A peer-reviewed version of this preprint was published in PeerJ on 29 February 2016.

<u>View the peer-reviewed version</u> (peerj.com/articles/1727), which is the preferred citable publication unless you specifically need to cite this preprint.

Letcher BH, Hocking DJ, O'Neil K, Whiteley AR, Nislow KH, O'Donnell MJ. 2016. A hierarchical model of daily stream temperature using air-water temperature synchronization, autocorrelation, and time lags. PeerJ 4:e1727 <u>https://doi.org/10.7717/peerj.1727</u>

A robust hierarchical model of daily stream temperature using air-water temperature synchronization, autocorrelation, and time lags

Benjamin H Letcher, Daniel J Hocking, Kyle O'Neill, Andrew R Whiteley, Keith H Nislow, Matthew J O'Donnell

Water temperature is a primary driver of stream ecosystems and commonly forms the basis of stream classifications. Robust models of stream temperature are critical as the climate changes, but estimating daily stream temperature poses several important challenges. We developed a statistical model that accounts for many challenges that can make stream temperature estimation difficult. Our model identifies the yearly period when air and water temperature are synchronized, accommodates hysteresis, incorporates time lags, deals with missing data and autocorrelation and can include external drivers. In a small stream network, the model performed well (RMSE = 0.59 °C), identified a clear warming trend (0.063 °C \cdot y⁻¹) and a widening of the synchronized period (2.9 d \cdot y⁻¹). We also carefully evaluated how missing data influenced predictions. Missing data within a year had a small effect on performance (~ 0.05% average drop in RMSE with 10% fewer days with data). Missing all data for a year decreased performance (~ 0.6 °C jump in RMSE), but this decrease was moderated when data were available from other streams in the network. Straightforward incorporation of external drivers (e.g. land cover, basin size) should allow this modeling framework to be readily applied across multiple sites and at multiple spatial scales.

1	
2	
3	
4	A robust hierarchical model of daily stream temperature using air-water temperature
5	synchronization, autocorrelation, and time lags
6	
7	
8	Benjamin H. Letcher ¹ , Daniel J. Hocking ¹ , Kyle O'Neill ¹ , Andrew R. Whiteley ² , Keith H.
9	Nislow ³ , Matthew J. O'Donnell ¹
10	
11	
12	¹ S.O. Conte Anadromous Fish Research Center
13	US Geological Survey/Leetown Science Center
14	Turners Falls, Massachusetts 01376 USA
15	
16	² Department of Environmental Conservation
17	University of Massachusetts, Amherst, MA 01003-4210 USA
18	
19	³ Northern Research Station
20	USDA Forest Service
21	University of Massachusetts, Amherst, MA 01003-4210 USA
22	
23	
24	Corresponding author: Benjamin H. Letcher, bletcher@usgs.gov
25	

26 Abstract

27 Water temperature is a primary driver of stream ecosystems and commonly forms the 28 basis of stream classifications. Robust models of stream temperature are critical as the climate changes, but estimating daily stream temperature poses several important 29 challenges. We developed a statistical model that accounts for many challenges that can 30 31 make stream temperature estimation difficult. Our model identifies the yearly period when air and water temperature are synchronized, accommodates hysteresis, incorporates time 32 lags, deals with missing data and autocorrelation and can include external drivers. In a 33 small stream network, the model performed well (RMSE = 0.59 °C), identified a clear 34 warming trend (0.063 °C \cdot y⁻¹) and a widening of the synchronized period (2.9 d \cdot y⁻¹). We 35 also carefully evaluated how missing data influenced predictions. Missing data within a 36 year had a small effect on performance ($\sim 0.05\%$ average drop in RMSE with 10% fewer 37 days with data). Missing all data for a year decreased performance (~ 0.6 °C jump in 38 39 RMSE), but this decrease was moderated when data were available from other streams in the network. Straightforward incorporation of external drivers (e.g. land cover, basin size) 40 41 should allow this modeling framework to be readily applied across multiple sites and at 42 multiple spatial scales.

43

44 Introduction

Accurate stream temperature predictions are increasingly important as human impacts on 45 streams and on the climate accelerate stream temperature change (Kaushal et al., 2010; 46 47 Rice & Jastram, 2014). Human activities influence stream temperatures directly via increased water withdrawals, altered channel engineering and dam operation (Poole & 48 Berman, 2001) and indirectly by altering landscape features (e.g. riparian cover) and by 49 50 affecting air temperatures at broad spatial scales via climate change (Hayhoe et al., 2007; Huntington et al., 2009). Understanding how stream temperatures are changing over time 51 and space and the ability to forecast future temperatures are important because stream 52 temperatures directly influence stream ecosystems (Quinn et al., 1994; Wenger et al., 53 54 2011) and because regulatory agencies commonly use stream temperature as a metric for

55 managing streams and their watersheds (e.g. Beauchene et al., 2014). Altered stream

56 temperatures are likely to have profound effects on the abundance and distribution of

57 stream biota (Isaak & Rieman, 2012; Eby et al., 2014), especially coldwater, ectothermic

58 species because many physiological and demographic rates are temperature-dependent

59 (Fry, 1971; Elliott & Elliott, 2010; Letcher et al., 2015).

60 The general importance of stream temperature has prompted the development of a

number of models for stream temperature (e.g. Mohseni, Stefan & Erickson, 1998; Caissie,

62 El-jabi & Satish, 2001; Hague & Patterson, 2014; Sun et al., 2014; Li et al., 2014). Stream

63 temperature models vary along several important gradients, including model type

64 (physical-statistical), temporal resolution (daily-yearly) and spatial resolution (local-broad

65 spatial coverage). As with all models of complex systems, tradeoffs among these gradients

66 usually limit models to highly-detailed, local models (Brown, 1969; Kim & Chapra, 1997;

67 Younus, Hondzo & Engel, 2000) or simple, general models (e.g. Crisp & Howson, 1982). The

68 detailed, local models typically produce good accuracy (RMSE ~ 1.0 °C) but may not predict

69 temperatures well outside of the local area, while the simple models generate moderate to

70 poor accuracy (RMSE ~ 1.5 °C to 3.0 °C) across a broad spatial range. Models that

71 aggregate over longer time intervals generally perform better (Stefan & Preud'homme,

1993; Pilgrim, Fang & Stefan, 1998; Webb, Clack & Walling, 2003; Morrill, Bales & Conklin,

73 2005), but even hourly models can perform well (Kanno, Vokoun & Letcher, 2013). A

careful consideration of six key temperature modeling issues may provide the basis for the

75 development of daily stream temperature models of medium complexity that provide good

76 predictions across space.

