

## Utilisation des réseaux de neurones et de la régularisation bayésienne en modélisation de la température de l'eau en rivière

### The bayesian-regularized neural network approach to model daily water temperature in a small stream

Mohamed Nohair, André St-Hilaire and Taha. B. Ouarda

Volume 21, Number 3, 2008

URI: <https://id.erudit.org/iderudit/018783ar>

DOI: <https://doi.org/10.7202/018783ar>

[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

Nohair, M., St-Hilaire, A. & Ouarda, T. B. (2008). Utilisation des réseaux de neurones et de la régularisation bayésienne en modélisation de la température de l'eau en rivière. *Revue des sciences de l'eau / Journal of Water Science*, 21(3), 373–382. <https://doi.org/10.7202/018783ar>

Article abstract

Understanding and predicting water temperatures is essential in order to help prevent or forecast high temperature problems. To attain this objective, we define in this work a model that predicts temperature variations in a small stream according to climatic variables, such as air temperature, water flow and quantity of rainfall in the river catchment. Static neural networks were used as a technique for evaluation of the relations among the various variables, with a mean error of 0.7°C.

In addition, we developed a forecasting model able to estimate the short-term and mid-term variations of water temperature, i.e., to forecast the temperature of days ( $j+i$ ,  $i=1,2,\dots$ ) from climatic parameters of day  $j$ . Two methods were used: the first one is iterative and uses the estimated value of day  $j$  to estimate the value of the water temperature for day  $j+1$ . The second method is much simpler, involving an estimate of the temperature of all days at once. The Levenberg-Marquardt algorithm implemented in the Matlab neural network toolbox allowed a marked improvement in the performance of the model. Very satisfactory results were then obtained by testing the validity by cross validation technique with a mean error of 1.5°C for long term prediction of 7 days.

# UTILISATION DES RÉSEAUX DE NEURONES ET DE LA RÉGULARISATION BAYÉSIENNE EN MODÉLISATION DE LA TEMPÉRATURE DE L'EAU EN RIVIÈRE

*The bayesian-regularized neural network approach to model daily water temperature in a small stream*

MOHAMED NOHAIR<sup>a\*</sup>, ANDRÉ ST-HILAIRE<sup>b</sup> ET TAHA. B. OUARDA<sup>ba</sup>

<sup>a</sup>Laboratoire de chimie organique et chimiométrie, Faculté des sciences et techniques de Mohammedia, Maroc

<sup>b</sup>Institut national de la recherche scientifique, Centre Eau Terre Environnement, Québec, Canada

Reçu le 13 octobre 2006, accepté le 5 mars 2008

---

## RÉSUMÉ

Dans ce travail, nous avons élaboré un modèle de prédiction des variations de la température d'un cours d'eau en fonction de variables climatiques, telles que la température de l'air ambiant, le débit d'eau et la quantité de précipitation reçue par le cours d'eau. Les réseaux de neurones statiques ont été utilisés pour approximer la relation entre ces différentes variables avec une erreur moyenne de 0,7 °C. Par ailleurs, nous proposons un modèle de prédiction de l'évolution de la température de l'eau à court et moyen termes pour les jours ( $j + i$ ,  $i = 1, 2, \dots$ ). Deux méthodes ont été appliquées : la première, de type itérative, utilise la valeur estimée du jour  $j$  pour prédire la valeur de la température de l'eau au jour  $j + 1$ ; la seconde méthode, beaucoup plus simple à mettre en œuvre, consiste à estimer la température de tous les jours considérés en une seule fois.

L'optimisation de la fonction de coût par l'algorithme de Levenberg-Marquardt, disponible dans l'outil « réseaux de neurones » de MATLAB a permis d'améliorer nettement la

performance des modèles. Des résultats très satisfaisants sont alors obtenus en testant la validité du modèle par la validation croisée avec des erreurs moyennes de prédiction à sept jours de 1,5 °C.

**Mots clés :** *cours d'eau, réseaux de neurones, prédiction et prévision, température d'eau, validation croisée.*

## ABSTRACT

Understanding and predicting water temperatures is essential in order to help prevent or forecast high temperature problems. To attain this objective, we define in this work a model that predicts temperature variations in a small stream according to climatic variables, such as air temperature, water flow and quantity of rainfall in the river catchment. Static neural networks were used as a technique for evaluation of

---

\*Auteur pour correspondance :

Téléphone:

Télécopie : 212-023-31-53-53

Courriel : [nohairmohamed@yahoo.fr](mailto:nohairmohamed@yahoo.fr)

the relations among the various variables, with a mean error of 0.7°C.

In addition, we developed a forecasting model able to estimate the short-term and mid-term variations of water temperature, i.e., to forecast the temperature of days ( $j+i$ ,  $i=1,2,\dots$ ) from climatic parameters of day  $j$ . Two methods were used: the first one is iterative and uses the estimated value of day  $j$  to estimate the value of the water temperature for day  $j+1$ . The second method is much simpler, involving an estimate of the temperature of all days at once. The Levenberg-Marquardt algorithm implemented in the Matlab neural network toolbox allowed a marked improvement in the performance of the model. Very satisfactory results were then obtained by testing the validity by cross validation technique with a mean error of 1.5°C for long term prediction of 7 days.

**Keywords:** *stream, neural network, forecasting and prediction, water temperature, cross validation.*

## 1. INTRODUCTION

La température de l'eau intervient de façon indirecte mais déterminante dans la pollution des eaux de surface. Par exemple, l'eau possède une capacité d'auto-épuration, qui est contrôlée essentiellement par sa température. L'auto-épuration de l'eau est limitée par la quantité d'oxygène dissoute. L'eau étant très polaire, elle ne peut dissoudre que des composés chimiques ayant une polarité différente de zéro, en établissant des liaisons de forte énergie de type hydrogène. Ainsi, l'oxygène est peu soluble dans l'eau, à cause de sa faible polarité par rapport à celle de l'eau. En eau douce, la solubilité de l'oxygène dans l'eau est de 14,4 mg/L à 0 °C. Elle diminue rapidement avec une élévation de la température. À 30 °C, elle n'est plus que de 7,53 mg/L.

