VOLUME 8 N°01 CONFLUENCE
Journal of Watershed Science and Management

Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period

R.D. Moore, *University of British Columbia*
Lilianne Callahan, *SLR Consulting Ltd., Vancouver, British Columbia*

Abstract

Stream temperature is widely considered a "master variable" in aquatic ecosystems. Stream temperature models are often used to evaluate historic thermal habitat suitability where measurements are lacking and to quantify the potential effects of future climate scenarios. A major barrier to the widespread application of models is the sparse coverage and limited duration of observed stream temperature records to support model calibration. This study quantified the sensitivity of the hybrid stream temperature model Air2stream to the length of the calibration record. The study involved 23 hydrometric stations in British Columbia, Canada, including rain-dominated, snow-dominated, hybrid rain-snow, and nivo-glacial regimes within regulated and unregulated systems. Air2stream was calibrated using two approaches. The first used all possible subsets of consecutive years ranging from one to eight years in length, depending on data availability at each site, followed by validation on the data not used for calibration. The second used the years 2021 and 2022 for validation after using all possible subsets of consecutive years up to and including 2020 for calibration. Air2stream generally performed best for rain-dominated and hybrid regimes and worst for nival regimes. Even just two years of calibration data produced validation root-mean-square errors of less than 2°C for all but one site, regardless of hydrologic regime or regulation status, and calibrating with three or four years provided similar performance in validation to longer calibration periods.

Keywords: *stream temperature, river temperature, Air2stream, calibration, validation*

Introduction

Stream temperature is widely considered to be a "master variable" in aquatic ecosystems, and is especially important in relation to thermal habitat suitability for fish and some amphibians (Jackson et al., 2001; Nelitz et al., 2007; Parkinson et al., 2016; Jones et al., 2017; McEwan et al., 2021; Johnson et al., 2024). Stream temperature varies spatially and temporally in response to the integrated effect of energy and water exchanges occurring throughout the channel network upstream of a point of interest (Leach et al., 2023). Energy and water exchanges are influenced by meteorological and climatic variability, catchment physiography, channel morphology, and land cover/land use, particularly the nature of riparian vegetation (Isaak et al., 2010; Moore et al., 2013; Fullerton et al., 2022).

Agencies and organizations responsible for managing aquatic resources require an understanding of the spatio-temporal variability of stream temperature and its implications for thermal habitat suitability for valued aquatic species (Weller et al., 2023; Fuller et al., 2025). A fundamental challenge to managers is that stream temperature is sparsely monitored in many regions (Dugdale et al., 2017; Daigle et al., 2023). Consequently, substantial effort over the last several decades has focused on developing and testing models for a variety of applications. These include the prediction of temperature for thermal habitat assessment at unmonitored sites (Moore et al., 2013; Weller et al., 2023) and the assessment of stream temperature response to land-cover changes such as forestry and urbanization (Chen et al., 1998; Sun et al., 2015; Fullerton et al., 2022), water management operations, including impoundments and withdrawals (Lee et al., 2020; Cheng et al., 2020), and potential effects of future climate change (Cheng et al., 2020; Lee et al., 2020; Fullerton et al., 2022; Weller et al., 2023).

Stream temperature models range from fully empirical approaches based on predictor variables such as air temperature, stream discharge, catchment snow cover, and reach-scale and catchment characteristics (Chu et al., 2009; Wehrly et al., 2009; Isaak et al., 2017; Jackson et al., 2018; Weller et al., 2023; Collins et al., 2025) to spatially distributed, process-based models that simulate the coupled influences of hydrological and thermal processes (Chen et al., 1998; St-Hilaire et al., 2000; van Vliet et al., 2012; Ficklin et al., 2012; MacDonald et al., 2014; Sun et al., 2015). A limitation of many empirical models is that they predict indices such as maximum weekly average temperature or mean August temperature (Moore et al., 2013; Isaak et al., 2017), which may not be applicable to a specific aquatic species and/or life-history stage. Process-based models typically generate time series of daily or sub-daily temperature, and thus provide flexibility in calculating temperature indices relevant to specific situations. However, process-based models require extensive input data that may not be available in operational contexts and also require highly trained analysts to set up and run them.

Toffolon and Piccolroaz (2015) introduced a hybrid physical-empirical model, Air2stream, which is based on a differential equation that represents simplified reach-scale heat and water budgets, with surface energy exchanges parameterized using air and water temperatures. The model has five variations, two of which require stream discharge as an input variable to represent the effects of stream depth and longitudinal advective fluxes. Air2stream predicts stream temperature as a time series with the same frequency as the input data, typically daily. Air2stream has been shown to outperform purely statistical models (Piccolroaz et al., 2016), including under extreme heatwave and extended drought conditions (Callahan & Moore, 2025a).

The Air2stream model has been used to hindcast historical temperatures (Islam et al., 2019; Zhu et al., 2022; Shrestha et al., 2024), and to make projections under future climate scenarios given projections of air temperature and stream discharge (Piotrowski & Napiorkowski, 2018). Time series of daily mean temperature generated by Air2stream could provide input to bioenergetic models, such as that applied by Nelitz et al. (2007), to evaluate the potential effects of climate change on fish growth. Alternatively, the daily time series could be used to compute thermal habitat suitability indices, such as maximum weekly average temperature or mean August stream temperature (Moore et al., 2013; Isaak et al., 2017). These indices could be combined with analyses, such as that by Parkinson et al. (2016), to predict species distributions under current or historical conditions or to evaluate the potential for shifts in fish species assemblages under future climate scenarios. These analyses could support the identification of thermally sensitive streams as inputs to decisions about allocation of habitat restoration efforts (Weller et al., 2023).

Callahan and Moore (2025a) identified two research directions required to support the broader application of Air2stream: to investigate whether modelled streamflow—for example, from a regionally calibrated model—can be used in place of observed streamflow for calibration, and to evaluate how long a period of stream temperature record is required to support a robust calibration. In relation to the latter research need, Piccolroaz et al. (2016) demonstrated that five years of data appear to be sufficient for robust calibration of Air2stream. However, stream temperature records are often shorter, frequently as short as one or two years (e.g., Moore et al., 2013). The objective of the current study is thus to quantify the sensitivity of Air2stream's predictive accuracy to the length of the data record used to calibrate the model, including consideration of records as short as one year.

Methods

Study domain

This study used the dataset produced by Callahan and Moore (2025a), which includes streamflow and stream temperature data collected at 23 Water Survey of Canada (WSC) hydrometric stations in British Columbia, Canada (Figure 1). Readers interested in seeing time series plots of the data should see Callahan and Moore (2025a) and the associated supporting information document (Callahan & Moore, 2025b).

Moore & Callahan. (2026). Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period.
http://confluence-jwsm.ca/index.php/jwsm/article/view/71/8 . doi:10.22230/jwsm.2026v8n1a71

Figure 1: Map of Water Survey of Canada hydrometric stations used in this study. Symbols indicate locations of the gauging stations and red lines indicate catchment boundaries, which were retrieved from Government of Canada (2024).

British Columbia (BC) covers almost one million square kilometres and has a diverse range of climatic regimes, topography, and vegetation, with streamflow regimes that include rain-dominated (pluvial), snow-dominated (nival), hybrid rain-snow, and nivo-glacial (Fleming et al., 2007; Déry et al., 2009). Rain-dominated regimes have minor streamflow contributions from snowmelt, and the seasonal distribution of streamflow is similar to that of rainfall. In snow-dominated catchments, a substantial portion of winter precipitation falls as snow that is stored within the catchment until it is released by melting in spring and early summer. Nivo-glacial regimes are similar to nival regimes, but nivo-glacial catchments tend to have higher elevations that lead to snowmelt contributions that persist later than those for nival catchments, along with contributions from glacier melt that sustain flows in late summer and early autumn (Moore et al., 2020). Hybrid regimes are characterized by high winter flows associated with rain events, particularly in lower-elevation portions of the catchments, as well as high spring and early summer flows from the melting of seasonal snow stored at higher elevations. In coastal areas, some nivo-glacial catchments occasionally exhibit high flows in autumn and winter during warm rain-on-snow events (Moore, 1993; Trubilowicz et al., 2016).

