

Article

« Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple »

M. Bélanger, N. El-Jabi, D. Caissie, F. Ashkar et J. M. Ribí

Revue des sciences de l'eau / Journal of Water Science, vol. 18, n° 3, 2005, p. 403-421.

Pour citer cet article, utiliser l'information suivante :

URI: <http://id.erudit.org/iderudit/705565ar>

DOI: 10.7202/705565ar

Note : les règles d'écriture des références bibliographiques peuvent varier selon les différents domaines du savoir.

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter à l'URI <https://apropos.erudit.org/fr/usagers/politique-dutilisation/>

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche. Érudit offre des services d'édition numérique de documents scientifiques depuis 1998.

Pour communiquer avec les responsables d'Érudit : info@erudit.org

Estimation de la température de l'eau en rivière en utilisant les réseaux de neurones et la régression linéaire multiple

Water temperature prediction using neural networks and multiple linear regression

M. BÉLANGER¹, N. EL-JABI^{1*}, D. CAISSIE², F. ASHKAR³, J.-M. RIBI⁴

Reçu le 8 avril 2003, accepté le 22 juillet 2005**.

SUMMARY

Water temperature is a parameter of great importance for water resources. For instance, modifications of the thermal regime of a river can have a significant impact on fish habitat. Therefore, understanding and predicting water temperatures is essential in order to help prevent or forecast high temperature problems. In order to predict water temperatures, data series are necessary. Many data series exist for air temperatures, but water temperature series are relatively scarce and those available are often short or have missing values. This study presents the modelling of water temperature using neural networks and multiple linear regression to relate water temperature to air temperature and discharge in Catamaran Brook, New Brunswick, Canada.

Catamaran Brook is a small stream (51 km²) where long-term multidisciplinary habitat research is being carried out. Many variables can impact water temperatures in a river, such as air temperature, solar radiation, wind speed, discharge, groundwater flow, etc. For this study, only air temperature and discharge were used. These were judged to be the most often available parameters for modelling temperatures in rivers, and to have the greatest impact on water temperature. More precisely, input variables included current air temperature (°C), air temperature of the previous day (°C), air temperature two days earlier (°C), discharge (m³/s) and a trigonometric function of time (days). Data used for the analysis were from 1991 to 2000. Data from 1991 to 1995 were used to calibrate the model while data from 1996 to 2000 were used for validation purposes. Observed and predicted water temperatures

1. Département de génie civil, Université de Moncton, Moncton, NB, Canada, E1A 3E9.
École de génie, Université de Moncton, Moncton, NB, Canada, E1A 3E9 - tél : (506) 858-4300, fax : (506) 858-4082.
2. Ministère des Pêches et Océans, Boîte postale 5030, Moncton, NB, Canada, E1C 9B6.
3. Département des mathématiques et statistiques, Université de Moncton, Moncton, NB, Canada, E1A 3E9.
4. Département de génie civil, École d'ingénieurs de Fribourg, Fribourg, Suisse.

* Correspondance: eljabin@umoncton.ca

** Les commentaires seront reçus jusqu'au 30 juin 2006.

for each model were presented for the calibration data and the validation data. The coefficient of determination, R^2 , was used to compare the efficiency of both models as well as the residual standard deviation and the bias. This is equivalent to basing the comparison on the standard deviation (or variance) of the residuals. Coefficients of determination for calibration were 94.2% for the neural networks and 92.6% for the multiple linear regression, which correspond to a residual standard deviation of 1.01°C for the neural networks and of 1.05°C for the multiple linear regression. For validation, coefficients of determination were 92.2% for the neural networks and 91.6% for the multiple linear regression, which correspond to a residual standard deviation of 1.10°C for the neural networks, and of 1.25°C for the multiple regression. The overall bias during the study period (1991-2000) was calculated at + 0.11°C for the neural network model and at - 0.26°C for the regression model. Results indicated that it was possible to predict water temperature for a small stream using air temperature, flow and time, as input variables, with neural networks and multiple linear regression. The residual series obtained by both models were very similar. Of the two models, neural networks gave slightly better results in terms of fit, but the small difference in results lets us believe that both approaches are equally good in predicting stream water temperatures.

Keywords: *small stream, water temperature, model, neural networks, multiple linear regression.*

RÉSUMÉ

La température de l'eau en rivière est un paramètre ayant une importance majeure pour la vie aquatique. Les séries temporelles décrivant ce paramètre thermique existent, mais elles sont moins nombreuses et souvent courtes, ou comptent parfois des valeurs manquantes. Cette étude présente la modélisation de la température de l'eau en utilisant des réseaux de neurones et la régression linéaire multiple pour relier la température de l'eau à celle de l'air et le débit du ruisseau Catamaran, situé au Nouveau-Brunswick, Canada. Une recherche multidisciplinaire à long terme se déroule présentement sur ce site. Les données utilisées sont de 1991 à 2000 et comprennent la température de l'air de la journée en cours, de la veille et de l'avant-veille, le débit ainsi que le temps transformé en série trigonométrique. Les données de 1991 à 1995 ont été utilisées pour l'entraînement ou la calibration du modèle tandis que les données de 1996 à 2000 ont été utilisées pour la validation du modèle. Les coefficients de détermination obtenus pour l'entraînement sont de 94,2 % pour les réseaux de neurones et de 92,6 % pour la régression linéaire multiple, ce qui donne un écart-type des erreurs de 1,01 °C pour les réseaux de neurones et de 1,05 °C pour la régression linéaire multiple. Pour la validation, les coefficients de détermination sont de 92,2 % pour les réseaux de neurones et de 91,6 % pour la régression linéaire multiple, ce qui se traduit en un écart-type des erreurs de 1,10 °C pour les réseaux de neurones et de 1,25 °C pour la régression linéaire multiple. Durant la période d'étude (1991-2000), le biais a été calculé à + 0,11 °C pour le modèle de réseaux de neurones et à - 0,26 °C pour le modèle de régression. Ces résultats permettent de conclure qu'il est possible de prévoir la température de l'eau de petits cours d'eau en utilisant la température de l'air et le débit, aussi bien avec les réseaux de neurones qu'avec la régression linéaire multiple. Les réseaux de neurones semblent donner un ajustement aux données légèrement meilleur que celui offert par la régression linéaire multiple, toutefois ces deux approches de modélisation démontrent une bonne performance pour la prédiction de la température de l'eau en rivière.

Mots clés : *petit cours d'eau, température de l'eau, modèle, réseaux de neurones, régression linéaire multiple.*

1 - INTRODUCTION

La température de l'eau est un paramètre des plus importants pour la vie aquatique en rivière. En fait, la majorité des paramètres de qualité de l'eau, tels les paramètres physiques, chimiques et biochimiques, sont dépendants de la température de l'eau (NEMEROW, 1985). De plus, une modification du régime thermique d'un cours d'eau peut avoir un impact significatif sur son habitat. Notamment les températures extrêmes peuvent être problématiques pour les poissons en augmentant leur niveau de stress (GARSIDE, 1973, LUND *et al.* 2002). L'augmentation de température entraîne aussi l'apparition d'algues flottantes et d'organismes aquatiques indésirables. Une meilleure connaissance du régime thermique d'une rivière ainsi que l'aptitude à prédire la température de l'eau sont donc nécessaires pour gérer les ressources aquatiques et effectuer des études sur l'impact environnemental.

