

Earth's Future

RESEARCH ARTICLE

10.1029/2024EF004815

Key Points:

- Thermal Infrared remote sensing from Landsat was used to estimate the surface water temperature of the highly regulated Columbia River Basin
- Using 42 years of data, the surface water temperature could be predicted on average with 0.7°C of absolute error regardless of the dam's potential thermal influence
- The reconstructed hydro-thermal behavior indicated a long-term downstream warming trend along the Columbia River

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

F. Hossain,
fhossain@uw.edu

Citation:

Darkwah, G. K., Hossain, F., Tchervenski, V., Holtgrieve, G., Graves, D., Seaton, C., et al. (2024). Reconstruction of the hydro-thermal behavior of regulated river networks of the Columbia River Basin using satellite remote sensing and data-driven techniques. *Earth's Future*, 12, e2024EF004815. <https://doi.org/10.1029/2024EF004815>

Received 24 APR 2024

Accepted 22 AUG 2024

Author Contributions:

Conceptualization: G. K. Darkwah, Faisal Hossain

Data curation: G. K. Darkwah, Victoria Tchervenski, Pritam Das

Formal analysis: G. K. Darkwah, Victoria Tchervenski

Funding acquisition: Gordon Holtgrieve, Sanchit Minocha

Investigation: G. K. Darkwah, Faisal Hossain, Victoria Tchervenski, David Graves, Charles Seaton

© 2024, The Author(s).

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Reconstruction of the Hydro-Thermal Behavior of Regulated River Networks of the Columbia River Basin Using Satellite Remote Sensing and Data-Driven Techniques

G. K. Darkwah¹ , Faisal Hossain¹ , Victoria Tchervenski², Gordon Holtgrieve³ , David Graves⁴, Charles Seaton⁴, Sanchit Minocha¹ , Pritam Das¹ , Shahzaib Khan¹ , and Sarath Suresh¹ 

¹Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA, ²Innovation Lab High School, Bothell, WA, USA, ³School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA,

⁴Watershed Department, Columbia River Inter-Tribal Fish Commission, Portland, OR, USA

Abstract The use of satellite-based thermal infrared remote sensing has facilitated the assessment of surface water temperature on a large scale. However, the inherent limitations of this remote sensing technique make it difficult to assess rivers unless ambient conditions are cloud-free, devoid of steep terrain and the rivers are at least 60 m wide. To address these challenges that limit the spatiotemporal continuity of satellite-based hydro-thermal data, we harnessed the extensive coverage from the Landsat missions' thermal infrared sensors and data-driven techniques to estimate surface water temperature of rivers. Out of the tested data-driven techniques, we selected the Random Forest Regressor as our prime non-linear approach for estimation of surface water temperature in rivers. Using the selected technique, proposed as THORR (Thermal History of Regulated Rivers), we successfully reconstructed a multi-decadal, continuous spatiotemporal surface water temperature record for regulated rivers in the Columbia River Basin. Using 42 years of data, the surface water temperature could be predicted on average with 0.71°C of absolute error regardless of the dam's potential thermal influence in the downstream reaches. The reconstructed hydro-thermal behavior generated from THORR revealed a long-term downstream warming trend along the Columbia River. The open-source THORR tool can be extended to any river system around the world that is not gauged with in-situ temperature measurements for the reconstruction of hydro-thermal behavior.

Plain Language Summary Surface water temperature of rivers, which is an important ecological parameter for more robust water management, can be estimated in a cost-effective and globally scalable way from satellite thermal infrared (TIR) remote sensing. However, such a technique is limited to rivers of sufficient width and cloud-free conditions. In this study, data-driven techniques were explored to overcome this limitation and reconstruct surface water temperature of rivers in the continuum of space and time. By using a comprehensive record spanning 42 years of satellite remote sensing, we demonstrate that it is possible to recreate robust estimates of long spatiotemporal trends of river temperature to understand how surface waters are being altered thermally due to water management and climate change.

1. Introduction

Water temperature is an important variable in riverine ecosystems that provides insight into water quality and ecological dynamics. The thermal regime of rivers is most often dynamic with significant variation at daily, monthly, seasonal, and yearly timescales. Human alteration of land and riverscapes can alter this natural variation in riverine water temperature through a myriad of processes. Changes in stream temperature affect aquatic biota in diverse ways. For example, even brief periods above or below thermal tolerance limits can be lethal, while more subtle long-term changes can affect the growth, population, and distribution of aquatic organisms (Caissie, 2006; Halverson et al., 2022). An increase in temperature can increase the nutrient load in streams which can lead to algal bloom (Duan & Kaushal, 2013). Harvey et al. (2011) also showed that water temperature can affect dissolved oxygen, which directly affects aquatic life.

Topography, atmospheric conditions, streambed, and stream discharge are some of the major influences on the temperature of rivers (Caissie, 2006). Although these factors can vary naturally, anthropogenic factors have disrupted or accelerated the impact of these drivers on stream water temperature, especially by modifying stream discharge and the heat fluxes flowing through the river network (Hester & Doyle, 2011). River regulation by dam

Methodology: G. K. Darkwah,
Faisal Hossain
Project administration: Faisal Hossain
Resources: G. K. Darkwah,
Victoria Tchervenski, Sanchit Minocha,
Pritam Das, Shahzaib Khan, Sarah Suresh
Software: G. K. Darkwah,
Sanchit Minocha, Pritam Das,
Shahzaib Khan, Sarah Suresh
Supervision: Faisal Hossain,
Gordon Holtgrieve
Validation: G. K. Darkwah
Visualization: G. K. Darkwah
Writing – original draft: G. K. Darkwah
Writing – review & editing:
G. K. Darkwah, Faisal Hossain,
Victoria Tchervenski, Gordon Holtgrieve,
David Graves, Charles Seaton,
Sanchit Minocha, Pritam Das,
Shahzaib Khan, Sarah Suresh

operation, either for water supply, flood management, or hydropower generation, is a notable example of the anthropogenic factors that affect streamflow and water temperature patterns with implications on aquatic ecosystems. Hydropower dams in particular tend to have deep reservoirs that can be thermally stratified in warmer seasons and sometimes in colder seasons. Discharging water through the penstocks of large hydropower dams during thermal stratification in the warmer often leads to both cooling and stabilization of downstream temperatures (Ahmad et al., 2021; Bonnema et al., 2020; Olden & Naiman, 2010; Wang et al., 2023). The converse warming of downstream water temperature may also happen during colder seasons when thermal layers turn over with colder water in the upper layer. Hereafter, we will use the terms ‘dam’ and ‘reservoir’ interchangeably to mean the same entity—the dam-reservoir system.

Human activities, coupled with the impact of climate change, have disrupted the natural variability of water temperature. Therefore, it is important to understand the hydro-thermal behavior of river networks because water temperature plays a role in maintaining or transforming aquatic ecosystems. Here, we define hydro-thermal behavior as the long-term overview of past temperature variations and trends in river networks. For example, the long-term assessment of the impact of changes in water temperature reported by Halverson et al. (2022) on native and invasive fish species of the San Francisco Estuary portrays the importance of hydro-thermal behavior.

Inland, historical hydro-thermal studies are limited by the availability and spatiotemporal coverage of in-situ water temperature measurements (Ouellet et al., 2020; Worrall et al., 2022). For example, the Snake River of the Columbia River Basin (USA) has a 285-km reach that is unmonitored by temperature gauges even though this reach has two dams that are a major driver of the hydro-thermal regime of the river (Figure 1). However, remote sensing of water temperature offers broader spatial coverage than direct point-based measurements of in-situ temperature probes. Thermal infrared (TIR) and microwave (MW) sensors aboard satellites can record the surface temperature of water bodies from space. The use of satellite remote sensing is well-established in oceanography for generating sea-surface temperature (Gaube et al., 2019; Handcock et al., 2012; Minnett et al., 2019). For terrestrial studies, MW sensing does not offer the spatial resolution that is required to estimate the temperature of rivers even though it is not impacted by cloud cover. This is because MW sensors have a very large km-scale spatial resolution that is inadequate for observing most rivers. Meanwhile, because of the spatial resolution of satellite TIR temperature products, satellite remote sensing of water temperature has focused more on lakes, reservoirs, and large rivers with widths greater than 60 m (Handcock et al., 2006; Lalot et al., 2015; Wawrzyniak et al., 2012). For example, Ling et al. (2017) used TIR observations from the Landsat 7 ETM + to assess the downstream thermal pollution caused by dams along the Qingjiang River, a tributary to the Yangtze River. The widths of the section of the Qingjiang River studied by Ling et al. (2017) ranged from 200–500 m. In a similar study on the Yangtze River, Shi et al. (2010) used the TIR observations on Landsat 5 TM and Landsat 7 ETM + to assess the thermal regime under dam influence for river widths ranging from 600–850 m. Narrow river widths remain a challenge with the use of satellite TIR data. However, this is the case for many dams: the upstream reservoir pool is spatially large enough to be monitored with satellite remote sensing, but the reach downstream can be narrower than the satellite TIR’s footprint due to the choking of the river by the dam. Cloud cover also masks out portions of satellite observations, leading to spatiotemporal gaps in water surface records. Hence, the use of airborne TIR is sometimes preferred for narrow streams. This is because the aircraft are flown below clouds and the nearness of the aircraft to the river body provides the advantage of observations with high spatial resolution. Mapping the surface temperature of rivers with airborne TIR sensing also makes it possible to longitudinally follow the meandering transects of rivers, especially when the sensor is mounted on a helicopter. However, the use of airborne TIR is expensive, short-term, and limited in spatial coverage (Torgersen et al., 2001).

Therefore, the overarching question we ask in this study is as follows: *Given the non-exhaustive set of potential drivers and geophysical factors that influence stream temperature, particularly those related to reservoir operations, is it possible to reconstruct the hydro-thermal behavior of regulated rivers in a spatiotemporal continuum?* To estimate river water temperature, a non-exhaustive list of potential drivers and geophysical ‘influencers’ are the reservoir’s dynamic state (surface area, storage change, elevation), land use/land cover, and land temperature conditions adjacent to river reaches, season, and climate.

The limited availability of in-situ data poses a challenge for extensive spatial studies. In areas where it is difficult to acquire in-situ data, satellite remote sensing provides a convenient and scalable way to estimate factors that affect river water temperature. For instance, satellite remote sensing can be used to determine reservoirs’ surface

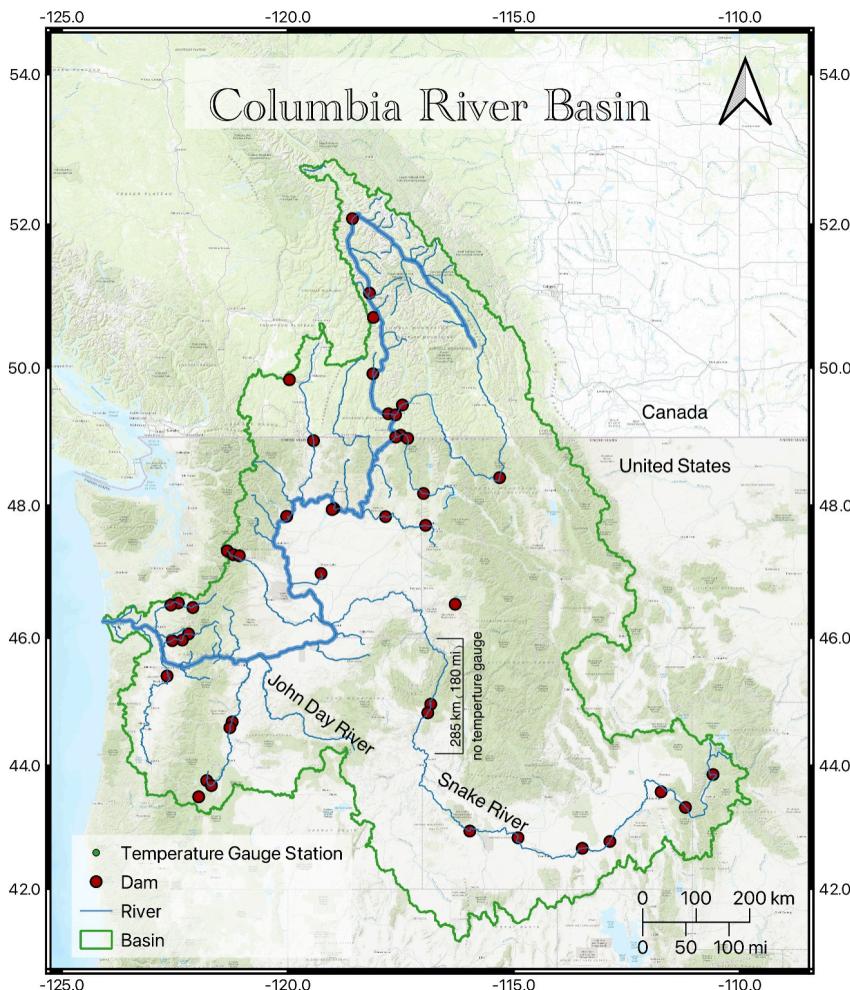


Figure 1. A map of the Columbia River Basin. This map indicates the dams (red points) and temperature gauge stations (green points) along the Columbia River and its major tributaries.

area, surface elevation, and storage change, allowing for the estimation of the reservoir outflow (Das et al., 2022). Remote sensing can also be used to classify the land use/land cover of the riverbanks. With these factors, among many, obtainable through remote sensing, it is possible to use machine learning techniques to fill the gaps in the spatial-temporal continuity of water temperature along an entire river. In situations where cloud cover limits the TIR sensing of water temperature in wider rivers, or when the river width is narrower (<60 m), machine learning can be used to estimate the water temperature. The machine learning approach offers the advantage of scalability where trained models can be used to estimate water temperature in other geographic locations and time periods. Machine learning models, after their development, require fewer computational resources and technical know-how to run compared to other physics-based models.

