Regional stream temperature modeling in pristine Atlantic salmon rivers: A hybrid deterministic–Machine Learning approach.

Authors:

Ilias Hani [1,2]

André St-Hilaire [1,2]

Taha B.M.J. Ouarda [1]

[1] Canada Research Chair in Statistical Hydro-Climatology, Institut national de la recherche scientifique, Centre Eau Terre Environnement. INRS-ETE, 490 De la Couronne, Québec City, QC, Canada

[2] Canadian Rivers Institute, University of New Brunswick, Fredericton, Canada

Corresponding author: Ilias.Hani@inrs.ca

Keywords

Regional model; Water temperature, Sensitivity analysis; Ungauged basins; deterministic model; Machine-learning; Reanalysis.

Abbreviations

Tw	Stream water temperature
Tair	Air temperature
MP	Model parameter
VARS	Variogram Analysis of Response Surfaces
IVARS	Integrated Variograms Across a Range of Scale
ML	Machine Learning
MLR	Multiple linear regression
SVR	Support vector regression
FS	Feature selection
RFE	Recursive feature elimination
SFS	Stepwise forward selection
LASSO	Least Absolute Shrinkage and Selection Operator
ENET	Elastic net
CMA-ES	Covariance matrix adaptation evolution strategy
RMSE	Root Mean Square Error

Abstract:

Study Region

This study focuses on the thermal regime of pristine Atlantic salmon rivers located across northeastern Canada and the USA. These rivers are critical habitats for Atlantic salmon and represent a diverse range of climatic and watershed conditions in the region.

Study Focus

To simulate water temperature in ungauged rivers, we explore the regionalization of thermal parameters within the CEQUEAU model—a deterministic, semi-distributed hydrological and water temperature model. Additionally, a global sensitivity analysis is conducted to identify the most sensitive thermal parameters within the study region. We employed the support vector regression algorithm (SVR), to map the dependence of these parameters with climatic and watershed characteristics.

New Hydrological Insights for the Region

Parameters controlling radiative and sensible heat fluxes are the most critical for CEQUEAU water temperature modeling within the study region. Key explanatory variables include low cloud coverage, high wind speed quantiles, upstream land cover percentages, distance to the coast, watershed orientation, and topographical features describing surface curvature and elevation. Using a leave-one-out cross validation, SVR outperformed the commonly used multiple linear regression (MLR) method, with a mean regional RMSE of 1.89 °C. The machine learning-based regionalization approach offers a robust alternative for deriving water temperature model parameters from watershed attributes, especially in ungauged regions where flow measurements are available.

1-Introduction

Adequate thermal conditions in rivers are essential for the well-being of aquatic ecosystems and the survival of various aquatic species (Dugdale al., 2017b; FitzGerald & Martin, 2022; Johnson et al., 2024). In many regions of the world, human activities and climate change are rapidly increasing river temperatures, which is particularly alarming for cold fish-bearing watersheds (Boyer et al., 2021; Ficklin et al., 2023; Isaak & Luce, 2023; St-Hilaire et al., 2023; Wanders et al., 2019). Among the many species affected, the plight of the iconic Atlantic salmon (*Salmo salar*) is particularly alarming. These species are experiencing heightened stress and mortality rates, posing a significant threat to their existence and the overall biodiversity of our rivers (Gillis et al., 2023; Hani et al., 2023; Railsback & Harvey, 2023; St-Hilaire et al., 2021).

To effectively model these habitats in the context of climate change, deterministic models are gaining increasing attention (Khorsandi et al., 2023; Oyinlola et al., 2023; Wade et al., 2024). This is particularly crucial in Canada, where most river systems lack sufficient water temperature (Tw) data (Boyer et al., 2016; Shrestha et al., 2024; St-Hilaire et al., 2018). When combined with climatic reanalysis data, deterministic models can expand their geographic applicability and improve discharge and Tw estimates (Bosmans et al., 2022; Gatien et al., 2022; Rincón et al., 2023). In many instances, these models encompass a wide range of parameters to assist in accurately representing hydrological and thermal dynamics (Hay et al., 2023; St-Hilaire et al., 2015). Water temperature deterministic modeling aims at solving a heat budget and therefore relies on various inputs, including hydrological and Tw observations, land cover, and climatic forcing. However, parameterizing a deterministic model on a large scale or in remote areas poses significant challenges due to the scarcity of Tw measurements. Nonetheless, research indicates that basin characteristics can elucidate broad-scale patterns in Tw, suggesting that these relationships can be effectively transferred between catchments (Jackson et al., 2017; St-Hilaire et al., 2019; Wade et al., 2023). Recent studies have shown that river thermal regimes result from a complex interplay of physio-climatic attributes, with each stream temperature reflecting a unique combination of these characteristics (Abidi et al., 2022; Charron et al., 2021; Loerke et al., 2023; Souaissi et al.,, 2023a). This concept is closely associated with parameter regionalization, which involves transferring deterministic model parameters from gauged to ungauged locations using regression models (Clark et al., 2017; Gallice et al., 2015). However, deterministic modeling can encounter significant challenges in relating landscape properties to hydrological behavior due to the problem of equifinality (Feigl et al., 2020; Hundecha & Bárdossy, 2004; Samaniego et al., 2010).

The process of regionalization is often considered as a twofold process: first, the identification of homogeneous regions based on physio-climatic and hydrological conditions, and second, the use of different statistical models, including machine learning (ML) regression techniques to transfer information from gauged to ungauged locations (Abidi et al., 2022; Ouali et al., 2016; Ouarda et al., 2001). This second step often requires a significant number of explanatory variables to achieve accurate predictions (Ouarda et al., 2018). However, a common pitfall in model design is the inclusion of redundant variables, known as feature selection bias, as this can negatively impact the model's effectiveness (Varadharajan et al., 2022). While many studies use expert knowledge-based feature selection and simple correlation tests to address collinearity, they often overlook the interactions between features.

In deterministic modeling, multiple linear regression (MLR) is commonly utilized to establish the relationship between basin characteristics and model parameters (Arsenault et al., 2019; Feigl et al., 2021; Pagliero et al., 2019; Song et al., 2022). However, the assumption of a linear relationship may not accurately capture the complex interactions between parameters and catchment attributes (Hundecha et al., 2008). Additionally, multicollinearity among predictors in MLR models can compromise their predictive effectiveness and interpretability, potentially leading to overfitting (Maxwell & Shobe, 2022). Hydrological studies often reveal intricate, nonlinear, and nonstationary relationships between physical catchment attributes and model parameters, indicating that simpler models may not adequately reflect underlying hydrological dynamics (Guo et al., 2021; Kuczera &

Mroczkowski, 1998; Yang et al., 2018). To address these challenges, feature selection algorithms (FS) are applied as a preprocessing step to refine the feature subset by mitigating multicollinearity issues (Gharib & Davies, 2021; Guyon & Elisseeff, 2003; Wade et al., 2023). Two potential downsides of incorporating irrelevant or redundant features are overfitting the model's parameters to the training data and compromising the model’s interpretability (Quilty & Adamowski, 2020).

Comparative studies on FS methods have been widely conducted for environmental variables. For example, Fouad & Loáiciga. (2020) compared seven FS methods for river flow quantile estimation in ungauged basins found that FS methods outperformed dimension reduction techniques, such as principal component analysis, in reducing multicollinearity in feature subsets. However, few studies have evaluated the performance of FS methods for river thermal regime estimations in ungauged watersheds, as the number of available predictors increases, along with the risk of redundancy and overfitting. Among the exceptions, Souaissi et al. (2023a) assessed the ability of Recursive Feature Elimination (RFE) and the Least Absolute Shrinkage and Selection Operator (LASSO) in a regional modeling context of Tw river quantiles in Switzerland.

ML models, such as artificial neural networks and Gradient Boosting Machines (GBM), have shown advantages over conventional MLR in regionalizing deterministic hydrological model parameters (Heuvelmans et al., 2006; Song et al., 2022)., Most studies focus solely on transferring thermal signatures from gauged to ungauged watersheds (Abidi et al., 2022; Charron et al., 2020; Souaissi, Ouarda, & St-Hilaire, 2023a, 2023b). Ouarda et al. (2022) were the first to develop a statistical regional framework for Tw modeling. They used Generalized Additive Model (GAM) to transfer temperature duration curves (TDCs) from gauged to ungauged locations in eastern Canada. The estimated TDCs were combined with spatial interpolation to obtain daily Tw estimates at ungauged sites, effectively combining daily data from multiple sources as a data augmentation technique. Building on this research, Siegel et al, (2023) used a GAM model and physio-climatic covariates to estimate daily Tw for over 200,000 stream reaches across the Pacific Northwest USA, achieving

RMSEs below 2°C. In another study, Weierbach et al. (2022) developed a regional ML framework to predict monthly Tw in pristine and human-impacted catchments across the Mid-Atlantic and Pacific Northwest USA regions. The study employed physio-climatic catchment attributes and ML models such as extreme Gradient Boost Machines (XGboost) and Support vector regression (SVR). These models outperformed traditional MLR and accurately predicted Tw in both temporal and spatial scenarios.

However, a significant knowledge gap persists in regional Tw modeling using deterministic models. The limited research using deterministic models includes innovative contributions like those from Gallice et al. (2015), who developed a hybrid regional model to predict temperatures at ungauged sites in Switzerland. Building on this work, Rahmani et al. (2023) recently introduced a cutting-edge framework that integrates neural networks with process-based models via differentiable programming, significantly enhancing the accuracy of Tw parameter estimation (Bindas et al., 2024; C. Shen et al., 2023; Tsai et al., 2021). This methodological evolution highlights a growing recognition of ML's potential to provide nuanced insights into complex hydrological processes. Incorporating spatially distributed geophysical catchment properties into model parameter definition would improve model performance, reduce uncertainty, and enable predictions in ungauged basins (Feigl et al., 2020).

As part of the Atlantic Salmon Research Joint Venture (www.asrjv.com), this study assesses the potential for transferring deterministic Tw model parameters from gauged to ungauged locations to estimate daily mean Tw in Atlantic salmon bearing watersheds. To achieve this, we used CEQUEAU a deterministic hydrological and thermal model and selected 35 pristine Atlantic salmon rivers in northeastern Canada and the U.S (Morin & Paquet, 1995; St-Hilaire et al., 2015). To guide our regionalization efforts, we conducted a sensitivity analysis to identify critical parameters within the study region. Subsequently, ML regression models are employed to establish relationships between key physio-climatic characteristics and CEQUEAU most critical thermal parameters.

2- Methods

Figure 1 provides an overview of the methodology employed in this study. The process begins with calibrating the CEQUEAU hydrological module to simulate daily streamflow. A sensitivity analysis is then conducted to identify the most critical CEQUEAU Tw model parameters (MP) for the study region. These parameters are regionalized using the Support Vector Regression (SVR) algorithm and physio-climatic attributes describing watershed characteristics. The regionalized parameters are subsequently injected into the CEQUEAU model to estimate daily mean water temperatures at ungauged locations. Leave-One-Out cross validation is used to evaluate the generalizability and performance of this approach.

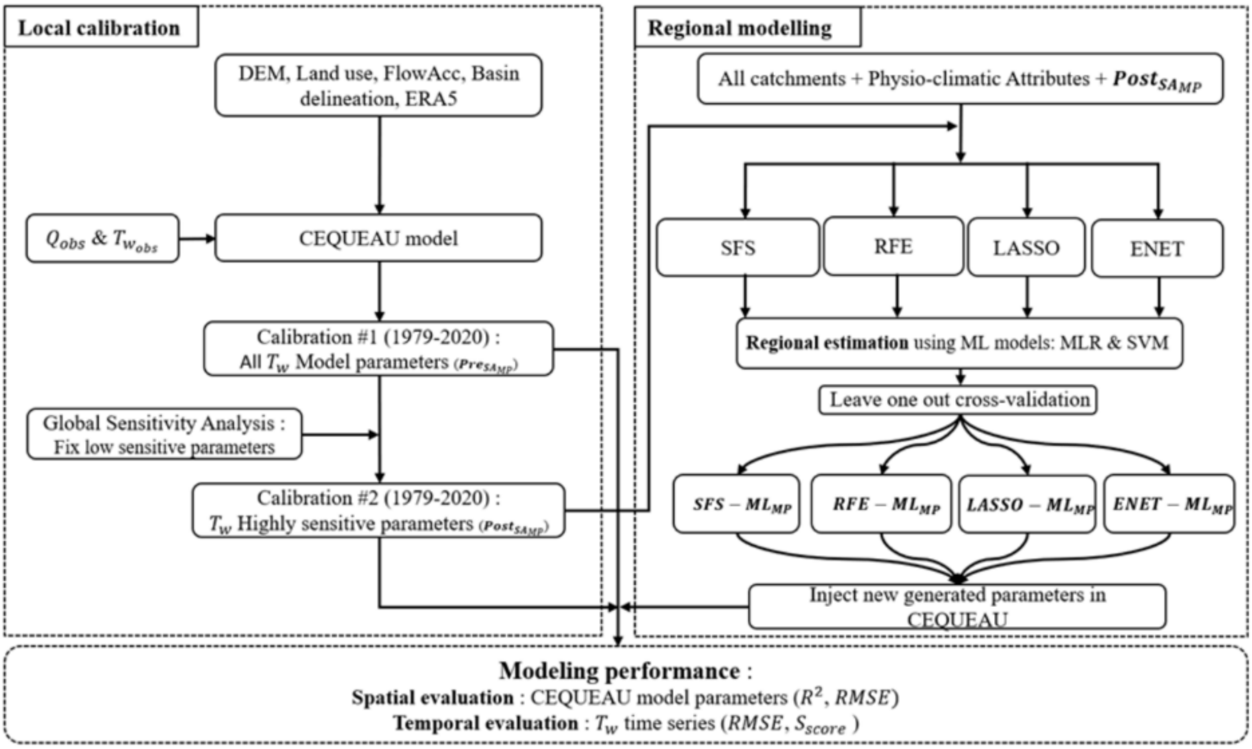

Figure 1: Flowchart of CEQUEAU thermal parameter calibration and regionalization. SFS, RFE, LASSO, and ENET– Feature selection algorithms, MLR –multiple linear regression; SVR –support vector regression; RMSE – Root mean square error; R^2– Coefficient of determination; S_{score} improvement skill score.

2.1- Study region and datasets

This study focuses on the thermal regimes of rivers that are home to significant Atlantic salmon populations. Catchments were selected based on two criteria: minimal human impact (near-pristine rivers) and availability of at least two summers of observed Tw data (May 1st to September 1st). Discharge and Tw stations were required to be on the mainstem as far downstream as possible, with only one station per stream channel within the same drainage network. Where multiple stations existed on the same river segment, the station with the longest record was retained to ensure data independence and prevent simplistic transfer of Tw data from upstream stations during regionalization. This process resulted in a set of 35 rivers, with Tw station locations and channel widths depicted in Figure 2a, and elevation and drainage area in Table 1. The study's geographical scope extends from the subarctic Canadian Aux Mélèzes (QC) and Reid (NL) rivers to the southern edge distribution of Atlantic salmon in North America, the Sheepscot River (ME) (Kocik et al., 2022), covering the north-south axis and from the Conne River (NL) to the Gouffre River (QC) for the east-west axis.

The catchments in this study represent a diverse range of environmental conditions (Figure S2). The region's landscape and physiography vary significantly, featuring subarctic rivers in northern Quebec and Labrador, islands such as Anticosti, Prince Edward Island, and Newfoundland, and peninsulas like Gaspésie, New Brunswick, and Nova Scotia. Catchment size varies between 14 km^2 and 40 000 km^2. The study region is characterized by diverse land uses, including crop and urban dominance in Prince Edward Island's watersheds, forested areas in Quebec and New Brunswick, abundant shrubland cover in Newfoundland's watersheds, and significant wetland presence in Maine's rivers.

Watershed name	Station ID	Drainage Area (Km^2)	Tw station elevation (m)
Miramichi SW	1	7255.29	34.50
Miramichi NW	2	3555.35	43.01
Restigouche	3	5330.54	221.57
Matapedia	4	3803.95	152.04
Upsaqluitch	5	2306.79	170.15
Moisie	6	18963.00	48.54
Natashquan	7	15814.57	26.71
Ste-Anne	8	824.08	24.85
Ste-Marguerite	9	1111.26	107.67
Ouelle	10	823.69	41.49
Nouvelle	11	1164.53	35.85
Bonaventure	12	1918.40	103.73
Petite-Cascapedia	13	1337.09	67.66
Jupiter	14	543.82	177.46
Gouffre	15	753.08	178.71
Godbout	16	1941.93	135.50
Dartmouth	17	907.59	44.93
Cascapedia	18	2133.21	130.32
Huile	19	177.18	13.20
Narragagus	20	586.52	29.41
Sheepscot	21	368.12	52.25
Ducktrap	22	43.07	53.38
Gilbert	23	420.48	92.59
Conne	24	607.71	81.47
St-Lewis	25	2153.09	86.74
Highland	26	162.25	18.65
Reid	27	147.06	18.76
LaHave	28	467.72	72.17
Sackville	29	130.96	18.03
Margaree	30	372.57	107.80
West	31	99.25	50.50
Carruthers	32	49.74	12.76
Wilmot	33	46.36	15.67
Bear	34	13.98	19.53
Aux Mélèzes	35	40027.57	138.03

187 Table 1: Watersheds name and weather stations elevation and drainage area

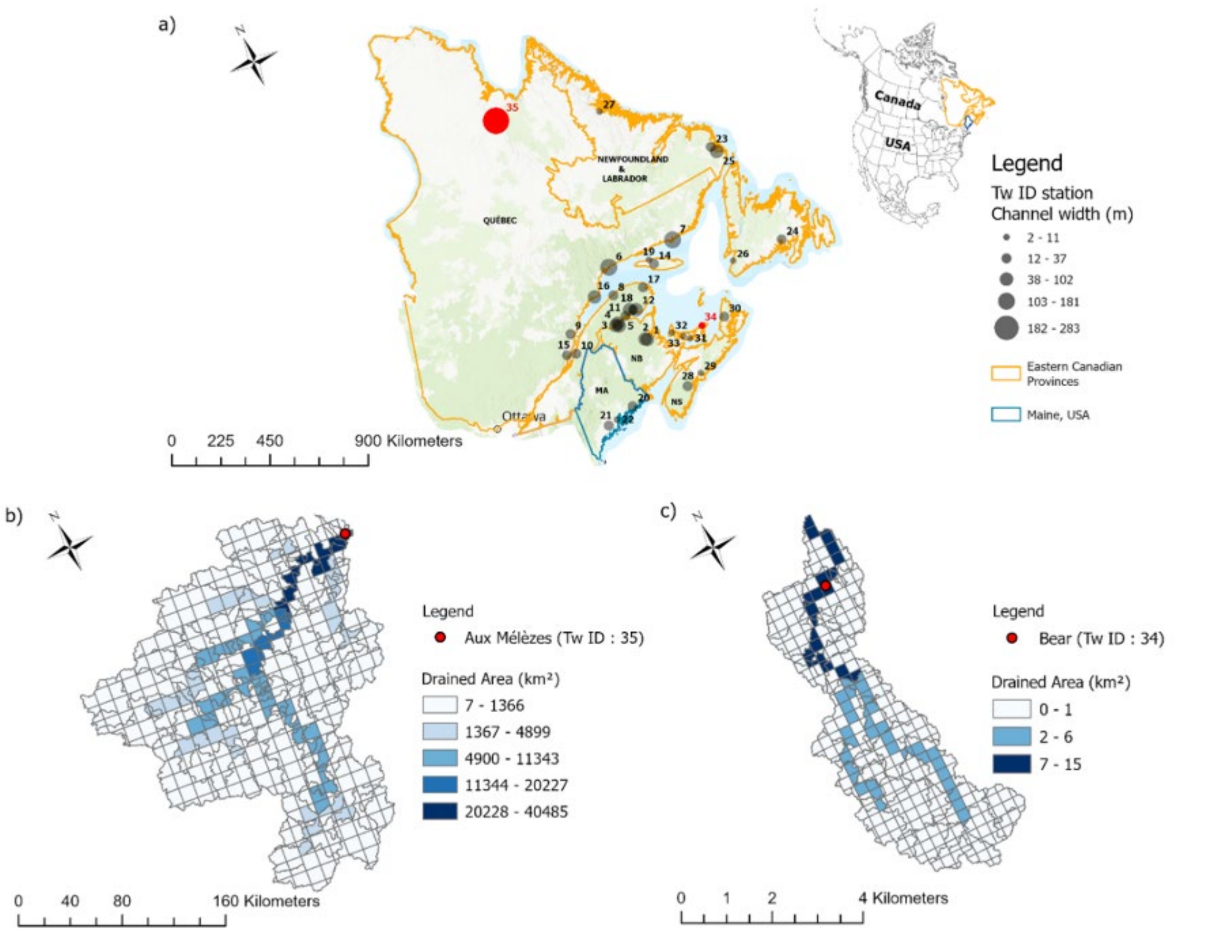

Figure 2: a) Stations locations and channel width at Tw station. b) The largest basin: Aux Mélèzes River, QC (ID:35), with a discretization of 13km x 13km. c) The smallest basin: Bear River (ID:34) with a discretization of 300m x 300m

A preliminary analysis of the chosen catchments sought to categorize the rivers based on their thermal behavior, to later determine if the models' performance was impacted by the river's thermal regime. Using the slope (thermal sensitivity) and intercept values of the Tair-Tw relationship for thermal classification, two distinct groups of catchments were identified. The first group consists of watersheds where a significant portion of discharge originates from deep aquifer infiltration (rivers id 31, 32, 33, and 34 in Table 1), labeled as "groundwater-fed streams" in Figure *3*. This group is characterized by low slope and high intercept values, consistent with findings from numerous studies (Boyer et al., 2021; Caissie, 2006; Gallice et al., 2015; Webb et al., 2008). The Baseflow Index (BFI) for these catchments was high, indicating a higher ratio of stream baseflow to total discharge volume (Figure

S2). The second group, which consists of most watersheds, relies on precipitation as the primary source of discharge. These watersheds are labeled as "thermally climate-driven" and are characterized by relatively low intercept and high slope values, indicating a strong correlation between stream and air temperatures. BFI is computed from observed discharge data using the Lynne-Hollick (LH) baseflow filter with the hydrostats R package (Bond & Bond, 2022; Ladson et al., 2013; Lyne & Hollick, 1979).

