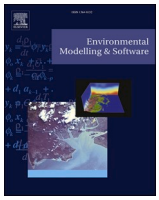
Environmental Modelling and Software 171 (2024) 105866

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Incorporating physically-based water temperature predictions into the National water model framework

Jeffrey Wade [a,*], Christa Kelleher [b], Barret L. Kurylyk [c]

[a] Department of Earth and Environmental Sciences, Syracuse University, Syracuse, NY, USA
[b] Department of Civil and Environmental Engineering, Lafayette College, Easton, PA, USA
[c] Department of Civil and Resource Engineering, Dalhousie University, Halifax, Nova Scotia, Canada

ARTICLE INFO

Handling Editor: Daniel P Ames

Keywords:
Stream temperature
Water temperature
Mechanistic model
National water model
Energy balance

ABSTRACT

While river water temperatures are a strong control on instream processes and aquatic ecosystems, monitoring networks for river water temperatures are often sparse. 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 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 during a six-week summer period. Our modeling framework, representing a first effort at pairing water temperature simulation with predictions from the NWM, suggests 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 and 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 and 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 and 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 temperate 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

* Corresponding author.
E-mail address: jswade@syr.edu (J. Wade).

https://doi.org/10.1016/j.envsoft.2023.105866
Received 2 May 2023; Received in revised form 1 September 2023; Accepted 26 October 2023
Available online 4 November 2023

'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 Model 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 several weeks of summer baseflow conditions to determine if the NWM framework is suitable for temperature prediction. Though we only considered temperature modeling in a single basin, our intention is that the strategies we have developed for the modeling framework are 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 during summer, when water temperatures are of greatest concern for aquatic species, 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 km^2 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., 2019a, 2019b; Wondzell et al., 2009), river corridor connectivity (McGuire and 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 (Fig. 1). While considered a fifth-order catchment by most field studies (Ward et al., 2019a,b), 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 m above sea level, and is primarily forested by Douglas fir trees (Ward et al., 2019a,b). 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 and Jones, 2015). Flows at the basin outlet (Lookout Creek) typically reach a maximum in December or January, and a minimum in September (Jennings and 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 and 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 (Fig. 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

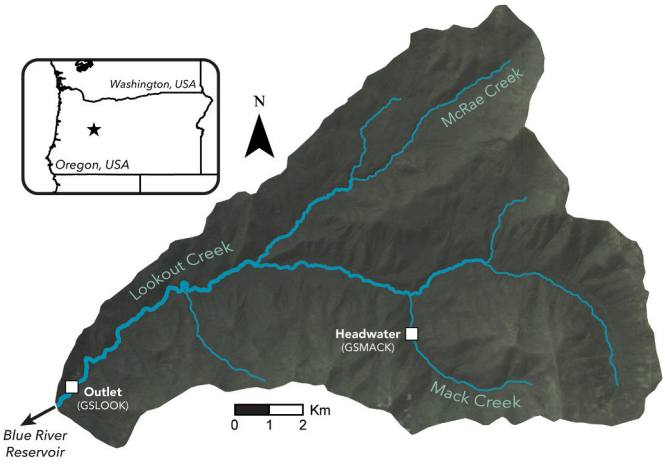


Fig. 1. Location of water temperature gages ('Headwater': GSMACK, 'Outlet': GSLOOK) with the H.J. Andrews Experimental Forest watershed in relation to channels identified by the National Water Model.

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, 2021a,b). 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, 2021a). 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., 1997). 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 (accessible at: https://registry.opendata.aws/nwm-archive/) 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 Fig. 2. These inputs can be divided into three categories: meteorological forcing data, hydrological model outputs, and channel geometry parameters (Fig. 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 (accessible at: https://www.nohrsc.noaa.gov/pub/staff/keicher/NWM_live/NWM_parameters/NWM_parameter_files.tar.gz).

2.3. Modeling approach

2.3.1. Model resolution

We simulated hourly water temperatures throughout the H.J. Andrews stream network during a six-week period of low flow from July 1 to August 15, 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 and 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):

$$\xi = x_o - \int_t^{t+\Delta t} u \, dt \tag{1}$$

where,

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

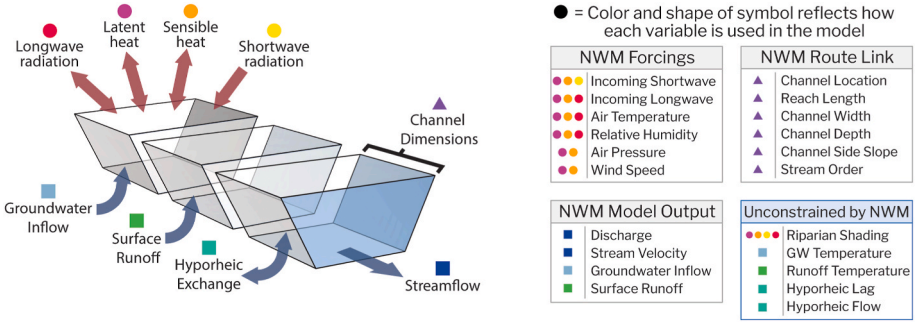


Fig. 2. Primary heat fluxes represented in the water temperature model. Model data sources fall into four broad categories: NWM gridded forcings, NWM model outputs, NWM channel route link files, and external data unconstrained by the NWM. The color and shape of symbology indicates how each variable contributes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

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):

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

where:
ξ = upstream location of the water parcel at time t, m
$u(j')$ = flow velocity in the jth model segment at time t, m s^{-1}
$\Delta(j')$ = time taken to traverse the jth model segment, s
j = index of model segment, unitless

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

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

where.