77 First, the relationship between air temperature and stream temperature is non-linear at high and low air temperatures (Mohseni, Stefan & Erickson, 1998), but for different 78 79 reasons. At high air temperatures, evaporative cooling slows warming of stream water, 80 while at low air temperatures, air temperatures can dip well below the water temperature freezing limit (Caissie, 2006; Webb et al., 2008). Air and water temperatures are no longer 81 synchronized when air temperatures are near and below 0 °C, which can generate a poor 82 83 relationship between air and stream temperatures and heterogeneity of variance across 84 temperatures. Many simple statistical models use a non-linear model to describe the

relationship between air and stream temperature (Mohseni, Stefan & Erickson, 1998; 85 Webb, Clack & Walling, 2003; Kanno, Vokoun & Letcher, 2013). Others use a linear model 86 and limit analysis to the summer (Hilderbrand, Kashiwagi & Prochaska, 2014) or to the ice-87 88 free period of the year (Stefan & Preud'homme, 1993; Erickson & Stefan, 2000), in an 89 attempt to avoid the non-linear portions of the air-water temperature relationship. Time series (Caissie, El-Jabi & St-Hilaire, 1998; Caissie, El-jabi & Satish, 2001; Benyahya et al., 90 91 2007) or non-parametric models (Benyahya et al., 2008; Li et al., 2014) of stream 92 temperature trends over time that include air temperature as a predictor as well as local, physical models (e.g. Sinokrot & Stefan, 1993) can accommodate the non-linearity. 93

94 Second, accuracy can be improved when models account for hysteresis, a different relationship between air and water temperature in the spring (rising temperatures) vs. the 95 fall (falling temperatures) (Mohseni, Stefan & Erickson, 1998; Caissie, El-jabi & Satish, 96 97 2001; Webb, Clack & Walling, 2003). Seasonal hysteresis is often caused by influx of cool snow melt or rain water in the spring (Lisi et al., 2015) which depresses spring stream 98 99 temperature/air temperature relationships relative to fall stream temperature/air temperature relationships (Webb & Nobilis, 1997). Mohseni et al. (1998) observed that 100 43% of their study streams exhibited hysteresis; they addressed hysteresis by fitting 101 separate non-linear curves to the rising and falling seasonal temperatures. Time series 102 103 models with non-symmetric seasonal functions account for hysteresis by default (e.g. Li et al., 2014). 104

Third, due to thermal inertia, stream temperature does not respond instantaneously to
changes in air temperature. Including lags in air temperature effects can improve estimates
for models with short time scales (Benyahya et al., 2008; Webb, Stewardson & Koster,
2010). The effects of time lags increase with stream depth (Stefan & Preud'homme, 1993)
and stream flow (Smith & Lavis, 1975; Webb, Clack & Walling, 2003). Time lags are a key
component of time series modeling (Shumway & Stoffer, 2006).

111 Fourth, while the amount of stream temperature data available worldwide is increasing

112 very rapidly (Webb et al., 2008), many sites have incomplete data. Very few study regions

113 have a complete matrix of sample sites and years: data may be missing for an entire year at

a site or may be incomplete within a year. Incomplete within-year data will have variable 114 effects on estimation depending on the extent and timing of the missing data. Effects of 115 missing data will also depend on model type. For simple linear models, within-year missing 116 117 data may not have a large effect on estimation because of the linear relationship between 118 stream and air temperature. For non-linear models, missing data could have dramatic effects on estimation as missing data fail to 'anchor' the curve. Other modeling approaches, 119 120 such as time series models, machine learning models (DeWeber & Wagner, 2014), and models with varying coefficients (Li et al., 2014) may be less sensitive to missing data. In 121 general, hierarchical models with random effects across space (sites, stream networks or 122 regions) and time (months, seasons, or years) can accommodate missing data as they 123 124 'borrow information' across units (Wagner, Hayes & Bremigan, 2006; Gelman & Hill, 2007).

Fifth, spatial and temporal autocorrelation can cause estimation problems (Caissie, 2006; 125 Benyahya et al., 2007; Hague & Patterson, 2014). Autocorrelation occurs when data points 126 127 in space or time are not independent, i.e. close points are similar or dissimilar to each other simply because they are close. For example, downstream temperatures can be similar to 128 129 upstream temperatures because water flows downstream or today's temperature can be similar to yesterday's temperature due to the combination of high heat capacity of water, 130 low density and heat transfer from air, and conduction of heat from surrounding 131 environment (i.e. thermal inertia) (Caissie, El-Jabi & St-Hilaire, 1998; Isaak et al., 2014). 132 This is a very common issue in estimation and a variety of time series models can 133 accommodate temporal autocorrelation (Shumway & Stoffer, 2006) and some newer 134 135 approaches are now available to deal with spatial autocorrelation (Peterson & Ver Hoef, 2010; Rushworth et al., 2015). 136

Finally, air temperature is not the only important predictor of stream temperature (Webb,
1996; Caissie, 2006). Many regression-based models have evaluated effects of landscape
and environmental drivers on stream temperatures (Hawkins et al., 1997; Isaak & Hubert,
2001; Hill, Hawkins & Carlisle, 2013). Important landscape drivers typically include
topography, riparian cover, impervious surface, and stream depth (Poole & Berman, 2001)
and environmental drivers often include stream flow, snow melt, groundwater input, and
humidity (Taylor et al., 2013; Lisi et al., 2015; Snyder, Hitt & Young, 2015). It is

straightforward to incorporate external drivers beyond air temperature into most classesof stream temperature models (Hague & Patterson, 2014).

146 Here, we develop a model for mean daily stream temperature that improves accuracy of statistical models by addressing most of the issues listed above. To avoid fitting a 147 relationship between stream and air temperature when there is none (e.g. winter), we 148 149 develop a metric that limits estimation to the days of the year that stream temperature and air temperature are synchronized (roughly spring to fall). This metric is flexible among 150 vears and sites. To address hysteresis, we estimate a non-symmetrical trend across the 151 synchronized days with a hierarchical structure to accommodate missing data. We also add 152 153 an autoregressive term to the model to deal with temporal autocorrelation and we estimate spatial covariance to accommodate spatial autocorrelation. Because data presented here 154 are spatially constrained to four sites in a small network, we do not include landscape 155 variables in the model, although their addition is straightforward in the model structure. 156 157 The two environmental drivers in the model are air temperature and stream flow. In addition to presenting the model, we analyze trends in estimates over time and conduct a 158 159 detailed missing observations analysis.

160 Methods

161 Study area

The study site was located in western Massachusetts, USA (42^o 25' N; 72^o39' W, Fig. 1) and 162 consisted of a third-order mainstem (West Brook, WB) and three second-order tributaries 163 (Open large, OL; Open small, OS; Isolated large, IL). A dense canopy of mixed hardwood 164 165 with some hemlocks provides cover throughout the watershed. Watershed area above our study area is 11.8 km² and landuse in the area is limited residential with some farming. 166 Average stream width of the WB is 4.5 m and is between 1-3 m for the tributaries. Water is 167 168 stored in two of the streams; a drinking water reservoir is upstream of the WB, and a large beaver dam complex is above OS (Fig. 1). OL and IL were free-flowing during the course of 169 the study. 170

171 We deployed four temperature loggers (± 0.1 C; Onset Computer Corporation, Pocasset, MA, USA, and ± 0.05 C; Solinst Canada Ltd., Georgetown, ON) in permanently watered 172 sections of the study area. All loggers recorded data every 15 minutes throughout the year. 173 174 The logger in the WB was deployed 1998 to 2013 and the loggers in the tributaries were 175 out from 2002 to 2013. We do not have continuous air temperature measurements from 2002 to 2013, so we used air temperature estimates for our study area from Daymet 176 177 (http://daymet.ornl.gov/). For the years that we do have West Brook air temperature data (2008-2013), the relationship between West Brook and Daymet air temperatures was 178 strong (p-value < 10^{-16} , $r^2 = 0.91$), suggesting that Daymet air temperatures are a good data 179 source for the study site. Additionally, stream water is thermally controlled by energy 180 181 sources over a large area, so the air temperatures in Daymet may have a stronger 182 relationship with water temperatures compared to any local, single-point air temperature measurement. Stream flow was estimated using a flow extension model (Nielsen, 1999) 183 based on data from a nearby USGS stream gage (Mill River, Northampton, MA, U.S.A.). See 184 (Xu, Letcher & Nislow, 2010) for details. 185

186

187 Statistical analysis

Descriptive statistics. As a coarse comparison of daily water temperatures, we calculated correlations among sites. We also explored patterns in water temperature over time and among sites by comparing cumulative residuals from a spline fit to all the data (function gam() in R, Fig. 2). We calculated residuals for each water temperature data point and then developed empirical cumulative curves over days of the year for each year and site combination.

