

**Université du Québec
Institut National de la Recherche Scientifique
Centre Eau Terre Environnement**

**MODÉLISATION DE LA TEMPÉRATURE DE L'EAU EN
RIVIÈRE À L'AIDE DU MODÈLE ADDITIF GÉNÉRALISÉ ET
COMPARAISON AVEC D'AUTRES APPROCHES
STATISTIQUES**

**Par
Fatima Laanaya**

**Mémoire présenté pour l'obtention du grade de
Maître ès sciences (M.Sc.)
en Sciences de la Terre**

Jury d'évaluation

| | |
|---------------------------------|--|
| Examineur externe | Khader Khadraoui Université Laval |
| Examineur interne | Normand Bergeron INRS-Eau-Terre-Environnement |
| Directeur de recherche | André St-Hilaire INRS-Eau-Terre-Environnement |
| Codirecteur de recherche | Erwan Gloaguen INRS-Eau-Terre-Environnement |

REMERCIEMENTS

La réalisation de ce mémoire a été possible grâce au concours de plusieurs personnes à qui je voudrais témoigner toute ma reconnaissance.

Je voudrais tout d'abord adresser toute ma gratitude à mon directeur de recherche Monsieur André St-Hilaire, professeur à l'INRS-ETE, de m'avoir accueillie dans son équipe et d'avoir accepté de diriger ce travail. Sa rigueur scientifique, sa disponibilité et ses qualités humaines m'ont profondément touchée.

Mes remerciements s'adressent également à mon codirecteur Monsieur Erwan Gloaguen, professeur à l'INRS-ETE, pour avoir accepté de co-superviser ce travail.

Je tiens à remercier également le directeur de programme Monsieur Pierre-Simon Ross, professeur à l'INRS-ETE, pour l'honneur qu'il m'a fait en acceptant de m'intégrer au programme de maîtrise en sciences de la terre dans un moment difficile de ma carrière de recherche.

Qu'ils puissent trouver dans ce travail le témoignage de ma sincère gratitude et de mon profond respect.

Mes remerciements les plus chaleureux vont à tous mes camarades à l'INRS pour leurs encouragements, soutiens et pour l'ambiance agréable tout au long de cette maîtrise.

Finalement, mes remerciements les plus sincères vont aussi à mon mari et à mes enfants, qui ont toujours su m'encourager et me soutenir le long de cette belle expérience de recherche.

RÉSUMÉ

La prévision du processus thermique en rivière implique le développement et l'application des approches appropriées aptes à produire des prédictions précises et stables. Cette prévision peut se faire à base de deux catégories de modèles, déterministes et statistiques. La modélisation statistique est un outil qui nécessite moins d'intrants. Elle permet de quantifier les effets des conditions hydrométéorologiques (ex. température de l'air, débit, ...) sur la température d'un cours d'eau et d'évaluer l'impact de cette dernière sur la qualité et la quantité des ressources aquatiques. Le but de cette étude est de tester l'applicabilité du modèle additif généralisé (MAG) à simuler la température de l'eau et de mettre en valeur son intérêt en le comparant à deux approches statistiques paramétriques déjà existantes, soit le modèle logistique et le modèle de régression des résidus¹. La revue de littérature a montré que le modèle MAG, un modèle non paramétrique, n'a pas été exploré par des études antérieures de la température de l'eau. Il a l'avantage d'être très flexible pour capter la tendance présentée par les données. C'est un outil de modélisation très populaire dans les études environnementales et épidémiques. Le modèle a été évalué sur des données d'un site de la rivière Ste-Marguerite (Canada), une rivière ayant une population de saumon Atlantique, qui préfère les eaux relativement froides. Les moyennes journalières enregistrées sur ce site de la température de l'eau, de l'air et le débit couvrent la période de 2007 à 2013. La comparaison des trois modèles MAG, logistique et régression des résidus est effectuée selon trois indices d'évaluation, l'erreur de prévision (the root mean square

¹ Ce modèle de régression de résidus a souvent été appelé stochastique, bien qu'il ne le soit pas statistiquement parlant. Cette nomination inappropriée est une confusion en modélisation thermique entre le modèle autorégressif d'ordre 2 et le modèle de régression linéaire des résidus de la température de l'eau et les effets retardés d'ordre 1 et 2 des résidus de la température de l'air.

error : RMSE), le biais (B) et le coefficient de Nash-Sutcliff (Nash-Sutcliffe coefficient of efficiency : NSC). En utilisant la technique de validation croisée, ces indices ont été calculés pour chaque année et pour toute la période d'étude. Les résultats de cette procédure ont montré que le modèle MAG performe mieux, suivi par la régression des résidus suivie par le modèle logistique.

TABLE DES MATIÈRES

| | | |
|----------|---|-----------|
| 1 | INTRODUCTION | 1 |
| 1.1 | MISE EN CONTEXTE | 3 |
| 1.2 | LES IMPACTS ANTHROPIQUES | 4 |
| 1.2.1 | <i>Changements climatiques (CC)</i> | 4 |
| 1.2.2 | <i>Les barrages</i> | 5 |
| 1.2.3 | <i>La coupe forestière</i> | 7 |
| 1.3 | VARIABLES INFLUENÇANT LA TEMPÉRATURE DE L'EAU | 8 |
| 1.4 | MODÉLISATION DE LA TEMPÉRATURE DE L'EAU | 9 |
| 1.5 | OBJECTIFS DE L'ÉTUDE..... | 12 |
| 2 | SYNTHÈSE DES RÉSULTATS | 15 |
| 2.1 | INTRODUCTION | 17 |
| 2.2 | MÉTHODOLOGIE | 18 |
| 2.1 | DONNÉES ET SITE D'ÉTUDE..... | 21 |
| 2.2 | RÉSULTATS | 22 |
| 2.3 | DISCUSSION..... | 26 |
| 3 | ARTICLE | 29 |
| | ABSTRACT | 33 |
| | RÉSUMÉ | 34 |
| 3.1 | INTRODUCTION | 35 |
| 3.2 | WATER TEMPERATURE MODELS | 37 |
| 3.3 | DATA AND STUDY AREA..... | 40 |
| 3.4 | STATISTICAL METHODS..... | 42 |

| | | |
|----------|--|-----------|
| 3.4.1 | <i>Generalized additive model</i> | 42 |
| 3.4.2 | <i>Comparison with other models</i> | 45 |
| 3.4.3 | <i>Model evaluation and validation</i> | 47 |
| 3.5 | RESULTS..... | 48 |
| 3.5.1 | <i>GAM model</i> | 49 |
| 3.5.2 | <i>Logistic model</i> | 55 |
| 3.5.3 | <i>Residual regression model</i> | 56 |
| 3.5.4 | <i>Comparative study</i> | 57 |
| 3.6 | DISCUSSION AND CONCLUSIONS | 61 |
| | REFERENCES | 65 |
| 4 | CONCLUSIONS ET RECOMMANDATIONS..... | 71 |
| 5 | RÉFÉRENCES..... | 73 |

LISTE DES TABLEAUX

| | |
|--|----|
| Table 3.1 : Number of non-missing data by year and period for the three variables water and air temperatures and flow | 41 |
| Table 3.2 : Significance of effects | 50 |
| Table 3.3 : Results of the cross validation water temperature (°C) modeling expressed by RMSE, bias error and NSC by GAM, logistic and residuals regression approaches | 59 |

LISTE DES FIGURES

| | |
|---|----|
| Figure 3.1 : Watershed Sainte-Marguerite River with location of hydrometric station, weather station and thermographs (s1, s2, s3, respectively) | 40 |
| Figure 3.2 : Average daily water and air temperatures from 2007 to 2013 | 49 |
| Figure 3.3 : Average daily flows from 2007 to 2013..... | 49 |
| Figure 3.4 : Estimated smoothing function, the x-axis shows the values of Air temperature and the y-axis represents the contribution of the fitted values | 51 |
| Figure 3.5 : Estimated smoothing function, the x-axis shows the values of flow and the y-axis represents the contribution of the fitted values..... | 51 |
| Figure 3.6 : Estimated smoothing function the x-axis shows the values of day of year and the y-axis represents the contribution of the fitted values | 52 |
| Figure 3.7 : Normal Q-Q plot of GAM residuals..... | 54 |
| Figure 3.8 : Normal distribution of GAM residuals | 54 |
| Figure 3.9 : Residuals versus estimated water temperature | 55 |
| Figure 3.10 : Relation between daily mean water temperature and mean air temperature and a fitted logistic function..... | 56 |
| Figure 3.11 : Mean daily water temperature, annual component estimated by sinusoidal function and the residual component | 57 |
| Figure 3.12 : Scatterplot of observed versus predicted values from the GAM for the season of May to October (2007 to 2013) | 60 |
| Figure 3.13 : Scatterplot of observed versus predicted values from the logistic for the season of May to October (2007 to 2013) | 60 |

Figure 3.14 : Scatterplots of observed versus predicted values from residuals regression for the
season of May to October 61

1 INTRODUCTION

1.1 Mise en contexte

Le régime thermique des cours d'eau est un facteur important pour les écosystèmes aquatiques (Acuña & Tockner, 2009; Cairns et al., 1978). En effet, les températures élevées diminuent les concentrations en oxygène (Ozaki et al., 2003; Secondat, 1952) et augmente celles des polluants (Ficke et al., 2007). En particulier, Whitehead et al., (2009) rapporte que la température de l'eau a un impact sur la qualité de l'habitat du poisson. Plusieurs études se sont intéressées aux impacts et conséquences du changement thermique des cours d'eau sur les poissons préférant les eaux froides du fait que leur niche d'habitat thermique est plus restreinte que les poissons plus tolérants aux variations thermiques (Brett, 1979; Britton et al., 2010; Eaton & Scheller, 1996). Ainsi, il a été prouvé que la variation de la température de l'eau a des impacts sur la répartition des espèces en milieu fluvial (Ebersole et al., 2003; Edwards & Cunjak, 2007; Magnuson et al., 1979; Vannote et al., 1980). De plus, les périodes les plus chaudes de l'année provoquent un stress thermique chez plusieurs espèces de poissons (Isaak et al., 2012). La réponse à ce stress chez les salmonidés ne dépend pas uniquement de la température absolue de l'eau, mais aussi de la rapidité avec laquelle la température de l'eau augmente (Quigley & Hinch, 2006). Par conséquent, lors d'une hausse importante de celle-ci, ces espèces vont chercher, lorsqu'ils sont disponibles, des refuges thermiques pour se protéger (Baird & Krueger, 2003; Dugdale et al., 2013). En revanche, les organismes aquatiques ont des limites de tolérance aux variations thermiques. Dans le cas de l'exposition à des températures dépassant ce seuil de tolérance les poissons sont soumis à un stress physiologique voire même une mortalité. Même en deçà du seuil léthal, une exposition à des températures élevées peut générer des conséquences sur leurs croissance, reproduction et résistances aux maladies (Crozier et al., 2008; Martins et al., 2012).

En outre, un taux important de mortalité des poissons, dû à la hausse des températures, peut avoir un impact négatif sur la pêche (Macdonald et al., 2010) et sur le reste de la chaîne alimentaire.

1.2 Les impacts anthropiques

De nombreuses études portent sur l'identification des impacts anthropiques modifiant le régime thermique des cours d'eau. Certains impacts ont une influence considérable comme l'urbanisation, l'agriculture, les effluents thermiques. Nelson et Palmer (2007) ont montré un accroissement de 7 °C de la température des cours d'eau urbains suite à un ruissellement de surfaces imperméables, avec une augmentation moyenne de 3.5 °C. Peu d'études ont traité l'impact de l'agriculture au Canada. Cependant, l'effet de l'aménagement agricole se traduit souvent par la coupe de la végétation ligneuse en bordure des cours d'eau ce qui les rend plus exposés à la radiation solaire incidente, ainsi qu'au ruissellement ce qui occasionne un réchauffement (Grégoire & Trencia, 2007). Concernant les effluents thermiques des stations énergétiques ou nucléaires, leur impact se résume à une augmentation de la température de l'eau causée par le rejet de l'eau ayant servi aux refroidissements (Prats et al., 2012), comme dans le cas de certains tronçons de la Loire en France (Moatar & Gailhard, 2006). Nonobstant, les principaux impacts anthropiques étudiés jusqu'à présent sont ceux associés aux changements climatiques, à la présence de barrages et à la coupe forestière.

1.2.1 Changements climatiques (CC)

Selon le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), le climat change, tant au Canada que partout ailleurs dans le monde. Ce fait est accepté par tous les spécialistes, ainsi que par la plupart des décideurs. L'impact de l'Homme sur ces changements est encore sujet à des controverses politiques, mais, encore une fois, l'ensemble des spécialistes

est unanime sur la causalité entre l'activité anthropique et les changements. Les dernières évaluations internationales confirment le changement de nombreux paramètres, par exemple l'augmentation des températures de l'air et des océans (GIEC, 2013). Plusieurs autres études mettent en évidence l'augmentation de la température des eaux des rivières (Arismendi et al., 2012; Ducharme, 2008; Ficklin et al., 2013; Kaushal et al., 2010; Morrill et al., 2005). L'impact écologique le plus notable du réchauffement climatique est la migration d'espèces intolérantes aux températures chaudes vers des altitudes et latitudes plus élevées, répondant à leurs préférences thermiques (Parmesan & Yohe, 2003; Walther et al., 2002). Une autre conséquence de ces CC est l'augmentation de la température des eaux douces à l'échelle planétaire. Ainsi, l'étude de Van Vliet et al., (2011) prévoit une hausse moyenne globale de la température de l'eau de 2.6 °C pour une augmentation de 4 °C de celle de l'air. Ce changement thermique de l'eau a une incidence directe sur les conditions de l'habitat des organismes poïkilothermes (Hari et al., 2006). En effet l'étude de Daufresne et al., (2009) démontre l'existence d'une relation directe entre l'augmentation de la température et la réduction de la taille de ces organismes. De même, l'étude de Caldwell et al., (2015), menée sur 61 sites au sud-est des É -U., a estimé une augmentation de 0.11 °C/décennie de la moyenne annuelle de la température de l'eau historique (1961-2010). Ceci semble peu, alors que les projections entre 2011 et 2060 ont suggéré une augmentation de 0.41 °C/décennie.

1.2.2 Les barrages

Des changements dans les signaux thermiques dus à la présence des réservoirs hydrauliques peuvent perturber le cycle de vie des organismes aquatiques (Lowney, 2000). En effet, certains barrages modifient le régime thermique des cours d'eau en aval de la structure de retenue (Olden & Naiman, 2010; Webb, 1996). Leurs effets dépendent du type d'ouvrage, du mode de gestion et

de la position de la prise d'eau, du volume du réservoir (Bartholow et al., 2004) et de la stratification de la colonne d'eau dans les réservoirs (Webb & Walling, 1997). Le mode de gestion notamment le moment du relâchement et le volume d'eau relâché influencent également le régime thermique (Toffolon et al., 2010). Ainsi, la position des prises d'eau modifie la température de l'eau en relâchant de l'eau chaude (épilimnique) ou froide (hypolimnique) issue, de la surface ou de la profondeur respectivement, des réservoirs stratifiés (Olden & Naiman, 2010). L'étude de Webb et Walling (1997) en Angleterre, a exploré les données de la température de l'eau d'une rivière régulée et a détecté une hausse de la moyenne de la température, une diminution des valeurs estivales maximales et une réduction des fluctuations journalières.

L'influence des retenues sur la température de l'eau des rivières est présentée par un effet de lissage des cycles journaliers et/ou annuels de la température de l'eau en aval (Liu et al., 2005). En Australie, Preece et Jones (2002) ont noté que la température maximale annuelle a diminué de 5°C et se produisait trois semaines plus tard par rapport à celle observée avant la construction du barrage. Par contre Singer et Gangloff (2011) ont remarqué un réchauffement de la température moyenne mensuelle tout au long de l'année suite à l'installation d'un petit barrage sur une rivière en Alabama. Aussi, l'étude de Maheu et al., (2015) à l'est du Canada a démontré le même constat concernant l'impact des barrages de stockage sur la température de l'eau à l'été et à l'automne.

Enfin, la simulation de l'étude de Poirel et al., (2010), qui porte sur l'influence des barrages-réservoirs sur la température de l'eau en aval des ouvrages, confirme une hausse de la température de l'eau estivale liée à la hausse des températures de l'air et d'une baisse marquée des débits.

1.2.3 La coupe forestière

De nombreuses études ont examiné l'impact de la coupe forestière sur la thermie des cours d'eau (St-Hilaire et al., 2015). Hannah et al., (2008) ont démontré que la coupe forestière riveraine influence la température de l'eau. La récolte de la bande riveraine mène à une augmentation de la radiation solaire incidente et par conséquent une hausse de 6 °C à 7 °C de la température de l'eau aura lieu (Moore et al., 2005). Généralement, le retour à une densité de végétation semblable à celle qui existait avant la coupe prend en moyenne 15 à 20 ans (Moore et al., 2005). Une comparaison, des cours d'eau exposés à la radiation solaire et ceux qui sont ombragés au Minnesota, révèle une différence de 2.5 °C de la température moyenne hebdomadaire (Blann et al., 2002). Aussi, une augmentation entre 1.4 °C et 4.4 °C de la moyenne hebdomadaire de la température maximale a été observée dans les cours d'eau du Maine où l'exploitation forestière a eu comme résultat l'élimination de la bande riveraine protectrice (Wilkerson et al., 2006). St-Hilaire et al., (2000) rapportent que dans le cas des événements météorologiques extrêmes, la coupe forestière peut faire accroître la température journalière moyenne de 1°C. Cependant, Jones et al., (2006) ont noté qu'une bande de 30 m aide à conserver une température adéquate pour les salmonidés dans 63% des cours d'eau de Géorgie (É.-U.). Par contre une bande de 15 m a permis de maintenir une température adéquate dans seulement 9% des cours d'eau, alors qu'une coupe à blanc a conduit à une augmentation de 2 °C à 8 °C des fluctuations de la température diurne en été (Johnson & Jones, 2000). La coupe forestière peut indirectement augmenter la thermie des cours d'eau en raison de l'érosion qui élargit son lit et réduit sa profondeur (Beschta, 1997). D'autres paramètres de végétation, à savoir l'indice de surface foliaire, la hauteur moyenne des arbres et la largeur de la bande riveraine, sont fortement corrélés avec la température maximale du cours d'eau (Sridhar et al., 2004).

1.3 Variables influençant la température de l'eau

Les variations spatiales de la température de l'eau de rivière sont influencées par les composantes locales du climat (Mohseni et al., 1998; Sinokrot & Stefan, 1993), la structure géomorphologique (Caissie, 2006; Mosley, 1983), la topographie (Chen et al., 1998), le débit (Caissie, 2006), les interactions entre l'eau souterraine et l'eau de surface (Ward, 1985; Webb et al., 2008), et le couvert végétal du bassin versant et riverain (St-Hilaire et al., 2000). À titre d'exemple, la régression linéaire multiple de Segura et al., (2014) a démontré que la surface de drainage exerce un fort contrôle durant toutes les saisons, alors que l'effet des eaux souterraines n'était apparent que pour le printemps et l'automne. Toutefois, les eaux souterraines influencent négativement la température des cours d'eau, sauf durant l'hiver. Ils constatent également que l'élévation et la pente moyenne du bassin versant ont des effets négatifs pour l'ensemble des saisons, ce qui signifie que les bassins plus pentus ont tendance à avoir un régime thermique plus frais en raison de la réduction des temps d'exposition à la chaleur solaire sur l'eau de ruissellement sur les versants.

