Ecological Indicators 122 (2021) 107229

Contents lists available at ScienceDirect

Ecological Indicators

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

Heed the data gap: Guidelines for using incomplete datasets in annual stream temperature analyses

Zachary C. Johnson [a,*], Brittany G. Johnson [b], Martin A. Briggs [c], Craig D. Snyder [d], Nathaniel P. Hitt [d], Warren D. Devine [e]

[a] Civil and Environmental Engineering, University of Washington, 3760 E Stevens Way NE, Seattle, WA 98195, USA
[b] School of Environmental and Forest Sciences, University of Washington, 3715 W Stevens Way NE, Seattle, WA 98195, USA
[c] U.S. Geological Survey, Earth System Processes Division, Hydrogeophysics Branch, 11 Sherman Place, Unit 5015, Storrs, CT 06269, USA
[d] U.S. Geological Survey, Leetown Science Center, 11649 Leetown Road, Kearneysville, WV 25430, USA
[e] Washington State, Department of Natural Resources, Forest Resources Division, 1111 Washington Street SE, Olympia, WA 98504, USA

ARTICLE INFO

Keywords:
Stream temperature monitoring
Thermal regimes
Sine-wave linear regression
Missing data
Paired air-water temperature
Data imputation

ABSTRACT

Stream temperature data are useful for deciphering watershed processes important for aquatic ecosystems. Accurately extracting signal trends from stream temperature is essential for predicting responses of environmental and ecological indicators to change. Missing data periods are common for various reasons, and pose a challenge for scientists using temperature signal analysis to support stream research and ecological management objectives. However, the sensitivity of estimated temperature signal patterns to missing data has not been thoroughly evaluated, despite the potentially large impact on interpretation. In this study, we explored the effects of simulated missing daily data on the characterization of annual water temperature signals measured at headwater sites in the Pacific Northwest and Mid-Atlantic regions of the USA. For each site, we used linear regressions of sine-waves fitted to complete (365-d) and partial (7–357 consecutive missing data points) annual datasets of daily mean water temperature and computed three thermal parameters (mean, phase, and amplitude), which together can indicate thermally and ecologically influential watershed processes (e.g., depth and magnitude of groundwater discharge). Expected values (derived from complete datasets) ranged from 7.0 to 12.6 °C, 205 to 254 d, and 1.9 to 9.5 °C for annual mean, phase, and amplitude, respectively. While annual phase and amplitude could be accurately estimated (i.e., within 95–99% confidence intervals of expected values) with up to approximately two months of consecutively missing data, annual mean temperature required more complete datasets. We found that datasets with less than seven weeks of consecutively missing data enabled estimation of all annual signal parameters with reasonable accuracy (>75% probability of being within the 95–99% confidence intervals of expected values). Imputation of missing data expanded this range to approximately 20 weeks, with the greatest improvements in parameter estimation between 9 and 27 weeks of imputed missing data. However, caution should be exercised when applying this technique. For example, imputation improved the accuracy of parameter estimation for most sites, but accuracy decreased for some sites exhibiting strong groundwater influence. The timing of consecutive missing data points within a year had inconsistent effects on annual thermal parameter estimates among regions, years, and individual parameters. Utilizing sites with more than approximately seven consecutive weeks of missing data or 20 weeks of imputed data increases the probability of mischaracterization of annual stream thermal regimes. Understanding this limitation is vital for identifying the potential of streams to serve as climate refugia for ecological indicator species and effective future management of stream systems.

1. Introduction

Temperature influences most of the physical, chemical, and biological processes of the stream corridor, determining habitat suitability for stream organisms (including important ecological indicator species) and driving patterns of fish development and metabolism and

* Corresponding author at: U.S. Geological Survey, Washington Water Science Center, 934 Broadway, Suite 300, Tacoma, WA 98402, USA.
E-mail address: zcjohnsonpubs@gmail.com (Z.C. Johnson).

https://doi.org/10.1016/j.ecolind.2020.107229
Received 9 June 2020; Received in revised form 24 November 2020; Accepted 27 November 2020
Available online 15 December 2020

biological production and decomposition. Channel water temperature is controlled by a balance of meteorological, hydrogeological, and biological factors, such as incoming solar radiation, air temperature, wind speed, humidity, stream channel dimensions (depth and width), flow volume, groundwater (GW) inputs (Anderson, 2005; CAISSIE, 2006; Cluis, 1972; Gallice et al., 2015; Kurylyk et al., 2019; Webb et al., 2008), and riparian shading (Dugdale et al., 2018). However, climate change and other anthropogenic alterations to watersheds are substantially transforming the natural thermal regime of streams and rivers (Arismendi et al., 2013; CAISSIE, 2006; Hrachowitz et al., 2010; Steel et al., 2017). Therefore, understanding the effect of these alterations on stream temperature regimes is essential for the effective management of stream ecosystems.

Recent advances in technology and techniques, such as remote sensing and inexpensive digital data loggers, have increased the capacity to monitor and predict stream temperature at local to global scales. These burgeoning data represent a unique opportunity to explore trends and controlling factors of stream temperature across spatial and temporal scales. However, satellite-based thermal infrared is not applicable to most headwater streams, and most in-situ stream temperature datasets are still collected seasonally, usually in summer, or over other limited periods of time (Arismendi et al., 2013; Leach and Moore, 2014). Long-term trends (i.e., annual and longer) in water temperature are driven by many interacting watershed processes that frequently vary over relatively short time scales (i.e., daily to seasonal) such as air temperature, GW influence, and surface water flow regimes. In order to understand and predict long-term trends, recent studies have established the importance of monitoring stream temperature for full annual cycles (Marsha et al., 2018; Steel et al., 2012), and paired annual air and water temperature signal patterns can indicate important watershed processes and climate change resilience of stream systems (Briggs et al., 2018a; Johnson et al., 2020).

Heat has long been used as a tracer of various watershed processes (Anderson, 2005; Constantz, 2008), and many statistics and metrics specific to study objectives are calculated in the literature to explore patterns in the thermal signals, most often only during the summer period. In total, these metrics cover the various facets of the annual stream thermal regime—including magnitude, variability, frequency, duration, and timing—but in many cases are highly correlated with each other (Arismendi et al., 2013; Dunham et al., 2005; Jones and Schmidt, 2018). In contrast, only recently have studies begun to explore patterns in distinct components of the annual temperature signal—such as differences in the mean, phase, and amplitude—or their relation to watershed processes (Maheu et al., 2016a). Together, these properties can indicate processes such as thermal inertia (Letcher et al, 2016), influence of GW discharge and source depth (Briggs et al., 2018a, 2018b; Johnson et al., 2020), riparian shading (Johnson and Jones, 2000; Wondzell et al., 2019), or upstream dam presence (Buccola et al., 2016; Kędra and Wiejaczka, 2018; Maheu et al., 2016b). In a recent study, Johnson et al. (2020) used sine-wave linear regressions to investigate local thermal drivers at local to continental scales. In that study, the authors found that local controls were important for watershed- to regional-scale patterns in stream temperature and that annual signal patterns are relatively stable for periods less than a decade. Exploration of these annual patterns across more locations and time periods would provide a broader understanding of the drivers of stream temperature dynamics.

The issue of missing data is already a common problem for stream temperature datasets (Letcher et al., 2016; Li et al., 2017; McNyset et al., 2015; Sowder and Steel, 2012) and can arise from a myriad of events such as logger failure, stream freezing and drying, dislocation of loggers due to high flows, human interference, or simply limited measurement time periods and locations. Unfortunately, there are currently no guidelines for determining the proportion of missing data points acceptable to maintain robust characterization of annual stream temperature signals. Some studies have defined an “open-water” period (i.e., $<$ 365 d) for regressions of seasonally frozen or uncoupled air-water temperature periods (Letcher et al., 2016; Maheu et al., 2016b), but these regressions are fundamentally different from full annual (365 d) regressions and the comparisons of coefficient values and subsequent watershed process interpretation and prediction may not be appropriate. Other studies have recommended incorporating spatial covariance structures to account for missing data in stream temperature records (Bal et al., 2014; Letcher et al., 2016), but such approaches may underperform if geophysical covariates lack inference for local GW processes (e.g., Snyder et al., 2015) and predictive error from missing data would be expected to increase with spatial extent.

Spatiotemporal varying coefficient model techniques that emphasize temporal correlation over spatial correlation (Li et al., 2017) may represent an improvement in this respect. Imputation (replacement with estimate) of missing values with multivariate data analysis (Josse et al., 2016) is another recently developed technique that shows promise and has been used in large regional studies of stream temperature (Isaak et al., 2018, 2020). However, this technique’s predictions are dependent on temporal correlations with data from surrounding sites having more continuous records that may not be available and may not share the same annual thermal characteristics as the site with missing data. Even where available, it is currently unknown how much missing data the imputation technique could predict before accuracy deteriorates.

The objective of this study was to systematically explore the effect of the proportion and timing of missing data points on the accuracy of annual thermal parameters for stream temperature estimated using sine-wave linear regression. We also evaluated the performance of a data imputation technique on improving the accuracy of estimated parameters. We used spatially intensive data collected from two disparate regions of the USA as case studies. This research outlines methods for researchers to make case-specific decisions regarding the appropriate level and timing of missing data for long-term signal modeling of stream temperature using common methodology. These techniques could also be applied to other stream temperature analyses dealing with missing data points.

2. Methods and site descriptions

We explored the effect of missing daily data on regressed annual (365 d) stream temperature signals from field data collected in mountainous headwater basins in the Pacific Northwest and Mid-Atlantic regions of the USA. Sine-wave linear regressions were used to represent annual air and water temperature signals (Briggs et al., 2018a; Johnson et al., 2020). The effects of the amount and timing of missing data periods on the accuracy of annual thermal parameters (inferred based upon differences with the full annual record) were evaluated and an approach for determining the appropriate level of missing data is outlined. The regional datasets used were chosen because of their robustness and fine spatial grain of sampling ($>$1 site per 2,000 ha and $>$1 site per 650 ha, respectively), which allowed for evaluations of the influence of local processes such as GW intrusion on thermal characteristics. The following subsections contain a brief description of the sites utilized for this study. Detailed site information and maps can be found in Johnson et al. (2020).

2.1. Pacific Northwest: Olympic Experimental State Forest (OESF)

The OESF is a 523,000-ha mostly forested planning area that contains 110,000 ha of state trust lands on the Olympic Peninsula in western Washington, USA (Martens et al., 2019). The 56 flow-disconnected OESF subwatersheds considered in this study (Johnson et al., 2020) fall within the Coast Range level III ecoregion (Omernik, 1987; Omernik and Griffith, 2014). Paired air and water hourly temperature data were collected at these sites between 1 October 2012 and 31 December 2018. Hourly data used were converted to daily mean values. Elevation in the OESF ranges from sea level to 1,155 m above mean sea level (a.m.s.l.).