La température de l'eau d'une rivière est déterminée par deux types de facteurs, soit des facteurs météorologiques et géophysiques. Les principaux facteurs météorologiques sont la température de l'air, l'intensité et la durée de la radiation solaire, la vitesse du vent à la surface et la pression de la vapeur d'eau dans l'air. Les principaux facteurs géophysiques sont la profondeur de la rivière, le débit, le degré de turbulence, les dimensions de surfaces libres et la température de l'écoulement souterrain (MORIN et COUILLARD, 1990).

Pour estimer la température de l'eau, plusieurs modèles ont été développés et utilisés (CLUIS, 1972; MORIN et COUILLARD, 1990; STEFAN et PREUD'HOMME 1993). MARCEAU *et al.* (1986) ont classifié les modèles utilisés pour estimer la température de l'eau en deux catégories : (i) modèles déterministes et (ii) modèles stochastiques. Les modèles déterministes utilisent une approche du genre conceptuel avec une relation de cause à effet entre les paramètres météorologiques, les caractéristiques physiques du cours d'eau et la température de l'eau en rivière. Ces modèles de température tiennent compte de la majorité des facteurs météorologiques et géophysiques. Toutefois, la disponibilité de tous ces facteurs étant parfois problématique, il est alors préférable d'élaborer des modèles de température plus simples. L'approche stochastique est basée quant à elle sur la structure de la série temporelle de la température de l'eau et estime la température à partir d'analyse de séries temporelles (EL-JABI *et al.*, 1995). Les paramètres nécessaires étant beaucoup moins nombreux que ceux nécessaires pour l'approche déterministe, les coûts d'exploitation sont moins élevés et la durée de développement est réduite.

Plusieurs travaux sur la modélisation stochastique de la température de l'eau ont été réalisés dans le début des années 1970 (CLUIS, 1972; KOTHANDARAMAN, 1971; KOTHANDARAMAN, 1972). Ces études ont démontré que les variations de températures de l'eau sont significativement reliées aux températures de l'air (CLUIS, 1972, STEFAN et PREUD'HOMME, 1993, MOHSENI et STEFAN, 1999). Par exemple, KOTHANDARAMAN (1971) a montré que les écarts de la température de l'eau étaient corrélés aux écarts de températures de l'air des deux jours précédents.

La plupart des études antérieures de modélisation de la température de l'eau en rivière se sont intéressées à l'application de modèles déterministes et stochastiques, modèles qui requièrent un niveau d'expertise important. Il existe

d'autres approches plus simples et tout aussi efficaces qui ont été très peu considérées dans la littérature, à savoir la régression linéaire multiple et les réseaux de neurones. L'objectif de cette étude est d'appliquer ces deux méthodes en utilisant une quantité minimale de facteurs météorologiques et géophysiques, et de comparer leurs performances à celles des modèles utilisés dans le cadre des études antérieures. Pour cette étude, les données utilisées pour modéliser la température de l'eau sont la température de l'air et le débit, puisque ces paramètres ont l'impact le plus important sur la température de l'eau et sont le plus souvent disponibles.

2 – RÉGION D'ÉTUDE

La région d'étude est le ruisseau Catamaran, un affluent de la rivière « Little Southwest Miramichi » situé au centre du Nouveau-Brunswick, Canada (figure 1). Ce ruisseau est localisé à la latitude $46^{\circ} 52,7' N$ et à la longitude $66^{\circ} 06,0' O$. La superficie de son bassin de drainage est de 51 km^2 . Son écoulement est naturel et aucune activité industrielle ou pollution thermique n'est présente. Le ruisseau Catamaran est le site d'une étude multidisciplinaire hydro-biologique visant à quantifier l'écosystème et à évaluer l'impact des coupes de bois dans la région (CUNJAK *et al.*, 1990). Le saumon de l'Atlantique est l'espèce de poisson la plus commune du ruisseau Catamaran (CUNJAK *et al.*, 1993) et des relevés de pêche électrique démontrent que le ruisseau est l'un des cours d'eau de la rivière Miramichi (RANDALL, 1981) qui produit le plus de saumons de l'Atlantique.

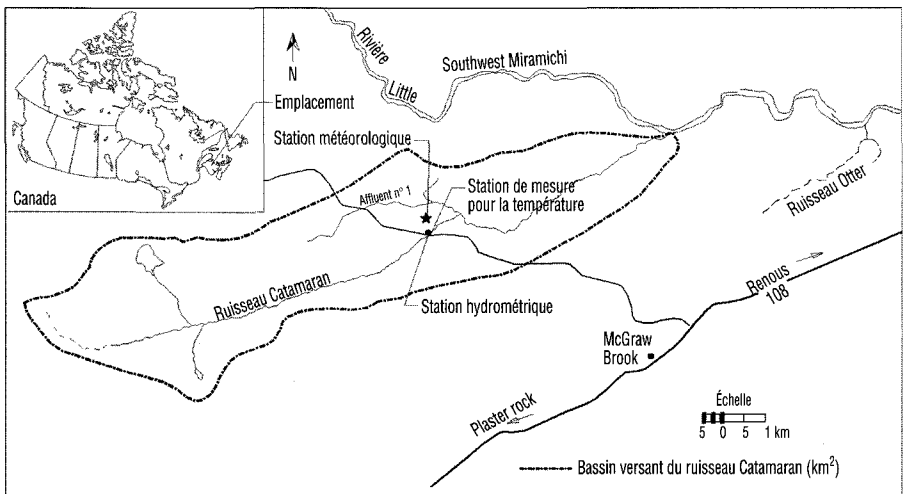


Figure 1 Bassin versant du ruisseau Catamaran avec emplacement des stations hydrométrique et météorologique.

Catamaran Brook basin showing the location of hydrometric and meteorological stations.

Le ruisseau Catamaran mesure 20,5 km de longueur, a une largeur d'environ 10 m et une profondeur de 0,3 m dans la partie médiane du bassin. Le niveau d'élévation se situe à environ 70 mètres au-dessus du niveau moyen de la mer, allant jusqu'à 335 mètres dans son cours supérieur. La densité de drainage est de $0,60 \text{ km}^{-1}$ (CUNJAK *et al.*, 1990). Le bassin versant compte deux lacs, dont le principal (côté Nord du bassin) est le lac Catamaran, possédant une superficie de $0,3 \text{ km}^2$ (figure 1). Le second (côté sud du bassin) est sans nom et a une superficie de $0,05 \text{ km}^2$. Les données historiques sur la température de l'air et la précipitation ont été enregistrées à partir de la station McGraw Brook (figure 1) opérée par le ministère des Ressources naturelles du Nouveau-Brunswick.

Les données de température de l'air utilisées dans la présente étude proviennent de la station météorologique du ruisseau Catamaran. L'analyse du débit au ruisseau Catamaran montre qu'une fréquence élevée des crues est observée au printemps avec une fréquence maximum au mois d'avril, qui résulte sans doute de la précipitation et surtout de la fonte des neiges. La fréquence des crues est généralement faible en hiver (EL-JABI *et al.*, 1999). En ce qui a trait aux statistiques sur les températures de l'air, le mois de janvier a la température moyenne la plus froide, soit $-11,8 \text{ }^\circ\text{C}$. Juillet est le mois le plus chaud, avec une moyenne de $18,8 \text{ }^\circ\text{C}$ (CUNJAK *et al.* 1993). Pour mesurer le débit, une station hydrométrique a été installée au milieu du bassin en octobre 1989. Le niveau d'eau enregistré a été utilisé pour calculer le débit de la rivière en continu. Le débit moyen annuel pour le ruisseau Catamaran est estimé à $1,3 \text{ m}^3/\text{s}$. Durant les périodes de crues de la fin avril au début mai, le plus grand débit enregistré a été de $25 \text{ m}^3/\text{s}$ le 3 mai 1991. Le plus petit débit enregistré est survenu du 31 juillet au 1^{er} août 1991 pour une valeur de $0,068 \text{ m}^3/\text{s}$ (CAISSIE *et al.*, 1998). Une sonde pour mesurer la température de l'eau a été installée en milieu de bassin pour enregistrer les températures à chaque heure. La température la plus élevée enregistrée entre 1991 et 2000 a été mesurée le 18 juillet 1999 à $21,05 \text{ }^\circ\text{C}$.