En revanche, les tendances temporelles dans la température de l'eau sont liées aux changements climatiques ainsi qu'aux activités humaines. En effet, la sensibilité de la rivière en réponse à ces facteurs est difficilement prévisible puisque les changements de la température peuvent être influencés par d'autres facteurs confondants (Garner et al., 2014). Cole et al. (2014) ont en effet introduit la température de l'air, la vitesse du vent, la pression atmosphérique, la radiation solaire, le point de rosée ainsi que l'humidité, et le couvert nuageux pour modéliser la température de l'eau. Néanmoins, la température de l'air demeure la variable la plus utilisée dans la modélisation de la température de l'eau en raison de la covariation spatiale et temporelle entre les deux températures (Guillemette et al., 2009; Li et al., 2014; Ouellet, 2011; Wehrly et al.,

2009). La relation température de l'eau – température de l'air n'est pas forcément linéaire, mais dépend aussi des conditions locales. Par exemple, la température de l'eau est influencée par le débit de la rivière (Caissie, 2006; Moatar & Gailhard, 2006; Webb et al., 2003). *A contrario*, les cours d'eau ayant de faibles débits sont plus affectés par les conditions atmosphériques (Van Vliet et al., 2011).

1.4 Modélisation de la température de l'eau

La modélisation et la prévision de la température de l'eau à l'échelle spatiale ou temporelle ont été explorées selon deux types de modèles. Les modèles physiques et les modèles statistiques (Benyahya et al., 2007; Caissie, 2006; Cole et al., 2014; Marceau et al., 1986). Les modèles physiques sont basés sur les processus physiques des échanges thermiques entre la rivière et son milieu naturel (Caissie et al., 2007; Cox & Bolte, 2007; Edinger et al., 1968; St-Hilaire et al., 2000). Ils sont typiquement développés pour des tronçons de cours d'eau, bien que certains l'aient été à l'échelle du bassin versant. Ils peuvent être utilisés pour prévoir les conséquences de modifications anthropiques sur les cours d'eau ou leur bassin versant, tels que la construction d'un barrage ou la coupe forestière. Cependant, ces modèles exigent un grand nombre d'intrants et la connaissance de plusieurs paramètres pour sa calibration (ex. la température de l'air, le point de rosée, la vitesse du vent, la radiation solaire, la géométrie du cours d'eau, ...). Malgré le fait que les modèles déterministes puissent fournir des estimations précises de la température de l'eau, l'acquisition des données nécessaires n'est pas toujours évidente à l'échelle régionale (Mohseni et al., 1998; Risley et al., 2003).

De plus, les relations physiques entre les différentes parties du modèle peuvent être difficiles à modéliser en raison de l'hétérogénéité spatiale. Les modèles statistiques ou empiriques

fournissent une alternative intéressante si l'échantillonnage est exhaustif et sans biais. Ces derniers ne prennent pas en considération directement les sources et les puits des flux de chaleur, mais ils représentent le concept physique implicitement en se basant sur les corrélations sur un grand nombre de mesures entre un nombre relativement restreint de variables environnementales d'entrée et la température de l'eau (Caissie, 2006; Webb et al., 2008). Ils ont l'avantage de nécessiter moins de paramètres, peu de covariables, une durée du développement moins importante et des coûts réduits des données et d'exploitation (Eaton & Scheller, 1996; Mohseni et al., 1999). Par contre ils nécessitent de longues séries d'observations (Marceau et al., 1986). De multiples travaux de modélisation statistique de la température de l'eau ont déjà été effectués avec chacun ses forces et ses faiblesses. Parmi les méthodes statistiques, on peut citer :

1) Les modèles de régression paramétriques pour lesquels la relation est spécifiée de manière paramétrique a priori, comme les modèles de régression linéaires, stochastiques² et le modèle non linéaire. Les modèles de régression linéaire basés sur l'hypothèse forte de linéarité, utilisent une seule variable explicative (souvent température de l'air) (Caissie et al., 2001; Erickson & Stefan, 2000; Segura et al., 2014; Smith, 1981; Webb et al., 2003) ou plusieurs variables explicatives (température de l'air, son effet retardé, débit, ...) (Ahmadi-Nedushan et al., 2007; Stefan & Preud'Homme, 1993b; Webb et al., 2003; Webb & Nobilis, 2007) pour prédire la température de l'eau. Ces derniers sont plus efficaces à l'échelle hebdomadaire et mensuelle qu'à l'échelle journalière du fait que les températures quotidiennes de l'eau sont susceptibles d'avoir une autocorrélation dans le temps (Caissie, 2006; Erickson & Stefan, 2000; Morrill et al., 2005; Pilgrim et al., 1998). Les modèles dits « stochastiques » décomposent les séries chronologiques

² Bien que ce terme ait été souvent utilisé de manière inappropriée en modélisation thermique car il s'agit d'une régression linéaire multiple.

en une composante saisonnière et une résiduelle afin de produire les prévisions journalières de la température de l'eau (Ahmadi-Nedushan et al., 2007; Caissie et al., 2001; Caissie et al., 1998; Cluis, 1972; Kothandaraman, 1971). Le modèle de régression paramétrique non linéaire le plus utilisé, reliant la température de l'eau et celle de l'air, est le modèle logistique (nommé aussi sigmoïde) (Benyahya et al., 2007; Caissie, 2006; Caissie et al., 2001; Mohseni & Stefan, 1999; Mohseni et al., 1998; Segura et al., 2014). Ce modèle est adéquat sur des données hebdomadaires par contre il n'a pas donné un bon ajustement pour des données quotidiennes (Caissie et al., 2001; Mohseni & Stefan, 1999; Mohseni et al., 2003). Le modèle logistique fournit une bonne explication de l'effet d'hystérèse des températures de l'eau et de l'air. La relation non linéaire est convenable puisque la température de l'eau est moins sensible que la température de l'air pendant les saisons froides (température de l'air est inférieure à 0°C) due à l'altération du débit et au gel du cours d'eau. En revanche, pour les saisons chaudes, quand la température de l'air augmente, la température de l'eau augmente du même ordre à cause du taux élevé de l'évaporation (Li et al., 2014). Afin de remédier aux limitations de ces derniers, on utilise :

2) les modèles de régression non paramétriques pour lesquels la relation n'est pas spécifiée a priori. On peut citer l'approche des analogues ou K voisins les plus proches (K-Nearest Neighbours) comme exemple de modèles non paramétriques utilisés pour la prévision de la température de l'eau (Benyahya et al., 2008; St-Hilaire et al., 2012). Ainsi, l'approche par réseaux de neurones artificiels (ANNs), qui sont des modèles non paramétriques, non additifs et assimilés à des modèles de régression non linéaires généralisés, ils apportent une nette amélioration des prévisions fournies par des modèles paramétriques (Bélanger et al., 2005; Chenard & Caissie, 2008; Daigle et al., 2014; Nohair et al., 2008; Sharma et al., 2008). Enfin, les modèles linéaires généralisés (GLM) combinés à la méthode non paramétrique des polynômes

locaux ont été utilisés pour modéliser la température de l'eau à la rivière Sacramento en Californie (Caldwell et al., 2014). Ils ont donné de bons résultats pour prédire le maximum journalier de la température de l'eau (É.-U.) (Caldwell et al., 2014).

1.5 Objectifs de l'étude

En général, les modèles de la température de l'eau sont prédictifs et non pas explicatifs. Cependant l'objectif visé par les études de la modélisation de la température est de trouver un modèle capable d'apprendre des données disponibles qu'on lui propose et d'offrir une bonne prédiction de la température de l'eau. Notre étude partage aussi cet objectif commun avec les études antérieures. Son objectif principal consiste à évaluer la performance du modèle MAG afin de maximiser la qualité de l'estimation de la température de l'eau en fonction de la température de l'air, tout en prenant en considération l'impact du débit sur la relation température de l'eau – température de l'air. Cette évaluation est basée sur la détermination de sa précision, son biais et son adéquation à travers sa comparaison à deux modèles populaires, le modèle logistique et le modèle de régression des résidus. Elle va permettre aussi de juger si le MAG peut être utilisé pour la gestion des ressources aquatiques et par les scientifiques qui s'intéressent à développer leur propre modèle.

Les travaux réalisés dans le cadre de cette maîtrise sont présentés sous forme d'un mémoire par article. Le second chapitre présente un résumé en français des travaux complétés et le contenu de l'article intitulé « Modeling the water temperature: comparison of the generalized additive model, logistic and residuals regression models ». le chapitre 3 présente l'article en anglais le chapitre 4 présente les conclusions et les recommandations du projet.

Cette étude a été réalisée sous la direction du professeur André St-Hilaire et la codirection du professeur Erwan Gloaguen. Le rôle de l'étudiante pendant la réalisation de ce projet a été de préparer les données (agrégation journalière, présentation graphique des données,...), procéder aux traitements statistiques, analyser et synthétiser les résultats et en dernier lieu rédiger l'article en tant que première auteure. Ces travaux ont été effectués sous l'encadrement du directeur et codirecteur de recherche, qui ont assuré un excellent suivi et orientation de la méthodologie utilisée. Ils ont évalué la précision et la fiabilité des résultats et des analyses statistiques en émettant des corrections, suggestions et commentaires.

2 SYNTHÈSE DES RÉSULTATS

2.1 Introduction

Comme il a déjà été mentionné, de nombreuses études ont montré l'existence d'un lien significatif entre les variations de la température de l'eau et celles de la température de l'air (Erickson & Stefan, 2000; Smith, 1981) et/ou débit (Moatar & Gailhard, 2006; Webb et al., 2003). Dans la présente étude, nous proposons d'utiliser le modèle MAG avec, comme variables d'entrée, la température de l'air et le débit. Ce dernier a un impact sur la relation entre la température de l'eau et celle de l'air (Caissie, 2006; Moatar & Gailhard, 2006; Webb et al., 2003). De plus, la prise en compte du débit a permis d'améliorer le pouvoir prédictif des modèles de régression et les gains qui lui ont été associés n'étaient pas négligeables (Bélangier et al., 2005; Moatar & Gailhard, 2006; Van Vliet et al., 2011; Webb et al., 2003). En outre, des séries temporelles pour ces deux variables sont souvent disponibles. À notre connaissance, le modèle MAG a été utilisé seulement pour modéliser les variations spatiales dans le cadre de la modélisation de la température de l'eau, tel que rapporté par la revue de Wehrly et al., (2009) qui a comparé la précision du modèle de régression linéaire multiple (MLR), de MAG, du krigeage ordinaire (OK) et du modèle linéaire mixte (LMM) pour prédire spatialement la température moyenne au mois de juillet des cours d'eau situés à Michigan et Wisconsin (É.-U.). Cette revue a montré que le LMM a donné les meilleures prévisions, suivi par le MAG, le LMR et le OK. Les variations spatiales de la température de l'eau sont importantes, mais en général elles sont inférieures aux variations temporelles (Gras, 1969; Webb et al., 2003). La présente étude vise à vérifier l'efficacité et la limite du modèle MAG à prédire la température moyenne journalière à une station de suivi de la rivière Sainte-Marguerite. Cette vérification est faite en comparant sa performance à celle de l'approche logistique proposée par Mohseni et al. (1998) et l'approche de

la régression des résidus de la température de l'eau et de l'air décrite par Caissie et al. (1998). Il est à noter que ces deux derniers modèles n'ont pas les mêmes intrants que le modèle MAG utilisé. Le but de l'exercice est vraiment de tester le modèle MAG avec des méthodes conventionnelles utilisées par les praticiens.

2.2 Méthodologie

Notre approche de modélisation non paramétrique utilise le MAG introduit par Hastie et Tibshirani (1986). Le MAG (équation 3.2) dispose de nombreux avantages par rapport au modèle GLM (McCullagh & Nelder, 1989), dont il est une extension. Contrairement au GLM (équation 3.1) qui se base sur l'hypothèse forte de linéarité entre les paramètres, le MAG ne suppose aucune forme de dépendance et la relation n'est pas obligatoirement linéaire. Il permet de tenir compte des relations non linéaires complexes entre les variables étant donné que la forme des relations est déterminée par les données. Il est aussi intéressant, car l'effet additif des variables explicatives est conservé et aucune hypothèse d'indépendance de ces variables n'est nécessaire. Son principe repose sur la modélisation de la réponse au moyen d'une somme de fonctions non linéaires, ce qui permet de détecter plus précisément l'influence de chaque variable explicative. Ce modèle permet une très grande flexibilité au niveau des fonctions de lissage (équation 3.4) qui relie la variable dépendante – température de l'eau – aux variables explicatives – température de l'air et débit. Les relations qui unissent la température de l'air et le débit respectivement à la température de l'eau ne sont pas nécessairement linéaires, ni nécessairement aussi simples que le modèle logistique. C'est pourquoi le MAG pénalisé a été

considéré dans cette étude (Wood, 2006), où les fonctions de lissage³ sont des splines⁴ cubiques (nommées aussi splines pénalisées).

L'objectif principal de la régression des séries chronologiques est d'étudier si une partie de la variation à court terme dans la variable réponse peut être expliquée par des changements dans les variables explicatives. Ces séries présentent souvent des variations de tendance à long terme comme des variations saisonnières. Par exemple, l'étude d'Orr et al., (2014) a utilisé le MAG pour détecter les tendances présentes dans des données irrégulières de la température de l'eau sur des sites en Angleterre et au pays de Galles sur une période de 17 ans. Pendant les 17 ans, en moyenne, la température de l'eau moyenne a augmenté de 0.03°C ($\pm 0.002^{\circ}\text{C}$) par année. Un changement positif dans la température de l'eau a été observé sur 86% des sites. Ces variations de tendances et saisonnières sont susceptibles de dominer les données ce qui peut confondre les associations à court terme cherchées entre variable réponse – variables explicatives et influencer l'interprétation des liaisons à court terme entre séries (Peng & Dominici, 2008). Comme notre intérêt est dans les associations à court terme, l'objectif est de contrôler ces variations à long terme en incluant dans le modèle un terme qui dépend du temps et qui permet de reproduire la quasi-périodicité saisonnière des séries (Bhaskaran et al., 2013). Dans la modélisation de la température de l'eau, le contrôle de la saisonnalité du signal permet de prendre en compte l'effet d'hystérésis qui lui est associé.

La mise en œuvre du modèle MAG a été effectuée avec le logiciel R (R Development Core Team, 2012) muni du progiciel « mgcv⁵ » (Wood, 2012) version 1.7-22. Ce progiciel utilise une

³ Un lissage est un outil non paramétrique qui contrôle la régularité d'une fonction (de régression ou de densité). La régularité est contrôlée par le degré de différentiabilité de la fonction.

⁴ Une spline cubique est une fonction polynômiale par morceaux de degré 3 (ou d'ordre 4) et de points de rupture appelés nœuds. Elle est de classe C^{4-2} à chaque nœud.

⁵ Minimizing Generalized Cross Validation

version modifiée par Wood (2006) de la méthode de GCV (generalized cross-validation) pour déterminer automatiquement le degré de lissage optimal associé à chaque variable explicative.

Le modèle MAG ci-dessus a été comparé au modèle logistique et au modèle de régression des résidus.

Le modèle logistique (équation 3.6) a quatre paramètres qui ont une signification physique (Mohseni et al., 1998). Il permet de modéliser la température de l'eau en fonction de la température de l'air. Ce modèle a été abondamment utilisé avec les données hebdomadaires, mais rarement utilisé avec des données mensuelles. Malgré cela, il n'a pas donné de bons résultats pour l'estimation de la température maximale journalière de l'eau (Caissie et al., 2001).

L'approche de la régression des résidus consiste à séparer les séries de température de l'eau et de l'air en deux composantes: la composante annuelle ou saisonnière à long terme et la composante à court terme constituée des résidus (après soustraction de la composante à long terme) en utilisant les équations (3.7) et (3.8) respectivement. La composante saisonnière est modélisée par une fonction sinusoïdale pour les deux séries de températures (eau et air) (Cluis, 1972) par les équations (3.9) et (3.10) respectivement. La composante résiduelle de la température de l'eau est modélisée à l'aide d'une régression multiple (équation 3.11) en fonction des délais (lag1 et lag2) de la composante résiduelle de la température de l'air (Kothandaraman, 1971).

L'ensemble des trois modèles a été validé par la technique de la validation croisée (Jackknife) (Quenouille, 1949). Cette procédure consiste à enlever temporairement les données d'une année de la température de l'eau du jeu de données et d'utiliser le reste comme données de calibration, pour ensuite estimer les valeurs de température de l'année enlevée. On répète cette opération pour toutes les années pour trouver une estimation de toutes les valeurs de la température de

l'eau permettant une comparaison entre les valeurs estimées et les valeurs observées via trois critères de comparaison : la racine de l'erreur quadratique moyenne (RMSE) (Janssen & Heuberger, 1995), le biais et le coefficient NSC (Nash & Sutcliffe, 1970) dont les formules sont représentées par les équations (3.12) (3.13) et (3.14) respectivement.

Le critère NSC mesure l'adéquation du modèle aux données. L'adéquation entre le modèle et les observations est parfaite pour un NSC=1 et le modèle commence à être satisfaisant à partir d'un NSC supérieur ou égal 0.80. Cependant, pour un NSC négatif, il serait préférable de prendre la moyenne des valeurs mesurées comme estimateur au lieu de celle estimée par le modèle.

2.1 Données et site d'étude

La rivière Sainte-Marguerite est une rivière à saumon Atlantique de la région du Saguenay, au Québec (Canada) (Guay et al., 2000). Elle est un affluent de la rivière Saguenay constituée de deux branches principales de 90 km et 85 km dans un bassin versant ayant une superficie totale de 2100 km², qui se joignent 2.5 km avant de se jeter dans la rivière Saguenay (voir Figure 3.1). Le site de l'étude se situe sur la branche nord-est de la rivière Sainte-Marguerite. La figure 3.1 montre l'emplacement de la station météorologique, la station hydrométrique gérée par le Ministère du Développement durable, de l'Environnement et de la Lutte contre les changements climatiques (MDDELCC) et la sonde S3 (thermographe S3) utilisée pour la température de l'eau. Les données enregistrées sur ce site de la température de l'eau, de l'air et le débit couvrent la période de 2007 à 2013 (6 années de données), cependant, aucune donnée n'est disponible pour l'année 2008. Les données horaires relevées des deux premières variables sont agrégées en moyennes journalières. Pour chaque année on ne garde que les mesures des mois de mai à octobre (184 jours). Le choix de ces mois se justifie par le climat (climat continental) de la zone

étudiée caractérisée par un régime hivernal où les variations thermiques des cours d'eau sont négligeables pendant l'hiver à cause du couvert de glace qui limite les échanges thermiques entre la rivière et son environnement (Marsh, 1990). Ces variations deviennent importantes suite à la fonte des neiges pour les saisons de printemps, d'été et d'automne ce qui correspond à notre période d'étude. Il faut noter que les données dont on dispose ne sont pas complètes pour certaines années à cause de pannes d'enregistrement du thermographe ou l'occurrence de périodes sèches durant lesquelles il a été exondé, ce qui rend les données enregistrées invalides (Table 3.1).