Annual precipitation in this maritime climate ranges between 203 and 355 cm with the majority falling as rain during the autumn and winter months (approximately October to March). Basin area for the 56 sites ranges from 15 to 789 ha.

2.2. Mid-Atlantic: Shenandoah National Park (SHEN)

SHEN is a 77,700-ha mostly forested protected area located along the spine of the Blue Ridge Mountains in northern Virginia, USA. The sites considered in this study are within the Blue Ridge level III ecoregion (Omernik, 1987; Omernik and Griffith, 2014) and include 120 flow-connected subwatersheds within 18 flow-disconnected watersheds (Johnson et al., 2020). Hourly water temperature data was collected at all 120 sites, whereas hourly air temperature were collected at a subset of the sites (n = 27) and modeled across sites using latitude and elevation as predictors (Johnson et al., 2017). For this study, all hourly data used were converted to daily mean values and collected between 23 June 2012 and 30 September 2016. The details of the temperature data collection can be found in Snyder et al. (2015) and Johnson et al. (2017). SHEN ranges in elevation from 162 to 1235 m a.m.s.l. and receives an average annual precipitation amount of 100 to 150 cm (Jastram et al., 2013) with a majority falling as rain at all but the highest elevations. SHEN experiences greater seasonal changes in air temperatures but smaller seasonal changes in precipitation than OESF (Johnson et al., 2020). Basin area for the 120 sites ranges from 27 to 3,628 ha.

2.3. Annual sine-wave modeling

To calculate the annual signal of mean daily air and water temperatures, we followed the method outlined in Johnson et al., (2020). In short, the annual signal (365 d) of mean daily air and water temperatures for each site were approximated as sine waves using only the 1st harmonic (Johnson et al., 2020; Kothandaraman, 1971) and linearized using Ptolemy's theorem:

$$T(t) = a^{*}sin(\omega t) + b^{*}cos(\omega t) + T_0 + \in(t) \quad (1)$$

where T is daily mean temperature (°C), a and b are the regression coefficients, ω is the angular frequency (rad d^{-1}), t is time (d), T_0 is the annual mean temperature (°C), and ∈ is the error term (°C). The regression coefficient a is equal to $A^{*}cos(\phi)$ and b is equal to $A^{*}sin(\phi)$, where A is the amplitude (°C) and ϕ is the phase (rad). The annual temperature signal amplitude is then determined as $\sqrt{a^2 + b^2}$ and phase is equal to $tan^{-1}(b/a)$. Calculating the day of the year corresponding to the annual maximum temperature varies with the annual window defined. For more information see Appendix A, but for this study the following shift in phase was applied for the calendar year (1 January to 31 December): $3\pi/2 - tan^{-1}(b/a)$. Phase was converted from radians into days by multiplying by $365/2\pi$.

The amplitude, phase, and mean represent the core parameters of a sine-wave signal—the variation, horizontal displacement, and vertical displacement, respectively. Comparing these core parameters between air and water temperature simultaneously has been shown to indicate various watershed processes, such as relative GW contribution to streamflow and effective depth of that contribution (Johnson et al., 2020). This comparison is achieved by calculating three combined air–water annual thermal metrics: amplitude ratio, phase lag, and mean ratio (Briggs et al., 2018a; Johnson et al., 2020). Amplitude ratio (A_R) is the water amplitude (A_w) divided by the air amplitude (A_a). Phase lag (Δϕ) was calculated as the difference between water temperature phase (ϕ_w) and air temperature phase (ϕ_a). Mean ratio (M_R) is the water mean temperature ($T_{0,w}$) divided by the air mean temperature ($T_{0,a}$).

2.4. Simulation of missing data

The decision for what missing data threshold to use, above which the corresponding time series would be excluded from further analysis, should be made on a case-by-case basis, based on the type of signal processing method used and acceptable level of accuracy. To determine an appropriate threshold of missing data for our temperature datasets using linear regression to derive annual temperature signal parameters, a common method for hydroecological stream temperature characterization and prediction, the following steps were taken. From each region, 20 sites were randomly selected (40 sites total) from a subset of sites containing less than 10% missing data among the original records within an approximately two year long time range (23 September 2016 to 19 September 2018 for OESF and 5 August 2014 to 21 September 2016 for SHEN). The value of 20 sites in each region was chosen as it was the minimum sample size required to include a representative range of A_R, Δϕ, and M_R values for each region, which were calculated from preliminary regressions that did not allow for any missing data.

Within a specific annual time window, each site's temperature record was subjected to a sequence of missing data points, m, from 0 to 357 days with 7 day steps (total of 52 m values), which was used to develop a series of regressions. We assumed that missing data on consecutive days are more likely to be representative of actual temperature datasets (e.g., missing data resulting from temperature logger failure, stream freezing or drying, or dislocation of loggers due to high flows and human interference) than if the missing data were randomly distributed throughout time. Therefore, in this study, we focus on consecutively missing data. For completeness, these same procedures were also tested using a random distribution of missing data points, but those results (Appendix B) are not the focus of the current study.

For each "site-m-annual window" combination (total of 55,120 for OESF and 63,440 for SHEN), we generated 100 unique regression models that were distinguished by a different randomly selected start date for missing data (t_r) within the given annual window. Temperature values for dates from t_r to $t_r + m$-1 were then removed. This process was iterated within the regional two-year time range for each site-m pairing by moving the annual time window seven days forward and repeating the process described above for all annual windows completely within the regional time range. Once the two-year time range iterations were completed for a site-m pairing, the next m value in the sequence was assigned and the entire site-m-annual window process was repeated. These steps were then repeated for each site.

We used the *missMDA* package (version 1.17) in R (Josse and Husson, 2016) to test the performance of data imputation of missing values on improving annual parameter estimates. The above steps were followed for four of the annual windows with starting dates separated by 13 weeks in each region, but imputation of simulated missing data for a given site-m-annual window combination was inserted prior to running the regression models. Specifically, principal component analysis was used to impute missing data for a given site utilizing the observed datasets from the other 19 regional sites (OESF and SHEN sites were handled separately). Imputation was conducted using the *imputePCA* function, where the number of dimensions was set by the results of running the *estim_ncpPCA* function with a possible maximum of six dimensions. This function uses an iterative principal components analysis (PCA) imputation technique (Isaak et al., 2018). The sequence of imputed missing data points, *im*, follows that of m summarized above.

An accuracy metric for non-imputed and imputed estimations was calculated as the relative difference between estimated and expected amplitude, phase, and mean values (i.e., estimated / expected) to determine missing data thresholds. A range in accuracy levels was assessed using the *confint* function in R (R Core Team, 2020) to calculate linear regression confidence intervals (CI) of expected values. Three CIs were computed: 95%, 99%, and 99.9%. The 95% CI is the narrowest range (i.e., most accurate) and the 99.9% CI is the widest range (i.e., least accurate) of the three levels. Estimated values were then compared to the minimum and maximum of a given CI level of the expected values.

The maximum allowable number of missing data points (m_{max}) was calculated for four "exceedance probability" levels (0.5, 0.75, 0.9, or

Table 1
Summary of the maximum number of missing data points acceptable (m_{max}) to satisfy various exceedance probability levels of estimated annual water temperature parameter values within a given confidence interval (CI) of the expected annual parameter values. The results for the annual mean ($T_{0,w}$) are shown on top, annual phase (ϕ_w) in the middle, and annual amplitude (A_w) on the bottom. Thresholds of m_{max} values are given for fractions of estimated annual values, within a given accuracy level (i.e. CI), exceeding 0.5, 0.75, 0.9, and 0.95. The m_{max} values were extracted from the mean of the site lines (thick lines in Fig. 1). OESF values are shown to the left of the pipe (|) and SHEN values are shown to the right. As an example, the first m_{max} value of 84 | 126 should be read as, "up to 84 for OESF and 126 for SHEN missing data points are acceptable in order to have at least half (0.5) of estimated annual mean values within the 99.9% CI of the expected annual mean."

Annual Parameter	Exceedance Probability	99.9% CI	99% CI	95% CI
$T_{0,w}$(°C)	0.5	84 | 126	70 | 119	56 | 98
	0.75	56 | 98	49 | 63	35 | 42
	0.9	42 | 56	35 | 42	28 | 28
	0.95	35 | 49	28 | 35	21 | 21
ϕ_w(d)	0.5	168 | 175	126 | 161	98 | 140
	0.75	91 | 133	77 | 112	63 | 70
	0.9	63 | 91	56 | 63	42 | 49
	0.95	56 | 70	42 | 56	35 | 42
A_w(°C)	0.5	133 | 175	112 | 154	91 | 133
	0.75	77 | 133	63 | 112	49 | 84
	0.9	56 | 105	42 | 56	35 | 35
	0.95	42 | 63	35 | 42	28 | 28

0.95) to assess the effect of various m values on the performance of annual linear regressions. These exceedance probability levels were defined for each parameter and accuracy level by calculating the fraction of the site-m regressions (i.e., all annual windows included) whose given estimated parameter value fell within the given accuracy level of the expected parameter value. For example, the 0.9 exceedance probability for $T_{0,w}$ at the 95% accuracy level for a given site occurs at the m value where the fraction of regressions producing $T_{0,w}$ estimates within the 95% CI of the expected $T_{0,w}$ values is above 0.9. While these fraction values were calculated for each site, we used the mean of all site fraction values at each m value to report exceedance probability m_{max} values.

To investigate seasonal effects of missing data, we applied sequentially changing start dates using a single missing data threshold (m_{max}). The objective was to choose an m_{max} value that would result in a high probability ($> \sim$90%) of estimating the annual parameter values within the 95–99% CI of the expected values. However, because the results varied by regional dataset and parameter, those conditions could not be satisfied for every case without excluding additional sites due to missing data. Therefore, a value of m_{max} was selected that was a balance between maximizing inclusion of sites within the two datasets and estimation accuracy of the three annual parameters: $m_{max} = 49$ (Johnson et al., 2020). With this threshold value, the estimated annual mean, phase, and amplitude values for both regions had a $>$ 75% probability of being within the 99% CI of the expected annual mean values (Table 1) and a $>$ 90% probability for annual phase and amplitude values. A slightly greater m_{max} threshold of 63 would result in a $<$ 75% probability of the estimated annual mean values being within the 99% CI of the expected parameters, so we decided to be more conservative and use a threshold of 49 in this study. Using this m_{max} threshold value, the effect of timing was assessed by calculating regression results with the first m_{max} dates of a given year assigned as missing data and then repeating regression calculations for sequentially later missing data periods of length m_{max} until the end of the year was reached for each site.

