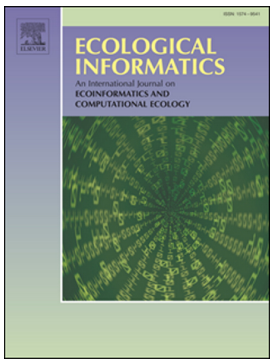

Regional thermal analysis approach: A management tool for predicting water temperature metrics relevant for thermal fish habitat

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

PII: S1574-9541(22)00142-X

DOI: https://doi.org/10.1016/j.ecoinf.2022.101692

Reference: ECOINF 101692

To appear in: *Ecological Informatics*

Received date: 24 January 2022

Revised date: 21 May 2022

Accepted date: 21 May 2022

Please cite this article as: O. Abidi, A. St-Hilaire, T.B.M.J. Ouarda, et al., Regional thermal analysis approach: A management tool for predicting water temperature metrics relevant for thermal fish habitat, *Ecological Informatics* (2021), https://doi.org/10.1016/j.ecoinf.2022.101692

Title: 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

May 21, 2022

George Arhonditsis

Editor-in-Chief, Ecological Informatics

Dear Editor,

My coauthors and I would like to thank you for considering our article for publication in Ecological Informatics. We are grateful to you and the reviewers for the valuable feedback and constructive suggestions which contributed to the improvement of the quality of the paper.

All formulated comments were addressed, and appropriate modifications were made in the revised version of the manuscript.

Sincerely,

Olfa Abidi (Olfa.Abidi@inrs.ca)

André St-Hilaire (Andre.St-Hilaire@inrs.ca)

Taha B. M. J. Ouarda (Taha.Ouarda@inrs.ca)

Abstract

The field of 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 is an important indicator 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). 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, the regional models were compared to models encompassing all stations (i.e., one region). For each regional estimation model and each thermal metric, a set of optimal explanatory variables were selected using a forward stepwise 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 efforts 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. 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 to high temperatures (e.g., Elliott, 1991; Jonsson and Jonsson, 2009). 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 egg incubation is the life stage that is most affected by temperature. 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, Heggenes et al. (2021) showed that regular occurrences of high stream temperatures (> 22.5 °C) 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 U.S. 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 U.S. 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 the 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 the 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. 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/thermal data at sites with scarce or no information is a very interesting alternative technique for water resources planning and integrated management (Babaei et al., 2019; Mehdizadeh et al., 2019).

Numerous previous studies have 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 few models were developed at the regional scale.

To overcome this shortcoming, some studies have recently attempted to develop regional approaches. 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 Ducharne (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 a 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 and often 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 homogeneous groups. A simple approach consists of defining groups of geographically contiguous stations with similar temperature regimes. An alternative to this approach is based on similarities other than location. This approach, used in the present study, allows us to define regions based on the climatic and physiographic characteristics of the watersheds (Ouarda et al., 2001). 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 identifying predictors which significantly affect water temperature metrics and defining which model best fits our response variables by evaluating combinations for regional thermal analysis prior to 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. The region of interest and associated datasets 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 selected climatic and physiographic variables. The standardized Euclidean distance is defined by:

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

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 using Ward's algorithm (Ward, 1963). It is based 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 completed in ascending order, starting with all stations in separate groups and coalescing them (Johnson, 1967).

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

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):

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

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 model 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:

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

With Dij 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 $v_0 \backsimeq \Lambda u_0$ with $\Lambda = diag(\rho_1, \dots, \rho_k)$. 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, the Mahalanobis distance is used (Equation (5)) in the multidimensional space defined by the canonical physiographic, meteorological and hydrological variables (Ribeiro-Corréa et al., 1995) as follows:

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

Where d represents 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 physiographic and meteorological canonical 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 useful 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 $\lambda_i = corr(U_i, V_i)$ is maximized by imposing a unit variance. It is important to mention that these are linear transformations of the original variables *X* and *Y*:

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

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

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

CCA makes it possible to identify the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ 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):

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

where $Tw\,(t)$ represents the water temperature metric; βi 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). The method does not assume any specific form of dependence

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

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

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 (Leitão 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. 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 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 Equation (13) and the percent bias (PBIAS) (Equation (14)), which are measures of overestimation or underestimation of a model.