In this study, we explore the reconstruction of the hydro-thermal behavior of regulated rivers in the Columbia River Basin, taking into consideration the potential factors that influence river water temperature. We also develop a scalable methodology that takes advantage of the global and multi-decadal coverage of satellite remote sensing data (including TIR) to estimate hydro-thermal history globally, even in rivers narrower than 60 m regardless of cloud cover conditions. The rest of the paper is organized as follows. Section 2 describes the methods used in this study, including the study area. Section 3 describes the data used in this study. In Section 4, we discuss findings from our approach to reconstruct the hydro-thermal behavior of regulated rivers. The outcome of our study provides a general spatial-temporal overview of the surface water temperature in the river network of the Columbia River Basin. By extension, the resulting data and the technique developed here can be applicable,

but not limited to, fisheries, reservoir management, and hydropower operations, and inform public health studies such as predicting waterborne diseases.

2. Materials and Methods

2.1. Study Area: The Columbia River Basin

The Columbia River Basin (CRB) is an international watershed that spans approximately a 673,000 km² area of the United States (U.S.) and Canada (Figure 1). It includes portions of seven states in the northwestern U.S. and the southeastern portion of the Canadian province of British Columbia (Yearsley, 2009). The Columbia River has an approximate length of 2,000 km and ultimately discharges into the Pacific Ocean. The Columbia River can be divided into three major sections: a) the Lower Columbia starting from the mouth of the river at the Pacific Ocean upstream toward the Bonneville dam; (b) the Middle Columbia from Bonneville dam upstream to the Grand Coulee Dam and (c) the Upper Columbia starting from the Grand Coulee Dam upstream to the headwaters at Fairmont in British Columbia.

The CRB has been extensively developed over the past decades with over 150 dams for hydropower generation, irrigation, and flood control. There are stream gauge stations within the CRB that record reservoir and stream properties such as discharge, water temperature, and changes in reservoir elevation. However, records at these gauge stations are point-based and unevenly distributed so they are unable to capture the continuous variation in water properties such as temperature. Some tributaries to the Columbia River do not have any temperature-measuring gauge stations along their reaches (Figure 1).

Apart from the limited spatial representation of temperature measurements in the CRB, the availability of daily water temperature records diminishes further back in history. This poses a challenge in studying the basin-wide behavior hydro-thermal regime of the rivers and streams in the CRB. Another apparent drawback is the difficulty in accessing temperature data across borders. In the U.S. for instance, access to point-based temperature data in the CRB is limited to the bounds of the country even though part of the basin lies within Canada. The use of satellite remote sensing data in this study makes it more practicable to conduct basin-wide and cross-boundary studies.

2.2. Water Temperature Estimation

The presence of clouds and narrow widths introduces gaps in TIR-based remotely sensed water temperature records. To estimate river temperature in a spatiotemporal continuum, we therefore considered three different data-driven techniques—linear regression (LR), random forest regression (RFR), and artificial neural networks (ANN)—to estimate the water temperature with the available data. In the initial variation of the models, we estimated the water temperature without factoring in the effect of river regulation. This variation was used to select the best estimation approach. Given the well-established knowledge that reservoirs impact downstream temperature (Ahmad et al., 2021; Bonnema et al., 2020; Olden & Naiman, 2010; Wang et al., 2023), we tested a second variation of the selected model by considering the upstream reservoir dynamics for downstream reaches up to 50 km. The upstream reservoir dynamics considered in the second variation are reservoir storage change and the relative distance downstream from the dam.

The reservoir storage change values were obtained from simulations from the Reservoir Assessment Tool (RAT) developed and improved by Das et al. (2022) and Minocha et al. (2023). RAT is a modeling framework that uses hydrologic modeling with satellite observations to track the complete reservoir state comprising inflow, storage change, surface area/elevation change, evaporative losses, and outflow. We used RAT version 3.0 for this study (see Section 3 and Minocha et al., 2023). Because the reaches in consideration are longitudinal sections of the river instead of specific points along the river, the distance in kilometers from the upstream dam is calculated from the center of the reach to the upstream dam (more details on river reach discretization are provided in Section 3). In the second variation, we included the reservoir dynamics in the inputs and trained the model with reaches within 50 km downstream of a dam.

The independent variables used as model inputs for estimating water temperature include remotely sensed land temperature surrounding the river reach, the land cover classification near the reach, climate classification, nominal river width of the reach, and the time of the year. The land temperature at the reach corridor was obtained from Landsat TIR. We used the average land NDVI along the river reach as a proxy for land cover classification.

Table 1
Selected Reaches for Local-Scale Assessment

Reach characteristic	Test reach 1	Test reach 2	Test reach 3	Test reach 4
Within 50 km below a dam	Yes	Yes	Yes	No
Average width >120 m (width)	Yes (432 m)	Yes (850 m)	No (<30 m)	No (72 m)
Climate Class (Code)	Arid	Cold	Arid	Temperate
Location description	Below Grand Coulee Dam, WA	Below Libby Dam, MT	Below Osoyoos Lake, WA	At Albany, OR
River	Columbia	Kootenai	Okanogan	Willamette
Existing Gauge	USBR-GCGW ^a	USGS-12301933 ^b	USGS-12439500 ^c	USGS-14174000 ^d

Note. These reaches represent varying characteristics based on relative distance from an upstream dam, average width, and climate class. ^aColumbia River below Grand Coulee Dam near Barry, WA (www.usbr.gov/pn-bin/inventory.pl?site=GCGW&ui=true). ^bKootenai River below Libby Dam near Libby, MT (<https://waterdata.usgs.gov/monitoring-location/12301933>); available test data from this location spans January 2020–January 2023. ^cOkanogan River at Oroville, WA (<https://waterdata.usgs.gov/monitoring-location/12439500>). ^dWillamette River at Albany, OR (<https://waterdata.usgs.gov/monitoring-location/14174000>).

The climate class for the reach location was based on the Köppen-Geiger climate classification extracted from a climate classification map at a resolution of 1 km (Beck et al., 2018). The target variable for the models is the in-situ surface water temperature.

Water temperature derived from Landsat TIR data represents the surface skin temperature while in-situ water temperature is representative of temperature in the upper subsurface layer of water. Landsat TIR temperature will be inherently and slightly higher than in-situ temperature. We believe the potential bias is addressed by training the machine learning technique on in-situ data which then transforms Landsat TIR data to mimic in-situ temperature. On the other hand, in-situ temperature data may not always be available for training, especially in developing regions where rivers are mostly ungauged for temperature. To take advantage of the ANN's flexible framework, we therefore modularized the ANN model such that in the absence of in-situ temperature data for training, part of the model can be used independently to generate an 'equivalent' Landsat water temperature for each river reach (see Section 2.2.3). While the reservoir dynamics are combined with the regular inputs for the RFR and LR in the second variation, the reservoir dynamics are added to the second block of the ANN model in the second variation. The first variation of the models was implemented on an irregular timescale depending on the availability of Landsat TIR observations, and the output was resampled to semi-monthly and monthly timescales.

We split the available data into two main sets for model development and testing. Approximately 80% of the available data was set aside for model development in each variation of the models. For the test set, we apportioned data from four handpicked reaches from the year January 2020 to December 2023 (Test reaches 1, 2, 3, and 4; Table 1; Figure 2) in addition to randomly selected instances constituting 20% of the available data. The four handpicked reaches represent unique characteristics of interest within the CRB to contextualize the results in this study. The characteristics of interest include the average reach width, climate classification, and the relative distance from an upstream dam, if any.

2.2.1. Linear Regression Model

To select the data-driven technique, we used a linear regression (LR) model from the suite of non-linear models (described later). This was done to verify if the selected non-linear technique was indeed justified when compared to simpler and linear modeling approaches. The LR model used in our study is the *Elastic Net* LR model (Zou & Hastie, 2005). The Elastic Net model combines the strengths of both ridge regularization (L_2) and lasso regularization (L_1) in estimating the model coefficients. Linear regression in its basic form estimates a target variable (\hat{y}) from a set of n predictors ($X = x_1, \dots, x_n$) using the relationship:

$$\hat{y} = \beta_0 + x_1\beta_1 + \dots + x_n\beta_n + \epsilon \quad (1)$$

where β_0 is the intercept, β_1, \dots, β_n are the coefficients (β) and ϵ is the error. Linear regression uses the ordinary least squares method to estimate the optimum coefficients by minimizing the sum of squared residuals. The L_2 regularization introduces a penalty in the objective function that controls the magnitude of the coefficients while

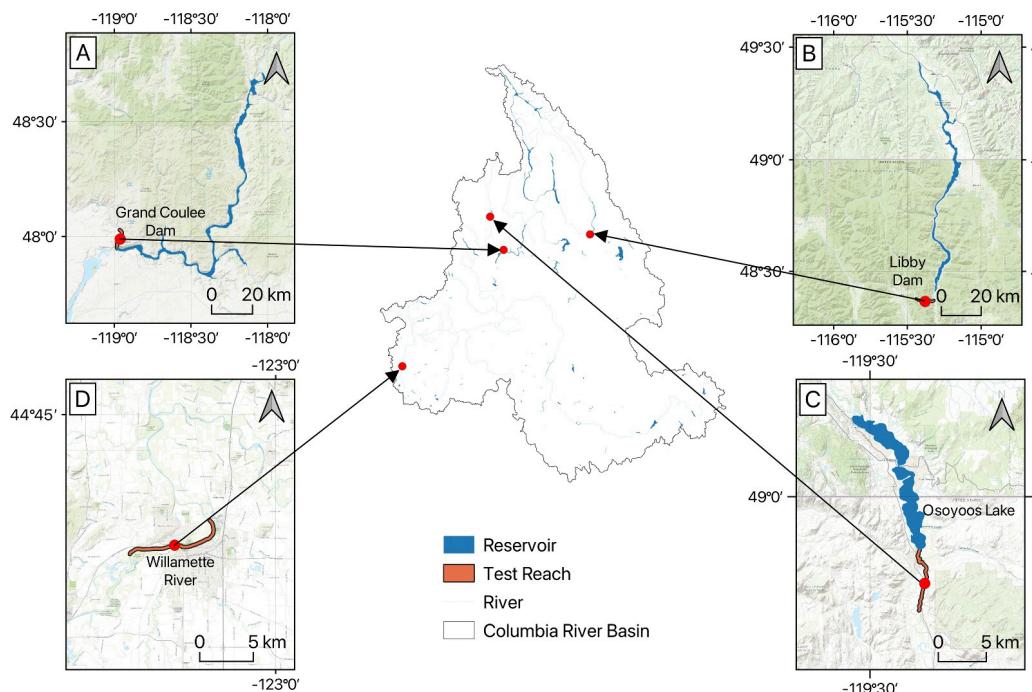


Figure 2. Locations of the selected test reaches within the Columbia River basin. (a) Test Reach 1—Below Grand Coulee Dam; (b) Test Reach 2—Below Libby Dam, MT; (c) Test Reach 3—Below the Osoyoos Lake, WA; (d) Test Reach 4—Willamette River at Albany, OR.

the l_1 regularization introduces a penalty that can control the redundancy of certain features. Using the Elastic Net which combines both the l_2 and l_1 regularizations, the objective function of the LR model is given by:

$$\min_{\beta} \frac{1}{2n} \|X\beta - y\|_2^2 + \alpha\rho\|\beta\|_1 + \frac{\alpha(1-\rho)}{2}\|\beta\|_2^2$$

where α is a multiplication factor for the regularization terms and ρ is the l_1 ratio. When α is 0, the optimization will be equivalent to the ordinary least squares method. A ρ value of 0 results in l_2 regularization while a value of 1 means that only l_1 regularization is used. In each variation of the LR model, we selected the best values for α and ρ by using a 5-fold cross-validation repeated 10 times. More details are provided on the LR as supplementary material (Table S1 in Supporting Information S1).

