

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

Examineur externe	Zoubeida Bargaoui ENIT
Examineur externe	Charles Lin Université McGill
Examineur 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		
	MODÈLES 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 autorégressives (ARMA) et les modèles périodiques autorégressives (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 °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 REQ_M 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 saisonnières, 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 ($\nu = 1, \dots, n$) 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 $\varepsilon_{\nu,\tau}$ 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 que 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.