$x_{j'}$ = locatliion 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 $t + \Delta dt$. 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 jth model segment at time $t + \Delta(j')$ by integrating radiative and hydrologic energy fluxes along each model segment:

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

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 jth 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, J kg^{-1} °C^{-1}
$D(t,x_{j'})$ = water 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, the above formula 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:

$$H = H_{SW} + H_{LW} + H_{SH} + H_{LH} \tag{5}$$

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 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:

$$Q_I = Q_S + Q_T + Q_{GW} \tag{6}$$

where.

Q_I = total inflow rate, m^3 s^{-1}
Q_S = surface water runoff rate, m^3 s^{-1}
Q_T = tributary inflow rate, m^3 s^{-1}
Q_{GW} = groundwater inflow rate, m^3 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):

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

where.

T_I = effective temperature of inflows,°C
T_S = temperature of surface water runoff,°C
T_T = temperature of tributary inflow,°C
T_{GW} = temperature of groundwater inflow,°C

Q_I = total inflow rate, $m^3\ s^{-1}$
Q_S = surface water runoff rate, $m^3\ s^{-1}$
Q_T = tributary inflow rate, $m^3\ s^{-1}$
Q_{GW} = groundwater inflow rate, $m^3\ s^{-1}$

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):

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

where.

Φ = effective advective heat flux, °C s^{-1}
Q_L = total lateral inflow rate, $m^3\ s^{-1}$
V = model reach volume, m^3
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 main-stem 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:

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

where.

$C_{AT\text{-}GW}$ = Air temperature–groundwater temperature coefficient, varying from 0 to 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:

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

where:
W = air temperature moving window duration, days.
AT_n = daily mean air temperature on day n, °C.

By tuning $C_{AT\text{-}GW}$ 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) (Fig. 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, Oregon State University, 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:

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

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:

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

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, $m^3\ s^{-1}$
V = model reach volume, m^3
T_{hyp} = hyporheic return flow temperature,°C
T = channel water temperature,°C

We tuned the hyporheic lag duration parameter between 2 and 24 h, simulating a range of hyporheic flow path velocities. Although hyporheic flow paths often have residence times longer than 24 h, the variability in the mean temperature of simulated streamflow (approximating hyporheic return temperature) over periods longer than 24 h is negligible. As such, we limited the lag duration to a maximum of 24 h 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 a 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).

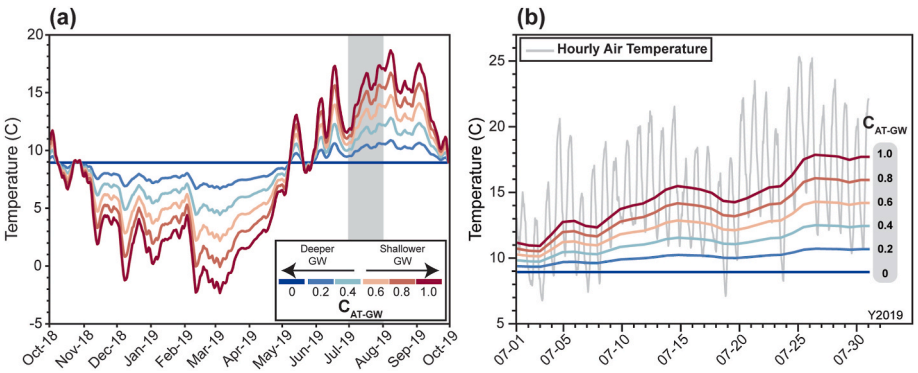


Fig. 3. Estimated groundwater inflow temperatures for (a) water year 2019 and (b) July 2019 study period at the headwaters of Mack Creek for a range of $C_{AT\text{-}GW}$ (air temperature scaling coefficient) values. Values of $C_{AT\text{-}GW}$ closer to 0 represent relatively deeper sourcing depths.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

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:

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

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 50 m
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):

$$h_r = \frac{h_v * 100}{r_w * 1.62} \tag{14}$$

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, 2020a,b). 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 performance*

Model performance was evaluated during distinct calibration and validation periods. Model calibration was performed during the first four weeks of the selected time period (July 1 to July 31). The model was then reinitialized and run during a two-week model validation period (August 1 to August 15). 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 during the calibration and validation periods. In the calculation of error metrics, we removed the first 48 h of simulated temperatures in both the calibration and validation periods to account for model spin-up. While this spin-up time is shorter than that of other hydrologic models, we found it sufficient, as the boundary condition temperatures rapidly equilibrated with radiative forcings downstream 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-h 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 5000 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 ($C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$, 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 ($C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$, 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 5000 model calibration runs of each configuration (M1, M2, M3, M4), we selected the top 1% of runs (50 runs) sorted by $RMSE_w$ to represent peak potential model performance. $RMSE_w$ is a weighted error metric, calculated as the weighted average of headwater RMSE (25% weight) and outlet RMSE (75% weight). These runs are highlighted amongst all calibration model runs in Fig. 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.

The parameter sets of the top 1% of calibration runs ranked by $RMSE_w$ were further evaluated during a two-week validation period (Fig. 4). By assessing model performance using only the top 1% of calibrated parameter combinations ranked by $RMSE_w$ (inferred to represent 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.