Breakpoints. The goal is to develop a robust model for the relationship between mean daily water and mean daily air temperature. A key limitation in developing this relationship is that lower water temperatures in the winter are bounded near 0 °C while air temperatures are not. This means that water and air temperatures can become decoupled when air temperatures are cold resulting in only a weak relationship, at best, between water and air temperature. In contrast, as air temperatures warm in the spring and before they get too cold in the autumn, water and air temperatures can be synchronized (Fig. 3 above),

suggesting the possibility of a strong relationship between water and air temperatureduring the synchronized portion of each year.

203 The key to the approach is identifying a breakpoint in the spring when water and air temperature become synchronized and a breakpoint in the autumn when temperatures 204 become desynchronized. To identify the synchronization breakpoints we calculated a 205 simple index (waterT – airT)/waterT (waterT > 0), where waterT was mean daily water 206 temperature and airT was mean daily air temperature (Fig. 3). This temperature index 207 (tempIndex) approaches 0 when water and air temperature are similar and is very 208 different from 0 when temperatures diverge (Fig. 3). While water and air temperatures are 209 210 synchronized, tempIndex flattens out (Fig. 3b), providing the opportunity to identify the beginning and end (breakpoints) of the flat period. 211

212 To identify the spring and autumn breakpoints, we used a runs analysis that determined 213 the first (spring) and last (autumn) day of the year that the tempIndex was consistently 214 within the flat period (Fig. 3). We established the range of tempIndex values that 215 comprised the flat period by calculating the 99.9% confidence interval (CI) for tempIndex 216 using the middle 150 days of the year (late April to mid-September). The middle 150 days of the year were always within the flat period based on visual observation of tempIndex 217 plots. Separate CI values were calculated for each year and stream. For the breakpoint 218 219 estimation, we used a moving average for tempIndex with a centered 10-day window to 220 help stabilize tempIndex values near the breakpoints. Temperatures were considered 221 synchronized when 10 consecutive days of the moving average fell within the 99.9% CI. Beginning on day 1 and moving towards day 150, the first time 10 consecutive days were 222 synchronized was used as the spring breakpoint and we moved from the end of the year to 223 day 150 to establish the fall breakpoint. Numbers of days in the synchronized period for 224 225 each stream and year are shown in Table 1.

226 We evaluated trends in fall and spring breakpoints by running three linear models with

227 breakpoint day of the year as the dependent variable and year alone or year + stream or

228 year * stream as independent variables. We estimated AIC to determine the most

229 parsimonious model.

NOT PEER-REVIEWED

Peer Preprints

230 *Water temperature model description*. With breakpoints established for each year and site,

- 231 we modeled the relationship between water temperature and air temperature for the
- 232 synchronized period using a hierarchical linear autoregressive model with a cubic trend
- across days within a year and covariation among sites. We fit the model using a Bayesianapproach.
- Observed water temperature $(t_{s,d,y})$ for each site $(s; s_1 = WB, s_2 = OL, s_3 = OS, s_4 = Is)$, day of year (*d*) and year (*y*) was assumed to derive from a normal distribution with mean $\mu_{s,d,y}$ and standard deviation *sd* (residual model error):

238
$$t_{s,d,y} \sim N(\mu_{s,d,y}, sd)$$
 Equation 1

We used a non-informative uniform prior [0,10] for *sd*. We modeled the mean with a linear trend ($\omega_{s,d,y}$) adjusted by an AR(1) autoregressive coefficient (δ_s) on the residual error from the previous day:

242
$$\mu_{s,d,y} = \omega_{s,d,y} + \delta_s(t_{s,d-1,y} - \omega_{s,d-1,y})$$
 Equation 2

243 We placed a hierarchical structure on δ_s :

244
$$\delta_s \sim N(\mu \delta_s, sd\delta) T(-1,1)$$
 Equation 3

where site-specific δ_s were drawn from a truncated normal distribution with mean $\mu \delta_s$ and standard deviation $sd\delta$. Values for δ_s were truncated to keep them within the admissible range for a correlation. Priors for the mean and standard deviation were non-informative; $\mu \delta_s \sim U(-1,1)$, and $sd\delta \sim U(0,2)$ (an upper limit of 2 for $sd\delta$ is non-informative for the truncated data).

When observed temperature data were not available for the previous day (beginning of a series or following a break in the series) we modeled the mean without the autoregressive component:

253
$$\mu_{s,d,y} = \omega_{s,d,y}$$
 Equation 4

PeerJ PrePrints | https://doi.org/10.7287/peerj.preprints.1578v1 | CC-BY 4.0 Open Access | rec: 10 Dec 2015, publ: 10 Dec 2015

254 We modeled the linear component with a combination of fixed and random effects:

255
$$\omega_{s,d,y} = \alpha + \beta_1 T_{s,d,y} + \beta_2 T_{s,d-1,y} + \beta_3 T_{s,d-2,y} + \beta_4 F_{s,d,y} + \beta_5 T_{s,d,y} \cdot F_{s,d,y} + \beta_{6:8} s + \beta_{9:11}$$
256 $s \cdot T_{s,d,y} + Y_y$ Equation 5

257 where α is the overall intercept, the β are the coefficients for the fixed effects (*T* is mean

- daily air temperature, F is mean daily stream flow, s is site) and Y_v represents random
- effects among years. Priors for the $\beta_{1:11}$ were independent and non-informative, N(0,100).
- 260 Y_v represented random effect temporal trends (cubic) across years where:

261
$$Y_{y} = \alpha_{y} + \beta_{12,y} D_{s,y,d} + \beta_{13,y} D_{s,y,d}^{2} + \beta_{14,y} D_{s,y,d}^{3}$$
 Equation 6

262 For convenience, this equation can be written in matrix notation as

263
$$Y_y = X_{s,d,y} B_y$$
 Equation 7

where X is a data matrix with *l* columns (l = 4; the number of year-level predictors) with the first column a vector of 1's for the intercept and B_y is the *y* x *l* matrix of year-level regression coefficients. Priors for the mean were non-informative, with

267
$$B_{\nu} \sim MVN(M_{\nu}\Sigma)$$
 Equation 8

where $M_l = (\mu_{\alpha}, \mu_{\beta 12}, \mu_{\beta 13}, \mu_{\beta 14})$ is a vector of length *l*, representing the mean of the distribution of intercept and slopes. The *l* x *l* covariance matrix is represented by Σ where the variance of each regression coefficient is on the diagonal and the covariance on the offdiagonals. The hyperprior for the means were non-informative with $\mu_{\alpha} = 0$ and $\mu_{\beta 12}, \mu_{\beta 13},$ $\mu_{\beta 14} \sim N(0,100)$. Standard deviation priors were also non-informative and were drawn from an inv-Wishart distribution:

274
$$\Sigma \sim inv - wish(diag(l), l+1).$$
 Equation 9

275 Parameter estimation

We used the program JAGS (http://mcmc-jags.sourceforge.net) to code the model and to draw posterior samples of the parameters (see supplemental material for JAGS code). We

called JAGS from R (3.1.2) using the package 'rjags' (V 3-14)(Plummer, 2014). We ran three