2.2.2. Non-linear Model: Random Forest Regressor (RFR) Model

The Random Forest Regressor (RFR) was one of the two non-linear machine learning models used in this study. RFR is a version of random forests that takes numerical inputs to generate an output based on the collective vote of outputs from independent decision trees (Breiman, 2001). Voting in random forest regression is done by taking the average output over all the independent trees. In general, the higher the number of decision trees in the random forest model, the better the model is in generalizing and generating accurate results. The loss function used to train the RFR was the mean squared error. For our model, we selected the optimal number of decision trees using 5-fold cross-validation repeated 10 times. Additional details on the RFR model are provided as supplementary material (Table S2 in Supporting Information S1).

2.2.3. Non-linear Model: Artificial Neural Network (ANN) Model

The second non-linear model used in this study was the classic artificial neural network (ANN) model. The structure of the ANN model consists of two main blocks (Figure 3). The inputs are initially passed through an n -unit dense layer before the first block. The first block (block 1) consists of an initial eight-unit layer followed successively by $2n$ -, $4n$ -, $4n$ -, $2n$ -, and 1-unit dense layers. The initial number of units, n , is a hyperparameter that

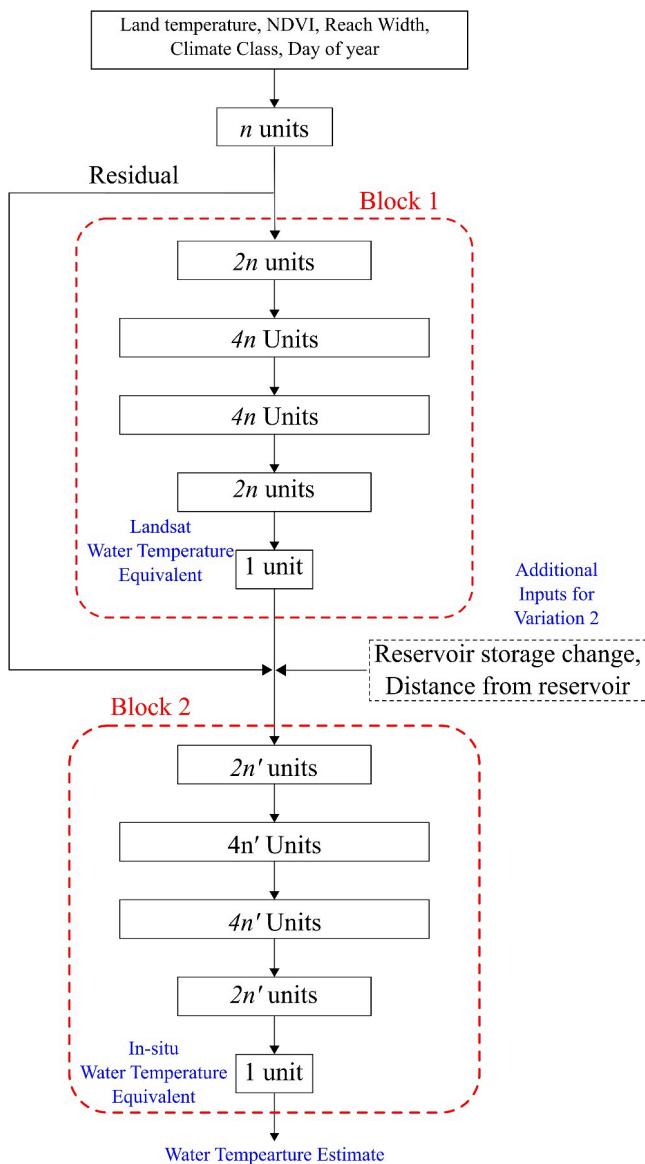


Figure 3. Configuration of the ANN model. There are two main blocks. The first block predicts the equivalence of Landsat water temperature which is then combined with a residual from the initial layer as inputs to the second block which then predicts water temperature close to the in-situ records. The base form of the model without the additional inputs is Machine Learning 1 (ML1). Additional inputs for Machine Learning 2 (ML2), a variation of the model to include reservoir storage dynamics, are introduced to the base model after Block 1.

subsets, and the model is trained iteratively on the subsets while one is reserved for validation. Each subset is used for validation once by the end of one complete round of cross-validation. We repeated the 5-fold cross-validation 10 times for each combination of hyperparameters, resulting in a total of 50 iterations of training and validation for each set of hyperparameters. The main reason for the repeated 5-fold cross-validation was to provide a statistically representative basis for inter-model comparison. The Hyperband algorithm uses the championship bracket approach to select the best combination of hyperparameters. The number of candidate combinations is halved by a factor after each round of iteratively increasing the number of epochs in training. We used a halving factor of three in the ANN model development. Halving at each round of brackets is done based on the MSE score

is tuned in the model development stage. Block 1 is trained independently with Landsat TIR as the target variables and hence can be used as a standalone to predict the equivalent Landsat water temperature. The output of the initial n -unit layer is carried over as a residual to the second block (block 2). Block 2 is configured similarly to the layers that come after the initial layer of block 1. However, the value of n -units (referred to as n' in block 2) may differ depending on the outcome of the hyperparameter tuning of block 2. The main function of block 2 is to correct the Landsat water temperature to mimic in-situ water temperature by training it separately from block 1. Block 2 is trained to predict the equivalent in-situ water temperature. When training block 2, block 1 is frozen since the layers were trained separately on Landsat water temperature. In the second variation of the ANN model, the reservoir dynamics are introduced as inputs before block 2 (Figure 3). This is to maintain the independence of block 1.

We constructed the ANN model using the TensorFlow framework (Abadi et al., 2016). Both blocks of the ANN model were compiled using the Adam optimizer (Kingma & Ba, 2014) and mean squared error (MSE) as the loss. For each variation of the ANN model, we optimized the model hyper-parameters using a grid search method for each block. The ANN hyper-parameters include the number of units and the activation function of the initial layer, the learning rate (α), and the batch size of both blocks. Due to the nature of the ANN configuration, we used the same version of block 1 for all the different variations. We controlled the number of iterations in the training process using the early stopping callback with a patience of 10. For each block, we used the training data for a 5-fold repeated cross-validation to assess the performance of the model.

2.3. Model Development

The model development involves tuning the model hyperparameters to obtain the optimum set of model hyperparameters that yields the best score when the model is validated. We used the mean squared error (MSE) as the score for the model development:

$$MSE = \frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2 \quad (3)$$

where O and P represent the observed water temperature and estimated water temperature respectively, and n is the number of samples used in the model evaluation. It is recommended to separate the development data into training and validation subsets; however, we used two different algorithms, the 5-fold repeated cross-validation for the LR and RFR models and the Hyperband algorithm (Li et al., 2016) for the ANN, that did not require us to divide the development set into training and validation sets categorically. In 5-fold cross-validation, the development data set is randomly divided into five

of the hyperparameter combinations. At the end of the hyperband algorithm, the set of model hyperparameters that yields the lowest MSE is selected.

2.4. Model Evaluation

After developing the models and selecting the optimum hyperparameters, we evaluated the final model by conducting a 5-fold cross-validation repeated 10 times to calculate the average mean absolute error (MAE), root mean squared error (RMSE), Nash–Sutcliffe model efficiency coefficient (NSE) (Nash & Sutcliffe, 1970) and the Kling–Gupta Efficiency (KGE) (Kling et al., 2012) using the formulae:

$$MAE = \sqrt{\frac{1}{n} \sum_{i=1}^n |O_i - P_i|}$$

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

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$$

$$KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}$$

where O and P represent the observed water temperature and estimated water temperature respectively, n is the number of samples used in the model evaluation, r is the coefficient of correlation, β is the bias ratio, that is, the ratio of the mean of P to the mean of O and γ is the variability ratio, that is, the ratio of the coefficient of variation of P to the coefficient of variation of O . The MAE and RMSE are an indication of the amount of error in the model estimates. However, the RMSE is sensitive to higher error values while the MAE provides an unweighted indication of error among the models. The closer the MAE or RMSE value is to 0, the better the model performance. The NSE indicates how well the observed and simulated plots fit on a 1:1 line (Moriasi et al., 2007). Ranging from $-\infty$ to 1, an NSE or KGE value closer to 1 represents a better fit. The KGE was formulated to overcome some of the limitations of the NSE such as handling the bias, correlation, and variability between observed and simulated data more effectively (Gupta et al., 2009). While both NSE and KGE are used to assess model performance, KGE offers a more comprehensive evaluation by considering multiple statistical properties simultaneously. We consider multiple evaluation metrics because they offer complementary insights into different aspects of model performance. While one metric may highlight the accuracy of predictions, another may emphasize the model's ability to capture variability or bias. By examining a range of metrics, we can gain a more comprehensive understanding of the model's strengths and weaknesses, leading to more informed decisions in model selection, refinement, and application.

2.5. Reconstruction of Long-Term Hydro-Thermal Behavior

After developing our model, we applied it to reconstruct the long-term seasonal hydro-thermal patterns along the Columbia River. We assessed the rate of change in temperature along the lower, middle, and upper sections of the Columbia River for each season, offering a longitudinal perspective on water temperature. To enhance our insights, we transformed this linear representation into a 2-dimensional overview across the CRB. As a result, we also created an online tool, facilitating convenient access to the reconstructed spatiotemporal water temperature data.

3. Data

The geophysical properties of a river, such as its width and temperature, vary along the river's length. To adequately capture such variations along the rivers, we've divided each river into smaller reach units that are no longer than 10 km in length based on an azimuthal equidistant projection (Figure 4). The mean width of each reach was extracted from the Global River Widths from Landsat (GRWL) Database (Allen & Pavelsky, 2018). The GRWL Database contains the estimated mean widths of rivers wider than 30 m at mean annual discharge.

