

Syracuse University

SURFACE at Syracuse University

Dissertations - ALL

SURFACE at Syracuse University

5-14-2023

Stream Temperature as a Tracer of Interactions Amongst Hydrological Processes, Atmospheric Exchange, and Human Activity

Jeffrey Wade
Syracuse University

Follow this and additional works at: <https://surface.syr.edu/etd>

Recommended Citation

Wade, Jeffrey, "Stream Temperature as a Tracer of Interactions Amongst Hydrological Processes, Atmospheric Exchange, and Human Activity" (2023). *Dissertations - ALL*. 1709.
<https://surface.syr.edu/etd/1709>

This Dissertation is brought to you for free and open access by the SURFACE at Syracuse University at SURFACE at Syracuse University. It has been accepted for inclusion in Dissertations - ALL by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.

Abstract

The water temperature of global river networks (also referred to as ‘stream temperature’ or ‘river temperature’) is an influential control on numerous aspects of water quality and riverine ecology, impacting rates of solute processing, dissolved oxygen content, and habitat viability for aquatic ecosystems. River water temperatures arise from the complex interplay of hydrological processes, atmospheric forcings, anthropogenic disturbances, making river thermal regimes challenging to understand and predict at the reach, regional, and global scale. In the absence of widespread water temperature observations, models are commonly used to simulate aspects of water temperature variability by integrating the influence of basin-specific controls and heat fluxes into and out of river systems. In addition to their role as a critical water quality parameter, water temperatures can also be leveraged as a practical tool to probe hydrologic interactions between stream channels and the underlying subsurface. This dissertation explores three diverse applications of water temperature modeling: 1) tracing groundwater-surface water interactions around stream restoration structures using water temperature observations; 2) leveraging machine learning to infer continental-scale drivers of river thermal behavior; and 3) predicting water temperatures at high spatial and temporal resolutions with coupled temperature-hydrologic models.

The first chapter of this dissertation uses water temperature heat tracing methods, in combination with other field observations, to characterize hyporheic exchange induced by beaver dam analogue restoration structures. Beaver dam analogues are process-based restoration structures designed to mimic the effects of natural beaver dams and stabilize degraded and incised river reaches. Despite their frequent application, the influence of these structures on groundwater-surface water hydrology remains unclear. Vertical heat tracing, measurements of

hydraulic head, and analyses of porewater biogeochemistry were used to investigate hydrologic behavior associated with three beaver dam analogues installed on Red Canyon Creek, WY, USA. These analyses demonstrated that while the restoration structures had a negligible effect on overall stream chemistry, beaver dam analogues were capable of producing heterogeneous and localized regions of hyporheic exchange. These results highlight the effectiveness of using water temperatures to trace vertical heat flow and related groundwater-surface water interactions in tandem with other field-based observations.

Given the demonstrated impacts of water temperatures on river water quality, it is critical to better understand how the dominant controls on river thermal regimes vary in time and across broad spatial scales in order to design more effective watershed management strategies. Machine learning models are well suited to this objective, as they can generate accurate predictions of environmental processes while revealing key interactions between variables in large datasets. In the second chapter of this dissertation, a suite of random forest models was used to predict metrics of river temperature variability across the conterminous US using watershed characteristics extracted from a publicly-available dataset. Variable importance metrics were then interpreted to infer the underlying controls on river temperatures. Regional climate forcings tended to most closely control river temperature magnitude, though those forcings were mediated by the influence of hydrological processes, watershed characteristics, and anthropogenic disturbances. Results from the random forest models underscored the challenge in predicting aspects of water temperature variability at continental scales, particularly when river thermal regimes are disrupted by dams and reservoirs. The presented machine learning approach to river temperature prediction illustrates how large environmental datasets can be leveraged to provide discerning insight into the drivers of hydrologic and thermal processes.

To supplement predictions of water temperatures at point locations along the river network, deterministic energy balance models are often applied to provide spatially distributed and temporally continuous water temperature simulations. Deterministic water temperature models function by quantifying radiative, turbulent, and advective heat fluxes into and out of a river at the air-water and water-streambed interfaces. While such water temperature models are often applied within single catchments, many watershed management applications require high resolution predictions of temperatures at a broader spatial extent. The third chapter of this dissertation focuses on the development of a coupled hydrological-water temperature energy balance model in a single test basin with the potential for expansion to the full conterminous US. Using forcings and outputs from the National Water Model, a continental-scale hydrologic model implemented by NOAA and NCAR, several water temperature model configurations of increasing complexity were tested to evaluate tradeoffs between performance and computational efficiency. Modeling efforts demonstrated that the National Water Model can be effectively leveraged to provide high-quality predictions of hourly water temperatures throughout a river network, though critical challenges remain in expanding coupled water temperature models to continental scales.

STREAM TEMPERATURE AS A TRACER OF INTERACTIONS AMONGST
HYDROLOGICAL PROCESSES, ATMOSPHERIC EXCHANGE, AND HUMAN ACTIVITY

by

Jeffrey Scott Wade

B.S., University of Wisconsin-Madison, 2019

Dissertation

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Earth Sciences.

Syracuse University
May 2023

Copyright © Jeffrey Scott Wade 2023

All rights reserved

Acknowledgements

This material is based upon work supported by the National Science Foundation under grant DGE-1449617, including an INTERN supplement. Financial support for this work was also provided by the Syracuse University Research Excellence Doctoral Funding (REDF) Program, the Syracuse University Department of Earth and Environmental Sciences Vincent E. McKelvey Scholarship Award, and the National Oceanic and Atmospheric Administration William M. Lapenta Student Internship Program. The Nature Conservancy of Wyoming provided support for site access, training, and construction of the beaver dam analogues at Red Canyon Creek, WY.

I would like to thank Dr. Laura Lautz, my initial advisor, for seeing my potential and guiding me in my first steps as a scientist. I would also like to thank my advisor Dr. Christa Kelleher for her unceasing optimism, encouragement, and dedication to her students. Our weekly meetings were the highlight of my time at Syracuse University. Thank you to my committee members, Drs. Stephen Shaw, Samuel Tuttle, and Tao Wen, for providing guidance and insightful feedback on my dissertation. Thank you to my collaborators, Drs. David Hannah, Adam Ward, Rebecca Schewe, Philippe Vidon, and Barret Kurylyk, for their invaluable assistance in guiding the directions of my research. I would also like to thank Julianne Davis, Julio Beltran and Casey Pearce for their support of field work at Red Canyon Creek. To my closest friends, Eliza Hurst, McKenzie Brannon, Carson Ruffledt, Briana Edgerton, Lachlan Wright, Courtney Gammon, and Harold Jones, thank you for supporting me during my years in Syracuse and tagging along on more camping trips, hikes, and bike rides than I can count. To my family, I am eternally grateful for the love and encouragement I've received and the opportunity I've been given to pursue a Ph.D.

Table of Contents

Abstract	i
Acknowledgements	vi
Table of Contents	vii
List of Figures	x
List of Tables	xiii
Chapter 1: Beaver Dam Analogues Drive Heterogeneous Groundwater-Surface Water Interactions	1
Abstract	2
1. Introduction	2
2. Methods	5
2.1. Study Site.....	5
2.2. Field Data Collection.....	7
2.3. Vertical Heat Tracing	8
2.4. Water Chemistry Principal Component Analysis.....	9
3. Results	10
3.1. Vertical Streambed Fluxes	10
3.2. Streambed Porewater Chemistry	11
3.3. Classification of Water Types	13
4. Discussion	15
4.1. How do BDAs affect patterns of hyporheic flow and biogeochemical cycling?	15
4.2. Do BDAs function similarly to natural beaver dams?.....	20
4.3. What factors influence the function of BDAs?	21
5. Conclusion.....	23
Figures	25
Tables	32
References	34
Chapter 2: Machine learning unravels controls on river water temperature regime dynamics	42
Abstract	43
1. Introduction	43
2. Materials and Methods	46

2.1. Site Selection	47
2.2. Stream Temperature Signatures	47
2.3. Site Predictors.....	48
2.4. Applying Random Forest Models to Infer Drivers of Stream Temperature Regimes ...	49
2.4.1. Training Random Forest Models	49
2.4.2. Identifying Influential Predictors of Stream Temperature Signatures	50
3. Results	52
3.1. Observed Spatial and Temporal Variability in Stream Thermal Regimes	52
3.2. RF Model Error	53
3.3. Exploring Inferred Controls on Stream Temperature Variability	53
4. Discussion	56
4.1. Revealing Relative Controls on Stream Temperature Variability.....	56
4.2. Anthropogenic Influence on Stream Thermal Regimes	60
4.3. Seasonal Variability in Key Controls of Stream Temperatures	63
4.4. Implications for Watershed Management	64
5. Conclusions	65
Figures.....	68
Tables	76
References	80
Chapter 3: Incorporating Physically-based Water Temperature Predictions into the National Water Model Framework: Application to a Western Forested Headwater Catchment.....	91
Abstract	92
1. Introduction	93
2. Methods.....	96
2.1. Study Site: H.J. Andrews Experimental Forest	96
2.2. Model Data	97
2.2.1. National Water Model Retrospective v2.1.....	97
2.3. Modeling Approach.....	99
2.3.1. Model Resolution.....	99
2.3.2. Computation of Water Temperatures.....	99
2.3.3. Heat Transfer Equations	103
2.3.4. Hydrologic Heat Fluxes	103

2.3.5. Estimation of Unknown Inflow Temperatures	105
2.3.6. Approximating the Thermal Effects of Hyporheic Exchange	107
2.3.7. Estimating Riparian Shading in the Absence of On-site Observations	108
2.3.8. Assessing Model Error.....	110
2.4. Sequential Evaluation of Model Configurations	111
3. Results	113
3.1. Calibrated Models	113
3.1.1. M1: Variable Groundwater Inflow Temperatures	113
3.1.2. M2: Variable Groundwater Inflow Rate	114
3.1.3. M3: Conceptual Hyporheic Zone.....	115
3.1.4. M4: Variable Groundwater Inflow Rate and Conceptual Hyporheic Zone	115
3.2. Optimal Calibrated Parameters.....	116
4. Discussion	117
4.1. Evaluating Performance of Water Temperature Model Configurations	117
4.2. Strategies for Constraining Uncertain Inputs	120
4.2.1. Estimating Groundwater Temperatures Key to Accurate Water Temperature Predictions.....	121
4.2.2. Is a Conceptual Hyporheic Zone Needed?.....	123
4.3. Challenges and Opportunities in Expanding from the Catchment to Continental Scale	126
5. Conclusion.....	128
Figures	129
Tables	136
References	139
Appendix A: Supplemental Information for Chapter 2	152
Figures	152
Appendix B: Supplemental Information for Chapter 3	158
1. Water Temperature Modeling Equations	158
1.1. Channel Geometry	158
1.2. Heat Transfer Equations	159
References	167
Curriculum Vita.....	168

List of Figures

Chapter 1

Figure 1. Aerial imagery of beaver dam analogue study reach, Red Canyon Creek, WY	25
Figure 2. Site maps of instrumentation surrounding beaver dam analogues.	26
Figure 3. Boxplots of vertical water flux derived from vertical heat tracing.....	27
Figure 4. Spatial interpolated maps of vertical hydraulic gradients	28
Figure 5. Bivariate plots of major ion chemistry collected from streambed porewater, stream water, and groundwater.....	29
Figure 6. Redox score (PC1) and nitrate score (PC2) results from biogeochemical porewater principal component analysis.....	30
Figure 7. Spatial distribution of porewater type classifications based on biogeochemical principal component analysis	31

Chapter 2

Figure 1. Graphical representation of steps used to assess the dominant drivers of stream temperature regimes.....	68
Figure 2. Site map of stream temperature gages used in random forest modeling.	69
Figure 3. Conceptual depiction of variables used to predict two metrics of water temperature variability	70
Figure 4. Spatial and temporal variability of selected stream temperature metrics.....	71
Figure 5. Variable importance heatmaps for maximum stream temperatures at all sites and sites without dams.....	72
Figure 6. Variables importance heatmaps for thermal sensitivity at all sites and sites without	

dams.....	73
Figure 7. Cumulative importance of predictor categories through time.....	74
Figure 8. Variable importance heatmaps for maximum stream temperatures in three selected hydrologic regions	75
 Chapter 3	
Figure 1. Site map of model river reaches and water temperature gages in H.J. Andrews Experimental Forest, OR, USA.....	129
Figure 2. Modeled atmospheric, radiative, and hydrologic heat fluxes and associated National Water Model input data	130
Figure 3. Estimated groundwater inflow temperatures using tuned source depth approach.	131
Figure 4. Simulated water temperature RMSE across model calibration runs.....	132
Figure 5. Comparison of model configuration performance across three metrics of error.....	133
Figure 6. Envelopes of well-calibrated water temperature simulations at headwater gage.....	134
Figure 7. Envelopes of well-calibrated water temperature simulations at the outlet gage	135
 Appendix A	
Figure 1. Variable importance heatmaps for thermal sensitivity in three selected hydrologic regions.....	152
Figure 2. Variable importance heatmaps for maximum stream temperatures and thermal sensitivity at sites with dams.....	153
Figure 3. Random forest model error in predicting maximum stream temperatures and thermal sensitivity at all sites	154
Figure 4. Random forest model error in predicting maximum stream temperatures and thermal sensitivity at sites without dams	155

Figure 5. Random forest model error in predicting maximum stream temperatures at sites in three selected hydrologic regions..... 156

Figure 6. Random forest model error in predicting thermal sensitivity at sites in three selected hydrologic regions 157

List of Tables

Chapter 1

Table 1. Sediment parameters used in vertical flux heat transport modeling 32

Table 2. Loadings, eigenvalues, and explained variance of biogeochemical porewater principal component analysis 33

Chapter 2

Table 3. Test set errors of monthly random forest models 76

Table 4. Summary statistics of maximum stream temperatures and thermal sensitivity. 77

Table 5. Descriptions of predictor variables used in random forest models 78

Chapter 3

Table 6. Summary of water temperature model formulations 136

Table 7. Parameter definitions and tuning ranges for model configurations 137

Table 8. Optimal parameter values for best model calibration runs. 138

Chapter 1: Beaver Dam Analogues Drive Heterogeneous Groundwater-Surface Water Interactions

This chapter has been published as:

Wade, J., Lautz, L., Kelleher, C., Vidon, P., Davis, J., Beltran, J., & Pearce, C. (2020). Beaver dam analogues drive heterogeneous groundwater–surface water interactions.

Hydrological Processes, 34(26). <https://doi.org/10.1002/hyp.13947>

Abstract

Beaver dam analogues (BDAs) are a cost-effective stream restoration approach that leverages the recognized environmental benefits of natural beaver dams on channel stability and local hydrology. While natural beaver dams are known to exert considerable influence on the hydrologic conditions of a stream system by mediating geomorphic processes, nutrient cycling, and groundwater-surface water interactions, the impacts of beaver-derived restoration methods on groundwater-surface water exchange are poorly characterized. To address this deficit, we monitored hyporheic exchange fluxes and streambed porewater biogeochemistry across a sequence of BDAs installed along a central Wyoming stream during the summer of 2019. Streambed fluxes were quantified by heat tracing methods and vertical hydraulic gradients. Biogeochemical activity was evaluated using major ion porewater chemistry and principal component analysis (PCA). Vertical fluxes of approximately 1.0 m/day were observed around the BDAs, as was the development of spatially heterogeneous zones of nitrate production, groundwater upwelling, and anaerobic reduction. Strong contrasts in hyporheic zone processes were observed across BDAs of differing sizes. This suggests that structures may function with size-dependent behavior, only altering groundwater-surface water interactions after a threshold hydraulic step height is exceeded. Patterns of hyporheic exchange and biogeochemical cycling around the studied BDAs resemble those around natural beaver dams, suggesting that BDAs may provide comparable benefits to channel complexity and near-stream function over a one-year period.

1. Introduction

In recent years, stream restoration techniques have increasingly been applied to degraded riparian systems to reduce reach-scale homogeneity and promote hyporheic exchange (Bernhardt

et al., 2005; Kasahara & Hill, 2006; Zimmer & Lautz, 2015). Shifts in land-use practices and livestock overgrazing in the western United States have reduced vegetation cover and accelerated surface runoff, leading to downcutting and stream incision (Chaney et al., 1993). The deterioration of streams in the western US has led to the development and application of a variety of stream restoration techniques that include the installation of log dams, debris dams, and cross-vanes (Fanelli & Lautz, 2008; Lautz et al., 2006; Lautz & Fanelli, 2008). These engineered restoration approaches generate bedforms that induce hydraulic head differentials, driving surface waters into the biologically active hyporheic zone (Brunke & Gonser, 1997; Kasahara & Hill, 2006). With increased exchange between surface water and groundwater, the hyporheic zone of a restored stream becomes a critical region of solute processing, contributing to improvements in near-stream hydrological and ecological function (Lautz & Fanelli, 2008).

Although engineered restoration structures alter hyporheic exchange (Daniluk et al., 2013; Fanelli & Lautz, 2008; Kasahara & Hill, 2006; Zimmer & Lautz, 2015), land managers in the western US have sought a less expensive and more dynamic solution to address stream homogenization and incision. New restoration approaches have taken inspiration from natural dam designs of the North American beaver, *Castor canadensis*. Aptly referred to as “ecosystem engineers”, the North American beaver drastically modifies riparian systems by building channel-spanning dams that impound water, elevate water tables, reduce discharge velocities, and trap organic matter and sediment (Naiman et al., 1988; Westbrook et al., 2006). These riparian zone alterations are particularly beneficial to semi-arid landscapes, where the hydrologic effects of beaver dams can provide resiliency to near-stream wetlands and vegetation threatened by drought (Fairfax & Small, 2018). Much like engineered restoration structures, beaver dams also create streambed pressure gradients that enhance hyporheic exchange and subsurface

biogeochemical activity (Briggs et al., 2012, 2013). However, by contrast, beaver dams are inherently transient structures. Their permeable nature, coupled with frequent breaching events, promotes complexity in stream morphology and minimizes channelization (Pilliod et al., 2018; Pollock et al., 2014; Wegener et al., 2017). Despite the recognized hydrologic benefits of beaver dams, the historical extirpation of beaver populations in the western US has left riparian systems at risk of degradation, increased channel erosion, and lowered groundwater tables (Lautz et al., 2019; Naiman et al., 1988).

To leverage the benefits of natural beaver dams in the absence of beaver populations, beaver dam analogues (BDAs) have emerged as a dynamic stream restoration solution in the western US (Lautz et al., 2019; Pilliod et al., 2018; Pollock et al., 2014). BDAs are temporary, semi-permeable, and inexpensive channel-spanning structures constructed with the intent of mimicking the effects of beaver dams on riparian systems (Pollock et al., 2014, 2015). A BDA consists of several wooden posts pounded vertically into the streambed, willow branches woven between these posts, and cobble fill material placed upstream of the structure (Pollock et al., 2015). Grass, sand, and mud are also packed into the woven willow branch fabric to improve the structure's ability to retain water. While a BDA requires consistent maintenance to sustain its structural integrity, the construction process is cost-effective and requires little machinery in comparison to engineered restoration designs.

The primary hydrologic goal of a BDA is to create an area of slow-moving water upstream of the dam (Shahverdian et al., 2019). The impoundment of water behind a BDA raises the local water table, reconnecting an incised stream to its floodplain (Pollock et al., 2015). BDAs are most effective when installed in sequence because, much like natural beaver dams, interactions between the structures amplify the net effect on a riparian system (Pollock et al.,

2015; Shahverdian et al., 2019). Unlike static engineered restoration structures, BDAs are intentionally dynamic features intended to allow a channel to return to its natural heterogeneous state (Pollock et al., 2014). Generally, BDAs are constructed with the geomorphic goal of inducing aggradation and limiting incision, but they may also contribute ancillary benefits such as increased hyporheic flow and subsurface biogeochemical cycling.

While the popularity of beaver-inspired restoration approaches in the western United States has grown swiftly in recent years (Lautz et al., 2019; Pilliod et al., 2018), fundamental questions about their effect on groundwater-surface water interactions, including hyporheic exchange, remain unanswered. Beaver dam analogues are inherently distinct from both engineered restoration structures and natural beaver dams and therefore may alter stream function in different ways. For example, it is unclear if BDAs can induce similar magnitudes of hyporheic exchange fluxes relative to other channel-spanning structures given that they may be present for only short, single-year time scales. To address this gap in knowledge, this study seeks to (1) characterize patterns of hyporheic exchange and biogeochemical redox zonation associated with BDAs and (2) examine the causes of variability between the hydrologic effects of BDAs at the reach scale.

2. Methods

2.1. Study Site

Red Canyon Creek is a third-order, meandering stream located near Lander, WY, east of the Wind River Range (Figure 1). The magnitude and frequency of peak streamflow events in this semi-arid region are generally controlled by spring snowmelt, resulting in large variance between average monthly stream discharges (Lautz et al., 2006). Streamflow rates at the time of this study were typical of summer baseflow conditions, at around 150 to 200 L/s (Jin et al.,

2012). The Red Canyon Creek watershed is underlain by the Permian Phosphoria Formation and the gypsum-rich Triassic Chugwater Formation, which source the silt and sandy gravels that make up the streambed substrate and alluvial valley floor. A previous study of Red Canyon Creek estimated that the streambed has a hydraulic conductivity between 10^{-5} and 10^{-6} m/s (Lautz et al., 2010). The majority of the 80 km² Red Canyon Creek watershed is sustainably managed as a cattle ranch by The Nature Conservancy (TNC) of Wyoming. Red Canyon Creek, like many other streams in the western US, has been degraded by stream incision, largely as a result of livestock activity. In an effort to counteract accelerated erosive processes, TNC has employed a variety of in-stream restoration structures including log dams and debris dams to disperse the focused energy of incisive flows. Restoration structures installed by TNC within the Red Canyon Creek watershed have been extensively studied to evaluate their ability to alter stream morphology, hyporheic exchange, and porewater geochemistry (Fanelli & Lautz, 2008; Jin et al., 2012; Lautz et al., 2006). The Red Canyon Creek watershed is part of a larger ecoregion in the western US that has historically sustained beaver populations (Gibson & Olden, 2014). As such, the watershed has been the subject of numerous studies exploring the hydrological and geochemical effects of natural beaver dams on riparian function and stability (Briggs et al., 2012, 2013, 2014; Jin et al., 2009). While isolated beaver populations still inhabit the upper tributaries of the watershed, few beaver have occupied the lower reaches of Red Canyon Creek since 2015.

To reduce stream incision and promote streambed aggradation, five beaver dam analogues were constructed on Red Canyon Creek from April to August of 2018 as part of a field workshop hosted by the Natural Resources Conservation Service and Utah State University, in collaboration with TNC. The BDA restoration structures (BDA 1-5) were installed along a

~200-m reach of Red Canyon Creek (Figure 1). BDAs 2, 3, and 4 were built in sequence with a spacing of approximately 30 m using best-practice construction techniques involving wooden posts, woven willow branches and other riparian vegetation, and streambed sediment (Figure 2). After their installation, the BDAs formed discontinuous hydraulic steps over the structures. BDA 2 and BDA 3 induced water surface drops of 0.14 m and 0.12 m respectively, while BDA 4 formed a larger water step of 0.40 m. BDAs 1 and 5 were excluded from analysis in this study, as they were built using experimental designs and failed before the summer of 2019. Field data were collected roughly one year after the installation of the BDAs on Red Canyon Creek.

2.2. Field Data Collection

To assess the magnitude and direction of vertical water fluxes in the streambed around the BDAs, 11 vertical temperature profiles were installed to measure the propagation of the diurnal temperature signal of the stream water into the subsurface (Figure 2). Each vertical temperature profile consisted of three ThermoChron iButton temperature loggers (DS1922L, Maxim Integrated Products, Inc., San Jose, California) imbedded into wooden dowels with diameters of 3.2 cm. The iButton temperature loggers have a manufacturer-reported accuracy of $\pm 0.5^{\circ}\text{C}$ and a resolution of 0.0625°C . The vertical profiles were driven into the streambed such that the iButton sensors were positioned at the streambed interface and at depths of 10 cm and 20 cm. Stream water and porewater temperatures were recorded at 10-minute intervals over 188 hours, spanning from July 23 to July 31, 2019.

The streambed surrounding the BDAs was also instrumented with 47 mini-piezometers to allow for the collection of hyporheic porewater samples and the measurement of vertical hydraulic gradients. The mini-piezometers were composed of 1.25 cm diameter polyvinyl chloride (PVC) pipe screened over the bottom 5 cm. The mini-piezometers were inserted into the

streambed such that the screened depth extended from 17.5 cm to 22.5 cm below the bed, allowing for a mean sampling depth of 20 cm. Vertical hydraulic gradients (VHGs) were determined at each mini-piezometer by measuring the difference in water height between the stream surface and the water level in the mini-piezometer (Lautz & Fanelli, 2008).

To collect streambed porewater, the mini-piezometers were purged and water samples were extracted using a syringe and plastic tubing. Porewater samples ($n = 47$) were filtered during extraction using $0.45 \mu\text{m}$ Millex-HN nylon filters and refrigerated in preparation for laboratory analysis. A streamwater sample was collected concurrently to the collection of porewater samples and filtered using the same process. Samples were analyzed for major ions and nutrients using ion chromatography (Dionex ICS-2000, Thermo Fisher Scientific Inc., Waltham, MA) and induced coupled plasma optical emission spectrometry (Perkin Elmer Optima 3300DV, Perkin–Elmer, Waltham, MA).

The bed topography, dam crests, and water surface elevations were surveyed using a Nikon Nivo 5.M total station. A total of 281 elevation points were measured along the study reach. Recorded elevations were used to generate thalweg topography cross-sections via inverse distance weighting interpolation in ArcGIS Pro. Water surface drops over the dams were assessed by calculating the difference in surveyed elevation between the wetted dam crest and the downstream water surface.

2.3. Vertical Heat Tracing

Vertical exchange fluxes were calculated at each temperature profile using the Vertical Fluid Heat Transport Solver (VFLUX 2; Gordon et al., 2012; Irvine et al., 2015) using analytical solutions to the one-dimensional heat transport equation (Hatch et al., 2006). The VFLUX script employs time-series measurements of water temperature to isolate the diurnal signals and

calculate vertical streambed flux using a variety of phase amplitude and lag-based solutions (Gordon et al., 2012). Vertical flux accuracy was optimized using a workflow introduced by Irvine et al. (2017) that leverages the sequential application of heat transport solutions developed by Hatch et al. (2006) and Luce et al. (2013) to minimize error, especially in the case of upwelling conditions (Irvine et al., 2017). The first and last 24 hours of time-series flux estimates were removed to prevent edge effects introduced by the VFLUX filtering process (Hatch et al., 2006). Vertical temperature profile sensor pairs were individually selected at each measurement location to minimize the uncertainty of VFLUX results for varying upwelling and downwelling conditions (Figure 3). The selected sensor pairs produced amplitude ratios and time lags within the expected bounds, ensuring VFLUX converged on real flux values in each location (Gordon et al., 2012). Thermal and physical parameter inputs to VFLUX (Table 1) were adapted from previous heat tracing studies within the watershed (Briggs et al., 2012).

2.4. Water Chemistry Principal Component Analysis

Principal component analysis (PCA) is a widely used linear transformation technique that aims to summarize interdependent variance within a complex set of variables and reduce the dimensionality of a dataset to ease interpretation (Abdi & Williams, 2010; Jolliffe & Cadima, 2016). PCA has been used extensively to classify hyporheic porewaters, particularly in studies concerning the effects of dams and restoration structures on biogeochemical cycling (Briggs et al., 2013; Gordon et al., 2013; Lautz & Fanelli, 2008; Zimmer & Lautz, 2014). A PCA procedure was applied to a set of redox-sensitive ion concentrations (NH_4^+ , NO_3^- , SO_4^{2-} , Mn, Fe) measured in Red Canyon Creek porewaters in order to group porewaters into distinct chemical categories. Prior to the application of PCA, the concentrations of each chemical species were normalized to a mean of 0 and a standard deviation of 1, such that the relative magnitudes of concentrations

were not unduly weighted during linear transformation (Zimmer & Lautz, 2014). First and second principal component scores with accompanying loadings and eigenvalues are reported for 47 porewater samples and one stream water sample.

3. Results

3.1. Vertical Streambed Fluxes

Estimates of vertical flux from heat tracing were calculated at each of the 11 vertical temperature profiles. Over the seven days of temperature measurements, hourly estimates of flux showed negligible variation. Given the stationarity of the flux estimates, median flux rates were used to summarize flux magnitude and direction at each measurement point in order to enable comparison between structures (Figure 3). These median fluxes along the reach ranged from -1.00 m/day (downwards) to 1.25 m/day (upwards). Magnitudes of temperature-derived vertical fluxes were spatially organized as a function of distance from the BDAs (Figure 3). At vertical temperature profiles distant from the dams, median flux values were consistently small ($-0.15 < q < 0.05$ m/day) with no clear directionality. These locations were considered to be outside the range of influence of the BDAs. In general, flux rates increased in magnitude with proximity to the dams in both the upstream and downstream directions. Flux measurements at P2 and P10 immediately upstream of BDAs 2 and 4 had strong downwelling signals, with median flux values of -0.64 and -1.00 m/day, respectively. Median upwelling fluxes of 0.70 and 1.25 m/day were observed at P7 and P11, downstream of BDAs 3 and 4. Flux measurements adjacent to BDA 4 were larger than those observed at the other beaver dam analogues in both the upstream and downstream directions.

The directions of vertical hydraulic gradients (VHG) obtained from direct measurements of hydraulic heads in piezometers generally coincided with the spatial distribution of thermally-

derived exchange fluxes (Figures 3 & 4). VHGs were either neutral or small and negative (downward) at all measurement locations near BDAs 2 and 3. Median vertical hydraulic gradients at BDAs 2 and 3 were -0.08 and -0.09 respectively, equating to head differences of approximately 2 cm between the stream and the streambed. No relationship was observed between VHG and distance from these BDAs. Conversely, gradient measurements at BDA 4 were spatially organized similar to the pattern observed in the thermally-derived vertical fluxes. Upstream of BDA 4, hydraulic gradients were consistently downward and larger in magnitude than those observed at BDAs 2 and 3. VHGs were an order of magnitude larger in a narrow region immediately upstream of BDA 4, with vertical gradients reaching values up to -2.0. This area of large, downward hydraulic gradients extended roughly 3 m upstream of BDA 4. Downstream of BDA 4, positive VHGs of up to 0.1 were observed at the three mini-piezometers closest to the dam. Point measurements of hydraulic gradients and thermally-derived exchange fluxes were not co-located and differed at some locations near BDAs 2 and 3, likely due to small-scale heterogeneity in streambed sediment conductivity. Head gradients are a function of both true downward flux and hydraulic conductivity. Therefore, the magnitude of large head gradients around the BDAs could either reflect significant downwelling, effectively measured by heat tracing, or regions of low hydraulic conductivity.

3.2. Streambed Porewater Chemistry

The collected stream water samples and a set of 33 historical groundwater samples (Lautz & Fanelli, 2008) provided two distinct end-member compositions for comparison to streambed porewater samples (Figure 5). Lautz & Fanelli (2008) collected groundwater samples in 2005 and 2006 from 33 wells spread throughout the meadow immediately south of the Red Canyon Creek study reach. The wells were located between 5 m and 125 m from the BDA reach and

were screened at depths ranging from 0.8 m to 3.7 m (Lautz & Fanelli, 2008). Both the stream water and groundwater were dominated by high concentrations of Ca^{2+} and SO_4^{2-} , attributed to the dissolution of gypsum ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) from the underlying bedrock. The stream and groundwater had sulfate concentrations of 206.4 and 434.7 mg/L, respectively. As the dissolution of gypsum produces Ca^{2+} and SO_4^{2-} ions in a 1:1 molar ratio, porewater samples falling near the gypsum dissolution line in Figure 5a represent direct mixing between in-stream waters and groundwater with little biogeochemical activity. Several streambed porewater samples had deficits in SO_4^{2-} with respect to Ca^{2+} , deviating from the gypsum dissolution line. The sulfate deficit in these samples was ascribed to biogeochemical reduction processes occurring within the hyporheic zone. Conversely, porewaters with elevated concentrations of sulfate in exceedance of the stream water end-member by >50% indicate an interaction with groundwater flow paths. Such conditions occurred exclusively downstream of BDA 4 at three mini-piezometers, where SO_4^{2-} ranged from 305.6 to 445.9 mg/L.

Concentrations of nitrate in the stream water and groundwater were moderately low at 0.16 and 0.18 mg NO_3^- /L, respectively. Figure 5b shows the range of nitrate present in porewater samples and the relationship between NO_3^- and SO_4^{2-} within the streambed. Nitrate production in porewater was observed within a narrow range of SO_4^{2-} concentrations centered around an in-stream value of 206.4 mg/L. In samples with low amounts of NO_3^- (<0.05 mg/L), SO_4^{2-} was often reduced from stream-like concentrations to values approaching zero, producing a characteristic “L-shaped” pattern (Gordon et al., 2013). Deficits in sulfate attributed to reduction processes were also associated with elevated concentrations of manganese (Figure 5c), iron, and ammonium.

3.3. Classification of Water Types

Amongst porewater samples, strong linear correlations between redox-sensitive ions (NH_4^+ , NO_3^- , Mn, Fe, and SO_4^{2-}) were observed. To summarize the relationships between correlated ion concentrations, principal component analysis (PCA) was used to reduce the dimensionality of the dataset into two principal components. Cumulatively, the first two principal components accounted for 78.1% of the total variance between redox-sensitive ions. Additional principal components accounting for less than 10% of total variance were excluded from further analyses. Loadings, eigenvalues, explained variance percentages, and correlations of the first two principal components are presented in Table 2.

The first principal component (PC1) was termed the “redox function” while the second principal component (PC2) was termed the “nitrate function”, following the convention of previous applications of PCA on porewater geochemistry at Red Canyon Creek (Lautz & Fanelli, 2008). PC1 was significantly correlated to all redox ions used in the principal component analysis (Table 2). The “redox function” serves as a strong proxy for summarizing the loss of nitrate and sulfate and the gain of soluble manganese, iron, and ammonia due to biogeochemical reduction in the streambed. While PC1 accounts for considerable variation within nitrate values, PC2 allows for further discrimination of porewaters based on nitrate production and reduction. Of the five ions included in PCA, the “nitrate function” is only significantly correlated to nitrate concentrations (Table 2).

Each of the 47 porewater samples was assigned two scores corresponding to the loading values of the first two principal components and plotted to show variation in “redox” and “nitrate scores” (Figure 6). Samples were classified into four porewater types based on the values of the principal component scores: reducing conditions, nitrate production, groundwater signature, and

stream water signature. Porewater samples within the reducing conditions bin were characterized by high redox scores, corresponding to low nitrate and sulfate concentrations and high manganese, iron, and ammonia concentrations. The reducing conditions classification accounted for the most porewater samples of the four bins ($n = 30$), containing 64% of all samples. The nitrate production class of porewater samples ($n = 11$) was distinguished by elevated nitrate scores, low redox scores, and high concentrations of nitrate. Samples in the groundwater signature bin ($n = 3$) had the lowest redox and nitrate scores, due to their high concentrations of sulfate and their relative lack of nitrate. The mean of previously collected groundwater samples (Lautz & Fanelli, 2008) fell within the bounds of the groundwater signature category. While the porewaters within this classification had chemical signatures most similar to those of groundwater, they generally had lower concentrations of Ca^{2+} and SO_4^{2-} as compared to the groundwater end member. This suggests that these porewaters reflect a mixing of surface water and groundwater, rather than the discharge of pure groundwater. The stream water signature bin ($n = 3$) was defined by stream-like concentrations of redox-sensitive ions and contained the stream water sample in the redox-nitrate space.

The porewater classifications at each mini-piezometer were plotted spatially to explore the distribution of water types at each BDAs (Figure 7). BDAs 2 and 3 were dominated by large regions of reducing conditions, with minor areas of nitrate production upstream and downstream at both dams. At BDA 4, porewater conditions were more heterogeneous and tended away from the widespread streambed reduction seen at the other dam sites. BDA 4 had a region of nitrate production immediately upstream of the dam, followed by a region with a stream water signature further upstream. A strong groundwater signature was present on the downstream side of BDA 4. Stream water and groundwater signatures were exclusively present at BDA 4 and were not

identified at BDAs 2 or 3.

4. Discussion

4.1. How do BDAs affect patterns of hyporheic flow and biogeochemical cycling?

To characterize how the beaver dam analogues altered hyporheic exchange, we first considered the state of the streambed prior to the installation of the BDAs. While this study did not monitor the channel in its undisturbed state, prior work on the same reach of Red Canyon Creek provides a baseline for hydrological conditions. Previous heat tracing results using the same methodology as this study found minimal seepage fluxes (< 0.15 m/day) where BDA 4 is presently located (Lautz et al., 2010). The historical observations on the reach indicate that background exchange fluxes were small in magnitude in the absence of BDAs. This is further confirmed by the observations at monitoring sites most distant from the BDAs (e.g., P1, P4, P5, P8), which had thermally-derived exchange fluxes of less than 0.1 m/day (Figure 3). Hyporheic exchange at these distant sites appears to reflect background conditions, rather than the effects of the BDAs. The presence of soluble manganese and iron and the depletion of nitrate and sulfate at these locations suggest that residence times were sufficiently long to enable the anaerobic reduction of alternate electron acceptors (Lautz & Fanelli, 2008). In a study of neighboring Cherry Creek, Briggs et al. (2013) found that the hyporheic zone remained oxic along flow paths with residence times up to 1 hour in length, after which the streambed tended towards anoxia and reducing conditions. We note that the length of this residence time threshold is highly variable across physical settings and is controlled by the supply of DO and DOC, water temperature, and the hydraulic properties of the hyporheic zone (Zarnetske et al., 2011). Further exploration of residence time thresholds and their biogeochemical effects across diverse settings is presented in Gomez et al. (2012).

If we assume conditions are consistent throughout the Red Canyon Creek watershed, Briggs et al. (2013)'s determination of residence times on Cherry Creek can be applied to the hyporheic zone surrounding the BDAs on Red Canyon Creek as a first order approximation. Using this method at Red Canyon Creek, a 1-hour residence time is equivalent to a vertical flux rate of roughly 1.2 m/day at a depth of 20 cm (Briggs et al., 2013). Streambed fluxes smaller in magnitude than this value will permit microbial reduction of terminal electron acceptors in the streambed. Therefore, the beaver dam analogues must produce hyporheic flow paths with vertical fluxes larger than this threshold to shift regions of the shallow streambed from anaerobic reduction to oxic nitrification.

At BDAs 2 and 3, regarded as smaller dams due to their small hydraulic steps, principal component analysis of porewaters indicated that the dams did not produce hyporheic fluxes of sufficient magnitude to overcome the natural background of reduction in the streambed (Figure 7). The presence of isolated regions of nitrification shows that in some locations, hyporheic fluxes were sufficiently strong to oxygenate the streambed. However, the spatially heterogeneous and limited extent of these areas suggests that their fluxes were more likely controlled by sub-meter scale variance in streambed conductivity rather than by the influence of the BDAs. Thermally-modeled fluxes show that BDAs 2 and 3 did induce exchange in a 1 m region upstream and downstream of the dams, but these flows did not exceed the 0.5 m/day threshold demonstrated by Briggs et al. (2013) to prompt widespread nitrification. The hyporheic flow cells generated around these smaller BDAs follow a general pattern of downwelling upstream and upwelling downstream of the dams, but the residence times along these flow paths do not appear considerably different than the reach's natural background.

In contrast to BDAs 2 and 3, spatial patterns of porewater geochemistry and exchange

fluxes at BDA 4 show evidence for a robust hyporheic flow cell with connection to regional groundwater flow paths. Strong downward hydraulic gradients upstream of BDA 4 (Figure 4) confirm the presence of a significant head differential over the structure. Furthermore, the latitudinal homogeneity of these gradients across the channel demonstrates that such gradients are likely controlled by the existence of the larger BDA rather than by heterogeneity in streambed sediment conductivity. Moderate downward fluxes upstream of BDA 4 enable nitrification and the accumulation of nitrate in the streambed, processes that can only occur in conditions with sufficient dissolved oxygen (Zarnetske et al., 2011).

In some regions upstream of BDA 4, the streambed had a geochemical signature similar to that of the stream with little evidence of nitrate production (Figures 6 & 7). Zimmer et al. (2015) found that along hyporheic flow paths with high rates of exchange (3 m/day), nitrification products were unable to accumulate in the streambed due to the brevity of residence times and rapid recycling of water through the hyporheic zone. A similar process of ‘hyporheic flushing’ occurred upstream of BDA 4, where downward fluxes were sufficiently strong to prevent the buildup of nitrate, resulting in a region with stream water signatures. Downstream of BDA 4, the ascending limb of the hyporheic flow cell reemerges with an upwelling flux exceeding 1 m/day, carrying with it significant concentrations of calcium and sulfate. The geochemistry of these upwelling waters suggests that the flow cell produced by BDA 4 extends deep enough into the subsurface to interact with regional groundwater paths, which are enriched with products of gypsum dissolution (Fanelli & Lautz, 2008; Jin et al., 2010). The spatial extent of nitrate production and groundwater signatures in the streambed at BDA 4 indicates that it influences a larger spatial range than BDAs 2 and 3, extending up to 4 m upstream and downstream of the dam.

