

Université du Québec
Institut national de la recherche scientifique
INRS-ETE

**MODÉLISATION STATISTIQUE DE LA TEMPÉRATURE DE L'EAU
EN RIVIÈRE ET EN RÉGIME NON-HIVERNAL**

Par

Loubna Benyahya

Thèse présentée
pour l'obtention du grade de Philosophae Doctor (Ph.D.)
en Sciences de l'Eau

Jury d'évaluation

Examinateur externe	Zoubeida Bargaoui ENIT
Examinateur externe	Charles Lin Université McGill
Examinateur interne	Monsieur Karem Chokmani INRS-ETE
Directeur de recherche	Monsieur André St-Hilaire INRS-ETE
Codirecteur de recherche	Monsieur Taha Ouarda INRS-ETE
Codirecteur de recherche	Monsieur Bernard Bobée INRS-ETE

Février 2007

Je dédie cette thèse à mon très regretté père Monsieur Driss Benyahya

REMERCIEMENTS

J'exprime mes profonds remerciements à mon directeur de thèse, le Professeur André St-Hilaire, pour l'aide très compétente qu'il m'a apportée, pour sa patience et son encouragement tout au long de mes années de doctorat. Mes remerciements vont aussi à mes Codirecteurs de thèse, les Professeurs Bernard Bobée et Taha Ouarda, de m'avoir encadré et pour avoir relu avec rigueur tous mes écrits. Je tiens à remercier Monsieur Daniel Caissie, hydrologue au ministère de Pêches et Océans Canada, pour ses commentaires constructifs. Je tiens à souligner l'aide de Monsieur Jacques Dumas, hydrologue à l'Institut National de la Recherche Agronomique (INRA) de St-Pée (France) dans l'acquisition des données sur le terrain.

Mes collègues et amis de la Chaire d'Hydrologie Statistique, Behrooz Ahmadi Nedushan et Salaheddine El Adlouni m'ont aussi appuyé, je les remercie sincèrement.

Finalement, mes remerciements les plus sincères vont à ma mère et à mes frères, qui ont toujours su m'encourager, et je leur en suis reconnaissante. Mes remerciements vont aussi à tous mes amis.

RÉSUMÉ

Le processus de prévision de la température de l'eau en rivière est entamé par le développement et l'application des modèles appropriés permettant une bonne description du régime thermique. La modélisation peut donc être un outil de gestion important, qui permet de prédire la température de l'eau pour des conditions hydrométéorologiques variées et donc d'en évaluer l'impact sur la qualité et la quantité d'habitat aquatique. Cette thèse a donc pour but de proposer une nouvelle approche appropriée de modélisation statistique de la température de l'eau.

La revue de littérature a montré que les modèles périodiques n'ont pas été employés jusqu'à présent. Pourtant, ils offrent l'avantage de modéliser la variation périodique de la fonction d'autocorrélation, une caractéristique souvent présente dans les séries hydro-climatiques à petit pas de temps (hebdomadaires ou mensuelles). De plus, il ressort de la revue de littérature que les modèles non-paramétriques (réseaux de neurones artificiels (RNA) et k-voisins les plus proches (VPP)) sont très peu considérés, et donc ils demeurent de nouvelles pistes pour déterminer s'ils pourront apporter une amélioration en terme de performance par rapport aux modèles périodiques.

Une première étude a été effectuée afin de vérifier l'applicabilité du modèle périodique autorégressif (PAR) et d'en souligner l'intérêt en comparant sa performance avec celle d'une approche autorégressive courante (AR). Les données utilisées dans la modélisation numérique sont les températures hebdomadaires de la rivière Deschutes (Oregon, États-Unis) entre 1963-1980. La performance des deux modèles a été évaluée à l'aide de trois critères numériques: la racine de l'erreur quadratique moyenne, le biais absolu et le

coefficient de Nash. Ces critères ont été calculés pour chaque année en utilisant une technique de validation croisée de type "Jackknife". Les résultats préliminaires ont montré que le modèle PAR et le modèle autorégressif existant présentent une performance similaire dans la simulation des températures hebdomadaires. Toutefois, le modèle PAR est capable de fournir pour chaque période de l'année des valeurs de biais plus stationnaires que le modèle AR.

Pour tenter d'obtenir une modélisation plus réaliste et plus opérationnelle pour les gestionnaires des ressources hydriques, une deuxième étude a été effectuée en incorporant la température de l'air et le débit comme variables affectant le régime thermique. Deux nouveaux modèles de température de l'eau sont alors proposés et comparés, soit le modèle périodique autorégressif avec variables exogènes (PARX) et la méthode des k-voisins les plus proches (VPP). Les données de température hebdomadaire de l'eau utilisées (1984-2004), proviennent de la rivière Nivelle (Pyrénées atlantiques, France). Les résultats obtenus dans ce volet de la thèse montrent que les deux méthodes sont équivalentes et conduisent aux meilleurs résultats en terme de critères de performance. De plus, les modèles VPP4 (modèle avec quatre variables explicatives) et PARX3 (modèle avec trois variables explicatives) se démarquent de tous les autres modèles considérés.

En conclusion, en vertu de leur paramétrisation, les modèles périodiques précisément développés dans le cadre de cette thèse, demeurent les plus appropriés, car de plus ils préservent la persistance des températures hebdomadaires de l'eau, ils offrent une équation explicitant la relation entre la température de l'eau et les variables explicatives.

TABLE DES MATIÈRES

CHAPITRE 1. INTRODUCTION	1
1.1 Mise en contexte.....	3
1.2 Objectifs de la thèse.....	4
CHAPITRE 2. RÉSUMÉ DES TRAVAUX DE RECHERCHE	9
2.1 Revue de littérature sur la modélisation de la température de l'eau.....	11
2.1.1 Modélisation de la température de l'eau.....	12
2.1.2 Conclusion.....	18
2.2 Modélisation stochastique de la température de l'eau : Étude de cas de la rivière Deschutes (États Unis).....	19
2.2.1 Méthodologie.....	19
2.2.2 Région d'étude.....	21
2.2.3 Résultats.....	22
2.2.4 Conclusion	24
2.3 Modélisation préliminaire de la température de l'eau par des approches paramétrique et non-paramétrique: Étude de cas de la rivière Nivelle (France).....	26
2.3.1 Méthodologie.....	26
2.3.2 Région d'étude.....	28
2.3.3 Résultats.....	29
2.3.4 Conclusion	32
CHAPITRE 3. REVUE DE LITTÉRATURE	35
Abstract.....	39
Résumé	40
3.1 Introduction	41

3.1.1	Importance of water temperature for fish habitat	41
3.1.2	Spatial and temporal variability.....	42
3.1.3	Study objectives.....	43
3.2	Water temperature models.....	44
3.2.1	Deterministic models.....	44
3.2.2	Statistical/Stochastic models	46
3.3	Parametric statistical models	48
3.3.1	Regression models.....	48
3.3.2	Autoregressive models	51
3.3.3	Periodic Autoregressive models	53
3.4	Non-parametric statistical models	55
3.4.1	Artificial Neural Networks	56
3.4.2	k-Nearest Neighbours (k-NN)	57
3.5	Discussion and Conclusion.....	59
	References	63
	CHAPITRE 4. MODÉLISATION DE LA TEMPÉRATURE DE L'EAU : LES MODEÈLES AR et PAR	73
	Abstract.....	77
4.2	Methods	82
4.2.1	Autoregressive model (AR) on residuals.....	82
4.2.2	Periodic Autoregressive Model (PAR).....	83
4.2.3	Model evaluation and validation	85
4.3	Case study area and data set	86
4.4	Results	88

4.5 Discussion and conclusion.....	91
References	95
CHAPITRE 5. MODÉLISATION DE LA TEMPÉRATURE DE L'EAU : LES MODÈLES VPP et PARX	109
Abstract.....	113
Résumé	114
5.1 Introduction	115
5.2 Methods	119
5.2.1 k-Nearest Neighbours (k-NN)	119
5.2.2 Periodic Autoregressive model with exogenous variables (PARX).....	121
5.2.3 Model evaluation and validation	122
5.3 Case study area and data set	124
5.4 Results	125
5.4.1 k-NN model	125
5.4.2 PARX model.....	129
5.5 Discussion and conclusion.....	131
References	135
CHAPITRE 6. CONCLUSIONS DE LA THÈSE.....	151

LISTE DES TABLEAUX

Table 3.1	Advantages and disadvantages of statistical models based on approach and identification of related studies.....	72
Table 4.1	Regression coefficients of AR(1) and AR(2) models, estimated by using Leave-One-Out technique for the Deschutes river.	100
Table 4.2	Sample mean, standard deviation and autocorrelation coefficients of lag 1 and lag 2 of average weekly maximum temperature series for the Deschutes river.	101
Table 4.3	Parameter estimates for PAR(1) model ($\hat{\phi}_1$) of average weekly maximum water temperature series for the Deschutes river.	102
Table 4.4	Results of the cross validation of the periodic and non-periodic approaches expressed by the root mean square error, bias error and Nash coefficient and their range at Deschutes river	103
Table 5.1	List of attributes included in the preliminary analysis of k-NN method	141
Table 5.2	Performance measures (Root Mean Square Error, Bias, NSC) and their range of best-subset k-NN and PARX models	142
Table 5.3	Model parameters estimates for PARX3 of mean weekly maximum temperatures of the validated year 2004.....	143

LISTE DES FIGURES

Figure 3.1	Artificial Neural Network representation	71
Figure 4.1	Location of the water temperature measuring site on Deschutes river.....	104
Figure 4.2	Average Weekly Maximum Temperatures at the gauging station, from 1963 through 1980	105
Figure 4.3	Average weekly maximum temperatures (1963) and annual component estimated by a sinusoidal function. The solid line represents sinusoidal function, the dashed line represents average observed weekly maximum temperatures and the dashed dot line represents the residuals.	106
Figure 4.4	Results of the water temperature modelling using periodic and non- periodic approaches for some validation years. The solid lines represent observed temperatures, the points those predicted by PAR(1) and dashed lines those predicted by AR(1)	107
Figure 4.5	Results of the water temperature modelling using periodic and non- periodic approaches. The solid lines represent variation of root mean square error and the dashed line represent variation of bias error.....	108
Figure 5.1	Nivelle river catchement area with the indication of the water temperature and streamflow stations	144
Figure 5.2	Weekly water temperature data collected in the Nivelle catchement (1984-2004) (a) weekly water temperature data including the periods from December to April, (b) detailed view of weekly water temperature data (1994-1998). (b1) represents the amplitude of the water temperature variation from December to April. (b2) represents the amplitude of the water temperature variation from May to November.	145

Figure 5.3	Water and air temperature and streamflow data excluding the periods from December to April collected in the Nivelle catchement from 1984 to 2004	146
Figure 5.4	Scatter plot of the attributes on the PC1, PC2 plane	147
Figure 5.5	Boxplot showing the correlation between Tw0 and each attributes.....	148
Figure 5.6	Sample mean, standard deviation and autocorrelation coefficients of lag 1 and lag 2 of mean weekly maximum temperature series for the Nivelle river.....	149
Figure 5.7	Time series of observed and predicted water temperatures obtained by k-NN4 and PARX3 for some validation years, the solid lines represent the observations.....	150

CHAPITRE 1. INTRODUCTION

1.1 Mise en contexte

Le régime thermique des cours d'eau a été le sujet de nombreuses recherches au cours des dernières décennies en raison de ses intérêts écologiques et économiques. La température de l'eau est considérée comme l'une des variables physiques les plus déterminantes de la qualité de l'eau car elle intervient dans un certain nombre de processus chimiques, physiques et biologiques. Par exemple, la température de l'eau modifie la solubilité de l'oxygène. À mesure que la température augmente, la solubilité de ce gaz diminue. De plus, les eaux chaudes peuvent permettre le développement accru de bactéries, et d'autres organismes pathogènes qui peuvent rendre le milieu hostile pour le développement de nombreuses espèces aquatiques plus particulièrement les poissons poïkilothermes (i.e., la température du sang varie avec celle du milieu). Les salmonidés par exemple, sont des espèces dont les activités métaboliques et comportementales sont dépendantes de la température de l'eau. Au Canada et au Québec, la pêche de ces poissons est une ressource précieuse puisqu'elle procure des avantages économiques et sociaux considérables.

Le régime thermique d'un cours d'eau est un phénomène complexe mettant en jeu plusieurs facteurs, comme par exemple les variations météorologiques et hydrologiques. En plus de cette variabilité naturelle, les impacts anthropiques tels que la déforestation, les rejets thermiques, la diminution du niveau d'eau due à l'installation des réservoirs, peuvent modifier en profondeur ce régime.

Étant donné l'importance de la température dans la qualité et l'écologie d'un cours d'eau, et étant donné les impacts humains significatifs, il est nécessaire de développer des outils de gestion des ressources aquatiques. Pour cela, les modèles de simulation ou de prévision de

température s'avèrent des outils importants. Dans la littérature, les approches habituellement utilisées dans la modélisation du régime thermique se regroupent en deux grandes classes : l'approche statistique/stochastique et l'approche déterministe. En général, la comparaison des deux catégories de modèles permet de conclure que les modèles déterministes de la température de l'eau sont plus nombreux et nécessitent souvent un grand nombre d'intrants qui sont parfois difficilement mesurables. Contrastant à la modélisation déterministe, la modélisation statistique a été relativement peu utilisée pour modéliser le régime thermique jusqu'à présent. Pourtant, elle offre l'avantage de nécessiter peu de données et un temps de développement habituellement moins long que les modèles déterministes. Pour ces raisons, l'approche privilégiée dans cette thèse sera de type statistique.

1.2 Objectifs de la thèse

Les objectifs de cette thèse sont :

L'identification des approches statistiques les plus prometteuses qui pouvant être adaptées à la modélisation de la température de l'eau : la synthèse bibliographique permet de souligner les avantages et les limites des différents modèles statistiques afin d'en choisir ceux qui permettent de bien représenter les températures de l'eau à une échelle de temps spécifiée. Les outils proposés devront intégrer un nombre limité d'intrants tout en restant proche de la réalité physique du régime thermique;

La vérification de l'applicabilité d'une nouvelle approche statistique : il s'agit d'adapter un modèle périodique autorégressif (PAR) précédemment utilisé dans d'autres domaines (économétrie, prévision des crues) au cas spécifique du régime thermique de l'eau et d'en

souligner l'intérêt en comparant sa performance à celle d'une approche couramment utilisée. Cette dernière consiste à extraire la composante déterministe du signal par ajustement d'une fonction sinusoïdale et de modéliser les résidus par une approche autorégressive;

L'introduction des variables explicatives dans la modélisation de la température de l'eau : pour rendre la modélisation de la température de l'eau plus applicable à l'étude et l'analyse d'impacts anthropiques (e.g. modification du climat ou du régime hydrologique), il est nécessaire d'introduire des variables hydro-climatiques. Conséquemment, le second modèle, qui est l'élément principal de cette thèse, est fondé sur le développement d'un modèle périodique autorégressif avec variables exogènes (PARX). L'emphase est mise sur l'introduction de la température de l'air et du débit comme variables affectant la température de l'eau. Les températures de l'eau simulées par le modèle PARX seront comparées à celles obtenues par un modèle non-paramétrique, le modèle des k-voisins les plus proches (VPP).

En fonction de ces objectifs, la thèse est organisée de la manière suivante. Le second chapitre résume les travaux de recherche qui ont été effectués afin de répondre aux objectifs principaux de la thèse. Il s'agit d'un résumé rédigé en français des chapitres 3, 4 et 5; qui sont des articles rédigés en anglais. Afin d'alléger le contenu de la thèse, les équations, les figures, les tableaux et les références bibliographiques utilisés se trouvent dans les dits chapitres.

Le chapitre 3 présente une synthèse bibliographique des différents modèles de la température de l'eau existant, et situe les approches utilisées parmi ces modèles. En soulignant les avantages et les limites des différents modèles, ce chapitre aborde en premier

lieu les modèles déterministes basés sur les équations du budget thermique, puis en deuxième lieu les modèles statistiques paramétriques et non-paramétriques

Pour représenter la variation périodique de la fonction d'autocorrélation dans les séries chronologiques de la température de l'eau à pas de temps hebdomadaire, le chapitre 4 propose une modélisation adaptée, en l'occurrence le modèle Périodique Autorégressif (PAR). Ce modèle est développé et appliqué, et sa performance est évaluée par rapport à celle d'un modèle existant dont l'approche consiste à décomposer le signal saisonnier et la composante non saisonnière de la température de l'eau.

La température de l'air et le débit sont les variables ayant l'impact le plus important sur la température de l'eau et sont aussi les variables explicatives les plus souvent disponibles. L'incorporation de ces variables explicatives dans un modèle périodique avec variables exogènes (PARX), sera mise à profit dans le chapitre 5. Ce modèle est présenté dans ce chapitre, et le modèle des k-voisins les plus proches (VPP) est proposé, afin de tenter de comparer la performance de ces deux approches et donc de mieux simuler les températures hebdomadaires de l'eau.

Ces trois derniers chapitres sont exposés sous forme d'articles rédigés en anglais. L'article portant sur la revue de littérature des modèles statistiques de la température de l'eau a été soumis à la revue *Canadian Water Resources Journal*. L'article portant sur la modélisation de la température de l'eau en utilisant une approche stochastique : Étude de cas de la Rivière Deschutes (États-Unis), est sous presse à la revue *Journal of Environmental Engineering and Science*. L'article portant sur la modélisation préliminaire de la température de l'eau en utilisant des approches paramétrique et non-paramétrique: Étude de cas de la rivière Nivelle (France), a été soumis à la revue *Hydrological Sciences Journal*.

Au chapitre 6, la conclusion permet de résumer les faits saillants des résultats obtenus et d'identifier de nouvelles voies de recherche.

CHAPITRE 2. RÉSUMÉ DES TRAVAUX DE RECHERCHE

2.1 Revue de littérature sur la modélisation de la température de l'eau

La température de l'eau est une variable très importante pour définir la qualité et la quantité d'habitat aquatique. En effet, étant des poissons à sang froid, les salmonidés par exemple, sont des espèces qui peuvent être défavorablement affectées par une augmentation de la température de l'eau. Par exemple, Bjornn et Reiser (1991) ont démontré que la température létale de certaines populations des truites est comprise entre 23°C et 25°C. Hodgson et Quinn (2002) ont expliqué qu'une augmentation de la température de l'eau de l'ordre de 19°C pendant la saison de fraye peut amener des migrations précoces de saumon. Lund et al. (2002) ont montré que les températures extrêmes peuvent être problématiques pour les poissons en augmentant leur niveau de stress.

Les activités humaines telles que la déforestation, l'installation des réservoirs peuvent avoir un impact significatif sur le régime thermique des cours d'eau. Par exemple, Holtby (1988) a examiné l'impact de la coupe forestière sur la température de l'eau. Il a montré qu'une coupe à blanc couvrant une superficie équivalente à 41% du bassin versant s'est traduite par une augmentation de la température moyenne mensuelle de l'eau pendant tous les mois de l'année. Ces accroissements par rapport aux températures d'avant la coupe allaient de 0,7°C, en décembre, à 3,2°C en août. Un autre exemple d'impact anthropique sur le régime thermique des cours d'eau est celui des changements climatiques. Pour un scénario climatique reposant sur un doublement de la concentration en CO₂, Eaton et Sheller (1996) ont montré qu'aux États-Unis, la densité des espèces d'eau froide pourrait diminuer de 50%, et on pourrait assister à leur migration vers le nord à la recherche de conditions climatiques plus favorables.

La température de l'eau d'une rivière est caractérisée par des variations spatiales et temporelles. En plus des facteurs géographiques (e.g. latitude et altitude), les variations spatiales peuvent résulter de la conjonction de facteurs météorologiques (e.g, température de l'air, l'intensité et la durée de la radiation solaire, la vitesse du vent à la surface, etc) et hydrologiques (profondeur de la rivière, débit, etc). Par exemple, en Amérique du Nord, les plus grandes fluctuations temporelles se produisent en été tandis que les plus basses fluctuations se produisent en hiver.

Étant donné l'importance de la température dans l'habitat aquatique, et étant donné les impacts humains actuels et potentiels sur le régime thermique des rivières, il s'avère nécessaire de développer des outils de gestion des ressources hydriques permettant de faire la simulation ou la prévision de la température de l'eau. La revue de littérature permet de présenter les différents modèles de température de l'eau existants, et situe les approches utilisées parmi ces modèles. En soulignant les avantages et les inconvénients des modèles déterministes et statistiques, ce chapitre aborde en premier lieu les modèles basés sur des lois physiques régissant la température de l'eau en milieu naturel, puis en deuxième lieu les modèles inférant à partir des données, la relation statistique entre la température de l'eau et les variables environnementales.

2.1.1 Modélisation de la température de l'eau

De nombreux modèles existent pour simuler la température de l'eau. Ils se regroupent essentiellement en deux grandes catégories: les modèles déterministes et les modèles statistiques. Bien que les modèles déterministes ne soient pas l'approche privilégiée dans ce travail, il est important de présenter les avantages et les inconvénients inhérents à cette approche.

Les modèles déterministes s'appuient sur des équations issues des connaissances thermodynamiques (e.g. conservation d'énergie; équation (3.1)). Ils requièrent souvent un nombre élevé de variables telles que la radiation solaire, la vitesse du vent, le débit, la pression de vapeur, etc. Parmi les modèles déterministes existants, on trouve par exemple le modèle hydrologique et de température de l'eau CEQEAU (Morin et Couillard, 1990; St-Hilaire et al. 2003) et le modèle SSTEMP (Bartholow, 1999). Toutefois, en raison de la multitude de données requises qui sont rarement disponibles sur les bassins versants, il est parfois préférable d'élaborer des modèles de température plus simples.

Alternativement, les modèles statistiques visent à identifier une relation mathématique entre les variables d'entrée (e.g. température de l'air, débit) et la température de l'eau sans nécessairement tenir compte explicitement des lois physiques. Bien qu'ils exigent de longues séries de données, les modèles statistiques ont l'avantage de nécessiter peu de données et un effort de développement souvent moins important. Plusieurs travaux de modélisation statistique de la température de l'eau ont été réalisés. Généralement, deux familles de modèles sont incluses dans la catégorie statistique :

1) Les modèles paramétriques

Ce sont des modèles prédéfinis qui nécessitent l'ajustement de paramètres dont les valeurs sont déterminées par la calibration desdits modèles. Cette approche statistique peut regrouper des méthodes telles la régression linéaire, la régression non-linéaire, les modèles autorégressifs (ARMA) et les modèles périodiques autorégressifs (e.g, PAR, PARMA) etc. La régression linéaire entre les températures de l'eau et de l'air (équation (3.3)) a fait l'objet de quelques études [Stefan et Preud'homme, 1993; Pilgrim et al., 1998; Erickson et Stefan, 2000]. Pilgrim et al. (1998) ont montré que d'une part la valeur de la pente de

l'équation de régression augmente en fonction de l'échelle de temps (quotidienne, hebdomadaire et mensuelle) et que d'autre part une amélioration de la performance du modèle pourrait être accomplie si on considérait une échelle mensuelle ($R^2=0,89$) plutôt qu'une échelle journalière ($R^2=0,70$). Ahmadi-Nedushan et al. (2007) ont modélisé les températures journalières de l'eau de la rivière Moisie (Québec) en utilisant une régression linéaire multiple. Afin de réduire l'impact de la colinéarité [i.e. corrélation entre les variables explicatives], ces auteurs ont utilisé la régression "Ridge" (équation (3.6)). L'utilisation de cette technique a été justifiée par le fait qu'elle provoque un changement dans les valeurs numériques des coefficients du modèle, dont les prévisions deviennent légèrement biaisées, mais plus stables, et donc plus crédibles. Les résultats ont montré que ce modèle donne des valeurs de la racine carré de l'erreur quadratique moyenne (REQM) inférieures à $0,65\text{ }^{\circ}\text{C}$. L'avantage de la régression linéaire est sa simplicité. Toutefois, pour des températures extrêmes de l'air (maximum et minimum), la relation linéaire n'est pas toujours adéquate. En effet, pour représenter un tel comportement de la température de l'eau, Mohseni et al. (1998) ont ajusté une fonction non-linéaire représentant une forme sigmoïde (équation (3.7)).

En utilisant une approche stochastique autorégressive, de nombreuses études, telles que celles de Kothandaraman (1971), Cluis (1972) et Caissie et al. (1998) ont séparé les températures de l'eau en une composante saisonnière et une composante résiduelle. La composante saisonnière représente les fluctuations qui se répètent plus ou moins régulièrement d'année en année. La composante résiduelle regroupe toutes les autres fluctuations non prises en compte dans la composante précédente. Pour représenter la composante annuelle, Cluis (1972) utilise une fonction sinusoïdale (équation (3.8)) pour

approximer les données de la température de l'eau. Pour représenter la série résiduelle à court terme, cet auteur a opté pour une chaîne de Markov d'ordre 2. Cette approche de décomposition des séries chronologiques a été appliquée par Caissie et al. (1998) et a fourni des résultats satisfaisants avec des valeurs de REQM de l'ordre de 0,9°C.

La transformation préliminaire par décomposition des séries chronologiques a pour but d'éliminer les variations saisonnières avant de tenter d'estimer un modèle stationnaire. Or, le modèle stationnaire ne tient pas compte de la variation saisonnière de la fonction d'autocorrelation périodique qui ne peut pas être extraite par simple décomposition des données. Une telle variation est souvent présente dans les séries hydro-climatiques à petit pas de temps (hebdomadaires ou mensuelles). D'un point de vue pratique, cette propriété d'autocorrélation favorise les modèles périodiques autorégressifs tels que les modèles PAR et PARMA. Ces derniers sont une extension du modèle ARMA (Box et Jenkins, 1970) permettant de modéliser une structure d'autocorrelation qui varie en fonction de la période de l'année. Ces modèles sont donc bien adaptés à la simulation des variables à une échelle inférieure à une année. Tout comme les modèles de régression, présentés précédemment, les modèles périodiques autorégressifs sont des modèles linéaires. Ces modèles ont fait l'objet d'application dans différentes disciplines scientifiques : l'économétrie (Osborn et Smith, 1989; Novales et de Frutto, 1997), la climatologie (Lund et al. 1995) et l'hydrologie (Salas et al. 1980; Vecchia 1985; Bartolini et al. 1988; Ouarda et al. 1997).

2) Les modèles non-paramétriques

Les modèles non-paramétriques ne supposent pas avoir une structure statistique prédéfinie, mais leurs paramètres caractéristiques sont plutôt inférés à partir des données disponibles sans hypothèses sur la distribution de la population. Ce sont des outils efficaces lorsque les

systèmes à modéliser sont non-linéaires. Il existe plusieurs modèles non-paramétriques, entre autres les réseaux de neurones artificiels (RNA) et les k-voisins les plus proches (VPP).

La méthode des RNA est un modèle dont la conception est schématiquement inspirée des neurones dans le cerveau. Les neurones sont repartis en couches successives, le calcul est effectué des entrées vers la sortie (Figure 3.1). Les entrées correspondent aux variables explicatives et la sortie correspond à la variable dépendante. Parmi les RNA, on distingue les réseaux à rétro-propagation. Ces derniers utilisent des algorithmes d'entraînement des poids. Chacune des entrées est pondérée et un poids total est ainsi calculé. En hydrologie, les RNA sont de plus en plus utilisés afin de simuler le débit et les paramètres de la qualité des eaux des rivières (e.g, Hsu et al. 1998; Conrads et Roehl, 1999). En modélisation de la température de l'eau, cette approche commence à être considérée. Bélanger et al. (2005) ont modélisé la température de l'eau du ruisseau Catamaran (Nouveau-Brunswick, Canada) en comparant les RNA et la régression linéaire multiple avec la température de l'air et le débit comme variables exogènes. Les résultats ont montré que les deux approches de modélisation sont performantes pour la prévision de la température de l'eau en rivière. Aussi, les RNA ont un énorme potentiel pour la modélisation des variables hydrologiques (Coulibaly et al. 2001).

La seconde méthode non-paramétrique est celle des k-voisins les plus proches. Comme les RNA, les VPP est une méthode qui ne fait aucune hypothèse quand aux relations analytiques entre les variables d'entrées et la variable de sortie. Cette méthode permet de déterminer parmi les situations passées celles qui ressemblent le plus à la situation

présente. Pour pouvoir obtenir une estimation ponctuelle de la sortie, les étapes sont les suivantes :

Définir les variables explicatives (attributs) qui caractérisent la variable dépendante;

Projeter ces attributs (de nombre de n) sous forme vectorielle X_i ($i = 1 \dots n$). En raison des unités de mesure différentes, il est préalablement nécessaire de standardiser les données (i.e. soustraire la moyenne et diviser par l'écart type);

Sélectionner les k voisins les plus proches de X_{i,j_1} (pour le jour j_1) au sens d'une distance euclidienne entre les normes pondérées N_{j_1} et N_{j_2} ($j_2 = 1 \dots, j_1 - 1, j_1 + 1, \dots, m$), telle que représentée par l'équation (3.12);

Chercher dans tout l'historique de X_i , les k voisins les plus proches correspondant à la situation actuelle X_{i,j_1} ;

Effectuer la somme pondérée des k voisins les plus proches.

Bien que la méthode des VPP soit une méthode facile à implanter, le temps de calcul associé peut s'avérer considérable si l'historique est important. On peut alors restreindre le voisinage. De plus, la performance de la méthode dépend du choix du nombre de voisins et du poids associé à chaque attribut, dont les valeurs doivent être ajustés par essai et erreur pour obtenir des valeurs simulées les plus rapprochées des valeurs observées.

En hydrologie, la méthode VPP a été utilisée dans le domaine de la prévision des débits journaliers à partir des précipitations (Karlsson et Yakowitz, 1987). Galeati (1990) a montré que les prévisions issues de la méthode des VPP sont comparables à celles produites par un modèle autorégressif avec variables exogènes (ARX).