Supplemental to the results for annual water temperature patterns, combined air-water annual thermal metrics (mean ratio, phase lag, and amplitude ratio) were calculated using existing air temperature data (no missing data manipulation) and these results are presented in Appendix C. This last procedure was used to assess the effect of missing data on the interpretation of potential GW influence, which recent studies have shown can be inferred from the paired air-water annual thermal metrics (Briggs et al., 2018a; Johnson et al., 2020). If missing data is an issue for a specific air temperature dataset, then the combined air-water results will reflect a combination of errors resulting from water and air temperature regressions.

3. Results

The relative difference between estimated fitted sine-wave regression coefficient values for various levels of incomplete datasets and the expected fitted sine-wave regression coefficient values using a complete 365-d dataset were compared for non-imputed and imputed datasets.

3.1. Accuracy

As anticipated, the relative difference between estimated and expected annual water temperature mean ($T_{0,w}$), phase (ϕ_w), and amplitude (A_w) values decreased as the proportion of missing data points decreased but the rate of this change (i.e., the slope of error) differed between parameters and region of field data collection (Figs. 1-2 and Table 1). Specifically, as the exceedance probability (0.5, 0.75, 0.9, 0.95) increased, the m_{max} value decreased for all accuracy levels (i.e., within the 99.9% CI, 99% CI, and 95% CI of the expected values) (Table 1). Accurate estimation of $T_{0,w}$ was the most sensitive to the number of missing data points, while ϕ_w and A_w allowed for less complete datasets (higher m_{max}) for accurate estimations in both regions. OESF sites generally had lower m_{max} thresholds (i.e., more sensitive to missing data) for estimates of the three annual thermal parameters than SHEN sites (Table 1), despite smaller annual variation (A_w) and smaller regression residuals for OESF sites (Suppl. Info. Johnson et al., 2020). For OESF, the expected values ranged between 7.0 and 9.2 °C for $T_{0,w}$, between 210 and 254 d for ϕ_w, and between 1.9 and 4.7 °C for A_w. For SHEN, the expected values ranged between 8.8 and 12.6 °C for $T_{0,w}$, between 205 and 224 d for ϕ_w, and between 4.0 and 9.5 °C for A_w.

The average relative (estimated / expected) CI ranges were similar between $T_{0,w}$ and ϕ_w, and three to four times wider (i.e., less accurate) for A_w for both datasets (Table 2). The relative differences in accuracy of estimated parameters between 99.9% and 99% CI ranges and between 99% and 95% CI ranges were 0.4–0.6% in $T_{0,w}$ and ϕ_w, and 1.4–2.3% in A_w. SHEN sites exhibited narrower CI ranges for ϕ_w and A_w, and slightly wider CI ranges for $T_{0,w}$ than OESF sites.

3.2. Timing of missing data effect

Timing of missing data points within the annual window affected the accuracy of estimated $T_{0,w}$, ϕ_w, and A_w values, but the magnitude of the effect varied between regions, years, and annual parameters (Fig. 3). Overall, accuracy was high, with most of the estimated annual parameter values falling within the 95% CI of the expected annual parameter values and few occasions exceeding 10% of estimates outside of the 99% CI, regardless of timing. For both OESF and SHEN sites, the estimated

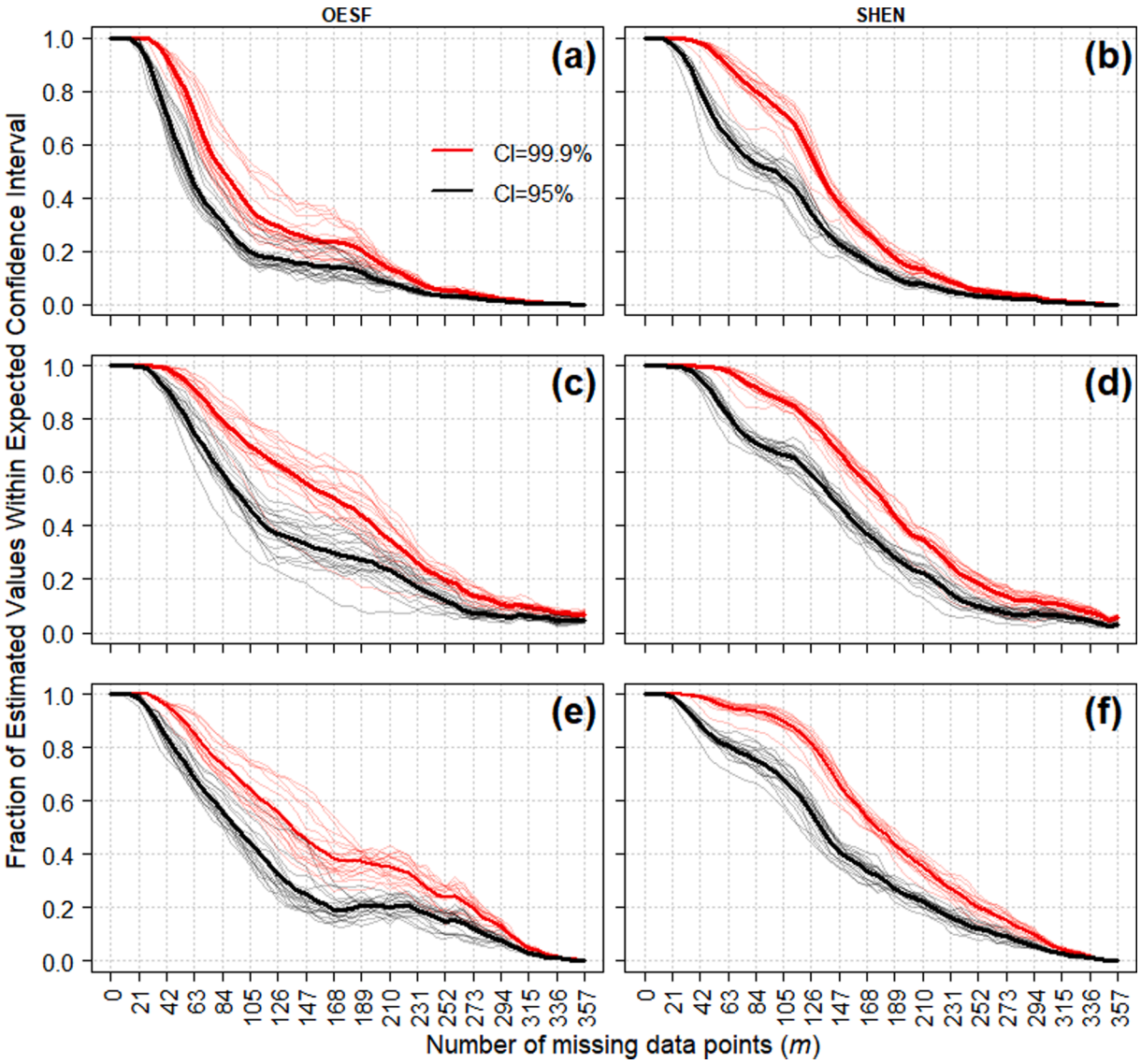

Fig. 1. Fraction of estimated annual water temperature parameter values within a given confidence interval (CI) of the expected annual parameter values versus the number of consecutively missing data points within a year (*m*). The results for the annual mean ($T_{0,w}$) are shown on top (a,b), annual phase (ϕ_w) in the middle (c,d), and annual amplitude (A_w) on the bottom (e,f). Results for OESF sites are shown on the left (a,c,e) and results for SHEN sites are shown on the right (b,d,f). Colors represent the different accuracy levels (95% CI > 99.9% CI) of estimated annual parameter values. Each thin line represents the results for an individual site (n = 20 for each region). Thick lines represent the mean of all the sites for a given parameter-region-*m*-CI combination. The 99% CI lines are not shown, but plot between the 95% and 99.9% CI lines for each site.

annual values were least accurate for $T_{0,w}$ given relatively tight CI ranges. The CI ranges for ϕ_w were similarly tight but in general the estimates were relatively more accurate. A_R exhibited the widest CI ranges and similarly accurate estimates as ϕ_w. However, for OESF sites, missing data beginning between approximately December and January produced less accurate estimates of all three annual parameters in at least one of the two years. For both OESF and SHEN sites, the patterns between individual years varied.

3.3. Imputation performance

Overall, imputation of missing data increased the accuracy of estimated annual temperature parameters (Fig. 4 and Table 3). Improvements were greatest between approximately 9–27 weeks of imputed missing data (i.e., *im* = 63–189 d) and extended exceedance probability thresholds by between 5 and 18 weeks. Selection of an imputed missing data threshold (im_{max}) using the same criteria as m_{max} described above, results in a value of 133 d with a less conservative upper limit of 154 d.

However, imputation was not equally successful among sites, as seen by the large variation among sites (Fig. 4). Additionally, in one OESF site and one SHEN site exhibiting late ϕ_w and moderate to low A_w (indicators of strong, shallow GW influence; Briggs et al., 2018a), accuracy of estimated parameters was worse, not better, when missing data were imputed. Conversely, imputation improved the accuracy of estimated parameters for another SHEN site exhibiting strong GW signals. Note that the results presented in Fig. 1 and Table 1 include all annual windows, but are virtually the same as the results produced by subsetting to the same annual windows used for the missing data imputation scenarios. Therefore, comparison between Tables 1 and 3 is consistent.