By considering the spatial distribution of porewater chemistry and estimates of downward flux in the context of BDA geometry, the size of the dams and the heights of their respective water drops appear to act as a first-order control on the degree of induced hyporheic exchange. Prior investigations of debris dams and other channel-spanning features show that the generation of hyporheic flows is strongly coupled to the magnitude of hydraulic head drop over a structure (Crispell & Endreny, 2009; Hester & Doyle, 2008; Janzen & Westbrook, 2011; Kasahara & Wondzell, 2003). Moreover, Janzen & Westbrook (2011) suggest that natural beaver dams may function with a size threshold behavior, altering hyporheic flows only after a certain dam height is exceeded. With a water surface drop of 0.40 m, BDA 4 was the only structure to produce a vertical, looping hyporheic cell that significantly affected spatial patterns of streambed porewater biogeochemistry. BDAs 2 and 3, with hydraulic steps of 0.14 m and 0.12 m respectively, were unable to replicate effects of BDA 4, generating minor hyporheic flows that did not overcome the streambed's natural background of reduction.

The disparity in behavior between the structures indicates that BDAs could act in a threshold-based manner similar to natural beaver dams. The positive relationship between dam height, measured by the hydraulic step, and hyporheic exchange magnitude may arise from the model of hydraulic pumping, where longitudinal gradients in streambed hydraulic head govern the magnitude of advective transport caused by a channel feature (Elliott & Brooks, 1997). Based on observations at this site, the threshold beaver dam analogue hydraulic step height required to produce a vertical, looping hyporheic flow cell is between the height of BDA 2 (0.14 m) and the height of BDA 4 (0.40 m). This size threshold is expected to differ across physical settings and is likely a function of local stream hydraulics (e.g., stream water level, streamflow velocity), streambed sediment properties, and dam construction techniques. The dependence of BDA

function on hydraulic step height is likely tied to the physical characteristics inherent to BDAs. Beaver dam analogues, composed of wooden posts, woven willow branches, and fill material, are intrinsically transient and leaky structures. Although BDAs do impound water and create upstream pools, their permeable nature does not promote the formation of fixed geomorphic conditions such as glides that are typically observed upstream of engineered restoration structures (Lautz et al., 2019; Lautz & Fanelli, 2008). Glides are a key geomorphic control on hyporheic exchange around other engineered restoration structures, as the change in streambed slope associated with these bedforms creates zones of contrasting hydraulic head that drive downward flows (Crispell & Endreny, 2009; Tonina & Buffington, 2009). If defined glide-slopes are not formed upstream of the BDAs, the dams will not produce the bedform-driven hydrodynamic flows that are often observed around impermeable, long-lasting restoration structures. Instead, hyporheic exchange will be driven solely by the hydrostatic head differentials created by hydraulic steps over the dams, accounting for the observed difference in exchange flux intensity between the individual BDAs. Though few field-based studies of BDAs exist, the link between BDA height and function has been affirmed in other work. Scamardo & Wohl (2020) found that BDA height was a significant predictor of sedimentation volume associated with the structures. These findings provide further confirmation that ability of a BDA to alter riparian processes is closely linked to BDA size.

Although BDAs at Red Canyon Creek did produce observable alterations of hyporheic exchange, the limited spatial extent of these changes compared to the overall length of the reach suggests BDAs did not impart a considerable influence on net surface water chemistry. While the hyporheic zone acts as a hotspot for biogeochemical activity, especially around in-stream structures, only a small proportion of total streamflow passes through the region, limiting its

effectiveness in mediating reach-scale solute concentrations (Briggs et al., 2013; Kasahara & Hill, 2006). Instead, the findings of this study suggest that BDAs promote moderate groundwater-surface water exchange and induce hyporheic flow paths with a wide spectrum of residence times. This may improve near-stream function by creating local heterogeneity in streambed biogeochemical cycling, evidenced by the presence of tightly-spaced regions of anaerobic reduction, nitrate production, and groundwater upwelling (Figure 7). The varied streambed environments induced by BDAs have the potential to support unique microhabitats for aquatic fauna, whose biodiversity is strongly coupled to the physical complexity of the hyporheic zone ecotone (Boulton et al., 1998; Brunke & Gonser, 1997).

4.2. Do BDAs function similarly to natural beaver dams?

As BDAs draw their primary inspiration from the designs of beaver, it is worth considering their ability to alter groundwater-surface water interactions in comparison to the function of natural beaver dams. The considerable breadth of work on natural beaver dams within the Red Canyon Creek watershed (Briggs et al., 2012, 2013, 2014) allows for a straightforward comparison with the BDAs of this study, assuming physical and hydrologic conditions are consistent throughout the watershed. The formerly observed beaver dams on neighboring Cherry Creek had water surface drops of 0.35 m and 0.75 m (Briggs et al., 2012), making them most closely comparable to BDA 4 in terms of size. At these sites, Briggs et al. (2012) estimated that downward hyporheic fluxes ranged from 1.2 m/day to 1.6 m/day at upstream glide locations and that fluxes were generally organized by streambed morphology, decreasing in magnitude with distance from the beaver dams. The presence of similar fluxes upstream of BDA 4 suggests that beaver dam analogues and natural beaver dams of equivalent size can produce hyporheic fluxes of comparable magnitudes.

The BDAs were also capable of reproducing spatial patterns of biogeochemical cycling observed around natural beaver dams. At locations closest to both the BDAs and the natural beaver dams, high flux rates diverted oxygenated stream water into the streambed, allowing shallow hyporheic flow paths with short residence times to remain oxic (Briggs et al., 2013). With increasing distance from both types of structures, limited downward fluxes resulted in a shift toward low dissolved oxygen conditions supporting anaerobic reduction (Figures 3 & 7). The signature of groundwater mixing present downstream of BDA 4 was not observed at the natural beaver dams on Cherry Creek, though this dissimilarity may be attributed to local differences in sediment hydraulic conductivity or the position of the structures within regional groundwater flow paths. The BDAs on Red Canyon Creek were capable of recreating the spatial distribution of hyporheic exchange created by similarly-sized natural beaver dams within a one-year period. However, it remains unclear if the effects of BDAs can be scaled up to replicate those of larger, maintained beaver dam complexes or extended to time scales longer than the one year of this study.

4.3. What factors influence the function of BDAs?

In their commentary on the widespread installation of beaver-related restoration approaches, Lautz et al. (2019) call for a thorough investigation into the underlying factors that regulate BDA efficacy. While this study identifies that dam size acts as a first-order control on the function of BDAs, there are a number of other factors that likely influence the effects of BDAs on groundwater-surface water interactions. Beaver dam analogues are intended to be installed in sequence, allowing for interaction between dams (Pollock et al., 2015), yet the effect of this cumulative behavior on hyporheic exchange remains largely unquantified. Although the most effective dam in this study was the most downstream structure in its sequence, it is difficult

to separate the cooperative contributions of BDAs 2 and 3 from the impact of the larger size of BDA 4. The upstream BDAs have the potential to affect hyporheic flows at downstream structures via multiple mechanisms. In the case of natural beaver dams, the first dam in a sequence has been shown to aggrade a larger proportion of pore-clogging, fine sediment before it can be transported and deposited further downstream (Curran & Cannatelli, 2014). A similar process may occur along a sequence of BDAs. This would result in higher streambed conductivity and larger magnitude hyporheic flows at more downstream dams due to the lack of available sediment. While this hypothesis is promising, further study of the BDAs' longitudinal effect on sediment character is required to verify its legitimacy. Waters diverted into the hyporheic zone by the upstream BDAs also have the potential to reemerge at downstream dams, contributing to the upwelling signal observed at the last BDA in the sequence. However, this scenario is likely implausible at Red Canyon Creek given the configuration and roughly 30 m spacing of the BDAs. At this site, hyporheic cells induced by restoration structures were observed to have flow paths that become horizontal at depths less than 1 m (Briggs et al., 2012; Lautz et al., 2006). The limited depths of these flow cells suggest that exchange fluxes generated by individual BDAs are unable to directly interact with other BDAs in a sequence due to their spacing. Additional study of the function of beaver dam analogues in sequence is required to isolate their additive interactions and to establish an optimal spacing interval between the structures.

The ability of a beaver dam analogue to effectively influence hyporheic exchange, as observed at BDA 4 (Figure 4), is also expected to strongly depend on the location of the structure in the broader spatial context of channel meanders and regional groundwater flow. Previous modeling of Red Canyon Creek demonstrates that the stream rapidly oscillates between gaining

and losing water over its length, as the complex geometry of its meanders interacts with the regional south to north trend in groundwater flow (Figure 1) (Lautz & Siegel, 2006). Lautz et al. (2006) show that hyporheic interactions around debris dams are particularly dependent on whether a reach is gaining or losing to the subsurface. At BDAs 2 and 3, where vertical hydraulic gradients and prior modeling suggest that the stream is losing, downward hyporheic flows were more likely to be captured by regional subsurface flows and not reemerge downstream (Lautz et al., 2006). By contrast, BDA 4 was located in a portion of the reach where regional flow paths are expected to rejoin the stream. When a debris dam is positioned along a gaining reach, water diverted into the subsurface by the structure may be quickly returned to the channel, potentially bearing with it the geochemical signature of groundwater (Lautz et al., 2006). This process may contribute to the upwelling of sulfate-rich waters downstream of BDA 4, though it remains challenging to decouple the magnitude of this effect from other explanatory factors such as dam size. The interactions between the BDAs and regional hydrology are also dependent on the streamflow conditions at the time of observation. While the results of this study hold for baseflow conditions, further work is required to extend these findings to other points in time throughout the annual hydrograph.

5. Conclusion

The results of this study indicate that, under certain circumstances, beaver dam analogues effectively promote spatial heterogeneity in hyporheic exchange fluxes and subsurface biogeochemical cycling. While the two smaller dams, BDAs 2 and 3, produced fluxes that exceeded those of pre-restoration conditions, residence times along these flow paths remained sufficiently long such that anaerobic reduction persisted as the dominant streambed process. Exchange fluxes associated with BDA 4 were larger in magnitude, prompting the development

of hotspots of varied biogeochemical activity via increases in the supply of dissolved oxygen to microbial communities in the hyporheic zone. This study attributes the difference in performance between the BDAs to a size threshold behavior, where patterns of streambed biogeochemical cycling are altered after a site-dependent dam height is reached. The functional similarity between BDAs and beaver dams of equivalent size demonstrates that the structures likely induce hyporheic exchange by a common hydrodynamic mechanism. We observed that BDAs effectively replicate the functions and impacts of their natural counterparts in the short term, though it remains unclear if they can maintain these effects over years to decades.

As one of the first field studies examining the influence of BDAs on hyporheic processes, these results guide the future design and implementation of beaver-inspired restoration approaches. In order to achieve heterogeneity in hyporheic conditions, this study suggests that BDAs be constructed to produce a hydraulic step exceeding a specified threshold height, though this threshold value is likely a function of local hydrology and streambed sediment character. The additive nature of BDAs in sequence is widely acknowledged, but our observations from the three structures on Red Canyon Creek were insufficient to isolate the degree to which this effect influenced hyporheic exchange. While the behavior of the dams in sequence, coupled with other factors such as regional hydrology, may affect the performance of the structures, these results indicate that the size of a BDA's hydraulic step acts as the primary control on its ability to alter groundwater-surface water interactions.

Figures

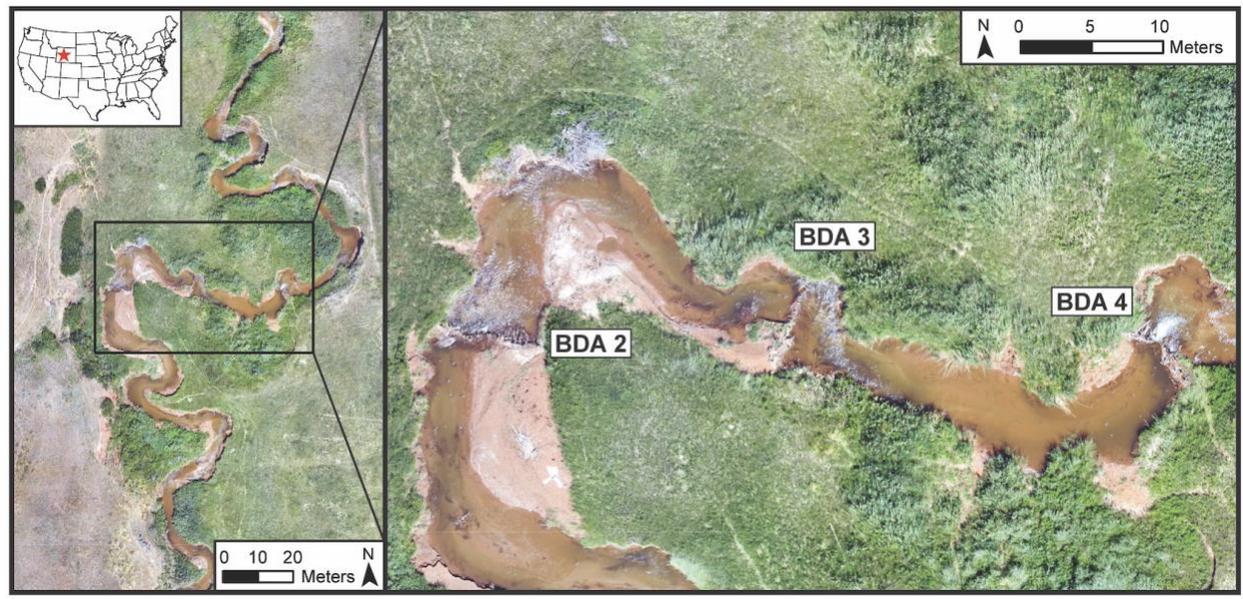


Figure 1. Aerial image of the study site and its location within the United States. The installation and monitoring of three BDAs occurred along a roughly 60 m segment of Red Canyon Creek.

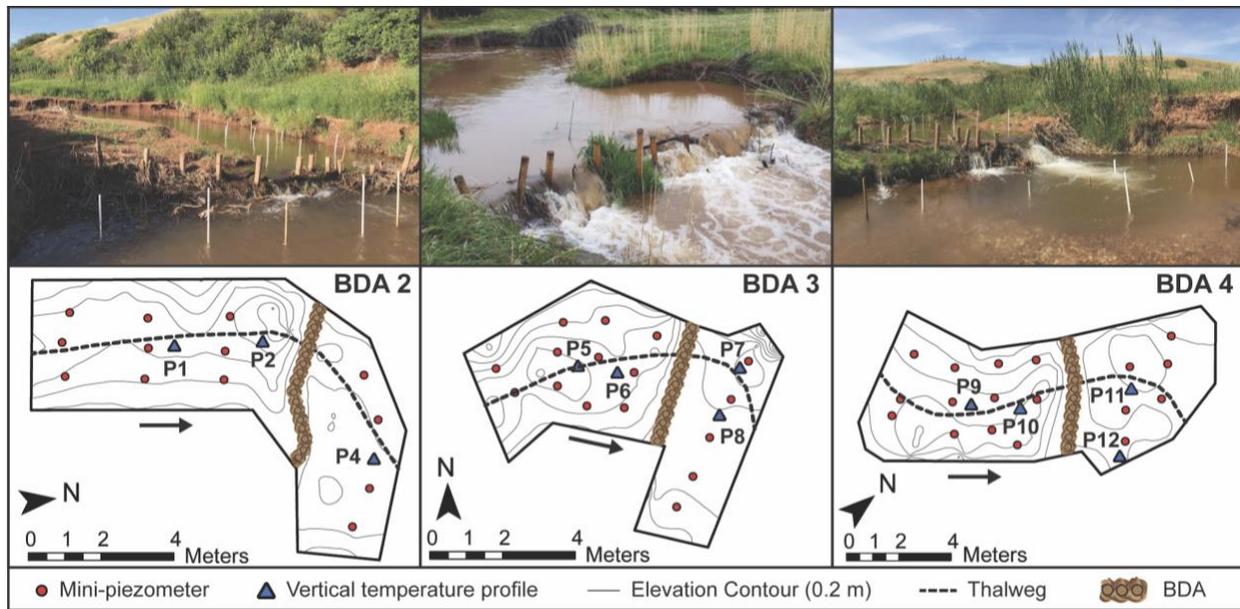


Figure 2. Site maps of beaver dam analogues with locations of instrumentation. Vertical temperature profiles are labeled from P1 (upstream, BDA2) to P12 (downstream, BDA4). The direction of streamflow is indicated by black arrows.

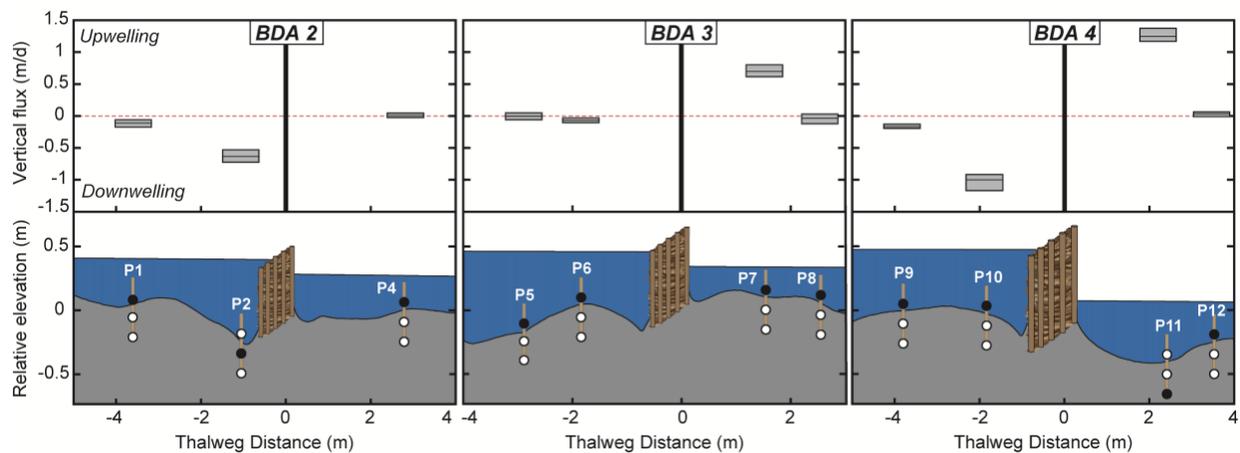


Figure 3. Simplified boxplots indicating the minimum, median, and maximum vertical water flux estimated at each temperature profile plotted against thalweg distance. Positive flux values indicate upwelling and negative flux values indicate downwelling. Vertical temperature profiles (P1-P12) are displayed with iButtons at depths of 0 m, 0.1 m, and 0.2 m. White highlighted iButtons indicate the optimal sensor pair used to obtain thermally-derived exchange fluxes. Relative elevations were standardized at each BDA by subtracting the mean elevation from each survey point. Flow is from left to right in all profiles.

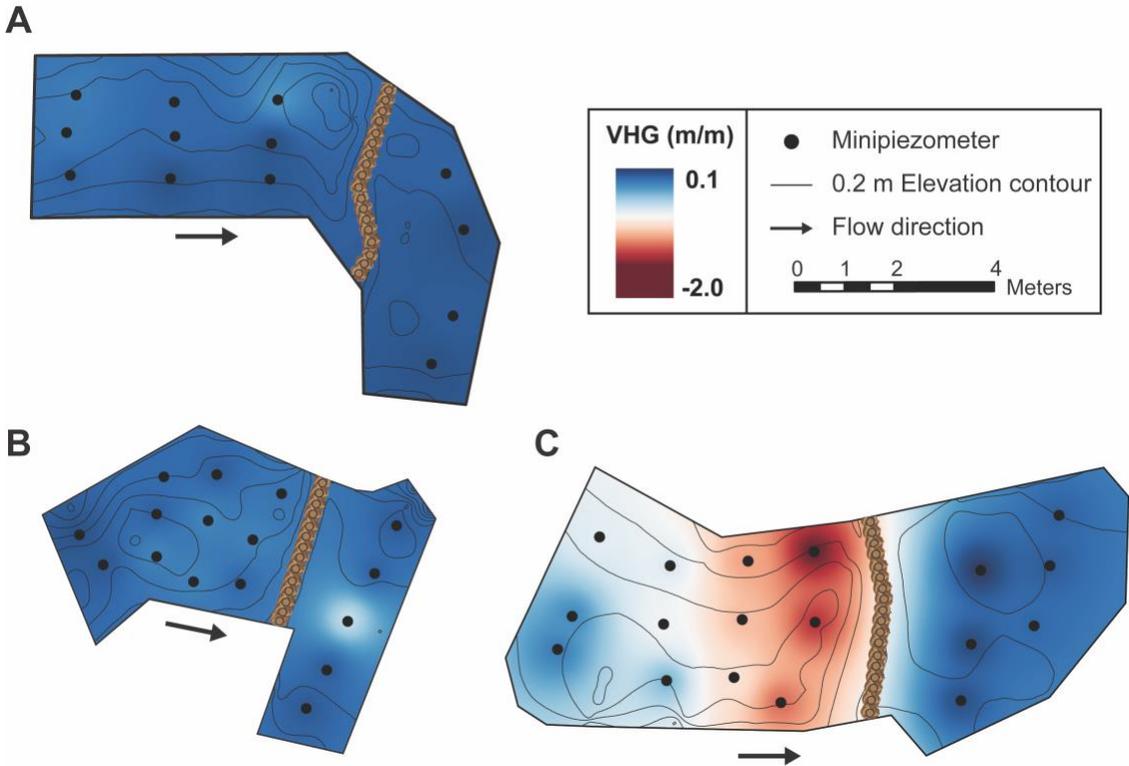


Figure 4. Spatial distribution of vertical hydraulic gradients (VHG) at (A) BDA 2, (B) BDA 3, and (C) BDA 4. Positive gradient values are upwards and negative gradient values are downwards. Interpolation of mini-piezometer hydraulic gradient measurements was performed using inverse distance weighting interpolation

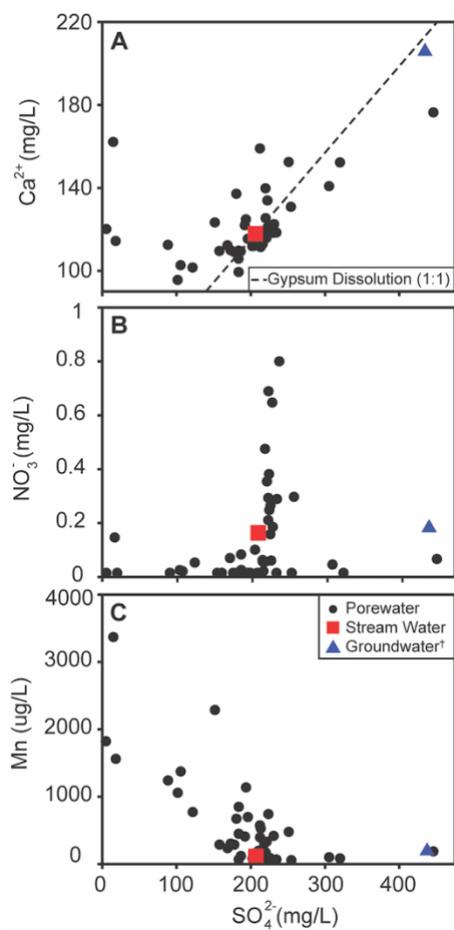


Figure 5. Bivariate plots of major ion chemistry collected from streambed porewater, stream water, and groundwater. The slope of the dissolution line in (A) represents the dissociation of gypsum at a 1:1 equivalence ratio. †Groundwater value from October 2005 (Lautz & Fanelli, 2008).

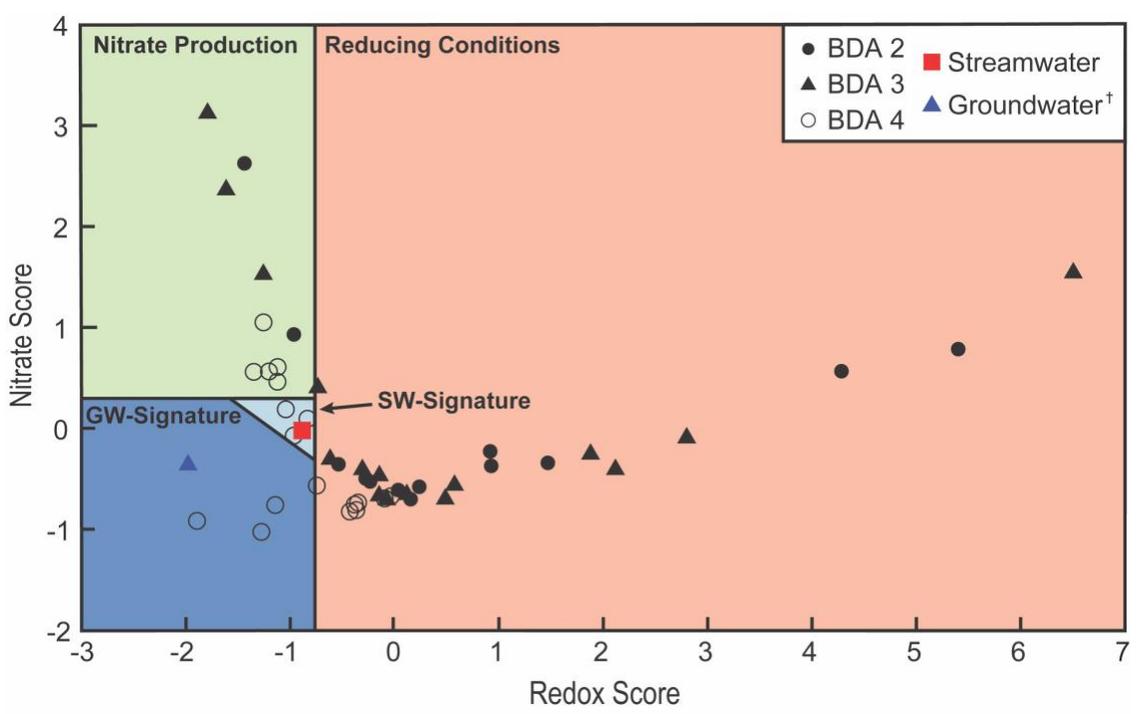


Figure 6. Redox score (PC1) and nitrate score (PC2) results from principal component analysis (PCA) plotted by beaver dam analogue. Samples were classified into four categories (nitrate production, reducing conditions, groundwater signature, and stream water signature) based on PCA results. †Groundwater value from October 2005 (Lautz & Fanelli, 2008).

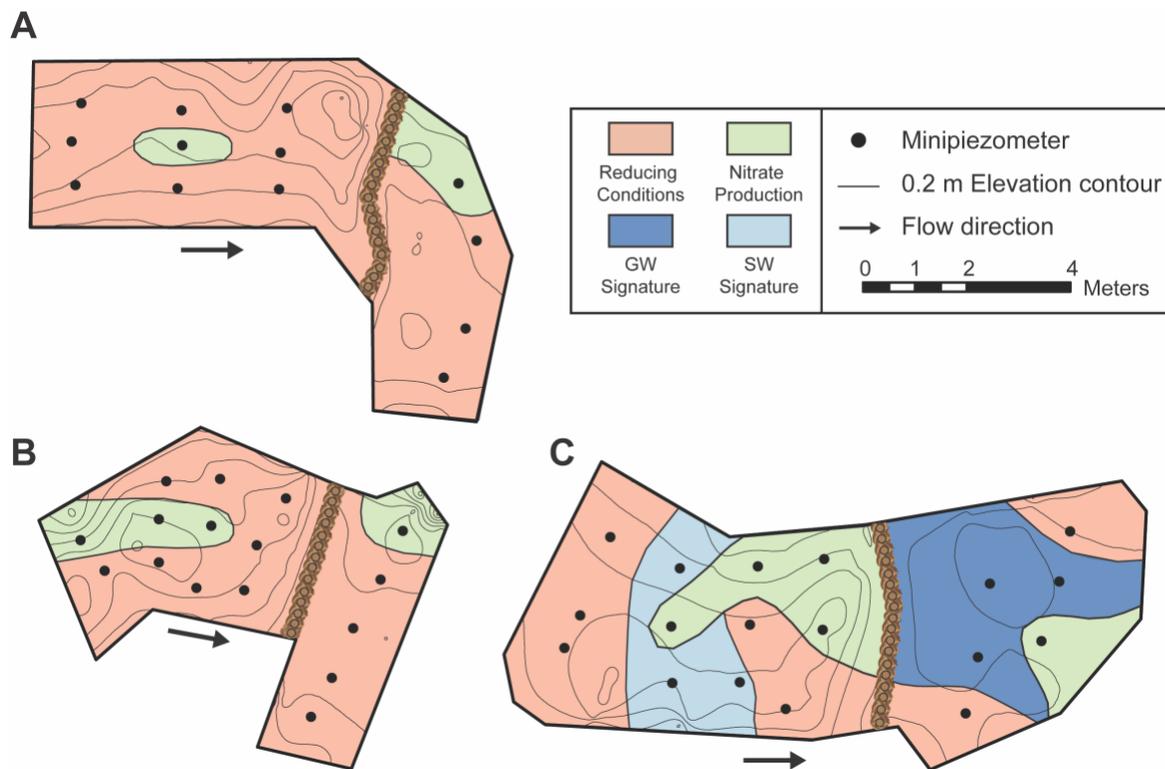


Figure 7. Classification of porewater type (collected and analyzed at each mini-piezometer) based on principal component analysis at (A) BDA 2, (B) BDA 3, and (C) BDA 4.

Tables

Table 1. Sediment parameters used in vertical flux heat transport modeling. [†]

Parameter	Value	Units
Porosity	0.35	Dimensionless
Thermal dispersivity (β)	0.001	m
Thermal conductivity of the sediment (λ_0)	0.0033	cal s ⁻¹ cm ⁻¹ °C ⁻¹
Volumetric heat capacity of the sediment (C_s)	0.5	cal cm ⁻³ °C ⁻¹
Volumetric heat capacity of the water (C_w)	1.0	cal cm ⁻³ °C ⁻¹

[†]Parameter values adapted from prior heat transport modeling in the watershed by Briggs et al. (2012.).

Table 2. Loadings, eigenvalues, and explained variance of the first two principal components.

Correlations presented between the principal components and the concentrations of individual redox species.

	PC1 (<i>Redox function</i>)	p	PC2 (<i>Nitrate function</i>)	p
<i>PCA loadings</i>				
NH ₄ ⁺	0.45	0.78	0.12	ns
NO ₃ ⁻	-0.24	-0.41	0.96	0.91
Mn	0.51	0.89	0.02	ns
Fe	0.50	0.87	0.18	ns
SO ₄ ²⁻	-0.47	-0.81	-0.15	ns
Eigenvalue	3.02		0.89	
Explained variance	60.3%		17.8%	
Cumulative explained variance	60.3%		78.1%	

ns = not significant at $p < 0.05$

References

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433–459. <https://doi.org/10.1002/wics.101>
- Bernhardt, E. S., Palmer, M. A., Allan, J. D., Alexander, G., Barnas, K., Brooks, S., Carr, J., Clayton, S., Dahm, C., Follstad-Shah, J., Galat, D., Gloss, S., Goodwin, P., Hart, D., Hassett, B., Jenkinson, R., Katz, S., Kondolf, G. M., Lake, P. S., ... Sudduth, E. (2005). Synthesizing U.S. River Restoration Efforts. *Science*, 308(5722), 636–637. <https://doi.org/10.1126/science.1109769>
- Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, 52(4), 603–679. <https://doi.org/10.1002/2012RG000417>
- Boulton, A. J., Findlay, S., Marmonier, P., Stanley, E. H., & Maurice Valett, H. (1998). The functional significance of the hyporheic zone in streams and rivers. *Annual Review of Ecology and Systematics*, 29, 59–81. <https://doi.org/10.1146/annurev.ecolsys.29.1.59>
- Briggs, M. A., Lautz, L. K., & Hare, D. K. (2014). Residence time control on hot moments of net nitrate production and uptake in the hyporheic zone. *Hydrological Processes*, 28(11), 3741–3751. <https://doi.org/10.1002/hyp.9921>
- Briggs, M. A., Lautz, L. K., Hare, D. K., & González-Pinzón, R. (2013). Relating hyporheic fluxes, residence times, and redox-sensitive biogeochemical processes upstream of beaver dams. *Freshwater Science*, 32(2), 622–641. <https://doi.org/10.1899/12-110.1>
- Briggs, M. A., Lautz, L. K., McKenzie, J. M., Gordon, R. P., & Hare, D. K. (2012). Using high-resolution distributed temperature sensing to quantify spatial and temporal variability in

- vertical hyporheic flux. *Water Resources Research*, 48(2), 1–16.
<https://doi.org/10.1029/2011WR011227>
- Brunke, M., & Gonser, T. (1997). The ecological significance of exchange processes between rivers and groundwater. *Freshwater Biology*, 37(1), 1–33. <https://doi.org/10.1046/j.1365-2427.1997.00143.x>
- Chaney, E., Elmore, W., & Platts, W. S. (1993). *Livestock Grazing on Western Riparian Areas*. Northwest Resource Information Center Inc. United States Environmental Protection Agency.
- Crispell, J. K., & Endreny, T. A. (2009). Hyporheic exchange flow around constructed in-channel structures and implications for restoration design. *Hydrological Processes*, 23(8), 1158–1168. <https://doi.org/10.1002/hyp.7230>
- Curran, J. C., & Cannatelli, K. M. (2014). The impact of beaver dams on the morphology of a river in the eastern United States with implications for river restoration. *Earth Surface Processes and Landforms*, 39(9), 1236–1244. <https://doi.org/10.1002/esp.3576>
- Daniluk, T. L., Lautz, L. K., Gordon, R. P., & Endreny, T. A. (2013). Surface water-groundwater interaction at restored streams and associated reference reaches. *Hydrological Processes*, 27(25), 3730–3746. <https://doi.org/10.1002/hyp.9501>
- Elliott, A. H., & Brooks, N. H. (1997). Transfer of nonsorbing solutes to a streambed with bed forms: Theory. *Water Resources Research*, 33(1), 123–136.
<https://doi.org/10.1029/96WR02784>
- Fairfax, E., & Small, E. E. (2018). Using remote sensing to assess the impact of beaver damming on riparian evapotranspiration in an arid landscape. *Ecohydrology*, 11(7), e1993.
<https://doi.org/10.1002/eco.1993>

- Fanelli, R. M., & Lautz, L. K. (2008). Patterns of Water, Heat, and Solute Flux through Streambeds around Small Dams. *Ground Water*, 46(5), 671–687.
<https://doi.org/10.1111/j.1745-6584.2008.00461.x>
- Gibson, P. P., & Olden, J. D. (2014). Ecology, management, and conservation implications of North American beaver (*Castor canadensis*) in dryland streams. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 24(3), 391–409. <https://doi.org/10.1002/aqc.2432>
- Gomez, J. D., Wilson, J. L., & Cardenas, M. B. (2012). Residence time distributions in sinuosity-driven hyporheic zones and their biogeochemical effects. *Water Resources Research*, 48(9), 1–17. <https://doi.org/10.1029/2012WR012180>
- Gooseff, M. N. (2010). Defining Hyporheic Zones - Advancing Our Conceptual and Operational Definitions of Where Stream Water and Groundwater Meet. *Geography Compass*, 4(8), 945–955. <https://doi.org/10.1111/j.1749-8198.2010.00364.x>
- Gordon, R. P., Lautz, L. K., Briggs, M. A., & McKenzie, J. M. (2012). Automated calculation of vertical pore-water flux from field temperature time series using the VFLUX method and computer program. *Journal of Hydrology*, 420–421, 142–158.
<https://doi.org/10.1016/j.jhydrol.2011.11.053>
- Gordon, R. P., Lautz, L. K., & Daniluk, T. L. (2013). Spatial patterns of hyporheic exchange and biogeochemical cycling around cross-vane restoration structures: Implications for stream restoration design. *Water Resources Research*, 49(4), 2040–2055.
<https://doi.org/10.1002/wrcr.20185>
- Hatch, C. E., Fisher, A. T., Revenaugh, J. S., Constantz, J., & Ruehl, C. (2006). Quantifying surface water-groundwater interactions using time series analysis of streambed thermal records: Method development. *Water Resources Research*, 42(10), 1–14.

<https://doi.org/10.1029/2005WR004787>

- Hester, E. T., & Doyle, M. W. (2008). In-stream geomorphic structures as drivers of hyporheic exchange. *Water Resources Research*, 44(3). <https://doi.org/10.1029/2006WR005810>
- Irvine, D. J., Briggs, Martin. A., Cartwright, I., Scruggs, C. R., & Lautz, L. K. (2017). Improved Vertical Streambed Flux Estimation Using Multiple Diurnal Temperature Methods in Series. *Groundwater*, 55(1), 73–80. <https://doi.org/10.1111/gwat.12436>
- Irvine, D. J., Lautz, L. K., Briggs, M. A., Gordon, R. P., & McKenzie, J. M. (2015). Experimental evaluation of the applicability of phase, amplitude, and combined methods to determine water flux and thermal diffusivity from temperature time series using VFLUX 2. *Journal of Hydrology*, 531, 728–737. <https://doi.org/10.1016/j.jhydrol.2015.10.054>
- Janzen, K., & Westbrook, C. J. (2011). Hyporheic Flows Along a Channelled Peatland: Influence of Beaver Dams. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, 36(4), 331–347. <https://doi.org/10.4296/cwrj3604846>
- Jin, L., Siegel, D. I., Lautz, L. K., & Lu, Z. (2012). Identifying streamflow sources during spring snowmelt using water chemistry and isotopic composition in semi-arid mountain streams. *Journal of Hydrology*, 470–471, 289–301. <https://doi.org/10.1016/j.jhydrol.2012.09.009>
- Jin, L., Siegel, D. I., Lautz, L. K., Mitchell, M. J., Dahms, D. E., & Mayer, B. (2010). Calcite precipitation driven by the common ion effect during groundwater-surface-water mixing: A potentially common process in streams with geologic settings containing gypsum. *Bulletin of the Geological Society of America*, 122(7–8), 1027–1038. <https://doi.org/10.1130/B30011.1>
- Jin, L., Siegel, D. I., Lautz, L. K., & Otz, M. H. (2009). Transient storage and downstream solute

- transport in nested stream reaches affected by beaver dams. *Hydrological Processes*, 23(17), 2438–2449. <https://doi.org/10.1002/hyp.7359>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Kasahara, T., & Hill, A. R. (2006). Hyporheic exchange flows induced by constructed riffles and steps in lowland streams in southern Ontario, Canada. *Hydrological Processes*, 20(20), 4287–4305. <https://doi.org/10.1002/hyp.6174>
- Kasahara, T., & Wondzell, S. M. (2003). Geomorphic controls on hyporheic exchange flow in mountain streams. *Water Resources Research*, 39(1), SBH 3-1-SBH 3-14. <https://doi.org/10.1029/2002wr001386>
- Krause, S., Lewandowski, J., Grimm, N. B., Hannah, D. M., Pinay, G., McDonald, K., Martí, E., Argerich, A., Pfister, L., Klaus, J., Battin, T. J., Larned, S. T., Schelker, J., Fleckenstein, J., Schmidt, C., Rivett, M. O., Watts, G., Sabater, F., Sorolla, A., & Turk, V. (2017). Ecohydrological interfaces as hot spots of ecosystem processes. *Water Resources Research*, 53(8), 6359–6376. <https://doi.org/10.1002/2016WR019516>
- Lautz, L. K., & Fanelli, R. M. (2008). Seasonal biogeochemical hotspots in the streambed around restoration structures. *Biogeochemistry*, 91(1), 85–104. <https://doi.org/10.1007/s10533-008-9235-2>
- Lautz, L. K., Kelleher, C., Vidon, P. G., Coffman, J., Riginos, C., & Copeland, H. (2019). Restoring stream ecosystem function with beaver dam analogues: Let's not make the same mistake twice. *Hydrological Processes*, 33(1), 174–177.