2.2 Résultats

Les séries de données sont présentées graphiquement (Figure 3.2 : température de l'eau et de l'air en haut, débit en bas). Cette figure montre que les amplitudes des variations de la température de l'air sont plus prononcées que celles de la température de l'eau qui sont naturellement lissées. En effet, l'inertie thermique de l'eau empêche les variations rapides de sa température. Il faut noter que la température de l'eau a atteint des moyennes journalières relativement élevées au cours des étés des années 2010, 2012 et 2013. Des températures de l'eau élevées telles qu'enregistrées pendant ces trois années sont souvent observées dans les cours d'eau larges et peu profonds (Caissie, 2006). La Figure 3.2 suggère une relation non linéaire entre la température de l'eau et les variables explicatives température de l'air et débit.

Une modélisation MAG a été effectuée en incorporant la température de l'air et le débit comme variables expliquant le régime thermique. L'introduction de la variable jour de l'année dans cette modélisation a permis de reproduire la saisonnalité des données.

Le MAG permet d'estimer l'influence d'une seule variable explicative, après avoir enlevé celle des autres variables explicatives, sur la variable réponse. Les résultats de ce modèle peuvent être représentés graphiquement par la relation de la température de l'eau avec la température de l'air et le débit. Les résultats analytiques de ce modèle montrent que les effets de la température de l'air, du débit et la saison sont significativement non linéaires avec une probabilité critique (P-value) < 0.00001 (voir Tableau 3.2). Cela veut dire que la part non linéaire n'est pas négligeable et tenir compte que de la composante linéaire de ces effets n'est pas suffisant. La déviance expliquée (voir Tableau 3.2) par le modèle a la même signification que le R^2 dans le cas linéaire. Le calcul de la déviance expliquée pour un modèle est effectué par rapport à la déviance du modèle nul. Le pourcentage de la variance totale expliquée par ce modèle est de 93% (voir Tableau 3.2). Les effets nets non linéaires de la température de l'air, du débit et l'effet saisonnier sont représentés par les Figures 3.3, 3.4 et 3.5. Le tracé des effets nous permet d'appréhender la nature de la relation entre la température de l'air et le débit avec la température de l'eau. On peut remarquer que le nuage de la courbe lisse de la température de l'air (voir Figure 3.3) dessine légèrement une forme en S. Le débit représente un effet plutôt décroissant (voir Figure 3.4). L'effet du débit est plus marqué par la présence des crues printanières. Cet effet est plus présent pendant le printemps à cause du gros volume d'eau issu de la fonte des neiges. L'effet de la température de l'air est plus marqué pendant l'été où la rivière présente un faible débit, ce qui signifie que la température de l'air est étroitement associée à la température de l'eau au cours de cette saison. L'inclusion du débit a permis de tenir compte implicitement des spécificités de chaque saison, tandis que l'inclusion de terme qui dépend du temps minimise le risque d'erreur associée à l'effet d'hystérèse.

La contribution à la déviance expliquée du modèle par une variable explicative est égale à la déviance du modèle complet moins la déviance du modèle réduit de cette variable explicative divisée par la déviance du modèle nul (où il n'y a aucune variable explicative) (Turgeon, 2012). En termes de pourcentage, la contribution associée à chaque variable explicative est rapportée à la somme des contributions de l'ensemble des variables explicatives incluses dans le modèle. Pour le modèle MAG proposé, la contribution à la déviance expliquée du modèle de la température de l'air, de débit et de la saison est de 44,6 %, 30.2 % et 24.1 % respectivement. Pour ce cas d'étude, le modèle MAG a donné un bon ajustement de la température de l'eau en fonction de la température de l'air et le débit avec un RMSE (équation 3.12) pour toute la période de l'étude de 1.79°C. L'analyse résiduelle a été menée afin de déterminer si le modèle MAG peut être accepté. Cette analyse a été effectuée sur les résidus de validation croisée pour éviter le problème de surajustement des données. Elle porte sur le test de la normalité des résidus et le diagnostic de la persistance des structures dans ces derniers. Elle a confirmé l'acceptation du modèle MAG (voir Figure 3.6 et 3.7).

Le modèle logistique (Mohseni et al., 1998) qui relie la température de l'eau à la température de l'air en utilisant l'équation 3.6 a été appliqué sur les moyennes journalières de la température de l'eau et de l'air. L'estimation de cette fonction logistique est donnée par l'équation 3.15. Les données journalières de la température de l'eau et de l'air montrent une forte dispersion (Figure 3.7). Cette fonction ne représente pas bien les données journalières. Elle est plus adaptée aux données hebdomadaires où la dispersion est réduite. La variabilité totale expliquée par ce modèle est de 78 %. Le RMSE pour toute la période de l'étude est de 2.75°C.

La régression des résidus commence par l'optimisation de la composante annuelle de la température de l'eau et de l'air en appliquant l'équation 3.8 et 3.9 respectivement. La

composante annuelle de la température de l'eau explique 80.3 % de la variabilité totale de la température de l'eau. La figure 3.8 représente l'estimation de cette composante pour l'année 2013 en utilisant l'équation 3.16. Une fois que la composante annuelle a été retirée des séries de la température de l'eau et de l'air, les résidus de la température de l'eau ont été estimés par un modèle de régression en fonction des délais (lag1 et lag2) des résidus de la température de l'air (équation 3.17). Les coefficients de lag1 et lag2 pour les résidus de température de l'air étaient significatifs et l'ordonnée à l'origine n'était pas significative. Ce modèle est meilleur que le modèle logistique pour représenter les données quotidiennes. Les lag 1 et lag2 des résidus de la température de l'air expliquent 44% de la variabilité totale des résidus de température de l'eau. Cependant, la variabilité totale expliquée par ce modèle lorsque la composante annuelle a été ajoutée est de 89% et le RMSE pour toute la période de l'étude est de 2.09 °C

La comparaison des trois modèles MAG, logistique et régression des résidus est effectuée et les résultats des indices d'évaluation, RMSE (équation 3.12), biais (équation 3.13) et NSC (équation 3.14) pour chaque année et pour toute la période sont présentés dans le Tableau 3.3.

Dans le cadre de cette étude, le MAG donne de meilleurs résultats avec un RMSE pour toute la période de l'étude de 1.79 °C qui varie entre 1.28 °C et 2.18 °C, selon les années. Le biais est centré à zéro, avec une valeur annuelle entre -1.35 et 1.32°C. Le critère NSC indique une bonne adéquation du modèle MAG à ces données avec une valeur de 0.90. Ce résultat est confirmé par la présentation graphique des valeurs estimées versus valeurs observées (voir Figure 3.10). Cette figure montre un alignement compact autour de la première bissectrice, indiquant que l'écart entre les valeurs observées et les valeurs prédites par le modèle MAG est minime et sans biais.

Pour la validation du modèle logistique et du modèle de régression des résidus, les RMSE pour toute la période d'étude étaient de 2.75 °C et 2.09 °C respectivement. Ces deux modèles sont

moins performants que le modèle MAG, mais la performance du modèle de régression est entre celle du MAG et du modèle logistique. Les valeurs annuelles du RMSE varient entre 1.83 °C et 3.40 °C pour le modèle logistique puis entre 1.04 °C et 2.64 °C pour le modèle de régression des résidus. Le biais pour toute la période de l'étude est très faible et centré en zéro pour ces deux derniers modèles, mais l'étendue est très large (i.e. entre -1.08 et 1.99°C pour le modèle logistique et entre -1.15 °C et 1.85 °C pour le modèle de régression des résidus). Le NSC est de 0.76 pour le modèle logistique et 0.86 pour le modèle de régression des résidus. La figure 3.11 présente les valeurs estimées par le modèle logistique versus les valeurs observées. Elle montre que ce modèle surestime les faibles valeurs de la température de l'eau. Cependant, la figure 3.12 démontre que le modèle de régression des résidus sous-estime les valeurs les plus élevées et surestime les faibles valeurs de la température de l'eau. Il est important de noter que ces deux derniers modèles ne tiennent pas compte de l'effet du débit. Cette omission est peut-être la cause des surestimations de la température de l'eau qui ont été observées généralement au printemps à des niveaux d'eau plus élevés et des sous-estimations qui ont été observées au cours de l'été, quand les niveaux d'eau sont plus faibles.

2.3 Discussion

L'objectif de cette évaluation est de déterminer si le MAG est un outil prometteur qui peut aider à la gestion des ressources aquatiques en fournissant des simulations et des prédictions précises de température de l'eau. La performance du modèle MAG a été comparée à celle d'approches existantes, soient le modèle logistique et le modèle de régression des résidus. Le modèle logistique a donné des résultats moins précis que les deux autres modèles pour les données quotidiennes (RMSE = 2,75 °C), ce qui concorde avec les résultats de Caissie et al., (2001) pour

la température maximale journalière de l'eau. Selon la revue de Benyahya et al. (2007), ce modèle est plus adapté aux données hebdomadaires pour lesquelles il a été largement utilisé. Le modèle de régression des résidus présente une performance relativement acceptable (RMSE = 2,09 °C). De meilleures performances ont été obtenues avec ce type de modèle dans des études antérieures telles que celle de Caissie et al., (1998). Ce dernier modèle ne prend pas en compte la structure de l'autocorrélation de la série de température de l'eau (mais prend en compte celle de l'air), qui est généralement importante (Caissie et al., 1998). L'effet du débit sur la température de l'eau est également omis dans ce modèle, comme dans le cas du modèle logistique. Il faut noter que la période de l'étude contient trois années avec des températures de l'eau relativement chaudes, ce qui est atypique pour la rivière Ste-Marguerite (voir Guillemette et al., (2009) pour maxima mensuels). Ces limitations ont une incidence sur la performance du modèle de régression des résidus pour cette rivière. En outre, la forte corrélation entre le lag 1 et lag2 des résidus de la température de l'air pose un problème de multicollinéarité, ainsi les résultats peuvent donc être moins précis. Les résultats du modèle MAG ont montré qu'il était très flexible pour ajuster les relations non linéaires entre la température de l'eau et les variables explicatives (température de l'air et le débit). Ces résultats n'étaient pas influencés par les valeurs atypiques présentes dans les données de cette étude. L'inclusion du débit a permis une amélioration du modèle, probablement due au fait que la crue printanière et les étiages estivaux modulent les extrêmes thermiques. L'analyse résiduelle du modèle MAG donne confiance dans son potentiel d'utilisation. En somme, le modèle MAG a établi des relations non linéaires très flexibles entre la température de l'eau et les variables explicatives, température de l'air et le débit. Il a donné un meilleur ajustement pour la température de l'eau de la rivière Sainte-Marguerite et surtout, il était plus robuste dans le cas de données atypiques. Pour valider ces résultats préliminaires de

cette étude, des applications sur d'autres sites sans valeurs atypiques seraient utiles. Sa validation sur d'autres sites permettra de juger s'il peut être utilisé comme un outil potentiel pour une gestion efficace des systèmes aquatiques.

3 ARTICLE

**Modeling the water temperature: comparison of the generalized additive
model, logistic and residuals regression models**

By

Fatima Laanaya^{1a}

André St-Hilaire¹

Erwan Gloaguen¹

¹ INRS-ETE, Université de recherche
490 da la couronne, Québec, G1K 9A9, Canada

Manuscript submitted to: Hydrological Sciences Journal

31 July 2015

^a Corresponding author: email: Fatima.laanaya@ete.inrs.ca

Abstract

Water temperature plays a fundamental role in the control of dissolved oxygen concentrations and it has a significant influence on the life cycle of aquatic organisms, including stenotherm fish such as salmonids. It is thus of prime importance to build reliable tools to forecast water temperature in order to preserve river fishes to water warming. This study proposes to evaluate a statistical scheme to model the average water temperature based on daily average air temperature and average flow rate at the Sainte-Marguerite River, North Canada. The aim of this work is to test a non-parametric water temperature regression model and to compare its performance to two previously developed approaches: the logistic and residuals regression models. The non-parametric generalized additive model (GAM) model allows to test whether the relationship between the response variable (water temperature) and the explanatory variables (air temperature, flow rate) is significantly different from a linear relation, and graphically illustrates the form of this relationship. Due to its flexibility, the GAM was able to capture some of the non-linear response between water temperature and the two explanatory variables. The shape of these effects is determined by the trends shown in the collected data. The three models were evaluated annually using a cross-validation technique. Three comparison criteria were calculated: the root mean square error (RMSE), the bias error and the Nash-Sutcliffe coefficient of efficiency (NSC). The goodness of fit of the three models was also compared graphically. The GAM is the best of among the three models with a RMSE = 1.79 °C, bias = 0.006 and NSC = 0.90 compared to the logistic model RMSE = 2.75 °C, bias = 0.01, and NSC = 0.76. Whereas the residuals regression model gave RMSE = 2.09 °C, bias = -0.03 and NSC = 0.86.

Key words: Mean Daily water temperature, logistic model, GAM, residuals regression.

Résumé

La température de l'eau exerce un contrôle fondamental de la quantité d'oxygène dissout. Elle a une influence significative sur le cycle de vie des systèmes aquatiques, tels que les salmonidés. Ce travail propose une modélisation statistique de la moyenne journalière de la température de l'eau en fonction des moyennes journalières de la température de l'air et le débit à la rivière St-Marguerite située à la région du Saguenay-Lac-Saint-Jean-Côte-Nord au Québec (Canada). Elle porte sur l'élaboration d'un modèle de régression non paramétrique de température de l'eau et sur la comparaison de sa performance avec le modèle logistique et le modèle de régression des résidus. Le modèle non paramétrique GAM permet de tester si la relation entre la variable réponse (température de l'eau) et les variables explicatives (température de l'air, débit) est significativement différente d'une relation linéaire et de visualiser graphiquement la forme de cette relation. Grâce à sa flexibilité, le modèle additif généralisé (GAM) a montré que les effets des deux variables explicatives sont non linéaires. La forme de ces effets est déterminée par les tendances présentées dans les données récoltées. Les trois modèles ont été évalués en utilisant la technique de validation croisée par année. Trois critères de comparaison ont été calculés, l'erreur quadratique moyenne (RMSE), le biais et le NSC. La qualité de l'ajustement des trois modèles a également été comparée graphiquement. Le modèle GAM est le meilleur des trois modèles avec un $RMSE = 1.79$, $\text{biais} = 0.006$ et $NSC = 0.90$ comparativement au modèle logistique avec $RMSE = 2.75$, $\text{biais} = 0.01$ et $NSC = 0.76$ et le modèle de régression des résidus avec $RMSE=2.09$, $\text{biais} = -0.03$ et $NSC = 0.86$.

Mots-clés : Moyenne journalière de la température de l'eau, modèle logistique, GAM, régression des résidus.

3.1 Introduction

Water temperature is one of the physical factors controlling water quality. For instance, it affects oxygen concentration that is inversely proportional to water temperature (Ozaki et al., 2003; Secondat, 1952), as well as the concentration of some pollutants (Ficke et al., 2007). Water temperature has direct or indirect effect on many other physical (e.g. density, viscosity, ...), chemical (e.g. conductivity, oxygen solubility, ...) or biological (Stevens et al., 1975) variables in lotic ecosystems.

The thermal regime of rivers is important for aquatic ecosystems (Acuña & Tockner, 2009; Cairns et al., 1978; Caissie, 2006). Aquatic species require suitable habitats in order to complete their full life cycle. Several studies have examined the impacts and consequences of changes in the thermal regime of streams on cold-water fish and they found out that these fish have less opportunity for a possible migration from their thermal habitat (Brett, 1979; Britton et al., 2010; Eaton & Scheller, 1996). Water temperature is also a determining ecological factor for many biological parameters such as metabolic activity (Demars et al., 2011), reproduction, survival (Connor et al., 2003; Eaton et al., 1995) and growth rate (Elliott & Hurley, 1997; Neuheimer & Taggart, 2007). Also, it has been proven that water temperature has an impact on the distribution of species in rivers (Ebersole et al., 2003; Edwards & Cunjak, 2007; Magnuson et al., 1979; Vannote et al., 1980). Higher than average water temperature may generate thermal stress for certain species of fish (Isaak et al., 2012). Indeed, the response to this stress for salmonids depends on the rate at which the water temperature increases (Quigley & Hinch, 2006). In some cases, these species will seek, when available, thermal refuges in order to avoid being exposed to high temperature conditions (Baird & Krueger, 2003; Dugdale et al., 2013).

Many studies focus on the identification of human impacts altering thermal regimes of rivers. Some activities may have significant influence on water temperature, such as urbanization (e.g. Nelson & Palmer, 2007) , agriculture (e.g. Grégoire & Trencia, 2007) and thermal effluents (e.g. Moatar & Gailhard, 2006; Prats et al., 2012). Other important anthropogenic impacts include climate change, dams and logging.

Worldwide, international assessments confirm the change in air and ocean temperatures (IPCC, 2013). Climate change (CC) represents a significant threat to the life of aquatic species (Crozier & Zabel, 2006; Eaton & Scheller, 1996; Ferrari et al., 2007; Mohseni et al., 2003; Webb et al., 2004). Several studies have highlighted the fact that water temperature of rivers has significantly increased (Arismendi et al., 2012; Ducharne, 2008; Ficklin et al., 2013; Kaushal et al., 2010; Morrill et al., 2005). Such a shift in the thermal regime of rivers would have a direct impact on the habitat conditions of poikilothermic organisms (Hari et al., 2006). The most notable ecological impact of climate change remains the migration of different species to higher altitudes and latitudes, as a response to their thermal preferences (Parmesan & Yohe, 2003; Walther et al., 2002). Indeed, the study of Daufresne et al., (2009) demonstrated the presence of a direct relationship between temperature increase and the size reduction of ectothermic aquatic organisms (bacteria, phyto- and zooplankton, and fish). Changes in thermal signals due to the presence of reservoirs disrupt the life cycle of aquatic organisms (Lowney, 2000). Indeed, dams alter the river thermal regime downstream of the impoundment (Olden & Naiman, 2010; Webb, 1996). Their effects depend on the type of the structure, the management mode and the position of the water intake. The influence of impoundment on river water temperature is presented by a smoothing of daily cycles downstream of the dam (Liu et al., 2005).

3.2 Water Temperature Models

The modeling and forecasting of the water temperature under different spatial or temporal scales has been conventionally tackled using two models: deterministic and statistic. (Benyahya et al., 2007; Caissie, 2006; Cole et al., 2014; Marceau et al., 1986). Deterministic or physical models are based on the mathematical representation of the physical processes of heat exchange between the river and its natural environment (Caissie et al., 2007; Cox & Bolte, 2007; Edinger et al., 1968; St-Hilaire et al., 2000). They are typically developed at the river reach scale. Although deterministic models can provide accurate estimates of water temperature, acquiring the necessary data is not always possible, especially at the regional level (Mohseni et al., 1998; Risley et al., 2003).

