PeerJ Preprints NOT PEER-REVIEWED

A hierarchical model of daily stream temperature for regional predictions

D. J. Hocking[1], K. O'Neil[2], and B. H. Letcher[2]

[1]Department of Biology, Frostburg State University, Frostburg, MD, USA.

[2]US Geological Survey, Leetown Science Center, Conte Anadromous Fish Research Center, Turners Falls, MA, USA.

Corresponding author: Daniel J. Hocking (djhocking@frostburg.edu)

Key Points:

- Flexible approach to modeling daily stream temperature across broad space
- Hierarchical modeling approach allows for inclusion of short observed stream temperature time series to fine-tune estimates to local conditions
- Air temperature effects on stream temperature were moderated by precipitation and drainage area

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

Abstract

Stream temperature is an important exogenous factor influencing populations of stream organisms such as fish, amphibians, and invertebrates. Many states regulate stream protections based on temperature. Therefore, stream temperature models are important, particularly for estimating thermal regimes in unsampled space and time. To help meet this need, we developed a hierarchical model of daily stream temperature and applied it across the eastern United States. Our model accommodates many of the key challenges associated with daily stream temperature models including the lagged response of water temperature to changes in air temperature, incomplete and widely-varying observed time series, spatial and temporal autocorrelation, and the inclusion of predictors other than air temperature. We used 248,517 daily stream temperature records from 1,352 streams to fit our model and 100,909 records were withheld for model validation. Our model had a root mean squared error of 0.61 C for the fitted data and 2.03 C for the validation data, indicating excellent fit and good predictive power for understanding regional temperature patterns. We then used our model to predict daily stream temperatures from 1980 - 2015 for all streams <200 km^2 from Maine to Virginia. From these, we calculated derived stream metrics including mean July temperature, mean summer temperature, and the thermal sensitivity of each stream reach to changes in air temperature. Although generally water temperature follows similar latitudinal and altitudinal patterns as air temperature, there are considerable differences at the reach scale based on landscape and land-use factors.

1 Introduction

Temperature is a critical factor in regulating the physical, chemical, and biological properties of streams. Warming stream temperatures decrease dissolved oxygen, and decrease water density of streams. Biogeochemical processes such as nitrogen and carbon cycling are also temperature dependent and affect primary production, decomposition, and eutrophication. Both physical properties and biogeochemical processes influence the suitability for organisms living in and using the stream habitat beyond just primary producers. Additionally, temperature can have direct effects on the biota, especially poikliotherms such as invertebrates, amphibians, and fish. Given commercial and recreational interests, there is a large body of literature describing the effects of temperature on fish, particularly the negative effects of warming temperatures on cool-water fishes such as salmonids [*Xu et al.*, 2010a, 2010b; *Al-Chokhachy et al.*, 2013a; *Kanno et al.*, 2013]. Finally, stream temperature can even affect electricity, drinking water, and recreation [*van Vliet et al.*, 2011]. Therefore, understanding and predicting stream temperatures is important for many of stakeholders.

Stream temperature models can be used for explanatory purposes (understanding factors and mechanisms affecting temperature) and for prediction. Predictions can be spatial and temporal including forecasting and hindcasting. Predictions across space are especially valuable because there is often a need for information at locations with little or no observed temperature data. For example, many states have regulations related to the management of streams classified as cold, cool, and warm waters [*Beauchene et al.*, 2014], but because of the tremendous number of headwater streams it is impossible classify most streams based on observed data. Therefore, modeled stream temperature is needed to classify most streams for regulatory purposes. Forecasting can provide immediate information such as the expected temperature the next hour, day, or week as well as long-term information about expected temperatures months, years, and decades in the future. Hindcasting can be used to examine temperature variability and trends

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

over time and for model validation. Both forecasting and hindcasting are useful for understanding climate change effects on stream temperature regimes.