<https://doi.org/10.1002/hyp.13333>

Lautz, L. K., Kranes, N. T., & Siegel, D. I. (2010). Heat tracing of heterogeneous hyporheic exchange adjacent to in-stream geomorphic features. *Hydrological Processes*, 24(21), 3074–3086. <https://doi.org/10.1002/hyp.7723>

Lautz, L. K., & Siegel, D. I. (2006). Modeling surface and ground water mixing in the hyporheic zone using MODFLOW and MT3D. *Advances in Water Resources*, 29(11), 1618–1633. <https://doi.org/10.1016/j.advwatres.2005.12.003>

Lautz, L. K., Siegel, D. I., & Bauer, R. L. (2006). Impact of debris dams on hyporheic interaction along a semi-arid stream. *Hydrological Processes*, 20(1), 183–196. <https://doi.org/10.1002/hyp.5910>

Lewandowski, J., Putschew, A., Schwesig, D., Neumann, C., & Radke, M. (2011). Fate of organic micropollutants in the hyporheic zone of a eutrophic lowland stream: Results of a preliminary field study. *Science of The Total Environment*, 409(10), 1824–1835. <https://doi.org/10.1016/j.scitotenv.2011.01.028>

Luce, C. H., Tonina, D., Gariglio, F., & Applebee, R. (2013). Solutions for the diurnally forced advection-diffusion equation to estimate bulk fluid velocity and diffusivity in streambeds from temperature time series. *Water Resources Research*, 49(1), 488–506. <https://doi.org/10.1029/2012WR012380>

Naiman, R. J., Johnston, C. A., & Kelley, J. C. (1988). Alteration of North American Streams by Beaver. *BioScience*, 38(11), 753–762. <https://doi.org/10.2307/1310784>

Pilliod, D. S., Rohde, A. T., Charnley, S., Davee, R. R., Dunham, J. B., Gosnell, H., Grant, G. E., Hausner, M. B., Huntington, J. L., & Nash, C. (2018). Survey of Beaver-related Restoration Practices in Rangeland Streams of the Western USA. *Environmental*

- Management, 61(1), 58–68. <https://doi.org/10.1007/s00267-017-0957-6>
- Pollock, M. M., Beechie, T. J., Wheaton, J. M., Jordan, C. E., Bouwes, N., Weber, N., & Volk, C. (2014). Using Beaver Dams to Restore Incised Stream Ecosystems. *BioScience*, 64(4), 279–290. <https://doi.org/10.1093/biosci/biu036>
- Pollock, M. M., Jordan, C. E., Lewallen, G., Woodruff, K., & Castro, J. (2015). *The Beaver Restoration Guidebook: Working with Beaver to Restore Streams, Wetlands, and Floodplains*. United States Fish and Wildlife Service.
<https://www.fws.gov/oregonfwo/promo.cfm?id=177175812>
- Sawyer, A. H., Bayani Cardenas, M., & Buttles, J. (2012). Hyporheic temperature dynamics and heat exchange near channel-spanning logs. *Water Resources Research*, 48(1), 1–11.
<https://doi.org/10.1029/2011WR011200>
- Scamardo, J., & Wohl, E. (2020). Sediment storage and shallow groundwater response to beaver dam analogues in the Colorado Front Range, USA. *River Research and Applications*, 36(3), 398–409. <https://doi.org/10.1002/rra.3592>
- Shahverdian, S. M., Wheaton, J. M., Bennett, S. N., Bouwes, N., Camp, R., Jordan, C. E., & Weber, N. (2019). Mimicking & Promoting Wood Accumulation & Beaver Dam Activity with PALS & BDAs.
- Storey, R. G., Howard, K. W. F., & Williams, D. D. (2003). Factors controlling riffle-scale hyporheic exchange flows and their seasonal changes in a gaining stream: A three-dimensional groundwater flow model. *Water Resources Research*, 39(2), 1–17.
<https://doi.org/10.1029/2002WR001367>
- Tonina, D., & Buffington, J. M. (2009). Hyporheic Exchange in Mountain Rivers I: Mechanics and Environmental Effects. *Geography Compass*, 3(3), 1063–1086.

<https://doi.org/10.1111/j.1749-8198.2009.00226.x>

Wegener, P., Covino, T., & Wohl, E. (2017). Beaver-mediated lateral hydrologic connectivity, fluvial carbon and nutrient flux, and aquatic ecosystem metabolism. *Water Resources Research*, 53(6), 4606–4623. <https://doi.org/10.1002/2016WR019790>

Westbrook, C. J., Cooper, D. J., & Baker, B. W. (2006). Beaver dams and overbank floods influence groundwater-surface water interactions of a Rocky Mountain riparian area. *Water Resources Research*, 42(6), 1–12. <https://doi.org/10.1029/2005WR004560>

Zarnetske, J. P., Haggerty, R., Wondzell, S. M., & Baker, M. A. (2011). Dynamics of nitrate production and removal as a function of residence time in the hyporheic zone. *Journal of Geophysical Research*, 116(G1), G01025. <https://doi.org/10.1029/2010JG001356>

Zimmer, M. A., & Lautz, L. K. (2014). Temporal and spatial response of hyporheic zone geochemistry to a storm event. *Hydrological Processes*, 28(4), 2324–2337. <https://doi.org/10.1002/hyp.9778>

Zimmer, M. A., & Lautz, L. K. (2015). Pre- and postrestoration assessment of stream water–groundwater interactions: effects on hydrological and chemical heterogeneity in the hyporheic zone. *Freshwater Science*, 34(1), 287–300. <https://doi.org/10.1086/679514>

**Chapter 2: Machine learning unravels controls on river water temperature regime
dynamics**

This chapter has been submitted as:

Wade J, Kelleher C, Hannah, D.M. Machine learning unravels controls on river water temperature regime dynamics, (in review), *Journal of Hydrology*.

Abstract

Water temperature is vitally important to the health of rivers and streams, influencing the integrity of ecosystems, aquatic biogeochemistry, and the habitability of waterways for a variety of species. While climate is often regarded as the primary driver of stream temperature regimes, other factors - including hydrology, watershed characteristics, and human impacts - add substantial complexity to the variability of water temperatures. However, it remains challenging to disentangle the influence of these drivers through time and across rivers spanning diverse settings. To quantify the underlying controls on river thermal regimes, we applied random forest algorithms to model maximum monthly temperatures and thermal sensitivities at 410 watersheds spanning the conterminous United States. We interpreted these random forest models using variable importance rankings, describing seasonal and spatial variability in the dominant controls on water temperatures. Although our empirical results confirm that climate is indeed a primary control on temperature magnitude, our models highlight the diversity in drivers of water temperature variability across seasons, hydrologic regions, and between metrics. By combining random forest models with process-based understanding of stream thermal regimes, we provide new insights on the dynamic controls of water temperature variability across broad geographical domains, informing region- and season-specific controls for tailored thermal watershed management and guiding the framing of future water temperature modeling.

1. Introduction

The water temperature of streams is a leading control on overall water quality, mediating the rate of solute processing and the viability of aquatic habitats (Caissie, 2006; Ouellet et al., 2020; Steel et al., 2017; Webb et al., 2008). As such, it is critical to understand the distribution of stream thermal regimes and to identify the underlying environmental drivers of water temperature

variability. Though there are many examples of studies that have characterized key drivers of river flow regimes at regional, national, and global scales (Hammond et al., 2021; Konapala & Mishra, 2020; Price et al., 2021; Singh & Basu, 2022), there are comparatively fewer studies that seek to categorize river thermal regimes over large geographical domains. Recent efforts to classify stream temperature behavior into distinct groups have emphasized that while general patterns in variability driven by regional climate gradients exist, thermal regimes are strongly heterogeneous - indicating the presence of other, second-order drivers (Garner et al., 2014; Hill et al., 2014; Maheu et al., 2016; Willis et al., 2021).

The controls on stream thermal regimes are diverse, spanning a range of meteorologic, landscape, and hydrologic processes across nested space and time domains (Dugdale et al., 2017; Hannah & Garner, 2015; Webb, 1996). Although climate appears to be the dominant macro-scale control on stream temperature, hydrologic and watershed characteristics likely moderate climate forcings, introducing complexity at the basin and reach scale (Hannah & Garner, 2015; Laizé et al., 2017). Supplementing these site-specific controls, the transformative effect of anthropogenic activity on earth system processes can perturb numerous aspects of stream thermal regimes (Croghan et al., 2019; Ficklin et al., 2023). Broadly, human activity alters climate forcings on streams (anthropogenic climate change), as well as the hydrologic and landscape controls that mediate those forcings (urbanization, deforestation, GW pumping, channelization, dam construction, wetland destruction) (Caissie, 2006; Poole & Berman, 2001). Given the close links between global climate patterns and stream thermal regimes, future anthropogenic warming is anticipated to propagate into warming of water temperatures (Caldwell et al., 2015; Isaak et al., 2012; Michel et al., 2022; van Vliet et al., 2011). However, trends in future water temperatures are likely to be geographically non-uniform due to the space-time variability of watershed and

hydrologic forcings (Kelleher et al., 2021). Therefore, it is remarkably important, both to the management of thermal regimes and to the ecosystem services provided by river networks, to better understand how climate and watershed processes interact in space and time to control water temperature behavior.

The relationships between thermal regimes and moderating environmental factors are likely nonlinear and interactive. As such, nonparametric data-driven methods are needed to disentangle the primary controls on stream temperatures (Konapala & Mishra, 2020). Despite large volumes of environmental data, many process-based models struggle to uncover complex macro- and basin-scale patterns inherent to hydrologic systems (Nearing et al., 2021). In this context, supervised machine learning algorithms are a popular tool to overcome this hurdle because such tools are capable of parsing large environmental datasets and inferring subtle relationships between input predictors and output variables without explicitly representing physical processes (Konapala & Mishra, 2020; Nearing et al., 2021; Tyrallis et al., 2019). While machine learning approaches have been increasingly applied to water resources prediction, only a limited number of studies use the technique to identify influential controls on hydrologic behavior (Addor et al., 2018; Gannon et al., 2022; Hammond et al., 2021; Hill et al., 2014; Konapala & Mishra, 2020).

Though our understanding of controls on river temperature regimes continues to develop, we still lack insight regarding how the dominant controls on thermal characteristics vary seasonally and across broad spatial scales. To address this, we leverage publicly available records of stream temperature and environmental covariates at 410 sites to explore the following questions:

1. What are the relative influences of climate, watershed characteristics, hydrologic

controls, and anthropogenic disturbances on stream temperature magnitude and variability in the conterminous US?

2. Do the dominant drivers of stream thermal regimes differ seasonally and across spatial domains?

While stream thermal regimes are complex and can be quantified in myriad ways, we focused on two primary metrics – maximum water temperature and thermal sensitivity – calculated monthly for all sites. To infer key drivers of water temperature, we applied a supervised machine learning approach to derive models predicting monthly water temperature metrics using a suite of environmental covariates. To our knowledge, our study represents one of the most comprehensive data driven investigations of empirical drivers of water temperature regimes, that not only accounts for regional variability, but also considers how drivers may vary at seasonal scales. Such insight is vital for not only improving our understanding of river temperature regimes but managing complex landscapes and the ecosystems they support now and into the future.

2. Materials and Methods

The general approach of our methodology is outlined in Figure 1 and explained in detail below. Across selected stream sites, we quantified two metrics of water temperature variability at a monthly time scale and extracted variables describing climate, watershed characteristics, hydrology, and human impacts thought to either directly or indirectly shape water temperatures. We then fit a suite of random forest models to predict water temperature metrics using these environmental covariates. Finally, by leveraging variable importance metrics to rank predictors by their relative influence, we inferred how dominant controls on river thermal regimes vary seasonally, within specific geographic domains, and across degrees of anthropogenic disruption.

2.1. Site Selection

We chose to investigate the dominant drivers of stream temperature regimes in the conterminous US given the availability of public data and the presence of strong gradients in climate, physiographic landscapes, and degrees of human impacts across US watersheds. Our analysis focuses on 410 U. S. Geological Survey (USGS) sites with relatively complete records (<10% missing) of daily stream temperature and discharge observations from 2016 to 2020 (Figure 2). This recent period was chosen to maximize the number of sites available, given the continual expansion of the USGS stream temperature monitoring network over time. All sites are present in the Geospatial Attributes of Gages for Evaluating Streamflow dataset (GAGES-II; Falcone et al., 2010), as this dataset contains quantitative descriptors of watershed characteristics, climate, and anthropogenic influence for USGS stream gages. The sites chosen span a broad range of watershed sizes, geographic settings, and degrees of human influence (Figure 2).

2.2. Stream Temperature Signatures

To explore and explain variability in stream temperature regimes, we adapted the concepts of hydrologic signatures (Hannah et al., 2000; McMillan, 2020) and thermal facets (Casado et al., 2013; Steel et al., 2017) to calculate two quantitative stream temperature signatures that capture aspects of a site's thermal behavior. Stream temperature signatures were calculated from mean daily stream temperature observations across 4 years of data, retrieved from the USGS National Water Information System (NWIS).

First, we extracted the maximum monthly stream temperature (abbreviated ST_{max}) at each gage. ST_{max} has clear ecological implications, with extreme water temperatures being linked to aquatic ecosystem mortality (O'Neal, 2002). Values were first calculated at a monthly scale and

then averaged for each month across the analysis period. We also calculated monthly thermal sensitivity, which represents the linear slope between daily air temperatures and stream temperatures (abbreviated *TS*). *TS* describes how water temperatures are coupled to air temperatures, and has been linked to stream and watershed characteristics (Kelleher et al., 2012). Air temperature observations at each site used in calculations of *TS* were derived from the PRISM gridded climate dataset (Daly et al., 2008). As with *ST_{max}*, *TS* values were taken as a monthly average across the analysis period.

2.3. Site Predictors

We selected 23 quantitative variables, primarily derived from the GAGES-II dataset, that could be linked to aspects of stream temperatures regimes (Figure 3). As the GAGES-II dataset does not include discharge data, we also included median monthly discharge from 2016-2020 in the predictor set, retrieved at each site using the USGS NWIS. With the exception of monthly air temperature, precipitation, and discharge, all selected predictor variables were static in time. We acknowledge that by using static GAGES-II variables as predictors in our models, we are not necessarily representing the time-varying states of selected watersheds. This choice may affect our ability to detect seasonal patterns in importance, and further inclusion of variables that track temporal changes in landscape states may improve model quality (Singh & Basu, 2022).

Variables were excluded from the final set of input predictors if they displayed strong multicollinearity with other variables of interest ($\rho > 0.8$). These included: the removal of annual mean air temperature, which was correlated with monthly air temperature; the removal of PET, which was correlated with monthly air temperature; the removal of developed land cover and road density, which were correlated with imperviousness. This left us with a set of 23 variables for further analysis.

We classified variables from our analysis into four broad categories: climate, watershed characteristics, hydrology, and human impacts. These categories enable a comprehensive evaluation of the relative and cumulative effect of each variable category. While many of the variables may span multiple potential categories, we placed them into the category that most closely described their relationship with stream thermal processes. Complete descriptions of the selected predictor variables are available in Table S2.

2.4. Applying Random Forest Models to Infer Drivers of Stream Temperature Regimes

2.4.1. Training Random Forest Models

To infer controlling drivers of stream thermal regimes, we applied random forest (RF) models, an ensemble learning algorithm based on the aggregation of numerous independent regression trees (Breiman, 2001; Tyrallis et al., 2019). The primary goal of an RF model is to use a set of input predictor variables (covariates) to simulate an observed value of interest. Random forest models have been used to predict streamflow in ungauged basins (White, 2017), assess flood severity and extent (Albers et al., 2016; Woznicki et al., 2019), and evaluate the dominant controls on hydrologic behavior (Konapala & Mishra, 2020; Price et al., 2021). The original random forest algorithm (Breiman, 2001), which is underlain by classification and regression trees, has been modified to function using conditional inference trees to reduce model bias (Strobl et al., 2007). While similar in structure to a typical regression tree, conditional inference trees differ in that their splitting criteria for recursive partitioning is based on hypothesis tests of independence between covariates and the response variable, rather than an arbitrary maximum tree depth (Hothorn et al., 2006).

Using the *cforest* tool in the R *party* package (Strobl et al., 2009), we generated a series of conditional inference RF models to predict stream temperature signatures (ST_{max} and TS) at a

monthly time scale using 23 quantitative variables as predictors. Each RF model was created using a geographically stratified training set (80%) and validated against a test set (20%). Model hyperparameters (*mtry* and *ntree*) were tuned using 5-fold cross-validation on input training sets to prevent overfitting. We fit the models using a *ntree* value of 1000, an *mtry* value of 10, a bootstrapped fraction of 0.63 without replacement, and a minimum splitting criterion of 0.1. For each stream temperature signature and unique subset of input sites, we created 12 independent models corresponding to each month of the year. Performance for each of these RF models was assessed using test set error, applied to 20% of sites not used in model training. Further analysis of error is reported in the results section.

To further investigate spatial drivers of stream temperature regimes, we subdivided the full dataset into smaller subsets based on geographic regions, representing gradients in climate and physiography, and the influence of dams. The geographic subsets included sites from three data-rich hydrologic regions: New England – Mid-Atlantic (HUC 01/02, $n = 58$), South Atlantic – Gulf (HUC 03, $n = 65$), and Pacific Northwest (HUC 17, $n = 96$). We also filtered the input sites into two subsets based on the presence of a major dam in each watershed (Major Dam, $n = 239$; No Major Dam, $n = 171$) to disentangle the relationship between dams and stream temperature regimes. In the GAGES-II dataset, major dams are defined as dams greater than 15 m in height or with a total storage greater than 6 million m³ (Falcone et al., 2010).

2.4.2. Identifying Influential Predictors of Stream Temperature Signatures

The output of random forest models, like other so-called ‘black box’ machine learning algorithms, may be challenging to interpret in isolation (Ribeiro et al., 2016). However, variable importance metrics help to translate the ensemble of trees into an understandable explanation of the model’s behavior. Permutation variable importance may be used to rank predictor variables

based on their relative utility in predicting model output. These importance algorithms function by randomly permuting each predictor variable while holding all others constant, mimicking the removal of that predictor from the ensemble of trees in the model (Strobl et al., 2008). After permuting each variable, it is possible to assess the decrease in model accuracy to measure the relative importance of a specific variable to model performance.

While random forest models tend to be resistant to multicollinearity in explanatory variables (Konapala & Mishra, 2020), permutation importance (hereby referred to as ‘marginal permutation importance’) may struggle to correctly identify important predictors in datasets with highly correlated variables (Strobl et al., 2008). When a variable is randomly permuted, its correlative links with other predictor variables are broken, disrupting the underlying structure in the dataset. As datasets containing environmental variables tend to have strong multicollinearity (e.g., elevation and precipitation), marginal permutation importance may be a sub-optimal option for assessing variable importance. Conditional permutation importance, introduced by Strobl et al. (2008), is an alternative RF importance metric that accounts for predictor correlation in a dataset. Rather than permuting each variable randomly while holding others constant, conditional permutation importance permutes predictor variables of a conditional inference tree model only within set ranges of other variables, preserving the dataset’s correlation structure (Strobl et al., 2008). Given the outlined benefits, we make use of this method in our study.

For all models, we computed the conditional permutation importance for each predictor variable, ranking predictors by their relative importance in simulating the stream temperature signatures (Strobl et al., 2008). Variable importance values were normalized within each RF model to enable comparisons in variable rankings between models. Although RF importance metrics indicate the relative importance of each variable in predicting a temperature signature,

they give no information on the direction of that importance. Therefore, we also quantified the Spearman's rank correlation, a measure of nonlinear association, between temperature signatures and each predictor variable to characterize the direction of impact for each variable of interest.

3. Results

3.1. Observed Spatial and Temporal Variability in Stream Thermal Regimes

At the 410 selected water temperature gages during the 2016-2020 study period, we observed considerable space-time variability in the calculated ST_{max} and TS metrics across the conterminous US (Figure 4). Across all sites and months, monthly ST_{max} had a mean value of 15.2°C and a range of 0°C (at many sites during winter months) to a maximum of 32.0°C. TS had a mean value of 0.3 across all months and sites, ranging from -0.4 to 0.9.

During winter months, both ST_{max} and TS were generally organized along a North-South latitudinal trend (Figure 4a, 4c). Sites in the Northern and Rocky Mountain regions of the US has ST_{max} values ranging from 0 to 5°C and TS values ranging from -0.1 to 0.2. In comparison, streams in the Southeastern US tended to have higher ST_{max} values and TS values closer to unity. TS appeared to be more heterogeneous at regional scales than ST_{max} during winter months.

In contrast, summer ST_{max} and TS values did not display the same latitudinal organization as observed during winter months (Figure 4b, 4d). Instead, ST_{max} tended to be relatively homogeneous and warmer (25-32°C) in the Eastern US, and generally cooler and more variable in the Western US. Summer TS displayed a similar, albeit weaker, east-to-west pattern as ST_{max} , though variability in TS appeared to be greater between proximal basins than at large regional scales. In the US Northeast, TS varied noticeably between neighboring watersheds. Across most sites selected for analysis, summer TS was larger in magnitude (nearer to unity) as compared with winter values.

3.2. RF Model Error

Test set error for all RF models is organized in Table 1. Error metrics are presented as RMSE normalized against the mean observed test set value (RMSEn, %) to enable comparisons between months and metrics. Additional plots of model training and test set error are included in Supporting Information.

Overall, models of monthly ST_{max} outperformed models of TS in all months for comparable subsets of input sites. In general, test set error tended to be higher in winter for models of ST_{max} and higher in summer for models of TS , though this pattern was not universal across all months and subsets of sites. Models of undammed sites had lower error than both models of dammed sites and all sites for each of the studied metrics. The normalized RMSE of ST_{max} predictions in each of the selected hydrologic regions was relatively low; however, other metrics of error suggest that these values may be artificially depressed by the lack of variability in the magnitude of observations within each region. We also concluded that the error of TS model predictions in the HUC-specific models was too large for further interpretation, and as such, the importance results of these models are excluded from this analysis.

3.3. Exploring Inferred Controls on Stream Temperature Variability

Using a combination of variable importance metrics and Spearman's rank correlations, we inferred the dominant drivers of variability of monthly maximum stream temperatures (ST_{max}) and thermal sensitivity (TS). Air temperature, the average annual percentage of precipitation falling as snow, baseflow index, and watershed slope were the most important variables in predicting ST_{max} at all sites (Figure 5a). Air temperature was the strongest predictor of ST_{max} in all months, though its importance diminished in June, July, and August. Across all months, climate was the leading predictor category for ST_{max} models (Figure 7a). Predictor importance

metrics of TS showed more variability than those of ST_{max} models.

In contrast, dam storage volume was the most important predictor of TS for all sites and across all months, with larger storage volumes corresponding to less variability in stream temperature with respect to air temperature (Figure 6a). This dominance of dam storage was most apparent in summer and early autumn, with this single predictor accounting for 76.3% and 81.2% of total August and September importance, respectively. Air temperature and snow percentage also emerged as important winter predictors of TS , accounting for 48.6% and 12.2% of total January importance, respectively.

Given the clearly identified influence of dams on stream thermal regimes both in this study and in published literature (Ahmad et al., 2021; Casado et al., 2013), we also generated RF models for monthly temperature signatures at sites without major upstream dams. Models of sites with major dams showed similar patterns of importance to models of the full dataset, and as such, are only presented in Supporting Information. Considering only sites with no major upstream dams, we observed a strong shift in the dominant predictors of TS (Figure 6b; 7d) and negligible change in the predictors of ST_{max} (Figure 5b; 7c). Much like the models of ST_{max} at all sites, air temperature and snow percentage made up the largest proportion of relative importance in predicting stream temperature magnitude at undammed sites. The aggregate predictor importance values of ST_{max} were relatively unchanged between models for all sites versus models for undammed sites (Figure 7a; 7b). Climate predictors became slightly more important and less seasonally variable in undammed ST_{max} models, along with a slight loss of importance in hydrological predictors during summer months. With the removal of dammed sites, the leading predictors in TS models shifted away from dam storage to a broader suite of variables (Figure 6b). These predictors, which tended to be temporally inconsistent, included air temperature,

measures of precipitation, snow percentage, baseflow index, and watershed slope. The relative magnitude of these importance metrics was highly variable from month to month. At sites without dams, we identified climate as the most important winter predictor class and hydrology variables as the dominant summer controls on *TS* (Figure 7d).

By modeling stream temperature signatures at a monthly time-step, we were able to discern seasonal patterns in the controls of variability that would not be revealed by annual-scale models. We observed these time-varying patterns in predictors of both ST_{max} and *TS* for each of the modeled subsets, though to differing degrees. While the two most important predictors of ST_{max} (air temperature and snow percentage) were generally consistent throughout the year, their relative importance decreased from late spring to early autumn. During these summer months, a broader set of predictors, including baseflow index, drainage area, watershed slope, and dam storage, had increased importance in predicting ST_{max} (Figure 5a). We observed this pattern in ST_{max} models of all sites and undammed sites, with winter variance primarily explained by air temperature and summer variance explained by a suite of predictors from each of the four variable categories.

Seasonal variability in predictor importance was notably more pronounced in models of *TS*. As noted above, dam storage was the leading predictor of *TS* during summer months, though its relative importance diminished during winter and early spring (Figure 6a). While dam storage did contribute roughly 20% of relative importance during these months, monthly air temperature became the most importance winter predictor of *TS* in December, January, and February. We found a similar seasonal contrast in predictors of *TS* at undammed sites, though summer models identified baseflow index, snow percentage, watershed slope, and watershed elevation as relatively important predictors instead of dam storage. Spearman rank correlations tended to

have little seasonal variability in direction for each predictor across months.

Using smaller scale geographic models (HUC-01/02: New England-Mid-Atlantic; HUC03: South Atlantic-Gulf; HUC17: Pacific Northwest), we observed regional variability in the leading controls of ST_{max} . While all three regional models identified air temperature as a relatively important predictor, each region also had unique predictors within the categories of human impacts, hydrology, and watershed characteristics. In addition to air temperature, influential predictors of ST_{max} included snow percentage, dam storage, watershed elevation, and forest land cover in the New England-Mid Atlantic region and snow percentage, wetland land cover, water table depth, and baseflow index in the South Atlantic-Gulf. In the Pacific Northwest, snow percentage was the leading predictor across all months, peaking in importance in January and February. Agricultural land cover (linked to warmer maximum STs), forested land cover, riparian forest land cover, and watershed slope (linked to cooler maximum STs) were also influential predictors of ST_{max} in the region. Our results from the three selected hydrologic regions highlight the complexity of basin-scale controls on stream temperature.

4. Discussion

4.1. Revealing Relative Controls on Stream Temperature Variability

In an era of changing climate and increasing human influence on the landscape, it has become increasingly necessary to better constrain the drivers of stream thermal variability across spatial and temporal scales (Steel et al., 2017). The results of our models confirm that river temperatures are influenced by a complex set of variables describing climate, hydrology, watershed characteristics, and human impacts (Caissie, 2006; Hannah & Garner, 2015; Webb, 1996). Within this nested predictor set, regional climate is often recognized as a first-order control on stream temperature magnitude, given the strong coupling between latitudinal variance

in net radiation and water temperatures (Garner et al., 2014; Hannah & Garner, 2015). As the leading predictor of ST_{max} in all our models, the high relative importance of monthly air temperature reflects this overarching control of climate on continental-scale variability (Figure 5; 7). Air temperature is often used as a predictor in models of water temperature magnitude when heat fluxes at a site are unknown, acting as a proxy for heat received by a stream from incoming solar radiation (Arismendi et al., 2014; Caissie, 2006; Casado et al., 2013; Lisi et al., 2015). As air temperature tends to vary along the same latitudinal gradient as stream temperatures (Figure 4), particularly during winter months, the variable is useful in predicting ST_{max} at a continental scale.

We found that the percentage of annual precipitation received as snow (snow percentage) was the second-most influential predictor of ST_{max} in models of all sites and undammed sites (Figure 5). Snowmelt runoff represents a considerable proportion of spring and summer streamflow in seasonally cold watersheds and can drastically alter stream thermal regimes (Leach & Moore, 2014; Michel et al., 2020; Yan et al., 2021). When streamflow is sourced from snowmelt, the relative temperature of its relatively cool source water can overcome radiative heat fluxes, resulting in buffered water temperatures (Blaen et al., 2013; Leach & Moore, 2014). We observed a similar effect, where predicted ST_{max} tended to be lower at sites that received higher proportions of precipitation as snowmelt. However, it's not clear whether the importance of snow in our models reflects the hydrologic relationship between streamflow source and water temperature, or if the predictor is simply another manifestation of climate gradients. Continental-scale patterns in snow percentage coincide closely those of air temperature, making it difficult to disentangle the spatial correlation between the variables and isolate the true source of the predictor's influence. Furthermore, within the GAGES-II dataset, the snow percentage variable

only represents a static, average annual proportion of precipitation rather than a more detailed record of monthly snowpack volume. Therefore, the snow percentage predictor likely derives some of its importance in models of ST_{max} by acting as a secondary proxy of regional climate. Nonetheless, our models show tentative evidence of a mechanistic link between snow and colder water temperatures.

While we attribute broad spatial patterns in stream thermal regimes to regional climate variability, basin- to reach-scale characteristics impart an additional layer of complexity on stream thermal regimes. These watershed characteristics, which include riparian shading, land cover, and subsurface hydrology among others, modify a stream's radiative and hydrologic inputs to produce varied thermal responses from a single set of climate forcings (Hannah & Garner, 2015; Orr et al., 2015). Our second stream temperature metric, the slope of the regression relationship between air temperatures and water temperatures (TS) is an explicit representation of how basin-scale processes moderate broad-scale climate conditions (Kelleher et al., 2012). Thermal sensitivity has been linked to measures of baseflow index, stream size, snowpack, watershed slope, and elevation (Beaufort et al., 2020; Kelleher et al., 2012; Lisi et al., 2015; Segura et al., 2015). Geographic trends in TS were weaker than those observed in ST_{max} (Figure 4), confirming prior findings that metrics of water temperature sensitivity tend to be highly variable at regional and sub-regional scales (Kelleher et al., 2012).

Our analysis captured not only the importance of climate, but equally the moderating effects of basin-scale controls in importance metrics of ST_{max} and TS models (Figure 5; 6). In models of ST_{max} , a combination of hydrologic and watershed characteristics variables gained in relative importance during summer months. The importance of baseflow to both ST_{max} (Fig 5) and TS at undammed sites (Figure 6b) aligns with our expectations of groundwater-induced

cooling of water temperatures. Summer stream temperatures are often depressed and decoupled from air temperatures along reaches receiving large fluxes of groundwater inflow due to the relatively cold temperature of groundwater (Beaufort et al., 2020; Kelleher et al., 2012). We also identified two proxies of hydrologic residence time, drainage area and watershed slope, as influential geomorphic predictors of ST_{max} and TS (at undammed sites) (Figure 5a, 6b). In large watersheds draining low-gradient terrain, long hydrologic residence times allow surface waters to accumulate radiative inputs and reach thermal equilibrium (Caissie, 2006; Garner et al., 2014; Lisi et al., 2015). In contrast, smaller and steeper watersheds tend to produce buffered and cooler thermal responses due to their shorter residence times (Lisi et al., 2015; Segura et al., 2015). Watershed elevation also imparted a buffering effect on models of TS at undammed sites, likely due to correlative links between elevation, watershed slope, and snowpack (Figure 6b). During winter and into early spring, TS had a strong positive relationship with air temperatures, where colder temperatures corresponded to decoupled air-water temperatures. While it was challenging to identify a definitive mechanism for this control, prior studies have observed that air and water temperatures tend to become disconnected when water temperatures approach 0°C (Kelleher et al., 2012; Mohseni et al., 1998). This effect could be attributed to river ice cover, which would insulate waters from radiative forcings, or to the presence of snow, that melting of which would provide an external source of cold water to buffer waters against temperature change (Brown & Hannah, 2008; Mellor et al., 2017). Rain-on-snow events could also contribute relatively cool water to streams, further decoupling air and water temperatures (Leach & Moore, 2014).

While our findings do agree with the existing literature, it is important to keep in mind that our findings are based on empirical inference. Though RF models can effectively detect influential predictors of temperature behavior, our analysis does not pinpoint how a certain

predictor affects a stream's thermal regime. To alleviate this shortcoming, deterministic water models, which explicitly quantify energy and fluid fluxes, could be paired with machine learning models to confirm the role of a predictor of interest.

4.2. Anthropogenic Influence on Stream Thermal Regimes

The river network of the continental US, through a wide range of anthropogenic modifications, has been heavily fragmented, disrupted, and impounded (Dynesius & Nilsson, 1994; Grill et al., 2019). The construction of dams, deforestation of riparian buffers, channelization, and urbanization have interrupted the natural hydrologic regimes of free-flowing rivers and in turn, dramatically altered their thermal behavior (Grill et al., 2019; Hester & Doyle, 2011). While each catchment is affected by a unique suite of anthropogenic influences, human disturbances generally tend to cause net warming of water temperatures through loss of shading, reduced groundwater exchange, and the effects of a warming global climate (Caissie, 2006; Hester & Doyle, 2011; Hill et al., 2014; Somers et al., 2013). The most notable exception to this general trend of anthropogenic warming is that of dams and their associated upstream reservoirs. Along river reaches downstream of large dams, the seasonal patterns of stream thermal regimes are altered, often resulting in cooler and more insensitive water temperatures in the summer and warmer temperatures in autumn and winter seasons (Ahmad et al., 2021; Angilletta et al., 2008; Casado et al., 2013; Kelleher et al., 2012). This effect is most pronounced for dams that bottom release from the relatively cold hypolimnion of deep, thermally stratified reservoirs (Cheng et al., 2020).

The anthropogenic impacts on stream thermal regimes are readily apparent in models of both ST_{max} and TS , overprinting other natural controls. Across all sites, the volume of dam storage (normalized to watershed area per the National Inventory of Dams) was the single most

important predictor of TS (Figure 6a; 7c). In our dataset, larger volumes of dam storage in watersheds corresponded to decoupled air-water temperatures and more insensitive water temperature responses. The importance of dam storage in models of TS likely illustrates the strong contrast in sensitivity between undammed streams, where water temperatures tend to be more atmospherically coupled, and impounded streams, which have relatively flat TS relationships (Kelleher et al., 2012). These effects were mirrored, though to a lesser degree, in models of monthly ST_{max} . Although dam storage was relatively uninfluential across all sites (Figure 5a), reservoir volume exhibited a strong negative correlation with predicted summer ST_{max} for dammed sites (Figure S2). The diminished importance of dam storage in models of ST_{max} relative to models of TS could indicate that extreme temperatures, although moderated by the effects of reservoirs, are still primarily controlled by atmospheric forcings.

Amongst sites in impounded watersheds, the apparent cooling and atmospheric-decoupling effects of dams may also reflect the contrast in hydrologic function between impoundments of varying sizes. Though large reservoirs often cause cool and insensitive downstream temperatures (Casado et al., 2013; Cheng et al., 2020), comparatively smaller storage dams can induce the opposite effect, resulting in downstream warming. For instance, Maheu et al. (2016) found that releases from shallow impoundments (depth < 6 m), which have a larger proportion of surface area exposed to radiative heating, caused warming of summer stream temperatures by up to 2°C. Given that our dataset contains a wide range of reservoir volumes, the importance of dam storage in our models may highlight these opposing end-member behaviors. We also emphasize that although the mean distance from gages effected by impoundments ($n = 239$) to the nearest major dam was 18.3 km, we still observed the alteration of stream thermal regimes by dams. This underscores the longitudinal persistence of dams' downstream thermal

effects, which, under some circumstances, can still be observed up to 100 km downstream (Angilletta et al., 2008; Lowney, 2000; Willis et al., 2021). The importance of dam storage in our models clearly indicates the necessity to better understand the thermal implications of dam operations on river networks to accurately predict future stream temperatures.

While dams may embody the most direct disruption of natural hydrologic systems, urbanization, climate change, and other human actions on the landscape also intersect with stream thermal regimes (Hester & Doyle, 2011; Kurylyk et al., 2015; Steel et al., 2017). Deforestation of riparian buffers (Wondzell et al., 2019), declines in groundwater inflow due to pumping and other diversions (Kurylyk et al., 2015), warmer air temperatures (Hill et al., 2014; Wondzell et al., 2019), and power plant effluent discharges (Miara et al., 2018) all been linked to warmer stream temperatures (Hester & Doyle, 2011). Mechanistically, anthropogenic activities alter hydrological and thermal process and yet, outside of dam storage, little of this influence is captured in our models. There are several potential explanations for this lack of importance given to human-related predictors. First, while our models contain key predictors that explicitly represent anthropogenic disruption, there are many other impacts associated with human activity that cannot be determined at the scale of our investigation. Additionally, anthropogenic impacts are also implicitly represented by their modification of predictors such as air temperature, riparian forest land cover, and baseflow index. As human activities on the landscape overprint natural hydrologic and thermal responses, it is difficult to discern whether the importance of a particular predictor such as baseflow index is controlled by natural patterns in subsurface flow or human disruption of hydrologic systems. Finally, the paucity of sites used in this study (due to limited ST observations) makes it likely that our models undersample urban streams and therefore underrepresent the influence of humans on stream thermal regimes. Of the urbanized

sites that we do include, our models may struggle to adequately represent stream temperature behavior given the inherent difficulty in modeling the complex variability of urban hydrologic systems.

4.3. Seasonal Variability in Key Controls of Stream Temperatures

The importance of individual predictors, as well as the aggregate importance of predictor categories, tended to vary seasonally for both modeled ST metrics in response to the relative strength of dynamic drivers (Figure 5; 6; 7; 8). In models of ST_{max} at all sites, this effect was best demonstrated by the relative rise in importance of baseflow index, drainage area, dam storage, and watershed slope during summer months (Figure 5a). We observed a similar summer decline in the influence of climate predictors accompanied by a shift to a more complex suite of basin-scale predictors for TS (Figure 6). Prior modeling studies of stream temperature regimes have found comparable summer-winter contrasts in influential model predictors. In monthly multiple linear regression models of maximum and mean stream temperature, respectively, Hrachowitz et al. (2010) and Imholt et al., (2013) noted that summer models of thermal behavior required more parameters than winter models. These results, coupled with findings from our models, suggest that the controls on both stream temperature magnitude and TS are more complex during summer months (Hrachowitz et al., 2010; Imholt et al., 2013).

While these seasonal patterns in importance are logically supported by our understanding of stream thermal processes, it is also possible that a component of this signal may be an artifact of ensemble learning algorithms and associated variable importance metrics. When applied to RF models, variable importance metrics assess which predictors are most useful in differentiating the behavior of a metric across a set of sites. In winter, we observe strong latitudinal contrasts in both air and stream temperatures in the conterminous US, which likely reflect differences in

incoming radiation received by between Northern and Southern sites (Figure 4a). These coincident spatial patterns make air temperature an effective winter predictor of ST_{max} (Figure 5). This does not mean that regional climate and radiative inputs are less influential controls on summer stream temperatures at individual sites, only that that differences in thermal behavior across sites are not easily explained by summer climate.

4.4. Implications for Watershed Management

Globally, climate change is expected to alter the thermal regimes of streams, with the warmest water temperatures projected to increase by up to 2°C (Ficklin et al., 2013; van Vliet et al., 2013; Wanders et al., 2019). Watershed management strategies, including riparian planting, preservation of natural land cover, and groundwater and flow manipulation, have been employed to resist this regime shift and to preserve cold-water refugia for sensitive instream ecosystems (Beaufort et al., 2020; Kurylyk et al., 2015). Machine learning and big-data investigations of stream temperature can supplement these efforts, as such approaches are capable of efficiently identifying the most influential atmospheric and landscape controls on thermal processes from large environmental datasets. We do, however, acknowledge that the conclusions drawn from continental-scale models may be challenging to translate into actionable management solutions at specific sites. The rivers of the US drain catchments that span a diverse spectrum of physiography and climate, and as such, it is probable that stream thermal regimes in each ecoregion are influenced by different landscape processes. The spatially-unfocused continental-scale RF models likely struggle to capture this regional heterogeneity in controls.

Regional models of water temperature behavior present a promising alternative to nationwide models, as they can discern relevant controls at scales better suited to the management of individual watersheds. In our models of ST_{max} in three distinct regions, we show that although

climate remained a dominant influence across the US, its influence was mediated by a suite of watershed-scale predictors unique to each geographic area (Figure 8). For example, wetland land cover and water table depth were relatively influential predictors of ST_{max} in the South Atlantic-Gulf region, yet the variables showed negligible importance in the other two selected regions (Figure 8b). Similarly, riparian forest land cover had a strong negative effect on spring water temperatures in the Pacific Northwest, an effect not observed in the US Northeast or South (Figure 8c). These contrasts confirm that much like the drivers of hydrologic behavior, the controls on stream thermal regimes are also non-uniform across the US, with considerable implications for the management of water temperatures (Gannon et al., 2022; Hammond et al., 2021). This spatial heterogeneity suggests that the efficacy of successful thermal management strategies in one geographic region may not translate into similar levels of success elsewhere. However, regional inference is limited in under-monitored regions such as the US Southwest. With additional stream temperature monitoring, machine learning models of stream temperatures at sub-regional scales could become increasingly practical. Future investigations of landscape controls on stream thermal regimes should attempt to discern influential processes at local spatial scales. By leveraging the information provided by data-intensive, yet regional, models, thermal management strategies can be tailored to target watershed modifications that are most effective for a specific region (Hrachowitz et al., 2010; Imholt et al., 2013).

