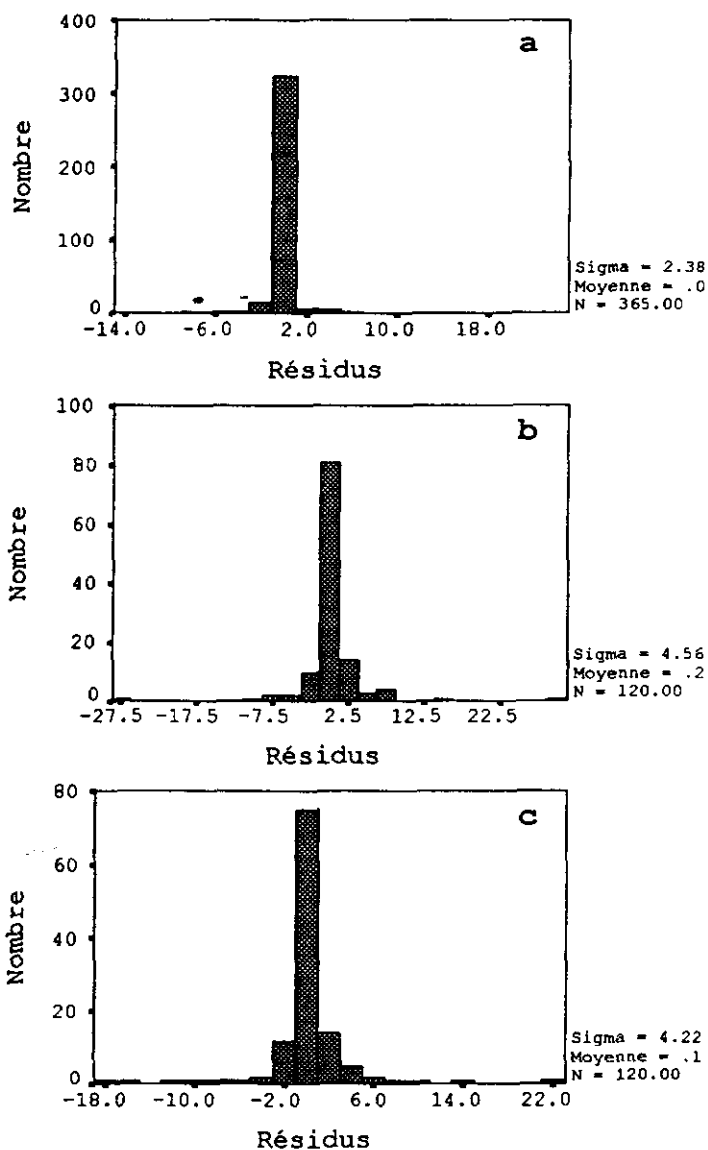


Figure 4 Distribution des résidus pour les trois modèles. a : prédiction « à l'instant » ; b : prédiction à « un pas » ; c : prédiction « multipas ».

Distribution of residuals for the three models. a : "at instant" prediction ; b : "one-step ahead" prediction ; c : "multi-step" prediction.

montre la figure 3c. Les coefficients de corrélation sont élevés ($R = 0,908$, $CR = 0,727$). La raison de la discordance entre les valeurs de R et CR est la même que pour la prévision à « un pas ».