Given the importance of temperature in aquatic systems, it is not surprising that there are a variety of models and approaches to understanding and predicting stream temperature. Stream temperature models are generally divided into three categories: deterministic (also called process-based or mechanistic), stochastic, and statistical [*Caissie*, 2006; *Benyahya et al.*, 2007; *Chang and Psaris*, 2013]. Deterministic models are based on heat transfer and are often modeled using energy budgets [*Caissie*, 2006; *Benyahya et al.*, 2007]. These models require large amounts of detailed information on the physical properties of the stream and adjacent landscape as well as hydrology and meteorology. They are useful for detailed assessments and local scenario testing. However, the data requirements preclude the models from being applied over large spatial extents.

Stochastic models attempt to combine pattern (seasonal and spatial trends) with the random deviations to describe and predict environmental data [*Kiraly and Janosi*, 2002; *Sura et al.*, 2006; *Chang and Psaris*, 2013]. Stochastic models of stream temperature generally rely on relationships between air and water temperature adding random noise and an autoregressive correlation, often decomposed by seasonal and annual components. These models are most commonly used to estimate daily temperature fluctuations because of their ability to address autocorrelation and approximate the near-random variability in environmental data [*Caissie et al.*, 2001; *Kiraly and Janosi*, 2002; *Ahmadi-Nedushan et al.*, 2007]. A limitation is that the physical processes driving temperature fluctuations are not elucidated with these models. They are generally used to describe characteristics and patterns in a system and to forecast these patterns in the future [*Kiraly and Janosi*, 2002]. Additionally, stochastic models rely on continuous, often long, time series from a single or a few locations. Inference cannot be made to other locations without assuming that the patterns and random deviations are identical at those locations.

As with stochastic models, statistical models generally rely on correlative relationships between air and water temperatures, but also typically include a variety of other predictor variables such as basin, landscape, and land-use characteristics. Statistical models are often linear with normally distributed error and therefore used at weekly or monthly time steps to avoid problems with temporal autocorrelation at shorter time steps (e.g. daily, hourly, sub-hourly). Parametric, nonlinear regression models have been developed to provide more information regarding mechanisms than traditional statistical models without the detail of physical deterministic models [*Mohseni et al.*, 1998]. Researchers have also developed geospatial regression models that account for spatial autocorrelation within dendritic stream networks [*Isaak et al.*, 2010; *Peterson et al.*, 2010, 2013]. However, due to the complexity of the covariance structure of network geostatistical models, they are best used for modeling single temperature values across space (e.g. summer maximum, July mean, etc.) rather than daily temperatures [*Peterson et al.*, 2010; *Ver Hoef and Peterson*, 2010]. Additionally, statistical machine learning techniques such as artificial neural networks have been used to model stream temperatures when unclear interactions, nonlinearities, and spatial relationships are of particular concern [*Sivri et al.*, 2007, 2009; *DeWeber and Wagner*, 2014].

In contrast with deterministic approaches, statistical models require less detailed site-level data and therefore can be applied over greater spatial extents than process-based models. They also can describe the relationships between additional covariates and stream temperature, which is a limitation of stochastic models. These relationships can be used to understand and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

predict anthropogenic effects on stream temperature such as timber harvest, impervious development, and water control and release [*Webb et al.*, 2008]. Quantifying the relationship between anthropogenic effects, landscape characteristics, meteorological patterns, and stream temperature allows for prediction to new sites and times using statistical models. This is advantageous for forecasting and hindcasting to predict and understand climate change effects on stream temperatures. This is critical because not all streams respond identically to air temperature changes and the idiosyncratic responses may be predicted based interactions of known factors such as flow, precipitation, forest cover, basin topology, impervious surfaces, soil characteristics, geology, and impoundments [*Webb et al.*, 2008].

Letcher et al. [2016] outline six general challenges of statistical stream temperature models including accounting for 1) the non-linear relationship between air and water temperature at high and low air temperatures, 2) different relationships between air and water temperature in the spring and fall (hysteresis), 3) thermal inertia resulting in lagged responses of water temperature to changes in air temperature, 4) incomplete time series data and locations with large differences in the amount of available data, 5) spatial and temporal autocorrelation, and 6) important predictors of stream water temperature other than air temperature. They developed a statistical model that addresses aspects of non-linear relationships, hysteresis, thermal inertia, and spatial and temporal autocorrelation but their analysis was limited to a single small network of one stream with three tributaries and multi-decadal time series [*Letcher et al.*, 2016].