$T_{0,w}$ remained the most sensitive to the number of imputed missing data points, while ϕ_w and A_w required less complete datasets (i.e., higher im_{max}) for accurate estimations in both regions. OESF sites still generally had lower im_{max} thresholds (i.e., more sensitive to imputed missing data) for estimates of the three annual thermal parameters than SHEN sites (Table 3). For OESF, the expected values ranged between 7.1 and 9.2 °C for $T_{0,w}$, between 211 and 253 d for ϕ_w, and between 1.9 and 4.7 °C for

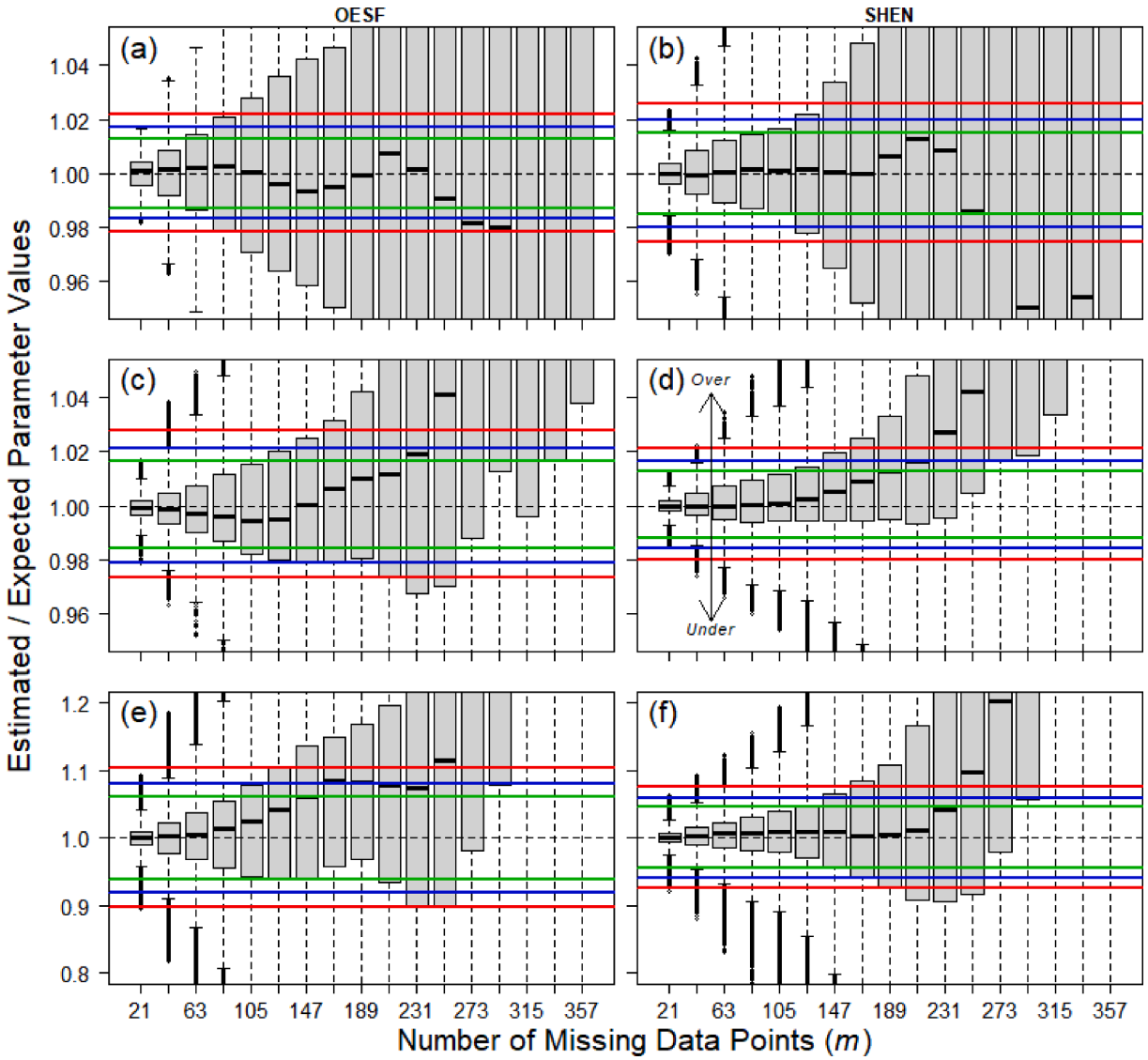

Fig. 2. Box plots of the ratio between estimated and expected annual water temperature parameter values (estimated / expected) versus the number of missing data points within a year (m). The results for the annual mean ($T_{0,w}$) are shown on top (a,b), annual phase (ϕ_w) in the middle (c,d), and annual amplitude (A_w) on the bottom (e,f). Results for OESF sites are shown on the left (a,c,e) and results for SHEN sites are shown on the right (b,d,f). Gray boxes show the interquartile range (IQR) with the median drawn as a horizontal black line, whiskers extend to 1.5·IQR above and below the boxes, and circles represent outliers (i.e., values beyond 1.5·IQR). Mean confidence interval ranges of all sites and annual periods are displayed as colored horizontal solid lines in each panel (99.9% = red, 99% = blue, and 95% = green). Each box plot consists of all results for each m value (i.e., 100 regressions × 20 sites × 53 (OESF) or 61 (SHEN) annual periods). Larger deviations from unity (horizontal dashed line) indicate less accurate estimates of annual water temperature parameter values (greater than unity is overestimation and less than unity is underestimation). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Summary of the average relative accuracy of confidence intervals (CI) calculated as the ratio between estimated and expected annual water temperature parameter values (estimated / expected). Each accuracy value is determined by taking the mean of all site-annual period pairings in each regional dataset for a specific CI range. The relative accuracy (%) of the estimations can be calculated as the difference from unity multiplied by 100 (i.e., *100(ratio −1)*).

Annual Parameter	Dataset	99.9% CI	99% CI	95% CI
$T_{0,w}$(°C °C^{-1})	OESF	0.978–1.022	0.983–1.017	0.987–1.013
	SHEN	0.974–1.026	0.980–1.020	0.985–1.015
ϕ_w(d d^{-1})	OESF	0.973–1.027	0.979–1.021	0.984–1.016
	SHEN	0.980–1.021	0.984–1.016	0.988–1.012
A_w(°C °C^{-1})	OESF	0.897–1.104	0.919–1.081	0.939–1.061
	SHEN	0.925–1.076	0.941–1.059	0.955–1.045

A_w. For SHEN, the expected values ranged between 8.8 and 12.6 °C for $T_{0,w}$, between 205 and 224 d for ϕ_w, and between 4.1 and 9.4 °C for A_w. Generally, for all annual parameters in both regions with less than approximately 231 days of imputed data, the degree of improved accuracy via imputation was greatest within the 95% CI and least within the 99.9% CI.

4. Discussion

Missing data is a common issue with stream temperature datasets, but currently there is little quantitative guidance for the use of incomplete datasets to model annual stream temperature signals. Determining a missing data threshold (m_{max}) depends on the specific research goals and thus establishing an exact value of m_{max} to use for every possible scenario, location, and purpose is not possible. The value of this threshold could determine the interpretation of potential GW influence,

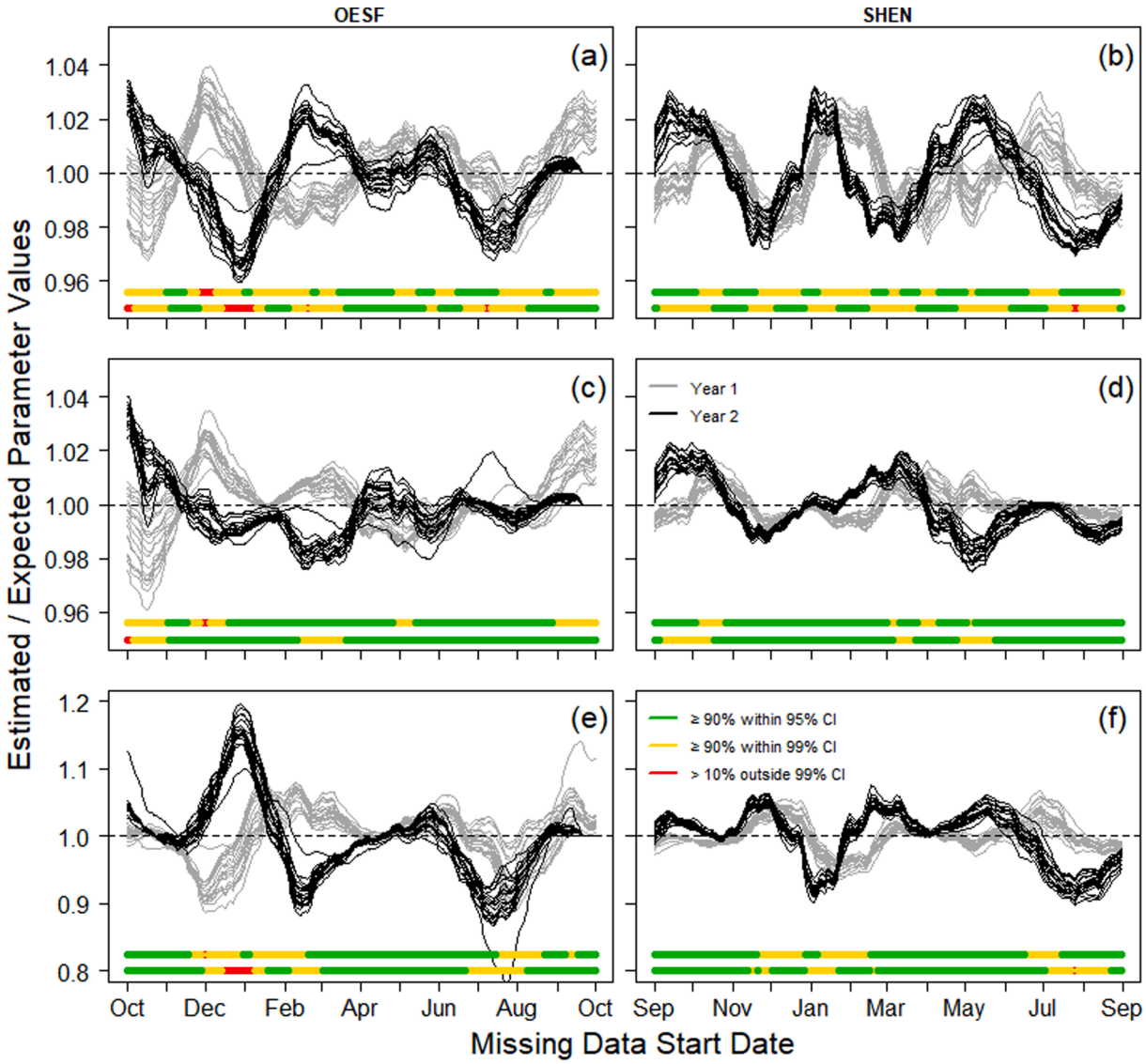

Fig. 3. Ratio between estimated ($m = 49$) and expected ($m = 0$) annual water temperature parameters (estimated / expected) versus the starting date of consecutive missing data for two consecutive years (October 2016–2018 for OESF and September 2014–2016 for SHEN). Ratio values of annual mean ($T_{0,w}$, °C) are shown in the top row (a,b), annual phase (ϕ_w, d) in the middle row (c,d), and annual amplitude (A_w, °C) in the bottom row (e,f). Results for OESF sites (n = 20) are shown in the left column (a,c,e) and for SHEN sites (n = 20) in the right column (b,d,f). Each line (gray = Year 1, black = Year 2) represents the results for a single site. The horizontal dashed line represents unity, where less accurate estimates produce greater deviations from unity (greater than unity is overestimation and less than unity is underestimation). The two green-yellow-red bands at the bottom of each panel represent dates in which at least 90% of estimated parameter values are within the 95% CI (green), at least 90% are within the 99% CI (yellow), or greater than 10% are outside the 99% CI (red) of the expected values, respectively (Year 1 is the top band and Year 2 is the bottom band). For OESF, the expected values range between 7.0 and 9.4 °C for $T_{0,w}$, between 211 and 254 d (Gregorian calendar) for ϕ_w, and between 1.9 and 4.7 °C for A_w. For SHEN, the expected values range between 8.8 and 12.7 °C for $T_{0,w}$, between 205 and 225 d for ϕ_w, and between 4.2 and 9.5 °C for A_w. Note that the x-axes differ between OESF and SHEN sites. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which recent studies have shown can be inferred from the paired air-water annual thermal metrics (Briggs et al., 2018a; Johnson et al., 2020). Results of this study showed that accurate estimates of all three annual thermal parameters were possible even when as much as seven to nine weeks (49–63 d) of temperature data are missing on consecutive days. In addition, imputation of missing data could extend this threshold to approximately 19 to 22 weeks (133–154 d) given complete temperature datasets are available from other sites with similar thermal signatures. However, variability in thermal parameter accuracy among sites suggests that caution should be exercised when using imputation approaches to fill temperature data gaps.