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

Model	Tuned Parameters	Number of Parameters
M1	$C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$, W, R_{shade}	5
M2	$C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$, W, R_{shade}, GW^1, GW^2, GW^3	8
M3	$C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$, W, R_{shade}, H_{lag}, H^1_{frac}, H^2_{frac}, H^3_{frac}	9
M4	$C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$, 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
$C^1_{AT\text{-}GW}$	AT-GW coefficient (first-order)	Unitless	0–1	M1, M2, M3, M4
$C^2_{AT\text{-}GW}$	AT-GW coefficient (second-order)	Unitless	0–1	M1, M2, M3, M4
$C^3_{AT\text{-}GW}$	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

3. Results

3.1. Calibrated and validated 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 ($C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$, 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 gages (Figs. 5–7). The 1st percentile of calibration 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 (Fig. 5). The performance of M1 decreased for validation runs, where the top calibrated parameter sets produced a mean headwater RMSE of 2.70 °C and a mean outlet RMSE of 2.47 °C. This set of best calibrated runs overestimated peak temperatures in the headwater reach, with a daily maxima error of 3.70 °C during the validation period (Figs. 5 and 6). Despite this strong positive bias in headwater reaches, M1 outlet predictions had a negative bias during validation, underestimating daily minima by −1.04 °C (Figs. 5 and 7). The best M1 validation run ranked by $RMSE_w$ had a headwater RMSE of

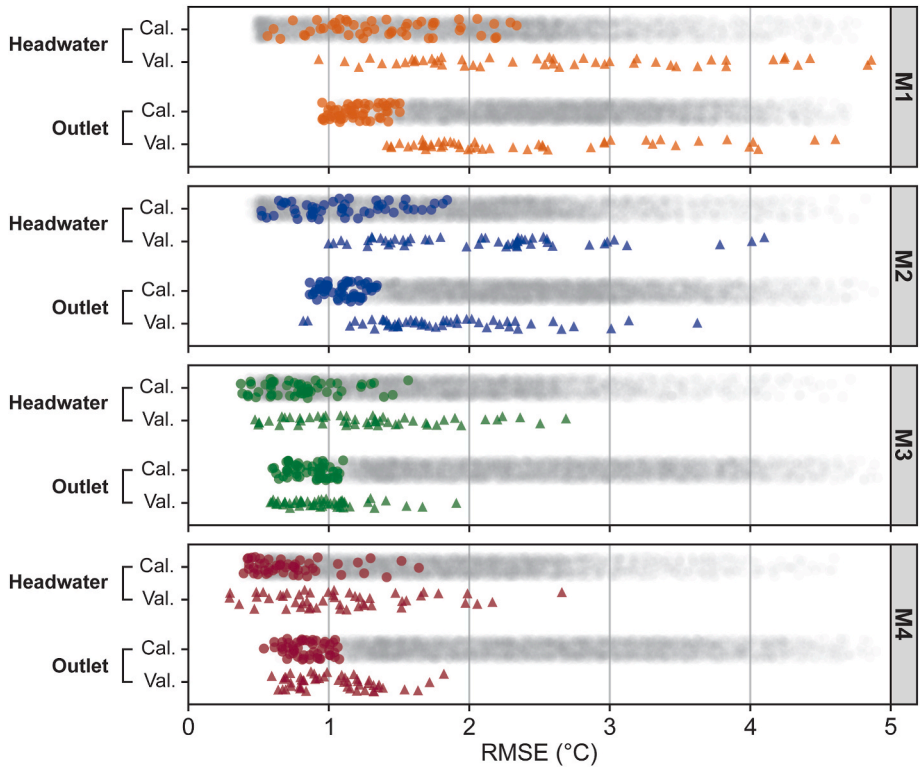


Fig. 4. Simulated water temperature prediction RMSE (°C) at headwater (Mack Creek) and outlet (Lookout Creek) gages. For the calibration (Cal.) of each model configuration, the top 1% of runs, ranked by $RMSE_w$ (weighted headwater (25%) and outlet (75%) RMSE), are highlighted as colored points amongst all calibration runs (represented by gray points). The validation (Val.) performance of the top 1% of calibrated parameter sets is represented by colored triangles.

1.48 °C and an outlet RMSE of 1.41 °C. In several validation runs of M1, modeled temperatures spiked to implausibly high levels (>30 °C; Figs. 6 and 7). These temperature spikes coincided with periods when relatively humidity values approached 100%.

3.1.2. M2: variable groundwater inflow rate

In addition to the variables tuned in M1 ($C_{AT\text{-}GW}$, 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 validation M1 runs, with mean RMSE values of 2.12 °C and 1.84 °C at the headwater and outlet gages, respectively (Fig. 5). However, much like M1, M2 struggled to reproduce diurnal temperature cycles throughout the stream network (Figs. 6 and 7). In the headwater reach, M2's lower mean error was largely driven by a narrowing of the diurnal cycle and a shift of predicted daily minima to cooler temperatures (Figs. 5 and 6). Headwater maxima error during validation was reduced to 2.78 °C while daily minima error was reduced to 0.20 °C. Improvements 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 (Figs. 5 and 7). The negative bias in outlet predictions observed for M1 also persisted for M2 (Fig. 7). The best validation run for M2 had a headwater RMSE of 1.00 °C and an outlet RMSE of 0.85 °C. M2 also experienced similar implausible temperature spikes during validation as observed in M1, again coinciding with periods of high relative humidity.