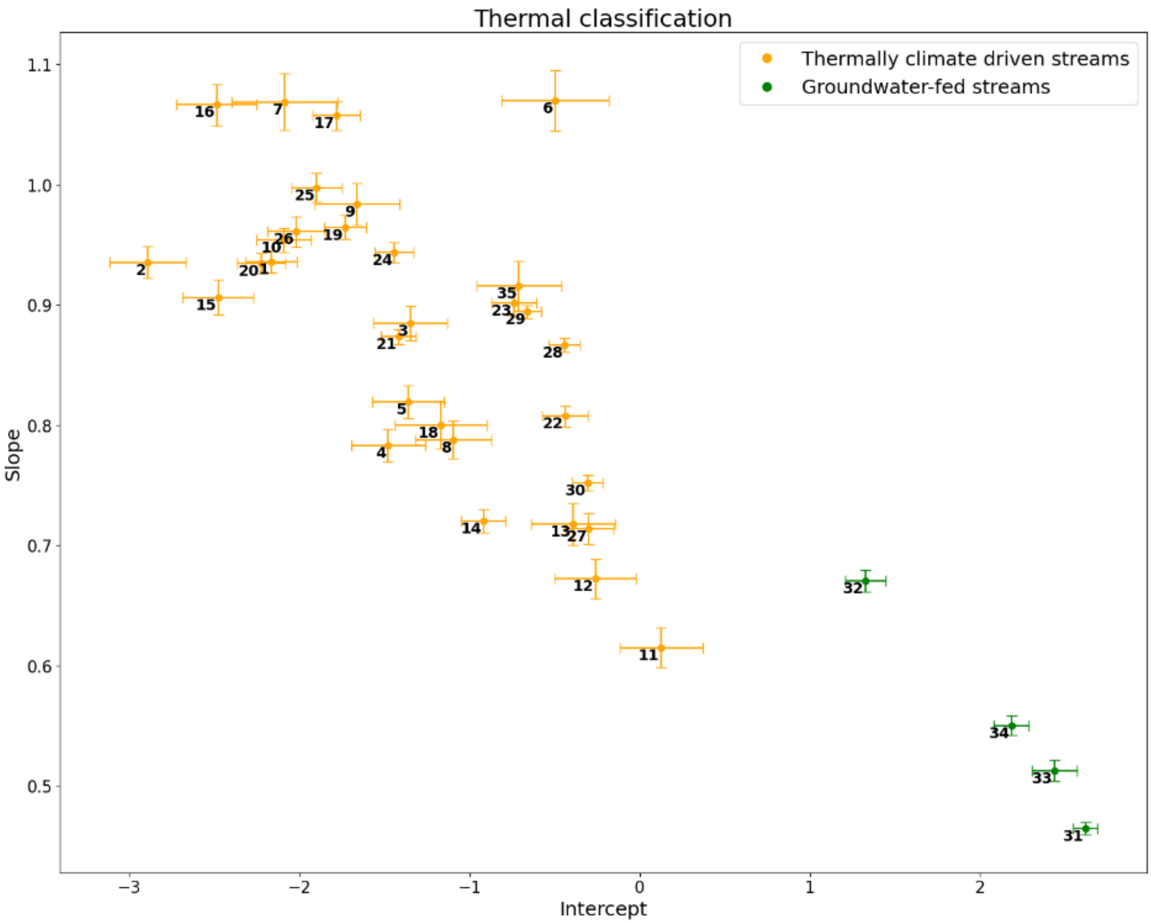

Figure 3: Slopes and intercepts of the regression lines fitted to the respective catchments' stream–air temperature points. All points with negative air temperature values have been discarded prior to fitting. The bars indicate the standard error estimates.

2.2.1- Calibration datasets

Discharge data for Canadian rivers were collected from the Ministry of the Environment and the Fight Against Climate Change website (https://www.cehq.gouv.qc.ca) and the HYDAT database via the ECCC data explorer (https://collaboration.cmc.ec.gc.ca/cmc/hydrometrics/www/). Tw time series for Canadian rivers were obtained from the RivTemp database (https://www.rivtemp.ca) (Boyer et al., 2016). The RivTemp database is the result of a partnership between universities, provincial and federal governments, watershed organizations and organizations dedicated to the conservation of Atlantic salmon. For the rivers considered in the U.S, discharge and Tw data were obtained from the USGS website (https://waterdata.usgs.gov/monitoring-location). Figure 4 illustrates the Tw data availability from January 1, 1979, to December 31, 2020.

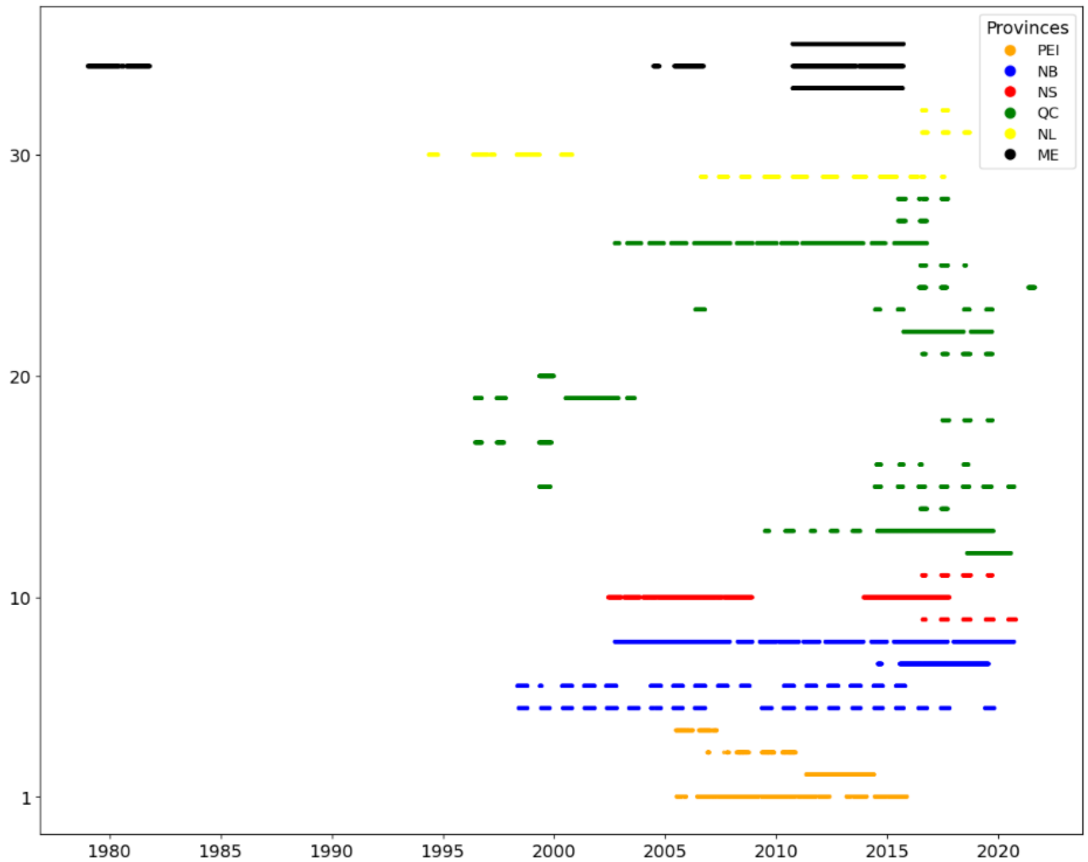

Figure 4: Daily Tw availability for the selected sites (35). For each site on the y-axis, a colored mark indicates instances where the daily mean Tw was recorded for a given site.

Due to limited meteorological data availability, we utilized daily ERA5 reanalysis data from the European Center for Medium-Range Weather Forecasts (ECMWF) website (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5) with a native grid resolution of 0.25°×0.25° (Hersbach et al., 2020). The meteorological inputs and their corresponding units required by the CEQUEAU model are enumerated in Table 2. To calculate the saturation vapor pressure of water, we employed the daily dew temperature from ERA5 and the Tetens equation (Tetens, 1930).

Variable	Description	Units
Ptot	Total precipitation	Mm
tMax & tMin	2m surface air temperature	◦C
SSR	Incoming surface solar radiation	MJ/m^2
VP	Vapour pressure	mmHg
VV	Wind speed	km/h
CC	Cloud cover	%

Table 2: ERA5 daily data used in this study.

2.2.2- Regional attributes

The physio-climatic features characterizing the watershed upstream of the Tw gauging station were extracted using ArcGIS Pro software. We have identified 23 relevant predictors for this study, defined in Table 3. These predictors include three related to the geographical location of the Tw station (LAT, LONG, and D2O), four related to climate forcing (90thPtot, 90thSSR, 25thCC, and 90thVV), four related to land uses (FOREST, LAKE, WET, and CROP), one related to topsoil texture (LOSA), and eight related to geomorphology and topography (MIND, DD, SLP, ORUP, ORWSH, ASPC, ELVT, DVME, DFME, GSCV, and DOGS). The refined set of covariates and key references used in this study can be found in Table 3.

Abbreviation	Description	Units	Key reference
LAT, LON	Tw station geographical coordinates	◦	(R. D. Moore, 2006; Wuebbles et al., 2017)
MIND	depth of the river reach	m	(Ebersole et al., 2003; Jackson et al., 2017; Story et al., 2003)
ELVT	Tw station elevation	m	(Mohseni et al., 1998; Souaissi et al., 2021)
DD	Drainage density	km^{-1}	(Godsey & Kirchner, 2014; T. B. Ouarda et al., 2022)
SLP	Tw station slope	-	(Caissie, 2006; Ouarda et al., 2022; Souaissi, Ouarda, St-Hilaire, et al., 2023)
ORUP	Upstream station channel orientation	◦	(Garner et al., 2017)
ORWSH	Watershed orientation	◦	(Arora et al., 2018)
D2O	Tw station distance to the ocean	km	(Collins, 2023; Jackson et al., 2018)
FOREST	Percentage of upstream forests	%	(Garner et al., 2014; St-Hilaire et al., 2000)
LAKE	Percentage of upstream lakes	%	(Abidi et al., 2022; Arora et al., 2018; J. A. Leach et al., 2021)
WET	Percentage of upstream wetlands	%	(O'Sullivan et al., 2019)
CROP	Percentage of upstream crops	%	(Charron et al., 2021; Essaid & Caldwell, 2017)

LOSA	upstream topsoil texture: loamy sand (USDA)	%	(Kandala et al., 2024; Kurylyk et al., 2014; Sepaskhah & Boersma, 1979)
90^{th}PTOT	Summer total precipitation (90^{th}Q)	mm	(Coffey et al., 2019; Raymondi et al., 2013)
90^{th}SSR	Summer insolation (90^{th}Q)	Mj	(Bray et al., 2017; Jeppesen & Iversen, 1987; Laizé et al., 2017)
25^{th}CC	Summer Cloud coverage (25^{th}Q)	-	(Girard et al., 2003; J. Leach & Moore, 2010; Sinokrot & Stefan, 1993)
90^{th}VV	Wind speed (90^{th}Q)	m/s	(Garner et al., 2014; Laizé et al., 2017; J. Leach & Moore, 2010; Sinokrot & Gulliver, 2000)
ASPC	Slope orientation	◦	(McCutchan & Fox, 1986; O'Sullivan et al., 2019)
DFME	Difference between the grid cell elevation and the mean of its neighbouring cells.	-	(Houndekindo & Ouarda, 2023, 2024; I. D. Moore et al., 1991; Wilson, 2018)
DVME	Difference between the grid cell elevation and the mean of its neighbouring cells normalized by the standard deviation.	-	(Houndekindo & Ouarda, 2023, 2024; I. D. Moore et al., 1991; Wilson, 2018)
GSCV	Product between the maximal and the minimal curvature. Measure of surface curvature	m^{-2}	(Florinsky, 2017; Houndekindo & Ouarda, 2023, 2024; I. D. Moore et al., 1991; Wilson, 2018)
DOGS	Difference between two copies of the DEM smoothed with two different gaussian kernel. Measure land surface curvature.	-	(Florinsky, 2017; Houndekindo & Ouarda, 2023, 2024)

Table 3: Physiographical and meteorological predictors

The river temperature at the Tw gauging station is significantly influenced by the physical characteristics of the drainage basin (Jackson et al., 2017; Ouarda et al., 2022; Souaissi et al., 2023a). These factors include watershed orientation, drainage density, land use, and soil types, with the latter two expressed as a percentage of the total drainage area. To prevent redundancy and overlapping effects, only selected land use and soil type variables are considered, as they do not sum up to 100%, and there is collinearity among these predictors.

Climate-related factors have a significant impact on vegetation, soils, and landforms, influencing hydrological and thermal processes (Caissie, 2006; Leach & Moore, 2010). Recent regionalization studies underscore the importance of considering these predictors (Charron et al., 2020; Ouarda et al., 2022; Souaissi et al., 2023a). For this study, we used high and low summer quantiles [25^{th}, 90^{th}]

calculated between May 1st and October 1st (1979-2020) extracted from ERA5 daily reanalysis data (Hersbach et al., 2020).

Additionally, we included the percentage of upstream dominant topsoil texture, as it affects soil properties such as hydraulic conductivity and soil matric potential, which in turn influence runoff, groundwater recharge, stream discharge, and temperature (Fernandez-Illescas et al., 2001; Kurylyk et al., 2014). The Harmonized World Soils Database v2 provided the topsoil texture features at a 1km spatial resolution (Nachtergaele et al., 2023). Moreover, land cover variables, such as lakes, croplands, wetlands, and vegetation, significantly impact river temperatures by affecting evaporation and surface runoff temperature (Abidi et al., 2022; Garner et al., 2014; O'Sullivan et al., 2019; St-Hilaire et al., 2000). We obtained information about land cover types from the North American Land Change Monitoring System (2020) available at (http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2020/) with a 30m spatial resolution.

Furthermore, topographical features like aspect, channel slope, gaussian curvature, and relative topographical position are crucial for modeling terrain landscape configuration and estimating various energy heat fluxes (Jackson et al., 2018; Maxwell & Shobe, 2022; O'Sullivan et al., 2019; Teutschbein et al., 2018; Tovar-Pescador et al., 2006). These variables were extracted from a 30m spatial resolution DEM using the NASA shuttle radar topography mission (Nasa, 2013) and computed with the *WhiteboxTools* developed at the University of Guelph, Canada (Lindsay, 2014).

2.2- CEQUEAU model

The CEQUEAU model is a conceptual and semi-distributed hydrological model designed to simulate stream discharge and Tw. Watersheds are divided into equal-sized whole squares (CE, from the French acronym), which are further subdivided into smaller partial squares (CP) based on drainage characteristics and sub-basin divisions (Morin & Paquet, 2007; St-Hilaire et al., 2015). The chosen grid size aims to strike a balance between resolution and model run-time (Dugdale et al., 2017a)

(Figure 2a and *2*b). The model operates through two core functions: a production function (Equation S1), which calculates a water balance for each CE at every time (t) and distributes precipitation among the conceptual storage units (e.g., rivers, lakes, marshes, upper and lower soil layers) based on land use data. Then a transfer function, routes available water downstream between CP units to simulate discharge over time (Figure S1). At the same time step, the hydrological simulation results are fed to the thermal module in addition to other meteorological data (solar radiation, wind speed, air vapor pressure and cloud cover) and for each CP, the change in Tw is given by:

$$\Delta T_{W_i} = \frac{\Delta Q_t}{V_t * \theta} \qquad (1)$$

ΔQ_t is the total change in enthalpy (MJ), V is the volume of water (m^3) (output of the hydrological calibration), θ is the heat capacity of water ($4.187\ MJ\ m^{-3} °C^{-1}$). ΔQ_t is computed by summing the various components of the heat budget including the advective and air-water interface heat fluxes on each CP as follows:

$$Q_t = Q_{sw_t} + Q_{lw_t} + Q_{e_t} + Q_{s_t} + Q_{a_t} \qquad (2)$$

Where Q_{sw_t} represent shortwave radiation heat flux, Q_{lw_t} is the longwave radiative heat flux, Q_{e_t} is the latent heat transfer, Q_{s_t} is the sensible heat transfer and Q_{a_t} represent the local advective heat transfer from the upstream grid cells, groundwater, and subsurface flow (Equations *S2*, S3, S4 and S5 in supplementary material). The model's capabilities have been improved with the introduction of 'pycequeau,' a new Python-based physiographical toolbox (https://github.com/erinconv/pycequeau), employed in this study. This toolbox enhances the model's functionality, making sub-basin divisions and routing more efficient and accurate.

The CEQUEAU model has consistently proven effective in simulating the hydrological and thermal regimes of various Canadian watersheds and conditions (Charron et al., 2021; Khorsandi et al., 2023; Kwak et al., 2017; Rincón et al., 2023; St-Hilaire et al., 2023).. Notably, the model has consistently achieved RMSE values below 2°C in its Canadian applications, showcasing its precision and reliability in Tw modeling.

This study represents a significant achievement as it is the first instance of applying the CEQUEAU model at a regional scale, covering 35 watersheds. Previous applications were predominantly confined to individual watersheds. We utilized a 30-meter spatial resolution DEM obtained from the NASA shuttle radar topography mission (Nasa, 2013) and a 30m spatial resolution raster from the North American Land Change Monitoring System accessible at (http://www.cec.org/north-american-environmental-atlas/land-cover-30m-2020/) to discretize catchments in CEQUEAU.

2.2.1- Sensitivity Analysis

.

A sensitivity analysis provides a thorough assessment by examining model responses across the entire feasible parameter range, involving extensive input factor samples and output variation analysis (Saltelli et al., 2008). It also helps mitigate equifinality risks, where multiple parameter sets produce identical results, underscoring its importance before calibration for more precise outcomes (Feigl et al., 2020; Klotz et al., 2017).

In this study, we used the Variogram Analysis Response Surfaces (VARS) framework to conduct a comprehensive sensitivity analysis for CEQUEAU's thermal model parameters (Razavi et al., 2019; Razavi & Gupta, 2016a, 2016b). This would not only provide a detailed evaluation of parameter influence but also plays a pivotal role in the regionalization process by identifying the most critical thermal parameters. By focusing calibration efforts on these key parameters and fixing the less sensitive ones, we enhance regionalization efficiency and streamline the modeling process.

VARS provides a robust and efficient approach, integrating variance and derivative-based methods and offering a wide range of sensitivity information. It incorporates directional variograms across various perturbation scales and generates sensitivity indices known as IVARS (Integrated Variograms Across a Range of Scales) (Razavi & Gupta, 2016a). Variograms display the variability of a function at different length scales. Let's suppose the function *f* that relates CEQUEAU thermal parameters (MP) to Tw:

$$T_w = f(\boldsymbol{MP}) \tag{3}$$

where $\boldsymbol{MP} = \{MP_1, \dots, MP_{10}\}$

Sensitivity analysis of T_w with respect to MP_i is a scale-dependent property that can be characterized by the multidimensional variogram (γ) and covariogram (C) functions:

$$\gamma\,(\boldsymbol{h}) = \frac{1}{2}V\big(T_w(\boldsymbol{MP}+\boldsymbol{h}) - T_w(\boldsymbol{MP})\big) \tag{4}$$

$$C(\boldsymbol{h}) = C(0) - \gamma\,(\boldsymbol{h}) = V\big(T_w(\boldsymbol{MP})\big) - \gamma\,(\boldsymbol{h}) \tag{5}$$

Where *V* represents variance $\mathbf{h} = \mathbf{MP^A}$- $\mathbf{MP^B}$ refers to the increment and represents the distance vector, $\mathbf{h} = \{\mathrm{h}_1, \dots, \mathrm{h}_{10}\}$ between any two points *A* and *B* in the parameter space. Greater values of $\gamma\,(h_i)$ indicate a higher sensitivity at that scale *h*. Razavi & Gupta. (2016a) introduced IVARS as an index to summarize global sensitivities, which is calculated as the average $\gamma\,(h_i)$ up to a certain scale limit (10%, 30%, 50%). For example, $IVARS_{50}$ is the average of $\gamma\,(h_i)$ up to 50% of the range of MP_i and it is defined as follows:

$$\Gamma(H_i) = \int_0^{H_i} \gamma(h_i)\, dh_i \tag{6}$$

In this study, we use H_i=50% ($IVARS_{50}$) of the parameter scale range. This approach led to a total of 9100 function evaluations for each watershed. The sensitivity analysis results are presented using normalized $IVARS_{50}$ values that sum up to 100% (ratio of sensitivity, see supplementary material for details), enabling straightforward interpretation of sensitivity indices and facilitating a consistent comparison across different metrics and watersheds.

