

ANALYSE RÉGIONALE STATIONNAIRE DES TEMPÉRATURES DE L'EAU EN RIVIÈRES

Par
Olfa Abidi

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

Jury d'évaluation

Président du jury et
Examineur interne

Saeid Homayouni
Professeur, INRS-ETE, Université de
Québec

Examineur externe

Mélanie Trudel
Professeure, Université de Sherbrooke

Directeur de recherche

André St-Hilaire
Professeur, INRS-ETE, Université de
Québec

Codirecteur de recherche

Taha B. M. J Ouarda
Professeur, INRS-ETE Université de
Québec

REMERCIEMENTS

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

J'offre premièrement de sincères et chaleureux remerciements à mon directeur de recherche, Pr. André St. Hilaire. Certes, le mérite d'un mémoire appartient à l'auteur, mais également à son directeur qui l'encadre. Je lui suis reconnaissante de l'accueil qu'il m'a fait au sein de son équipe. Je salue son amabilité, sa patience malgré tous les défis que j'ai eu à relever au début, sa disponibilité, sa souplesse d'esprit et surtout ses judicieux conseils, qui m'ont permis d'approfondir mes connaissances.

Je tiens également à exprimer ma gratitude à mon co-directeur Pr. Taha B. M. J Ouarda qui m'a soutenue, encouragée tout le long de la réalisation de ce projet de recherche. Sa rigueur professionnelle, ses critiques constructives et ses encouragements m'ont poussée à me dépasser moi-même et à donner une meilleure performance. Merci pour les nombreuses discussions qu'on a eues et qui ont contribué à me guider dans mes recherches.

Je souhaite également remercier toute l'équipe qui travaille dans le cadre de ce projet, qui a bien voulu répondre à mes différentes questions dans le but de permettre la facilité de l'élaboration de ce travail. Je dédie un merci particulier à Christian Charron pour m'avoir familiarisée avec les analyses statistiques et à Claudine Boyer et Anik Daigle de m'avoir facilité l'accès aux données.

J'adresse mes vifs remerciements aux membres du jury, Pr. Mélanie Trudel et Pr. Saeid Homayouni pour avoir gentiment accepté d'évaluer mon mémoire de recherche.

Je tiens aussi à adresser mes remerciements à ma famille, et plus précisément à mes parents. Leur confiance en moi et leur soutien constant m'assure des bases solides me permettant de persévérer et de me surpasser.

C'est certes avec joie et fierté que je dépose aujourd'hui ce mémoire, mais aussi avec un brin de nostalgie que je termine ce programme d'études et je conclus ce premier travail de recherche.

Merci beaucoup !

RÉSUMÉ

La température de l'eau est une variable de haute importance affectant la gestion des ressources en eau, contrôlant à la fois la qualité de l'eau ainsi que le métabolisme et la distribution des espèces aquatiques. En particulier, cette variable physique est fondamentale pour les poissons ectothermes tels que les salmonidés (truites, saumons, ombles) dont les processus physiologiques sont directement contrôlés par les températures du milieu ambiant.

L'objectif principal de la présente étude consiste à développer une approche nommée analyse thermique régionale (RTA), similaire à l'analyse fréquentielle régionale (RFA), permettant de prédire des métriques thermiques pertinentes pour la gestion des stocks de saumon atlantique dans l'Est du Canada. Ces métriques sont liées à la température maximale et sa date d'occurrence, ainsi qu'à la variabilité saisonnière. L'approche proposée va permettre d'estimer les métriques sélectionnées, notamment, MaxWaterTmax (moyenne interannuelle de la température maximale de l'eau, MaxNumDay (nombre maximum des jours consécutifs au-dessus du seuil potentiellement stressant pour le saumon atlantique avec $T_{max} > 25^{\circ}\text{C}$ et $T_{min} > 20^{\circ}\text{C}$). En plus de ces deux métriques, les trois paramètres de la fonction gaussienne, y compris Gaussian_a (maximum de la température journalière moyenne interannuelle), Gaussian_b (durée de la période chaude) et Gaussian_c (date d'occurrence du maximum annuel) ont été estimés dans des rivières où il y a peu ou pas de données disponibles. L'estimation régionale a été effectuée à l'aide des modèles statistiques paramétriques (le modèle de régression linéaire multiple MLR et le modèle additif généralisé GAM) et le modèle non paramétrique basé sur l'approche des splines de régression adaptatives multivariées MARS. 22 variables physiographiques et météorologiques ont été testées comme variables explicatives.

Ensuite, les modèles régionaux MLR, GAM et MARS ont été appliqués au sein des régions homogènes identifiées à l'aide de trois approches nommées, analyse de regroupement hiérarchique (HCA), les régions d'influence (ROI) et, finalement, l'approche d'analyse canonique des corrélations (CCA). De plus, ils ont été comparés et évalués au sein d'une seule région englobant toutes les stations pour pouvoir comparer leurs performances.

Au total, douze combinaisons ont été testées puis comparées. Généralement, les meilleures performances, en termes de R^2 , de la racine de l'erreur quadratique moyenne (RMSE) et de biais, sont obtenues avec les modèles utilisant l'analyse de regroupement hiérarchique (HCA) pour le regroupement des sites en régions potentiellement homogènes couplée avec les deux modèles

d'estimation régionale GAM et MARS pour l'estimation des métriques d'intérêt. La délimitation des régions ayant un comportement thermique relativement similaire est fonction de la taille d'échantillon et de la structure du modèle. En effet, pour le modèle simple MLR, trois régions potentiellement homogènes ont été identifiées pour prédire deux métriques thermiques (MaxWaterTmax et MaxNumDay) alors que seulement deux régions ont été sélectionnées pour prédire les trois paramètres de la fonction gaussienne, ajustée sur les moyennes journalières interannuelles des températures. Dans le cas du modèle GAM, seulement trois régions thermiques ont été identifiées pour prédire la métrique MaxNumDay tandis que deux régions ont été sélectionnées pour prédire les autres métriques (MaxWaterTmax, Gaussian_a, Gaussian_b, Gaussian_c). Pour l'approche MARS, bien qu'avec cet outil de modélisation, il est possible d'établir des estimations au sein des groupements de petite taille (par exemple $n < 10$), seulement deux régions homogènes ont été sélectionnées pour estimer les cinq métriques de températures de l'eau vu qu'elles ont abouti aux meilleures performances.

Mots-clés: Température de l'eau, métriques thermiques, analyse thermique régionale, modèle de régression linéaire multiple MLR, modèle additif généralisé GAM, approche des splines de régression adaptatives multivariées MARS, régions homogènes, analyse de regroupement hiérarchique (HCA).

ABSTRACT

Water temperature is a variable of great importance affecting water resources management, controlling both water quality as well as the metabolism and distribution of aquatic species. In particular, this physical variable is fundamental for ectothermic fish such as salmonids (trout, salmon, char) whose physiological processes are directly controlled by the ambient temperatures.

The main target of this study is to develop an approach called regional thermal analysis (RTA), similar to regional frequency analysis (RFA), to predict thermal metrics relevant for the management of Atlantic salmon stocks in Eastern Canada. The latter are related to maximum temperature and its date of occurrence, as well as the seasonal variability. The proposed approach will make it possible to estimate the selected metrics, counting MaxWaterTmax (Interannual mean of maximum summer temperature), MaxNumDay (Interannual mean of the number of consecutive days with maximum water temperature > 25 °C and minimum water temperature > 20 °C). In addition to these two indices, three other parameters of a Gaussian function including Gaussian_a (Maximum of interannual daily mean temperature), Gaussian_b (Mean duration of the warm period), as well as Gaussian_c (Date of occurrence of the interannual maximum temperature) were estimated in rivers where there is little or no data available. The regional estimation step was conducted using parametric statistical models (multiple linear regression model MLR and generalized additive model GAM) as well as a nonparametric model based on the multivariate adaptive regression splines approach MARS. 22 physiographic and meteorological variables were tested as explanatory variables.

Then, the regional models MLR, GAM and MARS were applied within homogeneous regions identified using three approaches namely, hierarchical clustering analysis (HCA), regions of influence (ROI) as well as canonical correlations analysis (CCA). In addition, they were compared and evaluated within a single region encompassing all stations to evaluate models' performance.

Overall, twelve combinations were tested and then compared. Generally, the best performances, in terms of R^2 , Root Mean Square Error RMSE and bias, are obtained with the models using the hierarchical clustering analysis (HCA) for grouping sites into potentially homogeneous regions coupled with the two regional estimation GAM and MARS models for estimating the metrics of interest. The delineation of regions having relatively similar thermal behavior is a function of the sample size as well as the model structure. Indeed, for the simple MLR model, three homogeneous regions were identified to predict two thermal variables (MaxWaterTmax and

MaxNumDay) whereas only two regions were selected to predict the three parameters of the gaussian function. In the case of the GAM model, only three thermal regions were identified to predict the MaxNumDay metric whereas two regions were selected to predict the other metrics (MaxWaterTmax, Gaussian_a, Gaussian_b, Gaussian_c). For the MARS approach, despite the fact that it is possible to derive estimates within small size clusters (for example $n < 10$), only two homogeneous regions were selected to estimate the five water temperature metrics as they resulted in the best model's performance.

Keywords: Water temperature, thermal metrics, regional thermal analysis, multiple linear regression model MLR, generalized additive model GAM, multivariate adaptive regression splines MARS, homogeneous regions, hierarchical clustering analysis (HCA).

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES FIGURES.....	XI
LISTE DES TABLEAUX.....	XIII
LISTE DES ABRÉVIATIONS.....	XV
1 INTRODUCTION.....	1
1.1 MISE EN CONTEXTE.....	1
1.1.1 <i>Les facteurs environnementaux affectant la température de l'eau.....</i>	<i>3</i>
1.1.2 <i>Les facteurs anthropiques affectant la température de l'eau</i>	<i>4</i>
1.1.3 <i>Problématique et Objectifs</i>	<i>5</i>
1.2 REVUE DE LA LITTÉRATURE.....	6
1.2.1 <i>Modèles déterministes</i>	<i>7</i>
1.2.2 <i>Modèles statistiques</i>	<i>7</i>
1.2.3 <i>Détermination des régions homogènes.....</i>	<i>9</i>
1.2.4 <i>Estimation régionale</i>	<i>10</i>
1.3 STRUCTURE DU MEMOIRE.....	11
2 REGIONAL THERMAL ANALYSIS APPROACH: A MANAGEMENT TOOL FOR PREDICTING WATER TEMPERATURE METRICS RELEVANT FOR THERMAL FISH HABITAT	13
HIGHLIGHTS.....	14
ABSTRACT	15
1 INTRODUCTION.....	17
2 METHODOLOGY.....	22
2.1 DELINEATION OF THERMALLY HOMOGENEOUS REGIONS (DHR)	22
2.1.1 <i>Hierarchical clustering analysis (HCA)</i>	<i>22</i>

2.1.2	<i>Region of influence (ROI)</i>	23
2.1.3	<i>Canonical correlations analysis (CCA)</i>	23
2.2	REGIONAL ESTIMATION (RE).....	25
2.2.1	<i>Multiple Linear Regression (MLR)</i>	25
2.2.2	<i>Generalized Additive Model (GAM)</i>	25
2.3	STEPWISE REGRESSION	26
2.4	VALIDATION.....	26
3	CASE STUDY	28
3.1	THERMAL WATER TEMPERATURE METRICS.....	29
3.2	CLIMATIC AND PHYSIOGRAPHIC VARIABLES	31
4	RESULTS	33
4.1	SELECTION OF EXPLANATORY VARIABLES FOR MLR.....	34
4.2	SELECTION OF EXPLANATORY VARIABLES FOR GAM	35
4.3	DELINEATION OF THERMALLY HOMOGENEOUS REGIONS BY HCA, CCA AND ROI.....	36
4.4	COMPARATIVE PERFORMANCE ANALYSIS	39
5	DISCUSSION	42
	CONCLUSION	44
	CREDIT AUTHOR STATEMENT	47
	ACKNOWLEDGMENTS	47
	DATA STATEMENT	47
3	ANALYSE SUPPLÉMENTAIRE: ÉTUDE COMPARATIVE ENTRE MARS ET GAM	48
3.1	INTRODUCTION	48
3.2	MÉTHODOLOGIE	49
3.3	RÉSULTATS	50
3.3.1	<i>Sélection des variables explicatives pour le modèle MARS</i>	50
3.3.2	<i>Validation</i>	51
	53	
4	DISCUSSION GÉNÉRALE ET CONCLUSION	61
5	BIBLIOGRAPHIE	65

LISTE DES FIGURES

FIGURE 1-1. FACTEURS DE CONTROLE DE LA TEMPERATURE DE L'EAU (SCHEMA MODIFIE ISSUE DE CAISSIE (2006)).	3
FIGURE 2-1: LOCATION OF THERMAL STATIONS ACROSS THE STUDY AREA.	29
FIGURE 2-2. SPLINES FOR THE PREDICTORS USED IN THE ALL-GAM MODEL (E.G., MAXWATER _{TMAX}).	36
FIGURE 2-3. DENDROGRAM WITH CUT-OFF THRESHOLD FOR THE VARIABLE MAXWATER _{TMAX} . NOTE THAT STATION NUMBERS ARE SHOWN ON THE X-AXIS.	37
FIGURE 2-4. MAPS OF STATIONS GROUPED USING HCA-MLR FOR EACH THERMAL METRIC.	38
FIGURE 2-5: SCATTER PLOTS OF OBSERVED VS SIMULATED WATER TEMPERATURE METRICS FOR MLR.	41
FIGURE 2-6: HISTOGRAMS SHOWING THE FREQUENCY DISTRIBUTION OF THE VARIABLES MAXNUMDAY AND THE PARAMETER C OF THE GAUSSIAN FUNCTION.	42
FIGURE 3-1. PROCEDURE DE SELECTION DES VARIABLES EXPLICATIVES POUR LES DEUX MODELE MARS (A) ET GAM (B).	50
FIGURE 3-2. EXEMPLES D'AJUSTEMENTS DE CERTAINES VARIABLES EXPLICATIVES DANS LES DEUX CAS : (A) CAS DU MODELE MARS, (B) CAS DU MODELE GAM.	53
FIGURE 3-3. DIAGRAMMES DES VALEURS ESTIMÉES EN FONCTION DES VALEURS OBSERVÉES POUR CHAQUE MÉTRIQUE THERMIQUE DANS LE CAS DU MODÈLE MARS.	54
FIGURE 3-4. DIAGRAMMES COMBINES MONTRANT LES VALEURS ESTIMEES ET LES VALEURS OBSERVES POUR LES VARIABLES MAXNUMDAY (A, C) ET LE PARAMETRE GAUSSIEN C (B, D).	55
FIGURE 3-5. REPRESENTATION SCHEMATIQUE DE LA VARIABILITE MOYENNE QUOTIDIENNE ET JOURNALIERE DE LA TEMPERATURE DE L'EAU EN FONCTION DE L'ECHELLE SPATIALE, C'EST-A-DIRE DE LA DIRECTION EN AVAL, QUI SE TRADUIT AUSSI GENERALEMENT PAR L'ORDRE DES COURS D'EAU (MODIFIE D'APRES CAISSIE, 2006 ; AUSTIN ET AL., 2015).	56
FIGURE 3-6. LOCALISATION GEOGRAPHIQUE DE LA RIVIERE SHELBURNE EN NOUVELLE ÉCOSSE (TIREE SHELBURNE CANADIAN HERITAGE RIVER MONITORING REPORT 2008-2018).	57

FIGURE 3-7. DIAGRAMME DE TAYLOR ILLUSTRANT L'EVALUATION CONJOINTE DE L'ECART TYPE (SD),
DE LA DIFFERENCE QUADRATIQUE MOYENNE CENTREE ET DU COEFFICIENT DE
CORRELATION POUR LES CINQ METRIQUES PREDITES PAR LES MEILLEURS MODELES.
.....58

FIGURE 3-8. GRAPHIQUES COMBINES MONTRANT L'ERREUR RELATIVE EN FONCTION DE CERTAINES
VARIABLES SIGNIFICATIVES JUGEES IMPORTANTES.....59

LISTE DES TABLEAUX

TABLEAU 1-1. SEUILS CRITIQUES DE TEMPERATURE (°C) POUR LA SURVIE DU SAUMON ATLANTIQUE (ELLIOT ET ELLIOT, 2010).	2
TABLE 2-1. SELECTED THERMAL METRICS ACROSS ALL 146 STATIONS AND SOME OF THEIR STATISTICAL CHARACTERISTICS.	30
TABLE 2-2: PHYSIOGRAPHIC AND CLIMATIC VARIABLES AND SOME OF THEIR STATISTICAL CHARACTERISTICS.	32
TABLE 2-3: LAND COVER AND SURFACE DEPOSITS VARIABLES AND SOME OF THEIR STATISTICAL CHARACTERISTICS.	33
TABLE 2-4: SELECTED PREDICTORS WHEN ALL STATIONS GROUPED TOGETHER.	34
TABLE 2-5: PERFORMANCE INDICATORS FOR DIFFERENT REGIONAL THERMAL ANALYSIS IN ATLANTIC SALMON RIVERS.	40
TABLEAU 3-1. VARIABLES EXPLICATIVES SELECTIONNEES POUR CHAQUE METRIQUE POUR LES DEUX MODELES GAM ET MARS.	51
TABLEAU 3-2. RESULTATS DE LA VALIDATION CROISEE POUR LES DIFFERENTES COMBINAISONS.	53

LISTE DES ABRÉVIATIONS

AIC	Akaike information criterion
AirTmax	Maximum air temperature
AirTmin	Minimum air temperature
ANN	Artificial neural networks
CCA	Canonical correlations analysis
DHR	Delineation of homogeneous regions
GAM	Generalized additive model
Gaussian_a	Parameter a of the Gaussian function
Gaussian_b	Parameter b of the Gaussian function
Gaussian_c	Parameter c of the Gaussian function
GCV	Generalized cross validation (GCV)
HCA	Hierarchical clustering analysis
MARS	Multivariate adaptive regression splines
MaxElevation	Maximum elevation of the catchment
MaxNumDay	Interannual mean of the number of consecutive days
MaxWaterTmax	Interannual mean of maximum temperature
MeanElevation	Mean elevation of the catchment
MinElevation	Minimum elevation of the catchment
MLR	Multiple linear regression
PBIAS	Percent bias
RE	Regional estimation
RF	Random Forest
RFA	Regional frequency analysis
RMSE	Root mean square error

ROI	Region of influence
RRMSE	Relative root mean square error
RSS	Residual sum of squares
RTA	Regional thermal analysis
TotPrecip	Total precipitation

1 INTRODUCTION

1.1 Mise en contexte

Un des effets prévisibles du changement climatique est une modification du régime thermique des milieux aquatiques. Ce dernier est soumis à des variations naturelles, qui rendent sa mesure difficile et à des influences anthropiques qui rendent sa modélisation délicate. Il est issu d'un mélange de plusieurs types de contributions hydrologiques en fonction des caractéristiques des bassins. La température de l'eau constitue une variable critique de haute importance affectant les processus écologiques des cours d'eau et des rivières, elle gère les processus physiques, chimiques et biologiques se déroulant en milieu aquatique (Ward, 1985; Caissie, 2006; Webb et al., 2008).

L'incidence de la température de l'eau sur la qualité des écosystèmes aquatiques n'est plus à démontrer (Fry, 1971; Ward, 1982). Ces dernières années, l'intérêt scientifique pour le thème de la température de l'eau s'est nettement accru. En effet, l'étude de la thermie des rivières est importante pour parvenir à concilier les enjeux économiques et environnementaux, en particulier, en contexte du changement climatique. Plusieurs études ont été menées pour mettre en évidence l'importance de la température de l'eau. Dans ce sens, une étude réalisée par Wehrly et al. (2003) a montré que les différences observées dans la structure des communautés des poissons des cours d'eau sont largement attribuables à la variation spatiale de la température moyenne.

En particulier, cette variable physique est fondamentale pour les poissons ectothermes dont les processus physiologiques sont directement contrôlés par les températures du milieu ambiant comme les salmonidés (Elliot et al., 1998; Lund et al., 2002; Mather et al., 2008). Ces organismes aquatiques sont incapables de contrôler eux-mêmes leur température corporelle et sont, donc, intolérants aux températures élevées (ex., Elliott, 1991; Jonsson et Jonsson, 2009). Conséquemment, l'exposition prolongée à des températures au-dessus du seuil critique supérieur entraîne la mort des individus (Elliott, 1991; Elliot et Elliot, 2010).

Le saumon atlantique (*Salmo salar*), espèce d'eau froide, est un exemple typique de la famille des salmonidés. Cette espèce a été largement étudiée ces dernières années dans le but de mieux comprendre les conditions de sa croissance et surtout sa distribution en rivières. Plusieurs études menées dans ce contexte, ont montré que le régime thermique des rivières a des impacts significatifs sur la croissance ainsi que la répartition spatiale des saumons atlantiques juvéniles (Nicieza et Metcalfe, 1997; Elliot et Elliot, 2010; Sundt Hansen et al., 2018). D'autres travaux se sont concentrés sur l'étude des seuils de température pour lesquels les espèces comme le saumon atlantique subissent des stress thermiques considérables au cours des mois chauds de l'été (Lund et al., 2002; Hodgson et Quinn, 2002). Dernièrement, une étude a été menée

par Heggens et al. (2021) dans laquelle les auteurs ont montré qu'une exposition régulière à des températures élevées au-dessus du seuil 22.5°C provoque des stress thermiques. De plus, des études ont montré que dans de nombreuses régions d'Amérique du Nord, les poissons connaissent déjà des épisodes chauds avec des températures approchant leur limite létale supérieure (Sinokrot et al., 1995; Eaton et al., 1995). Dans l'Est du Canada, les changements climatiques pourraient entraîner une perte globale de l'habitat du saumon atlantique juvénile (Minns et al., 1995).

Comme mentionné précédemment, les organismes aquatiques ont des limites physiologiques en ce qui a trait aux conditions de la température. Le dépassement de seuils létaux peut également mener à des événements de mortalité massive chez ces individus (Ouellet et al., 2010). Chez les salmonidés, la réponse à ce stress ne dépend uniquement pas de la température de l'eau, mais aussi de la rapidité avec laquelle la température augmente (Quigley et Hinch, 2006).

Le tableau 1-1 résume les seuils critiques de température de l'eau pour la survie du saumon Atlantique.

<i>Stade de vie</i>	<i>Inférieure</i>	<i>Supérieure</i>
Œufs	0	16
Alevins à long terme	0-2	23-24
Alevins à court terme	0-1	24-25
Létale initiale	0-2	22-28
Létale ultime	-0.8	30-33
Limite supérieure d'alimentation	0-7	22-28

Tableau 1-1. Seuils critiques de température (°C) pour la survie du saumon Atlantique (Elliot et Elliot, 2010).

Pour les saumons atlantiques juvéniles, la température optimale pour la croissance se situe entre 16 et 20 °C (Elliott et Elliott 2010). La température létale initiale, définie comme la température que le poisson peut tolérer pendant longtemps (habituellement sept jours), est de 27,8 °C (Elliott 1991). La température létale ultime pour les saumons juvéniles, définie comme la température que le poisson ne peut tolérer que pour de courtes périodes (habituellement 10 minutes), varie entre 30 et 33 °C. L'exposition à des températures dépassant les limites de tolérances provoquent un stress physiologique voire une mortalité. Même en deçà du seuil de tolérance, l'exposition à des températures élevées peut entraîner des conséquences sur leur croissance, reproduction et la résistance aux maladies (Crozier et al., 2008; Martins et al., 2012).

La température des rivières est très vulnérable aux nombreux facteurs environnementaux et aux impacts des activités humaines.

1.1.1 Les facteurs environnementaux affectant la température de l'eau

Les conditions atmosphériques (par exemple, la température de l'air, le rayonnement solaire) représentent la catégorie de forçage ayant le plus d'influence sur le régime thermique des cours d'eau et des rivières. En effet, plusieurs études ont montré qu'il existe une relation linéaire entre la température de l'eau et la température de l'air (Mohseni et Stefan, 1999; Kinouchi et al., 2007). D'autres facteurs d'ordre topographiques et morphologiques (ex., profondeur de la rivière, pente, degré de turbulence, dimensions de surfaces libres, géologie, substrat, végétation riveraine et ombrage) sont également importants vu qu'ils influencent les conditions régissant en partie le flux de chaleur à la surface, mais aussi avec le lit et les berges du cours d'eau.

Selon Caissie (2006), les facteurs de contrôle de la température sont classés en quatre catégories, à savoir les conditions atmosphériques, l'hydrologie, les échanges à l'interface eau/sédiment et la géomorphologie (Figure 1-1).