Many fish biologists have focused on weekly, monthly, or summer-only metrics of stream temperature to relate warm conditions to trout distributions [*Al-Chokhachy et al.*, 2013b; *Jones et al.*, 2014]. However, daily temperatures are useful because they can be used to understand and model activity or detection conditional on the thermal conditions at the time of sampling [*Kery*, 2010; *Hocking et al.*, 2013; *Milanovich et al.*, 2015]. They can also be summarized into any derived metrics of interest by scientists or managers [*Jackson et al.,* 2018]. Depending on the species, life-stage, or management options, decision makers and biologists might be interested in different metrics such as degree days since an event (e.g. oviposition, hatching), frequency of thermal excursions, magnitude of excursions, mean summer temperature, or variability in temperature of different time frames, all of which can be derived from daily temperature predictions. Daily temperatures can also relate more closely to state agency regulations such as the frequency of daily temperatures over a threshold when classifying cold, cool, and warm streams for legal protection [e.g. *MassDEP*, 2013]. Without knowing in advance all the potential uses of predicted stream temperatures, a daily model provides the flexibility to derive the values needed for particular decisions.

We describe a novel statistical model of daily stream temperature that incorporates features of stochastic models and extends the *Letcher et al.* [2016] framework to large geographic areas. This model handles time series data of widely varying duration from many sites using a hierarchical mixed model approach to account for autocorrelation at specific locations within watersheds. It incorporates catchment, landscape, and meteorological covariates for explanatory and predictive purposes. It includes an autoregressive function to account for temporal autocorrelation in the time series, a challenge with other statistical models at fine temporal resolution. Additionally, our hierarchical Bayesian approach readily allows for greater accounting of uncertainty. We use the model to predict daily stream temperature across the northeastern United States over a 36-year time record and derive metrics of thermal characteristics for each stream reach.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

