

Université du Québec  
Institut national de la recherche scientifique  
Eau, Terre et Environnement

**Modélisation de la température de l'eau basée  
sur une méthode d'interpolation spatiale combinée  
à une approche multivariée**

Par  
Nicolas Guillemette

Mémoire présenté pour l'obtention  
du grade de Maîtrise ès sciences (M.Sc.)  
en Sciences de l'eau

**Jury d'évaluation**

**Examineur externe**

**Daniel Peters  
Environnement Canada**

**Examineur interne**

**Fateh Chebana  
INRS-Eau-terre-Environnement**

**Directeur de recherche**

**André St-Hilaire  
INRS-Eau-terre-Environnement**

**Codirecteurs de recherche**

**Taha B.M.J. Ouarda  
INRS-Eau-terre-Environnement**

**Normand Bergeron  
INRS-Eau-terre-Environnement**



**À mon père Vianney**



## REMERCIEMENTS

Au cours des deux dernières années, plusieurs collaborateurs ont pris part à la réalisation de ce projet.

Une mention toute spéciale pour André St-Hilaire à qui je dois dans un premier temps mon passage à la maîtrise recherche, merci grandement de m'avoir dirigé vers ce parcours. Au cours des deux dernières années, André m'a permis de réaliser un projet d'envergure au cours duquel il m'a supporté, encadré, enseigné et légué un bagage scientifique important. Son dévouement, son extrême générosité, ainsi que sa personnalité font de lui un personnage unique à qui je dois une véritable source d'inspiration. Merci pour cette tranche de vie André !

Des remerciements à mes deux co-directeurs Taha Ouarda et Normand Bergeron pour leur encadrement et soutien scientifique dans la rédaction des articles. Je vous suis très reconnaissant pour votre aide et collaboration et vous remercie grandement d'avoir soutenu ce projet et participé à sa mise en œuvre.

Une reconnaissance particulière à tous mes collègues associés à l'équipe St-Hilaire qui ont, à diverses étapes du projet, collaboré au bon déroulement des activités, participé aux campagnes de terrain et coloré mon quotidien. Ce fût un privilège de travailler avec vous. Merci à Valérie Ouellet, Simon Tardif, Yves Tremblay et Anick Daigle.

Je désire remercier aussi l'équipe technique qui m'a accompagné quotidiennement sur le terrain au cours de l'été 2007, soit Jean-Nicolas Bujold, Patricia Johnston et André Boivin. Ce fût une expérience enrichissante et combien amusante. Merci également à tous les intervenants de près ou de loin qui œuvrent au sein du CIRSA, quel terrain de jeu magnifique que cette rivière Ste-Marguerite.

Je tiens à remercier spécialement Jimmy Poulin pour son implication et expertise en géomatique. Sa disponibilité et son aide ont contribué au succès de ce projet.

Finalement, merci à mes parents, ami(e)s et à ma copine adorée.



## RÉSUMÉ

La description des processus thermiques en rivière nécessite l'application de modèles capables de décrire la dynamique des cours d'eau, tout en identifiant les facteurs prédominants qui permettent la caractérisation de la variabilité thermique pour différentes échelles spatiales et temporelles. Les progrès récents en instrumentation thermographique permettent un meilleur suivi « *in situ* » en facilitant grandement l'acquisition d'information et résultent ultimement en une augmentation des capacités de modélisation du régime thermique des rivières et de la caractérisation des écosystèmes lotiques.

La littérature relate plusieurs approches et catégories de modèles appartenant, dans la grande majorité des cas, aux familles des modèles déterministes et statistiques. À une exception près, aucune étude ne porte sur l'utilisation d'un modèle géostatistique pour l'estimation des températures de l'eau. Cette catégorie de modèle offre l'avantage, dans un contexte de gestion des ressources hydriques, de caractériser la distribution spatiale de variables thermiques pour une zone d'intérêt spécifique. Les modèles géostatistiques usuels sont la plupart du temps développés dans des espaces euclidiens physiques à une ou deux dimensions. Le présent projet vise à documenter une approche alternative qui consiste à développer ce type de modèle dans un espace mathématique bidimensionnel. Cet espace est construit à l'aide d'outils multivariés telles que l'analyse en composante principale (ACP) et l'analyse canonique des corrélations (ACC) qui prennent en compte un nombre limité de variables physiographiques caractérisant le ou les bassins versants à l'étude.

Une première étude porte sur la faisabilité d'utiliser un modèle géostatistique combiné à une approche multivariée pour l'interpolation de températures maximales mensuelles à l'échelle d'un bassin versant relativement grand. L'objectif vise à évaluer les performances d'un tel modèle pour les deux approches multivariées proposées (ACP et ACC) en utilisant des critères de performances connus comme la racine de l'erreur quadratique moyenne relative et le biais relatif. Les résultats démontrent que les deux approches multivariées mènent à des performances similaires, cependant l'utilisation de

l'ACP semble plus adéquate pour la représentation des contrastes de température des régions ayant une signature thermique distincte, tandis que l'utilisation de l'ACC mène à une erreur d'estimation légèrement plus faible.

Une deuxième étude s'intéresse plus spécifiquement à l'interpolation de variables thermiques usuelles et dérivées à une échelle spatiale plus petite que la précédente, soit un tronçon de rivière. L'approche vise donc à démontrer la possibilité d'estimer des variables thermiques qui peuvent faciliter la caractérisation de la qualité et de la quantité des habitats de poissons en rivière et qui démontrent donc un potentiel intéressant comme outils opérationnels pour les gestionnaires en ressources hydriques. Les résultats illustrent l'existence de variables physiographiques locales différentes de celles utilisées à l'échelle du bassin, qui décrivent adéquatement les processus thermiques pour un tronçon de rivière.

Une discussion sur le développement et la pertinence du modèle est établie en fonction de divers critères de modélisation et en termes d'applications réalistes pour les gestionnaires du milieu. Finalement, une réflexion est portée sur les diverses pistes futures qui devraient être davantage prises en compte pour la modélisation des processus thermiques en rivière et la caractérisation des habitats aquatiques.



# TABLE DES MATIÈRES

<b>REMERCIEMENTS .....</b>	<b>v</b>
<b>RÉSUMÉ .....</b>	<b>vii</b>
<b>LISTE DES TABLEAUX.....</b>	<b>xi</b>
<b>LISTE DES FIGURES .....</b>	<b>xiii</b>
<b>CHAPITRE 1 INTRODUCTION.....</b>	<b>15</b>
1.1 Mise en contexte .....	17
1.2 Facteurs influençant la température de l'eau .....	18
1.3 Outils de modélisation .....	22
1.4 Approche proposée .....	23
1.5 Objectifs de la thèse .....	25
1.6 Structure du mémoire.....	26
<b>CHAPITRE 2 RÉSUMÉ FRANÇAIS DU PREMIER ARTICLE.....</b>	<b>29</b>
2.1 Introduction.....	31
2.2 Données et site d'étude .....	32
2.3 Approche statistique.....	33
2.4 Résultats.....	36
2.4.1 Krigeage dans l'espace ACP.....	37
2.4.2 Krigeage dans l'espace ACC .....	38
2.4.3 Performances de krigeage .....	39
2.5 Discussion.....	40
2.6 Conclusion et applications futures.....	42
<b>CHAPITRE 3 RÉSUMÉ FRANÇAIS DU DEUXIÈME ARTICLE.....</b>	<b>43</b>
3.1. Introduction.....	45
3.2. Données et site d'étude .....	46
3.3. Approche statistique.....	48
3.4. Performances de krigeage .....	49
3.5. Résultats.....	50
3.5.1. Krigeage des températures maximales mensuelles.....	51
3.5.2. Krigeage des caractéristiques thermiques dérivées.....	51
3.5.3. Krigeage probabiliste .....	52
3.6. Discussion.....	53
3.7. Conclusion .....	55
<b>CHAPITRE 4 PREMIER ARTICLE VERSION INTÉGRALE ANGLAISE .....</b>	<b>57</b>
4.1. Abstract.....	61
4.2. Introduction.....	62
4.3. Data and study site .....	65

4.4.	Statistical Methods.....	67
4.4.1.	Results.....	71
4.4.2.	Kriging in PCA physiographical space.....	71
4.4.3.	Kriging in CCA physiographical space .....	73
4.4.4.	Kriging performances .....	75
4.5.	Discussion .....	76
4.6.	Conclusions and future work .....	81
4.7.	Acknowledgements.....	82
<b>CHAPITRE 5 DEUXIÈME ARTICLE VERSION INTÉGRALE ANGLAISE .....</b>		<b>91</b>
5.1.	Abstract.....	95
5.2.	Introduction.....	96
5.3.	Data and study site.....	98
5.4.	Statistical Methods.....	100
5.5.	Kriging performances .....	103
5.6.	Results.....	104
5.6.1.	Monthly maximum temperatures.....	106
5.6.2.	Derived thermal characteristics.....	108
5.6.3.	Probability kriging.....	110
5.7.	Discussion .....	111
5.8.	Conclusions.....	116
<b>CHAPITRE 6 DISCUSSION GÉNÉRALES .....</b>		<b>133</b>
<b>CHAPITRE 7 CONCLUSIONS GÉNÉRALES ET PERSPECTIVES .....</b>		<b>143</b>
<b>BIBLIOGRAPHIE.....</b>		<b>149</b>

## LISTE DES TABLEAUX

<b>Table 4-1</b>	Station names and their associated broad geographical regions and monthly maximum temperatures. Data used to construct kriged maps of July 1996 and 1997.....	83
<b>Table 4-2</b>	Physiographical variables characterising each station estimated using GIS.....	83
<b>Table 4-3</b>	Correlation coefficients between physiographical variables and monthly maximum water temperatures of July 1997 and associated P-values.....	84
<b>Table 4-4</b>	Physiographical variables extracted to estimate monthly maximum stream temperatures using the PCA algorithm at ungauged stations (Figure 4-1) for July 1996.....	84
<b>Table 4-5</b>	Cross-validation results.....	84
<b>Table 5-1</b>	Geographical coordinates of the 22 thermographs deployed at 11 sites with their associated names and locations.....	119
<b>Table 5-2</b>	Physiographical variables characterising each station calculated using GIS. Significantly correlated variables (p-value < 0.05) are identified in bold.....	120
<b>Table 5-3</b>	Cross validation results using ordinary kriging.....	120
<b>Table 5-4</b>	Cross validation results using probability kriging.....	121



## LISTE DES FIGURES

<b>Figure 4-1</b>	Location and map of the Moisie and Ste-Marguerite rivers drainage basins with temperature monitoring stations grouped into three broad geographical regions. X marks indicate locations of ungauged stations .....	85
<b>Figure 4-2</b>	Projection of significantly correlated physiographical variables in PCA space work set loading. A) July 1996. B) July 1997 .....	86
<b>Figure 4-3</b>	Experimental semivariogram of July 1996 used to krige in PCA space. Exponential model with nugget = 0.0001. range = 1.725 and sill = 0.314.....	87
<b>Figure 4-4</b>	Interpolated monthly maximum water temperature in PCA space performed with ordinary kriging. <b>Triangle</b> : tributaries of the Ste-Marguerite River. <b>Circle</b> : northern tributaries of the Moisie River. <b>Square</b> : southern tributaries of the Moisie River (see Figure 4-1 to identify geographical regions). A) July 1996. B) July 1996. <b>Note</b> : station locations were not identical for both years (see Table 4-1) .....	88
<b>Figure 4-5</b>	Experimental semivariogram of July 1996 used to krige in CCA space. Gaussian model with nugget = 0.001. range = 6.639 and sill = 2.011. Search radius was limited to $h < 3.2$ .....	89
<b>Figure 4-6</b>	Interpolated monthly maximum water temperature in CCA space performed with ordinary kriging. <b>Triangle</b> : tributaries of the Ste-Marguerite River. <b>Circle</b> : northern tributaries of the Moisie River. <b>Square</b> : southern tributaries of the Moisie River (see Figure 4-1 to identify geographical regions). A) July 1996. B) July 1996. <b>Note</b> : station locations were not identical for both years (see Table 4-1) .....	90
<b>Figure 5-1</b>	Location and map of the Ste-Marguerite River drainage basin with temperature monitoring stations grouped by pairs at 11 sites.....	122
<b>Figure 5-2</b>	Experimental semivariograms illustrating spatial structure related to monthly maximum temperatures using 7 significantly correlated metrics. A) July 2007. B) August 2007. C) September 2007 .....	123
<b>Figure 5-3</b>	Interpolated monthly maximum water temperatures (°C) performed with ordinary kriging using a 7 metric space. A) July 2007. B) August 2007. C) September 2007 .....	125
<b>Figure 5-4</b>	Experimental semivariograms illustrating spatial structure related to derived thermal characteristics using 7 significantly correlated metrics. A) <i>NEV</i> exceeding a 19°C threshold. B) <i>DD</i> over a 19°C threshold. C) <i>DUR</i> exceeding a 19°C threshold .....	126
<b>Figure 5-5</b>	Interpolated derived thermal characteristics performed with ordinary kriging into a 7 metric space. A) <i>NEV</i> exceeding a 19°C threshold. B) <i>DD</i> over a 19°C threshold. C) <i>DUR</i> exceeding a 19°C threshold .....	128

<b>Figure 5-6</b>	Experimental semivariograms illustrating spatial structure related to monthly maximum temperatures using probability distribution with 7 significantly correlated metrics. A) July 2007. B) August 2007. C) September 2007 .....	129
<b>Figure 5-7</b>	Interpolated monthly maximum water temperatures (°C) performed with probability kriging in a 7 metric space. A) July 2007. B) August 2007. C) September 2007 .....	131

## **CHAPITRE 1**

### **INTRODUCTION**





## 1.1 Mise en contexte

Le comportement thermique des rivières suscite de l'intérêt à l'échelle du globe et a fait l'objet d'études approfondies au cours des dernières années. Trois articles révisent en détail la littérature sur ce sujet et identifient les principaux travaux de recherche portant sur la caractérisation du régime thermique des cours d'eau (Ward 1985, Caissie 2006, Webb *et al.*, 2008). Un des principaux aspects abordés par les scientifiques à travers leurs études se situe au niveau des implications et conséquences de la variabilité thermique des cours d'eau sur les écosystèmes aquatiques. En effet, la température de l'eau est une variable abiotique qui influence de façon déterminante la qualité des habitats aquatiques en altérant l'ensemble des processus physiques, chimiques et biologiques. Plusieurs études récentes ont permis de décrire différents processus concernant le régime thermique des rivières en conditions naturelles ou modifiées. On dénote entre autres, d'importants travaux portant sur l'aspect qualitatif et quantitatif de l'habitat thermique de plusieurs espèces de poissons poikilothermes (i.e., espèces dont la température corporelle varie en fonction de la température du milieu ambiant), notamment en ce qui concerne le saumon Atlantique (Elliot *et al.*, 1998; Lund *et al.*, 2002; Mather *et al.*, 2008; Breau *et al.*, 2008), ainsi que la truite arc-en-ciel et le saumon du Pacifique (Berman et Quinn, 1991; Torgersen *et al.*, 1999; Ebersole *et al.*, 2001; Hodgson et Quinn, 2002). Plus spécifiquement, la compréhension du régime thermique des cours d'eau associée aux aspects qualitatifs et quantitatifs des habitats de diverses espèces de poissons a aussi été prise en compte par de nombreux auteurs qui se sont intéressés entre autres à l'impact des changements climatiques (Stefan et Sinokrot, 1993; Eaton et Scheller, 1996; Mohseni *et al.*, 2003; Tung *et al.*, 2006; Hari *et al.*, 2006; Crozier et Zabel, 2006; Sharma *et al.*,

2007; Ferrari *et al.*, 2007) et l'influence des barrages et des faibles débits sur la modification du régime hydraulique et thermique (Sinokrot *et al.*, 1995; Sinokrot et Gulliver, 2000; Salomon et Sambrook, 2004; Sutton *et al.*, 2007, Khangaonkar et Zhaoqing, 2008). Divers autres travaux portant aussi sur l'aspect qualitatif et quantitatif des habitats de poissons ont davantage étudié l'impact des résurgences d'eau souterraine (Ebersole *et al.*, 2003; Caissie et Giberson, 2003, Burkholder *et al.*, 2008), ou encore les conséquences des coupes forestières et du rôle des zones tampons (corridors forestiers) sur la variabilité thermique des cours d'eau (Holtby, 1988; Chen *et al.*, 1998; Johnson and Jones, 2000; St-Hilaire *et al.*, 2000; Malcom *et al.*, 2008). Dans l'ensemble, que ces conséquences et impacts soient associés à la variabilité naturelle des écosystèmes aquatiques ou qu'il s'agisse de répercussions plus fortement associées aux activités de nature anthropique, il en découle la nécessité de mieux comprendre les enjeux reliés à l'hétérogénéité thermique des cours d'eau, et ce, à différentes échelles d'étude. Le développement de nos connaissances dans ce domaine vise directement une meilleure gestion des ressources aquatiques et des pêches, en plus d'impliquer dans certains cas, la gestion de la qualité de l'eau dans un contexte d'approvisionnement pour les secteurs municipaux, agricoles, industriels et récréatifs.

## **1.2 Facteurs influençant la température de l'eau**

Un des principaux objectifs visés par l'étude et l'analyse du comportement thermique des rivières se situe au niveau de l'identification des facteurs qui influencent significativement la variabilité spatiale et temporelle. On dénote en effet un bon nombre d'éléments qui contrôlent plus ou moins directement le régime thermique des cours

d'eau. Une partie importante de ces éléments définit les caractéristiques d'un bassin versant qu'on regroupe sous le nom de variables physiographiques. Elles sont classifiées selon quatre groupes : i) géographie physique du bassin versant; ii) climat et conditions atmosphériques; iii) hydrologie et hydraulique de surface; et iv) captage souterrain. La géographie physique du bassin versant est fonction du relief, ainsi que de la répartition, de la composition et de l'étendue des éléments qui composent le territoire à l'étude. Elle influence de façon significative les conditions hydrologiques. Parmi ce groupe se retrouvent des facteurs comme la géomorphologie du cours d'eau, le type et l'étendue de la végétation riveraine, la topographie du bassin versant, le pourcentage de territoire occupé par la végétation et les plans d'eau, la situation géographique (latitude, longitude, azimut du cours d'eau), ainsi que la superficie du bassin versant. Le climat et les conditions atmosphériques sont en grande partie responsables du bilan thermique associé aux échanges de chaleur à l'interface eau-air. Les facteurs pouvant être greffés à ce groupe d'éléments sont par exemple la température de l'air, les précipitations (eau et neige), l'humidité, la vitesse du vent, la radiation solaire, ainsi que l'évaporation et l'évapotranspiration. Par ailleurs, les conditions hydrologiques et hydrauliques de surface influencent la capacité des cours d'eau à se réchauffer ou à se refroidir suivant les volumes d'eau qui transitent à l'intérieur du bassin versant. On dénote des facteurs tels la longueur et le degré de ramification du réseau hydrographique. Des éléments mesurables ponctuellement le long des cours d'eau peuvent aussi être pris en compte tel le débit, le degré de turbulence et les dimensions des sections d'écoulement. L'évaluation du temps de concentration permet de fournir un indice du temps de résidence des volumes d'eau à l'intérieur du bassin versant. Cet élément est en effet plus difficilement mesurable

puisqu'il fait intervenir simultanément des variables comme la pente moyenne du bassin versant, le coefficient de ruissellement associé à l'utilisation du territoire et au type de sol, ainsi que l'intensité et la durée des précipitations qui dépendent entre autres de la situation géographique du bassin versant. Finalement, la température de l'eau peut aussi être fortement influencée par les flux d'eau souterraine provenant du lit de la rivière. Des facteurs tels la dimension et le type de granulats, la conductivité hydraulique, le taux de sédimentation ainsi que la pression hydrostatique (cette dernière influence la capacité des échanges à l'interface eau-sol) sont toutes des variables pouvant influencées de façon significative le régime thermique des rivières.

Quelques auteurs ont identifié certaines variables issues de ces quatre groupes et les ont prises en compte pour la caractérisation thermique des cours d'eau. Ces dernières ont été identifiées principalement à l'intérieur de bassins versants composés de forêts et de pâturages, donc peu urbanisés. Isaaks et Hubert (2001) ont par exemple identifié, à l'échelle d'un bassin versant alpin situé dans les montagnes rocheuses au États-Unis, différentes variables géographiques physiques qui sont susceptibles d'influencer les températures maximales d'un cours d'eau. On dénote l'élévation moyenne du bassin versant, l'abondance d'arbres le long de la bande riveraine, la pente moyenne du bassin versant, la morphologie de la vallée, la largeur du cours d'eau, ainsi que l'aire du bassin versant. Dans une autre étude réalisée par Lowney (2000), on a plutôt mis en évidence des variables climatiques et atmosphériques pour l'estimation de températures journalières maximales et minimales sur un tronçon de 250 km sur la rivière Sacramento en Californie. L'auteure a utilisé entre autres des métriques comme la température de l'air, la pression atmosphérique, la densité de l'eau et de l'air, la radiation solaire et la

vitesse du vent. Plus récemment, Gardner *et al.* (2003) ont considéré deux variables relatives au réseau hydrographique d'un bassin versant relativement petit (770 km<sup>2</sup>) situé dans les monts Catskill dans l'état de New-York pour l'estimation de températures journalières. Ces deux variables hydrologiques caractérisaient dans un premier temps une distance le long du réseau hydrographique et dans un deuxième temps une distance qui combinait à la fois la longueur du réseau hydrographique et le degré de ramification (ordre du cours d'eau) de ce dernier. Finalement, les études menées par Moore (2006) et Brown et Hannah (2008) ont permis l'identification et la combinaison de variables qui tiennent compte à la fois de la géographie physique du bassin versant, des conditions climatiques et de l'hydraulicité du cours d'eau. L'estimation de températures médianes mensuelles à grande échelle dans la province de la Colombie-Britannique, ainsi que l'estimation de températures journalières dans un très petit bassin versant de 7.7 km<sup>2</sup> situé dans le Pyrénées en France, représentaient les caractéristiques thermiques estimées. Ces deux ouvrages ont permis d'identifier des métriques tels que la température de l'air, les précipitations, l'altitude, le logarithme de la superficie du bassin versant, les pourcentages de superficie suivant le type d'occupation du sol (plan d'eau, milieux humides, glacier, urbain), l'azimut, la longueur du cours d'eau, l'ordre du cours d'eau et finalement le régime d'écoulement (régularisé ou non).

Dans l'ensemble, la totalité des facteurs énumérés ci-haut possèdent un pouvoir prédictif pour l'estimation des températures de l'eau. Le choix des variables explicatives à retenir dans l'analyse et l'estimation de caractéristiques thermiques, ainsi que pour la caractérisation des habitats fauniques, demeure donc très dépendant des échelles spatiales (Burnett *et al.*, 2006) et temporelles. Ces études démontrent bien que certaines variables

sont communes même si les échelles spatiales et temporelles diffèrent, cependant des analyses plus élaborées pour identifier certains facteurs prépondérants demeurent nécessaires à une meilleure compréhension de la variabilité thermique. Pour y arriver, plusieurs outils de modélisation sont disponibles et fournissent une certaine latitude quant au choix de l'échelle spatiale et de l'échelle temporelle à laquelle le modélisateur désire caractériser la signature thermique d'un cours d'eau.

### **1.3 Outils de modélisation**

Les types de modèles utilisés pour l'estimation des températures de l'eau se regroupent en deux grandes catégories : (i) les modèles déterministes ou physiques et (ii) les modèles statistiques. Les modèles déterministes sont basés sur des équations qui permettent de quantifier les flux d'énergie totaux décrivant les échanges de chaleur à l'interface entre l'eau et l'air, ainsi qu'à l'interface avec le lit de la rivière. L'objectif de ce type de modèle est donc de calibrer des composantes météorologiques telles la radiation solaire, la température de l'air, la vitesse du vent, l'humidité, la pression de vapeur, afin de minimiser l'erreur entre les températures observées et prédites. Plusieurs études relatent cette famille de modèles pour l'estimation unidimensionnelle des températures de l'eau (Morin *et al.*, 1994; Chen *et al.*, 1998; St-Hilaire *et al.*, 2003; Caissie *et al.*, 2007). Ce type d'approche a comme avantage d'être relativement flexible suivant la modification ou l'altération de certains paramètres, notamment la bande riveraine, la végétation ou le débit. Par contre, les modèles déterministes requièrent parfois un grand nombre de variables qui peuvent être difficiles à mesurer. De plus, les longues séries chronologiques requises sont relativement difficiles à trouver.

Une alternative aux modèles déterministes consiste à utiliser des modèles statistiques (Ahmadi-Nedushan, 2006). Ces derniers sont souvent utilisés pour la modélisation des températures de l'eau spécialement lorsque la température de l'air est la seule donnée disponible à l'intérieur du secteur à l'étude. Plus spécifiquement, on dénote à l'intérieur de cette catégorie de modèle, deux groupes distincts soit : (1) les modèles paramétriques et (2) les modèles non-paramétriques. Les modèles paramétriques nécessitent l'ajustement de paramètres basés sur des structures statistiques prédéfinies. On dénote dans cette sous-catégorie, les méthodes de régressions linéaires, de régressions non-linéaires, les modèles autorégressifs (ARMA) et les modèles périodiques autorégressifs (PARMA). Plusieurs études témoignent de l'utilisation des ces modèles paramétriques (Stefan et Preud'homme, 1993, Caissie *et al.*, 1998; Mosheni *et al.*, 1998; Ahmadi-Nedushan *et al.*, 2007; Benyahya *et al.*, 2007). Par ailleurs, les modèles non-paramétriques ne possèdent quant à eux aucune structure statistique prédéfinie. Efficace pour la modélisation de systèmes non-linéaires, ce type de modèle inclut entre autres les réseaux de neurones artificiels (RNA) qui ont aussi été utilisés pour la modélisation des températures de l'eau (Bélanger *et al.*, 2005; Chenard et Caissie, 2008).

#### **1.4 Approche proposée**

La présente étude vise à utiliser une méthode d'interpolation spatiale appelée krigeage combinée à une approche multivariée pour l'estimation de caractéristiques thermiques en rivière. Ce modèle vise à développer une alternative aux deux catégories de modèles précédents en tirant profit des développements récents en instrumentation thermographique. Dans le passé, un travail similaire a été réalisé par Gardner *et al.*

(2003) lorsque celle-ci a utilisé la technique de krigeage pour estimer des températures de l'eau journalières à l'aide d'un réseau relativement dense de 72 thermographes sur un bassin versant de 770 km<sup>2</sup>. Le modèle consistait à effectuer une interpolation unidimensionnelle dans l'espace physique (cours d'eau) en tenant compte des distances de séparation entre les stations. Trois métriques ont été élaborées afin de caractériser ces distances inter-stations. La première était la distance minimale entre deux stations (distance euclidienne), la deuxième calculait plutôt une distance le long du réseau hydrographique (kilomètre de rivière) et finalement, la troisième tenait compte de la distance le long du réseau pondérée par l'ordre du cours d'eau.