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

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

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

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

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

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 sample size.

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 1).

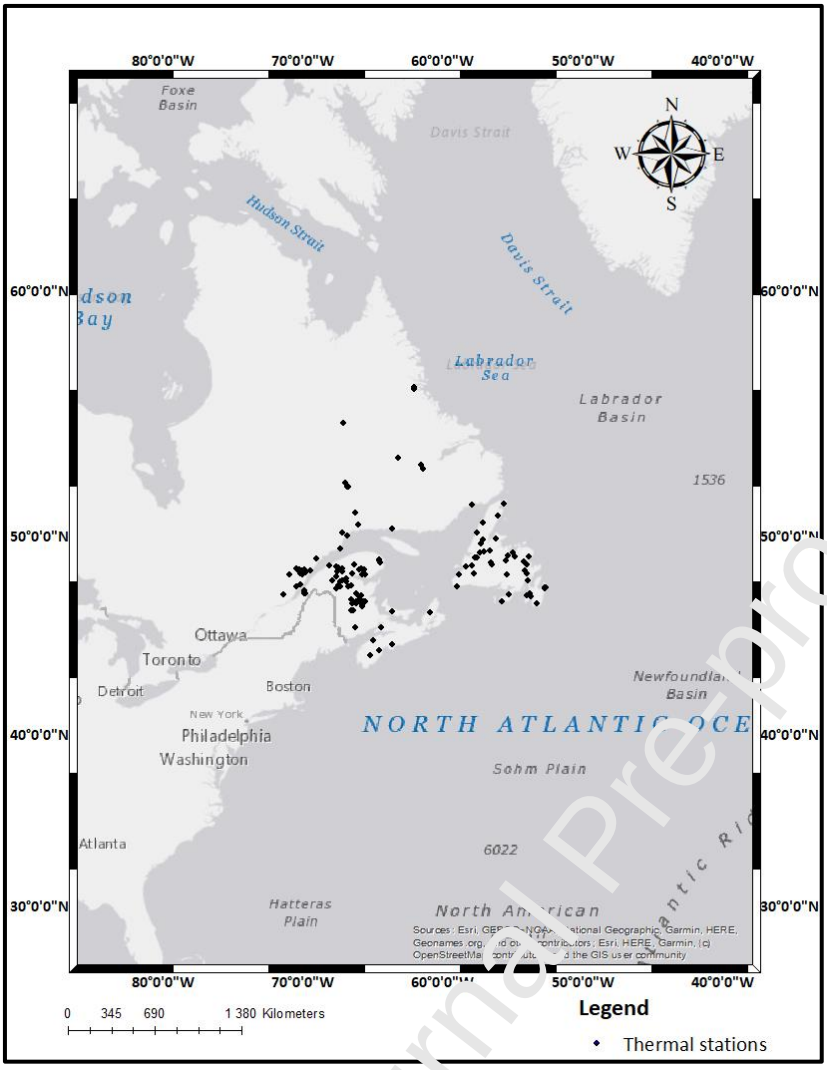

Figure 1: Location of thermal stations across the study area.

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 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 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 at 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 lists the selected physiographic and climatic variables as well as some of their statistical characteristics across all 146 stations.

Variables	Description	Unit	Mean	Median	Minimum	Maximum
Physiographic variables						
Basin Area	Catchment area	km^2	1099	360	0.75	25444
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	329.27	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: Physiographic and climatic variables and some of their statistical characteristics.

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

Variables	Description	Unit	Mean	Median	Minimum	Maximum
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	%	4.20	1.19	0.00	34.39

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

The variables presented in tables 2 and 3 were selected through a 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 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 carried out 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, individually for each model. The top six most important predictors were selected (Table 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 Station	Slope	GlacialDeposists

Table 4: Selected predictors when all stations are grouped together.

4.1 Selection of explanatory variables for MLR

As expected, at least one air temperature variable has been found to be substantially 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, throughout 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 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 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).