These streamflow regimes are associated with distinct thermal regimes (Quilty & Moore, 2010). Snow-dominated and nivo-glacial streams typically remain at 0°C between freeze-up in autumn through to the onset of snowmelt in late winter or early spring, when stream temperature rises rapidly to reach peak values in July and August. This peak is typically followed by a more gradual decline through autumn back to 0°C. In rain-dominated catchments, on the other hand, annual temperature cycles tend to appear like lagged and attenuated versions of the local air temperature cycle.

Water temperature typically remains above 0°C through winter, with occasional dips to 0°C during cold spells. Thermal regimes in hybrid rain-snow catchments lie on a spectrum between those for nival and rain-dominated hydrologic regimes.

Stream temperature data sources and processing

Hourly stream temperature data for all WSC hydrometric stations in BC with recorded temperatures were obtained from Environment and Climate Change Canada (ECCC). The values are reported with a resolution of 0.01°C, but visual examination and comparison with manual spot temperatures measured during site visits indicate that the data are less accurate, with time series sometimes exhibiting issues including spikes and multi-year sensor drift. Following a high-level automated check of the data to identify out-of-range values such as negative temperatures, Callahan and Moore (2025a) conducted a visual assessment of the recorded hourly stream temperature following Sowder and Steel (2012), which involved examining interactive plots of stream temperature (recorded and manual), air temperature, and stream discharge using functions in the plotly R package (Plotly Technologies Inc., 2015). Periods with suspected inaccurate data, particularly related to instrument drift and de-watering, were flagged and removed prior to analysis. See Callahan and Moore (2025a) for more details of the data cleaning and processing.

Daily mean stream temperature was computed from the hourly data for each stream for all days with at least 20 valid hourly measurements. Stations were selected for analysis based on having at least five total years of data after missing or erroneous periods were removed. Application of this criterion yielded a total of 23 streams, 15 with natural flow regimes, and eight regulated; see Figure 1 and Tables 1 and 2.

Table 1: Locations of Water Survey of Canada hydrometric stations used in the analysis.

Station ID	Station name	Latitude (°N)	Longitude (°W)	Station elevation (m)
07EA004	Ingenika River above Swannell River	56.73061	125.1050	685
07EB002	Ospika River above Aley Creek	56.52411	123.9360	739
08CG001	Iskut River Below Johnson River	56.73444	131.6690	21
08FF001	Kitimat River Below Hirsch Creek	54.04872	128.6905	17
08FF003	Little Wedeene River Below Bowbyes Creek	54.13639	128.6900	69
08GA077	Seymour River Below Orchid Creek	49.52030	123.0040	249
08HA002	Cowichan River at Lake Cowichan	48.82591	124.0530	164
08HA011	Cowichan River Near Duncan	48.77308	123.7145	20
08HA069	Renfrew Creek Near Port Renfrew	48.63675	124.2917	308
08HA070	Harris Creek Near Lake Cowichan	48.71938	124.2261	265
08HB023	Ash River Below Moran Creek	49.36992	124.9841	76
08JA017	Nechako River Below Cheslatta Falls	53.68550	124.8393	717
08JB008	Nadina River at Outlet of Nadina Lake	53.90250	126.9547	901
08KB001	Fraser River at Shelley	54.00367	122.6248	575
08KH006	Quesnel River Near Quesnel	52.84306	122.2253	540
08LB078	Lemieux Creek Near the Mouth	51.42759	120.2020	396
08LE031	South Thompson River at Chase	50.76311	119.7431	349
08LF051	Thompson River Near Spences Bridge	50.35463	121.3936	208
08MF040	Fraser River above Texas Creek	50.61367	121.8534	194
08NE074	Salmo River Near Salmo	49.04714	117.2943	594
08NG065	Kootenay River at Fort Steele	49.61203	115.6353	771
08NH118	Duncan River Below Lardeau River	50.23286	116.9552	552
08NK002	Elk River at Fernie	49.50347	115.0701	997

Moore & Callahan. (2026). Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period. http://confluence-jwsm.ca/index.php/jwsm/article/view/71/8 . doi:10.22230/jwsm.2026v8n1a71

Table 2: Characteristics of Water Survey of Canada hydrometric stations used in the analysis. Year 1 and Year n are the start and end years of the water temperature data, and n_{eff} is the effective number of years of the record, not counting missing values.

Station ID	Area (km^2)	Regulation	Regime	Year 1	Year n	n_{eff}
07EA004	4144	unregulated	snow	2011	2022	11.7
07EB002	2192	unregulated	snow	2011	2022	5.3
08CG001	9500	unregulated	nivo-glacial	2016	2022	5.0
08FF001	1993	unregulated	nivo-glacial	2014	2022	7.7
08FF003	179	unregulated	hybrid	2014	2022	8.4
08GA077	62	unregulated	hybrid	2012	2022	7.9
08HA002	594	regulated	rain	2011	2022	9.9
08HA011	826	regulated	rain	2015	2022	5.9
08HA069	7	unregulated	rain	2014	2022	8.2
08HA070	28	unregulated	rain	2016	2022	6.4
08HB023	387	regulated	rain	2016	2022	6.5
08JA017	15491	regulated	snow	2011	2022	11.2
08JB008	368	regulated	snow	2016	2022	6.1
08KB001	32436	unregulated	snow	2011	2022	10.1
08KH006	11546	unregulated	snow	2011	2022	9.0
08LB078	528	regulated	snow	2016	2022	6.8
08LE031	15819	unregulated	snow	2011	2022	9.7
08LF051	55417	unregulated	snow	2011	2022	11.2
08MF040	154232	regulated	snow	2011	2022	11.0
08NE074	1242	unregulated	snow	2013	2022	8.5
08NG065	11469	unregulated	snow	2011	2022	8.7
08NH118	4080	regulated	nivo-glacial	2011	2022	10.2
08NK002	3093	unregulated	snow	2011	2022	8.5

Stream discharge data sources and processing

Daily mean stream discharge data were obtained for the 23 WSC stations that met selection criteria based on stream temperature. Data from the beginning of each station's period of record until October 31, 2022, were obtained using functions in the tidyhydat package in R (Albers, 2017). Two stations each had a single gap in the streamflow record, the longest of which was eight days. These missing data were infilled by linear interpolation with the "na.approx" function in base R (R Core Team, 2021).

Stations were classified as having regulated or natural flow as reported in metadata provided by WSC. Streamflow regimes for unregulated catchments were based on visual inspection of daily streamflow hydrographs following the descriptions in the Study Domain section. A minimum glacier cover of 2% was used to distinguish nivo-glacial regimes from nival or hybrid regimes following Moore et al. (2009). For regulated catchments, the expected streamflow regime in the absence of regulation was based on regional context and catchment glacier cover. Note that Callahan and Moore (2025a) classified Duncan River below Lardeau River as glacial/hybrid. Based on the criteria applied in this study, it is classified as nivo-glacial.

Air temperature data sources and processing

Air temperature data were interpolated from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5 (ERA5) gridded surface product to the locations of the WSC stations. Daily mean 2-m air temperature data were downloaded through the Copernicus Climate Change Service (C3S) using the CDS API client in Python for the region bounded by the latitudes -145° to -105° and longitudes 46° to 62° from 1979 to 2022 (Copernicus, 2024). The temperatures and the surface geopotentials were projected onto a regular grid in the BC Albers Equal Area Conic projection via bilinear interpolation using the "projectRaster" function within the raster package in R (Hijmans & van Etten, 2012), creating a new resolution of 16.3 km 27.9 km.

The re-projected air temperature and geopotential data were used to generate air temperature series for each WSC station. The air temperatures at the four grid points surrounding each WSC station were adjusted to the station's elevation using monthly lapse rates for daily mean air temperature as reported by Stahl et al. (2006). These elevation-adjusted air temperatures were then bilinearly interpolated to the WSC station location. See Callahan and Moore (2025a) for further details. Callahan (2023) demonstrated that these ERA5-derived air temperatures performed similarly as predictors in regression models to air temperatures based on the Stahl et al. (2006) station-based interpolation used by Moore et al. (2013).

Air2stream model

The 8-parameter version of the Air2stream model was used in this study. The model is based on the following heat budget equation:

$$\rho c_p V \frac{dT_w}{dt} = AH + \rho c_p \left(\sum_i Q_i \, T_{w,i} - QT_w \right) \qquad (1)$$

where p is water density (kg m^{-3}), c_p is the specific heat capacity of water at constant pressure (J K^{-1} kg^{-1}), V (m^3) is the volume of water in a river reach and its tributaries, T_w is stream temperature (°C), t is time (expressed in days), A is the surface area of the reach (m^2), H is the net heat flux between the river and the atmosphere (W m^{-2}), Q is the streamflow at the temperature monitoring location (m^3 s^{-1}), Q_i is the discharge associated with the i^{th} contributing upstream water flux (m^3 s^{-1}), and $T_{w,i}$ is the temperature of the i^{th} contributing upstream water flux (°C).