Statistical or empirical models remain an alternative that requires less parameters than deterministic models (DeWeber & Wagner, 2014; Mohseni et al., 2003; Moore, 2006; Morrill et al., 2005; Smith, 1981). In addition to requiring less input data statistical approaches often require a shorter period of development and reduced costs of data and operation (Eaton & Scheller, 1996; Mohseni et al., 1999). However, they require long series of observations (Marceau et al., 1986). Statistical models do not take into account direct sources of heat flow. Instead, they represent implicitly the physical concept by taking into consideration the correlations between a relatively small number of environmental variables input and the water temperature (Caissie, 2006; Webb et al., 2008). Multiple statistical works of water temperature modeling have already been presented with their strengths and weaknesses (Benyahya et al., 2007). Among these statistical methods are the following:

1) The parametric regression models, for which the relationship is parametrically specified a priori. This category includes the single or multiple linear regression, stochastic and nonlinear models (Ahmadi-Nedushan et al., 2007; Erickson & Stefan, 2000; Smith, 1981; Stefan & Preud'homme, 1993a; Webb et al., 2003) to predict water temperature. They are more effective at weekly and monthly scales than at daily scales, given that the daily water temperatures have greater autocorrelation (Caissie, 2006; Erickson & Stefan, 2000; Morrill et al., 2005; Pilgrim et al., 1998). The term “stochastic model” has been used to describe water temperature models that decompose time series into a seasonal and a residual component so as to produce daily forecasts of water temperature (Ahmadi-Nedushan et al., 2007; Caissie et al., 2001; Caissie et al., 1998; Cluis, 1972; Kothandaraman, 1971). The most often used non-linear parametric regression model that links the water temperature to air temperature, remains the logistic model (also called sigmoid model) (Benyahya et al., 2007; Caissie, 2006; Caissie et al., 2001; Mohseni & Stefan, 1999; Mohseni et al., 1998; Segura et al., 2014). This model is suitable for weekly data whereas its application on daily data did not produce a good fit (Caissie et al., 2001; Mohseni & Stefan, 1999; Mohseni et al., 2003).

To overcome the limitations of parametric models, other models are used:

2) Non-parametric regression models for which the relationship is not specified a priori. The K-nearest neighbor approach is an example of non-parametric models that were previously used for predicting water temperature (Benyahya et al., 2008; St-Hilaire et al., 2012). Artificial Neural Networks (ANNs) often show improvement in forecasts when compared to parametric models (Bélanger et al., 2005; Chenard & Caissie, 2008; Daigle et al., 2014; Nohair et al., 2008; Sharma et al., 2008). And finally, the generalized linear model (GLM) combined with the non-parametric

method of local polynomial was used to model water temperature in the Sacramento River in California which gave a good results to predict daily maximum water temperature (USA) (Caldwell et al., 2014).

In this study, we propose to apply a generalized additive model (GAM), with air temperature and flow as the input variables. The flow represents an important factor influencing the actual relationship between water and air temperature (Caissie, 2006; Moatar & Gailhard, 2006; Webb et al., 2003).). Also, including the flow has improved the predictive power of some water temperature regression models (Bélanger et al., 2005; Moatar & Gailhard, 2006; Van Vliet et al., 2011; Webb et al., 2003).

The GAM offers the option to take into account the complex nonlinear relationships between parameters because the shape of the relationship is determined by the data. In fact, the relationship between air temperature, flow and the water temperature is known to be non-linear at the daily time step. This model allows great flexibility for functions that link the dependent variable-water temperature-to independent variables-air temperature and flow. For these reasons, the GAM model offers some potential for water temperature modelling. The GAM model was only used previously to model the spatial variations of water temperature, as reported by the review of Wehrly et al., (2009). This review compared the model precision and accuracy of multiple linear regression (MLR), generalized additive model (GAM), ordinary kriging and of linear mixed modeling (LMM) so as to spatially predict the average watercourses temperature in July in the states of Michigan and Wisconsin (USA). Their results showed that the LMM gave the best predictions followed by the GAM, LMR and krig. The spatial variations in water temperature are important, but in general they are inferior to the temporal variations (Gras, 1969; Webb et al., 2003).

This study allows to evaluate the performance of the GAM model by determining its precision, bias and fitting through its comparison with two popular models: the logistic and the residuals regression models. This assessment will determine whether the GAM could be a better option for the managers of aquatic resources and for scientists who are interested in developing their own model.

3.3 Data And Study Area

The study site is located on the Sainte-Marguerite River, an Atlantic salmon river (Guay et al., 2000) located in the Saguenay-Lac-Saint-Jean region of Quebec (Canada). The Sainte-Marguerite River is a tributary of the Saguenay River that consists of two main branches of 90 km and 85 km in a watershed with a total area of 2100 km², which join 2.5 km before emptying to the Saguenay River (see Figure 3.1).

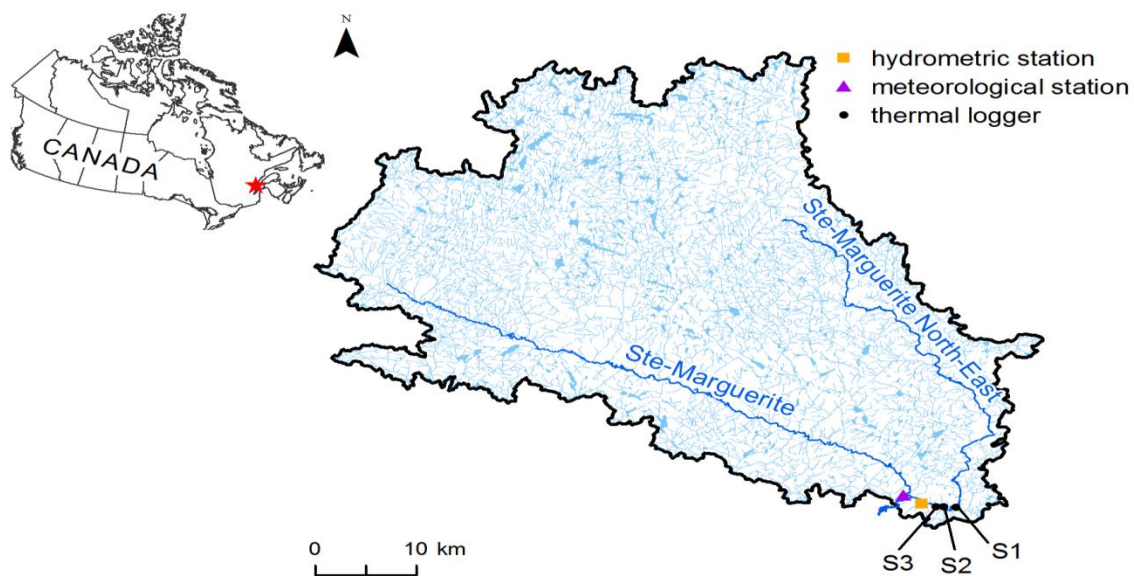


Figure 3.1 : Watershed Sainte-Marguerite River with location of hydrometric station, weather station and thermographs (s1, s2, s3, respectively)

Water temperature, air temperature and flow data were recorded on the study sites from 2007 to 2013. The observed hourly data of the first and second variable are aggregated to daily averages. Figure 3.1 shows the location of the weather station (latitude 48° 16' 37.2" N, longitude 69° 56' 2.399" O), the hydrometric station (latitude 48° 16' 5.016" N, longitude 69° 54' 33.012" O) managed by Ministry of Sustainable Development, Environment and the Fight against Climate change (https://www.cehq.gouv.qc.ca/depot/historique_donnees/fichier/062803_Q.txt) and the thermistor (107b temperature sensor, Campbell Scientific, ±0.5 °C) S3 (latitude 48° 0' 0" N, longitude 69° 0' 0" O) used to measure water temperatures. For each year we have only the measures from May to October, which is the period corresponding to the ice-free season, were used. For some years, water temperature time series are incomplete (see Table 3.1) because of equipment malfunction or because of low water levels, which exposed the thermograph to the air.

Table 3.1 : Number of non-missing data by year and period for the three variables water and air temperatures and flow

| Year | Early Period (JJ/MM) | End of period (JJ/MM) | Number of observation |
|-------------|-----------------------------|------------------------------|------------------------------|
| 2007 | 05/07 to | 03/10 | 90 |
| 2008 | 0 | 0 | 0 |
| 2009 | 03/07 to | 10/09 | 70 |
| 2010 | 23/05 to | 17/10 | 148 |
| 2011 | 01/05 to 2011/07 to | 18/07 31/10 | 177 |
| 2012 | 01/05 to 26/06 to | 18/05 31/10 | 146 |
| 2013 | 01/05 to 31/10 to | 27/10 31/10 | 183 |

3.4 Statistical Methods

3.4.1 Generalized additive model

This non-parametric modeling uses the generalized additive model (GAM) was introduced by (Hastie & Tibshirani, 1986). The GAM, which is an extension of the generalized linear model (GLM) (McCullagh & Nelder, 1989) has many advantages over this latter model. Unlike the GLM that is based on the strong assumption of linearity of the parameters, the GAM assumes no form of dependency and the relationship is not necessarily linear. Its principle is based on modelling of the response using a sum of nonlinear functions, which allows to model more precisely the influence of each explanatory variable. This specificity makes it a popular tool in the modelling the effects of environmental variables because these effects are often non-linear and are difficult to specify parametrically (Bruneau & Grégoire, 2011; Peng & Dominici, 2008). The review of the literature of Jbilou and El Adlouni (2012) presents the potential of the GAM in environmental health studies as a powerful technique to identify non-linear relationships between explanatory environmental variable and the health dependent variable.

If restricted to the GLM model, given a response variable y and p explanatory variables x_1, x_2, \dots, x_p , the GLM model is written as:

$$\text{Equation 3.1 : } g(E(y)) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

Where g is the link function which is a parametric function allowing for the extension of the Gaussian distribution to the exponential family. It connects the average of the dependent variable to a set of explanatory variables; $E(y)$ is the expected value of the response variable; ε is the error assumed to be normally distributed with variance σ_ε .

In the case of the standard linear regression, the link function g is the identity.

The GAM model still uses a link function g as in the case of the GLM, but the linear dependency is replaced by more general dependency functions. This model is written as follows:

$$\text{Equation 3.2 : } g(E(y)) = \beta_0 + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \varepsilon$$

The implementation of this model requires the estimation of the smooth nonlinear functions $f_i(x_i)$ $i = 1, \dots, p$, for each explanatory variable x_i .

To avoid the problem of overfitting, the penalized GAM was used (Wood, 2006), where smoothing⁶ functions f_i are cubic smoothing splines⁷ (named also penalized spline). These splines are defined as the solution to the following optimization problem: among the twice continuously differentiable functions, keeping only those minimizing the sum of the penalized squares which is called as the Penalized Residual Sum of Square (equation 3.3):

$$\text{Equation 3.3 : } \operatorname{argmin}_{f_i \in C^2} (\|y - f_i(x_i)\|^2 + \sum_i \lambda_i \int f_i''(x)^2 dx)$$

The first term of (equation 3.3) measures the fit to data and the second term penalizes curvature in function. Smoothing parameters $\lambda_1, \dots, \lambda_p$ represent the penalties on the roughness of the fitted function, associated with each explanatory variable x_i . They control the smoothing level of each function f_i and the compromise between the bias and the variance. The underlying idea is to position much more nodes than deemed necessary and by the use of a penalty eliminating those nodes providing insufficient information.

6 A smoothing is a nonparametric tool that controls the regularity of a function (regression or density). Regularity is controlled by the degree of differentiability of the function.

7 A cubic spline is a polynomial function of degree 3 (or about 4) piecewise and breaking points called nodes. It is C4-2 class at each node.

The optimization problem (equation 3.3) has a formulation parametric equivalent, based on the decomposition of the f_i functions based on the cubic splines (De Boor, 1978). Giving that b_j represents the cubic B-spline. Each f_i function will have the following form:

Equation 3.4 : $f_i = B^i \beta^i$ where $B\beta = \sum_j b_j \beta_j$

The optimization problem (equation 3.3) becomes:

Equation 3.5 : $\arg \min_{(\beta^1, \dots, \beta^p) \in \mathbb{R}^{k \times p}} \left(\left\| y - \beta_0 + \sum_{i=1}^p B^i \beta^i \right\|^2 + \sum_{i=1}^p \lambda_i \int \{(B^i \beta^i)''\}^2 dx \right)$

Where f_i is the i^{th} regression function in the additive model, f'' denote the second derivative of f , $\| \cdot \|$ is the Euclidean norm and $\lambda_1, \dots, \lambda_p$ are terms of penalty.

The resolution of the problem (equation 3.5) and the fit of the model are done by the iterative reweighted least squares algorithm (IRLS). The name of this algorithm comes from the fact that in each iteration, the problem solved is equivalent to a weighted least-squares problem.

The second derivative of the smoothing function f_i gives an idea about the form of the smoothing. A high value of f_i'' means that f_i is rough, however a straight line has a second derivative equals to zero. The optimal values of λ_i the penalties should avoid the two extremes and allow to have a good fit with a balanced minimization of bias and variance. The choice of these parameters is done by using generalized cross-validation (GCV) and performance metrics.

The implementation of the GAM model was made with R software (Team, 2014) provided with the package « mgcv⁸ » (Wood, 2013) version 1.7-22. This package uses a modified version by Wood (2006) of the GCV method to automatically determine the optimal smoothing associated

⁸ Minimizing Generalized Cross Validation

with each explanatory variable. The threshold of the smoothing functions of explanatory variables to be retained is 5%.

The main objective of the regression of the time series is to investigate whether some of the variation in the short term in the response variable can be explained by changes in the explanatory variables. Water and air temperature are known to have a strong seasonal signal. The warming rate in the spring and cooling rate in the fall are often different and can produce a hysteresis effect in the air-water temperature relationship. An additional term was included in the model to account for seasonality (Julian day) (Bhaskaran et al., 2013).

3.4.2 Comparison with other models

The above GAM model was compared to two other popular models that are used to predict water temperature. These are the logistic and residuals regression models.

3.4.2.1 Logistic model

This non-linear regression model based on the development of a logistic function (Equation 3.6) which has three parameters with physical significance (Mohseni et al., 1998). It allows modeling water temperature as a function air temperature, and provides a better fit than a linear regression for the extreme values. This model has been extensively used with weekly data and some applications were made at a daily time step (e.g. St-Hilaire et al., 2012).

$$\text{Equation 3.6 : } T_w = \frac{\alpha}{1 + e^{\gamma(\beta - T_a)}}$$

Where T_w and T_a represent water and air temperatures. α is a coefficient which estimates the highest water temperature, β is the air temperature at the inflexion point and γ represents the steepest slope of logistic function. Those are parameters to be estimated.

3.4.2.2 Residuals regression model

This approach consists to separate water and air temperatures time series into two components: the long-term annual component or seasonal and the short-term component or residuals:

$$\text{Equation 3.7 : } T_w = T_{wA(t)} + R_{w(t)}$$

$$\text{Equation 3.8 : } T_a = T_{aA(t)} + R_{a(t)}$$

Where $T_{wA(t)}$ and $T_{aA(t)}$ are the long term annuals components of water and air temperature respectively; $R_{w(t)}$ and $R_{a(t)}$ are the short-term components of water and air temperature respectively and where t represents Julian day or the day of year (e.g. 1 June=152).

The seasonal component $T_{wA(t)}$ and $T_{aA(t)}$ in water temperature and air temperature respectively can be explained by using a Fourier series analysis (Kothandaraman, 1971) or an optimized sine function (Cluis, 1972). Caissie et al. (1998) showed a slight insignificant difference in the annual component between Fourier series and sine function.

For this study, the seasonal component is modeled using a sine function for both time series of water and air temperatures:

$$\text{Equation 3.9 : } T_{wA(t)} = a_w + b_w \sin \left[\frac{2\pi}{365} (t + t_{0w}) \right]$$

$$\text{Equation 3.10 : } T_{aA(t)} = a_a + b_a \sin \left[\frac{2\pi}{365} (t + t_{0a}) \right]$$

Where $T_{wA(t)}$ and $T_{aA(t)}$ is the seasonal component of a water temperature and air temperature time series; a_w , b_w , t_{0w} , a_a , b_a , t_{0a} are fitted coefficients that can be estimated using a nonlinear regression approach minimizing the sum of square errors.

The residual water temperatures (departures from the long-term water annual component) were modelled using a multiple regression with air temperature residuals (departures from the long-

term air annual component) of lag 1 and 2 as input variables (Kothandaraman, 1971). The residuals regression model is given by:

$$\text{Equation 3.11 : } R_{w(t)} = \beta_1 R_{a(t-1)} + \beta_2 R_{a(t-2)} + \varepsilon_{(t)}$$

Where $R_{w(t)}$ is the residual water temperature at time t , $R_{a(t-1)}$, $R_{a(t-2)}$ are the residuals of air temperature at time $t-1$ and $t-2$; β_1 and β_2 represent the regression coefficients which are estimated using least square method and $\varepsilon(t)$ is the error assumed to be normally distributed

3.4.3 Model evaluation and validation

All three models have been validated by the technique of cross-validation (leave-one-out cross validation year-method or Jackknife technique; (Quenouille, 1949)). This procedure consists of temporarily removing the data of water temperature of one year from the database and using the rest of the base as calibration data. Then, the temperature values of the removed year are estimated. This operation is repeated for all years to find an estimate for all the water temperature values for a comparison between the estimated values and the observed values using three comparison criteria. These criteria are : the root mean square error (RMSE) (Janssen & Heuberger, 1995), the bias error (B) and the Nash-Sutcliffe coefficient of efficiency (NSC) (Nash & Sutcliffe, 1970) whose formulas are represented by equations (3.12) (3.13) and (3.14) respectively

$$\text{Equation 3.12 : } RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}$$

$$\text{Equation 3.13 : } B = \frac{1}{n} \sum_{i=1}^n (P_i - O_i)$$

$$\text{Equation 3.14 : } NSC = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (\bar{O} - O_i)^2}$$

Where n is the number of data points, P_i and O_i are the predicted and observed value of daily mean water temperature respectively, and \bar{O} is the mean of observed daily water temperature values.

3.5 Results

The daily averages were calculated from hourly water and air temperature data for the ice-free periods between 2007 and 2013, with no data available for 2008. Within each year, we dispose of observations of the period from May to October (184 days). The series are presented graphically (Figure 3.2 and Figure 3.3). This first figure shows that the amplitudes of variation of air temperature are more pronounced than those of water temperature. However, it should be noted that water temperature reached relatively high mean values (higher than mean air temperature values) during the summers of 2010, 2012 and 2013. This is not typical of most Eastern Canadian rivers, but is often observed in wide, shallow water courses (Caissie, 2006). Figure 3.2 and Figure 3.3, suggests a nonlinear relationship between water temperature and the explanatory variables air temperature and flow rate.