VARS is conducted for the first time to assess the sensitivity of semi-distributed Tw model parameters. We classified parameters with a normalized IVARS sensitivity ratio over 10% as highly sensitive ($nIVARS_{50} \geq 10\%$), while those below 10% were deemed low sensitive or insensitive (Abdelhamed et al., 2023). VARS-TOOL on MATLAB environment is employed for this purpose, see Razavi et al. (2019) for more details.

2.2.2- Calibration procedure

The CEQUEAU calibration procedure involves a sequential approach, calibrating the hydrological model followed by calibrating the thermal component. The hydrological component of the CEQUEAU model is governed by 31 parameters (Table S1), while the Tw component is controlled by 10 parameters, with specific details on Tw parameters provided in Table 4.

Description	**Low bound**	**Upper bound**	**Units**	**Name**
Fitting coefficient controlling the minimum depth of the river reach	0.01	2	-	COPROM
Fitting coefficient controlling the river width	0.01	2	-	COLARG
Fitting coefficient for incoming solar radiation	0.1	2	-	CRAYSO
Fitting coefficient for longwave radiation flux	0.1	2	-	CRAYIN
Fitting coefficient for latent heat flux	0.1	2	-	CEVAPO
Fitting coefficient for sensible heat flux	0.1	2	-	CCONV
Threshold for snow stock controlling Tw	0	250	mm	CRIGEL
Groundwater temperature	1	10	◦C	TNAP
Minimum precipitation to define days with low solar radiation	5	15	mm	BASSOL
Correction factor for BASSOL	0	1	-	CORSOL

Table 4: CEQUEAU water temperature model parameters

To ensure the capture of full variability in discharge and water temperature signals, the entire dataset was used for both calibration and validation. This approach is essential in our regionalization context where retaining the variability in the data is critical for parameter transferability, especially given the relatively short duration of many water temperature datasets, with some stations recording only two summers of data (Rakovec et al., 2016; H. Shen et al., 2022).

Once the local hydrological discharge calibration is completed, the thermal component is calibrated in two phases. First, an automatic calibration is performed for all thermal parameters at the catchment scale, resulting in an initial parameter set referred to as Pre_{SA}. In the second phase, calibration focuses solely on the most sensitive parameters (low-sensitivity parameters are fixed at their Pre_{SA} values), producing the benchmark parameter set, termed $Post_{SA}$. This approach allows for evaluating whether the identified low-sensitivity parameters truly have minimal influence within the study region and quantifying the impact of fixing these parameters between the two calibration steps (Abdelhamed et al., 2023; Feigl et al., 2020).

2.2.3- Optimization algorithm and objective functions

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) is used as a stochastic global optimization algorithm for both hydrological and thermal components (Auger & Hansen, 2012; Hansen & Ostermeier, 2001). CMA-ES has demonstrated state-of-the-art performance in deterministic discharge and water temperature modeling, proving particularly effective for CEQUEAU applications in recent years (Arsenault et al., 2014; Khorsandi et al., 2022, 2023; Oyinlola et al., 2023; Rincón et al., 2023).

For discharge calibration, the Kling-Gupta Efficiency (KGE) coefficient was maximized as the objective function (Gupta et al., 2009). However, the pronounced seasonality in Tw time series can inflate KGE values, as most temporal variance is driven by consistent seasonal patterns (Ouellet-Proulx et al., 2019). To address this, we minimized the RMSE between observed Tw and simulated values generated using the Pre_{SA} and $Post_{SA}$ parameter sets (Figure *6*). A termination criterion was established at 4000 evaluations, as this was found to be sufficient for achieving convergence for both discharge and Tw. For each station, the CMA-ES calibration was independently executed ten times. The final optimal parameters for each station were determined by averaging the results from these ten iterations (Khorsandi, 2024).

2.3- Regional modeling

In this study, all stations were considered without defining homogeneous regions or neighborhoods. Additionally, we tested an alternative approach, which involved using the two predefined thermal regions in Figure *3* as separate homogeneous regions. However, this approach did not yield conclusive results. Therefore, using the first scenario allowed us to leverage the full breadth of our dataset, maximizing the use of available data to inform the transfer of CEQUEAU thermal parameters. In the results, we present this approach where no homogeneous regions or neighborhoods were defined.

2.3.1- Model description

For this study, we selected a classical ML model : The support vector regression (SVR), a model widely used in hydrology and known to excel with small and tabular datasets (Deka, 2014; Lange & Sippel, 2020; Weierbach et al., 2022). This study uses the epsilon-insensitive SVR (ε-SVR) formulation (Vapnik, 2013). Unlike traditional least squares methods, ε-SVR employs the ε-insensitive tube concept, which combines mean absolute error and L-2 norm penalty. This approach allows for some samples to be at a specific distance (ξ or ξ*) from their correct margin boundary, acknowledging that problems are not always perfectly separable with a hyperplane. The goal is to fit the error within a threshold (ε) while minimizing the coefficient norm using the penalty term (C), which controls the strength of this penalty and acts as an inverse regularization parameter.

Given training vectors $x_i \in \mathbb{R}^p, i = 1, \ldots, n,$ and a vector $y \in \mathbb{R}^n$, our goal is to find the regression weight coefficient $w \in \mathbb{R}^p$ such as the prediction given by $(w^T\theta(x_i) + b)$ is accurate for most samples. ε-SVR solves the following primal problem:

$$min_{w,b,\xi,\xi^*} \ \frac{1}{2} \ w^T w + C \sum_{j=1}^{n}(\xi_i + \xi_i^{\ *}) \ \ subject\,to: \begin{cases} y_i - \ w^T\theta(x_i) - b \le \varepsilon + \xi_i \\ -y_i + w^T\theta(x_i) + b \le \varepsilon + \xi_i^{\ *} \\ \xi_i \xi_i^{\ *} \ge 0 \ i = 1, \ldots, n \end{cases} \quad (7)$$

We penalize samples whose predictions deviate from their true targets by at least ε. These samples incur a penalty of ξ or ξ*, depending on whether their predictions fall above or below the ε tube.

Although SVR handles non-linear decision boundaries of arbitrary complexity, by using kernel functions, we limit ourselves in this paper to linear kernel because of the nature of the data sets under investigation (Gallice et al., 2015). Using simple linear models helps address issues of parsimony and overfitting, especially given the limited dataset (stations and years). Moreover, the interpretability of linear models allows regression coefficients to clearly reflect the strength and direction of the relationship between predictors and sensitive CEQUEAU thermal parameters, avoiding multicollinearity (Houndekindo & Ouarda, 2023). For a linear kernel, the primal problem can be equivalently formulated as follows:

$$min_{w,b} \ \frac{1}{2} \ w^T w + C \sum_{j=1}^{n} max(0, \left| y_i - (w^T\theta(x_i) - b) - \varepsilon \right|) \ \ (8)$$

we used the LinearSVR Python function in the scikit-learn library sklearn.svm (Pedregosa et al., 2011), and equation 9 is the form that is directly optimized by this model. Through leave one out cross validation we will assess the generalizability and the trade-offs between complexity (C and ε hyperparameters) and predictive accuracy. We compared the performance of linear SVR with a standard MLR, commonly used as a benchmark in recent Tw regional studies (Charron et al., 2021; Gallice et al., 2015; T. B. Ouarda et al., 2022; Souaissi, Ouarda, & St-Hilaire, 2023b; Weierbach et al., 2022; Weller et al., 2024).

2.3.2- Input Feature Selection and Preprocessing

In this study, we compare four FS algorithms, presented as follows:

Recursive Feature Elimination

RFE is a wrapper backward elimination algorithm that involves fitting an estimator to remove the least important predictors iteratively until a specified minimum number of features is reached (Guyon et al., 2002). First, the estimator is trained on the initial set of features, and the importance of each feature is obtained through the magnitude of the weight coefficient vector or inherent feature importance in the model. Then, the least important features are pruned from the current set of features. This procedure is recursively repeated on the pruned set until the desired number of features to select is reached. The RFE requires a specified number of features to be retained and leave-one-out cross-validation is used to evaluate different subsets of features and select the best subset.

Sequential Forward Selection

Sequential forward feature (SFS) selection is a greedy wrapper method that progressively adds features that significantly enhance the model's performance (Derksen & Keselman, 1992; Hastie et al., 2017). Initially, the process begins with no features and identifies the one that maximizes a cross-validated score when the estimator is trained on this single feature. After selecting the first feature, the procedure is repeated by adding another feature to the existing set. The iteration continues until a stopping criterion is met, determined through leave-one-out cross-validation. However, SFS might be slower as it requires the evaluation of a larger number of models.

Least absolute shrinkage absolute operator

LASSO is a penalized linear regression approach introduced by Tibshirani. (1996), which can be used as an FS-embedded method. It applies an L1-norm penalty to the regression coefficients using the regularization parameter λ, which controls the severity of the penalty. This penalization effectively reduces some coefficients to zero, thus leading to a model that is both sparse and interpretable. One notable limitation highlighted for LASSO is the potential for reduced predictive accuracy when the predictors are highly correlated (Yamada et al., 2014; Zou & Hastie, 2005).

Elastic Net

Elastic Net (ENET) improves upon LASSO by adding an L-2 norm penalty (Zou & Hastie, 2005). This combination helps to overcome some of LASSO's limitations, particularly its tendency to select only one variable from a group of highly correlated variables while potentially ignoring others. It is a particularly useful approach when dealing with highly correlated data or when the number of predictors exceeds the number of observations. An additional parameter α is used to balance the L-1 and L-2 penalties, where α=1 corresponds to LASSO (pure L-1 penalty) and α=0 corresponds to Ridge (pure L-2 penalty).

In this study, selected covariates were standardized to a zero mean and unit variance prior to model fitting, and model parameter controlling solar radiation (CRAYSO) was log-transformed to reduce skewness and render the data more normally distributed (Jackson et al., 2018). The FS algorithms were implemented using the following functions: ElasticNetCV, SequentialFeatureSelector, RFECV, and LassoCV, available in the scikit-learn library sklearn.feature_selection (Pedregosa et al., 2011). We tested all combinations of FS algorithms (RFE, SFS, LASSO, ENET) and regional models (MLR, SVR), resulting in 8 model combinations: RFE-MLR, SFS-MLR, LASSO-MLR, ENET-MLR, RFE-SVR, SFS-SVR, LASSO-SVR, and ENET-SVR.

2.3.3- Spatial and temporal evaluation

Spatial evaluation

. To mimic the estimation of CEQUEAU most sensitive thermal MP at ungauged sites, a Leave-One-Out cross validation procedure was implemented. This approach systematically excludes one watershed from the calibration process, using it exclusively for validation (pseudo-ungauged). This ensures that the calibrated model is evaluated across diverse climatic and physiographical conditions, enhancing its robustness and suitability for regionalization. To compare between calibrated and regionalized thermal parameters ($\boldsymbol{Post}_{\boldsymbol{SA}_{MP}}$ and RE_{MP}) (Spatial evaluation in Figure 1), RMSE and the coefficient of determination (R^2) are used to assess the performance of ML regional models:

$$RMSE_{MP} = \sqrt{\frac{\sum_{i=1}^{n}(\boldsymbol{Post}_{\boldsymbol{SA}_{MP_i}} - RE_{MP_i})^2}{n}}$$

$$R_{MP}^2 = 1- \frac{\sum_1^n(\boldsymbol{Post}_{\boldsymbol{SA}_{MP_i}} - RE_{MP_i})^2}{\sum_1^n(\boldsymbol{Post}_{\boldsymbol{SA}_{MP_i}} - \overline{RE_{MP_i}})^2}$$

Where n, represents the total number of watersheds. Additionally, we decided to use the RMSE skill score to compare the performance of the SVR model with that of the MLR model. The RMSE-based skill score is calculated using the following equation:

$$S_{score} = 1 - \frac{RMSE_{SVR_{T_W}}}{RMSE_{MLR_{T_W}}} \quad (10)$$

If the prediction (SVR) has lower errors compared to the benchmark prediction (MLR), the skill score values are positive. Conversely, if the errors are higher, the skill score values are negative. If the errors are equal, the skill score values are zero.

<u>Temporal evaluation:</u>

. Subsequently, for each watershed, daily mean Tw time series produced by the most effective regional model ($T_{w_{RE}}$) are compared with the ones generated using local benchmark calibration parameter set $T_{w_{\boldsymbol{Post}_{\boldsymbol{SA}}}}$using RMSE (Temporal evaluation in Figure 1):

$$RMSE_{T_w} = \sqrt{\frac{\sum_{i=1}^{n}(T_{w_{\boldsymbol{Post}_{\boldsymbol{SA}}}} - T_{w_{RE}})^2}{n}} \quad (11)$$

Where n, represents the daily time step from 01-01-1979 to 31-12-2020.

3- Results

3.1- Global Sensitivity Analysis

The results from VARS indicated that COLARG, CRIGEL, TNAP, and BASSOL had low sensitivity ($nIVARS_{50} \leq 10\%$) (Figure 5). Additionally, certain parameters such as COPROM, CEVAPO, and CORSOL showed low variability during the calibration process and remained almost constant across all watersheds (Figure S3). Consequently, these parameters were given a fixed value based on the first calibration results (Pre_{SA}).

The VARS analysis further showed that parameters associated with the main heat fluxes exchange energy at the air-water interface (CRAYSO, CRAYIN, CEVAPO, CCONV) were the most sensitive across all rivers ($nIVARS_{50} \geq 10\%$). River 17 deviated from this trend, with COPROM, a parameter controlling the depth-width ratio of the mainstem channel, being identified as the most sensitive parameter.

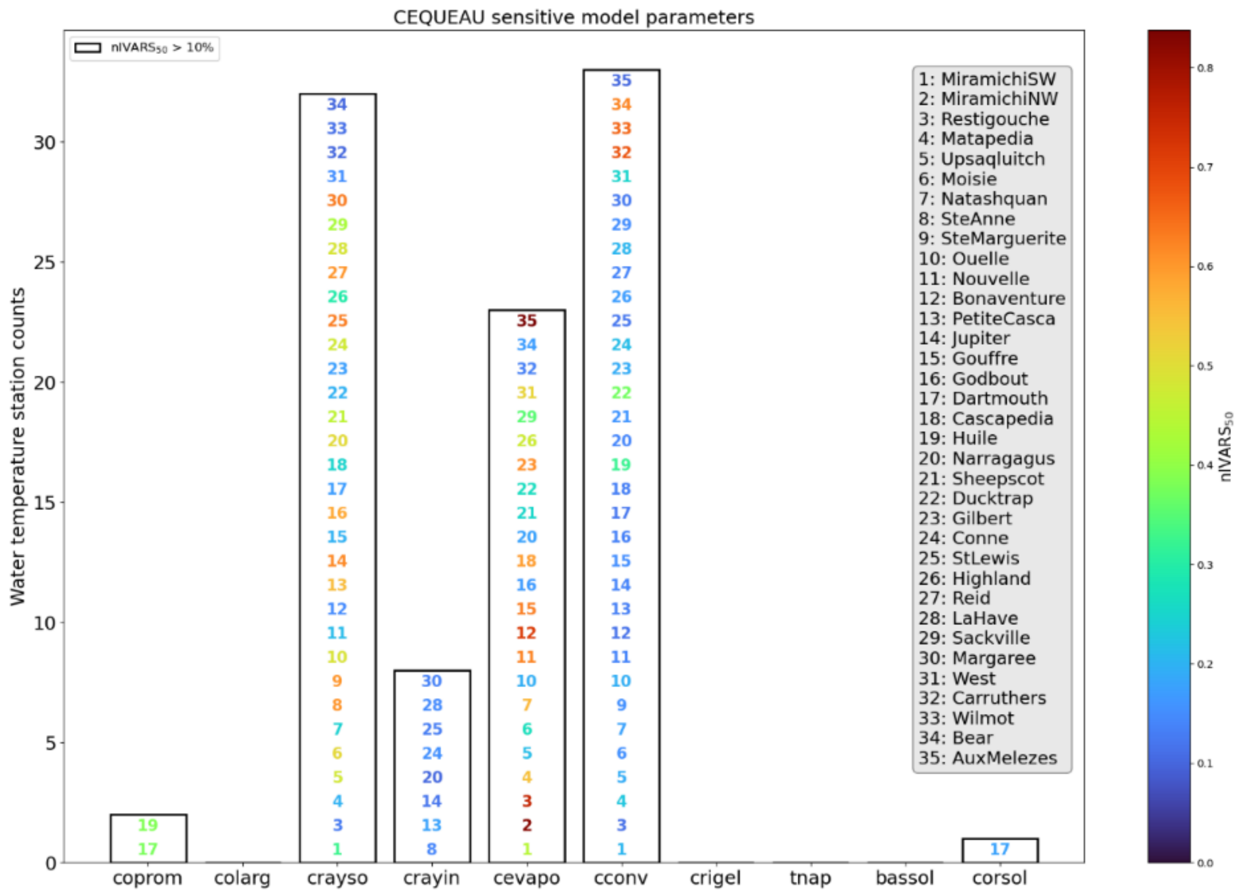

Figure 5: Sensitivity analysis summary for CEQUEAU Tw model parameters. The number within the bars refers to the river IDs, and the colors refer to the $nIVARS_{50}$ values

The sensitivity analysis findings in Figure 5 showed that CRAYSO, CEVAPO, and CCONV are sensitive parameters for most of the rivers in the study region. These parameters were found to be highly sensitive for sixteen, fourteen, and five rivers, respectively, with nIVARS > 0.4. CRAYIN was found to be sensitive to eight rivers but with low nIVARS values (< 0.2) and was fixed to its optimal value (Pre_{SA}). Subsequently, our calibration and regionalization efforts focused on highly sensitive parameters with high variability, namely CRAYSO and CCONV.

CRAYSO showed high sensitivity (nIVARS > 0.4) for rivers with wide mainstem channels at Tw stations (> 20m). These include rivers flowing into Chaleur Bay (5, 13), Quebec rivers flowing into the St. Lawrence estuary (6, 8, 14, 16), southern Quebec rivers (9, 10), rivers in Nova Scotia and Maine (28, 29, 30; 20, 21), coastal and subarctic Labrador rivers (25, 27), and the Conne River (24) in southern maritime Newfoundland.

CEVAPO, on the other hand, had low variability and remained nearly constant for most considered catchments. However, it was highly sensitive for rivers with wide mainstem channels at Tw stations, such as the Natashquan River (7), rivers flowing into Chaleur Bay (1, 2, 3, 4, 11, 12, 18), high latitude subarctic rivers (23, 35), and the Gouffre and Highland rivers (15, 26). The West River (31) in PEI, characterized by high cropland uses, is an exception in this regard, with channel width inferior to 10 m.

The Sackville River (29) was highly sensitive to both CRAYSO and CEVAPO, as the Tw station location is in a highly urbanized area with no riparian vegetation. On the other hand, CCONV showed high sensitivity for the smallest basins with channel widths less than 10 m, such as the Huile River in Anticosti Island, Ducktrap River in Maine, and groundwater-fed rivers in PEI (19, 22, 32, 33, 34). The Huile and Dartmouth rivers (17, 19) also showed high sensitivity towards CORPOM. Additionally, a considerable cumulative sensitivity effect was observed through parameters controlling surface solar radiation (CRAYSO, CORSOL) for the Dartmouth River.

3.2- Calibration performance

Figure 6 illustrates the discharge and Tw calibration performance using CEQUEAU model and compared results before and after the sensitivity analysis. The Pre_{SA} results showed the initial calibration state, whereas the $Post_{SA}$ results illustrated the recalibration made after identifying and fixing parameters with low variance and sensitivity for the Tw component as listed in Table S2 and S3. The CEQUEAU model demonstrated robust performance for discharge, achieving a mean Kling-Gupta Efficiency (KGE) value of 0.84 across the evaluated rivers, with individual performance values ranging from 0.65 to 0.94 However, Sackville, Margaree, and Wilmot rivers exhibited the lowest performance for discharge calibration, with KGE values below 0.7.