Specifically, although imputation of missing data improved estimations of annual thermal parameters for a vast majority of sites in this study, for some sites imputation had the opposite effect. This pattern seemed to be associated with some of the sites exhibiting strong apparent GW influence, evidenced by late ϕ_w (>15 d) or low A_w (<0.55), which have important implications for the thermal stability of streams in a warming climate (Briggs et al., 2018a, 2018b). However, it is not immediately clear why the accuracy of estimated parameters for some sites with strong GW influence benefit from imputation while others exhibiting similar characteristics suffer. The imputation techniques presented in this study are dependent upon correlations among sites and records in a dataset to estimate missing values. Therefore, caution is needed when applying these techniques in regions or stream networks where GW influence is spatially patchy or if unique annual thermal characteristics may be present. Accurate characterization of spatial

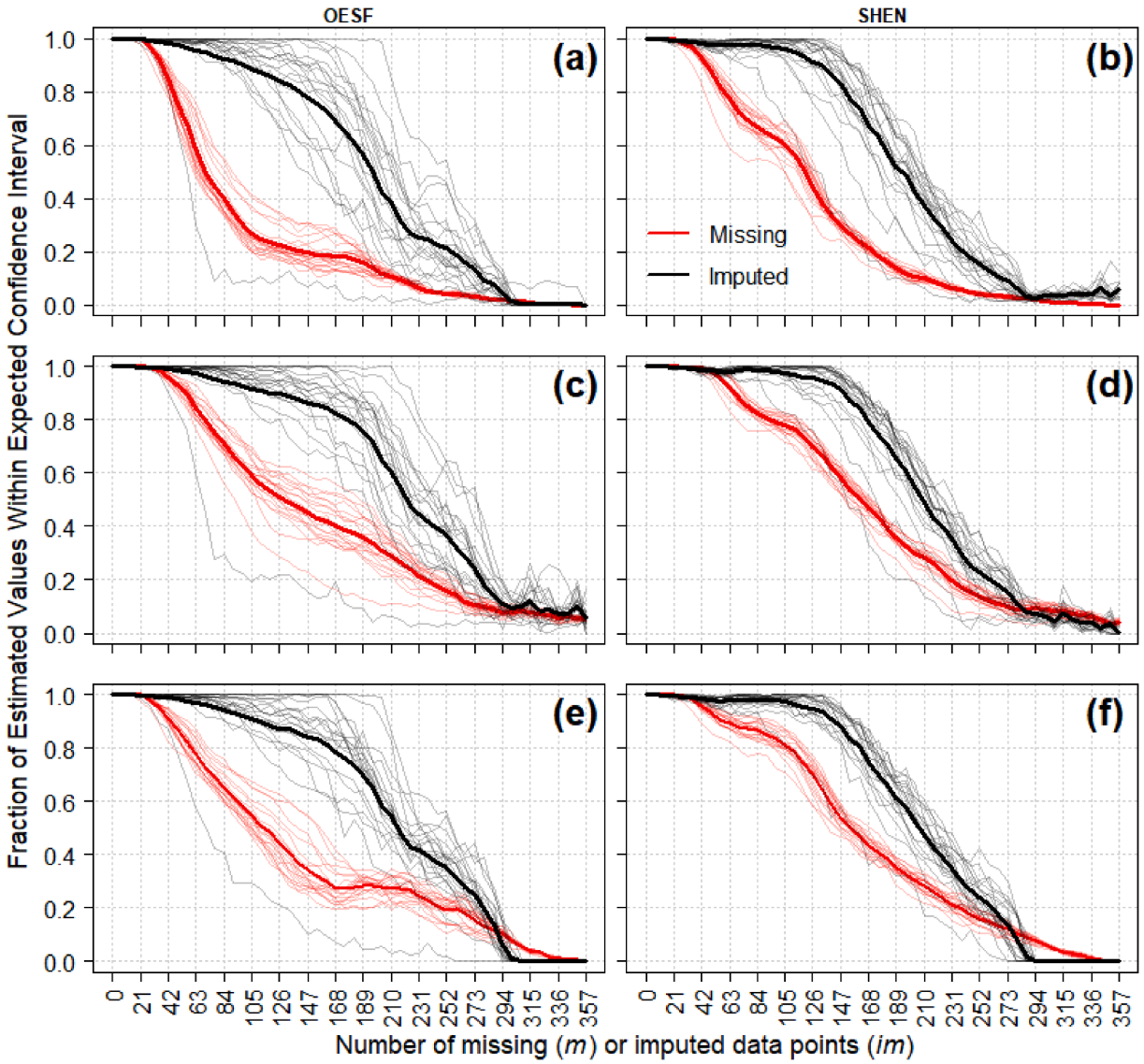

Fig. 4. Fraction of estimated water temperature annual parameter values within the 99% confidence interval (CI) of the expected annual parameter values versus the number of consecutively missing (*m*) or imputed (*im*) data points within a year. The results for the annual mean ($T_{0,w}$) are shown on top (a,b), annual phase (ϕ_w) in the middle (c,d), and annual amplitude (A_w) on the bottom (e,f). Results for OESF sites are shown on the left (a,c,e) and results for SHEN sites are shown on the right (b,d,f). Colors represent either missing (red) or imputed missing data (black) for the 99% CI of estimated annual parameter values. Each thin line represents the results for an individual site (n = 20 for each region). Thick lines represent the mean of all the sites for a given parameter-region-*m* (or *im*)-CI combination. The 95% and 99.9% CI's are not shown, but are slightly shifted below and above the 99% CI lines, respectively. See Fig. 1 for the missing data points 95% and 99.9% CI values for comparison. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

heterogeneity in stream thermal regimes is essential to predicting future thermal and climate refugia. Further research is needed to assess the performance of this imputation technique, such as the number of PCA dimensions needed for a given number of missing data points.

Using a m_{max} threshold of 49 d, the difference in estimated versus expected annual parameter values was minimized while simultaneously maximizing the number of applicable sites. A compromise in the chosen m_{max} value was necessary because of the different sensitivities to missing data of the three annual thermal parameters estimations. Therefore, $T_{0,w}$ estimates will be less accurate for *m* values approaching m_{max}, whereas ϕ_w and A_w will remain highly accurate in both regions. In contrast, estimated – expected differences in annual water temperature parameter values will fare worse for streams experiencing extended frozen or dry periods (e.g., $m > 63$ d) and far worse for datasets collected only seasonally (e.g., $m > 240$ d). Imputation could be considered for sites in these types of datasets if: 1) there are less than approximately 20 weeks of missing data within a given annual period for a subset of sites in the dataset, 2) the complete site time series are long enough to establish general relationships of annual thermal characteristics between sites (Jones and Schmidt, 2018), and 3) the site(s) with missing data do(es) not exhibit vastly different annual thermal characteristics from the rest of the sites in the dataset.

Missing data points that occurred on consecutive days were the focus of this analysis as this is a more likely scenario when collecting stream temperature data because of data logger failure, dewatering, wash-out, etc. The results of this study indicate that incomplete datasets may be used in annual stream temperature modeling and the magnitude of error will depend on the threshold set. Sites in OESF and SHEN, which are representative of remote headwater sites, are only occasionally visited throughout the year and may be at higher risk of exclusion from annual analyses because the chances of data loss occurring for more than seven to nine weeks prior to visiting individual sites increases with fewer visits within a given year. Conversely, if the stream of interest is easily accessible, multiple visits are warranted within any given year to limit

Table 3
Summary of the maximum number of imputed missing data points acceptable (im_{max}) to satisfy various exceedance probability levels of estimated water annual parameter values within a given confidence interval (CI) of the expected annual values. Thresholds of im_{max} values are given for fractions of estimated annual values, within a given accuracy level, exceeding 0.5, 0.75, 0.9, and 0.95. The im_{max} values were extracted from the mean of the site lines (e.g., thick lines in Fig. 4). OESF values are shown to the left of the pipe (|) and SHEN values are shown to the right.

Annual Parameter	Exceedance Probability	99.9% CI	99% CI	95% CI
$T_{0,w}$(°C)	0.5	203 | 196	196 | 189	182 | 182
	0.75	168 | 161	154 | 154	133 | 147
	0.9	105 | 140	98 | 133	91 | 112
	0.95	70 | 119	70 | 112	63 | 98
ϕ_w(d)	0.5	231 | 210	217 | 203	210 | 203
	0.75	196 | 175	189 | 168	175 | 168
	0.9	140 | 147	126 | 147	112 | 140
	0.95	84 | 140	84 | 133	77 | 119
A_w(°C)	0.5	217 | 203	210 | 203	210 | 196
	0.75	182 | 168	175 | 168	168 | 161
	0.9	112 | 147	112 | 140	105 | 133
	0.95	84 | 133	77 | 126	77 | 112

the length of potential missing data periods. The advent of automatic cloud-based data monitoring and access, such as those used at USGS streamgage stations (U.S. Geological Survey, 2020), also limits the length of potential missing data gaps by providing real-time access to data, which allows for faster responses to data failure.

Ward (1963) showed that 12 monthly averages for an example site give virtually the same regression fit as 365 days of data, but this does not translate into measuring daily mean temperature only once every month. Based on our results, data from a large majority of the year (>80% or >60% with imputation) are required for accurate estimates of all three annual thermal metrics, less for randomly distributed missing data (Appendix B). A similar threshold (>70% complete) was used in a recent study (Isaak et al., 2020), though this was for a five-year period and not specifically for each year. Therefore, seasonally collected temperature datasets (most commonly during summer) especially should not be used to produce annual water temperature parameter estimates (<15% probability of being within the 95% CI of expected values), nor should imputation be used to produce annual estimates for these datasets. More research is needed to explore how methods to fill in missing data (i.e., imputation) could benefit the applicability of sub-annually measured water temperature data in annual thermal pattern calculations in other regions.