L'approche proposée ici consiste à développer et poursuivre le travail effectué par Gardner *et al.* (2003) en procédant à l'interpolation bidimensionnelle dans un espace autre que l'espace physique. L'espace utilisé sera créé à l'aide d'une combinaison de variables physiographiques qui explique le mieux la variabilité thermique. Ces espaces d'interpolation combinent l'ensemble de l'information physiographique par des méthodes multivariées telles l'analyse en composantes principales (ACP) et l'analyse canonique des corrélations (ACC). Ces deux approches multivariées transforment les variables physiographiques originales de façon à obtenir des vecteurs (combinaisons linéaires) orthogonaux dans l'espace (non-corrélés). On peut alors procéder à l'interpolation de diverses caractéristiques thermiques à l'intérieur des espaces bidimensionnels ACP et ACC. Chokmani et Ouarda (2004) ont démontré la pertinence de cette approche en procédant à l'estimation de quantiles de débits dans l'espace ACP et ACC.

Les approches multivariées sont appropriées pour l'analyse de la qualité des habitats aquatiques, puisqu'elles considèrent l'interrelation et la structure de corrélation entre les



différentes variables explicatives retenues. Ce travail présente deux études de cas qui permettront d'appliquer le modèle proposé à deux échelles spatiales différentes, soit à l'échelle du bassin versant et à l'échelle du tronçon de rivière.

## **1.5 Objectifs de la thèse**

Le principal objectif visé par le projet consiste à utiliser un modèle basé sur l'utilisation de la géostatistique et des méthodes d'analyses statistiques multivariées pour rendre possible l'estimation et la caractérisation de variables thermiques en rivière.

Les objectifs spécifiques du travail sont :

- Démontrer la faisabilité d'utiliser un modèle géostatistique pour l'interpolation de températures maximales mensuelles à l'intérieur d'un espace multivarié construit à l'aide de l'ACP et de l'ACC.
- Déterminer les performances d'un tel modèle en termes d'erreurs d'estimation en comparant les deux approches multivariées (ACP et ACC)
- Identifier les variables explicatives prépondérantes qui influencent le régime thermique des cours d'eau
- Établir les différences observées lorsque le modèle est utilisé à l'échelle d'un grand bassin versant versus une modélisation effectuée à l'échelle du tronçon de rivière, en considérant la nature et le nombre de variables explicatives retenues, ainsi que la densité de thermographes utilisées

- Discuter de la pertinence du modèle en termes d'applications concernant la gestion des ressources aquatiques et des pêches.

## **1.6 Structure du mémoire**

L'ensemble des travaux effectués dans le cadre de cette maîtrise est présenté ici sous forme d'un mémoire par articles. Ainsi, les chapitres deux et trois sont des résumés en français des deux articles rédigés dans le cadre de ce projet; vient ensuite l'intégralité des deux articles rédigés en anglais dans les chapitres quatre et cinq. Le premier article s'intitule « Feasibility study of a geostatistical modelling of monthly maximum stream temperatures in a multivariate space » et a été accepté pour publication dans la revue « Journal of Hydrology ». Le deuxième article s'intitule « Statistical tools for thermal regime characterization at the river segment scale: case study of the Ste-Marguerite River » et sera soumis prochainement dans la revue « River Research and Applications ». Les références relatives à chacun des deux articles sont reportées dans la bibliographie présentée à la toute dernière section de ce mémoire. Finalement, le chapitre six présente la discussion et la conclusion du projet.

Le rôle de l'étudiant dans la réalisation des ces deux articles a été de réaliser les campagnes de terrain, de compiler et d'extraire les données utilisées, de procéder à l'ensemble des analyses statistiques et finalement de rédiger les deux articles en tant que premier auteur. L'encadrement de ces travaux de recherche a été assuré principalement par le directeur de recherche en collaboration avec les co-directeurs de recherche. Ceux-ci ont entre autres orienté la méthodologie de recherche employée, validé l'exactitude des

analyses statistiques et des résultats obtenus et finalement participé à la révision des deux articles en émettant corrections, commentaires, suggestions et notes techniques.



## **CHAPITRE 2**

### **RÉSUMÉ FRANÇAIS DU PREMIER ARTICLE**



## 2.1 Introduction

Dans ce premier article, il est question de l'estimation de températures de l'eau maximales mensuelles à l'échelle d'un bassin versant relativement grand. L'intérêt est tout d'abord porté sur l'importance de caractériser les patrons thermiques durant la période estivale, afin de favoriser l'identification de la disponibilité et de la distribution des habitats aquatiques durant cette période chaude de l'année. Différentes études sont relatées et ciblent entre autres les seuils de température pour lesquels des espèces poïkilothermes comme le saumon Atlantique ou le saumon du Pacifique subissent des stress thermiques considérables au cours des mois chauds de l'été (Lund *et al.*, 2002; Hodgson et Quinn, 2002). L'altération des habitats lotiques n'est certes pas sans conséquence pour la faune aquatique. La thermie étant une variable d'habitat essentielle, elle nécessite une meilleure caractérisation et compréhension via différents outils de modélisation. Un survol des principaux modèles utilisés expose les principales pistes explorées au cours des dernières années portant sur la caractérisation du régime thermique des cours d'eau. Les travaux de Gardner *et al.* (2003) ainsi que ceux de Chokmani et Ouarda (2004) sont spécialement considérés et présentent l'utilisation d'une méthode géostatistique appelée krigeage pour l'estimation de températures de l'eau et de quantiles de débit respectivement. Dans le deuxième cas, cet outil d'interpolation spatiale est combiné à deux approches multivariées, soit l'analyse en composantes principales (ACP) et l'analyse canonique des corrélations (ACC). La présente étude a pour but de poursuivre et de développer un modèle basé sur la combinaison des travaux de ces auteurs en intégrant différentes variables explicatives pour la construction d'un espace d'interpolation à l'aide d'outils multivariés. Les objectifs visés par cette étude sont : (1)

démontrer la faisabilité d'utiliser un modèle géostatistique combiné à une approche multivariée pour l'interpolation des températures de l'eau à l'échelle d'un grand bassin versant. (2) Comparer les résultats de l'utilisation de deux approches multivariées (ACP et ACC) basée sur les performances du modèle pour l'estimation de températures extrêmes dans un contexte de gestion des écosystèmes aquatiques.

## **2.2 Données et site d'étude**

L'information, provenant de 20 thermographes recueillie par Hydro-Québec sur les bassins versants de la rivière Moisie et Ste-Marguerite (Figure 4-1) entre 1989 et 1998, représente les données d'entrée utilisées dans le cadre de cette étude de faisabilité. La majorité des stations étaient situées sur le bassin versant de la rivière Moisie et la totalité de la superficie couverte par les deux bassins versants correspond à environ 25 900 km<sup>2</sup>. L'intérêt marqué pour ce territoire situé sur la côte-Nord du Québec vient du fait que la rivière Moisie est considérée comme la plus importante rivière à saumon Atlantique de cette région.

Les températures maximales mensuelles ainsi que les écarts maximums mensuels ont été calculés à partir de données horaires pour les mois les plus chauds de chaque année enregistrées sur ces deux bassins versants. La continuité et la disponibilité des mesures a permis de retenir les mois de juillet 1996 et de juillet 1997 comme mois d'intérêt pour l'estimation des températures extrêmes. Le choix des maximas de température comme information thermique repose en fait sur la pertinence d'obtenir un modèle capable de caractériser les zones de stress ou de refuges thermiques pour le saumon. Pour ces deux périodes retenues, un maximum de 12 stations a pu être utilisé. Pour chacune de ces



stations, 18 variables explicatives traduisant la physiographie du territoire ont été extraites à l'aide d'un système d'information géographique (Tableau 4-2). Le choix de ces métriques est basé essentiellement sur la capacité de ces dernières à influencer les températures de l'eau tout en considérant évidemment la disponibilité des données sur le site en question. Différentes connaissances en pratique de gestion par bassin versant rendent pertinent le choix des variables explicatives.

### **2.3 Approche statistique**

Cette étude est basée sur l'utilisation de données physiographiques pour la construction de l'espace d'interpolation servant à l'estimation des températures de l'eau. Basées sur l'étude de Chokmani et Ouarda (2004), deux approches multivariées sont utilisées pour la construction de l'espace d'interpolation, soit l'analyse en composantes principales (ACP) et l'analyse canonique des corrélations (ACC). Ces deux méthodes permettent de simplifier la complexité d'ensembles de données, en plus de pouvoir décrire la relation de dépendance existant entre les différentes variables explicatives. Une première utilisation de ces deux outils combinée à l'utilisation de variables abiotiques telles que la température de l'eau permet d'évaluer la faisabilité d'utiliser l'ACP et l'ACC pour la construction d'un système de coordonnées multivariées.

L'analyse en composantes principales (ACP) est une technique statistique qui permet de transformer un ensemble de données originales en un plus petit ensemble de variables non-corrélées. Ces dernières sont traduites sous forme de combinaisons linéaires (appelées composantes principales) qui expliquent un maximum d'information provenant

de l'ensemble de données originales. Algébriquement, la première composante principale est une combinaison linéaire des variables originales ( $x_i$ ) standardisées.

$$PC_1 = \sum_{i=1}^p a_{1i} x_i \quad (1)$$

Dans l'équation (1), la variance expliquée par  $PC_1$  est maximisée sous la contrainte que la somme des poids ( $a_i$ ) au carré est égale à un :

$$\sum_{i=1}^p a_{1i}^2 = 1 \quad (2)$$

Ainsi, la première composante principale explique le plus grand pourcentage de variabilité à l'intérieur du groupe de données originales. Cette variance expliquée diminue au fur et à mesure que les autres composantes principales ( $PC_2, PC_3, \dots, PC_i$ ) sont obtenues.

L'utilisation de l'analyse canonique des corrélations (ACC) permet plutôt d'examiner la relation entre deux ensembles de variables originales. Cette méthode vise à déterminer les combinaisons linéaires ( $U$  et  $V$ ) appelées variables canoniques qui permettent de maximiser la corrélation entre un premier groupe de variables  $X$ , telles que des statistiques de température de l'eau et un deuxième groupe de variables  $Y$ , qui inclut les variables physiographiques. Algébriquement, une paire de variables canoniques est une paire de combinaisons linéaires  $U$  et  $V$  :

$$\begin{aligned} U &= a'X \\ V &= b'Y \end{aligned} \quad (3)$$

qui permet de maximiser le coefficient canonique des corrélations donné par :

$$r_{U,V} = \frac{a' S_{12} b}{\sqrt{a' S_{11} a} \sqrt{b' S_{22} b}} \quad (4)$$

où ,  $a$  et  $b$  représentent les vecteurs des coefficients canoniques associés au premier et deuxième ensemble de variables originales ( $X$  et  $Y$ ) respectivement. Par ailleurs,  $S_{12}$ ,  $S_{11}$  and  $S_{22}$  représentent les matrices de covariance pour ces mêmes ensembles de variables originales. L'analyse canonique des corrélations (ACC) permet donc de maximiser la corrélation entre deux ensembles de données, tandis que l'analyse en composante principale (ACP) permet de maximiser la variance expliquée à l'intérieur d'un seul groupe de variables.

Une fois les espaces cartésiens décrits par les deux premières composantes pour l'ACP et les deux premiers vecteurs de variables canoniques physiographiques pour l'ACC , il devient possible d'interpoler les températures maximales mensuelles en utilisant la technique de krigeage soit dans l'espace  $PC_1-PC_2$  ou dans l'espace  $V_1-V_2$ . Cette approche géostatistique permet de quantifier la structure de corrélation spatiale entre les stations (thermographes) en fonction des distances inter-stations. Pour ce faire, la première étape consiste à construire un semi-variogramme expérimental en utilisant les valeurs de températures de l'eau, ainsi que les distances inter-stations mesurées à l'intérieur du système de coordonnées  $PC_1-PC_2$  ou  $V_1-V_2$ . L'équation du semi-variogramme expérimental est :

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (z_{x_i+h} - z_{x_i})^2 \quad (5)$$

Où  $\hat{\gamma}(h)$  estime la valeur de semivariance pour une distance donnée  $h$  et  $N(h)$  représente le nombre de paires de points séparées par cette même distance  $h$  dont la valeur observée est  $z_x$ . Une fois cette relation établie, une courbe théorique est ajustée au semi-variogramme expérimental et ensuite intégrée au système de krigeage composé d'équations linéaires incluant l'information de voisinage en termes de distances inter-stations.

$$\sum_{i=1}^{N(h)} \lambda_i \gamma(z_i - z_j) + \nu = \gamma(z_i - z_0) \quad (6)$$

Cette équation est résolue pour les poids  $\lambda_i$  de façon à minimiser la variance de l'erreur d'estimation sous la contrainte que la somme des poids soit égale à un.

$$\sum_{i=1}^{N(h)} \lambda_i = 1 \quad (7)$$

Finalement, l'estimateur de krigeage ( $\hat{z}$ ; équation 8) est utilisé pour calculer la valeur de la variable thermique d'intérêt aux points non jaugés ( $x_0$ ).

$$\hat{z}(x_0) = \sum_{i=1}^{N(h)} \lambda_i z(x_i) \quad (8)$$

## 2.4 Résultats

Pour les mois de juillet 1996 et juillet 1997, 12 stations (Tableau 4-1) ont été retenues pour l'estimation des températures maximales mensuelles. Au total, un maximum de 11 métriques (Tableau 4-3) étaient corrélées significativement (coefficient de Pearson,

$p$ -value < 0,05) avec les températures maximales. Des erreurs d'estimation plus faibles et une meilleure stabilité du modèle ont été obtenues en considérant 10 variables significativement corrélées pour la construction de l'espace ACP et les 8 variables les plus significativement corrélées pour la construction de l'espace ACC.

#### **2.4.1 Krigeage dans l'espace ACP**

Les figures 4-2, 4-3 et 4-4 illustrent l'ensemble des résultats obtenus pour l'ACP. Pour juillet 1996,  $PC_1$  et  $PC_2$  expliquent respectivement 60,4 pourcent et 26,2 pourcent de la variance totale. En juillet 1997, cette variance expliquée totalise 63,6 pourcent pour  $PC_1$  et 20,7 pourcent pour  $PC_2$ . Pour les deux mois,  $PC_1$  est dominée par les coordonnées géographiques (LAT et LONG) ainsi que par les caractéristiques du bassin tels le kilomètre de rivière (RK) et la superficie du bassin versant (DA). De l'autre côté,  $PC_2$  est plutôt associée au pourcentage du couvert forestier (FC), ainsi qu'à la topographie du bassin prise en compte par l'altitude à la station (A) et l'altitude maximum du bassin versant (MAXA).

La position des stations dans l'espace des composantes principales se divise en trois ensembles distincts qui représentent en fait les trois régions géographiques associées aux bassins versants de la rivière Moisie et Ste-Marguerite (Figure 4-1). Cette représentation des stations dans l'espace ACP aide à définir les régions géographiques qui sont davantage associées à des températures maximales plus fraîches ou plus chaudes durant cette période de l'année. Cet exercice permet d'identifier les zones de l'espace  $PC_1 - PC_2$  plus fortement associées à des zones potentielles de refuge ou de stress thermique. Finalement, une application du modèle démontre la possibilité d'estimer des températures

maximales mensuelles à des sites non jaugés (positions marquées par un X sur la figure 4-1). En effectuant l'extraction des variables physiographiques pour chaque point non-jaugé, il devient possible de calculer les coordonnées dans l'espace  $PC_1 - PC_2$  et d'estimer la variable thermique d'intérêt (maximum mensuel dans le présent travail). Le Tableau 4-4 donne les résultats de cet exemple d'application.

#### **2.4.2 Krigeage dans l'espace ACC**

L'algorithme de l'ACC vise à maximiser la corrélation entre deux groupes de données, soit un premier ensemble de variables contenant les statistiques de températures et un deuxième ensemble contenant les variables physiographiques retenues. Une contrainte limite le nombre de variables pouvant être incorporées à cette analyse multivariée. En effet, le nombre de métriques considérées doit être inférieur aux nombres de stations. Pour chacun des mois considérés, les variables physiographiques retenues ont été les 8 métriques les plus significativement corrélées ( $p$ -value les plus faibles) du Tableau 4-3. Par ailleurs, le premier ensemble de données associé aux variables thermiques inclut l'écart maximal mensuel et le maximum mensuel de juillet 1996 pour la construction de l'espace ACC de ce même mois, tandis que les maximums mensuels de juillet, août et septembre 1997 ont été retenus pour la construction de l'espace ACC nécessaire à l'interpolation des températures maximales mensuelles de juillet 1997. Les corrélations canoniques obtenues entre les deux ensembles de variables sont de 0,99 pour la première paire de variables canoniques ( $U_1$  et  $V_1$ ), et ce, pour chacun des mois, tandis que ces corrélations diminuent légèrement à 0,98 et 0,86 pour la deuxième paire de variables canoniques ( $U_2$  et  $V_2$ ) associée à juillet 1996 et juillet 1997 respectivement. Les figures

4-5 et 4-6 illustrent les semi-variogrammes et les cartes de krigeage associés à ces deux périodes. Les résultats démontrent que l'assemblage spatial des stations est moins bien structuré que dans l'espace ACP. En effet, les trois régions géographiques identifiées sur la figure 4-1 sont relativement moins bien définies sur la figure 4-6. Par contre, les stations demeurent groupées et localisées dans un même secteur de l'espace ACC.

### 2.4.3 Performances de krigeage

Une technique de rééchantillonnage de type « Leave-one-out » a été utilisée pour évaluer les erreurs associées à l'interpolation des températures maximales mensuelles dans l'espace ACP et ACC. Le biais relatif (rBIAS) ainsi que la racine de l'erreur quadratique moyenne relative (rRMSE) ont été utilisés comme critères pour l'évaluation de la performance du modèle.

$$\text{rBIAS} = \frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right) \quad (9)$$

$$\text{rRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right)^2} \quad (10)$$

Le Tableau 4-5 illustre les résultats obtenus. Les valeurs associées au rBIAS et au rRMSE ont varié de -2,5 % à 1,1% et de 5,2 % à 20,9 % respectivement. Des erreurs plus faibles ont été calculées pour l'approche ACC, spécialement pour juillet 1996.

## 2.5 Discussion

La présente étude a proposé un modèle géostatistique combiné à une approche multivariée pour la caractérisation des maximums thermiques d'une rivière à l'échelle du bassin versant. Les résultats démontrent qu'il est possible d'utiliser un espace multivarié pour l'interpolation de températures extrêmes, soulignant l'intérêt d'une telle méthode pour la gestion des habitats aquatiques.

L'espace ACP permet d'obtenir une meilleure représentation des contrastes entre les régions, puisque cette approche maximise la variance expliquée à l'intérieur des variables physiographiques. Les cartes de krigeage démontrent en effet que les trois regroupements de stations sont caractérisés par une signature thermique spécifique qui correspond en fait à trois régions distinctes sur les bassins versants à l'étude. L'ACP révèle que pour juillet 1996 et 1997, la latitude (LAT) ainsi que le kilomètre de rivière (RK) sont plus fortement associés à  $PC_1$  et représentent des variables explicatives pertinentes pour l'interpolation de maximums de températures à l'échelle du bassin versant. D'un autre côté,  $PC_2$  est plus fortement dominée par le pourcentage de couvert forestier (FC), soit une métrique qui caractérise la quantité de radiation solaire incidente, ainsi que par des variables explicatives caractérisant la topographie du territoire (A et MAXA) qui influencent directement la température de l'air. Il est intéressant de noter ici que toutes ces métriques n'auraient pas pu être incluses simultanément dans une modélisation 1-D.

L'utilisation de l'ACC ne permet pas de maximiser la variance physiographique expliquée, cependant les cartes de krigeage pour juillet 1996 et juillet 1997 illustrent généralement les mêmes regroupements de stations que ceux obtenus avec l'ACP. La



technique de l'ACC présente des valeurs de rRMSE plus faibles que pour l'ACP, puisque celle-ci est mieux adaptée pour associer l'information physiographique aux caractéristiques thermiques du cours d'eau. En effet, dans le cas de l'ACP, la variance physiographique est maximisée sans tenir compte de la variabilité thermique des températures de l'eau. Il est intéressant de noter que l'ACC présente de meilleurs résultats en utilisant que huit métriques, comparativement à 10 métriques pour l'ACP. Dans les deux cas, l'utilisation d'un nombre plus petit de variables explicatives que huit ou 10 a mené à une détérioration des performances du modèle basées sur les valeurs de rRMSE. Par ailleurs, les valeurs de rRMSEr et de rBIAS obtenues avec le présent modèle ont été comparées à celles obtenues à l'aide d'une régression linéaire multiple pour les mêmes variables physiographiques utilisées pour la construction de l'espace ACP et ACC. Des valeurs de rRMSE et de rBIAS de 99 % et de -42 % ont été obtenues respectivement pour juillet 1996, tandis que des valeurs de 47 % et de 10.9 % ont été trouvées pour juillet 1997. La performance supérieure de l'approche par krigeage a aussi été notée par Joseph *et al.* (2007) qui ont modélisé des quantiles de crue.

La présente étude a été effectuée avec un nombre limité de 12 stations localisées sur un large territoire. Ce nombre minimum de mesures rend l'ajustement du semi-variogramme difficile. Une première application du modèle permet l'estimation des températures à des sites non-jaugés en calculant les coordonnées physiographiques de ces derniers dans l'espace ACP. Des valeurs de températures similaires à celles mesurées aux stations jaugées ont été estimées, c'est-à-dire qu'aucune valeur aberrante n'a été prédite.

Finalement, un rappel concernant les principales sources d'incertitude souligne l'importance associée à la densité du réseau de stations, ainsi que la distribution spatiale

de ces dernières pour l'interpolation des températures de l'eau. Des implications au niveau de la structure de voisinage et de l'ajustement du semi-variogramme en découlent et influencent directement la variance de krigeage.

## **2.6 Conclusion et applications futures**

Cette étude a porté sur le développement d'un nouveau modèle d'estimation des températures de l'eau basé sur l'utilisation d'une approche géostatistique et multivariée. L'étude a démontré que l'utilisation des deux approches multivariées, c'est-à-dire l'ACP et l'ACC, peuvent être employées pour construire un espace physiographique servant à l'interpolation spatiale. Une meilleure performance a été obtenue avec l'utilisation de l'ACC en plus de nécessiter moins d'information en termes de variables physiographiques comparativement à l'utilisation de l'ACP. Une future application pourrait consister à répéter l'approche avec un plus grand nombre de stations et d'analyser la détérioration des performances du modèle suivant une diminution de la densité de thermographes utilisés. Une application de cette méthode à une échelle spatiale plus restreinte, comme par exemple un sous-bassin versant ou un tronçon de rivière, serait certes intéressante et permettrait d'identifier les variables d'influence prépondérantes et potentiellement différentes à cette échelle d'étude.

Il est permis de croire que le modèle peut devenir un outil intéressant pour une meilleure compréhension du régime thermique des rivières, lequel est souvent essentiel pour une gestion durable des ressources aquatiques et des pêches.

## **CHAPITRE 3**

### **RÉSUMÉ FRANÇAIS DU DEUXIÈME ARTICLE**



### 3.1. Introduction

Ce deuxième article s'intéresse, tout comme le premier, à la caractérisation de la variabilité thermique des cours d'eau. La présente étude reprend la méthodologie décrite dans le chapitre précédent, en proposant cette fois-ci une caractérisation thermique à l'échelle d'un tronçon d'une rivière. Il est entre autre mention de différents outils permettant de caractériser l'hétérogénéité thermique d'un cours d'eau, notamment l'utilisation de courbes température-durée (Brown et Hannah, 2008) ou du calcul cumulatif de degrés-jours (Mather *et al.*, 2008) au-dessus d'un seuil spécifique de température. Ce type d'outils possède une signification écologique importante en permettant une meilleure caractérisation de la distribution des habitats de poissons poïkilothermes et l'identification des refuges thermiques.

Plusieurs modèles statistiques utilisés dans les dernières années ont plutôt considéré et analysé l'autocorrélation qui existe à l'intérieur de séries chronologiques de températures de l'eau (Moatar *et al.*, 2006; Benyahya *et al.*, 2008). À l'exception de Gardner *et al.* (2003), ainsi que Peterson et Sickbert (2006), très peu d'études ont considéré la corrélation spatiale entre des mesures de température enregistrées à différents sites sur un même cours d'eau. L'identification de variables physiographiques capables d'expliquer la variabilité thermique d'un cours d'eau à différentes échelles d'études est une façon de caractériser cette structure de corrélation spatiale. Les études de Gardner *et al.* (2003) et de Moore (2006) sont respectivement deux exemples utilisant une combinaison de métriques à l'échelle d'un grand bassin versant et d'un tronçon de rivière.

L'étude vise donc à développer et à poursuivre les travaux effectués dans le cadre du premier article, mais se distingue de l'étude précédente par ses objectifs spécifiques. Ces derniers sont : (1) Élaborer un modèle permettant l'estimation de températures de l'eau et de variables thermiques dérivées à l'échelle du tronçon de rivière (< 100 km) en utilisant les variables physiographiques appropriées pour caractériser la variabilité thermique. (2) Estimer et comparer les performances du modèle entre les diverses variables krigées utiles à la caractérisation des habitats de poissons en rivières.

### **3.2. Données et site d'étude**

Le site échantillonné est situé sur un affluent de la rivière Saguenay portant aussi le nom de rivière Ste-Marguerite (différente de celle étudiée au précédent chapitre). Elle est située sur la Côte-Nord du Québec et s'écoule via deux principaux tronçons de 90 km (bassin versant de 1120 km<sup>2</sup>) et 85 km (bassin versant de 980 km<sup>2</sup>) respectivement, lesquels se rejoignent 2,5 km avant de se déverser dans la rivière Saguenay (Figure 5-1). Le bassin versant présente une topographie possédant des élévations allant de 100m jusqu'à 900m, ainsi que plusieurs vallées étroites où plusieurs cascades caractérisent le régime hydraulique. Le bassin est recouvert à 95 % de forêts et à 5 % de lacs et marais. Les différentes espèces de poissons présentes sont le saumon Atlantique (*Salmo Salar*), l'anguille d'Amérique (*Anguilla rostrata*), l'omble de fontaine (*Salvelinus fontinalis*), ainsi que diverses espèces de cyprinidées et de catostomidées.