For the Tw modeling, Pre_{SA} configuration yielded a mean RMSE of 1.59°C, with values ranging from 0.77 °C (Aux Mélèzes river) to 3.04 °C (Sheepscot river). Eight rivers had an RMSE over 2 °C, specifically, Maine rivers (Narragagus, Sheepscot, and Ducktrap), Nova Scotian rivers with an outlet in the north Atlantic shore (LaHave and Sackville), as well as Matapedia, Gouffre, and Highland rivers. The mean RMSE for the $Post_{SA}$ configuration slightly increased to 1.75°C, with values spanning from 0.77°C to 3.26°C, and nine rivers had an RMSE over 2°C. The results before (Pre_{SA}) and after ($Post_{SA}$) addressing the low and non-sensitive parameters showed a mean decrease in accuracy of less than 20% in terms of RMSE across all catchments. However, the Highland River stands out with a significant loss in accuracy before and after sensitivity analysis (43%, +0.9°C). Consistent performance across various watersheds indicates that the CEQUEAU model reliably captures the key processes governing water temperature dynamics, supporting its use for parameter transfer to ungauged locations.

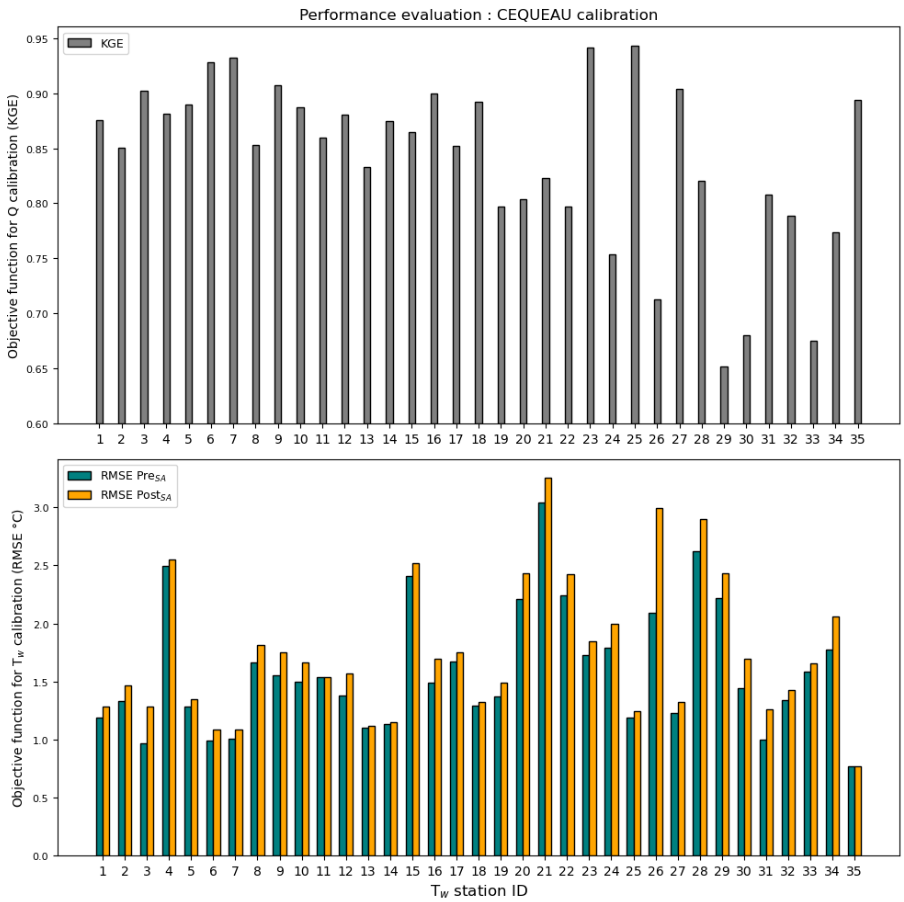

Figure 6: Discharge and Tw calibration performance. Pre_{SA} refers to the results of the first calibration considering all Tw model parameters. $Post_{SA}$ refers to the second calibration where low sensitive parameters are fixed to their previous optimal value.

The estimated mean daily discharge and Tw time series (1979–2020) were uploaded to the RivTemp database. These datasets have already proven valuable in a recently submitted paper that we co-authored, titled "Changes in size-at-age of juvenile Atlantic salmon cohorts over the past 50 years and linkages to environmental factors," the estimated time series effectively captured variability in the Atlantic salmon growth cycle within two of the most important Atlantic salmon watersheds in eastern Canada—the Restigouche and Miramichi rivers.

3.3- Regionalization modeling performance

3.3.1- Thermal parameters

Figure 7 presents a comparative analysis of regional models' performance across all FS methods. The results showed that SVR consistently outperforms MLR in both R^2 and RMSE metrics across all FS methods. This suggests SVR model is better suited for handling the complexities of the dataset. Given these findings, the decision to focus exclusively on SVR as the regional estimation model in the subsequent sections is justified.

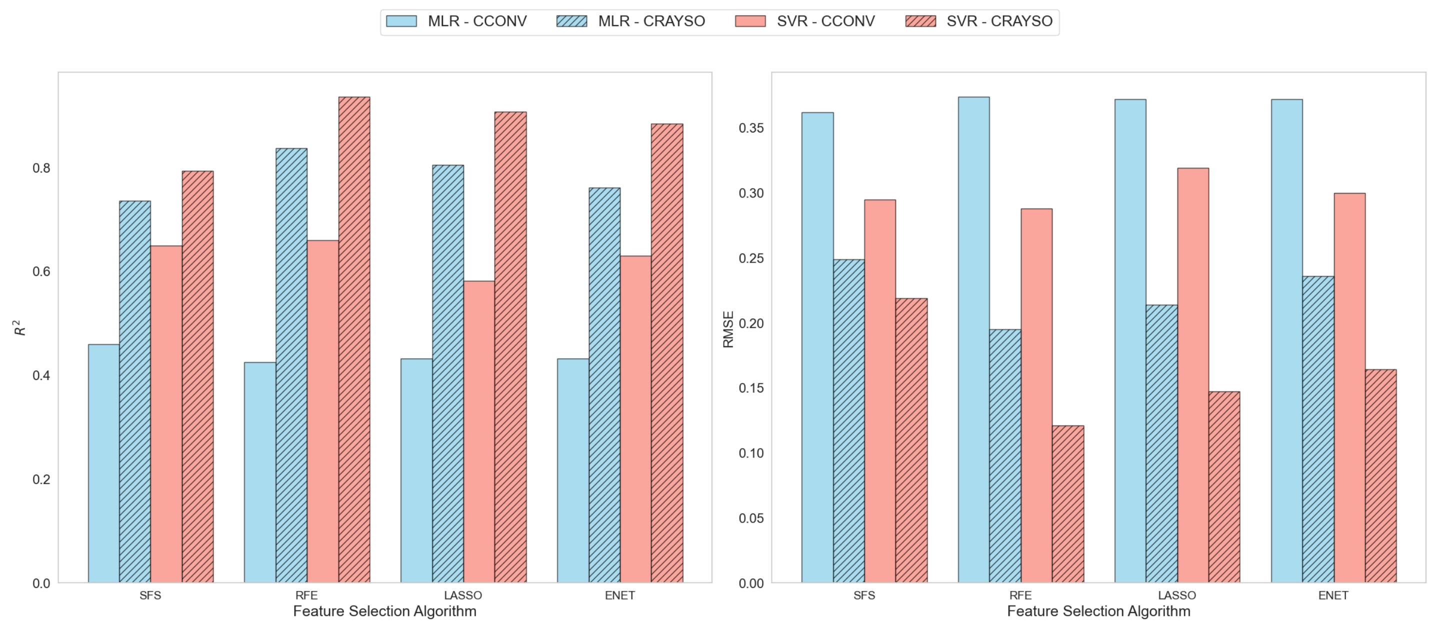

Figure 7: Leave-one-out cross validation results for the regional models MLR and SVR, using R^2 and RMSE as performance metrics

Our findings indicated that feature selection methods significantly impact the performance of SVR model in estimating the most critical parameters of CEQUEAU. Regarding CRAYSO, the following models: RFE-SVR, LASSO-SVR, and ENET-SVR demonstrated top performance, with R^2 values of 0.94, 0.91, and 0.88, respectively. The SFS-SVR method also showed good overall performance with an R^2 of 0.79. However, for CCONV, the models exhibited lower accuracy, with R^2 values ranging from 0.58 to 0.66. Once again, RFE-SVR outperformed all models, with SFS-SVR ranking second, while LASSO-SVR and ENET-SVR showed similar performance, with LASSO-SVR displaying fewer underestimated values. Our results indicated that our modeling procedure was able to explain a greater variance in CRAYSO compared to CCONV. Overall, using SVR, features selected using RFE consistently outperformed regularization and SFS selected features for both parameters.

Figure 8 illustrates observed versus predicted regression plots using the SVR model (Leave-one-out spatial evaluation in Figure 1). The figure revealed strong agreement between observed and estimated model parameters, particularly with RFE-SVR. The plots showed that high CRAYSO and CCONV values were typically associated with larger basins featuring wider channels (rivers: 1, 2, 6, 7, 9, 16, 23, 25, 35), while lower values were observed in smaller basins with narrower channels (rivers: 22, 27, 29, 31, 32, 33, 34). For CRAYSO, some minor deviations were noted for certain rivers, particularly for Gilbert River (23), LaHave River (28), and rivers with an outlet in Chaleur Bay, such as rivers 5, 12, 13, 15, and 17. Conversely, the data points for CCONV exhibited more scattering, indicating less consistent model performance. However, the LASSO-SVR plot demonstrated reduced scattering, particularly for underestimated CCONV values in rivers 1, 5, 8, 20, 23, 28, and 35.

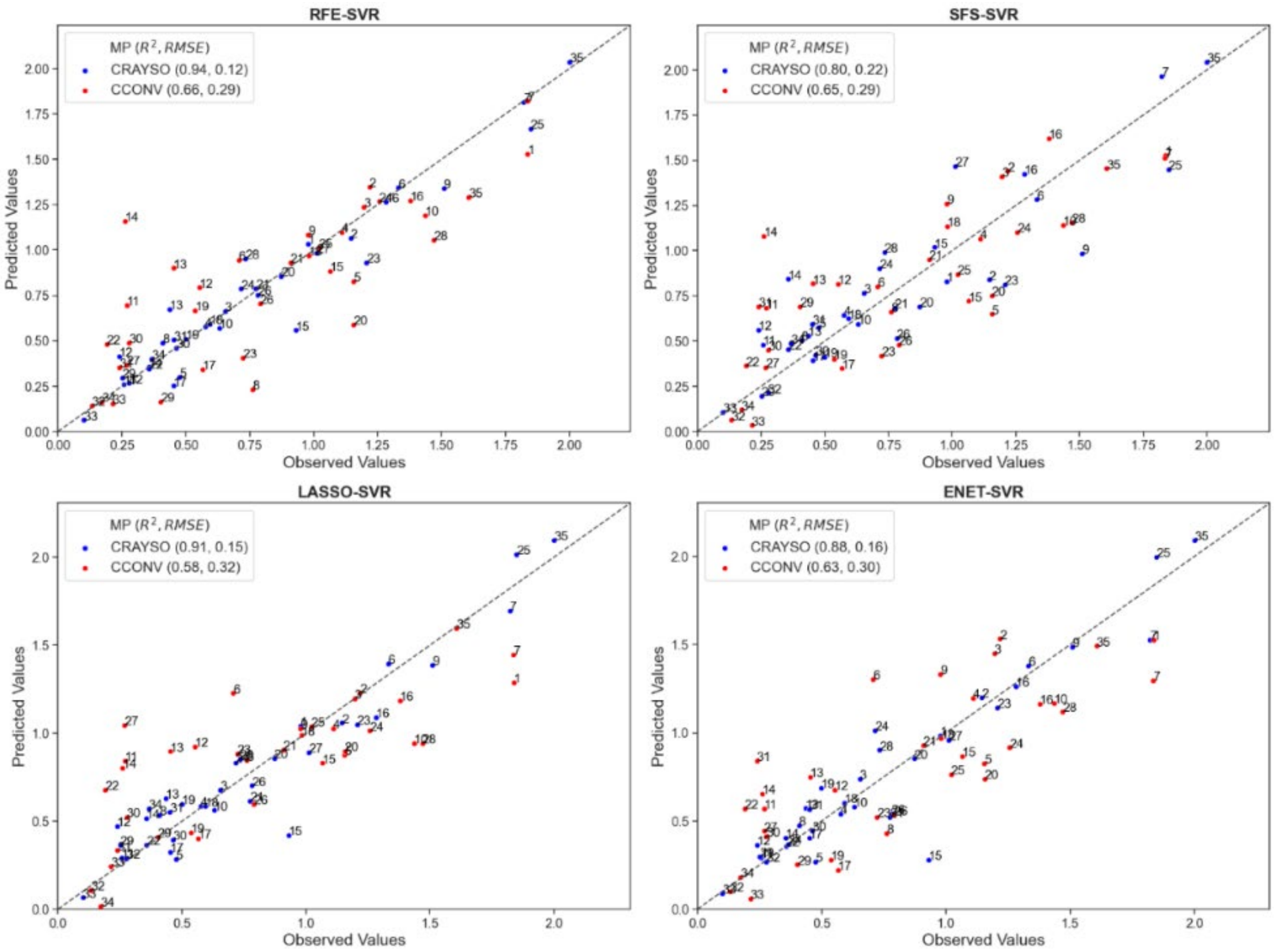

Figure 8: Regression plots for SVR as the regional estimation model across all feature selection algorithms.

Figure 9a displays the optimal number of selected features for CRAYSO and CCONV using all FS methods. On average, CRAYSO had 14 features selected, while CCONV had 11. It is worth noting that SFS-SVR selected the least features for CRAYSO, whereas RFE-SVR, LASSO, and ENET chose 13, 15, and 17 features, respectively. As for CCONV, LASSO resulted in the most parsimonious model, choosing 7 features, while SFS-SVR, ENET, and RFE-SVR selected 11, 12, and 14 features.

Figure 9b shows the condition number (C) obtained from the correlation matrix of the chosen feature sets. The condition number was used to evaluate multicollinearity among predictors. According to Chatterjee & Hadi. (2015), a threshold of 15 suggests potential multicollinearity and values above 30 necessitate corrective action. It is promising to note that all models constructed using the chosen features have C values below 15, indicating that they are likely to be more concise and less susceptible to multicollinearity issues.

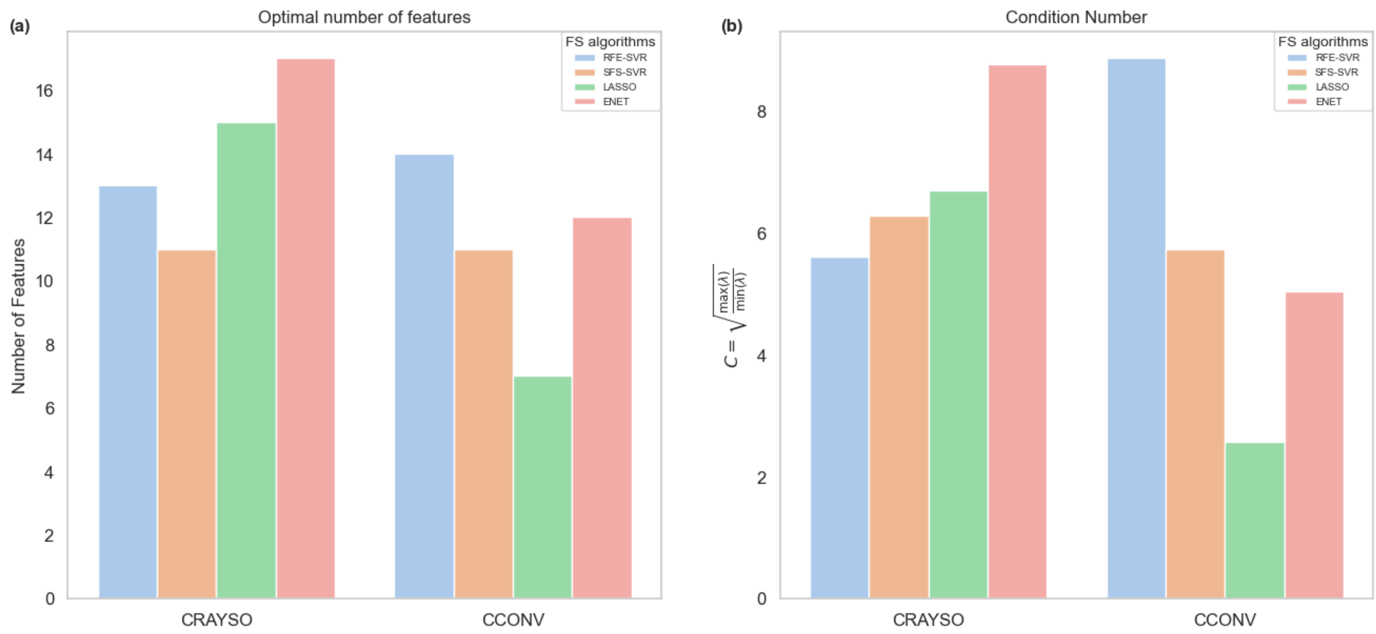

Figure 9: a) Optimal number of features for each FS method and b) the corresponding condition numbers for both the most sensitive parameters, CRAYSO and CCONV.

3.3.2- Water temperature

We injected SVR and MLR-generated parameters in the CEQUEAU model to compute the Tw time series. In Figure 10, we compared the performance of SVR and MLR using the skill score and found that SVR consistently outperformed the standard regression benchmark model (MLR), with a median skill score of 0.05 across all feature selection methods (Temporal evaluation in Figure 1). We will focus on SVR model going forward. For detailed MLR results, the reader is referred to Figure S4 in the supplementary materials.

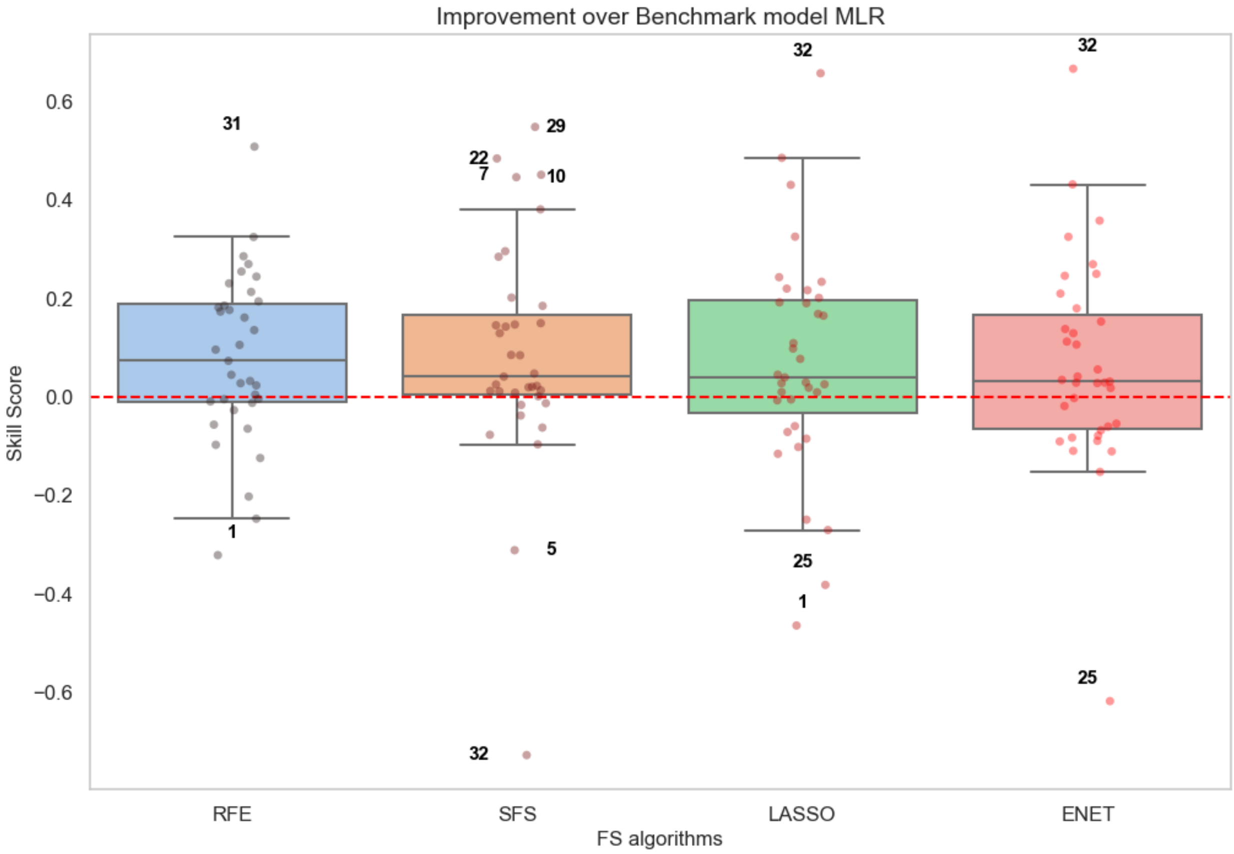

Figure 10: Skill Score of ML Models (SVR) in reference to benchmark MLR Model.