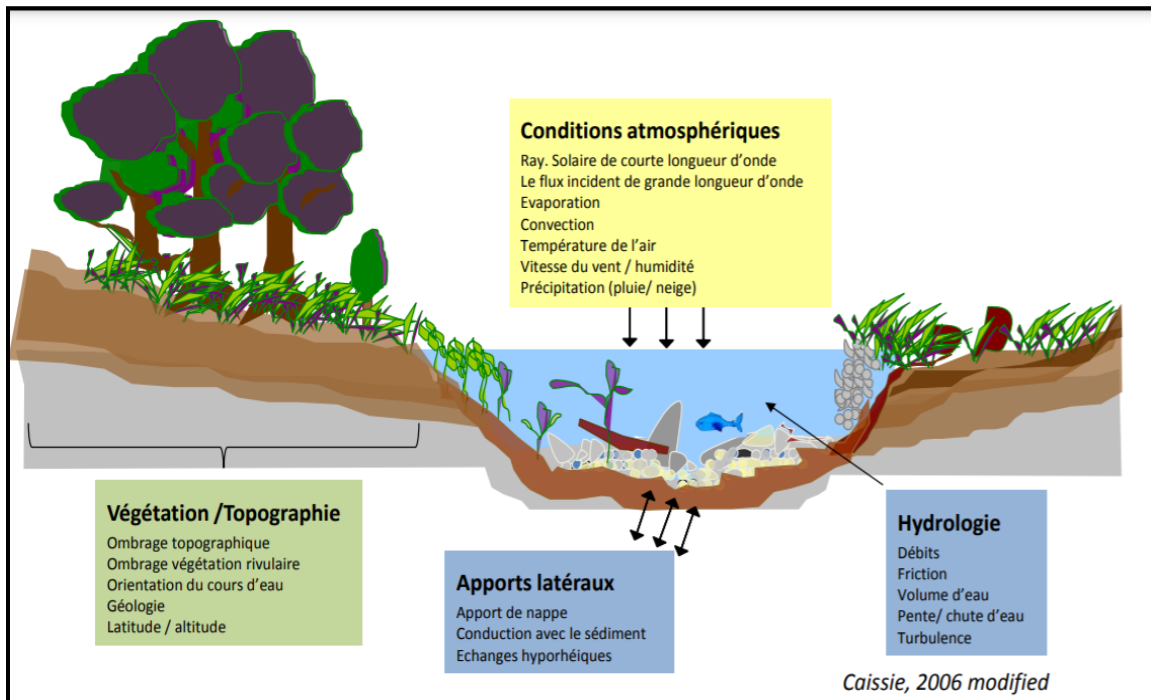


Figure 1-1. Facteurs de contrôle de la température de l'eau (Schéma modifié issue de Caissie (2006)).

1.1.2 Les facteurs anthropiques affectant la température de l'eau

La thermie peut être aussi régie par plusieurs activités anthropiques. Comme par exemple, la déforestation (Beschta et al., 1987; Chen et al., 1998a, b; Moore et al., 2005), la présence des barrages (Maheu et al., 2016b), et le réchauffement climatique (Isaak et Rieman, 2013; Van Vliet et al., 2013; Isaak et al., 2017). D'autres facteurs, y compris, la régulation du débit (Webb et Walling, 1993), les effluents thermiques, l'urbanisation et l'agriculture peuvent avoir un impact sur le régime thermique.

1.1.2.1 Déforestation

La déforestation a été retrouvée comme source importante de perturbation du régime thermique des rivières. À cet égard, de nombreux travaux se sont concentrés sur l'étude de l'impact de la coupe forestière sur la thermie des cours d'eau (ex., Hannah et al., 2008; St-Hilaire et al., 2015). En effet, l'élimination de la bande riveraine de végétation par des activités agricoles ou forestières peut entraîner un réchauffement des cours d'eau (Barton et al., 1985 ; Rutherford et al., 2004; Moore et al., 2005). Ceci entraîne une amplification de l'effet du rayonnement solaire incident et par conséquent une hausse de la température de l'eau et un changement de l'habitat aquatique (Beschta et al., 1987, St-Hilaire et al., 2000).

1.1.2.2 Barrages

Les barrages peuvent également avoir un impact significatif sur différents attributs du régime thermique (Olden et Naiman, 2010). L'impact de ces réservoirs hydrauliques sur la température de l'eau dépend de plusieurs paramètres, à savoir le type d'ouvrage, le mode de gestion, la position de la prise d'eau et du volume du réservoir (Bartholow et al., 2004). D'autres paramètres peuvent aussi influencer la thermie (ex., stratification de la colonne d'eau dans les réservoirs (Webb et Walling, 1997)). Ainsi, un effet de lissage des cycles journaliers et/ou annuels en aval a été mis en évidence (Liu et al., 2005). Cet effet dépend entre autres de la profondeur de la prise d'eau, du mode de fonctionnement, du type d'ouvrage, du positionnement dans le bassin versant et de la taille du réservoir (Webb, 1996, Webb et al., 1997, Bartholow et al., 2004).

D'une manière générale, on associe les barrages de grande taille à une pollution thermique froide et les barrages de petite taille à une pollution thermique chaude (Olden et Naiman, 2010). Dans ce contexte, une étude effectuée par Singer et al. (2011) a montré que l'installation d'un petit barrage sur une rivière en Alabama a occasionné une augmentation de la température moyenne mensuelle. Ceci a été confirmée à l'aide d'une étude effectuée par Maheu et al. (2015) dans l'Est du Canada. Aussi, les barrages et retenues peuvent modifier les communautés aquatiques des rivières régulées (Bunn et al., 2002, Poff et al., 2010).

1.1.2.3

Effets du changement climatique sur la température de l'eau

D'après le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), le climat change tant au Canada qu'ailleurs dans le monde. Ces changements, bien qu'attribuables en grande partie à des causes humaines, peuvent avoir des répercussions sur la thermie des rivières (Webb, 1996, Mohseni et al., 1999, Poirel et al., 2010, Van Vliet et al., 2011). Dans la majorité des cas, les différents scénarios d'un tel changement prévoient des conséquences sur la diversité ainsi que le risque d'extinction d'espèces. Dans le même contexte, Mohseni et al. (2003) ont montré que le réchauffement climatique aura une incidence directe sur les communautés piscicoles conduisant à une baisse de 36% au niveau de la quantité d'habitats thermiques favorables pour les poissons d'eau froide, au profit des poissons d'eau chaude qui augmenteront de 31% alors qu'une baisse de 15% pour les poissons d'eau tiède est prévue.

Plusieurs études révèlent l'augmentation de la température des eaux des rivières (ex., Ducharme, 2008; Arismendi et al., 2012; Ficklin et al., 2013). Le GIEC (2013) a révélé une augmentation des températures de l'air ainsi que des océans. Selon divers scénarios, d'ici la fin du 21^{ème} siècle, la température de l'air augmentera dans un intervalle de 1.1°C et 6.4°C (Meehl et al., 2007). Conséquemment, une augmentation de la température de l'air conduit habituellement à une augmentation de la température de l'eau selon un patron assez similaire. Également, une étude conduite par Van Vliet et al. (2011) prévoit une augmentation moyenne annuelle globale de la température de l'eau de 2.6°C pour une hausse de 4°C de la température de l'air. Daufresne et al. (2009) ont montré l'existence d'un lien direct entre l'augmentation de la température et la réduction de taille des organismes.

À la lumière de toutes ces preuves, la température de l'eau demeure une variable clé de haute importance pour des fins de gestion efficace des ressources en eau.

1.1.3 Problématique et Objectifs

Il est bien évident que les mesures directes de la température de l'eau sur les rivières à saumon de l'Atlantique sont utiles, voire nécessaires. Cependant, ces données demeurent limitées et sporadiques sur de nombreux sites et, donc, la connaissance et la compréhension du régime thermique demeurent difficiles. En conséquence, la modélisation a été introduite pour pallier entre autres aux lacunes décrites ci-haut et pour permettre aussi une meilleure compréhension de la dynamique thermique des rivières. À cet égard, plusieurs travaux antérieurs se sont concentrés sur le développement de plusieurs approches de modélisation à l'échelle locale et régionale.

De nombreux outils de modélisation ont été développés afin de pallier au manque relatif des données thermiques sur de nombreuses rivières à saumon de l'Atlantique au Canada et ailleurs dans le monde. En

outre, le présent projet de recherche a pour objectif de développer une approche de modélisation à l'échelle régionale pour estimer cinq métriques thermiques de l'eau en rivières. Ces métriques sont pertinentes pour la gestion des stocks de saumon atlantique dans l'Est du Canada. Elles sont liées à la température maximale et sa date d'occurrence, ainsi qu'à la variabilité saisonnière. L'approche proposée va permettre d'estimer les métriques sélectionnées dans des rivières où il y a peu ou pas de données disponibles à l'aide des modèles statistiques paramétriques linéaire (MLR) et non linéaire (GAM) ainsi qu'un modèle non paramétrique (MARS). Un certain nombre de variables physiographiques et météorologiques seront testées comme variables explicatives.

Pour ce faire, une approche nommée analyse thermique régionale (RTA), similaire à l'analyse fréquentielle régionale (RFA) couramment utilisée en hydrologie pour estimer les quantiles de crue ou de faible débit est développée.

En hydrologie, et pour estimer les débits extrêmes sur des sites non jaugés, l'approche RFA comporte deux étapes principales :

- Définition des groupes de stations qui ont un comportement hydrologique relativement similaire.
- Estimation des quantiles de débits d'intérêt à travers le transfert des informations hydrologiques des sites jaugés vers les sites cibles non jaugés.

Dans la présente étude, ces deux principales étapes sont adaptées à la problématique décrite ci-haut. Les regroupements de stations sont basés sur une similitude hydro-thermique et les quantiles de débit sont remplacés par les métriques de température de l'eau (RTA).

Les objectifs spécifiques pour la présente étude consistent à:

- Modéliser les métriques de température de l'eau à travers l'approche d'analyse thermique régionale;
- Évaluer et comparer la performance des modèles statistiques testés et ensuite sélectionner le modèle le plus performant;
- Identifier les variables explicatives qui ont le pouvoir de prédiction le plus important pour estimer les métriques d'intérêt.

1.2 Revue de la littérature

En raison de l'inquiétude croissante concernant les effets probables des activités anthropiques et des changements climatiques sur la vie aquatique (Caissie, 2006; Benyahya et al., 2007; Dugdale et al., 2017), la modélisation spatiale et temporelle de la température de l'eau a été introduite permettant une meilleure compréhension de la dynamique thermique des rivières. Cette technique permet de quantifier les effets des

conditions hydrométéorologiques (température de l'air, débit, ...) sur la température des cours d'eau et d'évaluer l'impact de cette dernière sur la quantité et la qualité des ressources aquatiques. Deux grandes catégories de modèles ont été mises en œuvre à savoir les modèles déterministes et les modèles statistiques.

1.2.1 Modèles déterministes

Cette catégorie des modèles est fondée sur des équations du bilan énergétique permettant de quantifier les flux d'énergie totaux décrivant les échanges de chaleur à l'interface entre l'eau et l'air, ainsi qu'à l'interface avec le lit de la rivière. Elle vise à calculer l'échange thermique entre l'atmosphère et les cours d'eau à travers les variables d'entrées comme le rayonnement solaire, la vitesse du vent, l'évaporation et la conduction (Chen et al., 1998a; Foreman et al., 2001; St-Hilaire et al., 2003; Caissie et al., 2007). Bien que ces modèles à base physique puissent fournir des estimations précises de la température des cours d'eau (Sinokrot et Stefan, 1993; Kim et Chapra, 1997; Gu et al., 1998; Younus et al., 2000; Horne et al., 2004; Caissie et al., 2007), ces derniers ne sont pas toujours faciles à mettre en œuvre à l'échelle régionale, surtout là où les données sont rares (Mohseni et al., 1998; Caissie et al., 2001; Risley et al., 2003; Iskra et Droste, 2008). Conséquemment, l'utilisation de ce type de modèles est donc limitée par la disponibilité des données. Dans de nombreuses situations, seuls quelques paramètres météorologiques sont disponibles. Il est donc particulièrement intéressant de développer des modèles à partir de méthodes statistiques pour établir des relations empiriques entre les données disponibles et la température de l'eau.

1.2.2 Modèles statistiques

Contrastant avec les modèles déterministes, cette catégorie de modèles est basée plutôt sur la corrélation entre la température des cours d'eau et les prédicteurs comme la température de l'air et les caractéristiques environnementales (Caissie, 2006; Webb et al., 2008). Plusieurs méthodes statistiques de modélisation de la température de l'eau ont été présentées dans la littérature (par exemple, Mohseni et al., 1998; Caissie et al., 2001; Ahmadi-Nedushan et al., 2007; Benyahya et al., 2007). Cette catégorie comprend deux sous-ensembles de modèles :

1.2.2.1 Modèles paramétriques

Les modèles statistiques paramétriques sont basés sur une relation statistique spécifiée. Ils sont fréquemment utilisés pour prédire la température de l'eau en fonction d'une ou de plusieurs variables indépendantes (Mohseni et al., 1998, Neumann et al., 2003). Un grand nombre de ces modèles empiriques sont basés sur des approches paramétriques comme les régressions linéaires et non linéaires, les modèles autorégressifs et les modèles périodiques autorégressifs (ex., Stefan et Preud'homme, 1993; Mohseni et al.,

1998; Caissie et al., 1998; Ahmadi-Nedushan et al., 2007; Benyahya et al., 2007). Cette catégorie de modèles est développée dans le but de prédire la température des cours d'eau dans les bassins versants non jaugés. La plupart de ces approches n'utilisent qu'une ou deux variables explicatives (plus souvent, la température de l'air et le débit). Pratiquement, ces modèles sont moins adaptés à des échelles de temps journalières, vu l'autocorrélation forte de la température de l'eau (Caissie, 2006, Laanaya et al., 2017).

Des études comparatives ont été également proposées pour différencier l'impact de l'approche de modélisation. En effet, Wehrly et al. (2009) ont montré, à travers l'étude de quatre approches de modélisation statistique (Régression linéaire multiple par moindres carrés ordinaires, modèle additif généralisé (GAM), Krigeage ordinaire, modèle linéaire mixte) que la différence de précision des estimations entre ces modèles était relativement faible. La même conclusion a été tirée par Daigle et al. (2010) avec les approches de régression linéaire multiple, régression ridge, régression robuste et régression pas à pas. Ceci amène à une flexibilité substantielle dans le choix des approches de modélisation. En conséquence, le développement des estimations plus précises de la variable réponse dépend fortement du choix des caractéristiques hydrologiques et physiographiques (Risley et al., 2003; Johnson, 2004). D'autres comparaisons entre les différents modèles reportés dans la littérature sont malheureusement limitées par la diversité des métriques de température de l'eau.

1.2.2.2 Modèles non paramétriques

Les modèles statistiques non paramétriques ont été utilisés pour prédire la température de l'eau, ils ont l'avantage d'être souples et de ne pas imposer de forme a priori (Benyahya et al., 2007). Parmi ces modèles, on peut mentionner la méthode des K plus proches voisins (St-Hilaire et al., 2000, Benyahya et al., 2007), et les réseaux de neurones artificiels (RNA) (Bélanger et al., 2005, Karacor et al., 2007, Jeong et al., 2013, Hadzima-Nyarko et al., 2014, Piotrowski et al., 2015). À titre d'exemple, dans l'étude de Bélanger et al. (2005), les auteurs ont montré que les réseaux de neurones semblent donner un ajustement aux données légèrement meilleur que celui offert par la régression linéaire multiple. En effet, pour la validation, les coefficients de détermination sont de 92,2 % pour les réseaux de neurones et de 91,6 % pour la régression linéaire multiple, ce qui se traduit en un écart-type des erreurs de 1,10 °C pour les réseaux de neurones et de 1,25 °C pour la régression linéaire multiple. Toutefois, ces deux approches de modélisation démontrent une bonne performance pour la prédiction de la température de l'eau en rivière. En suivant la même idée, Piotrowski et al. (2015) ont proposé et comparé un certain nombre de méthodes (ex., perceptron multicouches, systèmes d'inférence floue basés sur un réseau adaptatif et réseaux de neurones à ondelettes) dans le but de prédire la température des cours d'eau en fonction de diverses variables météorologiques et hydrologiques. D'une part, les résultats montrent que les réseaux de neurones perceptrons multicouches

simples ne sont, dans la plupart des cas, pas surpassés par des modèles plus complexes et avancés. D'autre part, les meilleurs résultats sont obtenus lorsque les températures quotidiennes moyennes, maximales et minimales de l'air des jours précédents sont utilisées comme entrées.

De nombreux outils de modélisation ont été développés, l'effort de modélisation a principalement porté sur des modèles locaux (ex., Caissie et al., 1998; Jeong et al., 2013; Boudreault et al., 2019). Cette catégorie nécessite un historique de mesures au site d'intérêt qui soit suffisamment long pour caler les modèles. En réponse au manque relatif des données thermiques sur de nombreux sites de mesure, certaines études ont récemment tenté de développer des modèles régionaux. Les deux travaux les plus complets sur la régionalisation de la relation linéaire existant entre la température des cours d'eau et la température de l'air ont été récemment menés par Ducharme (2008) et Segura et al. (2014). Dans le même contexte, Hrachowitz et al. (2010), Rivers-Moore et al. (2012) et Imholt et al. (2013) ont exprimé la température de l'eau comme une combinaison linéaire de variables climatiques et physiographiques pour chaque mois de l'année séparément. En effet, l'étude de Rivers-Moore et al. (2012) a été basée sur la corrélation entre les mesures de la température de l'eau et 16 variables environnementales à l'aide des modèles de régression multiples. Les auteurs ont montré que la corrélation avec des variables explicatives fournit la « Pierre de Rosette » pour représenter spatialement les mesures de la température de l'eau. Dans la même veine, l'analyse de régression dans l'étude d'Imholt et al. (2013) a montré que l'élévation et l'aire du bassin versant peuvent être des prédicteurs importants pour prédire les mesures de la température de l'eau. Ainsi, le couvert forestier était une variable explicative significative pendant les mois d'été.

Les approches non paramétriques ont été moins utilisées pour modéliser les caractéristiques du régime thermique des rivières.

Tel que mentionné précédemment, les approches menées dans le cadre RFA comportent deux étapes principales à savoir:

1.2.3 Détermination des régions homogènes

Le concept de base en analyse fréquentielle régionale (RFA) est celui des régions homogènes qui consiste à regrouper des sites ayant un comportement hydrologique similaire au site cible non jaugé par l'intermédiaire de diverses méthodes statistiques (ex., Burn, 1990; Cavadias et al., 2001; Ouarda et al., 2001). Cette délimitation est basée essentiellement sur les caractéristiques physiographiques et hydrologiques des bassins versants. Les régions homogènes peuvent être géographiquement contiguës, géographiquement non contiguës ou de type voisinage (Ouarda et al., 2001).

En RFA, différentes méthodes ont été introduites pour établir ces groupements homogènes. Une approche simple pourrait être de définir des groupes de stations géographiquement contiguës avec des régimes hydrologiques similaires. Une alternative à cette approche est basée sur des similitudes autres que l'emplacement et permettant de définir des régions à partir des caractéristiques climatiques et physiographiques des bassins versants. On a recours à cette approche dans la présente étude et des groupes de rivières potentiellement non contigus ont été construits pour appliquer l'approche d'analyse thermique régionale (RTA), en prenant en considération des variables physiographiques et climatiques susceptibles d'expliquer la variabilité thermique inter-rivières.

L'utilisation des régions non contiguës a été recommandée dans la littérature (par exemple, Ouarda et al., (2008); Haddad et Rahman, 2012). L'analyse de regroupement hiérarchique (HCA), basée sur les caractéristiques du site est l'une des méthodes les plus pratiques utilisées pour définir les régions non contiguës (Hosking et Wallis, 1997). Comme alternative à la définition des régions contiguës ou non contiguës, l'approche de voisinage comme la méthode des régions d'influence ROI (Burn, 1990) ou celle de l'analyse canonique des corrélations CCA (Ouarda et al., 2000; Cavadias et al., 2001) ont, également, été testées en RFA dans le passé (par exemple, Ouarda et al., 2018; Msilini et al., 2020). Un voisinage est défini comme une région centrée autour d'un site cible qui inclut les sites jaugés les plus près, au sens mathématique, de ce point de référence. Dans ce sens, des études comparatives ont montré que les voisinages conduiraient à des meilleures estimations régionales de quantiles de crue que les régions fixes (Burn, 1990; Tasker et al., 1996; Ouarda et al., 2008). Cette approche n'a pas été testée en RTA avant l'étude de Charron et al. (2019). La principale différence méthodologique entre la présente étude et celle de Charron et al. (2019) est la base de données de la température de l'eau. En effet, dans l'étude de Charron et al. (2019), le principal critère de sélection des stations thermiques pour l'analyse régionale était la durée de la série chronologique (cinq ans ou plus). Dans la présente étude, ce critère a été assoupli à quatre étés (du 1er juin au 30 septembre comme durée minimale). Le deuxième critère qui a été appliqué dans cette nouvelle étude était d'éliminer les stations redondantes. Afin de pallier ce problème, une seule station a été retenue sur la plupart des cours d'eau.

1.2.4 Estimation régionale

Cette étape nécessite un grand nombre de variables explicatives pour aboutir à des performances prédictives satisfaisantes. Ce nombre dépasse, généralement, cinq comme dans l'étude de Ouarda et al. (2018). Les techniques antérieures de modélisation telles que les modèles linéaires généralisés GLM (McCullagh et Nelder, 1989) se sont avérées limitées dans leur capacité à s'adapter aux relations complexes et non linéaires qui existent souvent entre les variables réponses et les prédicteurs environnementaux (ex., Austin et al.,

1990). Pour résoudre ce problème, des techniques alternatives sont maintenant disponibles, permettant une description plus réaliste du phénomène. Parmi celles-ci, les modèles additifs généralisés GAMs (Hastie et Tibshirani, 1990) sont les plus utilisés, que ce soit dans les études terrestres (ex., Leathwick, 1998) ou bien marines (ex., Gregr et Trites, 2001). Les GAMs sont des modèles de régression non linéaires flexibles qui ont été utilisés dans le contexte RFA entre autres par Chebana et al. (2014). Les auteurs ont constaté que les méthodes basées sur le GAM présentent les meilleures performances par rapport aux classiques. Dans le contexte d'analyse thermique régionale, une étude a été menée par Charron et al. (2019) dans laquelle deux modèles statistiques ont été comparés (MLR et GAM) couplés à deux approches de délimitation des régions homogènes (HCA et ROI). Cette étude aboutit à la conclusion que le modèle GAM en conjonction avec l'approche HCA est le meilleur modèle pour prédire les mesures des températures de l'eau. Cependant, le GAM présente encore un certain nombre d'inconvénients. En effet, il est difficile d'adapter le GAM à des bases de données complètes en raison des limitations de mémoire imposées par les complexités numériques de cette approche (Leathwick et al., 2006). Aussi, le GAM ne gère pas bien l'interaction entre les variables physiographiques au sein du bassin versant (par exemple, Ramsay et al., 2003). Conséquemment, l'introduction des termes d'interaction entre les variables explicatives utilisées qui est reconnue depuis longtemps (voir par exemple, Niehoff et al., 2002) semble être essentielle pour de meilleures estimations. De ce fait, l'approche des splines de régressions adaptatives multivariées (MARS) (Friedman, 1991) a été introduite comme solution ultime à la problématique décrite ci-haut. En suivant la même idée, une étude effectuée par El Mahdi et al. (2020), dans laquelle les modèles GAM et MARS ont été appliqués pour modéliser les variations des superficies de panache d'eau froide à l'échelle journalière. Les résultats de l'étude prouvent que le modèle MARS offre une meilleure performance dans la prévision et l'estimation de la variabilité de l'aire de refuge thermique que le GAM dans toutes les stations d'études. Cependant, compte tenu de ces résultats, les deux modèles présentent un grand potentiel avec un RMSE relatif variant entre 6,93 % et 38,39 %.

1.3 Structure du mémoire

Ce projet de recherche intitulé « Analyse régionale stationnaire des températures de l'eau en rivières » s'inscrit dans le cadre du programme de maîtrise en sciences de l'eau. Ce travail de recherche a bénéficié du financement du conseil de recherche en sciences naturelles et en génie du Canada (NSERC) et de la mission universitaire de Tunisie en Amérique du Nord (MUTAN).

Ce travail de recherche vise à développer une approche d'analyse régionale des métriques de température de l'eau pertinentes pour la gestion des stocks du saumon Atlantique dans l'Est du Canada. Ce mémoire de maîtrise se divise en quatre chapitres;

- Le chapitre 1 présente une introduction générale avec une mise en contexte afin de mieux connaître le cadre de ce projet de recherche. Puis, une revue de littérature des travaux antérieurs s'inscrivant dans le thème " modélisation de la température de l'eau ". Par la suite, une présentation de la structure du mémoire.
- Le chapitre 2 présente le travail de recherche sous forme d'article rédigé en anglais qui traite le développement d'une approche de modélisation régionale nommée Analyse thermique régionale.
- Le chapitre 3 présente une analyse supplémentaire effectuée pour tester la performance de l'approche des splines de régression adaptatives multivariées MARS. Ensuite, une comparaison de sa performance avec le modèle additif généralisé (GAM).
- Le chapitre 4 présente une discussion générale et une conclusion, servant à synthétiser et à mettre en relation les éléments principaux présentés dans l'article en faisant sortir les éléments originaux et novateurs de la recherche, d'identifier les éventuelles limites des travaux et de présenter des perspectives pour des travaux futurs.
- Le chapitre 5 présente une bibliographie pour l'ensemble du mémoire.

2 REGIONAL THERMAL ANALYSIS APPROACH: A MANAGEMENT TOOL FOR PREDICTING WATER TEMPERATURE METRICS RELEVANT FOR THERMAL FISH HABITAT

Olfa Abidi^a
André St-Hilaire^a
Taha B. M. J. Ouarda^a
Christian Charron^a
Claudine Boyer^a
Anik Daigle^a

^a Canada Research Chair in Statistical Hydro-Climatology, INRS-ETE, 490 de la couronne,
Québec, QC, G1K9A9, Canada

Corresponding author:

Olfa Abidi: Olfa.Abidi@inrs.ca

Tel: +1418-2719342

INRS-ETE, 490 de la couronne, Québec, QC, G1K9A9, Canada

Soumis au Journal Ecological Informatics

(Sous révision)

Contribution de l'auteur principal et des coauteurs

Olfa Abidi : la conception et la mise en place du plan de travail, méthodologie, analyse formelle, interprétation, rédaction-copie originale. André St-Hilaire : la conception et la mise en place du plan de travail, supervision, révision. Taha B. M. J. Ouarda : la conception et la mise en place du plan de travail, supervision, révision. Christian Charron: Contribue de manière substantielle au développement du code informatique et des algorithmes. Claudine Boyer et Anik Daigle : la collecte des données.