- chains with 1000 burn-in and 2500 evaluation iterations. Chains were thinned to keep
- 280 every fifth iteration. We checked convergence using the 'potential scale reduction factor'
- 281 (Brooks & Gelman, 1998) from the 'coda' package in R (Plummer et al., 2006) and also
- assessed chains visually.
- 283 Model assessment
- 284 Goodness of Fit and prediction. We assessed goodness of fit in two ways. First, we compared
- observed and predicted values for the complete dataset. Second, we ran a series of cross
- validation tests where we randomly left out a portion of the water temperature data,
- 287 estimated parameters with the remaining (training) data and compared predictions of the
- left out (testing) data to original values. This involved leave-p-out cross-validation where
- we randomly left out a proportion (p) of the data, where p = 0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5,
- 290 0.6, 0.7, 0.8. We ran 10 replicates for each value of p. For each condition, we calculated the
- 291 root mean square error (RMSE) of the residuals for the training and the test data sets.
- 292 Missing data

We also ran a series of tests to ask how the quantity, timing, and location of missing data influenced model performance (estimation and prediction). These tests can be used to help understand performance and to help design monitoring strategies. This set of analyses differed from the leave-p-out cross-validation (above) because data were not left out randomly. Rather, consecutive days of data were left out, either within a year or across streams, reflecting the character of missing field data.

299 <u>Quantity</u>: To evaluate how increasing the number of sampling days within a year affects 300 estimation and prediction, we left out increasing numbers of days on either side of the 301 median sampling date for each stream and year combination. Specifically, we started with 302 complete data and then conducted nine sets of runs where we left out data $15 \cdot d$ days from 303 the beginning and $15 \cdot d$ days from the end of each time series (where d = 1 to 9), generating 304 shorter time series by 30 days for each scenario.

305 <u>When</u>: We assessed how changing the timing of missing data affected predictions by

306 shifting the window of available data from the beginning to the end of the synchronized

307 period. To do this, we left out data for all but 30 consecutive days at a time for 13 non-

308 overlapping scenarios with scenario one starting at day of year 70 and scenario 13 starting

309 at day of year 310.

310 <u>Where</u>: To evaluate how well the model predicted stream temperatures when data were

311 missing from one or more streams, we ran the above analyses leaving out data yearly from

all streams or just the West Brook. For years with data just from the West Brook (1999 –

313 2002), we removed all data a year at a time. For years with data from the tributaries and

the West Brook (2003 - 2013), we either removed all the data for each year (all four

315 streams) or just the data from the West Brook for each year. Removing all the data for a

316 given year tests how well the model predictions work when there are no data for the year

317 (but there are data for other years), while removing data for just the West Brook tests how

318 well predictions work when data are missing for a stream (but there are data for other

319 streams and years).

For all tests, we compared RMSE of the residuals for the test (left out) data to the RMSE of the residuals of the full training set (base case).

322 **Results**

323 Descriptive statistics. Evaluation of the descriptive statistics suggested that water

324 temperatures were similar for OL and IL and for WB and OS and that the streams appear to

325 be warming over the duration of the study. Correlations of daily water temperatures

among the four sites were all between 0.96 and 0.97, except for the correlation between OL

and IL (0.99). Patterns in the cumulative water temperature residuals were generally

328 similar for WB and OS, with cooler years in the beginning of the time series and warmer

329 years later (Fig. 4). OS demonstrated the warmest temperatures, especially in 2010-2012.

330 Patterns were remarkably similar between OL and IL, also demonstrating generally cooler

temperatures earlier in the data time series (Fig. 4). Monthly distributions of water

temperature were highly variable across years and streams (Fig. A1).

333 *Parameter estimation*: Potential scale reduction factor (R-hat) values for all parameters

- were less than 1.01, indicating good convergence (Brooks & Gelman, 1998). Parameter
- estimates gave an overall mean of 15.1, with strong air temperature effects (1.52 unlagged,
- 0.20 lagged 1 day, 0.15 lagged 2 days), a positive effect of stream flow (0.36), and strong
- 337 site differences (OL = -0.50, OS = 0.59, IL = -0.54) (Fig. 5 and Table A1). The autoregressive
- 338 mean equaled 0.79 and there was little variation in the autoregressive terms among sites
- (Fig. 5). The estimate for residual model error from Eq. 1 was 0.77.
- 340 *Model assessment: Goodness of Fit and prediction*. Using the full dataset, predicted values
- 341 were very similar to observed values (Fig. S2). The slope of the relationship was 0.99 (s.e. =
- 342 0.0064) with an intercept of 0.15 (s.e. = 0.089) and an R^2 of 0.98. Overall RMSE was 0.59 ±
- 343 0.09 (Table 2). For the cross-validation tests where we randomly left out 30% of the data,
- 344 the RMSE increased to 0.69 ± 0.003 for the training data and to 0.86 ± 0.010 for the test
- 345 data (Table 2).
- Across a broader range of data randomly left out (0 to 0.8), the RMSE for the test data
- increased approximately linearly with a 0.025 increase in RMSE for each 0.1 increase in
- proportion of data left out ($r^2 = 0.98$; Fig. S3). RMSE for the training data set was largely
- insensitive to the proportion of data left out and had a mean value of 0.86 (s.d. = 0.016, Fig.S3).
- Break point trends. Break points appear to be getting later in the year in the fall and earlier 351 352 in the year in the spring (Fig. 6). In the fall, delta AIC values for the linear models were all within two so we selected the simplest model (year only). In the spring, the delta AIC value 353 for the simplest model was 5.2 so we also selected the simplest model (year only). Both 354 355 seasons showed significant changes in breakpoints over years (estimate = 1.33, F (1,40) = 4.68, p-value = 0.036 fall; estimate = -1.61, F (1.40) = 9.13, p-value = 0.0045 spring), but 356 year explained only 10% (fall) or 19% (spring) of the variation in the relationship. The 357 358 parameter estimates indicated that breakpoints are 1.6 days earlier in the spring and 1.3 days later per year in the fall, generating an estimated widening of the breakpoint window 359 of 2.9 days per year. 360

361 *Trends in cubic functions.* Predicted mean water temperatures based on the cubic function

- 362 (Eq. 7) varied among years (Fig. S4), mirroring the general trend in the raw data (Fig. 2).
- 363 Yearly maximum water temperature (white dots in S4) increased over the course of the
- study (F=5.34, df=1,13, p-value=0.037, $R^2 = 0.24$), with an estimated annual increase of
- 365 0.063 °C (Fig. 7, above). In contrast, the day of year of the temperature maximum did not
- change over the course of the study (F=0.030, df=1,13, p-value=0.86)(Fig. 7, below).
- 367 *Missing data: Quantity.* Adding more data to either side of the median date improved
- 368 predictions of the test data (filled circles in Fig. 8). The slope of the regression (-0.46, s.e. =
- 369 0.035) indicated that a 10% increase in data resulted in a reduction in RMSE of 0.046.
- *Missing data: When.* The timing of data availability had a threshold effect on RMSE, with
 relatively high and variable RMSE before day 160 and consistent lower RMSE after day 160
- 372 (triangles in Fig. 9).