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). M3 resulted in a considerable improvement in performance in comparison to both M1 and M2, with the top 1% of calibrated parameter sets producing mean validation RMSE values of 1.29 °C and 0.97 °C in the headwaters and outlet, respectively (Figs. 5–7). The M3 configuration greatly reduced the positive headwater bias observed in previous model configurations, reducing daily headwater maxima error to 1.44 °C during validation and more accurately representing the observed magnitude of diurnal variability (Figs. 5 and 6). We also observed improvements in predicting daily minimum temperatures at the outlet, where validation minima error improved to 0.28 °C (Fig. 5). Across 5000 model runs, M3's best run ranked by validation $RMSE_w$ had a headwater RMSE of 0.80 °C and an outlet RMSE of 0.76 °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 (Fig. 5). Predicted water temperature envelopes from M3 and M4 at the headwater and outlet were difficult to distinguish visually (Figs. 6 and 7). M4 had the lowest mean validation RMSE value amongst all model configurations at the headwater gage (1.10 °C). The configuration's validation RMSE at the outlet gage (1.04 °C) was comparable to that of M3 (Fig. 5). Headwater daily maxima and minima prediction error for M4 were low compared to other model configurations during validation, with values of 1.12 °C and −0.29 °C respectively (Fig. 5). The best performing M4 run had a headwater RMSE of 0.70 °C and an outlet RMSE of 0.59 °C over the validation period.

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

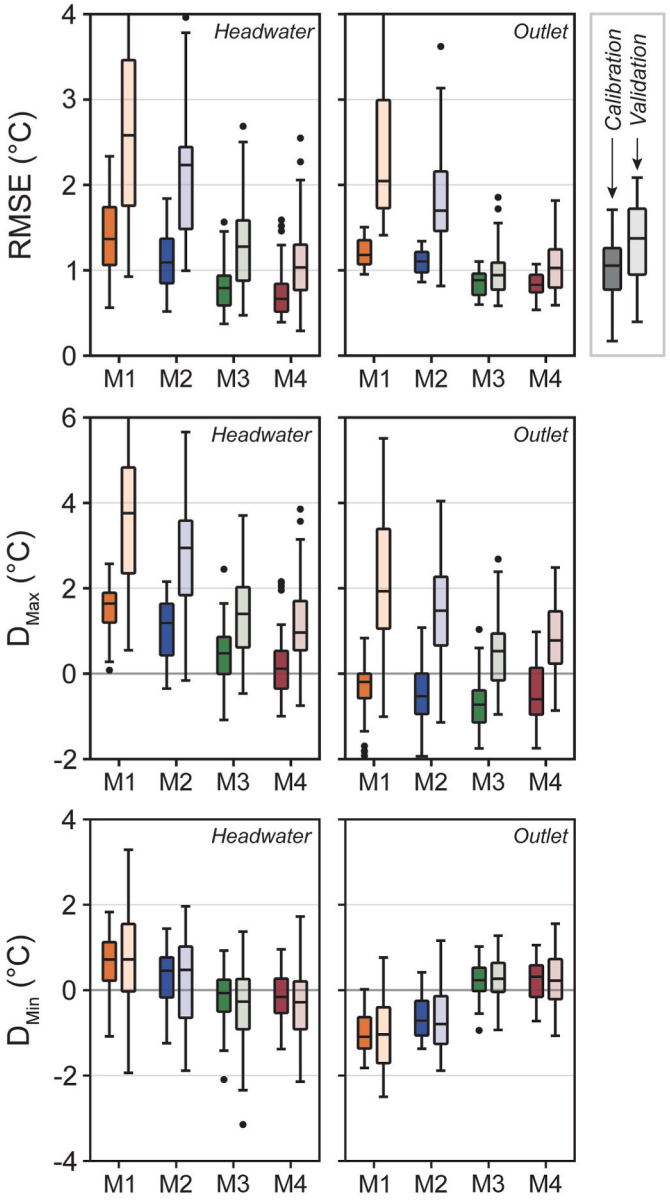


Fig. 5. Performance of four model configurations at the headwater (Mack Creek) and outlet (Lookout Creek) gages, evaluated across three metrics of model error (RMSE: root mean square error; D_{Max}: daily maxima error; D_{Min}: daily minima error). Paired boxplots show error metrics during calibration (left) and validation (right) periods for each configuration. Error metrics calculated using parameter sets from the top 1% of calibration runs, ranked by $RMSE_w$ (weighted headwater (25%) and outlet (75%) RMSE).

temperature of groundwater inflow for first-, second-, and third-order streams ($C^1_{AT\text{-}GW}$, $C^2_{AT\text{-}GW}$, $C^3_{AT\text{-}GW}$) were tuned for all model configurations. First-order coefficients ($C^1_{AT\text{-}GW}$) 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 $C_{AT\text{-}GW}$ 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 0.999 and 1.042 respectively, suggesting little bias in estimated riparian shading values. However, M3 and M4 had notably lower optimal R_{shade} values of 0.690 and 0.699, 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 1.141 to 1.433, 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 11.380 and 11.740 h, 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 0.666 and 0.682, 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 certain configurations of our model confirm that water temperatures can be successfully simulated using inputs derived from a continental-scale hydrologic model (in this case, the NWM). Of the four configurations tested, two (M3 and M4) produced calibrated simulations with validation RMSEs near or below 1.0 °C at both the headwater and outlet reaches. These prediction errors are well below the 2.0 °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 select configurations of our model 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 (Figs. 5–7). 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 reduction in error across most metrics over the previous model version (M1) (Fig. 5). By contrast, the addition of these calibrated parameters to M4 did not result in considerable improvement over M3 (Fig. 5). We hypothesize that the