Lorsque la température augmente, la quantité d'oxygène diminue, ce qui provoque le développement d'autres micro-organismes capables de vivre en anaérobiose (absence d'oxygène) avec pour conséquence :

1. la formation de gaz putrides malodorants;
2. la disparition de certains organismes modifiant la diversité des espèces; et
3. la réduction de la capacité de charge et d'auto-épuration de l'eau.

En somme, la température de l'eau est un paramètre très important pour toute étude qualitative d'un cours d'eau. Ainsi la connaissance, la prédiction et la prévision de la température de l'eau de rivière représentent un atout capital.

Les applications des réseaux de neurones dans le domaine hydrométéorologique, à des fins de prévision, concernent généralement la prévision des débits des rivières, comme l'estimation des précipitations, la prévision des apports naturels aux réservoirs d'irrigation. Dans le domaine de la prévision de la qualité de l'eau, les études réalisées portaient sur la prévision de la salinité (MAIER et DANDY, 1996), l'acidité et la conductivité (BASTARACHE *et al.*, 1997) et sur le pouvoir de dissolution de l'eau des produits chimiques à base de carbone et d'azote (CLAIR et ETHERMAN, 1998).

Dans ce travail, nous proposons un modèle de prédiction des variations de la température d'un cours d'eau en fonction de variables climatiques telles que la température de l'air ambiante, le débit d'eau et la quantité des précipitations se déversant dans le cours d'eau. Des réseaux de neurones ont été utilisés pour approximer la relation entre ces différentes variables. Nous avons considéré trois paramètres, disponibles dans nos données historiques relatives à la température de l'eau relevée sur le site étudié. Ces paramètres sont jugés pertinents : la température de l'air (CLUIS, 1972; MOHSENI et STEFAN, 1999; PIPER *et al.*, 1982; STEFAN et PREUD'HOMME, 1993), le débit d'eau et la quantité de précipitations enregistrés au cours de la journée (CAISSIE *et al.*, 2001).

Dans un premier temps, nous proposons un modèle purement statique, considérant que la température de l'eau ne dépend que des variables mesurées au cours de la journée. Ce modèle est simple car il est basé sur l'estimation de la relation entre les différentes variables, qui sont mesurées à chaque instant. D'autre part, les processus de variation de la température de l'eau et de l'air sont des systèmes évolutifs dans le temps. Il est alors judicieux d'en tenir compte en introduisant d'autres variables liées à la dynamique de ces régimes. Ainsi, nous proposons un autre type de modèle pour prédire la température de l'eau à différents horizons dans le futur jusqu'au jour ( $j + 7$ ).

## 2. DESCRIPTION DU SITE

Les données disponibles proviennent du site du ruisseau du Panola Mountain. La région est située à 25 km au sud-ouest d'Atlanta (E.-U.). Le climat de la région est considéré comme subtropical tempéré avec une température moyenne de 16 °C et une moyenne de précipitation de l'ordre de 1,24 m (HOPPER, 2001). Les données historiques relevées sur le site (la température et le débit d'eau de la rivière, la température de l'air et la quantité de précipitations) couvrent la période de 1998 à 2003. Notons la présence de données manquantes sur l'ensemble de cette période. Les valeurs horaires des trois premières variables relevées pendant la journée sont converties en moyenne journalière et la quantité de précipitations est transformée en une variable cumulative de la journée. La

journee commence à 1 heure du matin, et se termine à minuit. La figure 1 représente l'évolution de la température durant la période considérée. Quelques paramètres statistiques de ces données sont indiqués dans le tableau 1.

### 3. MÉTHODOLOGIE DES RÉSEAUX DE NEURONES

#### 3.1 Réseaux de neurones : introduction

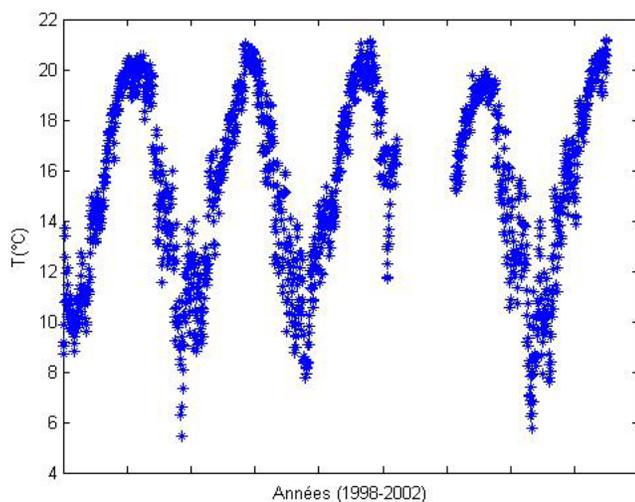
Le problème que se posent les scientifiques en recherchant des modèles déterministes de type « cause à effet » est le suivant : à partir d'un ensemble de mesures de variables d'un processus de nature quelconque (physique, chimique ou biologique) et des résultats de ce processus, comment trouver une relation déterministe entre ces variables et ce résultat, représentée par une équation mathématique, valable dans le domaine où les

mesures ont été effectuées. Dans notre étude, nous utilisons une technique qui a fait ses preuves en calcul statistique d'approximation fonctionnelle : les réseaux de neurones. Elle permet de faire des prévisions, d'élaborer des modèles, de reconnaître des formes ou des signaux, etc. (DREYFUS *et al.*, 2002).