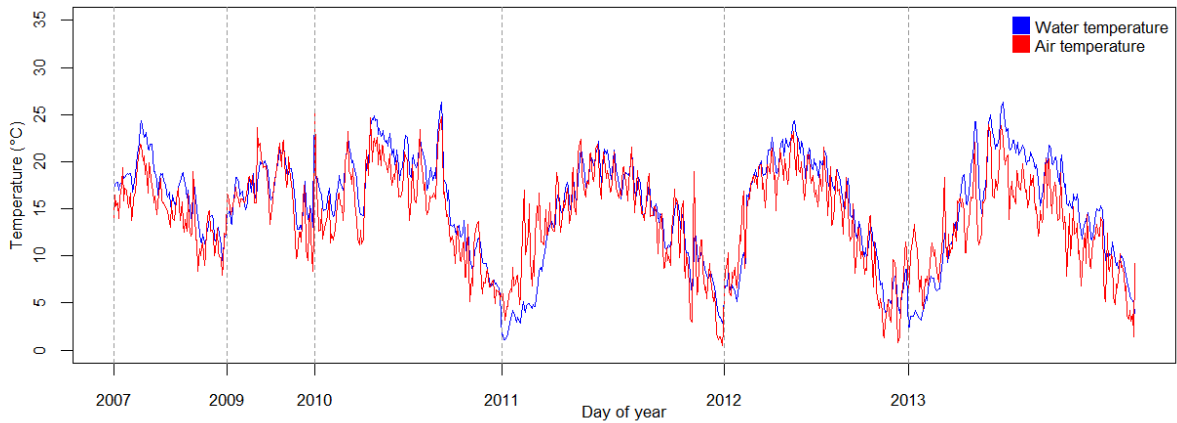


Figure 3.2 : Average daily water and air temperatures from 2007 to 2013

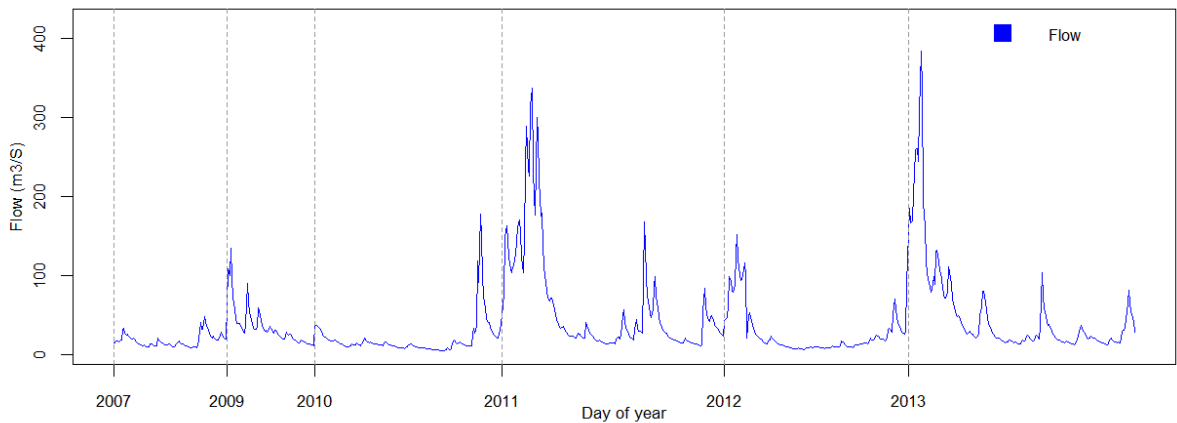


Figure 3.3 : Average daily flows from 2007 to 2013

3.5.1 GAM model

A GAM model was applied by incorporating air temperature and flow rate as variables affecting the thermal regime. The introduction of the day of the year variable in this model allowed to reproduce the seasonality of data.

After removing the other explanatory variables, the GAM allows to estimate the influence of a single explanatory variable on the response variable. It produces a visual representation of the relationship between water temperature and air temperature or flow rate, respectively. The analytical results of this model show that the effects of air temperature, the flow rate and the seasonality are significantly nonlinear with a critical probability (P-value) < 0.00001 (see Table 3.2). This means that the non-linear portion is not negligible considering just the linear component of these effects is not enough. The explained deviance (see Table 3.2) by the model has a similar meaning as R^2 in the linear case. The calculation of deviance explained for a model is made with respect to the deviation of the null (non-explanatory) model. The percentage of the total variance explained by this model is 93% (see Table 3.2). The nonlinear net effects of the air temperature, the flow rate and the seasonality are represented in Figures 3.4, Figure 3.5 and Figure 3.6. The shading corresponds to a confidence interval of 95%. The black dots, representing partial residuals, are the residuals that remain after removing the effect of all other predictors. The effects plot allows us to understand the nature of the relationship between air temperature and the flow rate with water temperature. One can notice that the cloud of the smooth curve of air temperature (see Figure 3.4) lightly draw an S shape. The flow rate represents a rather decreasing effect (see Figure 3.5).

Table 3.2 : Significance of effects

| Smoothing functions | Estimate degrees of freedom | F-test (Fisher test) | P-value |
|---------------------------|-----------------------------|----------------------|----------------------|
| S (Air temperature) | 3.4 | 168.3 | 2.2 e ⁻¹⁶ |
| S (Flow) | 3.5 | 75.9 | |
| S (Day of the year) | 5.8 | 72.2 | |
| R²_adj | | 0.93 | |
| Deviance explained | | 92.8% | |

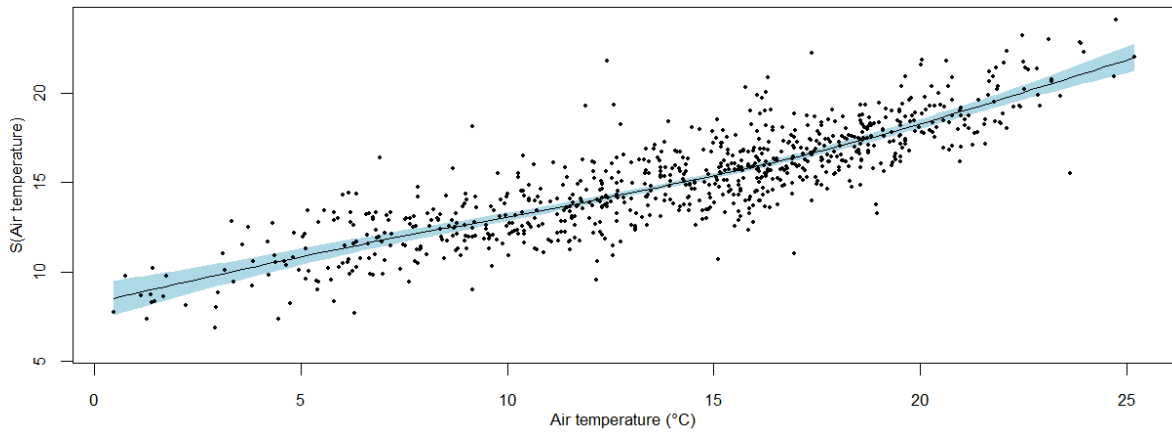


Figure 3.4 : Estimated smoothing function, the x-axis shows the values of Air temperature and the y-axis represents the contribution of the fitted values

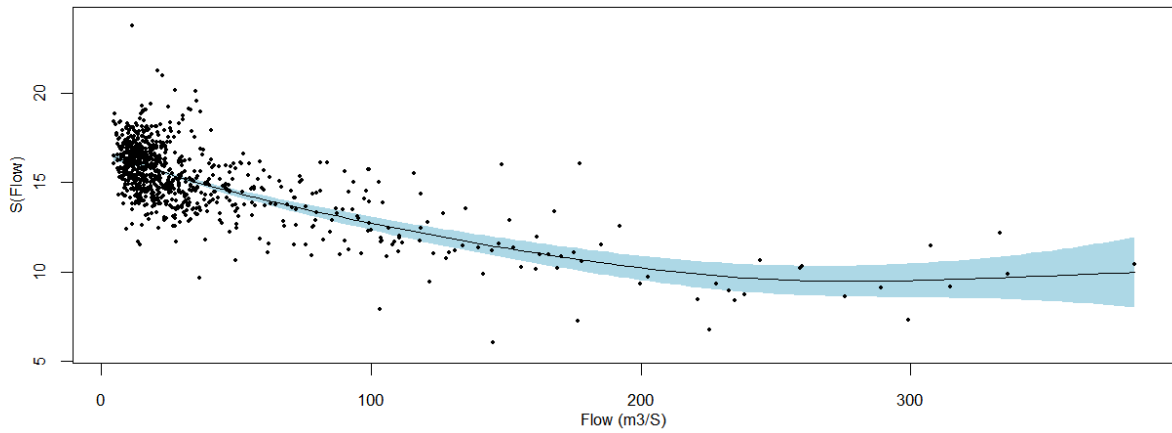


Figure 3.5 : Estimated smoothing function, the x-axis shows the values of flow and the y-axis represents the contribution of the fitted values

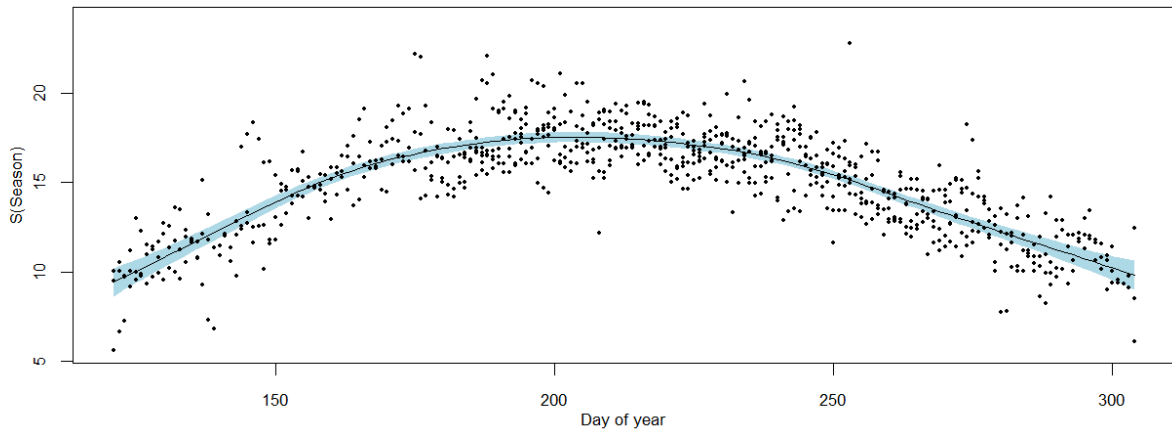


Figure 3.6 : Estimated smoothing function the x-axis shows the values of day of year and the y-axis represents the contribution of the fitted values

The effect of flow rate is marked by the presence of the spring floods. This effect is present in the spring because of the large volume of water resulting from snow melt. The air temperature effect is more pronounced during the summer when the river has a low flow rate, which means that air temperature is closely associated with water temperature during this season. The inclusion of a flow and air temperature implicitly takes into account the specificities of each season while the inclusion of the Julian Date, minimizes the risk of increased error associated with the hysteresis effect.

The contribution of an explanatory variable to the deviance explained of the model, is equal to the deviance of the full model minus the deviance of the reduced model of this explanatory variable divided by the deviance of the null model (Turgeon, 2012). In terms of percentage the contribution associated with each explanatory variable is relative to the sum of the contributions of all the explanatory variables included in the model. For the proposed GAM model, the contribution to the deviance of the air temperature, flow rate and season is 44.6%, 30.2% and 24.1% respectively. For this case study, the GAM model gave a good fit of water temperature

according to air temperature and the flow rate with overall prediction error of 1.79 °C. Thereafter, it was compared to two other popular models. Finally, residual analysis was conducted to determine if the GAM model can be accepted. This analysis was performed on the cross-validation residuals to avoid the data over-fitting problem. Figure 3.7 and Figure 3.8 confirm the normality of the distribution of these residuals. Furthermore, the graph of these residuals in terms of estimated water temperature shows a cloud of points that are uniformly dispersed and an equal spread everywhere (see Figure 3.9). This indicates that no structure remains in the residuals

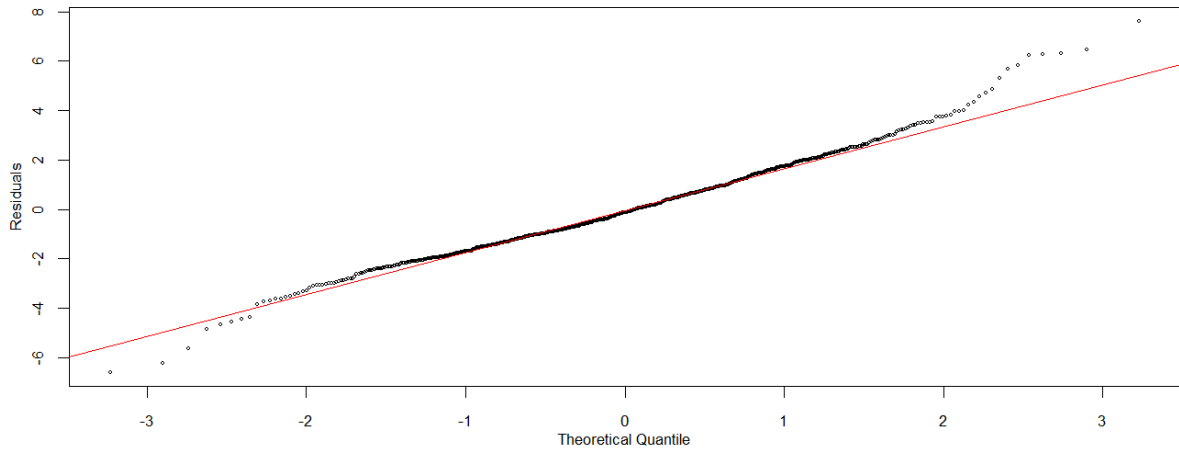


Figure 3.7 : Normal Q-Q plot of GAM residuals

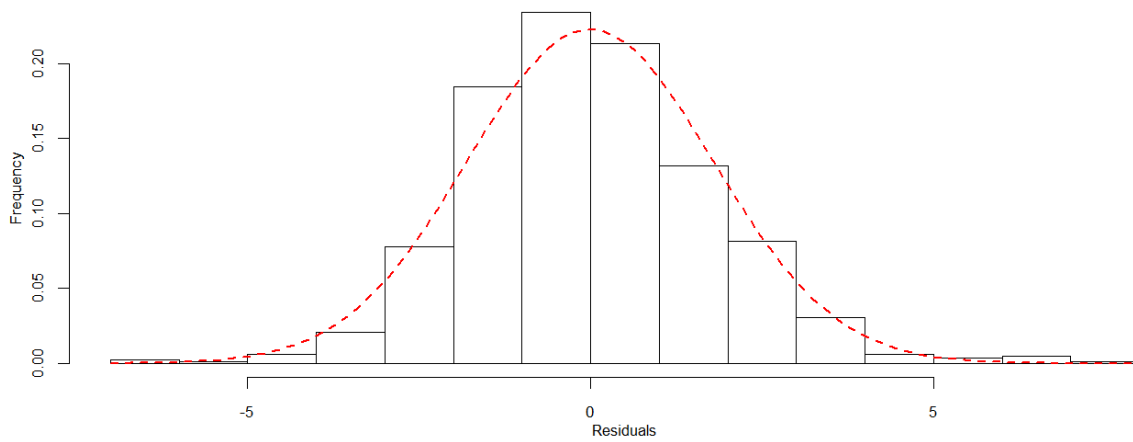


Figure 3.8 : Normal distribution of GAM residuals

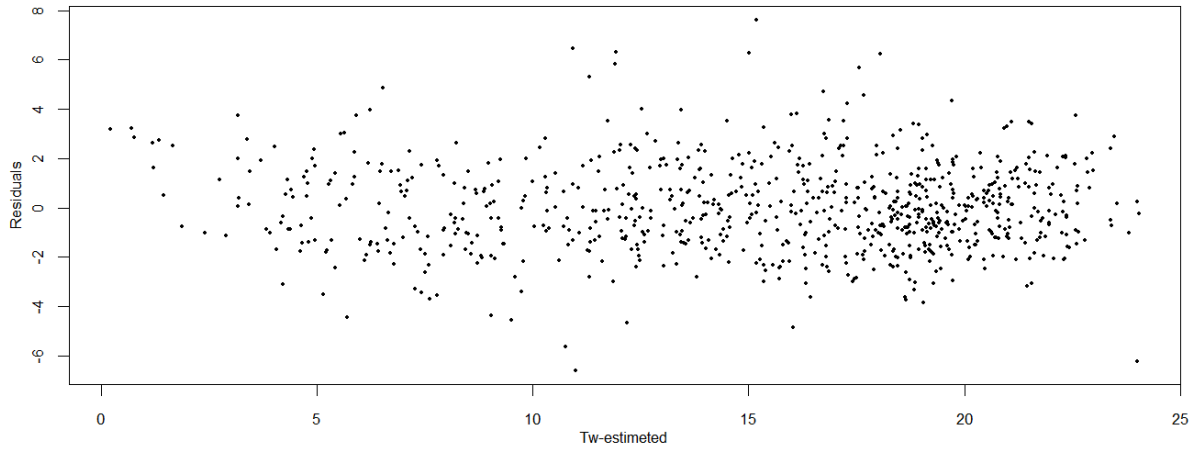


Figure 3.9 : Residuals versus estimated water temperature

3.5.2 Logistic model

The logistic model (Mohseni et al., 1998) which connects water temperature to air temperature by using Equation 3.6 was applied to the daily average of the water and the air temperature. The estimation of this logistic function is given by:

$$\text{Equation 3.15 : } T_w = \frac{26.72}{1 + e^{0.17(12.33 - T_a)}}$$

Water and air temperature daily means show a strong dispersion (Figure 3.10). The logistic function fits the general trend, but is not well suited for daily data, although it was shown to perform well at the weekly time step (Benyahya et al., 2007). The total variability explained by this model is 78%. The overall prediction error (RMSE) is 2.75 °C.