5. Conclusions

In this study, we evaluated the distribution of two water temperature metrics, ST_{max} and TS , across 410 rivers in the conterminous US and used RF models to infer the dominant space-time drivers of stream thermal regimes. Our results highlight that, above all else, there is not a single set of universal controls that comprehensively explain stream temperature behavior.

Instead, we found that the most influential controls on select water temperature metrics varied across a range of predictor categories (climate, watershed characteristics, hydrology and human influence). Broadly, maximum water temperatures were linked most closely to air temperatures, although the role of climate was mediated by basin-scale processes - especially in summer months.

The presence of major dams was the most influential predictor of thermal sensitivity within our dataset, illustrating the profound impact anthropogenic disturbances have on stream thermal processes. Often, these human impacts on river water temperature have been somewhat overlooked (Ficklin et al., 2023). Although many approaches to the study of stream temperatures focus solely on the warmest summer water temperatures, we highlight explicitly the need to explore the controls on additional modes of water temperature variability in both space and time. While we address maximum temperatures and thermal sensitivity at a monthly time step in our models, there are many other stream temperature metrics of ecological importance, including degree-day accumulation, diurnal range, and the timing and frequency of extreme events (Casado et al., 2013; Steel et al., 2017). These metrics will almost certainly be controlled by unique combinations of land surface processes, and a comprehensive understanding of their dominant drivers will allow for the development of more effective management approaches.

As well as providing new information on river water temperature controls, we believe our results make a methodological contribution. Our analysis exemplifies the discerning power of machine learning approaches when applied to large and complex environmental datasets. Random forest algorithms and related tree-based models are relatively simple ML applications, yet they allow us to resolve complex and seasonally variable controls on stream thermal regimes. While the findings of our analysis demonstrate the advantages provided by big-data approaches

to hydrology, they also emphasize that caution and nuance are needed in translating model results into human-interpretable insights. The types of models used in this study rely only on statistical relationships between water temperature behavior and landscape covariates, meaning that their implications are inferential rather than causal. To extend confidence in predictions and to support management predictions, we recommend that machine learning approaches should be supplemented by process-based hydrologic understanding and models to confirm the influence and interaction of variables across individual sites. Nonetheless, by powerful comparative analysis, our research yields new perspectives into the time-varying controls water temperature regimes across watersheds that span strong gradients in climate, hydrology, watershed characteristics, and degrees of human impacts.

Figures

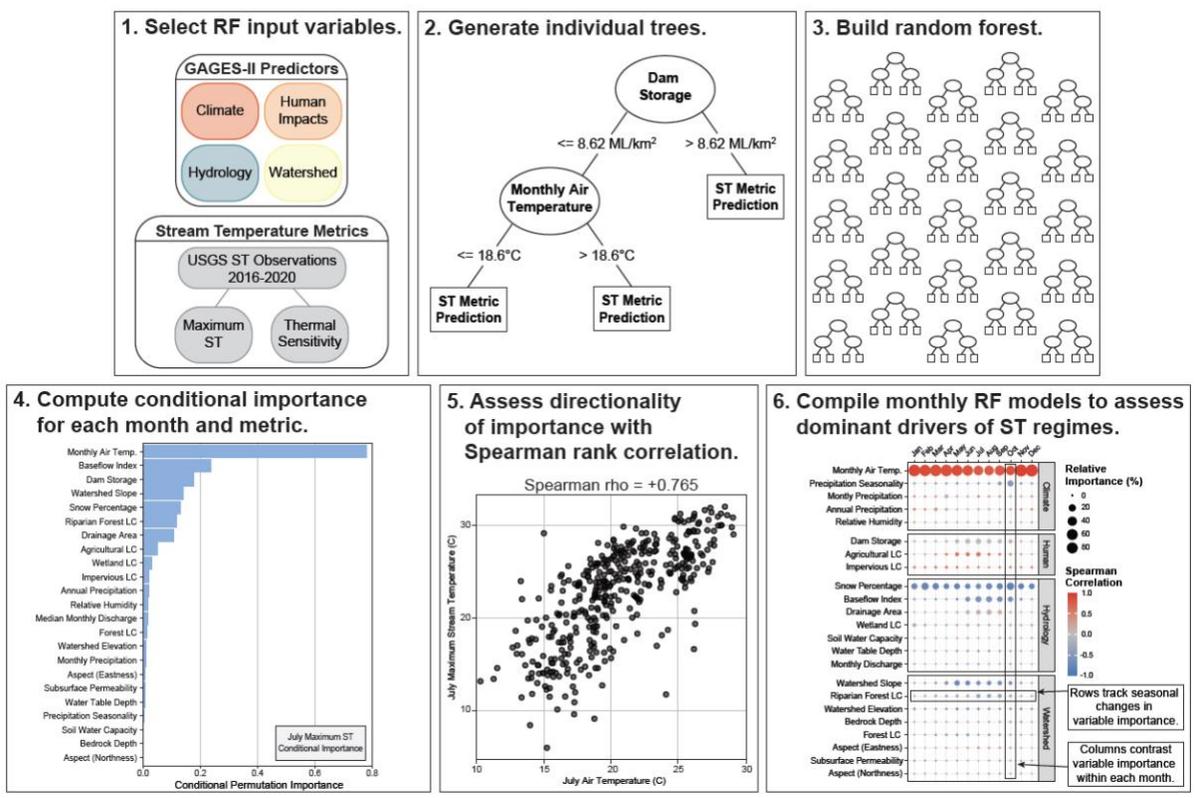


Figure 1. Graphical representation of the analytical steps used to fit RF models and assess the dominant drivers of stream temperature regimes through time.

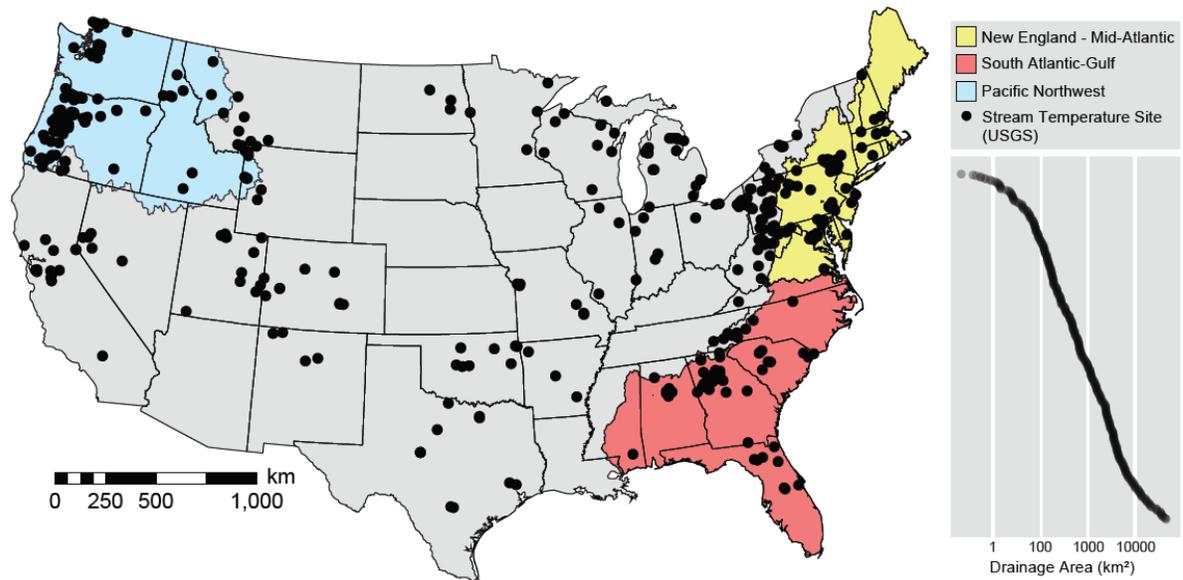


Figure 2. Locations of 410 USGS gages present in the GAGES-II dataset with complete records of daily stream temperature and discharge from 2016 to 2020. Hydrologic units selected for further analysis are highlighted. The watersheds of the selected gages span a wide and continuous range of contributing areas (see inset).

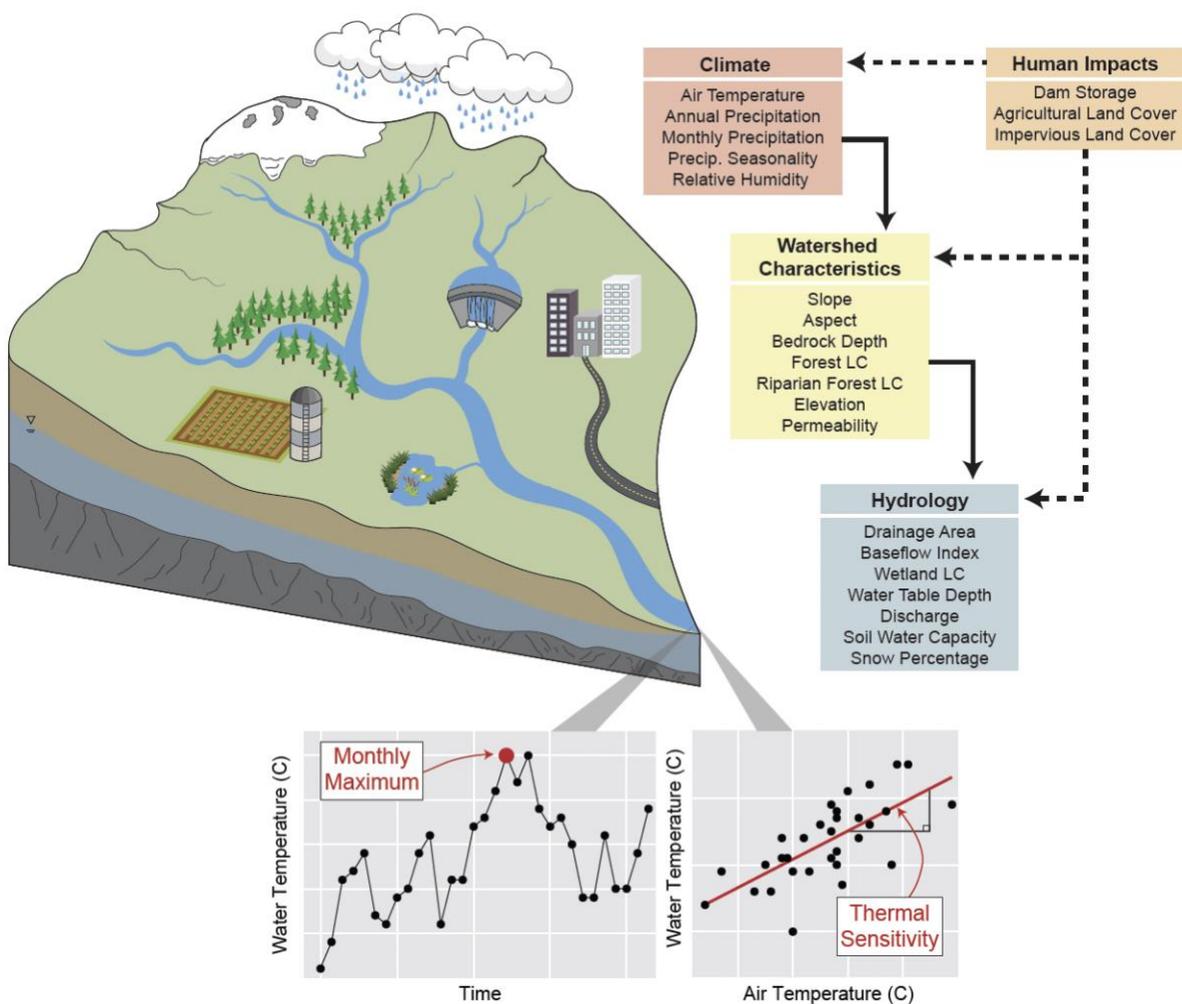


Figure 3. Conceptual depiction of the complex nature of the controls of stream temperature regimes. Random forest predictor variables selected from the GAGES-II dataset are displayed in their respective categories. The two quantitative stream temperature signatures, monthly maximum temperature (ST_{max}) and thermal sensitivity (TS), arise from the nested interactions of climate, watershed characteristics, hydrology, and human impacts.

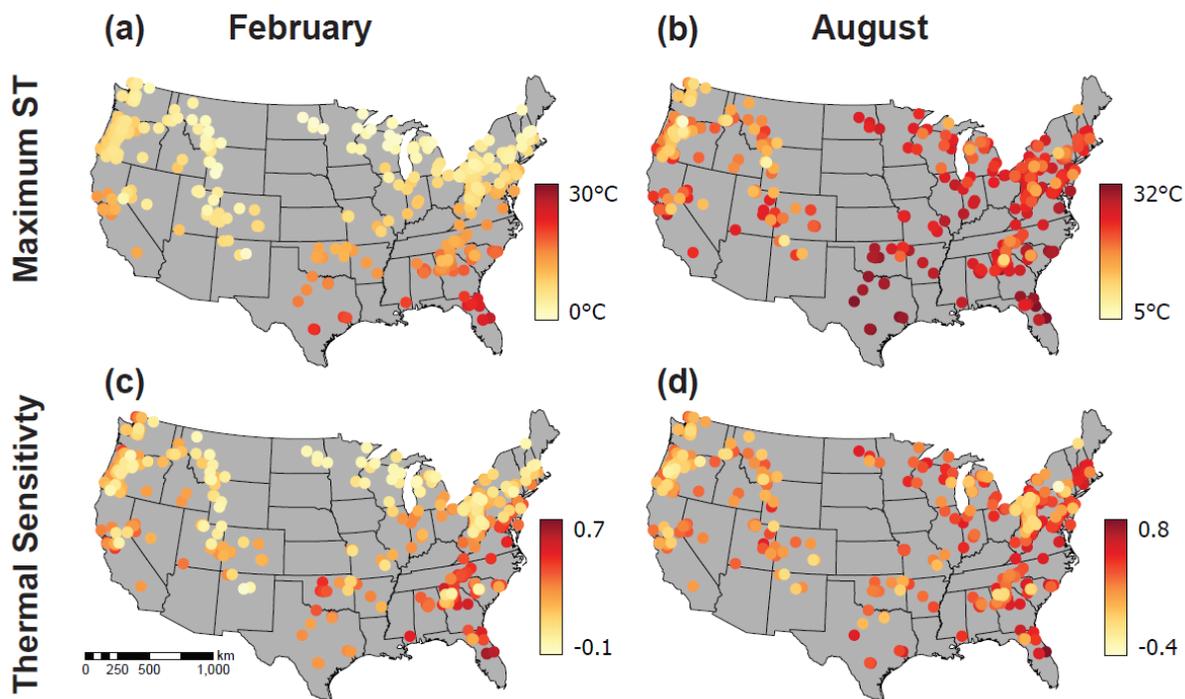


Figure 4. Spatial variability of (a, b) maximum stream temperature (ST_{max}) and (c, d) thermal sensitivity (TS) in February and August across the conterminous US.

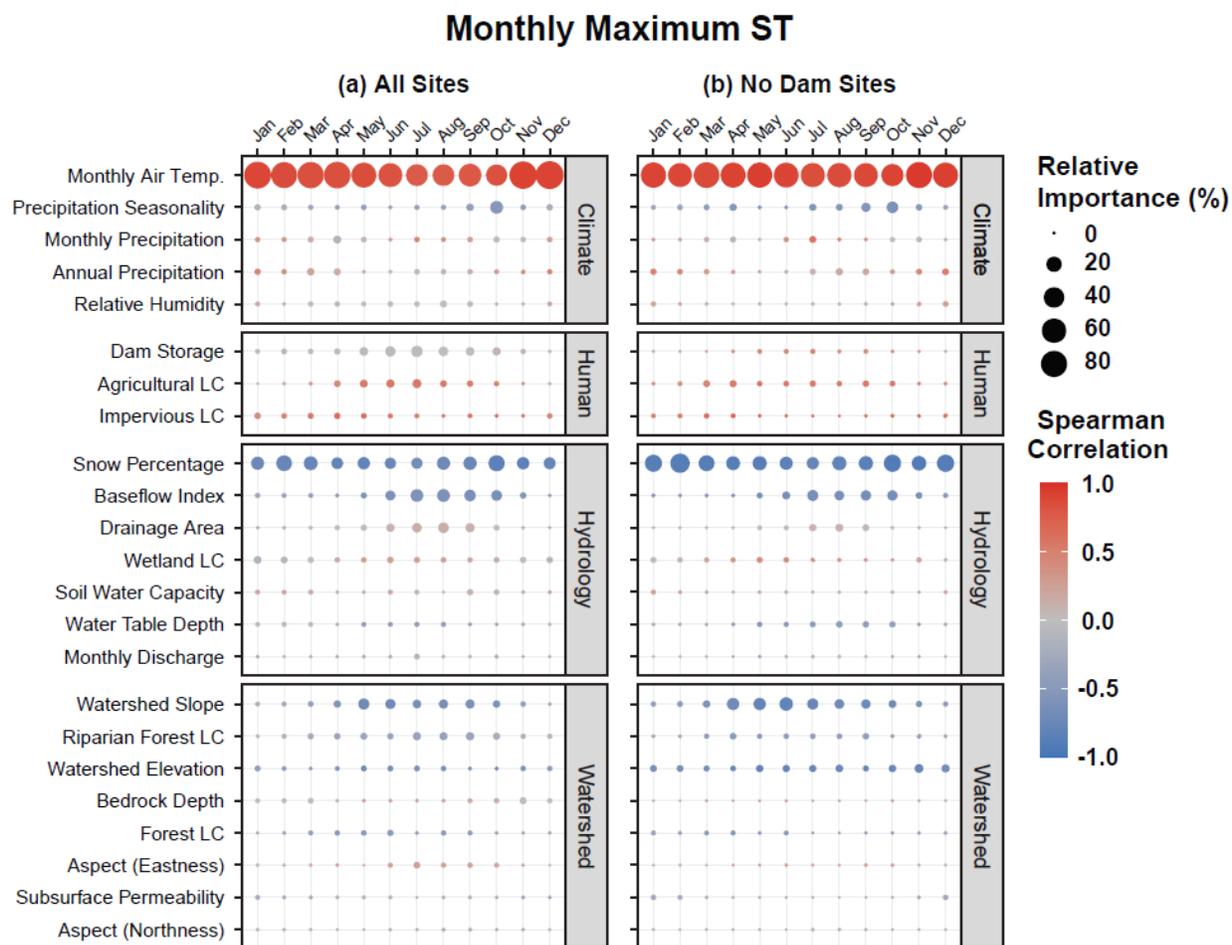


Figure 5. Aggregated heatmap of RF relative variable importance in predicting monthly maximum stream temperatures (ST_{max}) for (a) all sites and (b) sites with no major dam in their watersheds. The size of each point represents its magnitude of relative importance. The color of each point indicates the direction and strength of the relationship between a variable and the modeled metric as assessed by Spearman's rank correlations. Plot columns represent individual RF models for each month and display differences in relative importance between predictor variables. Plot rows track variability in the importance of a single predictor over months.

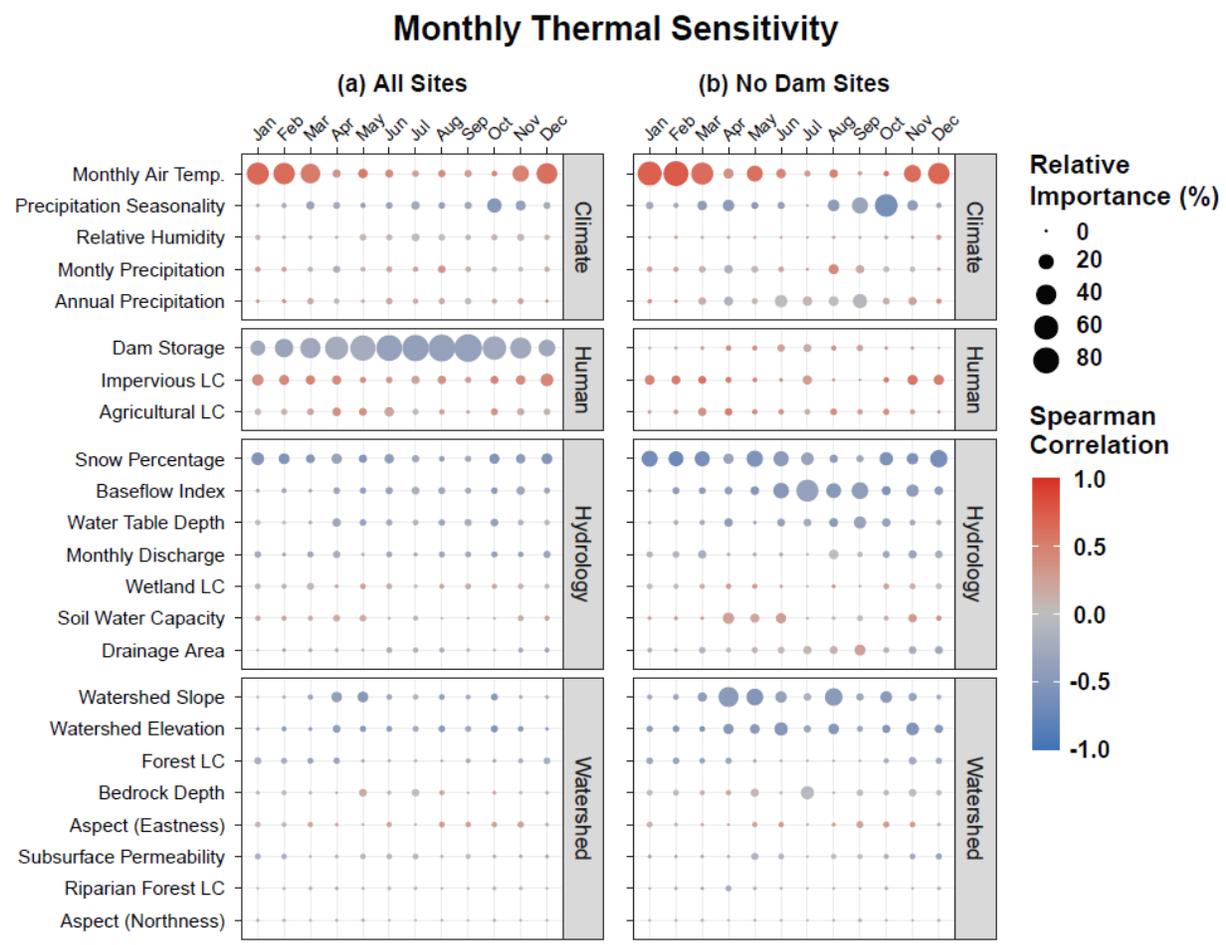


Figure 6. Aggregated heatmap of RF relative variable importance in predicting thermal sensitivity (TS; slope of linear relationship between daily air and water temperatures) for (a) all sites and (b) sites with no major dam in their watersheds. Symbology repeated from Figure 5.

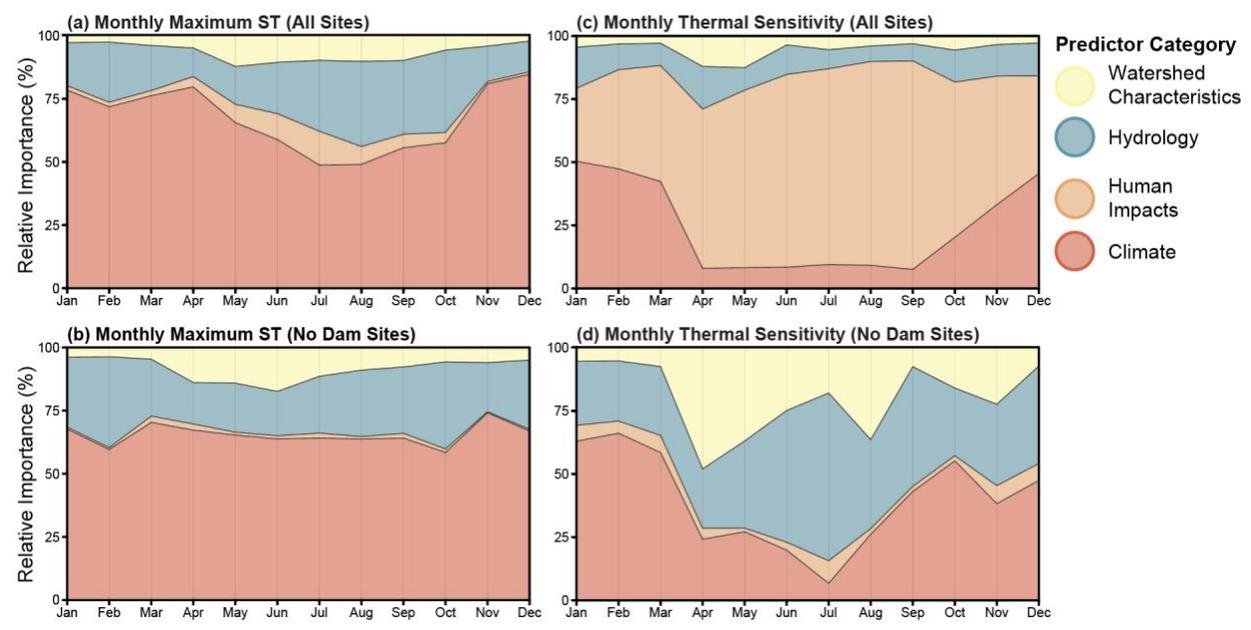


Figure 7. Cumulative relative importance of predictor variables classified within each of the four variable categories (climate, watershed characteristics, hydrology, and human impacts). Variable importance percentages over time displayed for monthly maximum stream temperatures (ST_{max}) at (a) all sites and (b) sites with no major dam in their watersheds, and for monthly thermal sensitivity (TS) at (c) all sites and (d) sites with no major dam in their watersheds.

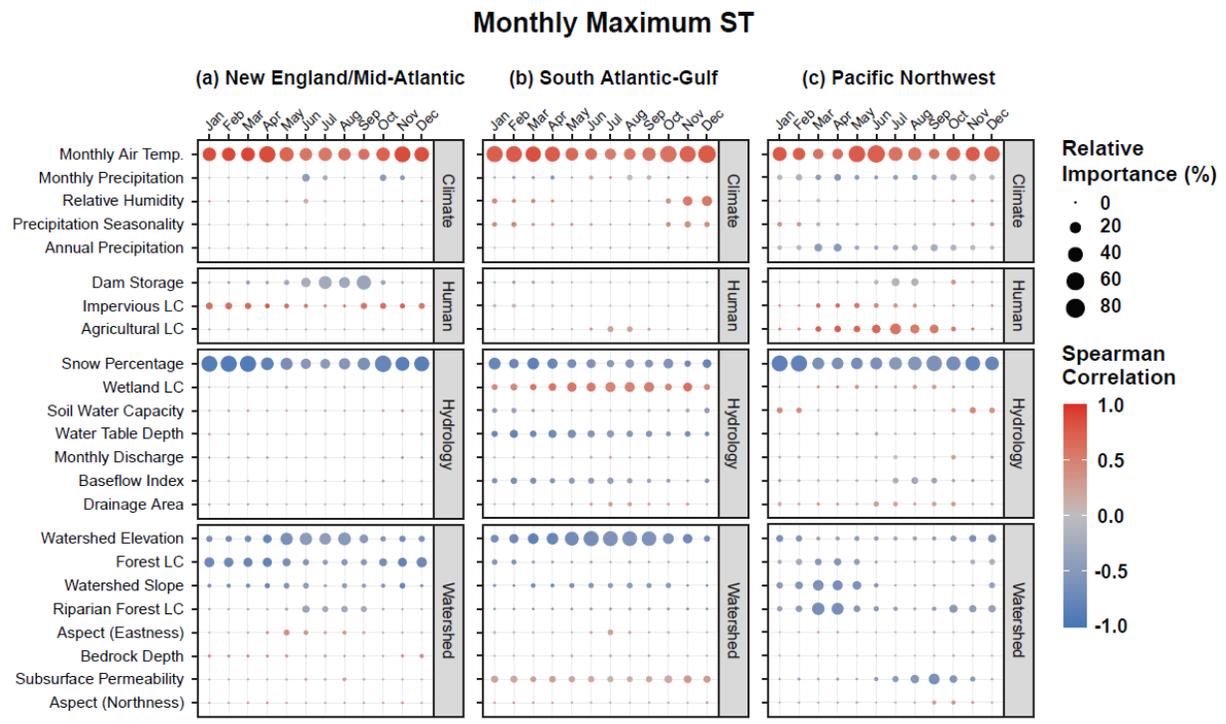


Figure 8. Aggregated heatmap of RF relative variable importance in predicting monthly maximum stream temperature (ST_{max}) for three hydrologic regions: (a) New England/Mid-Atlantic, (b) South Atlantic-Gulf, and (c) Pacific Northwest. Symbology repeated from Figure 5.

Tables

Table 1. Test set errors of monthly random forest models for maximum ST (ST_{max}) and thermal sensitivity (TS) across all sites, sites with major dams (Dam), sites with no major dams (No Dam), and sites in selected hydrologic unit code regions (HUC 01/02 - New England/Mid-Atlantic; HUC 03 - South Atlantic-Gulf; HUC 17 - Pacific Northwest). HUC-specific models of thermal sensitivity were excluded from analysis due to poor performance.

Month	Values in nRMSE (%)								
	Maximum ST						Thermal Sensitivity		
	All Sites	Dam	No Dam	HUC 01/02	HUC 03	HUC 17	All Sites	Dam	No Dam
Jan	21.7	27.2	20.4	22.4	13.5	14.6	49.2	48.3	39.9
Feb	19.9	26.2	23.3	23.4	14.7	15.8	46.0	43.5	30.3
Mar	16.7	23.1	18.7	18.5	12.5	20.1	38.5	40.9	27.6
Apr	15.6	18.3	13.3	14.1	13.3	20.3	39.0	44.7	23.2
May	15.4	16.0	9.0	13.2	16.4	18.2	37.8	37.9	18.6
Jun	16.2	15.8	7.5	13.2	17.0	17.6	42.9	44.8	27.0
Jul	17.2	15.3	8.6	14.1	18.2	18.3	51.5	65.1	30.4
Aug	16.4	14.5	8.6	12.4	17.3	17.3	49.6	59.7	32.2
Sep	14.7	13.5	8.8	13.2	15.2	15.5	45.4	51.8	26.1
Oct	11.9	12.9	11.3	10.8	13.4	16.5	33.3	33.4	19.9
Nov	13.2	16.0	13.6	9.7	8.3	13.1	36.3	31.6	23.0
Dec	18.0	23.2	20.4	17.6	13.0	17.3	47.7	49.0	28.6

Table 2. Summary statistics (minimum, mean, maximum) for maximum ST (ST_{max}) and thermal sensitivity (TS) across all sites.

Key: Min - Mean - Max		
Month	Maximum ST (°C)	Thermal Sensitivity
January	0.00 - 7.2 - 21.8	-0.06 - 0.2 - 0.8
February	0.00 - 8.2 - 24.6	-0.05 - 0.2 - 0.6
March	0.8 - 10.4 - 25.8	-0.02 - 0.3 - 0.6
April	2.4 - 13.8 - 27.6	-0.01 - 0.3 - 0.6
May	4.8 - 18.2 - 30.1	0.00 - 0.4 - 0.7
June	5.9 - 20.7 - 31.6	-0.08 - 0.4 - 0.8
July	6.0 - 22.7 - 32.0	-0.2 - 0.3 - 0.8
August	6.0 - 22.3 - 31.3	-0.4 - 0.3 - 0.8
September	6.3 - 20.8 - 31.0	-0.2 - 0.4 - 0.9
October	5.2 - 17.5 - 28.9	-0.05 - 0.4 - 0.7
November	3.0 - 12.1 - 25.9	-0.01 - 0.4 - 0.6
December	0.05 - 8.1 - 23.4	-0.1 - 0.3 - 0.6

Table 3. Descriptions of predictor variables used in RF models, adapted from Falcone et al.

(2010). (1) – (Falcone et al., 2010)

Variable Name	Category	Description (1)
Annual Precipitation	Climate	Mean annual precip (cm) at the gage location, from 800m PRISM data. 30 years period of record 1971-2000.
Relative Humidity	Climate	Site average relative humidity (percent), from 2km PRISM, derived from 30 years of record (1961-1990).
Precipitation Seasonality	Climate	Precipitation seasonality index (Markham, 1970; Dingman, 2002). Index of how much annual precipitation falls seasonally (high values) or spread out over the year (low values). Based on monthly precip values from 30 year (1971-2000) PRISM. Range is 0 (precip spread out exactly evenly in each month) to 1 (all precip falls in a single month).
Monthly Precipitation	Climate	Mean monthly precip (cm) for the watershed, from 800m PRISM data. 30 years period of record 1971-2000.
Monthly Air Temperature	Climate	Average monthly air temperature for the watershed, degrees C, from 800m PRISM data. 30 years period of record 1971-2000.
Dam Storage	Human Impacts	Dam storage in watershed ("NID_STORAGE"); megaliters total storage per sq km (1 megaliters = 1,000,000 liters = 1,000 cubic meters).
Agricultural LC	Human Impacts	Watershed percent "planted/cultivated" (agriculture), 2006 era. Sum of cultivated classes 81 and 82 in NLCD.
Impervious LC	Human Impacts	Watershed percent impervious surfaces from 30-m resolution NLCD06 data.
Snow Percentage	Hydrology	Snow percent of total precipitation estimate, mean for period 1901-2000. From McCabe and Wolock. (submitted, 2008), 1km grid.
Baseflow Index	Hydrology	Base Flow Index (BFI), The BFI is a ratio of base flow to total streamflow, expressed as a percentage and ranging from 0 to 100. Base flow is the sustained, slowly varying component of streamflow, usually attributed to ground-water discharge to a stream.
Wetlands LC	Hydrology	Watershed percent "wetlands", 2006 era. Sum of wetland classes 90 and 95 in NLCD.

Soil Water Capacity	Hydrology	Average value for the range of available water capacity for the soil layer or horizon (inches of water per inches of soil depth).
Water Table Depth	Hydrology	Average value of depth to seasonally high water table (feet).
Monthly Median Discharge	Hydrology	Median monthly discharge, averaged across a 4-year period from 2016-2020. Records downloaded from the USGS NWIS database.
Drainage Area	Watershed Characteristics	Watershed drainage area, sq km, as delineated in our basin boundary.
Forest LC	Watershed Characteristics	Watershed percent "forest", 2006 era. Sum of forested classes 41, 42, and 43 in NLCD.
Riparian Forest LC	Watershed Characteristics	Mainstem 100m buffer "forest", 2006 era. Sum of forested MAINS100_41, 42, and 43 in NLCD.
Subsurface Permeability	Watershed Characteristics	Average permeability (inches/hour).
Bedrock Depth	Watershed Characteristics	Average value of total soil thickness examined (inches).
Elevation	Watershed Characteristics	Mean watershed elevation (meters) from 100m National Elevation Dataset.
Watershed Slope	Watershed Characteristics	Mean watershed slope, percent. Derived from 100m resolution National Elevation Dataset, so slope values may differ from those calculated from data of other resolutions.
Aspect (Northness)	Watershed Characteristics	Aspect "northness". Ranges from -1 to 1. Value of 1 means watershed is facing/draining due north, value of -1 means watershed is facing/draining due south.
Aspect (Eastness)	Watershed Characteristics	Aspect "eastness". Ranges from -1 to 1. Value of 1 means watershed is facing/draining due east, value of -1 means watershed is facing/draining due west.

References

- Addor, N., Nearing, G., Prieto, C., Newman, A. J., le Vine, N., & Clark, M. P. (2018). A Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water Resources Research*, *54*(11), 8792–8812. <https://doi.org/10.1029/2018WR022606>
- 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*, 1–20. <https://doi.org/10.1029/2020ef001916>
- Albers, S. J., Déry, S. J., & Peticrew, E. L. (2016). Flooding in the Nechako River Basin of Canada: A random forest modeling approach to flood analysis in a regulated reservoir system. *Canadian Water Resources Journal*, *41*(1–2), 250–260. <https://doi.org/10.1080/07011784.2015.1109480>
- Angilletta, M. J., Steel, E. A., Bartz, K. K., Kingsolver, J. G., Scheuerell, M. D., Beckman, B. R., & Crozier, L. G. (2008). Big dams and salmon evolution: changes in thermal regimes and their potential evolutionary consequences. *Evolutionary Applications*, *1*(2), 286–299. <https://doi.org/10.1111/j.1752-4571.2008.00032.x>
- Arismendi, I., Safeeq, M., Dunham, J. B., & Johnson, S. L. (2014). Can air temperature be used to project influences of climate change on stream temperature? *Environmental Research Letters*, *9*(8). <https://doi.org/10.1088/1748-9326/9/8/084015>
- Beaufort, A., Moatar, F., Sauquet, E., Loicq, P., & Hannah, D. M. (2020). Influence of landscape and hydrological factors on stream–air temperature relationships at regional scale. *Hydrological Processes*, *34*(3), 583–597. <https://doi.org/10.1002/hyp.13608>
- Blaen, P. J., Hannah, D. M., Brown, L. E., & Milner, A. M. (2013). Water temperature dynamics in High Arctic river basins. *Hydrological Processes*, *27*(20), 2958–2972.

<https://doi.org/10.1002/hyp.9431>

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.

<https://doi.org/10.1201/9780429469275-8>

Brown, L. E., & Hannah, D. M. (2008). Spatial heterogeneity of water temperature across an alpine river basin. *Hydrological Processes*, 22(7), 954–967.