Dans notre cas, les données disponibles forment une série chronologique dont on peut effectuer des prévisions des événements futurs, mais dont les facteurs influençant les événements passés doivent rester les mêmes. À notre connaissance, aucun travail n'a été fait dans ce sens. Par contre, la problématique de la prévision en hydrologie a été largement étudiée, surtout le cas de la relation débit-précipitation (COULIBALY *et al.*, 2000; LUK *et al.*, 2000; TOTH *et al.*, 2000; ZEALAND *et al.*, 1999).

On distingue deux grands types d'architecture des réseaux de neurones pour résoudre ce type de problématique : les réseaux de neurones non bouclés et les réseaux de neurones bouclés. Dans notre cas, nous utiliserons une variante de la deuxième catégorie en introduisant comme variables d'entrée le résultat du processus de l'événement passé dans une nouvelle fonction. Une seule architecture de type réseaux à couches (MLP : un réseau de neurones de type perceptron multicouche) sera utilisée (Figure 2) : une couche de neurones est connectée à une couche de neurones « cachés », qui est reliée à une couche de neurones de sortie. L'activité des neurones d'entrée code l'information qui est présentée au réseau. Ce sont les variables explicatives externes citées précédemment. L'activité de chaque neurone caché est déterminée par l'activité des neurones d'entrée et les valeurs des coefficients de connexion ou poids. L'activité des neurones de sortie dépend de l'activité des neurones cachés et des coefficients de connexion reliant ces neurones à ceux de sortie. La couche de sortie contient un seul neurone, représentant la température de l'eau.

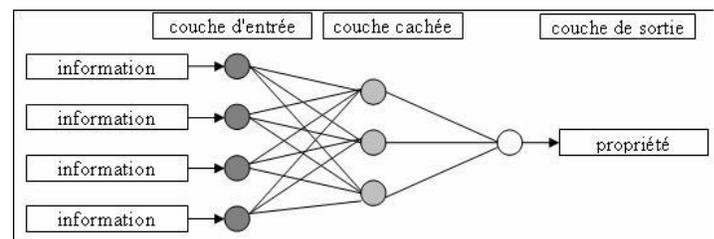
Un neurone est avant tout un opérateur mathématique, dont la valeur numérique se calcule par quelques lignes d'un programme. Un neurone réalise une somme pondérée suivie d'une fonction non linéaire  $f$ . Cette fonction  $f$  doit être bornée, continue et dérivable. Elle peut aussi avoir la forme d'une



**Figure 1.** Variation de la température de l'eau durant la période de l'étude.  
*Variation of the water temperature during the period under study.*

**Tableau 1.** Quelques paramètres statistiques descriptifs relatifs à la température de l'eau en fonction des années de l'étude.  
*Table 1. Statistical parameters of the data set corresponding to the water temperature during the period studied.*

Année Variables	1998	1999	2000	2001	2002
Nb de jours	365	360	301	238	236
Tmoyenne	15,3	15	15,8	16,2	16,2
Tmin-max	(8,7-20,7)	(5,5-21,6)	(7,7-21,1)	(7,0-19,9)	(7-20)



**Figure 2.** Représentation d'une simple architecture à couches de type MLP des réseaux de neurones.  
*Representation of a simple MLP architecture of the neural network.*

fonction seuil si le résultat recherché est de type booléen (soit 0 ou 1). Les fonctions les plus fréquemment utilisées sont les fonctions sigmoïdes.

### 3.2 Principe d'un modèle de réseaux de neurones

Le principe est le même que pour n'importe quel processus statistique caractérisé par une relation déterministe entre des causes et des effets. Des exemples d'apprentissage, c'est-à-dire des jeux de valeurs de neurones d'entrée et de valeurs des neurones de sortie correspondants, sont fournis au réseau de neurones. L'apprentissage consiste tout simplement à calculer les coefficients des connexions (poids) entre les différentes couches de telle manière que les sorties du réseau de neurones soient, pour les exemples utilisés, aussi proches que possible des sorties désirées. Pour ce faire, on utilise des algorithmes d'optimisation qui minimisent une fonction de coût constituée d'une mesure de l'écart entre la réponse réelle du réseau et la réponse désirée. Cette optimisation se fait de manière itérative, en modifiant les poids en fonction du gradient de la fonction du coût : le gradient est estimé par une méthode spécifique au réseau de neurones, dite méthode de rétropropagation de l'erreur. Les poids sont initialisés aléatoirement avant l'apprentissage, puis modifiés itérativement jusqu'à obtention d'un compromis satisfaisant entre la précision de l'approximation sur un ensemble de validation disjoint au précédent et celle obtenue sur l'ensemble des données d'apprentissage.

### 3.3 Mise en œuvre des modèles de réseaux de neurones

Pour réaliser un modèle de réseaux de neurones, trois étapes successives sont nécessaires :

- Tout d'abord, on choisit l'architecture du réseau, c'est-à-dire le nombre de neurones de la couche cachée. Le réseau ne doit être ni trop souple, ni trop rigide. Il existe maintenant plusieurs méthodes adaptées à ces considérations, telles que la régularisation bayésienne. Nous disposons de plus d'un millier de données pour un nombre de neurones d'entrée ne dépassant pas six variables. La fonction dans le modèle neuronal pourrait être définie sans surajustement, car le nombre de données est assez élevé et la température de l'eau ne varie pas de manière significative d'un jour à l'autre.
- Les poids sont calculés en minimisant l'erreur d'approximation sur les données de l'ensemble d'apprentissage, de telle manière que le réseau réalise la tâche désirée.
- Enfin, la qualité du réseau obtenu est estimée à partir d'exemples ne faisant pas partie de l'ensemble d'apprentissage.

Dans notre étude, les deux fonctions d'activation utilisées sont les suivantes : une fonction de type sigmoïde dans la

couche intermédiaire et une fonction linéaire dans la couche de sortie. Des travaux précédents ont montré que ce couple de fonctions permet d'approximer quasiment tous les types de relations non linéaires (DREYFUS *et al.*, 2002).