Additionally, these results provide guidance for sites that experience intermittent flow throughout the year or temporary ice or snow cover in the winter. For example, if the period of ice/snow cover or no-flow conditions for a specific stream is less than the specified m_{max} threshold, then researchers could simply ignore the data collected while ice/snow or no-flow is present. If ice/snow cover or no-flow conditions last longer than the specified m threshold, the use of a modeling approach where the sine-wave modeling time period is defined as the free-flowing time (Letcher et al., 2016; Maheu et al., 2016b), may be necessary. However, more research is needed on this topic to address the appropriateness of comparing annual temperature parameters derived from a period of 365 days to parameters derived from a period of less than 365 days. We hypothesize that imputation of these types of data could lead to erroneous inferences, but we have not tested this explicitly.

We found conflicting evidence regarding how the timing of missing data affect parameter accuracy. The magnitudes of these effects were highly variable, depending on the region, year, and annual parameter evaluated. Therefore, setting a more conservative maximum threshold of missing data based on the probability of exceedance for the entire year will likely be a better approach than setting a higher maximum threshold based on the inconsistent exceedance probability in a particular season. For example, assuming data are missing from winter months, annual estimates of one parameter may be highly accurate while annual estimates of another parameter may be poor. This inconsistency is likely driven by a combination of the variability in the leverage (influence of measured value on fitted value) of individual temperature data points in the linear regressions among different annual periods, which is related to the sine-wave ideality of the measured temperature data within that annual period, and the timing of those high leverage data points within the year. In this study, OESF temperatures exhibited less sine-wave ideality (i.e., slight temporal asymmetry) than the SHEN temperatures which may partly explain why m_{max} thresholds were generally lower for OESF sites than SHEN sites.

Outlying temperature values with high leverage have a large influence on fitted linear regressions solved via least squares. Therefore, missing data corresponding to these outlier temperature values may have a greater influence on estimated parameter values than missing data corresponding to more common temperature values. For datasets that are highly influenced by outlier values, one potential solution is to use an alternative solving method less influenced by outlier values such as the Theil-Sen estimator (Fernandes and Leblanc, 2005). Timing of missing data corresponding to outlier temperature values may, in part, explain differences in the effect on the three annual temperature parameters within a given region in this study. For example, missing data corresponding to anomalously high temperatures at or near the annual maxima may affect $T_{0,w}$ and A_R more than ϕ_w, whereas later timing of these same missing data points may affect ϕ_w more than $T_{0,w}$ or A_w. As the m_{max} threshold increases, the accuracy of estimated annual water temperature parameters will decrease and the importance of the timing of missing data will increase as the relative error in regression estimates approaches the level of acceptance. Understanding the effect of missing data timing is especially important for sites consistently influenced by seasonal ice cover or flow intermittency. Thus, further research is needed to assess whether these relationships are applicable to regions outside of the current study areas.

Previous research showed the utility of comparing paired air-water annual temperature $\Delta\phi$ and A_R values for deciphering watershed process (Briggs et al., 2018a; Johnson et al., 2020), where the depth of GW discharge is important for inferring a stream's thermal resilience to changes in climate. However, setting a maximum missing data threshold too high could result in false classification of streams based on annual thermal signals, which could lead to large errors in the predicted thermal stability of a stream given increasing air temperatures and therefore its interpreted capabilities to serve as a thermal or climate refugia (Ebersole et al., 2020). For example, Briggs et al. (2018a) found that, in the GW-dominated Quashnet River (Massachusetts), A_R values ranged from approximately 0.49–0.63 over a three-year period. Additionally, data from Shenandoah National Park, VA, and the models of Briggs et al. (2018a) indicated a $\Delta\phi$ of approximately 10 d or greater corresponded to shallow GW discharge comprising at least 25% of total streamflow. Deep GW discharge does not induce a large stream $\Delta\phi$ and is more resilient to changes in surface conditions. Therefore, an A_R threshold of 0.65 roughly separates strong atmospheric signals ($A_R > 0.65$) from strong GW signals ($A_R \leq 0.65$) and a $\Delta\phi$ threshold of 10 d roughly separates shallow ($\Delta\phi \geq 10$ d) from deep ($\Delta\phi < 10$ d) GW signals.

Using these approximate threshold guidelines for OESF and SHEN sites, it is clear that more missing data within a given annual period leads to a greater probability of mischaracterization of the dominant controls on the stream thermal regime (Appendix Figs. C1 and C2). At high missing data thresholds (e.g., greater than approximately half of the year), $\Delta\phi$ and A_R are overestimated, which leads to false indications of lagged strong atmospheric coupling where the expected values indicate shallow or deep GW influence (low A_R) or weaker atmospheric signal influence (lower $\Delta\phi$ with high A_R). This could lead to underestimations of the thermal resilience of these streams to changing climate conditions. Therefore, accurate estimations of $\Delta\phi$ and A_R are especially important when these values approach the approximate

classification thresholds for $\Delta\phi$ and A_R (Appendix C).

5. Conclusions

With the recent expansion in available temperature data around the world, greater understanding of stream temperature dynamics is possible. However, missing data points within these datasets will continue to be an omnipresent issue, resulting from logger malfunction, human error, or natural occurrences such as high flows, desiccation, or freezing. In fact, sensors and methods have been developed specifically to differentiate periods with and without flow in streams (Arismendi et al., 2017; Blasch et al., 2002; Chapin et al., 2014). Therefore, guidelines for the use and analysis of incomplete datasets are needed. In this study, we provide such guidelines for the analysis of annual paired air–water temperature datasets using linear regressions of sine-wave curves. The methods described here could be applied to other stream temperature modeling techniques that use incomplete datasets.

Annual water temperature signals, especially when paired with local air temperature signals, are useful in determining important watershed processes for stream ecosystems. We found that the simultaneous accurate (within the 95–99% CI of expected values) estimation of three annual water temperature parameters – mean, phase, and amplitude – could only withstand fewer than two months of consecutive missing data points for our sites in the Pacific Northwest and Mid-Atlantic regions of the USA. Imputation of missing data expanded this period to between four and five months for most sites, but also decreased accuracy for locations exhibiting unique annual thermal characteristics such as those with strong shallow GW influence. Nonetheless, the specific threshold of maximum allowable missing data points used in other projects should be set to the desired outcomes of the research or management objectives. In areas where both air and water temperature data are missing, utilizing modeled air temperature data, such as PRISM (PRISM Climate Group, 2019), may help to expand the usefulness of existing water temperature datasets via combined air–water annual thermal metrics (Briggs et al., 2018a; Johnson et al., 2020), but additional research in this field is required to determine the accuracy and applicability of these modeled air temperatures in other regions.

CRediT authorship contribution statement

Zachary C. Johnson: Conceptualization, Methodology, Data curation. **Brittany G. Johnson:** Conceptualization, Methodology. **Martin A. Briggs:** Conceptualization. **Craig D. Snyder:** Conceptualization, Data curation. **Nathaniel P. Hitt:** Conceptualization, Data curation. **Warren D. Devine:** Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the University of Washington School of Environmental and Forest Sciences. Some of this analysis was supported by the U.S. Geological Survey Toxic Substances Hydrology Program and by Department of Energy grant DE-SC0016412. Additional information and data are available as Appendices to this manuscript, which can be found in the Supplementary information. We thank Karen Rice, Robin White, Dan Isaak, and an anonymous reviewer for helpful suggestions and comments. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A

The method used to convert annual phase values into the day of the year corresponding to the annual maximum temperature varies with the annual window defined for the linear regressions (i.e., Eq. (1)). Given the two regression coefficients (*a* and *b*), there are four possible combinations of coefficient signs (Appendix Fig. A1): (1) both *a* and *b* are positive, (2) *a* is positive and *b* is negative, (3) *a* is negative and *b* is positive, and (4) both *a* and *b* are negative. In the first and second scenarios, the annual maximum temperature occurs prior to the annual window mid-point, with the first scenario maximum occurring approximately one to two months after the start of the annual window, and the second scenario maximum occurring approximately three months later. In the third and fourth scenarios, the annual maximum temperature occurs after the annual window mid-point. The third scenario maximum occurs approximately one to two months prior to the end of the annual window and the fourth scenario maximum occurs approximately three months earlier. Therefore, the third scenario corresponds with the water year (WY) calendar (1 October to 30 September) for northern hemisphere streams and the fourth scenario corresponds with the Gregorian year calendar for northern hemisphere streams. The first

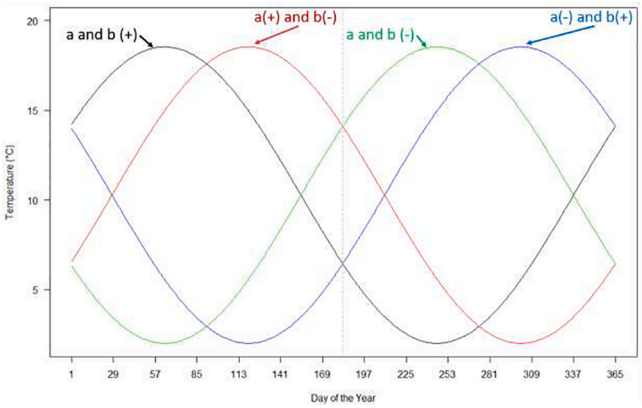

Fig. A1. Four combinations of sine-wave linear regression coefficient signs: (1) *a* and *b* (+) (black), (2) *a*(+) and *b*(−) (red), (3) *a*(−) and *b*(+) (blue), and (4) *a* and *b* (−) (green). The gray vertical dashed line represents the mid-point of the annual window.

scenario then corresponds with the Gregorian year calendar for southern hemisphere streams and the second scenario corresponds with the WY calendar for southern hemisphere streams. The first scenario could also correspond to an approximately June to May calendar year and the second scenario to an approximately April to March calendar year for northern hemisphere streams. Similarly, the third scenario could also correspond to an approximately April to March calendar year and the fourth scenario to an approximately June to May calendar year for southern hemisphere streams.

Fortunately, the conversion of phase values from radians into day of the year corresponding to the annual maximum temperature has only two variants, which apply to northern and southern hemisphere streams separately, despite the four possible combinations of regression coefficient signs. In other words, how the conversion is carried out is dependent only on the timing of the annual maximum in relation to the mid-point of the annual window (i.e. $t = \pi$ rad). For northern hemisphere streams (i.e., scenarios three and four, where the maximum occurs *after* the annual midpoint), phase (rad) is shifted by: $3\pi/2 - tan^{-1}(b/a)$. For southern hemisphere streams (i.e., scenarios one and two, where the maximum occurs *before* the annual midpoint), phase (rad) is shifted by: $\pi/2 - tan^{-1}(b/a)$. For all scenarios, phase is converted from units of radians to days by multiplying by: $365/2\pi$. Note that the timing of annual minimum temperature is shifted from the annual maximum by π for a symmetric sine-wave.