373 *Missing data: Where*. Compared to the base case (all data included), leaving data out of the

- 374 estimation one year at a time resulted in a mean increase in RMSE of 0.48 °C when just the
- WB data were left out and a mean increase of 0.57 °C when data for all four streams were left out (Table 2).
- As the amount of data was increased on either side of the median date, RMSE increased less from the base case when data were available from the other three streams than when data were not available for any of the streams. The slope of the relationship between the proportion of days included in the training data and the difference in mean RMSE was -0.46 when just the WB data were left out and the slope was -0.63 (s.e. = 0.20) when data from all streams were left out (Fig. 8). The slopes suggest either a 0.046 or a 0.063 decrease in RMSE with a 10% increase in days included in the estimation.
- When data were available for only 30 days, but the 30-day window of availability varied across the year, the presence of data from the other streams eliminated the variability in RMSE across scenarios (compare circles to triangles in Fig. 9). The resulting increase in RMSE was about 0.38 across 30-day window scenarios when data were present from other streams.

389 Discussion

390 We present a statistical model that accounts for many issues that can make stream 391 temperature estimation difficult. Our model limits analysis to days when air and water temperature are synchronized, accommodates hysteresis, incorporates time lags, can deal 392 with missing data and autocorrelation and can include external drivers. The result is quite 393 low bias with complete data (RMSE = 0.59 °C), and bias remains low (RMSE <1 °C) when 394 data from streams or years are missing. While we evaluated model performance for a 395 single small stream system, it is straightforward to extend the model to a broader spatial 396 scale to take full advantage of the rapidly increasing amount of available stream 397 398 temperature data.

399 A key feature of our model is a flexible way to identify the portion of days spring-to-fall when stream and air temperatures are synchronized. The air-water temperature 400 401 relationship breaks down during the winter, primarily, due to phase change 402 thermodynamics, insulating ice cover, snow melt, and other physical processes. Previous 403 researchers have omitted modeling winter temperatures or focused solely on summer 404 temperatures (Kanno, Vokoun & Letcher, 2013; e.g. DeWeber & Wagner, 2014; Snyder, Hitt & Young, 2015). However, defining the "winter" period that causes deviations in the air-405 water relationship depends on the conditions in a specific year and location; therefore, just 406 407 excluding the winter months based on calendar dates (21 December – 20 March in the northern hemisphere) is an imprecise cutoff with the potential to bias the model and the 408 409 resulting inference. For example, just as the amount of snow and duration of ice cover differs at 40° and 45° latitude, the physical properties that affect the air-water relationship 410 vary annually and from one location to another depending on the exact landscape 411 characteristics of the site, even when compared to nearby locations (Lisi et al., 2015). 412 Additionally, taking the opposite approach and limiting analyses to the summer period 413 excludes large amounts of data and prevents inference during other times of the year, 414 415 which are important in biological and biogeochemical processes. Our method of calculating the period of the year where the air-water relationship is synchronized alleviates these 416 417 issues of arbitrarily defining the winter period while maximizing the amount of data available for modeling linear effects of air on stream temperature. 418

Modeling the synchronized period of the year also provides additional information about 419 the spring and fall breakpoints and the duration between them. Despite considerable 420 random annual variation, we found that air-water relationships were getting synchronized 421 422 earlier in the spring and remaining synchronized later in the fall. This resulted in a 2.9 days 423 per year expansion in the synchronized period of the year or 44 days over the 15-year study period. This has implications for the growing season (e.g. algal growth, primary 424 425 productivity, nutrient cycling), which affects invertebrate (Ward & Stanford, 1982) and vertebrate growth and development (Neuheimer & Taggart, 2007; Venturelli et al., 2010). 426 Growing seasons worldwide have been expanding about 10-20 days over the last few 427 decades (Linderholm, 2006), which is slower than the expansion in the synchronized 428 429 period we observed. The relationship between plant-based growing season estimates and the width of the synchronized period is currently unknown, but application of our model 430 widely across space could establish this measure as an additional fundamental metric of 431 climate regime change in cold-temperate ecosystems. 432

Hysteresis is another challenge when modeling stream temperature (Webb & Nobilis, 433 434 1997). We allowed for the potential differences in seasonal warming and cooling with a cubic effect of day of the year on water temperature (Figs. 2 & A4). This can be understood 435 as the average expected water temperature on any day of the year during the synchronized 436 period. Then the effects of air temperature, flow, and site can be thought of as moving the 437 water temperature away from this mean expectation. We also allow this cubic effect to vary 438 randomly by year. This has two major benefits. First, it allows the idiosyncratic seasonal 439 440 temperature patterns to vary annually (Fig. A4). Otherwise it would be nearly impossible to have a parametric model describing the effects of a warm, wet spring followed by a cold 441 summer or three moderately cool weeks followed by one extremely hot week in the 442 443 autumn. The second benefit is, by having a random year effect, the pattern of hysteresis is variable and can be well-described when sufficient data are available, while in years with 444 little data the predictions move towards the mean across years. This borrowing effect 445 allows for good predictions even in years with minimal data. An alternative to the 446 parametric cubic function is a non-parametric smoothed function, but it can be challenging 447 to estimate hierarchical effects for smoothed functions. Li et al. (2014) present a stream 448

temperature model with time-varying smoothed functions which allows parameter
estimates to vary over time. The time varying coefficients can account for variation in the
air-water temperature relationship that is not included in the model. RMSE estimates (~1
°C) from the time-varying model are low and similar to the estimate from our model (0.59
°C).

454 Using the cubic function also provides information on the smoothed annual peak temperature and the date of the peak water temperature. We estimated that the peak 455 temperature increased at a rate of 0.063 °C per year, or 0.94 °C over the course of 15 years. 456 The stream temperature warming rate is within the range of rates identified in rivers and 457 streams across the US (0.007 - 0.077 °C per year, Kaushal et al., 2010), but three-fold faster 458 than the rate identified using simple linear models in the Chesapeake Bay watershed 459 (0.028 °C per year, Rice & Jastram, 2014). In contrast to the peak temperature, the day of 460 the year that the peak temperature was reached did not charge during the study. This 461 462 decoupling between the value of the peak and the day of the peak suggests that increased peak temperatures are not a result of a change in the timing of maximum temperatures, but 463 464 rather are driven primarily by increased air temperatures.

Changes in water temperature at a given location do not instantaneously follow changes in 465 air temperature. This is due to the movement of water, heat transfer time, and exchange 466 with thermally buffered below ground heat sources (and sinks)(Caissie, 2006). We 467 accounted for this by including one- and two-day lagged air temperature effects. In this 468 469 way, today's water temperature is influenced by a combination of the air temperature today, yesterday, and the day before yesterday. We found the strongest effect of today's air 470 temperature but significant effects of air temperature both the previous two days (Table 471 A1), suggesting that air temperature effects are operating on the time scale of several days 472 473 in our small stream system.