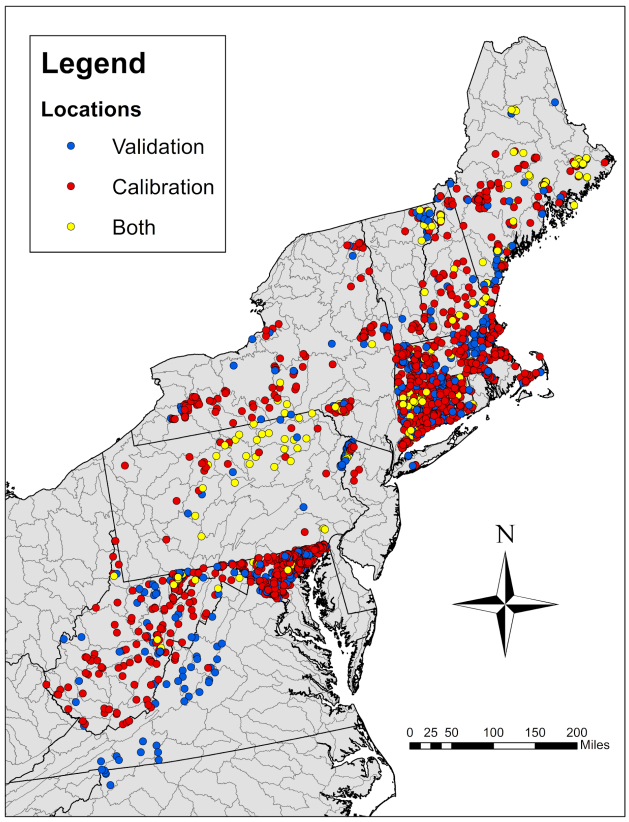
2 Data and Methods

2.1 Water temperature data

We gathered stream temperature data from state and federal agencies, individual academic researchers, and non-governmental organizations (NGOs) from Maine to Virginia (Figure 1). The data were collected using automated temperature loggers. The temporal frequency of recording ranged from every five minutes to once per hour. Organizations used a variety of different loggers but accuracies were generally reported as ±0.4 - 0.5 C with 0.1 C recording resolution. These data were consolidated in a PostgreSQL database linked to a web service at http://db.ecosheds.org. Data collectors could upload data at this website and choose whether to make the data publicly available or not. The raw data were stored in the database and users could flag problem values and time series. Only user-reviewed data were used in the analysis and flagged values were excluded. For our analysis, we performed some additional automated and visual quality assurance and quality control (QAQC) on the sub-daily values, summarized to mean daily temperatures and performed additional QAQC on the daily values. The QAQC was intended to flag and remove values associated with logger malfunctions, out-of-water events (including first and last days when loggers were recording but not yet in streams), and days with incomplete data which would alter the daily mean. Most QAQC was done through visual inspection of each time series. The QAQC webtool used for flagging questionable data can be found at http://db.ecosheds.org/qaqc. We also developed an R [*R Development Core Team*, 2016] package for analyzing stream temperature data from our database, including the QAQC functions which can be found at https://github.com/Conte-Ecology/conteStreamTemperature. The R scripts using these functions for our analysis are available at https://github.com/Conte-Ecology/conteStreamTemperature_northeast.

Figure 1. Map of the locations where temperature data were collected and indicating which locations were used for validation or model calibration (fitting). Many locations were used for both with some years withheld for validation. No individual datum was used for both calibration and validation.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018



Stream reach (stream section between any two confluences) was our finest spatial resolution for the analysis. In the rare case where we had multiple logger locations within the same reach recording at the same time, we used the mean value from the loggers for a given day. In the future, with sufficient within reach data, it would be possible to use our modeling framework to also estimate variability within reach by adding one more level to the hierarchical structure of the model (see 2.4 Statistical model description below).

2.2 Stream network delineation

Temperature logger locations were spatially mapped to the stream reaches of a high-resolution network of hydrologic catchments developed across the Northeastern United States. The National Hydrography Dataset High Resolution Delineation Version 2 (NHDHRDV2) maintains a higher resolution and catchment area consistency than the established NHDPlus Version 2 dataset. The main purpose of the higher resolution was to capture small headwaters that may be critical to ecological assessment. A summary of this dataset with links to detailed documentation can be found in the SHEDS Data project.

2.3 Meteorological and landscape data

The landscape and meteorological data were assembled from various sources. These variables were spatially attributed to the hydrologic catchments for incorporation into the model and include total drainage area, percent riparian forest cover, daily precipitation, daily air temperature, upstream impounded area, percent agriculture, and percent high-intensity

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

development. Air temperatures were gathered from Daymet (Table 1). These values are available on a 1 x 1 km grid across North America on a daily temporal resolution from 1980 – 2016. We used the mean temperature when multiple points fell within a single catchment. Air temperatures have relatively smooth transitions across the landscape and we used a weighted mean based on catchment area for all upstream drainage catchments because all of the upstream air temperatures the stream water is exposed to will influence the local water temperature.

We used air temperature of the day of the water temperature measurement to incorporate immediate responses along with 7-day air temperature to account for lagged changes in water temperature. Daily air temperature and prior 7-day air temperature were important predictors in other stream temperature models [*DeWeber and Wagner*, 2014; *Jackson et al.*, 2018]. This 7-day lag may result from a combination of factors including water flow from upstream, conductive heating from the ground which is slow to transfer heat and averages temperatures over extended periods, and through delayed surface and shallow ground-water inputs [*Benyahya et al.*, 2008; *Webb et al.,* 2010].