Highlights

- Regional analysis approach is used to derive estimates of five thermal metrics relevant for Atlantic salmon.
- Multiple linear regression model and generalized additive model were tested.
- Thermal homogeneous regions have been defined.
- Environmental drivers that significantly impact water temperature metrics were selected.
- The interannual mean of the number of consecutive days (MaxNumDay) above a potentially stressful threshold is a key metric for Atlantic salmon, is not modelled adequately due to the small sample size and the small number of stations with high values.

Abstract

Water resource management, including fisheries, is facing new challenges associated with climate change. This study sheds light on the modeling of water temperature indices (metrics) that describe critical thermal maxima of the Atlantic salmon (*salmo salar*). These thermal metrics include MaxWaterTmax (Interannual mean of maximum summer temperature), MaxNumDay (Interannual mean of the number of consecutive days with maximum water temperature > 25 °C and minimum water temperature > 20 °C). The latter are important indicators to evaluate thermal variability. Three other parameters of a Gaussian function fitted to the interannual daily mean temperatures characterizing the thermal regime of 146 stations located in Eastern Canada were estimated. These three parameters are Gaussian_a (Maximum of interannual daily mean temperature), Gaussian_b (Mean duration of the warm period), and Gaussian_c (Date of occurrence of the interannual maximum temperature). To meet the above-mentioned target, the classical Multiple linear regression model (MLR) and the non-linear Generalized additive model (GAM) were tested and compared to estimate the five thermal metrics. The regression-based approaches involve the identification of thermally homogeneous regions based on three approaches: hierarchical clustering analysis (HCA), regions of influence (ROI) as well as canonical correlation analysis (CCA). Then, the regional MLR and GAM models were applied within the delineated homogenous regions. Also, they were compared and evaluated with an application to a single region encompassing all stations. For each regional estimation model and each thermal metric, a set of optimal explanatory variables were selected using a forward stepwise regression procedure. The database consisted of 22 environmental predictors related to physiography, topography, climate, land cover and surface deposits. To assess performance of the models, the following statistical metrics were used: coefficient of determination R^2 , root mean square error (RMSE), Bias, Relative root mean square error (RRMSE) and percent bias (PBias). The results demonstrate that the non-linear GAM model was consistently better than the simpler MLR model for estimating the five thermal metrics. Results also show that the best practice consists in delineating homogeneous regions before applying the regional GAM model. According to all performance criteria, delineation of regions with the

HCA approach is considered to be more flexible and to lead to better performances than neighborhood-based approaches (CCA and ROI).

Keywords

River water temperature, Regional thermal analysis, Multiple linear regression model MLR, Generalized additive model GAM, Thermal homogeneous regions, HCA.

1 Introduction

In the last two decades, stream and river temperature has increasingly become a popular topic of research and has received great attention with enhanced monitoring effort and sharing of existing data due to the fundamental role that it plays on the environment (Ouellet et al., 2020). River temperature is the cornerstone of numerous chemical and biological processes that occur in rivers, as well as biodiversity. As a determining factor, water temperature is one of the most significant physical properties of freshwater systems. Degraded water quality can be harmful to the health of aquatic communities. Indeed, freshwater organisms have been found to be adversely affected by changes in the seasonal signal of water temperature (Ward and Stanford, 1982; McCullough, 1999). In this context, numerous studies have been carried out by several authors who contributed to a better understanding of water temperature impact on instream habitat (Breau et al., 2007; Elliott and Elliott, 2010; Dugdale et al., 2016).

In particular, water temperature is fundamental for ectothermic fish such as salmonids, whose physiological processes are directly controlled by ambient temperatures (Elliott et al., 1998; Lund et al., 2002; Mather et al., 2008) and who are often intolerant of high temperatures (e.g., Elliott, 1991; Jonsson and Jonsson, 2009). Consequently, prolonged exposure to temperature above the upper critical threshold results in mortalities (Elliott, 1991; Elliott and Elliott, 2010). The Atlantic salmon (*Salmo salar*), a cold-water iconic fish species in Eastern Canada and Northern Europe, is one of the focal fish species for which assessing thermal variability is a management requirement. This species has been widely studied in recent years. It has been found to be adversely affected by high maximum summer temperatures (e.g., Corey et al., 2017). Several studies carried out in this context have shown that the thermal regime of rivers has significant impacts on the growth, condition, and spatial distribution of juvenile Atlantic salmon (Nicieza and Metcalfe, 1997; Elliott and Elliott, 2010; Sundt Hansen et al., 2018). For instance, Danie et al. (1984) reported that eggs are the most temperature delicate Atlantic salmon life history stage. Other research projects have focused on studying the temperature thresholds at which species such as Atlantic salmon experience considerable heat

stress during the hot summer months (Lund et al., 2002; Hodgson and Quinn, 2002). Recently, Heggnes et al. (2021) showed that regular occurrences of high stream temperatures (> 22.5 °C) have been observed to cause thermal stress on juvenile Atlantic salmon.

Many previous studies focused on investigating the natural variability of river thermal regimes and determining the impact of different anthropogenic disturbances, including deforestation (Moore et al., 2005), dams (Maheu et al., 2016b) as well as global warming (Isaak et al., 2017; Isaak and Rieman, 2013; Van vliet et al., 2013). Other contributions focused on characterizing the thermal regime of streams and rivers across Canada and the US. For instance, Maheu et al. (2015) used a sinusoidal function fitted to historical time series to characterize and group rivers according to their thermal regime in conterminous US. Daigle et al. (2019) used a Gaussian function to characterize the thermal regime of rivers in Quebec (Canada). In most instances, relevant thermal metrics have been calculated at stations with relatively long time series. These metrics may be related to the magnitude (mean or extreme conditions), frequency of occurrence of warm or cold events, timing and duration of these events and rate of change in water temperatures (Chu et al., 2010; Olden and Naiman, 2010; Arismendi et al., 2013). Similarly, cumulative degree-days, a measure of magnitude and duration, has also been considered in biological models given the association of water temperature with the growth of aquatic organisms (Vannote and Sweeney, 1980; Neuheimer and Tggort, 2007). It is important to bear in mind that the selection of representative metrics can be partially subjective as a result of characteristics of the study area and its scale (Maheu et al., 2015). For example, metrics of magnitude and rate of change were identified as good discriminants of thermal regimes in the Great Lakes region, Canada (Chu et al., 2010), while metrics of magnitude and annual variability were identified as good discriminant variables in the Southern Cape region of South Africa (Rivers-Moore et al., 2013).

Expertise related to water temperature is increasingly becoming an essential tool for water resources management. For instance, direct measurements of water temperature in Atlantic salmon rivers are useful tools for assessing thermal variability and can provide insight for fisheries management. However, these

data remain limited and sporadic at many sites and, therefore, comprehension and characterization of the thermal regime of rivers remains a difficult mission. To cope with the existing challenges, predicting estimates of the hydrological data at sites with no scarce of information is a very interesting alternative technique for water resources planning and integrated management (Babaei et al., 2019; Mehdizadeh et al., 2019).

Consequently, a great deal of previous studies has focused on water temperature modeling in space and time (Caissie, 2006; Benyahya et al., 2007; Dugdale et al., 2017). In the past decades, physically based deterministic models and statistical models have been developed and successfully applied for predicting river water temperature (Benyahya et al., 2007; Cole et al., 2014; Kwak et al., 2017). In recent years, with the development of artificial intelligence, advanced machine learning techniques have gained attention and have been proven to be effective in river temperature modeling (Zhu and Piotrowski, 2020, Deweber and Wagner, 2014; Piotrowski et al., 2015; Zhu et al., 2018; Piotrowski and Napiorkowski, 2019; Zhu et al., 2019a; Zhu et al., 2019b; Feigl et al., 2021). In addition, process-based models (e.g., Read et al., 2019) or a combination of both approaches through physics-guided machine learning methods have been developed (e.g., Jia et al., 2021). The modeling effort has mainly focused on local models (e.g., Caissie et al., 1998; Jeong et al., 2013; Boudreault et al., 2019) and were seldom developed at the regional scale.

To overcome this shortcoming, some studies have recently attempted to develop regional models. Beaufort et al. (2021) focused on thermal peaks, Gallice et al. (2015) estimated monthly mean temperatures and DeWebber and Wagner (2014) developed a regional model for mean daily stream temperatures. The two most comprehensive works on the regionalization of the relationship between river temperature and air temperature have been carried out by Ducharme (2008) and Segura et al. (2014). In the same context, Hrachowitz et al. (2010), Rivers-Moore et al. (2012) and Imholt et al. (2013) expressed water temperature as linear combination of climatic and physiographic variables for each month of the year separately. However, previous studies that have been conducted around many rivers in Eastern Canada and elsewhere in the world were hindered by certain limitations. One of the greatest challenges is the relative paucity of

thermal data on many rivers, including those that host Atlantic salmon populations. Besides, the inhomogeneous spatial distribution of measurement sites, the variable, limited duration of annual time series as well as the small number of years with data for most sites represent another challenge for model implementation. In hydrology, the approach that allows to alleviate this problem for high or low flow quantiles is termed regional (or pooled) frequency analysis (RFA). The RFA approach involves two main steps, namely the delineation of homogeneous regions (DHR) and regional estimation (RE). The first step defines groups of stations based essentially on similarities in the physiographic, meteorological and hydrological characteristics of the watersheds. The second step allows to estimate metrics of interest through the transfer of information from the gauged sites to the ungauged target site. This step requires a potentially large number of explanatory variables (generally > 5 , according to Ouarda et al., 2018) to achieve satisfactory predictive performance.

In hydrological studies, different methods have been introduced to establish these homogeneous groups. A simple approach might be to define groups of geographically contiguous stations with similar temperature regimes. An alternative to this approach is based on similarities other than location. This approach allows us to define regions based on the climatic and physiographic characteristics of the watersheds (Ouarda et al., 2001). This approach is used in the present study. Groups of potentially non-contiguous rivers have been constructed. The use of non-contiguous regions has been recommended in the literature (e.g., Ouarda et al., 2008; Haddad and Rahman, 2012). Hierarchical clustering analysis (HCA) is based on site characteristics and is one of the most practical methods used to define non-contiguous regions (Hosking and Wallis, 1997). The regions of influence method ROI (Burn, 1990) and the canonical correlations analysis CCA (Cavadias et al., 2001) approaches were also tested. A neighborhood is defined as a group of stations identified for a specific, target site that includes the gauged sites with similarities in a mathematical sense (mathematical distance computed from physiographic, geographic, hydroclimatic variables) to that reference point. Comparative studies have shown that neighborhoods lead to better regional estimates than fixed regions (Burn, 1990; Tasker et al., 1996; Ouarda et al., 2008). For the second methodological step

(regional estimation), previous modeling techniques such as generalized linear GLM models (McCullagh and Nelder, 1989) have been shown to be limited in their ability to adapt to the complex, non-linear relationships that often exist between response variables and environmental predictors (e.g., Austin et al., 1990). To solve this problem, alternative techniques are now available, allowing for a more realistic description of the phenomenon. Among these, the generalized additive models GAMs (Hastie and Tibshirani, 1990) are perhaps the most frequently used in both terrestrial (e.g., Leathwick, 1998) and marine studies (e.g., Gregr and Trites, 2001). GAMs are flexible non-linear regression models that have been used in RFA by Chebana et al. (2014). The authors found that GAM-based methods show the best performance compared to more conventional approaches.

The primary objective of this work was to adapt RFA (e.g., Reed et al., 1999; Ouarda et al., 2000; Basu and Srinivas, 2014; Haddad et al., 2014) to derive estimates of water temperature metrics in an approach called regional thermal analysis (RTA). Five thermal metrics, known to be relevant for Atlantic salmon in Eastern Canada related to maximum temperature and its date of occurrence, as well as the seasonal variability were selected to test the RTA approach. Other objectives, include the identification of predictors which significantly affect water temperature metrics. We also aim at defining which model best fit our response variables by evaluating combinations for regional thermal analysis and analysing their advantages and limitations, prior using these approaches for fisheries-related water resources management.

The overall structure of this paper is as follows: Section 2 begins by laying out a brief overview of the regionalization methods that are considered in the current study. Case study and considered dataset are described in Section 3. The results are presented and discussed in sections 4 and 5, respectively. The last section provides the main conclusions and includes a discussion of the implication of the findings to future research.

2 Methodology

The methodology of the study integrates a number of methods for the development of the regional thermal analysis approaches and for their validation. The adopted statistical approaches are described as follows:

2.1 Delineation of thermally homogeneous regions (DHR)

2.1.1 Hierarchical clustering analysis (HCA)

This is the main approach used in RFA to group stations (Hosking and Wallis, 1997). It consists in grouping watersheds that have similar climatic and physiographic characteristics. This statistical method minimizes the differences within the group and maximizes differences between groups. It consists in calculating a mathematical distance (Euclidean in the present study, Equation. (1)) between each pair of stations in the multidimensional space defined by the climatic and physiographic variables selected. The standardized Euclidean distance is defined by:

$$d^2(r, s) = (x_r - x_s)D^{-1}(x_r - x_s)' \quad (1)$$

Where x_r and x_s are the vectors of coordinates in the physiographic and meteorological space for basin r and s respectively and D is the diagonal matrix for which the diagonal elements v_j^2 are the variances of the respective variables.

The clustering was performed through Ward's algorithm (Ward, 1963) based, primarily, on minimizing the sum of the squared distances between each site in each group and the group's centroid to ensure maximum similarity of the elements of the group (Equation. (2)). The Ward aggregation method was done in ascending order starting with all stations in separate groups and coalescing them (Johnson, 1967).

$$WSS_p = \sum_{i=1}^{n_p} d^2(x_{pi}, \bar{x}_p) \quad (2)$$

Where n_p is the size of the cluster p and \bar{x}_p is the centroid of the cluster p . The distance between cluster p and q is given by Equation. (3):

$$d_w(p, q) = WSS_{(p+q)} - (WSS_p + WSS_q) \quad (3)$$

The delineation of regions using the HCA method was carried out independently for each water temperature metric. The choice of the cut-off distance has an important impact on the number of stations in the regions and therefore on performance. In the present study, the level of truncation threshold chosen is the one that produced the lowest root mean square error (RMSE). The choice of the number of classes was made by visually selecting a truncation level through the tree diagram of possible groups also called dendrogram.

2.1.2 Region of influence (ROI)

To identify neighborhood-based regions, Burn (1990) proposed an approach called region of influence (ROI). It is used to separately identify, for each target site, the set of thermally similar sites to be used in the estimation of water temperature metrics. The ROI method uses the weighted Euclidean distance in a multidimensional space defined by physiographic and meteorological variables. The Euclidean distance is given by Equation. (4) as follow:

$$D_{ij} = \left[\sum_{k=1}^K (C_k^i - C_k^j)^2 \right]^{1/2} \quad (4)$$

With D_{ij} is a Euclidean distance; C_k^i, C_k^j are the standardized values of attribute k for stations i and j and K is the number of attributes.

2.1.3 Canonical correlations analysis (CCA)

CCA is a statistical method of multivariate analysis used to explore and describe the relationships that may exist between two groups of random variables. This approach was introduced in hydrology by Ouarda et al. (2000) and Cavadias et al. (2001) to identify hydrological neighborhoods. Indeed, due to the missing thermal information at the ungauged target location indicate $i = 0$, the CCA method estimates the

unavailable thermal variables in the form of linear combinations of site characteristics. It provides a linear estimate $v_0 \simeq \Lambda u_0$ with $\Lambda = \text{diag}(\rho_1, \dots, \rho_k)$. ρ_1, ρ_k represent the canonical correlations coefficients.

To delimit the neighborhoods, the CCA approach considers the canonical scores $u_i = (a_1, \dots, a_r)' x_i$ and $v_i = (b_1, \dots, b_r)' y_i$ which are, respectively, linear combinations of site characteristics x_i and thermal metrics y_i for site i . In hydrology, this technique uses the Mahalanobis distance (Equation. (5)) in the multidimensional space defined by the canonical physiographic, meteorological and hydrological variables (Ribeiro-Corréa et al., 1995) as follows:

$$d(v_i, \Lambda u_0) = (v_i, \Lambda u_0)' (I - \Lambda)^{-1} (v_i, \Lambda u_0) \quad (5)$$

Where d representing the Mahalanobis distance, I is an identity matrix, v_i denotes the corresponding values of canonical hydrological variables for the target site and u_i denotes the corresponding values of canonical physiographic and meteorological variables for the target site.

Hydrological variables are generally not continuous in geographic space. For instance, river temperature can change significantly downstream of the confluence between a river and its tributary. However, they are continuous in physiographic canonical space. One of the peculiarities of this method is the prediction of the neighborhood centers since the true thermal centers are unknown. The principle of the CCA technique consists in creating, from the thermal and physiographic variables of the watersheds: $X = X_1, X_2, \dots, X_q$ and $Y = Y_1, Y_2, \dots, Y_r$, respectively, new variables called canonical variables U and V , so that the canonical correlation $\lambda_i = \text{corr}(U_i, V_i)$ is maximum by imposing a unit variance. It is important to mention that these are linear transformations of the original variables X and Y :

$$U_i = a_{i1}X_1, a_{i2}X_2, \dots, a_{iq}X_q \quad (6)$$

$$V_i = b_{i1}Y_1, b_{i2}Y_2, \dots, b_{ir}Y_r \quad (7)$$

Where $i = 1, \dots, p$ and $p = \min(r, q)$.

CCA makes it possible to identify the vectors \mathbf{a} and \mathbf{b} for which $corr(U, V)$ is maximum. The CCA technique requires the normality of the thermal and the physiographic and meteorological variables.

2.2 Regional estimation (RE)

Once the clusters of thermally homogeneous rivers are determined, the next step is to estimate the temperature metrics of interest through the independent explanatory variables known to influence the water temperature. Each thermal metric was treated separately, and a statistical model was built to establish the link between the metric of interest and the selected predictors. In the present study, two statistical models were tested: Multiple Linear Regression model (MLR) and the Generalized Additive Model (GAM), they are described as follows:

2.2.1 Multiple Linear Regression (MLR)

This is a linear parametric model representing one of the simplest methods that can be used for information transfer to the ungauged target site. This method consists in establishing a direct linear relationship between the thermal variables and the physiographic and meteorological variables. It can be expressed by Equation. (8):

$$Tw(t) = \beta_0 + \sum_{i=1}^n \beta_i x_i(t) + \varepsilon \quad (8)$$

where $Tw(t)$ represents the water temperature metric; β_i is the coefficient to be adjusted.

x_i represents the explanatory variables and ε is an error term.

2.2.2 Generalized Additive Model (GAM)

This method is more flexible than the simple MLR model (Hastie and Tibshirani, 1990) since it is able to model a wide variety of non-linear relationships between the response variable Y and the explanatory variables x_i (McCullagh and Nelder, 1998). It does not assume any specific form of dependence between

the predicted variable and the predictors. It allows a more realistic description of the thermal process due to the nonparametric adjustment of the smooth functions (f).

$$g(Y) = \alpha + \sum_{j=1}^n f_j(X_j) + \varepsilon \quad (9)$$

Where g is the monotonic link function, f_j represents the smooth function giving the relationship between the response Y and explanatory variables X_j (here, a combination of cubic splines).

The GAM model is increasingly adopted in several fields such as hydro-climatology and environmental modeling (Wen et al., 2011; Rahman et al., 2018), public health (Leitte et al., 2009, Bayentin et al., 2010) as well as renewable energies (Ouarda et al., 2016). It has been used for modeling daily mean water temperature in a few studies (Laanaya et al., 2017; Boudreault et al., 2019). The two regression-based models mentioned above were processed in the R software, using the "mgcv" package (Wood, 2006).

2.3 Stepwise regression

Given the importance of using an optimal set of predictors as input in parametric regression models, it is very common to apply predictor selection algorithms. The forward stepwise selection procedure is applied in this work to select the optimal explanatory variables as in Charron et al. (2019). It consists of a gradual addition of the most efficient variables from an initial model without any candidate variable. The explanatory power of each predictor over the response variable was identified by the Akaike information criterion (AIC), an estimator of the relative quality of a statistical model for a given set of data. A lower AIC score is better. The variable which produced the lowest AIC was retained.

2.4 Validation

A jackknife cross-validation procedure commonly referred to as "Leave-One-Out" in hydrology was applied in order to compare the performance of models tested in the present work. This is done by estimating

the model parameters using all stations except one, which is left to be considered as the target site in the defined region. Then, using the fitted model, the metric of interest is estimated for the station that was left out and compared to that calculated from the observations. This procedure is repeated for each station in the region.

The validation step was performed for each algorithm separately. Based on this procedure, numerous standard performance criteria are used to assess the predictive power of each regional model (Ouali et al., 2016), such as the coefficient of determination (R^2) (Equation. (10)), the root mean square error (RMSE) (Equation. (11)) providing information about the precision of the estimator on an absolute scale, and the relative root mean square error (RRMSE) (Equation. (12)). Other statistical indicators used to assess the model's performance were the bias (BIAS), calculated according to the Equation. (13) and the percent bias (PBIAS) (Equation. (14)), which are measures of overestimation or underestimation of a model.

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (11)$$

$$RRMSE = 100 \times \left[\sqrt{\frac{1}{n} \times \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t)}} \right] \quad (12)$$

$$Bias = \frac{\sum_{t=1}^n (y_t - \hat{y}_t)}{n} \quad (13)$$

$$PBias = 100 \times \left[\frac{\sum_{t=1}^n (\hat{y}_t - y_t)}{\sum_{t=1}^n (y_t)} \right] \quad (14)$$

Where \hat{y}_t represents the simulated value of the water temperature metric for a period t ; y_t is the observed value of the metric of interest during a period t ; \bar{y}_t represents the average of the observed values of the metric and n is the size of the data sample.

3 Case study

This study is mainly based on water temperature data from the RivTemp database. RivTemp is a partnership between universities, provincial and federal governments, watershed groups and organizations dedicated to the conservation of the Atlantic salmon in Canada (<http://rivtemp.ca>; Boyer et al., 2016). RivTemp contains daily water temperature measurements for 433 monitoring stations installed on 158 rivers in Quebec and the Atlantic Provinces that were operational for a period of one summer to 28 years between 1985 and 2017. This study focused on 146-water temperature monitoring stations in Eastern Canada that have at least four years of data and that are distributed in Newfoundland-Labrador, New Brunswick and Quebec. In the latter province, most stations are located in the Gaspé Peninsula, the Saguenay region and the North Shore regions (Figure 2-1).

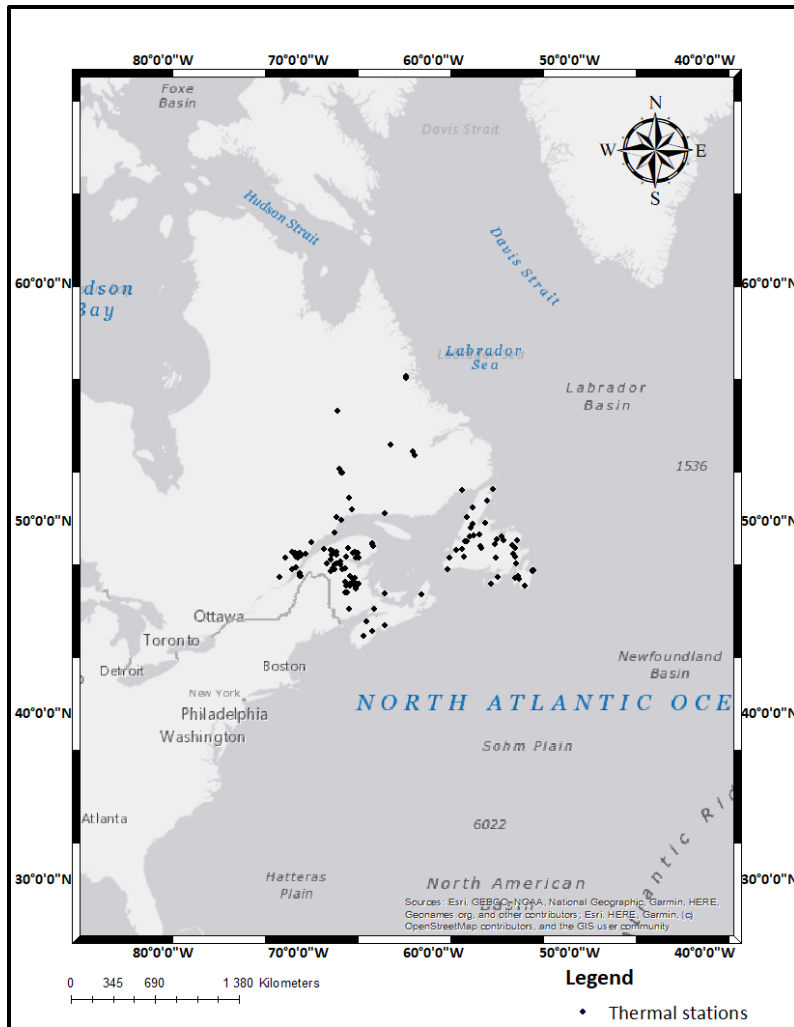


Figure 2-1: Location of thermal stations across the study area.

We present here a brief summary of the data used in this analysis.

3.1 Thermal water temperature metrics

To test the RTA approach, a limited number of water temperature metrics known to be relevant for Atlantic salmon were selected. The metrics of interest in the study are as follows:

- 1) Interannual mean of maximum temperature (MaxWaterTmax).
- 2) Interannual mean of the number of consecutive days (MaxNumDay) above a potentially stressful threshold for Atlantic salmon, i.e., with maximum water temperature > 25 °C and minimum water temperature > 20 °C.

In addition to these two metrics, three other variables were chosen based on the study of Daigle et al. (2019) in which the thermal regimes of Quebec rivers were characterized using a Gaussian function fitted to the interannual mean daily temperature T_w , as a function of the day of the year (d) as follow:

$$\hat{T}_w(d) = a \exp\left(-\frac{1}{2}\left(\frac{d-c}{b}\right)^2\right) \quad (15)$$

Where parameter a (Gaussian_a) is a scale factor representing the maximum value of interannual daily mean water temperature (°C). b (Gaussian_b) represents the standard deviation which is a measure of the duration of the warm period (days) and the parameter c (Gaussian_c) is the date of occurrence of the interannual maximum temperature (days).