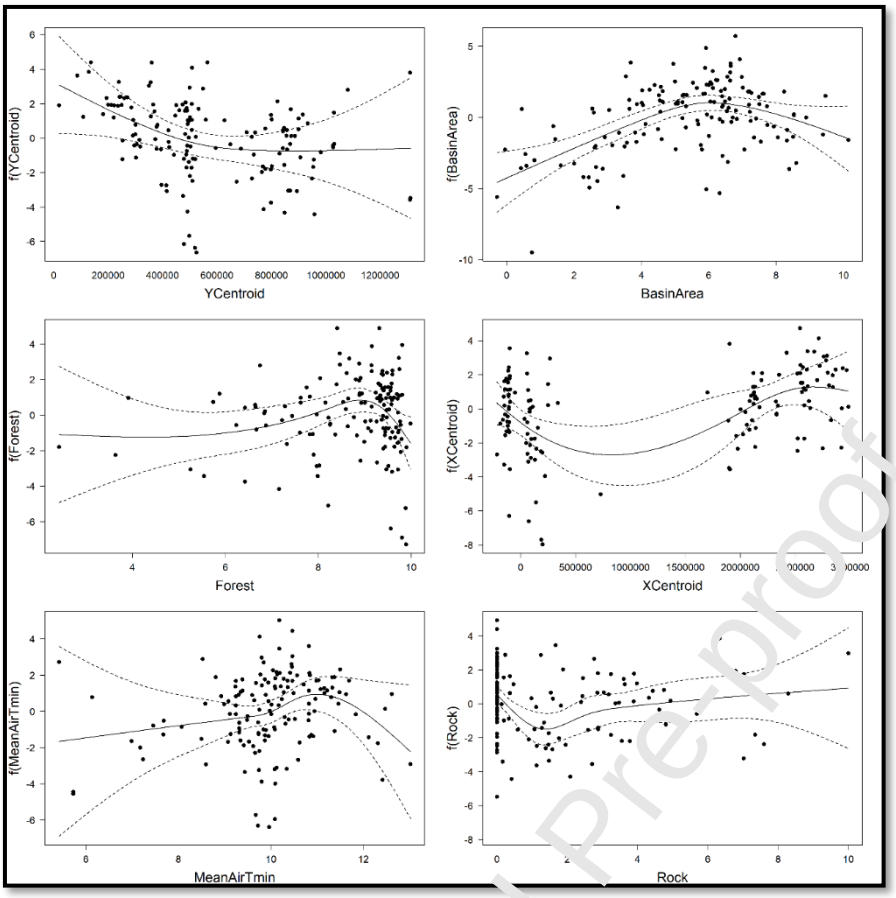

Figure 2. Splines for the predictors used in the ALL-GAM model (e.g., MaxWaterTmax).

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 3 provides an example of a dendrogram for the variable MaxWaterTmax in the case of MLR. When the truncation was performed at an Euclidean distance of 15.3, three thermally homogeneous regions were identified.