3 – MÉTHODOLOGIE

Les données utilisées pour la modélisation numérique (régression linéaire multiple et réseaux de neurones) de la température de l'eau sont : la température de l'air ($^\circ\text{C}$), la température de l'air de la veille ($^\circ\text{C}$), la température de l'air de l'avant-veille ($^\circ\text{C}$), le débit (m^3/s) et le temps (jours). Les températures de l'air de la journée précédente et des deux jours précédents ont été utilisées pour souligner le fait que la température de l'eau ne change pas aussi vite que celle de l'air (KOTHANDARAMAN, 1971 ; CLUIS, 1972). Le choix de ces données est basé sur deux critères, soit la disponibilité des données et la corrélation accrue entre ces paramètres telle que démontrée par plusieurs études précédentes (CLUIS, 1972, STEFAN et PREUD'HOMME, 1993, MOHSENI et STEFAN, 1999, SONG et CHIEN, 1977).

Le temps a été transformé en série trigonométrique (en utilisant des fonctions sinus et cosinus) dans le but de faciliter l'apprentissage des réseaux de

neurones (BASTARACHE, 1995). Les fonctions suivantes ont été utilisées pour transformer la variable temps, exprimée en jour de l'année :

$$\text{temps cos} = \cos\left(2\pi \times \frac{JJ}{365}\right) \quad (1)$$

$$\text{temps sin} = \sin\left(2\pi \times \frac{JJ}{365}\right) \quad (2)$$

Ce traitement permet de présenter le temps au réseau sous forme continue dans l'intervalle $[-1, 1]$, (BASTARACHE, 1995).

Les données utilisées dans cette étude sont des moyennes journalières calculées à partir des données prises à chaque heure (moyenne de 24 observations par jour) et couvrant la période de 1991 à 2000. Pour le choix de la période d'analyse et de validation du modèle, on se limite à la période du 3 mai (jour 123) au 17 octobre (jour 290), soit la période où la température de l'eau est supérieure à 0 °C. De plus, il s'agit de la plage de dates des années 1991-2000 où la température de l'eau est disponible. Pour tester la validité des deux modèles, la série de données a été divisée en une série d'entraînement (calibration) (1991-1995), et une série d'évaluation (validation) (1996-2000). Dans l'étape d'entraînement, les coefficients des réseaux de neurones sont ajustés afin que les vecteurs de sortie (température de l'eau) soient le plus près des vecteurs de sortie mesurés. Au total, 1 690 jours ont été utilisés. Toutes les années à l'étude ont été utilisées afin de représenter graphiquement les résultats des deux modèles, pour les valeurs d'entraînement et les valeurs de validation observées et estimées. Le but de cette étude consistait à comparer les coefficients de détermination (R^2), l'écart-type des erreurs et le biais pour chaque modèle.

4 – CONSIDÉRATIONS THÉORIQUES

4.1 Réseaux de neurones

La méthode des réseaux de neurones est basée sur une représentation schématique des neurones dans le cerveau. Cette méthode a été développée suite à des observations biologiques qui indiquent que les neurones font partie d'un réseau structuré. Les neurones sont organisés en couches, et il y a échange d'information entre les couches. Le problème de simulation consiste à trouver une relation satisfaisante entre un groupe de neurones représentant les données d'entrées et le groupe associé aux données de sorties (PAO, 1989). Le but des réseaux de neurones est donc d'entraîner un réseau avec les données disponibles afin d'avoir la meilleure corrélation possible entre les données de sortie mesurées et celles estimées. Il existe plusieurs modèles de réseaux de neurones, dont les modèles de Hopfield, Hamming, Carpenter, le perceptron à une couche et le perceptron multicouche (MOSELHI *et al.*, 1992). Les réseaux à une couche sont assez limités car ils n'utilisent que des relations linéaires avec une seule couche de transfert des informations. Pour faire des calculs plus complexes, les réseaux les plus utilisés sont les réseaux multicouches (figure 2). Ce sont ces réseaux qui sont utilisés dans la présente étude.

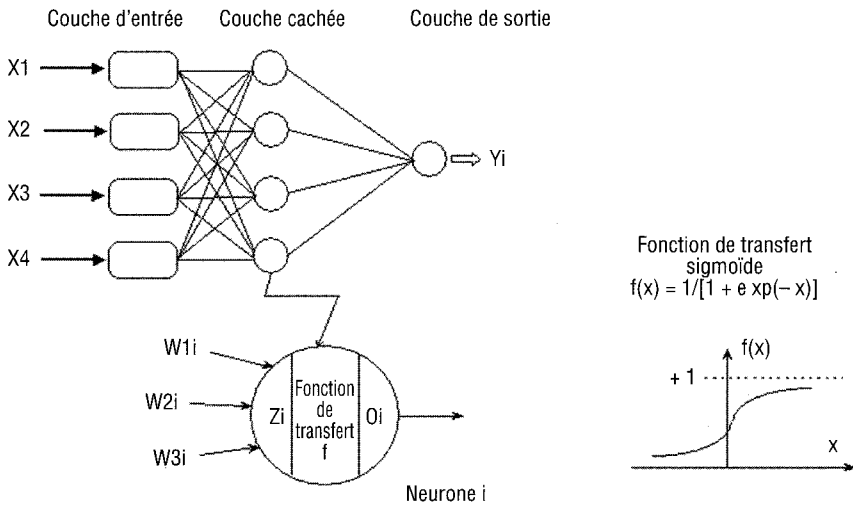


Figure 2 Réseaux multicouches (NNMODEL32 version 1.2 2.0. Copyright 1994-1998. Neural Fusion Shareware).

Multilayer networks.