Les températures de l'eau ont été mesurées en continu à 11 sites (Figure 5-1) à l'intérieur du bassin versant entre le 4 juillet 2007 et le 6 octobre 2007. À chacun des sites, deux thermographes étaient positionnés de façon à enregistrer les températures à la fois dans

un tributaire (généralement plus froid que le cours d'eau principal) et dans le cours d'eau principal directement. Ce dernier était disposé à quelques mètres en amont de la confluence avec le tributaire pour éviter toute influence thermique sur l'enregistrement des données. Au total, 22 thermographes ont été déployés sur le site d'étude (Tableau 5-1) et les températures ont été enregistrées à chaque 15 minutes.

Les variables thermiques estimées dans cette étude sont : températures maximales mensuelles dérivées des températures journalières, degrés-jours et durée en jours au dessus de seuils spécifiques de température, de même que le nombre d'événements et la probabilité au dépassement.

Le choix des diverses variables explicatives retenues dans cette étude est basé sur les critères suivants : 1) Impacts connus ou perçus des variables sur le régime thermique des cours d'eau et 2) disponibilité des données. Certaines métriques retenues avaient préalablement démontrées leur pertinence à expliquer la variabilité thermique (Gardner *et al.*, 2003; Guillemette *et al.*, 2008). D'autres ont plutôt été sélectionnées pour leur pertinence à l'échelle du tronçon de rivière. Au total, 18 variables ont été retenues (Tableau 5-2) en considérant leur possible habilité à influencer la signature thermique à une échelle moins importante que le bassin versant (tronçon de rivière) durant les mois les plus chauds de l'année. L'analyse a permis d'identifier huit métriques significativement corrélées ( $p\text{-value} < 0,1$ ) avec les variables thermiques estimées. Elles sont présentées au Tableau 5-2.

### 3.3. Approche statistique

La présente étude consiste à construire un espace mathématique orthogonal qui permet l'interpolation de caractéristiques thermiques en combinant différentes métriques via l'utilisation de l'analyse canonique des corrélations (ACC). Chockmani et Ouarda (2004) ainsi que Guillemette *et al.* (2008) ont respectivement utilisé cette approche pour l'interpolation de variables thermiques et hydrologiques dans un espace à deux dimensions. Dans la présente étude, l'utilisation de l'ACC a été retenue comme approche multivariée au dépend de l'ACP, puisque des erreurs d'estimation légèrement inférieures ont été obtenues avec l'utilisation de l'ACC dans le cadre du premier article (Table 4-5).

L'ACC est utilisée pour examiner la relation entre deux groupes de variables aléatoires. Dans notre cas, la première matrice X de variables  $x_1, x_2, \dots, x_i$  représente les variables thermiques à estimer, tandis que la seconde matrice Y de variables  $y_1, y_2, \dots, y_i$  représente les variables explicatives, telles des métriques géographiques et hydrologiques (voir Tableau 5-2).

L'utilisation de l'analyse canonique des corrélations a été décrite au chapitre précédent. Une fois l'ACC effectuée, les deux premiers vecteurs orthogonaux tirés des variables physiographiques ( $V_1$  et  $V_2$ ) sont retenus afin de construire un espace mathématique à deux dimensions. Les différentes variables thermiques peuvent par la suite être interpolées en utilisant la technique de krigeage. Pour ce faire, la première étape consiste à construire un semi-variogramme expérimental en utilisant les valeurs des variables thermiques à estimer, ainsi que les distances inter-stations mesurées à l'intérieur du système de coordonnées ACC. Les équations de base du krigeage ont été données au chapitre précédent.



Dans la présente étude, le krigeage probabiliste a aussi été utilisé afin d'estimer la probabilité au dépassement de seuils spécifiques de température. Pour ce faire, il s'agit de transformer la variable aléatoire  $z(x)$  en une variable d'indicatrice de la façon suivante :

$$I(x; z_k) = \begin{cases} 1, & \text{si } z(x) \geq z_k, \quad k = 1, 2, \dots, m \\ 0, & \text{autrement} \end{cases} \quad (6)$$

Finalement, la résolution des équations de krigeage permet d'estimer les caractéristiques thermiques en tenant compte des nouvelles variables transformées.

### 3.4. Performances de krigeage

L'erreur associée à l'interpolation spatiale de l'ensemble des variables thermiques estimées dans l'espace ACC est le principal critère d'évaluation de la performance du modèle. Une validation croisée utilisant la méthode de ré-échantillonnage appelée « leave-one-out » a été utilisée à cette fin. Les critères d'évaluation spécifiques utilisés pour estimer cette erreur sont définis ici:

$$\text{BIAS} = \frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i) \quad (7)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2} \quad (8)$$

$$\text{rBIAS} = \frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right) \quad (9)$$

$$\text{rRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right)^2} \quad (10)$$

### 3.5. Résultats

Les variables modélisées dans la présente étude ont été estimées en intégrant des métriques qui caractérisent bien la signature thermique à l'échelle du tronçon de rivière tels le débit (Q), l'altitude moyenne (MEANA), la distance du lac le plus près (CL), ainsi que le pourcentage de canopée à la station (CC). Afin d'augmenter la capacité du modèle à estimer l'ensemble des variables thermiques par la technique de krigeage, l'espace ACC a été construit en utilisant seulement des variables significativement corrélées (Tableau 5-2). Deux espaces ACC ont été construits en utilisant respectivement 7 et 8 métriques significativement corrélées. Des résultats satisfaisants ont été obtenus avec ce nombre spécifique de métriques, cependant une importante diminution des performances du modèle a été remarquée lorsqu'un nombre de variables inférieur à 7 est utilisé pour la construction de l'espace ACC.

Afin de déterminer et de mesurer la structure spatiale des variables à estimer dans l'espace ACC, la construction de semi-variogrammes expérimentaux a été effectuée. Une attention spéciale a été portée à l'hypothèse de stationnarité dans l'espace ACC. L'existence de tendances entre les variables canoniques et les variables thermiques a donc été vérifiée. Dans certains cas, une tendance linéaire entre ces deux groupes de variables était présente et a été retirée afin d'effectuer l'interpolation sur les résidus dans l'espace ACC. Cette tendance était par la suite réappliquée avant de construire les cartes finales de krigeage.

Les tableaux 5-3 et 5-4 présentent les erreurs d'estimation produites avec l'utilisation de la technique de krigeage dans un espace ACC construit avec 7 et 8 métriques respectivement. Les seuils spécifiques de température retenus comme valeurs de

référence pour certaines variables thermiques dérivées ont été fixés à 19 et 21 °C. Le choix de ces valeurs est strictement basé sur la signification biologique importante qu'ils représentent pour les habitats des espèces de poissons poïkilothermes (Lund *et al.*, 2002, Hodgson et Quinn, 2002).

### **3.5.1. Krigeage des températures maximales mensuelles**

Les températures maximales mensuelles de juillet, août et septembre 2007 représentent le premier ensemble de variables modélisées dans la présente étude. Les résultats démontrent que la première variable canonique (V1) est plus fortement associée à l'ordre du cours d'eau (SO), à l'altitude maximum du bassin versant (MAXA), ainsi qu'au kilomètre de rivière (RK). La deuxième variable canonique (V2) demeure tant qu'à elle plus influencée par l'aire du bassin versant (DA) et le pourcentage de canopée (CC). Les meilleurs résultats ont été obtenus avec un espace ACC à 7 métriques et présentent des valeurs de rRMSE oscillant entre 14,04 % et 19,31 % et des valeurs de rBIAS comprises entre -0,29 % et -1,29 % (Tableau 5-3). Les figures 5-2 et 5-3 illustrent respectivement les semi-variogrammes expérimentaux et les cartes de krigeage obtenues pour l'interpolation des températures maximales mensuelles dans un espace ACC construit avec 7 métriques.

### **3.5.2. Krigeage des caractéristiques thermiques dérivées**

Dans cette section, trois variables thermiques distinctes ont été interpolées en fonction des seuils de températures préalablement sélectionnés (19 et 21 °C). Ces variables sont :

- Nombre d'événements dépassant le seuil de température (*NEV*)
- Degrés-jours cumulés au dessus du seuil de température (*DD*)
- Durée en jours des événements qui dépassent le seuil de température (*DUR*)

Les résultats démontrent que *V1* est dominé à la fois par l'aire du bassin versant (*DA*) et par le kilomètre de rivière (*RK*), tandis que *V2* est plus fortement associé à l'altitude moyenne du bassin versant (*MEANA*) et le débit (*Q*). L'espace ACC construit avec 7 métriques, combiné à un seuil de température établi à 19 °C, procure les meilleurs résultats de krigeage pour ces trois caractéristiques thermiques dérivées. Des valeurs de *rRMSE* comprises entre 9,66 % et 15,08 %, ainsi que des valeurs de *rBIAS* oscillant entre -0,54 % et 0,62 % (Tableau 5-3) correspondent aux erreurs d'estimation obtenues pour les variables thermiques énumérées ci-dessus. Les figures 5-4 et 5-5 présentent l'ensemble des résultats graphiques.

### **3.5.3. Krigeage probabiliste**

Le résultat de ce type de krigeage permet de fournir des cartes qui attribuent pour une location donnée dans l'espace ACC, la probabilité de dépasser un seuil spécifique de température. L'analyse a été effectuée ici sur les températures maximales mensuelles de juillet, août et septembre 2007. Puisque les espaces ACC ont été construits avec les mêmes variables que pour la section 3.5.1, les métriques d'influence associées aux variables canoniques *V1* et *V2* sont identiques à celles énumérées dans cette section. Par contre, les erreurs associées à l'interpolation spatiale elle-même diffèrent de la section 3.5.1. Les meilleurs résultats ont une fois de plus été obtenus avec un espace ACC construit avec 7 métriques combiné à un seuil de température de 19 °C. Des valeurs de

RMSE comprises entre 0,22 et 0,31, ainsi que des valeurs de BIAS oscillant entre -0,02 et -0,04 (Tableau 5-4) ont été enregistrées avec l'utilisation du krigeage probabiliste. Les figures 5-6 et 5-7 présentent respectivement les semi-variogrammes expérimentaux et les cartes de krigeage obtenues.

### **3.6. Discussion**

Cette étude a permis d'estimer diverses caractéristiques thermiques à l'échelle du tronçon de rivière en utilisant un nombre relativement faible de thermographes et d'information physiographique. L'étude antérieure de Guillemette *et al.* (2008) avait démontré que l'utilisation d'un modèle géostatistique combinée à une approche multivariée telle l'ACC, représente une méthode adéquate pour l'interpolation de températures extrêmes sur un bassin versant relativement grand. Dans le présent projet, les caractéristiques thermiques ont été estimées adéquatement en incorporant des métriques qui décrivent bien les phénomènes locaux tels le débit (Q), le pourcentage de canopée à la station (CC), ainsi que la distance du lac le plus proche (CL). L'altitude moyenne du bassin versant (MEANA) a aussi été retenue et donne une bonne estimation de la topographie locale pour un secteur relativement petit, ce qui n'est pas nécessairement vrai pour un bassin versant de grande taille. D'un autre côté, il est intéressant de noter que quatre métriques demeurent corrélées significativement avec les variables thermiques pour chacune des échelles d'études : la superficie du bassin versant drainée (DA) à la station, l'altitude maximale du bassin versant (MAXA), l'ordre du cours d'eau (SO) et finalement, le kilomètre de rivière (RK). Une analyse des corrélations canoniques entre les variables canoniques ( $V_1$  et  $V_2$ ) et les statistiques de température révèle que  $V_1$  est plus fortement

associé aux variables décrivant le bassin versant (SO, RK, DA, MAXA), tandis que  $V_2$  est dominée par les variables associées aux phénomènes locaux observés à l'échelle du bassin versant (Q, MEANA, CC et CL).

Les valeurs de rRMSE et RMSE (Tableau 5-3 et 5-4 respectivement) indiquent dans la plupart des cas que le krigeage ordinaire effectué dans un espace à 7 métriques (excluant la métrique nommée CL), donne de meilleurs résultats qu'avec 8 métriques significativement corrélées (Tableau 5-2). Une des raisons potentielles qui explique cette différence est que même si la variable CL est significativement corrélée avec les caractéristiques thermiques estimées, cette dernière introduit des points discordants à l'intérieur du semi-variogramme expérimental. Les valeurs de rRMSE sont aussi plus faibles lorsqu'un seuil de température de 19 °C est pris en compte comparativement à un seuil de 21 °C. En effet, les variables thermiques présentent une plus faible probabilité au dépassement pour une valeur plus élevée de température. Lorsque très peu de valeurs mesurées dépassent ce seuil, il en résulte en une structure de corrélation spatiale moins robuste entre les stations dans l'espace ACC.

L'utilisation de caractéristiques thermiques dérivées (*NEV*, *DD* et *DUR*) ont mené à de bons résultats pour un espace multivarié construit avec 7 métriques. En effet, des valeurs de rRMSE variant entre 9,66 % et 15,08 % pour un seuil de 19°C et des valeurs de rRMSE variant entre 12,07 % et 19,75 % pour un seuil de température de 21°C ont été obtenues. Ces nouvelles variables estimées dans l'espace ACC reflètent la nature des caractéristiques thermiques (faible ou grande variabilité, constance) et fournissent une indication de l'hétérogénéité thermique parmi les stations du cours d'eau principal et des tributaires tout au long de l'été. Ces variables ont donc le potentiel d'indiquer et de

séparer les habitats de différente qualité pour des espèces de poissons spécifiques. Par ailleurs, le krigeage probabiliste semble plus approprié et plus avantageux lorsqu'il est utilisé pour cartographier le risque de dépasser ou non des seuils de température létaux ou sub-létaux pour certaines espèces de poissons.

Finalement, le nombre de stations (19 à 22) ainsi que le nombre de métriques utilisées (7 ou 8) ont permis de réduire la complexité du modèle. Un effort relativement faible a été nécessaire pour l'extraction des variables physiographiques dans un SIG et la densité de stations requises (22 ou moins sur deux tronçons totalisant 120 km en longueur) se compare avantageusement avec la densité de stations utilisée par Gardner *et al.* (> 70 sur un réseau hydrographique de 161 km) pour une interpolation unidimensionnelle dans un espace géographique.

### **3.7. Conclusion**

Cette étude a démontré que certaines variables physiographiques utilisées pour la construction de l'espace ACC à l'échelle du tronçon de rivière diffèrent de celles utilisées à l'échelle d'un grand bassin versant. Un nombre minimum de 7 métriques associées à une densité de stations comprise entre 19 et 22 ont mené à une structure de corrélation adéquate et par conséquent, à des résultats satisfaisants avec l'utilisation du krigeage ordinaire et probabiliste. Dans les deux cas, il a été démontré qu'il est possible d'obtenir des outils intéressants pour les gestionnaires des ressources en eau.





## **CHAPITRE 4**

### **PREMIER ARTICLE VERSION INTÉGRALE ANGLAISE**



# Feasibility study of a geostatistical modelling of monthly maximum stream temperatures in a multivariate space

By

Nicolas Guillemette<sup>\*1</sup>

André St-Hilaire<sup>1,4</sup>

Taha B.M.J. Ouarda<sup>1</sup>

Normand Bergeron<sup>2,4</sup>

Élaine Robichaud<sup>3</sup>

Laurent Bilodeau<sup>3</sup>

<sup>1.</sup> Chair in Statistical Hydrology, INRS-ETE, Université du Québec,  
490 de la Couronne, Québec, G1K 9A9, Canada

<sup>2.</sup> INRS-ETE, Université du Québec,  
490 de la Couronne, Québec, G1K 9A9, Canada

<sup>3.</sup> Hydro-Québec, 855 Ste-Catherine Street East,  
Montreal, Québec H2L 1A4. Canada

<sup>4.</sup> Centre Interuniversitaire de recherche sur le saumon Atlantique  
Sacré-Cœur, Québec

Manuscript published in the Journal of Hydrology,  
**Volume 364, Issues 1-2, 15 January 2009, Pages 1-12**

February 2008

\* Corresponding author: email: nicolas.guillemette@ete.inrs.ca Tel (418) 654-2530 (4457)

Fax number: (418)-654-2600



## **4.1. Abstract**

Healthy river conditions through optimal thermal regime controls water quality as well as the availability and distribution of fish habitat. A multivariate and geostatistical approach was developed to estimate maximum stream temperatures at a large basin scale. The methodology relies on the construction of a physiographical space using characteristics of gauging stations by testing two multivariate methods: principal components analysis (PCA) and canonical correlation analysis (CCA). Through that physiographical space, a geostatistical technique called ordinary kriging was then used to interpolate stream temperatures. Data from 12 temperature monitoring stations during July 1996 and July 1997 were used to estimate monthly maximum temperature. Results from the proposed approach were evaluated by comparing kriging performance obtained using both multivariate methods. Cross-validation technique has been performed on both approaches and satisfactory results were obtained. Kriging in the CCA physiographical space leads to better results because this approach seems more adapted to link physiographical information with specific water temperature. In addition, CCA required less physiographical information (i.e. 10 metrics for PCA vs 8 metrics for CCA) to provide more satisfactory results (up to 15 % decrease in rRMSE). In physiographical space, the gauging stations were found to cluster, potentially providing information to improve the accuracy of interpolation in that space. An example is provided to illustrate how to evaluate one of the stream temperature properties at ungauged stations using the PCA algorithm. The relevance of the results regarding the quality of fish habitats of the Moisie River is discussed.

**Keywords:** modelling, temperature, river, multivariate, geostatistical, kriging

## 4.2. Introduction

In recent years, global climate change is being considered as a potential threat for several fish species (Sinokrot *et al.*, 1995; Eaton and Scheller, 1996; Crozier et Zabel, 2006). The integrity of many physical and bio-chemical characteristics of an ecosystem is mainly controlled by temperature. The thermal regime of rivers controls water quality as well as the availability and distribution of fish habitat (Caissie, 2006). During warm months, unusually high stream temperature can occur naturally or as a result of human activities. Such extreme events are often linked to flow reduction, increased incident radiation and high air temperature (Sinokrot and Gulliver, 2000). Deforestation (Johnson and Jones, 2000, Chen *et al.*, 1998) has often been identified as having a negative impact on stream temperature (e.g. Holtby, 1988, St-Hilaire *et al.*, 2000).

A modified thermal regime can affect a number of poikilotherm fish such as Atlantic salmon (*Salmo salar*). For example, parr can be submitted to a thermal sublethal stress when water temperature exceeds 23°C through interrupted mRNA induction, which is an important process in protein synthesis (Lund *et al.*, 2002). Hodgson and Quinn (2002) found that Sockeye salmon (*Oncorhynchus nerka*) spawning can be interrupted or delayed when water temperature rises above a threshold of 19°C in northwestern USA. When this threshold is reached, adults start to seek thermal refugia. Therefore, improving our understanding of the thermal regime of rivers and our capacity to predict and simulate high temperatures by developing more accurate and flexible models applicable across watersheds are essential steps to identify the best management framework for aquatic resources and fisheries managers.

Many models have been developed and used to estimate stream temperature. Two broad categories are usually identified: (i) deterministic or physical models (e.g. Caissie *et al.*, 2007, St-Hilaire *et al.*, 2003, Gu and Li, 2002) and (ii) statistical/stochastic models (e.g. Caissie *et al.*, 1998, Mohseni *et al.*, 1998, Ahmadi-Nedushan *et al.*, 2007). Deterministic models use a conceptual approach which is based on the thermal exchange between atmosphere, the body of water and sometimes the river bed. Meteorological parameters such as air temperature, wind velocity and solar radiation are used as inputs to calculate energy budget equations and are also important to predict water temperature variations. For this reason, deterministic models are quite flexible in terms of input parameters modification, but also quite demanding in terms of model development and data requirement. As an alternative, statistical/stochastic approaches are based on a mathematical relationship between water temperature and independent variables such as air temperature and flow. This second model category requires fewer input data than deterministic models and model development can be relatively simple (e.g. linear or non-linear regressions; (Stefan and Preud'homme, 1993; Mosheni *et al.*, 1998)). Statistical/stochastic models are most often developed for a specific point or station on a river. As such, they cannot easily be transferred to another point in the river or another stream. Recent development of low cost temperature recorders makes it possible to sample water temperature at many locations with relatively good accuracy and high frequency.

Previous work on abiotic variables, such as water temperature, have attempted to make use of a number of independent physiographical/climatic variables via statistical analyses that explained the spatial and temporal variability (Collings, 1973; Mosley, 1982;

Miyazawa *et al.*, 1982; Hawkins *et al.*, 1997; Arscott, 2001; Scott *et al.*, 2002 ). Water temperature has also been characterized by spatial correlation (Peterson and Sickbert, 2006). One approach for modeling this type of spatial correlation is kriging (Isaaks and Srivastava, 1989). Ordinary kriging, a very popular geostatistical approach, consists of quantifying the spatial correlation structure between stations as a function of separation distances. The spatial interpolation at any point uses a weighted combination of neighbours. For example, Gardner *et al.* (2003) have considered a set of temperature recorders located throughout of Beaverkill Watershed in southwestern New-York to estimate river temperature at ungauged points on the same system. They used kriging to interpolate directly in a physical space by considering 3 metrics to calculate separation distances using a total of 72 temperature loggers. This resulted in a one-dimensional model of water temperature as a function of these 3 metrics: the shortest path between loggers, distances calculated along the stream network (river kilometer) and distances weighted by stream order.

The aim of the present study is to expand from the work of Gardner *et al.* (2003) and include a larger number of metrics to interpolate water temperature. In contrast with previous studies, a relatively sparse network of loggers is placed on a large scale hydrographical system and a physiographical space rather than a geographical interpolation space has been created by using multivariate approaches (Manly, 2004). This concept, first developed by Chokmani and Ouarda (2004) for flood quantiles, consists in combining physiographical and climatic information from drainage basin using principal component analysis (PCA) or canonical correlation analysis (CCA) to define a multivariate orthogonal interpolation space and then use ordinary kriging to



interpolate water temperature in the newly created principal components (PCA) or canonical correlation (CCA) space. Therefore, the objectives of this study are: (1) To elaborate a reliable spatial stream temperature interpolation model on a large scale basin using a multivariate geostatistical approach. (2) To compare results between two different multivariate approaches (PCA and CCA) to determine which is the most suitable for water resources managers in predicting stream temperature based on model performance.

### **4.3. Data and study site**

Thermographs were deployed on the Moisie and Ste-Marguerite rivers drainage basins on Québec North Shore, Canada (Figure 4-1), which are respectively large drainage basins of 19 197 km<sup>2</sup> and 6711 km<sup>2</sup>. The majority of the loggers were deployed on the drainage basin of the Moisie River which discharges a mean annual flow of 466 m<sup>3</sup> s<sup>-1</sup> into the St-Lawrence Estuary. It drains Lake Menistouc, which is located in the upper part of the drainage basin and then runs for 363 km. Main tributaries are Carheil, Nipissis and Aux Pécans rivers. In the province of Québec (Canada) the Moisie River is home to the most important spawning grounds for Atlantic salmon. Moreover, it is considered by anglers as one of the most important salmon rivers in the province because of the high average weight of individual adult salmon.

A total of 20 temperature monitoring stations were established by Hydro-Québec, the main provincial hydroelectric authority, during the period of 1989-1998. Daily temperatures were calculated from hourly observations provided by Hydro-Québec. As a test case for the methodology, we decided to focus on observations periods corresponding to warmer months of the year. There were a number of missing values during the first

part of the observation period (1989-1992) and for this reason we decided to focus on the second part of the observation period (1992-1998). Two years with the greatest number of concomitant time series for a minimum of 10 stations were selected for this study (1996 and 1997). Descriptive statistics used in the models are: monthly maximums of daily temperatures and maximum daily range, the latter being only used in CCA (see section 4.2). The focus on maxima stems from the potential interest of having a method that can be useful in defining areas of thermal stress or thermal refugia. Twelve stations with monthly maximum temperatures of July 1996 and July 1997 have been selected as test cases for modeling in multivariate space. Table 4-1 shows the station name with their associated monthly maximum temperatures for both time periods.

By extracting a large number of metrics, more information is known on the watershed. Multiple combinations of physiographical variables can then be used to influence the spatial variation of the thermal regime and to interpolate water temperatures. The choice of those variables is simply based on data availability and their potential to influence stream temperature across the watershed. In the present study, and in accordance with watershed management terminology, we identified geographical metrics such as the latitude, the longitude and the azimuth, physical variables such as the drainage area and the mean slope, and hydrographical variables such as the river kilometer and the stream order. Thus, a total of 18 physiographical variables characterizing each station were estimated using a Geographic Information System. They are reported in Table 4-2.

#### 4.4. Statistical Methods

In this study, the interpolation of stream temperature is based on the use of the basins coordinates in a physiographical space rather than a geographical space. Stream temperatures in the geographical space may change dramatically over adjacent drainage basins. In fact, while stream temperatures are discontinuous in the geographical space, they can be regarded as continuous variables in the physiographical space, thereby permitting the use of interpolation techniques. Different approaches are possible to construct the physiographical space. Chokmani and Ouarda (2004) proposed two multivariate approaches called principal component analysis (PCA) and canonical correlation analysis (CCA) to realize a study on regional flood frequency estimation. They used both methods respectively to simplify complex data sets as well as describing relationship of dependence existing between hydrological and physiographical variables. These approaches had never been tested on abiotic habitat variables such as water temperature. Therefore this study focuses on the feasibility of using PCA and CCA to establish a multivariate coordinate system.

Principal component analysis (PCA) is a statistical technique that linearly transforms an original set of variables into a substantially smaller set of uncorrelated (orthogonal) variates, called principal components, that represent most of the information of the original data set (Dunteman, 1989). Each principal component is a linear combination of the original variables. Geometrically, the first principal component is the line of closest fit to the  $n$  observations in the  $p$ -dimensional variable space. Algebraically, the first principal component,  $PC_1$ , is a linear combination of the original standardized variables  $x_i$ .

$$PC_1 = \sum_{i=1}^p a_{1i} x_i \quad (1)$$

In equation (1), the variance explained by  $PC_1$  is maximized under the constraint that the sum of the squared weights (e.g.  $a_{1i}$ ) is equal to one.

$$\sum_{i=1}^p a_{1i}^2 = 1 \quad (2)$$

The basic statistics of principal components analysis are the  $k$  variances (eigenvalues) and the associated variable weight vectors (eigenvectors)  $a_1, \dots, a_k$ . The correlations of the variables with a particular principal component are called the loadings. The sum of the squared correlations for each principal component equals the amount of variance explained by each PC. Therefore, the first principal component accounts for the greatest percentage of the variation in the original variables and the explained variance gets smaller as successive principal components are calculated.