Pour enseigner une tâche à un réseau, il faut ajuster les coefficients de chaque neurone et minimiser la différence entre la sortie désirée et la sortie effective. Ce procédé impose de calculer la dérivée d'une quantité nommée coût  $J$  par rapport aux coefficients de connexion, c'est-à-dire estimer la variation de l'erreur en fonction des variations de chaque coefficient de connexion. La méthode la plus utilisée pour déterminer ces dérivées est la méthode de rétropropagation de l'erreur.

Une fois que les activités de tous les neurones de sortie ont été déterminées pour l'ensemble des données d'apprentissage, le calcul du coût  $J$  est défini par l'expression suivante :

$$J(w) = \sum_{k=1}^N (y^k - g(x^k, w))^2 = \sum_{k=1}^N J^k(w) \quad (1)$$

où  $y$  est la température observée,  $g(x, w)$  est la température calculée,  $x$  et  $w$  représentent respectivement l'ensemble des variables explicatives et les coefficients de connexion ou poids,  $N$  est le nombre d'exemples et  $k$  est l'indice des données de l'ensemble d'apprentissage.

Évaluer le gradient consiste à évaluer le gradient du coût partiel  $J^k(w)$  relatif à l'observation  $k$ , et de faire ensuite la somme sur tous les exemples. L'algorithme de rétropropagation consiste essentiellement en l'application répétée de la règle des dérivées composées.

$$\left( \frac{\partial J^k}{\partial w_{ij}} \right) = \left( \frac{\partial J^k}{\partial v_i} \right) \left( \frac{\partial v_i}{\partial w_{ij}} \right)_k = \delta_i^k x_j^k \quad (2)$$

$\delta_j^k$  désigne la valeur du gradient du coût partiel par rapport au potentiel du neurone  $i$ , et ce pour l'exemple  $k$ . Il peut être exprimé de la manière suivante pour les neurones de la couche cachée :

$$\delta_j^k = \left( \frac{\partial J^k}{\partial v_i} \right) = -2g(x^k, w) \left( \frac{\partial g(x, w)}{\partial v_i} \right)_k \quad (3)$$

Le même type de calcul est appliqué aux neurones de la couche cachée. Finalement, les paramètres du réseau sont modifiés par la formule suivante à l'itération  $i$  de l'apprentissage.

$$w(i) = w(i-1) - \mu_i \nabla J(w(i-1)) \quad (4)$$

La quantité  $\mu_i$  représente le pas du gradient ou pas d'apprentissage.

Cette méthode représente des inconvénients quand le gradient de la fonction de coût devient trop petit. Pour éviter ce problème, des techniques basées sur la méthode de Newton sont utilisées en considérant que le pas d'apprentissage est égal à l'inverse de la dérivée de second ordre de la fonction de coût. Parmi ces techniques, citons l'algorithme de Levenberg-Marquardt.

Avant tout apprentissage, toutes les variables sont normalisées et centrées pour éviter la surestimation de l'effet des variables ayant des grandeurs importantes. Seules les variables représentant la température de l'air, le débit et la quantité des précipitations sont retenues dans une première phase. La corrélation entre les valeurs calculées (prédites) et observées (expérimentales) est examinée au moyen d'une relation linéaire :

$T_{obs} = a * T_{calc} + b$ .  $T$  désignant la température de l'eau,  $a$  et  $b$  représentent la pente de la droite et la valeur de température à l'origine.

Les paramètres statistiques sont calculés à partir des relations ci-dessous :

$$r^2 = 1 - \frac{\sum (T_{obs} - T_{calc})^2}{\sum (T_{obs} - T_{moyenne})^2} \quad (5)$$

$$s^2 = \frac{\sum (T_{obs} - T_{calc})^2}{n}$$

$r^2$  et  $s$  représentent respectivement le coefficient de corrélation et l'écart-type de l'erreur et  $n$  est le nombre d'exemples traités.

#### 4. RÉALISATION DU MODÈLE NEURONAL POUR LA PRÉDICTION ET LA PRÉVISION DE LA TEMPÉRATURE DE L'EAU

Le modèle consiste en une série de trois couches de neurones. La première couche représente les variables retenues précédemment : la température de l'air, le débit du cours d'eau et la quantité des précipitations pendant la journée. La dernière couche représente la température de l'eau. Les paramètres du réseau (les valeurs des poids) sont initialisés à l'aide d'une fonction qui fournit des valeurs aléatoires. L'apprentissage des données, c'est-à-dire la minimisation de la fonction de coût, est réalisé à l'aide de l'algorithme de Levenberg-Marquardt qui donne de meilleurs résultats que les algorithmes usuels à pas constant, cités précédemment. La couche intermédiaire est composée de sept neurones. L'architecture retenue est donc de

type 3-7-1. Pour un nombre de neurones dans la couche cachée supérieur à 7, le gain n'est pas significatif.

##### 4.1 Généralisation par validation croisée

Notre objectif est de choisir un modèle capable d'apprendre les données qu'on lui propose, mais aussi qui offre une bonne généralisation à des fins de prédiction, en d'autres termes qui évite le surajustement lors d'un test. Pour ce faire, une des techniques utilisées consiste à décomposer l'ensemble de données en trois sous-ensembles. Le premier sert à l'apprentissage du modèle et le deuxième à ajuster le modèle et éviter le surapprentissage en effectuant un arrêt prématuré quand son erreur augmente lors de l'apprentissage. Le troisième sous-ensemble sert à évaluer les performances du modèle. Cette technique est assez lourde à réaliser et pose la problématique du choix du deuxième sous-ensemble pour le calibrage du modèle. Dans notre cas, nous avons utilisé une méthode basée sur la technique de régularisation bayésienne par modération des poids. Elle consiste à pénaliser les valeurs élevées des poids en modifiant la fonction de coût (DREYFUS *et al.*, 2002). Cette technique force les paramètres (les poids) à ne pas prendre des valeurs trop élevées, et, par conséquent, à éviter le surajustement.