Parmi ces réseaux, on distingue les réseaux à rétro-propagation du gradient de l'erreur. Ces derniers utilisent des algorithmes d'entraînement des poids qui permettent d'entraîner des réseaux multicouches. La figure 2 montre un réseau multicouche. Chaque neurone *i* reçoit une série de signaux des neurones *j* situés aux couches précédentes. Le fonctionnement du réseau illustré est gouverné par l'équation suivante (TURKKAN, 1993) :

$$Z_i = \sum_{j=1}^{n_j} W_{ij} X_j + b_i \tag{3}$$

où n_j est le nombre de neurones d'entrée et X_j est la valeur du signal transmis par le neurone *j* de la couche précédente. Les W_{ij} représentent les poids respectifs des connexions entre les neurones *j* des couches précédentes et le neurone *i* de la couche actuelle. Les paramètres b_i sont des valeurs de biais permettant une fonction de transfert non nulle à l'origine. Les entrées X_j sont pondérées par les poids W_{ij} . Une fois l'entrée alimentée, le neurone *i* la transforme et produit une sortie. Dans ce cas, Z_i et la sortie O_i d'un neurone donné, sont reliées par une fonction de transfert de forme sigmoïde. L'avantage de cette fonction est sa continuité par rapport à une fonction d'activation linéaire (BASTARACHE, 1995) :

$$O_i = f(Z_j) = \frac{1}{1 + e^{-Z_j}} \tag{4}$$

Dans le cadre de cette recherche, on a utilisé le logiciel NNMODEL32 (version 1.2 2.0. Copyright 1994-1998 Neural Fusion Shareware). Ce logiciel est basé sur un modèle multicouche avec une seule couche cachée. En plus, ce logiciel utilise l'algorithme d'apprentissage de rétro-propagation du gradient supervisé. C'est

un algorithme de distribution des erreurs sur un réseau multicouche. L'erreur entre la sortie actuelle et la sortie désirée se propage, tout en ajustant les poids W_{ij} . Donc, l'objectif de cette méthode est d'apporter des corrections aux poids du réseau W_{ij} afin de réduire l'erreur globale \underline{E} . Pour minimiser l'erreur globale (entre les sorties obtenues et les sorties souhaitées), on utilise la méthode de gradient descendant basée sur l'équation suivante (figure 3) :

$$\underline{E} = \frac{1}{2} \sum_{k=1}^{n_k} (S_k - O_k)^2 \quad (5)$$

où S_k représente la valeur estimée et O_k la valeur observée et \underline{E} l'erreur globale.

L'avantage des réseaux de neurones est que la sortie est déterminée sans qu'il ne soit nécessaire de connaître la relation entre ce dernier et les paramètres d'entrée. L'inconvénient de cette approche est qu'il est difficile d'avoir une explication physique du résultat, dû au fait que les réseaux de neurones sont de type boîte noire.

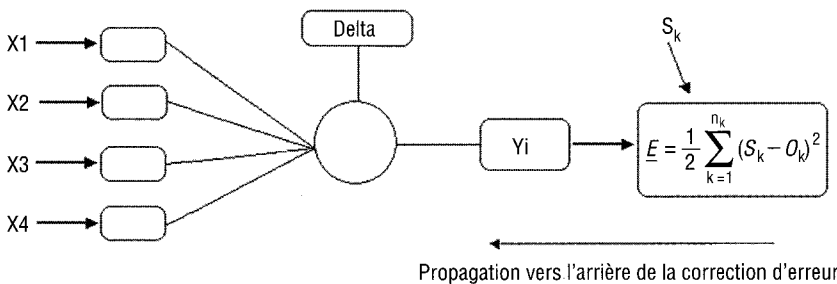


Figure 3 Détail de la couche cachée (NNMODEL32 version 1.2 2.0. Copyright 1994-1998. Neural Fusion Shareware).

Details of the hidden layer.

4.2 Régression linéaire multiple

Comme les réseaux de neurones, la régression linéaire multiple est utilisée pour prévoir les valeurs d'une variable dépendante à partir de variables explicatives ou indépendantes. Les régressions linéaires multiples sont utilisées pour trouver la relation linéaire la plus satisfaisante pour prévoir la valeur dépendante qui produit l'erreur type la moins grande. Dans un tel modèle, chaque variable indépendante est pondérée afin que la valeur des coefficients de régressions maximise l'influence de chaque variable dans l'équation finale. Il est possible de manipuler plusieurs variables indépendantes à partir des régressions linéaires multiples, mais seulement une variable dépendante.

4.3 Statistique de comparaison

Afin de comparer les résultats entre les différentes méthodes numériques (réseaux de neurones et régression linéaire multiple), trois indices de perfor-

mance ont été calculés pour chaque série : le coefficient de détermination (R^2); l'écart-type des erreurs (E) et le biais.

Le coefficient de détermination (R^2) est le pourcentage de l'erreur totale sur la variable dépendante y (température de l'eau) expliqué par le modèle. Ce coefficient s'exprime par :

$$R^2 = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2} \times 100 \quad (6)$$

où y est la valeur mesurée, \bar{y} est la moyenne des valeurs mesurées et \hat{y} est la valeur estimée par le modèle. L'écart-type des erreurs (E) est donné par l'équation suivante :

$$E = \sqrt{\frac{\sum (\hat{y} - y)^2}{n}} \quad (7)$$

tandis que le biais est donné par l'équation suivante :

$$\text{Biais} = \sum (\hat{y} - y) \quad (8)$$

5 – RÉSULTATS

5.1 Réseaux de neurones

L'architecture du réseau n'est pas fixée au préalable. Cependant, il est possible de définir les paramètres de cette architecture (nombre de couches cachées, le critère d'arrêt, etc.). Pour cette application, on a choisi une architecture contenant une seule couche cachée composée de dix neurones (figures 2 et 3).

Le modèle de réseaux de neurones a donné un coefficient de détermination pour l'entraînement de 94,2 %, ce qui équivaut à un écart-type des erreurs de 1,01 °C (figure 4). Pour la période de validation, le coefficient de détermination a été de 92,2 %, ce qui correspond à un écart-type des erreurs de 1,10 °C. La figure 4 montre les résidus en fonction des valeurs simulées, pour les données d'entraînement (figure 4a), et de validation (figure 4b) respectivement. Cette figure montre que les résidus varient généralement entre - 3 °C à + 3 °C. Aucune tendance des résidus n'a été notée en fonction de la température estimée. Par conséquent, ces résultats indiquent que le modèle de réseaux de neurones prédit la température de l'eau avec un niveau de certitude semblable sur toute la gamme de températures étudiée. La figure 5 permet de comparer les valeurs prévues et mesurées du modèle pour chaque année étudiée (1991-2000). Cette figure montre que le modèle de réseaux de neurones reproduit très bien la température de l'eau sur plusieurs années, en particulier durant l'année 1991, 1994, 1995 et 1998. Par contre, le modèle de réseaux de neurones ne reproduit pas aussi bien la température de l'eau en 1992 et 1999. En 1992, le modèle surestime la température de l'eau pendant une période assez

longue de l'été, tandis qu'en 1999 une sous-estimation de la température de l'eau a été observée. Il est aussi à noter que ces deux années constituent des conditions extrêmes de température (températures de l'air faibles en 1992 et élevées en 1999) durant la période en étude, ce qui pourrait expliquer la moins bonne performance du modèle.

Les résultats de la figure 5 indiquent que le modèle de réseaux de neurones prédit assez bien les périodes de température maximum durant l'été, surtout en 1991, 1995 et 1998. En effet, les températures de pointe estimées ainsi que la variabilité de la température durant ces années sont très proches de celles observées au ruisseau Catamaran.