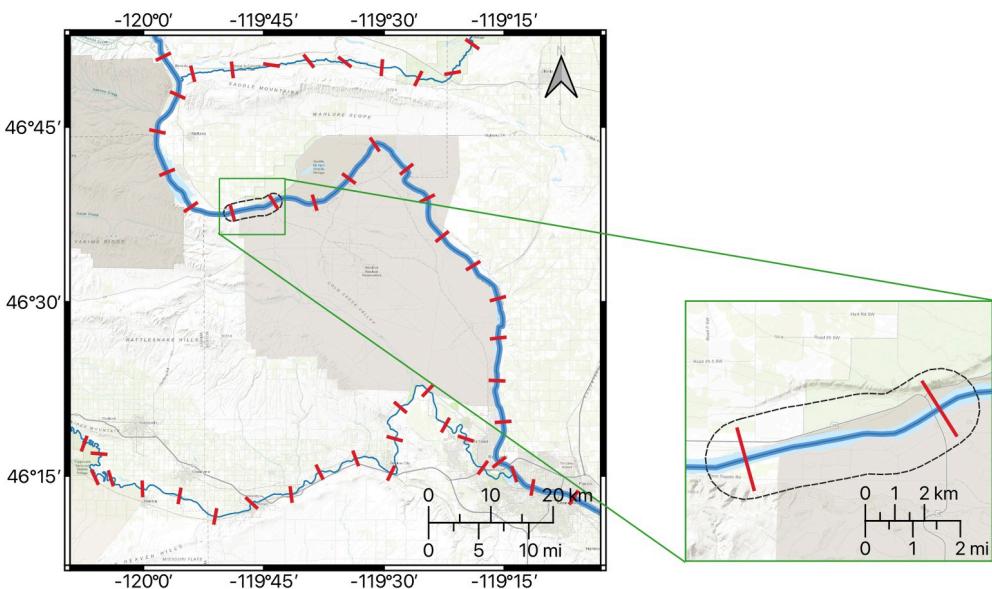


Figure 4. A map showing the division of rivers into 10-km reaches. The width of each reach is extended by 120 m on each side to create a reasonable area for land temperature and land cover extraction.

Therefore, for reaches narrower than 30 m, the nominal width of 30 m was assigned. To account for the effects of land temperature and land cover, the width of each reach was extended by a 120-m buffer on each side (Figure 4).

Remotely sensed geophysical data were obtained from Landsat 4, 5, 7, 8, and 9 Level 2, Collection 2, Tier 1 products, which contain atmospherically corrected surface reflectance and land surface temperature. The products also contain quality bands that were used to filter out clouds and high-saturation pixels. The surface temperature values in the Landsat products are precalculated based on the single-channel algorithm (Jimenez-Munoz et al., 2009). Landsat TIR sensors measure the surface water temperature while in-situ temperature probes are designed such that the probe remains submerged underwater (Heck et al., 2018). Therefore, we anticipate an inherent difference between the two sources of water temperature measurements. For each reach, we discriminated water pixels from non-water pixels using the Landsat Quality Assessment (QA). The QA band uses bit assignments that provide information that describes the Landsat image pixel such as the presence of clouds, snow, or water. The normalized difference water index (NDWI) can be used to classify water pixels, but this requires a careful determination of an optimum threshold for discrimination. Based on the water/non-water pixel classification, we derived the median water and non-water temperature for the reach. Here, we used the median value instead of the mean because the mean is sensitive to outliers and the presence of some erroneous pixel values within the buffers of the reach can translate into additional noise in the Landsat TIR temperature values. The median value pixel value tends to be more representative of the water temperature than the mean value (Cristea & Burges, 2009; Torgersen et al., 2001). The land cover of the reach corridor was classified using the averaged normalized difference vegetation index (NDVI) value of the non-water pixels. The NDVI was calculated using the NIR and red bands. The derived temperature and NDVI values were resampled temporally to monthly and semi-monthly averages.

All remotely sensed data were calculated and retrieved from the Google Earth Engine (GEE) platform (Gorelick et al., 2017). GEE is a cloud-based platform for geospatial analysis that hosts an array of accessible geospatial tools and data sets including Landsat products. GEE eliminates the need to download large data files and the need for high computational resources to analyze and retrieve the desired geospatial data.

The climatic conditions of a particular region determine the thermal response and interactions with geophysical factors. To consider the climatic variations from place to place, we included the Köppen-Geiger climate classification in our study. There are five main climate groups and 30 sub-divisions in the Köppen-Geiger system of climate classification which is “based on threshold values and seasonality of monthly air temperature and precipitation” (Beck et al., 2018, p. 2). For our study, we extracted the Köppen-Geiger climate classification from the “present” climate classification map developed by Beck et al. (2018) at a resolution of 1 km. According to the

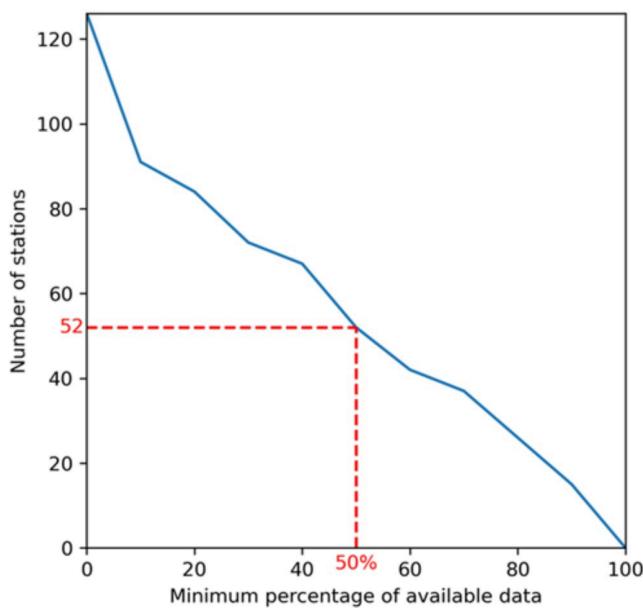


Figure 5. There are about 52 gauge stations along the Columbia River and its tributaries have at least 50% of available temperature records since 2000.

data from the U.S. Army Corps of Engineers and other agencies within the Columbia River Basin (CRB). We collated 42 years of in-situ data from August 1982 through February 2024. The in-situ data comprises temperature records from 134 gauge stations along the Columbia River and its tributaries. Out of the curated in-situ data, 52 of the gauge locations had at least 50% of data availability from the year 2000 with less frequency of availability in earlier records (Figure 5).

Because the data we used for water temperature estimation are of different ranges, we scaled the input data before using them in the machine learning models. In general, scaling the input data helps to achieve better performance. Data that tended to follow a distribution, such as the TIR land temperature and the NDVI, were scaled using a standard scaler while data that had a finite linear scale, such as the climate class and day of the year, were normalized into the range from 0 to 1 using a min-max scaler (Table 2).

Köppen-Geiger classification system, the arid (dry), temperate, and cold (continental) climate groups are the most prevailing in the CRB. The difference between the temperate and cold climate groups is that the mean air temperature in the coldest month is between 0 and 18°C in the temperate zone while that of the cold region is below 0°C (Beck et al., 2018).

The operation of large reservoirs can alter the hydro-thermal regime of rivers significantly (Ahmad et al., 2021). Therefore, information on reservoir operation can be critical in estimating downstream water temperature. For each dam within the study area, we obtained the reservoir dynamics such as the storage change and surface area using RAT (Das et al., 2022; Minocha et al., 2023). As mentioned earlier in Section 2, RAT combines satellite remote-sensing data and hydrological modeling to estimate reservoir dynamics. Data from optical and synthetic aperture radar sensors are used to determine the reservoir surface area, elevation, and changes in storage while the reservoir inflow is estimated using hydrologic models such as the Variable Infiltration Capacity model version 5 (VIC-5; Hammad et al., 2018).

We assessed the accuracy of water temperature estimates by comparing them with in-situ water temperature measurements obtained from the United States Geological Survey (USGS) and the Bureau of Reclamation (USBR). We also used in-situ data from the Columbia River DART (Data Access in Real Time <https://www.cbr.washington.edu/dart>), which collates and publishes river

Table 2
Summary of the Data Used in Developing the Water Temperature Estimation Models and Their Various Sources

Data	Unit	Scaler	Source
TIR land/water temperature	°C	Standard	Landsat 4 (August 1982–June 1993), Landsat 5 (March 1984–May 2012), Landsat 7 (May 1999–January 2024), Landsat 8 (March 2013–February 2024)
NDVI	–	Standard	Landsat 9 (October 2021–February 2024)
Climate Class	–	Min-Max	Köppen-Geiger Climate Classification Map (Beck et al., 2018)
Average reach width	m	Standard	GRWL (Allen & Pavelsky, 2018)
Reservoir Storage Change	m ³	Standard	RAT (Minocha et al., 2023)
Reservoir Surface Area	km ²	Standard	RAT (Minocha et al., 2023)
In-situ water temperature	°C	Standard	CBR-DART ^a , USBR ^b , USGS ^c
Day and Year	–	Min-Max	–

^aCBR-DART: https://www.cbr.washington.edu/dart/query/river_daily. ^bUSBR: <https://www.usbr.gov/pn/hydromet/arc-read.html>. ^cUSGS: <https://dashboard.waterdata.usgs.gov>.

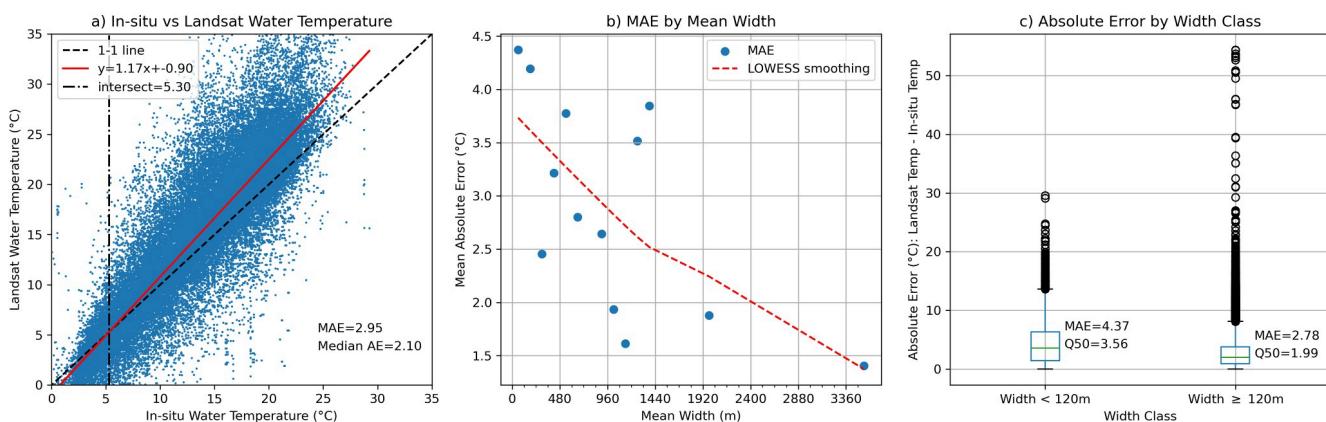


Figure 6. Comparison between Landsat water temperature and In-situ water temperature measurements. (a) 1-1 plot of the Landsat-based water temperature and in-situ water temperature. The 1-1 line is denoted by the dashed line. (b) The trend of mean absolute error between Landsat-based temperature and in-situ temperature with the reach width. (c) The interquartile distribution of the absolute deviations of Landsat-based temperature from in-situ temperature.

4. Results

4.1. Relationship Between Landsat TIR and In-Situ Water Temperature

Limitations due to the spatial resolution of the Landsat TIR sensor can affect the Landsat-based water temperature. In general, Landsat-based water surface temperature tends to be higher than in-situ temperature. The magnitude of deviation of the Landsat-based water temperature increases with an increase in temperature. For the data collected within the study area, the increasing positive deviation is reversed for temperatures below 5.30°C (Figure 6a). In terms of the widths of the reaches, there is a decreasing trend in the magnitude of error with an increase in mean width (Figure 6b). A comparison of the distribution of errors by width class shows higher errors in reaches that have widths narrower than 120 m than those wider than 120 m (Figure 6c). Overall, the mean and median absolute errors for the Landsat-based water temperature are 2.95°C and 2.10°C respectively (Figure 6a). Meanwhile, the mean and median absolute errors for reaches narrower than 120 m are 4.37°C and 3.56°C, respectively, and those of the reaches wider than 120 m are 2.78°C and 1.99°C (Figure 6c) which is closer to the population mean and median errors.