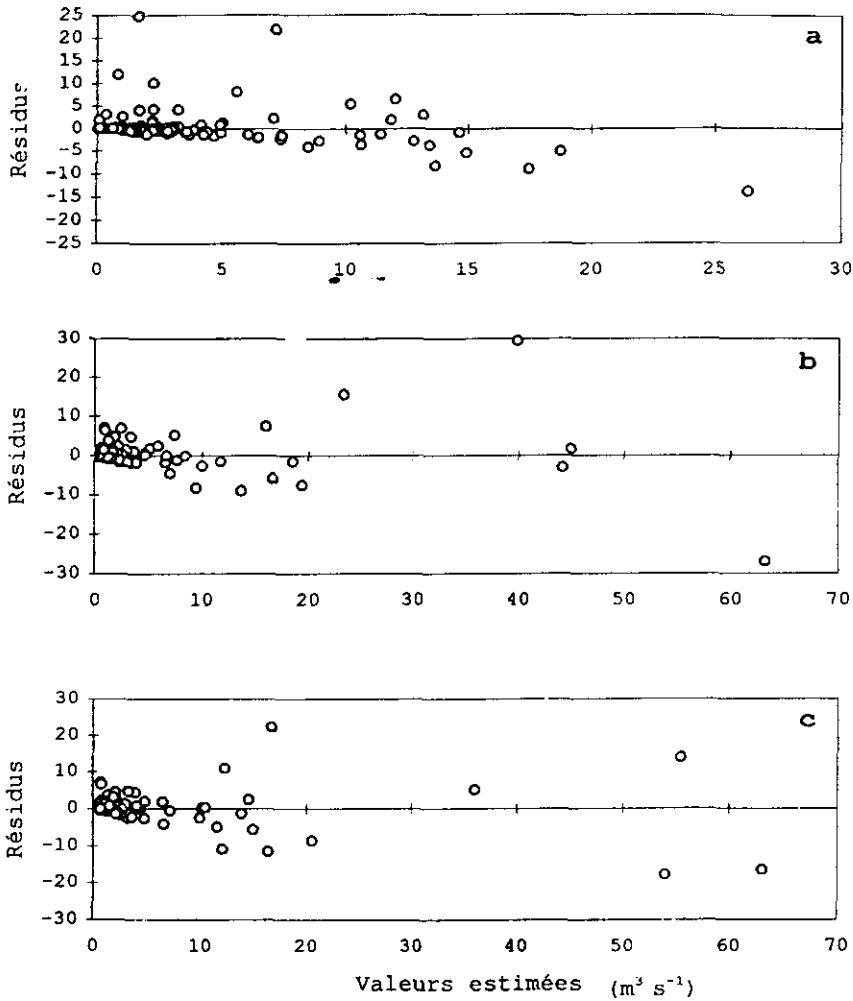


Figure 5 Relation entre les valeurs estimées et les résidus pour les trois modèles. a : prédiction « à l'instant»; b : prédiction à « un pas»; c : prédiction « multipas ».
 Relationship between residuals and estimated values for the three models. a : "at instant" prediction ; b : "one-step ahead" prediction ; c : "multi-step" prediction.

Le tableau 3 montre que les moyennes et écarts types sont pratiquement identiques. La moyenne des résidus est égale à 0,1. La figure 4c révèle une bonne distribution des résidus autour de la moyenne zéro, malgré quelques valeurs extrêmes. La relation entre les résidus et les valeurs estimées par le modèle (fig. 5c) montre de très bons ajustements de faibles valeurs qui constituent la majorité des points. Concernant les fortes valeurs, les erreurs sont plus importantes mais on ne constate pas de tendance marquée vers une sous-estimation ou une surestimation systématique.

Tableau 3 Paramètres statistiques de la prédiction « multipas » sur les données hebdomadaires (ensemble de validation).

Table 3 Statistical parameters of the prediction "multi-step" with weekly data (validation set).

Valeur	Observée (m ³ s ⁻¹)	Estimée (m ³ s ⁻¹)	Résidus (m ³ s ⁻¹)
Moyenne	4,582	4,518	0,064
Ecart-type	9,769	9,896	4,221