Appendix B

A secondary analysis using a random distribution of missing days of data found that dramatically more missing data was permissible than if the missing data existed on consecutive days (Fig. B1). Annual $T_{0,w}$ estimations were accurate with > 200 days of missing data, but annual ϕ_w and A_W (m > 250) remained accurate with even greater number of missing data (> 90% probability of being within the 95% CI of the expected values).

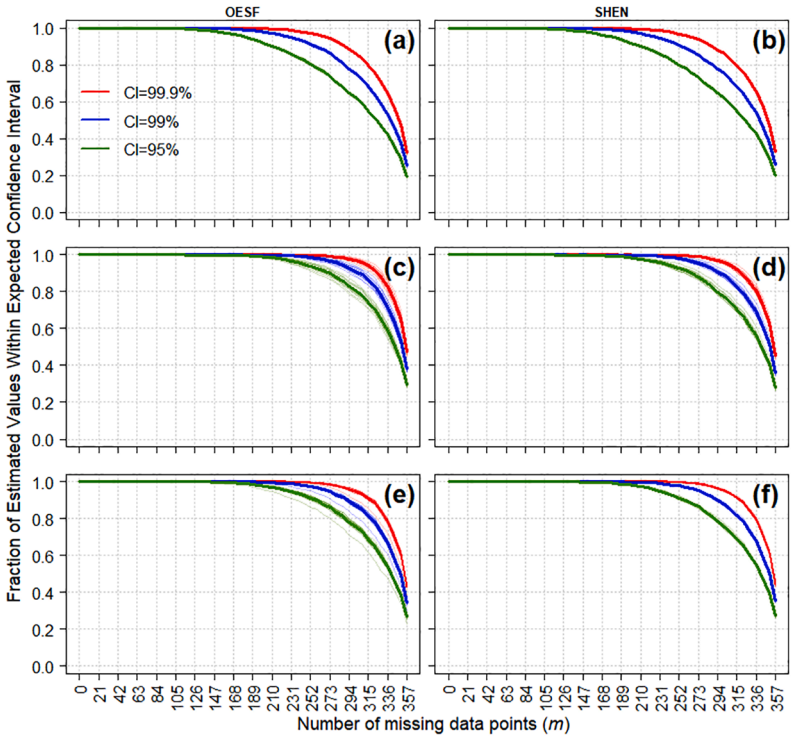

Fig. B1. Fraction of estimated annual water temperature parameter values within a given confidence interval (CI) of the expected annual parameter values versus the number of randomly missing data points within a year (m). The results for the annual mean ($T_{0,w}$) are shown on top (a,b), annual phase (ϕ_w) in the middle (c,d), and annual amplitude (A_w) on the bottom (e,f). Results for OESF sites are shown on the left (a,c,e) and results for SHEN sites are shown on the right (b,d,f). Colors represent the different accuracy levels (95% CI > 99% CI > 99.9% CI) of estimated annual parameter values. Each thin line represents the results for an individual site (n = 20 for each region). Thick lines represent the mean of all the sites for a given parameter-region-m-CI combination.

Appendix C

The difference between estimated and expected annual mean ratio (M_R), phase lag ($\Delta\phi$), and amplitude ratio (A_R) values follows that of annual water mean (M_w), phase (ϕ_w), and amplitude (A_w) (Figs. 1 and 2), assuming no missing air temperature data. Similar patterns were observed when plotting $\Delta\phi$ versus A_R (Appendix Figs. C1 and C2), where $m \leq 49$ corresponds to highly accurate estimates of expected annual thermal metric values. Previous research showed the utility of plotting $\Delta\phi$ versus A_R for deciphering watershed process (Briggs et al., 2018a; Johnson et al., 2020). Streams dominated by atmospheric signals exhibit low $\Delta\phi$ with high A_R, whereas streams dominated by deep GW signals exhibit low $\Delta\phi$ with low A_R, and streams dominated by shallow GW signals exhibit high $\Delta\phi$ with low A_R. Dams have been hypothesized to possibly induce large $\Delta\phi$ with high A_R (Johnson et al., 2020), but our results (Figs. C1 and C2) indicate missing data may also induce these signal patterns.

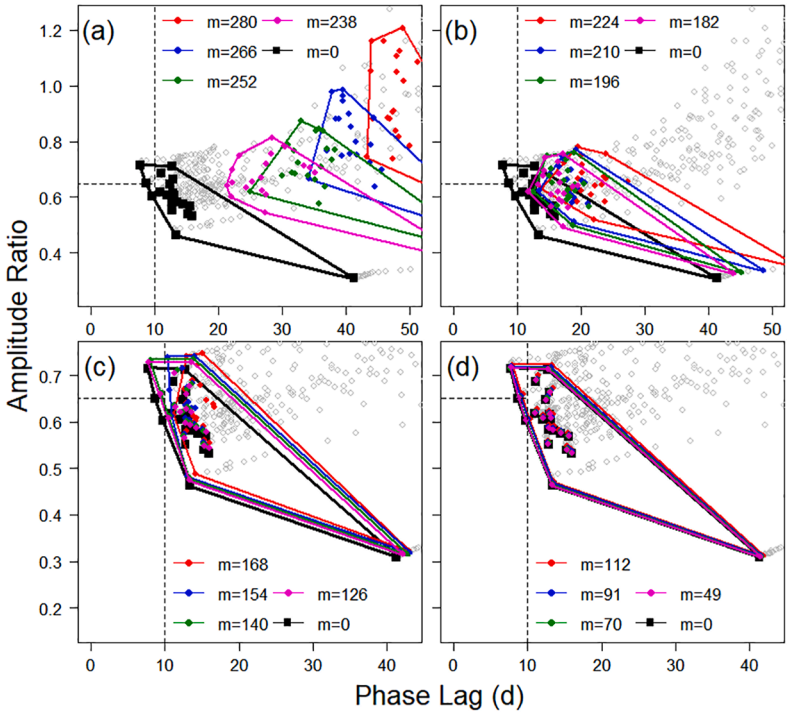

Fig. C1. OESF phase lag ($\Delta\phi$) versus amplitude ratio (A_R) for various sets of missing data points (m), progressing in descending order from panels (a) to (d). Only a subset of the missing data scenarios are colored and represented in the legend for each panel. Colored polygons represent the convex hull (i.e., smallest convex enclosure) for the five m values shown in each panel's legend. Open gray circles represent values from the other missing data scenarios not represented in the legend for each panel. Dashed lines ($\Delta\phi = 10$ d, $A_R = 0.65$) roughly divide the $\Delta\phi$-A_R space into three areas reflecting different dominant signals: top left area ($\Delta\phi < 10$ d, $A_R > 0.65$) reflects strong atmospheric signals, right area ($\Delta\phi \geq 10$ d) reflects strong shallow groundwater signals, and bottom left area ($\Delta\phi < 10$ d, $A_R \leq 0.65$) reflects strong deep groundwater signals.

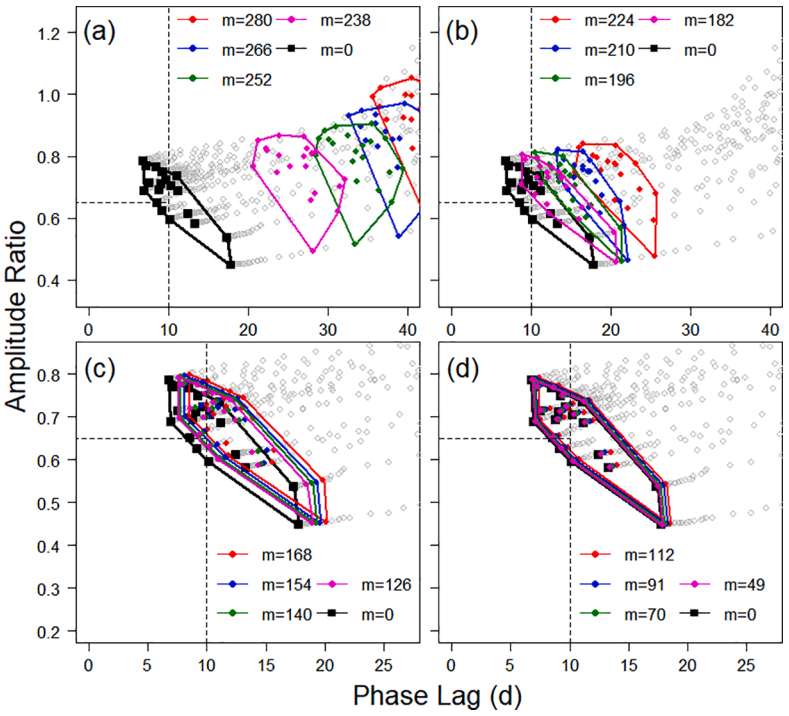

Fig. C2. SHEN phase lag ($\Delta\phi$) versus amplitude ratio (A_R) for various sets of missing data points (m), progressing in descending order from panels (a) to (d). Only a subset of the missing data scenarios are colored and represented in the legend for each panel. Colored polygons represent the convex hull (i.e. smallest convex enclosure) for the five m values shown in each panel's legend. Open gray circles represent values from the other missing data scenarios not represented in the legend for each panel. Dashed lines ($\Delta\phi$=10d, $A_R = 0.65$) roughly divide the $\Delta\phi$-A_R space into three areas reflecting different dominant signals: top left area ($\Delta\phi$<10d, $A_R > 0.65$) reflects strong atmospheric signals, right area ($\Delta\phi \geq$10d) reflects strong shallow groundwater signals, and bottom left area ($\Delta\phi$<10d, $A_R \leq 0.65$) reflects strong deep groundwater signals.

High m values resulted in overestimations of $\Delta\phi$ and A_R values (Fig. C1A and C2A), but as m decreases, $\Delta\phi$ and A_R decrease toward expected values. Therefore, high m values ($> \sim 182$) may indicate stronger lagged atmospheric coupling (high A_R with high $\Delta\phi$) in sites when it is more likely that they are influenced by shallow or deep GW. For OESF sites (Fig. C1), estimated A_R values are roughly representative of the expected A_R values as early as $m \approx 161$. However, $\Delta\phi$ values become underestimated as shown in the convex hull (i.e. smallest convex enclosure surrounding the data) shifted slightly to the left. For SHEN sites (Fig. C2), A_R and $\Delta\phi$ values are similarly overestimated and then underestimated, but the underestimation of $\Delta\phi$ is greater for SHEN sites. Initially, this results in estimated convex hulls shifted up and right from the expected, but then shifted slightly below and left for smaller values of m.

References

Anderson, M.P., 2005. Heat as a Ground Water Tracer. Groundwater 43, 951–968. https://doi.org/10.1111/j.1745-6584.2005.00052.x.