L'avantage des modèles non-paramétriques est leur souplesse relative et le fait qu'il n'y ait pas de paramètres à estimer. Cependant, leur inconvénient est qu'il est difficile d'avoir une fonction mathématique conventionnelle, et donc il sera difficile de les extrapoler. Ce problème devient évident quand le but de la modélisation est d'établir une prévision.

2.1.2 Conclusion

En soulignant les avantages et les inconvénients des modèles déterministes et statistiques, la revue de littérature a révélé qu'il n'y a pas de prédominance des uns par rapport aux autres. C'est donc à l'utilisateur de choisir, selon ses objectifs et les données disponibles, les modèles les mieux adaptés à ses besoins. Aussi, il apparaît que jusqu'à présent, il existe relativement peu de travaux réalisés sur la modélisation statistique de la température de l'eau. Pourtant, elle offre l'avantage de nécessiter relativement peu de données et un coût et un temps de développement habituellement plus bas que les modèles déterministes. Ces raisons font de l'approche statistique, l'approche privilégiée dans cette thèse. Les modèles statistiques existants sont tous basés sur l'hypothèse de stationnarité des statistiques des séries chronologiques de la température de l'eau. Or, une caractéristique importante et souvent observée des séries hydro-climatiques à petit pas de temps (hebdomadaires ou mensuelles) est la présence de variations saisonnière, plus particulièrement introduites par la variation périodique de leur fonction d'autocorrelation. Ceci motive l'introduction des modèles périodiques (e.g PAR, PARMA, PARX). Ces modèles ont été largement proposés pour des applications économétriques et hydrologiques mais ils n'ont pas été employés jusqu'à présent pour la modélisation de la température de l'eau.

Malgré que les modèles non-paramétriques (e.g, RNA, VPP) sont intuitifs et offrent une description peu claire de la relation entre les données d'entrées et de sortie (boîte noire), ils demeurent de nouvelles pistes pour déterminer s'ils pourront apporter une amélioration en terme de performance par rapport aux modèles précédents.

2.2 Modélisation stochastique de la température de l'eau : Étude de cas de la rivière Deschutes (États Unis)

La revue de littérature a permis de souligner l'importance de la température de l'eau pour la vie aquatique en rivière, ainsi que l'éventuelle application de nouvelles approches de modélisation du régime thermique. En utilisant les températures hebdomadaires de la rivière Deschutes (États-Unis), cette section vérifie l'applicabilité de l'approche périodique autorégressif (PAR) au cas spécifique du régime thermique de l'eau et d'en souligner l'intérêt en comparant sa performance à celle de l'approche décrite par Caissie et al. (1998).

2.2.1 Méthodologie

Tel que décrit dans la revue de littérature, l'approche présentée par Caissie et al. (1998) consiste à décomposer la série chronologique de la température de l'eau en une composante annuelle et une composante résiduelle (équation (4.1)). Pour représenter la composante annuelle, le modèle utilise une fonction sinus (équation (4.2)). Les coefficients sont calculés en minimisant la somme des carrés des erreurs entre les observations et les valeurs estimées par cette équation. Une fois que la variation saisonnière est soustraite, la série de résidus est décrite par une chaîne de Markov d'ordre 1 et 2 (équation (4.3)). AR (1) et AR

(2) sont les modèles autorégressifs dont les résidus sont représentés respectivement par une chaîne de Markov d'ordre 1 et 2.

Donc, le modèle AR est une représentation simple de la structure réelle des séries périodiquement corrélées, telles que la température de l'eau. Afin de simuler les températures de l'eau réalistes, il est important que le modèle préserve bien les caractéristiques de l'autocorrélation périodique. Une telle caractéristique peut être représentée par un modèle PAR (Salas, 1993) d'ordre p centré aux données corrigées par leur moyenne empirique μ (équation (4.5)). Il faut souligner que l'application du modèle PAR à un processus périodique, requiert une représentation matricielle (équation (4.4)) dont la ligne signifie l'évolution de l'année ($v = 1, \dots, n$) et la colonne signifie l'évolution de la période ($\tau = 1, \dots, \omega$). La constante p représente le nombre de termes autorégressifs, $\phi_{i,\tau}$ sont les paramètres autorégressifs et $\varepsilon_{v,\tau}$ est un bruit blanc. L'utilisation opérationnelle d'un modèle périodique pour lequel le paramètre p varie en fonction de la période, est difficile. On se limite donc, dans la présente étude à des modèles à faibles ordres, par exemple PAR (1) et PAR (2). Cette limitation est faite par souci de parcimonie. Concernant l'estimation des paramètres, une des méthodes les plus adéquates est la méthode des moindres carrés. Cette méthode consiste à trouver les paramètres qui minimisent la somme des carrés des erreurs (équation (4.6)).

Pour évaluer la performance des deux modèles (PAR et AR), trois indices ont été calculés : la racine de l'erreur quadratique moyenne (REQM) (équation (4.7)), le biais (équation (4.8)) et le coefficient de Nash (NSC) (équation (4.9)). Dans le cadre de cette thèse, on a

utilisé le logiciel MATLAB pour le développement des algorithmes et l'analyse des résultats (version 6.5 Copyright 1984-2002).

Les séries chronologiques de la température de l'eau existent mais elles sont moins nombreuses et souvent courtes. Afin de conserver le maximum d'information possible, la validation des deux modèles est réalisée en utilisant la procédure de "Leave-One-Out"(validation croisée). Cette technique consiste à supprimer d'une manière séquentielle une année et estimer les paramètres des modèles en utilisant les données des années restantes. Par la suite, le modèle calé sur les années restantes est utilisé pour simuler la température de l'année enlevée au départ et la comparaison est faite entre les températures observées et calculées par le modèle.

2.2.2 Région d'étude

La région d'étude est la rivière Deschutes située au centre nord de l'état d'Oregon (États-Unis) (Figure 4.1). La rivière prend sa source dans la chaîne des Cascades. La superficie de son bassin de drainage est 26 860 km². La rivière Deschutes suit son cours sur 300 km de longueur. Le débit moyen annuel varie entre 124 m³/s et 213 m³/s. La rivière Deschutes présente un intérêt écologique, elle est fréquentée par le saumon du Pacifique (*Oncorhynchus tshawytscha*) et la truite arc-en-ciel (*Oncorhynchus mykiss*). Elle se situe parmi les plus importantes rivières à saumon de l'état de l'Oregon. Ceci fait de la pêche une importante activité aquatique qui caractérise le bassin de la rivière Deschutes.

Les données de température de l'eau (°C) utilisées dans la présente étude, sont extraites de la banque de données de USGS (*U.S. Geological Survey*). La figure 4.1 montre l'emplacement de la station de mesure de la température de l'eau sur la rivière Deschutes.

Les données utilisées dans la modélisation numérique sont des températures hebdomadaires calculées à partir des températures maximales prises chaque jour (moyenne sur sept jours consécutifs) et couvrant la période de 1963 à 1980 (Figure 4.2). Le choix de pas de temps hebdomadaire est basé sur deux critères, soit le fait que la température hebdomadaire d'un cours d'eau soit un bon indicateur d'habitat de poissons (Eaton et Sheller, 1996), et le fait que les corrélations entre la température de l'eau et la température de l'air sont les plus fortes sur une échelle de temps hebdomadaire ou mensuelle que sur un pas de temps plus court ou plus long (Erickson et Stefan, 2000). Les analyses sont effectuées durant la période allant du mois de mai au mois de novembre (30 semaines), soit les mois où les poissons à sang froid sont plus sensibles aux fluctuations de la température de l'eau.

2.2.3 Résultats

Une comparaison entre les modèles AR et PAR est effectuée et les résultats des REQM, du biais et du NTD pour chaque année et pour chaque modèle, sont rapportés dans le tableau 4.4. Les REQM moyens obtenus par le modèle AR sont de l'ordre de 0,81°C et 0,82°C pour AR (1) et AR (2), respectivement. Pour chaque année de simulation, le modèle AR(1) produit des REQM qui varient de 0,62°C et 1,19°C tandis que le modèle AR(2) présente des valeurs allant de 0,62°C à 1,11°C. Le biais obtenu par le modèle AR sur l'ensemble des années est nul mais pour chaque année, il varie entre -0,07°C et -0,09°C pour AR(1) et de -0,06°C à 0,07°C pour AR(2). Les moyennes annuelles des coefficients de Nash révèlent une valeur de 0,94 aussi bien pour AR (1) que pour AR (2).

Dans le but de démontrer l'utilité du développement d'un modèle périodique (PAR), nous avons examiné la variation périodique de la moyenne, de l'écart type et de l'autocorrelation d'ordre 1 et 2 de la série des températures hebdomadaires (Tableau 4.2). L'autocorrelation

périodique du délai 1 est le lien qui existe entre les données de la semaine τ et celles de la semaine $(\tau-1)$ pour chaque semaine des mois analysés. On remarque un aspect périodique marqué de ces statistiques (Tableau 4.2), ce qui signifie qu'un modèle périodique pourrait servir à simuler la température hebdomadaire d'une manière efficace.

Comparativement au modèle précédent, le modèle PAR est associé à des valeurs de REQM légèrement plus élevées avec une différence maximale de $0,11^{\circ}\text{C}$ entre les modèles PAR et AR. Pour chaque année, les REQM varient entre $0,54^{\circ}\text{C}$ et $1,27^{\circ}\text{C}$ pour PAR (1) et entre $0,52^{\circ}\text{C}$ et $1,30^{\circ}\text{C}$ pour PAR (2). Comme le modèle AR, le biais moyen est nul mais pour chaque année les variations sont entre $-0,96^{\circ}\text{C}$ et $0,65^{\circ}\text{C}$ pour PAR (1) et entre $-0,91^{\circ}\text{C}$ à $0,64^{\circ}\text{C}$ pour PAR (2). On peut aussi constater du tableau 4.4 que le modèle PAR sous-estime les températures hebdomadaires pour les années 1977, 1978, 1979 et 1980. Les coefficients de Nash du modèle PAR ($\text{NSC} = 0,92$) sont légèrement inférieurs à ceux donnés par le modèle AR ($\text{NSC} = 0,94$).

D'une façon générale, selon les trois critères d'évaluation étudiés (REQM, biais et NSC), il semble que le modèle PAR produit une performance semblable (bien que légèrement plus faible) à celle du modèle AR. En effet, une comparaison graphique entre les températures observées et calculées (Figure 4.4) effectuée sur quelques années indique que les deux modèles sont capables de reproduire le cycle saisonnier des températures hebdomadaires et permettent de calculer des valeurs qui se rapprochent des valeurs observées. De plus, les températures hebdomadaires calculées et observées montrent que les deux modèles ne produisent pas d'erreur systématique. En effet, les valeurs calculées ne sont pas systématiquement supérieures ou inférieures aux températures observées.

L'analyse du tableau 4.4 et de la figure 4.4 ne permet pas une conclusion définitive sur le choix d'un modèle par rapport à l'autre. Cependant, en calculant les critères de la REQM et du biais pour chaque période (Figure 4.5), on remarque que le modèle PAR conduit à des valeurs de biais nuls et stationnaires tandis qu'ils sont plus hétérogènes lorsque le modèle AR est utilisé.

2.2.4 Conclusion

L'objectif de cette section est d'estimer la température de l'eau en utilisant une nouvelle approche et la comparer à une approche déjà existante, soit le modèle PAR et le modèle AR. Ceci permettra ensuite de prédire la température de l'eau pour des fins de gestion des ressources hydriques. Chacun des deux modèles utilisés a ses avantages et ses inconvénients. Le modèle AR tel que décrit par Caissie et al. (1998) est efficace, mais il nécessite l'identification de deux composantes: l'une annuelle à long terme, l'autre résiduelle à court terme. La modélisation par cette approche peut parfois s'avérer limitée en raison de l'hypothèse de stationnarité des résidus qu'elle implique. Une alternative à l'utilisation de cette approche simple est de considérer des séries brutes plutôt que des séries préalablement décomposées. De ce fait, le modèle PAR est approprié pour la modélisation des températures hebdomadaires de l'eau car il permet de représenter une structure d'autocorrélation qui varie selon les périodes, une caractéristique souvent observée en pratique.

Dans l'ensemble (1963-1980), les deux modèles AR et PAR donnent des résultats satisfaisants, tels que démontrés par la racine des erreurs quadratiques moyennes (REQM modèle AR(1)= 0,81°C et REQM modèle PAR(1)= 0,90°C). Le biais interannuel obtenu pour ces deux modèles est nul mais variant peu d'une année à l'autre. Les coefficients de

Nash sont semblables, c'est-à-dire de l'ordre de 0,92 et 0,94, respectivement. De plus, il est intéressant de noter que même si les modèles AR et PAR présentent une performance similaire, les critères de biais et de REQM calculés pour chaque période distinguent ces deux modèles. Plus précisément, on constate que le modèle PAR donne des valeurs de biais nuls avec une variance beaucoup moins élevée que celle donnée par le modèle AR.

Finalement, dans la présente étude, il a été démontré que dans un grand bassin versant et en utilisant un seul site de mesure de température de l'eau, le modèle PAR est tout aussi performant que le modèle AR dans la simulation des températures hebdomadaires de la rivière Deschutes. Cependant, en vertu de la paramétrisation, la performance du modèle PAR est jugée satisfaisante en raison de la préservation de la persistance des séries périodiques et de la stationnarité périodique des biais qu'il offre. Étant donné qu'aucun modèle statistique de type périodique n'a été appliqué aux températures hebdomadaires de l'eau, il est difficile de comparer la performance du modèle PAR à celle d'autres outils semblables.

Bien que le modèle PAR a certaines potentialités, il reste limité de point de vue pratique puisque aucune variables physiques (e.g, température de l'air, débit) n'a été introduite. Son application est donc limitée au cas où les caractéristiques climatiques et hydrologiques d'un cours d'eau demeurent constantes. Il serait important de développer un modèle périodique avec variables exogènes (PARX). Ce modèle permettrait aux gestionnaires des ressources hydriques de prédire et d'estimer la température de l'eau pour des conditions hydrométéorologiques variées.

2.3 Modélisation préliminaire de la température de l'eau par des approches paramétrique et non-paramétrique: Étude de cas de la rivière Nivelle (France)

À la lumière de la revue de littérature, plusieurs études ont démontré que les variations de températures de l'eau sont significativement reliées à celles de l'air (e.g, Stefan et Preud'homme, 1993; Mohseni et Stefan, 1999) et/ou au débit (e.g, Webb et al. 2003; Neumann et al. 2003) lesquelles sont souvent disponibles.

Cette étude présente la modélisation de la température de l'eau en utilisant la méthode des k-voisins les plus proches (VPP) et le modèle PAR avec variables exogènes (PARX) pour simuler les températures hebdomadaires de la rivière Nivelle (France).

2.3.1 Méthodologie

Tel que décrit dans la revue de littérature, la méthode VPP est une méthode non paramétrique, elle ne fait aucune hypothèse quand aux relations entre les variables d'entrées et de sortie. À partir d'un ensemble d'entrées, cette méthode consiste selon une distance euclidienne, à chercher dans tout l'historique, les k voisins les plus proches qui ressemblent à la température actuelle. La procédure proposée pour l'utilisation de la méthode des VPP est précédemment décrite dans la section 2.1.2. Une série d'attributs candidats ont été définis (tableau 5.1) :

Pour la température de l'eau :

Valeur de la semaine précédente ($Tw1$)

Valeur de l'avant-dernière semaine ($Tw2$)

Valeur de la moyenne des deux semaines précédentes ($Tw3$)

Pour la température de l'air :

Valeur de la semaine précédente ($Ta1$)

Valeur de l'avant-dernière semaine ($Ta2$)

Valeur de la moyenne des deux semaines précédentes ($Ta3$)

Valeur du degré-jour ($Ta4$)

Pour le débit :

Valeur de la semaine précédente ($Q1$)

Valeur de l'avant-dernière semaine ($Q2$)

Valeur de la moyenne des deux semaines précédentes ($Q3$)

Valeur du débit relatif ($Q4$)

Nous disposons donc d'un total de 11 attributs. Il n'est possible de considérer simultanément tous ces attributs et d'estimer leurs poids optimums. Dans cette étude, seulement cinq attributs jugés significatifs, sont considérés (cf. section 2.3.3).

Motivés par les résultats précédemment obtenus par le modèle PAR, nous avons décidé de réutiliser ce modèle en incluant la température de l'air et le débit comme variables exogènes au modèle. Comme le modèle PAR, le modèle PARX (équation 5.3) permet de modéliser la variation saisonnière de la fonction d'autocorrélation périodique, par conséquent, il est approprié pour la simulation des variables périodiques. Concernant, l'identification des modèles, on préfère une approche plus pragmatique qui consiste à choisir différents nombre du termes autorégressifs (p_1, p_2, p_3) des intrants, d'estimer les

paramètres du modèle ($\phi_{1,\tau}, \phi_{2,\tau}, \phi_{3,\tau}$), de simuler une série de température de l'eau et finalement d'examiner les résultats en fonction des critères de performance introduits dans la section 2.2.1. Le modèle qui répond mieux aux critères de performance est retenu.

2.3.2 Région d'étude

La région d'étude est la rivière Nivelle située au sud-ouest de la France (Figure 5.1). La superficie de son bassin de drainage est de 238 km². Elle prend sa source en Espagne et se déverse dans la baie de Saint Jean de Luz (France). La rivière Nivelle mesure 39 km de longueur et s'écoule avec un débit moyen annuel de 5.4 m³/s en aval de la confluence de son principal affluent et de 9 m³/s à son embouchure. Nivelle est le site d'une étude multidisciplinaire hydro-biologique visant à quantifier l'écosystème et évaluer l'impact des facteurs climatiques dans la région. Le saumon atlantique compte parmi les espèces les plus représentatives de Nivelle (Dumas et Haury, 1995; Dumas et Prouzet, 2003).

Les données de température de l'eau utilisées dans la présente étude proviennent de la station Ibarron (Site des laboratoires INRA à St Pée sur Nivelle (Pyrénées-Atlantiques)).

La figure 5.1 montre l'emplacement de la station de mesure de la température de l'eau sur la rivière Nivelle. Les données de température de l'air ont été enregistrées à partir de la station météorologique de l'aéroport de Biarritz-Parme opérées par Méteo-France à 13 km au nord d'Ibarron. Pour mesurer le débit moyen journalier une station de mesure gérée par la Direction Régionale de l'Environnement de l'Aquitaine (Cherchebruit) a été installée à 6 km en amont du site de mesure de la température de l'eau. L'analyse statistique de la température de l'air (1984-2004) montre que les mois les plus chauds de l'année, sont les mois entre juin et septembre avec une moyenne maximale de 19,96°C pour le mois de

juillet. En ce qui a trait au débit, des valeurs élevées ont été observées durant le printemps, ce qui résulte des précipitations fréquentes à cette saison (climat océanique doux et humide; Dumas et Haury, 1995).

Les données utilisées dans la modélisation (VPP et PARX) de la température de l'eau sont : les températures hebdomadaires de l'eau ($^{\circ}\text{C}$), les températures hebdomadaires de l'air ($^{\circ}\text{C}$) et les valeurs hebdomadaires du débit (m^3/s) (Figure 5.3). Ces données sont calculées à partir des températures maximales journalières (moyenne de 7 observations par semaine) et couvrant les années entre 1984 et 2004. Pour le choix de la période d'analyse et de validation, on se limite aux mois entre mai et novembre (30 semaines), soit les mois où les variations de la température de l'eau sont les plus significatives (entre 10°C et 12°C) (Figure 5.2.b2).

2.3.3 Résultats

Le modèle de VPP n'est pas préalablement déterminé. Pour cette application, on dispose d'un total de 11 attributs à tester (tableau 5.1). Comme il est difficile de considérer tous ces attributs simultanément, l'analyse en composante principale (ACP) a été utilisée afin d'examiner la corrélation entre les différents attributs et de vérifier ceux qui ont le même comportement (redondants), et par conséquent d'en choisir les plus utiles. La figure 5.4 est une représentation dans l'espace des deux premiers facteurs de l'ACP. On remarque dans cette figure que les attributs ($Tw0$, $Tw1$, $Tw2$, $Tw3$, $Ta1$, $Ta2$ et $Ta3$) sont regroupés ensemble et, par conséquent ils sont corrélés. De même pour les attributs ($Q1$, $Q2$ et $Q3$) lesquels sont fortement corrélés entre eux mais ils le sont inversement par rapport au premier groupe. De plus, l'attribut $Ta4$ (degrés-jours) se démarque de tous les autres attributs. Ceci a été confirmé par les diagrammes en boîte des corrélations entre l'attribut

$Tw0$ et tous les autres attributs (Figure 5.5). On remarque dans ce graphique que l'attribut $Tw0$ est fortement relié aux attributs $Tw1$, $Tw2$, $Ta1$ et $Ta4$ et inversement relié à $Q1$.

La méthode des VPP est appliquée en premier lieu en considérant un seul attribut à la fois (VPP1), il n'y a donc ainsi pas de poids à estimer. Le tableau 5.2 présente les valeurs des critères de performance (REQM, Biais, NSC) obtenus par la méthode de validation "Leave-One-Out" pour chacun des attributs sélectionnés par l'ACP. Une première analyse de ce tableau montre que les performances du modèle VPP1 données par chaque attribut fluctuent au sens de la racine de l'erreur quadratique moyenne ($2,52^{\circ}\text{C} < \text{REQM} < 4,02^{\circ}\text{C}$). Le biais obtenu par ce modèle et sur l'ensemble des années est élevé et varie de $-0,59^{\circ}\text{C}$ à $-0,53^{\circ}\text{C}$. Les moyennes annuelles des coefficients de Nash révèlent des valeurs entre $-0,83$ et $0,27$. D'après ces résultats, il ne semble pas y avoir un attribut qui contribue à une bonne performance. En générale, même s'il ne conduit pas à une valeur optimale des différents critères de performance, l'attribut $Tw1$ se démarque de tous les autres attributs considérés et conduit à des résultats relativement satisfaisants(REQM : $2,52^{\circ}\text{C}$; biais : $-0,54^{\circ}\text{C}$; NSC : $0,27$)

La méthode des plus proches voisins a été ensuite appliquée à plusieurs attributs (on considère les cas de 2, 3, 4 et 5 attributs dans la même norme). Les poids des différents attributs ont été estimés en minimisant les écarts entre les valeurs simulées par les différents modèles et les valeurs observées. Il est à noter que la procédure des VPP est effectuée à partir des données standardisées, ce qui permet de faire la somme des différents attributs sans problème d'unités. L'analyse des résultats montre que l'utilisation de plusieurs attributs améliore la performance de la méthode VPP. Ainsi dans les cas étudiés, il semble indiquer qu'il est préférable de considérer le modèle VPP4 ayant comme attributs

Tw1, Tw2, Ta1 et Q1. Le modèle VPP4 présente une REQM de l'ordre de 1,20°C avec des valeurs annuelles qui varient de 0,80°C à 1,50°C. Le biais moyen interannuel est nul, mais variant chaque année de -0,19°C à 0,19°C. La valeur moyenne interannuelle du coefficient de Nash est de 0,84.

Dans le but de comparer la méthode des voisins les plus proches avec une méthode paramétrique, le modèle PARX est proposé. Il fut déjà mentionné qu'en plus des paramètres autorégressifs du modèle original (PAR), le modèle PARX (équation 5.3) requiert l'ajustement des paramètres supplémentaires, soit ceux de la température de l'air et du débit. Dans un premier temps, le modèle PARX a été testé sur 18 variables explicatives dont seules 11 ont été retenues (tableau 5.1). Ceci est basé sur un processus d'optimisation qui cherche à retenir toutes les variables dont les valeurs simulées par le modèle PARX se rapprochent des températures observées.

Dans l'ensemble (1984-2004), le modèle PARX offre une performance presque similaire à celle du modèle VPP, tels que démontrés par les résultats des critères de performance (Tableau 5.2). Par exemple, pour le modèle PARX3 (modèle avec trois attributs), la moyenne de la racine des erreurs quadratiques moyennes est légèrement supérieure comparativement au modèle VPP4 (REQM=1,58°C; comparativement à 1,20°C pour VPP4). Le biais moyen est nul, ce qui est semblable à la valeur obtenue à partir du modèle VPP4. Cependant, le modèle PARX3 présente un biais variant chaque année de -0,99°C à 1,50°C. Les valeurs des coefficients de Nash du modèle PARX3 sont légèrement inférieures (NSC=0,70 comparativement au NSC du modèle VPP4 de 0,84).

L'analyse du tableau 5.2 démontre que les modèles PARX3 et VPP4 conduisent à des résultats satisfaisants en utilisant respectivement, [*Tw1, Tw2 et Ta4*] et [*Tw1, Tw2, Ta1 et*

QI] comme variables explicatives. Les séries chronologiques des températures hebdomadaires de l'eau observées et simulées par les modèles VPP4 et PARX3 permettent de conclure que les deux modèles peuvent reproduire l'allure générale du cycle annuelle des températures de l'eau (figure 5.7). De plus, les températures calculées et observées montrent que les deux modèles ne produisent pas d'erreur systématique. Autrement dit, les températures calculées ne sont pas systématiquement supérieures ou inférieures aux températures observées.

2.3.4 Conclusion

Dans l'intention de simuler la température de l'eau pour des conditions hydrométéorologiques différentes, l'objectif de cette étude est d'estimer la température de l'eau à partir de la température de l'air et du débit en utilisant des modèles peu ou pas utilisés dans la littérature, soit les modèles VPP et PARX. Chacun des deux modèles utilisés a ses avantages et ses inconvénients. La méthode des VPP est une méthode efficace en terme de performance, cependant elle est intuitive et offre une description peu claire (boîte noire) de la relation entre la température de l'eau et les variables explicatives. Par contre en dépit de sa paramétrisation, le modèle PARX offre une équation explicitant la relation entre la température de l'eau et les variables explicatives. Les modèles VPP4 et PARX3 donnent des résultats satisfaisants avec une REQM de 1,20°C et 1,58°C respectivement. Le biais interannuel associé à ces modèles est nul mais variant chaque année. Ces résultats sont semblables à ceux reportés dans la littérature (Galeati, 1990) pour les modèles VPP et ARX utilisés pour simuler d'autres variables hydrologiques tel que le débit.

En conclusion, même si les deux modèles peuvent simuler adéquatement les températures hebdomadaires de l'eau sur un site et sur la période (1984-2004), le modèle PARX demeure le plus approprié, en raison de la préservation de la persistance des séries périodiques. Enfin, il serait important de comparer les deux modèles sur plusieurs sites. Cette phase de l'étude permettrait de vérifier les résultats obtenus dans le cadre de la présente étude. Il est donc recommandé d'effectuer cette opération ultérieurement.

CHAPITRE 3. REVUE DE LITTÉRATURE

A review of statistical water temperature models

By

Loubna Benyahya*,¹

Daniel Caissie ²

André St-Hilaire ¹

Taha B.M.J. Ouarda ¹

Bernard Bobée ¹

¹. Chair in statistical hydrology, INRS-ETE, 490 De la Couronne Street, Québec City, Qc.

G1K 9A9

². Fisheries and Oceans Canada, Gulf Fisheries Centre, Moncton, N.B., E1C 9B6

Submitted to

Canadian Water Resources Journal

January 2007

Corresponding author. Email: andre_saint-hilaire@inrs-ete.quebec.ca

Tel. (418) 654-3113. Fax (418) 654-2600

Abstract

The use of statistical models to simulate or to predict stream water temperature is becoming an increasingly important tool in water resources and aquatic habitat management. This article aims to provide an overview of the existing statistical water temperature models. Different models have been developed and used to analyse water temperature-environmental variables relationship. Generally, deterministic models require numerous input data (e.g. depth, amount of shading, wind velocity). Hence, they are more appropriate for analyzing different impact scenarios due to anthropogenic effects (e.g presence of reservoirs, thermal pollution, deforestation). In contrast to the deterministic models, the main advantage of the statistical models is their relative simplicity and relative minimal data requirement. Parametric models such as linear and non-linear regression are popular methods often used for shorter time scales (e.g. daily, weekly). Ridge regression presents an advantage when the independent variables are highly correlated. The periodic models present advantages in dealing with seasonality that often exist in periodic time series. Non-parametric models (e.g. k-nearest neighbors, artificial neural networks) are better suited for analysis of nonlinear relationships between water temperature and environmental variables. Finally, advantages and disadvantages of existing models and studies are discussed.

Key words: water temperature models, parametric statistical models, non-parametric statistical models

Résumé

Étant donné l'importance de la température dans l'habitat aquatique, et étant donné les impacts humains actuels et potentiels sur le régime thermique des rivières, il s'avère nécessaire de développer des outils de gestion des ressources hydriques. Cet article propose une synthèse bibliographique des différents modèles utilisables pour la simulation ou la prévision de la température de l'eau. De nombreux modèles existent pour prédire la température de l'eau en fonction des conditions environnementales. Ils se regroupent essentiellement en deux grandes catégories: les modèles déterministes et les modèles statistiques/stochastiques. Les modèles déterministes de la température de l'eau nécessitent souvent un grand nombre d'intrants (e.g, couvert végétal, vitesse du vent, profondeur de la rivière) qui ne sont pas toujours disponibles. Alternativement, les modèles statistiques offrent l'avantage de nécessiter moins de données et un temps de développement habituellement moins long que les modèles déterministes. L'article réalise une étude comparative de l'application des différents modèles statistiques existant à savoir les modèles paramétriques (régression linéaire, non-linéaire, régression "Ridge", modèles périodiques, etc) et les modèles non-paramétriques (Réseaux de Neurones Artificiels et les k-Voisins les plus proches). Tous ces modèles ont donné des résultats encourageants voire satisfaisants et il est difficile d'exhiber la supériorité d'un modèle particulier. On conclue en présentant les avantages et les limites des différents modèles.

Mots-clés : modèles de température de l'eau, modèles paramétriques, modèles non-paramétriques.