Forest cover, impoundments and dam releases, human development, and flow have all been important in other stream temperature models [*DeWeber and Wagner*, 2014; *Jackson et al.*, 2018]. We did not include other predictors previously found to be important in statistical models because of correlation with existing covariates or a lack of variability in the potential predictor across the study area. For example, elevation can be a useful predictor of stream temperature [*DeWeber and Wagner*, 2014; *Detenbeck et al.*, 2016] but it lacks a specific mechanistic relationship or confounds multiple effects and covaries strongly with air temperature across the region. Similarly, human development and impervious surfaces can affect stream temperature. These variables exhibited high negative correlation with forest cover and both variables could not be included in the model. As more stream temperature data become available, it may be possible to separate the effects of forest cover and human development variables. Likewise, agricultural land-use can influence stream temperature or the effect of air temperature on stream temperature [*DeWeber and Wagner*, 2014], but there were insufficient observations over a range of agriculture in our data to include it in the current model. Agriculture can be added to a future version of the model as we expand coverage to the mid-Atlantic region of the U.S. and as more data accumulate. Direct stream shading can also influence local stream conditions [*Johnson*, 2004; *Caissie*, 2006; *Detenbeck et al.*, 2016], but is challenging to quantify over large regions. Ground water input is one of the few major remaining influences of headwater stream temperature that we do not account for because of a lack of currently reliable information on inputs over broad spatial ranges in this region.

Data continuous independent variables were standardized to Z-scores by subtracting the mean and dividing by the standard deviation. This is done primarily for computation efficiency and numeric stability but also to put the variables on similar scales for comparison. Standardization also reduces multicollinearity among independent variables in the model. This resulted in variables with little correlation excepting daily air temperature and 7-day air temperature (see Table 2S.). The correlation among these may slightly confound the effects of each air temperature variable. Further descriptions and data sources for each of these variables are described in Table 1. All of the variables referenced in the table refer to values calculated for the downstream point of each catchment (confluence pour point).

Table 1. Description and original source of variables used in the model.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

Variable	Description	Source
Total Drainage Area	The total contributing drainage area from the entire upstream network	The SHEDS Data project
Riparian Forest Cover	The percentage of the upstream 61 m (200 ft) riparian buffer area that is covered by trees taller than 5 meters	The National LandCover Database (NLCD)
Daily Precipitation	The mean daily precipitation for the individual local catchment	Daymet Daily Surface Weather and Climatological Summaries
Daily Air Temperature	The daily mean air temperature for the individual local catchment as the mean of the minimum and maximum daily temperature from Daymet	Daymet Daily Surface Weather and Climatological Summaries
Upstream Impounded Area	The total area in the contributing drainage basin that is covered by wetlands, lakes, or ponds that intersect the stream network	U.S. Fish & Wildlife Service (FWS) National Wetlands Inventory