Figure 11 compares the Tw regional modeling performance using SVR as the regional estimation model (Temporal evaluation). The benchmark model, $Post_{SA}$, showed a mean RMSE of 1.75 °C, with 26 rivers achieving an RMSE below 2 °C. The highest RMSE (3.26 °C) was observed in the southernmost watershed, the Sheepscot River (21), while the lowest RMSE (0.77 °C) was found in the northernmost watershed, the Aux Mélèzes River (35).

Models 2 and 4 had the highest mean RMSEs of 2.1°C and 2.03°C, respectively, with greater variability in their predictions. Model 3 demonstrated an improvement in reducing this variability, with a mean RMSE of 1.95°C. Model 1 performed the best, achieving a mean RMSE of 1.89°C, closely aligning with the benchmark model 5.

Some outliers were identified, with models 3 and 4 showing significant prediction errors for the Gouffre River (15), with RMSEs of 3.70°C and 4.47°C, respectively. Model 1 faced accuracy challenges for the Nova Scotian rivers, LaHave and Sackville (28, 29), although it slightly improved performance in the Sheepscot River (from 3.26 to 3.08 °C). Overall, Model 1 performed well, with 25 rivers achieving RMSE values below 2 °C, while Models 3 and 4 followed with 22 rivers, and Model 2 with 18 rivers.

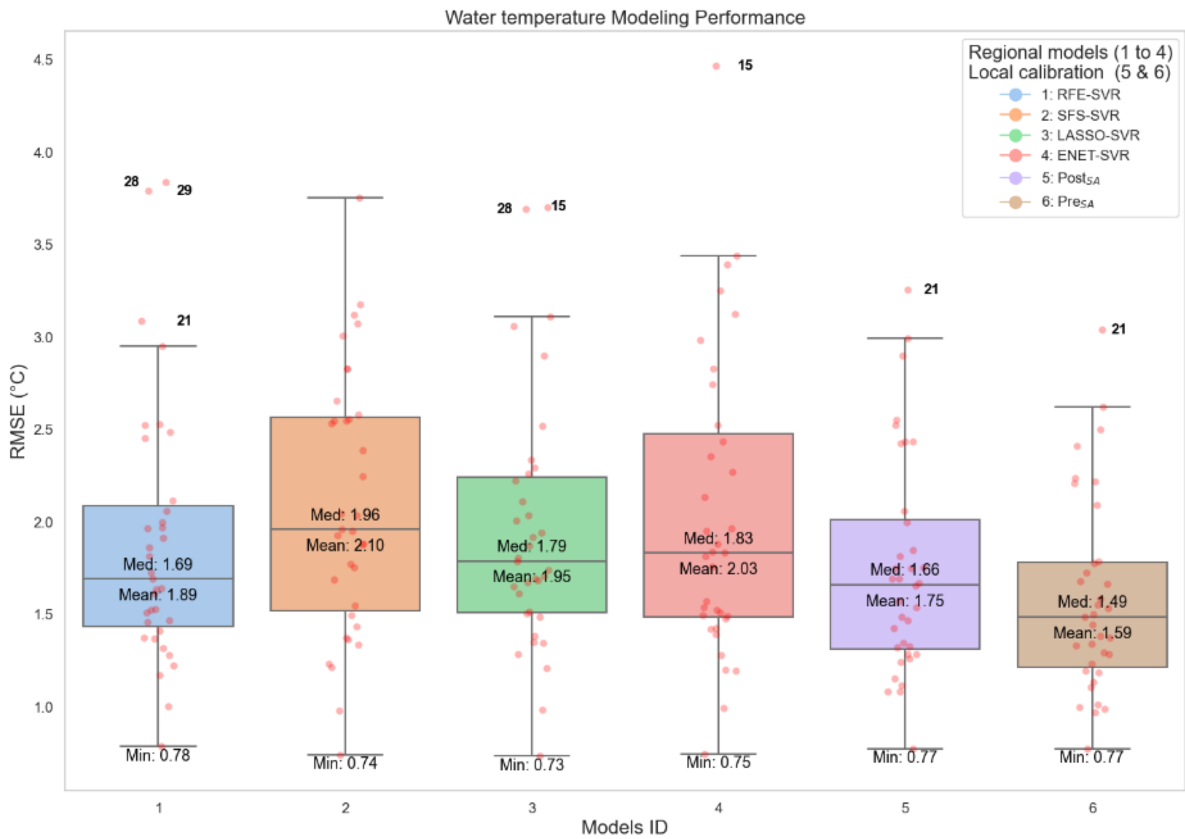

Figure 11: Water temperature modeling performance using regional models (1 to 4) and local calibrations (5 & 6).

3.4- Feature importance

The detailed heatmap in Figure 12 shows the selected predictors for CRAYSO and CCONV parameters. A selection count of four indicates unanimous choice across all models. For CRAYSO, five predictors (LAKE, FOREST, LOSA, D2O, DVME) were consistently selected by all models, and nine features were selected by at least three models. Each model discarded different features: RFE-SVR (cropland uses), SFS-SVR (elevation, upstream wetlands percentage, watershed orientation, low cloud coverage quantiles, aspect, and Gaussian surface curvature), LASSO (high insolation quantiles), and ENET (drainage density).

All models consistently selected three main land use features (LAKE, FOREST, CROP), except RFE-SVR, which chose WET over CROP. Each model included at least two climatic features (25^{th}CC, 90^{th}SSR, 90^{th}Ptot), although SFS-SVR exclusively used 90^{th}SSR, which impacted its prediction accuracy. Regarding topographical features, SFS-SVR selected upstream channel orientation over watershed orientation and used DOGS instead of GSCV for land surface curvature.

Five predictors were systematically selected to estimate CCONV: MIND, LAKE, ORWSH, D2O, and 90thVV, while four features were selected three times. All models included upstream cropland uses, but SFS-SVR uniquely chose upstream lake percentage as the land cover predictor. While all models used two orientation features (ORUP, ORWSH), LASSO solely used watershed orientation. For topographical features, all models selected at least two features from GSCV, DFME, and DVME. However, LASSO chose only DVME, and SFS-SVR preferred DFME over DVME.

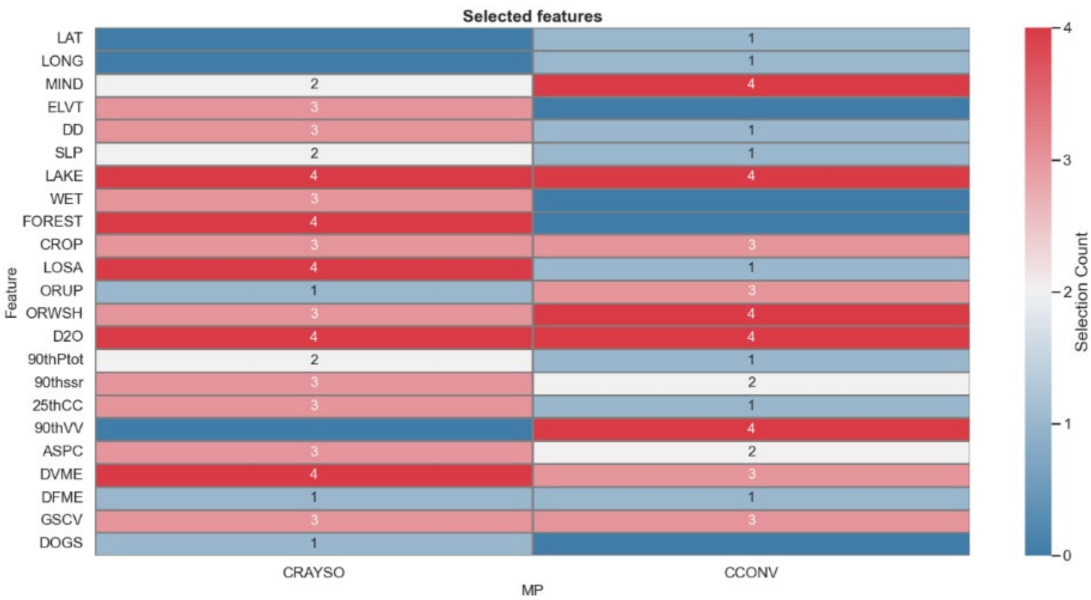

Figure 12: Selection count heatmap for selected features across FS methods for both most sensitive parameters CRAYSO and CCONV.

Linear regression models have the advantage of interpretability, providing a clear understanding of the relationships between predictors and the dependent variable. However, when using SVR with a linear kernel as the regional estimation model, the direction of the regression coefficients can be misleading due to the unitless nature of the target variables (CRAYSO and CCONV). To address this, feature importance was determined by the square of the regression coefficients in the SVR models (Guyon et al., 2002). Additionally, preprocessing is crucial in linear models, as standardization to zero mean and unit variance ensures that feature scales are comparable.

Figure 13 showed the importance of features for CRAYSO and CCONV. The figure consisted of two parts: Figure 13 (a1) and (a2) displays the average feature importance based on the frequency of feature selection, while Figure 13 (b1) and (b2) provided detailed feature importance plots for each regional estimation model, offering a comprehensive view of feature importance. Using the top-performing RFE-SVR model, the most influential features for CRAYSO (with absolute weight coefficients > 0.1) are: 25thCC, DVME, D2O, DD, FOREST, LOSA, GSCV, ELVT, LAKE, ORWSH, ASPC, and WET. In contrast, it was the least sparse model for CCONV (14 features), and the most important features using this model are MIND, 25thCC, 90thVV, LAKE, LONG, 90thSSR, ORWSH, ORUP, DD, CROP, DVME, GSCV, ASPC, and D2O. However, LASSO-SVR significantly reduced underestimations across most rivers with low to medium CCONV values, achieving an RMSE comparable to RFE-SVR while maintaining a more parsimonious model with only seven features, ranked as follows: 90thVV, MIND, LAKE, CROP, ORWSH, D2O, and DVME

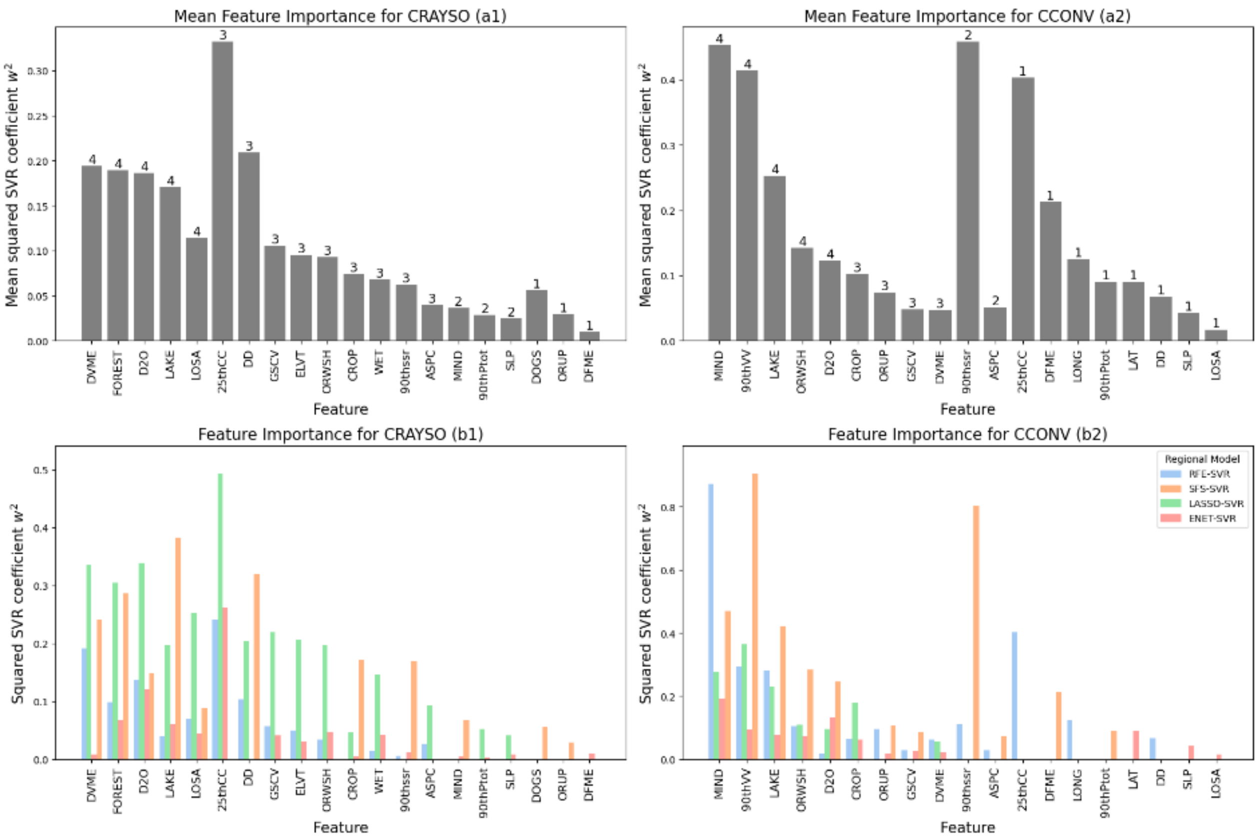

Figure 13: Squared regression coefficient of the SVR regional model to assess feature importance for both CRAYSO and CCONV.

This study tested the abilities of commonly used regression MLR and machine learning SVR for regional estimation of CEQUEAU's highly sensitive parameters, CRAYSO and CCONV. The findings revealed that for both parameters, SVR consistently outperformed the benchmark MLR model. Four different FS algorithms were evaluated to identify optimal features. RFE-SVR emerged as the most parsimonious model for CRAYSO, striking a balance between high predictive accuracy and complexity. While LASSO-SVR, ENET-SVR, and SFS-SVR also showed good performance for CRAYSO, they were less parsimonious and provided no significant improvement in accuracy compared to RFE-SVR. For CCONV, both RFE-SVR and SFS-SVR were top performers, but they tended to significantly underestimate values. In contrast, LASSO-SVR effectively balanced between accuracy and complexity by selecting a sparse feature set, roughly half the size of the top performers, and delivered a robust predictive performance for previously underestimated values. Overall, the condition number of the feature set correlation matrix revealed no collinearity issues among the relevant predictors for CRAYSO and CCONV, indicating that all models effectively selected a relevant and non-redundant feature set.

4- Discussion

This study aimed to assess the feasibility of regionalizing CEQUEAU thermal parameters to estimate Tw in ungauged locations. This includes pristine rivers spanning from the southernmost habitats of Atlantic salmon in the Gulf of Maine (USA) to the arctic climates of northern Quebec (Ungava Bay) and Labrador. A key challenge was defining parameter boundaries that account for diverse watershed characteristics while ensuring physically plausible relationships. Through sensitivity analyses and a two-phase calibration strategy, we effectively addressed this issue, highlighting the value of fixing low-sensitivity parameters within homogeneous regions to improve calibration efficiency and mitigate equifinality issues (Feigl et al., 2022).

CEQUEAU in northeastern America

Local calibration demonstrated a minimal RMSE increase of 0.25 °C after fixing low-sensitivity parameters, with an overall mean RMSE below 2 °C. These results indicate a robust calibration process suitable for regionalization purposes (Figure *11*). This finding highlights the suitability of the chosen parameter boundaries for regionalization. Most thermal parameters proved stable enough to be fixed, with CRAYSO and CCONV identified as the most sensitive parameters. This finding suggests that CEQUEAU may be over-parameterized for this region.

RFE-SVR and LASSO-SVR emerged as the most effective regional models for estimating CRAYSO and CCONV in ungauged locations offering both high predictive accuracy and simplicity, as demonstrated through a leave-one-out cross validation procedure within the study region. RFE and LASSO are particularly advantageous because they require tuning only a single parameter, making them efficient and easy to implement. SVR successfully captured the spatial variability of CRAYSO and CCONV, the most critical parameters, across the study region. However, the models performed better for CRAYSO than for CCONV, likely due to the exclusion of key predictors or the need for non-linear models to capture complex relationships between predictors and thermal parameters. Additionally, exploring surface elevation and curvature attributes at different spatial scales (resampling) may improve results (Houndekindo & Ouarda, 2024).

Advancement of regional Tw modeling

The leave-one-out cross validation suggested the model's performance is in line with advance in stream temperature regional modeling in northeastern America. Prior to this work, the only research developing a regional framework for estimating Tw in ungauged northeastern Canadian rivers was conducted by Ouarda et al. (2022). Their statistical approach showed the superiority of Generalized additive model (GAM) over MLR. They utilized data from over 120 stations, each with a minimum of four years of summer records, to model river thermal regime quantiles in ungauged rivers of eastern Canada, achieving RMSEs between 2°C and 3°C through a leave-one-out validation procedure. However, their study excluded streams from Prince Edward Island, Anticosti Island, and Maine, USA. Despite these limitations, their work represents a significant advancement in Tw stream modeling in ungauged sites within the study region.

In another study, Weierbach et al. (2022) shown that ML models such as SVR outperform traditional statistical approaches (MLR) in estimating monthly Tw in ungauged locations where discharge information is available in Pacific northwest and Mid Atlantic regions in the U.S. Using 78 stations and a minimum of 8 years of co-located Tw and discharge observations, they achieved a regional median RMSE of 1°C. More recently, Weller et al. (2024) developed a regression model to predict August mean stream Tw for British Columbia, Canada, using land cover, physiographic, and climatic characteristics from over 560 sites. They used a 10-fold cross-validation which yields an RMSE of 1.53 °C. Our regionalization approach focused on pristine catchments, achieving improved predictive performance in eastern Canadian rivers, while demonstrating comparable results to studies in the U.S. and western Canada.

Prior regional hydrological and Tw models have required extensive datasets with numerous watershed attributes (Kratzert et al., 2019; Rahmani et al., 2021). A notable contribution of our study is the efficient use of explanatory variables. Employing different FS methods allowed for a robust procedure, as the selected attributes for each parameter were consistently validated across multiple methods. Our findings are consistent with Souaissi et al. (2023a), who employed RFE and LASSO for feature selection in a regional modeling framework to estimate thermal river quantiles in Switzerland. Their study highlighted the comparable performance of these methods, emphasizing their effectiveness in identifying relevant and non-redundant features for Tw modeling at ungauged sites.

Key covariates for CRAYSO & CCONV

All models accurately captured that high CRAYSO and CCONV values were typically found in larger basins with wider channels, while smaller basins with narrower channels exhibited lower values. This could be explained by the increased solar radiation and wind exposure in wider rivers with relatively lower crown closure due to reduced shading from vegetation and topographical features (Jackson et al., 2018; Maheu & Caissie, 2023; Monk & Dugdale, 2023; O'Sullivan et al., 2019; St-Hilaire et al., 2023).

Incoming solar radiation is the dominant flux influencing stream heat budgets (Caissie, 2006; Leach et al., 2023). CRAYSO showed heightened sensitivity in climate-driven rivers with wide mainstem channels, where low cloud cover quantiles emerged as the most significant climatic predictor, reflecting solar radiation variability caused by weather patterns and wildfire smoke (David et al., 2018; Siegel et al., 2023). Topographical features, including DVME, aspect, channel slope, and GSCV, significantly influenced CRAYSO by altering exposure to solar radiation (Lookingbill & Urban, 2003; I. D. Moore et al., 1991; O'Sullivan et al., 2019). Land cover attributes also played a role. Upstream forest percentage regulated the thermal regime through shading (Garner et al., 2014; St-Hilaire et al., 2000) , while lakes and wetlands, warmed by solar exposure, impacted downstream Tw (Abidi et al., 2022; Arora et al., 2018; O'Sullivan et al., 2019). Coastal proximity influenced CRAYSO through climatic variability (teleconnection) driven by ocean-atmospheric interactions, such as the North Atlantic Oscillation and Atlantic Multidecadal Oscillation, affecting northeastern Canadian summers (Collins, 2023; Jackson et al., 2018; Ouarda et al., 2024; Ouarda & Charron, 2018). Watershed orientation was another critical factor, with north-south streams receiving more solar radiation compared to west-east streams shaded by equatorial-facing riparian vegetation (Garner et al., 2017; Jackson et al., 2018; Leach et al., 2023). Secondary factors like drainage density and loamy sand soils contributed by affecting shading and groundwater recharge (Jeong et al., 2013; Johnson et al., 2024; Kurylyk et al., 2015; O'Sullivan et al., 2020; T. B. Ouarda et al., 2022).