Nous avons testé le pouvoir prédictif du modèle neuronal par la technique de la validation croisée (leave-20%-out cross-validation method). Dans cette procédure, nous retirons 20 % de l'échantillon de base qui servira de test et le reste à l'apprentissage. Les données sont classées par ordre chronologique. Pour les cinq décompositions testées, le choix des valeurs initiales de poids n'a presque pas d'effet et la convergence de la fonction de coût est observée après 300 itérations en moyenne.

Les résultats présentés au tableau 2 sont plus que satisfaisants. Seuls 14 cas présentent une erreur supérieure à 3 °C et un seul cas avec une erreur supérieure à 4 °C. Il est important de signaler que ces paramètres statistiques proviennent d'une validation croisée dont le test porte sur l'ensemble des données. La figure 3 présente l'évolution de la température prédite en fonction de la température observée pour les cinq ensembles de test.

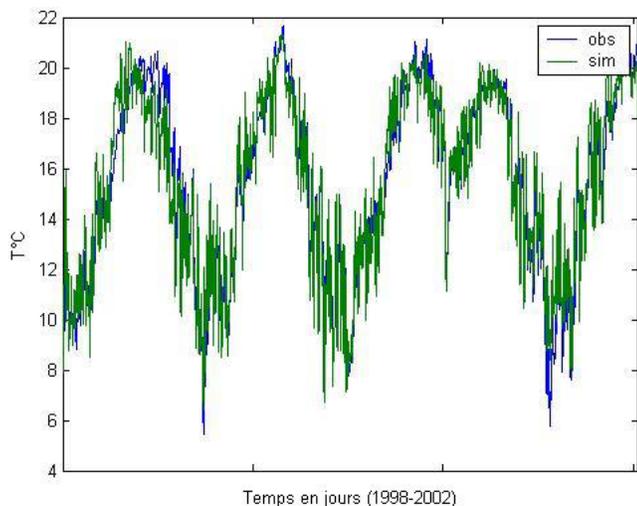
La nature de la décomposition de l'ensemble des données permet d'évaluer la performance du modèle. Lorsque les ensembles de test sont composés à partir de données prélevées une par une à intervalles égaux, les paramètres statistiques sont nettement améliorés ( $R^2 = 0,95$ ,  $s = 0,8$  °C), ceci est dû au fait que la température ne change pas de manière significative d'un jour à l'autre. La figure 4 présente l'erreur résiduelle pour l'ensemble des données en validation croisée.

L'analyse des données historiques de l'évolution de la température de l'eau en rivière montre que celle-ci ne varie

**Tableau 2. Résultats statistiques de la validation croisée pour une architecture 3-7-1.**

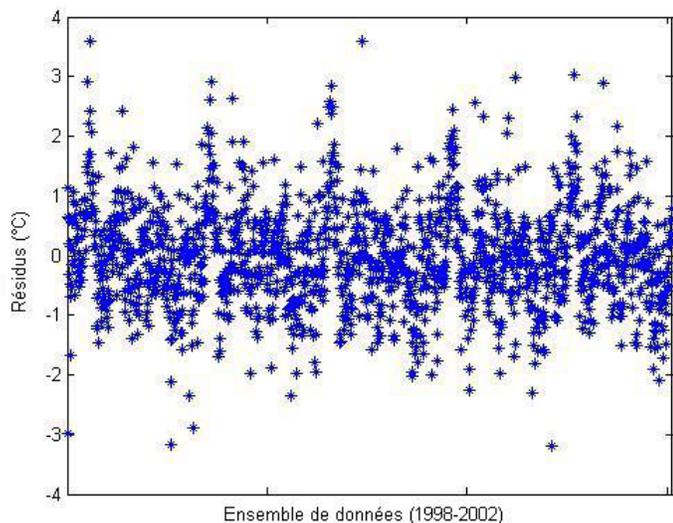
**Table 2. Statistical results of the cross validation for the architecture 3-7-1.**

Coefficient de détermination $R^2$	0,92
Déviatoin standard (RMSE, s)	1 °C
Histogramme de l'erreur (res = obs-calc)	
-1 < res < 1	70 %
-2 < res < 2	93 %
3 < res	0,9 %



**Figure 3. Températures observées et calculées pour l'ensemble des données en validation croisée.**

*Observed and calculated temperatures for the whole set by means of the cross validation technique.*



**Figure 4. Répartition des résidus pour l'ensemble des données en validation croisée.**

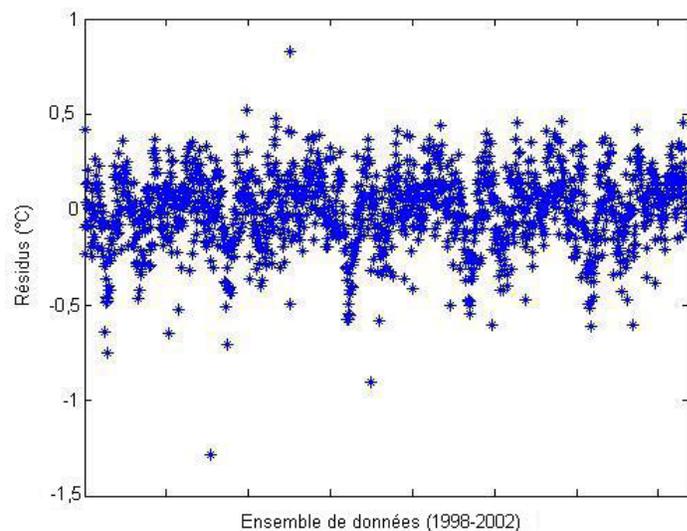
*Distribution of the residuals for the whole set by using cross validation.*