The net heat flux at the water surface is parameterized as a linear function of air and water temperatures, and the governing equation in Air2stream takes the following form:

$$\frac{dT_w}{dt} = \frac{1}{\delta}\left[a_1 + a_2 T_a - a_3 T_w + \theta \left[a_5 + a_6 cos\left(2\pi \left(\frac{t}{t_y} - a_7 \right) \right) - a_8 T_w \right] \right] \qquad (2)$$

where T_w is stream temperature (°C), T_a is air temperature (°C), t is time (days or other units), t_y is the duration of a year (same units as t), $\delta = \theta^{\wedge a_4}$, a_1 to a_8 are the parameters of the model, $\theta = Q\bar{Q}^{-1}$, Q is the streamflow (m^3 s^{-1}), and $\bar{Q}$ is the long-term mean streamflow. The terms associated with parameters a_1, a_2 and a_3 represent vertical energy fluxes, and the remaining terms represent advective heat inputs. The term δ is related to mean water depth, which is parameterized as a function of discharge.

Calibration and testing

Calibration and validation runs were conducted in two ways. In the first approach, sets of calibrations used $n_c = 1$ to n_{max} calendar years of data for each station, where n_{max} is the lesser of 8 and n - 1, and n is the length of the station's record in calendar years. Calibration data sets were limited to sequential calendar years. For each calibration, the remaining $n_v = n - n_c$ calendar years of data were used for validation. For example, for $n_c = 2$ and a data record spanning the years 2011 and 2022, the first iteration used 2011 and 2012 to calibrate and the years 2013 to 2022 to validate; the second iteration used 2012 and 2013 to calibrate and the years 2011 and 2014 to 2022 to validate; the third iteration used 2013 and 2014 to calibrate and the years 2011, 2012, and 2015 to 2022 to validate; and so on.

Calibration and testing were only conducted if the total number of days with observations in the calibration and validation periods each exceeded 300, so as to ensure that each calibration and validation data set captured the majority of an annual cycle. Some testing using shorter records indicated that calibrations were not robust when less than about one-half of a year was used.

A complication in interpreting the results from the first approach is that the validation period varies from run to run. To address this complication, a second approach used a common validation period comprising the data for 2021 and 2022, following Callahan and Moore (2025a). The year 2021 saw an extreme heatwave from June 25, 2021, to July 2, 2021, while 2022 saw an extended drought in

September and October. See Callahan and Moore (2025a) for more details about the meteorological conditions during those extreme weather events.

In this second approach, calibration periods comprised sets of sequential years up to and including the year 2020, with calibration periods ranging from 1 to n_{max}, where n_{max} is the lesser of 8 and n - 2. For each calibration run, *RMSE* values were computed for the whole validation period and also for the 2021 heatwave and 2022 drought periods. *RMSE* and mean bias error (*MBE*) values were also computed for the heatwave and autumn drought date ranges in the calibration period for comparison.

Parameter bounds used to constrain the calibrations are shown in Table 3. These are consistent with those specified in the example workflow available via the Air2stream Github site (Piccolroaz, 2017), with the exception of parameters a_3, a_4 and a_8. In this application, a_4 was constrained to lie in the range [0, 1] to ensure that the calibrated values are physically consistent with the concept of hydraulic geometry, while parameters a_3 and a_8 were constrained to be positive values to ensure that dT_w/dt was negatively related to T_w to be consistent with the positive relationship between heat loss from the water column and T_w.

Table 3: Upper and lower bounds for calibrated parameters.

Bound	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
upper	15	1.5	5	1	20	10	1	5
lower	-5	-5	0	0	0	0	0	0

Results

As seen in Figure 2, the *RMSE* value in the calibration period tended to increase with the number of years of data used for calibration, but generally remained below 1°C. Conversely, in the validation period, *RMSE* decreased when the number of years of data used for calibration increased from one to three years, then remained relatively constant. When two or more years of data were used to calibrate Air2stream, *RMSE* remained below 2°C for all but one site, which has a snow-dominated regime. *RMSE* was higher for snowmelt-dominated catchments and tended to be higher for regulated systems.

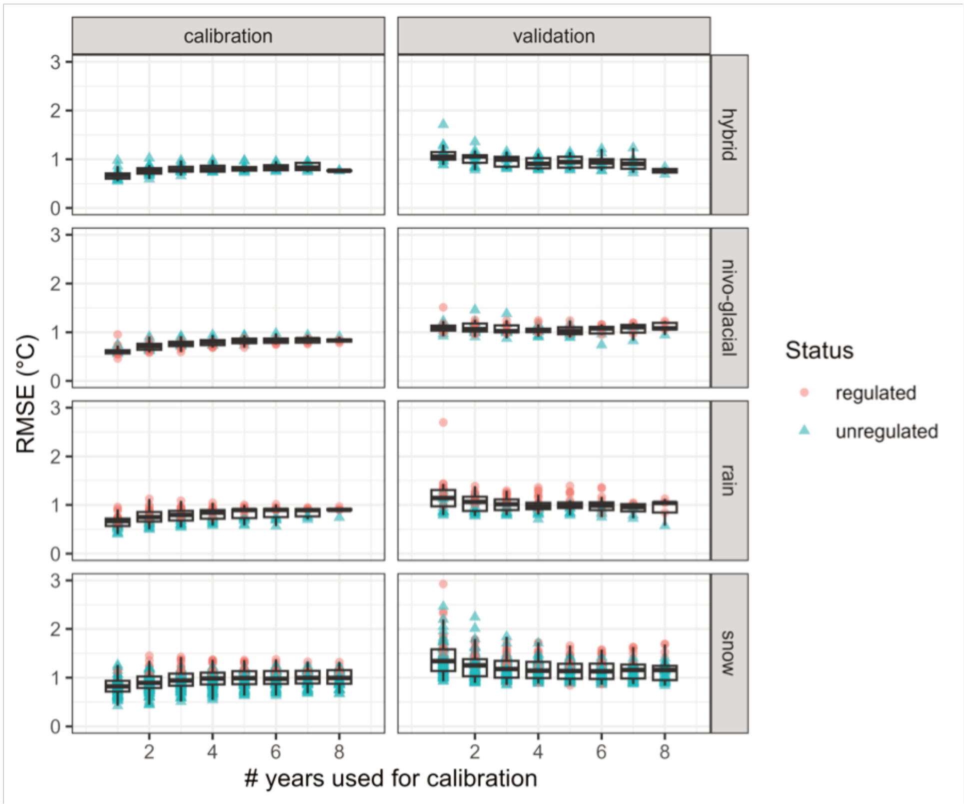

Figure 2: Boxplots showing root-mean-square error (RMSE) in the calibration and validation periods as a function of the number of years used to calibrate Air2stream. The central box indicates the upper hinge, median, and lower hinge of the data distribution, and the whiskers extend to largest and smallest values that lie within ± 1.5 interquartile ranges of the median.

Figure 3 shows the calibrated parameter values as a function of the number of years used to calibrate Air2stream, hydrological regime, and regulation status. The a_4 parameter calibrated to 0 for most cases, indicating that the δ term in Eq. 2 was unity. Some parameters displayed a tendency for their range to decrease with increasing length of calibration for some regimes, such as the parameter a_2 for hybrid regimes and parameter a_5 for nival regimes, but the majority of cases had no clear relationship. The distributions of some parameters appeared to vary somewhat among regimes, such as parameter a_3, which tended to be higher for hybrid regimes, and parameter a_7, which tended to be lower but have a greater range for nival regimes.