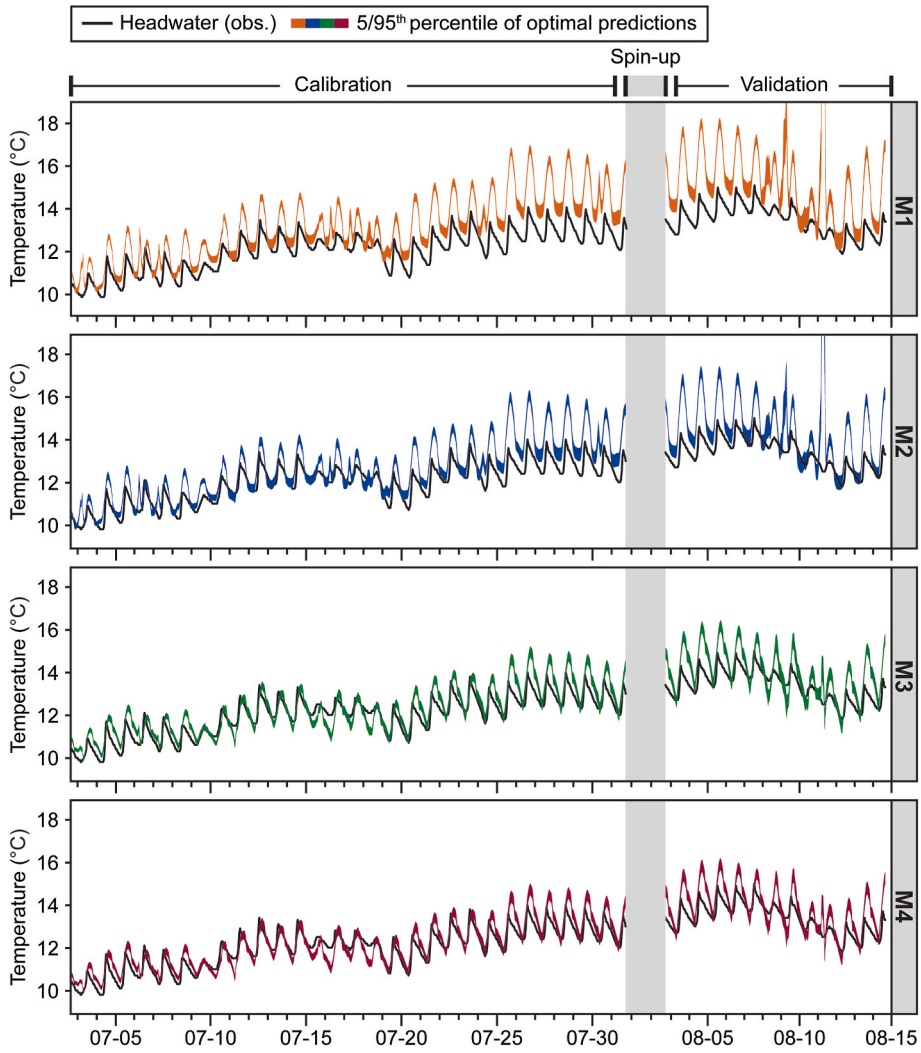


Fig. 6. Observed headwater temperatures (black) and 5/95th confidence envelope of water temperature predictions at the headwater gage across model configurations M1, M2, M3, and M4 for the top 50 calibration runs (1st percentile), ranked by weighted headwater and outlet RMSE ($RMSE_w$). Predictions displayed during a four-week calibration period and a two-week validation period, separated by a 48-h spin-up period for validation.

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 a 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.

All configurations of the model struggled to simultaneously generate accurate predictions at both the headwater and outlet gages to varying degrees. While model predictions were often capable of providing accurate predictions at the headwater gage (Fig. 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 (Figs. 5–7). 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 (Figs. 6 and 7). This effect was partially - though not entirely - alleviated by the inclusion of a conceptual hyporheic zone in M3 and M4 (Figs. 6 and 7), indicating that model configurations presented here may not fully capture all relevant heat fluxes in the system.

As we note in Section 3.1.1 and 3.1.2, a large proportion of M1 and M2 runs generated temperature spikes that exceeded plausible stream temperatures ranges typically observed at this site. These temperature spikes consistently coincided with periods of high relative humidity. Through further investigation of modeled heat fluxes, we found that the spikes were linked to high rates of sensible heat transfer into the stream, likely driven by our calculation of sensible heat using the Bowen Ratio (see Supporting Information Eq. 17 and 18). As Bowen's ratio is scaled