CCONV showed high sensitivity in island landscapes with groundwater-fed rivers and strong wind patterns, such as the rivers located in Prince Edward and Anticosti islands and Ducktrap River in Maine. Unique wind dynamics in these regions, like the channeling effect of the Gulf of St. Lawrence and coastal windstorms in Penobscot Bay, enhance convection processes (Beaucage et al., 2007) (Spicer et al., 2021; Townsend et al., 2023). Key climatic predictors include high wind speeds and low cloud coverage, with wind-induced mixing disrupting vertical stratification and enhancing heat exchange within the water column (Caissie, 2016; Ferchichi et al., 2021; Maheu et al., 2014). Topographical features such as mainstem channel depth also play a significant role, as wide and shallow channels promote efficient heat exchange, while deeper channels moderate temperature changes due to thermal inertia (Maheu et al., 2014; Sinokrot & Stefan, 1993). Land cover, including upstream lakes, influences CCONV as large lakes exposed to wind increases Tw and heat exchange (Abidi et al., 2022). Watersheds like the Ducktrap River and PEI catchments, characterized by high agricultural activity and flat terrain, have long wind fetches exposing channels to high wind speeds and amplifying convection processes (Dehghani-Sanij et al., 2022; Hall & Swingler, 2018). Additional factors like watershed orientation, distance to the coast, and topographical attributes (DVME, GSCV) affect wind sheltering and convective heat dynamics, reinforcing the importance of geographic and climatic variability in predicting convection processes (Garner et al., 2017; Jackson et al., 2021; O'Sullivan et al., 2019).

Convective heat transfer, though typically the smallest component of the total energy flux, remains significant (Maheu & Caissie, 2023; Morin & Couillard, 1990). Advances in wind speed modeling in Canada have highlighted the importance of relative topographical position (DVME) for estimating high wind speed quantiles and surface curvature (GSCV) for low wind speed quantiles (Houndekindo & Ouarda, 2023, 2024). These features play a crucial role in convection processes, supporting our findings on their importance for predicting CCONV. This understanding enhances our mechanistic view of stream Tw dynamics in Atlantic salmon habitats.

5- Conclusion

Our study represents a significant advancement in understanding and managing Atlantic salmon rivers in eastern North America by integrating deterministic and machine-learning approaches. For CEQUEAU, this study marks a significant milestone, showcasing its capability to accurately estimate hydrological processes in a regional context across a broad latitudinal gradient. A global sensitivity analysis, also a first for CEQUEAU, identified parameters controlling shortwave radiation and sensible heat fluxes as key drivers of thermal stream dynamics in pristine rivers northeastern Canada and Maine, USA. This work enabled the regional estimation of these parameters, enabling accurate mean daily Tw predictions for rivers with wide channels influenced by solar radiation and convection, provided discharge data is available.

Despite these advancements, certain limitations remain. The CEQUEAU model oversimplifies winter river thermal regimes by setting Tw to zero when air temperatures fall below freezing, which is physically inaccurate and affects snowmelt and spring freshet timing. However, as salmonids face the greatest thermal stress during high summer temperatures, our modeling focused on accurately simulating Tw during this critical period. The requirement for both discharge and Tw gauging stations at river outlets limited the database size, as few rivers have both measurements, restricting a more comprehensive regionalization. Nonetheless, the leave-one-out cross-validation produced satisfactory results. Model performance was notably better in regions with longer data series and lower geographic dispersion, such as Quebec, New Brunswick, and Prince Edward Island. These findings align with previous studies emphasizing the importance of data availability and quality for regional Tw modeling in eastern Canada (Charron et al., 2020; Ouarda et al., 2022).

For the future regionalization perspective, research should aim to extend this study to other geographic areas and databases with varying characteristics, such as river headstreams, rivers located on the Pacific Coast, or mountainous watersheds, where additional parameters may prove to be critical. This could necessitate regionalizing additional parameters and employing non-linear models to better capture the complex relationships between predictors and model parameters. Our study did not include dammed catchments. Incorporating these regulated catchments in future research could help explain additional variability in CEQUEAU thermal parameters, as the model is capable of accurately simulating hydrological conditions in regulated systems (Khorsandi et al., 2022, 2023; Oyinlola et al., 2023; Rahmati, 2023).

6- Author Contribution

Ilias Hani: Conceptualization, Methodology, Formal analysis, Software, Investigation, Writing—Original Draft.

André St-Hilaire: Conceptualization, Methodology, Writing—Review & Editing, Supervision, Funding acquisition.

Taha B. M. J. Ouarda: Conceptualization, Methodology, Review & Editing, Supervision, Funding acquisition.

7- Declaration of Competing Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

8- Acknowledgements

This work was partly funded by the MITACS, the Atlantic Salmon Research Joint Venture (ASRJV) and the Foundation for the Conservation of Atlantic Salmon (FCAS). The authors would like to thank the Ministry of Environment and Climate Change Canada (ECCC) and C. Beaulieu (RivTemp) for the access to datasets.

9- Data availability

Data will be made available on request.

10- References

Abdelhamed, M. S., Elshamy, M. E., Razavi, S., & Wheater, H. S. (2023). Challenges in Hydrologic-Land Surface Modeling of Permafrost Signatures—A Canadian Perspective. *Journal of Advances in Modeling Earth Systems*, *15*(3), e2022MS003013. https://doi.org/10.1029/2022MS003013

Abidi, O., St-Hilaire, A., Ouarda, T. B., Charron, C., Boyer, C., & Daigle, A. (2022). Regional thermal analysis approach: A management tool for predicting water temperature metrics relevant for thermal fish habitat. *Ecological Informatics*, *70*, 101692. https://doi.org/10.1016/j.ecoinf.2022.101692

Arora, R., Toffolon, M., Tockner, K., & Venohr, M. (2018). Thermal discontinuities along a lowland river: The importance of urban areas and lakes. *Journal of Hydrology*, *564*, 811–823. https://doi.org/10.1016/j.jhydrol.2018.05.066

Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G., & Romero-Lopez, R. (2019). Streamflow prediction in ungauged basins: Analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. *Hydrological Sciences Journal*, *64*(11), 1297–1311. https://doi.org/10.1080/02626667.2019.1639716

Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2014). Comparison of Stochastic Optimization Algorithms in Hydrological Model Calibration. *Journal of Hydrologic Engineering*, *19*(7), 1374–1384. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000938

Auger, A., & Hansen, N. (2012). *Tutorial CMA-ES: evolution strategies and covariance matrix adaptation*. 827–848. https://doi.org/10.1145/2330784.2330919

Beaucage, P., Glazer, A., Choisnard, J., Yu, W., Bernier, M., Benoit, R., & Lafrance, G. (2007). Wind assessment in a coastal environment using synthetic aperture radar satellite imagery and a numerical weather prediction model. *Canadian Journal of Remote Sensing*, *33*(5), 368–377. https://doi.org/10.5589/m07-043

Bindas, T., Tsai, W., Liu, J., Rahmani, F., Feng, D., Bian, Y., Lawson, K., & Shen, C. (2024). Improving River Routing Using a Differentiable Muskingum-Cunge Model and Physics-Informed Machine Learning. *Water Resources Research*, *60*(1), e2023WR035337. https://doi.org/10.1029/2023WR035337

Bond, N., & Bond, M. N. (2022). *Package 'hydrostats.'* https://github.com/nickbond/hydrostats

Bosmans, J., Wanders, N., Bierkens, M. F., Huijbregts, M. A., Schipper, A. M., & Barbarossa, V. (2022). FutureStreams, a global dataset of future streamflow and water temperature. *Scientific Data*, *9*(1), 307. https://doi.org/10.1038/s41597-022-01410-6

Boyer, C., St-Hilaire, A., Bergeron, N., Daigle, A., Curry, R., Caissie, D., & Gillis, C. (2016). RivTemp: A water temperature network for Atlantic salmon rivers in eastern Canada. *Water News*, *35*(2), 10–15.

Boyer, C., St-Hilaire, A., & Bergeron, N. E. (2021). Defining river thermal sensitivity as a function of climate. *River Research and Applications*, *37*(10), 1548–1561. https://doi.org/10.1002/rra.3862

Bray, E. N., Dozier, J., & Dunne, T. (2017). Mechanics of the energy balance in large lowland rivers, and why the bed matters. *Geophysical Research Letters*, *44*(17), 8910–8918. https://doi.org/10.1002/2017GL075317

Caissie, D. (2006). The thermal regime of rivers: A review. *Freshwater Biology*, *51*(8), 1389–1406. https://doi.org/10.1111/j.1365-2427.2006.01597.x

Caissie, D. (2016). River evaporation, condensation and heat fluxes within a first-order tributary of Catamaran Brook (New Brunswick, Canada). *Hydrological Processes*, *30*(12), 1872–1883. https://doi.org/10.1002/hyp.10744

Charron, C., St-Hilaire, A., Boyer, C., Daigle, A., Ouarda, T. B., & Bergeron, N. (2020). *Regional analysis and modelling of water temperature metrics for Atlantic Salmon (Salmo salar) in Eastern Canada. Final report.* https://espace.inrs.ca/id/eprint/11361/1/R1933.pdf

Charron, C., St-Hilaire, A., Ouarda, T. B., & Van Den Heuvel, M. R. (2021). Water temperature and hydrological modelling in the context of environmental flows and future climate change: Case study of the wilmot river (Canada). *Water*, *13*(15), 2101. https://doi.org/10.3390/w13152101

Chatterjee, S., & Hadi, A. S. (2015). *Regression analysis by example*. John Wiley & Sons.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., & Peters-Lidard, C. D. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, *21*(7), 3427–3440. https://doi.org/10.5194/hess-21-3427-2017

Coffey, R., Paul, M. J., Stamp, J., Hamilton, A., & Johnson, T. (2019). A review of water quality responses to air temperature and precipitation changes 2: Nutrients, algal blooms, sediment, pathogens. *JAWRA Journal of the American Water Resources Association*, *55*(4), 844–868. https://doi.org/10.1111/1752-1688.12711

Collins, S. (2023). *A spatio-temporal stream temperature model for the southern Coast Mountains of British Columbia*. https://doi.org/10.14288/1.0435654

David, A. T., Asarian, J. E., & Lake, F. K. (2018). Wildfire smoke cools summer river and stream water temperatures. *Water Resources Research*, *54*(10), 7273–7290. https://doi.org/10.1029/2018WR022964

Dehghani-Sanij, A., Al-Haq, A., Bastian, J., Luehr, G., Nathwani, J., Dusseault, M., & Leonenko, Y. (2022). Assessment of current developments and future prospects of wind energy in Canada. *Sustainable Energy Technologies and Assessments*, *50*, 101819. https://doi.org/10.1016/j.seta.2021.101819

Deka, P. C. (2014). Support vector machine applications in the field of hydrology: A review. *Applied Soft Computing*, *19*, 372–386. https://doi.org/10.1016/j.asoc.2014.02.002

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*(2), 265–282. https://doi.org/10.1111/j.2044-8317.1992.tb00992.x

Dugdale, S. J., Hannah, D. M., & Malcolm, I. A. (2017). River temperature modelling: A review of process-based approaches and future directions. *Earth-Science Reviews*, *175*, 97–113. https://doi.org/10.1016/j.earscirev.2017.10.009

Dugdale, S. J., St-Hilaire, A., & Allen Curry, R. (2017). Automating drainage direction and physiographic inputs to the CEQUEAU hydrological model: Sensitivity testing on the lower Saint John River watershed, Canada. *Journal of Hydroinformatics*, *19*(3), 469–492. https://doi.org/10.2166/hydro.2017.051

Ebersole, J. L., Liss, W. J., & Frissell, C. A. (2003). Cold water patches in warm streams: Physicochemical characteristics and the influence of shading 1. *JAWRA Journal of the American Water Resources Association*, *39*(2), 355–368. https://doi.org/10.1111/j.1752-1688.2003.tb04390.x

Essaid, H. I., & Caldwell, R. R. (2017). Evaluating the impact of irrigation on surface water–groundwater interaction and stream temperature in an agricultural watershed. *Science of the Total Environment*, *599*, 581–596. https://doi.org/10.1016/j.scitotenv.2017.04.205

Feigl, M., Herrnegger, M., Klotz, D., & Schulz, K. (2020). Function Space Optimization: A Symbolic Regression Method for Estimating Parameter Transfer Functions for Hydrological Models. *Water Resources Research*, *56*(10), e2020WR027385. https://doi.org/10.1029/2020WR027385

Feigl, M., Lebiedzinski, K., Herrnegger, M., & Schulz, K. (2021). Machine-learning methods for stream water temperature prediction. *Hydrology and Earth System Sciences*, *25*(5), 2951–2977. https://doi.org/10.5194/hess-25-2951-2021

Feigl, M., Roesky, B., Herrnegger, M., Schulz, K., & Hayashi, M. (2022). Learning from mistakes—Assessing the performance and uncertainty in process-based models. *Hydrological Processes*, *36*(2), e14515.

Ferchichi, H., St-Hilaire, A., Ouarda, T. B., & Lévesque, B. (2021). Impact of the future coastal water temperature scenarios on the risk of potential growth of pathogenic Vibrio marine bacteria. *Estuarine, Coastal and Shelf Science*, *250*, 107094. https://doi.org/10.1016/j.ecss.2020.107094

Fernandez-Illescas, C. P., Porporato, A., Laio, F., & Rodriguez-Iturbe, I. (2001). The ecohydrological role of soil texture in a water-limited ecosystem. *Water Resources Research*, *37*(12), 2863–2872. https://doi.org/10.1029/2000WR000121

Ficklin, D. L., Hannah, D. M., Wanders, N., Dugdale, S. J., England, J., Klaus, J., Kelleher, C., Khamis, K., & Charlton, M. B. (2023). Rethinking river water temperature in a changing, human-dominated world. *Nature Water*, *1*(2), 125–128. https://doi.org/10.1038/s44221-023-00027-2

FitzGerald, A. M., & Martin, B. T. (2022). Quantification of thermal impacts across freshwater life stages to improve temperature management for anadromous salmonids. *Conservation Physiology*, *10*(1), coac013. https://doi.org/10.1093/conphys/coac013

Florinsky, I. V. (2017). An illustrated introduction to general geomorphometry. *Progress in Physical Geography*, *41*(6), 723–752. https://doi.org/10.1177/0309133317733667

Fouad, G., & Loáiciga, H. A. (2020). Independent variable selection for regression modeling of the flow duration curve for ungauged basins in the United States. *Journal of Hydrology*, *587*, 124975. https://doi.org/10.1016/j.jhydrol.2020.124975

Gallice, A., Schaefli, B., Lehning, M., Parlange, M. B., & Huwald, H. (2015). Stream temperature prediction in ungauged basins: Review of recent approaches and description of a new physics-derived statistical model. *Hydrology and Earth System Sciences*, *19*(9), 3727–3753. https://doi.org/10.5194/hess-19-3727-2015

Garner, G., Malcolm, I. A., Sadler, J. P., & Hannah, D. M. (2014). What causes cooling water temperature gradients in a forested stream reach? *Hydrology and Earth System Sciences*, *18*(12), 5361–5376. https://doi.org/10.5194/hess-18-5361-2014

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. *Journal of Hydrology*, *553*, 471–485. https://doi.org/10.1016/j.jhydrol.2017.03.024

Gatien, P., Arsenault, R., Martel, J.-L., & St-Hilaire, A. (2022). Using the ERA5 and ERA5-Land reanalysis datasets for river water temperature modelling in a data-scarce region. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, 1–18. https://doi.org/10.1080/07011784.2022.2113917

Gharib, A., & Davies, E. G. (2021). A workflow to address pitfalls and challenges in applying machine learning models to hydrology. *Advances in Water Resources*, *152*, 103920. https://doi.org/10.1016/j.advwatres.2021.103920

Gillis, C.-A., Ouellet, V., Breau, C., Frechette, D., & Bergeron, N. (2023). Assessing climate change impacts on North American freshwater habitat of wild Atlantic salmon-urgent needs for collaborative research. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, 1–25.

Girard, P., Boisclair, D., & Leclerc, M. (2003). The effect of cloud cover on the development of habitat quality indices for juvenile Atlantic salmon (Salmo salar). *Canadian Journal of Fisheries and Aquatic Sciences*, *60*(11), 1386–1397. https://doi.org/10.1139/f03-118

Godsey, S., & Kirchner, J. W. (2014). Dynamic, discontinuous stream networks: Hydrologically driven variations in active drainage density, flowing channels and stream order. *Hydrological Processes*, *28*(23), 5791–5803. https://doi.org/10.1002/hyp.10310

Guo, Y., Zhang, Y., Zhang, L., & Wang, Z. (2021). Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. *WIREs Water*, *8*(1), e1487. https://doi.org/10.1002/wat2.1487

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(Mar), 1157–1182.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*, 389–422. https://doi.org/10.1023/A:1012487302797

Hall, M., & Swingler, A. (2018). Initial perspective on a 100% renewable electricity supply for Prince Edward Island. *International Journal of Environmental Studies*, *75*(1), 135–153. https://doi.org/10.1080/00207233.2017.1395246

Hani, I., St-Hilaire, A., & Ouarda, T. B. (2023). Machine-learning modeling of hourly potential thermal refuge area: A case study from the Sainte-Marguerite River (Quebec, Canada). *River Research and Applications*, *39*(9), 1763–1782. https://doi.org/10.1002/rra.4191

Hansen, N., & Ostermeier, A. (2001). Completely Derandomized Self-Adaptation in Evolution Strategies. *Evolutionary Computation*, *9*(2), 159–195. https://doi.org/10.1162/106365601750190398

Hastie, T., Tibshirani, R., & Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the lasso. *arXiv Preprint arXiv:1707.08692*. https://doi.org/10.48550/arXiv.1707.08692 Focus to learn more

Hay, L. E., LaFontaine, J. H., Van Beusekom, A. E., Norton, P. A., Farmer, W. H., Regan, R. S., Markstrom, S. L., & Dickinson, J. E. (2023). *Parameter estimation at the conterminous United States scale and streamflow routing enhancements for the National Hydrologic Model infrastructure application of the Precipitation-Runoff Modeling System (NHM-PRMS)* (2328–7055). US Geological Survey. https://doi.org/10.3133/tm6B10

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., & Schepers, D. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Heuvelmans, G., Muys, B., & Feyen, J. (2006). Regionalisation of the parameters of a hydrological model: Comparison of linear regression models with artificial neural nets. *Journal of Hydrology*, *319*(1), 245–265. https://doi.org/10.1016/j.jhydrol.2005.07.030

Houndekindo, F., & Ouarda, T. B. (2023). Comparative study of feature selection methods for wind speed estimation at ungauged locations. *Energy Conversion and Management*, *291*, 117324. https://doi.org/10.1016/j.enconman.2023.117324

Houndekindo, F., & Ouarda, T. B. (2024). Prediction of hourly wind speed time series at unsampled locations using machine learning. *Energy*, 131518. https://doi.org/10.1016/j.energy.2024.131518

Hundecha, Y., & Bárdossy, A. (2004). Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *Journal of Hydrology*, *292*(1), 281–295. https://doi.org/10.1016/j.jhydrol.2004.01.002

Hundecha, Y., Ouarda, T. B., & Bárdossy, A. (2008). Regional estimation of parameters of a rainfall-runoff model at ungauged watersheds using the “spatial” structures of the parameters within a canonical physiographic-climatic space. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2006WR005439

Isaak, D. J., & Luce, C. H. (2023). Elevation-dependent warming of streams in mountainous regions: Implications for temperature modeling and headwater climate refugia. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, *48*(2), 167–188. https://doi.org/10.1080/07011784.2023.2176788

Jackson, F. L., Fryer, R. J., Hannah, D. M., & Malcolm, I. A. (2017). Can spatial statistical river temperature models be transferred between catchments? *Hydrology and Earth System Sciences*, *21*(9), 4727–4745. https://doi.org/10.5194/hess-21-4727-2017

Jackson, F. L., Fryer, R. J., Hannah, D. M., Millar, C. P., & Malcolm, I. A. (2018). A spatio-temporal statistical model of maximum daily river temperatures to inform the management of Scotland's Atlantic salmon rivers under climate change. *Science of the Total Environment*, *612*, 1543–1558. https://doi.org/10.1016/j.scitotenv.2017.09.010

Jackson, F. L., Hannah, D. M., Ouellet, V., & Malcolm, I. A. (2021). A deterministic river temperature model to prioritize management of riparian woodlands to reduce summer maximum river temperatures. *Hydrological Processes*, *35*(8), e14314. https://doi.org/10.1002/hyp.14314

Jeong, D. I., Daigle, A., & 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 Research and Applications*, *29*(7), 805–821. https://doi.org/10.1002/rra.2574

Jeppesen, E., & Iversen, T. M. (1987). Two simple models for estimating daily mean water temperatures and diel variations in a Danish low gradient stream. *Oikos*, 149–155. https://doi.org/10.2307/3566020

Johnson, M. F., Albertson, L. K., Algar, A. C., Dugdale, S. J., Edwards, P., England, J., Gibbins, C., Kazama, S., Komori, D., & MacColl, A. D. (2024). Rising water temperature in rivers:

Ecological impacts and future resilience. *Wiley Interdisciplinary Reviews: Water*, e1724. https://doi.org/10.1002/wat2.1724

Kandala, R., Franssen, H. H., Chaudhuri, A., & Sekhar, M. (2024). The value of soil temperature data versus soil moisture data for state, parameter, and flux estimation in unsaturated flow model. *Vadose Zone Journal*, *23*(1), e20298. https://doi.org/10.1002/vzj2.20298

Khorsandi, M. (2024). *Adaptation d'outils de modélisation pour l'analyse d'impact sur le régime thermique des rivières dans le cadre de la gestion des barrages et du changement climatique* [Phd, Université du Québec, Institut national de la recherche scientifique]. https://espace.inrs.ca/id/eprint/15891/

Khorsandi, M., St-Hilaire, A., & Arsenault, R. (2022). Multisite calibration of a semi-distributed hydrologic and thermal model in a large Canadian watershed. *Hydrological Sciences Journal*, *67*(14), 2147–2174. https://doi.org/10.1080/02626667.2022.2132161

Khorsandi, M., St-Hilaire, A., Arsenault, R., Martel, J.-L., Larabi, S., Schnorbus, M., & Zwiers, F. (2023). Future flow and water temperature scenarios in an impounded drainage basin: Implications for summer flow and temperature management downstream of the dam. *Climatic Change*, *176*(12), 164. https://doi.org/10.1007/s10584-023-03634-w

Klotz, D., Herrnegger, M., & Schulz, K. (2017). Symbolic Regression for the Estimation of Transfer Functions of Hydrological Models. *Water Resources Research*, *53*(11), 9402–9423. https://doi.org/10.1002/2017WR021253

Kocik, J. F., Hayes, S. A., Carlson, S. M., & Cluer, B. (2022). A Resist-Accept-Direct (RAD) future for Salmon in Maine and California: Salmon at the southern edge. *Fisheries Management and Ecology*, *29*(4), 456–474. https://doi.org/10.1111/fme.12575

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, *23*(12), 5089–5110. https://doi.org/10.5194/hess-23-5089-2019

Kuczera, G., & Mroczkowski, M. (1998). Assessment of hydrologic parameter uncertainty and the worth of multiresponse data. *Water Resources Research*, *34*(6), 1481–1489. https://doi.org/10.1029/98WR00496

Kurylyk, B. L., MacQuarrie, K. T., Caissie, D., & McKenzie, J. M. (2015). Shallow groundwater thermal sensitivity to climate change and land cover disturbances: Derivation of analytical expressions and implications for stream temperature modeling. *Hydrology and Earth System Sciences*, *19*(5), 2469–2489.