Some statistical characteristics of the selected thermal metrics across all 146 stations are summarized in table 2-1.

Metric	Description	unit	Mean	Median	Minimum	Maximum
MaxWaterTmax	Interannual mean of maximum temperature	°C	23.26	23.46	13.01	28.86
MaxNumDay	Interannual mean of the number of consecutive days above a potentially stressful threshold for Atlantic salmon: With maximum water temperature > 25 °C and minimum water temperature > 20 °C	days	1.05	0.03	0.00	15.25
Gaussian_a	The maximum of interannual daily mean temperature	°C	18.41	18.55	9.55	24.36
Gaussian_b	The duration of the warm period	days	57.06	56.24	40.02	82.04
Gaussian_c	The date of occurrence of the interannual maximum temperature	days	213.91	213.27	204.94	238.39

Table 2-1. Selected thermal metrics across all 146 stations and some of their statistical characteristics.

MaxWaterTmax corresponds to the interannual mean of the annual maximum water temperature.

Gaussian_a corresponds to the height of the curve's peak (maximum value of interannual daily mean temperature)

3.2 Climatic and physiographic variables

To delineate groups of thermally homogeneous rivers, a certain number of climatic and physiographic variables were selected. In the current study, the meteorological data used were extracted from the ANUSPLIN database (Hutchinson et al., 2009). These data are interpolated into a 10km × 10km grid derived from observations made at Canadian meteorological stations. The available interpolated data are the maximum and the minimum daily air temperature (AirTmax, AirTmin) as well as daily precipitation (TotPrecip). Daily meteorological data were extracted using the ANUSPLIN grid point closest to each temperature station. For the analysis at hand, the annual values of AirTmax (maximum, mean, minimum), AirTmin (maximum, mean, minimum) as well as total precipitation were calculated from daily values. The physiographic variables characterizing the watersheds associated with each station in the network were compiled and added to the RivTemp database (Boyer et al., 2016). These variables were selected because they can have a significant influence on the dynamics of the thermal regime at various levels as well as explain the spatial variability of the selected thermal variables.

Table 2-2 lists the selected physiographic and climatic variables, some of their statistical characteristics as well as their transformations across all 146 stations.

Variables	Description	Unit	Mean	Median	Minimum	Maximum	Transformations
Physiographic variables							
Basin Area	Catchment area	km^2	1099	360	0.75	25444	Log
Xcentroid	X-axis location of the catchment centroid	m	12811727	1903793	216190	2984754	
Ycentroid	Y-axis location of the catchment centroid	m	572212	500531	19250	1310914	
Lake Area	Total lake area	%	4.56	2.91	0.00	26.25	
Drainage Density	Drainage density of the hydrological network	m^{-1}	1.55	1.48	0.43	3.02	
MinElevation	Minimum elevation of the catchment	m	99.25	67.50	-4.04	625.18	
MaxElevation	Maximum elevation of the catchment	m	586.27	599.5	83.00	1153.67	
MeanElevation	Mean elevation of the catchment	m	328.27	334.75	37.00	797.56	
Elevation Station	Elevation at the station	m	101.45	68.5	1.00	625.18	
Slope	River slope	%	0.0125	0.0050	0.0001	0.1634	√
Climatic variables							
TotPrecip	Total annual precipitation at the nearest grid point	mm	970.68	954.56	669.12	1298.35	
MeanAirTmax	Annual mean of maximum air temperatures at the nearest grid point	°C	8.00	8.31	0.17	12.13	
MaxAirTmax	Annual maximum of maximum air temperatures at the nearest grid point	°C	29.37	29.74	22.62	33.71	
MeanAirTmin	Annual mean of minimum air temperatures at the nearest grid point	°C	-1.28	-1.20	-8.78	2.90	
MinAirTmin	Annual minimum of minimum air temperatures at the nearest grid point	°C	-28.39	-29.55	-40.96	-15.88	

Table 2-2: Physiographic and climatic variables and some of their statistical characteristics.

Table 2-3 lists the land cover and surface deposits variables, some of their statistical characteristics as well as their transformations across all 146 stations.

Variables	Description	Unit	Mean	Median	Minimum	Maximum	Transformations
Land cover							
Shrubland	Percentage of shrubland area	%	7.36	2.81	0.00	47.76	√
Grassland	Percentage of grassland area	%	2.11	1.16	0.00	29.17	√
Wetland	Percentage of wetland area	%	1.85	0.75	0.00	13.71	√
Forest	Percentage of forest area	%	77.08	85.49	5.85	99.69	√
Surface deposits							
GlacialDeposits	Percentage of area covered by glacial deposits	%	61.05	66.81	0.00	100.00	√
Rock	Percentage of area covered by rock	%	6.33	0.02	0.00	100.00	√
FluvioGlacialDeposits	Percentage of area covered by fluvio-glacial deposits	%	4.00	1.19	0.00	34.39	√

Table 2-3: Land cover and surface deposits variables and some of their statistical characteristics.

The variables presented in tables 2-2 and 2-3 were selected through literature review. Some important variables which significantly affect water temperature were not included as predictors, given that their impact may be local (e.g., canopy cover over a reach) or not represented in the region (i.e., relatively low percentage throughout the study region, such as impervious area).

4 Results

Within this section, as starting point, we presented results of the selection of the physiographic and the climatic variables included in the two regional models MLR and GAM. Subsequently, results related to

delineation methods are discussed. Last, a comparison of the different combinations was made. Only a chosen sample of results will be presented here to avoid repetition.

Selection of explanatory variables

As mentioned in the Methodology Section, the MLR and GAM regional estimation models were applied with four different delineation methods: (1) all stations together without delineation of thermally homogeneous regions, (2) HCA, (3) CCA and (4) ROI.

The selection of predictor variables in each case was performed using a forward stepwise regression procedure. The forward procedure was applied on the total set of variables from the first procedure so that 22 subsets of predictors were tested for selection. The application of the forward stepwise is carried out individually for each model. The top six most important predictors were selected (Table 2-4).

model	metric	Selected variables					
MLR	MaxWaterTmax	MeanAirTmax	Xcentroid	Slope	Forest	Lake Area	Shrubland
	MaxNumDay	MeanAirTmax	Lake Area	TotPrecip	Ycentroid	Xcentroid	Forest
	Gaussian_a	MeanAirTmax	Lake Area	Slope	Ycentroid	Grassland	Shrubland
	Gaussian_b	MeanAirTmin	Forest	Basin Area	Wetland	TotPrecip	MeanElevation
	Gaussian_c	Ycentroid	Basin Area	MinElevation	TotPrecip	Forest	MaxElevation
GAM	MaxWaterTmax	Ycentroid	Basin Area	Forest	Xcentroid	MeanAirTmin	Rock
	MaxNumDay	Ycentroid	FluvioGlacialDeposists	Lake Area	MaxAirTmax	Xcentroid	MinAirTmin
	Gaussian_a	Ycentroid	Lake Area	Slope	MinAirTmin	Rock	Forest
	Gaussian_b	MeanAirTmin	Xcentroid	Basin Area	MinAirTmin	Forest	Wetland
	Gaussian_c	Ycentroid	Basin Area	Forest	Elevation	Slope	GlacialDeposists Station

Table 2-4: Selected predictors when all stations grouped together.

4.1 Selection of explanatory variables for MLR

As expected, at least one air temperature variable has been found to be substantially more important in predicting four out of five thermal metrics in the case of the MLR model. The outcome of this research is

consistent with the findings of Hill et al. (2013) who also identified air temperature to be a strong predictor of mean annual, summer and winter water temperature across the USA and of daily mean and maximum temperature in Canada (e.g., Caissie et al., 2001; Benyahya et al., 2007). When a model used no air temperature variable, the Ycentroid (i.e., latitude) and elevation were selected as predictors since they are highly correlated with air temperature. Other important and frequently selected explanatory variables were slope, basin area, total lake area, total precipitation, and forest cover.

We have observed that stations adjacent to lakes are characterized by the highest values of MaxNumDay and MaxWaterTmax. This could be explained, mostly, by the effect of lakes. Also, past research has confirmed these expectations and lake effects on temperature (see for example, Scott and Huff, 1996; Yang et al. 2019; Leach et al. 2021). Indeed, the latter are considered as effective sentinels for climate change (Adrian et al., 2009) because they are impacted by changes in air temperature and integrate information about changes in the catchment. For instance, all through the summer, lakes act as heat sinks storing up extra energy from the atmosphere. More generally, lake surface water tends to warm up over the summer faster than in small streams or rivers, because of high exposure to solar radiation and long water residence time.

4.2 Selection of explanatory variables for GAM

Considering that the GAM model has the capacity to model non-linear relationships that may exist between the response variable and predictors, a different selection of variables is expected compared to the MLR model. The results suggest that the Ycentroid (akin to latitude) is a particularly important variable for predicting four out of five thermal metrics. Other important variables are air temperature, basin area as well as lake area. Figure 2-2 is a representative example showing the relationship between each selected predictor and the water temperature metric (the variable MaxWaterTmax is shown as an example).

It is clear from Figure 2-2 that some variables present important non-linear relationships with the water temperature metric. The fact that most of these relationships are not linear indicates that the GAM model is a better-suited model for regionalization than the simple multiple linear regression model (MLR).

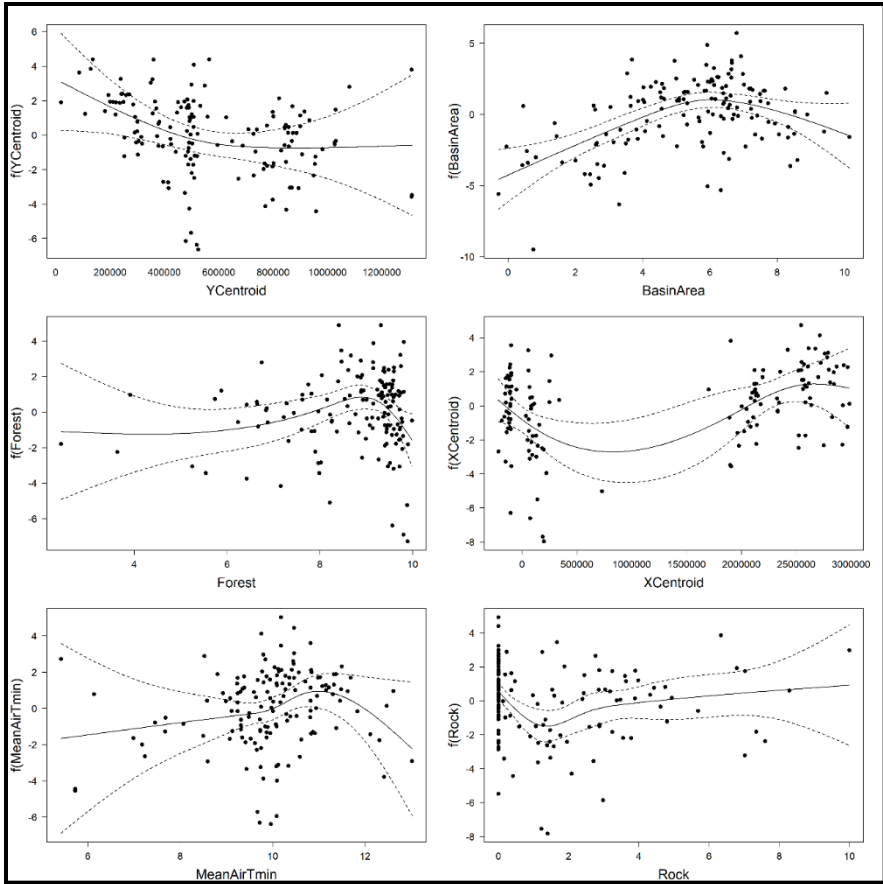


Figure 2-2. Splines for the predictors used in the ALL-GAM model (e.g., MaxWaterTmax).

4.3 Delineation of thermally homogeneous regions by HCA, CCA and ROI

Giving the example of the MLR, for the variables MaxWaterTmax and MaxNumDay, three homogeneous regions have been identified, whereas only two groups of thermally homogeneous stations were found for the parameters of the Gaussian function (Gaussian_a, Gaussian_b and Gaussian_c). In the case of GAM (not shown here), three homogeneous groups were identified for the variable MaxNumDay. Only two regions were distinguished for the metrics MaxWaterTmax, Gaussian_a, Gaussian_b and Gaussian_c. Figure 2-3 provides an example of a dendrogram for the variable MaxWaterTmax in the case of MLR. When the truncation was performed at a Euclidean distance of 15.3, three thermally homogeneous regions were identified.

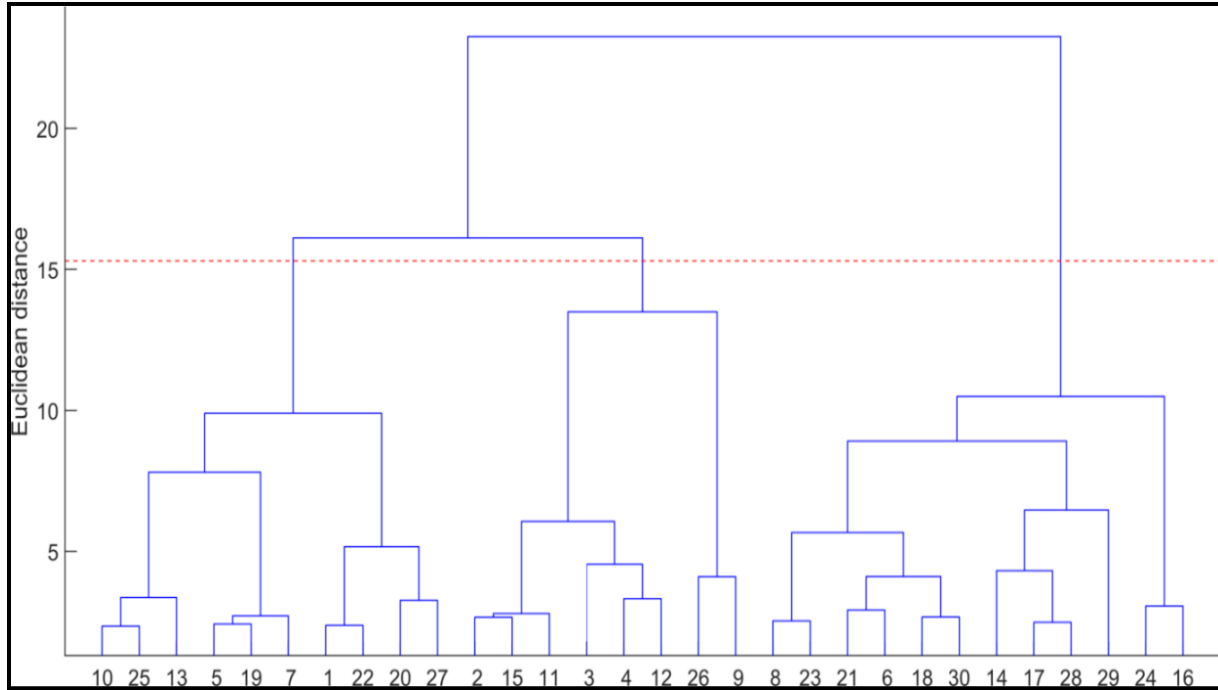


Figure 2-3. Dendrogram with cut-off threshold for the variable MaxWaterTmax. Note that station numbers are shown on the x-axis.

The spatial distribution of homogeneous groups generated by the HCA for each water temperature metric is illustrated in Figure 2-4. According to these maps, the distribution of homogeneous regions depends on the metric of interest. For the variable MaxWaterTmax, three regions were identified with one region comprising most of stations in Newfoundland and Labrador (Northeastern part of the study area). The other two regions are located in the Southwestern part, one of which includes stations in the Gaspé Peninsula and the Quebec North Shore, while the other region is located in New Brunswick and Nova Scotia with a few stations located in Newfoundland. The variable MaxNumDay also shows three thermally homogeneous regions. The stations in Newfoundland and Labrador form a homogeneous group. A second region includes stations located on the Gaspé Peninsula and the Quebec North shore. The third region is mostly located in New Brunswick and Nova Scotia. For parameters a and c of the Gaussian function, most of the stations of the first group are located in Newfoundland and Labrador with the addition of some stations in Nova Scotia and Prince Edward Island for parameter c . whereas, for parameter b , a first group is located in Newfoundland and Labrador and a second is located in New Brunswick-Gaspé Peninsula with a few

stations located in Labrador. These results indicate that there are distinct thermal regions in Eastern Canada and that for the most part, they can be defined geographically, except for a few stations. For the other two delineation methods CCA and ROI, the optimal neighborhood size was fixed at 97 and 90, respectively. Overall, ROI and CCA show practically the same neighborhoods to a reference site.

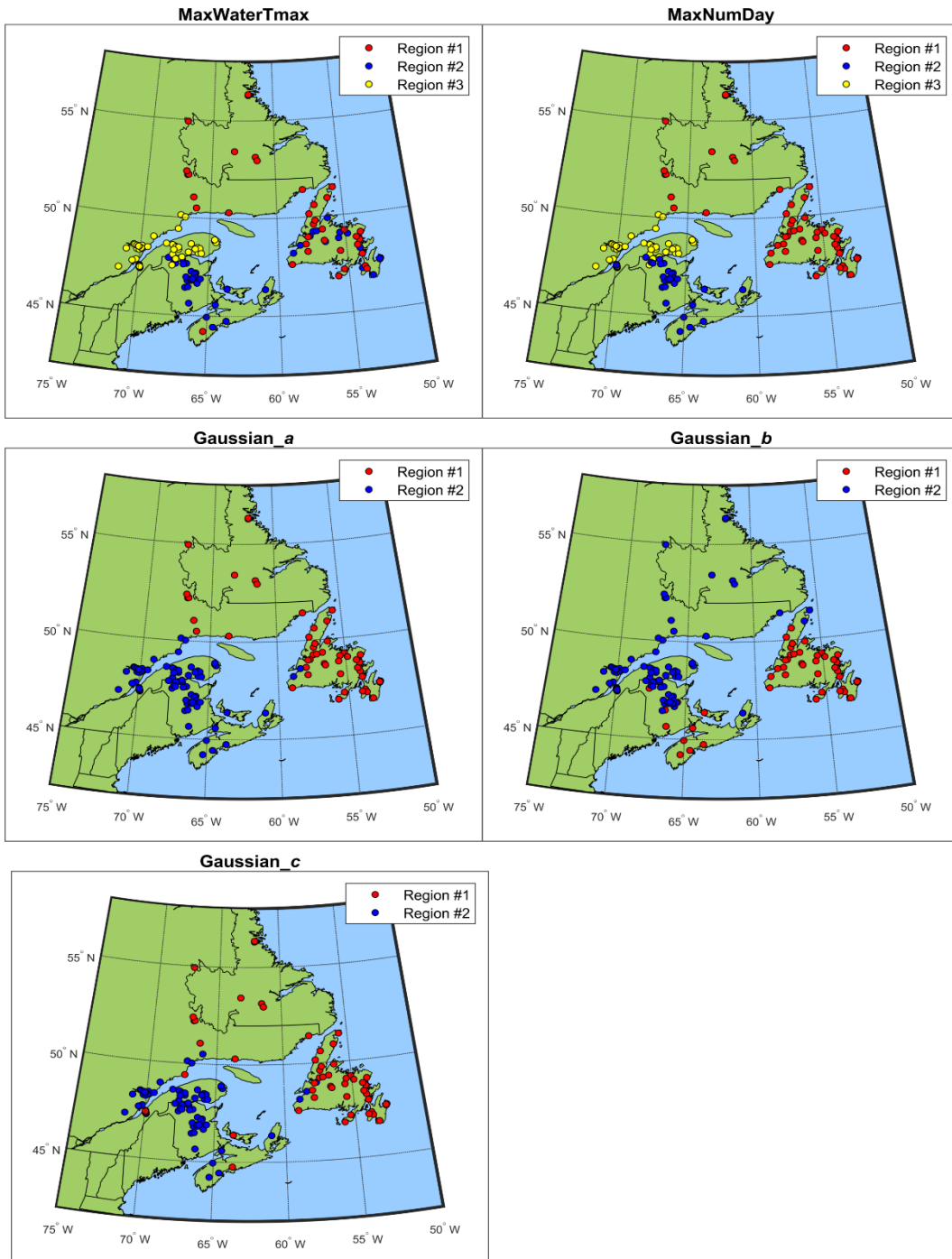


Figure 2-4. Maps of stations grouped using HCA-MLR for each thermal metric.

4.4 Comparative performance analysis

The best model performance was the one that showed adequate scores for the selected performance indicators, namely, R^2 value closer to 1, RMSE, RRMSE, BIAS and PBIAS closer to 0. In this subsection, the results of the different combinations of delineation methods and statistical regional estimation models is carried out. The performance criteria obtained from the cross-validation analysis are presented in Table 2-5.

Performance criteria	metrics	ALL+MLR	HCA+MLR	CCA+MLR	ROI+MLR	ALL+GAM	HCA+GAM	CCA+GAM	ROI+GAM
R²	MaxWaterTmax	0.38	0.48	0.37	0.42	0.39	0.39	0.42	0.34
	MaxNumDay	0.33	0.48	0.37	0.49	0.43	0.55	0.47	0.40
	Gaussian_a	0.33	0.45	0.37	0.42	0.43	0.59	0.45	0.39
	Gaussian_b	0.67	0.70	0.69	0.70	0.70	0.75	0.68	0.67
	Gaussian_c	0.39	0.42	0.35	0.39	0.40	0.48	0.33	0.36
Bias	MaxWaterTmax (°C)	-0.00	-0.08	0.33	-0.11	0.01	-0.05	0.17	0.03
	MaxNumDay (Days)	0.11	0.12	-0.05	0.08	0.08	0.09	0.05	0.11
	Gaussian_a (°C)	-0.01	-0.07	0.30	-0.08	-0.00	-0.05	0.17	-0.09
	Gaussian_b (Days)	0.00	0.14	-0.30	-0.06	-0.04	0.40	-0.39	0.27
	Gaussian_c (Days)	0.01	0.13	-0.63	0.21	-0.04	0.05	-0.78	-0.26
RMSE	MaxWaterTmax (°C)	2.54	2.31	2.55	2.44	2.51	2.50	2.45	2.61
	MaxNumDay (Days)	1.82	1.60	1.76	1.58	1.68	1.49	1.62	1.72
	Gaussian_a (°C)	2.24	2.03	2.17	2.08	2.06	1.74	2.02	2.14
	Gaussian_b (Days)	4.62	4.36	4.44	4.36	4.39	4.01	4.51	4.62

	Gaussian_c (Days)	4.26	4.12	4.40	4.25	4.22	3.91	4.46	4.35
PBias (%)	MaxWaterTmax	-0.01	-0.34	1.40	-0.74	0.05	-0.21	0.70	0.13
	MaxNumDay	10.89	11.35	-4.67	7.57	7.64	8.53	5.15	10.17
	Gaussian_a	-0.04	-0.36	1.62	-0.42	-0.00	-0.30	0.93	-0.51
	Gaussian_b	0.01	0.25	-0.52	-0.10	-0.07	0.70	-0.69	0.48
	Gaussian_c	0.00	0.06	-0.30	0.10	-0.02	0.02	-0.37	-0.12
RRMSE (%)	MaxWaterTmax	10.90	9.92	10.95	10.47	10.78	10.74	10.51	11.20
	MaxNumDay	172.76	151.96	167.47	150.19	159.46	141.92	153.86	163.27
	Gaussian_a	12.17	11.00	11.76	11.31	11.21	9.47	10.96	11.61
	Gaussian_b	8.10	7.64	7.79	7.65	7.69	7.03	7.91	8.10
	Gaussian_c	1.99	1.93	2.06	1.99	1.97	1.83	2.08	2.04

Table 2-5: Performance indicators for different regional thermal analysis in Atlantic salmon rivers.

Note: best results are in bold character.

It is clear from the table above that the GAM model applied to all stations in our study area without delineation of thermally homogeneous regions lead to a good performance according to the R^2 , the RMSE and the relative RMSE in comparison to the models using the MLR. This result mirrors the results obtained from the spline curves in Figure 2-2 indicating that a non-linear model is more suitable than the MLR model. Also, when the GAM model is tested in conjunction with the various regional delineation approaches, particularly with the HCA and CCA approaches, better performance has been obtained compared to the approach where GAM is applied to the whole study area. The HCA technique in conjunction with both MLR and GAM statistical models shows a higher R^2 and a lower RMSE compared to those provided by ROI, CCA or with all the stations pooled together. For instance, for the parameter a of the Gaussian function, R^2 is 59% with HCA+GAM, while it is 39% with ROI+GAM and 43% with ALL+GAM. RMSE is 1.74°C with HCA+GAM, while it is 2.14°C with ROI+GAM and 2.06°C with ALL+GAM. Therefore, the delineation method that benefits the most from the introduction of GAM is the HCA, where the performance obtained is comparable or superior to that of CCA+GAM. However, defining regions by HCA seems to give a bias that is slightly higher than using a single region including all stations.

The increase of the number of stations by relaxing the criterion related to the length of time series from five to four years resulted in slightly higher RMSEs and lower coefficient of determination R^2 .

The scatter plots in Figure 2-5 present the observed values versus the estimated values for all models using cross validation. They illustrate that, in most cases, the bias is more pronounced for the extreme values of the temperature metrics. This is notably the case for MaxNumDay and the parameter c of the Gaussian function.