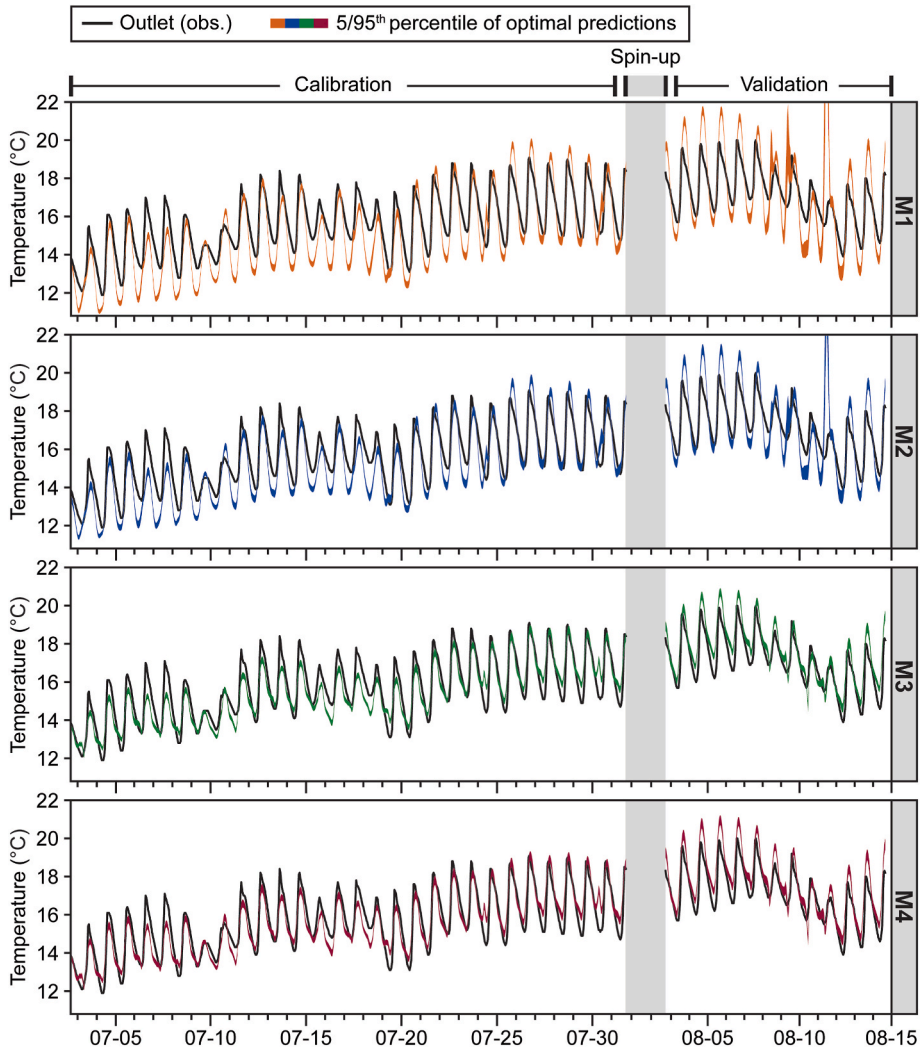


Fig. 7. Observed outlet temperatures (black) and 5/95th confidence envelope of water temperature predictions at the outlet gage across model configurations M1, M2, M3, and M4 for the top 50 calibration runs (1st percentile), ranked by weighted headwater and outlet RMSE ($RMSE_w$). Predictions displayed during a four-week calibration period and a two-week validation period, separated by a 48-h spin-up period for validation.

Table 3
Optimal mean parameter values for the top 50 calibration runs of each model (M1-M4), ranked by $RMSE_w$.

Parameter	M1	M2	M3	M4
$C^1_{AT\text{-}GW}$	0.633	0.563	0.261	0.269
$C^2_{AT\text{-}GW}$	0.828	0.837	0.487	0.554
$C^3_{AT\text{-}GW}$	0.640	0.745	0.525	0.589
W	7.200	7.220	7.880	7.420
R_{shade}	0.999	1.042	0.690	0.699
GW^1	–	1.141	–	1.262
GW^2	–	1.433	–	1.325
GW^3	–	1.339	–	1.279
H_{lag}	–	–	11.380	11.740
H^1_{frac}	–	–	0.666	0.682
H^2_{frac}	–	–	0.478	0.469
H^3_{frac}	–	–	0.175	0.142

by a relative humidity-dependent term (Bowen, 1926; Boyd and Kasper, 2003), high relative humidity values approaching 100% could lead to erroneous sensible heat fluxes, which we observe here. We attempted to alleviate these sensible heat errors by capping relative humidity at 97% in our model, but this only eliminated temperature spikes in some runs of M1 and M2. Alternative approaches to estimate sensible heat exchange during periods of high relative humidity are likely needed in catchments experiencing high humidity to remedy these errors.

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 and 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 and 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 and Garner, 2015). Rigorous on-site monitoring is required to determine the rate and temperature of groundwater inflow to channels (Caissie and 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 and 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 and 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 (Fig. 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 ($C_{AT\text{-}GW}$) to independently vary by stream order. The performance of our calibrated approach can be roughly assessed by evaluating the error in headwater daily minima temperatures (D_{Min}), as diurnal minima are closely coupled to the temperature of inflows. By this metric, our approach was successful when well-calibrated, with all configurations producing headwater D_{Min} values under 0.75 °C during the validation period (Fig. 5). 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 (Benz et al., 2022; Hannah and 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 and 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. We also note that while our approach did tune hyporheic lag time, we opted not to include parameters that would damp the temperatures of hyporheic return flows so as to minimize the risk of overparameterization. The exclusion of hyporheic damping, a recognized thermal process related to groundwater-surface water exchange (Briggs et al., 2018; Caissie et al., 2014; Caissie and Luce, 2017), could limit the peak potential performance of tested model configurations.