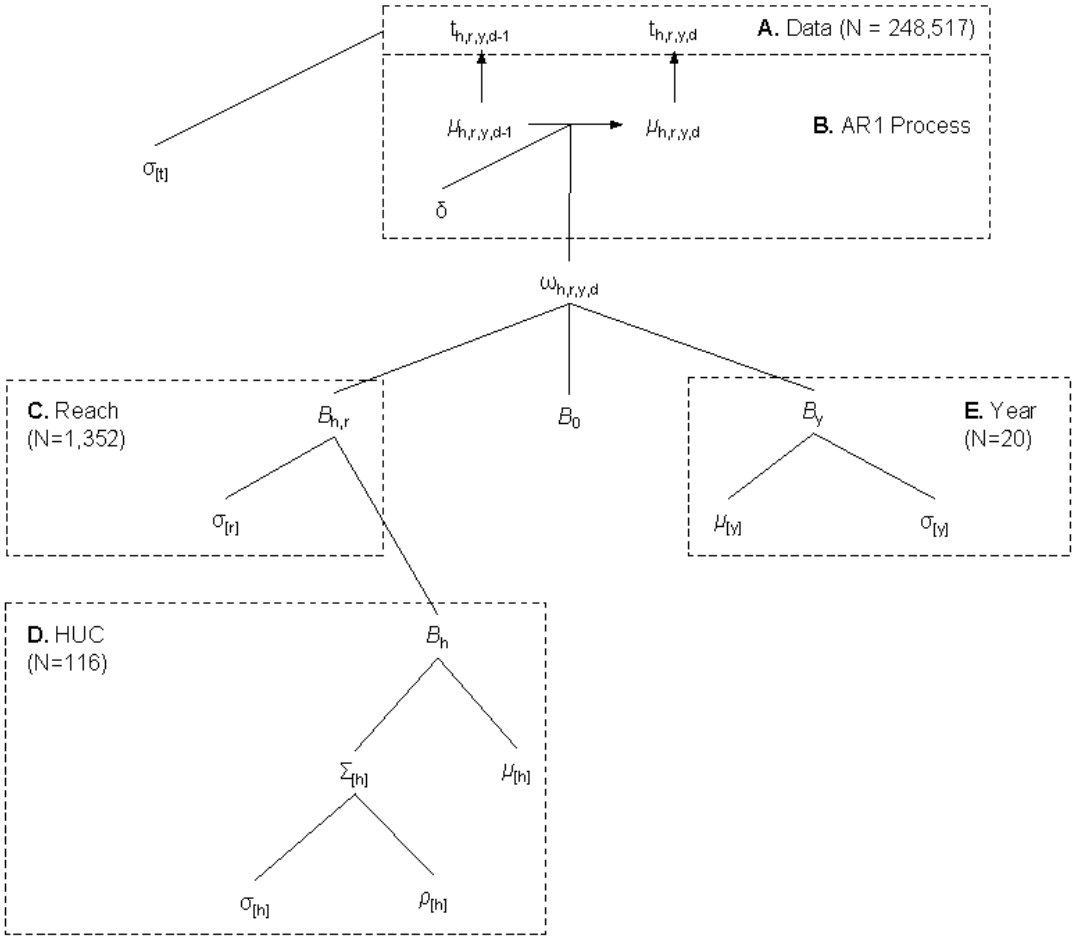
2.4 Statistical model

Statistical models of stream temperature often rely on the close relationship between air temperature and water temperature. However, this relationship breaks down during the winter in temperate zones, particularly as streams freeze, thereby changing their thermal properties. Many researchers and managers are interested in the non-winter effects of temperature. The winter period, when phase change and ice cover alter the air-water relationship, differs in both time (annually) and space. Our temperature model was for non-winter periods to avoid when phase-change and ice cover alter the air-water relationship. To determine the ice-free, non-winter portion of the year, we calculated an index of air-water synchrony [Letcher et al. 2016] for each stream reach across the region to truncated the time series accordingly.

We used a generalized linear mixed model to account for correlation in space (stream reach nested within the 8-digit Hydrologic Unit Code [HUC8; https://water.usgs.gov/GIS/huc.html]). This allowed us to incorporate short time series as well as long time series from different reaches and disjunct time series from the same reaches without risk of pseudoreplication [*Hurlbert*, 1984]. By limiting stream drainage area to <200 km^2 and only modeling the synchronized period of the year, we were able to use a linear model, avoiding the non-linearities that occur at very high temperatures due to evaporative cooling and near 0 C due to phase change [*Mohseni and Stefan*, 1999]. The general model structure is depicted in Figure 2.

Figure 2. Hierarchical structure of the daily stream temperature model. The observed daily temperatures are $t_{h,r,y,d}$ at HUC8 h and reach r in year y on day d. In general, μ represent means, σ represent standard deviations, B represent vectors of coefficients with subscripts representing the level of variation, Σ is the covariance matrix, ρ is the correlation matrix, ω is the expected temperature as a function of the deterministic components prior to inclusion of temporal autocorrelation, and δ is the autocorrelation coefficient. See details in the text for further description of the coefficients.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018



We assumed stream temperature measurements were normally distributed following,

$$t_{h,r,y,d} \sim \mathcal{N}(\mu_{h,r,y,d}, \sigma_{[t]})$$

where $t_{h,r,y,d}$ is the observed mean daily stream water temperature at the reach (r) within the sub-basin identified by the HUC8 (h) for each day (d) in each year (y). The expected mean temperature is $\mu_{h,r,y,d}$ and $\sigma_{[t]}$ is the standard deviation. Subscripts represent the levels at which the value varies. Bracketed subscripts are solely for additional naming purposes, for example to distinguish means and variances from different levels of the hierarchical model.