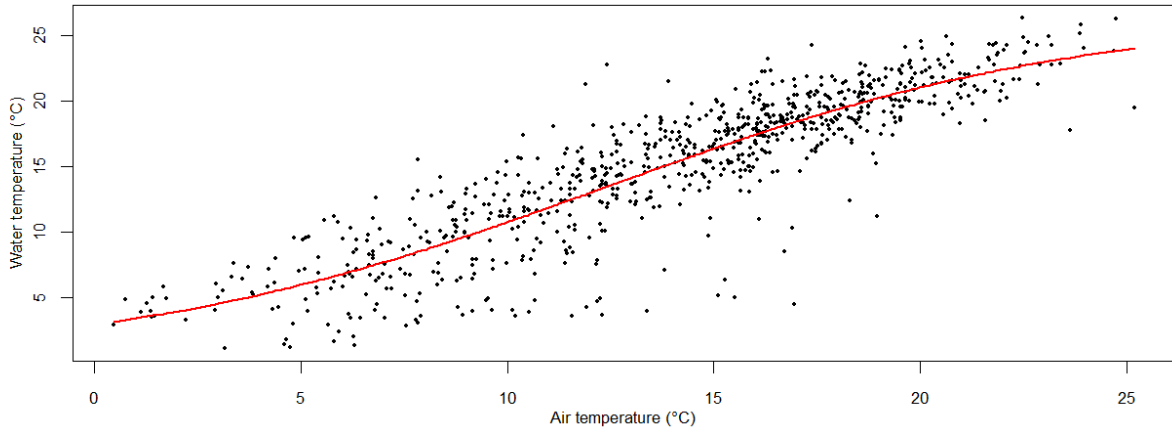


Figure 3.10 : Relation between daily mean water temperature and mean air temperature and a fitted logistic function

3.5.3 Residual regression model

Initially, the long-term seasonal component of water and air temperatures has been fitted with a sinusoidal function by applying Equations 3.9 and 3.10. It explains 80.3% and 67.5% of the total variability of water and air temperatures respectively. The highest daily average water temperature for the annual component is reached on August 1 (day 213 equivalent at $t=93$) and is 20.24 °C. In contrast, the highest daily average air temperature for the annual component is reached on July 24 (day 205 equivalent at $t=85$) and is 18.29 °C. For example, the Figure 3.11 shows the estimation of the water temperature components for 2013 using the following sinusoidal function:

$$\text{Equation 3.16 : } T_{A(t)} = 2.97 + 17.27 \sin \left[\frac{2\pi}{365} (t - 1.53) \right]$$

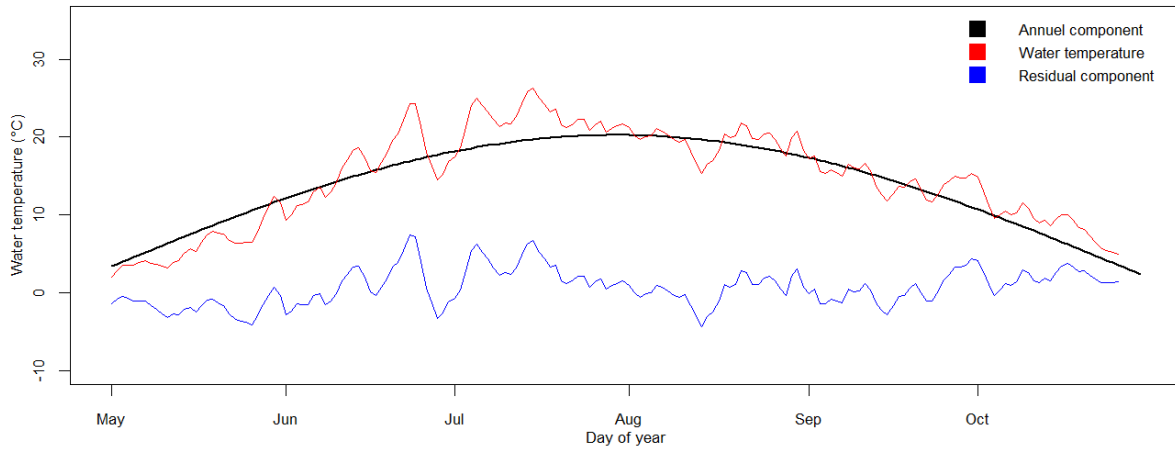


Figure 3.11 : Mean daily water temperature, annual component estimated by sinusoidal function and the residual component

Once the annual component was removed from both air and water temperature time series, the water temperature residuals were estimated by a regression model given by:

$$\text{Equation 3.17 : } R_{w(t)} = 0.51R_{a(t-1)} + 0.13R_{a(t-2)} + \varepsilon_{(t)}$$

The coefficients of lag1 and 2 for the air temperature residuals were significant in explaining the water temperature residuals and the intercept was insignificant. This model is better than the logistic model to represent the daily data. The air temperature residuals of lag 1 and 2 explained 44% of the total variability of water temperature residuals. However, the total variability explained by this model when the annual component was added is 89% and the overall prediction error (RMSE) is 2.09 °C.

3.5.4 Comparative study

The performance of all three models GAM, logistic and residuals regression was assessed using annual cross-validation and the three aforementioned indices: RMSE (equation 3.12), bias

(equation 3.13) and NSC (equation 3.14). These indices are calculated during the cross-validation for each year and for the entire period (see Table 3.3).

Overall, the GAM performs better than the other two models with a RMSE of 1.79 °C between predicted and observed values. Nevertheless, the logistic model and the residuals regression model, have an overall RMSE of 2.75 °C and 2.09 °C respectively. Both models perform less well than the GAM model but the residuals regression is between the GAM and logistic models. The annual values of RMSE vary between 1.28 °C and 2.18 °C for GAM model, between 1.83 °C and 3.40 °C for the logistic model and between 1.04 °C and 2.64 °C for the residuals regression model. The overall bias is very low and centered at zero but their range is wide for the two latter models (i.e. between -1.35 °C and 1.32 °C for GAM, between -1.08 and 1.99 °C for the logistic model and between -1.15 °C and 1.85 °C for the residuals regression model).

The NSC criterion indicates a good match of GAM model to these data with a value of 0.90. This result is confirmed by the graphical presentation (see Figure 3.12) of estimated values versus the observed values. This figure shows an alignment concentrated around the first bisector and indicates that the gap between the observed and the predicted values by the GAM model is less than with the other two models. The adequacy of logistic model for this case study is barely acceptable. The residuals regression results were also generally better than those of logistic model but not as good as GAM results. The NSC is 0.76 for the logistic and 0.86 for the residuals regression. Figure 3.13 shows the estimated values by the logistic model versus the observed values. It shows that this model overestimates the low values of water temperature. Nevertheless, Figure 3.14 demonstrates that the residuals regression model underestimates the highest values and overestimates the low values of water temperature.

This model does not account for flow rate so the overestimations of the water temperature were generally observed in the spring at higher water levels but the underestimations were observed during the summer when water levels are lower.

Table 3.3 : Results of the cross validation water temperature (°C) modeling expressed by RMSE, bias error and NSC by GAM, logistic and residuals regression approaches

| Model | GAM | | | Logistic | | | Residuals Regression | | |
|---------------|-----------------|------------------|-----------------|-----------------|------------------|-----------------|-----------------------------|------------------|------------------|
| Year | RMSE | Bias | NSC | RMSE | Bias | NSC | RMSE | Bias | NSC |
| 2007 | 1.28 | 0.47 | 0.86 | 1.86 | -0.88 | 0.69 | 1.33 | 0.53 | 0.85 |
| 2009 | 2.08 | 1.32 | 0.28 | 2.35 | 0.63 | 0.09 | 2.52 | 1.85 | -0.02 |
| 2010 | 1.83 | -0.28 | 0.87 | 1.83 | -1.08 | 0.87 | 2.55 | -1.05 | 0.74 |
| 2011 | 1.75 | 0.89 | 0.91 | 3.55 | 1.99 | 0.64 | 2.30 | 1.32 | 0.85 |
| 2012 | 1.32 | -0.01 | 0.95 | 2.03 | 0.06 | 0.89 | 1.08 | -0.45 | 0.97 |
| 2013 | 2.18 | -1.35 | 0.88 | 3.40 | -0.85 | 0.70 | 2.20 | -1.15 | 0.88 |
| Global | 1.79 | 0.006 | 0.90 | 2.75 | 0.01 | 0.76 | 2.09 | -0.03 | 0.86 |
| Range | 1.28 to 2.18 | -1.35 to 1.32 | 0.28 to 0.95 | 1.83 to 3.40 | -1.08 to 1.99 | 0.09 to 0.89 | 1.08 to 2.55 | -1.15 to 1.85 | -0.02 to 0.97 |

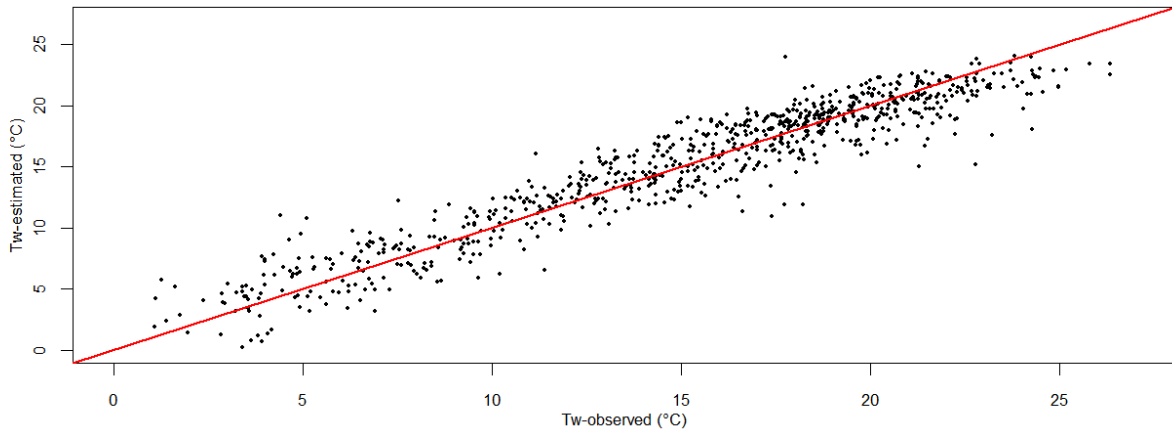


Figure 3.12 : Scatterplot of observed versus predicted values from the GAM for the season of May to October (2007 to 2013)

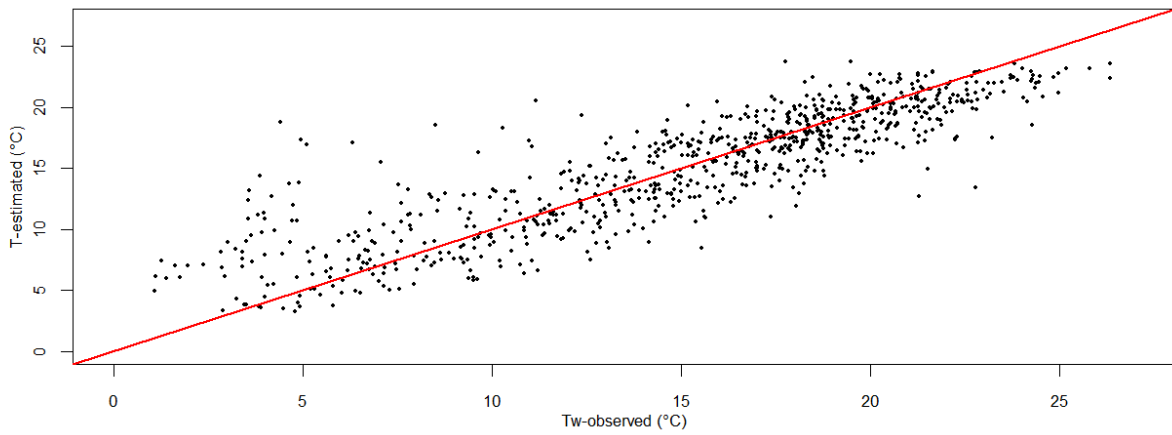


Figure 3.13 : Scatterplot of observed versus predicted values from the logistic for the season of May to October (2007 to 2013)

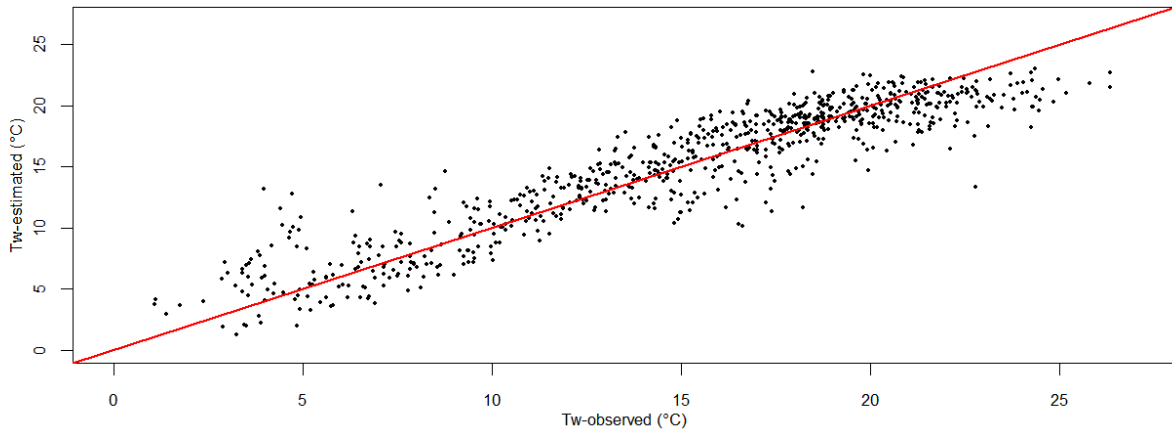


Figure 3.14 : Scatterplots of observed versus predicted values from residuals regression for the season of May to October

3.6 Discussion And Conclusions

This assessment will help to determine whether the GAM is a promising tool that can help for the management of aquatic resources by providing accurate temperature simulations/predictions. The performance of the GAM model was compared to the logistic and residuals regression models. The logistic model performed poorly for daily data (RMSE = 2.75 °C). This is consistent with the results of Caissie et al., (2001) on the maximum daily water temperature. Indeed, this model is more suited for longer time steps such as weekly data (Benyahya et al., 2007) for which it has been widely used. The residuals regression model that consists of separating the water and air temperatures times series into two components, seasonal component and residual component also had relatively acceptable performance (RMSE = 2.09 °C). Better performances were obtained with this type of model in previous studies such as that of Caissie et al., (1998). This model does not take into account the autocorrelation in the water temperature time series, which is usually significant (Caissie et al., 1998). The effect of the flow on water temperature is also

excluded from this model, as in the case of the logistic method. It should also be reminded that the period of this study contains three years with relatively warm water temperature, which is atypical for the Ste-Marguerite River (see Guillemette et al., (2009) for monthly maxima). Those limitations have impacted the performance of residuals regression for this river. Moreover, the strong correlation between lag 1 and lag 2 of the air temperature residuals causes multicollinearity problem for this regression, thus the results may be less accurate.

The results of the GAM model showed that it was very flexible to adjust non-linear relationships between water temperature and the explanatory variables (air temperature and the discharge), especially for atypical data as for this study. On an annual basis, this model gave a high RMSE for two years out of the six years of study. The first year for which the RMSE was relatively high is 2009 (RMSE = 2.08 °C). For this year, we adjusted the GAM model only for the warmer months due to missing data. Obtaining this RMSE is justified by the missing data during the cold period. The second year with a weaker performance, for the GAM model is 2013. This year is the warmest year and it differs from other years by the presence of the largest spring flood of the study period. This flood has generated a large volume of water that is added to the river base flow, which requires more energy to warm than years with smaller spring floods. The flow rate effect is marked by the presence of flooding but this effect is weak during the warm period when the air temperature effect dominates. The inclusion of flow allowed implicitly to reflect the specificities of each season. The residual analysis of the GAM model gives confidence in its use. To sum up, the GAM model has established nonlinear very flexible relationships between water temperature and the explanatory variables, air temperature and flow. It gave a better fit for the water temperature data of the Sainte-Marguerite River and especially it was more robust in the case of atypical data. To validate these preliminary findings of this study, applications on other

sites with no atypical data would be useful. Its validation on other sites will allow to judge whether it can be used as a potential tool for effective management of aquatic systems, to predict the impact of climate change on river temperature and to mitigate the effects of the latter on fish habitat.

3.7 References

- Acuña V & Tockner K (2009) Surface–subsurface water exchange rates along alluvial river reaches control the thermal patterns in an Alpine river network. *Freshwater Biology* 54(2):306-320.
- Ahmadi-Nedushan B, St-Hilaire A, Ouarda TB, Bilodeau L, Robichaud E, Thiemonge N & Bobee B (2007) Predicting river water temperatures using stochastic models: case study of the Moisie River (Quebec, Canada). *Hydrological Processes* 21(1):21-34.
- Arismendi I, Johnson SL, Dunham JB, Haggerty R & Hockman-Wert D (2012) The paradox of cooling streams in a warming world: regional climate trends do not parallel variable local trends in stream temperature in the Pacific continental United States. *Geophysical Research Letters* 39(10).
- Baird OE & Krueger CC (2003) Behavioral thermoregulation of brook and rainbow trout: comparison of summer habitat use in an Adirondack River, New York. *Transactions of the American Fisheries Society* 132(6):1194-1206.
- Bélanger M, El-Jabi N, Caissie D, Ashkar F & Ribí J (2005) Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple. *Revue des sciences de l'eau/Journal of Water Science* 18(3):403-421.
- Benyahya L, Caissie D, St-Hilaire A, Ouarda TB & Bobée B (2007) A review of statistical water temperature models. *Canadian Water Resources Journal* 32(3):179-192.
- Benyahya L, St-Hilaire A, Ouarda TB, Bobee B & Dumas J (2008) Comparison of non-parametric and parametric water temperature models on the Nivelles River, France. *Hydrological sciences journal* 53(3):640-655.
- Bhaskaran K, Gasparrini A, Hajat S, Smeeth L & Armstrong B (2013) Time series regression studies in environmental epidemiology. *International journal of epidemiology*:dyt092.
- Brett JR (1979) 10 Environmental Factors and Growth. *Fish Physiology*, W.S. Hoar DJR & Brett JR (Édit.) Academic Press, Vol Volume 8. p 599-675.
- Britton J, Cucherousset J, Davies G, Godard M & Copp G (2010) Non-native fishes and climate change: predicting species responses to warming temperatures in a temperate region. *Freshwater Biology* 55(5):1130-1141.
- Bruneau B & Grégoire F (2011) Étude de la distribution spatiale des données d'abondance de maquereau bleu (*Scomber scombrus*) et de capelan (*Mallotus villosus*) des relevés d'hiver aux poissons de fond des Divisions 4VW de l'OPANO à l'aide de modèles additifs généralisés. *Rapp. tech. can. sci. halieut. aquat.* 2930: vi + 22 pp.).
- Cairns JJ, Buikema ALJ, Heath AG & Parker BC (1978) Effects of temperature on aquatic organism sensitivity to selected chemicals. p Medium: X; Size: Pages: 92.
- Caissie D (2006) The thermal regime of rivers: a review. *Freshwater Biology* 51(8):1389-1406.
- Caissie D, El-Jabi N & Satish MG (2001) Modelling of maximum daily water temperatures in a small stream using air temperatures. *Journal of Hydrology* 251(1–2):14-28.
- 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(2):250-260.