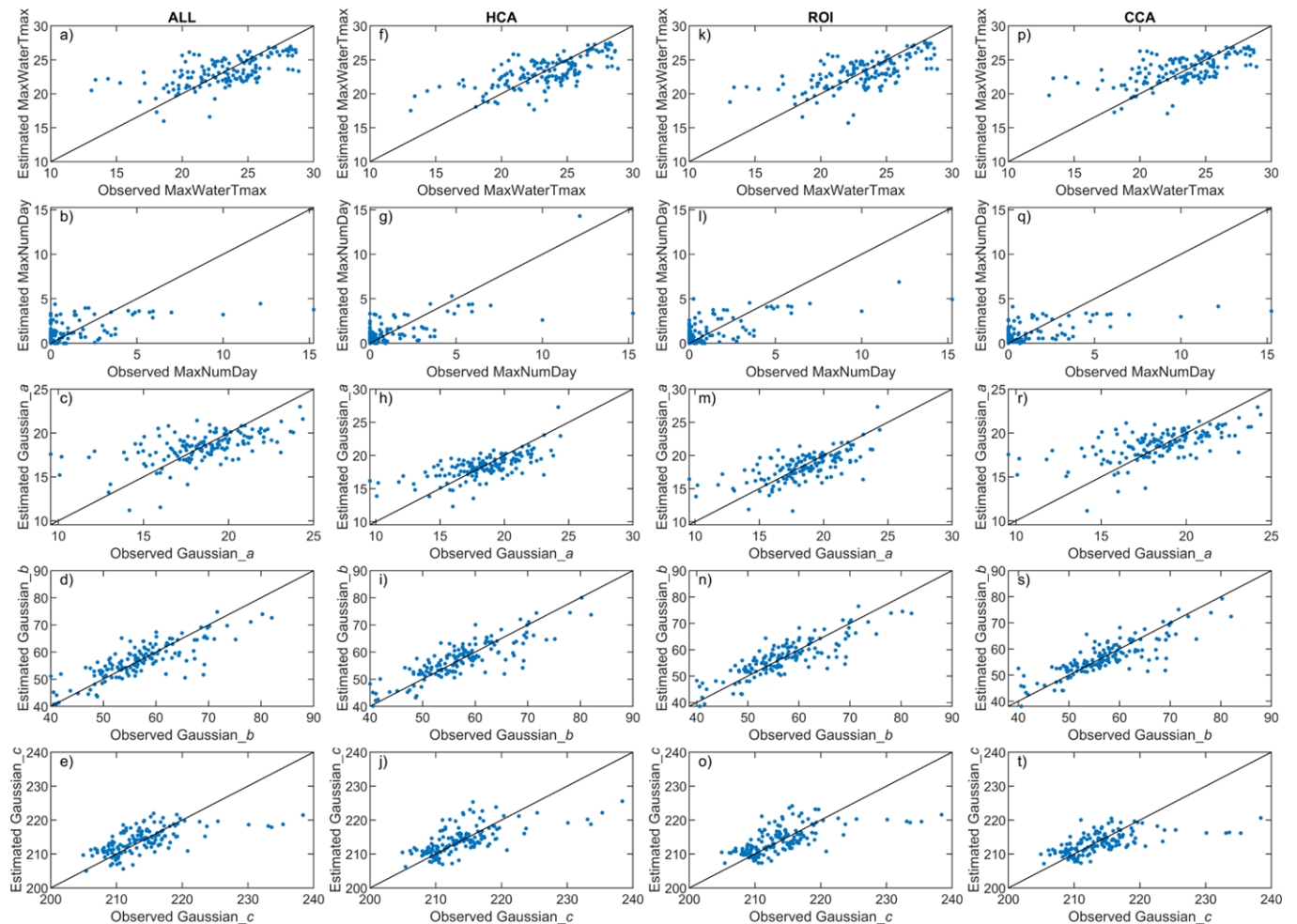


Figure 2-5: Scatter plots of observed vs simulated water temperature metrics for MLR.

Giving the example of the variable MaxNumDay, this variable is underestimated for values > 5 days. This could be attributed to the inequal distribution of this variable where most of the data are comprised between 0 and 3 days. In fact, 50 % of stations have a value of MaxNumDay = 0 (Figure 2-6). As well, data from

some thermal stations reveal detectable lake influences resulting in increasing of the maximum number of consecutive days with maximum water temperature $> 25\text{ }^{\circ}\text{C}$ and minimum water temperature $> 20\text{ }^{\circ}\text{C}$. As for the parameter c of the Gaussian function, beyond 220 days, this variable is underestimated. Again, Figure 2-6 shows that there are very few values > 220 days.

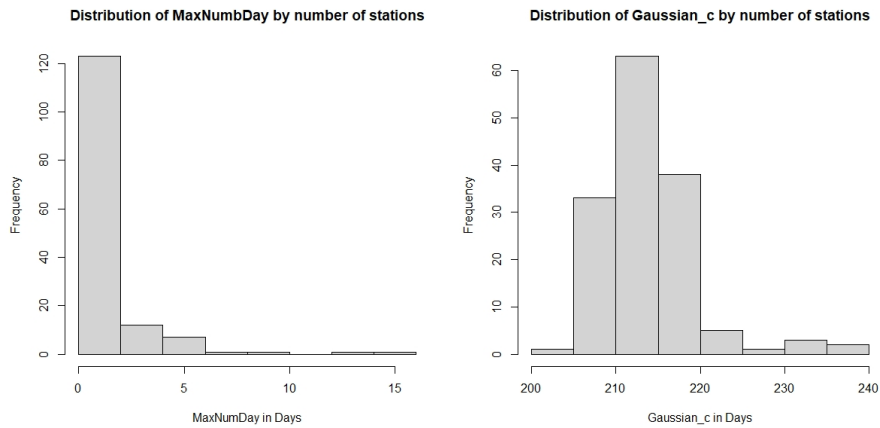


Figure 2-6: Histograms showing the frequency distribution of the variables MaxNumDay and the parameter c of the Gaussian function.

5 Discussion

The RivTemp database includes a large number of stations with only one to four years of data. In the present study, the selection criterion for inclusion of water temperature stations was a minimum of four years of data. This criterion has led to selecting more stations in Newfoundland and less in Nova Scotia and western Quebec. Despite this limitation, our analyses allowed for the definition of three thermally homogeneous regions, while potentially having a sufficiently large number of stations in each region to produce models with acceptable uncertainties. Given the high density of stations in Newfoundland and Labrador and New Brunswick, it is not surprising that the HCA separated the stations into groups primarily located in Newfoundland and Labrador, Gaspé Peninsula and New Brunswick. As station density improves in the study area, HCA may be used again to refine the delineation of thermal regions.

Nevertheless, the present finding is inconsistent with the results of previous studies (e.g., Makarowski, 2009; Chu and Jones, 2010) who claimed that, at the landscape scale, existing broad spatial classification

such as ecoregions have been found inadequate to capture spatial variability in stream water temperature. Also, the cross-validation results provide crucial insights into the relative strength of the GAM model and the need for more complex non-linear model for assessing thermal variability. Indeed, for a given delineation method, as well as for a regional estimation model applied to the entire study area, better performance is generally obtained with the GAM instead of MLR. This result mirrors the results of previous studies carried out on hydrological variables, in particular low flow quantiles (see for example Ouarda et al., 2018). This is basically due to the fact that this statistical model offers the possibility of modeling non-linear dependencies existing between the response variables and predictors.

Overall, the best results are obtained with HCA+GAM and CCA+GAM, as these two combinations lead to the best performance indices with reference to the coefficient of determination (R^2), the absolute and relative root mean square error (RMSE and RRMSE). Fixed regions approach (HCA) in conjunction with the two regionalization models MLR and GAM (HCA+GAM and HCA+MLR) present better performance than the neighborhood-based approaches (ROI+MLR, CCA+MLR, ROI+GAM and CCA+GAM).

This result is different than those obtained in previous RFA studies on hydrological variables (e.g., floods and low flows), which have shown that the neighborhood-based approaches lead to better performances than the non-contiguous fixed regions approach (see for example, Ouarda et al., 2008). Also, MaxNumDay a variable of high importance for Atlantic salmon, is not modelled adequately with RRMSE values of more than 100%. This can be attributed to the small sample size and the small number of stations with high values. In fact, many stations have a value of 0 for this metric. Model performance for this key metric will likely improve as the sample size increases and stations with high MaxNumDay values are included.

Besides catchment and climatic drivers, lake effects were also detected. In our study, all lakes are natural, i.e., there are no dams at their outlet. Regulated rivers can be thermally different than systems with natural lakes and thus, they would need to be treated separately. Eventual effort should be made to quantify the difference between the two regimes. Accordingly, it would be interesting in the future to introduce other

variables which differentiate reservoirs from natural lakes (e.g., morphometry characteristics, water level, flow at the outlet).

In general, the models error originates partly from the lack of appropriate field data. Several potential limitations of the GAM model should also be pointed out. The later may be fitting overly complex curves (smooth functions), which decreases the degrees of freedom to the model, making it relatively more complex. Also, the failure to identify and incorporate relevant interactions can be a major constraint. Solution could be provided by more complex statistical models such as the Bayesian optimal model proposed by Seidou et al. (2006). The latter has been defined as a parametric Bayesian combination of local and regional information in flood frequency analysis. It has been proven to be a promising and effective tool for the estimation of quantiles at sites with short to medium length flood records. Also, some of the selected predictors were found to have a strong relationship with the response variable. However, with the GAM model it is difficult to formally integrate the interactions between variables (see for example, Ramsay et al., 2003). There is a need for more sophisticated statistical modeling approach to overcome these shortcomings in future studies.

Conclusion

The current findings add to a growing body of literature on water temperature modeling, in which the RTA approach is tested to estimate five thermal metrics known to be relevant for Atlantic salmon and related to maximum temperature, its date of occurrence, as well as the seasonal variability. As the assumption of linearity is not always met, the GAM model was introduced and compared with the simple MLR model. It should be seen as a good technique, especially, because of its ability to use more predictor variability more optimally for a given water temperature metric than the simple MLR model.

The results of the current work confirmed those of previous studies regarding the performance of the GAM model for information transfer. Thus, using the GAM model instead of the simple multiple linear regression model MLR slightly improves model performance, especially for the parameters of the Gaussian function.

The delineation of homogeneous regions has proved to be central for RTA in Eastern Canada, as it leads to significant model improvement. As expected, climatic forcing was found to be the major driver of water temperature. While the linear model suggests a strong influence of air temperature on the prediction of the metrics of interest, GAM model demonstrated higher predictive power of the variables YCentroid (akin to latitude) and air temperature for predicting the metric of interest. Of the eight candidate models, the one using HCA for grouping stations and GAM for metric estimation had the best performance. When the GAM model is used to estimate the temperature metrics, only two regions were identified for four out of five thermal metrics. This could be explained by the fact that large sample sizes (i.e., station-years) are required to adequately fit a GAM model. In contrast, when the MLR model is used, three regions were identified for two out of five water temperature metrics. Consequently, as the database increases, these analyses will need to be revisited.

Such information represents a pertinent tool for managers and other groups concerned with the conservation and management of valuable stocks of Atlantic salmon. The results of this work will help planning to support the conservation of ectotherm aquatic species through:

- Identifying the environmental factors governing the thermal regime of salmon river. Then, restoration efforts to reduce river temperatures (e.g., restoration of riparian vegetation and channel narrowing as in Justice et al. (2017)) would improve the chances of successful spawning and would mitigate impacts of a future with warmer temperatures.
- Identifying the potential relatively homogeneous thermal regions for which common management tools and guidelines can be developed.

This study opens new horizons to continue to investigate potential predictors for thermal metrics. Likewise, more sophisticated models can be tested. Among these models, the non-parametric multivariate adaptive regression splines method (MARS) (e.g., Bond et al., 2017; Roy et al., 2018; Msilini et al., 2020), the more advanced machine learning technique such as Artificial Neural Networks ANN (e.g., Piotrowski et al., 2015) as well as Random Forest RF (e.g., Desai and Ouarda, 2021) are highly recommended to be

investigated. Also, modeling of other thermal metrics such as degree-days, extremes, and peaks over threshold (recently proposed by Caissie et al., 2019), could provide valued information for the management of freshwater ecosystems.

Considering that the GAM model performance improves in response to the increase in the size of the potentially homogenous groups, further studies are recommended for deeper investigation regarding the impacts of these models' performance when they are adjusted with different station densities. Besides, the success of such an approach could be expanded to other regions and would provide a practical tool to assist in the assessment of anthropogenic impacts on instream aquatic habitat. As well, the extension of the regional models compared in the present study to the multivariate case is also of great importance to carry out the estimation of a number of temperature metrics at the same time taking into consideration the linkage between these metrics. Considering the effect of lake on climatic variables and, therefore, on river water temperature, further studies should focus on investigating their potential impact.

In addition to the selected predictors in our study, several other factors controlling spatial and temporal variability of river water temperature could be investigated in future work to potentially offer increased accuracy and reliability in the predictions. They include, riverbed conditions, riparian vegetation, fluvial topography, drainage network characteristics, groundwater input and stream orientation (e.g., Caissie, 2006; Hester and Doyle, 2011; Kelleher et al., 2012; Lisi et al., 2015; Piccolroaz et al., 2016; Garner et al., 2017; Cai et al., 2018).

Studies have shown that in many parts of North America, fish are already experiencing warm episodes with temperatures approaching their upper lethal limit (Sinokrot et al., 1995; Eaton et al., 1995). In Eastern Canada, it was also estimated that climate change could result in an overall loss of juvenile Atlantic salmon habitat (Minns et al., 1995). The future of ectotherm aquatic species, under those circumstances, remains a focal point of researchers taking into consideration the challenges it raises. A good grasp of river water

temperature modeling techniques such as the ones developed in the present study is essential in the management of these resources in order to ultimately address global warming issues.

Credit author statement

Olfa Abidi: Conceptualization, Methodology, Formal analysis, Visualization, Writing-original draft; André St-Hilaire: Conceptualization, Supervision, Writing-review and editing; Taha Ouarda: Conceptualization, Supervision, Writing-review and editing; Christian Charron: helped in the implementation of the computer code; Claudine Boyer and Anik Daigle: Data collection.

Acknowledgments

This project would not have been possible without the tremendous effort that went into data collection by numerous Canadian agencies. This work was funded by NSERC Discovery Grant 2019-06701. Also, it was supported in part by the MUTAN (Mission Universitaire de Tunisie en Amérique du Nord).

Data Statement

All data used in this project are freely available and accessible. To begin with, the water temperature data used in this project were extracted from the RivTemp database and are already accessible via the RivTemp website (www.rivtemp.ca). In addition, the physiographic data originate from different public sources (e.g., the geological survey of Canada and the consortium for spatial information (CGIAR-CSI) and database from provincial departments of natural resources). The meteorological data used are daily air temperature and precipitation measurements interpolated on a 10km×10km grid (ANUSPLIN, Hutchinson et al., 2009) available upon request.

3 ANALYSE SUPPLÉMENTAIRE: ÉTUDE COMPARATIVE ENTRE MARS ET GAM

3.1 Introduction

Il est bien évident que le modèle GAM est un outil de modélisation représentant les meilleures performances par rapport aux méthodes classiques linéaires. Cependant, ce dernier présente un certain nombre d'inconvénients. En effet, il est difficile d'adapter le GAM à des bases de données complètes en raison des limitations de mémoire imposées par les complexités numériques de cette approche (Leathwick et al., 2006). Ces contraintes sont liées à l'ajustement des courbes de lissage (fonctions lisses) trop complexes. De plus, l'incapacité à identifier et à intégrer les interactions pertinentes entre les variables physiographiques au sein du bassin versant a été démontrée (Ramsay et al., 2003).

Dans cette section, on vise à présenter un outil de modélisation qui a gagné une popularité surtout en analyse fréquentielle régionale. L'approche des splines de régression adaptatives multivariées MARS est reconnue par sa performance significative. Le but de l'approche MARS est de surmonter certaines des limitations associées aux méthodes d'estimation décrites précédemment. Notamment, le nombre élevé des variables explicatives nécessaires pour décrire les processus de manière plus proche de la réalité, la non-linéarité ainsi que l'interaction entre les prédicteurs.

Cette approche a été présentée par Friedman (1991) comme une approche de régression non paramétrique flexible capable de modéliser des données de haute dimension (un grand nombre de variables d'entrées). Il a plus de pouvoir et de flexibilité pour modéliser des relations qui sont presque additives ou impliquent des interactions dans aux plus quelques variables. Elle a été appliquée avec succès dans divers domaines tels que l'écologie et l'environnement (par exemple, Leathwick et al., 2005; Leathwick et al., 2006; Balshi et al., 2009; Bond et al., 2017), la géologie (Zhang et al., 2015; Zhang et al., 2016), la finance (Lee et al., 2005; Lee et al., 2006), l'énergie (Li et al., 2016; Roy et al., 2018) et l'hydrologie (par exemple, Emamgolizadeh et al., 2015; Kisi, 2015; Kisi et al., 2016; Bond et al., 2017; Deo et al., 2017). Cette méthode peut facilement inclure des interactions entre les variables, ce qui permet de considérer n'importe quel degré d'interactions (Lee et al., 2006).

De point de vue structure, le modèle MARS prend la forme d'une combinaison linéaire des fonctions de base et de leurs interactions de la façon suivante:

$$f(X) = \beta_0 + \sum_{i=1}^N \beta_m BF(X) + \text{interactions} \quad (16)$$

Avec β_0 représente l'intercepte (ordonnée à l'origine), $BF(X)$ sont les fonctions de base et β_m est le coefficient des fonctions de base.

Les effets d'interactions peuvent être obtenus par le produit tensoriel des fonctions de base $BF(X)$. Les fonctions considérées sont des segments linéaires par morceaux, appelés splines linéaires. Ils sont définis par paires et séparés par un nœud représentant un point d'inflexion le long de la plage d'une variable explicative donnée. Selon Rounaghi et al. (2015), ces fonctions sont définies par:

$$(t - x)^- = \begin{cases} t - x & \text{si } t > x \\ 0 & \text{sinon} \end{cases} \quad (17)$$

$$(x - t)^+ = \begin{cases} x - t & \text{si } x > t \\ 0 & \text{sinon} \end{cases} \quad (18)$$

Avec t qui est la position des nœuds et x qui représente la variable explicative.

Récemment, cette approche a été testée par Msilini et al. (2020) dans le contexte RFA. Les auteurs ont montré que cette approche d'estimation régionale a un pouvoir prédictif significatif permettant une meilleure représentation du processus hydrologique. La même conclusion a été tirée par Leathwick et al. (2006).

3.2 Méthodologie

Nous avons d'abord voulu identifier les variables explicatives possédant le pouvoir de prédiction le plus important et pouvant influencer les métriques de la température de l'eau. Pour sélectionner l'ensemble des prédicteurs, une procédure de sélection pas à pas à deux phases en avant puis en arrière (Forward pass /Backward pass) a été effectuée. Cette procédure est différente de celle du modèle GAM qui est basée plutôt sur une procédure de sélection pas à pas en avant (Forward pass). Ainsi, l'application de l'approche MARS nécessite l'optimisation de trois paramètres principaux. Notamment, le nombre maximum de termes dans le modèle en phase forward (N_k), le degré d'interaction (degré), et le nombre maximum de termes en phase backward (N_{prune}). Une gamme de valeurs de ces paramètres a été testée puis évaluée afin de les optimiser en fonction des critères des modèles ajustés suivants : la validation croisée généralisée (GCV), la somme résiduelle des carrés (RSS) et le coefficient de détermination (R^2).

La figure suivante est un schéma explicatif décrivant la procédure de sélection pour les deux modèles MARS et GAM.

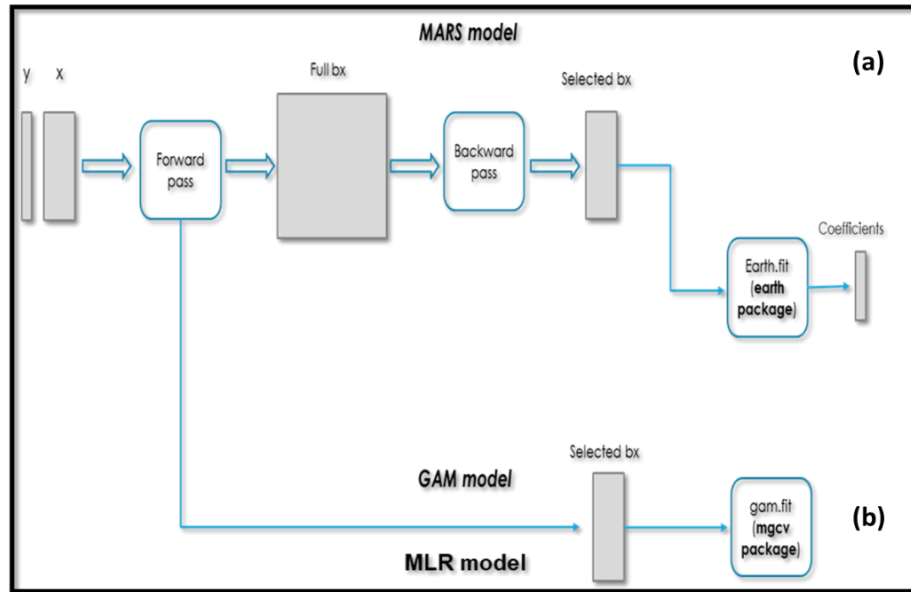


Figure 3-1. Procédure de sélection des variables explicatives pour les deux modèle MARS (a) et GAM (b).

Dans la présente étude, le modèle MARS a été ajusté à l'aide du package "**earth**" dans R.

3.3 Résultats

3.3.1 Sélection des variables explicatives pour le modèle MARS

Tenant en compte du fait que la procédure de sélection des variables explicatives est différente pour les deux modèles MARS et GAM, on s'attend à avoir une différence pour certains prédicteurs pour chaque métrique. Pour pouvoir comparer la performance des modèles testés, on a essayé de sélectionner presque les mêmes variables dans les deux cas.

Le tableau ci-dessous illustre l'ensemble des variables indépendantes sélectionnées pour chaque métrique de température de l'eau pour les deux modèles d'estimation MARS et GAM.

3.3.2 Validation

Modèle	Métrique	Variables sélectionnées					
GAM	MaxWaterTmax	Ycentroid	BasinArea	Forest	Xcentroid	MeanAirTmin	Rock
	MaxNumDay	Ycentroid	FluvioGlacialDeposists	LakeArea	MaxAirTmax	Xcentroid	MinAirTmin
	Gaussian_a	Ycentroid	LakeArea	Slope	MinAirTmin	Rock	Forest
	Gaussian_b	MeanAirTmin	Xcentroid	BasinArea	MinAirTmin	Forest	Wetland
	Gaussian_c	Ycentroid	BasinArea	Forest	ElevationStation	Slope	GlacialDeposists
MARS	MaxWaterTmax	YCentroid	BasinArea	MeanAirTmin	Forest	Rock	XCentroid
	MaxNumDay	YCentroid	LakeArea	MaxAirTmax	XCentroid	MinAirTmin	
	Gaussian_a	LakeArea	YCentroid	Slope	Rock	MinAirTmin	Forest
	Gaussian_b	MeanAirTmin	XCentroid	BasinArea	Wetland	Forest	
	Gaussian_c	GlacialDeposits	BasinArea	TotPrecip	ElevationStation	Forest	MaxElevation

Tableau 3-1. Variables explicatives sélectionnées pour chaque métrique pour les deux modèles GAM et MARS.

Les performances sont évaluées à l'aide d'une validation croisée. La comparaison entre les deux modèles a été basée sur les critères suivants : le coefficient de détermination R^2 , l'erreur quadratique moyenne (RMSE), le Biais, l'erreur quadratique moyenne relative (RRMSE) et le pourcentage de biais (PBias). D'après les résultats de la validation croisée :

- Avec toutes les stations sans délimitation des régions homogènes, les deux modèles MARS et GAM ont des performances comparables.
- La méthode de délimitation des régions non contiguës HCA améliore le plus le pouvoir de prédiction des estimations qu'avec les méthodes de voisinage CCA et ROI.

Avec la technique HCA, pour certaines variables thermiques, il était possible de regrouper les sites en trois et quatre régions qui ont un comportement relativement similaire. Seulement deux régions potentiellement homogènes ont été retenues en fonction des performances mesurées qui étaient les meilleures au sein de deux régions.

Pour les variables MaxWaterTmax et les trois paramètres du modèle gaussien, le modèle MARS surpasse légèrement le modèle GAM en termes de R², RMSE, RRMSE et PBias. Alors que, pour la variable MaxNumDay, le modèle GAM surpasse légèrement le modèle MARS. Au sein des régions définies par les méthodes de voisinage CCA et ROI, MARS couplé à la méthode CCA a une performance légèrement supérieure que celle obtenue avec la méthode ROI. Le tableau 3-2 illustre les résultats de la validation croisée pour les différentes combinaisons testées.