4.2. Water Temperature Estimates

4.2.1. Selection of Temperature Estimation Model

In this study, we set out to identify an effective data-driven approach that can be used to estimate the surface water temperature in regulated river networks and reconstruct a continuous spatiotemporal thermal behavior of rivers using satellite remote sensing and other widely available data. As mentioned in Section 2, we considered three different machine-learning approaches, LR, RFR, and ANN, to estimate the water temperature. Among the three different models, the RFR model performed best in estimating water temperature in the test data set, with $MAE = 0.71$, $RMSE = 1.18^\circ\text{C}$, $NSE = 0.96$, and $KGE = 0.97$ (Figure 7c). Although the ANN model performed better than the LR, the performance of the RFR outweighed the ANN's. In an analysis of the distribution of errors in the models, the RFR estimates resulted in the least variability ($\sigma = 1.13^\circ\text{C}$) compared to the ANN ($\sigma = 1.65^\circ\text{C}$) and LR ($\sigma = 3.31^\circ\text{C}$). When compared to the raw Landsat water temperature, the distribution of errors in the ML models was closer to a zero mean while the raw Landsat water temperature showed a general overestimation with $\mu = -1.57^\circ\text{C}$ (Figure 7e). Given the superior performance of the RFR to the ANN and LR, we selected the RFR (variation 1) for the estimation of water temperature. Details on the model development, including the selected optimum hyperparameters for each model, are presented in the Supporting Information Tables S1–S3 and Figures S1, S2 in Supporting Information S1.

In a second variation of the RFR model, we assessed the model performance after including reservoir dynamics. Although dam operation is known to affect downstream water temperature, the use of more extensive training data from Landsat collection spanning 42 years provides inconclusive evidence that a simplistic representation of reservoir storage dynamics in the RFR model can improve the accuracy of the model in reaches influenced by the

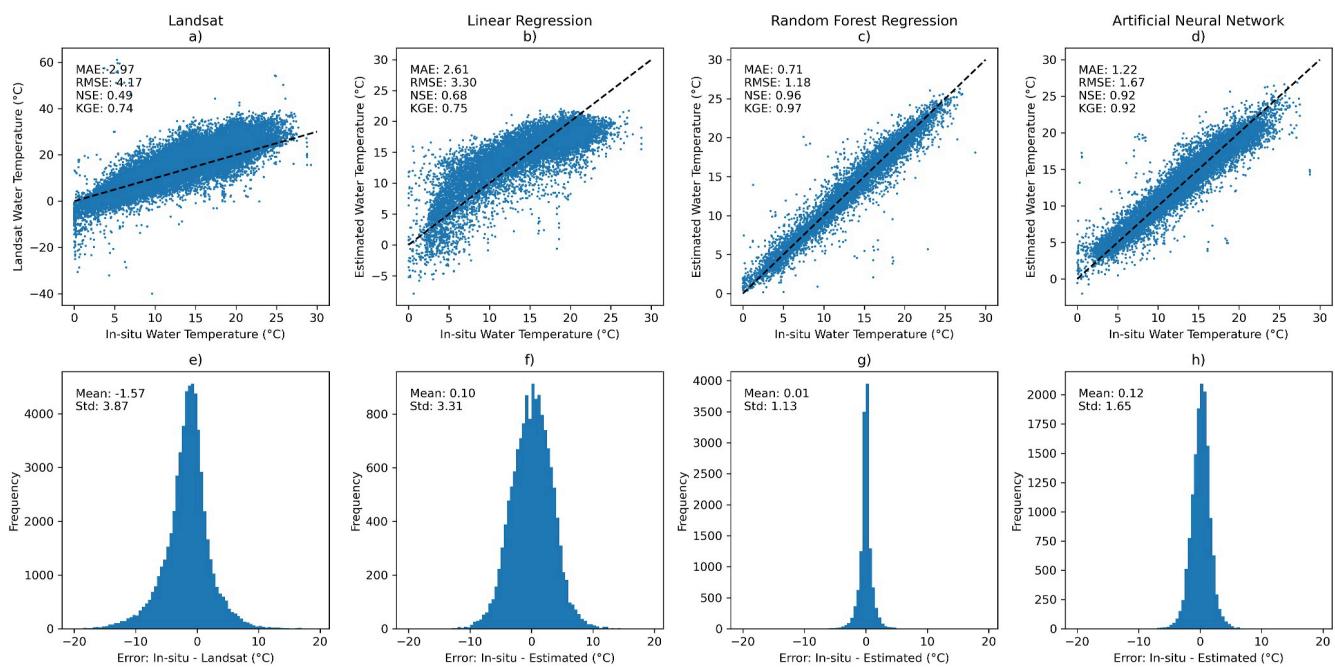


Figure 7. The performance of the water temperature estimates from Landsat and the various machine learning models. The top row (a, b, c, d) represents the scatter plot of the in-situ water temperature against the Landsat- and machine learning-based estimated water temperature. The bottom row (e, f, g, h) represents the distribution of error in the Landsat- and machine learning-based estimated water temperature. The models were evaluated using the test data sets and the estimates were generated using variation one of the models.

dam. The test results showed marginal differences in performance between the two variations (see Figures S3 and S4 in Supporting Information S1 for further details on the second variation). A key insight that emerges here is that a data-driven technique trained on a sufficiently long record improves predictability of surface water temperature even at locations that are plausibly influenced by upstream reservoir operations without the explicit accounting of reservoir storage dynamics in the training data (see Section 4.2.2). In other words, the longer the training data, the data-driven technique has less need for reservoir storage data due to the diminishing value in contributing to the learning of surface water temperature dynamics at reaches potentially influenced by dam operations.

4.2.2. Assessment of Surface Water Temperature Estimates at Selected Reaches

To assess the local performance of the selected data-driven estimation model (i.e. RFR variation 1), we selected reaches with unique characteristics such as reach width, climate class, and proximity to an upstream dam (see Table 1). Figure 8 shows a time series plot of estimated water temperature with in-situ water temperature records at the selected reaches. Although satellite remote sensing of water temperature is limited by cloud cover, our data-driven model estimates water temperature that tracks the seasonal variations, including the highs and lows, at all the test reaches, which is the intended goal of this study. This is also evident in the 1:1 plot with NSE values ranging from 0.86–0.98. Another limitation of using only satellite remote sensing for surface water temperature estimates is the spatial resolution which makes it difficult to obtain water temperature for narrower rivers solely based on satellite observations. The data-driven machine learning model makes it possible to overcome this limitation of river width and spatial resolution to generate estimates for river reaches narrower than 120 m, as seen in Test reaches 3 and 4 which have an average width of less than 30 and 72 m respectively (Figures 8c and 8d). However, the magnitude of error is least in the wider reaches and highest in the narrower reaches as indicated by the MAE values of 0.71, 1.44, 2.05, and 1.09 for test reaches 1–4, respectively (Figures 8e–8h).

4.3. Reconstruction of Long-Term Hydro-Thermal Behavior

Using our model (i.e. RFR variation 1), we reconstructed the water temperature for all reaches along the Columbia River, whether there is an in-situ temperature or not. Figure 9 shows the reconstructed behavior of water

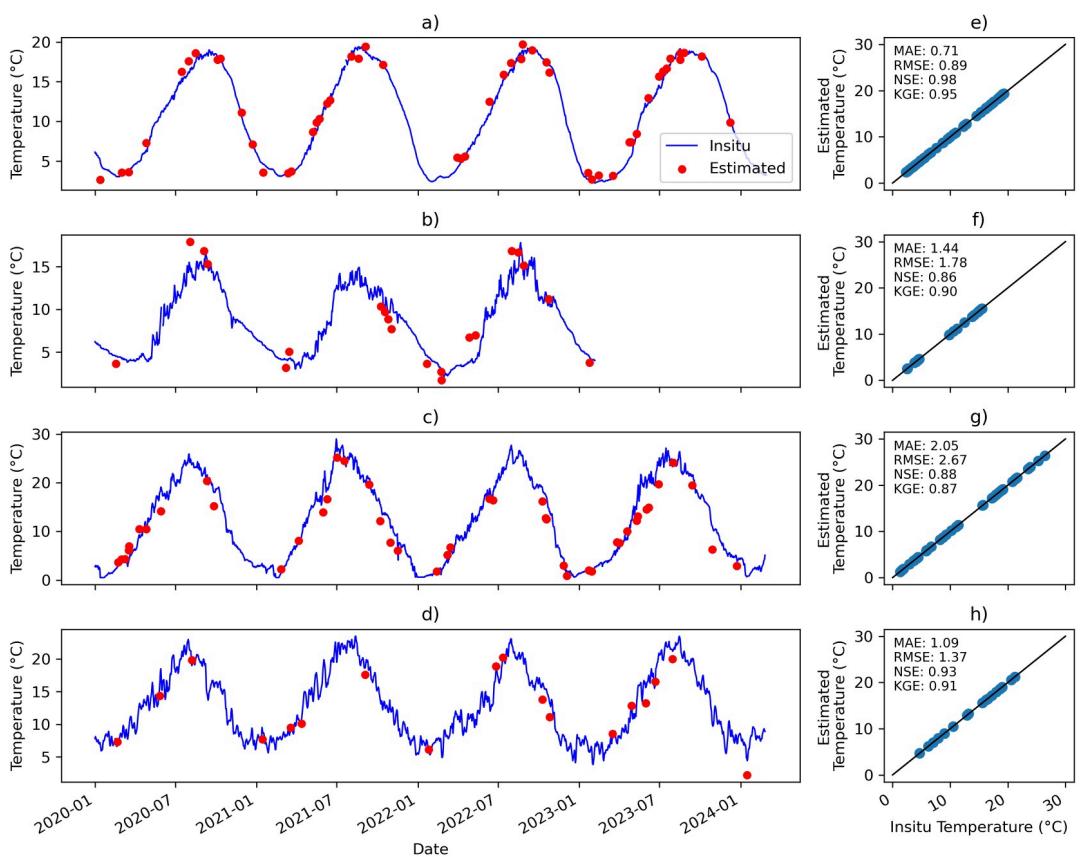


Figure 8. Estimated vs. in-situ water temperature at the selected test reaches. The left column shows the water temperature time series of selected reaches within the Columbia River Basin: (a) Test Reach 1—Columbia River, below Grand Coulee Dam, WA; (b) Test Reach 2—Kootenay River, below Libby Dam, MT; (c) Test Reach 3—Okanogan River, below Osoyoos Lake, WA; and (d) Test Reach 4—Willamette River, at Albany, OR. The in-situ daily temperature is represented by solid blue lines overlayed by the estimated water temperature represented by red solid dots. (e)–(h) Represent the 1:1 plots of their corresponding plots in (a)–(d) respectively. The plots were generated using the estimated values of the test data set and the estimates were generated using variation one of the RFR model.

temperature at the test reaches beyond the available in-situ records. Although our approach is limited to the Landsat observation days, it provides a long-term overview of the water temperature dynamics that can be used to assess climatological patterns and anthropogenic trends in the reaches in the absence of in-situ data. For example, there was no in-situ record for about half a decade (1988–1994) at Test Reach 1 but now, we can generate the hydro-thermal behavior during this period (Figure 9a). Available water temperature data at Test Reach 2 only started mid-2021 but we can now reconstruct the water temperature before the inception of in-situ data collection at this location (Figure 9b). Likewise, we reconstructed the water temperature at test reaches 3 and 4 (Figures 9c and 9d), as well as all other reaches within the CRB.

The reconstructed water temperature at the reaches within can be aggregated to generate spatial trends along any river or within a basin at large. For example, Figure 10 shows the reconstructed long-term mean water temperature profile for all four seasons along the Columbia River. The estimated means matched closely with the in-situ long-term means, except for a few outliers which were still within 1 standard deviation of uncertainty. Some of these outliers are due to the absence of enough data for comparison.