3.1 Introduction

3.1.1 Importance of water temperature for fish habitat

Water temperature is one of the most important parameters in ecosystem studies. Temperature can influence both chemical and biological processes such as dissolved oxygen concentrations, fish growth and even mortality. Many studies have addressed human influences or anthropogenic perturbations of river thermal regime in order to better understand their impacts on fish habitat. For example, Beschta et al. (1987) have studied the impact of forestry operations on river water temperature whereas Webb and Walling (1993) looked at the impact of reservoirs on downstream temperatures. Many biological conditions are linked to river thermal regime. For instance, Hodgson and Quinn (2002) demonstrated that the triggering of the spawning period for Sockeye Salmon (*Oncorhynchus nerka*) on the Northwestern coast of the U.S. was strongly influenced by water temperature. They found that when a threshold of 19 °C was reached, spawning was interrupted as individual fish sought thermal refuge. Bjornn and Reiser (1991) showed that high stream temperatures in the range of 23-25 °C can increase the mortality rate of salmonid fish. Johnson (1997) reported on the relation between the timing of emergence of Atlantic salmon (*Salmo salar*) fry and water temperature.

River and stream temperatures can also be impacted globally, by climate change. Notably, Eaton and Scheller (1996) suggested that under a doubling of the current atmospheric carbon dioxide concentration, densities of a number of cold water fish species could decline by as much as 50% in the United States. Mohseni et al. (2003) reported that, under this scenario, suitable thermal habitat for cold water fish will suffer a 36% decrease.

These and other studies show the critical role of river water temperature and the importance of understanding the thermal regime of rivers for effective aquatic management.

3.1.2 Spatial and temporal variability

The thermal regime of rivers is affected by heat exchanges in which both meteorological factors and physical characteristics are important. Ultimately, these factors will influence spatial and temporal variation in stream water temperature. Meteorological conditions are responsible for important heat exchange processes. Meteorological factors affecting the temporal variability of energy exchanges include the annual cycle of incoming solar radiation as well as short-term conditions of wind speed, humidity and air temperature among others. In North America, the greatest temporal variation in stream water temperature occurs in summer with an initial period of heating (generally from May to September) followed by a cooling period. The natural processes of heating and cooling highly depend on meteorological conditions as well as stream physical and hydrological characteristics. Stream characteristics affecting energy exchange include riparian vegetation, stream aspect, channel geomorphology, valley topography, location of tributaries and groundwater inflows. Spatial and temporal variations in water temperature are important for aquatic resources. For example, Vannote et al. (1980) have shown that in many cases both seasonal and daily variations of water temperature are important factors that influence the distribution and the composition of aquatic communities along river systems.

The thermal regime is naturally complex in itself; however, the spatial and temporal variability are often disturbed by human industrial use such as deforestation and regulation by dams (Webb and Walling, 1993). For instance, riparian vegetation, which can be affected by forestry operations, can have a major effect on various components of the stream energy budget (Brown and Krygier, 1967). Riparian vegetation has a dominant role in protecting the stream against heating although the cutting and clearing of trees within the basin can also impact river water temperature. Notably, a recent study suggested that increases in water temperature due to soil heating after forest removal remain low overall, but may be important during summer storm events (St-Hilaire et al. 2000). Flow modification and/or water withdrawal can also impact on spatial and temporal variability of water temperature. For example, Sinokrot and Gulliver (2000) showed that a reduction of river flow greatly influenced the thermal regime, by increasing water temperature within the affected reach. Correlations between air and water temperature were shown to be stronger when discharge was below the annual median (Webb et al., 2003).

3.1.3 Study objectives

Because of the ecological importance of stream water temperature and the potential impact of human activities on such temperatures, it is important to provide water resource and aquatic habitat managers with efficient assessment and modeling tools. In this study, a number of statistical models for predicting water temperature will be described, via a literature review. The emphasis of the review will be on statistical models as they are widely used in many fields of studies, mainly due to their relatively low data requirement and simplicity in application.

3.2 Water temperature models

Different models have been developed and used to predict river water temperature. These models have been classified into two major categories: deterministic models and statistical models. In specific application, each type of model has advantages and drawbacks. Although the aim of this review is to compare available statistical models, deterministic models will also be briefly described within the next section.

3.2.1 Deterministic models

Deterministic models are based on mathematical representation of the underlying physics of heat exchange between the river and the surrounding environment. Such models are generally carried out using an energy budget approach (Morin and Couillard 1990; Sinokrot and Stefan 1993; St-Hilaire et al. 2000). They require numerous input data including stream geometry, hydrology and meteorology. For example, physical characteristics of the stream such as the depth of water as well as the amount of shading and wind sheltering are often essential components of deterministic models in order to estimate the total energy exchanged within a river reach.

Deterministic models are based on energy conservation equations, thus estimating the changes in water temperature as a function of energy fluxes. For a well-mixed river, it is generally assumed that: 1) the heat flux from the streambed is small compared with the heat flux at the water surface (Sinokrot and Stefan 1994), and 2) the water temperature is fairly uniform with depth and laterally. Under these conditions, the general one-dimensional (x) model is expressed as follows:

$$\frac{\partial T_w}{\partial t} + v_x \frac{\partial T_w}{\partial x} = \frac{W}{\theta \rho A} H_{total} \quad (3.1)$$

where T_w : water temperature ($^{\circ}\text{C}$)

t : time (e.g. day)

A : cross sectional area (m^2)

W : river width (m)

θ : specific heat of water ($4.16 \times 10^{-3} \text{ MJ kg}^{-1} \text{ }^{\circ}\text{C}^{-1}$)

ρ : water density (1000 kg m^{-3})

v_x : mean water velocity (m day^{-1})

x : longitudinal direction

H_{total} : total heat flux from the environment to the river ($\text{MJ m}^{-2} \text{ day}^{-1}$)

The total heat flux is considered to be the summation of the different heat flux components and the main components in a heat budget are as follows:

$$H_{total} = H_s + H_l + H_e + H_c \quad (3.2)$$

where H_s : solar radiation or net short-wave radiation at the stream ($\text{MJ m}^{-2} \text{ day}^{-1}$)

H_l : net long-wave radiation ($\text{MJ m}^{-2} \text{ day}^{-1}$)

H_e : evaporative heat transfer ($\text{MJ m}^{-2} \text{ day}^{-1}$)

H_c : convective heat transfer ($\text{MJ m}^{-2} \text{ day}^{-1}$)

A positive energy flux means that energy is added to the river, while a negative one tends to cool the river system. For solving the equation (3.1), the numerical technique such as the Element Finite Method can be used.

Deterministic models are efficient tools when the users want to simulate modifications to some components of the heat budget (St-Hilaire et al. 2000). Consequently, they are very useful for analyzing and comparing different impact scenarios due to anthropogenic effects such as presence of reservoirs, thermal pollution, deforestation and others. For instance, the SHADE model simulates water temperature and includes shading effect (Chen et al. 1998). More recently, St-Hilaire et al. (2000) modified the CEQUEAU and hydrological and water temperature model (Morin et al. 1981) to include soil temperature and crown closure in its calculation of local advective terms in the heat budget. Others deterministic models include the US Fish and Wildlife SSTEMP model (Bartholow, 1999), as well as number of simpler models (e.g. Sinokrot and Stefan 1993; Gu et al. 1998; Gu and Li 2002; Younus et al. 2000). One potential limitation of this modelling approach results from the fact these tools can be quite demanding in terms of data requirement and their implementation can therefore be complex.

3.2.2 Statistical/Stochastic models

An alternative approach to deterministic models in predicting or simulating water temperatures is the use of statistical or stochastic models. In contrast to the deterministic models, the main advantage of the statistical models is their relative simplicity and minimal data requirement. Statistical models can generally be classified in two categories, namely parametric and non-parametric models. Among parametric models, a further classification can be made into regression models and stochastic models. Regression models are usually

applied for predicting or simulating water temperature at weekly, monthly and annual time steps, mainly due to the relatively high correlation between air and water temperature at those time scales. In some cases, non-linear regression models have been used to better capture the leveling off of water temperature at both high and low air temperature (Mohseni et al. 1998). This change in slope has mainly been attributed to groundwater inputs at low temperatures and to evaporative cooling at high temperatures.

When water temperatures are modeled for shorter time step than weekly, linear and non-linear regression models are generally more difficult to apply due to the autocorrelation within the water temperature time series. Hence, for shorter than weekly time scales, e.g. daily, stochastic models are often used as well as other types of models that account for the autocorrelation.

Although stochastic models linking water to air temperatures offer a simple means of predicting or simulating water temperature, other statistical models, such as parametric (e.g. Box Jenkins, ARMA, etc.) and non-parametric models (k-Nearest Neighbours, Artificial Neural Networks, etc.) are “statistically faithful” to the type of time series and represent well water temperature variability. Parametric models require estimation of parameters, the number of which depends on the number of independent variables. Non-parametric models differ from parametric models in that the model structure is not specified *a priori*, but is instead determined from data. The weights or model coefficients associated with each independent variable usually have little or no physical or statistical meaning. These methods are only beginning to be implemented and adapted to water temperature modeling. The following sections provide a more detailed review of the statistical approaches available to model water temperature.

3.3 Parametric statistical models

3.3.1 Regression models

Many studies have used a statistical approach to predict water temperature. Simple regression-based models have been successfully used to model water temperature as a function of one (usually air temperature) or more independent variables. The structure of these simple models can be depicted as follows:

$$Tw(t) = a_0 + a_1 * Ta(t) \quad (3.3)$$

where $Tw(t)$: water temperature for a given time period

$Ta(t)$: air temperature for a same time period as water temperature

a_0, a_1 : regression coefficients

Such linear regression models have been applied by Stefan and Preud'homme (1993) who demonstrated that the water-air temperature relationship becomes less scattered as the time interval of the data increases from 2-h, through daily averages to weekly means. Pilgrim et al. (1998) also used linear regression to relate stream water temperature to air temperature for a number of sites in Minnesota (U.S.A). They showed that the slope of the regression is increasing with time scale (daily, weekly and monthly). Erickson and Stefan (2000) showed that both the slope and intercept are a function of the time scale. They concluded that during open water periods, streams in Minnesota had a good linear air/water temperature relationship; however the warmer Oklahoma sites showed a non-linear structure when air temperature exceeded 25°C. This was most likely due to evaporative cooling. Moore (2006) showed that the linear models relating monthly median water

temperatures to climatic and catchment characteristics, explained up to 79% of the variance of water temperature.

Equation (3.3) specifies air temperature as the only independent variable, but the model has been generalized using multiple regression. Webb et al. (2003) noted that flow is another important variable that should be considered in water temperature models. Their study showed that air and water temperatures are more strongly correlated when flows are below median levels.

In situations where the predictor variables are highly cross-correlated amongst themselves (collinearity), the challenge is to minimize the possibility of including redundant variables in the model. Ridge regression is an attempt to deal with collinearity through use of a form of biased estimation in place of ordinary least squares (OLS) regression. The basic justification for ridge regression is that a slightly biased estimator with smaller variance may be more advantageous than an unbiased estimator having large variance. The ridge regression is a regression approach under constraint (Hoerl and Kennard, 1970) for which the equation is similar to the multiple regression model. In matrix form, the multiple regression can be written as:

$$Y = \beta X + \varepsilon \quad (3.4)$$

Where Y : vector representing the dependant variable;

β : vector of coefficients to be adjusted;

X : matrix of independent variables.

ε : error term

The estimator β can be given by the equation:

$$\beta = (X'X)^{-1} X'Y \quad (3.5)$$

In the ridge regression, a ridge constant K is included in order to avoid ill-conditioning of X :

$$\beta = (X'X + KI)^{-1} X'Y \quad (3.6)$$

Where I is the identity matrix.

The selected value for the ridge constant is the lowest value for which β is stabilized. Ahmadi-Nedushan et al. (2006) have shown that a ridge regression model can be used to simulate daily water temperature. Simulations on the Moisie river (Québec, Canada) showed good results with a root mean square error (RMSE) < 0.65 °C.

The assumption that the water-air temperature relationship is linear has been questioned. Indeed, Mohseni et al., (1998) also observed a non-linear behaviour between air and water temperatures at weekly intervals. Accordingly, these authors developed a model based on the logistic S-Shaped function to predict average weekly stream temperatures at different locations in the United States. The logistic function used by Mohseni et al., (1998) to determine the air to water relation is given by:

$$Tw = \frac{\alpha}{1 + e^{\gamma(\beta - Ta)}} \quad (3.7)$$

Where Tw and Ta represent water and air temperatures, α is a coefficient which estimates the highest water temperature, β is the air temperature at the inflection point and γ

represent the steepest slope of the logistic function. The advantage of this model over the linear regression is that it can better represent the tendency of water temperature in some water bodies to level off at higher air temperatures (Mohseni and Stefan, 1999). Equation (3.7) was also used by Webb et al. (2003) to model water-air temperature relationship at different time steps on the Exe river (UK). They found significant non-linear relations between these two variables at the hourly time step.

3.3.2 Autoregressive models

Among the autoregressive methods used to simulate and predict water temperature, some of models are labelled as “stochastic”. In this approach, the water temperature time series are generally divided into two components, namely the long-term annual component (seasonal variation) and the short-term variations or departure from the annual component (residuals). Then, time series models (e.g. Box Jenkins, ARMA, etc.) are fitted to water temperature residuals (short-term variations), once the seasonal component of the signal has been removed. Applications of this approach include those of Kothandaraman (1971), Cluis (1972) and Caissie et al. (1998; 2001).

Autoregressive (AR) models take into account the order of measurement and autocorrelation structure within the stream water temperature time series and can also account for the correlation with external variables (e.g. air temperature, streamflow) at various lag periods. The seasonal variation can be modeled by a Fourier series analysis (Kothandaraman, 1971) or even a simple sinusoidal function (Caissie et al., 1998). For example, the sinusoidal function may be written in the following form:

$$Tw_{\text{seasonal}}(t) = a + b \sin\left[\frac{2\pi}{365}(t + t_0)\right] \quad (3.8)$$

where $T_{W_{\text{seasonal}}}(t)$ is the seasonal component of a temperature time series; a , b and t_0 are fitted coefficients.

The short term component is generally modeled using a variety of approaches ranging from multiple regression analysis (Kothandaraman, 1971), to a second order markov chain (Cluis, 1972), or the use of Box-Jenkins time series analysis (Marceau et al., 1986). Caissie et al. (1998) compared these three different “stochastic” approaches to model mean and maximum daily water temperatures in a relatively small stream (Catamaran Brook, New Brunswick, 50 km² drainage area) using air temperature as the independent variable. Their preferred methodology, based on the comparison of goodness of fit (e.g. Nash coefficient) and error statistics (e.g. root mean square error) involved estimating an annual component in stream temperatures by fitting a Fourier series to the data and a second order Markov process model for the short term component. A similar study was also carried out by Caissie et al. (2001) in which they modeled maximum daily stream temperatures at the same study site (Catamaran Brook) using a stochastic model.

The stochastic approach is a method which requires relatively few parameters and thus its application is simpler. This approach can provide very good results. For instance, Caissie et al. (1998) have obtained RMSE less than 0.9 °C. Despite the good RMSE, this approach has some potential difficulties. For instance, a fixed sinusoidal function needs to be fitted to the time series. Thus, it can be argued that this may result in non-stationary residuals from year to year. Stationarity (i.e no seasonality in the data) is one of the underlying hypotheses of a number of time series models, including AR and Box Jenkins. As such, users can verify that residuals are indeed stationary prior to implementing such approaches.

3.3.3 Periodic Autoregressive models

Because of the thermal inertia of water bodies, stream temperatures and air temperatures are better correlated at the weekly and monthly timescale than at the hourly or daily scale (Pilgrim et al., 1998). Weekly stream temperature values have been used as criteria to characterize some fish habitat in a number of studies (e.g. Eaton and Scheller, 1996; Stefan et al., 2001), although it may be important in some studies to predict water temperature for shorter time scales. As stated in the previous section, the common procedure in modeling such periodic series is to deseasonalize the series prior to applying the stationary models (Salas et al., 1980; Vecchia, 1985; Salas, 1993; Chen and Rao, 2002). However, filtering time series may not yield stationary residuals due to periodic autocorrelations. In such cases, the resulting model may be misspecified for series in which periodic properties are present (Tiao and Grupe, 1980). To model periodicity in autocorrelations, periodic models can be advantageous. In such situations, an important class of periodic models consists of Periodic AutoRegressive (PAR) and Periodic AutoRegressive Moving Average (PARMA) models, which are extensions of commonly used ARMA models (Box and Jenkins, 1976) with the difference that the former use periodic parameters. ARMA models assume that the data are stationary (i.e no seasonality in the data). PARMA models are widely used for prediction of economic time series (Osborn and Smith, 1989; Novales and de Frutto, 1997) as well as in the field of hydrology (Salas et al., 1980; Vecchia, 1985; Bartolini et al., 1988; Ula and Smadi, 1997). PAR and PARMA models are usually applied to time series at monthly time steps or more, which limits the number of periods to 12 and hence the number of parameters.

Recently, to model the average weekly maximum temperatures (T_w), Benyahya et al. (2006) compared the performance of PAR and AR models. The PAR model of lag- i used was of the form:

$$T_{w_{v,\tau}} = \mu_\tau + \sum_{i=1}^p \phi_{i,\tau} (T_{w_{v,\tau-i}} - \mu_{\tau-i}) + \varepsilon_{v,\tau} \quad (3.9)$$

Where, v = year, τ = season (or period); μ_τ = mean of the water temperature process in season τ , and $\phi_{i,\tau}$ = autoregressive parameter which are estimated for each season by using the least square method. The error $\varepsilon_{v,\tau}$ is assumed to be normally distributed with mean zero and variance one. This model was calibrated using 18 years of average weekly maximum temperature series on the Deschutes river (Oregon, U.S.A.) and good modelling results have been obtained with a PAR(1) model with an average error (RMSE) less than 1°C.

PAR models rely strictly on the autocorrelation structure of the variable of interest (water temperature). For water resource and fisheries managers, it may be important to include other variables that may have an impact on the thermal regime of rivers, such as air temperature and flow. For this reason, the PAR method was extended by Benyahya et al. (2007) to incorporate other input variables (e.g. air temperature, streamflow) called exogenous variable and therefore the PAR model become a PARX model. A PARX model representing the water temperature series may be written in the following form:

$$T_{w_{v,\tau}} = \sum_{i_1=1}^{p_1} \phi 1_{i_1,\tau} T_{w_{v,\tau-i_1}} + \sum_{i_2=1}^{p_2} \phi 2_{i_2,\tau} T_{a_{v,\tau-i_2}} + \sum_{i_3=1}^{p_3} \phi 3_{i_3,\tau} Q_{v,\tau-i_3} + \varepsilon_{v,\tau}$$

and $\tau > i_1, i_2, i_3$ (3.10)

Where $\phi 1_{i_1, \tau}$, $\phi 2_{i_2, \tau}$ and $\phi 3_{i_3, \tau}$ are periodic parameters, p_1 , p_2 and p_3 are the lags of water temperature, air temperature and streamflow respectively and $\varepsilon_{\nu, \tau}$ is the error term. This model was calibrated using 21 years of weekly water temperatures of the Nivelle river (France) and results indicated that the PARX model performed relatively well with RMSE < 1.60°C.

The periodic models are particularly well adapted for weekly data with a good level of performance; however, these models require the estimation of a large number of parameters (i.e. one set of parameters for each period). Depending on the length of the time series available to calibrate the model this can violate the principle of parsimony (e.g select a model with as few parameters as possible).

3.4 Non-parametric statistical models

As stated earlier, the second main category of statistical models use the so-called non-parametric approaches. The structure of these models is highly dependent on available data and there is generally no judgement made by the modeller on the statistical structure of the model. Non-parametric models are considered to be good “data-learners” and their use should be limited to the range of values encountered in the past, i.e. their performance in the extrapolation range can be less reliable. Nonetheless, recent advances in computational algorithms and computing power of modern computers have made it possible to implement these models with relative ease. Artificial Neural Networks as well as the k-Nearest Neighbours are examples of such approaches that have been adapted to water temperature modeling.

3.4.1 Artificial Neural Networks

An Artificial Neural Network (ANN) model is a mathematical structure capable of describing complex nonlinear relations between input and output data. The architecture of ANN model is inspired by biological nervous systems (Figure 3.1). As in nature, independent variables (or predictors) are fed as inputs in the input layer through nodes (neurons) used during neural network training. The hidden layer is the location where the neural network is “trained” i.e. where weights are assigned to non-linear functions that combine the inputs (Ahmadi-Nedushan et al., 2007). The network connection weights are adjusted in order to minimize the error between the ANN outputs and the training set of the variable to be modeled. The weights of each node in the layers need to be adjusted. This can be done using several learning algorithms. One of the most popular learning algorithms is back propagation. In back propagation, a gradient descent is implemented to ensure that the direction of learning and rate of learning is appropriate.

In the field of hydrology, ANN modeling has been used for a variety of purposes. Hsu et al. (1998) used ANN for streamflow forecasting. In water-quality application, Conrads and Roehl (1999) used ANN models to simulate salinity, temperature, and dissolved oxygen. Coulibaly et al. (2001) applied the temporal neural networks to hydropower reservoir inflow forecasting. Risley et al. (2003) estimated water temperatures in small streams in Western Oregon using an ANN model. More recently, Belanger et al. (2005) compared two models of water temperature: artificial neural networks and multiple linear regression using air temperature and discharge as independent variables. Of these two models, results indicated that both approaches were equally good in predicting daily stream water

temperature with RMSE of 1.06°C for the regression model and 1.15°C for the ANN model.

ANN have often been criticized based on the fact that the contribution of the input variables in predicting the output is difficult to disentangle within the network, and explanations regarding the relative importance of each independent variable are not as straightforward as in the case of linear regression methods (Olden and Jackson, 2002). However, trained networks have been shown to perform very well within the interpolation range of the training data set and a number of packages exist that allow for relatively easy implementation of the approach.

3.4.2 k-Nearest Neighbours (k-NN)

A k-NN is a method that consists of finding for a given point in time, a small number of neighbours nearest to this value, and the prediction is estimated based on these neighbours.

The key steps in the k-NN algorithm are as following:

Step 1. Compile a feature vector: the vector X consists of values of the selected input attributes (e.g lagged water temperature, lagged air temperature and streamflow data) for which we are trying to find the k-Nearest Neighbours (usually $k < 4$);

Step 2. Find the weighted sum of the attributes: since the scales of water and air temperature units are different than that of streamflow, the weighted attributes can be generalized as a weighted standardized norm (N):

$$N = \sum_{i=1}^{N_{\text{attributes}}} w_i X_i \quad (3.11)$$

Where w_i are weights and X_i are the vectors of standardized (i.e. subtract mean value and divide by the standard deviation) values of the selected attributes.

Step 3. Calculate the Euclidean distance between the norm of the day of interest and the norm of all other available data: For two norms (N_{j_1}, N_{j_2}) calculated using vectors X_{j_1} for the day of interest and X_{j_2} ($j_2 = 1, \dots, j_1 - 1, j_1 + 1, \dots, m$) for the m other days in the data base, the Euclidean distance (δ) is defined as:

$$\delta(N_{j_1}, N_{j_2}) = |N_{j_1} - N_{j_2}| = \sum_{i=1}^{n_{\text{attributes}}} w_i * |X_{i,j_1} - X_{i,j_2}| \quad (3.12)$$

Step 4. Sort the distances δ in ascending order, and retain only the first k nearest neighbours: the strategy for choosing the optimal k is to try several successive values of k (e.g, 2, 3 and 4) and to select the combination for which the model gives the best prediction.

Step 5. Assign weight to each of the k neighbours, thus the predicted value of the final output is computed as a weighted sum of the values of neighbouring observations;

The k-NN is an approach that has been used in hydrology to model rainfall-runoff processes and has been compared with autoregressive moving average models with exogenous inputs (ARMAX) (Karlsson and Yakowitz, 1987; Yakowitz and Karlsson, 1987). In agreement with the conclusions reported by the aforementioned authors, Galeati (1990) showed that the k-NN method provides lower mean square error predictions of daily mean flow than an autoregressive model with exogenous inputs (ARX). Recently, Benyahya et al. (2007) compared the predictive ability of periodic autoregressive model (PARX) and k-NN to model weekly water temperature of the observation period (1984-

2004) in the Nivelle river, France. It was concluded that PARX is better suited to model the periodicity in autocorrelations; nevertheless, k-NN is equally an interesting statistical water temperature model. The simulations on the Nivelle river yielded a relatively small root mean square error of 1.20°C. However, one potential drawback of k-NN, and other non-parametric methods such as ANN, is that they do not produce a parametric function of the model and in such case, any given condition not previously observed in the historic record cannot be predict or simulated.

3.5 Discussion and Conclusion

This paper provided an overview of the existing methods used in water temperature modeling and compared their relative advantages and drawbacks. Both the deterministic and statistical models are relevant depending on the problem under investigation and data availability. Statistical models, which are the main focus of this review, are important tools for the prediction of water temperature based on few input variables. It was found that most of the statistical models reported in the literature are based on simple and multiple linear regression as well as logistic function. These models have been effective in predicting river water temperature for longer time scales, i.e., weekly, monthly and using annual means. For shorter time scales, stochastic models have been more effective in the prediction of river water temperatures. As noted, most statistical methods reviewed in this study present advantages and disadvantages in different contexts (Table 3.1). The selection of a particular statistical model depends on the modeling objective as well as the type of data available. In situations when the water temperature modeling is carried out at daily time steps and when air temperatures are the only available data, the so-called “stochastic” models have been shown to perform well. AR models applied on residuals (i.e. the non-

seasonal component, often called the stochastic component in the literature) are powerful tools that often produce root mean square errors (RMSE) that are less than 1°C. Users should be reminded that stationarity is assumed to exist in the time series of residuals when using this approach.

More recently other statistical approaches have been applied and these statistical models show promising results. For example, when periodicity is observed in the correlation structure, periodic models have been proposed. These models, such as PAR and PARX models are extensions of commonly used ARMA models with the difference that the former use periodic parameters. Despite of their parametrisation, the periodic models were shown to be an interesting tool for modeling weekly water temperatures based on their ability to model periodicity in autocorrelations (Benyahya et al., 2006; 2007).

The literature shows that non-parametric models are also a promising area of predictive statistical water temperature modeling owing to their ability in simulating the complex nonlinear relationships between response variable and environmental variables. Although non-parametric methods do not provide users with a conventional mathematical function, many readily applicable algorithms exist and their implementation has become easier with the advent of powerful computing technology. Users should be reminded that these approaches are site specific, as is the case in many statistical modelling approach and their performance ability in the extrapolation domain is usually limited.

In conclusion, this literature review shows that the selection of an appropriate statistical water temperature model depends on the comparison of advantages and disadvantages of different methods in a particular context. The context is defined by the following factors:

Time step required (e.g. daily, weekly, monthly, etc.);

Length of time series;

Statistical properties of time series (e.g. seasonality, normality of residuals, etc.);

The need to formalize the relationship between independent variables and water temperature.

Although most of the literature describes statistical model implementations in a simulation mode, these approaches could be used in forecasting as well, especially where forecasts of input variables exist (i.e. air temperature and/or flow). By developing a predictive relationship between water temperature and environmental variables, users could estimate how stream temperatures are likely to respond to these variables, and therefore how the health of many aquatic species could be threatened. Statistical models have played an important role, in the past, in studying water resource and fisheries management issues. Also, it is believed that with the new generation of statistical models this approach will remain an important approach, mainly because water temperature can be predicted with fewer input parameters than deterministic models.

References

- Ahmadi-Nedushan, B., A. St-Hilaire, M. Bérubé, É. Robichaud, N. Thiémonge and B. Bobée. 2006. A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment. *River Research and Applications*, 22(5): 503-523
- Ahmadi-Nedushan, B., A. St-Hilaire, T.B.M.J. Ouarda, L. Bilodeau, É. Robichaud, N. Thiémonge and B. Bobée. 2007. Predicting river water temperatures using stochastic models: case study of the Moisie river (Québec, Canada). *Hydrological Process*, 21(1): 21-34.
- Bartholow, JM. 1999. SSTEMP for Windows: The Stream Segment Temperature Model (Version 1.1.3). US Geological Survey computer model and help file.
- Bartolini, P., J.D. Salas and J.T.B. Obeysekera. 1988. Multivariate periodic ARMA(1,1) processes. *Water Resources Research*, 24: 1237-1246.
- Bélanger, M., N. El-Jabi, D. Caissie, F. Ashkar and J. M. Ribi. 2005. Estimation de la température de l'eau en rivière en utilisant les réseaux de neurones et la régression linéaire multiple. *Revue des sciences de l'Eau*, 18(3): 403-421.
- Benyahya, L., A. Saint-Hilaire, T.B.M.J. Ouarda, B. Bobée and B. Nedushan. 2006. Modeling of water temperatures based on stochastic approaches: Case study of Deschutes river (Oregon, USA). *Journal of Environmental Engineering and Science* (In press)

- Benyahya, L., A. St-Hilaire, T.B.M.J. Ouarda, B. Bobée and J. Dumas. 2007. Modeling of water temperatures based on stochastic approaches: Case study of the Nivelle river (St-Pée, France). Submitted to *Hydrological Sciences Journal*.
- Beschta, R.L., R.E. Bilby, G.W. Brown, L.B Holtby and T.D. Hofstra. 1987. Stream temperature and aquatic habitat: fisheries and forestry interactions. In: Streamside Management: Forestry and Fishery Interactions (Eds Salo, E.O., and T.W. Cundy). University of Washington, Institute of Forest Resources. Contribution No. 57. P. 191-232.
- Bjornn, T. C. and D. W. Reiser. 1991. Habitat requirements of salmonids in streams. Pages 83-138 in W. R. Meehan, editor. Influences of forest and rangeland management on salmonid fishes and their habitats. Special Publication 19. American Fisheries Society, Bethesda, Maryland
- Box, G. E. P. and G. M. Jenkins. 1976. Time series analysis: forecasting and control. Revised edition. (2nd edn). Holden-Day, San Francisco.
- Brown G.W and J.T. Krygier. 1967. Changing water temperatures in small mountain streams. *Journal of Soil and Water Conservation*, 22: 242–244.
- Caissie D, N. El-Jabi and M. Satish. 2001. Modeling of maximum daily water temperatures in a small stream using air temperatures. *Journal of Hydrology*, 251: 14-28.
- Caissie D, N. El-Jabi and A. St-Hilaire. 1998. Stochastic modelling of water temperatures in a small stream using air to water relations. *Canadian Journal of Civil Engineering*, 25: 250–260.