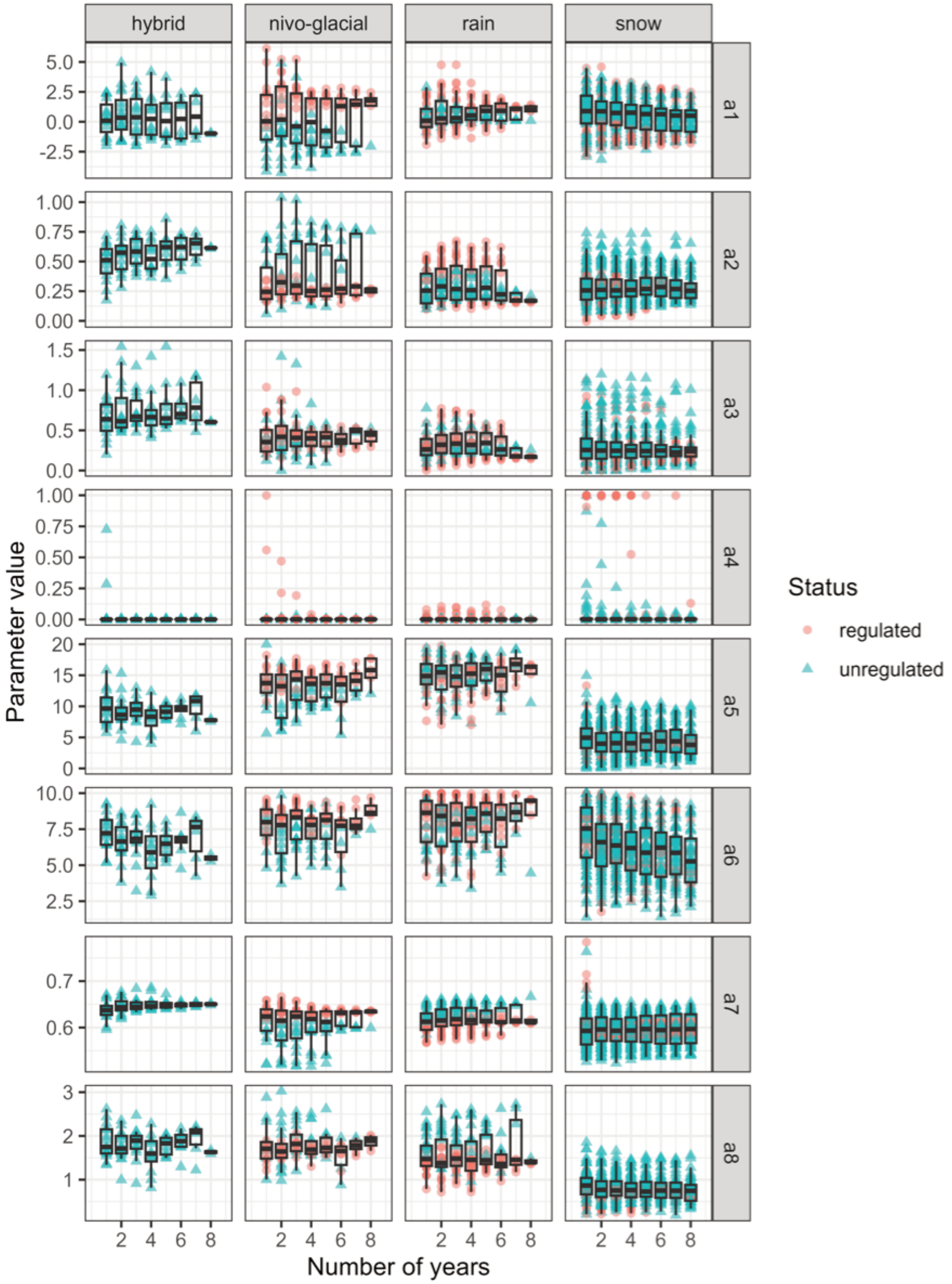

Figure 3: Boxplots showing calibrated parameter values as a function of the number of years used to calibrate Air2stream. See Figure 2 for a description of the boxplots.

When the validation period was fixed to the years 2021 and 2022, the relationships between *RMSE* and the number of calibration years were qualitatively similar to the situation shown in Figure 2, but with less variability (Figure 4). When considering just the date range for the heatwave event, *RMSE*

Moore & Callahan. (2026). Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period. http://confluence-jwsm.ca/index.php/jwsm/article/view/71/8 . doi:10.22230/jwsm.2026v8n1a71

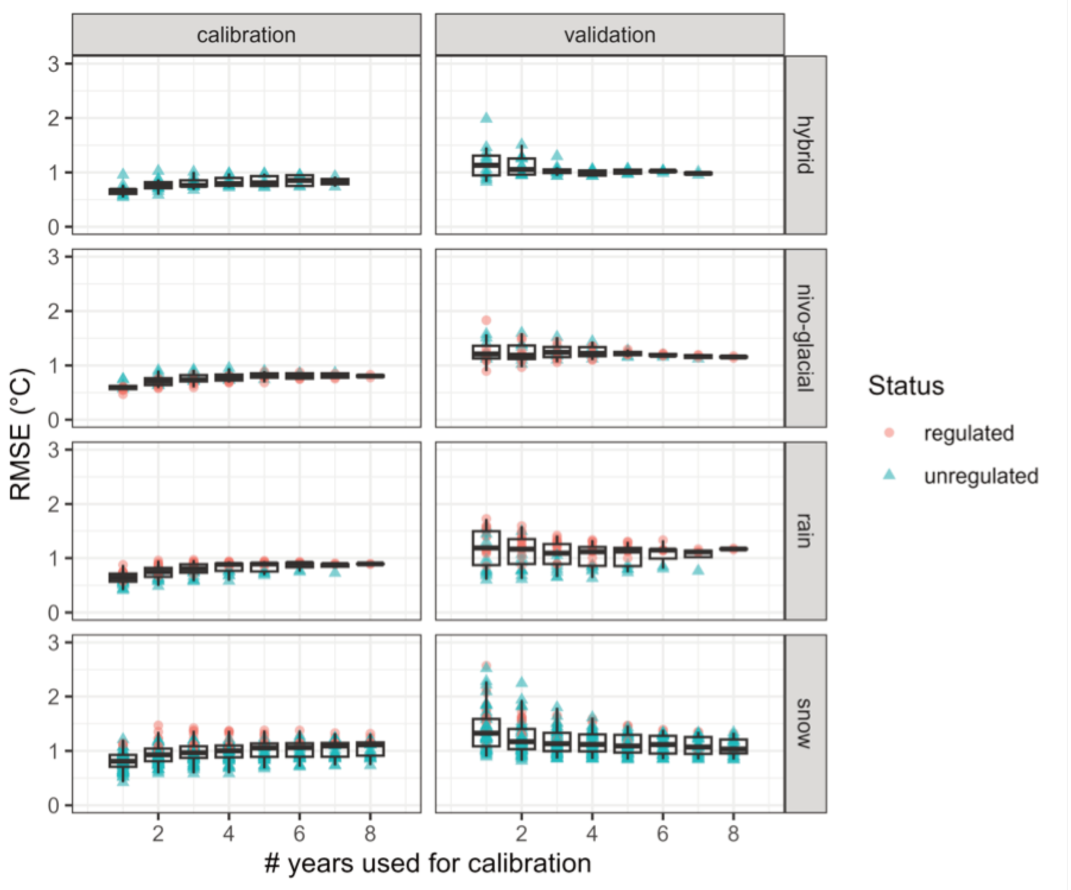

Figure 4: Boxplots showing root-mean-square error (RMSE) in the calibration and validation periods as a function of the number of years used to calibrate Air2stream for the case in which the validation period was fixed to the years 2021 and 2022. See Figure 2 for a description of the boxplots.