pas beaucoup à court terme, c'est-à-dire sur un ou deux jours. La température de l'air ambiante est la variable explicative, dont l'effet est le plus important sur ce processus. Une simple analyse statistique entre ces deux variables montre que la température de l'eau dépend non seulement de la température de l'air ambiant du même jour, mais aussi des jours précédents. Dans la perspective d'augmenter la capacité d'adaptabilité du modèle neuronal, nous avons considéré les températures de l'air ambiant des jours précédents comme des variables explicatives supplémentaires. Pour une meilleure flexibilité du réseau, seuls les deux jours précédents sont pris en considération. Le nombre de neurones dans la couche intermédiaire est maintenu égal à 7. Les paramètres statistiques de la validation croisée sont nettement meilleurs. Le coefficient de détermination est passé à 0,96, l'écart-type de l'erreur est de l'ordre de 0,7 °C. Comparativement à la figure 4, la figure 5 montre une nette amélioration de la répartition des résidus, seul un point présente une erreur entre -1 et -1,5.

D'une manière générale, les résultats sont très satisfaisants et justifient le recours à l'approche par les réseaux de neurones dans la prédiction de la température de l'eau en rivière.

#### 4.3 Modélisation dynamique par les réseaux de neurones

Les données historiques consistent en une série chronologique de l'évolution de la température de l'eau en fonction du temps. Les modélisations s'inscrivant dans la stratégie de Box et Jenkins (BOX et JENKINS, 1976) AR, MA, ARMA, ARIMA et SARIMA se prêtent bien à ce type de série,



**Figure 5. Résidus pour l'ensemble des données en validation croisée en considérant les valeurs des températures de l'air aux jours précédents.**

*Cross validation residuals for the whole set by considering the air temperature for the previous days.*

puisqu'elles considèrent que les données peuvent être fortement corrélées, comme le sont habituellement les séries successives d'une manière générale. L'observation est liée d'une manière ou d'une autre aux observations précédentes. Dans notre cas, un modèle de type ARMA se réduit tout simplement à une relation linéaire. La valeur de la température de l'eau  $\hat{y}(j)$  est reliée aux valeurs précédentes  $\hat{y}(j-k)$ , les variables explicatives précédemment retenues (température de l'air, précipitation et débit) et les erreurs de chaque estimation  $e(t-k)$  suivant la relation ci-dessous :

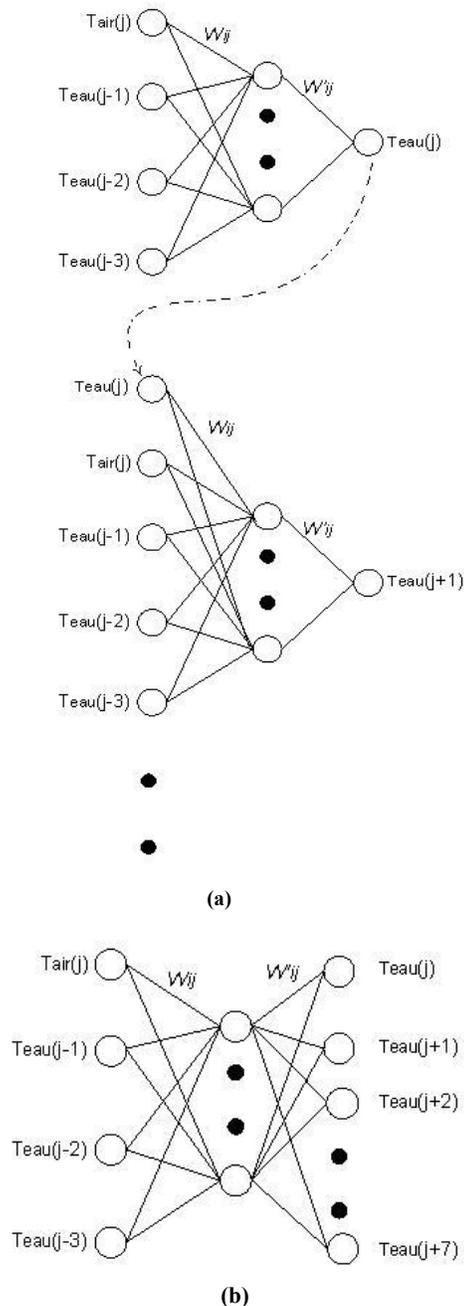
$$\hat{y}(t) = \sum_{k=1}^{n_a} a_k \hat{y}(t-k) + \sum_{l=1}^{n_b} b_l x(t-l) + \sum_{m=1}^{n_c} c_m e(t-m) \tag{6}$$

$a_k$ ,  $b_l$  et  $c_m$  représentent les nombres de jours précédents pris en considération et  $n_a$ ,  $n_b$  et  $n_c$  représentent le nombre des jours précédents pour chaque variable pris en compte dans le modèle. La température de l'air, les précipitations et le débit du cours d'eau sont représentés par l'élément  $x(t-1)$ , les valeurs estimées de la température de l'eau sont représentées par  $\hat{y}(t-k)$ . En général, la régression linéaire ne convient pas puisque les données sont supposées dépendantes, leur contribution étant alors arbitraire et ne reflétant pas la réalité de la relation (ZHANG, 2001).

Dans la première partie, nous nous sommes intéressés à la prédiction de la température de l'eau en rivière sans faire référence à un temps spécifique. Nous nous proposons dans cette partie de tenir compte de la chronologie des données. Il s'agit d'estimer la température de l'eau des jours à venir ( $j + i, i = 1, 2, \dots$ ). Notons que l'architecture du modèle neuronal se complique notablement à chaque étape, puisqu'avec notre technique, le nombre d'entrées augmente. Afin de préserver la parcimonie du modèle neuronal, nous avons jugé utile de réduire le nombre de paramètres d'entrée, en éliminant les deux variables représentant le cumul des précipitations et le débit ainsi que les variables représentant les températures de l'air ambiant des jours précédents. Les paramètres statistiques d'évaluation des modèles restent les mêmes.