Critères de performance	Métrique	ALL		HCA		CCA		ROI	
		GAM	MARS	GAM	MARS	GAM	MARS	GAM	MARS
R ²	MaxWaterTmax	0.39	0.38	0.39	0.46	0.42	0.48	0.34	0.38
	MaxNumDay	0.43	0.46	0.55	0.50	0.47	0.49	0.40	0.51
	Gaussian_a	0.43	0.36	0.59	0.57	0.45	0.43	0.39	0.25
	Gaussian_b	0.70	0.72	0.75	0.75	0.68	0.70	0.67	0.63
	Gaussian_c	0.40	0.46	0.48	0.55	0.33	0.37	0.36	0.40
Bias	MaxWaterTmax (°C)	0.01	0.03	-0.05	-0.15	0.17	0.16	0.03	-0.06
	MaxNumDay (JOUR)	0.08	-0.01	0.09	0.10	0.05	0.03	0.11	0.13
	Gaussian_a (°C)	-0.00	-0.08	-0.05	0.03	0.17	0.22	-0.09	-0.01
	Gaussian_b (JOUR)	-0.04	0.18	0.40	0.25	-0.39	-0.36	0.27	0.09
	Gaussian_c (JOUR)	-0.04	-0.04	0.05	0.10	-0.78	-0.59	-0.26	-0.21
RMSE	MaxWaterTmax (°C)	2.51	2.53	2.50	2.36	2.45	2.31	2.61	2.54
	MaxNumDay (JOUR)	1.68	1.64	1.49	1.58	1.62	1.60	1.72	1.56
	Gaussian_a (°C)	2.06	2.18	1.74	1.79	2.02	2.07	2.14	2.36
	Gaussian_b (JOUR)	4.39	4,22	4.01	4.00	4.51	4.35	4.62	4.89
	Gaussian_c (JOUR)	4.22	4.01	3.91	3.65	4.46	4.30	4.35	4.22
PBias (%)	MaxWaterTmax	0.05	0.10	-0.21	-0.60	0.74	0.70	0.13	-0.30
	MaxNumDay	7.64	-1.30	8.53	8.50	5.15	3.20	10.17	12.20
	Gaussian_a	-0.00	-0.40	-0.30	0.10	0.93	1.20	-0.51	-0.10
	Gaussian_b	-0.07	0.30	0.70	0.40	-0.69	-0.60	0.48	0.20
	Gaussian_c	-0.02	0.00	0.02	0.00	-0.37	-0.30	-0.12	-0.10
RRMSE (%)	MaxWaterTmax	10.78	10.86	10.74	10.16	10.51	9.92	11.20	10.90
	MaxNumDay	159.46	155.27	141.92	149.76	153.86	151.42	163.27	148.02
	Gaussian_a	11.21	11.82	9.47	9.69	10.96	11.23	11.61	12.82

Gaussian_b	7.69	7.39	7.03	7.01	7.91	7.62	8.10	8.57
Gaussian_c	1.97	1.87	1.83	1.71	2.08	2.01	2.04	1.97

Tableau 3-2. Résultats de la validation croisée pour les différentes combinaisons.

Graphiquement, les fonctions de lissage ajustées à certaines variables explicatives avec les deux modèles MARS et GAM (Figure 3-2) montrent que les splines linéaires ajustées à l'aide du modèle MARS se rapprochent étroitement des splines générées par le modèle GAM.

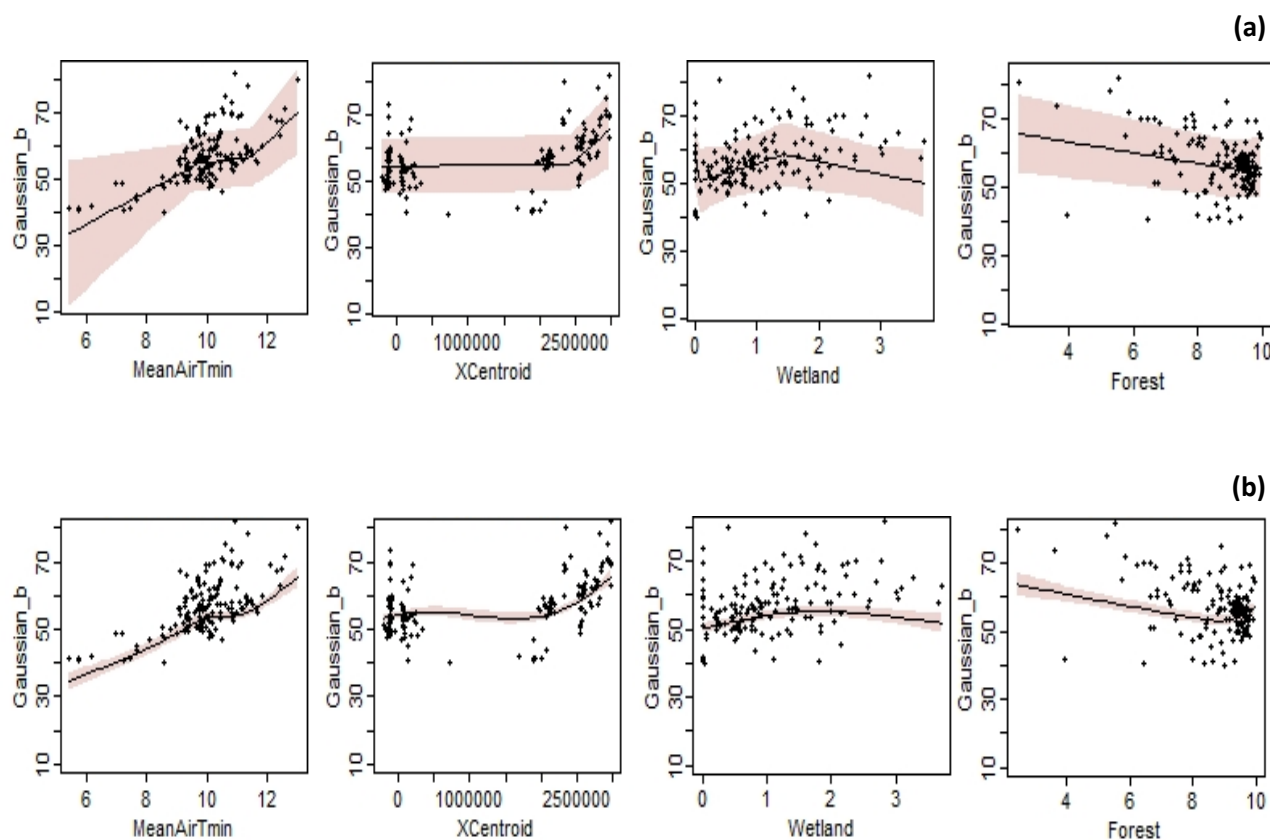


Figure 3-2. Exemples d'ajustements de certaines variables explicatives dans les deux cas : (a) cas du modèle MARS, (b) cas du modèle GAM.

La figure suivante illustre le diagramme des valeurs estimées en fonction des valeurs observées pour chaque métrique thermique dans le cas du modèle MARS. Selon cette figure, on peut voir que, pour les métriques MaxNumDay et le paramètre c du modèle gaussien, le biais est plus prononcé pour les valeurs extrêmes que pour les valeurs centrales.

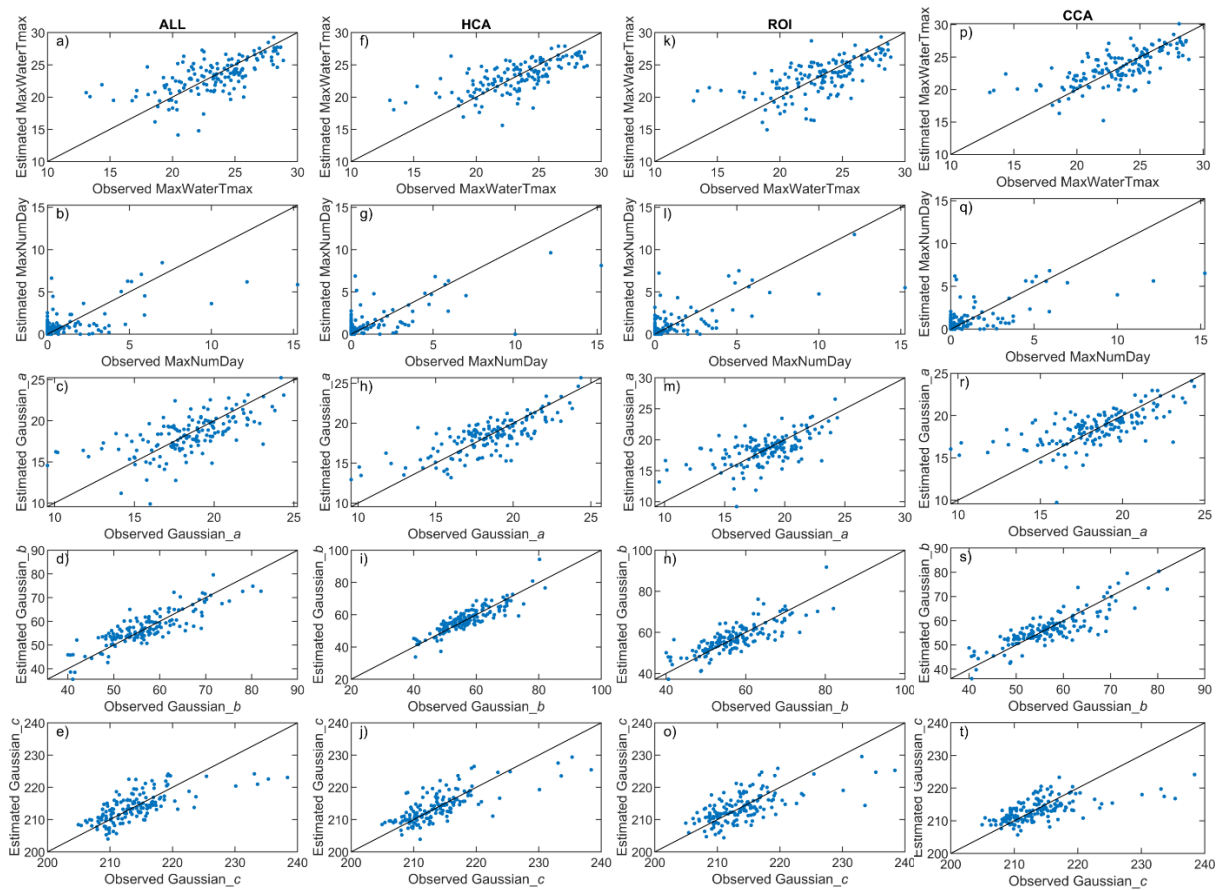


Figure 3-3. Diagrammes des valeurs estimées en fonction des valeurs observées pour chaque métrique thermique dans le cas du modèle MARS.

Les diagrammes combinés présentant les valeurs estimées et observées pour les variables MaxNumDay et le paramètre gaussien c (Figure 3-4) expliquent, en grande partie, la raison pour laquelle le modèle MARS exprime un biais et RMSE relativement plus élevés. En effet, on peut noter l'abondance des valeurs faibles avec peu de valeurs élevées qui sont estimées avec une erreur élevée.

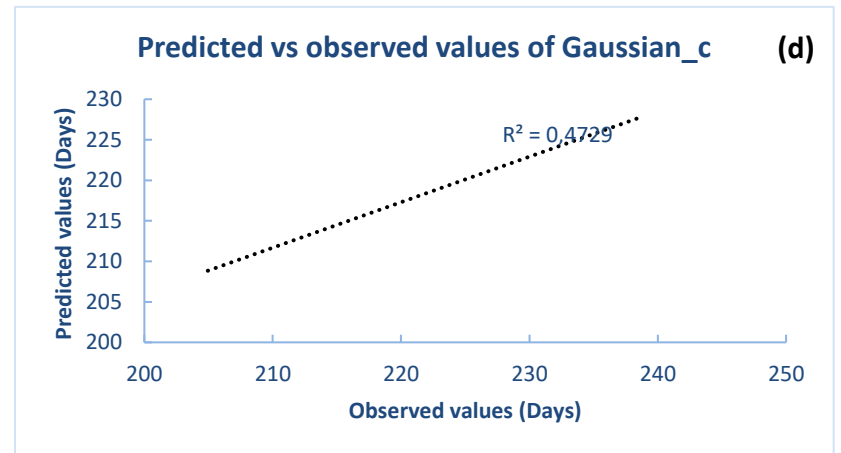
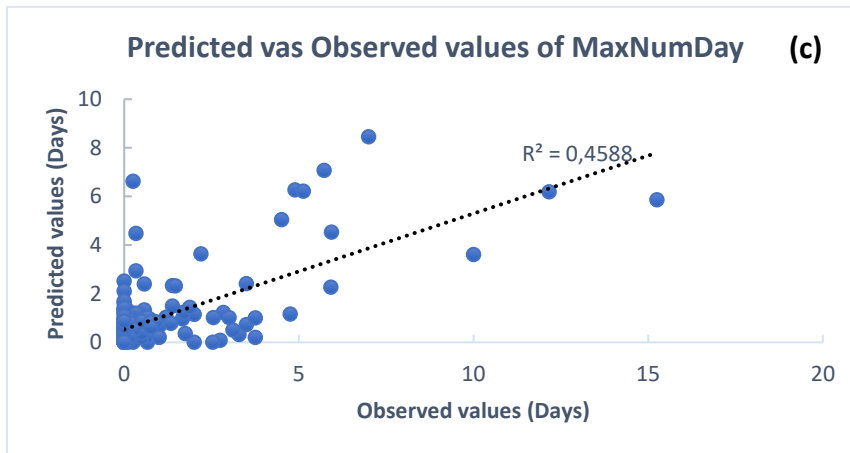
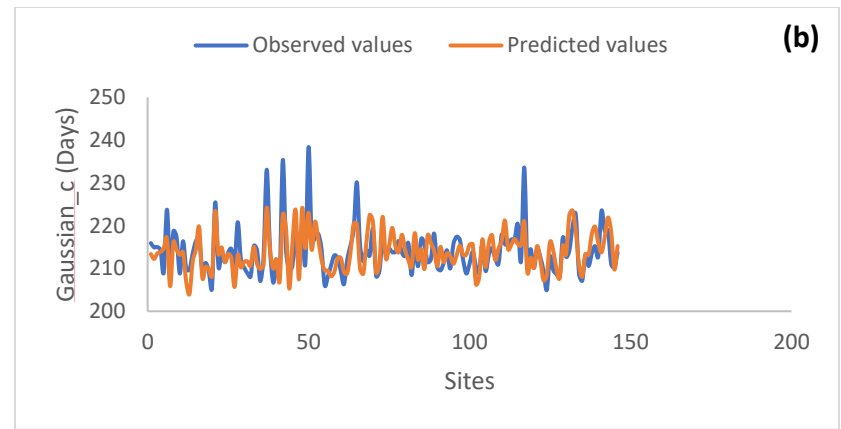
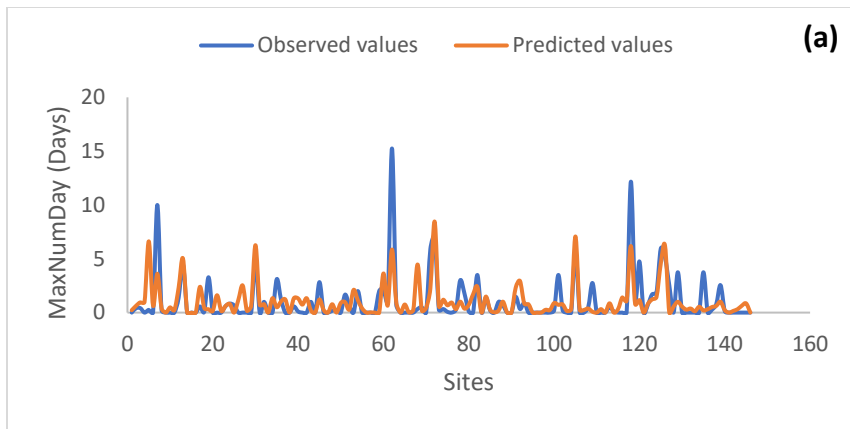


Figure 3-4. Diagrammes combinés montrant les valeurs estimées et les valeurs observées pour les variables MaxNumDay (a, c) et le paramètre gaussien c (b, d).

On a tenté de comprendre le processus qui peut expliquer la distribution des valeurs élevées de MaxNumDay. En se basant sur la figure 3-5 montrant que la température de l'eau a tendance à augmenter en direction de l'aval en atteignant un maximum estimé lorsque le cours d'eau a une largeur supérieure à 50 mètres et une profondeur inférieure à 1,5 mètre (Caissie, 2006), on a formulé les hypothèses suivantes :

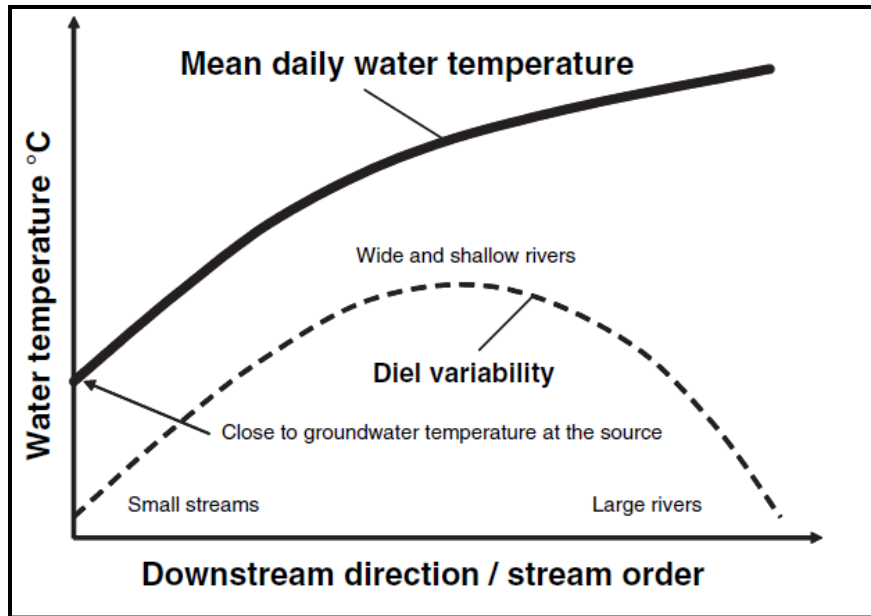


Figure 3-5. Représentation schématique de la variabilité moyenne quotidienne et journalière de la température de l'eau en fonction de l'échelle spatiale, c'est-à-dire de la direction en aval, qui se traduit aussi généralement par l'ordre des cours d'eau (Modifié d'après Caissie, 2006 ; Austin et al., 2015).

Au Nouveau Brunswick, les deux stations situées au bord de la rivière Miramichi sont situées sur des cours d'eau d'ordre 3. Ces cours d'eau sont peu profonds et plus larges et qui sont fortement exposés aux radiations solaires avec une couverture insuffisante pour les protéger de ces radiations. D'ailleurs la rivière Miramichi a connu des périodes de stress thermique surtout durant l'été, là où les températures de l'eau dépassent souvent les seuils de stress à haute température pour le saumon atlantique (Breau, 2013; Breau et Caissie, 2013). Les deux autres stations sont situées en Nouvelle Écosse (dans la partie Sud de notre zone d'étude). Les valeurs élevées de MaxNumDay peuvent être expliquées par la présence des lacs. Prenant comme exemple, la rivière Shelburne prend sa source dans le lac Buckshot. Son cours d'eau de 53km décrit un arc débutant en direction Nord, puis vers l'Est et ensuite vers le Sud vers le lac Rossignol (Figure 3-6).

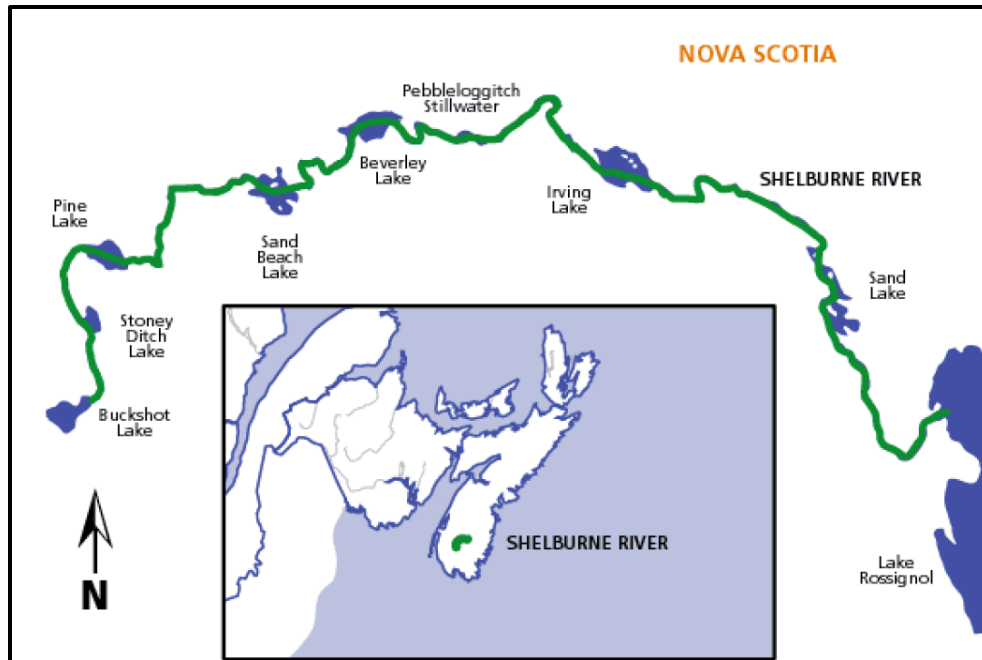


Figure 3-6. Localisation géographique de la rivière Shelburne en Nouvelle Écosse (tirée Shelburne Canadian Heritage River Monitoring Report 2008-2018).

Pour aller un peu plus loin, on a tenté de présenter le diagramme de Taylor pour chaque métrique thermique. Ce type de diagramme facilite l'évaluation comparative de différents modèles testés. Il est utilisé pour quantifier le degré de correspondance entre le comportement modélisé et observé en termes de trois statistiques :

- ✓ Le coefficient de corrélation de Pearson
- ✓ L'erreur quadratique moyenne RMSE
- ✓ L'écart type SD

Graphiquement, sur le diagramme de Taylor, les lignes noires pleines représentent l'écart type, les lignes pointillées correspondent aux corrélations entre les valeurs estimées et celles observées. Les lignes pleines vertes représentent la différence de l'erreur quadratique moyenne centrée entre les valeurs simulées et observées. La figure 3-7 illustre le diagramme de Taylor pour le paramètre a de la fonction gaussienne. On voit clairement que les deux modèles MARS et GAM montrent des performances comparables en termes de coefficient de corrélation, d'écart type et de RMSE.

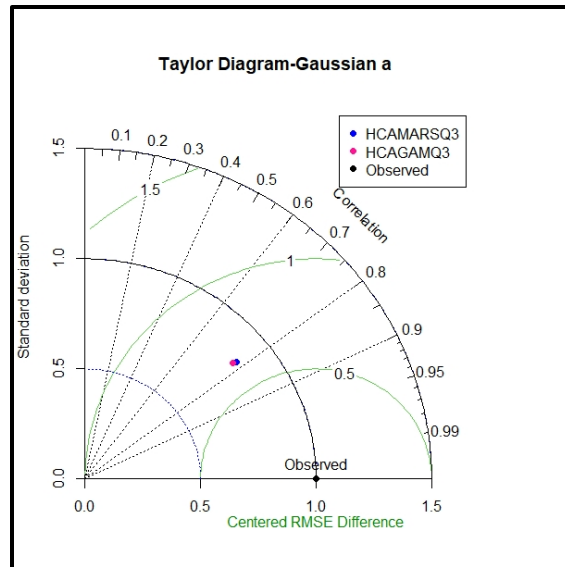


Figure 3-7. Diagramme de Taylor illustrant l'évaluation conjointe de l'écart type (SD), de la différence quadratique moyenne centrée et du coefficient de corrélation pour les cinq métriques prédites par les meilleurs modèles.

La figure 3-8 représente l'erreur relative en fonction de certaines variables explicatives jugées pertinentes pour prédire MaxWaterTmax (Q1). En général, le modèle MARS a une performance légèrement supérieure que le modèle non-linéaire GAM de point de vue erreur relative. En effet, les diagrammes combinés montrent que l'erreur relative est faible pour:

- Les latitudes les plus faibles du centroïde du bassin versant;
- Les bassins versants de taille moyenne;

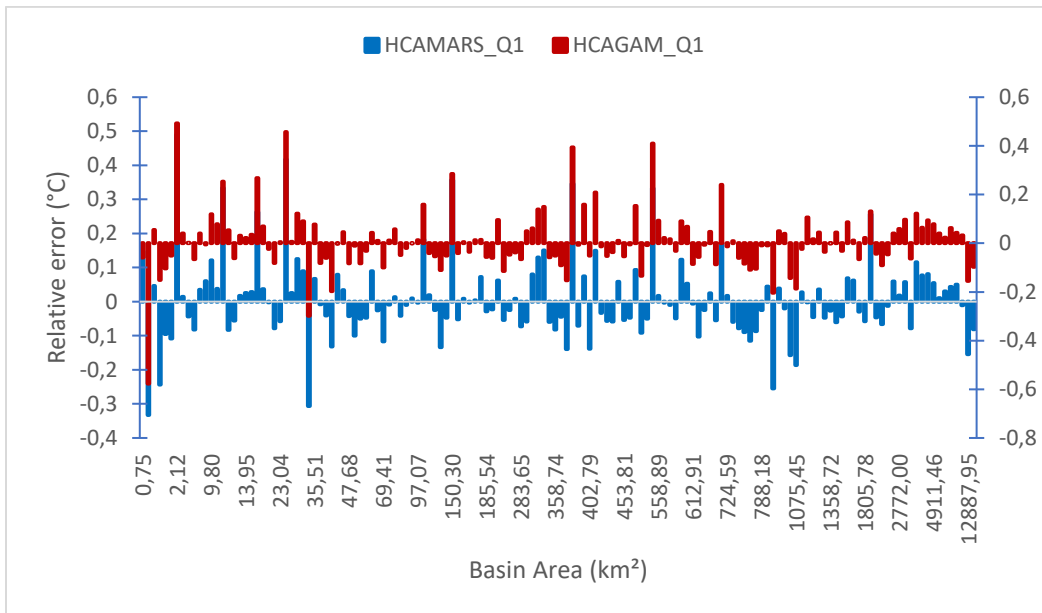
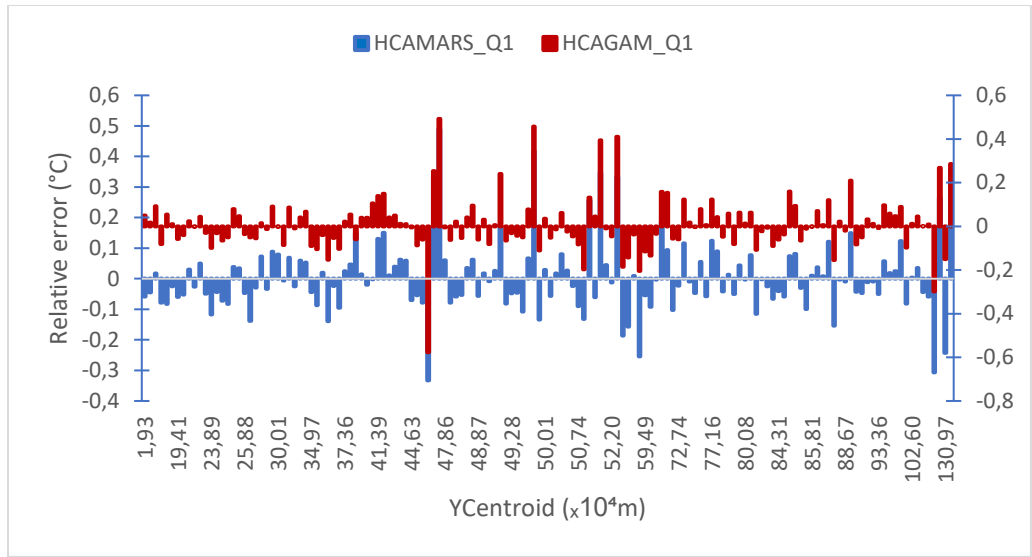


Figure 3-8. Graphiques combinés montrant l'erreur relative en fonction de certaines variables significatives jugées importantes.