- Chen Y.D., R.F. Carsel, S.C. McCutcheon and W.L. Nutter. 1998. Stream temperature simulation of forested riparian areas: 1. Watershed-scale model development. ASCE, *Journal of Environmental Engineering*, 124: 304–315.
- Chen, H.L and A.R. Rao. 2002. Testing hydrologic time series for stationarity. *Journal of Hydrologic Engineering*, 7(2): 129-136.
- 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.
- Conrads, P.A. and E.A. Roehl. 1999. Comparing physics-based and neural network models for simulating salinity, temperature, and dissolved oxygen in a complex tidally affected river basin *in* Proceedings of the 1999 South Carolina Environmental Conference, Myrtle Beach, March 15-16, 1999: Columbia, South Carolina, Water Environmental Association of South Carolina, 7 p.
- Coulibaly, P., F. Anctil and B. Bobée. 2001. Multivariate reservoirinflow forecasting using temporal neural networks. ASCE, *Journal of Hydrologic Engineering*, 6(5): 367-376
- Eaton, J.G. and R.M. Scheller. 1996. Effects of climate warming on fish thermal habitat in streams of the United States. *Limnology and Oceanography*, 41(5): 1109-1115
- Erickson, T.R. and H.G. Stefan. 2000. Linear air/water temperature correlations for streams during open water periods. ASCE, *Journal of Hydrologic Engineering*, 5(3): 317-321.
- Galeati, G. 1990. A comparison of parametric and non-parametric methods for runoff forecasting. *Hydrological Science Journal*, 35(1): 79-94

- Gu, R., S. Montgomery and T. Austin. 1998. Quantifying the effects of stream discharge of summer river temperature. *Hydrological Science Journal*, 43(6): 885-894.
- Gu, R. and Y. Li. 2002. River temperature sensitivity to hydraulic and meteorological parameters. *Journal of Environmental Management*, 66: 43-56.
- Hodgson S. and T.P. Quinn. 2002. The timing of adult sockeye salmon migration into fresh water: adaptations by populations to prevailing thermal regimes. *Canadian Journal of Zoology*, 80: 542-555.
- Horel A. and R. Kennard. 1970. Ridge Regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12: 55-67.
- Hsu, K., H.V. Gupta and S. Sorooshian. 1998. Streamflow forecasting using artificial neural networks, in Proceedings of the International Water Resources Engineering Conference, August 3-7, 1998, Memphis, Tennessee: Reston, Virginia, ASCE, 2: 1362-1367.
- Johnston T.A. 1997. Downstream movement of young-of-the-year fishes in Catamaran Brook and the Little Southwest Miramichi river, New Brunswick. *Journal of Fish Biology*, 51: 1047-1062.
- Karlsson, M. and S. Yakowitz. 1987. Nearest-Neighbor methods for non-parametric rainfall-runoff forecasting. *Water Resource Research*, 23:1300-1308
- Kothandaraman, V. 1971. Analysis of Water Temperature Variations in Large Rivers. ASCE, *Journal of the Sanitary Engineering Division*. 97 (SA1): 19 31.

Marceau, P., D. Cluis and G. Morin. 1986. Comparaison des performances relatives à un modèle déterministe et à un modèle stochastique de température de l'eau en rivière.

Canadian Journal of Civil Engineering, 13(3): 352-364.

Mohseni O., H.G. Stefan and T.R. Erickson. 1998. A nonlinear regression model for weekly stream temperatures. *Water Resources Research*, 34(10): 2685–2692.

Mohseni, O. and H.G. Stefan. 1999. Stream temperature/air temperature relationship : a physical interpretation. *Journal of Hydrology*, 218 :128-141.

Mohseni O., H.G .Stefan and J.G. Eaton. 2003. Global warming and potential changes in fish habitat in U.S. streams. *Climatic Change*, 59: 389–409.

Moore, D. 2006. Stream temperature patterns in British Columbia, Canada, based on routine spot measurements. *Canadian Water Resources Journal*. 31: 41-56.

Morin, G., J.-P. Fortin, J.-P. Lardeau, W. Sochanska and S. Paquette. 1981. Modèle CEQUEAU : Manuel d'utilisation. Rapport de recherche no R-93, INRS-Eau, Sainte-Foy

Morin, G. and D. Couillard, 1990. Predicting river temperatures with a hydrological model. Encyclopedia of Fluid Mechanics, 10, Gulf Publishing Compagny, Hudson, Texas, 171-209.

Novales, A. and R.F. de Frutto, 1997. Forecasting with periodic models: A comparison with time invariant coefficient models. *International Journal of Forecasting*, 13: 393-405.

Olden, J.D. and D.A. Jackson, 2002. Illuminating the “black box”: Understanding variable contributions in artificial neural networks. *Ecological Modelling*, 154: 135-150

- Osborn, D. and J. Smith, 1989. The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *Journal of Business and Economic Statistics*, 7: 117-127.
- Pilgrim, JM., X. Fang and HG. Stefan. 1998. Stream temperature correlations with air temperatures in Minnesota: implications for climate warming. *Journal of the American Water Resources Association*, 34(5): 1109-1121.
- Risley, J.C, E.A. Roehl and P.A. Conrads. 2003. Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models. *U.S Geological Survey Report 02-4218*
- Salas, J.D., J.W Delleur, V.Yevjevich and E.L. Lane. 1980. Applied modelling of hydrological time series. Water Resources Publications, Littleton, Colorado.
- Salas, J.D. 1993. Analysis and modelling of hydrologic time series. Handbook of Hydrology. Maidment, D. R., McGraw-Hill.
- Sinokrot B.A. and HG. Stefan. 1993. Stream temperature dynamics: measurements and modeling. *Water Resources Research*, 29(7): 2299-2312.
- Sinokrot, B.A. and H.G. Stefan. 1994. Stream Water-Temperature Sensitivity to Weather and Bed Parameters. *Journal of Hydraulic Engineering*, 120(6): 722-736.
- Sinokrot B.A. and J.S Gulliver. 2000. In-stream flow impact on river water temperatures. *Journal of Hydraulic Research*, 38: 339-349.
- St-Hilaire A., G. Morin, N. El-Jabi and D. Caissie. 2000. Water temperature modelling in a small forested stream: implication of forest canopy and soil temperature. *Canadian Journal of Civil Engineering*, 27: 1095-1108.

- Stefan, H.G. and E. Preud'homme, 1993. Stream temperature estimation from air temperature. *Water Ressources Bulletin*, 29(1): 27-45.
- Stefan H.G., X. Fang and J.G. Eaton. 2001. Simulated fish habitat changes in North American lakes in response to projected climate warming. *Transactions of the American Fisheries Society*, 130: 459–477.
- Tiao, G.C. and M.R. Grupe. 1980. Hidden periodic autoregres-sive moving average models in time series data. *Biometrika*, 67: 365-373.
- Ula, T.A. and A.A. Smadi. 1997. Periodic stationarity conditions for periodic autoregressive moving average processes as eigen-value problems. *Water Resource Research*, 33: 1929-1934.
- Vannote R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell and C.E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37:130–137.
- Vecchia, A.V. 1985. Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resource Bulletin*, 21: 721-730.
- Webb, B.W. and D.E. Walling. 1993. Temporal variability in the impact of river regulation on thermal regime and some biological implications. *Freshwater Biology*, 29: 167–182.
- Webb, B.W., P.D. Clack and D.E. Walling. 2003. Water-air temperature relationships in a Devon river system and the role of flow. *Hydrological Processes*, 17:3069-3084.

Yakowitz, S. and M. Karlsson. 1987. Nearest-Neighbor methods with application to rainfall/runoff prediction. p. 149–160. In J.B. Macneil and G.J. Humphries (ed.) Stochastic Hydrology. D. Reidel, Norwell, MA

Younus M., M. Hondzo and B.A. Engel. 2000. Stream temperature dynamics in upland agricultural watershed. *Journal of Environmental Engineering*, 126: 518–526.

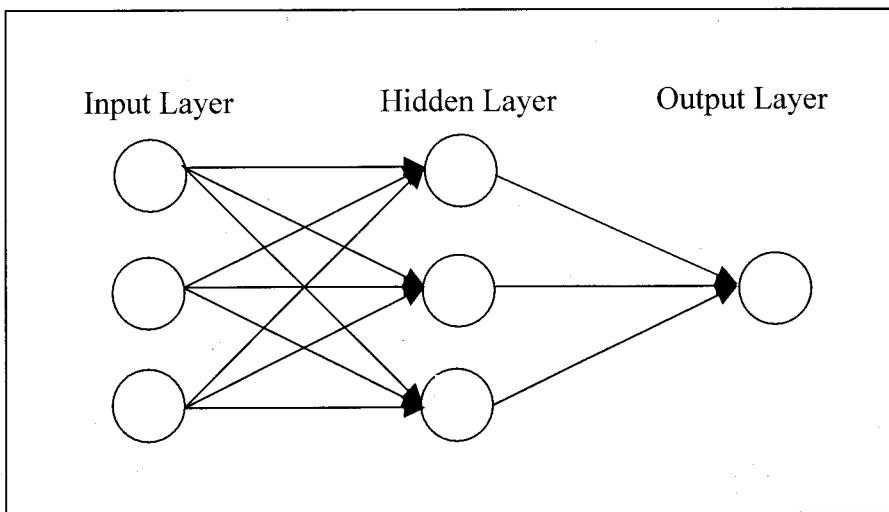


Figure 3.1 Artificial Neural Network representation

Table 3.1 Advantages and disadvantages of statistical models based on approach and identification of related studies

Modeling approach	Advantages	Disadvantages	Examples of water temperature modeling applications
Linear regression	Straightforward in application	Less appropriate when the assumption of the relationship is linear cannot be verified.	Pilgrim et al. (1998) Erickson and Stefan (2000) Ahmadi-Nedushan et al. (2007)
Logistic function	Appropriate when nonlinear relationship is observed in the data	Performed poorly in some cases when studying daily water temperature time series	Mohseni et al. (1998) Caissie et al. (2001) Webb et al. (2003).
Stochastic model	Appropriate when studying daily water temperature time series. Focuses on adequate modeling of residuals.	Appropriate when residuals are stationary	Cluis (1972) Caissie et al. (1998) Caissie et al. (2001)
Periodic Autoregressive models	Can capture the periodic autocorrelation in periodic time series	Necessitates the fitting of a greater number of parameters, which may violate the principle of parsimony.	Benyahya et al. (2006, 2007)
Artificial Neural Network K-Nearest Neighbours	No assumptions concerning statistical distributions and relationships are required; Ability in picking up the complex nonlinear relationships between input and output variables Completely data-driven.	Relatively more costly in computational time; Does not provide users with a conventional mathematical function and physical interpretation. Can be difficult to reproduce results in extrapolation. Requires relatively long time series. Dependency on the selection of a “good value” for k	Risley et al. (2003) Belanger et al. (2005) Galeati (1990) Benyahya et al. (2007)

**CHAPITRE 4. MODÉLISATION DE LA
TEMPÉRATURE DE L'EAU : LES
MODÈLES AR et PAR**

**Modeling of water temperatures based on stochastic
approaches: Case study of Deschutes river
(Oregon, USA)**

By

Loubna Benyahya*,¹

André St-Hilaire¹

Taha B.M.J. Ouarda¹

Bernard Bobée¹

Behrouz Ahmadi-Nedushan¹

¹ Chair in statistical hydrology, INRS-ETE, Université du Québec.

490 de la Couronne Street, Québec City, Québec. G1K 9A9 Canada

In press

Journal of Environmental Engineering and Science

August 2006

Word count: 6967

* Corresponding author. Email: Loubna_benyahya@ete.inrs.ca, Tel. (418) 654-4469. Fax (418) 654-

Abstract

Water temperature is an important physical variable in aquatic ecosystems. It can affect both chemical and biological processes, such as dissolved oxygen concentrations and the metabolism and growth of aquatic organisms. For water resource management, stream water temperature models that can accurately reproduce the essential statistical characteristics of historical data can be very useful. The present study deals with the modeling of Average Weekly Maximum Temperature (AWMT) series using univariate stochastic approaches in Deschutes river (Oregon, USA). Two different models were used to model AWMT data: an Autoregressive (AR) model and a Periodic Autoregressive (PAR) model. The AR model consisted of decomposing water temperature data into a long-term annual component and a residual component. The long-term annual component was modeled by fitting a sine function to the time series, while the residuals, which were the departure from the long-term annual component, were modeled using a Markov chain process. The PAR model was applied to the standardized data which were obtained by subtracting the AWMT series from interannual mean of each period. In order to test the performance of the above models, the Leave-One-Out (Jackknife) technique was used. The results indicated that both models have good predictive ability for a relatively large system such as the Deschutes river. On an annual basis from 1963 to 1980, the average Root Mean Square Error varied between 0.81°C and 0.90°C for AR(1) and PAR(1) respectively; and the mean bias remained near zero °C. Averaged Nash-Sutcliffe coefficient of efficiency (NSC) values obtained by AR and PAR models are close and thus comparable (0.94 and 0.92 respectively). Of the two models, the PAR(1) model seemed the most promising based on its performance and its ability to model periodicity in autocorrelations. Since no exogenous variables (i.e. air temperatures,

streamflow) were incorporated, the use of the PAR model limits the managerial decisions in natural streams and rivers.

Key words: Average Weekly Maximum Temperature, stochastic model, PAR, AR

4.1 Introduction

Water temperature is an important physical variable in aquatic ecosystems. It can affect both chemical and biological processes, such as dissolved oxygen concentrations and the metabolism and growth of aquatic organisms. In stream ecosystems, water temperature may influence both stream fish (Li et al. 1994; Peterson and Rabeni 1996) and aquatic insects (Vannote and Sweeney 1980; Ward and Stanford 1982). If a stream reach becomes increasingly warm and water temperature exceeds a certain tolerable limit, cold water fish (e.g. Salmonids) will leave their habitat and search for thermal refuge (Eaton and Scheller 1996). If no refuge exists, temperature extremes may cause stress and even fish mortality (Bjornn and Reiser 1991). For instance, Hodgson and Quinn (2002) have demonstrated that during the spawning season, some species of pacific salmon will suspend their spawning activity when water temperature exceeds 19°C.

The thermal regime of a watercourse is governed by the interaction of natural environment processes and human activities. Examples of the latter include thermal pollution (e.g. plant effluents) and deforestation, which has been linked to temperature rises in certain cases (Brown and Krygier 1970).

Given the importance of water temperature on lotic ecosystem, it is essential to provide efficient predictive tools to water resource managers. Two types of water temperature modeling approaches have been used in the past, namely the deterministic and stochastic approaches (Marceau et al. 1986). Deterministic approaches are based on mathematical representation of physical processes and calculate an energy budget to predict stream water temperature. Depending on the complexity of the model, the budget equations may require numerous input

parameters such as meteorological factors (e.g. air temperature, solar radiation, and wind velocity) and physical characteristics of the stream (e.g. depth of water, flow and forest cover). Examples of deterministic models include those of Morin and Couillard (1990), Bartholow (1999), and St-Hilaire et al. (2003). One potential limitation of this modelling approach results from the fact that some input variables can be difficult and/or expensive to measure. However, deterministic models are efficient tools when the user wants to simulate modifications to some components of the heat budget (St-Hilaire et al. 2000). Consequently, they are very useful for analyzing and comparing different impact scenarios due to anthropogenic sources such as presence of reservoirs, thermal pollution, and deforestation.

Alternatively, stochastic approaches are based on statistical functions which estimate water temperature with a limited number of independent variables (e.g. water temperature of previous days and air temperature). However, the disadvantage of these approaches is that they require long time series of water temperature for a particular stream. Examples of stochastic models include those of Pilgrim et al. (1998) who examined linear relationships between stream water temperature and air temperature for a number of sites in Minnesota. This study showed that the slope of the regression increased with increasing the time scale (daily, weekly and monthly). Mohseni et al. (1998) developed a nonlinear logistic function, to predict average weekly stream temperatures from air temperature at different locations in the United States that account for heat storage effects (hysteresis). More recently, Bélanger et al. (2005) modeled water temperatures in Catamaran Brook, a small stream catchment (51 km^2) in New Brunswick (Canada). They developed a neural network and a multiple linear regression to relate water

temperature to air temperature and discharge. They concluded that both approaches are equally adequate in predicting stream water temperatures.

The so-called stochastic models are generally applied for smaller time steps (e.g. daily). In this approach, the annual component of the water temperatures is first removed and time series models are then fitted to water temperature residuals. For instance, Caissie et al. (1998) and Caissie et al. (2001) modeled mean and maximum daily water temperatures in Catamaran Brook using air temperatures as the independent variable. Their preferred methodology involved estimating an annual component in water temperatures by fitting a sinusoidal function and/or Fourier series to the data and a second order Markov process to the residuals (Cluis, 1972).

It can be seen from this brief literature review that stochastic methods can be successfully applied to model water temperatures at different time scales. However, to assess changes in fish habitat and growth conditions, the weekly time scale is deemed most appropriate (Eaton and Scheller 1996; Oliver and Fidler 2001). Therefore, the general objective of this study is to develop stochastic approaches to accurately predict Average Weekly Maximum Temperatures (AWMT) at one site in the Deschutes river, a relatively large water course in the state of Oregon (USA).

The timescale is an important deciding factor of time series models. For example, many hydrological time series are periodic and show seasonal variation. In many cases, seasonality of the data cannot be removed by simple deseasonalization to achieve stationary residuals because the correlation structure of the series may be dependent on the period. In such situations, periodic models are more suitable as they are able to model periodicity in autocorrelations. Two popular periodic models are the PAR (periodic autoregressive) and PARMA (periodic autoregressive-moving

average) models, which are basically a group of ARMA (Autoregressive Moving Average) models (Box and Jenkins 1976) that allow periodic parameters. Such models have been widely used in economic applications (Osborn and Smith 1989; Novales and de Frutto 1997) as well as in the field of hydrology (Salas et al. 1980; Vecchia 1985; Bartolini et al. 1988; Rasmussen et al. 1996, Ula and Smadi 1997). However, to our knowledge, the periodic models have not been previously used for water temperature modelling. Therefore, the present study has the following specific objectives: (*i*) to study two different modelling approaches, namely an Autoregressive model (AR) on residuals and a Periodic autoregressive model (PAR), to predict Average Weekly Maximum Temperatures (AWMT) on relatively large water course (*ii*) to compare results of AR and PAR models (*iii*) to verify the efficiency of the two models to simulate water temperature.

4.2 Methods

4.2.1 Autoregressive model (AR) on residuals

In the present study, the Autoregressive model consists first of separating the water temperatures, $Tw(t)$, into two different components: the long-term annual component, $TA(t)$, and the short-term component or residuals, $Rw(t)$, such as:

$$Tw(t) = TA(t) + Rw(t) \quad (4.1)$$

where t represents the week of the year (e.g., $t = 18$ for first week of May).

Each term in equation (4.1) is modeled separately. The long-term annual component is computed by using the approach suggested by Cluis (1972): fitting a sinusoidal function to the AWMT series:

$$TA(t) = a + b * \sin\left[\frac{2\pi}{52}(t + t_0)\right] \quad (4.2)$$

where a and b are fitted coefficients so as to minimize the sum of squared errors; and t_0 is the mean time of departure from 0°C in the water temperature. The annual variation in water temperatures can also be calculated by using an empirical inter-annual mean value.

The stationary residual component, which is obtained from the difference between the original time series and $TA(t)$, is modeled by using first and second order Markov Chains. For instance, the second-order Markov Chain is given by

$$Rw(t) = A_1Rw(t-1) + A_2Rw(t-2) + \varepsilon(t) \quad (4.3)$$

Where, $Rw(t)$, $Rw(t-1)$ and $Rw(t-2)$ are the residuals of water temperature at time t , $t-1$ and $t-2$ respectively; A_1 and A_2 are regression coefficients which are estimated using the least square method and $\varepsilon(t)$ is the error term.

In the remainder of this article, the symbols AR(1) and AR(2) represent the autoregressive model with respectively first and second-order Markov Chain as models of residuals.

4.2.2 Periodic Autoregressive Model (PAR)

Given $Tw_{v,\tau}$ an AWMT series, in which v defines the year with $v=1, \dots, N$, and τ defines the period with $\tau=1, \dots, \omega$. The $Tw_{v,\tau}$ series can be written in the following matrix form:

$$\begin{bmatrix} Tw_{1,1} & Tw_{1,2} \dots Tw_{1,\omega} \\ Tw_{2,1} & Tw_{2,2} \dots Tw_{2,\omega} \\ \vdots & \vdots & \vdots \\ Tw_{N,1} & Tw_{N,2} \dots Tw_{N,\omega} \end{bmatrix} \quad (4.4)$$

Such a process may be represented by a PAR(p) model (Salas 1993), which can be expressed as

$$Tw_{v,\tau} = \mu_\tau + \sum_{i=1}^p \phi_{i,\tau} (Tw_{v,\tau-i} - \mu_{\tau-i}) + \varepsilon_{v,\tau} \quad (4.5)$$

where, p represents the order of the model. The model parameters $\phi_{i,\tau}$ and the mean μ_τ are all function of τ . $\varepsilon_{v,\tau}$ is the error term which is not included in the comparative study.

Since the autocorrelations vary for each period, the identification of the appropriate PAR model order from the plot of periodic autocorrelation can become a complicated problem. In such situations, the simplest PAR models can be used. In the present study, two PAR models, namely PAR(1) and PAR(2) are adjusted to the data. Low-order PAR models have been widely used in hydrology. PAR(1) model was originally introduced by Thomas and Fiering (1962) for modeling monthly streamflows.

Before applying the above models, it is necessary to ensure that the original data conforms to the hypothesis of normality. Otherwise, the original series should be transformed. In the present study, the Shapiro-Wilk test (Shapiro and Wilk 1965) is used to test the normality of the data. This test is appropriate when the sample size is lower than 50.

The parameters of the PAR model may be estimated by several methods, such as the method of moments, the least squares method or the method of maximum likelihood

(Salas 1993). In this study, the least squares (LS) method is adopted. Generally, estimates obtained by this method have optimal statistical properties: they are consistent, unbiased and efficient (Hsia 1977). Hence, the parameters $\{\phi_{i,\tau}\}$ of the model (Equation (4.5)) are estimated for each period independently by minimizing the sum of squared errors ($\varepsilon_{\nu,\tau}^2$) :

$$\sum_{\nu=1}^N \sum_{\tau=1}^{\omega} \varepsilon_{\nu,\tau}^2 = \sum_{\nu=1}^N \sum_{\tau=1}^{\omega} \left[(Tw_{\nu,\tau} - \mu_{\tau}) - \sum_{i=1}^p \phi_{i,\tau} (Tw_{\nu,\tau-i} - \mu_{\tau-i}) \right]^2 \quad (4.6)$$

4.2.3 Model evaluation and validation

As mentioned before, stochastic approaches require long time series of water temperature for a particular stream or river. However, long water temperature time series are scarce. Although the time series used in this study can be considered as a relatively long (18 years), it was deemed too short to validate the model using separate calibration and validation periods. Instead, the Leave-One-Out (Jackknife) technique (Quenouille 1949) is used to test the performance of the above models. This method consists of removing one year (e.g., $\{X_{1,1} \ X_{1,2} \dots \ X_{1,\omega}\}$) from the data set of matrix (4.4) and estimating the model parameters using the rest of data set. Then, the model is used to estimate the temperature for the left-out year and estimated values are compared with measurements. The idea of this method is to test if the model is able to make a correct prediction for the removed year.

To evaluate the performance of $AR(p)$ and $PAR(p)$ models, three criteria are used. The first is the Root Mean Square Error (RMSE) (Janssen and Heuberger 1995) which is calculated to combine the variance and the bias. It is given as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (4.7)$$

where, P_i and O_i are the predicted and observed weekly water temperatures, respectively, and n is the number of data points.

The second criterion is the Bias error (B) which is computed simply as the sum of the differences between predicted and observed values divided by n :

$$B = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (4.8)$$

where n is the number of data points.

The third criterion is the Nash-Sutcliffe coefficient of efficiency (NSC). Nash and Sutcliffe (1970) defined NSC, which is less than 1(equal to 1 when $P = O$), as:

$$NSC = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (\bar{O} - O_i)^2} \quad (4.9)$$

where

\bar{O} = Average of observed weekly water temperature values

These criteria are calculated for each year left out for interannual comparison and then for each period for interperiodic comparison.

4.3 Case study area and data set

Time series of water temperature are collected in the Deschutes river, located on north-central Oregon (USA) (Figure 4.1). It is a relatively large river with a drainage area of 26 860 km² (Hubbard et al. 1998). It takes its source near the Cascade Range along the western part of the basin and flows northerly for nearly 300 km before

joining the Columbia river. The climate in the basin is primarily semiarid. The average annual precipitation ranges from as high as 2030 mm in the high Cascade Range, to between 230mm and 510mm for much of the basin (O'Connor et al. 2003). The lower Deschutes river is characterized by its uniform flow. Near its confluence with the Columbia river, the Deschutes river has a mean monthly flow ranging from 124 m³/s in August to 213 m³/s in February (O'Connor et al. 2003). Peak flows result primarily from snowmelt in the Cascade Range.

The Deschutes river is home to both Chinook Salmon (*Oncorhynchus tshawytscha*) and Steelhead (*Oncorhynchus mykiss*) (Department of Environmental Quality, 2003). All of these fish are designated threatened by the National Marine Fisheries Service (NOAA , 2004).

Daily maximum water temperature data used in this study are obtained from the US. Geological Survey (USGS) for the period of 1962-1981 at station 14103000 located 2.25 km upstream from the confluence of the Deschutes river and the Columbia river (Figure 4.1). As there are some missing data for years 1962 and 1981, the daily maximum water temperatures are extracted for 18 years (1963-1980). The Average Weekly Maximum Temperature (AWMT) series, used in the present study, are obtained by averaging daily maximum water temperatures for seven consecutive days. This criterion is based on USEPA's recommendations in *Guidance for Pacific Northwest State and Tribal Temperature Water Quality Standards* (USEPA, 2003). The period of analysis within every year is chosen from May to November (30 weeks) because cold water fish at various life stages (fry emergence, egg incubation, emergence, rearing, migration and spawning) are particularly sensitive to variability in temperature and flow that occur within this study period.

4.4 Results

The data are obtained from daily maximum water temperature measurements in degree Celsius, averaged over seven consecutive days to obtain the AWMT series. The series contains 18 years of data from 1963 to 1980. The period of analysis within each year is chosen from May to November (30 weeks). The series is plotted in Figure 4.2.

By minimizing the square error, the optimized sine function (Equation (4.2)) is first fitted to the water temperature data. The leave one out method outlined in the previous section is used and a specific sine function is estimated for each year left out. Values for a , b and t_0 are thus estimated. For instance, the following sine curve is estimated for water temperature of the year 1963:

$$TA(t) = 11.90 - 7.54 * \sin\left[\frac{2\pi}{52}(t + 8.99)\right] \quad (4.10)$$

Where $TA(t)$ is the water temperature at time t (week) starting from week 18 (first week of May) to week 47 (last week of November). The maximum temperature (peak of curve) for the annual component occurs on week 30 (last week of July) at average weekly maximum temperature of 19.44°C (Figure 4.3). It should be noted that the sine function is fitted for a 52 week period and is not limited to zero values. Equation (4.10) explains 94.34% of the water temperature variations. The estimation of the annual component is also calculated using empirical inter-annual mean value. However, this analysis has been tried and it did not lead to any improvement.

Once the annual component is removed from the observed water temperatures of the year left out, the first and the second-order Markov Chains are carried out using the residuals of the rest of data. Estimates coefficients (A_1 and/or A_2 ; Equation (4.3))

obtained from this analysis are thus used to model the residuals of the left out year. As example, the estimation of the coefficients for the years 1964-1980 yielded the following equation for the second-order Markov process of the year 1963:

$$Rw(t) = 0.39Rw(t-1) + 0.17Rw(t-2) + \varepsilon(t) \quad (4.11)$$

Given (4.10) and (4.11), the water temperature is calculated for 1963 using equation (4.1). Table 4.1 provides the coefficients of AR(1) and AR(2) models to be applied for each year removed. From this table, it is clear that annual component coefficients show somewhat erratic variations from one year to the next. This, however, is not the case for the residual component coefficients which may be found more stationary.

The previous model does not take into account the significant periodic autocorrelation of water temperature residuals. Hence, the AR(1) and AR(2) models are compared with their periodic equivalent, the PAR(1) and PAR(2).

In the following analysis, the AWMT series can be written in the matrix form (Equation (4.4)), with $\tau=1$ corresponding to the first week of May (week 18) and $\tau=30$ corresponding to the last week of November (week 47). The expected cyclic behaviour of the AWMT series is shown in Figure 4.2. Results of the Shapiro-Wilk test with a 95% significance level revealed that the normality assumption is accepted for each period ($0.11 < p < 0.99$). Therefore, these data are not transformed. The sample mean ($\hat{\mu}$), standard deviation ($\hat{\sigma}$) and autocorrelation coefficients of lag 1 ($\hat{\rho}_1$) and lag 2 ($\hat{\rho}_2$) are given in Table 4.2. From these results, it can be seen that the statistical characteristics vary from week to week, indicating the pertinence of including periodically varying parameters in the model.

Once the presence of periodic correlation has been detected, PAR (1) and PAR(2) models are adjusted to the data. From Table 4.2, it is clear that although some periods may be estimated with strictly lag 1 model, a significant number of periods show relatively high autocorrelations for both lag 1 and 2. The performance of periodic model is also assessed using the Leave-One-Out technique. Parameters values resulting from PAR(1) and PAR(2) models are used to compute the predicted AWMT data. As an example, PAR(1) parameter estimates are given in Table 4.3.