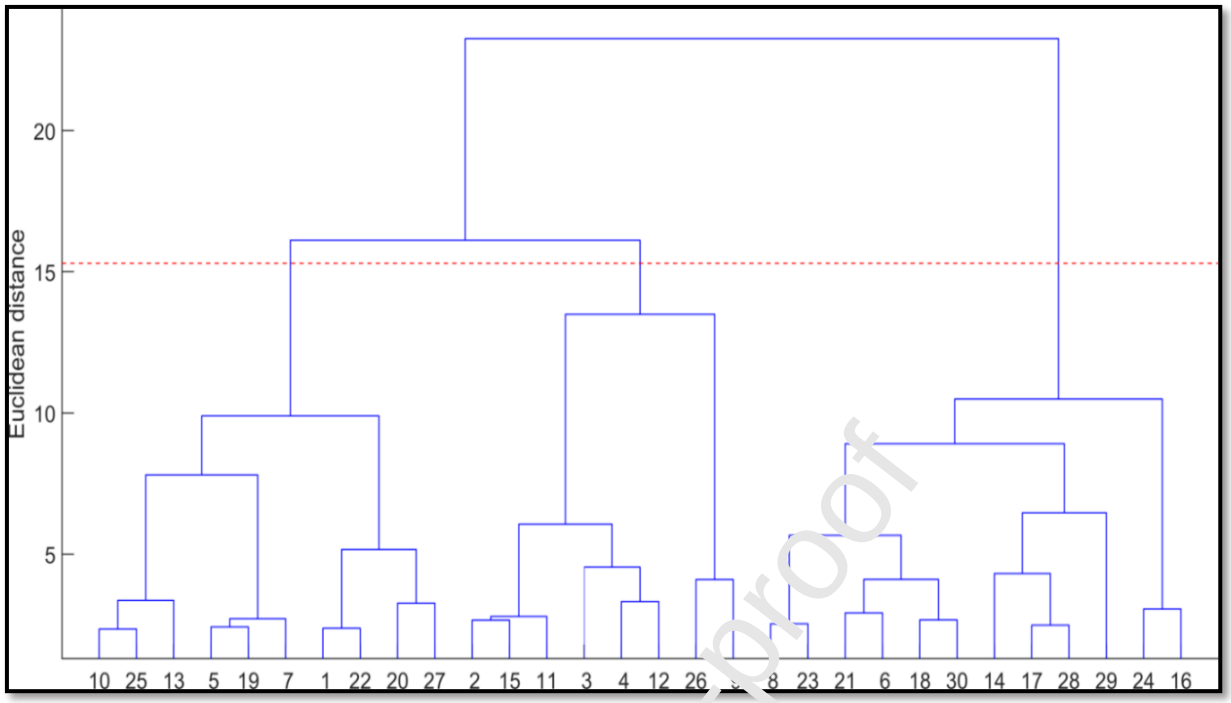

Figure 3. Dendrogram with cut-off threshold for the variable MaxWaterTmax. Note that station numbers are shown on the x-axis.

The spatial distribution of homogeneous groups generated by the HCA for each water temperature metric is illustrated in Figure 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 sizes were fixed at 97 and 90, respectively. Overall, ROI and CCA show practically the same neighborhoods to a reference site.