- 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(3):303-315.
- Caldwell J, Rajagopalan B & Danner E (2014) Statistical Modeling of Daily Water Temperature Attributes on the Sacramento River. *Journal of Hydrologic Engineering*.
- Chenard JF & Caissie D (2008) Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. *Hydrological Processes* 22(17):3361-3372.
- Cluis DA (1972) Relationship between stream water temperature and ambient air temperature. *Nordic Hydrology* 3(2):65-71.
- Cole JC, Maloney KO, Schmid M & McKenna JE (2014) Developing and testing temperature models for regulated systems: A case study on the Upper Delaware River. *Journal of Hydrology* 519:588-598.
- Connor WP, Piston CE & Garcia AP (2003) Temperature during incubation as one factor affecting the distribution of Snake River fall Chinook salmon spawning areas. *Transactions of the American Fisheries Society* 132(6):1236-1243.
- Cox MM & Bolte JP (2007) A spatially explicit network-based model for estimating stream temperature distribution. *Environmental Modelling & Software* 22(4):502-514.
- Crozier LG & Zabel RW (2006) Climate impacts at multiple scales: evidence for differential population responses in juvenile Chinook salmon. *Journal of Animal Ecology* 75(5):1100-1109.
- Daigle A, Jeong DI & Lapointe MF (2014) Climate change and resilience of tributary thermal refugia for salmonids in eastern Canadian rivers. *Hydrological Sciences Journal* 10.1080/02626667.2014.898121:1-20.
- Daufresne M, Lengfellner K & Sommer U (2009) Global warming benefits the small in aquatic ecosystems. *Proceedings of the National Academy of Sciences* 106(31):12788-12793.
- De Boor C (1978) *A practical guide to splines*. Springer Verlag.,
- Demars BO, Russell Manson J, Olafsson JS, Gislason GM, Gudmundsdottir R, Woodward G, Reiss J, Pichler DE, Rasmussen JJ & Friberg N (2011) Temperature and the metabolic balance of streams. *Freshwater Biology* 56(6):1106-1121.
- DeWeber JT & Wagner T (2014) A regional neural network ensemble for predicting mean daily river water temperature. *Journal of Hydrology* 517:187-200.
- Ducharne A (2008) Importance of stream temperature to climate change impact on water quality. *Hydrology and Earth System Sciences Discussions* 12(3):797-810.
- Dugdale SJ, Bergeron NE & St-Hilaire A (2013) Temporal variability of thermal refuges and water temperature patterns in an Atlantic salmon river. *Remote Sensing of Environment* 136:358-373.
- Eaton J, McCormick J, Goodno B, O'brien D, Stefany H, Hondzo M & Scheller R (1995) A field information-based system for estimating fish temperature tolerances. *Fisheries* 20(4):10-18.
- 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 (2003) Thermal heterogeneity, stream channel morphology, and salmonid abundance in northeastern Oregon streams. *Canadian Journal of Fisheries and Aquatic Sciences* 60(10):1266-1280.

- Edinger JE, Duttweiler DW & Geyer JC (1968) The Response of Water Temperatures to Meteorological Conditions. *Water Resources Research* 4(5):1137-1143.
- Edwards PA & Cunjak RA (2007) Influence of water temperature and streambed stability on the abundance and distribution of slimy sculpin (*Cottus cognatus*). *Environmental biology of fishes* 80(1):9-22.
- Elliott J & Hurley M (1997) A functional model for maximum growth of Atlantic salmon parr, *Salmo salar*, from two populations in northwest England. *Functional Ecology* 11(5):592-603.
- Erickson TR & Stefan HG (2000) Linear air/water temperature correlations for streams during open water periods. *Journal of Hydrologic Engineering* 5(3):317-321.
- 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):336-346.
- Ficke AD, Myrick CA & Hansen LJ (2007) Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries* 17(4):581-613.
- Ficklin DL, Stewart IT & Maurer EP (2013) Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California. *Water Resources Research* 49(5):2765-2782.
- Gras R (1969) Simulation du comportement thermique d'une rivière à partir des données fournies par un réseau classique d'observations météorologiques. *Proc. Int. Assoc. Hydraul. Res., Kyoto* 1:491-502.
- Grégoire Y & Trenchia G (2007) Influence de l'ombrage produit par la végétation riveraine sur la température de l'eau: un paramètre d'importance pour le maintien d'un habitat de qualité pour le poisson. Ministère des Ressources Naturelles et de la Faune, Secteur Faune Québec. Direction de l'aménagement de la faune de la région de la Chaudière-Appalaches.).
- Guay J, Boisclair D, Rioux D, Leclerc M, Lapointe M & Legendre P (2000) Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* 57(10):2065-2075.
- Guillemette N, St-Hilaire A, Ouarda TB, Bergeron N, Robichaud É & Bilodeau L (2009) Feasibility study of a geostatistical modelling of monthly maximum stream temperatures in a multivariate space. *Journal of hydrology* 364(1):1-12.
- Hari RE, Livingstone DM, Siber R, BURKHARDT-HOLM P & Guettinger 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.
- Hastie T & Tibshirani R (1986) Generalized additive models. *Statistical science*:297-310.
- IPCC (2013) Changements Climatiques 2013 Les éléments scientifiques, Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat.).
- Isaak D, Wollrab S, Horan D & Chandler G (2012) Climate change effects on stream and river temperatures across the northwest US from 1980–2009 and implications for salmonid fishes. *Climatic Change* 113(2):499-524.
- Janssen P & Heuberger P (1995) Calibration of process-oriented models. *Ecological Modelling* 83(1):55-66.
- Jbilou J & El Adlouni S (2012) *Generalized additive models in environmental health: a literature review*. INTECH Open Access Publisher,

- Kaushal SS, Likens GE, Jaworski NA, Pace ML, Sides AM, Seekell D, Belt KT, Secor DH & Wingate RL (2010) Rising stream and river temperatures in the United States. *Frontiers in Ecology and the Environment* 8(9):461-466.
- Kothandaraman V (1971) Analysis of water temperature variations in large river. *Journal of the Sanitary Engineering Division* 97(1):19-31.
- Liu B, Yang D, Ye B & Berezovskaya S (2005) Long-term open-water season stream temperature variations and changes over Lena River Basin in Siberia. *Global and Planetary Change* 48(1-3):96-111.
- 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.
- Magnuson JJ, Crowder LB & Medvick PA (1979) Temperature as an ecological resource. *American Zoologist* 19(1):331-343.
- Marceau P, Cluis D & Morin G (1986) Comparaison des performances relatives à un modèle déterministe et à un modèle stochastique de température de l'eau en rivière. *Canadian Journal of Civil Engineering* 13(3):352-364.
- McCullagh P & Nelder JA (1989) *Generalized linear models*. CRC press,
- Moatar F & Gailhard J (2006) Water temperature behaviour in the River Loire since 1976 and 1881. *Comptes Rendus Geoscience* 338(5):319-328.
- Mohseni O, Erickson TR & Stefan HG (1999) Sensitivity of stream temperatures in the United States to air temperatures projected under a global warming scenario. *Water Resources Research* 35(12):3723-3733.
- Mohseni O & Stefan H (1999) Stream temperature/air temperature relationship: a physical interpretation. *Journal of Hydrology* 218(3):128-141.
- Mohseni O, Stefan HG & Eaton JG (2003) Global warming and potential changes in fish habitat in US streams. *Climatic change* 59(3):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.
- Morrill JC, Bales RC & Conklin MH (2005) Estimating stream temperature from air temperature: implications for future water quality. *Journal of Environmental Engineering* 131(1):139-146.
- Nash J & Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology* 10(3):282-290.
- Nelson KC & Palmer MA (2007) *Stream Temperature Surges Under Urbanization and Climate Change: Data, Models, and Responses* 1. (Wiley Online Library).
- Neuheimer AB & Taggart CT (2007) The growing degree-day and fish size-at-age: the overlooked metric. *Canadian Journal of Fisheries and Aquatic Sciences* 64(2):375-385.
- Nohair M, St-Hilaire A & Ouarda T (2008) Utilisation des réseaux de neurones et de la régularisation bayésienne en modélisation de la température de l'eau en rivière. *Revue des sciences de l'eau/Journal of Water Science* 21(3):373-382.
- Olden JD & Naiman RJ (2010) Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. *Freshwater Biology* 55(1):86-107.

- Ozaki N, Fukushima T, Harasawa H, Kojiri T, Kawashima K & Ono M (2003) Statistical analyses on the effects of air temperature fluctuations on river water qualities. *Hydrological Processes* 17(14):2837-2853.
- Parmesan C & Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421(6918):37-42.
- Peng RD & Dominici F (2008) Statistical methods for environmental epidemiology with R. *R: a case study in air pollution and health*, Springer.
- Pilgrim JM, Fang X & Stefan HG (1998) STREAM TEMPERATURE CORRELATIONS WITH AIR TEMPERATURES IN MINNESOTA: IMPLICATIONS FOR CLIMATE WARMING1. *JAWRA Journal of the American Water Resources Association* 34(5):1109-1121.
- Prats J, Val R, Dolz J & Armengol J (2012) Water temperature modeling in the Lower Ebro River (Spain): Heat fluxes, equilibrium temperature, and magnitude of alteration caused by reservoirs and thermal effluent. *Water Resources Research* 48(5).
- Preece RM & Jones HA (2002) The effect of Keepit Dam on the temperature regime of the Namoi River, Australia. *River Research and Applications* 18(4):397-414.
- Quenouille MH (1949) Approximate tests of correlation in time-series 3. *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge Univ Press, p 483-484.
- Quigley JT & Hinch SG (2006) Effects of rapid experimental temperature increases on acute physiological stress and behaviour of stream dwelling juvenile chinook salmon. *Journal of Thermal Biology* 31(5):429-441.
- Risley JC, Roehl EA & Conrads PA (2003) Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models. in *U.S. GEOLOGICAL SURVEY, Water-Resources Investigations Report 02-4218* Portland, Oregon).
- Secondat M (1952) Les variations de la température et de la concentration en oxygène dissous des eaux lacustres et des eaux courantes. Leur retentissement sur la distribution des poissons. *Bull. Fr. Piscic.* (167):52-59.
- Segura C, Caldwell P, Sun G, McNulty S & Zhang Y (2014) A model to predict stream water temperature across the conterminous USA. *Hydrological Processes*.
- Sharma S, Walker SC & Jackson DA (2008) Empirical modelling of lake water-temperature relationships: a comparison of approaches. *Freshwater biology* 53(5):897-911.
- Smith K (1981) The prediction of river water temperatures/Prédiction des températures des eaux de rivière. *Hydrological Sciences Journal* 26(1):19-32.
- 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(6):1095-1108.
- St-Hilaire A, Ouarda TBMJ, Bargaoui Z, Daigle A & Bilodeau L (2012) Daily river water temperature forecast model with a k-nearest neighbour approach. *Hydrological Processes* 26(9):1302-1310.
- St-Hilaire A, Ouarda TB, Bargaoui Z, Daigle A & Bilodeau L (2012) Daily river water temperature forecast model with ak-nearest neighbour approach. *Hydrological Processes* 26(9):1302-1310.
- Stefan HG & Preud'homme EB (1993) STREAM TEMPERATURE ESTIMATION FROM AIR TEMPERATURE. *JAWRA Journal of the American Water Resources Association* 29(1):27-45.

- Stevens HH, Ficke JF & Smoot GF (1975) *Water temperature-influential factors, field measurement, and data presentation*. US Government Printing Office,
- Team RC (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. (ISBN 3-900051-07-0).
- Turgeon S (2012) Modélisation de l'utilisation de l'habitat du béluga du Saint-Laurent en fonction de ses proies à l'embouchure de la rivière Saguenay et à la baie Sainte-Marguerite.
- Van Vliet M, Ludwig F, Zwolsman J, Weedon G & Kabat P (2011) Global river temperatures and sensitivity to atmospheric warming and changes in river flow. *Water Resources Research* 47(2).
- Vannote RL, Minshall GW, Cummins KW, Sedell JR & Cushing CE (1980) The river continuum concept. *Canadian journal of fisheries and aquatic sciences* 37(1):130-137.
- Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJ, Fromentin J-M, Hoegh-Guldberg O & Bairlein F (2002) Ecological responses to recent climate change. *Nature* 416(6879):389-395.
- Webb B (1996) Trends in stream and river temperature. *Hydrological processes* 10(2):205-226.
- Webb B, Walsh A, Webb B, Acreman M, Maksimovic C, Smithers H & Kirby C (2004) Changing UK river temperatures and their impact on fish populations. *Hydrology: science and practice for the 21st century, Volume II. Proceedings of the British Hydrological Society International Conference, Imperial College, London, July 2004*. British Hydrological Society, p 177-191.
- Webb BW, Clack PD & Walling DE (2003) Water-air temperature relationships in a Devon river system and the role of flow. *Hydrological Processes* 17(15):3069-3084.
- 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.
- Wehrly KE, Brenden TO & Wang L (2009) A Comparison of Statistical Approaches for Predicting Stream Temperatures Across Heterogeneous Landscapes. (Wiley Online Library).
- Wood S (2006) *Generalized additive models: an introduction with R*. CRC press,
- Wood S (2013) mgcv: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package version 1.7-22.).

4 CONCLUSIONS ET RECOMMANDATIONS

L'objectif principal de la présente étude consistait à évaluer la performance de la modélisation non paramétrique MAG à maximiser la qualité de la prévision de la température de l'eau. Cette évaluation permettait de vérifier l'applicabilité du modèle MAG comme une nouvelle approche de modélisation statistique appropriée pour une prédiction précise de la température de l'eau, de mettre en valeur son intérêt en le comparant à deux approches statistiques déjà existantes. La prise en compte de l'impact du débit sur la relation température de l'eau – température de l'air dans le modèle MAG donne lieu à une quantification de son effet afin de pouvoir l'inclure comme variable explicative en situations aménagées où le débit est contrôlé (ex. barrages). Les résultats de cette étude ont permis de vérifier la performance du modèle MAG en déterminant sa précision, son biais et son adéquation à travers sa comparaison aux deux modèles populaires, le logistique et le modèle de régression des résidus. La mise en œuvre de ces trois approches a été complétée en utilisant les séries de données d'un site de la rivière Ste-Marguerite (Canada). Le modèle MAG a donné des prévisions plus précises que celles du modèle logistique et du modèle de régression des résidus. Le modèle logistique est plus approprié pour présenter les données hebdomadaires. Ainsi, ces deux modèles de comparaison ignorent l'effet du temps présent dans les données (Li et al., 2014). En effet, la décomposition de la série en composante annuelle et résiduelle par le modèle de régression des résidus ne permet pas tout à fait de reproduire la saisonnalité des données. Le MAG était plus adéquat (NSC=0.90) pour représenter les données atypiques (années chaudes). Ce modèle a montré la présence et la significativité de l'effet du débit à côté de celui de la température de l'air. Cependant, ce modèle ne prend pas en considération les effets retardés de la température de l'air qui étaient significatifs dans l'étude de Bélanger et al., (2005) et de Benyahya et al., (2008). L'inclusion des effets retardés crée la

multicolinéarité des variables explicatives qui peut influencer la précision des résultats. Afin d'éviter ce problème de multicolinéarité, la régression Ridge (Hoerl & Kennard, 1970) pourrait être intéressante. Par ailleurs, puisque ces effets peuvent être non linéaires, l'application de l'approche des modèles à effets retardés distribués non linéaires (Distributed Lag Non Linear Model : DLNM), introduits par Almon (1965) en économétrie, puis par Schwartz (Schwartz, 2000) et Gasparrini et al., (2010) en épidémiologie, serait également une méthode prometteuse. Ces modèles expliquent la variable d'intérêt au temps en fonction des effets retardés des variables explicatives. Il serait très intéressant aussi de comparer le MAG à une approche non paramétrique performante déjà utilisée tels que les réseaux de neurones ou les k-plus proche voisins, appliqués sur d'autres séries de données contenant moins d'observations atypiques que dans la présente étude de cas (trois années atypiquement chaudes sur 6). Enfin, on peut conclure que le MAG est une approche prometteuse qui peut être utilisée pour la gestion des ressources aquatiques et par les scientifiques qui s'intéressent à développer leurs propres modèles.

5 RÉFÉRENCES

- Acuña V & Tockner K (2009) Surface–subsurface water exchange rates along alluvial river reaches control the thermal patterns in an Alpine river network. *Freshwater Biology* 54(2):306-320.
- Ahmadi-Nedushan B, St-Hilaire A, Ouarda TB, Bilodeau L, Robichaud E, Thiemonge N & Bobee B (2007) Predicting river water temperatures using stochastic models: case study of the Moisie River (Quebec, Canada). *Hydrological Processes* 21(1):21-34.
- Almon S (1965) The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*:178-196.
- Arismendi I, Johnson SL, Dunham JB, Haggerty R & Hockman-Wert D (2012) The paradox of cooling streams in a warming world: regional climate trends do not parallel variable local trends in stream temperature in the Pacific continental United States. *Geophysical Research Letters* 39(10).
- Baird OE & Krueger CC (2003) Behavioral thermoregulation of brook and rainbow trout: comparison of summer habitat use in an Adirondack River, New York. *Transactions of the American Fisheries Society* 132(6):1194-1206.
- Bartholow JM, Campbell SG & Flug M (2004) Predicting the thermal effects of dam removal on the Klamath River. *Environmental Management* 34(6):856-874.
- Bélangier M, El-Jabi N, Caissie D, Ashkar F & Ribí J (2005) Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple. *Revue des sciences de l'eau/Journal of Water Science* 18(3):403-421.
- Benyahya L, Caissie D, St-Hilaire A, Ouarda TB & Bobée B (2007) A review of statistical water temperature models. *Canadian Water Resources Journal* 32(3):179-192.
- Benyahya L, St-Hilaire A, Ouarda TB, Bobee B & Dumas J (2008) Comparison of non-parametric and parametric water temperature models on the Nivelles River, France. *Hydrological sciences journal* 53(3):640-655.
- Beschta RL (1997) Riparian shade and stream temperature: an alternative perspective. *Rangelands*:25-28.
- Bhaskaran K, Gasparrini A, Hajat S, Smeeth L & Armstrong B (2013) Time series regression studies in environmental epidemiology. *International journal of epidemiology*:dyt092.
- Blann K, Frost Nerbonne J & Vondracek B (2002) Relationship of riparian buffer type to water temperature in the driftless area ecoregion of Minnesota. *North American Journal of Fisheries Management* 22(2):441-451.
- Brett JR (1979) 10 Environmental Factors and Growth. *Fish Physiology*, W.S. Hoar DJR & Brett JR (Édit.) Academic Press, Vol Volume 8. p 599-675.
- Britton J, Cucherousset J, Davies G, Godard M & Copp G (2010) Non-native fishes and climate change: predicting species responses to warming temperatures in a temperate region. *Freshwater Biology* 55(5):1130-1141.
- Bruneau B & Grégoire F (2011) Étude de la distribution spatiale des données d'abondance de maquereau bleu (*Scomber scombrus*) et de capelan (*Mallotus villosus*) des relevés d'hiver aux poissons de fond des Divisions 4VW de l'OPANO à l'aide de modèles additifs généralisés. Rapp. tech. can. sci. halieut. aquat. 2930: vi + 22 pp.).
- Cairns JJ, Buikema ALJ, Heath AG & Parker BC (1978) Effects of temperature on aquatic organism sensitivity to selected chemicals., p Medium: X; Size: Pages: 92.
- Caissie D (2006) The thermal regime of rivers: a review. *Freshwater Biology* 51(8):1389-1406.