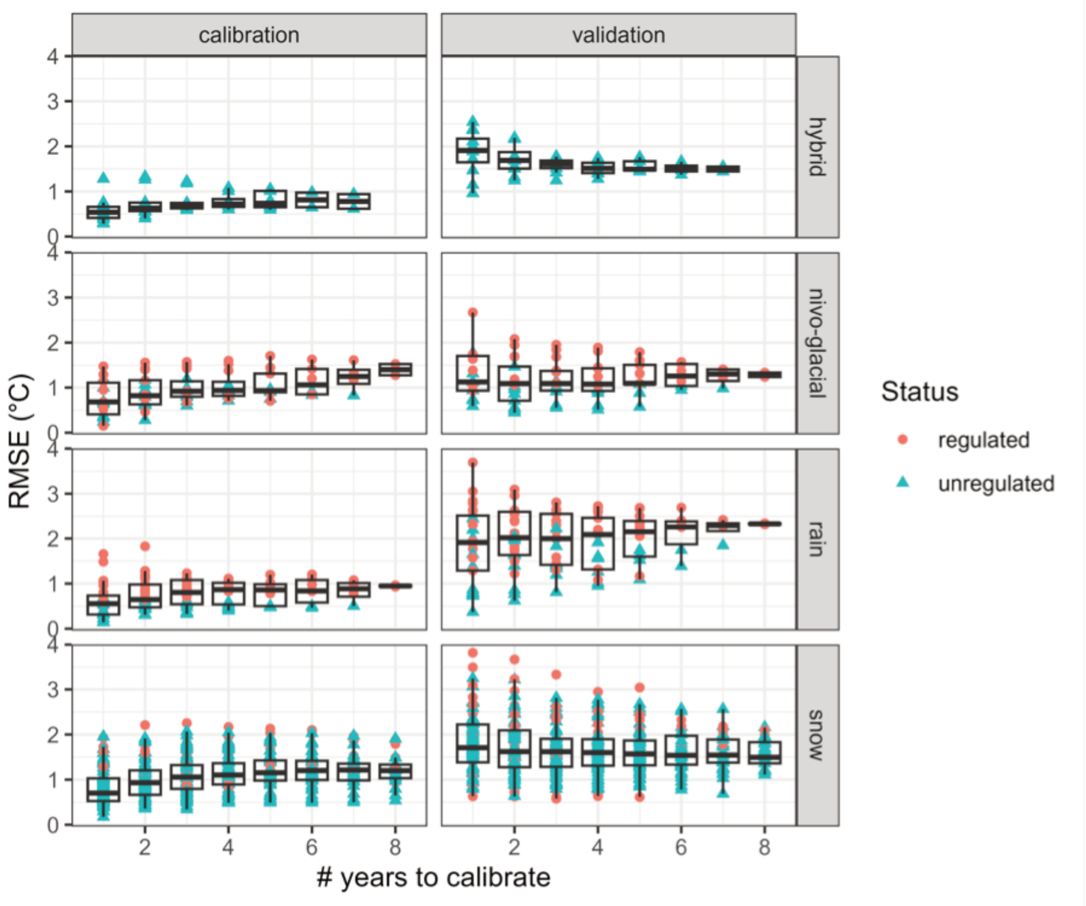

Figure 5: Boxplots showing root-mean-square error (RMSE) for the 2021 heat-wave and the heatwave dates during the calibration period as a function of the number of years used to calibrate Air2stream for the case in which the validation period was fixed to the years 2021 and 2022. See Figure 2 for a description of the boxplots.

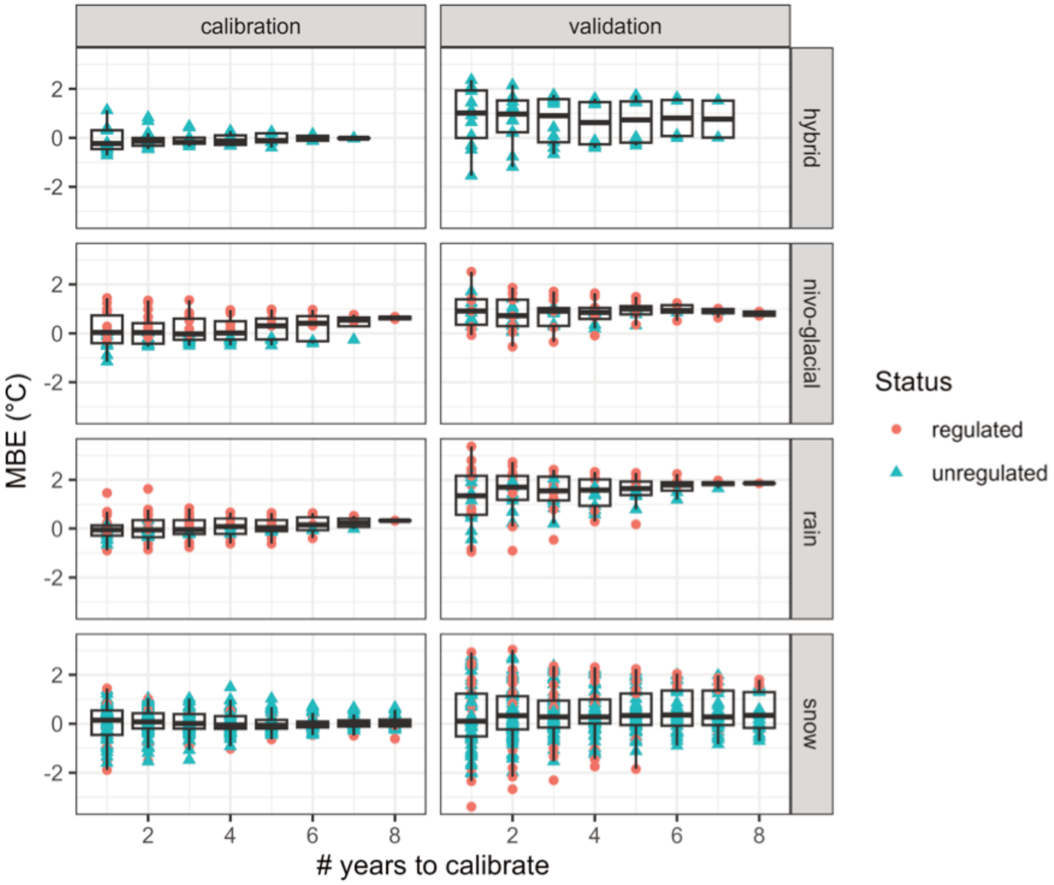

Figure 6: Boxplots showing mean bias error (MBE) for the 2021 heatwave and the heatwave dates during the calibration period as a function of the number of years used to calibrate Air2stream for the case in which the validation period was fixed to the years 2021 and 2022. See Figure 2 for a description of the boxplots.

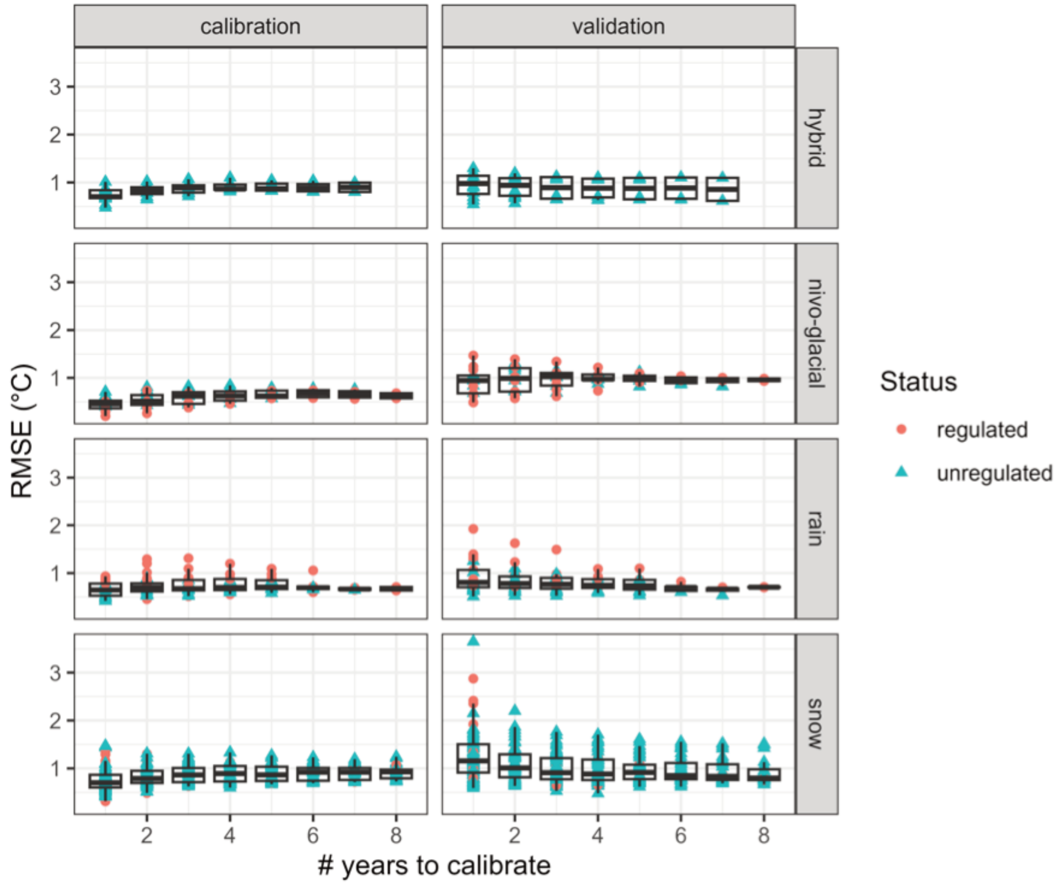

Figure 7: Boxplots showing root-mean-square error (RMSE) for the 2022 autumn drought period and the autumn drought dates during the calibration period as a function of the number of years used to calibrate Air2stream for the case in which the validation period was fixed to the years 2021 and 2022. See Figure 2 for a description of the boxplots.