4 DISCUSSION GÉNÉRALE ET CONCLUSION

La dynamique des cours d'eau et des rivières est complexe, régie par plusieurs facteurs à la fois naturels et anthropiques. Le besoin d'une meilleure connaissance du régime thermique est à l'origine du développement de la modélisation au cours des dernières années. La température de l'eau est un facteur déterminant les aires de répartition des organismes et, en grande partie, leur cycle de vie. Les variations de la température de l'eau peuvent modifier la structure, la distribution spatiale et l'abondance des populations au cours des saisons.

L'objectif principal de ce travail était de développer une approche novatrice nommée analyse thermique régionale afin d'estimer cinq métriques de température de l'eau importantes pour la gestion des stocks du saumon atlantique dans l'Est du Canada. L'étude a été effectuée à l'aide des données de température de l'eau issues essentiellement du réseau RivTemp.

Les objectifs spécifiques dans cette étude visent à :

- Montrer l'importance de l'approche de délimitation des régions homogènes;
- Prouver la nécessité d'approches non linéaires pour une meilleure estimation des métriques thermiques;
- Identifier les facteurs déterminants la température de l'eau;
- Identifier le modèle le plus parcimonieux pour prédire les variables d'intérêts.

Les résultats de cette étude offrent des informations précieuses sur le thème de la modélisation de la température de l'eau. Dans ce travail, on a montré que les modèles non linéaires sont plus favorables pour modéliser les processus thermiques complexes que le modèle classique linéaire. En effet, un modèle linéaire essaye d'ajuster la meilleure droite qui passe au milieu des données, cela ne fonctionne donc pas pour tous les enjeux de données. En revanche, les modèles non linéaires pouvant faire ceci en ajustant une fonction de lissage non linéaire, notamment des splines cubiques dans le cas du modèle GAM et des splines linéaires par morceaux dans le cas du modèle MARS.

Aussi, on a prouvé que la délimitation des régions thermiques potentiellement homogènes avant d'appliquer les modèles d'estimation régionale améliore l'estimation des métriques thermiques. Dans l'Est du Canada, il existe des régions thermiques distinctes dont leurs distributions varient en fonction de la métrique d'intérêt. Aussi, la taille de l'échantillon de données est un paramètre important dans la délimitation des groupements homogènes et par conséquent sur la performance du modèle d'estimation.

La comparaison entre les trois approches de modélisation montre qu'avec toutes les stations (ALL), l'approche MARS permet d'obtenir des résultats à la précision semblable à ceux obtenus à l'aide du modèle

GAM en termes de R^2 et RMSE. Ces deux approches sont trouvées plus performantes que le modèle simple MLR. Au sein des régions délimitées à l'aide de la technique analyse de regroupement hiérarchique HCA, les deux combinaisons HCA+MARS et HCA+GAM sont les meilleures combinaisons pour l'estimation des métriques thermiques en termes de R^2 , RMSE, RRMSE, Bias et PBias.

Cette étude permet aussi d'identifier les variables explicatives qui ont un pouvoir de prédiction significatif. La température de l'air et l'emplacement du centroïde sur l'axe y (lié à la latitude) sont des variables de haute importance pour l'estimation des métriques de température de l'eau.

Toutefois, l'exactitude et la fiabilité d'une telle approche de modélisation ne peuvent pas être garanties. Ces résultats restent donc à démontrer dans d'autres cas en vue d'enrichir et d'approfondir la compréhension de la dynamique thermique des rivières.

Les perspectives de travaux futurs à donner à ce mémoire de recherche sont :

- Inclure d'autres variables indépendantes comme par exemple, les caractéristiques du réseau du drainage pour dériver des estimations des indices thermiques.
- Compte tenu de l'effet des lacs sur les variables climatiques et, par conséquent, sur la température de l'eau de la rivière, l'étude de leur impact potentiel est indispensable.
- Explorer l'impact d'autres facteurs anthropiques, naturels, de changement climatique pouvant influencer le régime thermique.
- Considérant que les performances du modèle GAM s'améliorent en réponse à l'augmentation de la taille des groupes potentiellement homogènes, d'autres études sont recommandées pour une enquête plus approfondie concernant les impacts des performances de ces modèles lorsqu'ils sont ajustés avec différentes densités de stations.
- Par rapport à l'approche MARS, bien que cette approche soit un outil d'estimation efficace et simple pouvant être utilisé pour estimer les métriques thermiques, certaines contraintes telles que le nombre maximal de termes et le degré d'interaction maximal autorisé dans la passe avant doivent être spécifiés par l'utilisateur. Ceux-ci dépendent du problème à résoudre et doivent être examinés avec soin.

Divers scénarios de changements climatiques prévoient une tendance à la hausse des températures au cours des prochaines années (Isaak et al., 2012 ; Mohseni et al., 2003). De plus, les températures élevées des rivières deviennent de plus en plus problématiques pour plusieurs espèces de poissons (Fang et Stefan 1999, Crozier et Zabel 2006). Une bonne maîtrise des techniques de modélisation de la température des eaux en rivières telles que développées dans la présente étude pour des fins de gestion efficace des ressources aquatiques s'avère nécessaire, surtout en contexte du réchauffement climatique. Les méthodes présentées

dans ce mémoire pourront aider les gestionnaires des pêches pour la planification et les décisions opérationnelles.

5 BIBLIOGRAPHIE

- Adrian, R., O'Reilly, C. M., Zagarese, H., Baines, S. B., Hessen, D. O., Keller, W., ... & Winder, M. (2009). Lakes as sentinels of climate change. *Limnology and oceanography*, 54(6part2), 2283-2297. https://doi:10.4319/lo.2009.54.6_part_2.2283.
- Ahmadi-Nedushan, B., St-Hilaire, A., Ouarda, T. B. M. J., Bilodeau, L., Robichaud, E., Thiémonge, N., Bobée, B., 2007. Predicting river water temperatures using stochastic models: Case study of the Moisie River (Québec, Canada). *Hydrol. Process.* 21: 21-34. <https://doi.10.1002/hyp.6353>.
- Arismendi, I., Johnson, S. L., Dunham, J. B., and Haggerty, R., 2013. Descriptors of natural thermal regimes in streams and their responsiveness to change in the Pacific Northwest of North America. *Freshw. Biol.* 58, 880–894.
- Arismendi, I., Johnson, S. L., Dunham, J. B., Haggerty, R., and 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).
- Austin, H., Bradley, D., Steward-Rousson, I., and Milner, N., 2015. Literature review of the influence of large impoundments on downstream temperature, water quality and ecology, with reference to the water framework directive. APEM scientific report. Scottish Environmental Protection Agency, Stirling, Scotland. p. 84.
- Austin, M.P., Nicholls, A. O., and Margules, C. R., 1990. Measurement of the realized qualitative niche: environmental niches of five eucalypt species. *Ecol. Monogr.* 60, 161–177.
- Babaei, M., Moeini, R., Ehsanzadeh, E., 2019. Artificial neural network and support vector machine models for inflow prediction of dam reservoir (case study: zayandehroud dam reservoir). *Water Resour. Manag.* 33, 2203–2218. <https://doi.org/10.1007/s11269-019-02252-5>.
- Balshi, M. S., McGUIRE, A. D., Duffy, P., Flannigan, M., Walsh, J., and Melillo, J., 2009. Assessing the response of area burned to changing climate in western boreal North America using a Multivariate Adaptive Regression Splines (MARS) approach. *Global Change Biology* 15(3):578-600.
- Bartholow, J. M., Campbell, S. G., and Flug, M., 2004. Predicting the thermal effects of dam removal on the Klamath River. *Environ. Manag.* 34(6):856-874.
- Barton, D. R., Taylor, W. D., Biette, R. M., 1985. Dimensions of riparian buffer strips required to maintain trout habitat in southern Ontario streams. *North American Journal of Fisheries Management* 5(3A): 364-378.
- Basu, B., and Srinivas, V. V., 2014. Regional flood frequency analysis using kernel-based fuzzy clustering approach. *Water Resour. Res.* 50, 3295-3316. <https://doi:10.1002/2012WR012828>.

- Bayentin, L., El Adlouni, S., Ouarda, T. B. M. J., Gosselin, P., Doyon, B., and Chebana, F., 2010. Spatial variability of climate effects on ischemic heart disease hospitalization rates for the period 1989-2006 in Quebec, Canada. *Int. J. Health Geogr.* 9, 5. <https://doi.org/10.1186/1476-072X-9-5>.
- Beaufort, A., Diamond, J. S., Sauquet, E., and Moatar, F., 2021. The thermal peak: A simple stream temperature metric at regional scale, *Hydrol. Earth Syst. Sci. Discuss.* [preprint]. <https://doi.org/10.5194/hess-2021-218>. in review.
- Bélanger, M., El-Jabi, N., Caissie, D., Ashkar, F., and 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/J. Water Sci.* 18(3):403-421.
- Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B. M. J. and Bobée, B., 2007. A review of statistical water temperature models. *Can. Water Resour. J.* 32, 179–192.
- Beschta, R., Bilby, R., Brown, G., Holtby, L. 1987. Stream Temperature and Aquatic Habitat: Fishery and Forestry Interactions.” In *Streamside Management: Forestry and Fishery Interactions*, edited by E.O. Salo and T.W. Cundy, 191–232. Seattle: University of Washington, Institute of Forest Resources.
- Bond, N. R., and Kennard, M. J., 2017. Prediction of hydrologic characteristics for ungauged catchments to support hydroecological modeling. *Water Resour. Res.* 53(11), 8781-8794.
- Boudreault, J., St-Hilaire, A., Bergeron, N., and Chebana, F., 2019. Stream temperature modelling using functional regression models. *J. Amer. Water Resour. Assoc.* <https://doi.org/10.1111/1752-1688.12778>.
- Boyer, C., St-Hilaire, A., Allen Curry, R., Caissie, D., and Gillis, C. A., 2016. Technical Report: RivTemp: A Water Temperature Network for Atlantic Salmon Rivers in Eastern Canada. *Water News, Canadian Water Association Newsletter*, Spring edition.
- Breau, C., Weir, L. K., and Grant, J. W. A., 2007. Individual variability in activity patterns of juvenile Atlantic salmon (*Salmo salar*) in Catamaran Brook, New Brunswick. *Can. J. Fish. Aquat. Sci.* 64 (3), 486-494. <https://doi.org/10.1139/F07-026>.
- Breau, C. and Caissie, D., 2013. Adaptive management strategies to protect salmon (*Salmo salar*) under environmentally stressful conditions. *DFO Canadian Science Advisory Secretariat Research Document, 2012/164*. Ottawa: Fisheries and Oceans Canada.
- Breau, C., 2013. Knowledge of fish physiology used to set water temperature thresholds for in-season closures of Atlantic salmon (*Salmo salar*) recreational fisheries. *DFO Canadian Science Advisory Secretariat Research Document, 2012/163*. Ottawa: Fisheries and Oceans Canada.
- Bunn, S. E., and Arthington, A. H., 2002. Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environ. Manag.* 30(4):492-507.

- Burn, D. H., 1990. Evaluation of regional flood frequency analysis with a region of influence approach, *Water Resour. Res.* 26, 2257–2265.
- Cai, H., Piccolroaz, S., Huang, J., Liu, Z., Liu, F., Toffolon, M., 2018. Quantifying the impact of the three gorges dam on the thermal dynamics of the Yangtze River. *Environ Res Lett* 13:054016.
- Caissie, D., 2006. The thermal regime of rivers: A review. *J. Freshw. Biol.* 51, 1389-1406.
- Caissie, D., Ashkar, F., and El - Jabi, N., 2019. Analysis of air/river maximum daily temperature characteristics using the peaks over threshold approach. *Ecohydrology*. <https://doi:10.1002/eco.2176>.
- Caissie, D., El-Jabi, N., and Satish, M., 2001. Modeling of Maximum Daily Water temperatures in a small stream using air temperatures. *J. Hydrol*, 251, 14–28.
- Caissie, D., El-Jabi, N., and St-Hilaire, A., 1998. Stochastic modelling of water temperatures in a small stream using air to water relations. *Can. J. Civil. Eng.*, 25, 250–260.
- Caissie, D., Satish, M. G., and El-Jabi, N., 2007. Predicting water temperatures using a deterministic model: application on Miramichi River catchments (New Brunswick, Canada). *J. Hydrol.* 336, 303–315.
- Cavadias, G., Ouarda, T. B. M. J., Bobée, B., and Girard, C., 2001. A canonical correlation approach to the determination of homogeneous regions for regional flood estimation of ungauged basins. *Hydrol. Sci. J.* 46(4): 499-512.
- Charron, C., Boyer, C., St-Hilaire, A., Ouarda, T. B. M. J., Daigle, A., and Bergeron, N. E., 2019. Regional analysis and modelling of water temperature metrics for Atlantic salmon (*Salmo Salar*) in eastern Canada. INRS Scientific Report #1855, 29 pages.
- Chebana, F., Charron, C., Ouarda, T. B. M. J., and Martel, B., 2014. Regional frequency analysis at ungauged sites with the generalized additive model. *J. Hydrometeor.* 15(6), 2418-2428.
- Chen, Y. D., Carsel, R. F., McCutcheon, S. C., Nutter, W. L., 1998a. Stream temperature simulation of forested riparian areas: I. Watershed-scale model development. *Journal of Environmental Engineering* 124: 304–315.
- Chen, Y. D., McCutcheon, S. C., Norton, D. J., Nutter, W. L., 1998b. Stream temperature simulation of forested riparian areas: II. Model application. *Journal of Environmental Engineering* 124: 316–328.
- Chu, C., and Jones, N. E., 2010. Do existing ecological classifications characterize the spatial variability of stream temperatures in the Great Lakes Basin, Ontario? *J. Great Lakes. Res.* 36, 633–640.
- Chu, C., Jones, N. E., and Allin, L., 2010. Linking the thermal regimes of streams in the Great Lakes Basin, Ontario, to landscape and climate variables. *River. Res. Applic.* 26, 221–241. <https://doi:10.1002/rra.1259>.
- Cole, J.C., Maloney, K.O., Schmid, M., McKenna Jr., J.E., 2014. Developing and testing temperature models for regulated systems: a case study on the Upper Delaware River. *J. Hydrol.* 519, 588-598. DOI 10.1016/j.jhydrol.2014.07.058.

- Corey, E., Linnansaari, T., Cunjak, R. A., and Currie, S., 2017. Physiological effects of environmentally relevant, multi-day thermal stress on wild juvenile Atlantic salmon (*Salmo salar*). *Conservation Physiology* 5 (January). [https://doi: 10.1093/conphys/cox014](https://doi.org/10.1093/conphys/cox014).
- Crozier, L. G., Hendry, A., Lawson, P. W., Quinn, T., Mantua, N., Battin, J., Shaw, R., and 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, L.G. and Zabel, R.W., 2006. Climate impacts at multiple scales: evidence for differential population responses in juvenile Chinook salmon. *Journal of Animal Ecology*, 75, 1100–1109. [https://doi:10.1111/j.1365-2656.2006.01130.x](https://doi.org/10.1111/j.1365-2656.2006.01130.x).
- Daigle, A., St-Hilaire, A., and Boyer, C., 2019. A standardized characterization of river thermal regimes in Québec (Canada). *J. Hydrol.* 577, 123963. [https://doi:10.1016/j.jhydrol.2019.123963](https://doi.org/10.1016/j.jhydrol.2019.123963).
- Daigle, A., St-Hilaire, A., Peters, D., and Baird, D., 2010. Multivariate water temperature modeling in a semi-arid watershed. Accepted for publication, *Can. Water Resour. J.* 35(3), 237–258.
- Danie, D.S., Trial, J. G., and Stanley, J. G., 1984. Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (North Atlantic)—Atlantic salmon. U.S. Fish Wildl. Serv. FWS/ OBS-82/11.22, and U.S. Army Corps of Engineers, TR EL-82-4.19 pp.
- Daufresne, M., Lengfellner, K., and Sommer, U., 2009. Global warming benefits the small in aquatic ecosystems. *Proceedings of the National Academy of Sciences* 106(31):12788-12793.
- Deo, R. C., Kisi, O., and Singh, V. P., 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. *Atmospheric Research* 184:149-175.
- Desai, S., and Ouarda, T. B. M. J., 2021. Regional hydrological frequency analysis at ungauged sites with random forest regression. *J. Hydrol.* 594, 125861. [https://doi:10.1016/j.jhydrol.2020.125861](https://doi.org/10.1016/j.jhydrol.2020.125861).
- DeWeber, J. T., and Wagner, T., 2014. A regional neural network ensemble for predicting mean daily river water temperature, *J. Hydrol.* 517, 187–200. [https://doi:10.1016/j.jhydrol.2014.05.035](https://doi.org/10.1016/j.jhydrol.2014.05.035).
- Ducharne, A., 2008. Importance of stream temperature to climate change impact on water quality, *Hydrol. Earth Syst. Sci.* 12, 797–810. [https://doi: 10.5194/hess-12-797-2008](https://doi.org/10.5194/hess-12-797-2008).
- Dugdale, S. J., Franssen, J., Corey, E., Bergeron, N. E., Lapointe, M., and Cunjak, R. A., 2016. Main stem movement of Atlantic salmon parr in response to high river temperature. *Ecol. Freshw. Fish.* 25 (3), 429–445. [https://doi: 10.1111/eff.12224](https://doi.org/10.1111/eff.12224).
- Dugdale, S. J., Hannah, D. M., and Malcolm, I. A., 2017. River temperature modelling: a review of process-based approaches and future directions. *Earth Sci. Rev.* 175, 97–113. <https://doi.org/10.1016/j.earscirev.2017.10.009>.

- Eaton, J. G., McCormick, J. H., Stefan, H. G., and Hondzo, M., 1995. Extreme value analysis of a fish /Temperature field database. *Ecol. Eng.* 4 (4), 289-305.
- Elliott, J. M., 1991. Tolerance and resistance to thermal stress in juvenile Atlantic salmon, *Salmo salar*. *Freshw. Biol.* 25(1), 61–70.
- Elliott, J.M., and Elliott, J. A., 2010. Temperature requirements of Atlantic salmon *Salmo salar*, brown trout *Salmo trutta* and Arctic charr *Salvelinus alpinus*: Predicting the effects of climate change. *J. Fish. Biol.* 77 (8), 1793–1817. [https://doi: 10.1111/j.1095-8649.2010.02762](https://doi:10.1111/j.1095-8649.2010.02762).
- Elliott, S. R., Coe, T. A., Helfield, J. M., and Naiman, R. J., 1998. Spatial variation in environmental characteristics of Atlantic salmon (*Salmo salar*) rivers. *Can. J. Fish. Aquat. Sci.* 55, 267-280.
- Emamgolizadeh, S., Bateni, S., Shahcavani, D., Ashrafi, T., and Ghorbani, H., 2015. Estimation of soil cation exchange capacity using genetic expression programming (GEP) and multivariate adaptive regression splines (MARS). *J. Hydrol.* 529:1590-1600.
- Fang, X., and Stefan, H.G., 1999. Projections of climate change effects on water temperature characteristics of small lakes in the contiguous U.S. *Climatic Change*, 49, 377–412. <https://doi:10.1023/A:1005431523281>.
- Ficklin, D. L., Stewart, I. T., and Maurer, E. P., 2013. Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California. *Water Resour. Res.* 49(5):2765-2782.
- Foreman, M. G. G., Lee, D. K., Morrison, J., Macdonald, S., Barnes, D., and S. Williams IV, S., 2001. Simulations and retrospective analyses of Fraser watershed flows and temperatures. *Atmosphere-Ocean*, 39, 89–105.
- Friedman, J. H., 1991. Multivariate adaptive regression splines. *Annals of Statistics*, 19, 1–141.
- Fry, F. E. J., 1971. The effect of environmental factors on the physiology of fish. *Fish Physiology*, 6, 1–98. [https://doi.org/10.1016/S1546-5098\(08\)60146-6](https://doi.org/10.1016/S1546-5098(08)60146-6).
- Feigl, M., Lebedzinski, K., Herrnegger, M., and Schulz, K., 2021. Machine-learning methods for stream water temperature prediction. *Hydrol. Earth Syst. Sci.*, 25, 2951–2977. <https://doi.org/10.5194/hess-25-2951-2021>,
- Gallice, A., Schaepli, B., Lehning, M., Parlange, M. B., and Huwald, H., 2015. Stream temperature prediction in ungauged basins: Review of recent approaches and description of a new physics-derived statistical model. *Hydrol. Earth Syst. Sci.* 19, 3727–3753. <https://doi.org/10.5194/hess-19-3727-2015>.
- GIEC., 2013. *Changements Climatiques 2013 Les éléments scientifiques, Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat*).

- Garner, G., Malcolm, I. A., Sadler, J. P., Hannah, D. M., 2017. The role of riparian vegetation density, channel orientation and water velocity in determining river temperature dynamics. *J. Hydrol.* 553:471–485. <https://doi.org/10.1016/j.jhydrol.2017.03.024>.
- Gregg, E. J., and Trites, A. W., 2001. Predictions of critical habitat for whale species in the waters of coastal British Columbia. *Can. J. Fish. Aquat. Sci.* 58, 1265–1285.
- Gu, R.R., Montgomery, S., Austin, T. 1998. Quantifying the Effects of Stream Discharge on Summer River Temperature. *Hydrological Science Journal* 43:885-894.
- Haddad, K. and Rahman, A., 2012. Regional flood frequency analysis in eastern Australia: Bayesian GLS regression-based methods within fixed region and ROI framework—Quantile Regression vs. Parameter Regression Technique. *J. Hydrol.* 430, 142-161.
- Haddad, K., Rahman, A., and Ling, F., 2014. Regional flood frequency analysis method for Tasmania, Australia: A case study on the comparison of fixed region and region-of-influence approaches. *Hydrol. Sci. J.* null null. <https://doi:10.1080/02626667.2014.950583>.
- Hadzima-Nyarko, M., Rabi, A., and Šperac, M., 2014. Implementation of Artificial Neural Networks in Modeling the Water-Air Temperature Relationship of the River Drava. *Water Resour. Manag.* 28(5) :1379-1394.
- Hannah, D. M., Malcolm, I. A., Soulsby, C., and Youngson, A. F., 2008. A comparison of forest and moorland stream microclimate, heat exchanges and thermal dynamics. *Hydrol. Process.* 22(7):919-940.
- Hastie, T., and Tibshirani, R. J., 1990. *Generalized Additive Models. Monographs on Statistics and Applied Probability*, vol. 43. Chapman and Hall, London, 335 pp.
- Heggenes, J., Stickler, M., Alfredsen, K., Brittain, J. E., Adeva-Bustos, A., and Huusko, A., 2021. Hydropower-driven thermal changes, biological responses and mitigating measures in northern river systems. *River Res Applic.* 1–23. <https://doi.org/10.1002/rra.3788>.
- Hester, E. T., and Doyle, M. W., 2011. Human impacts to river temperature and their effects on biological processes: a quantitative synthesis. *J. Am. Water Resour. Assoc.* 47:571–587.
- Hill, R. A., Hawkins, C. P., and Carlisle, D. M., 2013. Predicting thermal reference conditions for USA streams and rivers. *Freshw. Sci.* 32, 39–55.
- Hodgson, S., and Quinn, T. P., 2002. The timing of adult sockeye salmon migration into fresh water: Adaptations by populations to prevailing thermal regimes. *Can. J. Zool.* 80, 542-555.
- Horne, B.D., Rutherford, E.S., Wehrly, K.E. 2004. Simulating Effects of Hydro-Dam Alteration on Thermal Regime and Wild Steelhead Recruitment in a Stable-Flow Lake Michigan Tributary. *River Research and Applications* 20:185-203.
- Hosking, J. R. M., and Wallis, J. R., 1997. *Regional Frequency Analysis: An Approach Based on LMoments.* Cambridge University Press, New York, 224 pp.