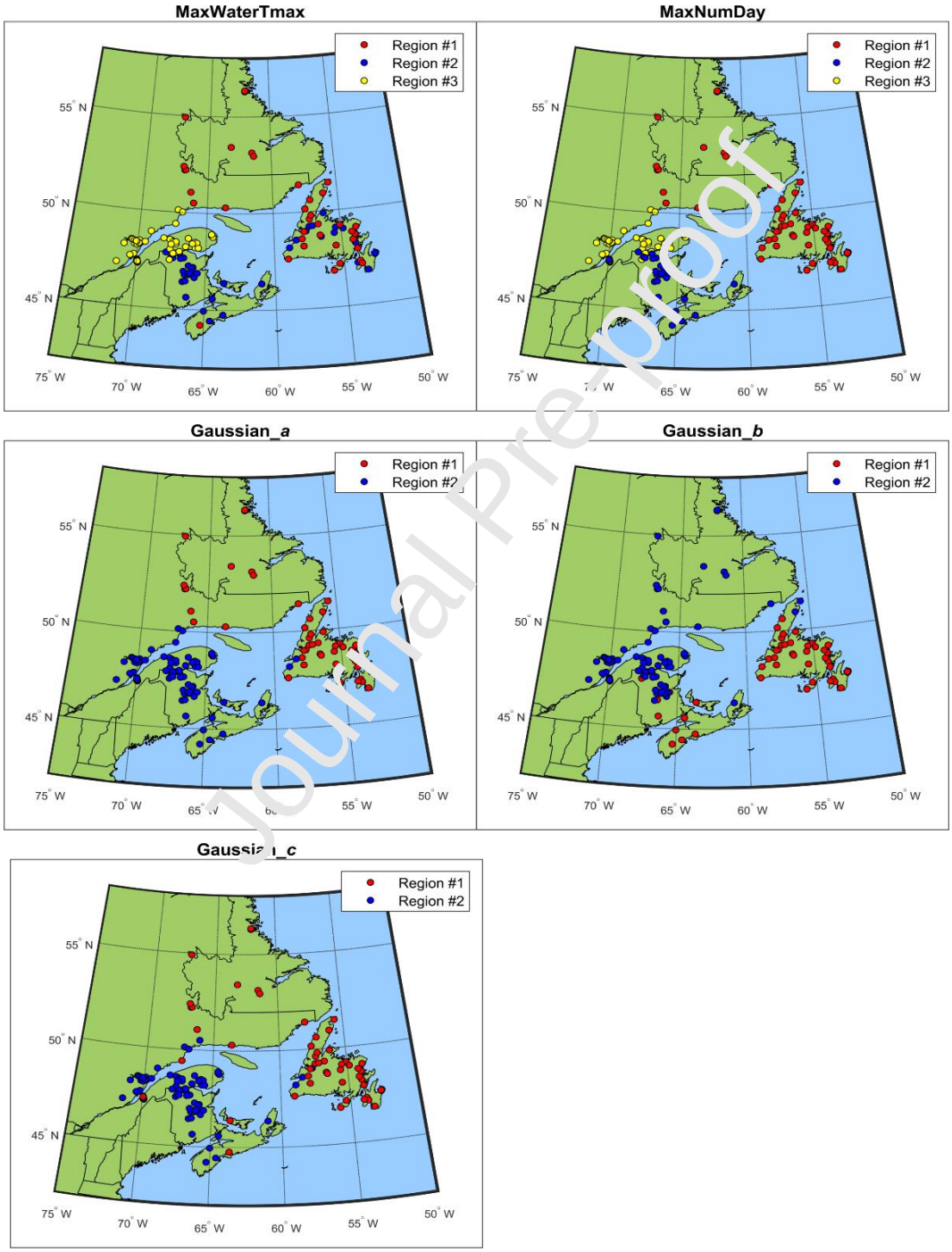

Figure 4. Maps of stations grouped using HCA-MLR for each thermal metric.

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 5.

Performance criteria	metrics	ALL+ MLR	HCA+MLR	CCA+MLR	ROI+MLR	ALL+GAM	HCA+GAM	CCA+GAM	ROI+GAM
R^2	MaxWaterTmax	0.38	**0.48**	0.37	0.42	0.39	0.39	0.42	0.34
	MaxNumDay	0.33	0.48	0.37	0.41	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 5: Performance indicators for different regional thermal analysis in Atlantic salmon rivers.

Note: best results are in bold character.

It is clear from Table 5 that the GAM model applied to all stations in our study area, i.e., without delineation of thermally homogeneous regions leads 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 is in accordance with the results obtained from the spline curves in Figure (2) indicating that a non-linear model is more suitable than the MLR model. Also, when the GAM model is tested with the various regional delineation approaches, a 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 coefficients of determination R^2 .

The scatter plots in Figure 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.