- Caissie D, El-Jabi N & Satish MG (2001) Modelling of maximum daily water temperatures in a small stream using air temperatures. *Journal of Hydrology* 251(1–2):14-28.
- 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(2):250-260.
- 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(3):303-315.
- Caldwell J, Rajagopalan B & Danner E (2014) Statistical Modeling of Daily Water Temperature Attributes on the Sacramento River. *Journal of Hydrologic Engineering*.
- Caldwell P, Segura C, Gull Laird S, Sun G, McNulty SG, Sandercock M, Boggs J & Vose JM (2015) Short-term stream water temperature observations permit rapid assessment of potential climate change impacts. *Hydrological Processes* 29(9):2196-2211.
- Chen YD, Carsel RF, McCutcheon SC & Nutter WL (1998) Stream temperature simulation of forested riparian areas: I. Watershed-scale model development. *Journal of Environmental Engineering* 124(4):304-315.
- Chenard JF & Caissie D (2008) Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. *Hydrological Processes* 22(17):3361-3372.
- Cluis DA (1972) Relationship between stream water temperature and ambient air temperature. *Nordic Hydrology* 3(2):65-71.
- Cole JC, Maloney KO, Schmid M & McKenna JE (2014) Developing and testing temperature models for regulated systems: A case study on the Upper Delaware River. *Journal of Hydrology* 519:588-598.
- Connor WP, Piston CE & Garcia AP (2003) Temperature during incubation as one factor affecting the distribution of Snake River fall Chinook salmon spawning areas. *Transactions of the American Fisheries Society* 132(6):1236-1243.
- Cox MM & Bolte JP (2007) A spatially explicit network-based model for estimating stream temperature distribution. *Environmental Modelling & Software* 22(4):502-514.
- Crozier LG, Hendry A, Lawson PW, Quinn T, Mantua N, Battin J, Shaw R & Huey R (2008) Potential responses to climate change in organisms with complex life histories: evolution and plasticity in Pacific salmon. *Evolutionary Applications* 1(2):252-270.
- Crozier LG & Zabel RW (2006) Climate impacts at multiple scales: evidence for differential population responses in juvenile Chinook salmon. *Journal of Animal Ecology* 75(5):1100-1109.
- Daigle A, Jeong DI & Lapointe MF (2014) Climate change and resilience of tributary thermal refugia for salmonids in eastern Canadian rivers. *Hydrological Sciences Journal* 10.1080/02626667.2014.898121:1-20.
- Daufresne M, Lengfellner K & Sommer U (2009) Global warming benefits the small in aquatic ecosystems. *Proceedings of the National Academy of Sciences* 106(31):12788-12793.
- De Boor C (1978) *A practical guide to splines*. Springer Verlag.,
- Demars BO, Russell Manson J, Olafsson JS, Gislason GM, Gudmundsdottir R, Woodward G, Reiss J, Pichler DE, Rasmussen JJ & Friberg N (2011) Temperature and the metabolic balance of streams. *Freshwater Biology* 56(6):1106-1121.
- DeWeber JT & Wagner T (2014) A regional neural network ensemble for predicting mean daily river water temperature. *Journal of Hydrology* 517:187-200.

- Ducharne A (2008) Importance of stream temperature to climate change impact on water quality. *Hydrology and Earth System Sciences Discussions* 12(3):797-810.
- Dugdale SJ, Bergeron NE & St-Hilaire A (2013) Temporal variability of thermal refuges and water temperature patterns in an Atlantic salmon river. *Remote Sensing of Environment* 136:358-373.
- Eaton J, McCormick J, Goodno B, O'brien D, Stefany H, Hondzo M & Scheller R (1995) A field information-based system for estimating fish temperature tolerances. *Fisheries* 20(4):10-18.
- 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 (2003) Thermal heterogeneity, stream channel morphology, and salmonid abundance in northeastern Oregon streams. *Canadian Journal of Fisheries and Aquatic Sciences* 60(10):1266-1280.
- Edinger JE, Duttweiler DW & Geyer JC (1968) The Response of Water Temperatures to Meteorological Conditions. *Water Resources Research* 4(5):1137-1143.
- Edwards PA & Cunjak RA (2007) Influence of water temperature and streambed stability on the abundance and distribution of slimy sculpin (*Cottus cognatus*). *Environmental biology of fishes* 80(1):9-22.
- Elliott J & Hurley M (1997) A functional model for maximum growth of Atlantic salmon parr, *Salmo salar*, from two populations in northwest England. *Functional Ecology* 11(5):592-603.
- Erickson TR & Stefan HG (2000) Linear air/water temperature correlations for streams during open water periods. *Journal of Hydrologic Engineering* 5(3):317-321.
- 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):336-346.
- Ficke AD, Myrick CA & Hansen LJ (2007) Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries* 17(4):581-613.
- Ficklin DL, Stewart IT & Maurer EP (2013) Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California. *Water Resources Research* 49(5):2765-2782.
- Garner G, Hannah DM, Sadler JP & Orr HG (2014) River temperature regimes of England and Wales: spatial patterns, inter-annual variability and climatic sensitivity. *Hydrological Processes* 28(22):5583-5598.
- Gasparrini A, Armstrong B & Kenward M (2010) Distributed lag non-linear models. *Statistics in medicine* 29(21):2224.
- GIEC (2013) Changements Climatiques 2013 Les éléments scientifiques, Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat.).
- Gras R (1969) Simulation du comportement thermique d'une rivière à partir des données fournies par un réseau classique d'observations météorologiques. *Proc. Int. Assoc. Hydraul. Res., Kyoto* 1:491-502.
- Grégoire Y & Trencia G (2007) Influence de l'ombrage produit par la végétation riveraine sur la température de l'eau: un paramètre d'importance pour le maintien d'un habitat de qualité pour le poisson. Ministère des Ressources Naturelles et de la Faune, Secteur Faune Québec. Direction de l'aménagement de la faune de la région de la Chaudière-Appalaches.).

- Guay J, Boisclair D, Rioux D, Leclerc M, Lapointe M & Legendre P (2000) Development and validation of numerical habitat models for juveniles of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences* 57(10):2065-2075.
- Guillemette N, St-Hilaire A, Ouarda TB, Bergeron N, Robichaud É & Bilodeau L (2009) Feasibility study of a geostatistical modelling of monthly maximum stream temperatures in a multivariate space. *Journal of hydrology* 364(1):1-12.
- Hannah DM, Malcolm IA, Soulsby C & Youngson AF (2008) A comparison of forest and moorland stream microclimate, heat exchanges and thermal dynamics. *Hydrological Processes* 22(7):919-940.
- Hari RE, Livingstone DM, Siber R, BURKHARDT-HOLM P & Guettinger 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.
- Hastie T & Tibshirani R (1986) Generalized additive models. *Statistical science*:297-310.
- Hoerl AE & Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1):55-67.
- IPCC (2013) Changements Climatiques 2013 Les éléments scientifiques, Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat.).
- Isaak D, Wollrab S, Horan D & Chandler G (2012) Climate change effects on stream and river temperatures across the northwest US from 1980–2009 and implications for salmonid fishes. *Climatic Change* 113(2):499-524.
- Janssen P & Heuberger P (1995) Calibration of process-oriented models. *Ecological Modelling* 83(1):55-66.
- Jbilou J & El Adlouni S (2012) *Generalized additive models in environmental health: a literature review*. INTECH Open Access Publisher,
- 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(S2):30-39.
- Jones KL, Poole GC, Meyer JL, Bumback W & Kramer EA (2006) Quantifying expected ecological response to natural resource legislation: a case study of riparian buffers, aquatic habitat, and trout populations. *Ecology and Society* 11(2):15.
- Kaushal SS, Likens GE, Jaworski NA, Pace ML, Sides AM, Seekell D, Belt KT, Secor DH & Wingate RL (2010) Rising stream and river temperatures in the United States. *Frontiers in Ecology and the Environment* 8(9):461-466.
- Kothandaraman V (1971) Analysis of water temperature variations in large river. *Journal of the Sanitary Engineering Division* 97(1):19-31.
- Li H, Deng X, Kim D-Y & Smith EP (2014) Modeling maximum daily temperature using a varying coefficient regression model. *Water Resources Research* 50(4):3073-3087.
- Liu B, Yang D, Ye B & Berezovskaya S (2005) Long-term open-water season stream temperature variations and changes over Lena River Basin in Siberia. *Global and Planetary Change* 48(1–3):96-111.
- 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.
- Macdonald JS, Patterson DA, Hague MJ & Guthrie IC (2010) Modeling the influence of environmental factors on spawning migration mortality for sockeye salmon fisheries

- management in the Fraser River, British Columbia. *Transactions of the American Fisheries Society* 139(3):768-782.
- Magnuson JJ, Crowder LB & Medvick PA (1979) Temperature as an ecological resource. *American Zoologist* 19(1):331-343.
- Maheu A, St-Hilaire A, Caissie D & El-Jabi N (2015) *Développement d'outils de caractérisation et de modélisation du régime thermique des rivières naturelles et régulées*. (Université du Québec, Institut national de la recherche scientifique, Centre Eau-Terre-Environnement).
- Marceau P, Cluis D & Morin G (1986) Comparaison des performances relatives à un modèle déterministe et à un modèle stochastique de température de l'eau en rivière. *Canadian Journal of Civil Engineering* 13(3):352-364.
- Marsh P (1990) Modelling water temperature beneath river ice covers. *Canadian Journal of Civil Engineering* 17(1):36-44.
- Martins EG, Hinch SG, Patterson DA, Hague MJ, Cooke SJ, Miller KM, Robichaud D, English KK, Farrell AP & Jonsson B (2012) High river temperature reduces survival of sockeye salmon (*Oncorhynchus nerka*) approaching spawning grounds and exacerbates female mortality. *Canadian journal of fisheries and aquatic sciences* 69(2):330-342.
- McCullagh P & Nelder JA (1989) *Generalized linear models*. CRC press,
- Moatar F & Gailhard J (2006) Water temperature behaviour in the River Loire since 1976 and 1881. *Comptes Rendus Geoscience* 338(5):319-328.
- Mohseni O, Erickson TR & Stefan HG (1999) Sensitivity of stream temperatures in the United States to air temperatures projected under a global warming scenario. *Water Resources Research* 35(12):3723-3733.
- Mohseni O & Stefan H (1999) Stream temperature/air temperature relationship: a physical interpretation. *Journal of Hydrology* 218(3):128-141.
- Mohseni O, Stefan HG & Eaton JG (2003) Global warming and potential changes in fish habitat in US streams. *Climatic change* 59(3):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 R, Spittlehouse D & Story A (2005) RIPARIAN MICROCLIMATE AND STREAM TEMPERATURE RESPONSE TO FOREST HARVESTING: A REVIEW1. (Wiley Online Library).
- Moore RD (2006) Stream temperature patterns in British Columbia, Canada, based on routine spot measurements. *Canadian Water Resources Journal* 31(1):41-56.
- Morrill JC, Bales RC & Conklin MH (2005) Estimating stream temperature from air temperature: implications for future water quality. *Journal of Environmental Engineering* 131(1):139-146.
- Mosley MP (1983) Variability of water temperatures in the braided Ashley and Rakaia rivers. *New Zealand Journal of Marine and Freshwater Research* 17(3):331-342.
- Nash J & Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology* 10(3):282-290.
- Nelson KC & Palmer MA (2007) Stream Temperature Surges Under Urbanization and Climate Change: Data, Models, and Responses1. (Wiley Online Library).
- Neuheimer AB & Taggart CT (2007) The growing degree-day and fish size-at-age: the overlooked metric. *Canadian Journal of Fisheries and Aquatic Sciences* 64(2):375-385.

- Nohair M, St-Hilaire A & Ouarda T (2008) Utilisation des réseaux de neurones et de la régularisation bayésienne en modélisation de la température de l'eau en rivière. *Revue des sciences de l'eau/Journal of Water Science* 21(3):373-382.
- Olden JD & Naiman RJ (2010) Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. *Freshwater Biology* 55(1):86-107.
- Orr HG, Simpson GL, Clers S, Watts G, Hughes M, Hannaford J, Dunbar MJ, Laizé CL, Wilby RL & Battarbee RW (2014) Detecting changing river temperatures in England and Wales. *Hydrological Processes*.
- Ouellet V (2011) *Température de l'eau du fleuve Saint-Laurent: extrêmes hydroclimatiques, modélisation 2D et application à l'habitat du poisson*. (Université du Québec).
- Ozaki N, Fukushima T, Harasawa H, Kojiri T, Kawashima K & Ono M (2003) Statistical analyses on the effects of air temperature fluctuations on river water qualities. *Hydrological Processes* 17(14):2837-2853.
- Parmesan C & Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421(6918):37-42.
- Peng RD & Dominici F (2008) Statistical methods for environmental epidemiology with R. *R: a case study in air pollution and health*, Springer.
- Pilgrim JM, Fang X & Stefan HG (1998) STREAM TEMPERATURE CORRELATIONS WITH AIR TEMPERATURES IN MINNESOTA: IMPLICATIONS FOR CLIMATE WARMING1. *JAWRA Journal of the American Water Resources Association* 34(5):1109-1121.
- Poirel A, Gailhard J & Capra H (2010) Influence des barrages-réservoirs sur la température de l'eau: exemple d'application au bassin versant de l'Ain. *La Houille Blanche* (4):72-79.
- Prats J, Val R, Dolz J & Armengol J (2012) Water temperature modeling in the Lower Ebro River (Spain): Heat fluxes, equilibrium temperature, and magnitude of alteration caused by reservoirs and thermal effluent. *Water Resources Research* 48(5).
- Preece RM & Jones HA (2002) The effect of Keepit Dam on the temperature regime of the Namoi River, Australia. *River Research and Applications* 18(4):397-414.
- Quenouille MH (1949) Approximate tests of correlation in time-series 3. *Mathematical Proceedings of the Cambridge Philosophical Society*. Cambridge Univ Press, p 483-484.
- Quigley JT & Hinch SG (2006) Effects of rapid experimental temperature increases on acute physiological stress and behaviour of stream dwelling juvenile chinook salmon. *Journal of Thermal Biology* 31(5):429-441.
- R Development Core Team (2012) R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0*.
- Risley JC, Roehl EA & Conrads PA (2003) Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models. in *U.S. GEOLOGICAL SURVEY, Water-Resources Investigations Report 02-4218* Portland, Oregon).
- Schwartz J (2000) The distributed lag between air pollution and daily deaths. *Epidemiology* 11(3):320-326.
- Secondat M (1952) Les variations de la température et de la concentration en oxygène dissous des eaux lacustres et des eaux courantes. Leur retentissement sur la distribution des poissons. *Bull. Fr. Piscic.* (167):52-59.
- Segura C, Caldwell P, Sun G, McNulty S & Zhang Y (2014) A model to predict stream water temperature across the conterminous USA. *Hydrological Processes*.

- Sharma S, Walker SC & Jackson DA (2008) Empirical modelling of lake water-temperature relationships: a comparison of approaches. *Freshwater biology* 53(5):897-911.
- Singer EE & Gangloff MM (2011) Effects of a small dam on freshwater mussel growth in an Alabama (USA) stream. *Freshwater Biology* 56(9):1904-1915.
- Sinokrot BA & Stefan HG (1993) Stream temperature dynamics: measurements and modeling. *Water Resources Research* 29(7):2299-2312.
- Smith K (1981) The prediction of river water temperatures/Prédiction des températures des eaux de rivière. *Hydrological Sciences Journal* 26(1):19-32.
- Sridhar V, Sansone AL, LaMarche J, Dubin T & Lettenmaier DP (2004) PREDICTION OF STREAM TEMPERATURE IN FORESTED WATERSHEDS1. (Wiley Online Library).
- St-Hilaire A, Duchesne S & Rousseau AN (2015) Floods and water quality in Canada: A review of the interactions with urbanization, agriculture and forestry. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques* (ahead-of-print):1-15.
- 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(6):1095-1108.
- St-Hilaire A, Ouarda TBMJ, Bargaoui Z, Daigle A & Bilodeau L (2012) Daily river water temperature forecast model with a k-nearest neighbour approach. *Hydrological Processes* 26(9):1302-1310.
- St-Hilaire A, Ouarda TB, Bargaoui Z, Daigle A & Bilodeau L (2012) Daily river water temperature forecast model with ak-nearest neighbour approach. *Hydrological Processes* 26(9):1302-1310.
- Stefan HG & Preud'homme EB (1993a) STREAM TEMPERATURE ESTIMATION FROM AIR TEMPERATURE. *JAWRA Journal of the American Water Resources Association* 29(1):27-45.
- Stefan HG & Preud'Homme EB (1993b) STREAM TEMPERATURE ESTIMATION FROM AIR TEMPERATURE1. (Wiley Online Library).
- Stevens HH, Ficke JF & Smoot GF (1975) *Water temperature-influential factors, field measurement, and data presentation*. US Government Printing Office,
- Team RC (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012. (ISBN 3-900051-07-0).
- Toffolon M, Siviglia A & Zolezzi G (2010) Thermal wave dynamics in rivers affected by hydropeaking. *Water Resources Research* 46(8).
- Turgeon S (2012) Modélisation de l'utilisation de l'habitat du béluga du Saint-Laurent en fonction de ses proies à l'embouchure de la rivière Saguenay et à la baie Sainte-Marguerite.
- Van Vliet M, Ludwig F, Zwolsman J, Weedon G & Kabat P (2011) Global river temperatures and sensitivity to atmospheric warming and changes in river flow. *Water Resources Research* 47(2).
- Vannote RL, Minshall GW, Cummins KW, Sedell JR & Cushing CE (1980) The river continuum concept. *Canadian journal of fisheries and aquatic sciences* 37(1):130-137.
- Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJ, Fromentin J-M, Hoegh-Guldberg O & Bairlein F (2002) Ecological responses to recent climate change. *Nature* 416(6879):389-395.
- Ward J (1985) Thermal characteristics of running waters. *Perspectives in Southern Hemisphere Limnology*, Springer. p 31-46.

- Webb B (1996) Trends in stream and river temperature. *Hydrological processes* 10(2):205-226.
- Webb B & Walling D (1997) Complex summer water temperature behaviour below a UK regulating reservoir. *Regulated rivers: research & management* 13(5):463-477.
- Webb B, Walsh A, Webb B, Acreman M, Maksimovic C, Smithers H & Kirby C (2004) Changing UK river temperatures and their impact on fish populations. *Hydrology: science and practice for the 21st century, Volume II. Proceedings of the British Hydrological Society International Conference, Imperial College, London, July 2004.* British Hydrological Society, p 177-191.
- Webb BW, Clack PD & Walling DE (2003) Water–air temperature relationships in a Devon river system and the role of flow. *Hydrological Processes* 17(15):3069-3084.
- 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.
- Webb BW & Nobilis F (2007) Long-term changes in river temperature and the influence of climatic and hydrological factors. *Hydrological Sciences Journal* 52(1):74-85.
- Wehrly KE, Brenden TO & Wang L (2009) A Comparison of Statistical Approaches for Predicting Stream Temperatures Across Heterogeneous Landscapes. (Wiley Online Library).
- Whitehead P, Wilby R, Battarbee R, Kernan M & Wade AJ (2009) A review of the potential impacts of climate change on surface water quality. *Hydrological Sciences Journal* 54(1):101-123.
- Wilkerson E, Hagan JM, Siegel D & Whitman AA (2006) The effectiveness of different buffer widths for protecting headwater stream temperature in Maine. *Forest Science* 52(3):221-231.
- Wood S (2006) *Generalized additive models: an introduction with R*. CRC press,
- Wood S (2012) *mgcv* 1.7-22. <http://cran.r-project.org/web/packages/mgcv/mgcv.pdf>
- Wood S (2013) *mgcv*: mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation. R package version 1.7-22.).