Regardless of the model used to predict water temperatures, all performance measures (RMSE, Bias and NSC) are calculated for water temperature values, not residuals. In general, both models seem to be adequate for simulating AWMT data. Table 4.4 gives the error estimates obtained for AWMT data of the two different models. For AR model, the mean RMSE between predicted and measured water temperatures is calculated as 0.81°C for AR(1) and 0.82°C for AR(2). The annual values of RMSE showed the range from 0.62°C to 1.19°C for AR(1) and from 0.62°C to 1.11°C for AR(2). The mean Bias error is centered on 0°C, with small annual values between -0.07°C and 0.09°C for AR(1) and between -0.06°C and 0.07°C for AR(2). The average NSC is of 0.94 for AR(1) and AR(2) models.

For the validation of PAR model, the mean RMSE values between measured and predicted water temperature are generally similar than those for AR model, with an increase of 0.09°C for PAR(1) and 0.11°C for PAR(2) (Table 4.4). The annual values of RMSE ranged from 0.54°C to 1.27°C for PAR(1) and varied between 0.52°C and 1.30°C for PAR(2). The mean Bias errors are centered on 0°C, while the annual values showed the ranges from -0.96°C to 0.65°C for PAR(1) and from -0.91°C to 0.64°C for PAR(2). Also, it can be seen from results presented in Table 4.4 that the

PAR models are more negatively biased during the years between 1977 and 1980.

The average NSC is of 0.92 for PAR(1) and PAR(2) models.

The slightly better performance of AR(1) and PAR(1) is an indication that parsimony should remain a guiding criterion when choosing a model. These results also suggest that there is no need to consider higher order AR and PAR models. The efficiency of the AR (1) and PAR(1) fit for some years, is illustrated in Figure 4.4. It can be seen that both models are generally able to reproduce the annual cycle and follow weekly data very closely.

The performance measures (RMSE and Bias) also varied from week to week and the overall results of the AR and PAR models showed that the highest values of RMSE are observed during the spring, with values of over 1°C (Figure 4.5). A higher value of RMSE in the spring is probably due to higher variation in water levels during that season. Inversely to interannual comparison, it is also observed from this figure that the Bias errors associated to PAR model are centered on 0°C, while those obtained by AR model showed a highest variation with the ranges from -0.59°C to 0.48°C for AR(1) and from -0.62°C to 0.45°C for AR(2). This is to be expected as the PAR model adds an important component to the corresponding AR model.

4.5 Discussion and conclusion

The prediction of water temperature data is important for the adequate management of water resources systems. Given the importance of water temperature to lotic ecosystem, it is essential to provide water resources managers with simple and good predictive models. For this purpose, the present study deals with the modeling of Average Weekly Maximum Temperature (AWMT) series using univariate stochastic approaches in Deschutes river (Oregon, USA). Two different models are then used to

model AWMT data: an Autoregressive (AR) model and a Periodic Autoregressive (PAR) model.

In the present study, the analysis of water temperatures by Autoregressive model consists first of calculating the long-term annual component by fitting a sinusoidal function to water temperatures. After the annual component is removed from the time series, the residuals series are used to model water temperatures using first and second order Markov Chains. However, this AR tools have their own drawback: since their application requires that a fixed sinusoidal function be fitted to the time series, it can be argued that this may result in non-stationary residuals from year to year. On the other hand, this approach does not take into account the periodic autocorrelation in the weekly water temperature series, which is significant. To alleviate this problem, the PAR model is used. However, the major drawback of PAR model is assuming that the distributions are normal; therefore, PAR model can not be applied to fit periodic time series of which density functions exhibit departures from Gaussian distributions. The PAR model is applied to the standardized data. To better compare models, the RMSE and Bias error are calculated for each validation year (Table 4.4) and when the years are predicted, these criteria were recalculated for each week (Figure 4.5).

The results based on the RMSE, Bias and NSC criteria indicated that both models are equally good in predicting AWMT on a relatively large river such as the Deschutes river. This is important, as AR models have been mostly developed for smaller water courses (e.g Caissie et al. (1998)). For the AR model, mean RMSE values are ranged from 0.81°C to 0.82°C. The mean Bias errors are centered on 0°C with smallest annual values. The AR model leads to a NSC of 0.94 on average. Compared to AR model, the PAR model showed an increase in mean RMSE less than 0.12°C, which is minimal. The mean Bias errors are also centered on 0°C, while the annual values

showed a highest variation. In terms of NSC, the average value is of 0.92. When studying the model performances from week to week, it is found that periodic model and non-periodic model both showed the highest values of RMSE during the spring, with values of over 1°C. A higher RMSE in the spring could be related to fluctuations in water levels during this flood season. Inversely to interannual comparison, the Bias errors associated to PAR model are more stationary than those obtained by the corresponding AR model. It would be therefore seen that a positive feature of PAR model is a decrease in Bias. This feature should be further investigated.

In conclusion, this study showed that AR (applied to the residuals) and PAR models behaved similarly, although with one site and shorter time series. However, the accuracy of the PAR model would be better if much longer continuous series of water temperature would exist. Such long series are difficult to find, but as temperature monitoring becomes increasingly important and easy to implement, PAR models may gain popularity. The PAR model seemed as an interesting alternative tool for modeling weekly water temperatures based on its performance and its ability to model periodicity in autocorrelations. It is clear that the PAR model will be computationally more intensive than that AR counterpart (Table 4.3) but this is the price to pay for getting a more suitable model. As no exogenous variables (i.e. air temperatures, streamflow) are incorporated, the use of the PAR model is limited in managerial decisions in natural streams and rivers. This is an important observation considering that several studies, such as those reported in the literature, have shown that the variations in water temperature are controlled by air temperature and/or flow. The suitability of PAR model suggests that future studies on water temperature modeling should focus on the incorporation of exogenous variables. Managers will also want to

test the forecast capability of the PAR models in real time, i.e. given a suite of parameters calibrated on all available data, provide an estimate of AWMT for the following weeks. This was not done as part of this initial study, because of the limited length of the time series and the large number of parameters required in the model. As well, other promising avenues include the comparison of PAR and deterministic models and the inclusion of the winter period in the model, for rivers that remain unfrozen during this season.

Acknowledgements

This paper is one part of a Ph. D thesis completed at National Institute of Scientific Research (INRS, University of Quebec). The research was supported by the National Sciences and Engineering Research Council of Canada (NSERC), and the United Nations Educational, Scientific and Cultural Organization (UNESCO). The authors would like to thank three anonymous reviewers whose comments significantly improved the manuscript.

References

- Bartholow, J. M. 1999. SSTEMP for Windows: The Stream Segment Temperature Model (Version 1.1.3). U.S. Geological Survey Computer Model and Help File.
- Bartolini, P., Salas, J.D., and Obeysekera, J.T.B. 1988. Multivariate periodic ARMA(1,1) processes. *Water Resour. Res.* **24**: 1237-1246.
- Bélanger, M., El-Jabi, N., Caissie, D., Ashkar, F., and Ribi, J. M. 2005. Estimation de la température de l'eau en rivière en utilisant les réseaux de neurones et la régression linéaire multiple (In press in *Rev. Sci. Eau*).
- Bjornn, T. C., and Reiser, D. W. 1991. Habitat requirements of salmonids in streams. *Am. Fish. Soc. Spec. Publ.* **19**: 83-138
- Box, G. E. P., and Jenkins, G. M. 1976. Time series analysis: forecasting and control. Revised edition. (2nd edn). Holden-Day, San Francisco.
- Brown, G.W., and Krygier, J.T. 1970. Effects of clear-cutting on stream temperature. *Water Resour. Res.* **6**: 1133-1139
- Caissie, D., El-Jabi, N., and St-Hilaire, A. 1998. Stochastic modelling of water temperatures in a small stream using air to water relations. *Can. J. Civ. Eng.* **25**: 250-260.
- Caissie, D., El-Jabi, N., and Satish, M.G. 2001. Modelling of maximum daily water temperatures in a small stream using air temperatures. *J. Hydrol.* **251**: 14-28.
- Cluis, D.A. 1972. Relationship between stream water temperature and ambient air temperature – A simple autoregressive model for mean daily stream water temperature fluctuations. *Nord. Hydrol.* **3**: 65-71.

Department of Environmental Quality (DEQ). 2003. Figure 130A: Fish use designations-Deschutes basin [online]. Available from http://www.deq.state.or.us/wq/standards/FishUseMapsFinal/FFigure130A_Deschutes.pdf.

Eaton, J.G., and Scheller, R.M. 1996. Effects of climate warming on fish thermal habitat in streams of the United States. *Limnol. & Oceanogr.* **41**: 1109-1115.

Hodgson, S, Quinn, T. 2002. The timing of adult sockeye salmon migration into fresh water: adaptations by populations to prevailing thermal regimes. *Can. J. Zool.* **80**: 542-555.

Hsia T. J., 1977. System Identification. Massachusetts: Lexington Books.

Hubbard, L.E., Herrett, T.A., Poole, J.E., Ruppert, G.P., and Courts, M.L. 1998. Water resources data, Oregon, water year 1997. Water-Data Report OR-97-1. U.S.Geol. Surv.

Janssen, P.H.M., and Heuberger, P.S.C. 1995. Calibration of process oriented models. *Ecol. Model.* **83**: 55-66.

Li, H. W., Lamberti, G. A., Pearson, T. N., Tait, C. K., Li, J.L., and Buckhouse, J. C. 1994. Cumulative effects of riparian disturbances along high desert trout streams of the John Day Basin, Oregon. *Trans. Am. Fisheries Soc.* **123**: 627-640.

Marceau, P., Cluis, D., and 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., and Erickson, T.R. 1998. Nonlinear regression model for weekly stream temperatures. *Water Resour. Res.* **34**: 2685-2692.

- Morin, G., and Couillard, D. 1990. Predicting river temperatures with a hydrological model. *Encyclopaedia of Fluid Mechanics*, 10, Gulf Publishing Company, Hudson, Texas.
- Nash, JE., and Sutcliffe, JV. 1970. River flow forecasting through conceptual models. Part A. Discussion of principles. *J. Hydrol.* **10**: 282–290.
- NOAA's National Marine Fisheries Service (NOAA Fisheries Service). 2004. Endangered species act: Status reviews and listing information [online]. Available from <http://www.nwr.noaa.gov/1salmon/salmesa/index.htm>
- Novales, A., and de Frutto, R.F. 1997. Forecasting with periodic models: A comparison with time invariant coefficient models. *Int. J. Forecasting.* **13**: 393-405.
- O'Connor, J.E., Grant, G.E., and T.L. Haluska. 2003. Overview of Geology, Hydrology, Geomorphology, and Sediment Budget of the Deschutes river Basin, Oregon. J.E. O'Connor and G.E. Grant eds. *A Peculiar river. Water Science and Application* 7. Am. Geophys. Union: 7-30. Available from http://www.fsl.orst.edu/wpg/pubs/OConnor_Paper.01.pdf
- Oliver, G.G., and Fidler, L.E. 2001. Towards a Water Quality Guideline for Temperature in the Province of British Columbia. Available from <http://wlapwww.gov.bc.ca/wat/wq/BCguidelines/temptech/>
- Osborn, D., and Smith, J. 1989. The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *J. Bus. Econom. Statist.* **7**: 117-127.
- Peterson, J. T., and Rabeni, C. F. 1996. Natural thermal refugia for temperate warm water stream fishes. *North. Am. J. Fish. Manage.* **16**: 738-746

- Pilgrim, J. M., Fang, X., and Stefan, H. G. 1998. Stream temperature correlations with air temperature in Minnesota: implications for climate warming, 1998. *J. Am. Water Resour. As.* **34**: 1109-1121.
- Quenouille, M. 1949. Approximate tests of correlation in time series. *J. Roy. Sta. Soc. B.* **11**:18-84
- Rasmussen, P.F, Salas, J.D. Fagherazzi, L., and Rassam, J.C. 1996. Estimation and validation of contemporaneous PARMA models for streamflow simulation. *Water Resour. Res.* **32**: 3151-3160.
- Salas, J.D., Delleur, J.W. Yevjevich, V., and Lane, E.L. 1980. Applied modelling of hydrological time series. Water Resources Publications, Littleton, Colorado.
- Salas, J. D. 1993. Analysis and modelling of hydrologic time series. Handbook of Hydrology. Maidment, D. R., McGraw-Hill.
- Shapiro, S.S., and Wilk, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*. **52**: 591-611.
- St-Hilaire, A., Morin, G., El-Jabi, N., and Caissie, D. 2000. Water temperature modelling in a small forested stream: implication of forest canopy and soil temperature. *Can. J. Civ. Eng.* **27**: 1095-1108.
- St-Hilaire, A., Morin, G., El-Jabi N., and Caissie D. 2003. Sensitivity analysis of a deterministic water temperature model to forest canopy and soil temperature in Catamaran Brook (New Brunswick, Canada). *Hydrol. Process.* **17**: 2033-2047.
- Thomas, H. A., and Fiering, M.B. 1962. Mathematical synthesis of Streamflow Sequences for analysis of river Basins by Simulation. *The Design of Water Resources Systems*, Ed. Maas, A. et al. Eds. Harvard University Press, Cambridge, Mass.

Ula, T.A., and Smadi, A.A. 1997. Periodic stationarity conditions for periodic autoregressive moving average processes as eigen-value problems. *Water Resour. Res.* **33**: 1929-1934.

U.S. Environmental Protection Agency (USEPA). 2003. Region 10 Guidance for Pacific Northwest State and Tribal Temperature Water Quality Standards. Available from <http://www.epa.gov>.

U.S. Geological Survey (USGS). 2004. Available from <http://waterdata.usgs.gov/nwis>.

Vannote, R. L., and Sweeney, B. W. 1980. Geographic analysis of thermal equilibria: A conceptual model for evaluating the effect of natural and modified thermal regimes on aquatic insect communities. *Am. Nat.* **115**: 667-695.

Vecchia, A.V. 1985. Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water. Resour. Bull.* **21**: 721-730.

Ward, J.V., and Stanford, J.A. 1982. Thermal responses in the evolutionary ecology of aquatic insects. *Annu. Rev. Entomol.* **27**: 97-117.

Table 4.1 Regression coefficients of AR(1) and AR(2) models, estimated by using Leave-One-Out technique for the Deschutes river.

	<i>Annual component</i>			<i>Residual component (second order Markov Chain)</i>		<i>Residual component (first order Markov Chain)</i>
Year of validation	<i>a</i>	<i>b</i>	<i>t₀</i>	<i>A₁</i>	<i>A₂</i>	<i>A₁</i>
1963	11.90	-7.54	8.99	0.39	0.17	0.47
1964	10.77	-8.60	8.70	0.39	0.16	0.46
1965	11.99	-6.11	9.35	0.39	0.14	0.45
1966	11.71	-7.13	8.30	0.39	0.17	0.46
1967	11.66	-7.53	8.69	0.39	0.16	0.47
1968	11.96	-6.99	9.52	0.40	0.17	0.49
1969	11.78	-7.24	9.21	0.39	0.17	0.46
1970	11.80	-6.91	9.57	0.42	0.14	0.49
1971	10.70	-8.17	8.98	0.37	0.16	0.44
1972	11.95	-6.14	9.96	0.38	0.15	0.45
1973	11.66	-7.49	9.34	0.39	0.16	0.47
1974	10.79	-7.64	8.50	0.38	0.16	0.45
1975	11.41	-7.68	8.52	0.40	0.17	0.47
1976	11.41	-7.68	8.52	0.40	0.17	0.47
1977	12.36	-7.69	9.37	0.38	0.17	0.45
1978	12.41	-7.59	9.58	0.38	0.16	0.45
1979	12.40	-8.25	9.28	0.37	0.14	0.42
1980	12.25	-7.57	8.82	0.39	0.16	0.46

Table 4.2 Sample mean, standard deviation and autocorrelation coefficients of lag 1

and lag 2 of average weekly maximum temperature series for the Deschutes river.

<i>Period</i> (τ)	$\hat{\mu}$ ($^{\circ}C$)	$\hat{\sigma}$ ($^{\circ}C$)	$\hat{\rho}_1(\tau)$	$\hat{\rho}_2(\tau)$
1	13.00	1.08	-0.23	0.20
2	13.78	0.89	0.17	0.14
3	14.41	1.33	0.36	0.23
4	14.95	0.91	0.40	-0.03
5	15.95	1.38	0.31	0.50
6	16.44	1.30	0.66	0.37
7	16.90	1.19	0.37	0.32
8	17.61	1.24	0.22	0.49
9	17.72	1.10	0.58	0.47
10	18.65	1.14	0.13	0.28
11	18.92	0.93	0.42	0.53
12	19.49	1.23	0.60	0.11
13	19.91	0.80	0.67	0.53
14	19.61	0.84	0.58	0.53
15	19.51	0.80	0.62	0.66
16	18.60	1.02	0.52	0.27
17	18.00	0.72	0.44	0.36
18	17.48	0.78	-0.09	-0.17
19	17.29	1.03	0.15	-0.12
20	16.46	0.90	0.67	-0.10
21	15.75	0.92	0.47	0.61
22	15.30	1.13	0.64	0.48
23	14.07	1.02	0.53	0.44
24	13.34	0.72	0.64	0.37
25	12.43	0.88	0.48	0.29
26	11.40	0.90	0.51	0.22
27	10.83	0.81	0.41	0.24
28	9.94	0.73	0.54	0.31
29	9.15	0.88	0.54	0.18
30	8.50	0.55	0.25	0.36

Table 4.3 Parameter estimates for PAR(1) model ($\hat{\phi}_t$) of average weekly maximum water temperature series for the Deschutes river.

Year of validation	Period(τ)																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1963	0.41	-0.17	0.18	0.25	0.05	0.37	0.33	0.11	0.38	-0.09	0.04	0.30	-0.04	0.38	0.30	-0.27	-0.09	0.18	0.17	0.24	0.11	0.32	0.47	0.52	0.58	0.41	0.52	-0.13	-0.13	-0.27
1964	0.39	-0.17	0.15	0.25	0.03	0.45	0.03	0.09	0.39	-0.09	0.14	0.36	0.00	0.37	0.35	-0.26	0.08	0.02	-0.04	0.19	0.05	0.13	0.35	0.48	0.48	0.44	0.45	-0.08	0.02	-0.14
1965	0.37	0.04	0.15	0.11	0.03	0.51	0.10	0.09	0.36	-0.19	-0.03	0.37	-0.10	0.37	0.33	-0.24	-0.23	0.03	0.14	0.23	0.25	0.36	0.41	0.51	0.46	0.34	0.46	-0.01	-0.17	-0.42
1966	0.57	-0.10	0.15	0.07	-0.10	0.33	0.14	0.07	0.41	-0.14	0.00	0.32	-0.10	0.43	0.32	-0.33	-0.24	-0.01	0.09	0.21	-0.04	0.34	0.65	0.52	0.46	0.41	0.47	-0.11	-0.19	-0.40
1967	0.47	-0.12	0.14	0.06	-0.10	0.35	0.14	0.04	0.46	-0.26	0.09	0.27	-0.10	0.41	0.29	-0.26	-0.11	0.11	0.05	0.24	-0.03	0.27	0.39	0.63	0.52	0.45	0.46	-0.09	-0.19	-0.22
1968	0.49	-0.17	0.16	0.06	0.03	0.34	0.14	0.15	0.35	-0.18	0.02	0.33	-0.05	0.44	0.32	-0.10	-0.08	0.11	0.06	0.24	0.16	0.25	0.38	0.50	0.46	0.43	0.52	-0.06	-0.13	-0.25
1969	0.52	-0.13	0.20	0.06	0.03	0.39	0.10	0.28	0.37	0.02	0.04	0.40	-0.14	0.41	0.27	-0.22	-0.14	0.15	-0.02	0.28	0.13	0.24	0.37	0.42	0.46	0.30	0.60	-0.16	-0.18	-0.29
1970	0.49	0.14	0.23	0.07	0.10	0.34	0.05	0.28	0.38	-0.02	0.02	0.33	0.02	0.39	0.27	-0.25	-0.18	0.15	-0.03	0.24	0.06	0.19	0.38	0.57	0.46	0.28	0.31	-0.13	-0.10	-0.29
1971	0.49	-0.20	0.18	0.07	0.36	0.40	0.04	-0.00	0.36	-0.23	0.01	0.37	0.36	0.54	0.58	-0.18	-0.22	0.16	0.12	0.22	0.08	0.05	0.42	0.48	0.45	0.28	0.41	0.06	0.08	-0.29
1972	0.49	-0.17	0.08	0.05	0.22	0.37	0.28	0.08	0.35	-0.24	0.01	0.48	0.21	0.69	0.37	-0.34	-0.25	0.20	0.07	0.22	0.03	0.06	0.37	0.52	0.35	0.30	0.47	0.08	-0.02	-0.18
1973	0.49	-0.19	0.40	0.07	-0.02	0.40	0.17	0.13	0.41	-0.06	0.31	0.46	0.15	0.52	0.37	-0.21	-0.21	0.24	0.10	0.30	0.04	0.34	0.37	0.52	0.57	0.45	0.46	-0.12	-0.32	-0.28
1974	0.44	0.02	0.15	0.14	0.04	0.48	0.15	0.06	0.34	-0.08	0.30	0.31	-0.04	0.40	0.28	-0.31	-0.15	0.30	0.10	0.38	0.10	0.28	0.37	0.51	0.51	0.40	0.41	-0.05	-0.33	-0.30
1975	0.43	-0.09	0.12	-0.02	0.02	0.38	0.15	0.03	0.32	-0.20	0.03	0.29	-0.07	0.42	0.29	-0.27	-0.20	0.10	0.09	0.10	0.10	0.25	0.40	0.52	0.48	0.40	0.42	-0.10	-0.15	-0.29
1976	0.43	-0.09	0.12	-0.02	0.02	0.38	0.15	0.03	0.32	-0.20	0.03	0.29	-0.07	0.42	0.29	-0.27	-0.20	0.10	0.09	0.10	0.10	0.25	0.40	0.52	0.48	0.40	0.42	-0.10	-0.15	-0.29
1977	0.47	-0.11	0.05	0.09	0.10	0.23	0.10	0.06	0.42	-0.12	-0.01	0.28	-0.07	0.28	0.22	-0.22	-0.15	0.09	0.10	0.16	0.11	0.33	0.43	0.47	0.37	0.49	0.43	-0.04	-0.16	-0.29
1978	0.50	-0.10	0.04	0.11	-0.12	0.21	0.12	0.05	0.35	-0.18	-0.07	0.18	-0.19	0.27	0.22	-0.13	-0.08	0.15	0.26	0.35	0.05	0.20	0.28	0.45	0.36	0.49	0.52	-0.05	-0.20	-0.11
1979	0.51	-0.10	0.03	0.03	-0.11	0.25	0.13	0.03	0.31	-0.16	-0.07	-0.00	-0.23	0.31	0.23	-0.30	-0.35	0.06	0.03	0.20	0.11	0.20	0.23	0.44	0.51	0.43	0.45	-0.15	-0.15	-0.20
1980	0.47	-0.09	0.19	0.11	0.05	0.40	0.15	0.13	0.36	-0.16	0.01	0.03	-0.23	0.42	0.19	-0.28	-0.36	0.17	0.06	0.19	0.13	0.22	0.22	0.56	0.47	0.47	0.48	-0.13	-0.15	-0.21

Table 4.4 Results of the cross validation of the periodic and non-periodic approaches expressed by the root mean square error, bias error and Nash coefficient and their range at Deschutes river

Year of validation	PAR(1)			PAR(2)			AR(1) *			AR(2) *			AR(1) †		
	RMSE	B	NSC												
1963	0.84	-0.18	0.94	0.83	-0.28	0.94	0.79	-0.03	0.94	0.83	-0.06	0.94	0.81	-0.18	0.94
1964	0.87	0.21	0.94	0.83	0.26	0.95	0.80	-0.03	0.95	0.80	-0.01	0.95	0.76	0.11	0.96
1965	0.91	0.26	0.89	0.91	0.24	0.90	0.72	-0.07	0.93	0.65	-0.03	0.95	0.80	0.18	0.92
1966	0.87	0.11	0.92	0.84	0.06	0.93	0.80	0.09	0.93	0.82	0.07	0.93	0.88	0.09	0.92
1967	0.75	-0.06	0.95	0.79	-0.08	0.94	0.70	0.05	0.95	0.71	0.04	0.95	0.77	0.01	0.94
1968	0.87	0.07	0.93	1.00	0.15	0.91	1.04	-0.00	0.9	1.06	0.01	0.9	0.93	0.04	0.92
1969	0.54	0.00	0.97	0.52	0.06	0.98	0.62	0.00	0.96	0.63	0.04	0.96	0.54	0.03	0.97
1970	1.11	0.22	0.89	1.20	0.14	0.87	1.19	0.01	0.87	1.11	-0.06	0.89	1.16	0.16	0.88
1971	1.15	0.57	0.91	1.30	0.64	0.88	0.94	0.05	0.94	0.95	0.07	0.94	1.00	0.39	0.93
1972	0.97	0.44	0.89	1.02	0.55	0.89	0.71	0.00	0.94	0.71	0.02	0.94	0.86	0.34	0.92
1973	0.75	0.13	0.95	0.77	0.19	0.95	0.83	0.02	0.94	0.85	0.05	0.94	0.74	0.07	0.95
1974	1.04	0.65	0.90	1.00	0.53	0.91	0.81	0.04	0.94	0.84	0.02	0.94	0.95	0.43	0.92
1975	0.66	0.09	0.96	0.73	0.08	0.95	0.84	-0.02	0.93	0.85	0.01	0.93	0.72	0.01	0.95
1976	0.66	0.09	0.96	0.73	0.08	0.95	0.84	-0.02	0.93	0.85	0.01	0.93	0.72	0.01	0.95
1977	1.11	-0.68	0.90	1.12	-0.71	0.90	0.85	-0.00	0.94	0.91	0.00	0.93	1.01	-0.41	0.92
1978	1.04	-0.57	0.91	1.06	-0.50	0.91	0.87	-0.02	0.94	0.90	-0.03	0.94	0.95	-0.39	0.93
1979	1.27	-0.96	0.88	1.23	-0.91	0.89	0.62	-0.01	0.97	0.62	-0.01	0.97	0.93	-0.63	0.94
1980	0.87	-0.55	0.93	0.87	-0.54	0.93	0.72	0.06	0.95	0.73	0.05	0.95	0.72	-0.30	0.95
Mean	0.90	-0.01	0.92	0.93	0.00	0.92	0.81	0.00	0.94	0.82	-0.01	0.94	0.85	-0.04	0.93
Range	[0.54; 1.27]	[-0.96; 0.65]	[0.88; 0.97]	[0.52; 1.30]	[-0.91; 0.64]	[0.87; 0.97]	[0.62; 1.19]	[-0.07; 0.09]	[0.87; 0.97]	[0.62; 1.11]	[-0.06; 0.07]	[0.89; 0.97]	[0.54; 1.16]	[-0.81; 0.43]	[0.88; 0.97]

*Annual component estimated by using a sine function

† Annual component estimated by using an inter-annual mean

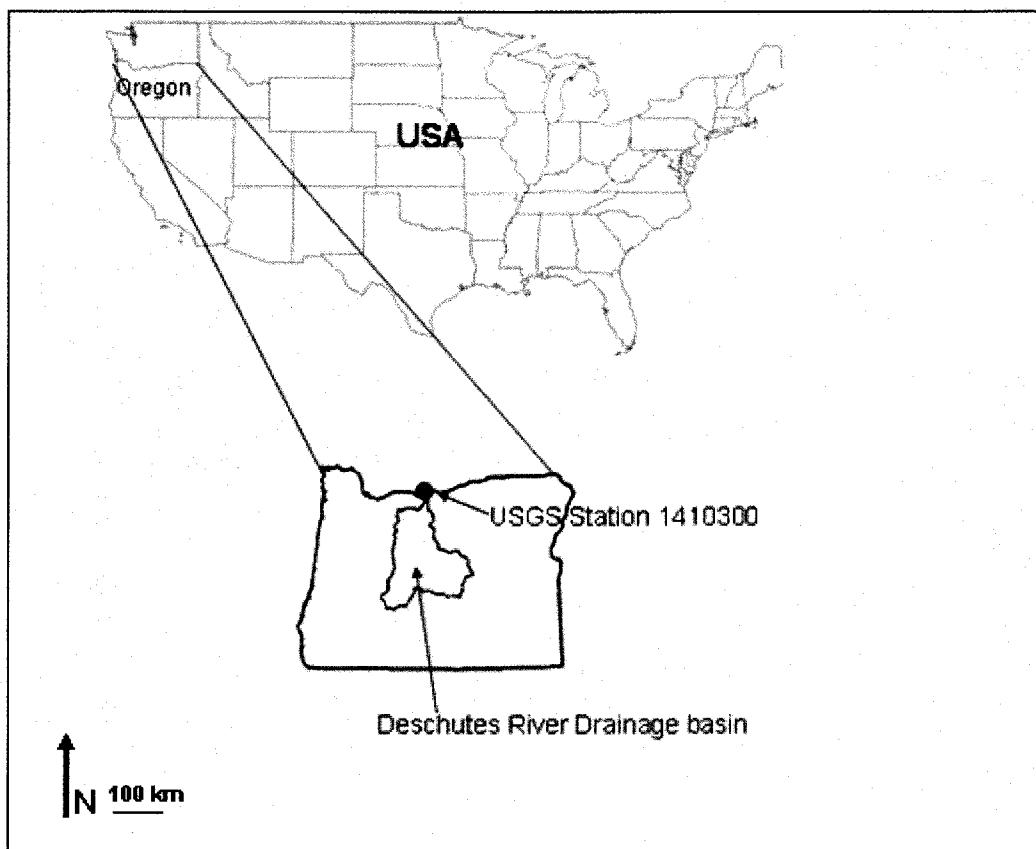


Figure 4.1 Location of the water temperature measuring site on Deschutes river

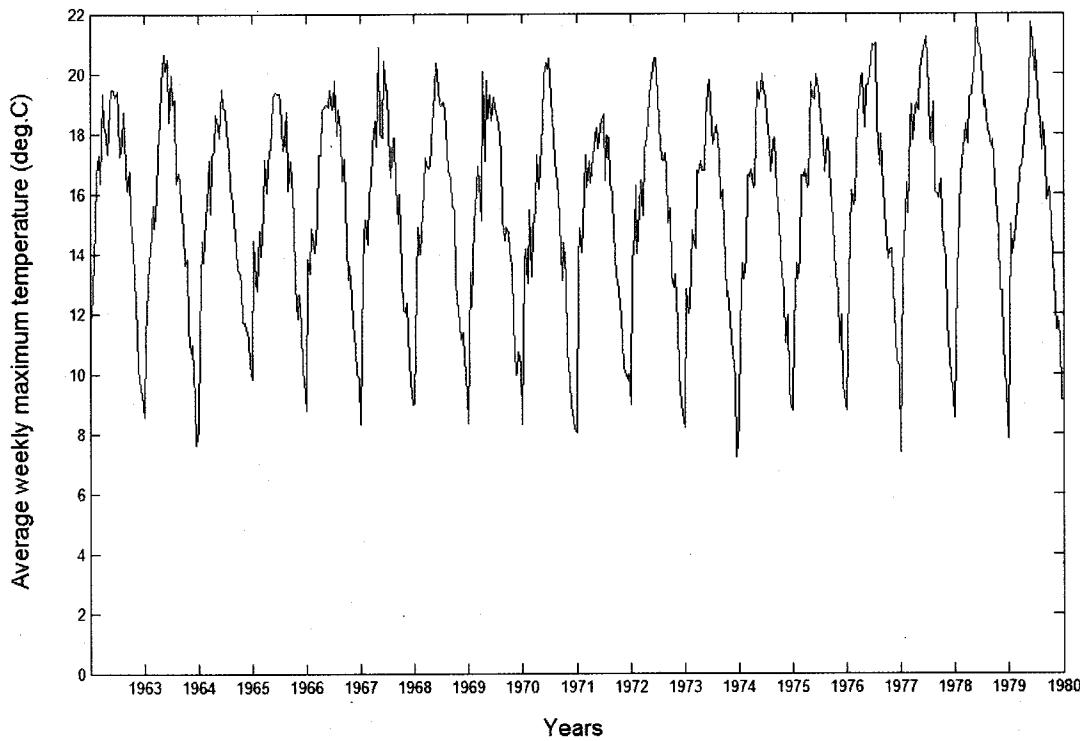


Figure 4.2 Average Weekly Maximum Temperatures at the gauging station, from 1963 through 1980