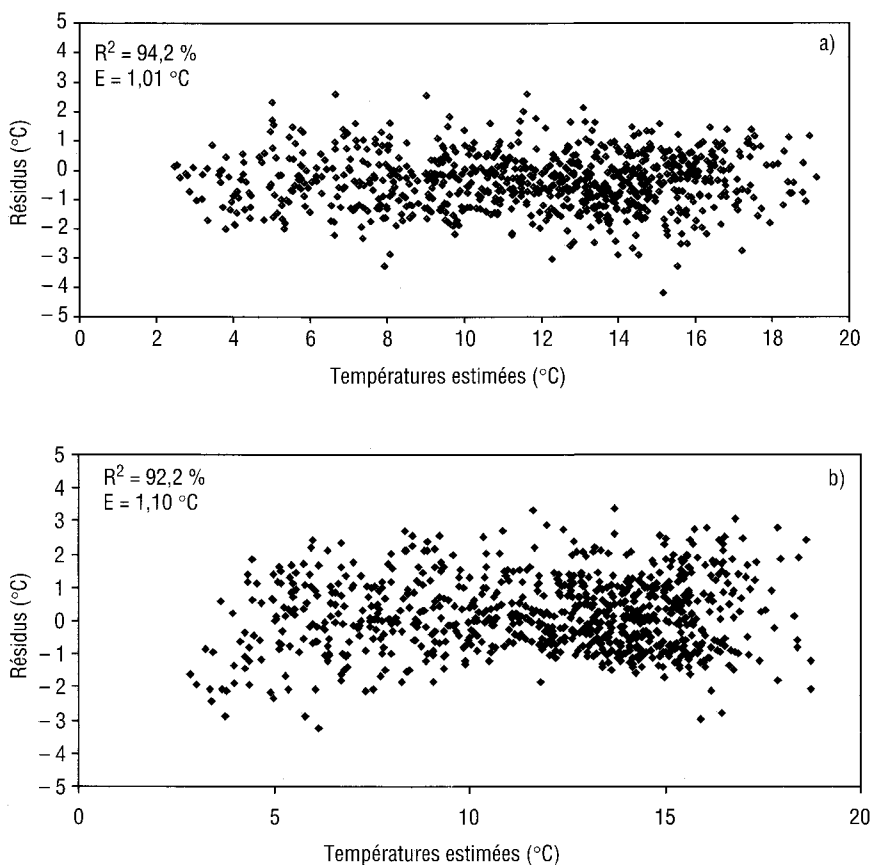


Figure 4 Résidus de la température de l'eau pour le modèle de réseaux de neurones. a) entraînement ; b) validation.

Water temperature residuals of the neural network model. a) training; b) validation.

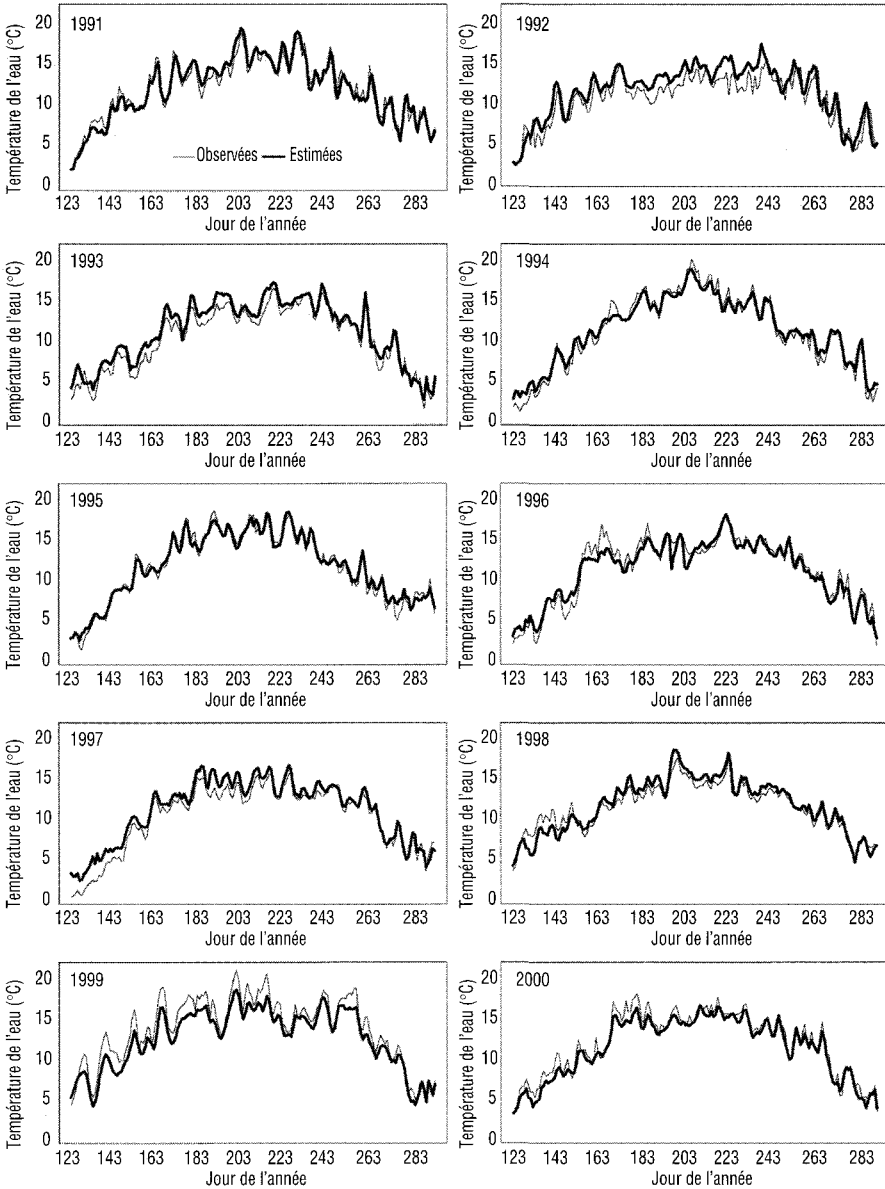


Figure 5 Températures de l'eau observées et estimées avec le modèle de réseaux de neurones au ruisseau Catamaran (1991-2000).
Observed and predicted water temperatures calculated with the neural network model at Catamaran Brook (1991-2000).

5.2 Régression linéaire multiple

Le tableau 1 donne les coefficients de la régression linéaire et l'erreur type de chaque coefficient. D'après ce tableau, on constate qu'il y a une relation étroite entre la température de l'eau et les variables dépendantes (T° air au temps (t), T° air au temps (t-1), T° air au temps (t-2), Débit au temps (t), Temps (cosinus), Temps (sinus)) puisque pour tous les coefficients, le test t de Student a donné des valeurs de probabilité très faibles (inférieures à 1 %). Ceci signifie que chaque variable a une contribution significative sur la température de l'eau.

Tableau 1 Coefficients du modèle de régression avec leurs erreurs types.

Table 1 Coefficients and standard error of coefficients for the regression model.

Variable	Coefficient	Erreur type du coefficient
Constante	1,08	0,103
T° air (t)	0,232	0,010
T° air (t-1)	0,161	0,012
T° air (t-2)	0,123	0,010
Débit (t)	- 0,245	0,036
Temps (cosinus)	- 3,91	0,167
Temps (sinus)	- 2,56	0,085

L'équation de régression obtenue est la suivante :

$$T^\circ \text{ eau (t)} = 1,08 + 0,232 * T^\circ \text{ air(t)} + 0,161 * T^\circ \text{ air(t-1)} + 0,123 * T^\circ \text{ air(t-2)} - 0,245 * \text{Débit (t)} - 3,91 * \text{Temps (cosinus)} - 2,56 * \text{Temps (sinus)} \quad (9)$$

où les variables *Temps (cosinus)* et *Temps (sinus)* sont données par les équations (1) et (2).