4.4. THORR—An Online Tool for Tracking Thermal History of Regulated Rivers

Based on our systematic investigation to develop a scalable data-driven technique and create continuous spatiotemporal river temperature estimates using TIR data, we developed an interactive online tool called the Thermal History of Regulated Rivers (THORR). We present THORR for public access at <https://depts.washington.edu/saswe/thorr/> to visualize the answer to our key research question of the study in the form of a tangible

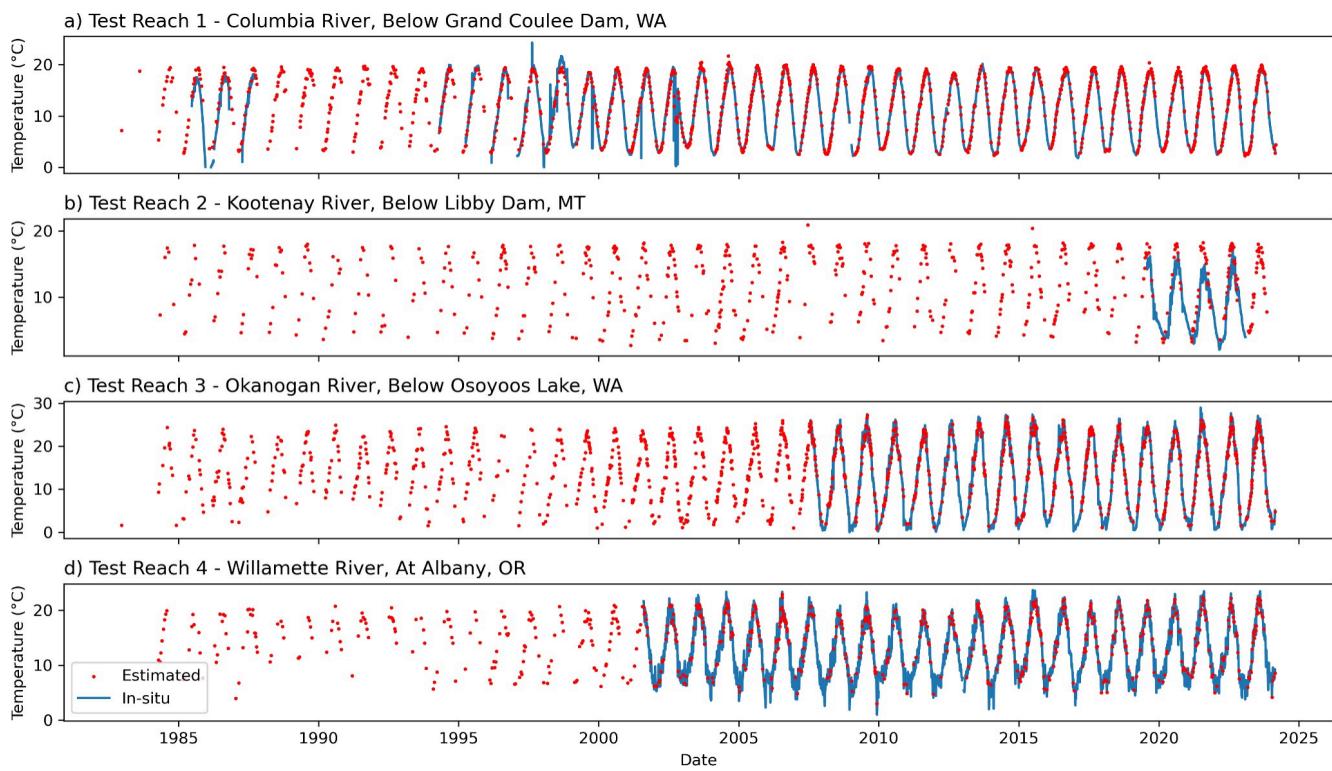


Figure 9. Reconstructed hydro-thermal behavior at (a) Test Reach 1—Columbia River, below Grand Coulee Dam, WA; (b) Test Reach 2—Kootenay River, below Libby Dam, MT; (c) Test Reach 3—Okanogan River, below Osoyoos Lake, WA; and (d) Test Reach 4—Willamette River, at Albany, OR. The water temperature estimates were generated using variation one of the RFR model.

end product (Figure 11). THORR provides easy access to observed Landsat and estimated water temperature for reservoirs and 10-km reaches within the CRB. The tool provides a spatiotemporal overview of water temperature and overcomes data gaps due to cloud cover or limited river width by using the data-driven RFR model. For each reach or reservoir, the tool can display the historical time series, long-term averages, and deviations at semi-monthly and monthly timescales. The tool also provides access to temperature estimates for dams within the CRB. Although the first version of the tool is focused on the CRB, future upgrades are set to expand the domain to include other regulated basins around the globe.

5. Discussion

Landsat surface water temperature observations are subject to various sources of uncertainty that lead to a general overestimation (Figure 6a). One such factor is the spatial resolution of the satellite remote sensors, leading to mixed pixel values, particularly in areas near riverbanks. As depicted in Figure 5b, narrower river sections tend to exhibit higher errors due to challenges in distinguishing pure water pixels and the frequent contamination from adjacent land temperatures. Additionally, a discrepancy arises from the nature of Landsat-based water temperature being an area aggregate of the radiant temperature at the water surface, contrasting with the in-situ measurement being the bulk temperature taken at a point location beneath the water surface. This inherent contrast introduces uncertainty (Handcock et al., 2006). The tendency for Landsat to overestimate water temperature, along with its deviation behavior correlated with the width of the river reach, echoes the conclusions drawn by Handcock et al. (2006) in the CRB, and a broader generalization to Landsat and other remote sensing observations of water temperature in other regions (Cherkauer et al., 2005; Martí-Cardona et al., 2019; Murphy et al., 2021).

The assessment of various data-driven techniques offers insight into the effectiveness and limitations of the different architectures in predicting riverine surface water temperature based on our study's inputs. Notably, the relatively poor performance of the LR model underscores its reliance on an assumed linear relationship between water temperature and input variables. This assumption can inaccurately infer linear correlations, as seen in the LR estimates, where certain input features like land temperature and day of the year exhibit stronger linear

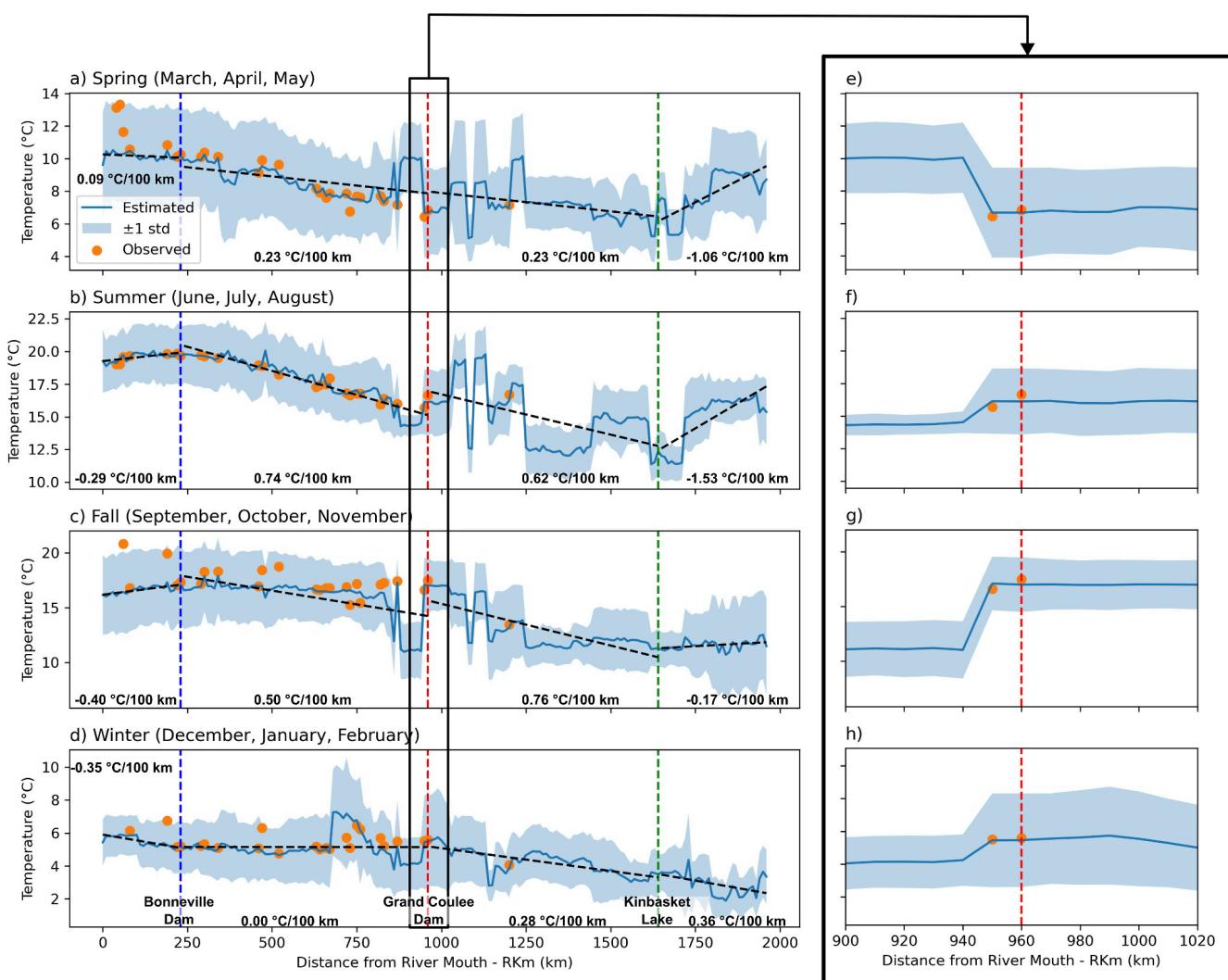


Figure 10. Longitudinal profile of the long-term seasonal hydro-thermal behavior along the Columbia River. The long-term means were reconstructed for (a) Spring (March, April, May), (b) Summer (June, July, August), (c) Fall (September, October, November), and (d) Winter (December, January, February) seasons using reconstructed water temperature from 1982 to 2024. The plots in the right column (e)–(h) are zoomed-in hydro-thermal profiles at the upstream and downstream reaches of the Grand Coulee Dam.

correlations than the in-situ water temperature itself. In contrast, the non-linear models demonstrate correlations more akin to the in-situ data. The RFR model performed best in learning the intricate non-linear relationships within the data, surpassing both the LR and ANN models. Its ensemble structure notably enhances output accuracy. While the ANN model also outperformed LR, the superiority of the RFR model highlights its robustness in capturing complex data dynamics, offering valuable insights into riverine water temperature estimation.

Despite the nuances of performance, all the data-driven techniques addressed the general overestimation in the Landsat-based water temperature. This is seen in the assessment of the model errors where the distributions of errors have mean values much closer to 0 than that of the Landsat-based water temperature. The distribution of errors in the RFR model yields the smallest range of errors ($\pm 1.13^\circ\text{C}$ std) followed by the ANN and LR, respectively. However, unlike the RFR and ANN, the LR generally overestimated in the mid-range of temperature while underestimating in the extreme lower- and upper-range temperatures (Figure 7b), hence a balance in the distribution of errors (Figure 7f). Regarding long-term means along the Columbia River, the in-situ water temperature mostly aligned with the estimated water temperature within 1°C of standard deviation. Higher deviations from the estimated long-term means were observed at locations without enough in-situ data to generate long-term means. For example, the available in-situ temperature records for comparison at 40 km of the Columbia River spanned April–July 2014, hence the high deviation from the estimated long-term mean.