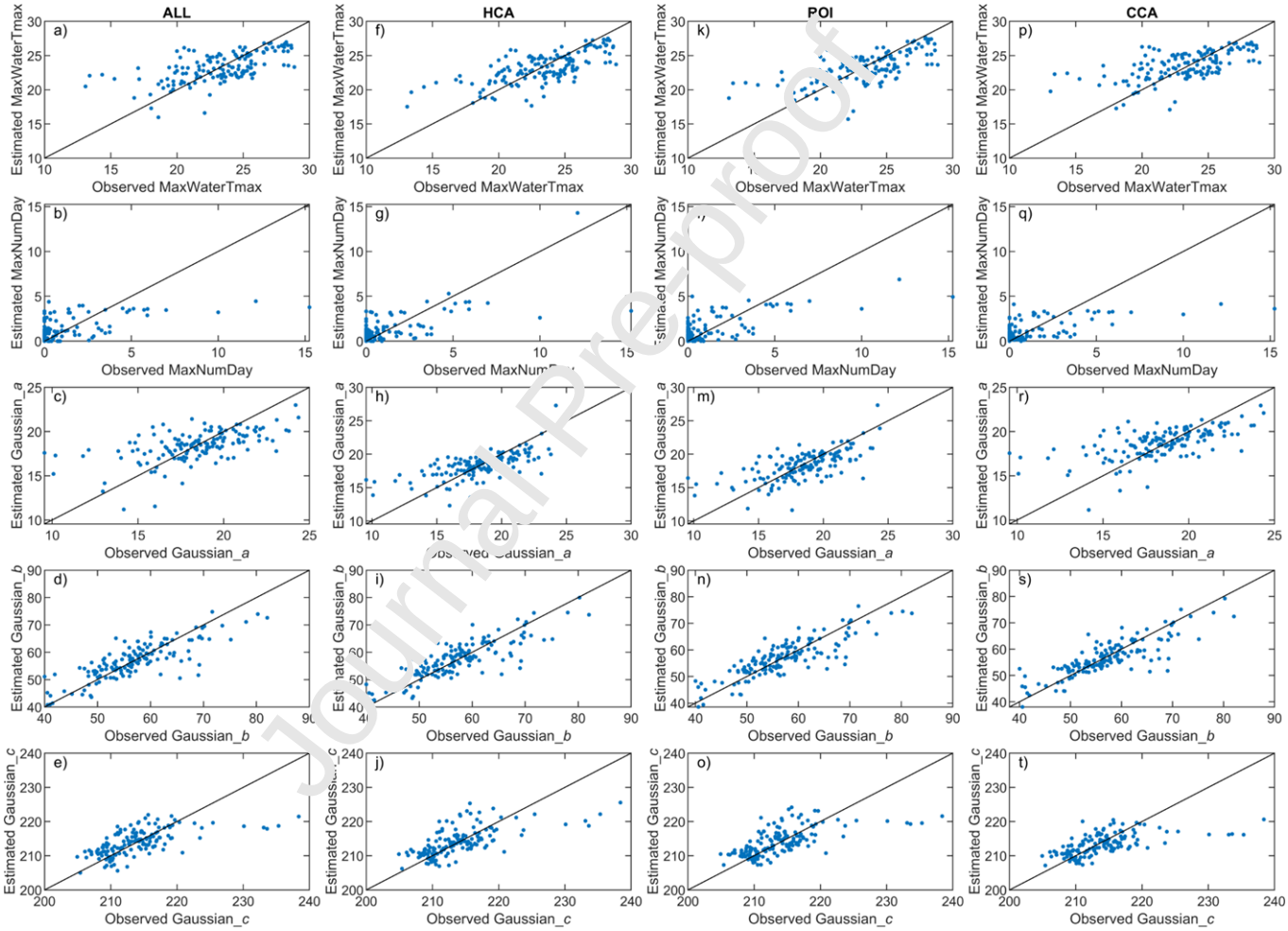

Figure 5: Scatter plots of observed vs simulated water temperature metrics for MLR.

For instance, MaxNumDay 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 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 °C and minimum water temperature > 20 °C. The parameter *c* of the Gaussian function is underestimated for values above 220 days. Again, Figure 6 shows that there are very few values > 220 days.

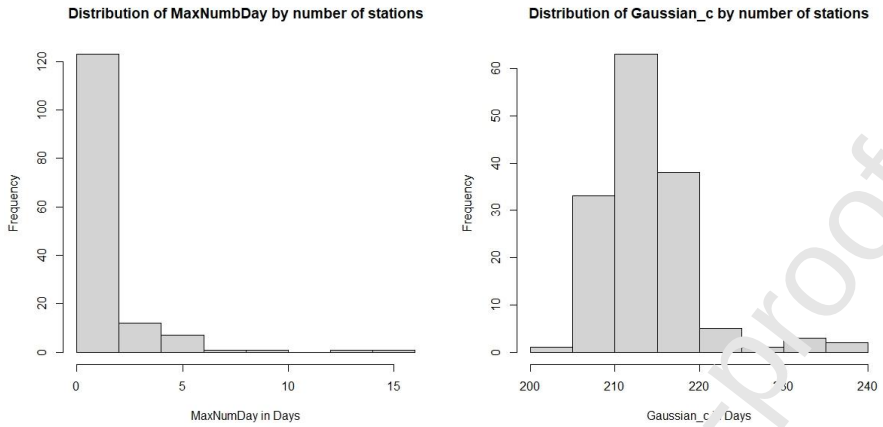

Figure 6: Histograms showing the frequency distribution of the variables MaxNumDay and the parameter c of the Gaussian function.