Arismendi, I., Dunham, J., Heck, M., Schultz, L., Hockman-Wert, D., 2017. A Statistical Method to Predict Flow Permanence in Dryland Streams from Time Series of Stream Temperature. Water 9, 946. https://doi.org/10.3390/w9120946.

Arismendi, I., Johnson, S.L., Dunham, J.B., Haggerty, R., 2013. Descriptors of natural thermal regimes in streams and their responsiveness to change in the Pacific Northwest of North America. Freshwater Biology 58, 880–894. https://doi.org/10.1111/fwb.12094.

Bal, G., Rivot, E., Baglinière, J.-L., White, J., Prévost, E., 2014. A Hierarchical Bayesian Model to Quantify Uncertainty of Stream Water Temperature Forecasts. PLOS ONE 9, e115659. https://doi.org/10.1371/journal.pone.0115659.

Blasch, K.W., Ferré, T.P.A., Christensen, A.H., Hoffmann, J.P., 2002. New field method to determine streamflow timing using electrical resistance sensors. Vadose Zone J. 1 (2), 289–299. https://doi.org/10.2136/vzj2002.2890.

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., 2018a. 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.

Briggs, M.A., Lane, J.W., Snyder, C.D., White, E.A., Johnson, Z.C., Nelms, D.L., Hitt, N.P., 2018b. Shallow bedrock limits groundwater seepage-based headwater climate refugia. Limnologica 68, 142–156. https://doi.org/10.1016/j.limno.2017.02.005.

Buccola, N.L., Turner, D.F., Rounds, S.A., 2016. Water Temperature Effects from Simulated Dam Operations and Structures in the Middle Fork Willamette River, Western Oregon: U.S. Geological Survey Open-File Report 2016-1159, 39 p. https://dx.doi.org/10.3133/ofr20161159.

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

Chapin, T.P., Todd, A.S., Zeigler, M.P., 2014. Robust, low-cost data loggers for stream temperature, flow intermittency, and relative conductivity monitoring. Water Resour. Res. 50 (8), 6542–6548. https://doi.org/10.1002/2013WR015158.

Cluis, D.A., 1972. Relationship between stream water temperature and ambient air temperatures: A simple autoregressive model for mean daily stream water temperature fluctuations. Hydrology Research 3, 65–71. https://doi.org/10.2166/nh.1972.0004.

Constantz, J., 2008. Heat as a tracer to determine streambed water exchanges. Water Resour. Res. 44 (4) https://doi.org/10.1029/2008WR006996.

Dugdale, S.J., Malcolm, I.A., Kantola, K., Hannah, D.M., 2018. Stream temperature under contrasting riparian forest cover: understanding thermal dynamics and heat exchange processes. Sci. Total Environ. 610-611, 1375–1389. https://doi.org/10.1016/j.scitotenv.2017.08.198.

Dunham, J., Chandler, G., Rieman, B.E., and Martin, D.J., 2005. Measuring stream temperature with digital data loggers: A user's guide. General Technical Report RMR-SGTR-150WWW, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO. https://doi.org/10.2737/RMRS-GTR-150.

Ebersole, J.L., Quiñones, R.M., Clements, S., Letcher, B.H., 2020. Managing climate refugia for freshwater fishes under an expanding human footprint. Front. Ecol. Environ. 18 (5), 271–280. https://doi.org/10.1002/fee.2206.

Fernandes, R., G. Leblanc, S., 2005. Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. Rem. Sens. Environ. 95 (3), 303–316. https://doi.org/10.1016/j.rse.2005.01.005.

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

Hrachowitz, M., Soulsby, C., Imholt, C., Malcolm, I.A., Tetzlaff, D., 2010. Thermal regimes in a large upland salmon river: a simple model to identify the influence of landscape controls and climate change on maximum temperatures. Hydrol. Process. 24 (23), 3374–3391. https://doi.org/10.1002/hyp.7756.

Isaak, D.J., Luce, C.H., Chandler, G.L., Horan, D.L., Wollrab, S.P., 2018. Principal components of thermal regimes in mountain river networks. Hydrol. Earth Syst. Sci. 22, 6225-6240. https://doi.org/10.5194/hess-22-6225-2018.

Isaak, D.J., Luce, C.H., Horan, D.L., Chandler, G.L., Wollrab, S.P., Dubois, W.B., Nagel, D. E., 2020. Thermal regimes of perennial rivers and streams in the Western United States. J Am Water Resour Assoc 56 (5), 842–867. https://doi.org/10.1111/1752-1688.12864.

Jastram, J.D., Snyder, C.D., Hitt, N.P., Rice, K.C., 2013. Synthesis and interpretation of surface-water quality and aquatic biota data collected in Shenandoah National Park, Virginia, 1979-2009: U.S. Geological Survey Scientific Investigations Report 2013-5157, 77 p. https://pubs.usgs.gov/sir/2013/5157/.

Johnson, S.L., Jones, J.A., 2000. Stream temperature responses to forest harvest and debris flows in western Cascades, Oregon. Can. J. Fish. Aquat. Sci. 57 (S2), 30–39. https://doi.org/10.1139/f00-109.

Johnson, Z.C., Johnson, B.G., Briggs, M.A., Devine, W.D., Snyder, C.D., Hitt, N.P., Hare, D.K., Minkova, T.V., 2020. Paired air-water annual temperature patterns reveal hydrogeological controls on stream thermal regimes at watershed to continental scales. J. Hydrol. 587, 124929. https://doi.org/10.1016/j.jhydrol.2020.124929.

Johnson, Z.C., Snyder, C.D., Hitt, N.P., 2017. Landform features and seasonal precipitation predict shallow groundwater influence on temperature in headwater streams. Water Resour. Res. 53 (7), 5788–5812. https://doi.org/10.1002/2017WR020455.

Jones, N.E., Schmidt, B.J., 2018. Thermal regime metrics and quantifying their uncertainty for North American streams. River Res. Appl. 34 (4), 382–393. https://doi.org/10.1002/rra.3257.

Kędra, M., Wiejaczka, Ł., 2018. Climatic and dam-induced impacts on river water temperature: assessment and management implications. Sci. Total Environ. 626, 1474–1483. https://doi.org/10.1016/j.scitotenv.2017.10.044.

Kothandaraman, V., 1971. Analysis of water temperature variations in large River. J. Sanitary Eng. Division 97, 19–31.

Kurylyk, B.L., Irvine, D.J., Bense, V.F., 2019. Theory, tools, and multidisciplinary applications for tracing groundwater fluxes from temperature profiles. WIREs Water 6 (1), e1329. https://doi.org/10.1002/wat2.1329.

Leach, J.A., Moore, R.D., 2014. Winter stream temperature in the rain-on-snow zone of the Pacific Northwest: influences of hillslope runoff and transient snow cover. Hydrology and Earth System Sciences 18, 819–838. https://doi.org/10.5194/hess-18-819-2014.

Letcher, B.H., Hocking, D.J., O'Neil, K., Whiteley, A.R., Nislow, K.H., O'Donnell, M.J., 2016. A hierarchical model of daily stream temperature using air-water temperature synchronization, autocorrelation, and time lags. PeerJ 4, e1727. https://doi.org/10.7717/peerj.1727.

Li, H., Deng, X., Smith, E., 2017. Missing data imputation for paired stream and air temperature sensor data: missing data imputation for stream and air temperature. Environmetrics 28 (1), e2426. https://doi.org/10.1002/env.2426.

Maheu, A., Poff, N.L., St-Hilaire, A., 2016a. A Classification of stream water temperature regimes in the conterminous USA: classification of stream temperature regimes. River Res. Appl. 32 (5), 896–906. https://doi.org/10.1002/rra.2906.

Maheu, A., St-Hilaire, A., Caissie, D., El-Jabi, N., 2016b. Understanding the thermal regime of rivers influenced by small and medium size dams in Eastern Canada: thermal regime of rivers influenced by small and medium size dams. River Res. Appl. 32 (10), 2032–2044. https://doi.org/10.1002/rra.3046.

Marsha, A., Steel, E.A., Fullerton, A.H., Sowder, C., 2018. Monitoring riverine thermal regimes on stream networks: insights into spatial sampling designs from the Snoqualmie River, WA. Ecol. Indic. 84, 11–26. https://doi.org/10.1016/j.ecolind.2017.08.028.

Martens, K.D., Devine, W.D., Minkova, T.V., Foster, A.D., 2019. Stream conditions after 18 years of passive riparian restoration in small fish-bearing watersheds. Environ. Manage. 63 (5), 673–690. https://doi.org/10.1007/s00267-019-01146-x.

McNyset, K., Volk, C., Jordan, C., 2015. Developing an Effective Model for Predicting Spatially and Temporally Continuous Stream Temperatures from Remotely Sensed Land Surface Temperatures. Water 7, 6827–6846. https://doi.org/10.3390/w7126660.

Omernik, J.M., 1987. Ecoregions of the conterminous United States. Ann. Assoc. Am. Geogr. 77 (1), 118–125. https://doi.org/10.1111/j.1467-8306.1987.tb00149.x.

Omernik, J.M., Griffith, G.E., 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. Environ. Manage. 54 (6), 1249–1266. https://doi.org/10.1007/s00267-014-0364-1.

PRISM Climate Group, 2019. Oregon State University [WWW Document]. accessed 10.1.19. http://prism.oregonstate.edu.

Snyder, C.D., Hitt, N.P., Young, J.A., 2015. Accounting for groundwater in stream fish thermal habitat responses to climate change. Ecol. Appl. 25 (5), 1397–1419. https://doi.org/10.1890/14-1354.1.

Sowder, C., Steel, E.A., 2012. A Note on the Collection and Cleaning of Water Temperature Data. Water 4, 597–606. https://doi.org/10.3390/w4030597.

Steel, E.A., Beechie, T.J., Torgersen, C.E., Fullerton, A.H., 2017. Envisioning, Quantifying, and Managing Thermal Regimes on River Networks. BioScience 67, 506–522. https://doi.org/10.1093/biosci/bix047.

Steel, E.A., Tillotson, A., Larsen, D.A., Fullerton, A.H., Denton, K.P., Beckman, B.R., 2012. Beyond the mean: the role of variability in predicting ecological effects of stream temperature on salmon. Ecosphere 3 (11), art104. https://doi.org/10.1890/ES12-00255.1.

U.S. Geological Survey USGS Current Water Data for the Nation [WWW Document]. Current Water Data for the Nation 2020 (accessed 6.2.20).

Ward, J.C., 1963. Annual variation of stream water temperature. J. Sanitary Eng. Division 89, 1–16.

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

Wondzell, S.M., Diabat, M., Haggerty, R., 2019. What matters most: are future stream temperatures more sensitive to changing air temperatures, discharge, or riparian vegetation? J. Am. Water Resour. Assoc. 55 (1), 116–132. https://doi.org/10.1111/1752-1688.12707.