Pour réaliser une multitude d'estimations par les réseaux de neurones, nous utilisons deux méthodes. La première procède par itération, elle utilise la valeur estimée du jour  $j$  pour estimer la valeur de la température de l'eau du jour  $j + 1$ . Nous réalisons ainsi autant de relations que de valeurs à estimer, en utilisant les relations suivantes :

$$\begin{aligned} \hat{y}(j) &= f_0(x(j), y(j-1), y(j-2), y(j-3)) \\ \hat{y}(j+1) &= f_1(\hat{y}(j), x(j), y(j-1), y(j-2), y(j-3)) \\ \hat{y}(j+2) &= f_2(\hat{y}(j+1), \hat{y}(j), x(j), y(j-1), y(j-2), y(j-3)) \\ &\bullet \\ &\bullet \\ \hat{y}(j+7) &= f_7(\hat{y}(j+6), \hat{y}(j+5), \\ &\dots \hat{y}(j+1), \hat{y}(j), x(j), y(j-1), y(j-2), y(j-3)) \end{aligned} \tag{7}$$



**Figure 6.** (a) Architecture neuronale pour la prédiction par itération, (b) Architecture neuronale pour la prédiction par bloc. Architecture for the forecast with iteration (a) and with blocks (b).

$x(j)$  représente la température de l'air du jour  $j$ , les fonctions  $f_0, f_1, f_2, \dots, f_7$  sont générées par les modèles neuronaux de chaque étape. La figure 6a illustre cet ensemble de fonctions.

La deuxième méthode est beaucoup plus simple, elle consiste à réaliser une estimation de tous les jours en une seule fois. Une seule fonction est utilisée (Figure 6b) :

$$\begin{aligned} \hat{y}(j), \hat{y}(j+1), \dots, \hat{y}(j+7) \\ = f(x(j), y(j-1), y(j-2), y(j-3)) \end{aligned} \quad (8)$$

Le nombre de neurones dans la couche intermédiaire a été fixé à sept, la technique de minimisation de la fonction de coût étant celle de Levenberg-Marquardt. Pour tester le modèle, une simple décomposition de la base de données initiale a été effectuée. En effet, l'ensemble des trois premières années a servi à l'apprentissage du modèle neuronal, les deux dernières années servant au test. Les paramètres statistiques sont résumés au tableau 3.

En général, le coefficient de détermination diminue subitement dès que l'on prédit au-delà du premier jour. Néanmoins, les valeurs restent acceptables. Les deux méthodes conduisent à des résultats semblables. L'erreur standard augmente aussi. La figure 7 présente la dispersion de l'erreur pour les quatre jours avec la méthode (b). Nous remarquons qu'elle est presque la même d'un jour à l'autre. Cette dispersion ne reflète pas la suite de la chronologie des événements. Les données correspondant aux intervalles, allant du mois de mai jusqu'au mois de décembre 2001 et du mois de janvier jusqu'au mois d'août 2002, sont manquantes. Les valeurs des résiduels sont faibles au cours de la période chaude. En effet, la température de l'eau en rivière varie faiblement en période chaude.

Dans la bibliographie, aucune étude n'a été réalisée pour expliquer la différence entre les deux méthodes. Les avis sont contradictoires quant aux performances d'une méthode par rapport à l'autre. Dans notre cas, la méthode (a) ne présente aucun avantage, malgré le fait qu'on procède par itération. La brusque diminution des paramètres statistiques est due principalement à la propagation de l'erreur, puisque nous nous servons de la valeur calculée de la température de l'eau du jour  $j$  pour réaliser celle du jour  $j+1$ . De surcroît, la nature du modèle connexionniste amplifie cette erreur, en raison de la liaison qui existe entre les paramètres.

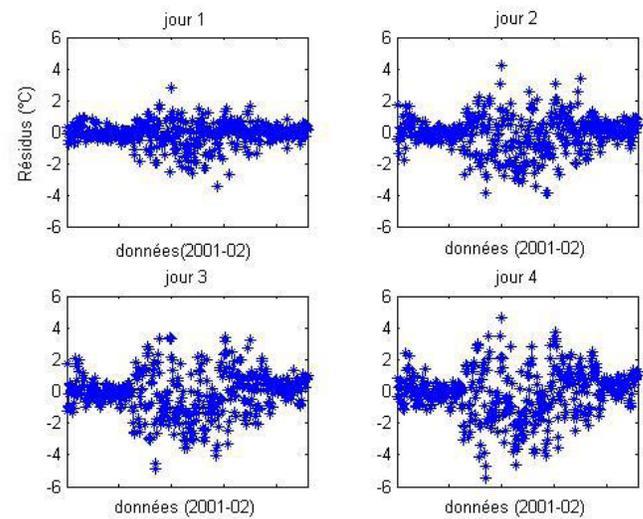


Figure 7. Résidus de la température de l'eau pour les quatre jours (méthode b).

*Residuals of the temperature for the four days (method b).*

Tableau 3. Coefficient de détermination et erreur standard. (a) Prédiction en procédant par itération, (b) prédiction par bloc.  
Table 3. Coefficient of determination and standard error: (a) forecast by iteration, (b) forecast by block.