Le coefficient de détermination obtenu par le modèle de régression pour la série d'entraînement a été de 92,6 %, avec un écart-type des erreurs de 1,05 °C (figure 6). Le coefficient de détermination pour la série de validation a été de 91,6 %, avec un écart-type des erreurs de 1,25 °C. La figure 6 montre les résidus en fonction des valeurs estimées pour l'entraînement et la validation respectivement. Comme pour le modèle de réseaux de neurones, les résidus du modèle de régression varient généralement de - 3 °C à + 3 °C avec quelques résidus dépassant ces températures. Pour le modèle de régression durant la validation, il est à noter que les résidus sont légèrement plus élevés aux températures extrêmes, tandis que ceux-ci sont généralement plus faibles à des températures moyennes (figure 6b; 9-13 °C). La figure 7 présente les valeurs estimées et mesurées au ruisseau Catamaran pour le modèle de régression pour la période d'observation de 1991 à 2000. Cette figure renseigne sur la performance du modèle pour chaque année. Les résultats de ce modèle sont similaires à ceux observés avec le modèle de réseaux de neurones, c'est-à-dire une bonne performance durant les années 1991, 1995 et 1998. Le modèle de régression linéaire multiple permet d'améliorer légèrement la qualité de l'estimation de la température de l'eau en 1997, comparativement au modèle précédent. Comme pour le modèle de réseaux de neurones, une légère surestimation de la température de l'eau est notée en 1992, et une sous-estimation assez importante a été notée en 1999 et 2000 avec le second modèle. Tel que mentionné pour le modèle de réseaux de neurones, la piètre

performance notée en 1999 peut être attribuée à une température de l'air plus élevée que la normale durant cette année. Les conditions météorologiques étaient effectivement très différentes en 1999, comparativement à celles utilisées durant l'entraînement du modèle.

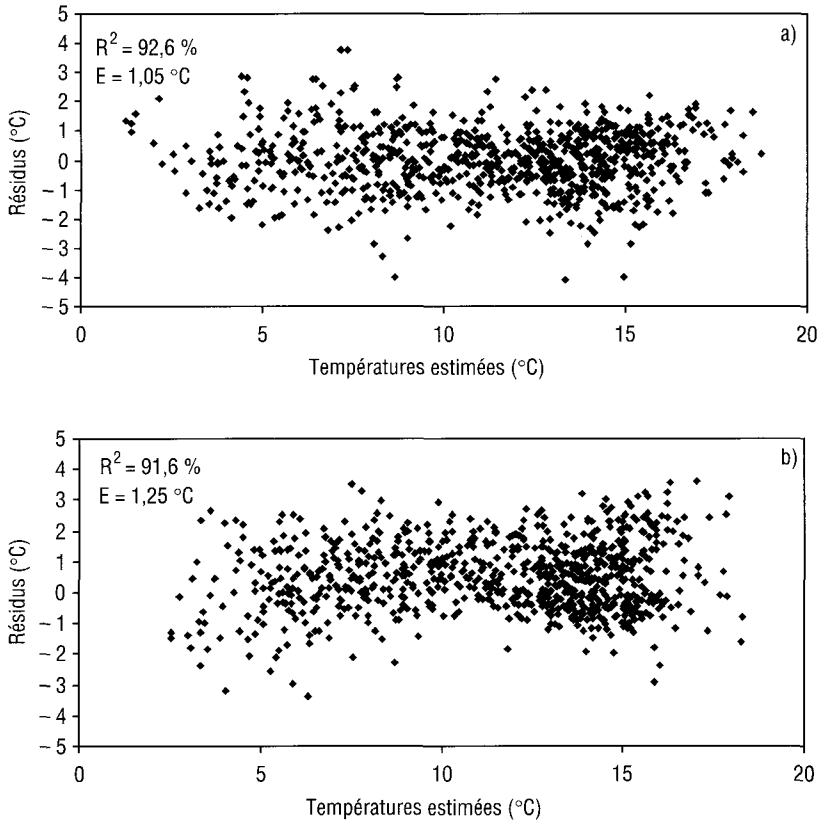


Figure 6 Résidus de la température de l'eau pour le modèle de régression. a) entraînement ; b) validation.

Water temperature residuals for the regression model. a) training; b) validation.

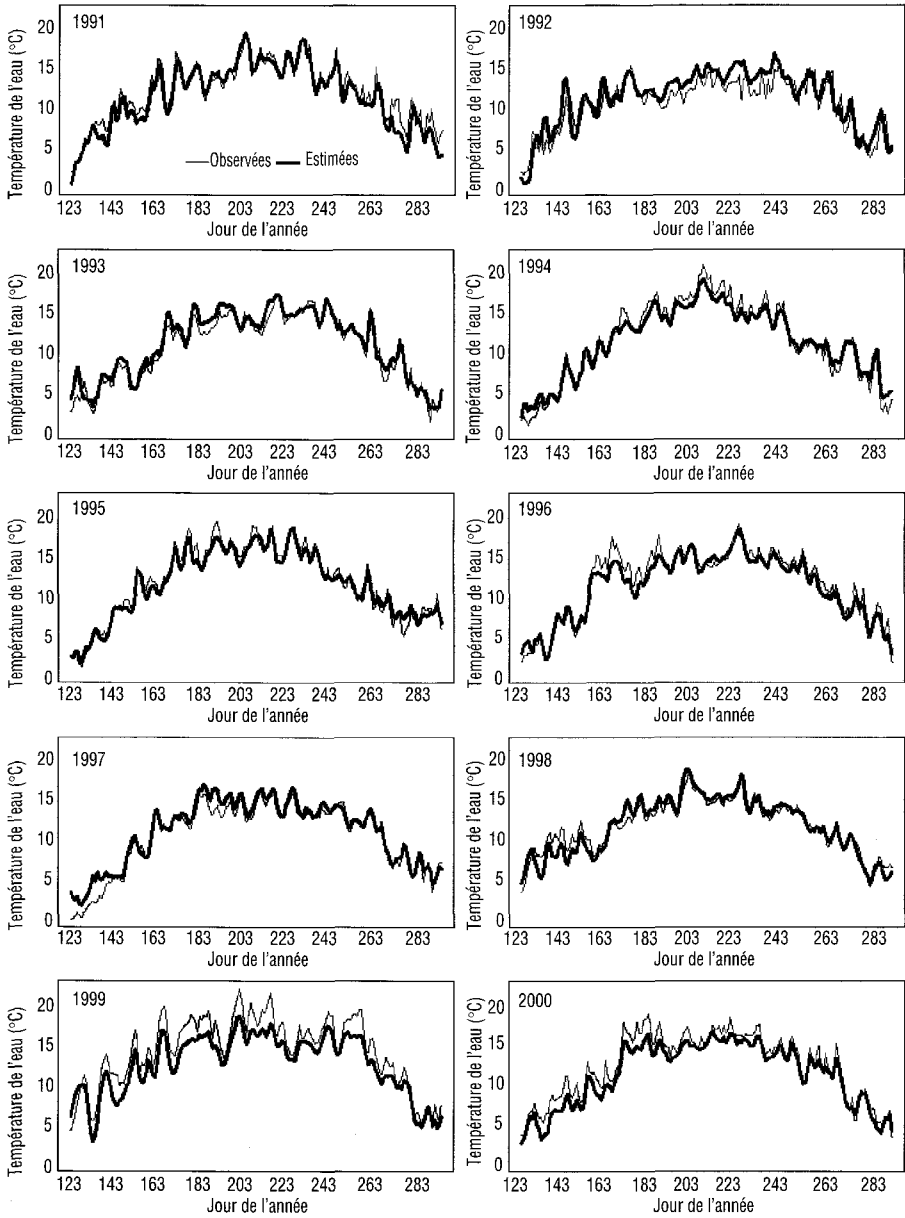


Figure 7 Températures de l'eau observées et estimées avec le modèle de régression au ruisseau Catamaran (1991-2000).