<https://doi.org/10.1002/hyp.6982>

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>

Caldwell, P., Segura, C., Gull Laird, S., Sun, G., McNulty, S. G., Sandercock, M., Boggs, J., & Vose, J. M. (2015). Short-term stream water temperature observations permit rapid assessment of potential climate change impacts. *Hydrological Processes*, 29(9), 2196–2211. <https://doi.org/10.1002/hyp.10358>

Casado, A., Hannah, D. M., Peiry, J. L., & Campo, A. M. (2013). Influence of dam-induced hydrological regulation on summer water temperature: Sauce Grande river, Argentina. *Ecohydrology*, 6(4), 523–535. <https://doi.org/10.1002/eco.1375>

Cheng, Y., Voisin, N., Yearsley, J. R., & Nijssen, B. (2020). Reservoirs Modify River Thermal Regime Sensitivity to Climate Change: A Case Study in the Southeastern United States. *Water Resources Research*, 56(6). <https://doi.org/10.1029/2019WR025784>

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., & Pasteris, P. P. (2008). Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology*, 28(15), 2031–2064. <https://doi.org/10.1002/joc.1688>

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(October), 97–113. <https://doi.org/10.1016/j.earscirev.2017.10.009>
- Dynesius, M., & Nilsson, C. (1994). Fragmentation and flow regulation of river systems in the northern third of the world. *Science*, 266(5186), 753–762. <https://doi.org/10.1126/science.266.5186.753>
- Falcone, J. A., Carlisle, D. M., Wolock, D. M., & Meador, M. R. (2010). GAGES: A stream gage database for evaluating natural and altered flow conditions in the conterminous United States. *Ecology*, 91(2), 621–621. <https://doi.org/10.1890/09-0889.1>
- Ficklin, D. L., Stewart, I. T., & Maurer, E. P. (2013). Effects of climate change on stream temperature, dissolved oxygen, and sediment concentration in the Sierra Nevada in California. *Water Resources Research*, 49(5), 2765–2782. <https://doi.org/10.1002/wrcr.20248>
- 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>
- Gannon, J. P., Kelleher, C., & Zimmer, M. (2022). Controls on watershed flashiness across the continental US. *Journal of Hydrology*, 609(March), 127713. <https://doi.org/10.1016/j.jhydrol.2022.127713>
- Garner, G., Hannah, D. M., Sadler, J. P., & Orr, H. G. (2014). River temperature regimes of England and Wales: Spatial patterns, inter-annual variability and climatic sensitivity. *Hydrological Processes*, 28(22), 5583–5598. <https://doi.org/10.1002/hyp.9992>
- Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., Babu, S., Borrelli, P.,

- Cheng, L., Crochetiere, H., Ehalt Macedo, H., Filgueiras, R., Goichot, M., Higgins, J., Hogan, Z., Lip, B., McClain, M. E., Meng, J., Mulligan, M., ... Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221.
<https://doi.org/10.1038/s41586-019-1111-9>
- Hammond, J. C., Zimmer, M., Shanafield, M., Kaiser, K., Godsey, S. E., Mims, M. C., Zipper, S. C., Burrows, R. M., Kampf, S. K., Dodds, W., Jones, C. N., Krabbenhoft, C. A., Boersma, K. S., Datry, T., Olden, J. D., Allen, G. H., Price, A. N., Costigan, K., Hale, R., ... Allen, D. C. (2021). Spatial Patterns and Drivers of Nonperennial Flow Regimes in the Contiguous United States. *Geophysical Research Letters*, *48*(2), 1–11.
<https://doi.org/10.1029/2020GL090794>
- Hannah, D. M., & Garner, G. (2015). River water temperature in the United Kingdom: Changes over the 20th century and possible changes over the 21st century. *Progress in Physical Geography*, *39*(1), 68–92. <https://doi.org/10.1177/0309133314550669>
- Hannah, D. M., Smith, B. P. G., Gurnell, A. M., & McGregor, G. R. (2000). An approach to hydrograph classification. *Hydrological Processes*, *14*(2), 317–338.
[https://doi.org/10.1002/\(SICI\)1099-1085\(20000215\)14:2<317::AID-HYP929>3.0.CO;2-T](https://doi.org/10.1002/(SICI)1099-1085(20000215)14:2<317::AID-HYP929>3.0.CO;2-T)
- Hester, E. T., & Doyle, M. W. (2011). Human impacts to river temperature and their effects on biological processes: A quantitative synthesis. *Journal of the American Water Resources Association*, *47*(3), 571–587. <https://doi.org/10.1111/j.1752-1688.2011.00525.x>
- Hill, R. A., Hawkins, C. P., & Jin, J. (2014). Predicting thermal vulnerability of stream and river ecosystems to climate change. *Climatic Change*, *125*(3–4), 399–412.
<https://doi.org/10.1007/s10584-014-1174-4>

- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Hrachowitz, M., Soulsby, C., Imholt, C., Malcolm, I. A., & Tetzlaff, D. (2010). Thermal regimes in a large upland salmon river: A simple model to identify the influence of landscape controls and climate change on maximum temperatures. *Hydrological Processes*, *24*(23), 3374–3391. <https://doi.org/10.1002/hyp.7756>
- Imholt, C., Soulsby, C., Malcolm, I. A., Hrachowitz, M., Gibbins, C. N., Langan, S., & Tetzlaff, D. (2013). Influence of Scale on Thermal Characteristics in a Large Montane River Basin. *River Research and Applications*, *29*(4), 403–419. <https://doi.org/10.1002/rra.1608>
- Isaak, D. J., Wollrab, S., Horan, D., & Chandler, G. (2012). Climate change effects on stream and river temperatures across the northwest U.S. from 1980-2009 and implications for salmonid fishes. *Climatic Change*, *113*(2), 499–524. <https://doi.org/10.1007/s10584-011-0326-z>
- Kelleher, C. A., Golden, H. E., & Archfield, S. A. (2021). Monthly river temperature trends across the US confound annual changes. *Environmental Research Letters*, *16*(10). <https://doi.org/10.1088/1748-9326/ac2289>
- Kelleher, C., Wagener, T., Gooseff, M., McGlynn, B., McGuire, K., & Marshall, L. (2012). Investigating controls on the thermal sensitivity of Pennsylvania streams. *Hydrological Processes*, *26*(5), 771–785. <https://doi.org/10.1002/hyp.8186>
- Konapala, G., & Mishra, A. (2020). Quantifying Climate and Catchment Control on Hydrological Drought in the Continental United States. *Water Resources Research*, *56*(1), 1–25. <https://doi.org/10.1029/2018WR024620>

- Kurylyk, B. L., Macquarrie, K. T. B., Linnansaari, T., Cunjak, R. A., & Curry, R. A. (2015). Preserving, augmenting, and creating cold-water thermal refugia in rivers: Concepts derived from research on the Miramichi River, New Brunswick (Canada). *Ecohydrology*, 8(6), 1095–1108. <https://doi.org/10.1002/eco.1566>
- Laizé, C. L. R., Meredith, C. B., 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>
- Leach, J. A., & Moore, R. D. (2014). Winter stream temperature in the rain-on-snow zone of the Pacific Northwest: Influences of hillslope runoff and transient snow cover. *Hydrology and Earth System Sciences*, 18(2), 819–838. <https://doi.org/10.5194/hess-18-819-2014>
- Lisi, P. J., Schindler, D. E., Cline, T. J., Scheuerell, M. D., & Walsh, P. B. (2015). Watershed geomorphology and snowmelt control stream thermal sensitivity to air temperature. *Geophysical Research Letters*, 42(9), 3380–3388. <https://doi.org/10.1002/2015GL064083>
- Lowney, C. L. (2000). Stream temperature variation in regulated rivers: Evidence for a spatial pattern in daily minimum and maximum magnitudes. *Water Resources Research*, 36(10), 2947–2955. <https://doi.org/10.1029/2000WR900142>
- Maheu, A., Poff, N. L., & St-Hilaire, A. (2016). A Classification of Stream Water Temperature Regimes in the Conterminous USA. *River Research and Applications*, 32(5), 896–906. <https://doi.org/10.1002/rra.2906>
- McMillan, H. (2020). Linking hydrologic signatures to hydrologic processes: A review. *Hydrological Processes*, 34(6), 1393–1409. <https://doi.org/10.1002/hyp.13632>
- Mellor, C. J., Dugdale, S. J., Garner, G., Milner, A. M., & Hannah, D. M. (2017). Controls on Arctic glacier-fed river water temperature. *Hydrological Sciences Journal*, 62(4), 499–

514. <https://doi.org/10.1080/02626667.2016.1261295>

Miara, A., Vörösmarty, C. J., Macknick, J. E., Tidwell, V. C., Fekete, B., Corsi, F., & Newmark, R. (2018). Thermal pollution impacts on rivers and power supply in the Mississippi River watershed. *Environmental Research Letters*, *13*(3). <https://doi.org/10.1088/1748-9326/aaac85>

Michel, A., Brauchli, T., Lehning, M., Schaepli, B., & Huwald, H. (2020). Stream temperature and discharge evolution in Switzerland over the last 50 years: annual and seasonal behaviour. *Hydrology and Earth System Sciences*, *24*(1), 115–142. <https://doi.org/10.5194/hess-24-115-2020>

Michel, A., Schaepli, B., Wever, N., Zekollari, H., Lehning, M., & Huwald, H. (2022). Future water temperature of rivers in Switzerland under climate change investigated with physics-based models. *Hydrology and Earth System Sciences*, *26*(4), 1063–1087. <https://doi.org/10.5194/hess-26-1063-2022>

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>

Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., & Gupta, H. v. (2021). What Role Does Hydrological Science Play in the Age of Machine Learning? *Water Resources Research*, *57*(3). <https://doi.org/10.1029/2020WR028091>

O’Neal, K. (2002). *Effects of global warming on trout and salmon in U.S. streams* (Issue May). https://defenders.org/sites/default/files/publications/effects_of_global_warming_on_trout_and_salmon.pdf

- Orr, H. G., Simpson, G. L., des Clers, S., Watts, G., Hughes, M., Hannaford, J., Dunbar, M. J., Laizé, C. L. R., Wilby, R. L., Battarbee, R. W., & Evans, R. (2015). Detecting changing river temperatures in England and Wales. *Hydrological Processes*, 29(5), 752–766. <https://doi.org/10.1002/hyp.10181>
- 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. <https://doi.org/10.1016/j.scitotenv.2020.139679>
- Poole, G. C., & Berman, C. H. (2001). An ecological perspective on in-stream temperature: Natural heat dynamics and mechanisms of human-caused thermal degradation. *Environmental Management*, 27(6), 787–802. <https://doi.org/10.1007/s002670010188>
- Price, A. N., Jones, C. N., Hammond, J. C., Zimmer, M. A., & Zipper, S. C. (2021). The Drying Regimes of Non-Perennial Rivers and Streams. *Geophysical Research Letters*, 48(14), 1–12. <https://doi.org/10.1029/2021GL093298>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should i trust you?” Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 13-17-August*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Segura, C., Caldwell, P., Sun, G., McNulty, S., & Zhang, Y. (2015). A model to predict stream water temperature across the conterminous USA. *Hydrological Processes*, 29(9), 2178–2195. <https://doi.org/10.1002/hyp.10357>
- Singh, N. K., & Basu, N. B. (2022). The human factor in seasonal streamflows across natural and managed watersheds of North America. *Nature Sustainability*.

<https://doi.org/10.1038/s41893-022-00848-1>

- Somers, K. A., Bernhardt, E. S., Grace, J. B., Hassett, B. A., Sudduth, E. B., Wang, S., & Urban, D. L. (2013). Streams in the urban heat island: Spatial and temporal variability in temperature. *Freshwater Science*, *32*(1), 309–326. <https://doi.org/10.1899/12-046.1>
- Steel, E. A., Beechie, T. J., Torgersen, C. E., & Fullerton, A. H. (2017). Envisioning, Quantifying, and Managing Thermal Regimes on River Networks. *BioScience*, *67*(6), 506–522. <https://doi.org/10.1093/biosci/bix047>
- Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*, 1–11. <https://doi.org/10.1186/1471-2105-9-307>
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*. <https://doi.org/10.1186/1471-2105-8-25>
- Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on! *R Journal*, *1*(2), 14–17. <https://doi.org/10.32614/rj-2009-013>
- Tyralis, H., Papacharalampous, G., & Langousis, A. (2019). A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources. *Water*, *11*(5), 910. <https://doi.org/10.3390/w11050910>
- van Vliet, M. T. H., Franssen, W. H. P., Yearsley, J. R., Ludwig, F., Haddeland, I., Lettenmaier, D. P., & Kabat, P. (2013). Global river discharge and water temperature under climate change. *Global Environmental Change*, *23*(2), 450–464. <https://doi.org/10.1016/j.gloenvcha.2012.11.002>
- van Vliet, M. T. H., Ludwig, F., Zwolsman, J. J. G., Weedon, G. P., & Kabat, P. (2011). Global

- river temperatures and sensitivity to atmospheric warming and changes in river flow. *Water Resources Research*, 47(2). <https://doi.org/10.1029/2010WR009198>
- Wanders, N., van Vliet, M. T. H., Wada, Y., Bierkens, M. F. P., & van Beek, L. P. H. (Rens). (2019). High-Resolution Global Water Temperature Modeling. *Water Resources Research*, 55(4), 2760–2778. <https://doi.org/10.1029/2018WR023250>
- Webb, B. W. (1996). Trends in stream and river temperature. *Hydrological Processes*, 10(2), 205–226. [https://doi.org/10.1002/\(SICI\)1099-1085\(199602\)10:2<205::AID-HYP358>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-1085(199602)10:2<205::AID-HYP358>3.0.CO;2-1)
- 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*, 22(7), 902–918. <https://doi.org/10.1002/hyp.6994>
- White, E. (2017). Predicting Unimpaired Flow in Ungauged Basins: “Random Forests” Applied to California Streams. *ProQuest Dissertations and Theses*, 69. <http://search.proquest.com.libraryproxy.griffith.edu.au/docview/2026286173?accountid=14543%0Ahttp://hy8fy9jj4b.search.serialssolutions.com/directLink?&atitle=Predicting+Unimpaired+Flow+in+Ungauged+Basins%3A+%22Random+Forests%22+Applied+to+California+Strea>
- Willis, A. D., Peek, R. A., & Rypel, A. L. (2021). Classifying California’s stream thermal regimes for cold-water conservation. *Plos One*, 16(8), e0256286. <https://doi.org/10.1371/journal.pone.0256286>
- Wondzell, S. M., Diabat, M., & Haggerty, R. (2019). What Matters Most: Are Future Stream Temperatures More Sensitive to Changing Air Temperatures, Discharge, or Riparian Vegetation? *Journal of the American Water Resources Association*, 55(1), 116–132.

<https://doi.org/10.1111/1752-1688.12707>

Woznicki, S. A., Baynes, J., Panlasigui, S., Mehaffey, M., & Neale, A. (2019). Development of a spatially complete floodplain map of the conterminous United States using random forest. *Science of the Total Environment*, *647*, 942–953.

<https://doi.org/10.1016/j.scitotenv.2018.07.353>

Yan, H., Sun, N., Fullerton, A., & Baerwalde, M. (2021). Greater vulnerability of snowmelt-fed river thermal regimes to a warming climate. *Environmental Research Letters*, *16*(5).

<https://doi.org/10.1088/1748-9326/abf393>

**Chapter 3: Incorporating Physically-based Water Temperature Predictions into the
National Water Model Framework: Application to a Western Forested Headwater
Catchment**

This chapter is in preparation for submission as:

Wade, J., Kelleher, C., Kurylyk, B.L. Incorporating Physically-based Water Temperature Predictions into the National Water Model Framework: Application to a Western Forested Headwater Catchment, (in prep).

Abstract

While river water temperatures are a strong control on instream processes and aquatic ecosystem integrity, monitoring networks for river water temperatures are often sparse. Water temperature models, which simulate thermal behavior along rivers, are useful tools for filling such spatial and temporal gaps in the monitoring network and furthering our understanding of dominant pathways of heat and water transfer. Despite recent advancements in water temperature modeling strategies, current models struggle to provide real-time and reach-specific predictions across broad spatial domains. We developed a physically-based water temperature model coupled to the National Water Model (NWM) to assess the potential for water temperature prediction to later be incorporated into the NWM at the continental scale. Using model forcings and outputs from the NWM v2.1 retrospective, we evaluated the ability of four model configurations of increasing complexity to simulate hourly water temperatures in the forested headwaters of H.J. Andrews Experimental Forest, Oregon, USA. After calibration, our NWM-coupled model produced water temperature simulations with root-mean-square error values under 0.7°C compared to two observed temperature records. We found that model performance generally improved with the addition of parameters controlling advective heat fluxes at the streambed, including hyporheic and groundwater flow. Model development highlighted several potential challenges in pairing water temperature prediction with the NWM, including estimating missing inputs in the absence of site-specific observations and uncertainty due to NWM's simplified representation of stream network extent. Our modeling framework, representing a first effort at pairing water temperature simulation with the high-resolution predictions of the NWM, confirms that the NWM can be leveraged to give insight into other water quality variables.

1. Introduction

River water temperature is often referred to as a ‘master’ water quality variable, as a wide range of chemical and biological processes are closely linked to in-stream thermal regimes (Caissie, 2006; Hannah & Garner, 2015; Ouellet et al., 2020). The temperature of rivers controls algal and bacterial growth rates, dissolved oxygen content, solute processing, and the integrity of ecosystems (Havens & Paerl, 2015; Isaak et al., 2012). Water temperatures are of particular economic interest to management agencies, as the viability of salmonid fisheries and the efficiency of river-side power plants are both threatened by warmer rivers (Ficke et al., 2007; Förster & Lilliestam, 2010). With future climate change expected to give rise to heightened river water temperatures (Caldwell et al., 2015; van Vliet et al., 2013; Wanders et al., 2019), it is critical to better understand, observe, and predict reach-scale water temperature dynamics at continental to global scales.

In comparison to records of discharge, observations of river water temperature are sparse, particularly outside of the world’s major river basins (Wanders et al., 2019). Without knowledge of river thermal regimes in unmonitored basins, it is exceedingly challenging to manage the threats warmer rivers pose to aquatic fauna and the productivity of fisheries (Ficke et al., 2007). Models of water temperature offer unique insight into the spatial and temporal dynamics of river thermal regimes, both within individual basins and at a global-scale, helping to bridge gaps between gages in a sparse temperature monitoring network. A wide range of modeling strategies are commonly applied to water temperature prediction and can generally be grouped into statistical and physically-based models (Caissie, 2006; Dugdale et al., 2017). The applications of statistical and physically-based water temperature models are extensively reviewed in Benyahya et al. (2007) and Dugdale et al. (2017), respectively.

Physically-based (or ‘mechanistic’) models are particularly well suited to water

temperature prediction (Dugdale et al., 2017) and function by calculating energy fluxes at the air-water and water-streambed interfaces and transporting mass and stored thermal energy downstream (Caissie, 2006). Despite these advantages, physically-based models require site-specific data (e.g., discharge, groundwater inflow, radiation flux) in order to resolve land surface and hydrologic processes (Dugdale et al., 2017). The need for high-quality input data can make physically-based models difficult to apply to unmonitored catchments, where knowledge of hydrologic behavior is often uncertain.

Coupled temperature-hydrological models (van Beek et al., 2012; van Vliet et al., 2012; Wanders et al., 2019) (hereby referred to as ‘coupled models’), which concurrently simulate both hydrological and thermal river processes, are an effective tool for overcoming limitations related to a lack of observational data (Dugdale et al., 2017). By leveraging calibrated hydrologic predictions, coupled models can accurately simulate water temperatures in unmonitored basins at adaptable spatial and temporal resolutions (Sun et al., 2015). Coupled models are well-suited to simulating thermal dynamics in settings where advective heat fluxes are influential, such as headwater reaches. Along these reaches, the ability of coupled models to divide inflows into multiple source water components (e.g., surface runoff, groundwater inflow) is crucial for properly estimating hydrologic heat fluxes (Dugdale et al., 2017).

Recent advances in continental-scale hydrologic modeling have introduced newfound prediction capabilities at unprecedented spatial and temporal scales (Lin et al., 2019; Salas et al., 2018). This progress presents new opportunities for expanding the extent and accuracy of river temperature predictions through the development of coupled hydrologic-temperature models. One such broad-scale hydrologic model with potential for coupling to a water temperature model is the National Water Model (NWM). The NWM is a hydrologic model developed by the

National Oceanic and Atmospheric Administration (NOAA), the National Weather Service (NWS), and the Office of Water Prediction (OWP) that forecasts hourly streamflow at 2.7 million river reaches in the conterminous US (CONUS) (Lahmers et al., 2021; NOAA, 2016). The NWM simulates components of the terrestrial water cycle, including land surface water and energy fluxes, soil moisture, subsurface flow, and channel routing, using a particular configuration of the NCAR-supported Weather Research and Forecasting hydrological modeling system (WRF-Hydro; Gochis et al., 2021) and the Noah Multi-Parameterization land surface model (Noah-MP; Niu et al., 2011). As the model provides high resolution (1 km) predictions of land surface and hydrologic states over a range of forecast lead times (NOAA, 2023), the framework of the NWM is well suited to be coupled to a continental-scale water temperature model.

Despite the ability of the NWM to accurately represent hydrological processes in catchments across the US, application of the modeling framework to river water temperatures remains unexplored. A coupled NWM-water temperature model could resolve thermal dynamics at reach scales relevant to watershed management along all conterminous US catchments and allow for forward-looking temperature forecasts. Using data derived from the NWM and other publicly available sources, we sought to develop a proof-of-concept water temperature model in a single test basin over a month of baseflow conditions to determine if the NWM framework is suitable for temperature prediction. While we only considered temperature modeling in a single basin, the strategies we developed for the modeling framework were intended to be transferable to broader spatial scales and potentially, with modifications, to other water quality variables. In this study, we aimed to (1) assess if forcings and outputs from the National Water Model can be leveraged to accurately simulate hourly river water temperatures in a forested headwater

catchment and (2) evaluate how model configurations of increasing complexity represent thermal processes influencing water temperatures.

2. Methods

2.1. Study Site: H.J. Andrews Experimental Forest

We selected the H.J. Andrews Experimental Forest (H.J. Andrews), a 64 km² forested headwater catchment located in the western Cascade Mountains, Oregon, USA, to serve as a test basin for this study. H.J. Andrews has been subject to continuous and extensive hydrologic monitoring since 1948 (Johnson et al., 2021), providing insight into hyporheic exchange processes (Becker et al., 2023; Ward et al., 2012; Ward et al., 2019; Wondzell et al., 2009), river corridor connectivity (McGuire & McDonnell, 2010; Ward et al., 2018a), and water temperature dynamics (Johnson, 2004). The breadth of past hydrologic research at the site, coupled with the availability of historical observations, make H.J. Andrews an ideal catchment to explore the performance of a water temperature model.

The H.J. Andrews watershed is drained by several streams, including McRae Creek, Mack Creek, and Lookout Creek, the latter of which drains downstream to the Blue River Reservoir (Figure 1). While considered a fifth-order catchment by most field studies (Ward et al., 2019), H.J. Andrews is represented as a third-order basin by the NWM (and NHD), which often does not resolve small headwater reaches. H.J. Andrews is characterized by high relief topography, with elevations ranging from 410 to 1630 meters above sea level, and is primarily forested by Douglas fir trees (Ward et al., 2019). Annual precipitation at this site is strongly seasonal and varies between 1900 and 2900 mm, with most falling in winter months between November and April (Jennings & Jones, 2015). Flows at the basin outlet (Lookout Creek) typically reach a maximum in December or January, and a minimum in September (Jennings &

Jones, 2015). H.J. Andrews' streams are home to a diversity of aquatic species, including cutthroat trout and coastal giant salamanders (Kaylor et al., 2019). The watershed is generally unimpacted by anthropogenic disturbances, with the exception of experimental logging in select catchments.

While H.J. Andrews contains a number of water temperature gaging stations, only two gages coincided with reaches represented by the National Water Model. These gages, providing records of water temperature and discharge (Gregory & Johnson, 2019), are located on the upper reaches of Mack Creek (GSMACK; drainage area: 580 ha) and the lower reaches of Lookout Creek (GSLOOK; USGS 14161500; drainage area: 6242) near the basin outlet (Figure 1). We took GSMACK to represent headwater behavior (hereby referred to as 'headwater') and GSLOOK to represent higher order stream behavior (hereby referred to as 'outlet') in the basin.

2.2. Model Data

2.2.1. National Water Model Retrospective v2.1

The NWM retrospective is a backwards-looking long-duration model run forced with observational meteorological data. While the NWM retrospective analyses are typically used to evaluate model performance (Dyer et al., 2022; Salas et al., 2018; Wan et al., 2022), the historical continuity of their predictions makes them useful in the testing and development of models coupled to the NWM. In this study, we used data from the 42-yr NWM retrospective version 2.1 (v2.1) (February 1979 to December 2020), a run of NWM version 2.1 forced by near surface meteorological conditions from the Analysis of Record for Calibration (AORC) dataset (NOAA, 2021). The v2.1 retrospective configuration uses version 5.2.0 of WRF-Hydro (Gochis et al., 2021) and does not assimilate observed discharge data from stream gages.

The AORC supplies gridded atmospheric forcing data to WRF-Hydro. This forcing data

includes hourly records of precipitation, air temperature, specific humidity, air pressure, downward shortwave radiation flux, downward longwave radiation flux, and u- and v-components of wind speed (NOAA, 2021). These near surface conditions are used by the Noah-MP land surface model to simulate vertical energy and water fluxes at a 1-km spatial resolution (Gochis et al., 2021). Vertical moisture fluxes through the land surface are then passed to subsurface routing modules, which influence the lateral flow of water across the model's surface, soil, and saturated domains. Using a 250-m grid, WRF-Hydro routes subsurface flow through a 2-m thick soil column and an unconfined groundwater aquifer, approximating hydraulic gradients using a D8 steepest descent method (Gochis et al., 2021; Lahmers et al., 2019). When the subsurface storage of a grid cell is exceeded, excess water is routed to channels as overland surface runoff using a diffusive wave approach (Julien et al., 1995; Lahmers et al., 2019; Rojas et al., 2003). The location and extent of NWM channels are derived from National Hydrography Dataset (NHD) Plus Version 2 (NHDPlusV2) river reaches (McKay et al., 2012; Salas et al., 2018). These channels can receive inflows either from surface runoff or from groundwater recharge, represented by empirically-tuned discharge from a conceptual exponential groundwater bucket (Gochis et al., 2021). Downstream flow is transported through trapezoidal NWM channels using Muskingum-Cunge routing (Gochis et al., 2021; Lahmers et al., 2019). Parameters related to channel geometry are empirically derived using relationships to each reach segment's drainage area (Gochis et al., 2021). These computations are then integrated to deliver hourly values of streamflow, stream velocity, surface water runoff, and groundwater bucket inflow at each reach segment.

We retrieved NWM retrospective v2.1 forcing and output data from a publicly available AWS repository and extracted hourly values of relevant variables at stream segments within the

H.J. Andrews study basin. The sources and respective applications of NWM data used in this study are presented in Figure 2. These inputs can be divided into three categories: meteorological forcing data, hydrological model outputs, and channel geometry parameters (Figure 2). From the meteorological forcing data, we extracted incoming shortwave radiation, incoming longwave radiation, air temperature, specific humidity, air pressure, and wind speed. Gridded meteorological forcings were assigned to vector stream reaches based on the centroid location of each reach. From the hydrological model outputs, we retrieved discharge, stream velocity, flux from the groundwater bucket, and runoff from terrain routing (surface runoff) corresponding to each model reach segment. We also retrieved channel geometry parameters, including location, reach length, width, side slope angle, and stream order, from the NWM Routelink dataset.

2.3. Modeling Approach

2.3.1. Model Resolution

We simulated hourly water temperatures throughout the H.J. Andrews stream network during a one-month period of low flow from July 1 to July 31 2019. We selected this time period because it was the most recent period where observed water temperatures in the basin and predictions from the NWM Retrospective v2.1 overlapped. We subdivided channels identified by the NWM into a series of 1-km long reach segments, beginning at the channel head of each tributary. Water temperature predictions were made at 46 model nodes, located at the beginning and end of each of these segments. Additional model nodes were added at the location of the two observed water temperature gages so as to not introduce error via spatial interpolation when assessing model performance against observations.

2.3.2. Computation of Water Temperatures

We adapted a semi-Lagrangian model formulation following Yearsley (2009), also

implemented in the DHSVM-RBM water temperature model, to develop a Python script that simulates water temperatures using primarily NWM forcings. Semi-Lagrangian approaches are widely used in the field of numerical weather prediction (Husain & Girard, 2017) and have also been applied extensively to water temperature modeling (Lee et al., 2020; Yan et al., 2021; Yearsley, 2009, 2012). This frame of reference combines aspects of Eulerian and Lagrangian approaches, coupling a fixed model grid with longitudinal particle tracking to gain efficiency over a strictly Eulerian method (Yearsley, 2009). Semi-Lagrangian models are numerically stable across broad ranges of space and time steps, facilitating simulations at time steps considerably longer than possible under models limited by the Courant condition (Yearsley, 2009).

In this semi-Lagrangian approach, unknown temperatures at a future time step were determined by applying reverse particle tracking to simulate the longitudinal paths of water parcels originating from model nodes where water temperatures are simulated (Yearsley, 2009). From a given node at time $t+\Delta t$, where Δt is equal to the computational time step, the upstream Lagrangian coordinate (ξ) at time t was equal to (Yearsley, 2009):

$$\xi = x_0 - \int_t^{t+\Delta t} u \, dt \quad (1)$$

where:

u = longitudinal velocity field of traversed river reaches, m s^{-1}

t = model time step, s

Δt = computational time step of the model, s

x_0 = starting position of the water parcel along the reach, m

As the origin location of each water parcel did not always coincide with a model node, we used second-order Lagrangian polynomials to interpolate the temperature at the origin point at time t

using known water temperatures from surrounding nodes (Yearsley, 2009).

Once the starting water temperature at the origin point was known, the particle was tracked as it traveled back downstream to the starting model node. As the water parcel passed downstream from time t to $t+\Delta t$, the location of a water parcel along its trajectory (x_j) was tracked by (Yearsley, 2009):

$$x_j = \xi + \sum_{j_0}^j u(j')\Delta(j') \quad (2)$$

where:

ξ = upstream location of the water parcel at time t , m

$u(j')$ = flow velocity in the j^{th} model segment at time t , m s⁻¹

$\Delta(j')$ = time taken to traverse the j^{th} model segment, s

j = index of model segment, unitless

The length of time $\Delta(j')$ to traverse the j^{th} model segment was equal to (Yearsley, 2009):

$$\Delta(j') = \frac{x_{j'} - x_j}{u(t, j')} \quad (3)$$

where:

$x_{j'}$ = location of the downstream boundary of the j' 'th node, m

x_j = location of the water parcel along its trajectory, m

$u(t, j')$ = flow velocity along the j' 'th model segment at time t , m s⁻¹

As water parcels traversed model segments downstream, radiative and hydrologic heat inputs were calculated and integrated over time to update the water temperature at each model node until the water parcel reached its final location at time $t+\Delta t$. The calculated water temperature was then inserted at the unknown node and the cycle repeated, either at the next time step or for the next node in the sequence. Following an approach based on Yearsley (2009),

water temperatures were updated at the downstream end of the j^{th} model segment at time $t+\Delta(j')$ by integrating radiative and hydrologic energy fluxes along each model segment:

$$T(t + \Delta(j'), x_{j'}) = T(t, x_{j'}) + \Delta(j') \left[\frac{H(t, x_{j'})}{\rho C_p D(t, x_{j'})} + \Phi(t, x_{j'}) \right] \quad (4)$$

where:

$T(t, x_{j'})$ = known water temperature at the current time step, °C

$T(t + \Delta(j'), x_{j'})$ = unknown water temperature at the future time step, °C

$\Delta(j')$ = time taken for the water parcel to traverse the j^{th} model segment, s

$H(t, x_{j'})$ = thermal energy flux across the air-water interface, W m^{-2}

ρ = density of water, kg m^{-3}

C_p = specific heat capacity of water, $\text{J kg}^{-1} \text{°C}^{-1}$

$D(t, x_{j'})$ = channel depth, m

$\Phi(t, x_{j'})$ = effective advected heat flux from hydrologic inflows, including groundwater and tributaries, °C s^{-1}

If a water parcel traversed more than one model segment during a computational time step, Equation 7 was computed at the end of each segment crossed.

To calculate how radiative and hydrologic forcings result in changes in water temperatures, the cross-sectional area, depth, width, and the volume of each reach must be known. For each reach, the NWM Routelink file supplied the reach length, channel side slope, bottom width, and top width to define a trapezoidal channel geometry (Gochis et al., 2021). As our model simulated water temperatures during summer low flow conditions, we only considered flows through the primary channel and disregarded overbank flow into the floodplain. A further description of our derivation of cross-sectional area, water depth, and reach volume is presented

in Supporting Information.

2.3.3. Heat Transfer Equations

The total thermal energy flux across the air-water interface (H) summarizes the radiative and atmospheric forcings to water parcels as they traverse model reaches. These energy fluxes include incoming shortwave radiation, net longwave radiation, sensible heat exchange, and latent heat exchange. The total thermal energy flux across the air-water interface was calculated by:

$$H = H_{SW} + H_{LW} + H_{SH} + H_{LH} \quad (5)$$

where:

H_{SW} = shortwave radiation flux, W m^{-2}

H_{LW} = net longwave radiation flux, W m^{-2}

H_{SH} = sensible heat exchange flux, W m^{-2}

H_{LH} = latent heat exchange flux, W m^{-2}

Full equations for the calculation of each of these heat balance components, including the integration of riparian vegetative shading, is described in Supporting Information. We did not include bed conduction in the net energy balance, as streambed temperatures would be difficult to quantify when expanding the model to broader scales. The bed conduction flux is generally small compared to other heat fluxes, though it can be an influential process along headwater reaches (Benyahya et al., 2012; Caissie et al., 2014; Johnson, 2004).

2.3.4. Hydrologic Heat Fluxes

In addition to the radiative and atmospheric heat fluxes to the water column, hydrologic inflows, including groundwater inflow, surface water runoff, and tributary inflow, contribute heat to the stream based on the relative temperature difference between the stream and inflows.

We aggregated the relative effects of these three inflows to generate a single advective heat flux to each model reach over time. The total hydrologic inflow rate was calculated by:

$$Q_I = Q_S + Q_T + Q_{GW} \quad (6)$$

where:

Q_I = total inflow rate, $\text{m}^3 \text{s}^{-1}$

Q_S = surface water runoff rate, $\text{m}^3 \text{s}^{-1}$

Q_T = tributary inflow rate, $\text{m}^3 \text{s}^{-1}$

Q_{GW} = groundwater inflow rate, $\text{m}^3 \text{s}^{-1}$

Once the total inflow rate and individual inflow components were known, the effective temperature of the inflows was calculated by a flow-weighted arithmetic mean (Glose et al., 2017):

$$T_I = \frac{Q_S T_S}{Q_I} + \frac{Q_T T_T}{Q_I} + \frac{Q_{GW} T_{GW}}{Q_I} \quad (7)$$

where:

T_I = effective temperature of inflows, $^{\circ}\text{C}$

T_S = temperature of surface water runoff, $^{\circ}\text{C}$

T_T = temperature of tributary inflow, $^{\circ}\text{C}$

T_{GW} = temperature of groundwater inflow, $^{\circ}\text{C}$

Q_I = total inflow rate, $\text{m}^3 \text{s}^{-1}$

Q_S = surface water runoff rate, $\text{m}^3 \text{s}^{-1}$

Q_T = tributary inflow rate, $\text{m}^3 \text{s}^{-1}$

Q_{GW} = groundwater inflow rate, $\text{m}^3 \text{s}^{-1}$

The advective heat flux to the stream was then derived by computing the difference between the aggregate inflow temperature and current water column temperature, scaled by the

proportion of lateral inflow to total channel volume. The effective advective flux was calculated by (Glose, 2013):

$$\Phi = \frac{Q_L}{V} (T_L - T) \quad (8)$$

where:

Φ = effective advective heat flux, °C s⁻¹

Q_L = total lateral inflow rate, m³ s⁻¹

V = model reach volume, m³

T_L = effective temperature of lateral inflows, °C

T = channel water temperature, °C

2.3.5. Estimation of Unknown Inflow Temperatures

In Equation 7, the water temperature of each inflow component (groundwater inflow, surface water runoff, and tributary inflow) was unknown. Following Wanders et al. (2019), we set the temperature of surface water runoff as 1.5 °C less than the current air temperature (though we note that this parameter was not expected to be influential given the lack of surface runoff that occurred during our simulation period). Simulated tributary temperatures were implicitly treated as lateral inflow to the model reach where the tributary joined the mainstem river.

Groundwater temperatures are particularly influential to modeled water temperatures, but are often unknown for the purposes of water temperature modeling. Our proposed approach relied on the assumption that the net water temperature of groundwater inflow must be bounded by the temperature of deep groundwater (approximated by annual mean air temperature) and the ground surface temperature (approximated by continuous air temperature). The ground surface temperature can be coarsely estimated by a smoothed mean air temperature, reflecting correlative links between patterns in solar radiation, air temperature, and ground surface temperature. By

scaling the magnitude of variability of a smoothed daily air temperature signal between the bounds of deep groundwater and the ground surface temperature, we estimated the effective inflow temperature of groundwater inflow. At each model time step and model reach segment, we calculated the groundwater inflow temperature (T_{GW}) by:

$$T_{GW} = C_{AT-GW} * (AT_D - \overline{AT}) + \overline{AT} \quad (9)$$

where:

C_{AT-GW} = Air temperature–groundwater temperature coefficient, varying from 0 – 1, unitless

AT_D = Mean daily air temperature smoothed over a variable duration moving window, °C

\overline{AT} = Mean annual air temperature, °C

The smoothed daily air temperature, AT_D , for a given time t was calculated by:

$$AT_{D,t} = \frac{1}{W} * \sum_{n=t-W-1}^t AT_n \quad (10)$$

where:

W = air temperature moving window duration, days

AT_n = daily mean air temperature on day n , °C

By tuning C_{AT-GW} between 0 and 1, we simulated effective sourcing of inflows from temporally-invariant deep groundwater (value closer to 0) or from more variable shallow groundwater (value closer to 1) (Figure 3). We derived mean annual air temperatures along the network using 4 km gridded PRISM means over a 4-year period from 2016 to 2019 (PRISM Climate Group, 2022). Daily air temperatures were retrieved from NWM forcings at each model reach and smoothed to a mean value using a backward-looking moving window, tuned to vary between a duration of 2 and 14 days. This moving window simulated the unknown response time of shallow groundwater

to radiative forcings (reflected by air temperature). We used the calculated groundwater inflow temperature time series at each model node to supply boundary conditions to the model, both to set the first time-step temperature along the entire network and the time-varying temperature of streamflow initiation at all reach heads during the study period.

2.3.6. Approximating the Thermal Effects of Hyporheic Exchange

To conceptually represent the thermal effect of hyporheic exchange in our model, we used a simplified approach that stores water temperatures from previous time steps and returns them at a later, lagged time at a rate proportional to a tuned fraction of discharge. The hyporheic return flow temperature (T_{hyp}) at time t along a given reach calculated by:

$$T_{hyp,t} = \frac{1}{H_{lag}} * \sum_{n=t-H_{lag}}^{t-1} T_n \quad (11)$$

where:

$T_{hyp,t}$ = hyporheic return flow temperature at time t , °C

H_{lag} = hyporheic lag duration, hours

T_n = simulated water temperature at previous time n , °C

In a similar manner to the computation of advective heat transfer due to groundwater inflow, the effective hyporheic heat flux was calculated by:

$$\Phi_{hyp} = \frac{H_{frac} * Q}{V} (T_{hyp} - T_t) \quad (12)$$

where:

Φ_{hyp} = effective advective heat flux, °C s⁻¹

H_{frac} = fraction of streamflow returned to channel as hyporheic flow, varying between 0 and 1, unitless

Q = discharge, $\text{m}^3 \text{s}^{-1}$

V = model reach volume, m^3

T_{hyp} = hyporheic return from temperature, $^{\circ}\text{C}$

T = channel water temperature, $^{\circ}\text{C}$

We tuned the hyporheic lag duration parameter between 2 and 24 hours, simulating a range of hyporheic flow path velocities. Although hyporheic flow paths often have residence times longer than 24 hours, the variability in the mean temperature of simulated streamflow (approximating hyporheic return temperature) over periods longer than 24 hours is negligible. As such, we limited the lag duration to a maximum of 24 hours to conserve computational runtime. The hyporheic flow fraction coefficient represented the amount of water returned to the stream at a given point in time and space as a proportion of discharge (e.g., an H_{frac} value of 0.4 equates to 40% of discharge returned to the stream as hyporheic flow). We allowed this fraction to vary independently by stream order (first, second, and third order reaches), as we generally expect stream order and hyporheic flow to demonstrate negative relationship. As stream order increases and stream slope decreases down-valley, the effects of hyporheic flow relative to other channel processes tends to decrease (Boano et al., 2014; Wondzell, 2011).

2.3.7. Estimating Riparian Shading in the Absence of On-site Observations

Riparian shading is a crucial variable in water temperature modeling, as it controls the proportion of radiation that reaches the water's surface. However, the NWM does not constrain shading of channels by streambank vegetations. In the absence of model data and on-site observations, we derived riparian shading values along the river network using an empirical formula (Vegetation-shading index; VSI) presented by Kalny et al. (2017) that relates vegetation height, vegetation buffer width, and vegetations density to riparian shading. VSI has been shown

to accurately characterize riparian shading in the absence of on-site observations, with estimated values displaying correlations of up to 0.9 with shading values derived from hemispherical photos (Kalny et al., 2017). VSI is calculated by:

$$VSI = \left(\frac{h_r}{h_{max}} + \frac{w}{w_{max}} + \frac{d}{d_{max}} \right) \div 3 \quad (13)$$

where:

VSI = vegetation-shading index, varying between 0 and 1, unitless

h_r = relative vegetation height, %

h_{max} = maximum vegetation height, equal to 100%

w = vegetation buffer width, m

w_{max} = maximum vegetation buffer width affecting water temperature, m, assumed to equal 50m

d = vegetation density, %

d_{max} = maximum vegetation density, equal to 100%

Given the dense and contiguous forest cover adjacent to river reaches in the H.J. Andrew's watershed, we assumed that the vegetation buffer width was equal to the maximum 50 m width value for all reach segments. Relative vegetation height (h_r) was calculated by scaling vegetation height by river width using the equation (Kalny et al., 2017):

$$h_r = \frac{h_v * 100}{r_w * 1.62} \quad (14)$$

$$\text{if } h_r \geq 100, h_r = 100$$

where:

h_r = relative vegetation height, %

h_v = vegetation height, m

r_w = river width, m

We modified the original relative vegetation height formula presented by Kalny et al. (2017) by multiplying the river width term by 1.62 to account for differences in latitude between our study site and the site where the above formula was derived. The 1.62 scalar value indicates that due to the mean solar angle between 10 a.m. and 2 p.m. (maximum solar incidence) at H.J. Andrews during our study period (July 2019), a tree would cast a shadow roughly 1.62 times its length (Kalny et al., 2017).

In the absence of in-situ observations of canopy cover, we retrieved values of existing vegetation height and forest canopy cover from 30-m resolution gridded US LANDFIRE datasets (LANDFIRE, 2020). We then calculated mean values of canopy variables along reach segments using 50 m buffers perpendicular to the centerline of stream, excluding water pixels from calculated means. To account for differences in shading due to the geographic aspect of each reach segment, we calculated canopy values using buffers on only the right bank for eastward-flowing segments (45° - 135°), only the left bank for westward-flowing segments (225° - 315°), and both banks for northward-flowing (45° - 315°) and southward-flowing segments (135° - 225°) (Kalny et al., 2017). In tuning scenarios where riparian shading exceeded 100% along a reach, its value was set to equal 100%.

2.3.8. Assessing Model Error

We assessed the error of model simulations by comparing predictions to observed temperatures at two gages ('headwater': Mack Creek, 'outlet': Lookout Creek) within the basin. In the calculation of error metrics, we removed the first 48 hours of simulated temperatures as a model spin-up period. While this spin-up time is shorter than that of other hydrologic models, water, the boundary condition temperatures rapidly equilibrated with radiative forcings after a

single diel cycle.