Jour j	0	1	2	3	4	5	6	7
(a) R <sup>2</sup> (s)	0,99(0,39)	0,95(0,83)	0,89(1,20)	0,86(1,36)	0,84(1,43)	0,83(1,47)	0,83(1,50)	0,82(1,53)
erreur_max	1,38	3,37	3,99	4,90	5,41	6,14	5,53	5,50
nb (res>2 °C)	0 cas	12	47	64	72	79	88	86
(b) R <sup>2</sup> (s)	0,99(0,40)	0,95(0,82)	0,89(1,16)	0,86(1,34)	0,84(1,43)	0,83(1,47)	0,82(1,51)	0,84(1,53)
erreur_max	1,38	3,38	4,10	5,08	5,00	5,68	5,09	5,03
nb (res>2 °C)	15 cas	15	48	71	76	75	82	87

res = Température calculée - Température observée

nb (res >2°C) : nombre d'observations avec une erreur supérieure à 2

R<sup>2</sup> et s sont respectivement le coefficient de détermination et l'écart-type

## 5. CONCLUSION

Dans ce travail, nous avons montré que le modèle connexionniste basé sur les réseaux de neurones est approprié à la modélisation de la température de l'eau d'un ruisseau. La température de l'air ambiante et les températures de l'eau des jours précédents suffisent pour construire ce modèle.

D'autre part, nous avons utilisé deux techniques pour réaliser des prévisions dans le futur. Elles ont conduit à des résultats identiques jugés très satisfaisants. La technique qui procède par itération est prometteuse si l'amplification de la propagation de l'erreur est contrôlée.

Dans la pratique, le modèle serait nettement meilleur si l'on disposait d'un modèle de prédiction de la température de l'air des jours  $j + 1, j + 2, \dots, j + 6$ . Les fonctions qui permettent de prévoir la température de l'eau auraient la forme suivante :

$$\begin{aligned} \hat{y}(j) &= f_0(x(j), y(j-1), y(j-2), y(j-3)) \\ \hat{y}(j+1) &= f_1(\hat{y}(j), \hat{x}(j+1), x(j), y(j-1), y(j-2), y(j-3)) \\ \hat{y}(j+2) &= f_2(\hat{y}(j+1), \hat{y}(j), \hat{x}(j+2), x(j), y(j-1), y(j-2), y(j-3)) \\ &\bullet \\ &\bullet \\ \hat{y}(j+7) &= f_7(\hat{y}(j+6), \hat{y}(j+5), \dots) \\ \hat{y}(j+1), \hat{y}(j), \hat{x}(j+7), x(j), y(j-1), y(j-2), y(j-3) \end{aligned} \quad (9)$$

$\hat{x}(j)$  représente le modèle de prédiction de la température de l'air.

## REMERCIEMENTS

Nous remercions l'Agence Universitaire de la Francophonie qui a apporté son support financier à ce travail, et ce, en accordant une bourse de mobilité scientifique entre la faculté des sciences et techniques de Mohammedia et l'INRS-ETE. Le bénéficiaire de la bourse est le professeur Mohamed Nohair, enseignant au département de chimie.

## RÉFÉRENCES BIBLIOGRAPHIQUES

BASTARACHE, D., N. EL JABI, N. TURKKAN et T.A. CLAIR (1997). Predicting conductivity and acidity for small streams using neural network. *Rev. Can. Gén. Civ.*, 24, 1030-1039.

BOX, G.E. P. et G.M. JENKINS (1976). *Times series analysis: forecasting and control*. 2<sup>e</sup> éd. Holden-Day, San-Francisco, Californie, (édition révisée).

CAISSIE, D., N. EL-JABI et M.G. SATISH (2001). Modelling of maximum daily water temperatures in small stream using air temperature. *J. Hydrol.*, 251, 14-28.

CLAIR, T. A. et J.M. ETHERMAN (1998). Using neural networks to assess the influence of changing seasonal climates in modifying discharge dissolved organic carbon, and nitrogen export in eastern Canadian rivers, *Water Resour.*, 34, 3, 447-455.

CLUIS, D.A.(1972). Relationship between stream water temperature and ambient air temperature – A simple autoregressive model for mean daily stream water temperature fluctuations. *Nordic Hydrol.*, 3, 2, 65-71.

COULIBALY, P., F. ANCTIL et B. BOBÉE (2000). Daily reservoir forecasting using artificial neural networks with stopped training approach. *J. Hydrol.*, 230, 244-257.

DREYFUS, G., J. MARTINEZ, M. SAMUELIDES, M. GORDON, F. BADRAN, S. THIRIA et L. HÉRAULT (2002). *Réseaux de neurones - Méthodologie et applications*. Édition Eyrolles, Paris, France.

HOPPER, P. H (2001). Applying the scientific method to small catchment studies: a review of the Panola Mountain experience. *Hydrol. Proc.*, 15, 2019-2050.

LUK, K.C., J.E. BALL et A. SHARMA (2000). A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *J. Hydrol.*, 227, 56-65.

MAIER, H.R. et G.C. DANDY (1996). The use of artificial neural network for the prediction of water quality parameters. *Water Resour. Res.*, 32, 4, 1013-1022.

MOHSENI, O. et H.G. STEFAN (1999). Stream temperature/ air temperature relationship: a physical interpretation. *J. Hydrol.*, 218, 128-141.

PIPER, R.G., I.B. MCELWAIN, L.E. ORME, J.P. MCCRAREN, L.G. FOWLER et J.R. LEONARD (1982). *Fish hatchery management*. U.S. Fish and Wildlife Service, Department of the Interior, 517 p.

STEFAN, H.G. et E.B. PREUD'HOMME (1993). Stream temperature estimation from air temperature. *Water Resour. Bull.*, 29, 1, 27-45.

TOTH, E., A. BRATH et A. MONTANARI (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *J. Hydrol.*, 239, 132-147.

ZEALAND, C.M., D.H. BURN et S. PSIMONOVIC (1999). Short term streamflow forecasting using artificial neural network. *J. Hydrol.*, 214, 32-48.

ZHANG, G.P. (2001). An investigation of neural networks for linear time-series forecasting. *Comp. Oper. Res.*, 28, 1183-1202.