- Hrachowitz, M., Soulsby, C., Imholt, C., Malcolm, I. A., and Tetzlaff, D., 2010. Thermal regimes in a large upland salmon river: A simple model to identify the influence of landscape controls and climate change on maximum temperatures. *Hydrol. Process.* 24(23), 3374–3391.
- Hutchinson, M. F., Mckenney, D. W., Lawrence, K., Pedlar, J. H., Hopkinson, R. F., Milewska, E., and Papadopol, P., 2009. Development and Testing of Canada-Wide Interpolated Spatial Models of Daily Minimum-Maximum Temperature and Precipitation for 1961-2003. *J. Appl. Meteor. Climatol.* 48(4), 725-741.
- Imholt, C., Soulsby, C., Malcolm, I. A., Hrachowitz, M., Gibbins, C. N., Langan, S., and Tetzlaff, D., 2013. Influence of scale on thermal characteristics in a large montane river basin, *River Res Applic.* 29, 403–419. [https://doi: 10.1002/rra.1608](https://doi.org/10.1002/rra.1608).
- Isaak, D. J., and Rieman, B. E., 2013. Stream isotherm shifts from climate change and implications for distributions of ectothermic organisms. *Glob. Change Biol.* 19 (3), 742–751. <https://doi.org/10.1111/gcb.12073>.
- Isaak, D. J., Wenger, S. J., Peterson, E. E., Ver Hoef, J. M., Nagel, D. E., Luce, C. H., Hostetler, S. W., Dunham, J. B., Roper, B. B., and Wollrab, S. P., 2017. The NorWeST Summer Stream Temperature Model and Scenarios for the Western US: A Crowd-Sourced Database and New Geospatial Tools Foster a User Community and Predict Broad Climate Warming of Rivers and Streams. *Water Resour. Res.* 53 (11), 9181–205.
- Isaak, D., Wollrab, S., Horan, D., and 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.
- Iskra, I., Droste, R., 2008. Parameter uncertainty of a watershed model. *Canadian Water Resources Journal* 33(1): 5-22.
- Jeong, D. I., Daigle, A., and St-Hilaire, A., 2013. Development of a stochastic water temperature model and projection of future water temperature and extreme events in the Ouelle River Basin in Québec, Canada. *River Res Applic.* 29, 805-821. [https://doi: 10.1002/rra.2574](https://doi.org/10.1002/rra.2574).
- Jia, X., Zwart, J., Sadler, J., Appling, A., Oliver, S., Markstrom, S., ... and Kumar, V., 2021. Physics-guided recurrent graph model for predicting flow and temperature in river networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM)* (pp. 612-620). Society for Industrial and Applied Mathematics.
- Johnson, S. C., 1967. Hierarchical clustering schemes. Springer ed. s.l.: Psychometrika.
- Johnson, S. L., 2004. Factors influencing stream temperatures in small streams: Substrate effects and a shading experiment. *Can. J. Fish. Aquat. Sci.* 61, 913–923.

- Jonsson, B., and Jonsson, N., 2009. A review of the likely effects of climate change on anadromous Atlantic salmon *Salmo salar* and brown trout *Salmo trutta*, with particular reference to water temperature and flow. *J. Fish. Biol.* 75(10), 2381–2447. <https://doi.org/10.1111/j.1095-8649.2009.02380.x>.
- Justice, C., White, S. M., McCullough, D. A., Graves, D. S., and Blanchard, M. R., 2017. Can stream and riparian restoration offset climate change impacts to salmon populations. *J. Environ. Manag.* 188, 212–227. <https://doi:10.1016/j.jenvman.2016.12.005>.
- Karacor, A. G., Sivri, N., and Ucan, O. N., 2007. Maximum stream temperature estimation of Degirmendere River using artificial neural network. *J. Sci. Ind. Res. India.* 66(5):363-366.
- Kim, K. S., Chapra, S. C., 1997. Temperature Model for Highly Transient Shallow Streams. ASCE, *Journal of Hydraulic Engineering* 123:30-40.
- Kinouchi, T., Yagi, H., Miyamoto, M., 2007. Increase in stream temperature related to anthropogenic heat input from urban wastewater. *J. Hydrol.* 335: 78-88.
- Kisi, O., 2015. Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *J. Hydrol.* 528:312-320.
- Kisi, O., Parmar, K. S., 2016. Application of least square support vector machine and multivariate adaptive regression spline models in long-term prediction of river water pollution. *J. Hydrol.* 534:104-112.
- Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K., Marshall, L., 2012. Investigating controls on the thermal sensitivity of Pennsylvania streams. *Hydrol. Process.* 26:771_785 DOI 10.1002/hyp.8186.
- Kwak, J., St-Hilaire, A., Chebana, F., 2017. A comparative study for water temperature modelling in a small basin, the Fourchue River, Quebec, Canada. *Hydrol. Sci. J.* 62 (1), 64-75.
- Laanaya, F., St-Hilaire, A., and Gloaguen, E., 2017. Water temperature modelling: comparison between the generalized additive model, logistic, residuals regression and linear regression models. *Hydrol. Sci. J.* 62 (7), 1078–1093. <https://doi:10.1080/02626667.2016.1246799>.
- Leach, J. A., Neilson, B. T., Buahin, C. A., Moore, R. D., and Laudon, H., 2021. Lake outflow and hillslope lateral inflows dictate thermal regimes of forested streams draining small lakes. *Water Resour. Res.* 57, e2020WR028136. <https://doi.org/10.1029/2020WR028136>.
- Leathwick, J. R., 1998. Are New Zealand's *Nothofagus* species in equilibrium with their environment? *J. Veg. Sci.* 9, 719-732.
- Leathwick, J., Elith, J., Hastie, T., 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Model.* 199(2):188-196.

- Leathwick, J.R., Rowe, D., Richardson, J., Elith, J., and Hastie, T., 2005. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshwater Biol.* 50, 2034–2052.
- Lee, T. S., Chen, I. F., 2005. A two-stage hybrid credit-scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications* 28(4):743-752.
- Lee, T. S., Chiu, C. C., Chou, Y. C., Lu, C. J., 2006. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data Analysis* 50(4):1113-1130.
- Leitte, A. M., Petrescu, C., Franck, U., Richter, M., Suci, O., Ionovici, R., Herbarth, O., and Schlink, U., 2009. Respiratory health, effects of ambient air pollution and its modification by air humidity in Drobeta-Turnu Severin, Romania. *Sci. Total Environ.* 407 (13), 4004–4011. <https://doi.org/10.1016/j.scitotenv.2009.02.042>.
- Li, Y., He, Y., Su, Y., and Shu, L., 2016. Forecasting the daily power output of a grid –connected photovoltaic system based on multivariate adaptive regression splines. *Applied Energy* 180:392-401.
- Liu, B., Yang, D., Ye, B., and 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.
- Lisi, P. J., Schindler, D. E., Cline, T. J., Scheuerell, M. D., Walsh, P. B., 2015. Watershed geo- morphology and snowmelt control stream thermal sensitivity to air temperature. *Geophysical Research Letters* 42:3380_3388 DOI 10.1002/2015GL064083.
- Lund, S. G., Caissie, D., Cunjak, R. A., Vijayan, M. M., and Tufts, B. L., 2002. The effects of environmental heat stress on heat-shock mRNA and protein expression in Miramichi Atlantic salmon (*Salmo salar*) parr. *Can. J. Fish. Aquat. Sci.* 59, 1553-1562.
- Maheu, A., Poff, N. L., and St-Hilaire, A., 2015. A classification of stream water temperature regimes in the conterminous United States. Published online in *River Research and Applications*, 32(16), 896-906. <https://doi: 10.1002/rra2906>.
- Maheu, A., St-Hilaire, A., Caissie, D., El-Jabi, N., Bourque, G., and Boisclair, D., 2016b. A regional analysis of the impact of dams on water temperature in medium-size rivers in eastern Canada. *Can. J. Fish. Aquat. Sci.* 73 (12), 1885–1897. <https://doi.org/10.1139/cjfas-2015-0486>.
- Makarowski, K. E., 2009. An investigation of spatial and temporal variability in stream temperature in several of Montana's reference streams: working toward a more holistic management strategy. Unpublished MSc. Thesis, University of Montana.
- Martins, E. G., Hinch, S. G., Patterson, D. A., Hague, M. J., Cooke, S. J., Miller, K. M., Robichaud, D., English, K. K., Farrell, A. P., and Jonsson, B., 2012. High river temperature reduces survival of sockeye

- salmon (*Oncorhynchus nerka*) approaching spawning grounds and exacerbates female mortality. *Can. J. fish. Aquat. Sci.* 69(2):330-342.
- Mather, M. E., Parrish, D. L., Campbell, C. A., McMenemy, J. R., and Smith, J. M., 2008. Summer temperature variation and implications for juvenile Atlantic salmon. *Hydrobiologia.* 603(1), 183-196.
- McCullagh, P., and Nelder, J. A., 1989. *Generalized Linear Models*. Second edition. Chapman and Hall, London, UK.
- McCullagh, P., and Nelder, J. A., 1998. *Generalized Linear Models*. 2nd ed. Monographs on Statistics and Applied Probability, 37. Boca Raton: Chapman & Hall/CRC.
- McCullough, D. A., 1999. A review and synthesis of effects of alterations to the water temperature regime on freshwater life stages of salmonids, with special reference to Chinook Salmon. U.S. Environmental Protection Agency: Seattle.
- Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F., Stouffer, R. J., and Taylor, K. E., 2007. The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Am. Meteorol. Soci.* 88(9) :1383-1394.
- Mehdizadeh, S., Fathian, F., Safari, M. J. S., Adamowski, J. F., 2019. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: a local and external data analysis approach. *J. Hydrol.* 579, 124225. <https://doi.org/10.1016/j.jhydrol.2019.124225>.
- Minns, C. K., Randall, R. G., Chadwick, E. M. P., Moore, J. E., and Green, R., 1995. Potential impact of climate change on the habitat and production dynamics of juvenile Atlantic salmon (*salmo salar*) in eastern Canada. In: Beamish, R.J. (Ed.). *Climate change and northern fish population*. *Can. Spec. Publ. Fish. Aquat. Sci.* 121, 699-708.
- Mohseni, O., and Stefan, H., 1999. Stream temperature/air temperature relationship: a physical interpretation. *J. Hydrol.* 218(3):128-141.
- Mohseni, O., Erickson, T. R., and Stefan, H. G., 1999. Sensitivity of stream temperatures in the United States to air temperatures projected under a global warming scenario. *Water Resour. Res.* 35(12):3723-3733.
- Mohseni, O., Stefan, H. G., and Eaton, J. G., 2003. Global warming and potential changes in fish habitat in US streams. *Climatic change* 59(3):389-409.
- Moore, R. D., Spittlehouse, D. L., and Story, A., 2005. Riparian microclimate and stream temperature response to forest harvesting: a review. *J. Amer. Water Resour. Assoc.* 41 (4), 813–834.
- Msilini, A., Masselot, P., and Ouarda, T.B.M.J., 2020. Regional Frequency Analysis at Ungauged Sites with Multivariate Adaptive Regression Splines. *J. Hydrometeor.* <https://doi:10.1175/jhm-d-19-0213.1>.
- Neuheimer, A. B., and Taggart, C. T., 2007. The growing degree-day and fish size-at-age: the overlooked metric. *Can. J. Fish. Aquat. Sci.* 64, 375–385.

- Neumann, D. W., Rajagopalan, B., and Zagona, E. A., 2003. Regression Model for Daily Maximum Stream Temperature. *J. Environ. Eng.* 129(7): 667-674.
- Nicieza, A. G., and Metcalfe, N. B., 1997. Growth Compensation in Juvenile Atlantic Salmon: Responses to Depressed Temperature and Food Availability, 78 (8), 2385–2400.
- Niehoff, D., Fritsch, U., and Bronstert, A., 2002. Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany. *J. Hydrol.* 267(1-2):80-93.
- Olden, J. D., and Naiman, R. J., 2010. Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. *Freshw. Biol.* 55, 86–107.
- Ouali, D., Chebana, F., Ouarda, T. B. M. J., 2016. Non-linear canonical correlation analysis in regional frequency analysis. *Stoch. Environ. Res. Risk Assess.* 30 (2), 449–462. <https://doi.org/10.1007/s00477-015-1092-7>.
- Ouarda, T. B. M. J., Ba, K. M., Diaz-Delgado, C., Carstenu, A., Chokmani, K., Gingras, H., Quentin, E., Trujillo, E., and Bobée, B., 2008. Regional flood frequency estimation at ungauged sites in the Balsas River Basin, Mexico. *J. Hydrol.* 348, 40-58. <https://doi: 10.1016/j.hydrol.2007.09.031>.
- Ouarda, T. B. M. J., Charron, C., Hundedcha, Y., St-Hilaire, A., and Chebana, F., 2018. Introduction of the GAM model for regional low-flow frequency analysis at ungauged basins and comparison with commonly used approaches. *Environ. Modell. Softw.* 109, 256-271. <https://doi.org/10.1016/j.envsoft.2018.08.031>.
- Ouarda, T. B. M. J., Charron, C., Marpu, P. R., and Chebana, F., 2016. The generalized additive model for the assessment of the direct, diffuse, and global solar irradiances using SEVIRI images, with application to the UAE. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 9 (4), 1553–1566. <https://doi.org/10.1109/jstars.2016.2522764>.
- Ouarda, T. B. M. J., Girard, C., Cavadias, G. S., and Bernard, B., 2001. Regional flood frequency estimation with canonical correlation analysis, *J. Hydrol.* 254, 157–173.
- Ouarda, T. B. M. J., Haché, M., Bruneau, P., and Bobée, B., 2000. Regional flood peak and volume estimation in northern Canadian basin, *J. Cold Reg. Eng.* 14, 176–190.
- Ouellet, V., Mingelbier, M., Saint-Hilaire, A., Morin, J., 2010. Frequency analysis as a tool for assessing adverse conditions during a massive fish kill in the St. Lawrence River, Canada. *Water Quality Research Journal of Canada* 45(1): 47-57.
- Ouellet, V., St-Hilaire, A., Dugdale, S. J., Hannah, D. M., Krause, S., and Proulx-Ouellet, S., 2020. River temperature research and practice: Recent challenges and emerging opportunities for managing thermal habitat conditions in stream ecosystems. *Sci. Total Environ.* 139679. <https://doi:10.1016/j.scitotenv.2020.1396>.

- Piccolroaz, S., Calamita, E., Majone, B., Gallice, A., Siviglia, A., and Toffolon, M., 2016. Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. *Hydrol. Process.*, 30, 3901–3917. <https://doi.org/10.1002/hyp.10913>.
- Piotrowski, A. P., and Napiorkowski, J. J., 2019. Simple modifications of the nonlinear regression stream temperature model for daily data. *J. Hydrol.*, 572, 308–328. <https://doi.org/10.1016/j.jhydrol.2019.02.035>.
- Piotrowski, A. P., Napiorkowski, M. J., Napiorkowski, J. J., and Osuch, M., 2015. Comparing various artificial neural network types for water temperature prediction in rivers. *J. Hydrol.* 529, 302–315. <https://doi.org/10.1016/j.jhydrol.2015.07.044>.
- Poff, N. L., Richter, B. D., Arthington, A. H., Bunn, S. E., Naiman, R. J., Kendy, E., et al., 2010. The ecological limits of hydrologic alteration (ELOHA): new framework for developing regional environmental flow standards. *Freshwater Biology* 55(1): 147-170.
- Poirel, A., Gailhard, J., and 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.
- Quigley, J. T., and Hinch, S. G., 2006. Effects of rapid experimental temperature increases on acute physiological stress and behaviour of stream dwelling juvenile chinook salmon. *J. Thermal Biol.* 31(5):429-441.
- Rahman, A., Charron, C., Ouarda, T. B. M. J., and Chebana, F., 2018. Development of regional flood frequency analysis techniques using generalized additive models for Australia. *Stoch. Environ. Res. Risk Assess.* 32 (1), 123–139. <https://doi.org/10.1007/s00477-017-1384-1>.
- Ramsay, T., Burnett, R., and Krewski, D., 2003. Exploring bias in a generalized additive model for spatial air pollution data. *Environ Health Perspect.* 111, 1283–1288.
- Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., et al. 2019. Process-guided deep learning predictions of lake water temperature. *Water Resour Res.* 55. <https://doi.org/10.1029/2019WR024922>.
- Reed, D., Faulkner, D., Robson, A., Houghton-Carr, H., and Bayliss, A., 1999. *Flood Estimation Handbook: Procedures for Flood Frequency Estimation*, vol. 3, Statistical Procedures for Flood Frequency Estimation, Inst. of Hydrol., Wallingford, U. K.
- Ribeiro-Corréa, J., Cavadias, G. S., Clément, B., and Rousselle, J., 1995. Identification of hydrological neighborhoods using canonical correlation analysis, *J. Hydrol.* 173, 71–89.
- Risley, J. C., Roehl, E. A., and Conrads, P. A., 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).

- Rivers-Moore, N. A., Dallas, H. F., and Morris, C., 2013. Towards setting environmental water temperature guidelines: a South African example. *J. Environ. Manag.* 128, 380–392.
- Rivers-Moore, N., Mantel, A., and Dallas, H., 2012. Prediction of water temperature metrics using spatial modelling in the Eastern and Western Cape, South Africa, *Water SA*, 32, 167–176. [https://doi: 10.4314/wsa.v38i2.2](https://doi.org/10.4314/wsa.v38i2.2).
- Rounaghi, M. M., Abbaszadeh, M. R., and Arashi, M., 2015. Stock price forecasting for companies listed on Tehran stock exchange using multivariate adaptive regression splines model and semi-parametric splines technique. *Physica A: Statistical Mechanics and its Applications* 438:625-633.
- Roy, S. S., Roy, R., and Balas, V. E., 2018. Estimating heating load in buildings using multivariate regression splines, extreme learning machine, a hybrid model of MARS and ELM. *Renew. Sust. Energ. Rev.* 82, 4256-4268.
- Rutherford, J. C., Marsh, N. A., Davies, P. M., and Bunn, S. E., 2004. Effects of patchy shade on stream water temperature: how quickly do small streams heat and cool? *Marine and Freshwater Research* 55: 737-748.
- Saadi, A. M., Msilini, A., Charron, C., St-Hilaire, A., and Ouarda, T. B. M. J., 2022. Estimation of the area of potential thermal refuges using generalized additive models and multivariate adaptive regression splines: A case study from the Ste-Marguerite River. *River Res Applic*, 38 (1), 23– 35. <https://doi.org/10.1002/rra.3886>.
- Scott, R.W., Huff, F. A., 1996. Impacts of the Great Lakes on Regional Climate Conditions. *J. Great Lakes Res.* Volume 22, Issue 4, 1996, Pages 845-863, ISSN 0380-1330. [https://doi.org/10.1016/S0380-1330\(96\)71006-7](https://doi.org/10.1016/S0380-1330(96)71006-7).
- Segura, C., Caldwell, P., Sun, G., McNulty, S., and Zhang, Y., 2014. A model to predict stream water temperature across the conterminous USA. *Hydrol. Process.* 29, 2178–2195. [https://doi: 10.1002/hyp.10357](https://doi.org/10.1002/hyp.10357).
- Seidou, O., Ouarda, T. B. M. J., Barbet, M., Bruneau, P., and Bobée, B., 2006. A parametric Bayesian combination of local and regional information in flood frequency analysis. *Water Resour. Res.* 42 (11), W11408. <https://doi.org/10.1029/2005wr004397>.
- Singer, E. E., and Gangloff, M. M., 2011. Effects of a small dam on freshwater mussel growth in an Alabama (USA) stream. *Freshw. Biol.* 56(9):1904-1915.
- Sinokrot, B. A., Stefan, H. G., 1993. Stream Temperature Dynamics: Measurement and Modeling. *Water Resources Research* 29:2299-2312.
- Sinokrot, B. A., Stefan, H. G., McCormick, J. H., and Eaton, J. G., 1995. Modeling of climate change effects on stream temperatures and fish habitats below dams and near groundwater inputs. *Climatic change*, 30 (2), 181-200.

- Stefan, H. G., Preud'homme, E., 1993. Stream temperature estimation from air temperature. *Water Resources Bulletin* 29: 27–45.
- St-Hilaire, A., Duchesne, S., and Rousseau, A. N., 2015. Floods and water quality in Canada: A review of the interactions with urbanization, agriculture and forestry. *Can. Water Resour. J/Revue canadienne des ressources hydriques (ahead-of-print)*:1-15.
- St-Hilaire, A., El-Jabi, N., Caissie, D., and Morin, G., 2003. Sensitivity analysis of a deterministic water temperature model to forest canopy and soil temperature in Catamaran Brook (New Brunswick, Canada). *Hydrol. Process.* 17, 2033–2047.
- St-Hilaire, A., Morin, G., El-Jabi, N., and Caissie, D., 2000. Water temperature modelling in a small forested stream: implication of forest canopy and soil temperature. *Can. J. Civil Eng.* 27(6):1095-1108.
- Sundt-Hansen, L. E., Hedger, R. D., Ugedal, O., Diserud, O. H., Finstad, A. G., Sauterleute, J. F., Tøfte, L., Alfredsen, K., and Forseth, T., 2018. Modelling climate change effects on Atlantic salmon: Implications for mitigation in regulated rivers. *Sci. Total Environ.* 631–632, 1005–1117. [https://doi: 10.1016/j.scitotenv.2018.03.058](https://doi.org/10.1016/j.scitotenv.2018.03.058).
- Tasker, G. D., Hodge, S. A., and Barks, C. S., 1996. Region of influence regression for estimating the 50-year flood at ungauged sites. *Water Resour Res.* 1(32), 163–170.
- Van Vliet, M. T. H., Franssen, W. H. P., Yearsley, J. R., Ludwig, F., Haddeland, I., Lettenmaier, D. P., and Kabat, P., 2013. Global river discharge and water temperature under climate change. *Global Environ Chang.* 23(2), 450–464. <https://doi.org/10.1016/j.gloenvcha.2012.11.002>.
- Van Vliet, M., Ludwig, F., Zwolsman, J., Weedon, G., and Kabat, P., 2011. Global river temperatures and sensitivity to atmospheric warming and changes in river flow. *Water Resour. Res.* 47(2).
- Vannote, R. L., and Sweeney, B. W., 1980. Geographic analysis of thermal equilibria: a conceptual model for evaluating the effect of natural and modified thermal regimes on aquatic insect communities. *The American Naturalist*, 115, 667–695.
- Ward, J., 1963. Annual variation of stream water temperature. *ASCE, J. Sanit. Eng. Div.* 89, 3710–3732.
- Ward, J., 1982. Ecological aspects of stream regulation: responses in downstream lotic reaches. *Water Pollution and Management Reviews*, (2), 1–26.
- Ward, J.V. 1985. Thermal characteristics of running waters. *Hydrobiologia*, 125, 31–46.
- Webb, B., 1996. Trends in stream and river temperature. *Hydrol. Process.*, 10 (2), 205–226. [https://doi: 10.1002/ \(ISSN\) 1099-1085](https://doi.org/10.1002/(ISSN)1099-1085).
- Webb, B., and Walling, D., 1997. Complex summer water temperature behaviour below a UK regulating reservoir. *Regulated rivers: research & management* 13(5):463-477.
- Webb, B., Walling, D., 1993. Temporal Variability in the Impact of River Regulation on Thermal Regime and Some Biological Implications. *Freshw. Biol.* 29 (1): 167–82.

- Webb, B.W., Hannah, D. M., Moore, R. D., Brown, L. E., and Nobilis, F., 2008. Recent advances in stream and river temperature research. *Hydrol. Process.* 22(7), 902-918.
- Wehrly, K. E., Brenden, T. O., and Wang, L., 2009. A comparison of statistical approaches for predicting stream temperatures across heterogeneous landscapes. *J. Am. Wat. Res. Assoc.* 45:986–997.
- Wehrly, K. E., Wiley, M. J., Seelbach, P.W., 2003. Classifying Regional Variation in Thermal Regime Based on Stream Fish Community Patterns. *Transactions of the American Fisheries Society* 132:18-38.
- Wen, L., Rogers, K., Saintilan, N., and Ling, J., 2011. The influences of climate and hydrology on population dynamics of waterbirds in the lower Murrumbidgee River floodplains in Southeast Australia: implications for environmental water management. *Ecol. Model.* 222, 154–163. <https://doi.org/10.1016/j.ecolmodel.2010.09.016>.
- Wood, S. N., 2006. *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC Press, London.
- Yang, K., Yu, Z., Luo, Y., Zhou, X., and Shang, C., 2019. Spatial-Temporal Variation of Lake Surface Water Temperature and its Driving Factors in Yunnan-Guizhou Plateau. *Water Resour Res.* <https://doi:10.1029/2019wr025316>.
- Younus, M., Hondzo, M., Newton, M., 2000. Stream Temperature Dynamics in Upland Agricultural Watersheds. *ASCE, Journal of Environmental Engineering* 126:518-526.
- Zhang, W., Goh, A., 2016. Evaluating seismic liquefaction potential using multivariate adaptive regression splines and logistic regression. *Geomech. Eng.* 10(3):269-280.
- Zhang, W., Goh, A.T., Zhang, Y., Chen, Y., and Xiao, Y., 2015. Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines. *Eng. Geol.* 188: 29-37.
- Zhu, S., and Piotrowski, A. P., 2020. River/stream water temperature forecasting using artificial intelligence models: a systematic review, *Acta Geophysica*, 1–10, Springer. <https://doi.org/10.1007/s11600-020-00480-7>, 2020.
- Zhu, S., Hadzima-Nyarko, M., Gao, A., Wang, F., Wu, J., and Wu, S., 2019a. Two hybrid data-driven models for modeling waterair temperature relationship in rivers. *Environ. Sci. Pollut. R.*, 26, 12622–12630. <https://doi.org/10.1007/s11356-019-04716-y>.
- Zhu, S., Heddiam, S., Nyarko, E. K., Hadzima-Nyarko, M., Piccolroaz, S., and Wu, S., 2019b. Modeling daily water temperature for rivers: comparison between adaptive neuro-fuzzy inference systems and artificial neural networks models. *Environ. Sci. Pollut. R.*, 26, 402–420. <https://doi.org/10.1007/s11356-018-3650-2>.
- Zhu, S., Nyarko, E. K., and Hadzima-Nyarko, M., 2018. Modelling daily water temperature from air temperature for the Missouri River, *PeerJ*, 6, e4894. <https://doi.org/10.7717/peerj.4894>.