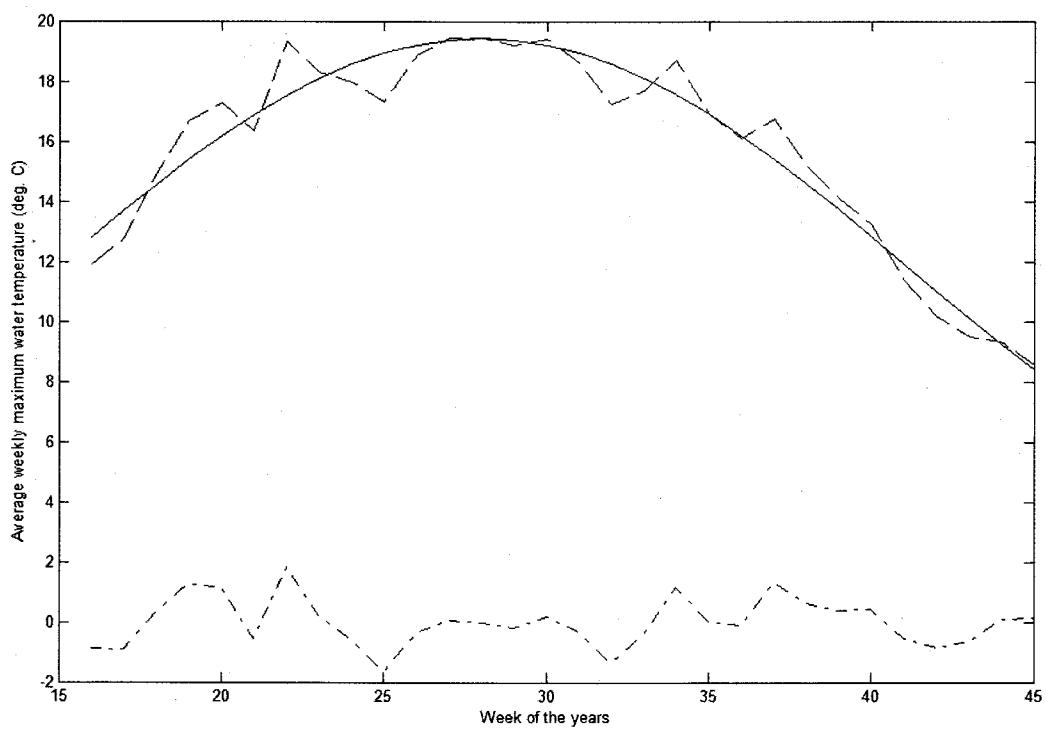


Figure 4.3 Average weekly maximum temperatures (1963) and annual component estimated by a sinusoidal function. The solid line represents sinusoidal function, the dashed line represents average observed weekly maximum temperatures and the dashed dot line represents the residuals.

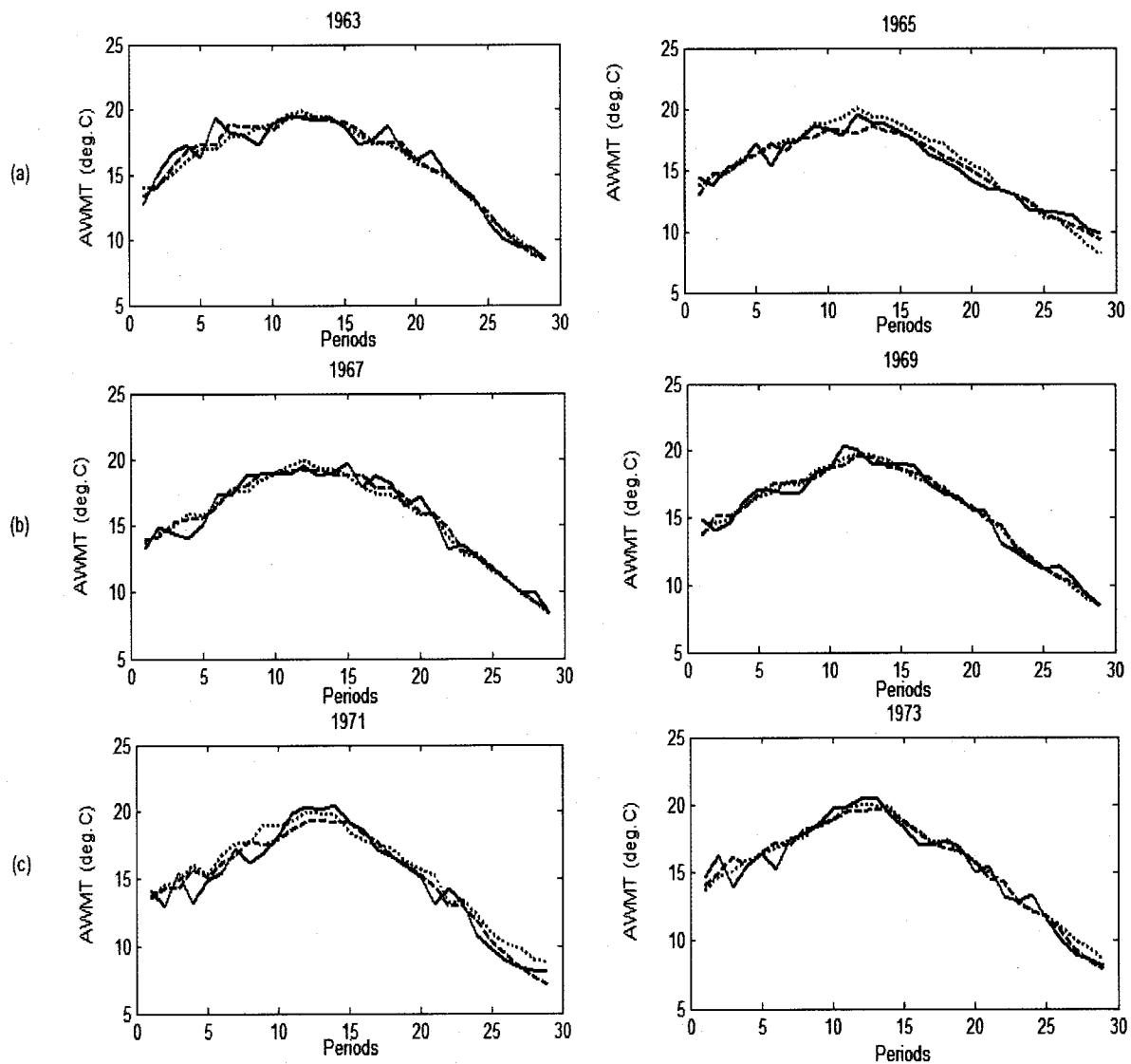


Figure 4.4 Results of the water temperature modelling using periodic and non-periodic approaches for some validation years. The solid lines represent observed temperatures, the points those predicted by PAR(1) and dashed lines those predicted by AR(1)

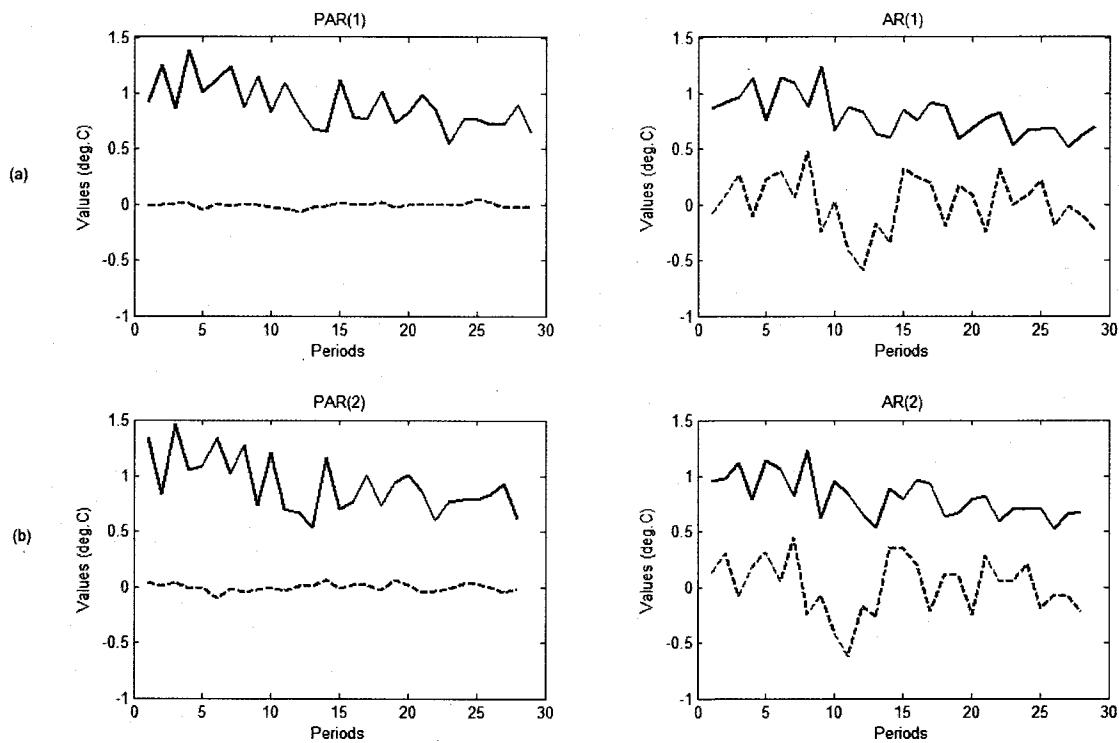


Figure 4.5 Results of the water temperature modelling using periodic and non-periodic approaches. The solid lines represent variation of root mean square error and the dashed line represent variation of bias error.

**CHAPITRE 5. MODÉLISATION DE LA
TEMPÉRATURE DE L'EAU : LES
MODÈLES VPP et PARX**

A comparison of non-parametric and parametric water temperature models: Case study of the Nivelle river (France)

By

Loubna Benyahya*,¹

André St-Hilaire¹

Taha B.M.J. Ouarda¹

Bernard Bobée¹

Jacques Dumas²

¹. Chair in statistical hydrology, INRS-ETE, Université du Québec

490 de la Couronne Street, Québec City, Québec. G1K 9A9 Canada

². INRA, Institut National de la Recherche Agronomique, UMR ECOBIOP, Pôle d'Hydrobiologie, Quartier Ibarron, 64310 St Péé sur Nivelle, France

Submitted to:

Hydrological Sciences Journal

February 2007

Corresponding author. Email: andre_saint-hilaire@inrs-ete.quebec.ca,

Tel. (418) 654-3113. Fax (418) 654-2600

Abstract

The present study deals with the modeling of stream water temperatures using statistical approaches with air temperature and streamflow as exogenous variables. These models are implemented in the Nivelle river, located in southern France. Two different models are used to model mean weekly maximum temperature data: a non-parametric approach, the k-Nearest Neighbours method (k-NN) and a parametric approach, the Periodic Autoregressive model with exogenous variables (PARX). The k-NN is a data-driven method, it consists of finding at each point of interest, a small number of neighbours nearest to this value, and the prediction is estimated based on these neighbours. The PARX is an extension of commonly used autoregressive models in which periods within the years are modeled separately. Different variants of air temperature and flow are used in the models developments. In order to test the performance of the above models, a Jackknife technique is used whereby model goodness of fit is assessed separately for each year. Results indicate that both models give good performances, but the PARX model should be preferred, due to its good estimation of the individual weekly temperatures and its ability to capture the periodic autocorrelation of the weekly water temperature series.

Key words: mean weekly maximum temperature, stochastic models, PARX, k-NN

Résumé

La température de l'eau est une variable très importante pour définir la qualité et la quantité d'habitat aquatique. En effet, étant des poissons à sang froid, les salmonidés par exemple, sont des espèces qui peuvent être défavorablement affectées par une augmentation de la température de l'eau. Cet article présente une modélisation statistique de la température de l'eau en utilisant la température de l'air et le débit comme variables explicatives. Deux nouveaux modèles de température de l'eau sont alors proposés et comparés, soit le modèle périodique autorégressif avec variables exogènes (PARX) et la méthode des k-voisins les plus proches (VPP). Les données utilisées dans la modélisation numérique sont les températures hebdomadaires de la rivière Nivelle (France) entre 1984-2004. La performance des deux modèles a été évaluée à l'aide de trois critères numériques: la racine de l'erreur quadratique moyenne, le biais et le coefficient de Nash. Ces critères ont été calculés pour chaque année en utilisant une technique de validation croisée de type "Jackknife". Plusieurs attributs (un total de 11) jugés significatifs sont considérés. Les résultats préliminaires ont montré que le modèle PARX et le modèle VPP présentent une performance similaire dans la simulation des températures hebdomadaires. Toutefois, le modèle PARX demeure le plus approprié, car de plus il préserve la persistance des séries périodiques, il offre une équation explicitant la relation entre la température de l'eau et les variables explicatives.

Mots-clés : températures hebdomadaires de l'eau, modèles statistiques, PARX, k-NN

5.1 Introduction

Water temperature is recognized as an important water quality parameter. It plays a major role in influencing many chemical and biological processes present in a stream, and hence influences the health and the distribution of aquatic ecosystems. Water temperatures can have adverse impacts on aquatic habitat especially when they are outside the optimal thermal range (Coutant, 1977). For example, warm waters have been observed to affect the mortality of trout (Lee & Rinne, 1980; Bjornn & Reiser, 1991). In addition, Lund et al. (2002) have shown that high water temperatures can have an impact on the development of juvenile salmonids. They showed that high summer water temperatures can cause significant protein damage and induce a heat-shock response as well.

Climate change is also expected to result in warmer air temperature and reduced summer flows which ultimately result in high water temperatures. These changes to the thermal regime of rivers could potentially impact on the distribution of fish as well as thermal habitat structure (Schindler, 2001). Eaton & Sheller (1996) have suggested that under a doubled atmospheric CO₂ scenario, increase in water temperatures would affect densities of a number of cold water fish species and they could decline by as much as 50% in the United States.

Water temperature is a variable for which changes are governed by the interaction of natural environment processes and human activities such as thermal pollution and deforestation (Brown & Krygier, 1970). Holtby (1988) examined the effects of extensive (41%) clearcuts on Carnation Creek in British Columbia. Although he found that every month exhibited an increase in mean monthly water temperature, ranging from 0.7°C in

December to 3.2°C in August. Therefore, water temperature modeling is a fundamental tool for the planning and management of water resources.

Existing water temperature models can be categorized in two types: deterministic models and stochastic models (Marceau *et al.*, 1986). Deterministic conceptual models are generally based on thermal budget calculations, which require numerous inputs (eg. physiographic, hydrological and meteorological parameters). Examples of deterministic models include the US Fish and Wildlife SNTEMP model (Bartholow, 1989), the SSTEMP model (Bartholow, 1999) and the CEQUEAU hydrological and water temperature model (Morin & Couillard, 1990; St-Hilaire *et al.*, 2003). Alternatively, statistical models are based on functions that estimate water temperature based on the correlation structure. This structure includes both autocorrelation (i.e. with using previous measurements to simulate present or future temperatures) and cross-correlation with independent variables (e.g. air temperature). Hence, statistical models tend to require fewer inputs than deterministic models. This can be an interesting feature for applications in data sparse areas.

Water-air temperature regression models have been successfully used for data relating to different time periods, including 2-hours (Stephan & Preud'homme, 1993) daily, weekly, monthly and annual means (e.g. Stephan & Preud'homme, 1993; Webb & Walling, 1993; Erickson & Stefan, 2000; Webb & Nobilis, 1997; Pilgrim *et al.*, 1998).

Other models have also been developed to assess relations between water temperature and air temperature. For example, the so-called stochastic models are generally applied for relatively small time steps (e.g. daily or sub-daily). In this approach, the annual component of water temperatures is first removed and time series models are then fitted

to water temperature residuals. These studies estimated an annual component in water temperatures by fitting a sinusoidal function and/or Fourier series to the data and a second order Markov process to the residuals. Early developments of this approach include those of Kothandaraman (1971) and Cluis (1972). More recently, Caissie *et al.* (1998, 2001) modeled mean and maximum daily water temperatures in Catamaran Brook (New Brunswick, Canada) using such models with air temperatures as the independent variable.

A more detailed review, leads to the following statements:

- (1) Most stochastic models only incorporate air temperature as an exogenous variable and do not include the potentially important influence of discharge on water temperature. Previous studies have shown that water temperature is also controlled by stream discharge. In fact, Webb *et al.* (2003) showed that in low flow situations (i.e. discharge below the annual median) in small catchments, variations in water temperature are more influenced by air temperature while flow plays a more important role in larger drainage basins during higher flow periods. Also, Neumann *et al.* (2003) developed a daily maximum regression model relating water temperature to flow and air temperature.
- (2) Stochastic models can be successfully applied to model water temperatures at different time scales, especially at smaller time steps. However, to assess changes in fish habitat and growth conditions, the weekly time scale is sometimes deemed more appropriate (Eaton & Scheller, 1996; Oliver & Fidler, 2001). Moreover, previous research showed that weekly and monthly averages of stream temperature and air temperature are more correlated than are daily values (Pilgrim *et al.*, 1998; Erickson &

Stefan, 2000). For these reasons, the analysis presented herein is conducted on weekly values.

The statistical characteristics of hydrologic series are important deciding factors in the selection of the type of model to be used. In the case of seasonal series, the most commonly used approach (e.g. Cluis, 1972; Caissie *et al.*, 1998, 2001) requires deseasonalization prior to applying the stationary models (e.g ARMA). However, this process does not guarantee that the seasonality of the data is completely removed as the correlation structure of the series may be dependent on the period. In these cases, the resulting model is misspecified (Tiao & Grupe, 1980). An alternative approach to model periodicity in autocorrelations is to consider periodic models. PAR (periodic autoregressive) and PARMA (periodic autoregressive-moving average) models are variants of ARMA models (Box & Jenkins, 1976) that allow periodic parameters. Periodic models have been widely and successfully used in econometrics applications (Osborn & Smith, 1989; Novales & de Frutto, 1997) as well as in the field of hydrology (Vecchia, 1985; Bartolini *et al.*, 1988, Ouarda *et al.*, 1997; Benyahya *et al.*, 2006).

An alternative approach, seldom used in water temperature modeling, consists of developing non-parametric models. One such approach is the k-nearest neighbours (k-NN) method (Yakowitz & Karlsson, 1987). This approach is based on the principle that if time series are sufficiently long, proper estimators of certain water to be simulated or forecasted exist in the past record. The approach is labelled as non-parametric because it is based on calculating a certain attribute without using a parameterised statistical model. The method consists in searching among past observations for the k events which are most similar to the present situation. A prediction is then built from the water

temperatures which are associated with these k events. In a hydrology context, the technique has been used to analyse rainfall-runoff processes and has been advantageously compared with autoregressive moving average models with exogenous inputs (ARMAX) (Karlsson & Yakowitz, 1987; Yakowitz & Karlsson, 1987). Galeati (1990) evaluated and compared the performances of the k-NN and autoregressive model with exogenous inputs (ARX), to predict daily mean discharge. The performances of the two methods are substantially equal.

Since these two approaches (i.e. a parametric Periodic ARX model and a non-parametric k-NN approach) have never been adapted to model water temperature, the present study aims to predict the weekly mean water temperature in a small stream using these tools.

More specifically, the present study has two objectives: (1) to evaluate the possibility of employing the k-Nearest Neighbours method (k-NN) and (2) to compare its performance with that of a PARX model with air temperature and streamflow as exogenous inputs.

These two approaches are implemented using data from a river where salmonid habitat is important namely the Nivelle river in South-western France.

5.2 Methods

5.2.1 k-Nearest Neighbours (k-NN)

In the present study, the key steps in the k-NN algorithm are as following:

Step 1. Compile a feature vector: the vector X consists of values of the selected input attributes (e.g lagged water temperature, lagged air temperature and streamflow data) for which we are trying to find the k-Nearest Neighbours;

Step 2. Find the weighted sum of the attributes: since the scales of water and air temperature units are different than that of streamflow, the weighted attributes can be generalized as a weighted standardized norm (N);

$$N = \sum_{i=1}^{N_{\text{attributes}}} w_i X_i \quad (5.1)$$

Where w_i are the optimized weights and X_i are the vectors of standardized (i.e. subtract mean value and divide by the standard deviation) values of the selected attributes. In this study, weights are varied between 1 and 1000 and all possible combinations are tried by increments of 100.

Step 3. Calculate the Euclidean distance between the norm of the period of interest (e.g., day, week) and the norm of all other available data: For two norms (N_{j1}, N_{j2}) calculated using vectors X_{j1} for the period of interest and X_{j2} ($j_2 = 1, \dots, j_1 - 1, j_1 + 1, \dots, m$) for the m other periods in the data base, the Euclidean distance (δ) is defined as:

$$\delta(N_{j1}, N_{j2}) = |N_{j1} - N_{j2}| = \sum_{i=1}^{n_{\text{attributes}}} w_i * |X_{i,j1} - X_{i,j2}| \quad (5.2)$$

Step 4. Sort the distances δ in ascending order, and retain only the first k nearest neighbours: In this study, k was limited to a maximum value of 3. The strategy for choosing the optimal k is to try several successive values of k (e.g., 2, 3 and 4) and to select the combination for which the model gives the best prediction.

Step 5. Assign weight to each of the k neighbours, thus the predicted value of the final output is computed as a weighted sum of the values of neighbouring observations;

Step 6. Steps 1 through 5 are repeated for each of the time step;

Obviously, this approach can be computationally intensive. However, applications indicate that a solution, which gives good results, can be obtained in a relatively short time.

5.2.2 Periodic Autoregressive model with exogenous variables (PARX)

Parametric models, on the other hand, can predict values without seeing in the historical record past. One of the used parametric models for periodic time series modeling is the periodic ARX model. Following the earlier work of Salas (1985), let us consider $Tw_{v,\tau}$, $Ta_{v,\tau}$ and $Q_{v,\tau}$, the time series of water temperature, air temperature and streamflow respectively. The subscripts v and τ denote the year and the season (e.g month, week), respectively, where $\tau = 1, \dots, \omega$ with ω being the number of seasons in the year. A PARX model representing the water temperature series may be written in the following form:

$$Tw_{v,\tau} = \sum_{i_1=1}^{p_1} \phi 1_{i_1,\tau} Tw_{v,\tau-i_1} + \sum_{i_2=1}^{p_2} \phi 2_{i_2,\tau} Ta_{v,\tau-i_2} + \sum_{i_3=1}^{p_3} \phi 3_{i_3,\tau} Q_{v,\tau-i_3} + \varepsilon_{v,\tau} \quad (5.3)$$

and $\tau > i_1, i_2, i_3$

Where, $\phi 1_{i_1,\tau}$, $\phi 2_{i_2,\tau}$ and $\phi 3_{i_3,\tau}$ are periodic parameters, p_1 , p_2 and p_3 are the lags of water temperature, air temperature and streamflow respectively. $\varepsilon_{v,\tau}$ is the error term which is excluded from the comparative study. The model by construction preserves the periodic autocorrelation structure. In the present study, the time series of each of the variables are first standardized. Subsequently, to obtain the final estimated value of water temperature, a back transformation is applied.

Once a model is identified, its parameters may be estimated for each time step (period) from the data by a number of techniques such as the method of moments, the least squares method or the method of maximum likelihood (Salas, 1993). In this study, the least squares method is used. Generally, estimates obtained by this method are consistent, unbiased and efficient (Hsia, 1977). Furthermore, modeling time series generally requires that the underlying series follows a normal distribution (Salas, 1993). Otherwise, the original series should be transformed to a Gaussian shape via log, square root or Box-Cox transformations. The predicted values from the model are then back-transformed into the original space. In the present study, the Shapiro-Wilk test (Shapiro and Wilk, 1965) is used to test the normality of the data. This test is appropriate when the sample size is lower than 50.

5.2.3 Model evaluation and validation

To compare the predicted (P) and the observed (O) water temperatures, the root mean square error (RMSE, Equation (5.4)), the bias (Bias, Equation (5.5)) and the Nash-Sutcliffe coefficient of efficiency (NSC Equation (5.6)) are used (Nash & Sutcliffe, 1970; Janssen & Heuberger 1995). They are respectively given by:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (5.4)$$

where

n = number of season water temperature observations

The RMSE defined as the square root of the sum of the variance and the square of the bias, is calculated to inform on the average magnitude of the water temperature errors.

This criterion is often used in water temperature modeling studies (e.g., Ahmadi-Nedushan *et al.*, 2007, St-Hilaire *et al.*, 2003, Caissie *et al.*, 1998, 2001).

The Bias error is computed simply as the sum of the differences between predicted and observed values divided by n :

$$Bias = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (5.5)$$

Efficiency of fit for both the k-NN method and the PARX model is determined with the NSC:

$$NSC = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (\bar{O} - O_i)^2} \quad (5.6)$$

where

\bar{O} = mean weekly water temperatures for the period n

The NSC compares model performances with the sample mean. A NSC ranges from minus infinity (poor model: model does not perform better than using the sample mean) to 1 (perfect model).

To test the performance of the above models, the Leave-One-Out (Jackknife) validation technique (Quenouille, 1949) is used. This method consists of removing one year from the data and estimating the model using the rest of data set. The performance is judged by testing the selected model on the data left out. Thereafter, the performance criteria (Equations 5.4 to 5.6) are calculated per year and an average of the annual values is also calculated. All models and performance estimation are done using Matlab 6.5.

5.3 Case study area and data set

Time series of water temperature are obtained from the Nivelle river, located in South-western France (Figure 5.1). It is a relatively small river with a drainage area of 238 km². It takes its source in Spain into the Bay of Biscay at Saint-Jean-de-Luz after travelling 39 km. This catchment is well known for its population of Atlantic salmon (*Salmo salar*) (Dumas & Prouzet, 2003).

The oceanic climate, mild and wet (1700 mm/year in St-Pée-sur-Nivelle), provides a mean annual discharge of 5.4 m³/s downstream from the confluence of the main tributary, the Lurgorrieta.

Water temperatures are measured daily at one stream location (Ibarron; Figure 5.1) by the National Institute of the Agronomic Research (INRA) using Tidbit, Minilog Vemco and Jules Richard temperature data loggers (accuracy \pm 0.2°C, \pm 0.3°C and \pm 0.4°C, respectively). Water temperature data are available from 1984 to 2004. The data used in the present study are Mean Weekly Maximum Temperatures (MWMT), which are calculated from daily maximum temperatures (mean of 7 consecutive calendar days). During the year 1994, to fill in some missing values from 18th July to 31st August, the data are interpolated using an interannual mean. The present study focuses on the period within every year spanning from May 1st to November 31st (30 weeks). Outside this period (i.e. December-April), the water temperature variations are typically less than 5°C on the Nivelle river (Figure 5.2 (b1)). It was therefore decided to study the period during which the largest temperature variations (magnitude between 10 and 12 °C during the warmer period of May to November; Figure 5.2 (b2)). One of the reasons for this selection is the fact that thermal stress for salmonids are much more likely to occur

during the warmer period, when maximum water temperature exceeds a certain tolerable limit. Indeed, Hodgson and Quinn (2002) have demonstrated that during the spawning season, some species of pacific salmon will suspend their spawning activity when water temperature exceeds 19°C. Moreover, Crozier and Zabel (2006) showed that juvenile Chinook salmon survival is negatively correlated with summer temperature. Daily air temperatures are obtained from the Biarritz airport weather station which is located 13 km north of the water temperature station. Streamflow for the same period are measured at Nivelle streamflow station (Figure 5.1), which is located in the downstream portion of the drainage area.

5.4 Results

Time series of water temperature, air temperature and flow used in this study are plotted in Figure 5.3. Water temperature data in the Nivelle river showed a good correlation with air temperatures and streamflow ($r = 0.93$ and -0.48 , respectively). This result indicates the relevance of including air temperature and streamflow as exogenous variables in the water temperature modeling. The results are organized as follows: first, k-NN results are shown and are thereafter compared to that of the PARX model with air temperature and flow as exogenous predictor variables.