Observed and predicted water temperatures calculated with the regression model (1991-2000).

5.3 Comparaison entre les deux modèles de température de l'eau

Une comparaison entre le modèle de réseaux de neurones et celui de la régression multiple a été effectuée et les écarts-types des erreurs et du biais sont présentés au tableau 2. Par exemple, les deux modèles présentent une performance similaire durant la période d'entraînement avec un écart-type des erreurs de 1,01 °C (réseaux de neurones) et de 1,05 °C (régression) respectivement. Par contre, le biais calculé par le modèle de réseaux de neurones est plus grand (0,39 °C) que celui du modèle de régression (0,00 °C). Il est important de noter que le biais du modèle de régression est nul sur toute la période de calibration, mais peut varier d'une année à l'autre. Ceci est une caractéristique de calibration des modèles de régression. Durant la période de validation, une différence plus importante a été notée au niveau de la performance des deux modèles. D'une part, le modèle de réseaux de neurones conduit à un écart-type des erreurs de 1,10 °C, soit une valeur inférieure à celle du modèle de régression (1,25 °C). Les résultats des deux modèles présentent un biais négatif durant la validation, soit 0,17 °C pour le modèle de réseaux de neurones et de -0,52 °C pour le modèle de régression, ce qui signifie une légère sous-estimation de la température de l'eau durant cette période.

Tableau 2 Résultats (écart-type des erreurs, °C) de la modélisation de la température de l'eau au ruisseau Catamaran par le modèle de réseaux de neurones et de régression.

Table 2 *Water temperature modeling results at Catamaran Brook (residual standard deviation, °C) by the neural network and regression models.*

Années	Écart-type des erreurs du modèle de réseaux de neurones	Biais du modèle de réseaux de neurones	Écart-type des erreurs du modèle de régression	Biais du modèle de régression
1991-1995	1,01	0,39	1,05	0,00
1996-2000	1,10	-0,17	1,25	-0,52
1991	0,80	0,02	1,10	-0,56
1992	1,50	0,85	1,46	0,54
1993	1,08	0,84	0,88	0,50
1994	0,80	0,18	0,92	-0,18
1995	0,64	0,08	0,78	-0,31
1996	1,00	-0,16	0,78	-0,44
1997	1,16	0,85	0,95	0,53
1998	0,92	0,25	0,90	-0,08
1999	1,47	-1,21	1,87	-1,59
2000	0,87	-0,60	1,26	-1,03
1991-2000	1,06	0,11	1,15	-0,26

La comparaison graphique (par année) effectuée antérieurement (figure 5) suggère une bonne performance du modèle de réseaux de neurones pour les années 1991, 1994-95 et 1998. En utilisant les critères du biais et de l'écart-type des erreurs, on remarque également une bonne performance durant ces années, quoique l'année 2000 possède aussi un écart-type des erreurs inférieur à 1 °C. Le biais durant ces années était généralement positif (4 années

sur 5; surestimation du modèle), et particulièrement faible en 1991 (0,02 °C) et 1995 (0,08 °C). Les années moins performantes du modèle sont les années 1992 et 1999 avec un écart-type des erreurs de 1,50 °C et 1,47 °C respectivement. Par ailleurs, le biais durant ces années est de + 0,85 °C (1992) et de - 1,21 °C (1999), démontrant aussi une faible performance du modèle.

Pour le modèle de régression, la comparaison graphique par année (figure 7) montre une bonne performance du modèle pour les années 1991, 1995 et 1998. L'écart-type des erreurs estimé va dans le même sens (valeurs de 1,10 °C, 0,78 °C et 0,90 °C respectivement pour les années 1991, 1995 et 1998 ; tableau 2). À noter que les écarts-types des erreurs de plusieurs autres années suggèrent une performance acceptable du modèle. En particulier, les écarts-types des erreurs associées aux années 1993 et 1996 sont de seulement 0,88 °C et 0,78 °C. Ces résultats démontrent que les deux modèles étudiés offrent généralement des performances semblables chaque année. Le biais en valeur absolue pour le modèle de régression durant les années de bonne performance est généralement inférieur à 0,5 °C, ce qui est semblable aux valeurs obtenues à partir du modèle de réseaux de neurones. Cependant, les années où apparaît un biais très faible présentent des différences entre les deux modèles. Par exemple, le modèle de réseaux de neurones conduit à un biais plus faible en 1991 (+ 0,02) et 1995 (+ 0,08) tandis que les années 1998 (- 0,08 °C) et 1994 (- 0,18 °C) présentent un biais plus faible lorsque le modèle de régression est utilisé.

6 – DISCUSSION ET CONCLUSIONS

L'objectif de cette étude était d'estimer la température de l'eau à partir de paramètres hydrométéorologiques facilement disponibles en utilisant des modèles très peu utilisés dans la littérature, soit le modèle de réseaux de neurones et la régression linéaire multiple. L'utilisation de tels modèles permet par la suite de prédire ou d'estimer la température de l'eau pour des conditions hydrométéorologiques variées ou durant des périodes de données manquantes (série historique ou intermittente). Chacun des deux modèles utilisés pour la prévision a ses avantages et inconvénients. En premier lieu, les réseaux de neurones sont très efficaces mais offrent une description peu claire de la relation entre les données d'entrée et de sortie. Pour la régression linéaire multiple, une équation est déterminée explicitant la relation entre les variables explicatives d'entrée et la température de l'eau. Tel que démontré par nos résultats, ces deux modèles donnent des résultats tout à fait acceptables avec un écart-type des erreurs global de 1,06 °C (réseaux de neurones) et 1,15 °C (régression linéaire multiple) (tableau 2). Un plus grand écart-type des erreurs du modèle de régression linéaire multiple est attribué en grande partie à une faible performance du modèle en 1999.

Des études précédentes ont démontré que les modèles de régression utilisés en modélisation de la température de l'eau sont moins performants sur une base journalière que sur une base hebdomadaire ou mensuelle (ERICKSON et

STEFAN, 2000; CAISSIE *et al.*, 2001). Par exemple, CAISSIE *et al.* (2001) ont obtenu un écart-type des erreurs supérieur à 2 °C sur une base journalière en utilisant un modèle de régression (régression logistique). ERICKSON et STEFAN (2000) ne présentent pas les écarts-types des erreurs de leur analyse de régression linéaire, mais il est clair que ceux-ci sont supérieurs à 2 °C, avec un coefficient de détermination de seulement 0,83 (données journalières). La présente étude montre que, par un choix approprié de paramètres, la régression linéaire multiple et les réseaux de neurones peuvent être très performants sur une base journalière, en plus d'être des modèles simples dans leur application. De plus, les résultats nous permettent de conclure que les écarts-types des erreurs sont semblables à ceux rapportés dans la littérature pour les modèles stochastiques et déterministes sur une base journalière, c'est-à-dire de l'ordre de 0,8 à 1,5 °C (KOTHANDARAMAN 1972 ; MARCEAU *et al.*, 1986 ; CAISSIE *et al.*, 1998). L'avantage des modèles présentés dans cette étude réside surtout au niveau de leur relative simplicité (développement, application et mise à jour du modèle), par rapport aux modèles stochastiques et déterministes. Par exemple, les modèles stochastiques requièrent plusieurs niveaux de calcul, soit au niveau de la composante saisonnière (représentant le cycle de température à long terme) et de l'étude de l'autocorrélation de la série temporelle. D'autre part, les modèles déterministes requièrent plusieurs paramètres hydrométéorologiques qui ne sont pas toujours disponibles, c'est-à-dire les données d'une station météorologique à plusieurs paramètres, le niveau d'eau de la rivière, ainsi que plusieurs facteurs géophysiques.