In some multivariate analyses, the variables divide naturally into two groups. A canonical correlation analysis (CCA) can then be used to examine the relationship between these two sets of random variables. Given a set of hydrological variables,  $X$ , (stream temperature in our case) and a set of physiographical variables,  $Y$ , characterizing each station location of  $X$ , CCA aims to link two sets using vectors of canonical variables. It involves searching for linear combinations of  $X$  variables ( $U_1, U_2, \dots, U_i$ ) that have the maximum possible correlation with linear combinations of  $Y$  variables ( $V_1, V_2, \dots, V_i$ ) (Manly, 2004). This is somewhat similar to the concept behind a principal components analysis, except that in this case, correlation is maximized instead of explained variance. Algebraically, a pair of sample canonical variates is a pair of linear combinations  $U, V$

(equation 3) having unit sample variances that maximize the ratio called the sample canonical correlation (equation 4).

$$\begin{aligned} U &= a'X \\ V &= b'Y \end{aligned} \tag{3}$$

$$r_{U,V} = \frac{a'S_{12}b}{\sqrt{a'S_{11}a}\sqrt{b'S_{22}b}} \tag{4}$$

In equation (3),  $a$  and  $b$  represent the canonical coefficients vectors for the first and second set of random variables ( $X$  and  $Y$ ) respectively. In addition,  $S_{12}$ ,  $S_{11}$  and  $S_{22}$  are the sample covariance matrices consistent with the case of the initial variables. In general, the  $k^{\text{th}}$  pair of sample canonical variates is the pair of linear combinations  $U_k$ ,  $V_k$  among those linear combinations uncorrelated with the previous  $k-1$  sample canonical variates (Johnson and Wichern, 2007). Once CCA is performed, we retained the first two physiological orthogonal vectors ( $V_1$   $V_2$ , which are most highly correlated with thermal variables) built from  $Y$  to construct a two-dimensional physiological space where stream temperature characteristics can be interpolated using kriging.

Once PCA and CCA spaces are obtained, monthly maximum temperatures can be projected into these physiological spaces to be interpolated using kriging. This geostatistical approach quantifies the spatial correlation structure between stations as a function of separation distance. An experimental correlogram, or its inverse, the semivariogram, is first established using water temperature measurements and Euclidian distances measured using the coordinates establishes by the first two PCA or CCA variates. The semi-variance (or covariance) structure is estimated by the experimental semivariogram:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (z_{x_i+h} - z_{x_i})^2 \quad (5)$$

Where  $N(h)$  is the number of data pairs at a separation distance  $h$  which have an observed value  $z_{x_i}$ . A model is then fitted to the experimental semivariogram such as spherical, exponential, or Gaussian functions with three parameters: the nugget effect ( $C_0$ ), the sill ( $c$ ) and the range ( $a$ ). The nugget effect describes the occurrence of discontinuity at the origin of the semivariogram that may be caused by dissimilar sample values at short inter-station distances (Isaaks and Srivastava, 1989). The sill is the plateau reached by  $\hat{\gamma}(h)$ , which indicates a value of semi-variance that is a threshold beyond which there is essentially no spatial structure in the data. Finally, the range represents the distance over which the observed values are correlated.

The kriging estimator is a weighted average of the observed values  $z(x_i)$  which is used to estimate the value of  $z(x_0)$ , identified at a specific location  $x_0$  where there are no measured values. The model is denoted

$$\hat{z}(x_0) = \sum_{i=1}^{N(h)} \lambda_i z(x_i) \quad (6)$$

Where  $\lambda_i$  are the weights of the estimator that minimize the variance of the estimation error (ordinary kriging weights). By using the spatial structure defined by the theoretical semivariogram, a kriging system of linear equations combining neighbouring information can be defined as

$$\sum_{i=1}^{N(h)} \lambda_i \gamma(x_i - x_j) + \nu = \gamma(x_i - x_0) \quad (7)$$

under the constraint on weights:

$$\sum_{i=1}^{N(h)} \lambda_i = 1 \quad (8)$$

where  $\nu$  is the Lagrange multiplier for the constraint on the weights. The values of  $\lambda_i$  are obtained by solving this linear kriging system. In the present study, the geostatistical software GS + (Gamma Design Software, 2007) has been used to solve the kriging system.

#### **4.4.1. Results**

During July 1996 and July 1997, 12 stations were available to perform the multivariate analysis (Table 4-1). In order, to construct the CCA and PCA spaces, the first step was to choose a maximum number of physiographical variables significantly correlated with monthly maximum temperatures for these two specific time periods. As an example, significant (p-value < 0.05) correlation coefficients for July 1997 are reported in Table 4-3 and a maximum of 11 physiographical variables were selected for this month. Similar results were found for July 1996. Smaller errors and stable models for both years were obtained by using 10 of the most significant correlated variables for PCA and the 8 most significantly correlated variables for CCA.

#### **4.4.2. Kriging in PCA physiographical space**

Figures 4-2, 4-3 and 4-4 illustrate the PCA results. Figure 4-2 shows the loadings, i.e. the projection of the physiographical information on the PCA space of the two first principal components, which explain respectively 60.4 percent and 26.2 percent of the total

variance, for July 1996. Similarly, the two first principal components for July 1997 explain respectively 63.6 percent and 20.7 percent of the total variance. PC1 (the x axis) is dominated in both cases by the geographical coordinates (LAT and LONG) and by drainage basin characteristic such as river kilometer (RK) and drainage area (DA). On the other hand, PC2 in both cases is most strongly associated with forest cover (FC) and drainage basin topography such as altitude at the station (A) and the maximum altitude of the drainage basins (MAXA).

In order to determine and measure the multivariate spatial structure of maximum water temperatures, the isotropic experimental semivariograms were calculated in accordance with work set scores, i.e. the projection of temperature monitoring station on the PCA physiological space of the two first principal components. Figure 4-3 presents an example of the semivariogram calculated for July 1996 using separation distances in PCA coordinates. The adjusted fitted semivariogram is an exponential function. Monthly maximum water temperatures for each station were projected in PCA space for each period and ordinary kriging was performed to obtain interpolated temperatures for the entire PCA space. Figure 4-4 shows the results obtained for July 1996 and July 1997. The stations were grouped into three broad geographical regions (Figure 4-1):

- **triangle:** Tributaries of the Ste-Marguerite River
- **circle:** Northern tributaries of the Moisie River
- **square:** Southern tributaries of the Moisie River

The position of each station in the PCA space helps to define which stations or broad geographical regions are more associated with warmer or colder maximum water



temperature in July. This exercise can help characterizing the area of thermal stress or identifying the area of thermal refuge under various hydrological conditions. The kriged maps (Figure 4-4) provide the basis for estimating water temperature maxima for any location in the drainage basins under study and probably on adjacent basins within region. For example, in July 1996, lower maximum temperature conditions, which may eventually become thermal refugia for poikilotherm fish, are found for PC1 values ranging between 2.08 and 3.36 and for PC2 values located between -0.73 and -2.18. PC1 values ranging between -2.40 and -1.12 and PC2 values between -0.52 and 0.52 characterize areas with the highest temperatures that could become sites of potential thermal stress. Conversely, Figure 4-1 shows a few ungauged points (X mark) of the river system where the stream temperatures were not monitored. By extracting physiographical information associated with each ungauged stations of Figure 4-1, it becomes possible to find the PCA coordinates (score values) needed to estimate monthly maximum water temperatures using kriged maps. Table 4-4 shows the physiographical variables extracted as well as the estimated monthly maximum water temperatures at each ungauged points using PCA algorithm and kriged maps. The range of estimated temperatures at ungauged stations is in accordance with measured values in the geographical sub-region of July 1996 (Table 4-1).

#### **4.4.3. Kriging in CCA physiographical space**

In comparison to the construction of the PCA space, the CCA space is relatively different. First, because CCA involves searching for linear combinations between two sets of variables by maximising the correlation instead of variance, it becomes impossible to use all the significantly correlated physiographical variables. Canonical correlation

analysis must restrict the number of degrees of freedom. In fact, model parsimony dictates that the number of metrics be as small as possible without deteriorating the correlation structure between the two groups of variables (Stevens, 1986). Best results were obtained using 10 original variables for July 1996 and 11 original variables for July 1997 to construct the CCA space. In comparison with PCA analysis where score values are obtained only with physiographical variables, CCA analysis requires water temperatures and associated variates (first set of variables) as well as physiographical variables (second set of variables) to extract score values. The July 1996 water temperature variables used in CCA analysis were monthly maximum temperature and monthly maximum daily range. Because a more comprehensive data set was available in 1997, the July 1997 CCA uses the monthly maximum water temperature of July, August and September as the first set of variables. In both cases, the 8 physiographical variables used as second group of variables to compute the canonical correlation analysis were the 8 most significantly correlated variables (smallest  $p$ -value) of Table 4-3.

Sample canonical correlations (i.e. correlations between the linear combinations of the two sets of variables) are always largest for the first pair of canonical variates ( $U_1, V_1$ ). Canonical correlations were 0.99 for July 1996 and July 1997. For the second pair of canonical variates ( $U_2, V_2$ ), the canonical correlations were 0.86 and 0.98 for July 1996 and July 1997 respectively. Hence, the  $V_1$  and  $V_2$  cartesian space is highly correlated with water temperature and it was selected as the CCA kriging space.

As for PCA, semivariograms were calculated using the inter-station distances in CCA space. Figure 4-5 shows the semivariogram obtained for July 1996, using a gaussian function as fitted model. Ordinary kriging was performed for the entire CCA space by

projecting monthly maximum water temperatures for July 1996 and July 1997. Figure 4-6 presents the results with groups of stations (tributaries of the Ste-Marguerite River, northern tributaries of the Moisie River and southern tributaries of the Moisie River) identified and projected in the interpolation space. The spatial pattern of stations is less structured than in PCA space. The three broad geographical regions are still recognizable but the three clusters are less defined with CCA. The cluster which represents the northern tributaries of the Moisie River (circle) is well circumscribed in contrast with the other stations characterising the southern tributaries of the Moisie River (square) and the tributaries of Ste-Marguerite River (triangle) which are less ordered. However, they are still located in a same area of the CCA space.

#### 4.4.4. Kriging performances

A cross validation using leave-one-out resampling was used to estimate the error associate with the interpolation of monthly maximum water temperature in PCA and CCA space. This validation technique eliminates temporarily a station from the sample and the value for this observation is then estimated using remaining stations. This procedure was repeated for the whole station set. The relative mean bias (rBIAS) as well as the relative root mean square error (rRMSE) were used as performance evaluation criteria. These indicators are defined as follows:

$$\text{rBIAS} = \frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right) \quad (9)$$

$$\text{rRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right)^2} \quad (10)$$

where  $z_i$  and  $\hat{z}_i$  are respectively the observed value and the estimated value at station  $i$ . Because the range of observed maximum temperature was relatively small in July 1996 and July 1997, the rBIAS and the rRMSE were redefined using this maximum range ( $\Delta z_{\max} = z_{i \max} - z_{i \min}$ ) of both years as the denominator. Equation (9) and (10) are a more conservative interpretation of error than the usual definition of rBIAS and rRMSE, because  $\Delta z_{\max}$  is of typically less than 1.5°C, while the typical temperature in July was more than 10 times  $\Delta z_{\max}$ . According to Table 4-5, rBIAS and rRMSE varied respectively between -2.5% and 1.1% and 5.2% and 20.9% of the range of measured values (i.e.  $\Delta z_{\max}$ , see equations 9 and 10). Kriging in CCA space produced lower rRMSE than PCA, especially for July 1996.

#### **4.5. Discussion**

This study focused on the characterization of the river thermal regime at the drainage basin scale using multivariate and geostatistical approach. The use of PCA and CCA was shown by Chokmani and Ouarda (2004) to provide an adequate means of summarizing climatic and/or physiographical variability of drainage basins, while producing at the same time a cartesian space in which interpolation of hydrological extremes can be performed. The results shown in the present study confirm that a similar space can be used to interpolate water temperature extremes, which can be of interest for many water resources management issues such as fish habitat. The present study focused only on monthly maximum temperatures of July 1996 and July 1997 as a test case to construct kriged maps. However, other temperature statistics could be used such as daily mean,

daily minimum or daily range to characterize any temperature events of short or long duration.

PCA kriging appears to be somewhat less accurate than interpolation in CCA space. This result is similar to those obtained by Fernandez and Saenz (2003) as well as Chokmani and Ouarda (2004) with hydrological variables. However, as expected PCA provides more definite clusters of stations because the underlying criterion is the maximization of the explained physiographical variance. An examination of Figure 4-4 reveals the three clusters which define specific areas of the drainage basin of the Moisie and Ste-Marguerite rivers. Therefore, the PCA space is a better representation of physiographical contrasts between regions. When comparing July 1996 and July 1997, it can be seen that each cluster was characterized by a specific range of temperatures. Stations from the same cluster have a similar thermal regime associated with a specific area of the drainage basin. The first two broad geographical regions constituted by the stations from tributaries of the Ste-Marguerite River (triangle) and southern Moisie River (square) are located on both maps of Figure 4-4 in a part of the PCA space where the monthly maximum temperature varies between 16.6 and 17.7 °C. On the other hand, the region defined by the stations from the northern tributaries of the Moisie River (circle) were found to be projected in an area where the monthly maximum water temperatures were warmer than the two other clusters with temperatures oscillating between 17.3 and 18.5 °C. This reverse north to south gradient can be partly explained by the proximity of the northern stations to the upper lakes which become relatively warm during hot summer months. Moreover, the southern stations from Ste-Marguerite and Moisie rivers are more influenced by a good number of upstream cold tributaries during the same period.

For July 1996 and 1997, PCA reveals that the latitude (LAT) and the river kilometer (RK) are most strongly associated with PC1, while the percentage of forest cover (FC) as well as metrics related to altitude (A and MAXA) are associated to PC2. Hence the geographical position as well as the degree of ramification of stream within each drainage basin appears to be suitable metrics for the interpolation of temperature maxima. This conclusion is in accordance with the results of Gardner *et al.* (2003) who used similar metrics in their 1-D interpolation. In addition, the percentage of forest cover, strongly associated with PC2 is in fact, an important metric to characterize the quantity of solar radiation. On the other hand, the maximum altitude of the drainage basin (MAXA) and the altitude at the station (A) are two metrics that influence directly air temperature and therefore indirectly affect water temperature. Such additional information could not readily be included in a 1-D interpolation scheme.

It is interesting to note that, despite of the fact that CCA does not maximise explained physiographical variance, the CCA kriged maps of July 1996 and 1997 (Figure 4-6) generally show the same station clusters than those obtained with PCA. Therefore, both approaches were able to reproduce three regions with distinct thermal signatures. For both time periods analyzed in our study, CCA presented rRMSE values lower than those obtained with PCA (Table 4-5). One potential reason for this difference is that PCA is restrained to maximising the variance along the physiographical space irrespectively of water temperature variability. For this reason, the CCA technique is more adapted to link physiographical information with specific stream temperature. Canonical correlations can be a very useful tool for a priori estimation of the links between physiographic features and thermal regime. It is interesting to note that CCA kriging produced better results with

8 metrics while PCA required 10 explanatory variables. The extraction of metrics using GIS can be time consuming and thus, it becomes relatively advantageous to optimize the number of metrics used and to reduce the degree of model complexity. It should be noted that attempts to use less than 8 or 10 metrics respectively for CCA and PCA lead to a decrease in model performance, in terms of rRMSE values. In fact, model parsimony dictates that the number of metrics must be as small as possible without deteriorating the correlation structure which is directly associated with semivariogram calibration and kriging performances.

To demonstrate the improvement of this modelling technique over more traditional approaches, we performed multiple linear regressions to estimate maximum water temperature for July of both years using the same independent variables (Table 4-3) than those selected for PCA and CCA. We found respectively rRMSE and rBIAS values of 99 % and -42 % for July 1996 and values of 47 % and 10.9 % for July 1997. Moreover, in the study realised by Joseph *et al.* (2007), it was shown that kriging in CCA space with 21 stations led to better results than a simple linear regression (rRMSE values of 54.4 % versus 59.6 % and rBIAS of -13 % versus 20 %) for the estimation of the mean annual streamflow. In both cases, kriging in multivariate space with a relatively low number of stations gives better result than regression analysis in terms of error estimations and shows less bias in the model.

In this present study, kriging was performed with a sparse network of stations (12) on a relatively large territory. With so few measurements, it becomes difficult fitting a semivariogram model. Attempts to further decrease the number of stations led to a very poor fit of the theoretical semivariogram and therefore no modeling capability. In fact,

the variability in the performances of the model observed for CCA between 1996 and 1997 is explained by the quality of the semivariogram fitting between these two years. Even if all metrics were significantly correlated with monthly maximum temperatures, with so few measurements, even one station with poorer metric measures is sufficient to introduce outliers in the experimental semivariograms. Therefore, the kriging performances can decrease largely between years because the correlation structure is highly dependant on all observations with only 12 stations.

In addition to the cross validation provided, Table 4-4 illustrates how to predict stream temperature at ungauged stations. By extracting physiographical information for each ungauged stations and by using the PCA algorithm of July 1996 constructed with only 12 stations in our case, PCA coordinates (score values) can be calculated for any location and stream temperatures can be estimated. In fact, once the physiographical information is known it becomes possible to use the PCA algorithm or the CCA algorithm for water temperature estimation anywhere in the study area, at any time period for which we have a minimum (e.g. 12 in our case) number of measurements. There were no temperature measurements to validate the estimated temperatures in Table 4-4. However, estimated values appeared to be in the same range as those observed in the nearby station belonging to the same cluster. For this reason we can assume that each cluster represents a relatively homogenous geographical region in terms of physiographical characteristics. For ungauged stations outside of these 3 regions, water temperature estimation seems to be appropriate, i.e. no unreasonable values were predicted.

Finally, it is important to recall that one of the main sources of uncertainty in the interpolation stage is network density and spatial distribution of stations, which can be



related to kriging variance (St-Hilaire *et al.*, 2003). As seen in Figure 4-1, the stations used in the present study were not uniformly dispersed on the drainage basins but rather formed clusters with relatively large regions with no measurements in between these clusters. This has direct implications on neighbourhood definition and semivariogram calibration. The user must find the appropriate trade-off between good local interpolation and high variance in low density areas.

#### **4.6. Conclusions and future work**

This study focused on developing a new stream temperature model by using a multivariate geostatistical approach. A physiographical space-based estimation technique was used for the interpolation of stream temperature rather than the usually employed geographical space. It was demonstrated that both multivariate methods, i.e. PCA and CCA, can be employed to construct the physiographical space and then used to build semivariograms for characterizing the correlation structure and ultimately, to perform spatial interpolation. A better performance was observed using the CCA algorithm. In addition, CCA required less information to provide more satisfactory results. Future work should include the possibility of testing the method on a much denser network of temperature loggers and compare the performance of each algorithm as network density decreases. The optimal choice of metrics may also change as a function of network density and a sensitivity analysis may lead to an optimal design.

The present study focused on two relatively large contiguous drainage basins (Moisie and Ste-Marguerite) located on the Quebec North shore. Hence the physiographical features of both basins are somewhat similar. A future study should test the approach when there

is potentially greater physiographical variability (e.g. various drainage basin areas, different climate, etc.). At the other end of the physiographical spectrum, the method could also be tested on a sub basin or river reach for which the selection of variables may be different.

Other multivariate approaches should be considered in subsequent analyses to construct a more informative space in which kriging could be performed. As example, nonlinear methods such as Principal Curves Analysis, Curvilinear Component Analysis, Nonlinear Canonical Correlation Analysis, Nonlinear Redundancy Analysis or Nonlinear Principal Predictor Analysis could be used (Yin, 2007). Most of these methods are explicitly designed for dimensionality reduction.

It is believed that this model can become a powerful tool to understand thermal regime of rivers which is often essential for best management of aquatic resources including fisheries.

#### **4.7. Acknowledgements**

The authors are grateful to J. Poulin for the extraction of physiographical data and to the two anonymous reviewers and associate editor for their insightful comments. This project was financed by NSERC and Hydro-Québec.

**Table 4-1** Station names and their associated broad geographical regions and monthly maximum temperatures. Data used to construct kriged maps of July 1996 and 1997

July 1996			July 1997		
Station name	Geographical region	T (°C)	Station name	Geographical region	T (°C)
TCAR_0543	North Moisie	18.36	TCAR_0544	North Moisie	18.35
TPEK_0352	North Moisie	17.50	TCAR_0543	North Moisie	18.67
TPEK_0350	North Moisie	18.55	TPEK_0352	North Moisie	17.14
TPEK_0007	North Moisie	18.29	TPEK_0350	North Moisie	17.99
TMOI_0008	North Moisie	18.06	TPEK_0007	North Moisie	18.42
TMOI_0509	South Moisie	17.19	TMOI_0008	North Moisie	18.50
TMOI_0507	South Moisie	16.58	TMOI_0509	South Moisie	17.72
TMOI_0555	South Moisie	17.63	TMOI_0507	South Moisie	16.41
TNIP_0554	South Moisie	17.49	TNIP_0554	South Moisie	16.88
TSMA_0586	Ste-Marguerite	17.03	TSMA_0586	Ste-Marguerite	17.43
SMA_0586	Ste-Marguerite	16.99	SMA_0586	Ste-Marguerite	17.42
SMA_0346	Ste-Marguerite	17.22	TSMA_0588	Ste-Marguerite	17.06

**Table 4-2** Physiographical variables characterising each station estimated using GIS

METRICS	UNITS	NOTATION
Latitude	° ' "	LAT
Longitude	° ' "	LONG
Drainage area	km <sup>2</sup>	DA
Distance to the closest tributary from station	m	CT
Distance to the closest lake from station	m	CL
Area of the closest lake	m <sup>2</sup>	ACL
Mean stream azimuth	°	MSA
Mean stream azimuth of all upstream tributaries	°	MSAT
Azimuth between station and closest lake	°	ASL
Altitude at the station	m	A
Maximum altitude of drainage basin	m	MAXA
Mean altitude of the drainage basin	m	MEANA
Mean slope of the drainage basin	%	MS
Local slope at the station	%	S
Stream order		SO
River kilometer	km	RK
Forest cover on drainage basin	%	FC
Percentage of area covered by lakes and marches	%	PLM

**Table 4-3** Correlation coefficients between physiographical variables and monthly maximum water temperatures of July 1997 and associated P-values

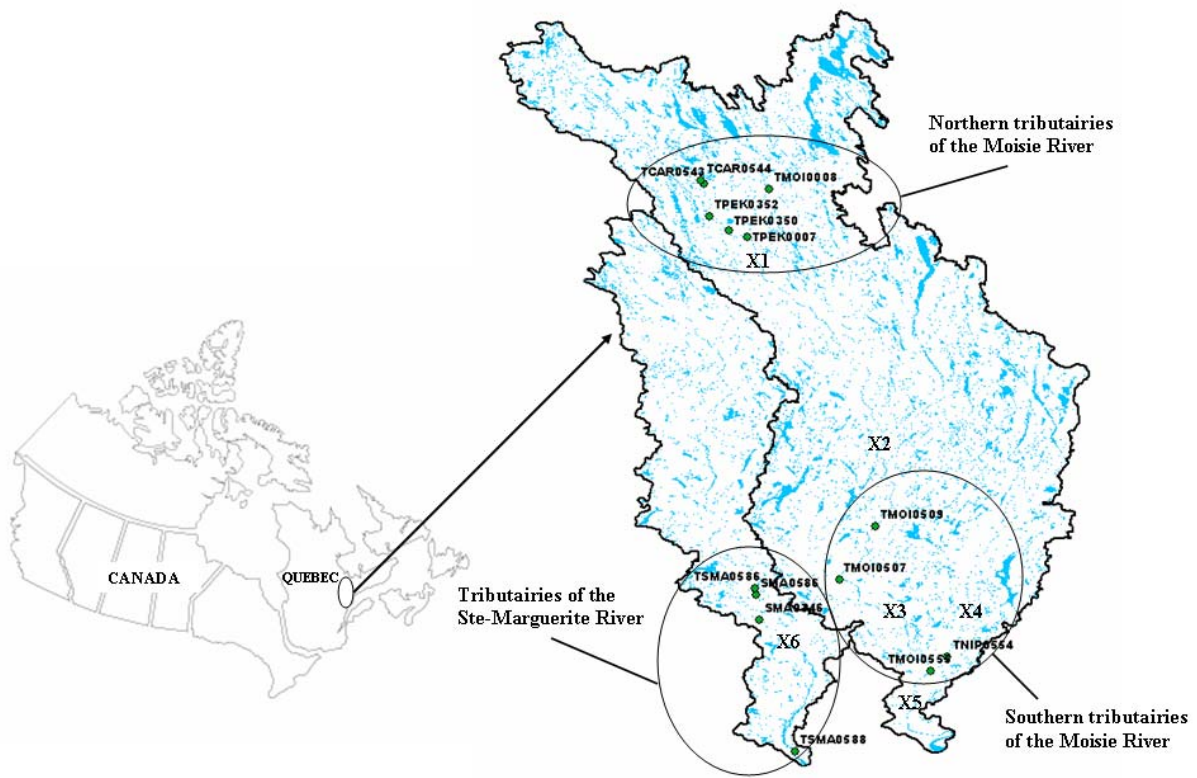
METRICS	NOTATION	CORRELATION COEFFICIENTS	P-VALUE
Latitude	LAT	0.74	0.0055
Longitude	LONG	-0.59	0.0417
Drainage area	DA	-0.61	0.0337
Area of the closest lake	ACL	0.70	0.0109
Altitude	A	0.58	0.0489
Maximum altitude of drainage basin	MAXA	-0.66	0.0198
Mean slope of the drainage basin	MS	-0.67	0.0176
Stream order	SO	-0.60	0.0392
River kilometer	RK	-0.65	0.0220
Forest cover on drainage basin	FC	-0.66	0.0189
Percentage of area covered by lakes and marches	PLM	0.69	0.0139

**Table 4-4** Physiographical variables extracted to estimate monthly maximum stream temperatures using the PCA algorithm at ungauged stations (Figure 4-1) for July 1996