Kurylyk, B. L., MacQuarrie, K. T., & McKenzie, J. M. (2014). Climate change impacts on groundwater and soil temperatures in cold and temperate regions: Implications, mathematical theory, and emerging simulation tools. *Earth-Science Reviews*, *138*, 313–334. https://doi.org/10.1016/j.earscirev.2014.06.006

Kwak, J., St-Hilaire, A., & Chebana, F. (2017). A comparative study for water temperature modelling in a small basin, the Fourchue River, Quebec, Canada. *Hydrological Sciences Journal*, *62*(1), 64–75. https://doi.org/10.1080/02626667.2016.1174334

Ladson, A. R., Brown, R., Neal, B., & Nathan, R. (2013). A standard approach to baseflow separation using the Lyne and Hollick filter. *Australasian Journal of Water Resources*, *17*(1), 25–34. https://doi.org/10.7158/13241583.2013.11465417

Laizé, C. L., Bruna Meredith, C., Dunbar, M. J., & Hannah, D. M. (2017). Climate and basin drivers of seasonal river water temperature dynamics. *Hydrology and Earth System Sciences*, *21*(6), 3231–3247. https://doi.org/10.5194/hess-21-3231-2017

Lange, H., & Sippel, S. (2020). Machine learning applications in hydrology. *Forest-Water Interactions*, 233–257.

Leach, J. A., Kelleher, C., Kurylyk, B. L., Moore, R. D., & Neilson, B. T. (2023). A primer on stream temperature processes. *Wiley Interdisciplinary Reviews: Water*, *10*(4), e1643.

Leach, J. A., Neilson, B. T., Buahin, C. A., Moore, R. D., & Laudon, H. (2021). Lake outflow and hillslope lateral inflows dictate thermal regimes of forested streams draining small lakes. *Water Resources Research*, *57*(6), e2020WR028136. https://doi.org/10.1029/2020WR028136

Leach, J., & Moore, R. (2010). Above-stream microclimate and stream surface energy exchanges in a wildfire-disturbed riparian zone. *Hydrological Processes*, *24*(17), 2369–2381. https://doi.org/10.1002/hyp.7639

Lindsay, J. (2014). *The whitebox geospatial analysis tools project and open-access GIS*. 16–18. https://jblindsay.github.io/ghrg/pubs/LindsayGISRUK2014.pdf

Loerke, E., Pohle, I., Wilkinson, M. E., Rivington, M., Wardell-Johnson, D., & Geris, J. (2023). Long-term daily stream temperature record for Scotland reveals spatio-temporal patterns in warming of rivers in the past and further warming in the future. *Science of the Total Environment*, *890*, 164194. https://doi.org/10.1016/j.scitotenv.2023.164194

Lookingbill, T. R., & Urban, D. L. (2003). Spatial estimation of air temperature differences for landscape-scale studies in montane environments. *Agricultural and Forest Meteorology*, *114*(3–4), 141–151. https://doi.org/10.1016/S0168-1923(02)00196-X

Lyne, V., & Hollick, M. (1979). *Stochastic time-variable rainfall-runoff modelling*. *79*(10), 89–93. https://d1wqtxts1xzle7.cloudfront.net/39814987/Stochastic_Time-Variable_Rainfall-Runoff20151108-28652-1m3nhta-libre.pdf?1447046893=&response-content-disposition=inline%3B+filename%3DStochastic_time_variable_rainfall_runoff.pdf&Expires=1719958762&Signature=gvVldATk4mEf9U6ikNS3DWJfpMvYGJXJiKbrIdoK-45x7dlVf2r3pVbiIQaBGYruDgOcSG2n~KLUCUUr-IN2XI1Q89jsQj8uHBGK8sm3oa5Kz492CqcuIfqGjFFRcFmnz6K86IPKf4vFHjT3OjrTLO5VfzEFSDJErrDpDOtPaPz07mwBSoyrPPQ~g3SnetGZjELnc9HdKXDSJElpolohajU3msf~S

VdzzUfVm-eCVci9GvXDVyFX5w7kmkpuN-ypMEM25jCWiD~j5Z1M0Z4gFI7ScziPyWbSbxMhWAF5zAUr2360XSIAAgB~JLt~LtJDczEaBSDovBMnSxW2GRZS4w__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Maheu, A., & Caissie, D. (2023). Spatial and temporal variability of the solar radiation heat flux in streams of a forested catchment. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, *48*(2), 206–221. https://doi.org/10.1080/07011784.2023.2167609

Maheu, A., Caissie, D., St-Hilaire, A., & El-Jabi, N. (2014). River evaporation and corresponding heat fluxes in forested catchments. *Hydrological Processes*, *28*(23), 5725–5738. https://doi.org/10.1002/hyp.10071

Maxwell, A. E., & Shobe, C. M. (2022). Land-surface parameters for spatial predictive mapping and modeling. *Earth-Science Reviews*, *226*, 103944. https://doi.org/10.1016/j.earscirev.2022.103944

McCutchan, M. H., & Fox, D. G. (1986). Effect of elevation and aspect on wind, temperature and humidity. *Journal of Applied Meteorology and Climatology*, *25*(12), 1996–2013. https://doi.org/10.1175/1520-0450(1986)025<1996:EOEAAO>2.0.CO;2

Mohseni, O., Stefan, H. G., & Erickson, T. R. (1998). A nonlinear regression model for weekly stream temperatures. *Water Resources Research*, *34*(10), 2685–2692. https://doi.org/10.1029/98WR01877

Monk, W. A., & Dugdale, S. J. (2023). River temperature: Perspectives, applications, and future directions for research in Canada. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, *48*(2), 67–71. https://doi.org/10.1080/07011784.2023.2222697

Moore, I. D., Grayson, R., & Ladson, A. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, *5*(1), 3–30. https://doi.org/10.1002/hyp.3360050103

Moore, R. D. (2006). Stream temperature patterns in British Columbia, Canada, based on routine spot measurements. *Canadian Water Resources Journal*, *31*(1), 41–56. https://doi.org/10.4296/cwrj3101041

Morin, G., & Couillard, D. (1990). Predicting river temperatures with a hydrological model. *Encyclopedia of Fluid Mechanic, Surface and Groundwater Flow Phenomena*, *10*, 171–209.

Morin, G., & Paquet, P. (1995). *Le modèle de simulation de quantité et de qualité CEQUEAU: Guide de l'utilisateur vesion 2.0 pour Windows.* https://espace.inrs.ca/id/eprint/1099/1/R000435.pdf

Morin, G., & Paquet, P. (2007). *Modèle hydrologique CEQUEAU.* https://espace.inrs.ca/id/eprint/1098/1/R000926.pdf

Nachtergaele, F., van Velthuizen, H., Verelst, L., Wiberg, D., Henry, M., Chiozza, F., Yigini, Y., Aksoy, E., Batjes, N., & Boateng, E. (2023). *Harmonized world soil database version 2.0*. Food and Agriculture Organization of the United Nations. https://openknowledge.fao.org/server/api/core/bitstreams/149f1562-bf6a-439f-9d3a-eb93940f39cf/content

Nasa, J. (2013). NASA shuttle radar topography mission global 1 arc second. *NASA EOSDIS Land Processes DAAC*, *10*.

O'Sullivan, A. M., Devito, K. J., & Curry, R. A. (2019). The influence of landscape characteristics on the spatial variability of river temperatures. *Catena*, *177*, 70–83. https://doi.org/10.1016/j.catena.2019.02.006

O'Sullivan, A. M., Devito, K. J., Ogilvie, J., Linnansaari, T., Pronk, T., Allard, S., & Curry, R. A. (2020). Effects of Topographic Resolution and Geologic Setting on Spatial Statistical River Temperature Models. *Water Resources Research*, *56*(12), e2020WR028122. https://doi.org/10.1029/2020WR028122

Ouali, D., Chebana, F., & Ouarda, T. B. M. J. (2016). Non-linear canonical correlation analysis in regional frequency analysis. *Stochastic Environmental Research and Risk Assessment*, *30*(2), 449–462. https://doi.org/10.1007/s00477-015-1092-7

Ouarda, T. B., Charron, C., Hundecha, Y., St-Hilaire, A., & Chebana, F. (2018). Introduction of the GAM model for regional low-flow frequency analysis at ungauged basins and comparison with commonly used approaches. *Environmental Modelling & Software*, *109*, 256–271. https://doi.org/10.1016/j.envsoft.2018.08.031

Ouarda, T. B., Charron, C., & St-Hilaire, A. (2022). Regional estimation of river water temperature at ungauged locations. *Journal of Hydrology X*, *17*, 100133. https://doi.org/10.1016/j.hydroa.2022.100133

Ouarda, T. B. M. J., & Charron, C. (2018). Nonstationary Temperature-Duration-Frequency curves. *Scientific Reports*, *8*(1), 15493. https://doi.org/10.1038/s41598-018-33974-y

Ouarda, T. B. M. J., Girard, C., Cavadias, G. S., & Bobée, B. (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, *254*(1), 157–173. https://doi.org/10.1016/S0022-1694(01)00488-7

Ouarda, T. B. M. J., Masselot, P., Campagna, C., Gosselin, P., Lavigne, É., St-Hilaire, A., Chebana, F., & Valois, P. (2024). Prediction of heatwave related mortality magnitude, duration and frequency with climate variability and climate change information. *Stochastic Environmental Research and Risk Assessment*, *38*(11), 4471–4483. https://doi.org/10.1007/s00477-024-02813-0

Ouellet-Proulx, S., St-Hilaire, A., & Boucher, M. (2019). Implication of evaporative loss estimation methods in discharge and water temperature modelling in cool temperate climates. *Hydrological Processes*, *33*(22), 2867–2884. https://doi.org/10.1002/hyp.13534

Oyinlola, M. A., Khorsandi, M., Penman, R., Earhart, M. L., Arsenault, R., Brauner, C. J., & St-Hilaire, A. (2023). Hydrothermal impacts of water release on early life stages of white sturgeon in the Nechako river, BC Canada. *Journal of Thermal Biology*, *117*, 103682. https://doi.org/10.1016/j.jtherbio.2023.103682

Pagliero, L., Bouraoui, F., Diels, J., Willems, P., & McIntyre, N. (2019). Investigating regionalization techniques for large-scale hydrological modelling. *Journal of Hydrology*, *570*, 220–235. https://doi.org/10.1016/j.jhydrol.2018.12.071

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825–2830.

Quilty, J., & Adamowski, J. (2020). A stochastic wavelet-based data-driven framework for forecasting uncertain multiscale hydrological and water resources processes. *Environmental Modelling & Software*, *130*, 104718. https://doi.org/10.1016/j.envsoft.2020.104718

Rahmani, F., Appling, A., Feng, D., Lawson, K., & Shen, C. (2023). Identifying structural priors in a hybrid differentiable model for stream water temperature modeling. *Water Resources Research*, *59*(12), e2023WR034420. https://doi.org/10.1029/2023WR034420

Rahmani, F., Lawson, K., Ouyang, W., Appling, A., Oliver, S., & Shen, C. (2021). Exploring the exceptional performance of a deep learning stream temperature model and the value of streamflow data. *Environmental Research Letters*, *16*(2), 024025.

Rahmati, N. (2023). *Modélisation et analyse hydrologique et thermique d'un cours d'eau aménagé: La rivière Tobique (Nouveau-Brunswick, Canada).* https://espace.inrs.ca/id/eprint/13787/

Railsback, S. F., & Harvey, B. C. (2023). Can thermal refuges save salmonids? Simulation of cold-pool benefits to salmonid populations. *Transactions of the American Fisheries Society*, *152*(4), 383–396. https://doi.org/10.1002/tafs.10411

Rakovec, O., Kumar, R., Attinger, S., & Samaniego, L. (2016). Improving the realism of hydrologic model functioning through multivariate parameter estimation. *Water Resources Research*, *52*(10), 7779–7792. https://doi.org/10.1002/2016WR019430

Raymondi, R. R., Cuhaciyan, J. E., Glick, P., Capalbo, S. M., Houston, L. L., Shafer, S. L., & Grah, O. (2013). Water resources: Implications of changes in temperature and precipitation. *Climate Change in the Northwest: Implications for Our Landscapes, Waters, and Communities*, 41–66. https://doi.org/10.5822/978-1-61091-512-0_3

Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, *52*(1), 423–439.

Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, *52*(1), 440–455. https://doi.org/10.1002/2015WR017559

Razavi, S., Sheikholeslami, R., Gupta, H. V., & Haghnegahdar, A. (2019). VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. *Environmental Modelling & Software*, *112*, 95–107.

Rincón, E., St-hilaire, A., Bergeron, N. E., & Dugdale, S. J. (2023). Combining Landsat TIR-imagery data and ERA5 reanalysis information with different calibration strategies to improve simulations of streamflow and river temperature in the Canadian Subarctic. *Hydrological Processes*, *37*(10), e15008. https://doi.org/10.1002/hyp.15008

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). *Global sensitivity analysis: The primer*. John Wiley & Sons.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5). https://doi.org/10.1029/2008WR007327

Sepaskhah, A., & Boersma, L. (1979). Thermal conductivity of soils as a function of temperature and water content. *Soil Science Society of America Journal*, *43*(3), 439–444. https://doi.org/10.2136/sssaj1979.03615995004300030003x

Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., & Li, L. (2023). Differentiable modelling to unify machine learning and physical models for geosciences. *Nature Reviews Earth & Environment*, *4*(8), 552–567. https://doi.org/10.1038/s43017-023-00450-9

Shen, H., Tolson, B. A., & Mai, J. (2022). Time to update the split-sample approach in hydrological model calibration. *Water Resources Research*, *58*(3), e2021WR031523. https://doi.org/10.1029/2021WR031523

Shrestha, R. R., Pesklevits, J. C., Bonsal, B. R., Brannen, R., Guo, T., & Hoffman, S. (2024). Rising summer river water temperature across Canada: Spatial patterns and hydroclimatic controls. *Environmental Research Letters*, *19*(4), 044058. https://doi.org/10.1088/1748-9326/ad365f

Siegel, J. E., Fullerton, A. H., FitzGerald, A. M., Holzer, D., & Jordan, C. E. (2023). Daily stream temperature predictions for free-flowing streams in the Pacific Northwest, USA. *PloS Water*, *2*(8), e0000119. https://doi.org/10.1371/journal.pwat.0000119

Sinokrot, B. A., & Gulliver, J. S. (2000). In-stream flow impact on river water temperatures. *Journal of Hydraulic Research*, *38*(5), 339–349. https://doi.org/10.1080/00221680009498315

Sinokrot, B. A., & Stefan, H. G. (1993). Stream temperature dynamics: Measurements and modeling. *Water Resources Research*, *29*(7), 2299–2312. https://doi.org/10.1029/93WR00540

Song, Z., Xia, J., Wang, G., She, D., Hu, C., & Hong, S. (2022). Regionalization of hydrological model parameters using gradient boosting machine. *Hydrology and Earth System Sciences*, *26*(2), 505–524. https://doi.org/10.5194/hess-26-505-2022

Souaissi, Z., Ouarda, T. B., & St-Hilaire, A. (2021). River water temperature quantiles as thermal stress indicators: Case study in Switzerland. *Ecological Indicators*, *131*, 108234. https://doi.org/10.1016/j.ecolind.2021.108234

Souaissi, Z., Ouarda, T. B., & St-Hilaire, A. (2023a). Non-parametric, semi-parametric, and machine learning models for river temperature frequency analysis at ungauged basins. *Ecological Informatics*, *75*, 102107.

Souaissi, Z., Ouarda, T. B., & St-Hilaire, A. (2023b). Regional thermal index model for river temperature frequency analysis in ungauged basins. *Environmental Modelling & Software*, *164*, 105709. https://doi.org/10.1016/j.envsoft.2023.105709

Souaissi, Z., Ouarda, T. B., St-Hilaire, A., & Ouali, D. (2023). *Regional frequency analysis of stream temperature at ungauged sites using non-linear canonical correlation analysis and generalized additive models*.

Spicer, P., Matte, P., Huguenard, K., & Rickard, L. N. (2021). Coastal windstorms create unsteady, unpredictable storm surges in a fluvial Maine estuary. *Shore & Beach*, *89*(2), 3. https://doi.org/10.34237/1008921

St-Hilaire, A., Boucher, M.-A., Chebana, F., Ouellet-Proulx, S., Zhou, Q. X., Larabi, S., Dugdale, S., & Latraverse, M. (2015). *Breathing a new life to an older model: The CEQUEAU tool for flow and water temperature simulations and forecasting*. Proceedings of the 22nd Canadian Hydrotechnical Conference.

St-Hilaire, A., Boyer, C., Bergeron, N., & Daigle, A. (2018). Water temperature monitoring in Eastern Canada: A case study for network optimization. *WIT Transactions on Ecology and the Environment*, *228*, 269–275. https://doi.org/10.2495/WP180251

St-Hilaire, A., Caissie, D., Bergeron, N. E., Ouarda, T. B., & Boyer, C. (2021). Climate change and extreme river temperature. In *Climate Change and Extreme Events* (pp. 25–37). Elsevier. https://doi.org/10.1016/B978-0-12-822700-8.00011-1

St-Hilaire, A., Morin, G., El-Jabi, N., & Caissie, D. (2000). Water temperature modelling in a small forested stream: Implication of forest canopy and soil temperature. *Canadian Journal of Civil Engineering*, *27*(6), 1095–1108. https://doi.org/10.1139/l00-021

St-Hilaire, A., Ouarda, T., Charron, C., Boyer, C., Daigle, A., & Bergeron, N. (2019). *Regional analysis of water temperature for the estimation of thermal indices at ungaged sites*. AGU Fall Meeting 2019.

St-Hilaire, A., Oyinlola, M. A., Rincón, E., & Ferchichi, H. (2023). Six decades of research on river temperature in Canada. *Canadian Water Resources Journal/Revue Canadienne Des Ressources Hydriques*, *48*(4), 450–474. https://doi.org/10.1080/07011784.2023.2283241

Story, A., Moore, R., & Macdonald, J. (2003). Stream temperatures in two shaded reaches below cutblocks and logging roads: Downstream cooling linked to subsurface hydrology. *Canadian Journal of Forest Research*, *33*(8), 1383–1396. https://doi.org/10.1139/x03-087

Tetens, O. (1930). Uber einige meteorologische Begriffe. *Z. Geophys*, *6*, 297–309.

Teutschbein, C., Grabs, T., Laudon, H., Karlsen, R. H., & Bishop, K. (2018). Simulating streamflow in ungauged basins under a changing climate: The importance of landscape characteristics. *Journal of Hydrology*, *561*, 160–178. https://doi.org/10.1016/j.jhydrol.2018.03.060

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tovar-Pescador, J., Pozo-Vázquez, D., Ruiz-Arias, J., Batlles, J., López, G., & Bosch, J. (2006). On the use of the digital elevation model to estimate the solar radiation in areas of complex topography. *Meteorological Applications*, *13*(3), 279–287. https://doi.org/10.1017/S1350482706002258

Townsend, D. W., Pettigrew, N. R., Thomas, M. A., & Moore, S. (2023). Warming waters of the Gulf of Maine: The role of shelf, slope and Gulf Stream water masses. *Progress in Oceanography*, *215*, 103030. https://doi.org/10.1016/j.pocean.2023.103030

Tsai, W.-P., Feng, D., Pan, M., Beck, H., Lawson, K., Yang, Y., Liu, J., & Shen, C. (2021). From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. *Nature Communications*, *12*(1), 5988. https://doi.org/10.1038/s41467-021-26107-z

Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media.