474 Our hierarchical approach to modeling handles years and sites with varying amounts of

475 incomplete data. A hierarchical model can accommodate missing data for one year and site

476 by 'borrowing' information from other years and sites (Bolker et al., 2009). If there is

477 enough local (site and year) information, the influence of other sites and years on

parameter estimates will be minimal. If data are missing, however, estimates with missing 478 data will tend (shrink) towards the hierarchical mean (Gelman & Hill, 2007). We evaluated 479 how missing data influenced prediction bias across years and sites. When data for a single 480 481 year and all sites were left out, bias (increase in RMSE) was higher (+0.57) than when just the West Brook data were left out (+0.48), demonstrating how data from nearby streams 482 can inform estimates. It will be important in the future to identify the strength of the spatial 483 484 decay function to understand how close sites (on the network) should be to allow effective information sharing. 485

We also evaluated how missing data within years influenced predictions. First, we added 486 487 data from the middle of the year in both directions and found that a 10% increase in data resulted in approximately a 10% improvement in RMSE. Clearly, more data during the 488 synchronized period will provide better predictions, but predictions can still be reasonable 489 490 with limited data during the year. This may be especially true when data from more nearby 491 streams are available, as stream temperature monitoring becomes increasingly common. 492 Second, we evaluated how data availability during the year affected predictions by 493 retaining 30 days of data and shifting the window of availability across the year. When data 494 were available from the other three streams, WB predictions with missing data were insensitive to the timing of available data (consistent 0.38 increase in RMSE). However, 495 when data from the other three streams were not available, predictions were poorer when 496 497 data were only available early in the year compared to late in the year. When data are available from nearby streams, the local data can help define the annual cubic pattern in 498 499 the model, but when they are not available the higher variability in daily stream 500 temperature in the spring compared to the autumn likely results in some years with cubic patterns that are a poor fit to autumn stream temperatures. 501

We used a simple autoregressive term to model the temporal autocorrelation in the residuals. This is critical in a regression-based daily temperature model because the error at time step *i* is likely to be correlated with the error at time *i*+1 due to some small temporal variation not accounted for by the regression parameters. Any autocorrelation or patterning in the residuals violates the assumptions of a linear regression model. This is a classic problem in time series analysis (Shumway & Stoffer, 2006). In our model, the AR1

508term adequately corrected for temporal autocorrelation such that the resulting residuals

- 509 displayed homogeneity and were normally distributed. No additional lagging or moving
- 510 average was needed in this case, but it would be easy to add these additional ARIMA
- 511 parameters to the model if necessary. The estimate of 0.79±0.05 (mean±s.d., Table A1) for
- 512 the autoregressive term indicates strong effect of the previous day's residual on stream
- 513 temperature.

514 Air temperature can be used as the primary variable predicting water temperature in small streams. However, additional factors can influence water temperature directly or affect the 515 air-water temperature relationship (Caissie, 2006; Sun et al., 2014). We found that the 516 517 effect of air temperature was reduced as stream flow increased (significant negative coefficient; Appendix A1). This corresponds to our expectations because a larger volume of 518 water will require more energy to heat and at high flows the streams are generally deeper 519 resulting in a lower relative surface area in contact with the air. Additionally, higher flow is 520 521 often a result of surface and ground water inputs originating over the previous days and weeks and therefore influenced by heat transfer over that time and less on that day's 522 523 current air temperature. Here, flow was our only external variable. Our model can easily 524 accommodate additional factors such as forest cover, agriculture, impervious surfaces, impoundments, and ground water when these data are available and vary over the streams 525 of interest. The model could also easily be extended to model daily minimum (Hughes, 526 Subba Rao & Subba Rao, 2007) or maximum (Caissie, El-jabi & Satish, 2001; Li et al., 2014) 527 stream temperature in addition to the daily mean modeled here. 528

Local variation of environmental drivers at very small spatial scales can have a strong 529 influence on stream temperatures. For example, ground water input can moderate air 530 531 temperature effects in the summer and winter (Poole & Berman, 2001; Kanno, Vokoun & Letcher, 2013; Westhoff & Paukert, 2014; Snyder, Hitt & Young, 2015). We did not model 532 groundwater effects because we lack information on the spatial distribution of 533 groundwater inputs in our small system, but we did observe marked differences in water 534 temperature across streams (Fig. 4). Water temperatures in the WB and OS were 535 536 considerably warmer than temperatures in OL and IL. Stream-specific intercepts reflect the raw stream temperature data, with a range of 1 °C across streams ('mu.year' parameters in 537

Table A1). The most likely explanation for the temperature differences is the presence of 538 upstream impoundments (Webb & Walling, 1996; Dripps & Granger, 2013); the warmer 539 streams have either an upstream reservoir with a surface release (WB) or a beaver 540 541 impoundment (OS). The cooler streams do not have any impounded water. Even small 542 temperature differences among streams can have important consequences for production and phenology of stream biota (Quinn et al., 1994; Miller et al., 2011; Wheeler et al., 2014; 543 544 Letcher et al., 2015), reinforcing the value of statistically robust stream temperature models. 545

By accounting for many of the issues that make stream temperature estimation difficult, 546 547 our stream temperature model provides robust estimates with low error. Most current stream temperature models do not address all of these issues and generally report higher 548 error rates, especially models of daily stream temperature. One reason error rates of our 549 model are low is that we limit analysis to the synchronized period, but this has the added 550 551 benefit of generating data for the beginning and end of the synchronized period which can be very useful for evaluating shifting stream phenology. Our model can also accommodate 552 553 missing data which, unfortunately, is common in streams as temperature logger availability 554 limits data to incomplete spatial coverage and often incomplete temporal coverage within a year. The structure of our model is flexible enough that a data time series even as short as 555 10 days could contribute important information in an analysis of multiple (100's) sites. 556

559	References
00)	nerer enrees

560	Beauchene M, Becker M, Bellucci CJ, Hagstrom N, Kanno Y. 2014. Summer Thermal
561	Thresholds of Fish Community Transitions in Connecticut Streams. North American
562	Journal of Fisheries Management 34:119–131.
563	Benyahya L, Caissie D, St-Hilaire A, Ouarda TBM., Bobée B. 2007. A Review of Statistical
564	Water Temperature Models. Canadian Water Resources Journal 32:179–192.
565	Benyahya L, St-Hilaire A, Ouarda TBMJ, Bobée B, Dumas J. 2008. Comparison of non-
566	parametric and parametric water temperature models on the Nivelle River, France.
567	Hydrological Sciences Journal 53:640–655.
568	Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS. 2009.
569	Generalized linear mixed models: a practical guide for ecology and evolution. Trends in
570	ecology & evolution 24:127–35.
571	Brooks S, Gelman A. 1998. General methods for monitoring convergence of iterative
572	simulations. Journal of computational and graphical 7:434–455.
573	Brown GW. 1969. Predicting temperatures of small streams. Water Resources Research
574	5:68–75.
575	Caissie D. 2006. The thermal regime of rivers: a review. <i>Freshwater Biology</i> 51:1389–1406.
576	Caissie D, El-jabi N, Satish MG. 2001. Modelling of maximum daily water temperatures in a
577	small stream. Journal of Hydrology 251:14–28.
578	Caissie D, El-Jabi N, St-Hilaire A. 1998. Stochastic modelling of water temperatures in a
579	small stream using air to water relations. Canadian Journal of Civil Engineering
580	25:250-260.

581	Crisp DT, Howson G. 1982. Effect of air temperature upon mean water temperature in
582	streams in the north Pennines and English Lake District. Freshwater Biology 12:359–
583	367.
584	DeWeber JT, Wagner T. 2014. A regional neural network ensemble for predicting mean
585	daily river water temperature. Journal of Hydrology 517:187–200.
586	Dripps W, Granger SR. 2013. The impact of artificially impounded, residential headwater
587	lakes on downstream water temperature. Environmental Earth Sciences 68:2399–
588	2407.
589	Eby L a., Helmy O, Holsinger LM, Young MK. 2014. Evidence of Climate-Induced Range
590	Contractions in Bull Trout Salvelinus confluentus in a Rocky Mountain Watershed,
591	U.S.A. <i>PLoS ONE</i> 9:e98812.
592	Elliott J, Elliott J a. 2010. Temperature requirements of Atlantic salmon Salmo salar, brown
593	trout Salmo trutta and Arctic charr Salvelinus alpinus: predicting the effects of climate
594	change. <i>Journal of Fish Biology</i> 44:no–no.
595	Erickson TR, Stefan HG. 2000. Linear Air/Water Temperature Correlations for Streams
596	during Open Water Periods. Journal of Hydrologic Engineering 5:317–321.
597	Fry F. 1971. The effect of environmental factors on the physiology of fish. In: Fish
598	physiology: environmental relations and behaviour. New York: Academic Press, 1–98.
599	Gelman A, Hill J. 2007. Data analysis using regression and multilevel/hierarchical models.
600	Cambridge University Press.
601	Hague MJ, Patterson D a. 2014. Evaluation of Statistical River Temperature Forecast Models
602	for Fisheries Management. North American Journal of Fisheries Management 34:132-
603	146.