Stations	LAT	LONG	DA	A	MAXA	MS	SO	RK	FC	PLM	PC1 (score)	PC2 (score)	T(°C)
X1	5771721	652543	6272.4	325.0	904	5.30	7	5967230	82.1	12.6	-3.36	-1.32	<b>18.1</b>
X2	5683390	689416	10507.7	152.0	915	6.79	7	10451370	85.2	10.6	-0.61	-1.50	<b>17.6</b>
X3	5623021	693833	13786.4	46.0	986	9.09	7	14154560	86.6	10.2	1.24	-1.22	<b>17.4</b>
X4	5606334	709213	4124.7	46.0	901	12.69	6	4573507	87.8	10.9	-0.16	0.11	<b>17.6</b>
X5	5590114	700899	18871.3	18.0	1009	10.50	7	19594020	87.1	10.2	2.55	-1.49	<b>16.7</b>
X6	5607757	666115	5196.9	65.0	1055	11.36	6	5893806	7.2	91.6	0.67	2.37	<b>17.1</b>

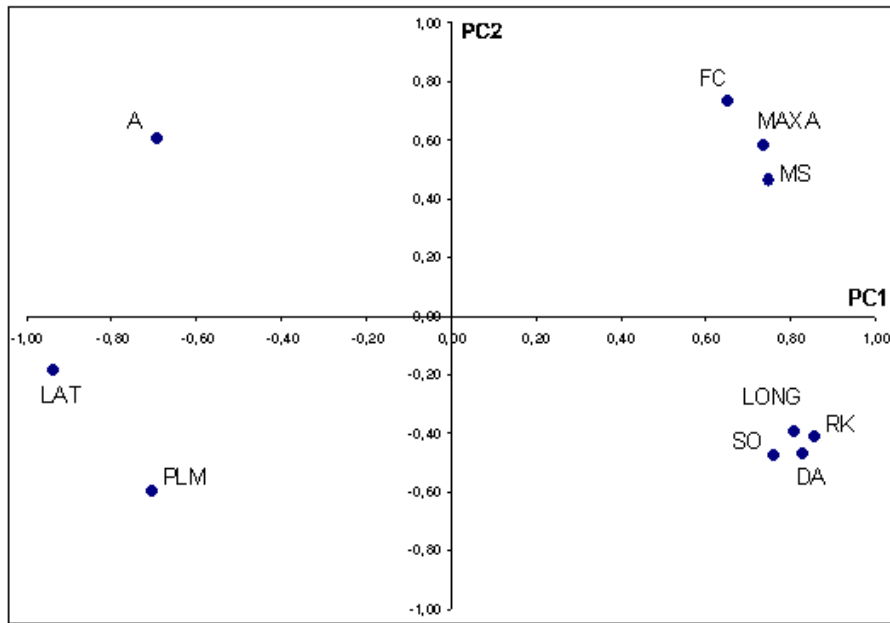
**Table 4-5** Cross-validation results

	Kriging in the PCA Physiographical Space		Kriging in the CCA Physiographical Space	
	July 1996	July 1997	July 1996	July 1997
rBIAIS (%)	-0.38	-2.03	1.14	-2.54
rRMSE (%)	20.89	19.71	5.17	18.49

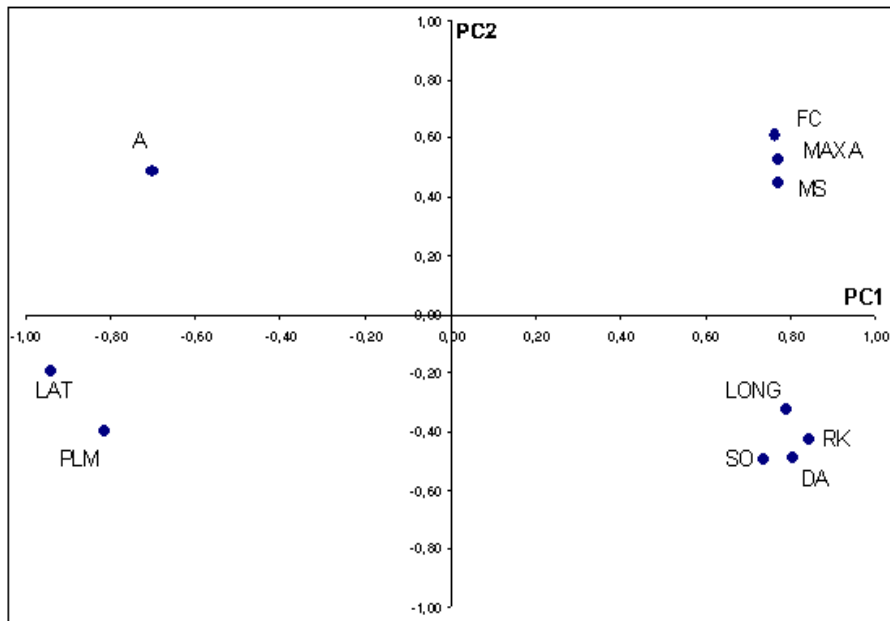


**Figure 4-1** Location and map of the Moisie and Ste-Marguerite rivers drainage basins with temperature monitoring stations grouped into three broad geographical regions. X marks indicate locations of ungauged stations

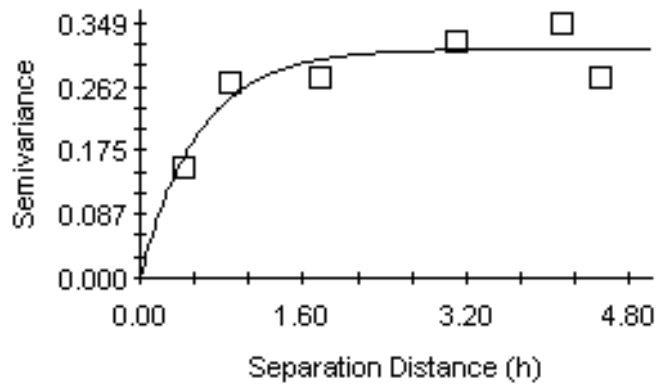
A)



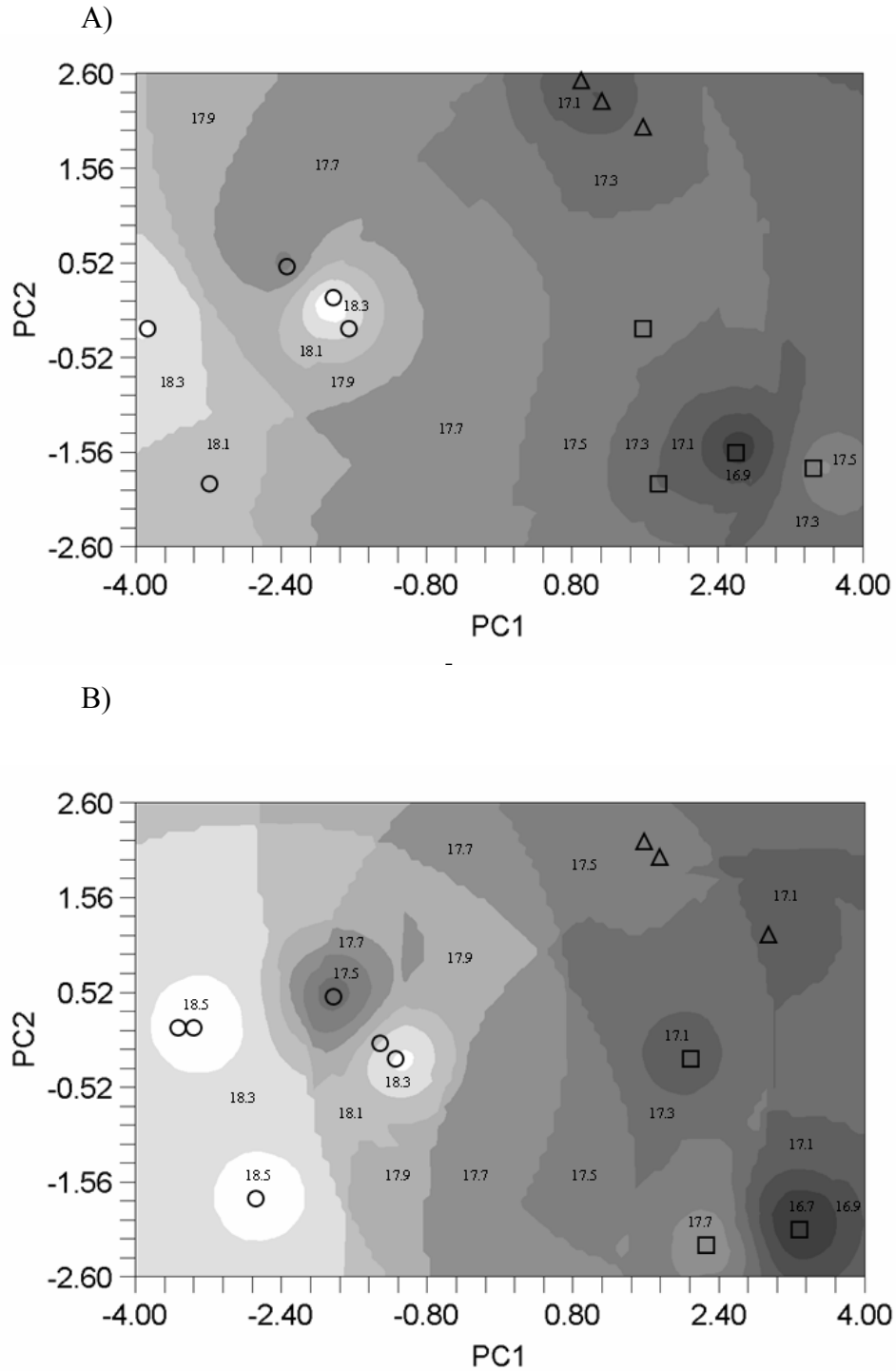
B)



**Figure 4-2** Projection of significantly correlated physiological variables in PCA space work set loading. A) July 1996. B) July 1997

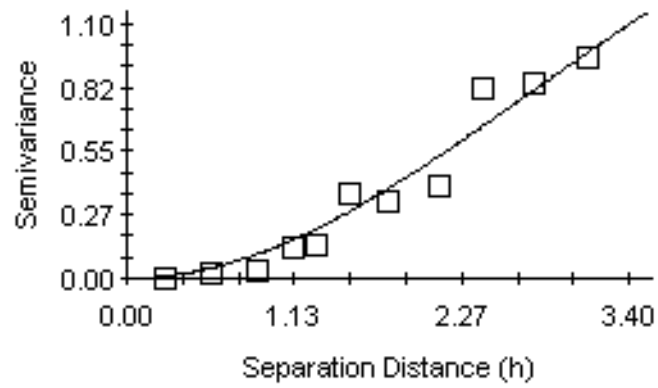


**Figure 4-3** Experimental semivariogram of July 1996 used to kriging in PCA space. Exponential model with nugget = 0.0001, range = 1.725 and sill = 0.314



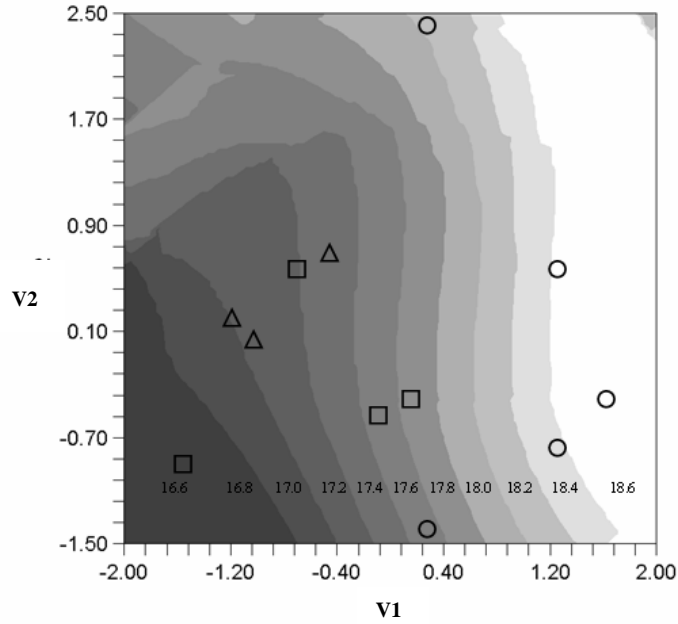
**Figure 4-4** Interpolated monthly maximum water temperature in PCA space performed with ordinary kriging. **Triangle:** tributaries of the Ste-Marguerite River. **Circle:** northern tributaries of the Moisie River. **Square:** southern tributaries of the Moisie River (see Figure 4-1 to identify geographical regions). A) July 1996. B) July 1996. **Note:** station locations were not identical for both years (see Table 4-1)



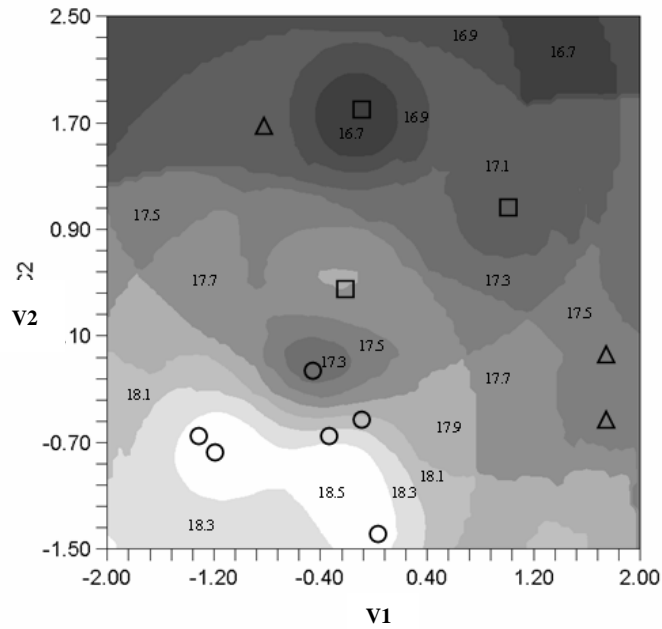


**Figure 4-5** Experimental semivariogram of July 1996 used to kriging in CCA space. Gaussian model with nugget = 0.001, range = 6.639 and sill = 2.011. Search radius was limited to  $h < 3.2$

A)



B)



**Figure 4-6** Interpolated monthly maximum water temperature in CCA space performed with ordinary kriging. **Triangle:** tributaries of the Ste-Marguerite River. **Circle:** northern tributaries of the Moisie River. **Square:** southern tributaries of the Moisie River (see Figure 4-1 to identify geographical regions). A) July 1996. B) July 1996. **Note:** station locations were not identical for both years (see Table 4-1)

## **CHAPITRE 5**

### **DEUXIÈME ARTICLE VERSION INTÉGRALE ANGLAISE**



# Statistical tools for thermal regime characterization at river segment scale: case study of the Ste-Marguerite River

By

Nicolas Guillemette<sup>\*1</sup>

André St-Hilaire<sup>1,3</sup>

Taha B.M.J. Ouarda<sup>1</sup>

Normand Bergeron<sup>2,3</sup>

<sup>1.</sup> Chair in Statistical Hydrology, INRS-ETE,  
490 de la Couronne, Québec, G1K 9A9, Canada

<sup>2.</sup> INRS-ETE,  
490 de la Couronne, Québec, G1K 9A9, Canada

<sup>3.</sup> Centre Interuniversitaire de recherche sur le saumon Atlantique  
Québec

Manuscript to be submitted to:

River Research and Applications

September 2008

\* Corresponding author: email: [nico\\_guillemette@hotmail.com](mailto:nico_guillemette@hotmail.com) ; Tel (418) 654-2530 (4457)

Fax number: (418)-654-2600



## 5.1. Abstract

Suitable thermal fish habitats are constrained by both maximum and minimum temperature tolerances. A multivariate and geostatistical approach was developed to estimate stream thermal characteristics at the river segment scale. Data from 22 temperature monitoring stations obtained during summer 2007 were used to estimate monthly maximum temperature as well as thermal characteristics such as the number of events, the cumulative degree-days and the associated duration over specific temperature thresholds of 19°C and 21°C. The probability of exceeding these temperature thresholds has also been interpolated. The methodology relies on the construction of a physiographical space using physiographic and hydrological characteristics of gauging stations as inputs in a multivariate approach called canonical correlation analysis (CCA). A geostatistical interpolation technique, ordinary kriging, was subsequently used to perform interpolation in the physiographical space constructed using CCA. Results from this study were obtained for thermal characteristics estimated into two different interpolation spaces: 1) an 7 metrics space, and 2) an 8 metrics space. Cross-validation technique was performed and satisfactory results were obtained. Kriging thermal characteristics into the 7 metric space for a 19°C threshold lead to best results with Relative Root Mean Square Error (rRMSE) ranging between 9.66 % and 15.08 %. The study shows that kriging in a multivariate space is a promising tool for water resources managers, especially in cases where risk mapping for lethal or sub-lethal temperature thresholds may be required for a specific fish species.

## 5.2. Introduction

The role of thermal regime is important in fisheries management (Caissie 2006). For a given fish species, suitable thermal habitats are constrained by both maximum and minimum temperature tolerances (Mohseni *et al.*, 2003). For instance, Lund *et al.* (2002) showed that Atlantic Salmon (*Salmo salar*) parr have a preference for cool to moderate water temperatures in the range of 9 to 19°C while Elliot (1991) showed that the same juvenile species can survive for more than 7 days at a maximum temperature of 27.8°C but feeding ceases when temperature exceeds approximately 22°C. A number of studies have shown that climate change in North America may lead to an increase in the frequency of exceedance of suitable temperature range for cold water fish (Eaton and Sheller, 1996, Mather *et al.*, 2008). Recent attention has focused on the impact of climate change on stream temperature and potential fish habitat changes (Tung *et al.*, 2006; Hari *et al.*, 2006; Crozier and Zabel, 2006).

Different studies have also demonstrated the potential importance of thermal refugia in relation with flow regime regulated by dams (Sutton *et al.*, 2007) while other authors tried to establish relation with thermal refugia and fish abundance by considering upwelling groundwater and local influx of cooler subsurface waters (Ebersole *et al.*, 2001; Ebersole *et al.*, 2003).

This recent research effort has demonstrated that the need for water temperature measurements is likely going to increase rapidly. Unfortunately, the network of temperature gauging stations is relatively sparse in countries like Canada. Stream temperature prediction/estimation tools are therefore necessary to characterize the actual and future temperature variability needed to describe or predict the distribution of



freshwater species. Different approaches for establishing stream temperature guidelines such as temperature-duration curves (Brown and Hannah, 2008) or degree-day curves (Mather *et al.*, 2008) over a specific threshold have a great ecological significance in terms of fish distribution, thermal refuges characteristics and temperature variability. In fact, stream ecology modellers are faced with the possibility of choosing from a number of water temperature models to represent this precious information. Two broad categories are usually identified: (i) deterministic models (Morin *et al.*, 1994; Gu and Li, 2002; St-Hilaire *et al.*, 2003; Caissie *et al.*, 2007) and (ii) statistical/stochastic models (Mohseni *et al.*, 1998; Caissie *et al.*, 1998; Benyahya *et al.*, 2007; Chenard and Caissie, 2008).

Most statistical models developed in recent years have made use of the temporal or autocorrelation that exists in relatively long water temperature time series (e.g. Moatar *et al.*, 2006, Benyahya *et al.*, 2008). Very few studies have attempted to use the spatial correlation between water temperature measured at different sites within a river system, with the notable exception of Peterson and Sickbert (2006).

In addition to past temperature measurements, a number of other independent variables that explain water temperature availability have been identified. These explanatory variables may differ at different spatial scales. A study realized by Moore (2006) demonstrated that a set of predictor variables such as monthly maximum air temperature, annual precipitation, mean catchment elevation, logarithm of drainage area, percent glacier cover, percent lake cover and flow regime (regulated or not) leave an imprint on stream temperature, even for large catchments. At the relatively smaller scale of the river reach, Gardner *et al.* (2003) used a combination of distance metrics on the river network, including stream order to provide predictive capability to a 1-D geostatistical model on a

small drainage basin of 700 km<sup>2</sup>. Guillemette *et al.* (2008) expanded the use of geostatistics in water temperature estimation and demonstrated the feasibility of a 2-D geostatistical modeling of monthly maximum temperatures in multivariate space using 8 to 10 physiographical variables on a large basin system of 25 900 km<sup>2</sup> with a minimum number (10) of temperature recorders.

The aim of this study is to expand from the work of Guillemette *et al.* (2008) and Gardner *et al.* (2003). The present study aims to estimate water temperatures at ungauged sites at the stream segment scale (i.e. < 100 km of river length) and to identify appropriate physiographical variables as predictors of water temperature at that scale. The specific objectives of this study are: (1) To elaborate a reliable geostatistical model at the river segment scale using the appropriate physiographical variables as predicting factors of temperature variability. (2) To estimate model performances for different kriged variables such as monthly maximum temperatures as well as derived thermal characteristics that may be useful in stream habitat characterization.

### **5.3. Data and study site**

The Ste-Marguerite River is located on the Québec North Shore (Canada) and runs via two principal segments of 90 km (drainage area of 1120 km<sup>2</sup>) and 85 km (drainage area of 980 km<sup>2</sup>) respectively which merge 2.5 km before discharging a total drainage area of 2100 km<sup>2</sup> in the Saguenay River (Figure 5-1). Both segments take their source from the high peaks of the Valin Range located in the heart of the Saguenay – Lac St-Jean region. The drainage basin presents high topographic relief i.e. from 100 m to over 900 m elevation with relatively narrow and steep valleys where a number of cascades

characterizes the hydraulic regime. The land use of the basin is homogenous with approximately 95 % of vegetated area and 5 % of lakes and marches. Vegetation is characterized mainly by boreal forest with a predominance of coniferous trees. In some areas of the basin, vegetative patterns have been altered through fire and logging, the latter being considered as the principal land use throughout the drainage basin. Fish species of concern include Atlantic salmon (*Salmo salar*), Brook trout (*Salvelinus fontinalis*), American eel (*Anguilla rostrata*), as well as a number of cyprinidae and catostomidae.

Stream temperatures were measured continuously at 11 sites (Figure 5-1) within the study segments from July 4, 2007 to October 6, 2007. Each site was selected because it was at the confluence of the main river and a tributary that was believed to be a potential thermal refuge. Tributaries represent a relatively high temperature deviation with the main channel during hot summer months. At each site, one thermograph was moored in the tributary, a few meters upstream of the confluence, while a second one was moored upstream of the confluence, in the main stem. Temperatures were recorded every 15 minutes. Hence, a total of 22 thermographs were deployed. The name and geographical coordinates of each site (tributary + main channel) are reported in Table 5-1.

Descriptive water temperature statistics used in the study are: monthly maximums of daily temperatures, as well as thermal characteristics derived from hourly means such as cumulative degree-days and their associated duration, the number of events where a specific threshold is exceeded and finally the probability of exceeding specific thresholds. We chose two specific temperature thresholds of 19 and 21 °C as reference values.

A number of physiographical predictors were estimated using a Geographical Information System (GIS). They are reported in Table 5-2. These predictors were selected based on the following criteria: 1) known or perceived impact of the variable on river thermal regime and 2) data availability. Some of the predictors have been shown to explain spatial variability of temperature in previous studies (Gardner *et al.*, 2003; Guillemette *et al.*, 2008). Others were tested because of their relevance at the river segment scale. For example, the influence of geographical metrics such as latitude and longitude can be negligible at the segment scale, while hydraulic parameters or local topographical variables become more important to explain temperature variability at that scale. In the present study, 18 variables were selected (Table 5-2) by considering their ability to impact the temperature signature on a relatively small drainage basin (river segment scale) during the warmest months of the year. From the 18 pre-selected variables, eight metrics were significantly correlated ( $p$ -value  $< 0.05$ ) with selected temperature statistics. They are identified in bold character in Table 5-2.

#### **5.4. Statistical Methods**

The present study used spatial interpolation technique to estimate stream temperature in a mathematical multivariate space rather than a geographical space, as proposed by Guillemette *et al.* (2008). The approach consisted of constructing an orthogonal space by combining physiographical variables through Canonical Correlation Analysis (CCA). Chokmani and Ouarda (2004) used this approach to reduce the dimension of physiographical data sets while maximising the correlation with hydrological variables in a study on regional flood frequency analysis. The interpolation of temperature is

performed in a two-dimensional space defined by the first two physiographical canonical variates.

Canonical correlation analysis (CCA) is used to examine the relationship between two sets of random variables. In our case, the first matrix  $X$  of variables  $x_1, x_2, \dots, x_i$  includes temperature-derived variables while the second matrix  $Y$  of variables  $y_1, y_2, \dots, y$  represents the predictors such as geographical, hydrographical and physiographical variables (see Table 5-2). CCA involves searching for linear combinations ( $U_1, U_2, \dots, U_i$ ) of variables in  $X$  that have the maximum possible correlation with linear combinations ( $V_1, V_2, \dots, V_i$ ) of variables in  $Y$  (Manly, 2004).  $U_i$  and  $V_i$  is a pair of canonical variates defined by equation (1) that maximises the sample canonical correlation (equation 2) and have unit sample variances.

$$\begin{aligned} U &= a'X \\ V &= b'Y \end{aligned} \tag{1}$$

$$r_{U,V} = \frac{a'S_{12}b}{\sqrt{a'S_{11}a}\sqrt{b'S_{22}b}} \tag{2}$$

The canonical coefficients vectors  $a$  and  $b$  in equations (1) and (2) are respectively associated with the first and second set of variables ( $X$  and  $Y$ ).  $S_{12}$ ,  $S_{11}$  and  $S_{22}$  are the sample covariance matrices consistent with the case of the initial variables. Each pair of sample canonical variates is uncorrelated with the previous  $k-1$  sample canonical variates (Johnson and Wichern, 2007).