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 were 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, a 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 that exist 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 from 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.

Aside from 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. Eventually, additional 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 model 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. The 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 approaches 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 proven 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 in the 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, and
- 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, the 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. The success of such an approach could also 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 lakes 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 B.M.J. 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). The authors wish to express their appreciation to Dr. George Arhonditsis, Editor in Chief, Associate Editors, and two anonymous reviewers for their insightful comments and suggestions which helped considerably improve the quality of the paper.

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.

References

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. Limnol. Oceanogr., 54(6part2), 2283-2297. https://doi.org/10.4319/lo.2009.54.6_part_2.2283.

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.

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.

Basu, B., and Srinivas, V. V., 2014. Regional flood frequency analysis using kernel-based fuzzy clustering approach. Water Resour. Res. 50, 3205–3216. https://doi.org/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.

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.

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.

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.

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.org/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. https://doi.org/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.org/10.1093/conphys/cox014.

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.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.

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.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.org/10.1016/j.jhydrol.2014.05.035, 2014.

Ducharne, A., 2008. Importance of stream temperature to climate change impact on water quality, Hydrol. Earth Syst. Sci. 12, 797–810. 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.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.org/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.

Feigl, M., Lebiedzinski, 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., Schaefli, 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.

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.jhydr ol.2017.03.024.

Gregr, E. J., and Trites, A. W., 2001. Predictions of critical habitat for whale species in the waters of coastal British Colombia. Can. J. Fish. Aquat. Sci. 58, 1265–1285.

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.org/10.1080/02626667.2014.950583.

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.

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.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.

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.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.org/10.1016/j.jenvman.2016.12.005.

Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K., Marshall, L., 2012. Inves- tigating controls on the thermal sensitivity of Pennsylvania streams. Hydrol. Process. 26:771_785. https://doi.org/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.org/10.1080/02626667.2016.1246799.

Leach, J. A., Neilson, B. T., Bechtel, 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.

Leitte, A. M., Petrescu, C., Franck, U., Richter, M., Suciu, 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.

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. https://doi.org/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.org/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.

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: Seatte.

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.

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.org/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.

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.

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.org/10.1016/j.hydrol.2007.09.031.

Ouarda, T. B. M. J., Charron, C., Hundecha, 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., 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.org/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.

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.

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., 2011. Prediction of water temperature metrics using spatial modelling in the Eastern and Western Cape, South Africa, Water SA, 32, 167–176. https://doi.org/10.4314/wsa. v38i2.2.

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

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.

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.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.

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.

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.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(3), 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.

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.

Webb, B., 1996. Trends in stream and river temperature. Hydrol. Process., 10 (2), 205–226, https://doi.org/10.1002/ (ISSN) 1099-1085.

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.org/10.1029/2019wr025316.

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., Heddam, 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.

Figure captions

Supplementary Materials

Acronyms

AIC: Akaike information criterion

AirTmax: Maximum air temperature

AirTmin: Minimum air temperature

ANN: Artificial neural networks

C

CCA: Canonical correlations analysis

D

DHR: Delineation of homogeneous regions

G

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

H

HCA: Hierarchical clustering analysis

M

MARS: Multivariate adaptive regression splines

MaxElevation: Maximum elevation of the catchment

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

MaxWaterTmax: Interannual mean of maximum temperature

MeanElevation: Mean elevation of the catchment

MinElevation: minimum elevation of the catchment

MLR: Multiple linear regression

P

PBIAS: Percent bias

R

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

RTA: Regional thermal analysis

T

TotPrecip: Total precipitation

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

- Significance of the regional analysis approach
- Identification of thermally homogeneous regions
- Selection of significant environmental drivers
- Limits of the models for the estimation of the interannual mean of the number of consecutive days above a potentially stressful threshold for Atlantic salmon