At each gauge, we calculated a suite of error metrics that capture a range of modes of variability, including RMSE, daily maxima error (D_{Max}), daily minima error (D_{Min}). RMSE was calculated using the full hourly time series of prediction. D_{Max} and D_{Min} were calculated as the mean difference between predicted and observed daily maxima and minima during each 24-hour period. By using multiple error metrics in tandem to evaluate model performance, we gained additional insight into how the model resolved radiative and hydrologic processes. Daily maxima error, which described how the model represents peak temperatures, is an indicator of the model's ability to accurately simulate radiative heat fluxes that typically dominate net heat transfer during daytime hours. Daily minima error, which quantified how the model captures nighttime and early morning temperatures, is closely linked to hydrologic heat fluxes that become more influential in the absence of solar radiation.

2.4. Sequential Evaluation of Model Configurations

Using a flexible model development framework (Fenicia et al., 2011; Hrachowitz et al., 2014), we tested the ability of 4 model formulations of increasing complexity and representation of physical processes to simulate water temperatures in our test basin. In general, we sought to develop model configurations that were parsimonious, representing physical behavior using the simplest formulation (or degrees of freedom) possible to avoid overparameterization and retain computational efficiency (Hrachowitz et al., 2014; Jakeman et al., 2006). With this in mind, we attempted to design modeling strategies such that they had sufficient complexity to produce accurate predictions, while matching the availability (or uncertainty) of model inputs (Wagener et al., 2001).

Models M1, M2, M3, and M4 each progressively incorporated additional degrees of

freedom, tuning a broader suite of parameters that reflect uncertainty in hydrologic and thermal processes (Table 1). M1, the simplest configuration, only tuned parameters related to groundwater inflow temperatures and riparian shading. This formulation excluded hyporheic flow and used NWM estimates for rates of groundwater flow. M2 built on the M1 configuration, tuning the NWM estimates for the rate of groundwater inflow along the network and again excluding hyporheic flow. M3 built on the M1 configuration, adding a conceptual representation of hyporheic exchange (see Section 2.3.6). M4 combined the complexity of M2 and M3, tuning parameters related to groundwater inflow temperatures, groundwater inflow rate, riparian shading, and hyporheic exchange. Our model configurations (M1, M2, M3, M4) were not intended to resolve every physical process controlling water temperatures and instead sought to balance gains in performance against computational cost and uncertainty.

We calibrated each model configuration using 5,000 uniform Monte Carlo samples of parameters (Table 2), totaling 20,000 model runs across all configurations. Parameters descriptions and their sampled plausible ranges are shown in Table 2. We intentionally defined wide parameter ranges to more fully explore all possible model outcomes. These parameters can be grouped into two categories: those that are tuned for the full network, and those that are tuned independently by stream order. We assumed that the full network parameters (air temperature moving window duration, riparian shading coefficient, and hyporheic lag duration) represent processes or sources of model error that are likely uniform throughout the basin. Parameters tuned by stream order (C^1_{AT-GW} , C^2_{AT-GW} , C^3_{AT-GW} , GW^1 , GW^2 , GW^3 , H^1_{frac} , H^2_{frac} , and H^3_{frac}) were assumed to represent processes that scale in relation to relative stream size. As the basin contains reaches up to third order, each of these variables was tuned independently across three degrees of freedom (first-, second-, and third-order reaches). The riparian shading coefficient (R_{shade}) and

groundwater inflow rate coefficients (GW^1 , GW^2 , GW^3) were unitless coefficients used to tune existing estimates of riparian shading and groundwater inflow, reflecting our uncertainty in the characterization of these processes. The coefficients used in the tuning of groundwater inflow temperatures (C^1_{AT-GW} , C^2_{AT-GW} , C^3_{AT-GW} , W) and hyporheic flow (H_{lag} , H^1_{frac} , H^2_{frac} , H^3_{frac}) were used in equations described in sections 2.3.5 and 2.3.6, respectively.

From the 5,000 model runs for each configuration (M1, M2, M3, M4), we selected the top 1% of runs sorted by $RMSE_w$, to represent peak potential model performance. $RMSE_w$ is a weighted error metric, calculated by the weighted average of headwater RMSE (25% weight) and outlet RMSE (75% weight). These runs are highlighted amongst all model runs in Figure 4. We prioritized runs with low RMSE values at the outlet because we expect that model error will decrease down-network as radiative forcings, which tend to be better-characterized than hydrologic forcings in water temperature models, become more influential. Therefore, we assumed that prediction quality at the outlet is relatively more valuable than at headwater reaches. By assessing model performance using only these top 1% of runs ranked by $RMSE_w$ (inferred to be feasible solutions), we aimed to compare the potential performance of each model formulation when well-calibrated, discarding model runs where randomly sampled parameters did not reflect the physical reality of the basin.

3. Results

3.1. Calibrated Models

3.1.1. M1: Variable Groundwater Inflow Temperatures

M1 was the simplest of all model configurations tested, allowing variability only in parameters related to the temperature of groundwater inflow (C^1_{AT-GW} , C^2_{AT-GW} , C^3_{AT-GW} , W) and riparian shading (R_{shade}) (Table 1). The model configuration struggled to reproduce the

magnitude and variability of observed temperature time series at both the headwater and outlet (Figure 5, 6, 7). The 1st percentile of runs of M1 ranked by $RMSE_w$ had a mean RMSE of 1.41°C at the headwater gage and a mean RMSE of 1.20°C at the outlet, the worst of any model configuration (Figure 5). This set of best calibrated runs generally overestimated peak temperatures in the headwater reach, with a daily maxima error of 1.55°C (Figure 5, 6). Despite this strong positive bias in headwater reaches, M1 outlet predictions had a negative bias, underestimating daily minima by -0.99°C (Figure 5, 7). Although M1 failed to simulate the magnitude of diurnal temperature cycles, it captured outlet daily maxima the best of any model configuration (Figure 5, 7). The top overall M1 run ranked by $RMSE_w$ had a headwater RMSE of 0.75°C and an outlet RMSE of 1.08°C.

3.1.2. M2: Variable Groundwater Inflow Rate

In addition to the variables tuned in M1 (C^1_{AT-GW} , C^2_{AT-GW} , C^3_{AT-GW} , W), M2 added three further degrees of freedom, tuning the rate of groundwater inflow in first-, second-, and third-order streams (GW^1 , GW^2 , GW^3) (Table 1). M2 showed an improvement over calibrated M1 runs, with mean RMSE values of 1.11°C and 1.10°C at the headwater and outlet gages, respectively (Figure 5). In the headwater reach, lower mean error was largely driven by a narrowing of the diurnal temperature cycle and a shift of predicted daily minima to cooler temperatures (Figure 5, 6). Headwater maxima error was reduced to 1.04°C while daily minima error was reduced to 0.33°C. Increases in model performance at the outlet were linked to a more accurate simulation of minima temperature magnitude, though M2 showed little improvement over M1 in predicting daily maxima (Figure 5, 7). The negative bias in outlet predictions observed for M1 also persisted for M2 (Figure 7). The best calibration run for M2 had a headwater RMSE of 0.52°C and an outlet RMSE of 0.93°C.

3.1.3. M3: Conceptual Hyporheic Zone

M3 introduced considerable complexity to the M1 configuration, adding a conceptual hyporheic zone tuned by hyporheic lag time (H_{lag}) and hyporheic flow fraction (H^1_{frac} , H^2_{frac} , H^3_{frac}) parameters (Table 1). By adding a single additional degree of freedom over the M2 configuration (Table 1), M3 resulted in a considerable improvement in performance in comparison to both M1 and M2, with RMSE values of 0.81°C and 0.86°C in the headwaters and outlet, respectively (Figure 5, 6, 7). The M3 configuration greatly reduced the positive headwater bias observed in previous model configurations, reducing daily headwater minima error to -0.16°C and accurately representing the observed magnitude of diurnal variability (Figure 5, 6). We also observed gains in predicting daily minimum temperatures at the outlet, where minima error improved to 0.22°C (Figure 5). Although M3 had notable gains in performance over M1 and M2 at the outlet, outlet maxima error for M3 was the poorest of all model configurations tested (Figure 5). Across 5000 model runs, M3's best run by $RMSE_w$ had a headwater RMSE of 0.56°C and an outlet RMSE of 0.65°C.

3.1.4. M4: Variable Groundwater Inflow Rate and Conceptual Hyporheic Zone

M4 was the most complex configuration tested, combining aspects from both M2 and M3 to tune hyporheic flow parameters (H_{lag} , H^1_{frac} , H^2_{frac} , H^3_{frac}) and groundwater inflow parameters (GW^1 , GW^2 , GW^3) (Table 1). Despite increased complexity and additional degrees of freedom, M4 did not show a marked improvement in performance over M3, providing only marginal decreases in RMSE (Figure 5). Predicted water temperature envelopes from M3 and M4 at the headwater and outlet were difficult to distinguish visually (Figure 6, 7). M4 had the lowest RMSE values amongst all model configurations at both the headwater (0.74°C) and outlet (0.84°C) gages (Figure 5). Headwater daily maxima and minima prediction error for M4 were low,

with values of 0.21°C and -0.12°C respectively (Figure 5). M4 generally underestimated the magnitude of diurnal variability at the outlet, predicting daily maxima that were too cool and daily minima that were too warm (Figure 7). The best calibrated M4 run had a headwater RMSE of 0.65°C and an outlet RMSE of 0.61°C .

3.2. Optimal Calibrated Parameters

Mean optimal parameter values across the top 1% of calibrated model runs gave additional insight into differences in performance between model configurations (Table 3). Coefficients controlling the temperature of groundwater inflow for first-, second-, and third-order streams (C^1_{AT-GW} , C^2_{AT-GW} , C^3_{AT-GW}) were tuned for all model configurations. First-order coefficients (C^1_{AT-GW}) were calibrated to lower values than second- and third-order coefficients for all configurations, reflecting cooler inflow temperatures in upland areas of the catchment. M1 and M2, models without hyporheic flow, had optimal C_{AT-GW} values that were considerably higher than M3 and M4, models that did represent hyporheic flow. W , representing the number of days of mean air temperatures that were incorporated into estimates of groundwater temperatures, was consistently tuned to a value between 7 and 8 days for all model configurations. We also tuned R_{shade} , a coefficient used to adjust the degree riparian shading along the network, for all model versions, reflecting uncertainty in our estimates of riparian cover derived from gridded datasets. In M1 and M2, R_{shade} was tuned to 1.00 and 1.04 respectively, suggesting little bias in estimated riparian shading values. However, M3 and M4 had notably lower optimal R_{shade} values of 0.69 and 0.70, respectively.

We tuned coefficients used to adjust the rate of groundwater inflow for first-, second-, and third-order streams (GW^1 , GW^2 , GW^3) for configurations M2 and M4. In both model configurations, optimal coefficients ranged from 1.14 to 1.43, representing increased

groundwater inflow along all reaches in the stream network relative to NWM values. There was no clear relationship between optimal *GW* values and stream order for M2 and M4. Coefficients describing processes governing flow through a conceptual hyporheic zone, including hyporheic lag duration (H_{lag}) and hyporheic flow fraction (H^1_{frac} , H^2_{frac} , H^3_{frac}), were calibrated for M3 and M4. Hyporheic zone parameters were tuned to relatively similar values between the two configurations. In M3 and M4, the mean optimal hyporheic lag duration, controlling the time delay before hyporheic flow is returned to the channel, was equal to 11.38 and 11.74 hours, respectively. Hyporheic flow fraction, describing the proportion of streamflow that is routed into the conceptual hyporheic zone, had a strong negative relationship with stream order for both tested model configurations. For both M3 and M4, first-order reaches had the highest proportion of hyporheic flow, with coefficients of 0.67 and 0.68, respectively. Optimal hyporheic flow fractions then sequentially decreased for second- and third- order reaches.

4. Discussion

4.1. Evaluating Performance of Water Temperature Model Configurations

The quality of predictions made by our models confirm that water temperatures can be successfully simulated using inputs derived from a continental-scale hydrologic model (in this case, the NWM). All the model configurations we tested produced calibrated simulations with RMSEs near or under 1°C at both the headwater and outlet reaches. These errors are well under the 2°C RMSE threshold estimated by Yearsley (2012) as an acceptable measure of performance for water temperature modeling and compare well to other studies using similar modeling strategies (Sun et al., 2015; Yan et al., 2021; Yearsley, 2012; Yearsley et al., 2019). The ability of our model configurations to adequately predict water temperatures in H.J. Andrews, a complex forested headwater catchment, is promising for the incorporation of water temperature

modeling into the NWM framework.

Each of the model configurations we explored in this study represents a unique hypothesis for our understanding of how radiative and hydrologic processes combine to influence river thermal regimes. As expected, the addition of degrees of freedom to configurations progressively improved model performance amongst most error metrics, though this relationship was not strictly linear (Figure 5, 6, 7). The strongest contrast in model performance existed between configurations that represented the thermal effects of hyporheic exchange (M3, M4) and those that only tuned parameters related to groundwater inflow rate and temperature (M1, M2). M3 and M4 demonstrated clear advantages over M1 and M2 in all error metrics excluding headwater daily maxima, suggesting that the influence of hyporheic flow on temperatures in this basin is too large to disregard. As a high-relief mountain headwater catchment, it is unsurprising that hyporheic exchange is an influential thermal process in H.J. Andrews, and its role in hydrologic function in the region has been thoroughly documented (Becker et al., 2023; Herzog et al., 2019; Schmadel et al., 2017; Ward et al., 2017).

Notably, the addition of parameters controlling the rate of groundwater inflow to configurations M2 and M4 resulted in improvements in model error, but to differing degrees (Table 1). When we added inflow tuning parameters to M2, we observed a clear reduction in error across most metrics over the previous model version (M1) (Figure 5). By contrast, the addition of these calibrated parameters to M4 did not result in considerable improvement over M3 (Figure 5). We hypothesize that the difference in the marginal reductions in error between M2 and M4 is likely attributable to the presence of other tuned hydrologic parameters (hyporheic exchange) in the M4 configuration. In M2, tuning to groundwater inflow represented the only pathway for the model to account for uncertain hydrologic processes, including hyporheic

exchange. This flexibility gave M2 an greater advantage over M1. When the model included hyporheic exchange, as it did in M3 and M4, it appeared less crucial to tune groundwater inflow rate. Although the groundwater inflow parameters were tuned to route additional inflow into the channel (Table 3), the magnitude of these increases did not exceed 144% of NWM inflows. This suggests that the NWM's estimated groundwater contributions, at least in this basin, are roughly of the correct magnitude to accurately simulate thermal processes.

Although all configurations tested had the potential to simulate water temperatures with RMSEs below 1°C, the models struggled to simultaneously generate accurate predictions at both the headwater and outlet gages. While model predictions were often capable of providing accurate predictions at the headwater gage (Figure 4), many of these runs translated into poor outcomes at the outlet. For example, despite overestimating temperatures at the headwater, both M1 and M2 predicted outlet temperatures that were colder than observed (Figure 5, 6, 7).

We highlight two possible explanations for the model's inability to fit both the headwater and the outlet concurrently. First, because we tuned several parameters independently by stream order (Table 2), random variations in parameter values for second- and third-order reaches only influenced predictions at the outlet and not at the headwater. This could be alleviated by narrowing calibrated parameter ranges or by enforcing a constrained sampling strategy informed by process-based knowledge (e.g., hyporheic flow fraction in second- and third-order reaches must be tuned to be less than that of first-order reaches), as has been implemented in hydrological modeling studies (e.g., Hrachowitz et al., 2014). Tradeoffs in fitting the headwater and outlet could also be caused by a mischaracterization of heat fluxes along the network. This was most evident in configurations M1 and M2, where unrealistically warm headwaters were required to achieve the reasonable predictions at the outlet (Figure 6, 7). This effect was partially

- though not entirely - alleviated by the inclusion of a conceptual hyporheic zone in M3 and M4 (Figure 6, 7), indicating that model configurations presented here may not fully capture all relevant heat fluxes in the system.

Of the four configurations tested, the M3 and M4 configurations best approximated water temperature behavior in the H.J. Andrews catchment during this specific time period. However, this does not necessarily give insight into the efficacy of our modeling frameworks in other locations or at broader scales. Thermal regimes and their controlling processes are remarkably diverse, both within single catchments and across the North American continent (Fullerton et al., 2015; Maheu et al., 2016). As such, the optimal model configuration in one basin may not translate to a neighboring catchment or to a different geographic region. This potential heterogeneity in model performance presents challenges in extending water temperature modeling from individual catchments to the continental US.

4.2. Strategies for Constraining Uncertain Inputs

Despite the wealth of hydrologic data provided by NWM runs, several key inputs required to force our water temperature model were uncertain or altogether unknown. These included but are not limited to parameters governing the water temperature of groundwater inflows, headwater initiation water temperatures, riparian shading of channels, and hyporheic exchange. If water temperatures are to be accurately predicted, particularly in a physically-based modeling framework, approaches must be developed to estimate these parameters at a high spatial resolution (1 km) and at nationwide scales. Leveraging publicly available data external to the NWM and Monte Carlo calibration, we designed strategies to overcome data limitations that both enabled us to fit temperature behavior in the study catchment and that we envisioned could be easily scalable to broader modeling domains. We explore the viability of our proposed

strategies to estimate groundwater temperatures and hyporheic exchange.

4.2.1. Estimating Groundwater Temperatures Key to Accurate Water Temperature

Predictions

The heat flux associated with groundwater inflow, although often smaller in magnitude than fluxes at the air-water interface, can be a strong control on the water temperature of streams (Caissie, 2006; Caissie & Luce, 2017; Kurylyk et al., 2016). Groundwater inflow is particularly influential to water temperatures in forested headwater streams, as the magnitude of other radiative and turbulent heat fluxes are diminished (Caissie & Luce, 2017; Ouellet et al., 2020). In reaches where flows are primarily sourced from relatively cold groundwater, water temperatures are cooler and typically have narrower diel ranges (Hannah & Garner, 2015). Rigorous on-site monitoring is required to determine the rate and temperature of groundwater inflow to channels (Caissie & Luce, 2017), making advective fluxes challenging to quantify at a broader spatial extent. As such, the temperature of groundwater fluxes to streams represents a substantial source of uncertainty in physically-based models.

In water temperature models, the temperature of groundwater inflow is generally set to the mean annual air temperature, mimicking the temperature of deep groundwater (Kurylyk et al., 2016; MacDonald et al., 2014). Perhaps counterintuitively, the temperatures of groundwater and inflows to streams are not always equivalent. The temperature of subsurface inflow when it enters a stream, whether sourced by shallow flow paths, warmed through bed conduction, or mixed with hyporheic waters, is often warmer than that of deep groundwater in summer and cooler in winter (Kurylyk et al., 2016; Leach & Moore, 2014). Advected inflow temperatures are also more temporally variable than that of deep groundwater and are loosely coupled to daily mean air temperatures (Leach & Moore, 2014). Past modeling studies have attempted to account

for the time-varying nature of inflow temperatures using non-linear regression (Mohseni et al., 1998) to predict inflow temperatures from smoothed air temperatures (van Vliet et al., 2012; Yearsley, 2012).

As groundwater temperature in our modeling approach was critical not only for forcing subsurface fluxes but also for setting the upstream boundary condition, our approach to estimating groundwater temperatures (Section 2.3.5) was intended to incorporate both variable sourcing depth and a lagged relationship between inflow temperature and air temperature (Figure 3). Under the assumption that the sourcing depth of inflow may vary down-network, we allowed the coefficient governing the effective source depth of inflows (C_{AT-GW}) to independently vary by stream order. The performance of our calibrated approach can be roughly assessed by evaluating error in modeled headwater temperatures, which are closely coupled to the temperature of inflows. By this metric, our approach was successful when well-calibrated, with headwater RMSE values approaching a minimum of 0.5°C for all configurations, though many of these runs corresponded with large errors at the outlet (Figure 4). However, it remains challenging to disentangle the true effectiveness of our groundwater temperature approach from other potentially mischaracterized or absent streambed processes, including bed conduction and hyporheic flow. For example, groundwater inflow temperatures were consistently tuned warmer for models without hyporheic flow processes to fit behavior at the outlet (M1/M2) (Table 3). This suggests that for the simplified model configurations, groundwater inflow temperatures may be tuned to compensate for missing thermal processes, resulting in poor performance in certain regions of the stream network.

Given the degree of heterogeneity across all US catchments, our calibration for groundwater inflow temperature parameters may be cumbersome to apply to a continental-scale

domain. The successes we observed in reproducing water temperatures using a tuned groundwater temperature approach are specific only to the study catchment during a period of low flow, and do not necessarily indicate transferability to other basins or time periods. By randomly tuning inflow temperature parameters, we sought to demonstrate that our model was capable of simulating water temperature behavior given well-calibrated parameters. This contrasts with a typical approach to modeling physical processes across broad spatial domains, where focus is instead placed on achieving acceptable mean model performance with uncertain parameter estimates. Expanding our water temperature model beyond the test basin would require a more complex approach to accurately simulate the broad diversity of subsurface flow dynamics across catchments and climates. Simulated groundwater temperatures at the continental scale would need to be temporally and spatially variable, reflecting site, basin, and regional controls on water temperature processes (Hannah & Garner, 2015). This approach would also need to incorporate the influence of seasonal snowpack on groundwater temperatures. Spatial statistical models or machine learning techniques could be an efficient and effective tool to predict variability in groundwater temperatures across the US, generating both upstream boundary conditions and inflow temperatures to drive a physically-based model (Dugdale et al., 2017).

4.2.2. Is a Conceptual Hyporheic Zone Needed?

Hyporheic flow, characterized by flow paths that originate in the stream, travel through the subsurface, and eventually return to the stream, is an important process controlling the magnitude and timing of water temperature variability (Arrigoni et al., 2008; Boano et al., 2014; Hannah et al., 2009). Particularly in high relief headwaters like those of H.J. Andrews, a considerable portion of streamflow can pass through the hyporheic zone, returning flows with

temperatures that are lagged and buffered compared to instream waters (Arrigoni et al., 2008; Schmadel et al., 2017; Ward et al., 2016; Wondzell, 2011). In some cases, hyporheic advective heat fluxes may comprise 25% of total net radiation in headwater streams (Moore et al., 2005). Although hyporheic flow can be an influential thermal process, many broad-scale hydrologic models, including the NWM, do not include hyporheic processes in their representation of river networks. Similarly, water temperature models also often neglect heat fluxes related to hyporheic exchange (Kurylyk et al., 2016).

Hyporheic flow and its associated effects on water temperatures are remarkably difficult to characterize, even when employing field observations and flow tracing techniques. In the absence of field measurements, our approach (outlined in Section 2.3.6) was intended to be conceptual rather than to give insight into true hyporheic behavior at this or any other study site. Our strategy for approximating hyporheic exchange did not represent physical mass transfer within the model and treated flow paths as point features, returning flow to the stream at the same point it originated. This simplification ignores the complex, 3D nature of hyporheic flow cells that can travel a considerable distances down-valley (Tonina & Buffington, 2009). For this reason, the calibrated hyporheic flow fraction values and time lags used in our model (Table 3) should not be taken as explicit estimations of hyporheic flow processes at H.J. Andrews.

The contrast in model performance between configurations that included hyporheic processes (M3 and M4) and those that did not (M1 and M2) suggests that incorporating, or at least mimicking, hyporheic exchange is critical to simulating water temperatures in the study basin (Figure 5, 6, 7). This finding was expected, given the multitude of studies describing the influence of hyporheic exchange on hydrological processes in H.J. Andrews (Becker et al., 2023; Kasahara & Wondzell, 2003; Schmadel et al., 2017; Ward et al., 2018b, 2019). The gains in

performance we observed when including hyporheic processes were primarily linked to an improved estimation of daily minimum water temperatures (M3 and M4; Figure 5). This likely indicates that our representation of hyporheic flow, which returned warmer daytime waters roughly 12 hours later (Table 3), served to address missing nighttime streambed fluxes. The influence of hyporheic flow in our model was tuned to decrease with increasing stream order (Table 3), matching our understanding of how hyporheic exchange evolves down-network (Ward et al., 2019). We note that because our model configurations did not resolve bed conduction heat fluxes, parameters associated with hyporheic exchange may be simultaneously accounting for the effects of both hyporheic flow and bed conduction. If bed conduction remains absent from future model configurations, it may be beneficial to instead tune a single time-varying conceptual term that integrates all lagged streambed heat fluxes.

Our results indicate that in H.J. Andrews, and likely other basins in similar settings, the inclusion of hyporheic flow processes can improve predictions of hourly water temperatures throughout the stream network. However, incorporating the thermal effects of hyporheic flow into a water temperature model at the continental scale of the NWM may pose challenges. Hyporheic flow dynamics are patchy and site-specific, varying considerably within stream networks, between physiographic regions, and across different flow conditions (Wondzell, 2011). Clearly, it is not feasible to simulate hyporheic flow paths for all US river reaches, particularly given the absence of field observations along many rivers. Nevertheless, flexible modeling strategies could be designed to incorporate hyporheic processes only where they are the most influential to water temperatures. In such a framework, hyporheic flow could be tuned to improve water temperature predictions only in select regions and along low-order streams where hyporheic advective fluxes have the strongest influence on hourly water temperatures.

Higher-order streams could then be represented by conceptually simpler modeling frameworks, improving computational efficiency.

4.3. Challenges and Opportunities in Expanding from the Catchment to Continental Scale

Though our study focused on developing modeling capabilities in a single catchment, our primary motivation was to evaluate the capacity for water temperature prediction to be coupled to the NWM at broader scales. We see several major challenges facing the application NWM to the simulation of water temperatures. These generally stem from the NWM's simplified representation of hydrological processes and represent areas of needed future study.

Foremost, the predictions made by our NWM-based water temperature model are ultimately limited by the accuracy and uncertainty of NWM simulations. Beyond discharge, the target variable for NWM calibration (Gochis et al., 2019), our physically-based temperature model is also reliant on several NWM states and parameters, including channel dimensions, groundwater inflow, surface runoff, and stream velocity. Despite the NWM's demonstrated ability to produce reasonable predictions of discharge, particularly in large river basins (Boyd & Kasper, 2003; Hansen et al., 2019; Salas et al., 2018), it can in some cases struggle to reproduce variability in other model states (e.g., soil moisture, snowpack; Garousi-Nejad & Tarboton, 2022; Wan et al., 2022). As these model states are not explicit targets for calibration in the NWM, their mischaracterization could propagate error into predicted water temperatures.

The NWM's representation of river network extent, derived from NHD Plus flow paths (McKay et al., 2012; Salas et al., 2018), is another potential source of uncertainty to water temperature modeling. Although the NHD provides exceptional spatial coverage of river networks across the US, it has been shown to systematically underestimate the true extent of river density (Elmore et al., 2013). In the H.J. Andrews catchment, the NWM models streamflow

along only 34.5 km of river length. By contrast, Ward et al. (2019) estimated the total river length in H.J. Andrews was 242 km using on-site lidar assessments and flow accumulation modeling. The omission of numerous headwater reaches by the NWM could have implications for the prediction of water temperatures in low stream order catchments. The amount of time water is exposed to radiation at the surface, which is influenced by the location of channel initiation, is a strong control on water temperature magnitude and variability (Yearsley, 2012). Due to this uncertainty, headwater temperatures may be difficult to accurately simulate in catchments where the true location of streamflow initiation is mischaracterized.

Despite these limitations, the unique framework of the NWM presents promising opportunities for the prediction of water temperatures at broad scales. While our models coupled with NWM version 2.1, NOAA is set to introduce the Next Generation Water Resources Modeling Framework (NextGen) in the coming years, with exciting implications for water temperature prediction. Based on the understanding that certain model configurations may perform better in specific catchments, the flexible and interoperable NextGen framework will enable domains to be simulated using model conceptualizations that best match the dominant hydrologic controls in a particular region (NOAA, 2021). By leveraging the NextGen framework, the same principle could be applied to tailor water temperature model configurations to specific catchments or regions. For example, in forested headwater catchments, a more parameterized model configuration could be used to resolve complex hyporheic heat fluxes. In contrast, water temperature predictions in large high order streams could be made using comparatively simpler and more efficient models. The potential flexibility offered by the NextGen framework could bring about profound advances in prediction quality, resolution, and extent across the US.

5. Conclusion

In this study, we developed and evaluated the capabilities of physically-based high resolution water temperature model driven by forcings and outputs from the National Water Model (NWM). Through the sequential calibration of four model configurations of increasing complexity, we demonstrated that the inclusion of heat fluxes at the streambed interface (e.g., hyporheic flow) is critical for simulating hourly water temperatures in a forested headwater catchment. The performance of the best-fitting model configuration was comparable to or better than other physically-based water temperature models, suggesting that the NWM can be an effective foundation for water temperature prediction.

While this work focuses on model development in a single catchment, the expansion of NWM-based water temperature modeling to broader spatial domains would improve understanding and management of the complex mosaic of US river thermal regimes. Hourly water temperature forecasts along all US river reaches could provide actionable information that would inform the management of fisheries and other sensitive aquatic ecosystems. Such a model would present a clear improvement over the patchwork of water temperature monitoring stations currently active across the continent. With the introduction of the NextGen NWM framework on the horizon, we recommend the continued development, exploration, and evaluation of NWM-coupled water temperature models to expand predictions from single catchments to all US watersheds.

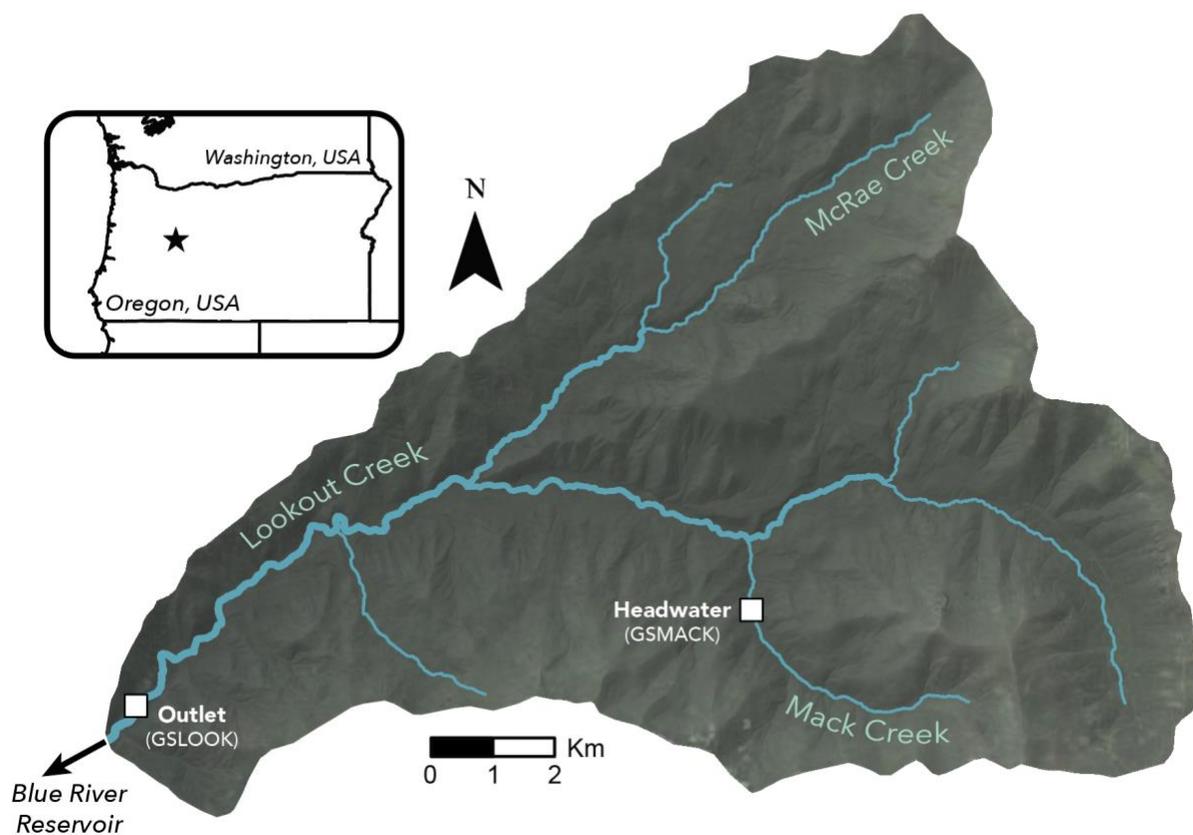
Figures

Figure 1. Location of water temperature gages (‘Headwater’: GSMACK, ‘Outlet’: GSLOOK) with the H.J. Andrews Experimental Forest watershed in relation to channels identified by the National Water Model.

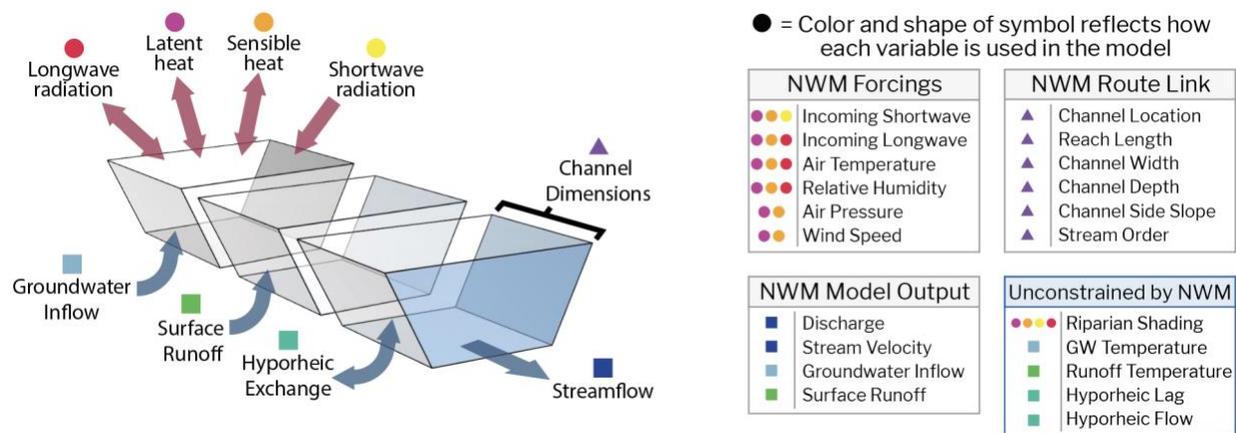


Figure 2. Primary heat fluxes represented in the water temperature model. Model data sources fall into four broad categories: NWM gridded forcings, NWM model outputs, NWM channel route link files, and external data unconstrained by the NWM. The color and shape of symbology indicates how each variable contributes to calculated heat fluxes in the model.

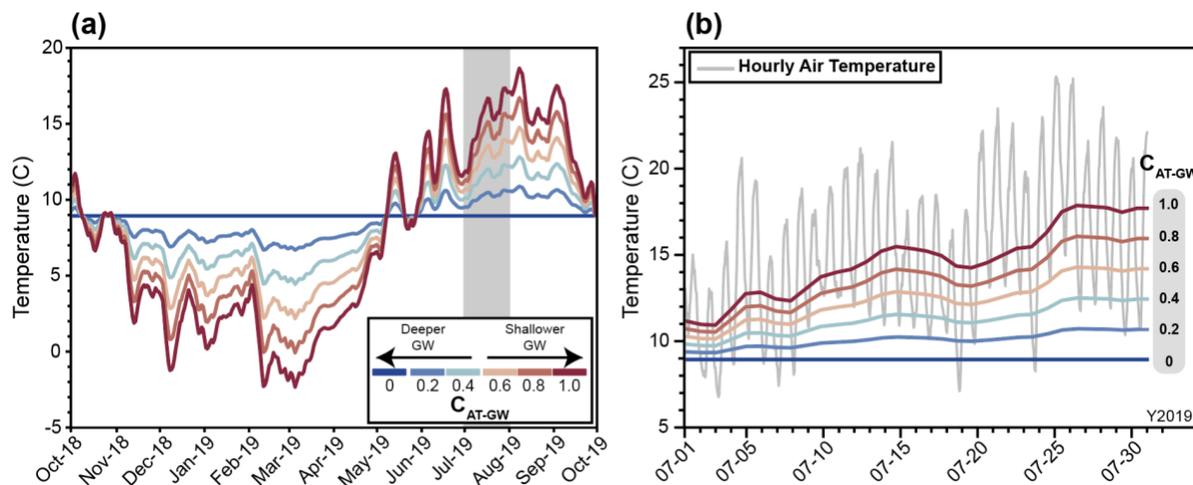


Figure 3. Estimated groundwater inflow temperatures for (a) water year 2019 and (b) July 2019 study period at the headwaters of Mack Creek for a range of C_{AT-GW} (air temperature scaling coefficient) values. Values of C_{AT-GW} closer to 0 represent relatively deeper sourcing depths while values of C_{AT-GW} closer to 1 represent relatively shallower sourcing depths.

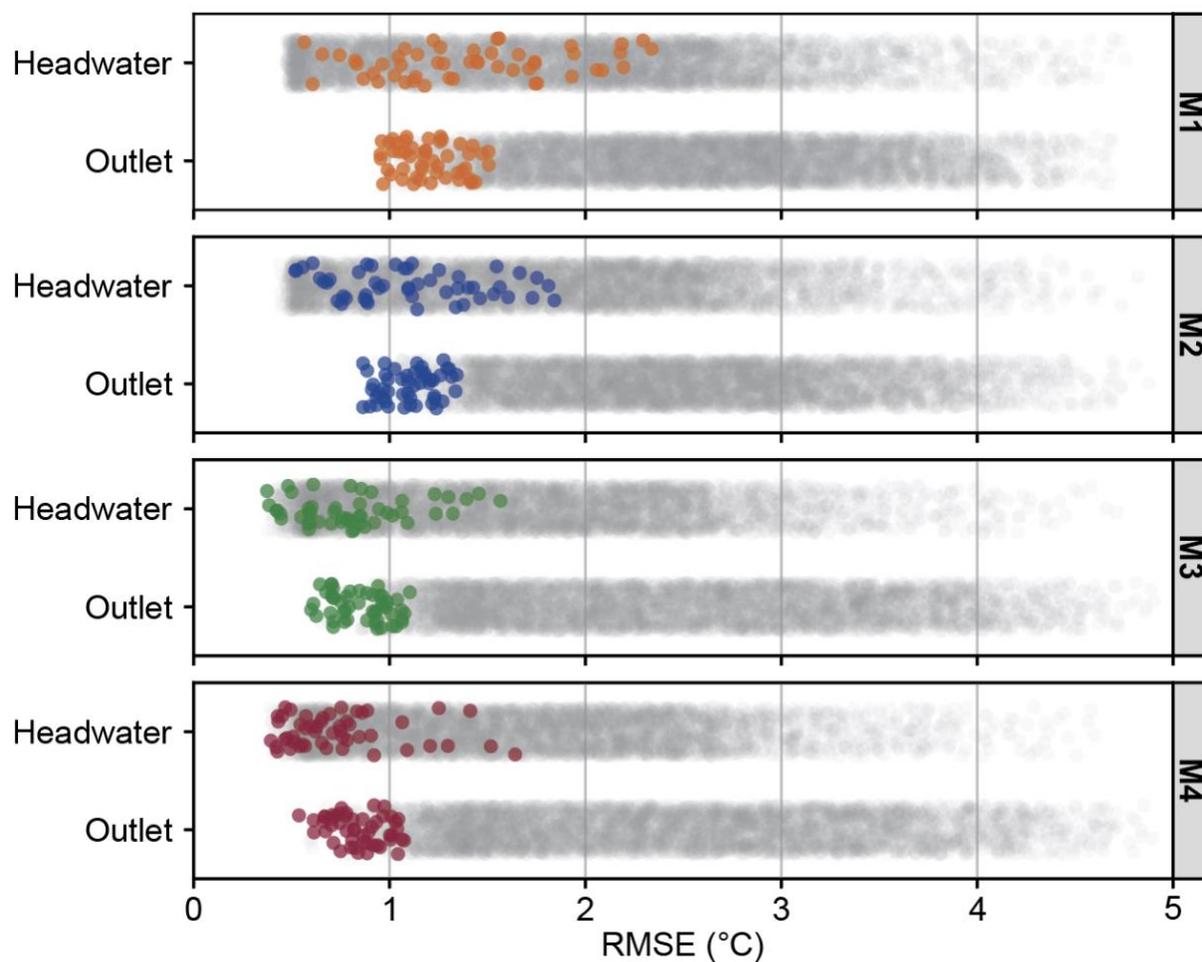


Figure 4. Simulated water temperature RMSE at headwater (Mack Creek) and outlet (Lookout Creek) gages for 5,000 Monte Carlo calibration runs of each model configuration (gray). Top 50 (1st percentile) runs of each model configuration, ranked by $RMSE_w$ (weighted headwater (25%) and outlet (75%) RMSE), are highlighted.