Once CCA is performed, we retained the first two physiographical orthogonal vectors ( $V_1$   $V_2$ , which are most highly correlated with thermal variables) built from  $Y$  to construct a two-dimensional physiographical space where stream temperature characteristics can be

interpolated using kriging. The first step is to model the spatial correlation structure between stations as a function of separation distances with the experimental semivariogram:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (z_{x_i+h} - z_{x_i})^2 \quad (3)$$

The semi-variance model obtained with equation (3) was constructed using the multivariate coordinates established with the Euclidian distances of the first two CCA variates ( $V_1$  and  $V_2$ ) and with estimated variable of interest ( $z$ ). In equation (3),  $N(h)$  is the number of data pairs at a separation distance  $h$  which have an observed values  $z(x)$ . Once the experimental semivariogram is obtained, a theoretical model can be fitted (spherical, exponential, or gaussian functions) using three parameters: the nugget effect ( $C_0$ ), the sill ( $c$ ) and the range ( $a$ ). The nugget effect describes the occurrence of discontinuity at the origin of the semivariogram that may be caused by dissimilar sample values at short inter-station distances (Isaaks and Srivastava, 1989). The sill is the plateau reached by  $\hat{\gamma}(h)$ , which indicates a value of semi-variance beyond which there is essentially no spatial correlation structure. Finally, the range represents the distance over which the observed values are correlated.

The kriging estimator is a weighted average of the observed values  $z(x_i)$  which is used to estimate the value of  $z(x_0)$ , identified at a specific location  $x_0$  where there are no measured values. The model is denoted

$$\hat{z}(x_0) = \sum_{i=1}^{N(h)} \lambda_i z(x_i) \quad (4)$$

Where  $\lambda_i$  are the weights of the estimator that minimize the variance of the estimation error under the constraint on weights (for ordinary kriging):

$$\sum_{i=1}^{N(h)} \lambda_i = 1 \quad (5)$$

The values of  $\lambda_i$  are obtained by solving this linear kriging system. In the present study, we also used probability kriging to estimate the probability of exceeding a specific threshold  $z_k$  at a given location. To achieve this, the random variable  $z(x)$  is transformed into a dichotomic indicator variable, as follows:

$$I(x; z_k) = \begin{cases} 1, & \text{if } z(x) \geq z_k, \quad k = 1, 2, \dots, m \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

For probability kriging, a semivariogram ( $\gamma_I$ ) is adjusted for indicator values as a function of separation distance. If a multitude of thresholds are used, a discrete distribution of the original variable can be obtained. To perform kriging on derived thermal characteristics and

## 5.5. Kriging performances

The error associated with the spatial interpolation of all estimated variables in the CCA space was the principal criterion to evaluate model performance. A cross validation was performed using a leave-one-out resampling technique. The relative mean bias (rBIAS) as well as the relative root mean square error (rRMSE) were used as specific performance evaluation criteria. These indicators are defined as follows:

$$\text{rBIAS} = \frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right) \quad (7)$$

$$\text{rRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{z_i - \hat{z}_i}{\Delta z_{\max}} \right)^2} \quad (8)$$

where  $z_i$  and  $\hat{z}_i$  are respectively the observed value and the estimated value of the thermal variable of interest (i.e. the sum of kriged residual and trend component) at station  $i$ . As described in Guillemette *et al.* (2008), equations (7) and (8) differ from the usual definition of rRMSE and rBIAS because the maximum range ( $\Delta z_{\max} = Z_{i \max} - Z_{i \min}$ ) of each original data set is used as denominator. The estimation errors associated with probability kriging cannot be calculated using equations (7) and (8) because of the binary nature of the variable. In this case, the performance was assessed by considering BIAS and RMSE:

$$\text{BIAS} = \frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i) \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2} \quad (10)$$

## 5.6. Results

Temperature variables were best modeled at the segment scale by adding metrics that describe local phenomena such as discharge (Q), mean altitude (MEANA), distance to closes lake (CL) and the canopy cover at the station (CC). All thermal characteristics were best estimated using significantly correlated metrics (Table 5-2) because of the semivariograms correlation structure improvement obtained with those variables in



comparison with uncorrelated variables. For this reason, kriging was performed in CCA space constructed using significantly correlated variables to increase modeling capability. Satisfactory results were found using two different CCA spaces. The first one was constructed using the maximum number (8) of significantly correlated variables (Table 5-2), while in the second one, only seven significantly correlated variables were used to define the canonical space. In fact, model parsimony dictates that the number of metrics be as small as possible without deteriorating the correlation structure between the two groups of variables (Stevens 1986). The metric called *CL* was the least significantly correlated metric (albeit slightly less than the others) with all modeled variables. Therefore, we removed this metric and examined the correlation structure between physiographical variables and temperature statistics as well as the estimation errors (*rRMSE*). Satisfactory results were found but attempts to use less than 7 metrics to construct CCA space lead to an important decrease in model performances.

As mention earlier, we chose two specific temperature thresholds of 19 and 21 °C as reference values to performed kriging on derived thermal characteristic and indicator variables. In both cases, threshold values have been selected because of their potential biological significance considering poikilotherm fish habitats (Lund *et al.*, 2002, Hodgson and Quinn, 2002). Therefore, the kriging maps associated with these thermal characteristics allow to identify areas of potential thermal stress or thermal refugia as a function of these specific temperatures thresholds.

To determine and measure the spatial structure of all kriging input (thermal variables) data into the CCA space, isotropic experimental semivariograms were calculated. Special attention was paid to the hypothesis of stationarity in CCA space. One of the basic

assumptions in ordinary kriging is that the modeled random variable has a constant mean at the neighbourhood scale (Arnaud and Emery, 2000). The existence of trends of temperature variables with CCA variates was therefore checked. The use of CCA, which maximises correlation between water temperature and physiographical variables made it more likely that a trend would exist. This was confirmed by significant correlations between each modeled water temperature variable and  $V_1$  and/or  $V_2$ . For this reason, the linear trend was first removed and the interpolation in CCA space was performed on residuals. Once semivariograms and associated kriging maps have been established using residual values, the trend was added to construct the final kriging map.

Table 5-3 and 5-4 provide all kriging error estimations using 7 and 8 metrics space. Results were slightly better for all modeled variables estimated in the 7 metric space for a 19°C threshold. They are introduced in the next section.

### **5.6.1. Monthly maximum temperatures**

The first set of modeled water temperature variables used in the CCA were maximum monthly temperatures of warm months (i.e. July, August and September) of 2007. The stream order (SO), the maximum altitude of the drainage basin (MAXA) and the river kilometer (RK) represents respectively the highest canonical factor loadings associated with  $V_1$ , while  $V_2$  is dominated by drainage area (DA) and the canopy cover (CC) respectively. When comparing results obtained in CCA space constructed with 7 versus 8 variables, the best results were obtained with the former with rRMSE values ranging between 14.04% and 19.31% and rBIAS values ranging between -0.29% and -1.29% (Table 5-3).

Figures 5-2 and 5-3 show associated results using the 7 metric space. Figure 5-2 illustrates isotropic semivariograms adjusted to define the spatial correlation structure of residual values for these three warm months. All semivariograms showed a relatively strong correlation at short distances in CCA space (0 to 2.5) and reached a specific sill at different range values. The fact that in most cases, the sill is well defined is an indication that trend removal was successful. The kriged maps shown in Figure 5-3 illustrate the estimated values as well as the stations location with their associated names. The position of stations in the CCA space for the three months is identical because orthogonal vectors ( $V_1$  and  $V_2$ ) representing the CCA coordinates of each station are obtained from the same analysis using the first and second set of variables described above. The right side of each map of Figure 5-3 defined for  $V_1$  values ranging between 0 and 1.6 contained 10 of the 11 stations located in the main stem while the 10 other stations (located in cooler tributaries) are located on the left side of kriging maps for  $V_1$  values ranging between -2.5 and 0. Only two stations (8T and 8P; Table 5-1) do not respect this pattern. Station 8T, the Xavier Tributary is the widest brook and has the largest drainage area of the 11 stations located in tributaries. For this reason, the stream temperatures of both stations (8T and 8P) are similar during hot summer months explaining their respective positions in CCA space.

Each maps of monthly maxima show lower temperature conditions for  $V_1$  values ranging between -1.4 and -0.7 and  $V_2$  value between -1.7 and 0.6, which corresponds to the area in CCA space where potential thermal refugia for poikilotherm fish could be located. Stations “6T”, “9T” and “11T” are posted into this cold area with monthly maximum temperatures oscillating between 17.03 and 18.08 °C for the warmest month which is

July. In contrast, sites defined as potential stressful area for fish are located in *VI* values ranging between 0.8 and 1.6 and *V2* value between 1.5 and 2 where stations called “9P”, “11P” and “10P” are posted. During July, maximum temperature of these stations reached values between 25.91 and 26.97 °C.

### **5.6.2. Derived thermal characteristics**

In this section, kriging results are reported for three different water temperature variables estimated as a function of selected temperature thresholds (19 and 21°C). These variables used as the first set in the CCA are:

- Number of events exceeding a specific temperature threshold (*NEV*)
- Cumulative degree-days over a specific temperature threshold (*DD*) in °C-days
- The duration in days of events exceeding the temperature threshold (*DUR*) in days

Analyses have been performed using hourly moving averages calculated from temperature values recorded at a 15 minutes interval. Hence, a threshold exceedance was identified when the hourly mean was greater than the threshold. The units of *DD* and *DUR* were respectively °C-days and days. Again, the interpolation was performed in the CCA space using 7 and 8 original physiographical variables that were significantly correlated (Table 5-2) with derived thermal variables. The events that occurred during the months of July, August and September were used to estimate the three aforementioned variables. The CCA indicated that *VI* is dominated by the drainage area (DA), the river kilometer (RK) while *V2* is more associated with the mean altitude of the drainage basin (MEANA) and the discharge (Q). Best results were found with the 7

metric space for a 19 °C threshold. We found rRMSE values ranging between 9.66% and 15.08% and rBIAS values ranging between -0.54% and 0.62% (Table 5-3). The *DD* map gives the best estimations errors with rRMSE of 9.66 % and rBIAS of 0.32 %

Figure 5-4 and 5-5 show the isotropic semivariograms as well as kriging maps for these three derived thermal characteristics using a 7 metric space and a 19°C threshold. The CCA performed to define the interpolation space was the same for the three estimated variables. For this reason again, the stations locations into the CCA space are the same for the three kriged maps of figure 5-5. These maps can be used to define thermal habitat as a function of stream temperature exceeding 19 °C during the three warmest months of summer 2007. The three maps of Figure 5-5 show respectively 19 stations on Figure 5-5A and 20 stations on Figure 5-5B and 5-5C. The list of stations included in the analysis changed slightly (one to three stations were removed) because they were considered as outliers and prevented proper variogram fitting. The reason for this erratic behaviour is not fully understood. It is believed that the physiographical data gathered at this site were either incorrect or insufficient to fully describe the local thermal regime for this station.

With a *NEV* exceeding 19°C ranging between 0 to 6, *DD* oscillating between 0 to 4.85 and a *DUR* ranging between 0 to 4.71 days, stations *10T*, *2T*, *9T* and *6T* constituted the coldest sites of the studied river system. In contrast, the warmest sites were stations *9P*, *10P* and *11P*, with a *NEV* oscillating between 23 and 37, *DD* ranging between 48 and 49.2 °C-days and a *DUR* ranging between 20.4 and 21.8 days. These thermal characteristics indicate variability across sites and provide significant information to qualify fish habitats considering specific temperature thresholds.

### 5.6.3. Probability kriging

Probability kriging is another tool to characterize thermal regime of rivers. In this case, the variable of interest (e.g. temperature statistics such as daily maximum, monthly maximum or daily mean temperature) was transformed into a dichotomic indicator variable. Indicators were used as the first set of variables in the CCA and the kriging interpolation was performed. The resulting kriging maps provide the probability of exceeding a specific threshold at a given location. Monthly maximum temperatures of July, August and September were used as modeled variables. As in the previous section, model performance was assessed with the same criteria for CCA space built using 7 and 8 metrics and for the same two selected thresholds (19 and 21 °C). Because monthly maximum temperatures were used to construct the indicators, station locations in CCA space were the same as those obtained in the first result section. For the same reason, metrics of influence associated with *V1* are the stream order (SO), the maximum altitude of the drainage basin (MAXA) and the river kilometer (RK), while *V2* is dominated by drainage area (DA) and the canopy cover (CC) respectively. Once again, best results were obtained for a threshold of 19°C using a 7 metric space. In that case, RMSE values ranged between 0.22 and 0.31 and BIAS values ranging between -0.02 and -0.04 (Table 5-4).

Figures 5-6 and 5-7 illustrate respectively semivariograms and probability kriging maps for a 7 metric space and a 19°C temperature threshold. During July and August 2007, probability kriging were identical because the 19°C temperature thresholds were exceeded at exactly the same stations. This situation illustrates the importance of

threshold selection. At higher thresholds, the spatial distribution of exceedance probability differed for the two months.

On the first two maps shown in figure 5-7, the majority of stations have approximately 100% chance of exceeding a temperature of 19°C while the three stations called *11T*, *10T* and *6T* have a probability of 0 % of exceeding this same threshold during July and August respectively. In contrast, the third map of Figure 5-7 illustrates the probability distribution for September. The majority of stations have a 0% probability of exceeding 19°C while only stations *9P* and *10P* have 100% chance of exceeding this same temperature threshold.

## **5.7. Discussion**

This study focused on the estimation of various thermal characteristics at the river segment scale, using a relatively small number of monitoring stations and physiographical information. The use of a geostatistical model was shown by Guillemette *et al.* (2008) to be an appropriate tool to interpolate temperature extremes at a much larger scale, i.e. the entire drainage basin, by creating a CCA physiographical space in which kriging can be performed.

The present study confirmed that the same methodology can be used at a river segment scale. The study focused on temperature extremes but the approach could be generalized to other variables of interest to characterize stream habitats such as potential thermal refugia. One of the main methodological differences between temperature modeling at the two different scales (drainage basin versus river segment) is the choice of the physiographical variables capable of influencing the thermal regime during hot summer

months. As described in the results, temperature variables were best modeled at the segment scale by adding metrics that describe local phenomena such as discharge (Q), the canopy cover at the station (CC) and the distance to closest lake (CL) when using a 8 metrics space (see section below). The mean altitude of the drainage basin (MEANA) has also been added and gives a good average of the topography area at this scale, which is not necessary true for a large drainage basin scale. On the other hand, four metrics are common to both modeling scales and remain significantly correlated with thermal variables: drainage area (DA), the maximum altitude of the drainage area (MAXA), the stream order (SO) and finally, the river kilometer (RK). An analysis of the canonical factor loadings, i.e. the correlation between the canonical variates (*V1* and *V2*) and the temperature statistics reveals that *V1* is dominated by the stream order (SO), the river kilometer (RK), the drainage area (DA) and the maximum altitude (MAXA). It is interesting to note that communality of metrics at both scales is more strongly associated with *V1*. Most of these four metrics help to characterize the size of the river at a given location (stream order, river kilometre, drainage area). Consequently, adding metrics used only at the river segment scale (Q, MEANA, CC and CL) are strongly associated with *V2*. We note here that *V1* links variables associated to drainage basin and *V2* reflects variables associated to local phenomena observed for the river segment scale.

Only the variable Q shows a temporal relation. In the present study, an average discharge was measured at each station that represents typical hydraulic conditions during hot summer months. This average measure was used to contrast the stations and was therefore imposed as time-invariant. In fact, considering temporal metrics such as discharge, air temperature or precipitation could be beneficial for semivariogram fitting



and may decrease estimation error. However, the resulting thermal estimations become time-dependant and would only be reliable for a specific short period and hardly transferable to other temporal scales if the variability of metrics is large throughout the year.

The rRMSE and rRMSE values (Table 5-3 and 5-4 respectively) indicate that in most of the cases, ordinary kriging performed into the 7 metric space excluding the variable called distance to the closest lake (CL) give better results than modeling with the eight significantly correlated metrics of Table 5-2. In fact, model performances are relatively similar (rRMSE around 15 %, except for two values > 20 %) but the theoretical semivariograms fitting was better with seven than with eight metrics. One potential reason for this difference is even if the metric *CL* is significantly correlated with thermal characteristics, it introduced outliers in the experimental semivariograms of temperature residuals. These outliers may be associated with the fact that *CL* does not account for lake volume, lake outflow or lake water temperature, which may have an equal or greater impact on river temperature than distance. The rRMSE values are also lower for a 19°C threshold in comparison with the 21°C threshold. The spatial correlation structure between stations in CCA space is altered depending of the selected temperature threshold. One simple explanation is that there are fewer exceedances for the 21 °C threshold and hence a relatively large number of zero values are found when this threshold is used. A sensitivity analysis on appropriate metrics that must be incorporated to the model considering multiple temperature thresholds could be an interesting first step when using a peak over threshold approach to define events. Finally, detailed analysis reveals that high RMSE values are not directly associated with station spatial distribution in the

physical space. Outliers introduced in the experimental semivariogram are in most of the case associated with smaller correlation between the thermal variable and a specific metric.

Changing the original data set to a dichotomic variable may be seen as a loss of information. However, it may allow for the inclusion of a greater number of stations. There may still be a number of locations where temperature monitoring is done using old max-min thermometers, or some sort of indirect empirical observations measure that are related to threshold exceedance. RMSE values for probability kriging are also expressed in terms of probability. Hence, a RMSE of approximately 0.2 (1 chance out of 5) is considered acceptable, but RMSE values of 0.47 or nearly 50% chance of error is definitely too high. The advantage of this approach is that probability kriging can be used in hazard mapping considering lethal or sub-lethal temperature thresholds.

The choice of 19°C and 21 °C temperature thresholds correspond to interesting options for stream ecologists who study thermal stress on poikilotherm fish and thermal mapping. In the study area, the simultaneous study of Figures 5-3, 5-5 and 5-7 indicates that any site with physiographical characteristics similar to those measured at stations *9T*, *6T* and *11T* (the coldest sites) may be identified as potential thermal refugia for salmonids. Stations *9P*, *10P* and *11P* define the warmest areas that could lead to thermal stress if fish are present during hot summer months and any other location on the main stem with similar hydrological and physiographical characteristics should also be considered as potentially stressful areas. The measurement pattern established on the study site led to the conclusion that colder areas are evidently into tributaries and warmer areas into the

main stem, as expected. High and low values of metrics DA, RK and Q mainly distinguish warmer from colder sites into the system.

Using derived thermal characteristics which are *NEV*, *DD* and *DUR* as modeled variables lead to good results for a 7 metric space with rRMSE comprised between 9.66 % to 15.08 % for a 19°C threshold and rRMSE ranging between 12.07 % to 19.75 % for a 21°C threshold. These new variables estimated in the CCA space reflect the nature of thermal characteristics (high or low variability, consistency) and provide an indication of the thermal heterogeneity among main stem and tributaries throughout the summer. The three derived thermal characteristics (*NEV*, *DD*, *DUR*) associated with the seven metrics used to construct the CCA space have the potential to indicate and separate habitat of different quality for specific fish species.

Perhaps the most interesting feature of the maps is firstly to bring stations all together in a mathematical space considering simultaneously seven or eight variables that impact significantly thermal signature which is impossible with so few stations at a physical scale of 2100 km<sup>2</sup>. Moreover, kriging maps allow the identification of thermal refugia at ungauged stations. By extracting significantly correlated metrics (Table 5-2) at ungauged sites, it becomes possible to identify the location of such a site into CCA space (Figures 5-3, 5-5 or 5-7) and thereby acquire an estimate of the temperature variable of interest (monthly maximum, derived thermal characteristics or probability). In fact, all thermal information extracted from gauged and ungauged stations provide untimely insights into thermal patterns of the studied river system.

Finally, it is important to recall that CCA is a useful method to reduce complex data sets and to maximise the correlation between two sets of variables. The CCA space constructed to interpolate thermal variables leads to satisfactory results and illustrates the possibility of reproducing regions with distinct or similar thermal signatures. The number of stations (19 to 22) as well as the number of metrics used (7 or 8) reduce the model complexity. Indeed, a relatively small effort is required in terms of metrics extraction from GIS and the required logger density at the stream segment level (22 on two 60 km segments) compares advantageously with the density required by Gardner *et al.* (> 70 on 161 km hydrographic network) who interpolated in geographical space. The user must find an appropriate trade-off between measurements pattern in terms of number of loggers used to get a good local and global interpolation and use a minimum number of metrics without deteriorating the correlation structure which is directly associated with semivariogram calibration and kriging performances.

## **5.8. Conclusions**

This study focused on the use of a multivariate geostatistical model to estimate different thermal characteristics at the river segment scale. It was demonstrated that some of the physiographical variables, different from those used at the basin level to construct the CCA interpolation space, are required at the segment scale. An optimal number of metrics (7) associated with a logger density between 19 and 22 led to adequate correlation structures and consequently to satisfactory results using ordinary and probability kriging. The latter may prove to be an interesting tool for water resources managers, especially in cases where risk mapping for lethal or sub-lethal temperature

thresholds may be required for a specific fish species. The leave-one-out cross validation has shown that estimates have an acceptable error range.

Finally, the approach may be used to estimate temperature at ungauged stations.



**Table 5-1** Geographical coordinates of the 22 thermographs deployed at 11 sites with their associated names and locations

Site	Location Name	Notation	Latitude	Longitude
1	Resimond T*	1T	48° 25.888' N	70° 26.101' W
	Resimond P*	1P	48° 25.872' N	70° 26.018' W
2	Fosse 58 T	2T	48° 25.805' N	70° 25.683' W
	Fosse 58 P	2P	48° 25.799' N	70° 25.666' W
3	Epiphane T	3T	48° 23.933' N	70° 16.118' W
	Epiphane P	3P	48° 23.936' N	70° 16.151' W
4	EpINETTE T	4T	48° 21.621' N	70° 08.110' W
	EpINETTE P	4P	48° 21.562' N	70° 08.217' W
5	Allaire T	5T	48° 21.308' N	70° 07.097' W
	Allaire P	5P	48° 21.263' N	70° 07.111' W
6	Des Monts T	6T	48° 20.944' N	70° 05.877' W
	Des Monts P	6P	48° 20.998' N	70° 05.894' W
7	Morin T	7T	48° 20.809' N	70° 03.654' W
	Morin P	7P	48° 20.201' N	70° 03.541' W
8	Xavier T	8T	48° 25.270' N	69° 53.801' W
	Xavier P	8P	48° 25.307' N	69° 53.920' W
9	Froid T	9T	48° 15.831' N	69° 51.764' W
	Froid P	9P	48° 15.873' N	69° 51.801' W
10	Joselito T	10T	48° 15.854' N	69° 52.858' W
	Joselito P	10P	48° 15.878' N	69° 52.831' W
11	Joe savard T	11T	48° 16.471' N	69° 55.132' W
	Joe savard P	11P	48° 16.438' N	69° 55.112' W

\* T: Tributary, P: Main stem

**Table 5-2** Physiographical variables characterising each station calculated using GIS. Significantly correlated variables (p-value < 0.05) are identified in bold

METRICS	UNITS	NOTATION
Latitude	UTM	LAT
Longitude	UTM	LONG
<b>Drainage area</b>	<b>km<sup>2</sup></b>	<b>DA</b>
<b>Canopy cover*</b>	<b>%</b>	<b>CC</b>
<b>Distance to the closest lake from station*</b>	<b>m</b>	<b>CL</b>
Area of the closest lake	m <sup>2</sup>	ACL
Azimuth at the station	°	AS
<b>Discharge</b>	<b>m<sup>3</sup>/s</b>	<b>Q</b>
Hydraulic radius	m	HR
Altitude at the station	m	A
<b>Maximum altitude of drainage basin</b>	<b>m</b>	<b>MAXA</b>
<b>Mean altitude of the drainage basin</b>	<b>m</b>	<b>MEANA</b>
Mean slope of the drainage basin	%	MS
Froude number		FN
<b>Stream order</b>		<b>SO</b>
<b>River kilometer</b>	<b>km</b>	<b>RK</b>
Forest cover on drainage basin	%	FC
Percentage of area covered by lakes and marches	%	PLM

\* P-value ranging between 0.05 and 0.1 were found for these two metrics

**Table 5-3** Cross validation results using ordinary kriging

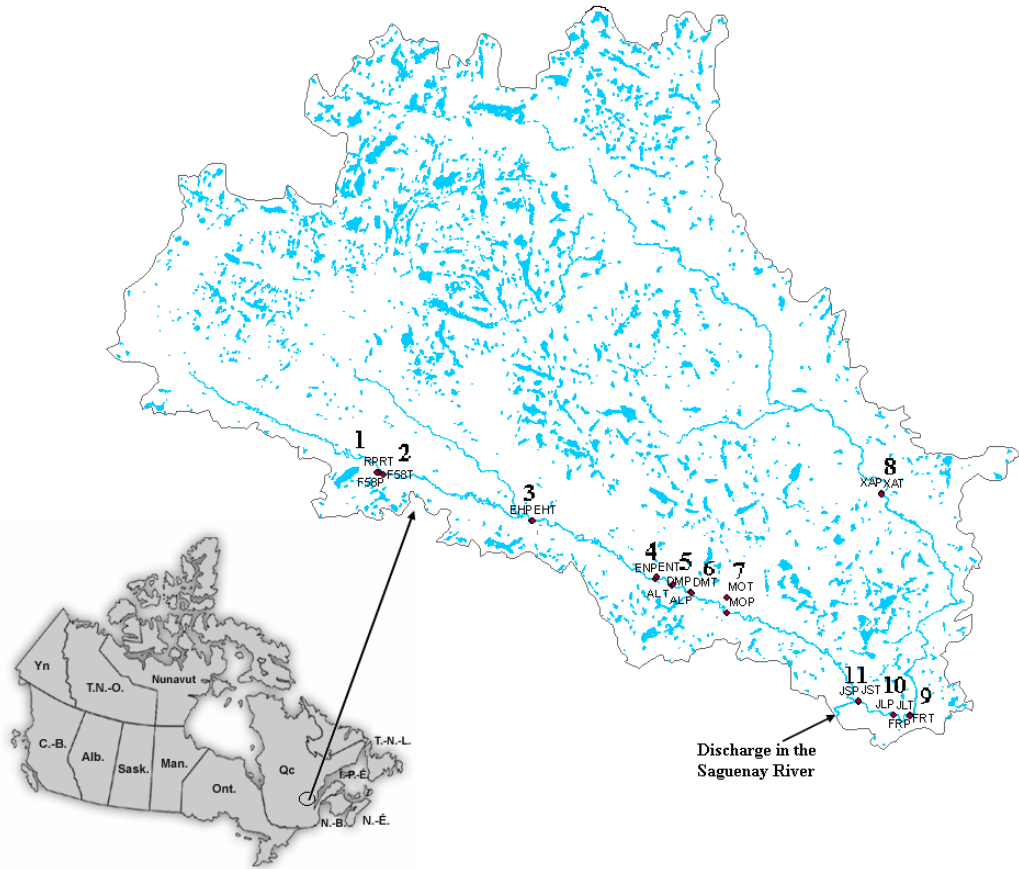
	7 metric space excluding "CL"		8 metric space	
	rBIAS (%)	rRMSE (%)	rBIAS (%)	rRMSE (%)
<b>Monthly maximum temperatures</b>				
July	-0.29	16.75	0.91	19.06
August	-0.83	14.04	0.90	16.58
September	-1.29	19.31	0.1	25.25
<b>Derived thermal characteristics (19°C)</b>				
<i>NEV</i>	-0.54	15.08	-0.57	15.40
<i>DD</i>	0.32	9.66	-0.88	10.48
<i>DUR</i>	0.62	11.86	-0.71	15.84
<b>Derived thermal characteristics (21°C)</b>				
<i>NEV</i>	-0.02	19.75	-2.11	22.97
<i>DD</i>	0.69	12.07	-0.60	12.44
<i>DUR</i>	-0.17	18.02	-0.92	14.82