604	Hawkins CP, Hogue JN, Decker LM, Feminella JW. 1997. Channel morphology , water
605	temperature , and assemblage structure of stream insects. Journal of the North
606	American Benthological Society1 16:728–749.
607	Hayhoe K, Wake CP, Huntington TG, Luo L, Schwartz MD, Sheffield J, Wood E, Anderson B,
608	Bradbury J, DeGaetano A, Troy TJ, Wolfe D. 2007. Past and future changes in climate
609	and hydrological indicators in the US Northeast. <i>Climate Dynamics</i> 28:381–407.
610	Hilderbrand RH, Kashiwagi MT, Prochaska AP. 2014. Regional and local scale modeling of
611	stream temperatures and spatio-temporal variation in thermal sensitivities.
612	Environmental Management 54:14–22.
613	Hill R a., Hawkins CP, Carlisle DM. 2013. Predicting thermal reference conditions for USA
614	streams and rivers. <i>Freshwater Science</i> 32:39–55.
615	Hughes GL, Subba Rao S, Subba Rao T. 2007. Statistical analysis and time-series models for
616	minimum/maximum temperatures in the Antarctic Peninsula. Proceedings of the Royal
617	Society A: Mathematical, Physical and Engineering Sciences 463:241–259.
618	Huntington TG, Richardson AD, McGuire KJ, Hayhoe K. 2009. Climate and hydrological
619	changes in the northeastern United States: recent trends and implications for forested
620	and aquatic ecosystemsThis article is one of a selection of papers from NE Forests
621	2100: A Synthesis of Climate Change Impacts on Forests of th. Canadian Journal of
622	Forest Research 39:199–212.
623	Isaak DJ, Peterson EE, Ver Hoef JM, Wenger SJ, Falke J a., Torgersen CE, Sowder C, Steel EA,
624	Fortin M-J, Jordan CE, Ruesch AS, Som N, Monestiez P. 2014. Applications of spatial
625	statistical network models to stream data. Wiley Interdisciplinary Reviews: Water:n/a-
626	n/a.

- 627 Isaak DJ, Hubert W a. 2001. A hypothesis about factors that affect maximum summer
- stream temperatures across montane landscapes. *Journal of the American Water Resources Association* 37:351–366.

NOT PEER-REVIEWED

Peer Preprints

630	Isaak DJ, Rieman BE. 2012. Stream isotherm shifts from climate change and implications for
631	distributions of ectothermic organisms. <i>Global Change Biology</i> :n/a-n/a.
632	Kanno Y, Vokoun JC, Letcher B. 2013. Paired stream-air temperature measurements reveal
633	fine-scale thermal heterogeneity within headwater brook trout streams networks.
634	River Research and Applications 10.1002/rr.
635	Kaushal SS, Likens GE, Jaworski N a, Pace ML, Sides AM, Seekell D, Belt KT, Secor DH,
636	Wingate RL. 2010. Rising stream and river temperatures in the United States. Frontiers
637	in Ecology and the Environment 8:461–466.
638	Kim K, Chapra S. 1997. Temperature model for highly transient shallow streams. Journal of
639	Hydraulic Engineering 123:30–40.
640	Letcher BH, Schueller P, Bassar RD, Nislow KH, Coombs JA, Sakrejda K, Morrissey M,
641	Sigourney DB, Whiteley AR, O'Donnell MJ, Dubreuil TL. 2015. Robust estimates of
642	environmental effects on population vital rates: an integrated capture-recapture
643	model of seasonal brook trout growth, survival and movement in a stream network.
644	Journal of Animal Ecology 84:337–352.
645	Li H, Deng X, Kim D-Y, Smith EP. 2014. Modeling maximum daily temperature using a
646	varying coefficient regression model. <i>Water Resources Research</i> 50:3073–3087.
647	Linderholm HW. 2006. Growing season changes in the last century. Agricultural and Forest
648	Meteorology 137:1–14.
649	Lisi PJ, Schindler DE, Cline TJ, Scheuerell MD, Walsh PB. 2015. Watershed geomorphology
650	and snowmelt control stream thermal sensitivity to air temperature. Geophysical
651	Research Letters 42:3380–3388.
652	Miller MR, Brunelli JP, Wheeler P a., Liu S, Rexroad CE, Palti Y, Doe CQ, Thorgaard GH. 2011.
653	A conserved haplotype controls parallel adaptation in geographically distant salmonid

654 populations. *Molecular Ecology*:no–no.

655	Mohseni O, Stefan HG, Erickson TR. 1998. A nonlinear regression model for weekly stream
656	temperatures. Water Resources Research 34:2685–2692.
657	Morrill JC, Bales RC, Conklin MH. 2005. Estimating stream temperature from air
658	temperature: Implications for future water quality. <i>Journal of Environmental</i>
659	Engineering-Asce 131:139–146.
660	Neuheimer AB, Taggart CT. 2007. The growing degree-day and fish size-at-age: the
661	overlooked metric. Canadian Journal of Fisheries and Aquatic Sciences 64:375–385.
662	Nielsen JP. 1999. Record extension and streamflow statistics for the Pleasant River, Maine.
663	US Department of the Interior, US Geological Survey Information Services. Available at:
664	http://me.water.usgs.gov/reports/finalreport.pdf.
665	Peterson EE, Ver Hoef JM. 2010. A mixed-model moving-average approach to geostatistical
666	modeling in stream networks. <i>Ecology</i> 91:644–51.
667	Pilgrim JM, Fang X, Stefan HG. 1998. Stream Temperature Correlations With Air
668	Temperatures in Minnesota: Implications for Climate Warming. Journal of the
669	American Water Resources Association 34:1109–1121.
670	Plummer M, Nest N, Cowles K, Vines K. 2006. CODA: convergence diagnosis and output
671	analysis for MCMC. <i>R News</i> 6:7–11.
672	Plummer M. 2014. rjags: Bayesian graphical models using MCMC.
673	Poole GC, Berman CH. 2001. An ecological perspective on in-stream temperature: Natural
674	heat dynamics and mechanisms of human-caused thermal degradation. Environmental
675	Management 27:787–802.
676	Quinn J, Steele G, Hickey C, Vickers M. 1994. Upper thermal tolerances of twelve
677	
	NewZealand stream invertrbrate species. New Zealand Journal of Marine and