Moore & Callahan. (2026). Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period. http://confluence-jwsm.ca/index.php/jwsm/article/view/71/8 . doi:10.22230/jwsm.2026v8n1a71

was more variable and generally higher for both calibration and validation periods than when considering performance over a full annual cycle (Figure 5). Air2stream tended to be positively biased during the heatwave for hybrid, nivo-glacial, and rain-dominated systems and less biased for snow-dominated systems (Figure 6). In terms of both *RMSE* and *MBE*, a calibration period of three years generated similar performance to that provided by longer calibration periods. Considering just the autumn drought period (September and October), Air2stream's performance was similar to and even slightly better than when considering the full year (Figure 7), and bias was lower than for the heatwave period (Figure 8).

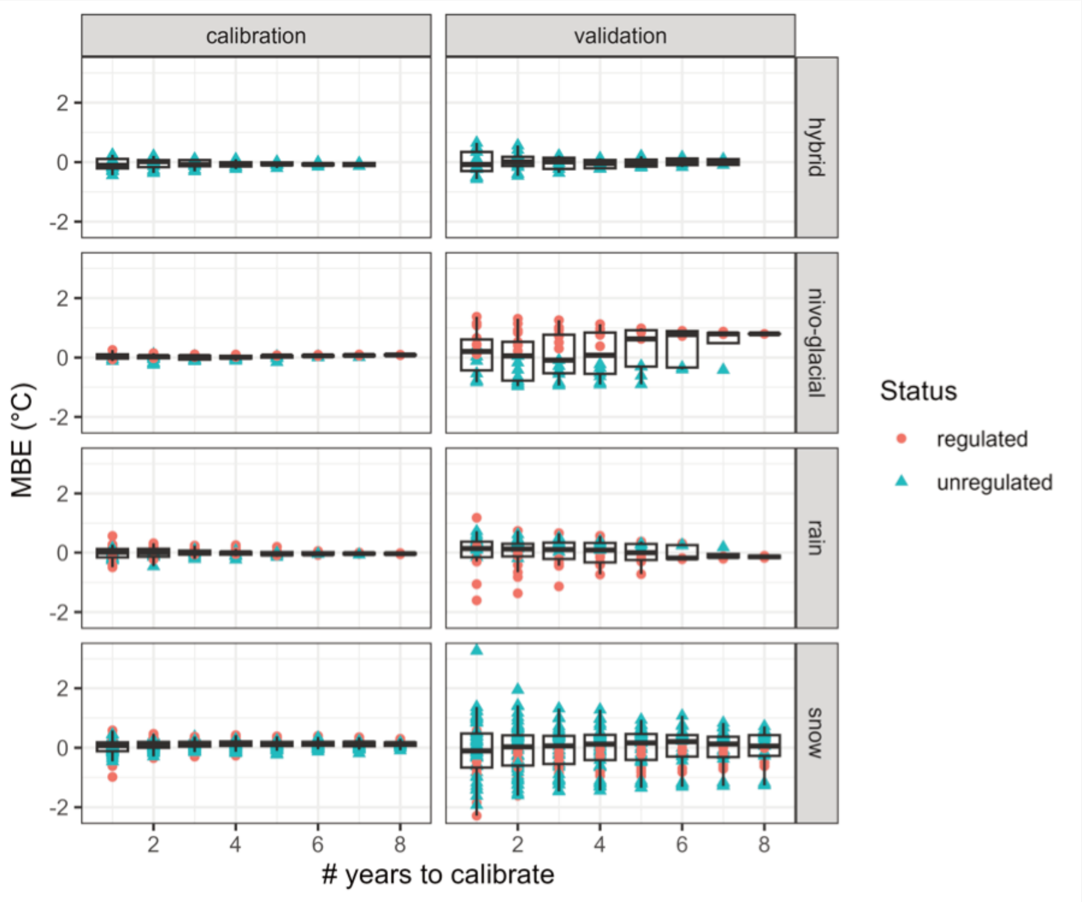

Figure 8: Boxplots showing mean bias error (MBE) for the 2022 autumn drought period and the autumn drought dates during the calibration period as a function of the number of years used to calibrate Air2stream for the case in which the validation period was fixed to the years 2021 and 2022. See Figure 2 for a description of the boxplots.

Discussion

The *RMSE* value for the calibration period tended to increase with the length of the calibration period, and decreased in the validation period. This finding illustrates the tradeoff between minimizing errors in the calibration period at the expense of the model's ability to extrapolate to different conditions (i.e., overfitting), which is more severe with shorter calibration periods.

The analyses demonstrate that, in most cases, a three-year calibration period yields performance comparable to that of a longer calibration period, even for extreme weather events such as the 2021 heatwave and the 2022 autumn drought. However, it should be noted that *RMSE* was higher when considering just the heatwave date range (June 25 to July 7) than the full annual cycle. The results suggest that, for many catchments, just two years would be sufficient to keep *RMSE* below 2°C in the validation period when considering the full annual cycle.

The calibration approach applied in this study was restricted to using subsets of sequential years. This approach was used as an alternative to iterating through all possible subsets of calibration periods of different lengths, which would have required substantially more computational time. The sequential-years sampling approach is consistent with the fact that many temporary monitoring programs sample contiguous blocks of time. An implication of the sequential-years approach is that the sampled period used for calibration may not represent a broad diversity of hydrometeorological conditions (e.g., if all sampled years lie within one phase of El Niño-Southern Oscillation or Pacific

Moore & Callahan. (2026). Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period. http://confluence-jwsm.ca/index.php/jwsm/article/view/71/8 . doi:10.22230/jwsm.2026v8n1a71

Decadal Oscillation) and thus may be biased relative to conditions outside the calibration period. It would be useful to assess the robustness of calibrations to varying hydrometeorological conditions using a differential split-sample approach, as described by Klemeš (1986)—for example, calibrate the model using years with cool summers and validate for years with warm summers.

Air2stream tended to perform worst for catchments with nival regimes. This finding may reflect that streamflow is more variable during the period of high summer temperatures for snowmelt-dominated systems than for other regimes. As a consequence, limitations associated with the simplified representation of hydrological influences on thermal dynamics for nival systems are more likely to manifest as prediction errors than for rain-dominated, hybrid, or nivo-glacial regimes, for which summer flows are less variable.

The dependence of the calibrated parameters on the length of the calibration period was notably less clear than the dependence of *RMSE*. This finding suggests that Air2stream's structure is subject to equifinality: different combinations of parameters can reproduce the same water temperature time series regardless of differences in hydrometeorological conditions. Nevertheless, this equifinality did not appear to hinder the model's validation performance when calibrated using three or more years of data.

The parameter a_4 was calibrated to the lower bound of 0 for most cases, regardless of the length of the calibration period. This parameter controls the extent to which the mean depth of the water column influences the response of water temperature to heat inputs, and a value of 0 indicates no influence. This finding is consistent with results reported by Toffolon and Piccolroaz (2015). Therefore, in the structure of the 8-parameter version of Air2stream, the dominant influence of hydrology appears to be via advection, as represented by parameters a_5 to a_8.

There appeared to be systematic variations in parameter distributions in relation to hydrological regime and regulation status. This result suggests that it may be possible to develop regionalized relationships to generate *a priori* parameter estimates to apply Air2stream to sites lacking sufficient data to enable a robust calibration. Alternatively, these distributions could be used as prior distributions to support Bayesian approaches to calibrating Air2stream. Bayesian approaches provide a formal method for using parameters derived from previous model calibrations to support calibration at a new site (e.g., Reitan & Petersen-Øverleir, 2008). However, this study draws upon a limited set (23) of locations, which leads to limited or no replication when data are subset by hydrologic regime and regulation status. Further analysis with more extensive data sets is required.