6 – CONCLUSION

Les résultats présentés dans cet article indiquent clairement que les réseaux de neurones artificiels peuvent apporter des réponses satisfaisantes pour la prédiction des relations pluie et débit dans une grande variété de configurations (données journalières pour la prévision « à l'instant », hebdomadaires pour la prévision à « un pas » ou « multipas ») et dans des zones climatiques aride ou semi-aride (le bassin de l'oued Beth) dont les pluies et les débits sont très irréguliers (CLARKE, 1994). Par rapport à des méthodes plus classiques, on constate que les réseaux de neurones possèdent un meilleur pouvoir prévisionnel. En effet, le tableau 4 nous montre les résultats obtenus par deux méthodes de prédiction appliquées sur les mêmes jeux de données : un modèle linéaire autorégressif avec des variables exogènes (pluies), ARX (MELARD, 1990) et les réseaux de neurones. Les réseaux de neurones sont nettement plus performants, dans tous les types de prévision. Ils présentent ainsi une alternative intéressante et prometteuse par rapport aux méthodes statistiques classiques pour étudier des séries chronologiques dans le domaine des sciences de l'environnement.

Tableau 4 Coefficient de corrélation de Pearson R entre valeurs observées et estimées des deux types de modèles sur l'ensemble de validation concernant les trois types de prévision. (RN : réseaux de neurones artificielles ; ARX : modèle autorégressif).

Table 4 Pearson correlation coefficient R between observed values and the values estimated by three models in three prediction types with the validation set. (RN : artificial neural networks ; ARX : autoregressive model).

Prédiction	« à l'instant »	« un pas »	« multipas »
RN	0,722	0,887	0,908
ARX	0,562	0,771	0,784

REMERCIEMENTS

Les auteurs remercient les trois lecteurs anonymes pour les critiques et conseils qu'ils ont formulé avant la publication de cette article.

RÉFÉRENCES BIBLIOGRAPHIQUES

- BEAUDEAU P., LÉBOULANGER T., HANETON S., 1994. Prévision des crues turbides : apport des méthodes neuronales aux démarches classiques sur l'exemple du forage d'Yport (Seine-Maritime). *TSM.*, 2 : 97-109.
- CARPENTER G.A., 1989. Neural network models for pattern recognition and associative memory. *Neural Networks*, 2 : 243-257.
- CHAKRABORTY K., MEHROTRA K., MOHAN C.K., RANKA S., 1992. Forecasting the behavior of multivariate time series using neural networks. *Neural Networks*, 5 : 961-970.
- CHEN S., BILLINGS S.A., GRANT P.M., 1990. Non-linear system identification using neural networks. *Int. J. Control*, 51 (6) : 1191-1214.
- CLARKE R.T., 1994. Statistical modelling in hydrology. Wiley, Chichester, 412p.
- CUBERO R.G., 1991. Neural networks for water demand time series forecasting. In Artificial Neural Networks, International Workshop, IWANN'91, Granada, Spain. Proceeding E.A. Prieto, 453-460.
- DERRAZ M., 1995. Etude de l'eutrophisation de la retenue du barrage El Kansera. Th. Doct. D'Etat es Sciences Biologie, Fac. Sciences Meknès, 150 p.
- DIMOPOULOS I., LEK S., LAUGA J. (1996). Modélisation de la relation pluie-débit par les réseaux connexionnistes et le filtre de Kalman. *Hydrological Sciences Journal.*, 41 : 179-193.
- GALLANT S.I., 1993. Neural Network Learning and Expert Systems. M.I.T. Press, London, 365 pp.
- HOPTRUFF R.G., 1993. The principles and practice of time series forecasting and business modelling using neural nets. *Neural Comput. & Applic.*, 1 : 59-66.
- MELARD G., 1990. Méthode de prévision à court terme. SMA, Bruxelles, 468 pp.
- RAMAN H., SUNILKUMAR N., 1995. Multivariate modelling of water resources times series using artificial neural networks. *Hydrological Sciences Journal*, 40, 145-163.
- RUMELHART D.E., HINTON G.E., WILLIAMS R.J., 1986. Learning representations by back-propagating error. *Nature*, 323 : 533-536.
- WEIGEND A.S., HUBERMAN B.A., RUMELHART D.E., 1992. Predicting sunspots and exchange rates with connectionist networks. In M. Casdagli & S. Eubank (Editors), Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity, Proc. vol XII, Addison-Wesley, Redwood City, 395-432.