5.4.1 k-NN model

In this analysis, a total of eleven candidate independent variables, all derived from the three observed variables (T_w , T_a and Q), are initially considered (Table 5.1). Attributes are defined as T_w , T_a and Q when referring to water temperature, air temperature and streamflow respectively. The potential inputs include autoregressive terms, weekly

degree-days, lagged air temperature, lagged streamflow and relative flow change. Relative flow change is used as opposed to direct flow measurements because it was shown by Ahmadi-Nedushan et al. (2007) to yield better results for daily water temperature models. In the present study, the weekly degree-days are defined as the cumulative sum of the air temperature over the period of analysis.

In order to select the best set of these explanatory variables, Principal Component Analysis (PCA) is used. PCA is a multivariate data analysis method, which is performed on the correlation matrix. The basic idea of PCA is to find some correlated variables by reducing the redundancy in the information provided by these variables. The method generates a new set of variables, called principal components. Each principal component is a linear combination of the original variables, that are orthogonal and that maximise explained variance. The first principal component (PC1) contains the greatest variation; the second one (PC2) contains the amount of variation unexplained by the first, and so on. Prior to using PCA, in order to avoid scale problems associated with variables measured by different units, the original variables are standardized into numbers with zero mean and unit standard deviation. The result of the analysis shows that the first two principal components explain more than 65% of the variance in the data, leading us to take a closer look at them. Figure 5.4 shows the plot of attributes in the plane of the first two components. This plot reveals that, as expected, most temperature attributes ($Tw0$, $Tw1$, $Tw2$, $Tw3$, $Ta1$, $Ta2$, and $Ta3$) are aggregated at one end, while flow attributes ($Q1$, $Q2$ and $Q3$) are at the other extreme. Consequently, each group of variables is highly correlated. Surprisingly, $Ta4$ which is the weekly degree-days of air temperature seems

to be anomalous and closely associated with PC2; therefore, it can be helpful to study this exceptional point also.

Figure 5.4 is used to avoid selecting redundant attributes. The selection process consists of choosing, from each group, the attributes which are representative of a cluster on Figure 5.4 and are deemed suitable for predicting MWMT. The simplest way to determine such a set is to calculate the correlation between $Tw0$ and each attributes independently. The correlation is calculated by taking, for each year, the correlation coefficient between the aforementioned variables. Figure 5.5 shows the correlation boxplots. Box boundaries represent the interquartile range and the whiskers represent the 10th and 90th percentile. The median for each correlation coefficient is indicated by the black center line. The extreme values are shown as individual points. From Figures 5.4 and 5.5, $Tw1$, $Tw2$, $Ta1$, $Ta4$ and $Q1$ are found to be the best attributes. Among the selected attributes, as expected, lag1 air ($Ta1$) and water temperature ($Tw1$) have the highest correlations with the present water temperature ($Tw0$). Median correlation coefficient values are 0.86, 0.73, 0.83, -0.55 and -0.38 for $Tw1$, $Tw2$, $Ta1$, $Ta4$, and $Q1$, respectively. Therefore, these five attributes are considered in the following analysis. Result of the analysis shows that there is a negative relationship between water temperature ($Tw0$) and flow of the past week ($Q1$), which can be explained by the effect of thermal inertia of the river [i.e. a time lag between flow and water temperature variations caused by the ability of a river to conduct heat].

In order to build the best k-NN model, single attributes are first tested independently. Then, the so-called best subset approach is used, i.e. all possible combinations of attributes are considered and the subset yielding the best results is selected as the final

model. The k-NN models which included one, two, three, four and five attributes are denoted k-NN1, k-NN2, k-NN3, k-NN4 and k-NN5. In order to compare the performance of these models, we used the three different criteria described in section II.3, i.e. the RMSE, Bias and NSC. Results are presented in Table 5.2. These results indicate that the performance associated with one attribute models (k-NN1) is modest. The mean values of RMSE ranged from 2.52°C to 4.02°C. All attributes slightly underestimate the water temperatures with annual mean bias values between -0.59°C and -0.53°C. The interannual average NSC ranged from -0.83 to 0.27.

The performance of subsets k-NN2, k-NN3, k-NN4 and k-NN5 is also compared (Table 5.2). These results indicate that these models outperformed the k-NN1. The interannual mean values of RMSE varied between 1.20°C and 2.14°C. Also, the NSC values calculated for all models are higher (> 0.49) than the value reported for k-NN1. It can be seen from results presented in Table 5.2 that interannual RMSE values are similar for k-NN3, k-NN4 and k-NN5 (between 1.20°C and 1.34°C) among which the model k-NN4 provides the lowest RMSE value (1.20°C). Small interannual mean bias values are observed for these models ($-0.02^{\circ}\text{C} < \text{bias} < 0.01^{\circ}\text{C}$) and NSC values are above 0.80. The results indicated that k-NN method performs well in prediction of weekly water temperatures of the Nivelle river. The k-NN4 model has the best performance among the models based on the lowest RMSE and bias and the maximum NSC values. This model achieves this level of performance with *Tw1*, *Tw2*, *Ta1* and *Q1* as exogenous variables. These results indicated also that streamflow (*Q1*) can be included in the model as an independent variable, and that considering streamflow in k-NN models improves performance; however, the improvement is modest ($<0.12^{\circ}\text{C}$)

5.4.2 PARX model

The models developed in the first part of the study have excellent descriptive ability; however they don't take into account the significant periodic autocorrelation of water temperatures. A plot of the original data, excluding December-April periods, shows the cyclic behaviour of the weekly air and water temperatures (Figure 5.3). The nonstationarity of water temperatures is apparent since the mean, standard deviation and autocorrelation functions vary from week to week (Figure 5.6). Therefore, considering a periodically model is appropriate. Hence, in the second part of article, the k-NN method is compared to the PARX model.

Before applying the PARX model, it is necessary to ensure that the original data are normally distributed. Results of the Shapiro-Wilk test with a 5% significance level revealed that the normality assumption is accepted for each period ($0.08 < p < 0.99$).

Therefore, these data are not transformed.

The second part of the study is devoted to testing and the development of PARX models that include air temperature and flow. Over 15 different variants of water temperature, air temperature flow were initially tested in the model development. However, only the subset of the most important variables (in terms of the performance obtained by the models) are shown in Table 5.1 and used in the final analysis to identify the best PARX models which have good descriptive ability. As in k-NN method, the PARX model which included two, three, four and five variables are denoted respectively PARX2, PARX3, PARX4 and PARX5. MWMT series are broken down into 30 periods with $\tau = 1$ corresponding to the first week of May and $\tau = 30$ corresponding to the last week of November.

Results for the performance evaluation are presented in Table 5.2. This Table indicates that all models perform well, with RMSE values between 1.58°C and 1.74°C among which the PARX3 provides the lowest value (1.58°C). All models have very similar interannual NSC means ($0.62 < \text{NSC} < 0.70$). PARX models are found to have small relative mean bias (centered on 0°C). The range of annual biases are also similar for all PARX models (Table 5.2), among which the annual bias values of PARX3 ranged from -0.99°C to 1.50°C. While the models performed comparably and the differences are minimal ($\leq 0.16^\circ\text{C}$), model PARX3 is selected as the best model based on minimum RMSE and maximum NSC coefficient. This level of performance is achieved with two autoregressive terms ($Tw1$, $Tw2$) and one exogenous independent variable: weekly degree days ($Ta4$). Model PARX3 is defined as follows:

$$Tw_{v,\tau} = \phi 1_{1,\tau} Tw_{v,\tau-1} + \phi 1_{2,\tau} Tw_{v,\tau-2} + \phi 2_{1,\tau} Ta4_{v,\tau-1} \quad (5.7)$$

and $\tau \geq 2$

It should be noted that the inclusion of streamflow did not bring improvement (PARX4).

As an example, parameter estimates for the validated year 2004 are displayed in Table 5.3. In order to illustrate the general quality of those results, Figure 5.7 shows the time series of predicted values by the k-NN4 and PARX3 models versus observed water temperatures for some verification years, which have the lowest values of RMSE. It can be seen that both models are suitable to capture the seasonal variation of water temperature and provide relatively good estimates of measured values.

5.5 Discussion and conclusion

This work was motivated by the need to develop a robust model to predict mean weekly maximum temperature with air temperature and streamflow as exogenous inputs. To this end, two stochastic approaches are used to relate water temperatures to air temperatures and streamflow in the Nivelle river, located on South-western France. The first model is a non-parametric k-NN method which is compared to the parametric model of PARX. Because of its non-parametric nature (data driven method), the k-NN method does not make any assumption about the underlying statistical distributions. This approach does not take into account the periodic autocorrelation in the weekly water temperature series, which is significant. Moreover, k-NN method cannot be easily extrapolated outside of the temperature range encountered in the dataset used to calibrate it. The PARX can thus be seen as more appropriate. The most striking feature of the PARX model, widely used in practice, is that it preserves the periodic correlation structure in the seasonal data. However, the disadvantage of this model is assuming that the distributions are normal; therefore, the PARX model can not be applied to fit periodic time series for which density functions exhibit departures from Gaussian distributions. In this case, k-NN method is a good alternative. However, this drawback is minor when considering temperature time series, which seldom depart from normality.

Different variants of water temperature, air temperature and streamflow are used in model development. PCA and boxplots of correlation coefficients showed that the best attributes included lagged water temperature (lags 1 and 2 weeks) and lagged air temperature (lag 1 and weekly degree days) and streamflow (lag 1). The results (Table 5.2) indicated that the best models (PARX3 and k-NN4) require respectively, one input

(weekly degree days of air temperature) and both lagged air temperature (lag 1) and streamflow (lag 1). However, the improvement in k-NN performance by including streamflow variable is modest.

A major conclusion of this first comparison, although with only one study site and relatively short time series (i.e, 21 year), is that the k-NN method, which is a data-driven method, may provide an accurate result in terms of RMSE, NSC and bias error. It should be remembered that for all multi-attribute k-NN models, the weight of the neighbours was also optimised. In agreement with the conclusion reported by Galeati (1990), the result shows the substantial equivalence of the two methods. However, the PARX model may be preferred, due to its good estimation of the individual weekly temperatures and its ability to capture the periodic autocorrelation of the weekly water temperature series. Once the PARX model (Equation (5.7)) is developed for this site, it could be used to estimate future water temperatures using weekly degree days of air temperature and lagged water temperature. In so doing, it might be a greater practical interest in utilizing these predicted water temperatures to anticipate the health of aquatic habitat.

In this study we used parametric and non-parametric approaches to model water temperature. There are still many methods which have been used successfully in many forecasting applications in engineering and science; however they have not yet been fully explored in water temperature modeling. For further research, it is necessary to explore other non-parametric methods to find suitable applications in water temperature modeling. To fully validate the preliminary conclusions of the present study, applications to other sites with longer time series would be useful.

Acknowledgements

This paper is one part of a Ph. D thesis completed at National Institute of Scientific Research (INRS, University of Quebec). The research was supported by the National Sciences and Engineering Research Council of Canada (NSERC) and Hydro-Quebec.

References

- Ahmadi-Nedushan, B., St-Hilaire, A., Ouarda, T.B.M.J., Bilodeau, L., Robichaud, É., Thiémonge, N. & Bobée, B. (2007) Predicting river water temperatures using stochastic models: case study of the Moisie river (Québec, Canada). *Hydrol. Processes.* **21**, 21-34.
- Bartholow, J.M. (1989) Stream Temperature Investigations: Field and Analytic Methods. Instream Flow Information. US Fish and Wildlife Service Paper No. 13.
- Bartholow, J.M. (1999). SSTEMP for Windows: The Stream Segment Temperature Model (Version 1.1.3). US Geological Survey computer model and help file.
- Bartolini, P., Salas, J.D. & Obeysekera, J.T.B. (1988) Multivariate periodic ARMA(1,1) processes. *Water Resour. Res.* **24**, 1237-1246.
- Benyahya, L., Saint-Hilaire, A., Ouarda, T.B.M.J., Bobée, B. & Nedushan, B. 2006. Modeling of water temperatures based on stochastic approaches: Case study of Deschutes river (Oregon, USA). *J. Environ. Eng. Sci* (In press)
- Bjornn, J.R. & Reiser, G.W. (1991) Habitat Requirements of Salmonids in Streams. Influences of Forest and Rangeland Management in Salmonid Fishes and Their Habitats. American Fisheries Society, Bethesda, Maryland. pp. 73-158 in Special Publication 15.
- Box, G.E.P. & Jenkins, G.M. (1976) Time series analysis: forecasting and control. Revised edition. (2nd edn). Holden-Day, San Francisco.
- Brown, G.W. & Krygier, J.T. (1970) Effects of clear-cutting on stream temperature. *Water Resour. Res.* **6**, 1133-1139

Caissie, D., El-Jabi, N. & St-Hilaire, A. (1998) Stochastic modelling of water temperatures in a small stream using air to water relations. *Can. J. Civ. Eng.* **25**, 250–260.

Caissie, D., El-Jabi, N. & Satish, M. (2001) Modeling of maximum daily water temperatures in a small stream using air temperatures. *J. Hydrol.* **251**, 14-28.

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

Coutant, C.C. (1977) Compilation of temperature preference data. *J. Fish. Res. Board Can.* **34**, 739-747

Crozier, L. & Zabel, R.W. (2006) Climate impacts at multiple scales: evidence for differential population responses in juvenile Chinook salmon. *J. Anim. Ecol.* **75**, 1100-1109

Dumas, J. & Prouzet, P. (2003) Variability of demographic parameters and population dynamics of Atlantic salmon (*Salmo salar* L.) in a southwest French river. *ICES J. Mar. Sci.* **60**, 356-370

Eaton, J.G. & Sheller, R.M. (1996) Effects of climate warming on fish thermal habitat in streams of the United States. *Limnol. & Oceanogr.* **41**, 1109-1115.

Erickson, T.R. & Stefan, H.G. (2000) Linear air/water temperature correlations for streams during open water periods. *ASCE, J. Hydrol. Eng.* **5**, 317–321.

Galeati, G. (1990) A comparison of parametric and non-parametric methods for runoff forecasting. *Hydrol. Sci. J.* **35**, 79-94.

Hodgson S. & Quinn, T.P. (2002). The timing of adult sockeye salmon migration into fresh water: adaptations by populations to prevailing thermal regimes. *Can. J. Zool.* **80**: 542-555.

Holtby, L.B. (1988) Effects of logging on stream temperatures in Carnation Creek, British Columbia, and associated impacts on the coho salmon (*Oncorhynchus kisutch*). *Can. J. Fish. Aquat. Sci.* **45**, 502-515.

Hsia, T. J. (1977) System Identification. Massachusetts: Lexington Books.

Janssen, P.H.M & Heuberger, P.S.C. (1995) Calibration of process-oriented models. *Ecol. Modelling.* **83**, 55–66.

Karlsson, M. & Yakowitz, S. (1987) Nearest-Neighbor methods for non-parametric rainfall-runoff forecasting. *Water Resour. Res.* **23**, 1300–1308

Kothandaraman, V. (1971) Analysis of water temperature variations in large river. *J. Sanit. Engng Div. ASCE.* **97**(SA1), 19–31.

Lee, R. M., & Rinne, J. N. (1980) Critical thermal maxima of five trout streams in the southwestern United States. *Trans. Amer. Fish. Soc.* **109**, 632-635.

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.

Morin, G., & Couillard, D. (1990) Predicting river temperatures with a hydrological model. *Encyclopaedia of Fluid Mechanics*, 10, Gulf Publishing Company, Hudson, Texas.

Nash, J.E. & Sutcliffe, J.V. (1970) River flow forecasting through conceptual models. Part A. Discussion of principles. *J. Hydrol.* **10**, 282–290.

Neumann, D.W., Zagona, E.A. & Rajagopalan, B. (2003) A Regression Model for Daily Maximum Stream Temperature, *J. Environ. Eng.* **129**, 667-674

Novales, A. & de Frutto, R.F. (1997) Forecasting with periodic models: A comparison with time invariant coefficient models. *Int. J. Forecasting.* **13**, 393-405.

Oliver, G. & Fidler, L. (2001) Towards a water quality guideline for temperature in the province of British Columbia. Pages 74. Ministry of Environment, Lands and Parks, Victoria, BC.

Osborn, D. & Smith, J. (1989) The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *J. Bus. Econom. Statist.* **7**, 117-127.

Ouarda, T.B.M.J, Labadie, J.W. & Fontane, D.G. (1997) Indexed sequential hydrologic modeling for hydropower capacity estimation. *J. Am. Water Resour. As.* **33**, 1337-1349

Pilgrim, J.M., Fang, X. & Stefan, H.G. (1998) Stream temperature correlations with air temperatures in Minnesota: implications for climate warming. *J. Am. Water Resour. As.* **34**, 1109–1121.

Quenouille, M. (1949) Approximate tests of correlation in time series. *J. Roy. Sta. Soc. B* **11**, 18-84

- Salas, J.D. (1985) Analysis and modelling of hydrologic time series. *Handbook of Hydrology*. Maidment, D. R., McGraw-Hill, New York 19.1:19.72.
- Salas, J.D. (1993) Analysis and modelling of hydrologic time series. *Handbook of Hydrology*. Maidment, D. R., McGraw-Hill.
- Schindler, D.W. (2001) The cumulative effects of climate warming and other human stresses on Canadian freshwaters in the new millennium. *Can. J. Fish. Aquat. Sci.* **58**, 18-29
- Shapiro, S.S. & Wilk, M. B. (1965) An analysis of variance test for normality (complete samples). *Biometrika*. **52**, 591-611.
- St-Hilaire, A., Morin, G., El-Jabi, N. & Caissie, D. (2003) Sensitivity analysis of a deterministic water temperature model to forest canopy and soil temperature in Catamaran Brook (New Brunswick, Canada). *Hydrol. Processes*. **17**, 2033-2047.
- Stefan, H.G. & Preud'homme, E.B. (1993) Stream temperature estimation from air temperature. *Water Resour. Bull.* **29**, 27-45.
- Tiao, G.C. & Grupe, M.R. (1980) Hidden periodic autoregressive moving average models in time series data. *Biometrika* **67**, 365-373.
- Vecchia, A.V. (1985) Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Water Resour. Bull.* **21**, 721-730.
- Webb, B.W. & Walling, D.E. (1993) Temporal variability in the impact of river regulation on thermal regime and some biological implications. *Freshw. Biol.* **29**, 167-182.

Webb, B.W. & Nobilis, F. (1997) A long-term perspective on the nature of the air-water temperature relationship: a case study. *Hydrol. Processes.* **11**, 137-147.

Webb, B.W., Clack, P.D. & Walling, D.E. (2003) Water-air temperature relationships in a Devon river system and the role of flow. *Hydrol. Processes.* **17**, 3069-3084.

Yakowitz, S., & Karlsson, M. (1987) Nearest-Neighbor methods with application to rainfall/runoff prediction. p. 149-160. In J.B. Macneil and G.J. Humphries (ed.) Stochastic Hydrology. D. Reidel, Norwell, MA.

Table 5.1 List of attributes included in the preliminary analysis of k-NN method

Attributs	Description	Formula
$Tw0$	Water temperature of the present week (t)	$Tw(t)$
$Tw1$	Water temperature of the week ($t-1$)	$Tw(t-1)$
$Tw2$	Water temperature of the week ($t-2$)	$Tw(t-2)$
$Tw3$	Mean of water temperature of the two past weeks	$mean[Tw(t-1), Tw(t-2)]$
$Ta1$	Air temperature of the week ($t-1$)	$Ta(t-1)$
$Ta2$	Air temperature of the week ($t-2$)	$Ta(t-2)$
$Ta3$	Mean of air temperature of the two past weeks	$mean[Ta(t-1), Ta(t-2)]$
$Ta4$	Weekly degree-days	Cumulative sum of air temperature
$Q1$	flow of the week ($t-1$)	$Q(t-1)$
$Q2$	flow of the week ($t-2$)	$Q(t-2)$
$Q3$	Mean of flow of the two past weeks	$mean[Q(t-1), Q(t-2)]$
$Q4$	Relative flow change	$(Q(t) - Q(t-1))/Q(t)$

Table 5.2 Performance measures (Root Mean Square Error, Bias, NSC) and their range of best-subset k-NN and PARX models

Model description	Variables	RMSE (°C)		Bias (°C)		NSC	
		Mean	Range	Mean	Range	Mean	Range
k-NN1	<i>Tw1</i>	2.52	[2.14, 2.94]	-0.54	[-0.79, -0.24]	0.27	[-0.47, 0.60]
	<i>Tw2</i>	3.15	[2.67, 3.57]	-0.53	[-0.74, -0.13]	-0.12	[-0.86, 0.27]
	<i>Ta1</i>	2.76	[1.75, 3.81]	-0.57	[-1.10, 0.12]	0.14	[-0.71, 0.59]
	<i>Ta4</i>	2.85	[1.99, 3.78]	-0.56	[-0.97, -0.17]	0.08	[-0.43, 0.67]
	<i>Q1</i>	4.02	[2.86, 4.98]	-0.59	[-1.14, -0.17]	-0.83	[-2.48, -0.05]
Best subset of k-NN2	<i>Tw1, Ta1</i>	2.14	[1.76, 2.61]	-0.44	[-0.67, -0.16]	0.49	[0.05, 0.70]
Best subset of k-NN3	<i>Tw1, Tw2, Ta1</i>	1.31	[0.78, 1.60]	0.01	[-0.09, 0.24]	0.81	[0.64, 0.89]
Best subset of k-NN4	<i>Tw1, Tw2, Ta1, Q1</i>	1.20	[0.80, 1.50]	-0.01	[-0.19, 0.19]	0.84	[0.67, 0.90]
k-NN5	<i>Tw1, Tw2, Ta1, Ta4, Q1</i>	1.34	[0.94, 1.67]	-0.02	[-0.19, 0.17]	0.80	[0.64, 0.90]
Best subset of PARX2	<i>Tw1, Ta4</i>	1.59	[1.19; 2.44]	-0.01	[-1.28; 1.79]	0.70	[0.04; 0.85]
Best subset of PARX3	<i>Tw1, Tw2, Ta4</i>	1.58	[1.19; 2.37]	-0.01	[-0.99; 1.50]	0.70	[0.12; 0.86]
Subset of PARX4	<i>Tw1, Tw2, Ta4, Q1</i>	1.68	[1.16; 2.76]	0.00	[-1.00; 1.38]	0.66	[-0.18; 0.88]
PARX5	<i>Tw1, Tw2, Ta1, Ta4, Q1</i>	1.74	[1.08; 2.98]	0.01	[-0.80; 1.39]	0.62	[-0.38; 0.90]

Table 5.3 Model parameters estimates for PARX3 of mean weekly maximum temperatures of the validated year 2004

Parameters	Period(τ)																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
$\phi_{1,\tau}^1$	1.04	-0.34	-1.19	-0.11	0.5	-0.37	-0.04	0.17	-0.16	0.24	0.31	0.02	0.25	0.04	0.21	0.27	-0.08	0.05	0.89	0.3	0.82	0.46	0.3	0.04	-0.3	0.28	0.21	-0.5	-0.28	0.16
$\phi_{1,\tau}^2$	-0.27	-0.21	-0.24	-0.2	0.33	0.22	-0.11	-0.16	-0.29	-0.13	-0.35	-0.1	-0.59	-0.22	-0.27	-0.37	0.1	-0.4	-0.55	-0.29	0.12	0.04	-0.05	0.23	0.04	-0.01	-0.67	-0.21	-0.12	-0.15
$\phi_{2,\tau}^1$	-0.86	0.26	-0.18	-0.39	0.39	0.21	0.05	0.60	0.12	0.36	-0.13	0.17	-0.16	0.08	-0.20	-0.24	0.38	0.27	0.19	-0.27	-0.49	-0.24	-0.44	-0.37	-0.59	-0.33	0.36	-0.21	-0.19	-0.33

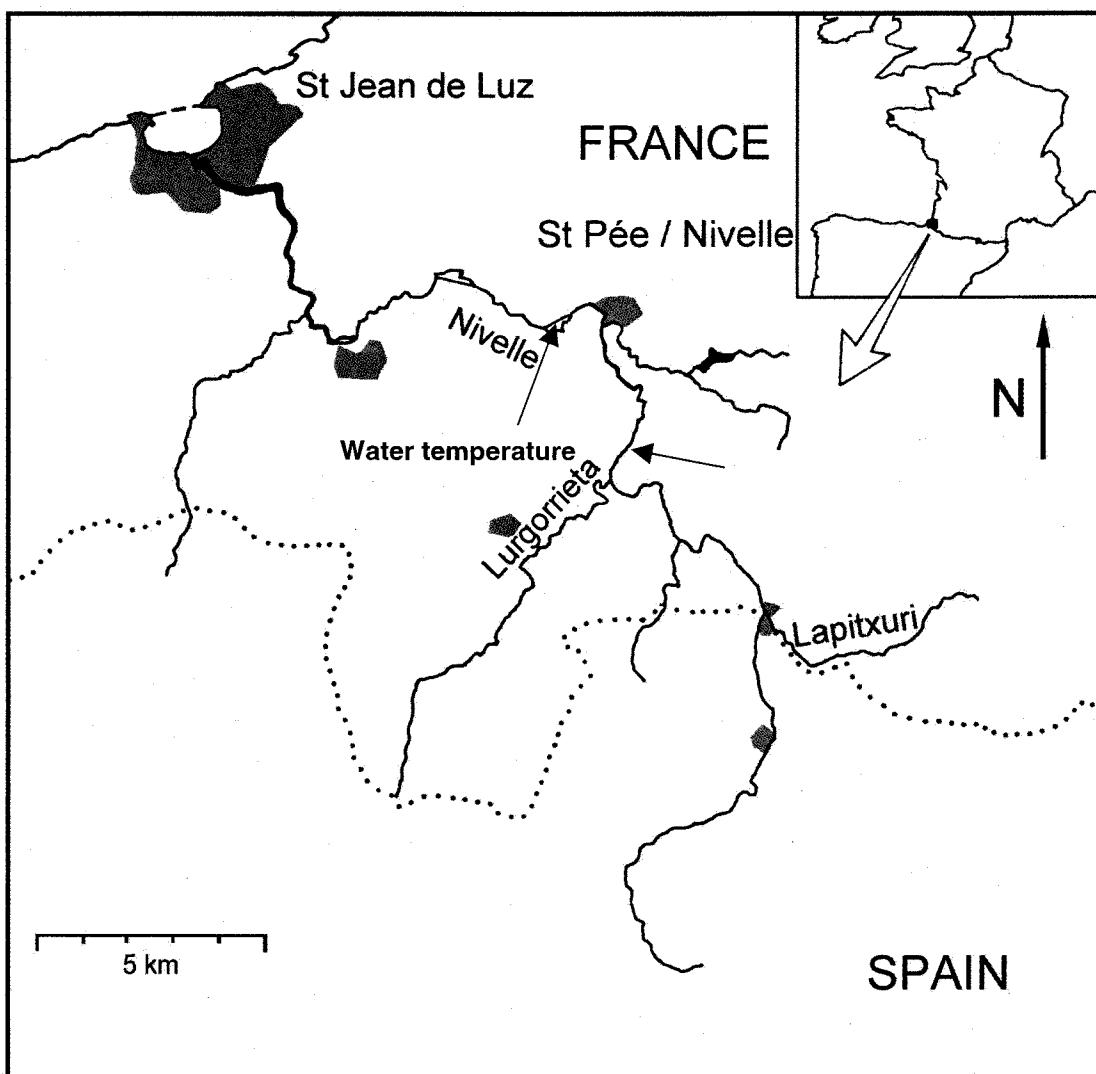


Figure 5.1 Nivelle river catchement area with the indication of the water temperature and streamflow stations

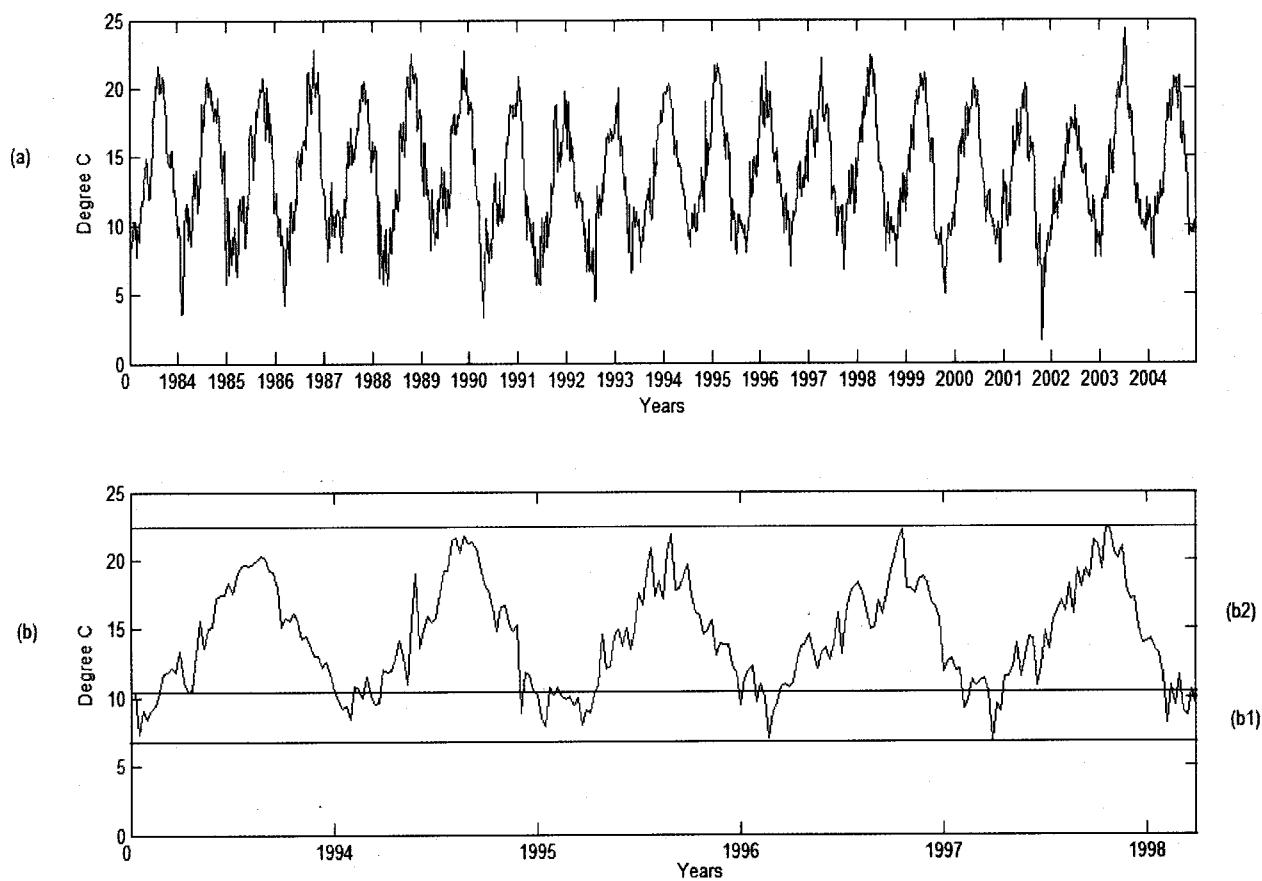


Figure 5.2 Weekly water temperature data collected in the Nivelle catchement (1984-2004) (a) weekly water temperature data including the periods from December to April, (b) detailed view of weekly water temperature data (1994-1998). (b1) represents the amplitude of the water temperature variation from December to April. (b2) represents the amplitude of the water temperature variation from May to November.