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 (Figs. 5–7). 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 and 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; Fig. 5). This likely indicates that our representation of hyporheic flow, which returned warmer daytime waters roughly 12 h 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., 2019a,b). 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

Our study details the development and implementation of a water temperature model for a single catchment during a summer period. Expanding this model to the continental scale will require further model calibration and validation for catchments in different climates and hydrologic settings. At present, the processes included in the model are most relevant for summer periods. It is therefore unlikely that the model will perform well in catchments at high latitudes during winter periods. Further work would be needed to incorporate additional hydrologic and heat exchange processes that will likely vary across catchments and throughout the year. At present, we also caution that the model includes processes that have not yet been tested, such as the incorporation of surface water inflow and associated water temperatures of this inflow, and reserve such testing for future work.

Ultimately, we found that our ability to test this model in different catchments was limited by the availability of water temperature observations concentrated within a single catchment. This need is what led us to select the location of the case study presented here. Watershed-scale water temperature monitoring occurs infrequently, and such data is often shared in disparate repositories, if shared at all. In contrast, monitoring by the USGS aims to cover a breadth of watersheds and rivers, though expanded river temperature monitoring networks in the Delaware River Basin show promise for assessing water temperature models. Our hope in making this model publicly available is that other river temperature modelers will apply it in their watersheds, incorporating their process-based understanding into their assessment of model fidelity, leading to future improvements in the model framework. While such benchmarking and community building is common in the watershed modeling space, it is rarer in water quality modeling. Future improvements in water quality modeling would certainly benefit from a community-based approach to assessing the ability of models to simulate water temperature (as well as other constituents), to evaluate to what extent our understanding of the hydrology informs this modeling, and to shape improvements in the model framework.

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. In many ways, the NWM offers a framework for river networks and detailed information on hydrologic fluxes and meteorological inputs that make it an ideal foundation for a temperature model. For comparison, other gridded water temperature models (e.g., DHSVM-RBM) require the modeler to gather meteorological observations at weather stations (that must be interpolated in some way across the study domain), define a river network (based on a digital elevation model), and estimate or tune key hydrologic fluxes. While there are certainly benefits to using the NWM as a framework for building a water temperature model, 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 and 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 and 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. (2019a,b) 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 model is 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, 2021b). 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 and validation 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. It is equally worth noting that our simulations were performed for a low flow period during summer, given the relevance of such periods to ecological management. Model performance during other periods of the year should be investigated to fine-tune process representation and to further generalize the model framework.

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.

Software availability

Name of the software: NWM-WT.
Developer: Jeffrey Wade.
Contact Information: jswade@syr.edu.
Year first available: 2023.
Program language: Python.
Source code availability: https://github.com/jswade/NWM-water-temperature.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Jeffrey Wade reports financial support was provided by National Oceanic and Atmospheric Administration.

Data availability

I have shared a link to my data and code in the Software Availability section of the manuscript.

Acknowledgements

Funding for this work was provided in part by the NOAA William M. Lapenta Internship Program and the Syracuse University Research Excellence Doctoral Funding Program. This material is based upon work supported by the H.J. Andrews Experimental Forest and Long-Term Ecological Research (LTER) program under the NSF grant LTER8 DEB-2025755. We thank Dr. Fred Ogden for his invaluable contributions in guiding the first steps of this research. Comments from two anonymous reviewers greatly improved this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2023.105866.

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 Resour. Res. 59 (1) https://doi.org/10.1029/2022WR032718.

Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T.B.M.J., Bobée, B., 2007. A review of statistical water temperature models. Can. Water Resour. J. 32 (3), 179–192. https://doi.org/10.4296/cwrj3203179.

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). Hydrol. Process. 26 (4), 475–484. https://doi.org/10.1002/hyp.8141.

Benz, S.A., Menberg, K., Bayer, P., Kurylyk, B.L., 2022. Shallow subsurface heat recycling is a sustainable global space heating alternative. Nat. Commun. 13 (1) https://doi.org/10.1038/s41467-022-31624-6.

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. Rev. Geophys. 52 (4), 603–679. https://doi.org/10.1002/2012RG000417.

Bowen, I.S., 1926. The ratio of heat losses by convection and evaporation from any water surface. Phys. Rev. 2 (27), 779–787.

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.

Briggs, M.A., Johnson, Z.C., Snyder, C.D., Hitt, N.P., Kurylyk, B.L., Lautz, L., Irvine, D.J., Hurley, S.T., Lane, J.W., 2018. Inferring watershed hydraulics and cold-water habitat persistence using multi-year air and stream temperature signals. Sci. Total Environ. 636, 1117–1127. https://doi.org/10.1016/j.scitotenv.2018.04.344.

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

Caissie, D., Luce, C.H., 2017. Quantifying streambed advection and conduction heat fluxes. Water Resour. Res. 53 (2), 1595–1624. https://doi.org/10.1002/2016WR019813.

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. J. Hydrol. 519 (PB), 1441–1452. https://doi.org/10.1016/j.jhydrol.2014.09.034.