**Table 5-4** Cross validation results using probability kriging

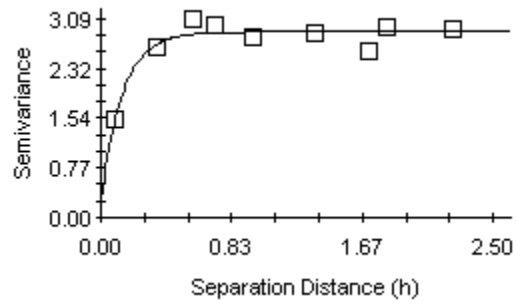
	7 metric space excluding "CL"		8 metric space	
	BIAS (% of probability)	RMSE (% of probability)	BIAS (% of probability)	RMSE (% of probability)
<b>Monthly maximum temperatures (19°C)</b>				
July	-0.04	0.31	-0.02	0.34
August	-0.04	0.31	-0.02	0.34
September	-0.02	0.22	-0.03	0.26
<b>Monthly maximum temperatures (21°C)</b>				
July	-0.02	0.44	-0.03	0.47
August	0.03	0.46	0.02	0.38
September	N/A*	N/A*	N/A*	N/A*

\* Any station exceeds a monthly maximum temperature of 21°C for September 2007

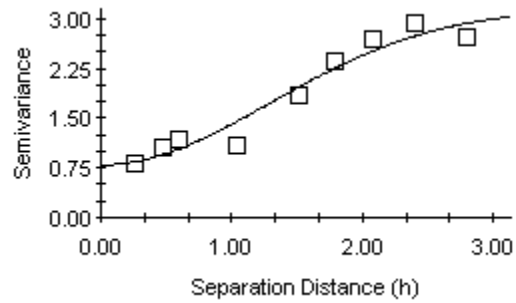


**Figure 5-1** Location and map of the Ste-Marguerite River drainage basin with temperature monitoring stations grouped by pairs at 11 sites

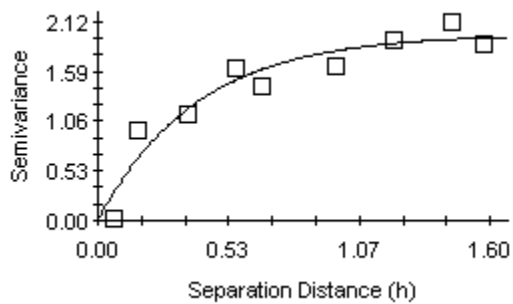
A)



B)

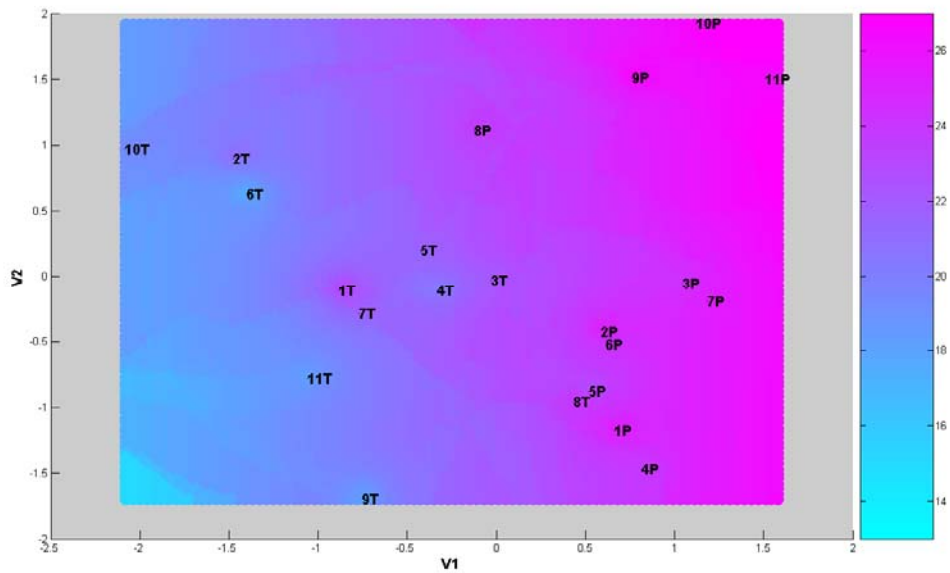


C)

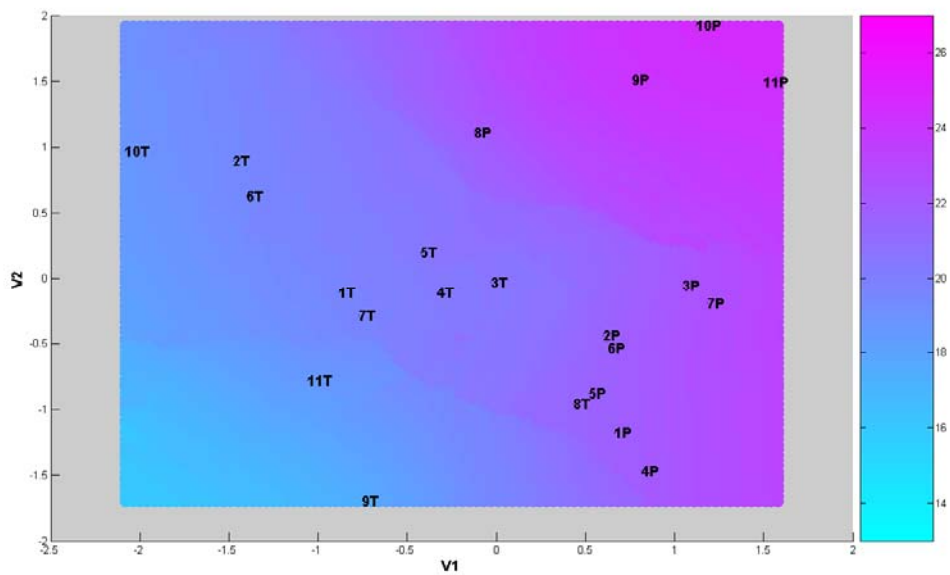


**Figure 5-2** Experimental semivariograms illustrating spatial structure related to monthly maximum temperatures using 7 significantly correlated metrics. A) July 2007. B) August 2007. C) September 2007

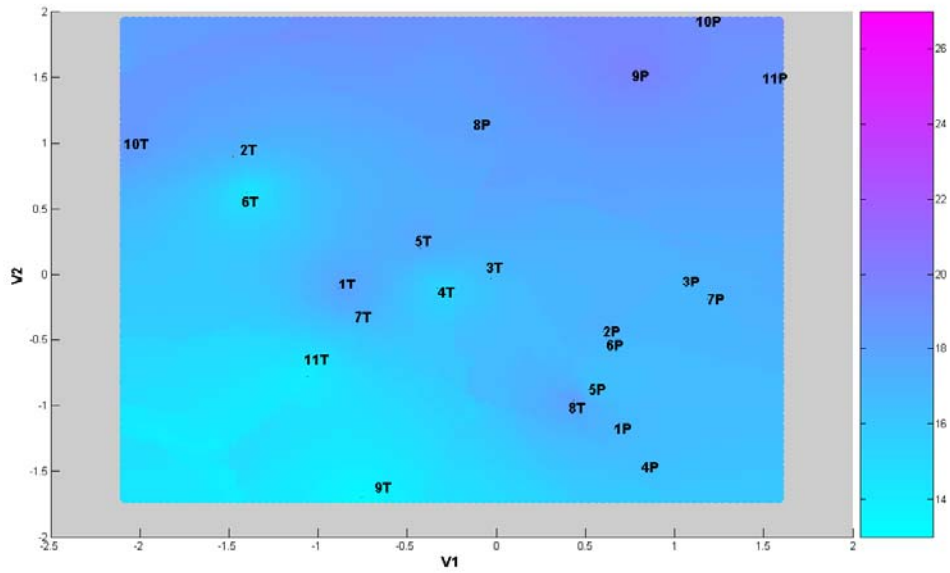
A)



B)

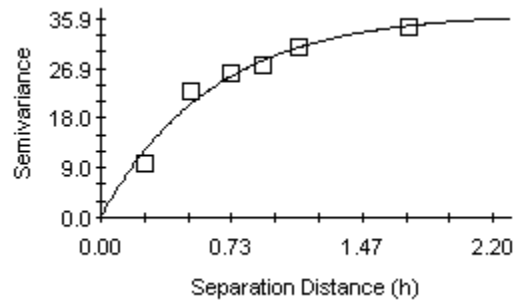


C)

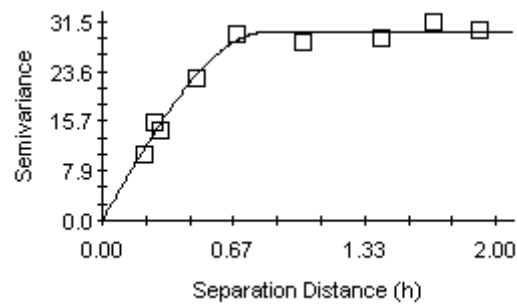


**Figure 5-3** Interpolated monthly maximum water temperatures ( $^{\circ}\text{C}$ ) performed with ordinary kriging using a 7 metric space. A) July 2007. B) August 2007. C) September 2007

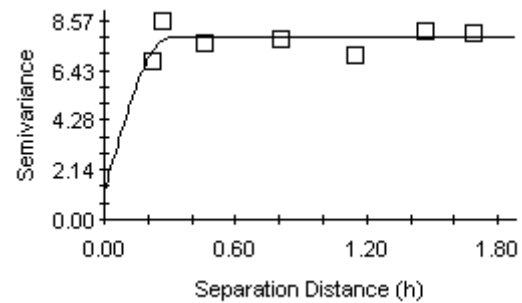
A)



B)

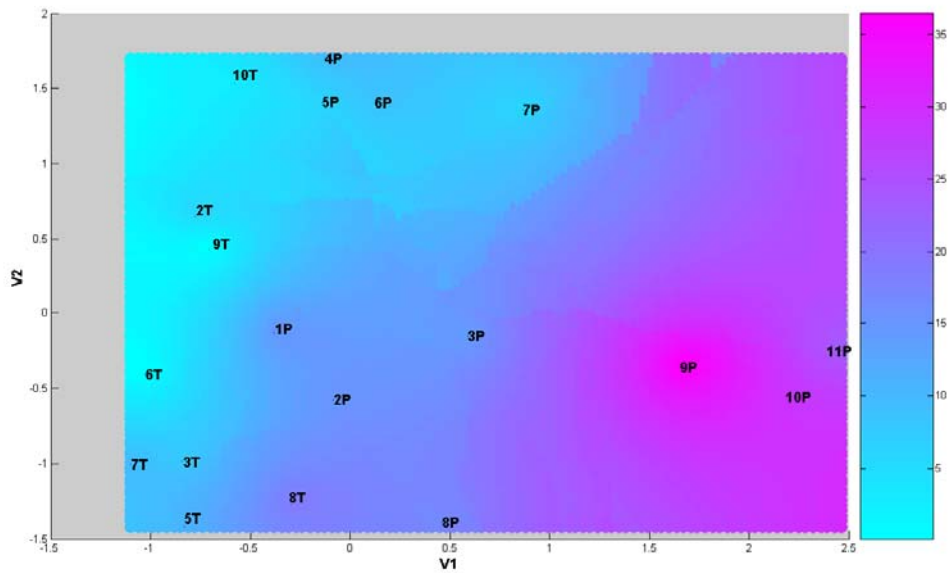


C)

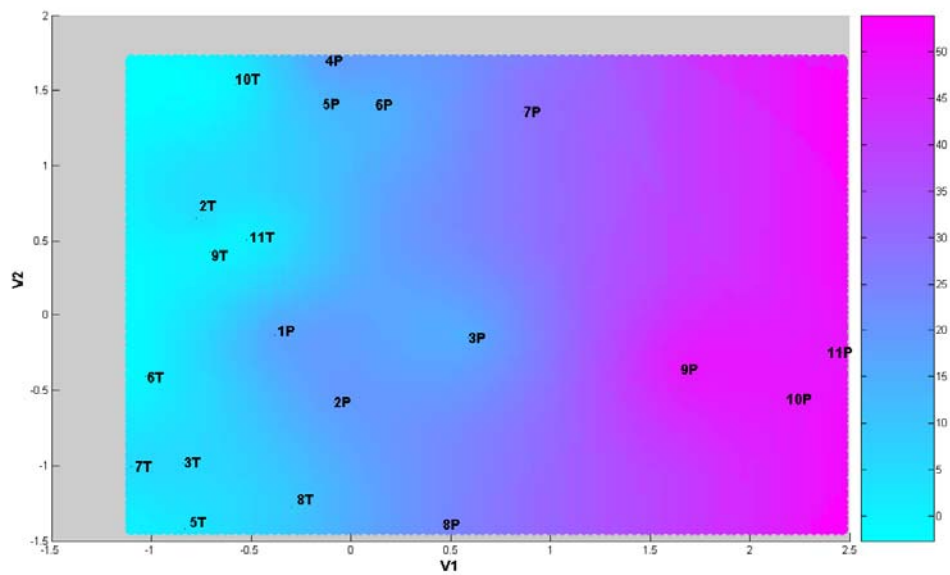


**Figure 5-4** Experimental semivariograms illustrating spatial structure related to derived thermal characteristics using 7 significantly correlated metrics. A) *NEV* exceeding a 19°C threshold. B) *DD* over a 19°C threshold. C) *DUR* exceeding a 19°C threshold

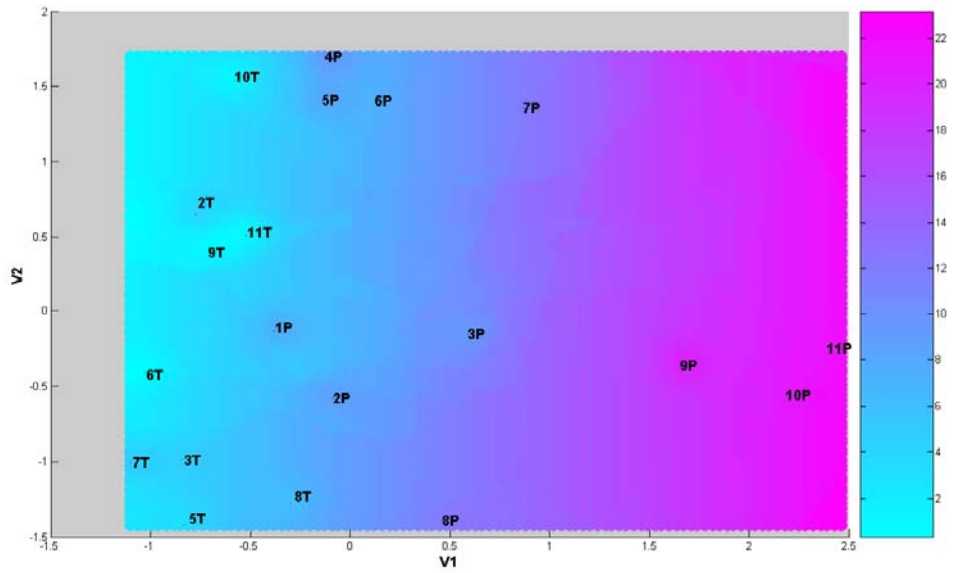
A)



B)



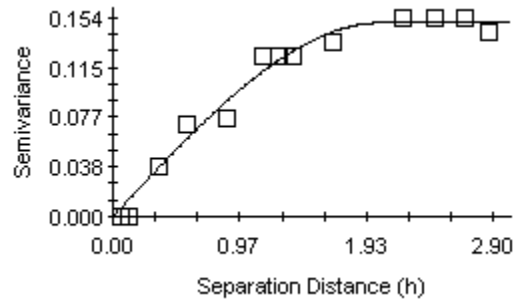
C)



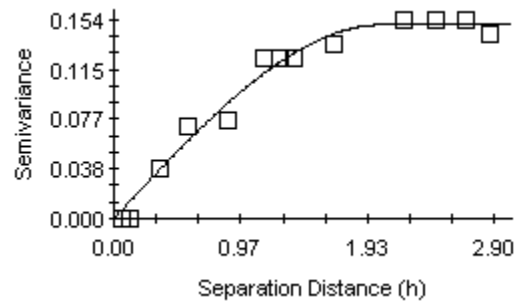
**Figure 5-5** Interpolated derived thermal characteristics performed with ordinary kriging into a 7 metric space. A) *NEV* exceeding a 19°C threshold. B) *DD* over a 19°C threshold. C) *DUR* exceeding a 19°C threshold



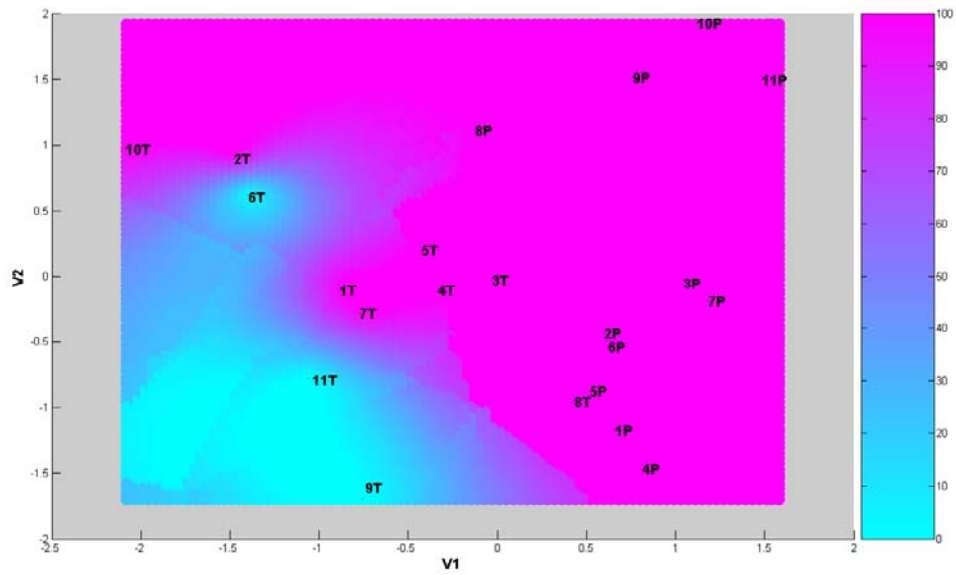
A)



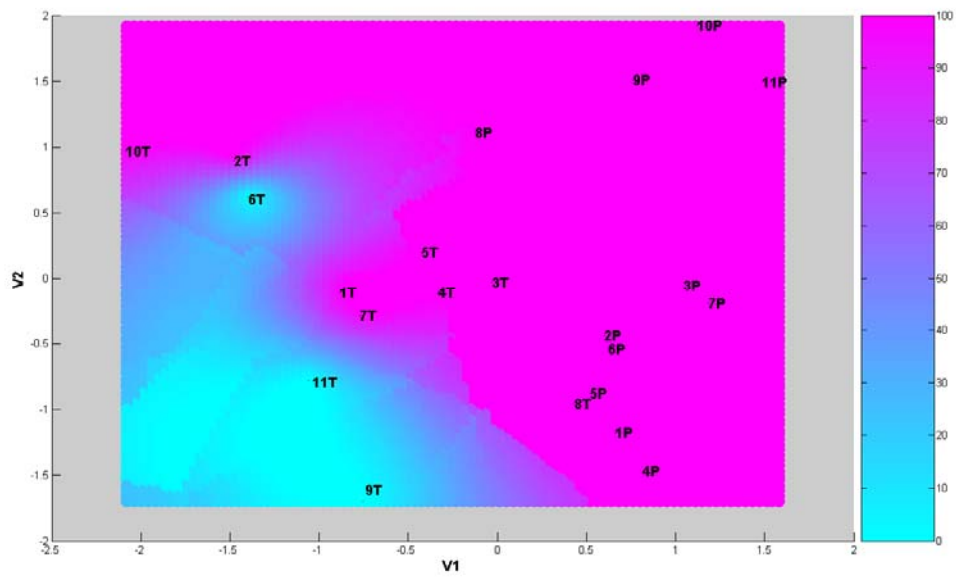
B)



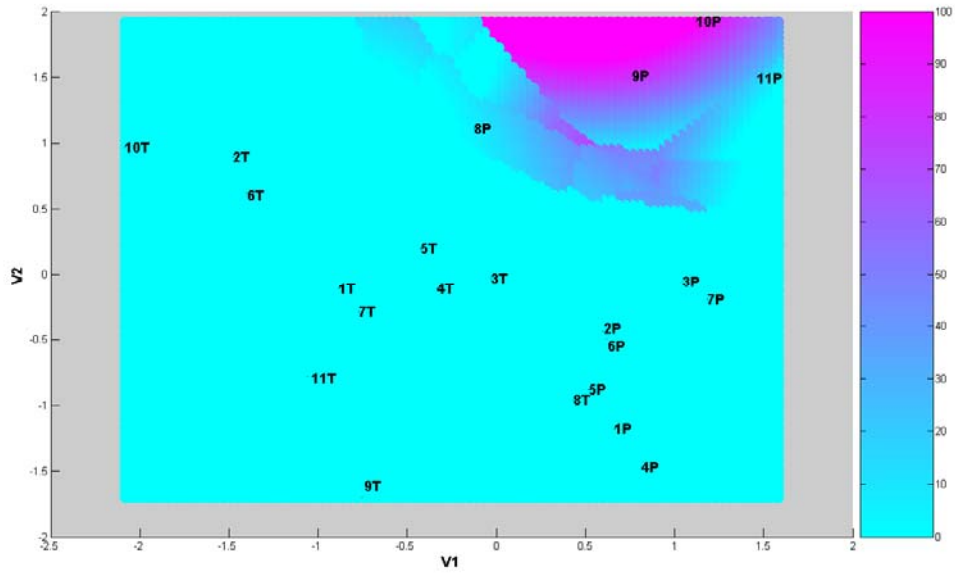
A)



B)



C)



**Figure 5-7** Interpolated monthly maximum water temperatures (°C) performed with probability kriging in a 7 metric space. A) July 2007. B) August 2007. C) September 2007



**CHAPITRE 6**  
**DISCUSSION GÉNÉRALES**



L'étude réalisée dans le cadre de ce mémoire a permis de tester et d'analyser la technique d'interpolation spatiale appelée krigeage combinée à une approche multivariée pour l'estimation de variables thermiques en rivière. Ce travail visait principalement à évaluer la pertinence d'un modèle géostatistique pour la caractérisation du régime thermique des cours d'eau dans un contexte d'évaluation de la qualité et de la disponibilité des habitats de poissons. Pour ce faire, plusieurs éléments ont été pris en compte pour évaluer la robustesse du modèle et permettent de bien saisir le cadre d'application dans lequel cet outil de modélisation peut être utilisé et optimisé. Les principaux éléments sont discutés dans la présente section.

L'utilisation des outils multivariés tel que l'ACP et l'ACC s'est révélée pertinente et appropriée dans un contexte de caractérisation d'habitats aquatiques. Ces deux approches permettent de considérer la relation entre les variables environnementales retenues soit en examinant la variance expliquée en les combinant, soit en utilisant leur structure de corrélation avec les variables thermiques. Dans les deux cas, l'avantage est d'obtenir un espace bidimensionnel construit à l'aide des deux premières composantes principales ou des deux premières variables canoniques physiographiques. À la différence d'un espace géographique (physique) unidimensionnel, on tire avantage d'une interpolation 2D puisqu'il devient possible de caractériser directement sur les cartes krigées, la variabilité spatiale de la variable régionalisée. En effet, l'utilisation du krigeage dans un espace multivarié permet d'identifier des structures « spatiales » relativement évidentes tel que des « zones » de températures élevées qui correspondent à une gamme précise de valeurs de variables physiographiques. Ainsi, il devient possible de synthétiser les principales caractéristiques structurales de la variable régionalisée, soit d'observer sa régularité dans

l'espace et mieux décrire la zone d'influence d'une observation. Cet exercice permet ultimement de juger de la qualité et de la disponibilité d'un habitat aquatique dans l'espace physiographique, ce qui peut correspondre à un travail relativement laborieux dans un espace physique où la structure spatiale de la variable observée n'est pas toujours évidente. Néanmoins, il est important de comprendre que l'information thermique associée à une coordonnée physiographique de l'espace multivarié ne correspond pas nécessairement à un point unique du bassin versant à l'étude. Par contre, une application du modèle décrite dans le premier article démontre bien la possibilité inverse qui consiste à déterminer l'information thermique pour un point non jaugé du cours d'eau en utilisant ses coordonnées physiographiques uniques de l'espace multivarié.

L'obtention d'une cartographie thermique est certes intéressante pour la caractérisation des habitats aquatiques. Cependant le cadre d'application dans lequel le modèle peut fournir une information relativement précise demeure primordial. Une première variable à considérer repose strictement sur le nombre de thermographes à déployer sur le site à l'étude, afin de capter au mieux la structure de corrélation spatiale. Le premier article a démontré la faisabilité d'utiliser un petit réseau (12 stations) sur un territoire relativement grand (25 000 km<sup>2</sup>) tout en conservant des valeurs de rRMSE (5 à 20 %) et de rBIAIS (- 2.5 à 1.1 %) acceptables. Il résulte de cette analyse que l'obtention d'une structure de corrélation spatiale cohérente via la construction du semi-variogramme expérimental demeure très sensible à l'utilisation d'un réseau relativement faible de thermographes. De toute évidence, le fait d'augmenter la densité du réseau permet d'augmenter dans la plupart des cas, les performances de krigeage. En effet, une structure de régionalisation mieux définie et plus robuste est observée lors de l'analyse variographique sur un réseau



plus dense de thermographes. Par contre, il demeure important de comprendre ici qu'une des forces de la technique de krigeage dans un espace multivarié réside dans la possibilité d'utiliser un réseau de thermographes relativement faible comparativement au krigeage dans l'espace physique. L'étude de Gardner *et al.* (2004) témoigne bien en ce sens si l'on tient compte parallèlement du nombre de thermographes (72 contre 12), des performances de krigeage similaires (RMSE entre 0.1°C et 1.3°C) et de l'espace d'interpolation utilisé (espace physique versus multivarié).