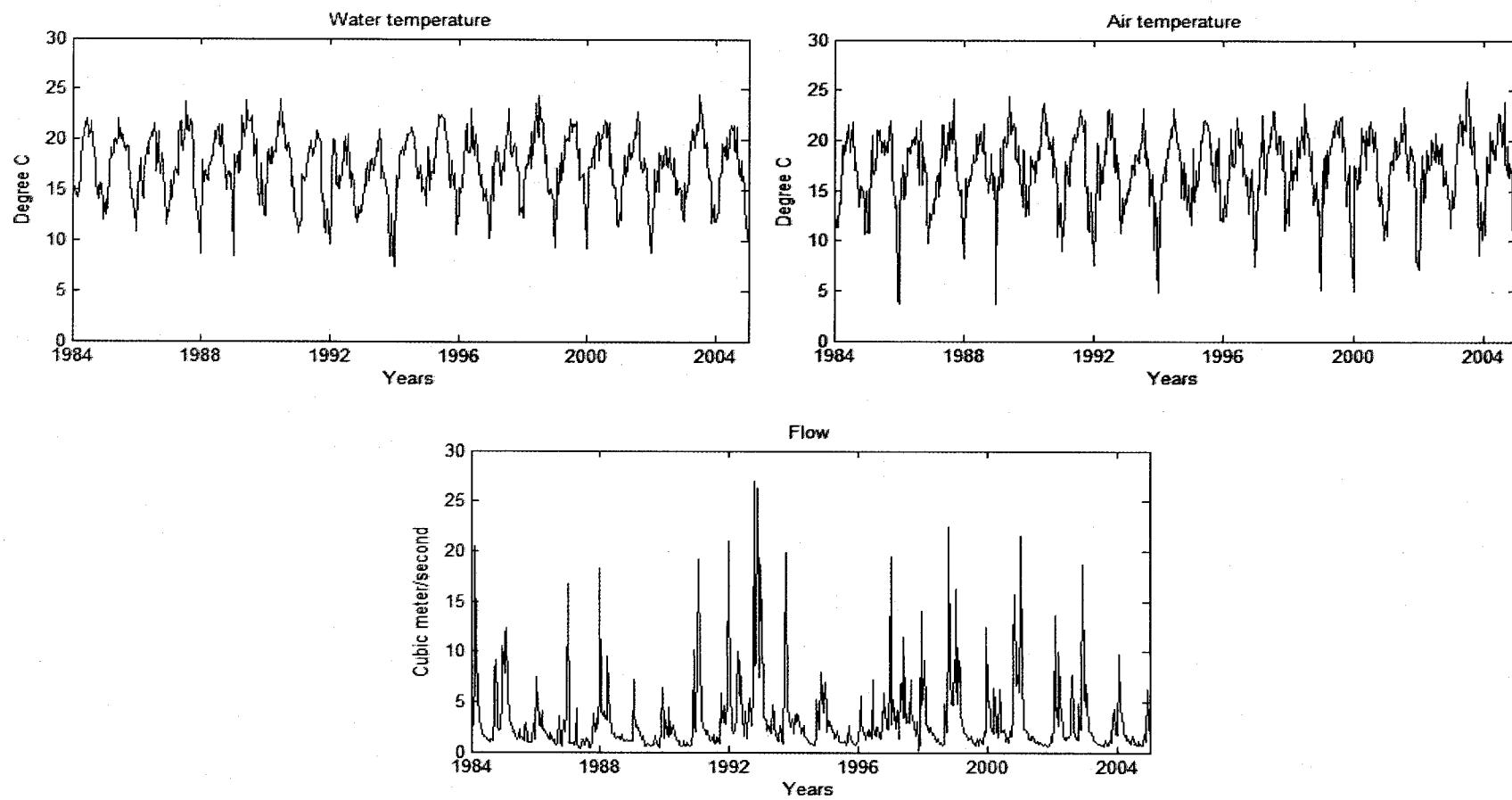


Figure 5.3 Water and air temperature and streamflow data excluding the periods from December to April collected in the Nivelle catchement from 1984 to 2004

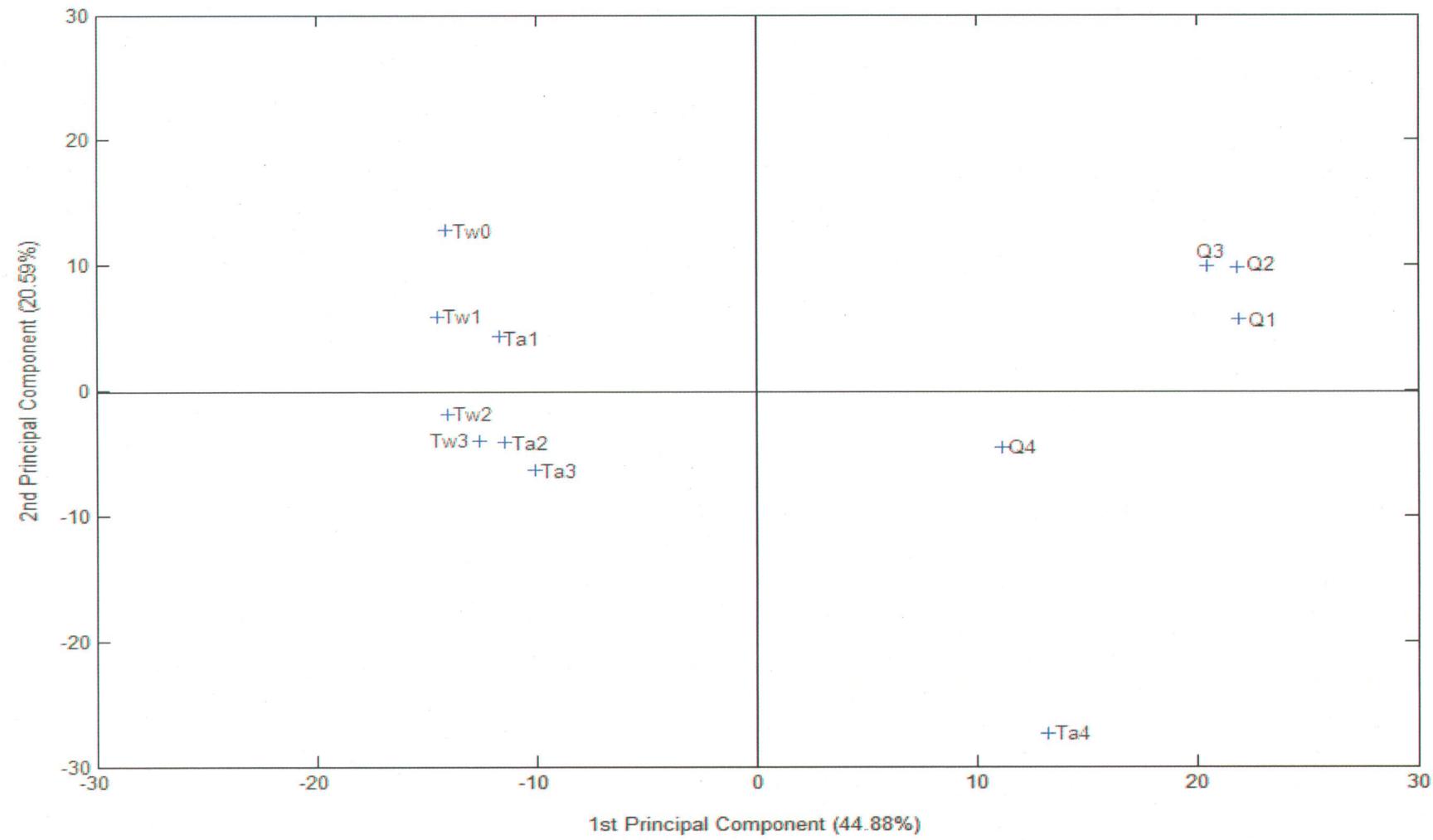


Figure 5.4 Scatter plot of the attributes on the PC1, PC2 plane

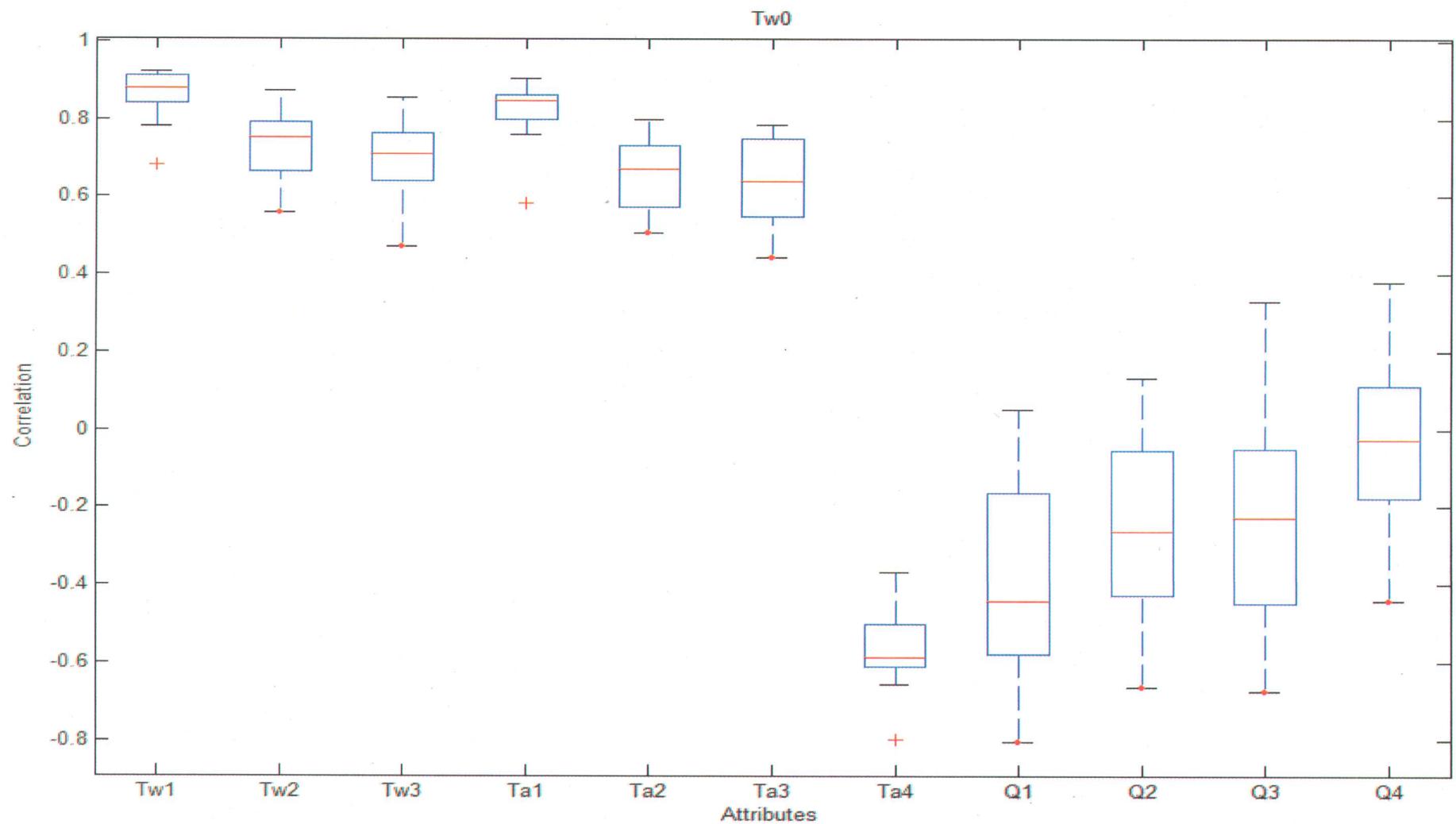


Figure 5.5 Boxplot showing the correlation between Tw0 and each attributes

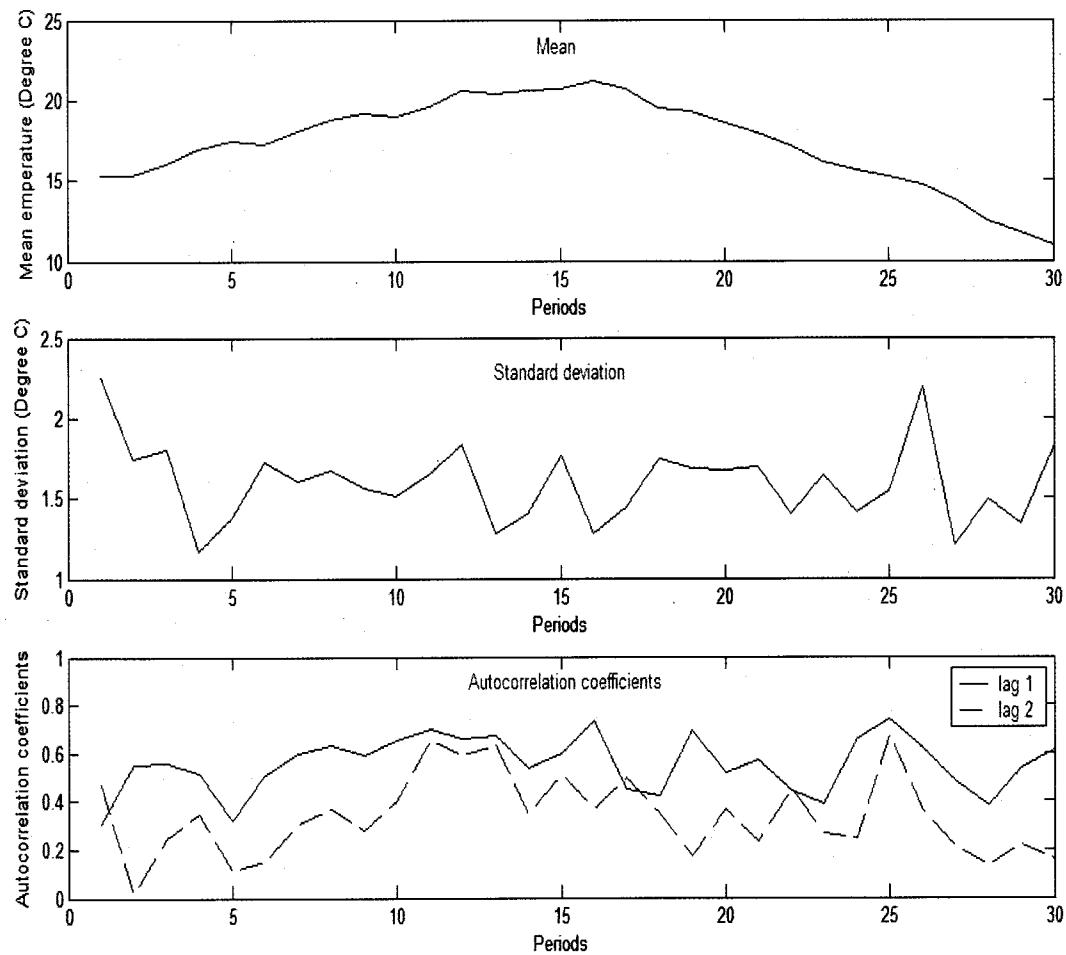


Figure 5.6 Sample mean, standard deviation and autocorrelation coefficients of lag 1 and lag 2 of mean weekly maximum temperature series for the Nivelle river.

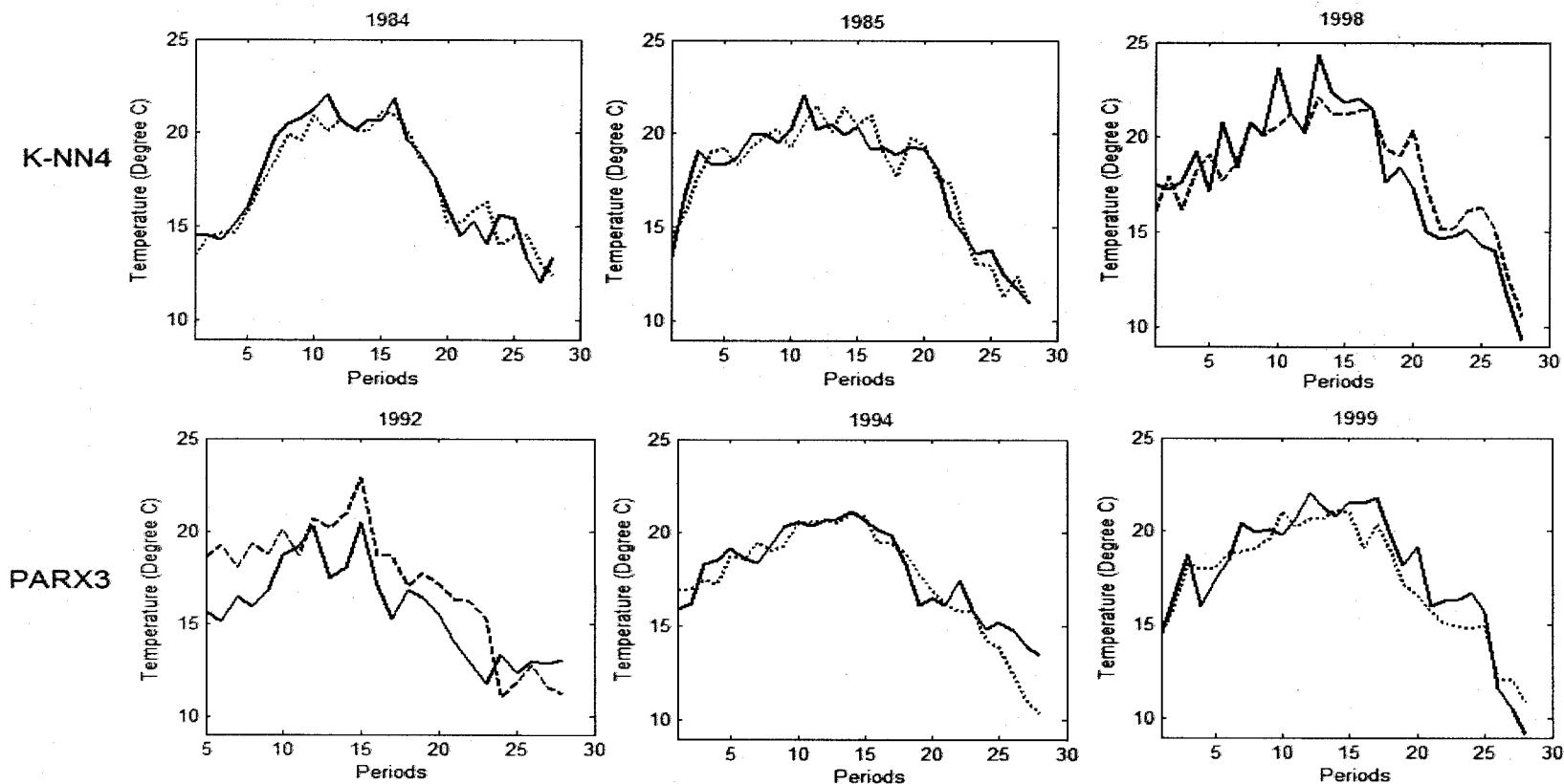


Figure 5.7 Time series of observed and predicted water temperatures obtained by k-NN4 and PARX3 for some validation years, the solid lines represent the observations.

CHAPITRE 6. CONCLUSIONS DE LA THÈSE

L'objectif principal de ce travail de thèse était de proposer une nouvelle approche appropriée de modélisation statistique de la température de l'eau en rivière dans le but de développer des outils de simulation et de prévision qui seront éventuellement utiles à la gestion des ressources aquatiques. Dans cette thèse, trois objectifs spécifiques ont été abordées : 1) l'identification, via une synthèse bibliographique, des approches statistiques les plus prometteuses qui pouvant être adaptées à la modélisation de la température de l'eau à une échelle de temps hebdomadaire, 2) la vérification de l'applicabilité d'une nouvelle approche statistique et la mise en valeur de son intérêt en comparant sa performance avec celle d'une des approches existantes, 3) l'introduction de variables explicatives de température de l'eau pour tenter d'obtenir une modélisation plus réaliste et plus opérationnelle pour les gestionnaires des ressources hydriques.

1. L'identification des approches statistiques les plus prometteuses

La revue de littérature a permis de souligner l'importance de la température de l'eau pour la vie aquatique en rivière, ainsi que l'éventuelle application de nouvelles approches de modélisation statistique. En effet, bien qu'il existe certains modèles permettant de modéliser les données périodiques (hebdomadaires, mensuelles), l'approche habituelle, relativement simple, consiste à éliminer les variations saisonnières souvent présentes dans ces données avant de procéder à leur modélisation. Une alternative à cette approche, proposée dans cette thèse, consiste à développer des modèles plus complexes qui par leur structure, permettent de modéliser la variation saisonnière sans nécessairement l'éliminer. Bien que les modèles périodiques soient utilisés dans plusieurs domaines, les travaux présentés dans les chapitres précédents sont, à notre connaissance, la première adaptation

de ces outils à la simulation de la température de l'eau en rivière. De plus, il ressort de la revue de littérature que les modèles non-paramétriques (réseaux de neurones artificiels (RNA) et k-voisins les plus proches (VPP)) sont très peu considérés, pourtant ils offrent l'avantage de ne pas nécessiter l'élaboration d'hypothèse a priori quand aux relations théoriques ou analytiques entre les intrants et la sortie et donc il n'y a pas de paramètres à estimer.

2. La vérification de l'applicabilité d'une nouvelle approche statistique

Ces travaux ont permis de vérifier la performance du modèle périodique autorégressif (PAR) et de la comparer à celle d'un modèle autorégressif existant (AR). Ce dernier consiste à extraire la composante saisonnière par ajustement d'une fonction sinusoïdale et faire la modélisation autorégressive des résidus. L'application de ces deux approches est rendue possible grâce à la disponibilité de séries chronologiques relativement longues de températures hebdomadaires de l'eau (1963-1980) d'un site sur la rivière Deschutes (États-Unis). Les analyses sont effectuées pour la période allant du mois de mai au mois de novembre de chaque année. Les trois critères d'évaluation utilisés pour évaluer la performance de ces modèles sont la racine de l'erreur quadratique moyenne (REQM), le biais et le coefficient de Nash (NSC). Ces critères ont été calculés pour chaque année en utilisant une technique de validation croisée de type "Jackknife". Les résultats ont montré que les deux approches présentent une performance similaire, avec une REQM de l'ordre de $0,81^{\circ}\text{C}$ pour le modèle AR(1) et de $0,90^{\circ}\text{C}$ pour le modèle PAR(1). Le biais moyen obtenu par ces deux modèles, en validation, est nul. Les coefficients de Nash moyens sont semblables, ils sont respectivement de l'ordre de 0,92 et 0,94. Toutefois, les critères de biais et de REQM calculés pour chaque période distinguent ces deux modèles.

En effet, le modèle PAR donne des valeurs de biais nuls et plus stationnaires que le modèle AR. On peut donc conclure que les résultats du modèle PAR sont prometteurs. Cependant, il faut se rappeler que ce modèle est entièrement dépendant de la structure d'autocorrélation de la série chronologique de température de l'eau, puisque aucune variable physique (e.g, température de l'air, débit) n'a été introduite. De ce fait, pour des conditions hydrométéorologiques variées, il serait important de développer un modèle périodique avec variables exogènes (PARX).

3. L'introduction des variables explicatives dans la modélisation de la température de l'eau

L'introduction de variables explicatives était une perspective logique pour rendre la modélisation plus réaliste et plus opérationnelle pour les gestionnaires des ressources hydriques. En effet, l'ajout de variables hydro-climatiques permet entre autres aux gestionnaires d'évaluer certains impacts anthropiques (e.g. modification du climat ou du régime hydrologique) sur le régime thermique. Dans cette thèse, deux nouveaux modèles de température de l'eau sont proposés, soit le modèle VPP et le modèle PARX (PAR avec variables exogènes). Le modèle VPP est une méthode non-paramétrique, elle consiste à calculer une distance euclidienne à partir d'attributs jugés importants pour la température de l'eau et à chercher dans tout l'historique, les k plus proches voisins qui serviront à estimer la température actuelle. Comme le modèle PAR, le modèle PARX permet de modéliser les données périodiques sans soustraire au préalable la composante saisonnière. Les paramètres de chaque variable indépendante sont estimés indépendamment pour chaque période.

Les données de température hebdomadaire de l'eau utilisées (1984-2004), proviennent de la station Ibarron (Site des laboratoires INRA à St Pée sur Nivelle dans la région des Pyrénées-Atlantiques, France). Plusieurs attributs (un total de 11) jugés significatifs sont considérés. Les résultats montrent que, pour ce qui est des critères de performance (REQM, biais et NSC), les modèles VPP4 (modèle avec quatre attributs) et PARX3 (modèle avec trois variables explicatives) se démarquent de tous les autres modèles considérés et conduisent aux meilleurs résultats. En effet, le modèle VPP4 présente une REQM de l'ordre de 1,20°C, elle est légèrement inférieure à celle obtenue à partir du modèle PARX3 (REQM=1,58°C). Le biais moyen annuel est nul pour les deux modèles mais variant d'une année à validation à l'autre. La valeur moyenne du coefficient de Nash du modèle PARX3 est légèrement inférieure (NSC=0,70) comparativement à celle obtenue par le modèle VPP4 (NSC= 0,84). En conclusion, même si les deux modèles peuvent simuler adéquatement les températures hebdomadaires de l'eau sur un petit bassin versant, le modèle PARX demeure le plus approprié, car de plus il préserve explicitement la persistance des séries périodiques, il offre une équation explicitant la relation entre la température de l'eau et les variables explicatives.

Les nouvelles approches de modélisation de la température de l'eau réalisées dans cette thèse ont donné des résultats très prometteurs. Toutefois, il faut noter certaines limites des modèles périodiques (PAR, PARX) qui peuvent être vite atteintes, étant donné le nombre relativement important de paramètres à estimer et de la rareté de longues séries chronologiques de la température de l'eau. De plus, le modèle VPP ne peut pas complètement rendre compte de toutes les caractéristiques des séries périodiques, y compris la périodicité de la fonction d'autocorrelation. Cependant, une telle

caractéristique est précieuse et son exploitation devrait permettre de mieux prédire la température de l'eau sur un pas de temps supérieur à celui pour lequel le modèle est conçu pour fonctionner. À cet égard, l'idéal serait sans doute d'obtenir un modèle qui pourrait éventuellement tenir compte des caractéristiques statistiques des températures hebdomadaires de l'eau sans nécessairement exiger une grande paramétrisation.

À la lumière de la revue de littérature et des résultats présentés de la modélisation des températures hebdomadaires de l'eau, plusieurs autres approches paramétriques et non-paramétriques pourront être proposées et comparées, en l'occurrence le modèle régression "*Ridge*" ou encore la méthode des RNA. Comme il fut mentionné, dans les situations où les variables environnementales sont nombreuses et fortement corrélées, la régression "*Ridge*" permet de réduire l'impact de la colinéarité et provoque un changement dans les valeurs numériques des coefficients du modèle, dont les prévisions deviennent légèrement biaisées, mais plus stables, et donc plus crédibles. Aussi, la méthode RNA permet entre autres la modélisation d'un système non-linéaire et la compression d'une multitude de variables en un réseau unique. L'application des approches de dépassement (au dessus ou au dessous d'un seuil) serait également prometteuse.

La présente étude est réalisée sur des bassins versants qui ne possèdent que peu de stations de mesures de température de l'eau. Il serait donc important dans les travaux futurs, d'inclure si possible, les données de cette variable disponibles sur des bassins versants comprenant plusieurs sites. Une approche multi-sites s'avère nécessaire afin d'obtenir des résultats plus précis sur la modélisation de la température de l'eau sur un bassin donné. Aussi, il serait intéressant de comparer tous les modèles (avec et sans

variables exogènes) sur plusieurs cours d'eau. Afin de discuter plus en détails les résultats obtenus, il serait également intéressant dans les travaux futurs de faire des analyses de sensibilité et d'effectuer une modélisation de l'incertitude associée aux estimations. De plus, il serait intéressant d'adopter un autre critère de comparaison de modèles, comme le Critère d'Information Bayésienne (CIB). Enfin, les modèles proposés sont des modèles ponctuels (0-D), il serait, donc, intéressant, de tester une modélisation spatio-temporelle.

Quoiqu'il en soit, il apparaît que les modèles périodiques (e.g. PARX) sont très appropriés et ont donné des résultats satisfaisants précisément dans ce cas. Ils trouveront sans doute plusieurs autres domaines d'application dans le futur.