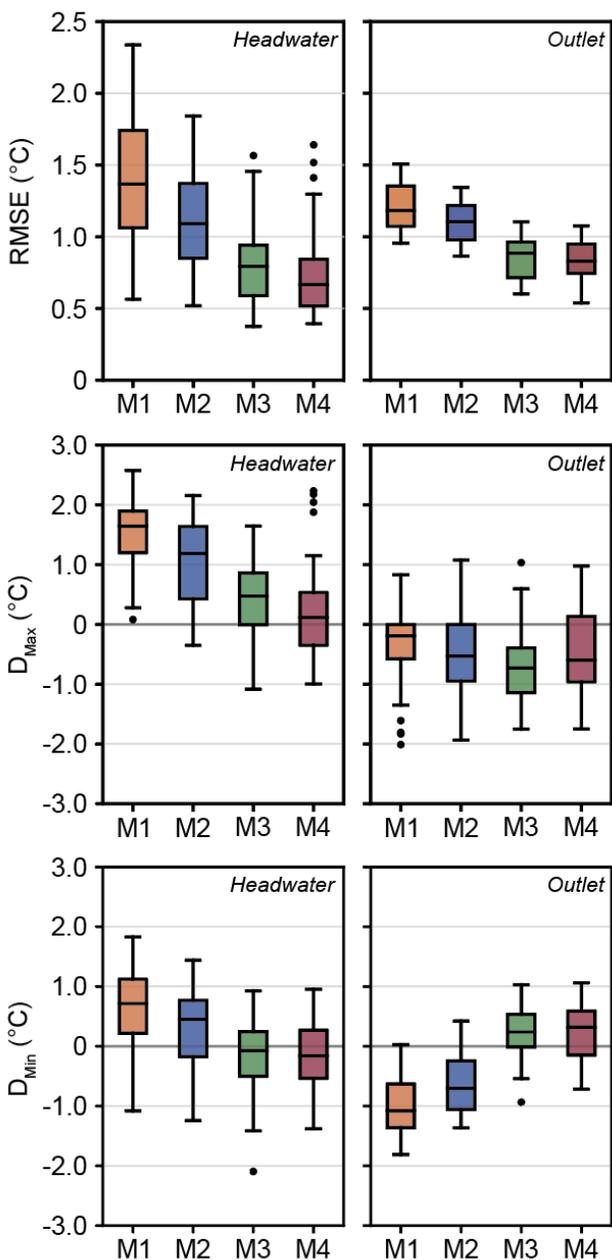


Figure 5. Performance of four model configurations at the headwater (Mack Creek) and outlet (Lookout Creek) gages, evaluated across three metrics of model error (RMSE: root mean square error; D_{Max} : daily maxima error; D_{Min} : daily minima error). Error metrics calculated using top 50 runs (1st percentile) ranked by weighted headwater and outlet RMSE for each configuration.

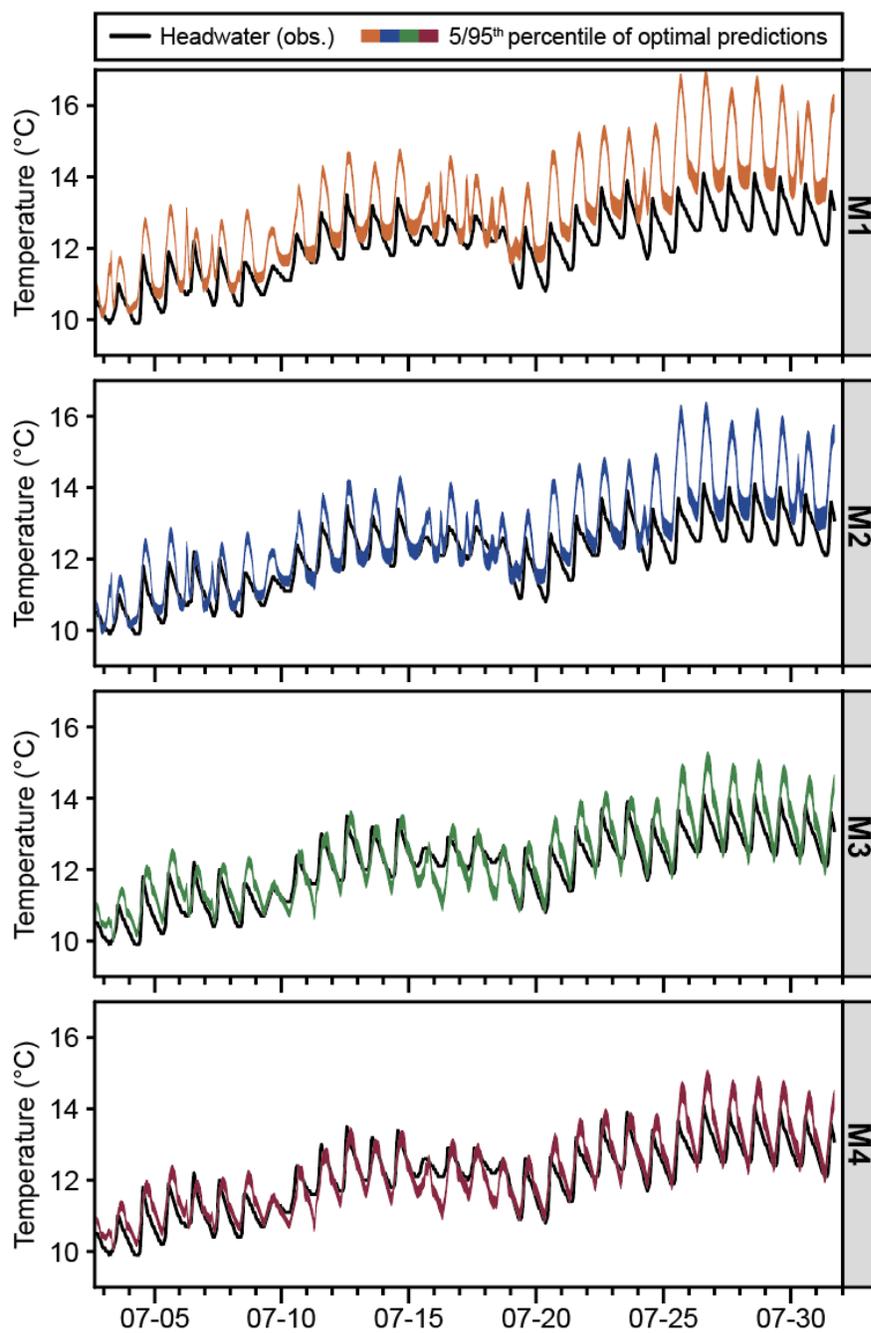


Figure 6. Observed headwater temperatures (black) and 5/95th confidence envelope of water temperature predictions at the headwater gages across model configurations M1, M2, M3, and M4 for the top 50 runs (1st percentile) ranked by weighted headwater and outlet RMSE.

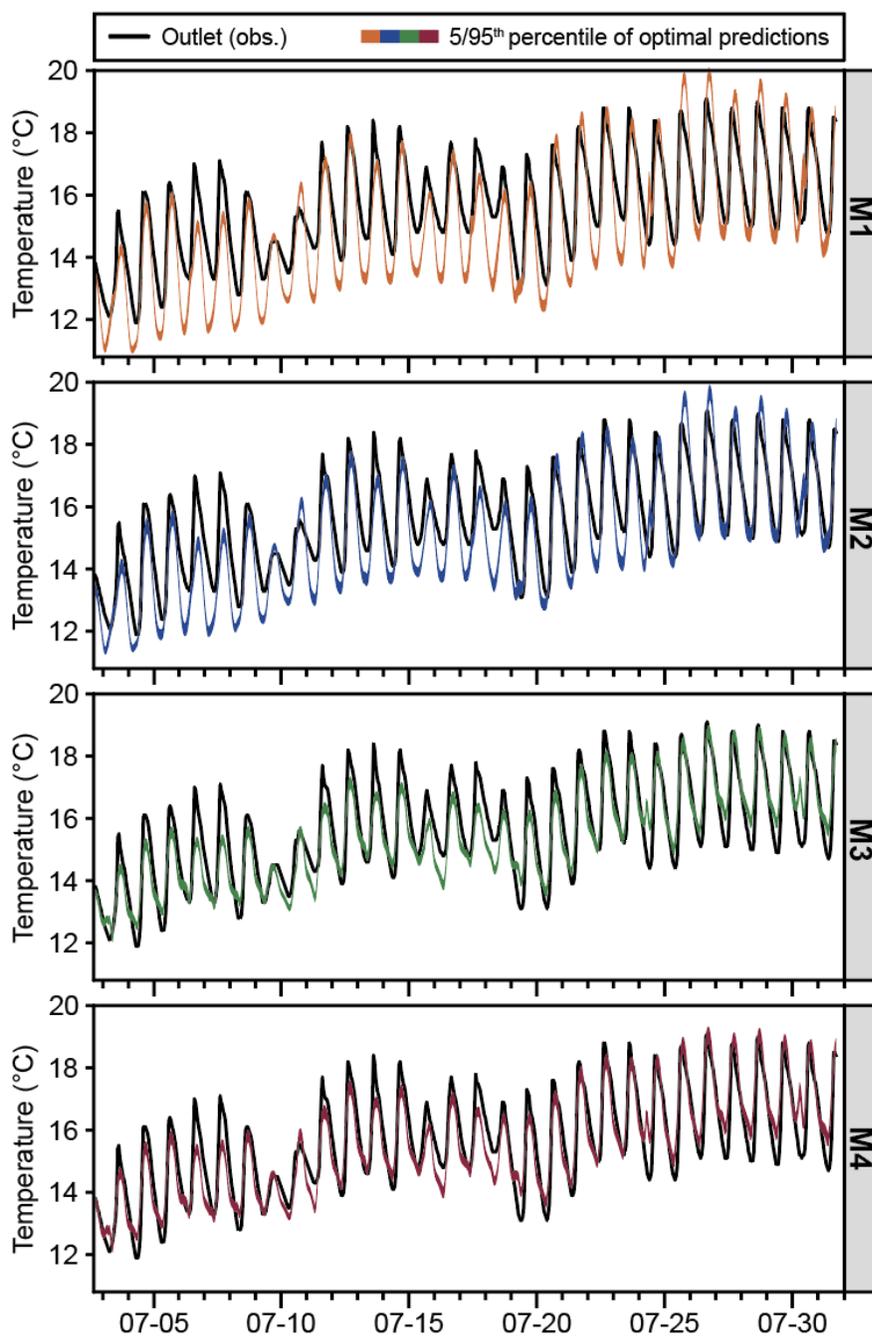


Figure 7. Observed outlet temperatures (black) and 5/95th confidence envelope of water temperature predictions at the outlet gage across model configurations M1, M2, M3, and M4 for the top 50 runs (1st percentile) ranked by weighted headwater and outlet RMSE.

Tables

Table 1. Water temperature model formulations, tuned parameters, and number of parameters.

Model	Tuned Parameters	Parameter Number
M1	$C^1_{AT-GW}, C^2_{AT-GW}, C^3_{AT-GW}, W, R_{shade}$	5
M2	$C^1_{AT-GW}, C^2_{AT-GW}, C^3_{AT-GW}, W, R_{shade}, GW^1, GW^2, GW^3$	8
M3	$C^1_{AT-GW}, C^2_{AT-GW}, C^3_{AT-GW}, W, R_{shade}, H_{lag}, H^1_{frac}, H^2_{frac}, H^3_{frac}$	9
M4	$C^1_{AT-GW}, C^2_{AT-GW}, C^3_{AT-GW}, W, R_{shade}, GW^1, GW^2, GW^3, H_{lag}, H^1_{frac}, H^2_{frac}, H^3_{frac}$	12

Table 2. Parameter definitions and tuning ranges for models M1-M4.

Notation	Parameter	Units	Calibration Range	Model
C^1_{AT-GW}	AT-GW coefficient (first-order)	unitless	0 - 1	M1, M2, M3, M4
C^2_{AT-GW}	AT-GW coefficient (second-order)	unitless	0 - 1	M1, M2, M3, M4
C^3_{AT-GW}	AT-GW coefficient (third-order)	unitless	0 - 1	M1, M2, M3, M4
W	Air temperature moving window duration	days	2 - 14	M1, M2, M3, M4
R_{shade}	Riparian shading coefficient	unitless	0.5 - 2	M1, M2, M3, M4
GW^1	Groundwater inflow rate coefficient (first-order)	unitless	0.5 - 2	M2, M4
GW^2	Groundwater inflow rate coefficient (second-order)	unitless	0.5 - 2	M2, M4
GW^3	Groundwater inflow rate coefficient (third-order)	unitless	0.5 - 2	M2, M4
H_{lag}	Hyporheic lag duration	hours	2 - 24	M3, M4
H^1_{frac}	Hyporheic flow fraction (first-order)	unitless	0 - 1	M3, M4
H^2_{frac}	Hyporheic flow fraction (second-order)	unitless	0 - 1	M3, M4
H^3_{frac}	Hyporheic flow fraction (third-order)	unitless	0 - 1	M3, M4

Table 3. Optimal mean parameter values for top 50 runs of each model (M1-M4).

Parameter	M1	M2	M3	M4
C^1_{AT-GW}	0.63	0.56	0.26	0.27
C^2_{AT-GW}	0.83	0.84	0.49	0.55
C^3_{AT-GW}	0.64	0.75	0.53	0.59
W	7.20	7.22	7.88	7.42
R_{shade}	1.00	1.04	0.69	0.70
GW^1	-	1.14	-	1.26
GW^2	-	1.43	-	1.33
GW^3	-	1.34	-	1.28
H_{lag}	-	-	11.38	11.74
H^1_{frac}	-	-	0.67	0.68
H^2_{frac}	-	-	0.48	0.47
H^3_{frac}	-	-	0.18	0.14

References

- Arrigoni, A. S., Poole, G. C., Mertes, L. A. K., Daniel, S. J. O., Woessner, W. W., & Thomas, S. A. (2008). *Buffered, lagged, or cooled? Disentangling hyporheic influences on temperature cycles in stream channels*. *44*, 1–13. <https://doi.org/10.1029/2007WR006480>
- Becker, P. S., Ward, A. S., Herzog, S. P., & Wondzell, S. M. (2023). Testing Hidden Assumptions of Representativeness in Reach-Scale Studies of Hyporheic Exchange. *Water Resources Research*, *59*(1). <https://doi.org/10.1029/2022WR032718>
- Benyahya, L., Caissie, D., Satish, M. G., & El-Jabi, N. (2012). Long-wave radiation and heat flux estimates within a small tributary in Catamaran Brook (New Brunswick, Canada). *Hydrological Processes*, *26*(4), 475–484. <https://doi.org/10.1002/hyp.8141>
- Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B. M. J., & Bobée, B. (2007). A Review of Statistical Water Temperature Models. In *Canadian Water Resources Journal* (Vol. 32, Issue 3, pp. 179–192). <https://doi.org/10.4296/cwrj3203179>
- Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, *52*(4), 603–679. <https://doi.org/10.1002/2012RG000417>
- Boyd, M. S., & Kasper, B. (2003). *Analytical Methods for Dynamic Open Channel Heat and Mass Transfer Methodology for the Heat Source Model Version 7.0*. <http://www.deq.state.or.us/wq/TMDLs/tools.htm>
- 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., Kurylyk, B. L., St-Hilaire, A., El-Jabi, N., & MacQuarrie, K. T. B. (2014).

- Streambed temperature dynamics and corresponding heat fluxes in small streams experiencing seasonal ice cover. *Journal of Hydrology*, 519(PB), 1441–1452.
<https://doi.org/10.1016/j.jhydrol.2014.09.034>
- Caissie, D., & Luce, C. H. (2017). Quantifying streambed advection and conduction heat fluxes. *Water Resources Research*, 53(2), 1595–1624. <https://doi.org/10.1002/2016WR019813>
- Caldwell, P., Segura, C., Gull Laird, S., Sun, G., McNulty, S. G., Sandercock, M., Boggs, J., & Vose, J. M. (2015). Short-term stream water temperature observations permit rapid assessment of potential climate change impacts. *Hydrological Processes*, 29(9), 2196–2211. <https://doi.org/10.1002/hyp.10358>
- 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(October), 97–113. <https://doi.org/10.1016/j.earscirev.2017.10.009>
- Dyer, J., Mercer, A., & Raczyński, K. (2022). Identifying Spatial Patterns of Hydrologic Drought over the Southeast US Using Retrospective National Water Model Simulations. *Water (Switzerland)*, 14(10). <https://doi.org/10.3390/w14101525>
- Elmore, A. J., Julian, J. P., Guinn, S. M., & Fitzpatrick, M. C. (2013). Potential Stream Density in Mid-Atlantic U.S. Watersheds. *PLoS ONE*, 8(8), 1–15.
<https://doi.org/10.1371/journal.pone.0074819>
- Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, 47(11). <https://doi.org/10.1029/2010WR010174>
- Ficke, A. D., Myrick, C. A., & Hansen, L. J. (2007). Potential impacts of global climate change on freshwater fisheries. *Reviews in Fish Biology and Fisheries*, 17(4), 581–613.

<https://doi.org/10.1007/s11160-007-9059-5>

Förster, H., & Lilliestam, J. (2010). Modeling thermoelectric power generation in view of climate change. *Regional Environmental Change*, 10(4), 327–338.

<https://doi.org/10.1007/s10113-009-0104-x>

Fullerton, A. H., Torgersen, C. E., Lawler, J. J., Faux, R. N., Steel, E. A., Beechie, T. J., Ebersole, J. L., & Leibowitz, S. G. (2015). Rethinking the longitudinal stream temperature paradigm: Region-wide comparison of thermal infrared imagery reveals unexpected complexity of river temperatures. *Hydrological Processes*, 29(22), 4719–4737. <https://doi.org/10.1002/hyp.10506>

Garousi-Nejad, I., & Tarboton, D. G. (2022). A comparison of National Water Model retrospective analysis snow outputs at snow telemetry sites across the Western United States. *Hydrological Processes*, 36(1). <https://doi.org/10.1002/hyp.14469>

Glose, A. (2013). *Stream Heat Budget Modeling with the HFLUX Stream Temperature Solver: Model Development, Verification, and Applications*. Syracuse University.

Glose, A., Lautz, L. K., & Baker, E. A. (2017). Stream heat budget modeling with HFLUX: Model development, evaluation, and applications across contrasting sites and seasons. *Environmental Modelling and Software*, 92, 213–228.

<https://doi.org/10.1016/j.envsoft.2017.02.021>

Gochis, D. J., Barlage, M., Cabell, R., Casali, M., Dugger, A., Fitzgerald, K., Mcallister, M., McCreight, J., Rafieeinassab, A., Read, L., Sampson, K., Yates, D., & Zhang, Y. (2021). *The NCAR WRF-Hydro® Modeling System Technical Description*.

<https://ral.ucar.edu/sites/default/files/public/WRFHydroV511TechnicalDescription.pdf>.

Gochis, D.J., Yates, D., Sampson, K., Dugger, A., McCreight, J., Barlage, M., RafieeiNasab, A.

(2019). *Overview of National Water Model Calibration General Strategy and Optimization*. National Center for Atmospheric Research.

https://ral.ucar.edu/sites/default/files/public/9_RafieeiNasab_CalibOverview_CUAHSI_Fall019_0.pdf

Gregory, S.; Johnson, S. 2019. Stream and air temperature data from stream gages and stream confluences in the Andrews Experimental Forest, 1950 to present. Long-Term Ecological Research. Forest Science Data Bank, Corvallis, OR. [Database]. Available: <http://andlter.forestry.oregonstate.edu/data/abstract.aspx?dbcode=HT004>.
<https://doi.org/10.6073/pasta/9437d1603044f5b92189110dd8343763>.

Hannah, D. M., & Garner, G. (2015). River water temperature in the United Kingdom: Changes over the 20th century and possible changes over the 21st century. *Progress in Physical Geography*, 39(1), 68–92. <https://doi.org/10.1177/0309133314550669>

Hannah, D. M., Malcolm, I. A., & Bradley, C. (2009). Seasonal hyporheic temperature dynamics over riffle bedforms. *Hydrological Processes*, 23(15), 2178–2194.
<https://doi.org/10.1002/hyp.7256>

Hansen, C., Shafiei Shiva, J., McDonald, S., & Nabors, A. (2019). Assessing Retrospective National Water Model Streamflow with Respect to Droughts and Low Flows in the Colorado River Basin. *Journal of the American Water Resources Association*, 55(4), 964–975. <https://doi.org/10.1111/1752-1688.12784>

Havens, K. E., & Paerl, H. W. (2015). Climate Change at a Crossroad for Control of Harmful Algal Blooms. *Environmental Science and Technology*, 49(21), 12605–12606.
<https://doi.org/10.1021/acs.est.5b03990>

Herzog, S. P., Ward, A. S., & Wondzell, S. M. (2019). Multiscale Feature-feature Interactions

Control Patterns of Hyporheic Exchange in a Simulated Headwater Mountain Stream.
Water Resources Research, 55(12), 10976–10992.

<https://doi.org/10.1029/2019WR025763>

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G., & Gascuel-Oudou, C. (2014). Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resources Research*, 50(9), 7445–7469. <https://doi.org/10.1002/2014WR015484>

Husain, S. Z., & Girard, C. (2017). Impact of consistent semi-Lagrangian trajectory calculations on numerical weather prediction performance. *Monthly Weather Review*, 145(10), 4127–4150. <https://doi.org/10.1175/MWR-D-17-0138.1>

Isaak, D. J., Wollrab, S., Horan, D., & Chandler, G. (2012). Climate change effects on stream and river temperatures across the northwest U.S. from 1980-2009 and implications for salmonid fishes. *Climatic Change*, 113(2), 499–524. <https://doi.org/10.1007/s10584-011-0326-z>

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software*, 21(5), 602–614. <https://doi.org/10.1016/j.envsoft.2006.01.004>

Jennings, K., & Jones, J. A. (2015). Precipitation-snowmelt timing and snowmelt augmentation of large peak flow events, western Cascades, Oregon. *Water Resources Research*, 51(9), 7649–7661. <https://doi.org/10.1002/2014WR016877>

Johnson, S. L. (2004). Factors influencing stream temperatures in small streams: Substrate effects and a shading experiment. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(6), 913–923. <https://doi.org/10.1139/F04-040>

- Johnson, S. L., Henshaw, D., Downing, G., Wondzell, S., Schulze, M., Kennedy, A., Cohn, G., Schmidt, S. A., & Jones, J. A. (2021). Long-term hydrology and aquatic biogeochemistry data from H. J. Andrews Experimental Forest, Cascade Mountains, Oregon. *Hydrological Processes*, 35(5). <https://doi.org/10.1002/hyp.14187>
- Julien, P. Y., Saghafian, B., & Ogden, F. L. (1995). Raster-based Hydrologic Modeling of Spatially-Variied Surface Runoff. *Journal of the American Water Resources Association*, 31(3), 523–536. <https://doi.org/10.1111/j.1752-1688.1995.tb04039.x>
- Kalny, G., Laaha, G., Melcher, A., Trimmel, H., Weihs, P., & Rauch, H. P. (2017). The influence of riparian vegetation shading on water temperature during low flow conditions in a medium sized river. *Knowledge and Management of Aquatic Ecosystems*, 418. <https://doi.org/10.1051/kmae/2016037>
- Kasahara, T., & Wondzell, S. M. (2003). Geomorphic controls on hyporheic exchange flow in mountain streams. *Water Resources Research*, 39(1), SBH 3-1-SBH 3-14. <https://doi.org/10.1029/2002wr001386>
- Kaylor, M. J., VerWey, B. J., Cortes, A., & Warren, D. R. (2019). Drought impacts to trout and salamanders in cool, forested headwater ecosystems in the western Cascade Mountains, OR. *Hydrobiologia*, 833(1), 65–80. <https://doi.org/10.1007/s10750-019-3882-2>
- Kurylyk, B. L., Moore, R. D., & Macquarrie, K. T. B. (2016). Scientific briefing: Quantifying streambed heat advection associated with groundwater-surface water interactions. *Hydrological Processes*, 30(6), 987–992. <https://doi.org/10.1002/hyp.10709>
- Lahmers, T. M., Gupta, H., Castro, C. L., Gochis, D. J., Yates, D., Dugger, A., Goodrich, D., & Hazenberg, P. (2019). Enhancing the Structure of the WRF-Hydro Hydrologic Model for Semiarid Environments. *Journal of Hydrometeorology*, 20(4), 691–714.

<https://doi.org/10.1175/JHM-D-18-0064.1>

- Lahmers, T. M., Hazenberg, P., Gupta, H., Castro, C., Gochis, D., Dugger, A., Yates, D., Read, L., Karsten, L., & Wang, Y.-H. (2021). Evaluation of NOAA National Water Model Parameter Calibration in Semi-Arid Environments Prone to Channel Infiltration. *Journal of Hydrometeorology*. <https://doi.org/10.1175/jhm-d-20-0198.1>
- LANDFIRE, (2020). Existing Vegetation Cover Layer, LANDFIRE 2.2.0, U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture. accessed 17 October 2022, <http://landfire.cr.usgs.gov/viewer>.
- LANDFIRE, (2020). Existing Vegetation Height Layer, LANDFIRE 2.2.0, U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture. accessed 17 October 2022, <http://landfire.cr.usgs.gov/viewer>.
- Leach, J. A., & Moore, R. D. (2014). Winter stream temperature in the rain-on-snow zone of the Pacific Northwest: Influences of hillslope runoff and transient snow cover. *Hydrology and Earth System Sciences*, 18(2), 819–838. <https://doi.org/10.5194/hess-18-819-2014>
- Lee, S. Y., Fullerton, A. H., Sun, N., & Torgersen, C. E. (2020). Projecting spatiotemporally explicit effects of climate change on stream temperature: A model comparison and implications for coldwater fishes. *Journal of Hydrology*, 588(April). <https://doi.org/10.1016/j.jhydrol.2020.125066>
- Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., Gleason, C. J., & Wood, E. F. (2019). Global Reconstruction of Naturalized River Flows at 2.94 Million Reaches. *Water Resources Research*, 55(8), 6499–6516. <https://doi.org/10.1029/2019WR025287>
- MacDonald, R. J., Boon, S., & Byrne, J. M. (2014). A process-based stream temperature

- modelling approach for mountain regions. *Journal of Hydrology*, 511, 920–931.
<https://doi.org/10.1016/j.jhydrol.2014.02.009>
- Maheu, A., Poff, N. L., & St-Hilaire, A. (2016). A Classification of Stream Water Temperature Regimes in the Conterminous USA. *River Research and Applications*, 32(5), 896–906.
<https://doi.org/10.1002/rra.2906>
- McGuire, K. J., & McDonnell, J. J. (2010). Hydrological connectivity of hillslopes and streams: Characteristic time scales and nonlinearities. *Water Resources Research*, 46(10).
<https://doi.org/10.1029/2010WR009341>
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., & Rea, A. (2012). *NHDPlus Version 2: User Guide*.
- 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>
- Moore, R. D., Sutherland, P., Gomi, T., & Dhakal, A. (2005). Thermal regime of a headwater stream within a clear-cut, coastal British Columbia, Canada. *Hydrological Processes*, 19(13), 2591–2608. <https://doi.org/10.1002/hyp.5733>
- Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research Atmospheres*, 116(12). <https://doi.org/10.1029/2010JD015139>
- NOAA, (2016). *National Water Model: Improving NOAA's Water Prediction Services*.
<http://water.noaa.gov>

- NOAA, (2021). *Analysis of Record for Calibration: Version 1.1 Sources, Methods, and Verification*. <https://hydrology.nws.noaa.gov/aorc-historic/Documents/AORC-Version1.1-SourcesMethodsandVerifications.pdf>
- NOAA, (2021). OWP NWM NextGen Framework. Office of Water Prediction, <https://www.weather.gov/media/owp/oh/docs/2021-OWP-NWM-NextGen-Framework.pdf>
- NOAA, (2023). The National Water Model. Office of Water Prediction, <https://water.noaa.gov/about/nwm>.
- Rojas, R., Julien, P., Johnson, B. (1997). *CASC2D v1.0 Reference Manual*. Colorado State University. https://www.engr.colostate.edu/~pierre/ce_old/Projects/CASC2D-SED%20Web%20site%20082506/Download_files/CASC2D-SED-Reference-Manual.pdf
- 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. <https://doi.org/10.1016/j.scitotenv.2020.139679>
- PRISM Climate Group, Oregon State University, <https://prism.oregonstate.edu>, data created Dec 2022, accessed 4 Oct 2022.
- Salas, F. R., Somos-Valenzuela, M. A., Dugger, A., Maidment, D. R., Gochis, D. J., David, C. H., Yu, W., Ding, D., Clark, E. P., & Noman, N. (2018). Towards Real-Time Continental Scale Streamflow Simulation in Continuous and Discrete Space. *Journal of the American Water Resources Association*, 54(1), 7–27. <https://doi.org/10.1111/1752-1688.12586>
- Schmadel, N. M., Ward, A. S., & Wondzell, S. M. (2017). Hydrologic controls on hyporheic

- exchange in a headwater mountain stream. *Water Resources Research*, 53(7), 6260–6278.
<https://doi.org/10.1002/2017WR020576>
- Sun, N., Yearsley, J., Voisin, N., & Lettenmaier, D. P. (2015). A spatially distributed model for the assessment of land use impacts on stream temperature in small urban watersheds. *Hydrological Processes*, 29(10), 2331–2345. <https://doi.org/10.1002/hyp.10363>
- Tonina, D., & Buffington, J. M. (2009). Hyporheic Exchange in Mountain Rivers I: Mechanics and Environmental Effects. *Geography Compass*, 3(3), 1063–1086.
<https://doi.org/10.1111/j.1749-8198.2009.00226.x>
- van Beek, L. P. H., Eikelboom, T., van Vliet, M. T. H., & Bierkens, M. F. P. (2012). A physically based model of global freshwater surface temperature. *Water Resources Research*, 48(9). <https://doi.org/10.1029/2012WR011819>
- van Vliet, M. T. H., Franssen, W. H. P., Yearsley, J. R., Ludwig, F., Haddeland, I., Lettenmaier, D. P., & Kabat, P. (2013). Global river discharge and water temperature under climate change. *Global Environmental Change*, 23(2), 450–464.
<https://doi.org/10.1016/j.gloenvcha.2012.11.002>
- van Vliet, M. T. H., Yearsley, J. R., Franssen, W. H. P., Ludwig, F., Haddeland, I., Lettenmaier, D. P., & Kabat, P. (2012). Coupled daily streamflow and water temperature modelling in large river basins. *Hydrology and Earth System Sciences*, 16(11), 4303–4321.
<https://doi.org/10.5194/hess-16-4303-2012>
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. v, & Sorooshian, S. (2001). A framework for development and application of hydrological models A framework for development and application of hydrological models. In *Hydrology and Earth System Sciences* (Vol. 5, Issue 1).

- Wan, T., Covert, B. H., Kroll, C. N., & Ferguson, C. R. (2022). An Assessment of the National Water Model's Ability to Reproduce Drought Series in the Northeastern United States. *Journal of Hydrometeorology*, 23(12), 1929–1943. <https://doi.org/10.1175/JHM-D-21-0226.1>
- Wanders, N., van Vliet, M. T. H., Wada, Y., Bierkens, M. F. P., & van Beek, L. P. H. (Rens). (2019). High-Resolution Global Water Temperature Modeling. *Water Resources Research*, 55(4), 2760–2778. <https://doi.org/10.1029/2018WR023250>
- Ward, A. S., Fitzgerald, M., Gooseff, M. N., Voltz, T. J., Binley, A. M., & Singha, K. (2012). Hydrologic and geomorphic controls on hyporheic exchange during base flow recession in a headwater mountain stream. *Water Resources Research*, 48(4). <https://doi.org/10.1029/2011WR011461>
- Ward, A. S., Schmadel, N. M., & Wondzell, S. M. (2018a). Simulation of dynamic expansion, contraction, and connectivity in a mountain stream network. *Advances in Water Resources*, 114, 64–82. <https://doi.org/10.1016/j.advwatres.2018.01.018>
- Ward, A. S., Schmadel, N. M., & Wondzell, S. M. (2018b). Time-Variable Transit Time Distributions in the Hyporheic Zone of a Headwater Mountain Stream. *Water Resources Research*, 54(3), 2017–2036. <https://doi.org/10.1002/2017WR021502>
- Ward, A. S., Schmadel, N. M., Wondzell, S. M., Gooseff, M. N., & Singha, K. (2017). Dynamic hyporheic and riparian flow path geometry through base flow recession in two headwater mountain stream corridors. *Water Resources Research*, 53(5), 3988–4003. <https://doi.org/10.1002/2016WR019875>
- Ward, A. S., Schmadel, N. M., Wondzell, S. M., Harman, C., Gooseff, M. N., & Singha, K. (2016). Hydrogeomorphic controls on hyporheic and riparian transport in two headwater

- mountain streams during base flow recession. *Water Resources Research*, 52(2), 1479–1497. <https://doi.org/10.1002/2015WR018225>
- Ward, A. S., Wondzell, S. M., Schmadel, N. M., Herzog, S., Zarnetske, J. P., Baranov, V., Blaen, P. J., Brekenfeld, N., Chu, R., Derelle, R., Drummond, J., Fleckenstein, J. H., Garayburu-Caruso, V., Graham, E., Hannah, D., Harman, C. J., Hixson, J., Knapp, J. L. A., Krause, S., ... Wisnoski, N. I. (2019). Spatial and temporal variation in river corridor exchange across a 5th-order mountain stream network. *Hydrology and Earth System Sciences*, 23(12), 5199–5225. <https://doi.org/10.5194/hess-23-5199-2019>
- Ward, A. S., Zarnetske, J. P., Baranov, V., Blaen, P. J., Brekenfeld, N., Chu, R., Derelle, R., Drummond, J., Fleckenstein, J. H., Garayburu-Caruso, V., Graham, E., Hannah, D., Harman, C. J., Herzog, S., Hixson, J., Knapp, J. L. A., Krause, S., Kurz, M. J., Lewandowski, J., ... Wondzell, S. M. (2019). Co-located contemporaneous mapping of morphological, hydrological, chemical, and biological conditions in a 5th-order mountain stream network, Oregon, USA. *Earth System Science Data*, 11(4), 1567–1581. <https://doi.org/10.5194/essd-11-1567-2019>
- Wondzell, S. M. (2011). The role of the hyporheic zone across stream networks. *Hydrological Processes*, 25(22), 3525–3532. <https://doi.org/10.1002/hyp.8119>
- Wondzell, S. M., LaNier, J., & Haggerty, R. (2009). Evaluation of alternative groundwater flow models for simulating hyporheic exchange in a small mountain stream. *Journal of Hydrology*, 364(1–2), 142–151. <https://doi.org/10.1016/j.jhydrol.2008.10.011>
- Yan, H., Sun, N., Fullerton, A., & Baerwalde, M. (2021). Greater vulnerability of snowmelt-fed river thermal regimes to a warming climate. *Environmental Research Letters*, 16(5). <https://doi.org/10.1088/1748-9326/abf393>

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>

Yearsley, J. R. (2012). A grid-based approach for simulating stream temperature. *Water Resources Research*, 48(3). <https://doi.org/10.1029/2011WR011515>

Yearsley, J. R., Sun, N., Baptiste, M., & Nijssen, B. (2019). Assessing the impacts of hydrologic and land use alterations on water temperature in the Farmington River basin in Connecticut. *Hydrology and Earth System Sciences*, 23(11), 4491–4508.

<https://doi.org/10.5194/hess-23-4491-2019>

Appendix A: Supplemental Information for Chapter 2

Figures

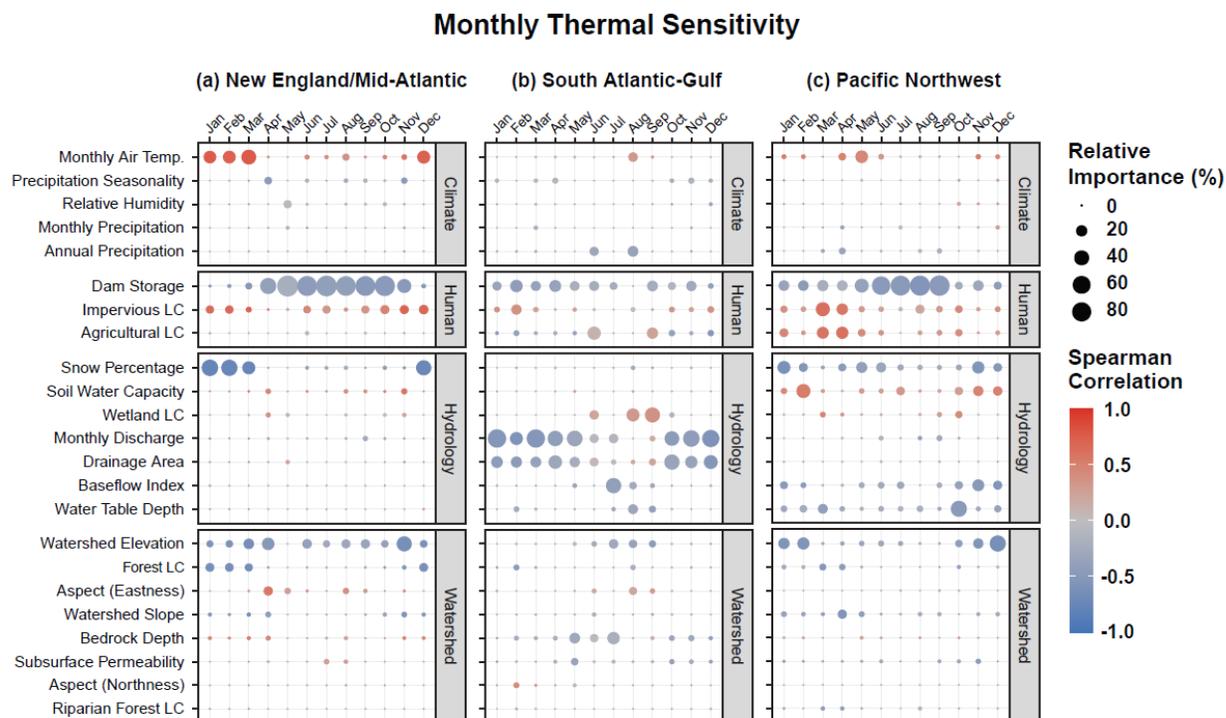


Figure 1. Aggregated heatmap of RF relative variable importance in predicting thermal sensitivity (TS: slope of linear relationship between daily air and water temperatures) for three hydrologic regions: (a) New England/Mid-Atlantic, (b) South Atlantic-Gulf, and (c) Pacific Northwest. The size of each point represents its magnitude of relative importance. The color of each point indicates the direction and strength of the relationship between a variable and the modeled metric as assessed by Spearman's rank correlations. Plot columns represent individual RF models for each month and display differences in relative importance between predictor variables. Plot rows track variability in the importance of a single predictor over months. Results were not presented in main text due to poor model error.