Par ailleurs, le deuxième article a permis de bien saisir l'influence des variables physiographiques retenues afin de conserver une structure spatiale cohérente de la variable régionalisée à une échelle d'étude dite locale. En effet, l'étude a tout d'abord discutée de l'influence d'une tendance prépondérante entre la variable estimée et les variables canoniques (V1 et V2) servant à la construction de l'espace multivarié. L'analyse a démontré clairement que lorsque cette tendance est significative, l'utilisation du krigeage ordinaire dans un cadre non-stationnaire demeure possible seulement si cette technique d'interpolation est appliquée à l'étude des résidus afin de respecter les hypothèses de quasi-stationnarité (moyenne constante à l'échelle du voisinage de krigeage). À noter que dans la plupart des cas, le fait de retirer la tendance observée sur les variables originales dégrade plus ou moins fortement la structure de corrélation. Néanmoins, l'étude effectuée dans le cadre du deuxième article a démontré clairement que même si une tendance doit être retirée selon l'une des deux variables canoniques (V1 ou V2), il est tout de même possible d'obtenir des valeurs de rRMSE (12 à 20 %) et de rBIAS (-1.29 à 0.69 %) acceptables avec l'utilisation du krigeage ordinaire. En effet, il ressort principalement de cette étude que les performances de krigeage mesurées ont été

obtenues grâce à l'utilisation de métriques significativement corrélées avec l'une des deux premières variables canoniques (V1, V2). Plus spécifiquement, l'utilisation des variables physiographiques locales telles le pourcentage de canopée (CC), le débit (Q) et l'altitude moyenne du bassin versant (MEANA) ont permis de conserver une structure de régionalisation cohérente, puisque ces dernières ont été fortement corrélées à la variable canonique V2. Aucune tendance significative des variables thermiques en fonction de V2 n'a été décelée. Les trois métriques locales (CC, Q, MEANA) ont permis le maintien de la structure de corrélation spatiale et l'analyse variographique a pu être réalisée sur les résidus stationnaires. Évidemment, il demeure possible que la soustraction d'une tendance selon V1 ou V2 détériore considérablement la structure de corrélation spatiale, ce qui pourrait engendrer une diminution notable des performances de krigeage. Dans cette situation, puisqu'il devient impossible de respecter les hypothèses de quasi-stationnarité du krigeage ordinaire tout en conservant une structure spatiale cohérente de la variable régionalisée, une solution réside dans l'utilisation d'une autre technique de krigeage qu'on nomme krigeage universel (Arnaud et Emery, 2000). Cette alternative non explorée dans le présent travail consiste à utiliser le krigeage dans le cadre non-stationnaire en prenant en compte la tendance explicitement dans le modèle. Bref que l'on utilise le krigeage ordinaire en respectant l'hypothèse de quasi-stationnarité ou que l'on se tourne vers l'utilisation du krigeage universel, le fait de retirer la dérive estimée pour se ramener à l'étude du résidu stationnaire entraîne inévitablement un problème de biais. Ce biais subsiste puisqu'on retire une dérive estimée et non la dérive réelle. Arnaud et Emery (2000) propose un moyen simple pour respecter une hypothèse de non-biais qui réside à identifier dans l'espace, une direction selon laquelle la dérive est constante. Cette

façon de faire permet d'obtenir un semi-variogramme directionnel sans biais. Par contre, une hypothèse additionnelle d'isotropie du résidu doit être imposée afin d'obtenir un modèle valide dans toutes les directions de l'espace. En général, il est possible de supposer que l'anisotropie du phénomène est prise en compte par la dérive. Ajoutons finalement que dans la plupart des cas, les conséquences du biais liées à l'estimation du semi-variogramme en krigeage avec modèle de tendance demeurent difficiles à évaluer.

Finalement, il demeure important de comparer les performances de krigeage obtenues dans le cadre de cette étude avec les performances de diverses autres catégories de modèles pour l'estimation de variables thermiques. Les comparaisons sont difficiles puisque aucun autre modèle spatial n'a été développé avec les mêmes variables thermiques. À titre indicateur, rappelons que pour l'ensemble des résultats produits, l'utilisation d'un modèle géostatistique combiné à une approche multivariée a produit des valeurs de RMSE allant de 0.1°C et 1.9 °C et des valeurs de BIAS comprises entre -0.08 °C et 0.09°C en ce qui concerne l'estimation de températures maximales mensuelles. En regard des études réalisées au cours des dernières années, une première étude réalisée par Chenard et Caissie (2008) indique que l'utilisation d'un réseau de neurones artificiels pour l'estimation de températures maximales journalières a produit des valeurs optimales de RMSE allant de 0.81°C à 1.64°C et des valeurs de BIAS comprises entre -0.71°C et 1.35°C. Par ailleurs, une étude réalisée par Benyahya *et al.* (2007) démontrait l'utilisation d'un modèle autorégressif (AR) et d'un modèle autorégressif périodique (PAR) pour l'estimation d'une série de températures maximales hebdomadaires moyennes (AWMT). L'utilisation des deux modèles a permis d'obtenir des valeurs de RMSE allant de 0.52 °C à 1.30 °C et des valeurs de BIAS comprises entre -0.96 °C et 0.65°C. Une autre étude

effectuée par Ahmadi-Nedushan *et al.* (2007) proposait la comparaison entre une régression multiple, un modèle autorégressif du second ordre et un modèle de type Box et Jenkins pour l'estimation de températures moyennes journalières. Dans l'ensemble, les trois modèles ont produits des valeurs de RMSE qui variaient entre 0.53 °C et 1.7 °C. L'étude de Bélanger *et al.* (2005) proposait quant à elle, l'estimation de températures moyennes journalières à l'aide d'un réseau de neurones et d'une régression multiple. Les deux approches ont produit des valeurs de RMSE variant entre 0.64°C et 1.50 °C et des valeurs de BIAS comprises entre -1.59 °C et 0.85°C. D'autre part, Caissie *et al.* (2007) ont utilisé un modèle déterministe pour l'estimation de températures moyennes journalières. Dans l'ensemble, les résultats de leur étude ont permis d'obtenir des valeurs de RMSE variant entre 1.12 °C et 2.04 °C. Finalement, rappelons que Gardner *et al.* (2003) ont obtenu des RMSE compris entre 0.1°C et 1.3 °C avec l'utilisation de 72 thermographes comparativement à un maximum de 22 dans notre cas.

Il faut noter en dernier lieu que l'estimation de variables thermiques dérivées, tels le nombre cumulatif de degrés-jours, le nombre d'événements, ainsi que la durée totale en jours au dessus de seuils spécifiques de températures, s'est avérée tout à fait pertinente pour la caractérisation d'habitats aquatiques. Des valeurs de rRMSE (9.6 à 19.8 %) et de rBIAS (-1.29 à 0.69 %) satisfaisantes ont été obtenues avec l'utilisation de sept variables physiographiques. L'estimation relativement précise de ces variables a permis d'évaluer un indice de l'hétérogénéité thermique du site à l'étude tout au long de l'été (de juillet à septembre) et peut permettre de juger de la qualité des habitats de poissons pour cette même période.

Il est clair qu'à la lumière des résultats présentés ci-haut, l'utilisation d'un modèle géostatistique combiné à une approche multivariée permet d'obtenir des erreurs d'estimation satisfaisantes et similaires aux autres études portant sur l'estimation de variables thermiques. Dans l'ensemble, il est difficile d'affirmer qu'un gain notable est observé par rapport à un ou l'autre des modèles présentés, puisque chaque modèle est utilisé dans un cadre spécifique qui caractérise à la fois le site d'étude, l'échelle d'étude et les données thermiques utilisées. Par contre, la modélisation bidimensionnelle qu'offre l'approche géostatistique et multivariée est certes l'avantage qui distingue nettement cette technique des autres types de modèles, en plus du faible niveau d'effort que constitue le déploiement d'un petit réseau de thermographes.



**CHAPITRE 7**

**CONCLUSIONS GÉNÉRALES**  
**ET PERSPECTIVES**





Les objectifs de départ fixés dans le cadre de cette étude ont nécessité le développement et l'utilisation d'un modèle géostatistique qui permet l'estimation et la caractérisation de variables thermiques en rivière. Le premier article a démontré la faisabilité d'utiliser une approche géostatistique pour l'estimation de températures extrêmes à l'intérieur d'un espace multivarié. La qualité de l'interpolation a été discutée en décrivant et comparant les performances des deux approches multivariées (ACP et ACC) en termes de biais et d'erreur d'estimation. Le deuxième article a par ailleurs permis d'évaluer l'importance du choix des métriques suivant l'échelle spatiale retenue, en plus de proposer un ensemble unique de variables physiographiques pour la construction de l'espace multivarié. L'importance de la densité du réseau de thermographes a aussi été documentée. Dans le cadre des deux articles, l'approche a été testée pour des variables thermiques extraites sur une base mensuelle. L'utilisation du modèle à une échelle temporelle journalière serait tout à fait envisageable, mais nécessiterait probablement l'insertion de variables hydro-climatiques afin de capter la variance thermique nécessairement plus bruitée à cette échelle d'étude. Finalement, l'utilisation du modèle représente un outil intéressant pour les gestionnaires des ressources hydriques et des pêches qui désirent évaluer la qualité thermique d'un habitat aquatique via l'estimation de multiples caractéristiques, en plus de leur fournir des cartes des variables régionalisées.

Pour conclure, des applications futures dans le domaine de la modélisation thermique devraient davantage prendre en compte les possibilités qu'offre une représentation bidimensionnelle de la variable estimée. Dans un premier temps, une avenue intéressante serait d'étudier davantage l'utilisation des techniques de krigeage moins courantes pour l'interpolation de caractéristiques thermiques à l'intérieur d'un espace multivarié. Par

exemple, le krigeage intrinsèque généralisé (Chauvet, 1999; Arnaud et Emery, 2000) pourrait être envisagé pour l'étude du cadre non-stationnaire et pourrait de plus prendre en compte les problèmes de biais rencontrés avec le krigeage universel. Ce type de krigeage est similaire au krigeage universel à l'exception que le semi-variogramme est remplacée par un outil structural plus général qu'on nomme covariance généralisée. Une façon plus abstraite de prendre en compte le problème de non-stationnarité pourrait aussi être envisagée avec l'utilisation du krigeage Bayésien (Omre, 1987). Par ailleurs, puisque les techniques de krigeage ordinaire et universel permettent l'étude d'une seule variable régionalisée à la fois, il serait très intéressant de se pencher sur la géostatistique multivariée qui a l'avantage d'améliorer la qualité des prévisions en exploitant les corrélations entre les variables régionalisées. Le cokrigeage constitue un bon exemple de cette catégorie et présente une avenue intéressante pour l'estimation de variables thermiques.

D'autre part, un outil très intéressant permettant de documenter la variabilité spatiale des températures de l'eau réside dans l'utilisation de l'imagerie infrarouge. Du côté déterministe, une application future utilisant cette technologie serait de jumeler cette information thermique bidimensionnelle à un modèle hydraulique possédant le même domaine physique. Une évaluation plus précise de l'indice de qualité des habitats de poissons deviendrait donc possible en tenant compte à la fois des conditions d'écoulement, des caractéristiques thermiques et du type de substrat pour un site d'étude spécifique. Du point de vue statistique, l'utilisation de l'imagerie infrarouge permet d'obtenir une très grande densité de points, ce qui facilite l'élaboration des modèles et l'identification de bons prédicteurs. Finalement, l'instrumentation d'un cours d'eau

expérimental sur plusieurs années permettrait d'acquérir des données thermiques de qualité, en grande quantité pour permettre une étude comparative plus exhaustive entre les différents modèles offerts dans la littérature.



## BIBLIOGRAPHIE

- Ahmadi-Nedushan B, St-Hilaire A, Ouarda TBMJ, Bilodeau L, Robichaud E, Thiemonge N, Bobee B. 2007. Predicting river water temperatures using stochastic models: case study of the Moisie River (Québec, Canada). *Hydrological Processes* **21**:21-34.
- Ahmadi-Nedushan B, St-Hilaire A, Berube M, Robichaud E, Thiemonge N, Bobee B. 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.
- Arnaud M, Emery X. 2000. *Estimation et interpolation spatiale*. Hermès Science Publications, Paris, 221 p.
- Arcott DB, Tockner K, Ward JV. 2001. Thermal heterogeneity along a braided floodplain river (Tagliamento River, northeastern Italy). *Canadian Journal of Fisheries and Aquatic Sciences*. **58** : 2359–2373.
- Bélangier M, El-Jabi N, Caissie D, Ashkar F, Ribí JM. 2005. Water temperature prediction using neural networks and multiple linear regression. *Revue des Sciences de l'Eau*. **18**(3) : 403-421.
- Benyahya L, St-Hilaire A, Ouarda TBMJ, Bobee B, Ahmadi-Nedushan B. 2007. Modeling of water temperatures based on stochastic approaches: case study of the Deschutes River. *Journal of Environmental Engineering and Science*. **6**(4) :437-448.
- Berman CH, Quinn TP. 1991. Behavioural thermoregulation and homing by spring chinook salmon, *Oncorhynchus tshawytscha* (Walbaum), in the Yakima River. *Journal of Fish Biology* **39**(3) : 301-312.
- Breau C, Cunjak RA, Bremset G. 2007. Age-specific aggregation of wild juvenile Atlantic salmon *Salmo salar* at cool water sources during high temperature events. *Journal of Fish Biology*. **71**(4) : 1179-1191.
- Brown LE, Hannah DM. 2008. Spatial heterogeneity of water temperature across an alpine river basin. *Hydrological Processes*. **22**(7) : 954-967.
- Burkholder BK, Grant GE, Haggerty R, Khangaonkar T, Wampler, PJ. 2008. Influence of hyporheic flow and geomorphology on temperature of a large, gravel-bed river, Clackamas River, Oregon, USA. *Hydrological Processes*. **22**(7) : 941-953.
- Burnett KM, Reeves GH, Clarke SE, Christiansen KR. 2006. Comparing riparian and catchment influences on stream habitat in a forested, montane landscape. *American Fisheries Society Symposium*. **48** : 175-197.
- Caissie D. 2006. The thermal regime of rivers: A review. *Journal of Freshwater Biology*. **51** : 1389-1406.

- Caissie D, Satish MG, El-Jabi N. 2007. Predicting water temperatures using a deterministic model: Application on Miramichi River catchments (New Brunswick, Canada). *Journal of Hydrology*. **336** : 303-315.
- Caissie D, El-Jabi N, St-Hilaire A. 1998. Stochastic modelling of water temperatures in a small stream using air to water relations. *Canadian Journal of Civil Engineering*. **25** : 250-260.
- Caissie D, Giberson DJ. 2003. Temporal variation of stream and intragravel water temperatures in an Atlantic salmon (*Salmo salar*) spawning area in Catamaran Brook (New Brunswick). *Canadian Technical Report of Fisheries and Aquatic Sciences*. 2464, 26 p.
- Chauvet P. 1999. *Aide mémoire de la géostatistique linéaire*. Cahiers de Géostatistique, Fascicule 2. École Nationale Supérieure des Mines de Paris, Centre de Géostatistique, Fontainebleau. 312 pages.
- Chen YD, McCutcheon SC, Norton DJ, Nutter WL. 1998. Stream temperature simulation of forested Riparian areas: II. Model application. *Journal of Environmental Engineering*. **124** : 316-328.
- Chenard JF, Caissie D. 2008. Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New-Brunswick, Canada. *Hydrological Processes* (sous presse).
- Chokmani K, Ouarda TBMJ. 2004. Physiographical space-based kriging for regional flood frequency estimation at ungauged sites. *Water Resources Research*. **40**(12) : art. no. W12514, pp. 1-13
- Collings MR. 1973. Generalization of stream temperature data in Washington. *US Geological Survey Water-Supply Paper*. **2029-B** : B1-B45.
- Crozier L, Zabel RW. 2006. Climate impacts at multiple scales: evidence for differential population responses in juvenile Chinook salmon. *Journal of Animal Ecology*. **75** : 1100-1109.
- Dunteman GH. 1989. *Principal Component analysis*. Sage Publications: Newbury Park, California, 96 p.
- Eaton JG, Scheller RM. 1996. Effects of climate warming on fish thermal habitat in streams of the United States. *Limnology and oceanography*. **41**(5) : 1109-1115.
- Ebersole JL, Liss WJ, Frissell CA. 2001. Relationship between stream temperature, thermal refugia and rainbow trout *Oncorhynchus mykiss* abundance in arid-land streams in the northwestern United States. *Ecology of Freshwater Fish*, **10**(1) : 1-10.
- Ebersole JL, Liss WJ, Frissell CA. 2003. Thermal heterogeneity, stream channel morphology, and salmonid abundance in northeastern Oregon streams. *Canadian Journal of Fisheries and Aquatic Sciences*. **60**(10) : 1266-1280.

- Elliot JM. 1991. Tolerance and resistance to thermal stress in juvenile Atlantic salmon, *Salmo salar*. *Freshwater Biology*. **25**(1) : 61–70.
- Elliott SR, Coe TA, Helfield JM, Naiman RJ. 1998. Spatial variation in environmental characteristics of Atlantic salmon (*Salmo salar*) rivers. *Canadian Journal of Fisheries and Aquatic Sciences*. **55** : 267-280.
- Fernandez J, Saenz J. 2003. Improved field reconstruction with the analog method: searching the CCA space. *Climate Research*. **24**(3) : 199-213.
- Ferrari MR, Miller JR, Russell GL. 2007. Modeling changes in summer temperature of the Fraser River during the next century, *Journal of Hydrology*. **342**(3-4) : 336-346.
- Gamma Design Software. 2007. GS+ Geostatistics for the Environmental Sciences, Professional Edition, vers. 7.0, *Gamma design Software*, Plainwell, Michigan.
- Gardner B, Sullivan PJ, Lembo AJ. 2003. Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. *Canadian Journal of Fisheries and Aquatic Sciences*. **60** : 344-351.
- Goovaerts P. 1997. *Geostatistics for natural resources evaluation*. Oxford University Press: New York, pp. 259-368.
- Guillemette N, St-Hilaire A, Ouarda TBMJ, Bergeron N, Robicahud É, Bilodeau L. 2008. Feasibility study of a geostatistical modelling of monthly maximum stream temperatures in a multivariate space. *Journal of Hydrology*. **364**(1-2) : 1-12.
- Gu RR, Li Y. 2002. River temperature sensitivity to hydraulic and meteorological parameters. *Journal of Environmental Management*. **66** : 43-56.
- Hari RE, Livingstone DM, Siber R, Burkhardt-Holm P, Güttinger H. 2006. Consequences of climatic change for water temperature and brown trout populations in Alpine rivers and streams. *Global Change Biology*. **12**(1) : 10–26.
- Hawkins CP, Hogue JN, Decker LM, Feminella JW. 1997. Channel morphology, water temperature, and assemblage structure of stream insects. *Journal of the North American Benthological Society* **16** : 728–749.
- Hodgson S, Quinn TP. 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.
- Holtby LB. 1988. Effects of logging on stream temperatures in Carnation Creek, British Columbia, and associated impacts on the coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences*. **45** : 502-515.

Isaaks EH, Srivastava RM. 1989. *An Introduction to Applied Geostatistics*. Oxford University Press: New-York, 561 p.

Isaak DJ, Hubert WA. 2001. A hypothesis about factors that affect maximum stream temperatures across montane landscapes. *Journal of the American Water Resources Association*. **37**(2) : 351-366.

Johnson RA, Wichern DW. 2007. *Applied multivariate statistical analysis*. Pearson Education, Inc: Upper Saddle River. New-Jersey.

Johnson SL, Jones JA. 2000. Stream temperature responses to forest harvest and debris flows in western Cascades, Oregon. *Canadian Journal of Fisheries and Aquatic Sciences*. **57** (Suppl.2) : 30-39.

Joseph G, Chokmani K, Ouarda TBMJ, St-Hilaire A. 2007. Une evaluation de la robustesse de la méthode du krigeage canonique pour l'analyse régionale des débits. *Revue des Sciences de l'Eau*. **20**(4) : 367-380.

Khangaonkar TP, Zhaoqing Y. 2007. Dynamic response of stream temperatures to boundary and inflow perturbation due to reservoir operations. *River Research and Applications*. **24**(4) : 420-433.

Lowney CL. 2000. Stream temperature variation in regulated rivers: Evidence for a spatial pattern in daily minimum and maximum magnitudes. *Water Resources Research*. **36**(10) : 2947-2955.

Lund SG, Caissie D, Cunjak RA, Vijayan MM, Tufts BL. 2002. The effects of environmental heat stress on heat-shock mRNA and protein expression in Miramichi Atlantic salmon (*Salmo salar*) parr. *Canadian Journal of Fisheries and Aquatic Sciences*. **59** : 1553-1562.

Malcolm IA, Soulsby C, Hannah DM, Bacon PJ, Youngson AF, Tetzlaff D. 2008. The influence of riparian woodland on stream temperatures: Implications for the performance of juvenile salmonids. *Hydrological Processes*. **22**(7) : 968-979.

Manly BFJ. 2004. *Multivariate Statistical Methods: A Primer, Third Edition*. Chapman and Hall, CRC Press Company: New-York, 228 p.

Mather ME, Parrish DL, Campbell CA, McMenemy JR, Smith JM. 2008. Summer temperature variation and implications for juvenile Atlantic salmon. *Hydrobiologia*. **603**(1) : 183-196.

Miyazawa T, Yamashita S, Kitagawa M, Sekine K. 1982. An evaluation and mapping of topographical factors affecting river water temperature in the upstream area of the Nagara River/ Japan. *Beitrag zur Hydrologie*. **3** : 83-95.

Mohseni O, Stefan HG, Eaton JG. 2003. Global warming and potential changes in fish habitat in U.S. streams. *Climatic Changes* **59** : 389-409.



- Mohseni O, Stefan HG, Erickson TR. 1998. A nonlinear regression model for weekly stream temperatures. *Water Resources Research*. **34**(10) : 2685-2692.
- Moore RD. 2006. Stream temperature patterns in British Columbia, Canada, based on routine spot measurements. *Canadian Water Resources Journal* **31**(1) : 41-56.
- Morin G, Nzakimuena TJ, Sochanski W. 1994. Predicting river water temperature using a conceptual model: the case of the Moisie River. *Canadian Journal of Civil Engineering*. **21** : 63-75.
- Mosley MP. 1982. New Zealand river temperature regimes. *Water and soil Miscellaneous Publication No. 36*. Water and Soil Division Ministry of Works and Development for the National Water and Soil Conservation Organisation: Christchurch.
- Omre H. 1987. Bayesian kriging - merging observations and qualified guesses in kriging. *Mathematical Geology*. **19**(1) : 25-39.
- Peterson EW, Sickbert TB. 2006. Stream water bypass through a meander neck, laterally extending the hyporheic zone. *Hydrogeology Journal*. **14** : 1443-1451.
- Solomon DJ, Sambrook HT. 2004. Effects of hot dry summers on the loss of Atlantic salmon, *Salmo salar*, from estuaries in South West England. *Fisheries Management and Ecology*. **11**(5) : 353-363.
- Scott MC, Helfman GS, McTammany ME, Benfield EF, Bolstad PV. 2002. Multiscale influences on physical and chemical stream conditions across Blue Ridge landscapes. *Journal of the American Water Resources Association*. **38** : 1379-1392.
- Sharma S, Jackson DA, Minns CK, Shuter, BJ. 2007. Will northern fish populations be in hot water because of climate change? *Global Change Biology*. **13**(10) : 2052-2064.
- Sinokrot BA, Stefan HG, McCormick JH, Eaton JG. 1995. Modeling of climate change effects on stream temperatures and fish habitats below dams and near groundwater inputs. *Climate change*. **30** : 181-200.
- Sinokrot BA, Gulliver JS. 2000. In-stream flow impact on river water temperatures. *Journal of Hydraulic Research*. **38** : 339-349.
- Stefan HG, Preud'homme E. 1993. Stream temperature estimation from air temperature. *Water Resources Bulletin*. **29** : 27-45.
- Stefan HG, Sinokrot BA. 1993. Projected global climate change impact on water temperatures in five north central U.S. streams. *Climatic Change*. **24**(4) : 353-381.
- Stevens J. 1986. *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Erlbaum, 660 p.

- St-Hilaire A, El-Jabi N, Caissie D, Morin G. 2003. Sensitivity analysis of a deterministic water temperature model to forest canopy and soil temperature in Catamaran Brook (New Brunswick, Canada). *Hydrological Processes*. **17** : 2033-2047.
- St-Hilaire A, Morin G, El-Jabi N, Caissie D. 2000. Water temperature modelling in a small forested stream: Implication of forest canopy and soil temperature. *Canadian Journal of Civil Engineering*. **27** : 1095-1108.
- St-Hilaire A, Ouarda TBMJ, Lachance M, Bobée B, Gaudet J, Gignac C. 2003. Assessment of the impact of meteorological network density on the estimation of basin precipitation and runoff: a case study. *Hydrological Processes*. **17** : 3561-3580.
- Sutton RJ, Deas ML, Tanaka SK, Soto T, Corum A. 2007. Salmonid observations at a Klamath River thermal refuge under various hydrological and meteorological conditions. *River Research and Applications*. **23**(7) : 775-785.
- Torgersen CE, Price DM, Li HW, McIntosh BA. 1999. Multiscale thermal refugia and stream habitat associations of chinook salmon in northeastern Oregon. *Ecological Applications*. **9**(1) : 301-319.
- Tung CP, Lee TY, Yang YC. 2006. Modelling climate-change impacts on stream temperature of Formosan landlocked salmon habitat. *Hydrological processes*. **20**(7) : 1629-1649.
- Yin H. 2007. Nonlinear dimensionality reduction and data visualization: A review. *International Journal of Automation and Computing*. **4**(3) : 294-303.
- Ward JV. 1985. Thermal characteristics of running waters. *Hydrobiologia*. **125**(1) : 31-46.
- Webb BW, Hannah DM, Moore RD, Brown LE, Nobilis F. 2008. Recent advances in stream and river temperature research. *Hydrological Processes*. **22**(7) : 902-918.