Varadharajan, C., Appling, A. P., Arora, B., Christianson, D. S., Hendrix, V. C., Kumar, V., Lima, A. R., Müller, J., Oliver, S., & Ombadi, M. (2022). Can machine learning accelerate process understanding and decision-relevant predictions of river water quality? *Hydrological Processes*, *36*(4), e14565. https://doi.org/10.1002/hyp.14565

Wade, J., Kelleher, C., & Hannah, D. M. (2023). Machine learning unravels controls on river water temperature regime dynamics. *Journal of Hydrology*, *623*, 129821. https://doi.org/10.1016/j.jhydrol.2023.129821

Wade, J., Kelleher, C., & Kurylyk, B. L. (2024). Incorporating physically-based water temperature predictions into the National water model framework. *Environmental Modelling & Software*, *171*, 105866. https://doi.org/Incorporating physically-based water temperature predictions into the National water model framework

Wanders, N., van Vliet, M. T., Wada, Y., Bierkens, M. F., & van Beek, L. P. (2019). High-resolution global water temperature modeling. *Water Resources Research*, *55*(4), 2760–2778. https://doi.org/10.1029/2018WR023250

Webb, B. W., Hannah, D. M., Moore, R. D., Brown, L. E., & Nobilis, F. (2008). Recent advances in stream and river temperature research. *Hydrological Processes: An International Journal*, *22*(7), 902–918. https://doi.org/10.1002/hyp.6994

Weierbach, H., Lima, A. R., Willard, J. D., Hendrix, V. C., Christianson, D. S., Lubich, M., & Varadharajan, C. (2022). Stream Temperature Predictions for River Basin Management in the Pacific Northwest and Mid-Atlantic Regions Using Machine Learning. *Water*, *14*(7), 1032. https://doi.org/10.3390/w14071032

Weller, J. D., Moore, R. D. (Dan), & Iacarella, J. C. (2024). Stream thermalscape scenarios for British Columbia, Canada. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *49*(2), 233–252. https://doi.org/10.1080/07011784.2023.2267028

Wilson, J. P. (2018). *Environmental applications of digital terrain modeling*. John Wiley & Sons.

Wuebbles, D., Fahey, D., Takle, E., Hibbard, K., Arnold, J., DeAngelo, B., Doherty, S., Easterling, D., Edmonds, J., & Edmonds, T. (2017). *Climate science special report: Fourth national climate assessment (NCA4), Volume I*.

Yamada, M., Jitkrittum, W., Sigal, L., Xing, E. P., & Sugiyama, M. (2014). High-dimensional feature selection by feature-wise kernelized lasso. *Neural Computation*, *26*(1), 185–207. https://doi.org/10.1162/NECO_a_00537

Yang, X., Magnusson, J., Rizzi, J., & Xu, C.-Y. (2018). Runoff prediction in ungauged catchments in Norway: Comparison of regionalization approaches. *Hydrology Research*, *49*(2), 487–505. https://doi.org/10.2166/nh.2017.071

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00527.x

11- Captions (Figures, tables, and equations)

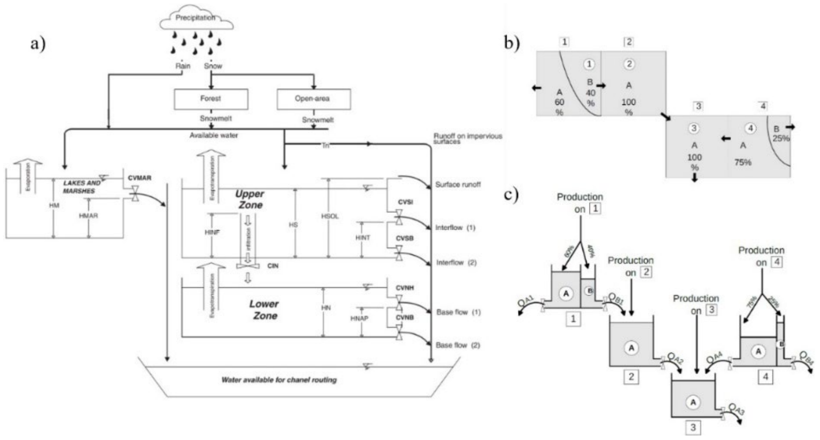

Figure S1: Schematic of CEQUEAU's production and transfer functions. a) Production function showing the different reservoirs in the CEQUEAU model. b) Plan view of two CEs with their respective CPs (the thick black arrows represent the flow direction). c) Transfer scheme between CEs (percentage indicates the amount of water drained into each CP). Adapted from Morin and Paquet *(Morin & Paquet, 2007)*

$$Q_{t_n} = P_{t_n} - ETP_{t_n} + \left(HU_{t_n} - HU_{t_{n-1}}\right) + \left(HL_{t_n} - HL_{t_{n-1}}\right) \qquad (S1)$$

$$Q_{sw_t} = CRAYSO * A_{CP} * Q_{in} \qquad (S2)$$

$$Q_{lw_t} = CRAYIN * A_{CP} * \sigma * (\beta T_{air}^4 - T_w^4) \qquad (S3)$$

$$Q_{e_t} = CEVAPO * A_{CP} * l_{et} * E \qquad (S4)$$

$$Q_{s_t} = CCONV * A_{CP} * [0.2 * W * (T_{air} - T_w)] \qquad (S5)$$

Parameter name	Description	LB	UB
Hydrological model			
CIN	Coefficient of percolation from the upper zone to the lower zone	0.05	0.8
CVMAR	Lakes and marshes drainage coefficient	0.01	1
CVNB	Lower-zone lower drainage coefficient	0.001	0.2
CVNH	Lower-zone upper drainage coefficient	0.001	0.2
CVSB	Upper-zone lower drainage coefficient	0	0.5
CVSI	Upper-zone intermediate drainage coefficient	0	0.5
XINFMA	Daily maximum infiltration	0.001	40
HINF	Threshold of percolation from the upper to the lower zone	10	100
HINT	Upper-zone intermediate drainage threshold	10	200
HMAR	Lakes and marshes drainage threshold	100	500
HNAP	Lower-zone upper threshold	20	1000
HPOT	Threshold of evaporation at the potential rate	0.01	200
HSOL	Upper-zone runoff threshold	100	300
HRIMP	Upper-zone runoff threshold for impermeable surfaces	0	10
EXXKT	Routing coefficient fitting parameter	0	0.8
Snowmelt model			
STRNE	Snow-rain temperature threshold	-2	3
TFC	Potential melting rate in forest	0.1	10
TFD	Potential melting rate in open (no canopy) areas	0.1	10
TSC	Minimum temperature threshold to initiate snowmelt in forest	-4.5	3
TSD	Minimum temperature threshold to initiate snowmelt in open areas	-2	5
TTD	Heat deficit coefficient	0.1	3
TTS	Minimum temperature for snow stock ripening	-5	2
Evapotranspiration model			
EVNAP	Fraction of evapotranspiration taken for the lower reservoir	0.1	1
XAA	Thornthwaite exponent	0.2	9
XIT	Thornthwaite Index	5	50

Table S1: CEQUEAU hydrological model parameters with respective lower and upper bounds

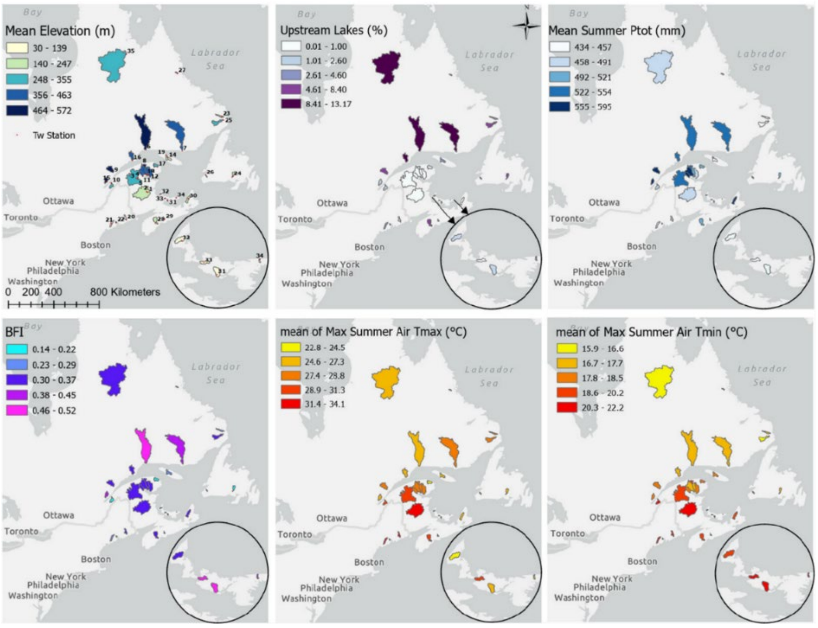

Figure S2: Watersheds environmental conditions

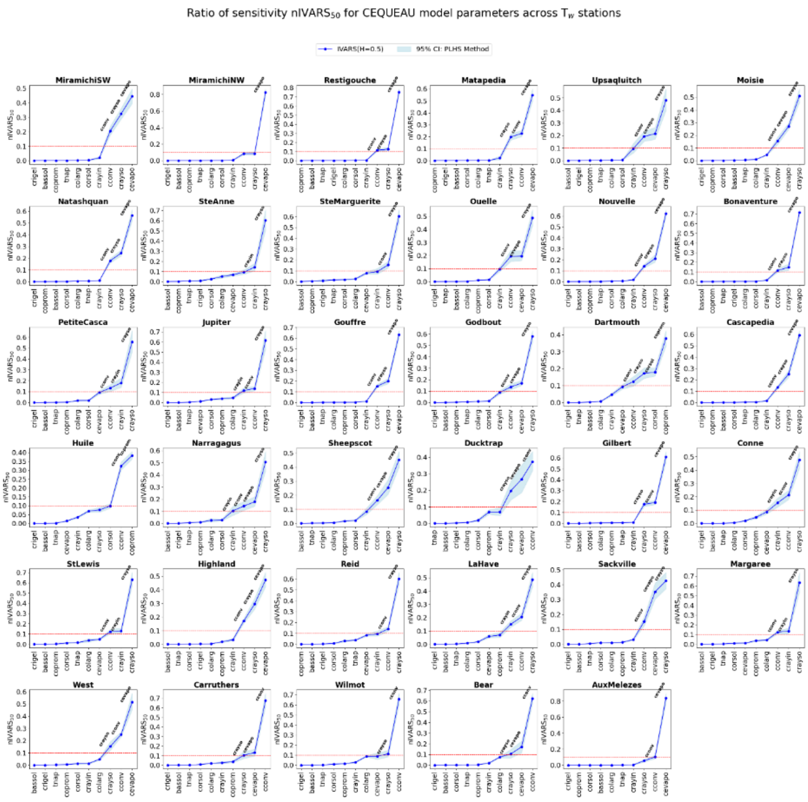

Figure S3: Ratio of sensitivity (n$IVARS_{50}$). Each parameter's sensitivity is calculated as the ratio of its respective sensitivity to the sum of the sensitivity indices of all model parameters (the sensitivity ratios sum to one).

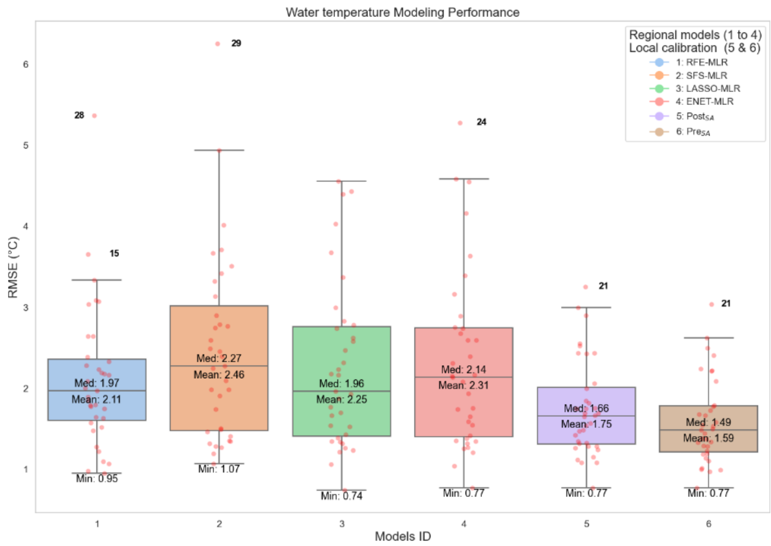

Figure S4: Water temperature model performance using MLR as the regional

Watershed	Province	COPROM	COLARG	CRAYSO	CRAYIN	CEVAPO	CCONV	CRIGEL	TNAP	BASSOL	CORSOL
1	NB	2.00	2.00	0.98	0.34	0.10	1.84	206.58	8.00	12.10	0.10
2	NB	2.00	2.00	1.15	0.79	0.10	1.22	198.80	8.00	12.49	0.10
3	NB	2.00	2.00	0.65	0.10	0.14	1.20	249.38	5.30	10.28	0.11
4	NB	2.00	2.00	0.48	0.38	0.10	1.16	178.51	6.35	10.98	0.10
5	QC	2.00	2.00	0.58	0.10	0.31	1.11	214.31	6.60	6.70	0.10
6	QC	2.00	2.00	1.33	0.10	0.10	0.71	125.73	8.00	11.08	0.10
7	QC	2.00	2.00	1.82	0.19	0.10	1.84	123.16	8.00	5.46	0.10
8	QC	2.00	2.00	0.41	0.10	0.10	0.76	250.00	6.97	7.39	0.10
9	QC	2.00	2.00	1.51	0.94	0.10	0.98	247.82	8.00	10.85	0.10
10	QC	2.00	2.00	0.63	0.40	0.25	1.44	83.75	8.00	5.11	0.10
11	QC	2.00	2.00	0.26	0.10	0.10	0.27	96.02	8.00	15.00	0.10
12	QC	2.00	2.00	0.24	0.14	0.25	0.55	248.04	6.22	14.40	0.10
13	QC	2.00	2.00	0.44	0.10	0.44	0.45	101.05	8.00	11.52	0.10
14	QC	2.00	2.00	0.36	0.93	0.10	0.26	175.48	8.00	5.90	0.10
15	QC	2.00	2.00	0.93	0.94	0.12	1.07	109.42	8.00	12.07	0.44
16	QC	2.00	2.00	1.28	0.28	0.10	1.38	75.46	8.00	12.42	0.10
17	QC	2.00	2.00	0.45	0.39	0.10	0.57	230.36	7.05	7.58	0.10
18	QC	2.00	2.00	0.59	0.10	0.10	0.98	162.64	6.97	12.63	0.17
19	QC	2.00	2.00	0.50	0.84	0.10	0.54	57.39	8.00	14.85	0.10
35	QC	2.00	2.00	2.00	0.11	0.10	1.61	64.6	4.0	13.93	0.10

20	MA	2.00	2.00	0.87	0.89	0.10	1.16	63.12	8.00	14.56	0.10
21	MA	2.00	2.00	0.77	0.85	0.17	0.91	216.97	8.00	10.37	0.10
22	MA	2.00	2.00	0.36	0.72	0.10	0.19	120.77	8.00	14.95	0.10
23	NL	2.00	2.00	1.21	0.85	0.10	0.72	214.08	8.00	11.25	0.10
24	NL	2.00	2.00	0.72	0.38	0.10	1.26	202.40	8.00	12.67	0.10
25	NL	2.00	2.00	1.85	1.63	0.10	1.02	226.34	8.00	12.29	0.10
26	NL	2.00	2.00	0.78	0.49	0.19	0.79	250.00	8.00	10.47	0.10
27	NL	2.00	2.00	1.01	1.25	0.10	0.27	196.20	8.00	7.62	0.10
28	NS	2.00	2.00	0.73	0.10	0.10	1.47	246.35	8.00	14.18	0.10
29	NS	2.00	2.00	0.25	0.10	0.10	0.40	76.46	8.00	11.09	0.10
30	NS	2.00	2.00	0.46	1.07	0.10	0.28	249.84	8.00	10.56	0.10
31	PEI	2.00	2.00	0.45	0.90	0.10	0.24	215.62	7.93	6.99	0.10
32	PEI	2.00	2.00	0.28	0.49	0.11	0.13	63.59	8.00	5.41	0.10
33	PEI	2.00	2.00	0.10	0.10	0.14	0.21	80.59	7.29	9.55	0.10
34	PEI	2.00	2.00	0.37	0.68	0.10	0.17	244.64	6.05	9.11	0.10

Table S2: $Post_{SA}$ water temperature model parameters

Watershed	Province	COPROM	COLARG	CRAYSO	CRAYIN	CEVAPO	CCONV	CRIGEL	TNAP	BASSOL	CORSOL
1	NB	2.00	2.00	0.88	0.34	0.10	1.69	206.58	8.0	12.10	0.10
2	NB	2.00	2.00	1.04	0.79	0.10	1.08	198.80	8.0	12.49	0.10
3	NB	2.00	2.00	0.78	0.10	0.14	1.27	249.38	5.3	10.28	0.11
4	NB	2.00	2.00	0.44	0.38	0.10	1.03	214.31	6.6	6.70	0.10
5	QC	2.00	2.00	0.46	0.10	0.31	0.99	178.51	6.3	10.98	0.10
6	QC	2.00	2.00	1.03	0.10	0.10	0.69	125.73	8.0	11.08	0.10
7	QC	2.00	2.00	1.47	0.19	0.10	1.66	123.16	8.0	5.46	0.10
8	QC	2.00	2.00	0.33	0.10	0.10	0.57	250.00	6.9	7.39	0.10
9	QC	2.00	2.00	1.04	0.94	0.10	0.87	247.82	8.0	10.85	0.10
10	QC	2.00	2.00	0.56	0.40	0.25	1.28	83.75	8.0	5.11	0.10
11	QC	2.00	2.00	0.25	0.10	0.10	0.24	96.02	8.0	15.00	0.10
12	QC	2.00	2.00	0.42	0.14	0.25	0.66	248.04	6.2	14.40	0.10
13	QC	2.00	2.00	0.43	0.10	0.44	0.49	101.05	8.0	11.52	0.10

14	QC	2.00	2.00	0.27	0.93	0.10	0.26	175.48	8.0	5.90	0.10
15	QC	2.00	2.00	0.60	0.94	0.12	0.94	109.42	8.0	12.07	0.44
16	QC	2.00	2.00	0.92	0.28	0.10	1.19	75.46	8.0	12.42	0.10
17	QC	2.00	2.00	0.37	0.39	0.10	0.46	230.36	7.0	7.58	0.10
18	QC	2.00	2.00	0.58	0.10	0.10	0.98	162.64	6.9	12.63	0.17
19	QC	2.00	2.00	0.39	0.84	0.10	0.43	57.39	8.0	14.85	0.10
35	QC	2.00	2.00	1.99	0.11	0.10	1.62	64.60	4.0	13.93	0.10
20	MA	2.00	2.00	0.66	0.89	0.10	0.99	63.12	8.0	14.56	0.10
21	MA	2.00	2.00	0.59	0.85	0.17	0.66	216.97	8.0	10.37	0.10
22	MA	2.00	2.00	0.35	0.72	0.10	0.20	120.77	8.0	14.95	0.10
23	NL	2.00	2.00	0.91	0.85	0.10	0.63	214.08	8.0	11.25	0.10
24	NL	2.00	2.00	0.62	0.38	0.10	1.13	202.40	8.0	12.67	0.10
25	NL	2.00	2.00	1.43	1.63	0.10	0.91	226.34	8.0	12.29	0.10
26	NL	2.00	2.00	0.59	0.49	0.19	0.65	250.00	8.0	10.47	0.10
27	NL	2.00	2.00	0.74	1.25	0.10	0.29	196.20	8.0	7.62	0.10
28	NS	2.00	2.00	0.60	0.10	0.10	1.26	246.35	8.0	14.18	0.10
29	NS	2.00	2.00	0.22	0.10	0.10	0.49	76.46	8.0	11.09	0.10
30	NS	2.00	2.00	0.37	1.07	0.10	0.24	249.84	8.0	10.56	0.10
31	PEI	2.00	2.00	0.30	0.90	0.10	0.18	215.62	7.9	6.99	0.10
32	PEI	2.00	2.00	0.23	0.49	0.11	0.11	63.59	8.0	5.41	0.10
33	PEI	2.00	2.00	0.10	0.10	0.14	0.14	80.59	7.3	9.55	0.10
34	PEI	2.00	2.00	0.25	0.68	0.10	0.19	244.64	6.0	9.11	0.10

1396 Table S3: Pre_{SA} water temperature model parameters