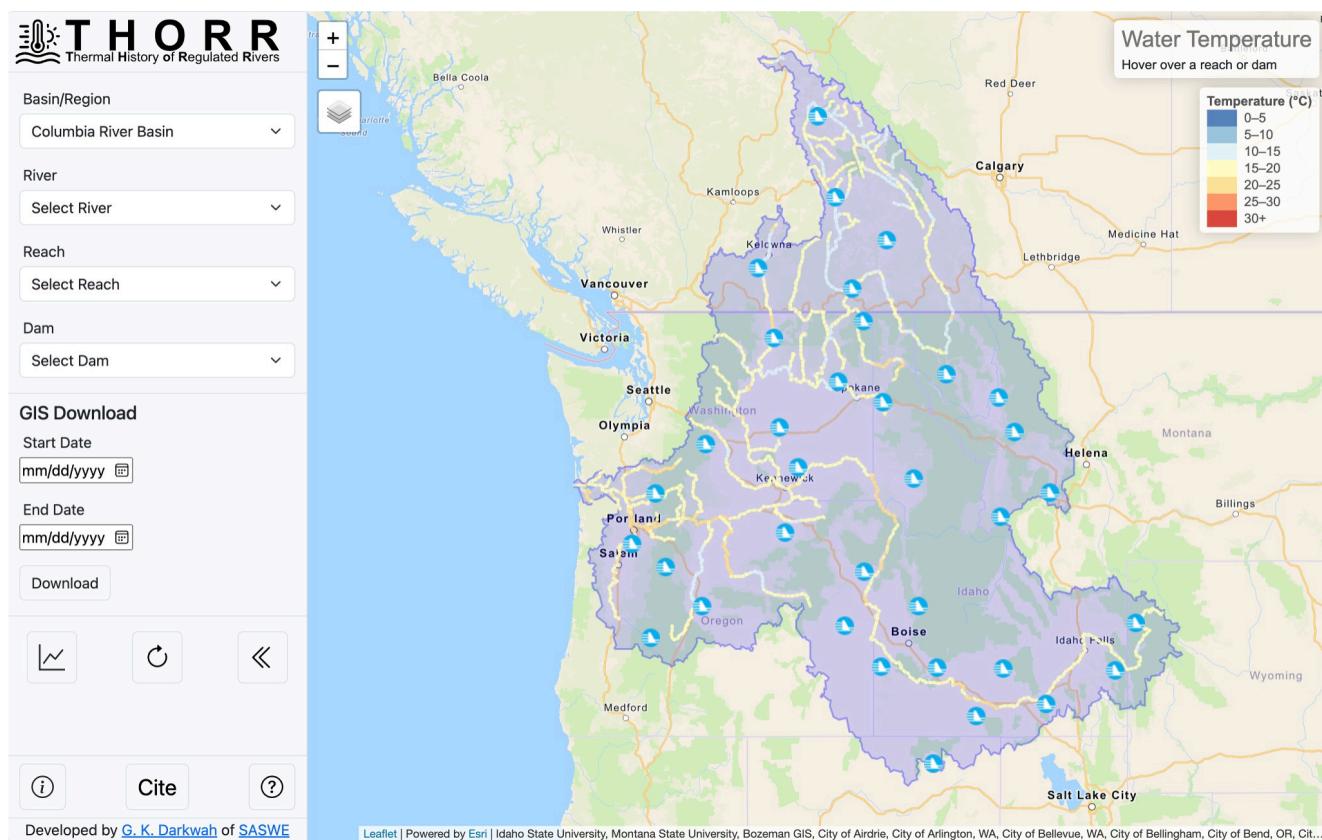


Figure 11. Thermal History of Regulated Rivers (THORR) online tool at <http://depts.washington.edu/saswe/thorr>.

In general, there is downstream warming along the Columbia River, especially from Kinbasket Lake to Bonneville Dam (Figure 10). This is expected as the Columbia River flows from the high elevation and colder region of the Kootenay mountain range and the river collects warmer runoff draining more from downstream tributaries and warmer land surface. The downstream warming is more significant in the summer with a heating gradient of 0.62–0.74°C every 100 km. As cold water from snowmelt flows from the mountain ranges, the water warms due to a combined interaction between air temperature and solar radiation. Such downstream warming is consistent with the findings of Cristea and Burges (2010) where there was a summer downstream heating trend in the Yakima River, also located in the CRB. In another study by Petersen and Kitchell (2001), the July water temperature observed at Bonneville Dam (230 km along the Columbia River) was consistently higher than the water temperature observed further upstream at Rock Island Dam (730 km along the Columbia River).

Although contrary to the general observed downstream warming trend, the negative gradient between the headwaters and Kinbasket Lake in Spring, Summer, and Fall could be attributed to the headwater temperature. The Columbia River originates from the Columbia Lake which is a warm lake at Fairmont, BC, rife with hot springs that contribute to the lake (Grasby & Hutcheon, 2001). To the best of our knowledge, there is no peer-reviewed study that investigates such unexpected hydro-thermal behavior. Therefore, using our approach to reconstruct water temperature can be a gateway to understanding the hydro-thermal behavior of such natural phenomena in the absence of in-situ water temperature measurements.

There are several dams along the Columbia River that can influence the water temperature. The Grand Coulee Dam, located 960 km along the river, is a prime example of how dam operations and the thermal dynamics of reservoirs interact both upstream and downstream (Figures 10e–10h). The reconstructed water temperature shows a difference between the upstream and downstream temperatures during the four seasons. In the summer and fall, the downstream reaches experience cooler temperatures due to thermal stratification when cooler water is drawn from the bottom layer of the reservoir through the penstock. In the fall, the Grand Coulee Dam is operated at maximum elevation for power generation which is a recipe for thermal stratification, hence the high difference

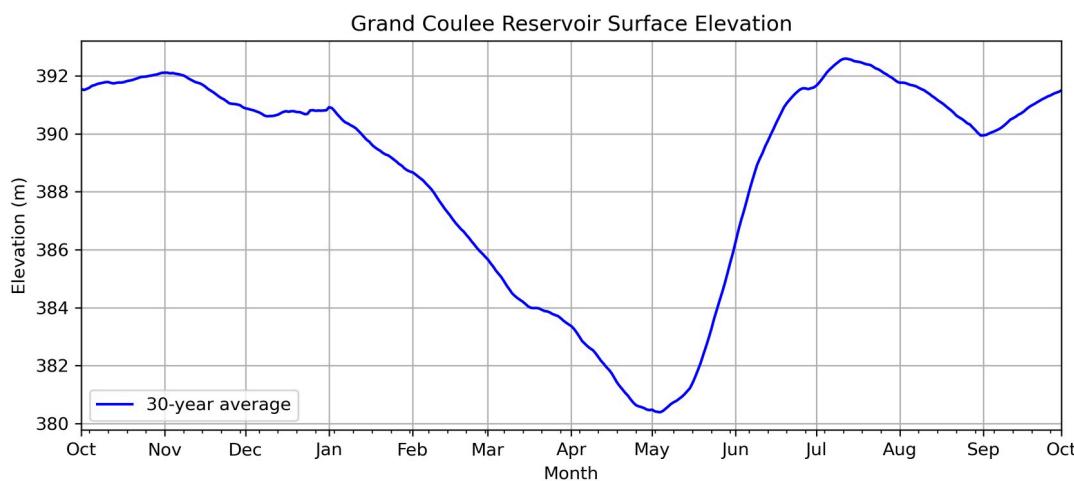


Figure 12. Long-term water surface elevation at the forebay of the Grand Coulee dam. The average was calculated over 30 years from 1 October 1992, to 30 September 2023. The reservoir is drafted from January to May for flood risk management in anticipation of high flows in the spring and early summer months.

between the upstream and downstream temperatures. Starting from the winter through spring (January–May), the dam is operated to reduce the reservoir water level for flood risk management (FRM) in anticipation of high inflows during spring and early summer (Figure 12). Later in the spring, water is impounded to prioritize hydropower generation in the summer and fall. This mode of operation could contribute to the reversed temperature difference in the spring, prompting the need for further research into the downstream temperature impacts of dam operations which is possible using our approach to reconstruct water temperature in the absence of in-situ temperature records.

There are two major lessons that emerge from our study that are of broad and global relevance to the scientific and water management community for stewardship of Earth's water in a warming planet. The first lesson is that by creating synergy between the complementary strengths of satellite remote sensing of surface water temperature and data-driven techniques, it is possible to create a powerful tool for a wide range of scientific inquiries and for development of water management strategies for most places of the world. For example, developing a skillful methodology, such as THORR, to estimate river temperature in regulated rivers and for the narrower reaches regardless of cloud cover conditions, can provide the following benefits for river management. First, such a technique can allow us to explore the potential of the reservoir to act as a "thermometer" for downstream water in the absence of direct in-situ temperature probes. Second, the methodology can help us reconstruct the hydro-thermal behavior of an entire river basin comprising streams of various orders and widths and not just those that are dammed or of water management relevance. Third, the methodology could be used in the few river basins that are not yet dammed or where the impact of reservoirs on stream temperature is minimal. Fourth, the methodology could be used to study the impact of dam addition or removal on the thermal regime of a river. The ability to recreate the long-term thermal behavior of rivers can pave the way for better management of rivers and reservoirs by understanding where and when the potentially more anomalous or harmful effects have taken place and why. Consequently, such reconstruction can lead to the development of more eco-sensitive reservoir operations optimized for maintaining aquatic habitats. Lastly, such history can be the foundation for predicting the hydro-thermal response of reservoirs for various scenarios leading to more proactive reservoir operations that are less harmful to aquatic habitats and public health (such as harmful algal bloom).

The second lesson, which has implications for future design of scalable data-driven techniques is that, with multi-decadal training data spanning 42 years from remote sensing platforms such as the Landsat mission, data-driven techniques can predict the water temperature dynamics with predictability that holds even at locations that are plausibly influenced by upstream reservoir storage dynamics without explicit accounting of reservoir storage dynamics in the training data. In fact, the incorporation of reservoir storage dynamics in the training is found to have a marginal or inconclusive impact on predictability. This lesson has important scalability considerations as reservoir storage change data are often unavailable or inaccessible from dam operators in most places of the world. It appears that the use of satellite-based techniques such as RAT tool to estimate simplistic reservoir

dynamics (i.e., storage change and surface area) would have value in improving predictability of data-driven techniques such as THORR when training data are limited. It also appears that a more comprehensive estimate of reservoir dynamics in terms of additional parameters of inflow and outflow for capturing the thermodynamic heat balance and further improvement of storage change with the newly launched Surface Water and Ocean Topography (SWOT; <http://swot.jpl.nasa.gov>) are research areas worthy of future study for the continual improvement of tools like THORR.

6. Conclusions

Around the world, prediction of surface water temperature of regulated rivers is largely limited by the extent of (or lack of) coverage and the temporal continuity of in-situ temperature measurements. Most regulated river basins around the world do not have water temperature gauging let alone sufficient gauging for streamflow. Thus, the vantage of space using remote sensing for tracking temperature and reservoir operations (such as via RAT) seems to be the only viable option for regulated rivers. Although TIR-based satellite remote sensing provides an opportunity to observe and assess water bodies, cloud cover and the limited spatial resolution of TIR sensors result in gaps in satellite water temperature observations. We systematically developed a data-driven approach based primarily on satellite TIR data to reconstruct a continuous spatiotemporal water temperature time series of rivers. Given the impacts of thermal stratification by dams on rivers, we also explored the value of including reservoir storage dynamics information in improving water temperature estimation downstream of dams.

Among the data-driven techniques explored, the RFR (Random Forest Regressor) was found to be the most robust and accurate for the intended goal of creating continuous temperature estimates. In our study, the RFR showed a superior performance to the LR and ANN approaches in estimating water temperature. The RFR model was used to reconstruct the history of hydro-thermal behavior of reaches within the Columbia River Basin based on which long-term spatiotemporal patterns were identified along the Columbia River. This approach also provided a close-up insight into the long-term upstream and downstream patterns in the Grand Coulee Dam. Beyond the results of this study, the approach presented in this study can serve as a frontier to unlocking and demystifying hydro-thermal dynamics of regulated rivers at a broader scale regardless of the availability of in-situ water temperature records.

Finally, we developed an interactive online tool called the Thermal History of Regulated Rivers (THORR). We present THORR for public access at <https://depts.washington.edu/saswe/thorr/> to help readers visualize the answer to our key research question of the study in the form of a tangible end product. To recall our key research question posed in our study was as follows:

Given the non-exhaustive set of potential drivers and geophysical factors that influence stream temperature, particularly those related to reservoir operations, is it possible to reconstruct the hydro-thermal behavior of regulated rivers in a spatiotemporal continuum?