- Rice KC, Jastram JD. 2014. Rising air and stream-water temperatures in Chesapeake Bay
 region, USA. *Climatic Change*.
- Rushworth a. M, Peterson EE, Ver Hoef JM, Bowman a. W. 2015. Validation and comparison
 of geostatistical and spline models for spatial stream networks. *Environmetrics*:n/a–
 n/a.
- Shumway RH, Stoffer DS. 2006. *Time Series Analysis and Its Applications: With R Examples*.
 Springer Science+Business Media.
- Sinokrot B a., Stefan HG. 1993. Stream temperature dynamics: measurements and
 modeling. *Water Resources Research* 29:2299–2312.
- Smith K, Lavis M. 1975. Environmental influences on the temperature of a small upland
 stream. *Oikos* 26:228–236 VN readcube.com.
- Snyder C, Hitt N, Young J. 2015. Accounting for groundwater in stream fish thermal habitat
 responses to climate change. *Ecological Applications* 25:1397–1419.
- Stefan HG, Preud'homme EB. 1993. Stream temperature estimation from air temperature.
 Water Resources Bulletin 29:27–45.
- Sun N, Yearsley J, Voisin N, Lettenmaier DP. 2014. A spatially distributed model for the
 assessment of land use impacts on stream temperature in small urban watersheds.
 Hydrological Processes 2345:n/a–n/a.
- Taylor RG, Scanlon B, Doll P, Rodell M, van Beek R, Wada Y, Longuevergne L, Leblanc M,
- 698 Famiglietti JS, Edmunds M, Konikow L, Green TR, Chen J, Taniguchi M, Bierkens MFP,
- 699 MacDonald A, Fan Y, Maxwell RM, Yechieli Y, Gurdak JJ, Allen DM, Shamsudduha M,
- 700 Hiscock K, Yeh PJ-F, Holman I, Treidel H. 2013. Ground water and climate change.
- 701 *Nature Clim. Change* 3:322–329.
- 702 Venturelli P a., Lester NP, Marshall TR, Shuter BJ. 2010. Consistent patterns of maturity and
- 703 density-dependent growth among populations of walleye (Sander vitreus): application

- of the growing degree-day metric. *Canadian Journal of Fisheries and Aquatic Sciences*67:1057–1067.
- Wagner T, Hayes DB, Bremigan MT. 2006. Accounting for Multilevel Data Structures in
 Fisheries Data using Mixed Models. *Fisheries* 31:180–187.
- Ward J V, Stanford JA. 1982. Thermal Responses in the Evolutionary Ecology of Aquatic
 Insects. *Annual Review of Entomology* 27:97–117.
- Webb B. 1996. Trends in stream and river temperature. *Hydrological Processes* 10:205–
 226.

Webb BW, Hannah DM, Moore RD, Brown LE, Nobilis F. 2008. Recent advances in stream
and river temperature research. *Hydrological Processes* 22:902–918.

Webb BW, Clack PD, Walling DE. 2003. Water-air temperature relationships in a Devon
river system and the role of flow. *Hydrological Processes* 17:3069–3084.

716 Webb BW, Nobilis F. 1997. A long term perspective on the nature of the air-water

temperature relationship: A case study. *Hydrol. Proc.* 11:137–147.

718 Webb JA, Stewardson MJ, Koster WM. 2010. Detecting ecological responses to flow

variation using Bayesian hierarchical models. *Freshwater Biology* 55:108–126.

720 Webb B, Walling D. 1996. Long-term variability in the thermal impact of river

impoundment and regulation. *Applied Geography* 16:211–223.

722 Wenger SJ, Isaak DJ, Luce CH, Neville HM, Fausch KD, Dunham JB, Dauwalter DC, Young MK,

723 Elsner MM, Rieman BE, Hamlet a. F, Williams JE. 2011. Flow regime, temperature, and

- biotic interactions drive differential declines of trout species under climate change.
- 725 Proceedings of the National Academy of Sciences:1–6.
- Westhoff JT, Paukert CP. 2014. Climate change simulations predict altered biotic response
 in a thermally heterogeneous stream system. *PloS one* 9:e111438.

728 Wheeler CA, Bettaso JB, Ashton DT, Welsh HH. 2014. Effects of water tempera

- 729 breeding phenology, growth, and metamorphosis of foothill yellow-legged frogs (Rana
- boylii): a case study of the regulated mainstem and unregulated tributaries of
- 731 California's trinity river. *River Research and Applications* 24.
- 732 Xu CL, Letcher BH, Nislow KH. 2010. Context-specific influence of water temperature on
- brook trout growth rates in the field. *Freshwater Ecology* 55:2342–2369.
- Younus M, Hondzo M, Engel B. 2000. Stream temperature dynamics in upland agricultural
 watersheds. *Journal of Environmental Engineering*:618–628.

737 Tables

- Table 1. Number of days with stream temperature data for each combination of year andsite.
- 740

				West
	Open Large	Open Small	Isolated	Brook
1999	0	0	0	233
2000	0	0	0	256
2001	0	0	0	230
2002	0	81	0	237
2003	179	183	180	191
2004	214	215	214	222
2005	203	102	200	235
2006	0	83	214	210
2007	192	204	192	0
2008	198	199	197	121
2009	243	251	247	0
2010	245	265	246	273
2011	205	248	205	233
2012	234	237	235	210
2013	218	0	212	226

741

743 Table 2. Root mean square error (RMSE, °C) for various scenarios described in the text. The

scenarios involved a training dataset and a test dataset (data left out).

745

Scenario	RMSE train	Test streams	RMSE test	RMSE
				difference
All data	0.59 ± 0.09	_	_	_
30% data	0.69 ± 0.003	All	0.86 ± 0.010	0.17
randomly				
left out				
For each	0.59 ± 0.09	West Brook	1.07 ± 0.26	0.48
year, West				
Brook left				
out				
		4.11		0 FF
For each	0.59 ± 0.09	All	1.16 ± 0.35	0.57
year, all				
streams left				
out				

746

748

- Fig. 1. Map of the study area. Dots indicate locations of temperature loggers and shading
- represents elevation (range approximately 100 m -250 m).









Fig. 3. Examples of raw air (red) and water (black) temperatures from the WB (above) and the temperature index (below)
used to calculate the temperature breakpoints (vertical lines). Horizontal lines in the lower panels are the 99% confidence
intervals of the temperature index for day of year 125 to 275. Vertical axis on the lower panels are truncated to -20 to 20.

759





Fig. 4. Cumulative residuals from the spline in Fig. 2 for each site and year combination. Curves on or near the horizontal line
 indicate 'typical' years whereas curves above the line indicate warm years and below the line indicate cool years.





Fig. 5. Parameter estimates from the stream temperature model. 'B[x]' stands for the β_x in Eq. 5, the 'ar1[x]' are the δ_x from Eq. 3, and the 'mu.year[x]' and the μY_x from Eq. 7, where x=1=WB, x=2=OL, x=3=OS, and x=4=I.



Fig. 6. Fall and spring breakpoints across years for the four streams.



Fig. 7. Predicted maximum temperature for each year (y axis value of dot in Fig. S4) and predicted day of the maximum
temperature (x axis value of dot in Fig. S4).



Fig. 8. Root mean square error (RMSE) difference from the base case (all data included) for the WB for the cross-validation
analyses changing the proportion of days included in estimation. Estimation data included either no data from any of the
streams for each year (triangles, dashed line) or data from the three other streams but no data for the WB for each year
(circles, solid line).

778



Fig. 9. Root mean square error (RMSE) difference from the base case (all data included) for the WB for the cross-validation
analyses changing the starting day of a non-overlapping 30-d moving window. Estimation data included either no data from
any of the streams for each year (triangles) or data from the three other streams but no data for the WB for each year (circles,
solid line).

NOT PEER-REVIEWED