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. Hydrol. Process. 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 Sci. Rev. 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 Resour. Res. 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. Rev. Fish Biol. Fish. 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. Reg. Environ. 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. Hydrol. Process. 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. Hydrol. Process. 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. Environ. Model. Software 92, 213–228. https://doi.org/10.1016/j.envsoft.2017.02.021.
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.
Gochis, D.J., Barlage, M., Cabell, R., Casali, M., Dugger, A., Fitzgerald, K., Mcallister, M., Mccreight, J., Rafieeinasab, 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.
Gregory, S., Johnson, S., 2019. Stream and air temperature data from stream gages and stream confluences in the Andrews Experimental Forest, 1950 to present. In: Long-Term Ecological Research. Forest Science Data Bank. https://doi.org/10.6073/pasta/9437d1603044f5b92189110dd8343763. Corvallis, OR. [Database]. Available: http://andlter.forestry.oregonstate.edu/data/abstract.aspx?dbcode=HT004.
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. Prog. Phys. Geogr. 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. Hydrol. Process. 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. J. Am. Water Resour. Assoc. 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. Environ. Sci. Technol. 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 Resour. Res. 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-Odoux, C., 2014. Process consistency in models: the importance of system signatures, expert knowledge, and process complexity. Water Resour. Res. 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. Mon. Weather Rev. 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. Environ. Model. 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 Resour. Res. 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. Can. J. Fish. Aquat. Sci. 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. Hydrol. Process. 35 (5) https://doi.org/10.1002/hyp.14187.
Julien, P.Y., Saghafian, B., Ogden, F.L., 1995. Raster-based hydrologic modeling of spatially-varied surface runoff. J. Am. Water Resour. Assoc. 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. Knowl. Manag. Aquat. Ecosyst. 418 https://doi.org/10.1051/kmae/2016037.
Kasahara, T., Wondzell, S.M., 2003. Geomorphic controls on hyporheic exchange flow in mountain streams. Water Resour. Res. 39 (1) https://doi.org/10.1029/2002wr001386. SBH 3-1-SBH 3-14.
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. Hydrol. Process. 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. J. Hydrometeorol. 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. J. Hydrometeorol. https://doi.org/10.1175/jhm-d-20-0198.1.
LANDFIRE, 2020a. Existing Vegetation Cover Layer, LANDFIRE 2.2.0. U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture. http://landfire.cr.usgs.gov/viewer. (Accessed 17 October 2022).
LANDFIRE, 2020b. Existing Vegetation Height Layer, LANDFIRE 2.2.0. U.S. Department of the Interior, Geological Survey, and U.S. Department of Agriculture. http://landfire.cr.usgs.gov/viewer. (Accessed 17 October 2022).
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. Hydrol. Earth Syst. Sci. 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. J. Hydrol. 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 Resour. Res. 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. J. Hydrol. 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 Res. Appl. 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 Resour. Res. 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 Resour. Res. 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. Hydrol. Process. 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. J. Geophys. Res. Atmos. 116 (12) https://doi.org/10.1029/2010JD015139.
NOAA, 2016. National Water Model: Improving NOAA's Water Prediction Services. http://water.noaa.gov.
NOAA, 2021a. 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, 2021b. OWP NWM NextGen Framework. Office of Water Prediction. https://www.weather.gov/media/owp/oh/docs/2021-OWP-NWM-NextGen-Framework.pdf. (Accessed 13 February 2023).
NOAA, 2023. The National Water Model. Office of Water Prediction. https://water.noaa.gov/about/nwm. (Accessed 30 January 2023).
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. Sci. Total Environ. 736 https://doi.org/10.1016/j.scitotenv.2020.139679.
PRISM Climate Group, Oregon State University, 2022. Data created Dec 2022. https://prism.oregonstate.edu (Accessed 4 October 2022).
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.
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. J. Am. Water Resour. Assoc. 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 Resour. Res. 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. Hydrol. Process. 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 Resour. Res. 48 (9) https://doi.org/10.1029/2012WR011819.
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. Hydrol. Earth Syst. Sci. 16 (11), 4303–4321. https://doi.org/10.5194/hess-16-4303-2012.

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 Environ. Change 23 (2), 450–464. https://doi.org/10.1016/j.gloenvcha.2012.11.002.

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. Hydrol. Earth Syst. Sci. 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. J. Hydrometeorol. 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 Resour. Res. 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 Resour. Res. 48 (4) https://doi.org/10.1029/2011WR011461.

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 Resour. Res. 52 (2), 1479–1497. https://doi.org/10.1002/2015WR018225.

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 Resour. Res. 53 (5), 3988–4003. https://doi.org/10.1002/2016WR019875.

Ward, A.S., Schmadel, N.M., Wondzell, S.M., 2018a. Simulation of dynamic expansion, contraction, and connectivity in a mountain stream network. Adv. Water Resour. 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 Resour. Res. 54 (3), 2017–2036. https://doi.org/10.1002/2017WR021502.

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., et al., 2019a. Spatial and temporal variation in river corridor exchange across a 5th-order mountain stream network. Hydrol. Earth Syst. Sci. 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., et al., 2019b. Co-located contemporaneous mapping of morphological, hydrological, chemical, and biological conditions in a 5th-order mountain stream network, Oregon, USA. Earth Syst. Sci. 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. Hydrol. Process. 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. J. Hydrol. 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. Environ. Res. Lett. 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 Resour. Res. 45 (12) https://doi.org/10.1029/2008WR007629.

Yearsley, J.R., 2012. A grid-based approach for simulating stream temperature. Water Resour. Res. 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. Hydrol. Earth Syst. Sci. 23 (11), 4491–4508. https://doi.org/10.5194/hess-23-4491-2019.