We believe we have answered this question in the affirmative, albeit with some caveats. Our study and the developed tool THORR are not without limitations. The first limitation to keep in mind is that the water temperature in question pertains to the surface or skin temperature of the water as that is what the satellite TIR data is able to capture. The temperature reconstructed here should not be mistaken as a depth-averaged temperature. The results we show may often not be indicative of the thermal regime near the river bed. Also, to improve the global performance of our model across dams of varying depths, we need to include training with sufficient data that covers all reservoir depth scenarios ranging from shallow (<10 m) to deep (>10m) dams and different climate classes. The inclusion of reservoir fluxes (inflow and outflow) with reservoir storage change data for constraining the heat balance may improve the performance of THORR further. We hope to report improvements to our data-driven technique of THORR in a future study in the spirit of open and collaborative science to facilitate the study of water quantity and thermal quality simultaneously for better water management in ungauged and regulated river basins.

Data Availability Statement

The data and code used in this study are available at <https://doi.org/10.5281/zenodo.10246699> (George Darkwah, 2023). The RAT 3.0 software with all documentation, data sources and tutorials is available at <http://ratdocs.com>.

io and <https://www.satellitedams.net>. In-situ temperature data are public-domain and available as follows: USGS-
<https://dashboard.waterdata.usgs.gov>; USBR- <https://www.usbr.gov/pn/hydromet/arcread.html>; CBR-DART:
https://www.cbr.washington.edu/dart/query/river_daily.

Acknowledgments

This study was supported by a grant from NASA Applied Science Water program (80NSSC22K0918) awarded to the corresponding author (Faisal Hossain). Generous support from Columbia River Intertribal Fish Commission (CRITFC) in the form of engagement, feedback and potential decision-making needs is also acknowledged. Authors Gordon Holtgrieve and Faisal Hossain also acknowledge the support from National Science Foundation Graduate Training program called Future Rivers awarded to the University of Washington.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (/OSDI) 16* (pp. 265–283). Retrieved from <https://www.tensorflow.org/>
- Ahmad, S. K., Hossain, F., Holtgrieve, G. W., Pavelsky, T., & Galelli, S. (2021). Predicting the likely thermal impact of current and future dams around the world. *Earth's Future*, 9(10). <https://doi.org/10.1029/2020EF001916>
- Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, 361(6402), 585–588. <https://doi.org/10.1126/science.aat0636>
- Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5(1), 180214. <https://doi.org/10.1038/sdata.2018.214>
- Bonnema, M., Hossain, F., Nijssen, B., & Holtgrieve, G. (2020). Hydropower's hidden transformation of rivers in the Mekong. *Environmental Research Letters*, 15(4), 044017. <https://doi.org/10.1088/1748-9326/ab763d>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- 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>
- Cherkauer, K. A., Burges, S. J., Handcock, R. N., Kay, J. E., Kampf, S. K., & Gillespie, A. R. (2005). Assessing satellite-based and aircraft-based thermal infrared remote sensing for monitoring pacific northwest river temperature1. *JAWRA Journal of the American Water Resources Association*, 41(5), 1149–1159. <https://doi.org/10.1111/j.1752-1688.2005.tb03790.x>
- Cristea, N. C., & Burges, S. J. (2009). Use of thermal infrared imagery to complement monitoring and modeling of spatial stream temperatures. *Journal of Hydrologic Engineering*, 14(10), 1080–1090. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000072](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000072)
- Cristea, N. C., & Burges, S. J. (2010). An assessment of the current and future thermal regimes of three streams located in the Wenatchee River basin, Washington State: Some implications for regional river basin systems. *Climatic Change*, 102(3–4), 493–520. <https://doi.org/10.1007/s10584-009-9700-5>
- Das, P., Hossain, F., Khan, S., Biswas, N. K., Lee, H., Piman, T., et al. (2022). Reservoir Assessment Tool 2.0: Stakeholder driven improvements to satellite remote sensing based reservoir monitoring. *Environmental Modelling and Software*, 157, 105533. <https://doi.org/10.1016/j.envsoft.2022.105533>
- Duan, S.-W., & Kaushal, S. S. (2013). Warming increases carbon and nutrient fluxes from sediments in streams across land use. *Biogeosciences*, 10(2), 1193–1207. <https://doi.org/10.5194/bg-10-1193-2013>
- Gaube, P., Chickadel, C. C., Branch, R., & Jessup, A. (2019). Satellite observations of SST-induced wind speed perturbation at the oceanic submesoscale. *Geophysical Research Letters*, 46(5), 2690–2695. <https://doi.org/10.1029/2018GL080807>
- George, D. (2023). UW-SASWE/THORR: Initial release (v1.0.0) [Dataset]. Zenodo. <https://doi.org/10.5281/ZENODO.10246699>
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2017.06.031>
- Grasby, S. E., & Hutcheon, I. (2001). Controls on the distribution of thermal springs in the southern Canadian Cordillera. *Canadian Journal of Earth Sciences*, 38(3), 427–440. <https://doi.org/10.1139/e00-091>
- 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>
- Halverson, G. H., Lee, C. M., Hestir, E. L., Hulley, G. C., Cawse-Nicholson, K., Hook, S. J., et al. (2022). Decline in thermal habitat conditions for the endangered delta smelt as seen from Landsat satellites (1985–2019). *Environmental Science and Technology*, 56(1), 185–193. <https://doi.org/10.1021/acs.est.1c02837>
- Hamman, J. J., Nijssen, B., Bohn, T. J., Gergel, D. R., & Mao, Y. (2018). The Variable Infiltration Capacity model version 5 (VIC-5): Infrastructure improvements for new applications and reproducibility. *Geoscientific Model Development*, 11(8), 3481–3496. <https://doi.org/10.5194/gmd-11-3481-2018>
- Handcock, R. N., Gillespie, A. R., Cherkauer, K. A., Kay, J. E., Burges, S. J., & Kampf, S. K. (2006). Accuracy and uncertainty of thermal-infrared remote sensing of stream temperatures at multiple spatial scales. *Remote Sensing of Environment*, 100(4), 427–440. <https://doi.org/10.1016/j.rse.2005.07.007>
- Handcock, R. N., Torgersen, C. E., Cherkauer, K. A., Gillespie, A. R., Tockner, K., Faux, R. N., & Tan, J. (2012). Thermal infrared remote sensing of water temperature in riverine landscapes. In *Fluvial remote sensing for science and management* (pp. 85–113). <https://doi.org/10.1002/9781119940791.ch5>
- Harvey, R., Lye, L., Khan, A., & Paterson, R. (2011). The influence of air temperature on water temperature and the concentration of dissolved oxygen in newfoundland rivers. *Canadian Water Resources Journal*, 36(2), 171–192. <https://doi.org/10.4296/cwrj3602849>
- Heck, M. P., Schultz, L. D., Hockman-Wert, D., Dinger, E. C., & Dunham, J. B. (2018). Monitoring stream temperatures—A guide for non-specialists. *Techniques and Methods*. <https://doi.org/10.3133/tm3A25>
- Hester, E. T., & Doyle, M. W. (2011). Human impacts to river temperature and their effects on biological processes: A quantitative Synthesis1. *JAWRA Journal of the American Water Resources Association*, 47(3), 571–587. <https://doi.org/10.1111/j.1752-1688.2011.00525.x>
- Jimenez-Munoz, J. C., Cristobal, J., Sobrino, J. A., Soria, G., Ninyerola, M., Pons, X., & Pons, X. (2009). Revision of the single-channel algorithm for land surface temperature retrieval from Landsat thermal-infrared data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(1), 339–349. <https://doi.org/10.1109/TGRS.2008.2007125>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. In *3rd international conference on learning representations, ICLR 2015 - conference track proceedings* (pp. 1–15).
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 264–277. <https://doi.org/10.1016/j.jhydrol.2012.01.011>
- Lalot, E., Curie, F., Wawrzyniak, V., Baratelli, F., Schombergk, S., Flipo, N., et al. (2015). Quantification of the contribution of the Beauce groundwater aquifer to the discharge of the Loire River using thermal infrared satellite imaging. *Hydrology and Earth System Sciences*, 19(11), 4479–4492. <https://doi.org/10.5194/hess-19-4479-2015>

- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., & Talwalkar, A. (2016). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18, 1–52.
- Ling, F., Foody, G. M., Du, H., Ban, X., Li, X., Zhang, Y., & Du, Y. (2017). Monitoring thermal pollution in rivers downstream of dams with Landsat ETM+ thermal infrared images. *Remote Sensing*, 9(11), 11. <https://doi.org/10.3390/rs9111175>
- Martí-Cardona, B., Prats, J., & Niclòs, R. (2019). Enhancing the retrieval of stream surface temperature from Landsat data. *Remote Sensing of Environment*, 224, 182–191. ISSN 0034-4257. <https://doi.org/10.1016/j.rse.2019.02.007>
- Minnett, P. J., Alvera-Azcárate, A., Chin, T. M., Corlett, G. K., Gentemann, C. L., Karagali, I., et al. (2019). Half a century of satellite remote sensing of sea-surface temperature. *Remote Sensing of Environment*, 233, 111366. <https://doi.org/10.1016/j.rse.2019.111366>
- Minocha, S., Hossain, F., Das, P., Suresh, S., Khan, S., Darkwah, G., et al. (2023). Reservoir assessment tool version 3.0: A scalable and user-friendly software platform to mobilize the global water management community. *Geoscientific Model Development Discussions*, 2023, 1–23. <https://doi.org/10.5194/gmd-2023-130>
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Murphy, R. D., Hagan, J. A., Harris, B. P., Sethi, S. A., Scott Smeltz, T., & Restrepo, F. (2021). Can Landsat thermal imagery and environmental data accurately estimate water temperatures in small streams? *Journal of Fish and Wildlife Management*, 12(1), 12–26. <https://doi.org/10.3996/JFWM-20-048>
- Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Olden, J. D., & Naiman, R. J. (2010). Incorporating thermal regimes into environmental flows assessments: Modifying dam operations to restore freshwater ecosystem integrity. *Freshwater Biology*, 55(1), 86–107. <https://doi.org/10.1111/j.1365-2427.2009.02179.x>
- Ouellet, V., St-Hilaire, A., Dugdale, S. J., Hannah, D. M., Krause, S., & Proulx-Ouellet, S. (2020). River temperature research and practice: Recent challenges and emerging opportunities for managing thermal habitat conditions in stream ecosystems. *Science of the Total Environment*, 736, 139679. <https://doi.org/10.1016/j.scitotenv.2020.139679>
- Petersen, J. H., & Kitchell, J. F. (2001). Climate regimes and water temperature changes in the Columbia River: Bioenergetic implications for predators of juvenile salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(9), 1831–1841. <https://doi.org/10.1139/f01-111>
- Shi, N., Liu, X., & Guan, Y. (2010). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *3rd international symposium on intelligent information technology and security informatics, IIITSI 2010* (pp. 63–67). <https://doi.org/10.1109/IIITSI.2010.74>
- Torgersen, C. E., Faux, R. N., McIntosh, B. A., Poage, N. J., & Norton, D. J. (2001). Airborne thermal remote sensing for water temperature assessment in rivers and streams. *Remote Sensing of Environment*, 76(3), 386–398. [https://doi.org/10.1016/S0034-4257\(01\)00186-9](https://doi.org/10.1016/S0034-4257(01)00186-9)
- Wang, L., Xin, Z., Xu, B., Chen, X., Zheng, Y., Zhang, J., & Zhang, C. (2023). Impacts of reservoir operation on hydrodynamic and thermal conditions in a large deep reservoir and the potential ecological impacts. *Journal of Water Resources Planning and Management*, 149(7), 04023028. <https://doi.org/10.1061/JWRMD5.WRENG-568>
- Wawrzyniak, V., Piégay, H., & Poirel, A. (2012). Longitudinal and temporal thermal patterns of the French Rhône River using Landsat ETM+ thermal infrared images. *Aquatic Sciences*, 74(3), 405–414. <https://doi.org/10.1007/s00027-011-0235-2>
- Worrall, F., Howden, N. J. K., Burt, T. P., & Hannah, D. M. (2022). River water temperature demonstrates resistance to long-term air temperature change. *Hydrological Processes*, 36(11), e14732. <https://doi.org/10.1002/hyp.14732>
- Yearsley, J. R. (2009). A semi-Lagrangian water temperature model for advection-dominated river systems. *Water Resources Research*, 45(12). <https://doi.org/10.1029/2008WR007629>
- 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.1467-9868.2005.00503.x>