Il est intéressant de noter que, même si les deux modèles ont des écarts-types des erreurs similaires sur toute la période étudiée (1991-2000), le critère du biais distingue ces deux ensembles de résultats. Plus précisément, on constate que le modèle de régression sous-estime légèrement la température (-0,26 °C ; 1991-2000), tandis que le modèle de réseaux de neurones surestime la température (+ 0,11 °C ; 1991-2000). La variabilité des données hydro-météorologiques durant les années études suggère un bon fonctionnement des deux modèles dans de telles conditions, de façon globale. Par contre, cette étude démontre également que les deux années avec une performance plus faible étaient des années de températures de l'air extrêmes (faible en 1992 et élevée en 1999). Aucune tendance des résidus n'a été notée (en fonction de la température estimée). Par contre, la variabilité semblait moindre pour les températures moyennes, surtout en période de validation du modèle de régression (figure 6b). Ceci suggère que les modèles étudiés peuvent être légèrement plus performants dans des conditions de températures moyennes que de températures extrêmes.

L'ensemble des résultats obtenus nous permet de conclure que les deux méthodes proposées peuvent être exploitées pour l'estimation de la température de l'eau. En effet, les réseaux de neurones ont démontré une capacité importante d'apprentissage et de prédiction de la température. Quant à l'approche de la régression linéaire multiple, elle permet aussi d'avoir des résultats similaires et performants.

REMERCIEMENTS

Cette étude a été rendue possible grâce à l'aide financière du Conseil de recherche en sciences naturelles et génie du Canada, du programme CCAF du ministère des Ressources naturelles et du ministère de Pêches et Océans du Canada. Un sincère remerciement s'adresse aussi au professeur Noyan Turkkan, de la faculté d'ingénierie de l'Université de Moncton, pour son aide concernant les réseaux de neurones, et aux réviseurs pour leurs commentaires constructifs.

RÉFÉRENCES BIBLIOGRAPHIQUES

- BASTARACHE D., 1995. Analyse des données hydro-environnementales du parc Kejimikujik par réseaux de neurones et réseaux polynomiaux. Thèse de maîtrise. Université de Moncton, N.B., Canada.
- CAISSIE D., EL-JABI N., SATISH M.G., 2001. Modelling of maximum daily water temperatures in a small stream using air temperatures. *Journal of Hydrology*, 251, 14-28.
- CAISSIE D., EL-JABI N., ST-HILAIRE A., 1998. Stochastic Modelling of Water Temperature in a Small Stream Using Air to Water Relations. *Canadian Journal of Civil Engineering*, 25, 250-260.
- 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 Hydrology*, 3(2), 65-71.
- CUNJAK R.A., CAISSIE D., EL-JABI N., 1990. Projet de recherche sur l'habitat du ruisseau Catamaran : description et champs d'étude générale. *Rapport Technique Canadien des Sciences Halieutique et Aquatique*, n° 1751.
- CUNJAK R.A., CAISSIE D., EL-JABI N., HARDIE P., CONLON J.H., POLLOCK T.L. GIBERSON D.J., KOMADINA-DOU-THWRIGHT S., 1993. The Catamaran Brook (New Brunswick) habitat research project: biological, physical and chemical conditions (1990-1992). *Canadian Technical Report of Fisheries and Aquatic Sciences*, 1914: 81 p.
- EL-JABI N., EL-KOURDAHI G., CAISSIE D., 1995. Modélisation stochastique de la température de l'eau en rivière. *La Revue des Sciences de l'Eau*, 8, 77-95.
- EL-JABI N., LOUA P. CAISSIE D., 1999. Hydro-météorologie du ruisseau Catamaran. *Revue de l'université de Moncton, Numéro spécial Génie et Technologie*, 5-17.
- ERICKSON T.R., STEFAN H.G., 2000. Linear air/water temperature correlations for streams during open water periods. *Journal of Hydrologic Engineering*, ASCE 5(3), 317-321.
- GARSDALE E.T., 1973. Ultimate upper lethal temperature of Atlantic salmon (*Salmo salar* L.). *Canadian Journal of Zoology*, 51, 898-900.
- KOTHANDARAMAN V., 1971. Analysis of Water Temperature Variations in Large Rivers. *ASCE Journal of the Sanitary Engineering Division*, 97(SA1). 19-31.
- KOTHANDARAMAN V., 1972. Air-Water Temperature Relationship in Illinois River. *Water Resources Bulletin*, 8(1), 38-45.
- LUND S.G., CAISSIE D., CUNJAK R.A., VIJAYAN M.M., TUFTS B.L., 2002. The effects of environmental heat stress on heat-shock mRNA and protein expression in Miramichi Atlantic salmon (*Salmo salar*) parr. *Can. J. Fish. Aquat. Sci.*, 59, 1553-1562.
- MARCEAU P., CLUIS D., MORIN G., 1986. Comparaison des performances relatives à un modèle déterministe et à un modèle

- stochastique de température de l'eau en rivière. *Can. J. Civ. Eng.*, 13, 352-364.
- MOHSENI O., STEFAN H.G., 1999. Stream temperature/air temperature relationship: a physical interpretation. *Journal of Hydrology*, 218, 128-141.
- MORIN G., COUILLARD D., 1990. Predicting river temperatures with a hydrological model, In: *Encyclopedia of Fluid Mechanics, Surface and Groundwater Flow Phenomena*, Chapter 5, N.P. Cheremisinoff [ed.], Gulf Publishing Company, Houston, Texas, Vol. 10, pp. 171-209.
- MOSELHI O., HEGAZY T., FAZIO P., 1992. Potential applications of neural networks in construction. *Can. J. Civ. Eng.*, 19(3), 521-529.
- NEMEROW N.L., 1985. Stream, lake, estuary and ocean pollution. Van Nostrand Reinhold Company Inc., New York, 444 p.
- NNMODEL32 version 1.2 2.0. Copyright 1994-1998. Application Help. Neural Fusion Shareware <http://www.neuralfusion.com>
- PAO Y., 1989. Adaptive pattern recognition and neural networks. Addison-Wesley, Reading, Mass.
- RANDALL R.G., 1981. Production rate of juvenile Atlantic Salmon (*Salmo Salar L.*) in relation to available food in two Miramichi River, N.B., nursery streams. Ph.D. thesis. University of New Brunswick, Fredericton, N.-B.
- SONG C.C.S., CHIEN C.Y., 1977. Stochastic properties of daily temperature in rivers. *Journal of the Environmental Engineering Division*, Proceedings of The American Society Of Civil Engineers, Paper N 12858, 103 (EE2) p. 217-231.
- STEFAN H.G., PREUD'HOMME E.B., 1993. Stream temperature estimation from air temperature. *Water Resources Bulletin*, 29(1), 27-45.
- TURKKAN N, 1993. Génie, gènes et neurones. *Revue de l'Université de Moncton*, 26(1), 205-221.