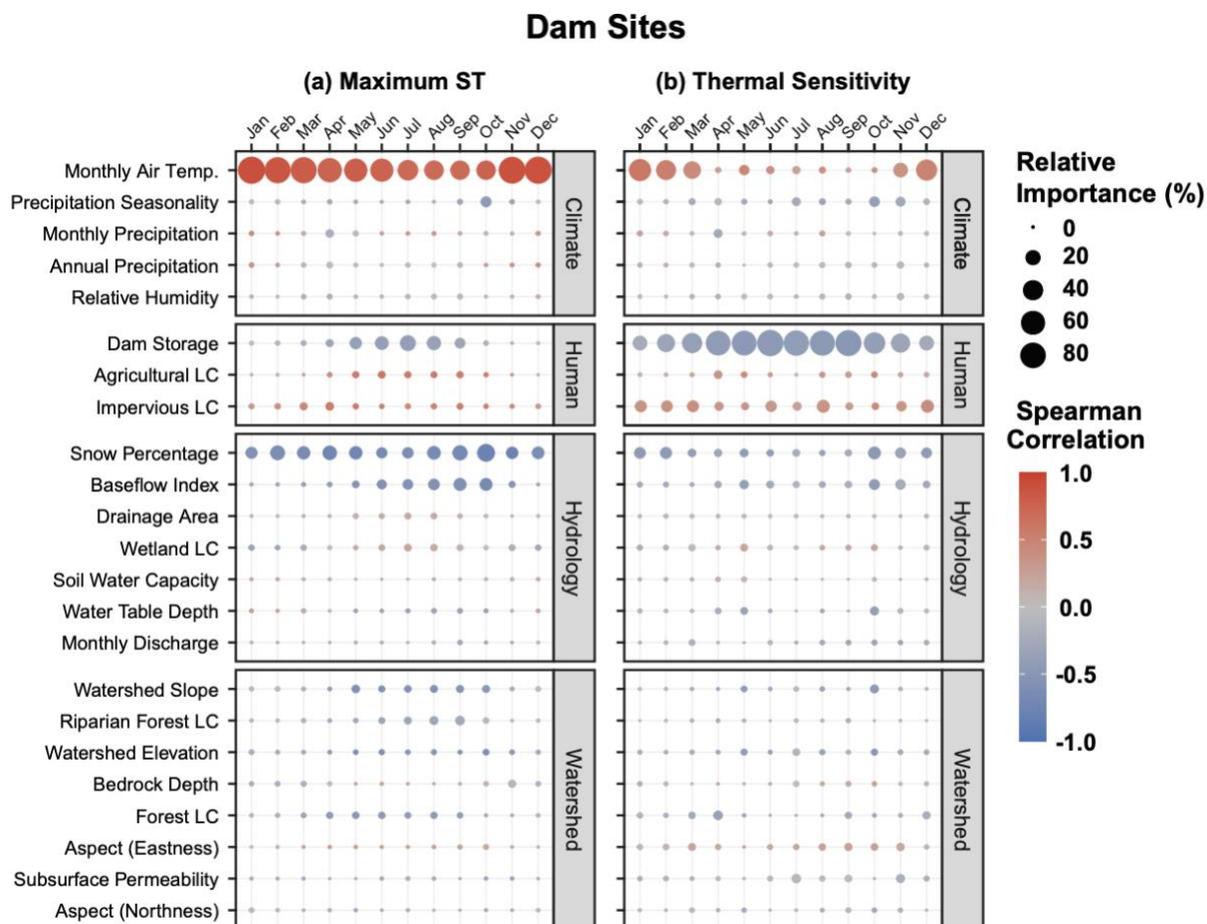


Figure 2. Aggregated heatmap of RF relative variable importance in predicting (a) maximum stream temperatures (ST_{max}) and (b) thermal sensitivity (TS) for sites with major dams in their watersheds. The size of each point represents its magnitude of relative importance. The color of each point indicates the direction and strength of the relationship between a variable and the modeled metric as assessed by Spearman's rank correlations. Plot columns represent individual RF models for each month and display differences in relative importance between predictor variables. Plot rows track variability in the importance of a single predictor over months.

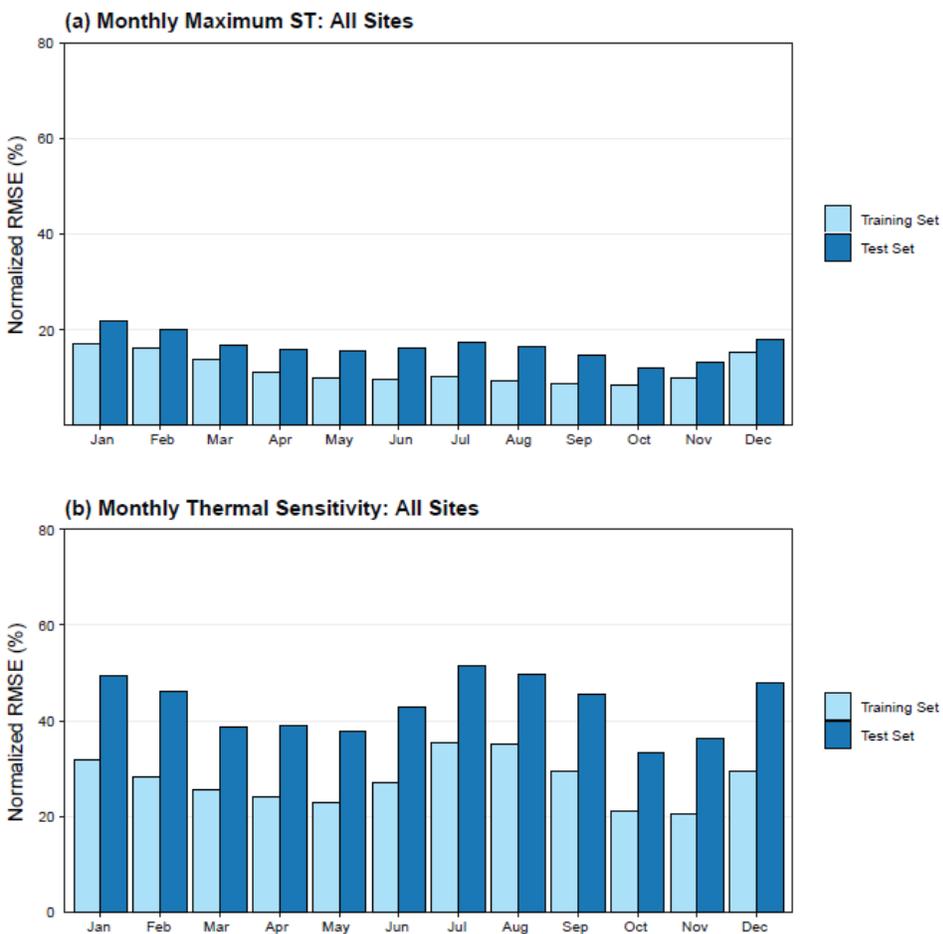


Figure 3. Comparisons of monthly training set (80% of data) and test set (20% of data) error in normalized RMSE for the prediction of (a) maximum stream temperatures (ST_{max}) and (b) thermal sensitivity (TS) for models of all sites.

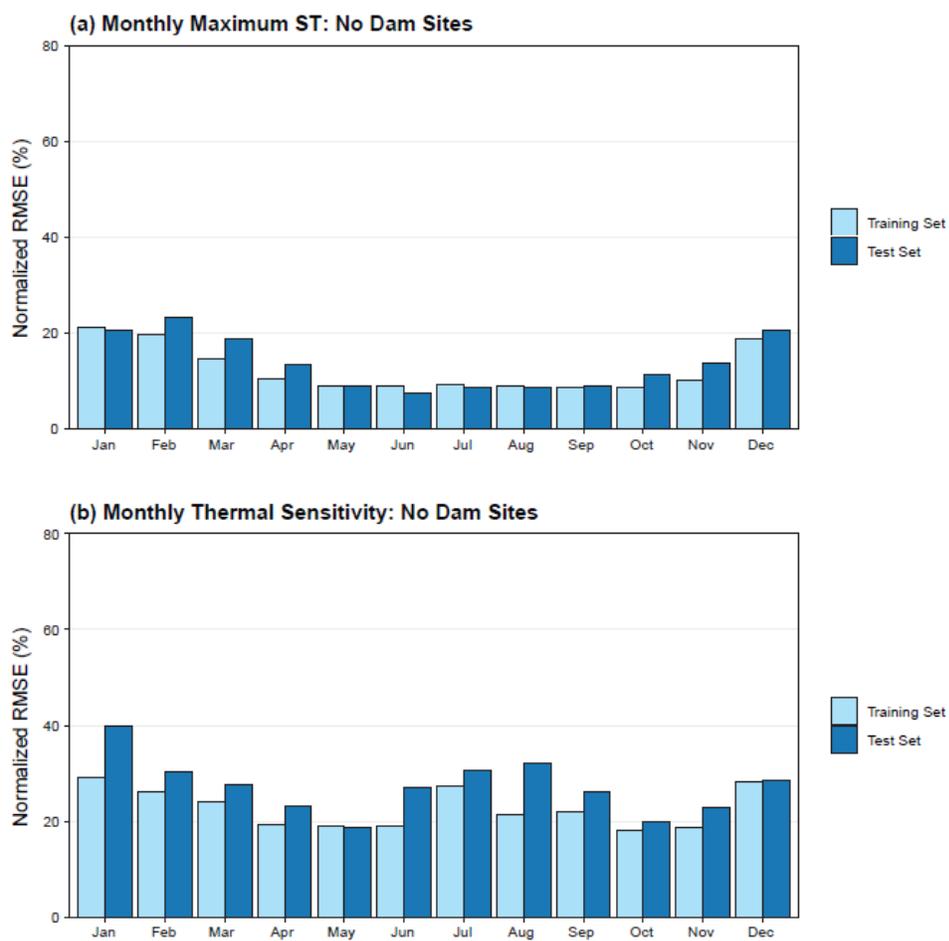


Figure 4. Comparisons of monthly training set (80% of data) and test set (20% of data) error in normalized RMSE for the prediction of (a) maximum stream temperatures (ST_{max}) and (b) thermal sensitivity (TS) for models of sites with no major dams in their watersheds.

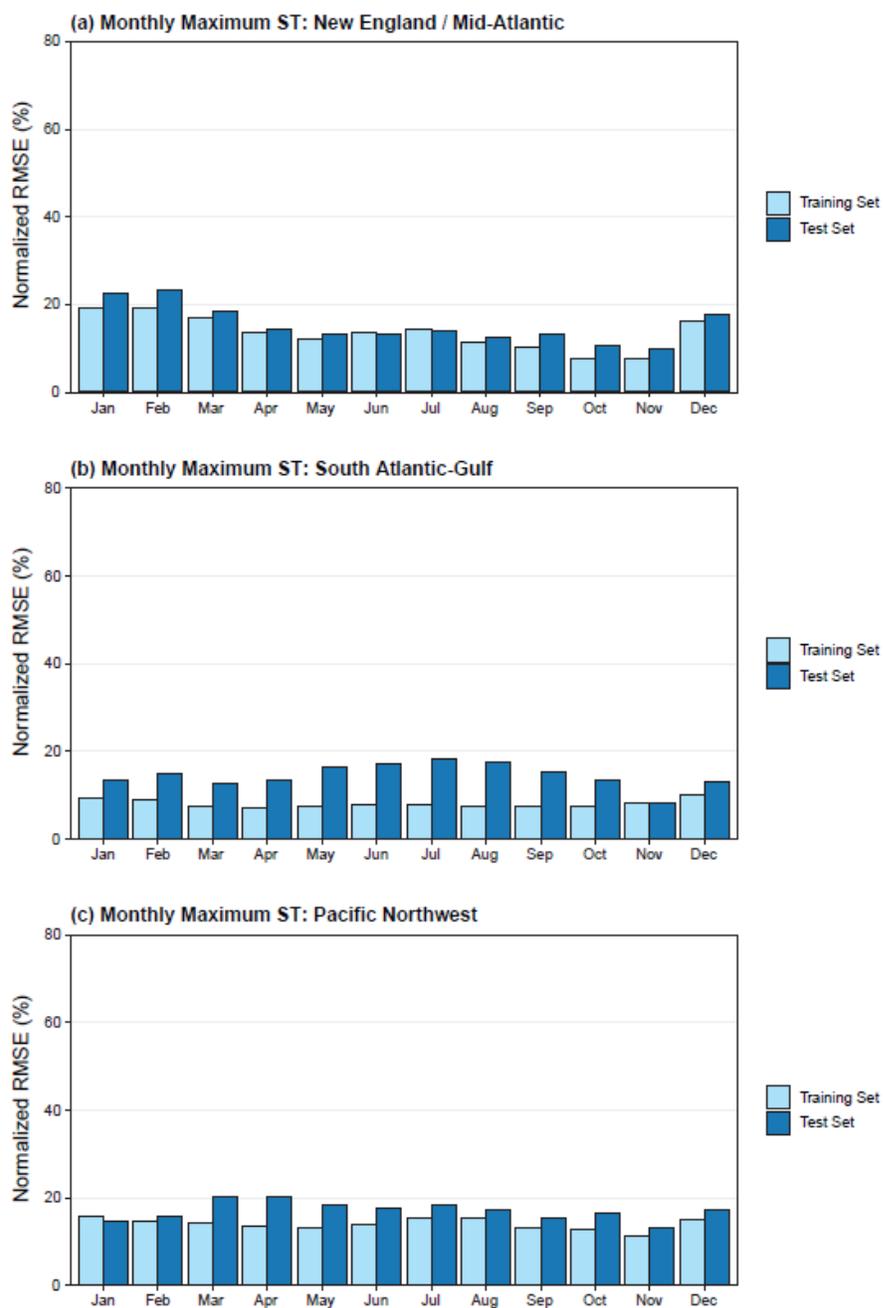


Figure 5. Comparisons of monthly training set (80% of data) and test set (20% of data) error in normalized RMSE for the prediction of maximum stream temperatures (ST_{max}) for sites in the following hydrologic regions: (a) New England/Mid-Atlantic, (b) South Atlantic-Gulf, and (c) Pacific Northwest.

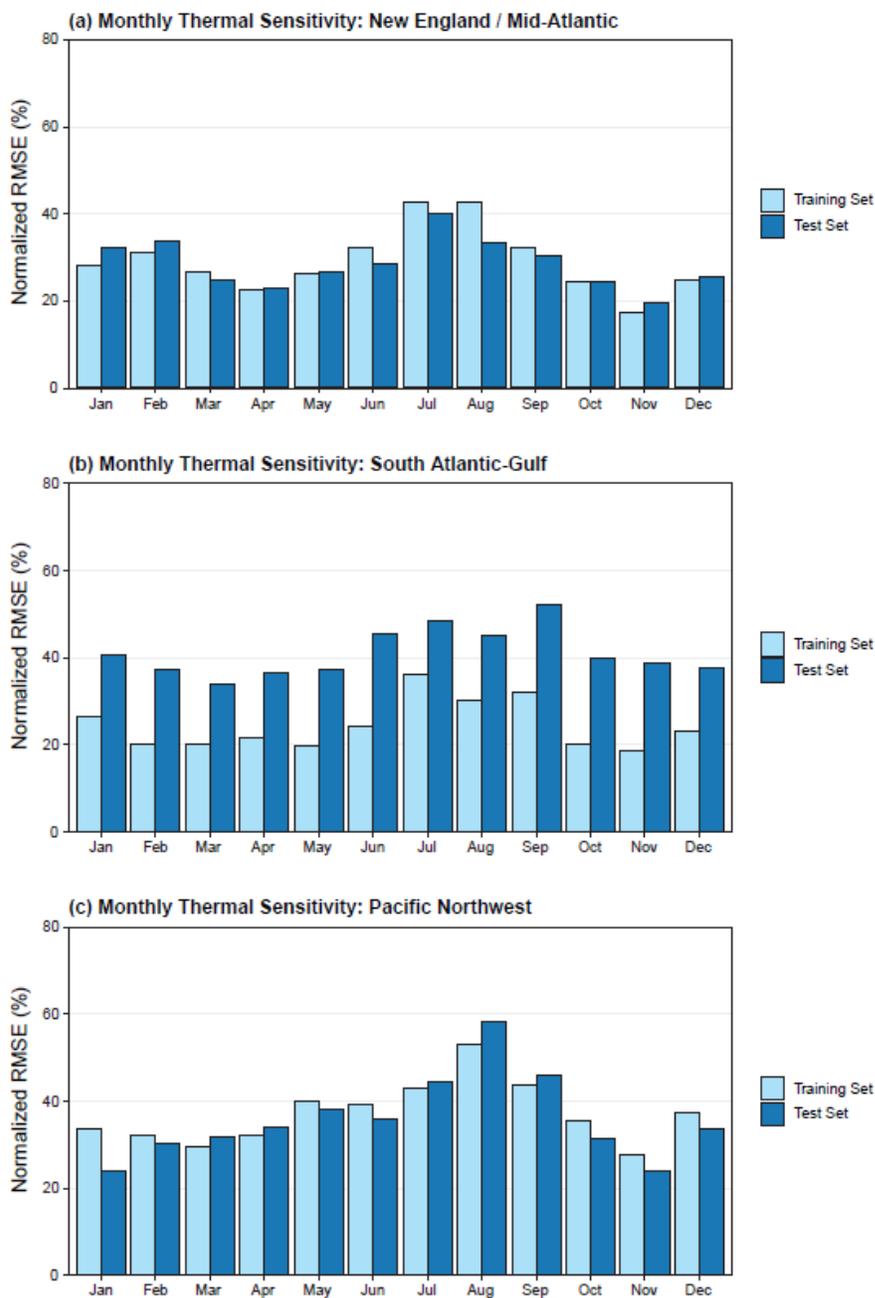


Figure 6. Comparisons of monthly training set (80% of data) and test set (20% of data) error in normalized RMSE for the prediction of thermal sensitivity (*TS*) for sites in the following hydrologic regions: (a) New England/Mid-Atlantic, (b) South Atlantic-Gulf, and (c) Pacific Northwest.

Appendix B: Supplemental Information for Chapter 3

1. Water Temperature Modeling Equations

1.1. Channel Geometry

The cross-sectional area of water in the channel (A , m^2) was calculated by:

$$A = \frac{Q}{v} \quad (1)$$

where:

Q = discharge, $\text{m}^3 \text{s}^{-1}$

V = channel velocity, m s^{-1}

Using the calculated area of water in the channel, the width of the water's surface was derived from geometric relationships by:

$$w = \sqrt{\frac{4A}{s} + BW^2} \quad (2)$$

where:

w = water surface channel width, m

A = cross-sectional area of water in channel, m^2

s = slope of main channel edge, unitless

BW = bottom width of main channel, m

The depth of water in the channel was then calculated by:

$$d_w = 0.5s(w - BW) \quad (3)$$

where:

d_w = depth of water, m

s = slope of main channel edge, unitless

w = water surface channel width, m

BW = bottom width of main channel, m

The volume of water within the main channel of a model node was calculated by:

$$V = l * A \quad (4)$$

where:

V = model reach volume, m³

l = length of model reach, m

A = cross-sectional area of water in channel, m²

1.2. Heat Transfer Equations

Incoming shortwave radiation energy flux to the water column was derived from the surface downward short-wave radiation flux (SWDOWN) in the NWM 1-km grid forcing data. Incoming shortwave fluxes were adjusted based on a constant water surface albedo and an approximated channel riparian shading. Incoming shortwave radiation was calculated by (Glose et al., 2017; Magnusson et al., 2012):

$$H_{SW} = (1 - \alpha)(1 - R)SW_{IN} \quad (5)$$

where:

H_{SW} = shortwave radiation flux, W m⁻²

α = reflectance of water surface, assumed to equal 0.04, unitless (Boyd, 1996; Magnusson et al., 2012)

R = riparian shading, ranging from 0 (no shading) to 1 (completely shaded), unitless

SW_{IN} = incoming shortwave radiation, W m⁻²

Net longwave radiation energy flux to the water column was equal to the sum of downward atmospheric longwave radiation, landcover radiation, and emitted blackbody radiation from the water column proportional to water temperature. Net longwave heat flux was calculated by (Westhoff et al., 2007):

$$H_{LW} = LW_{ATMOS} + LW_{BACK} + LW_{LANDCOVER} \quad (6)$$

where:

LW_{ATMOS} = atmospheric longwave radiation flux, $W m^{-2}$

LW_{BACK} = back longwave radiation flux, $W m^{-2}$

$LW_{LANDCOVER}$ = landcover longwave radiation flux, $W m^{-2}$

The atmospheric longwave radiation flux to the water column was derived using the surface downward long-wave radiation flux (LWDOWN) from the NWM 1-km grid forcing data and adjusted using the reflectance of water and riparian shading by:

$$LW_{ATMOS} = (1 - \alpha)(1 - r)LW_{IN} \quad (7)$$

where:

α = reflectance of water surface, assumed to equal 0.04 (Boyd, 1996; Magnusson et al., 2012)

R = riparian shading, ranging from 0 (no shading) to 1 (completely shaded), unitless

LW_{IN} = incoming atmospheric longwave radiation, $W m^{-2}$

Back longwave radiation is emitted from the stream's surface in proportion to simulated water temperature and was calculated by (Boyd & Kasper, 2003; Westhoff et al., 2007):

$$LW_{BACK} = \varepsilon\sigma_{sb}(T_w + 273.2)^4 \quad (8)$$

where:

ε = emissivity of water, assumed to equal 0.96 (Boyd & Kasper, 2003)

σ_{sb} = Stefan-Boltzman constant, $\text{W m}^{-2} \text{ } ^\circ\text{C}^{-4}$

T_w = water temperature, $^\circ\text{C}$

Landcover longwave radiation from surrounding vegetation is emitted in proportion to air temperature and riparian shading and is calculated by (Boyd & Kasper, 2003; Westhoff et al., 2007):

$$LW_{LANDCOVER} = (1 - \alpha)\varepsilon R\sigma_{sb}(T_{air} + 273.2)^4 \quad (9)$$

where:

α = reflectance of water surface, assumed to equal 0.04 (Boyd, 1996; Magnusson et al., 2012)

ε = emissivity of water, assumed to equal 0.96 (Boyd & Kasper, 2003)

R = riparian shading, ranging from 0 (no shading) to 1 (completely shaded), unitless

σ_{sb} = Stefan-Boltzman constant, $\text{W m}^{-2} \text{ } ^\circ\text{C}^{-4}$

T_{air} = air temperature, $^\circ\text{C}$

Evaporation from the water's surface and the associated latent heat transfer were calculated using the Penman combined method, which is often applied in stream temperature models to quantify evaporation rates from open water surfaces (Boyd, 1996; Glose et al., 2017; Westhoff et al., 2007). Evaporation rate was derived from several NWM forcing inputs, including air temperature, wind speed (magnitude combined from U and V components), specific humidity, and air pressure, by the following (Maidment, 1993; Westhoff et al., 2007):

$$E = \frac{s\Phi}{\rho_w L_e (s + \gamma)} + \frac{c_{air} \rho_{air} (e_s - e_a)}{\rho_w L_e r_a (s + \gamma)} \quad (10)$$

where:

E = evaporation rate, m s^{-1}

s = slope of the saturated vapor pressure curve, $\text{kPa } ^\circ\text{C}^{-1}$

Φ = sum of shortwave (H_{SW}) and longwave (H_{LW}) radiation, W m^{-2}

ρ_w = density of water, kg m^{-3}

L_e = latent heat of vaporization, J kg^{-1}

γ = psychometric constant, $\text{J kg}^{-1} ^\circ\text{C}^{-1}$

c_{air} = specific heat of air, $\text{J kg}^{-1} ^\circ\text{C}^{-1}$

ρ_{air} = density of air, kg m^{-3}

e_s = saturated vapor pressure, kPa

e_a = actual vapor pressure, kPa

r_a = aerodynamic resistance, s m^{-1}

The saturated vapor pressure of the air is dependent on air temperature and was calculated by (Dingman, 1994; Westhoff et al., 2007):

$$e_s = 0.611 e^{\frac{17.27 T_{air}}{237.3 + T_{air}}} \quad (11)$$

where:

e_s = saturated vapor pressure, kPa

T_{air} = air temperature, $^\circ\text{C}$

The actual vapor pressure of the air was adjusted from the saturated vapor pressure using specific humidity values by (Bolton, 1980; Dingman, 1994):

$$e_a = \frac{qP_a}{0.622 + 0.378q} \quad (12)$$

where:

e_a = actual vapor pressure, kPa

q = specific humidity, kg kg⁻¹

P_a = surface air pressure, kPa

Aerodynamic resistance was calculated from U- and V- components of wind speed by (Glose et al., 2017; Westhoff et al., 2007):

$$r_a = \frac{245}{0.54\sqrt{v_{U-wind}^2 + v_{V-wind}^2} + 0.5} \quad (13)$$

where:

r_a = aerodynamic resistance, s m⁻¹

v_{U-wind} = U-component of wind speed, m s⁻¹

v_{V-wind} = V-component of wind speed, m s⁻¹

The slope of the saturated vapor pressure curve was calculated from saturated vapor pressure and air pressure by (Westhoff et al., 2007):

$$s = \frac{4100e_s}{(237 + T_{air})^2} \quad (14)$$

where:

s = slope of the saturated vapor pressure curve, kPa °C⁻¹

T_{air} = air temperature, °C

The latent heat of vaporization was calculated from water temperature by (Maidment, 1993; Glose et al., 2017):

$$L_e = 10^6(2.501 - 0.002361T_w) \quad (15)$$

where:

L_e = latent heat of vaporization, J kg⁻¹

T_w = water temperature, °C

Following the calculation of evaporation rate at a given time step and model reach, latent heat transfer was then derived by (Westhoff et al., 2007):

$$H_{LH} = -\rho_w L_e E \quad (16)$$

where:

H_{LH} = latent heat exchange flux, W m⁻²

ρ_w = density of water, kg m⁻³

L_e = latent heat of vaporization, J kg⁻¹

E = evaporation rate, m s⁻¹

Note that the latent heat flux is negative when evaporation is occurring and positive when condensation is occurring.

Sensible heat exchange is the transfer of heat between the water's surface and the atmosphere due to temperature contrasts between the two mediums (Boyd & Kasper, 2003; Glose et al., 2017; Westhoff et al., 2007). Sensible heat was calculated by scaling latent heat fluxes using the Bowen Ratio by (Boyd & Kasper, 2003; Westhoff et al., 2007):

$$\Phi_{sensible} = B_r \Phi_{latent} \quad (17)$$

where:

$\Phi_{sensible}$ = sensible heat flux, W m⁻²

Φ_{latent} = latent heat flux, W m⁻²

B_r = Bowen Ratio, unitless

The Bowen Ratio was derived from water temperature, air temperature, air pressure, saturated vapor pressure at the air-water interface, and actual vapor pressure at the air-water interface by (Boyd & Kasper, 2003; Westhoff et al., 2007):

$$B_r = 6.1e^{-4} P_a \frac{T_w - T_{air}}{e_s^w - e_a^w} \quad (18)$$

where:

P_a = air pressure, kPa

T_w = water temperature, °C

T_{air} = air temperature, °C

e_s^w = saturated vapor at the air-water interface, kPa

e_a^w = actual vapor at the air-water interface, kPa

As Bowen's ratio is scaled by a relative humidity-dependent term, high relative humidity values approaching 100% can lead to implausibly large sensible heat fluxes. To prevent anomalous Bowen's ratio values and sensible heat fluxes in the model, we set all relative humidity values exceeding 97% to equal exactly 97%. This introduces a small error on estimates of sensible heat and latent heat fluxes during the most humid periods. As opposed to previous calculations of saturated and actual vapor pressure that are based on air temperature, e_s^w and e_a^w were instead calculated using water temperatures. The saturated vapor pressure at the air-water interface was calculated by (Boyd & Kasper, 2003; Westhoff et al., 2007):

$$e_s^w = 0.611e^{\frac{17.27T_w}{237.3+T_w}} \quad (19)$$

where:

e_s^w = saturated vapor pressure at the air-water interface, kPa

T_w = water temperature, °C

Similarly, the actual vapor pressure at the air-water interface was derived from the saturated vapor pressure at the air water interface, scaled by relative humidity (Boyd & Kasper, 2003; Westhoff et al., 2007):

$$e_a^w = \frac{H}{100\%} e_s^w \quad (20)$$

where:

e_a^w = actual vapor pressure at the air-water interface, kPa

H = relative humidity, expressed as a percentage, %

e_s^w = saturated vapor pressure at the air-water interface, kPa.

References

- Bolton, D. (1980). The Computation of Equivalent Potential Temperature. *Monthly Weather Review*, 108, 1046–1053.
- Boyd, M. S. (1996). *Heat Source: Stream, River and Open Channel Temperature Prediction*.
Title: Heat Source: Stream.
- Boyd, M. S., & Kasper, B. (2003). *Analytical Methods for Dynamic Open Channel Heat and Mass Transfer Methodology for the Heat Source Model Version 7.0*.
<http://www.deq.state.or.us/wq/TMDLs/tools.htm>
- Glose, A., Lautz, L. K., & Baker, E. A. (2017). Stream heat budget modeling with HFLUX: Model development, evaluation, and applications across contrasting sites and seasons. *Environmental Modelling and Software*, 92, 213–228.
<https://doi.org/10.1016/j.envsoft.2017.02.021>
- Magnusson, J., Jonas, T., & Kirchner, J. W. (2012). Temperature dynamics of a proglacial stream: Identifying dominant energy balance components and inferring spatially integrated hydraulic geometry. *Water Resources Research*, 48(6).
<https://doi.org/10.1029/2011WR011378>
- Westhoff, M. C., Savenije, H. H. G., Luxemburg, W. M. J., Stelling, G. S., van de Giesen, N. C., Selker, J. S., Pfister, L., & Uhlenbrook, S. (2007). A distributed stream temperature model using high resolution temperature observations. *Hydrol. Earth Syst. Sci.*, 11, 1469–1480. <https://doi.org/10.5194/hessd-4-125-2007>

Curriculum Vita

EDUCATION

Ph.D., Earth Sciences *Anticipated May 2023*

Syracuse University, Syracuse, NY

Dissertation: Stream Temperature as a Tracer of Interactions Amongst Hydrological Processes, Atmospheric Exchange, and Human Activity

B.S., Geology and Geophysics, Honors in the Liberal Arts 2019
University of Wisconsin-Madison, Madison, WI

PUBLICATIONS

Wade, J., Kelleher, C., Ogden, F. (In preparation), Incorporating physically-based river water temperature predictions into the National Water Model framework.

Wade, J., Kelleher, C., Hannah, D.M. (In review), Machine learning unravels controls on river water temperature regime dynamics, *Journal of Hydrology*.

Ward, A.S., **Wade, J., Kelleher, C., Schewe, R.L.** (2023), Clarify Jurisdiction of US Clean Water Act. *Science*, 379(6628), 148-148. DOI: <https://doi.org/10.1126/science.adf7391>

Wade, J., Kelleher, C., Ward, A. S., Schewe, R. L. (2022), The fluid definition of the ‘waters of the United States’: Non-uniform effects of regulation on US wetland protections. *Hydrological Processes*, 36(11), e14747. DOI: [10.1002/hyp.1474](https://doi.org/10.1002/hyp.1474)

Wade, J., Lautz, L., Kelleher, C., Vidon, P., Davis, J., Beltran, J., Pearce, C. (2020), Beaver dam analogues drive heterogeneous groundwater–surface water interactions. *Hydrological Processes*, 34, 5340–5353. DOI: [10.1002/hyp.13947](https://doi.org/10.1002/hyp.13947)

Orio, M., Luna, G. J., Kotulla, R., Gallager, J. S., Zampieri, L., **Wade, J.** . . . Zemko, P., (2017), CXO J004318.8 412016, a steady supersoft X-ray source in M 31. *Monthly Notices of the Royal Astronomical Society*, 470(2), 2212-2224. DOI: [10.1093/mnras/stx1355](https://doi.org/10.1093/mnras/stx1355)

COMMENTARIES

Wade, J., Kelleher, C., Ward, A. S., Schewe, R. L. (2023). Which wetlands are vulnerable? Mapping recent shifts in protection under US federal wetland law. Global Water Forum, UNESCO Chair in Water Economics and Transboundary Water Governance. <https://globalwaterforum.org/2023/02/02/which-wetlands-are-vulnerable-mapping-recent-shifts-in-protection-under-us-federal-law/>

RESEARCH EXPERIENCE

Ph.D. Dissertation Research, Syracuse University
Department of Earth and Environmental Sciences
Advisors: Dr. Christa Kelleher, Dr. Laura Lautz
Hyporheic Exchange & Stream Restoration

May 2019 – May 2023

- Collaborated with key stakeholders to design and implement a multi-year monitoring campaign to assess the hyporheic effects of stream restoration structures, informing land management practices in the Western US
- Applied vertical heat tracing models, hydraulic head measurements, and biogeochemical analyses to quantify rates of groundwater-surface water interaction

Stream Temperature Machine Learning

- Trained 144 machine learning models in R to predict metrics of stream temperature variability at 400+ US streams
- Interpreted variable importance metrics to identify leading controls on stream temperatures, providing insight into effective approaches for regulating river thermal regimes in a changing climate
- Catalogued and published 250+ files containing code, scientific data, and research workflows to GitHub

Wetland Policy Analysis

- Translated redefinitions of federal wetland policy into quantitative impact assessments at 253,000+ wetlands in New York State using a computationally-intensive ArcGIS Pro workflow (*published in Hydrologic Processes, commentaries accepted in Science and Global Water Forum*)
- Emphasized potential threats to geographically-isolated wetlands in policy-focused articles published in *Science* and *Global Water Forum*

Stream Temperature Model Development

- Independently led the conceptualization and development of a physically-based river temperature model within the framework of the National Water Model
- Developed coupled modeling framework that achieving hourly temperature predictions with errors under 1°C RMSE.

Research Assistant, University of Wisconsin-Madison

Apr. 2018 – May 2019

Department of Geoscience

- Characterized the behavior of quartz fluid inclusions using laboratory thermobarometry
- Organized geologic samples and maintained detailed records of laboratory experiments and results
- Interpreted sample temperature and pressure results in relation to modern seismic activity, leading to a presentation at the Geological Society of America 2018 Fall meeting

Research Assistant, University of Wisconsin-Madison

Nov. 2015 – May 2016

Department of Astronomy

- Performed data acquisition to examine supersoft X-rays emissions from a binary system in the Andromeda Galaxy
- Developed a Linux-based workflow to identify and extract spectra from XMM-Newton telescope observations

PROFESSIONAL EXPERIENCE

William M. Lapenta Intern, Office of Water Prediction, NOAA

Jun. 2022 – Aug. 2022

- Developed advective heat transport models to assess how water temperature predictions could best be incorporated into the National Water Model

- Developed and version-controlled a reproducible modeling workflow consisting of ~20,000 lines of Python code
- Authored a 25-page preliminary report detailing the methodology and scientific implications of modeling results, which were presented at the 2022 William M. Lapenta Professional Conference and the 2022 AGU Fall Meeting

Hydrologist, New England Water Science Center, USGS

May 2019 – Aug. 2019

NAGT-USGS Cooperative Summer Fellowship

- Budgeted, planned and directed a water sampling field campaign to assess urban nitrate and PFAS flux to coastal embayments in Cape Cod, MA
- Processed groundwater chemistry data and produced scientific data visualizations for multiple USGS Scientific Investigations Reports
- Visualized spatial-temporal trends in urban water table levels using ArcGIS and Adobe Illustrator

TEACHING AND MENTORING

Graduate Teaching Assistant, Syracuse University

Fall 2020, Spring 2022

EAR 205: Water and Our Environment

- Taught discussion sections for undergraduate-level hydrology and atmospheric science course (100 students)

Research Mentor, Syracuse University

Summer 2021

- Mentored a visiting undergraduate researcher in computational hydrology, leading to a poster presentation at the 2021 AGU Fall Meeting
- Advised student in research methods, data organization, coding, and hydrological principles

AWARDS AND HONORS

Research Excellence Doctoral Funding Fellowship, Syracuse University, \$23,590	2022
Chair's Award, Syracuse University, \$400	2022
NSF Graduate Research Fellowships Program (GRFP) Honorable Mention	2021
Vincent E. McKelvey Scholarship Award, Syracuse University, \$350	2021
NSF Grant, American Meteorological Society Science Policy Colloquium, \$5,600	2020
NSF NRT Fellowship, Energy Model Program on Water-Energy, \$32,00	2019-2020
Mack C. Lake Distinguished Geoscience Undergraduate, UW-Madison	2019
NAGT-USGS Cooperative Field Training Program Nominee	2019
Dexter Environmental Scholarship, UW-Madison, \$1,000	2018
William F. Vilas Scholarship, UW-Madison, \$400	2018
Wasatch-Uinta Field Camp Scholarship, UW-Madison, \$1,500	2018
L.R. Ingersoll Prize for Excellence in Physics, UW-Madison	2017

Schoenleber Scholarship, UW-Madison, \$32,000	2015-2019
Academic Excellence Scholarship, UW-Madison, \$10,000	2015-2019
Letters and Science Honors College, UW-Madison	2015-2019

PRESENTATIONS

Wade, J. (2023). Water Temperatures and the National Water Model: Insights from a Physically-based Modeling Approach, NOAA Physical Science Laboratory, Remote.

Wade, J., Ogden, F., Kelleher, C. (2022). Incorporating Physics-Based Temperature Predictions into the National Water Model Framework, H45I-1484, American Geophysical Union 2022 Fall Meeting, Chicago, IL.

Wade, J. (2022). Can the National Water Model be used to predict river water temperatures?: A continental-scale perspective, Waggoner Geology Graduate Seminar, Syracuse University.

Wade, J., Ogden, F. (2022). Incorporating Physics-based Water Temperature Predictions into the National Water Model, 2022 William M. Lapenta Professional Conference.

Wade, J., Kelleher, C., Hannah D.M. (2022). Untangling nested spatiotemporal controls on stream thermal regimes using machine learning methods, 2022 Central New York Earth Science Student Symposium.

Wade, J., Kelleher, C., Hannah D.M. (2021). Exploring continental-scale relationships between stream temperature signatures and watershed characteristics, H25X-1295, American Geophysical Union 2021 Fall Meeting, New Orleans, LA.

Ndlovu, W., **Wade, J.**, Kelleher, C., Gannon, J., Zimmer, M. (2021). Examining streamflow flashiness trends across the northeastern USA, H45C-1195, American Geophysical Union 2021 Fall Meeting, New Orleans, LA.

Hurst, C., Kelleher, C., Shaw, S., Jones, H., Davis, J., **Wade, J.**, Cero, A.D., Vidon, P. (2021) Assessing Changes in Channel Morphology Three Years Following Beaver Dam Analogue Installation, EP25B-1310, American Geophysical Union 2021 Fall Meeting, New Orleans, LA.

Jones, H., Kelleher, C., Shaw, S., Hurst, E., Larson, C., Coffman, J., Cero, A.D., **Wade, J.** (2021). Effects of Beaver Dam Analogues on Late-summer In-stream Water Level Variability, EP25B-1311, American Geophysical Union 2021 Fall Meeting, New Orleans, LA.

Wade, J., Kelleher, C. (2021). Continental-scale relationships between watersheds characteristics and stream temperature signatures: An ensemble learning approach, Waggoner Geology Graduate Seminar, Syracuse University.

Wade, J., Kelleher, C. (2020). Wetlands and WOTUS: Regulating New York's Vulnerable Waters, Waggoner Geology Graduate Seminar, Syracuse University.

Kelleher, C., Vidon, P., **Wade, J.** (2020). A multidisciplinary approach to assess the impact of beaver dam analogues on stream and floodplain hydrology and geomorphology. Red Canyon Ranch Scientific Brief to The Nature Conservancy.

Jones, A., Goodwin, L., Brown, P., **Wade, J.** (2018). Do small faults in the Baraboo Syncline preserve evidence of episodic tremor and slip (ETS)?, Geological Society of America 2018 Fall Meeting, Indianapolis, IN.

SERVICE

Faculty Search Student Representative, Syracuse University *Jan. 2022 – Mar. 2022*

- Represented the priorities of the graduate student body on the faculty search committee for the Department of Earth and Environmental Sciences
- Conducted 15 detailed interviews with applicants and evaluated candidate resumes, cover letters, and publications

Manuscript Reviewer, JAWRA *Sept. 2021*

- Provided peer review feedback for manuscripts submitted to the Journal of the American Water Resources Association

Unlearning Racism in Geoscience, Syracuse University *Jan. 2021 – May 2021*

- Participated in a geoscience community-wide program designed to oppose inequitable policies in academic institutions
- Drafted and refined inclusive policies that were presented to academic department leaders

Graduate Student Officer, Syracuse University *Aug. 2020 – May 2021*

- Served as the Treasurer Secretary for an organization representing the interest of Syracuse University Earth and Environmental Science graduate students
- Audited past organization financial records, developed an annual budget, and disbursed funds for organization events
- Prepared leadership meeting agendas and maintained public records of meeting minutes

PROFESSIONAL TRAINING

American Meteorological Society Summer Policy Colloquium *Jul. 2020*

- Applied for and received an NSF grant to fully fund attendance to the Summer Policy Colloquium
- Attended 4 weeks of discussions with members of Congress, Congressional staff, and industry leaders on topics related to federal science and environmental policy
- Received instruction on the federal R&D budget and the respective roles of the legislative and executive branches in the science-policy process

Education Model Program on Water Energy Research Fellow *Aug. 2019 – Dec. 2021*

- Engaged in a selective interdisciplinary graduate program that provided training in policy, communication, management, and law at the water-energy nexus
- Contributed to weekly student seminars that emphasized scientific and professional collaboration across research disciplines
- Completed a 2-week summer field course and a capstone research project on water quality concerns in Western US lakes

Alan Alda Center for Communicating Science Workshop *Jan. 2020*

- Expanded science communication through instruction in improvisation and message-development
- Applied storytelling skills to communicate academic research to broader audiences

Kathy Lambert Science Communication Workshop *Oct. 2019*

- Developed plain language translations of academic research and received training on effectively communicating science to public media

Wasatch-Uinta Summer Field Camp*Jun. 2018 – Jul. 2018*

- Completed a 6-week geology field course consisting of geologic mapping, subsurface structural inference, and rock classification
- Received a nomination to the NAGT-USGS Cooperative Field Training Program as a top student in the program

TECHNICAL SKILLS

Software: R, MATLAB, Python, Geospatial Python, ArcGIS, Linux, GitHub, Amazon Web Services, Numerical Geoscience Modeling, Microsoft Office Suite, Microsoft Access, Adobe Illustrator

Field Skills: Experience with designing field monitoring strategies, well measurements, mini-piezometer installation, discharge measurements using an acoustic Doppler velocimeter, topographic surveying, water sample collection, data logger programming