A limitation of Air2stream for some management applications is that it simulates the daily mean stream temperature. While daily data are useful for a number of purposes, sub-daily variability can also be important. For example, daily maxima are often critical in relation to many fish species, and the maximum weekly temperature, computed as the annual maximum of a seven-day running average of daily maximum temperatures, is used by many agencies as a water quality parameter (e.g., Province of British Columbia, 2001). Another limitation of Air2stream is that it does not account for changes in catchment characteristics, such as changes in riparian forest cover, which can have a profound effect on stream thermal regimes (Moore & MacDonald, 2024).

Conclusions

The key finding of this study is that a three-year calibration period appears sufficient to generate robust parameter estimates. This finding builds on the work of Piccolroaz et al. (2016), who found that five-year periods were sufficient, but did not assess the effect of calibrating using shorter record lengths. Furthermore, with the exception of nival systems, a calibration period of two years appears to be sufficient to produce *RMSE* values less than 2°C in the validation period when considering the full annual cycle.

There appeared to be some relationships between the distributions of calibrated parameters and the hydrological regime and/or regulation status of the systems. However, the sample of catchments used in this study is too small to draw firm conclusions. Further studies with larger samples should explore this point further. If such relationships are found to exist, at least within a region, then it may be possible to develop regionalized parameter sets to allow Air2stream to be applied at sites lacking

Moore & Callahan. (2026). Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period.
http://confluence-jwsm.ca/index.php/jwsm/article/view/71/8 . doi:10.22230/jwsm.2026v8n1a71

sufficient data for calibration. Such regionalized parameter sets could also be used to specify prior distributions to support Bayesian approaches to calibration, which may be useful when only limited calibration data are available.

Acknowledgements

This research was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant to RDM, as well as funding from the University of British Columbia in the form of graduate student stipends to LC. David Hutchinson, Regional Chief, Pacific and North Hydrometric Operations that provided the stream temperature data. Eric Leinberger produced the map presented as Figure 1. The authors thank the editor and two anonymous reviewers for their constructive comments, which helped improve the manuscript.

Data availability

The data used in this study are publicly available via doi:10.5281/zenodo.14502248. Analysis code is available upon reasonable request to the first author.

References

Albers, S. (2017). Tidyhydat: Extract and tidy Canadian hydrometric data. *The Journal of Open Source Software*, *2*(20), 1–4. doi:10.21105/joss.00511

Callahan, L. (2023). *Predicting daily stream temperature in British Columbia, Canada, using reanalysis air temperature in statistical models and the hybrid air2stream model* [Master's thesis, the University of British Columbia]. https://open.library.ubc.ca/soa/cIRcle/collections/ubctheses/24/items/1.0435697 [May 13, 2026].

Callahan, L., & Moore, R.D. (2025a). Evaluation of the hybrid Air2stream model for simulating daily stream temperature during extreme summer heat wave and autumn drought conditions. *Hydrological Processes*, *39*(1), e70033. doi:10.1002/hyp.70033

Callahan, L., & Moore, R.D. (2025b). *Supporting information for "Evaluation of the hybrid Air2stream model for simulating daily stream temperature during extreme summer heat wave and autumn drought conditions"*. https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fhyp.70033&file=hyp70033-sup-0001-Supinfo.docx

Chen, Y.D., Carsel, R.F., McCutcheon, S.C., & Nutter, W.L. (1998). Stream temperature simulation of forested riparian areas: I. Watershed-scale model development. *Journal of Environmental Engineering*, *124*(4), 304–315. doi:10.1061/(asce)0733-9372(1998)124:4(304)

Cheng, Y., Voisin, N., Yearsley, J. R., & Nijssen, B. (2020). Reservoirs modify river thermal regime sensitivity to climate change: A case study in the southeastern United States. *Water Resources Research*, *56*(6), e2019WR025784. doi:10.1029/2019wr025784

Chu, C., Jones, N.E., & Allin, L. (2009). Linking the thermal regimes of streams in the Great Lakes Basin, Ontario, to landscape and climate variables. *River Research and Applications*, *26*(3), 221–241. doi:10.1002/rra.1259

Collins, S.G., Pelto, B.M., Callahan, L.M., Friele, P., & Moore, R.D. (2025). Quantifying the effect of catchment snow cover on stream temperature dynamics in a mountainous region. *Hydrological Processes*, *39*(9), e70261. doi:10.1002/hyp.70261

Copernicus. (2024). *Climate data store*. https://cds.climate.copernicus.eu/how-to-api [May 13, 2026].

Daigle, A., Boyer, C., & Légaré, A. (2023). Modeling of the thermal regime of rivers subject to seasonal ice cover using data from different sources and temporal resolutions. *Canadian Water Resources Journal / Revue Canadienne Des Ressources Hydriques*, *48*(2), 132–148. doi:10.1080/07011784.2022.2140604

Déry, S.J., Stahl, K., Moore, R.D., Whitfield, P.H., Menounos, B., & Burford, J.E. (2009). Detection of runoff timing changes in pluvial, nival, and glacial rivers of western Canada. *Water Resources Research*, *45*(4). doi:10.1029/2008wr006975

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

Ficklin, D.L., Luo, Y., Stewart, I.T., & Maurer, E.P. (2012). Development and application of a hydroclimatological stream temperature model within the Soil and Water Assessment Tool. *Water Resources Research*, *48*, W01511. doi:10.1029/2011wr011256

Fleming, S.W., Whitfield, P.H., Moore, R.D., & Quilty, E. (2007). Regime-dependent streamflow sensitivities to Pacific climate modes cross the GeorgiaPuget transboundary ecoregion. *Hydrological Processes*, *21*(24), 3264–3287. doi:10.1002/hyp.6544

Fuller, M.R., Detenbeck, N.E., Leinenbach, P., Labiosa, R., & Isaak, D. (2025). Scenario planning management actions to restore cold water stream habitat: Comparing mechanistic and statistical modeling approaches. *River Research and Applications*, *41*(2), 382–401. doi:10.1002/rra.4381

Fullerton, A.H., Sun, N., Baerwalde, M.J., Hawkins, B.L., & Yan, H. (2022). Mechanistic simulations suggest riparian restoration can partly counteract climate impacts to juvenile salmon. *JAWRA Journal of the American Water Resources Association*, *58*(4), 525–546. doi:10.1111/1752-1688.13011

Government of Canada. (2024). *HydrometricNetworkBasinPolygons.* https://collaboration.cmc.ec.gc.ca/cmc/hydrometrics/www/HydrometricNetworkBasinPolygons [May 13, 2026].

Hijmans, R. ., & van Etten, J. (2012). *Raster: Geographic analysis and modeling with raster data*. http://CRAN.R-project.org/package=raster [May 13, 2026].

Isaak, D.J., Luce, C.H., Rieman, B.E., Nagel, D.E., Peterson, E.E., Horan, D.L., Parkes, S., & Chandler, G.L. (2010). Effects of climate change and wildfire on stream temperatures and salmonid thermal habitat in a mountain river network. *Ecological Applications*, *20*(5), 1350–1371. doi:10.1890/09-0822.1

Isaak, D.J., Wenger, S.J., Peterson, E.E., Ver Hoef, J.M., Nagel, D.E., Luce, C. H., Hostetler, S.W., Dunham, J.B., Roper, B.B., Wollrab, S.P., Chandler, G.L., Horan, D.L., & Parkes-Payne, S. (2017). The NorWeST summer stream temperature model and scenarios for the western U.S.: A crowd-sourced database and new geospatial tools foster a user community and predict broad climate warming of rivers and streams. *Water Resources Research*, *53*, 9181–9205. doi:10.1002/2017WR020969

Islam, S.U., Hay, R.W., Déry, S.J., & Booth, B.P. (2019). Modelling the impacts of climate change on riverine thermal regimes in western Canada's largest Pacific watershed. *Scientific Reports*, *9*(1), 11398. doi:10.1038/s41598-019-47804-2

Jackson, C.R., Sturm, C.A., & Ward, J.M. (2001). Timber harvest impacts on small headwater stream channels in the Coast Ranges of Washington. *JAWRA Journal of the American Water Resources Association*, *37*(6), 1533–1549. doi:10.1111/j.1752-1688.2001.tb03658.x

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

Johnson, M.F., Albertson, L.K., Algar, A.C., Dugdale, S., Edwards, P., England, J., Gibbins, C., Kazama, S., Komori, D., MacColl, A.D.C., Scholl, E.A., Wilby, R.L., de Oliveira Roque, F., & Wood, P.J. (2024). Rising water temperature in rivers: Ecological impacts and future resilience. *WIREs Water, 11*(4), e1724. doi:10.1002/wat2.1724

Jones, L.A., Muhlfeld, C.C., & Hauer, F.R. (2017). Temperature. In *Methods in stream ecology — Volume 1: Ecosystem structure* (pp. 109–120). Elsevier Science Publishing Co.

Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal*, *31*(1), 13–24. doi:10.1080/02626668609491024

Leach, J.A., Kelleher, C., Kurylyk, B.L., Moore, R.D., & Neilson, B.T. (2023). A primer on stream temperature processes. *WIREs Water*, *10*(4), 1–18. doi:10.1002/wat2.1643

Lee, S.-Y., Fullerton, A.H., Sun, N., & Torgersen, C.E. (2020). Projecting spatiotemporally explicit effects of climate change on stream temperature: A model comparison and implications for coldwater fishes. *Journal of Hydrology*, *588*, 125066. doi:10.1016/j.jhydrol.2020.125066

MacDonald, R., Boon, S., & Byrne, J.M. (2014). A process-based stream temperature modelling approach for mountain regions. *Journal of Hydrology*, *511*, 920–931. doi:10.1016/j.jhydrol.2014.02.009

McEwan, A.L., Johnson, C.J., Todd, M., & Govindarajulu, P. (2021). Resource selection and movement of the coastal tailed frog in response to forest harvesting. *Forest Ecology and Management*, *497*, 119448. doi:10.1016/j.foreco.2021.119448

Moore, R.D. (1993). Application of a conceptual streamflow model in a glacierized drainage basin. *Journal of Hydrology*, *150*(1), 151–168. doi:10.1016/0022-1694(93)90159-7

Moore, R.D., Fleming, S.W., Menounos, B., Wheate, R., Fountain, A., Stahl, K., Holm, K., & Jakob, M. (2009). Glacier change in western North America: Influences on hydrology, geomorphic hazards and water quality. *Hydrological Processes*, *23*(1), 42–61. doi:10.1002/hyp.7162

Moore, R.D., & MacDonald, R.J. (2024). Quantifying the influence of forestry and forest disturbance on stream temperature: Methodologies and challenges. *Hydrological Processes*, *38*(7), e15223. doi:10.1002/hyp.15223

Moore, R.D., Nelitz, M., & Parkinson, E. (2013). Empirical modelling of maximum weekly average stream temperature in British Columbia, Canada, to support assessment of fish habitat suitability. *Canadian Water Resources Journal*, *38*(2), 135–147. doi:10.1080/07011784.2013.794992

Moore, R.D., Pelto, B., Menounos, B., & Hutchinson, D. (2020). Detecting the effects of sustained glacier wastage on streamflow in variably glacierized catchments. *Frontiers in Earth Science*, *8*, 136. doi:10.3389/feart.2020.00136

Nelitz, M.A., MacIsaac, E.A., & Peterman, R.M. (2007). A science-based approach for identifying temperature-sensitive streams for rainbow trout. *North American Journal of Fisheries Management*, *27*(2), 405–424. doi:10.1577/M05-146.1

Parkinson, E.A., Lea, E.V., Nelitz, M.A., Knudson, J.M., & Moore, R.D. (2016). Identifying temperature thresholds associated with fish community changes in British Columbia, Canada, to support identification of temperature sensitive streams. *River Research and Applications*, *32*, 330–347. doi:10.1002/rra.2867

Piccolroaz, S. (2017). *air2stream*. https://github.com/spiccolroaz/air2stream [May 13, 2026].

Piccolroaz, S., Calamita, E., Majone, B., Gallice, A., Siviglia, A., & Toffolon, M. (2016). Prediction of river water temperature: A comparison between a new family of hybrid models and statistical approaches. *Hydrological Processes*, *30*(21), 3901–3917. doi:10.1002/hyp.10913

Piotrowski, A.P., & Napiorkowski, J.J. (2018). Performance of the air2stream model that relates air and stream water temperatures depends on the calibration method. *Journal of Hydrology*, *561*, 395–412. doi:10.1016/j.jhydrol.2018.04.016

Plotly Technologies Inc. (2015). *Collaborative data science*. https://plot.ly [May 13, 2026].

Province of British Columbia. (2001). *Water quality guidelines for temperature*. https://www2.gov.bc.ca/assets/gov/environment/air-land-water/water/waterquality/water-quality-guidelines/approved-wqgs/temperature-or.pdf [May 13, 2026].

Quilty, E., & Moore, R.D. (2010). Physical water quality — stream temperature. In R.G. Pike, T.E. Redding, R.D. Moore, R.D. Winkler, and K.D. Bladon (Eds.), *Compendium of forest hydrology and geomorphology in British Columbia. Land management handbook 66* (pp. 606–612). B.C. Ministry of Forest and Range, Forest Science Program, Victoria, B.C. and FORREX Forum for Research and Extension in Natural Resources. https://www.for.gov.bc.ca/hfd/pubs/Docs/Lmh/Lmh66.htm [May 13, 2026].

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/ [May 13, 2026].

Reitan, T., & Petersen-Øverleir, A. (2008). Bayesian power-law regression with a location parameter, with applications for construction of discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, *22*(3), 351–365. doi:10.1007/s00477-007-0119-0

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

Sowder, C., & Steel, E.A. (2012). A note on the collection and cleaning of water temperature data. *Water*, *4*(3), 597–606. doi:10.3390/w4030597

Stahl, K., Moore, R.D., Floyer, J.A., Asplin, M.G., & McKendry, I.G. (2006). Comparison of approaches for spatial interpolation of daily air temperature in a large region with complex topography and highly variable station density. *Agricultural and Forest Meteorology*, *139*(3), 224–236. doi:10.1016/j.agrformet.2006.07.004

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

Sun, N., Yearsley, J., Voisin, N., & Lettenmaier, D.P. (2015). A spatially distributed model for the assessment of land use impacts on stream temperature in small urban watersheds. *Hydrological Processes*, *29*(10), 2331–2345. doi:10.1002/hyp.10363

Toffolon, M., & Piccolroaz, S. (2015). A hybrid model for river water temperature as a function of air temperature and discharge. *Environmental Research Letters*, *10*(11), 114011. doi:10.1088/1748-9326/10/11/114011

Trubilowicz, J.W., Shea, J.M., Jost, G., & Moore, R.D. (2016). Suitability of North American Regional Reanalysis (NARR) output for hydrologic modelling and analysis in mountainous terrain. *Hydrological Processes*, *30*(13), 2332–2347. doi:10.1002/hyp.10795

van Vliet, M.T.H., Yearsley, J.R., Franssen, W.H.P., Ludwig, F., Haddeland, I., Lettenmaier, D.P., & Kabat, P. (2012). Coupled daily streamflow and water temperature modelling in large river basins. *Hydrology and Earth System Sciences*, *16*(11), 4303–4321. doi:10.5194/hess-16-4303-2012

Wehrly, K.E., Brenden, T.O., & Wang, L. (2009). A comparison of statistical approaches for predicting stream temperatures across heterogeneous landscapes. *JAWRA Journal of the American Water Resources Association*, *45*(4), 986–997. doi:10.1111/j.1752-1688.2009.00341.x

Weller, J.D., Moore, R.D., & Iacarella, J.C. (2023). Stream thermalscape scenarios for British Columbia, Canada. *Canadian Water Resources Journal*, *49*, 1–20. doi:10.1080/07011784.2023.2267028

Zhu, S., Luo, Y., Graf, R., Wrzesiński, D., Sojka, M., Sun, B., Kong, L., Ji, Q., & Luo, W. (2022). Reconstruction of long-term water temperature indicates significant warming in Polish rivers during 1966–2020. *Journal of Hydrology: Regional Studies*, *44*, 101281. doi:10.1016/j.ejrh.2022.101281

Moore & Callahan. (2026). Sensitivity of Air2stream Stream Temperature Prediction Accuracy to the Length of the Calibration Period.
http://confluence-jwsm.ca/index.php/jwsm/article/view/71/8 . doi:10.22230/jwsm.2026v8n1a71