The expected mean temperature ($\mu_{h,r,y,d}$) was adjusted based on the residual error from the previous day

$$\mu_{h,r,y,d} = \begin{cases} \omega_{h,r,y,d} + \delta(t_{h,r,y,d-1} - \omega_{h,r,y,d-1}) & \text{for } t_{h,r,y,d-1} \text{ is real} \\ \omega_{h,r,y,d} & \text{for } t_{h,r,y,d-1} \text{ is not real} \end{cases}$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

where δ is an autoregressive (AR1) coefficient and $\omega_{h,r,y,d}$ is the mean expected temperature before accounting for temporal autocorrelation in the error structure (Figure 2B). This mean temperature is modeled to follow a linear trend

$$\omega_{h,r,y,d} = X_{[0]}B_{[0]} + X_{h,r}B_{h,r} + X_{h}B_{h} + X_{y}B_{y}$$

where $B_{[0]}$ is the vector of K coefficients and K is the number of fixed effects parameters including the overall intercept. $X_{[0]}$ is the $n \times K$ matrix of predictor values where n is the observation number for a particular reach-HUC-year-day combination. We used 15 fixed effect parameters including interaction terms but not the overall intercept. These were 2-day total precipitation, 30-day cumulative precipitation, drainage area, upstream impounded area, percent riparian forest cover, and various two- and three-way interactions (Table 2). We assumed the following distributions and vague priors for the fixed effects coefficients

$$B_{[0]} = \begin{pmatrix} \beta_{[1]} \\ \vdots \\ \beta_{[K]} \end{pmatrix} \sim \mathcal{N}(0,100)$$

$B_{h,r}$ is the $R \times L$ matrix of regression coefficients where R is the number of unique reaches and L is the number of regression coefficients that vary randomly by reach within HUC8 (Figure 2C). In this case, we included a random intercept, and random slopes for the air temperature and 7-day air temperature (L = 3; Table 2). We assumed prior distributions of

$$B_{h,r} = \begin{pmatrix} \beta_{h,r,[0]} \\ \beta_{h,r,[1]} \\ \beta_{h,r,[2]} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{[r0]} & 0 & 0 \\ 0 & \sigma^2_{[r1]} & 0 \\ 0 & 0 & \sigma^2_{[r2]} \end{pmatrix} \right)$$

where $B_{h,r}$ is an $R \times L$ matrix, $\beta_{h,r}$ are normally distributed vectors of coefficients with a mean of 0 and length of R, for the intercept ($\beta_{h,r,[0]}$) and random slopes. We assumed an independent uniform prior on each standard deviation [Gelman2006],

$$\sigma_{[r]} \sim uniform(0,100)$$

For the random HUC8 level component, X_h is the matrix of parameters that vary by HUC8. B_h is the $H \times L$ matrix of coefficients where H is the number of HUC8 groups (Figure 2D). We allowed for correlation among the effects of these HUC8 coefficients as described by *Gelman and Hill* [2007]. As such, we assumed priors distributions of

$$B_{h} \sim \mathcal{N}(M_{[h]}, \Sigma_{[h]}), \text{ for h=1,...,H}$$

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

where $M_{[h]}$ is a vector of the means of length L and $\Sigma_{[h]}$ is the $L \times L$ covariance matrix. We assumed the means followed a multivariate normal distribution,

$$M_{[h]} \sim MVN(\mu_{[h(1:L)]}, \sigma_{[h(1:L)]})$$

with a vague normally distributed prior on the means,

$$\begin{pmatrix} \mu_{[h0]} \\ \mu_{[h1]} \\ \mu_{[h2]} \end{pmatrix} \sim \mathcal{N}(0,100)$$

We used a vague inverse-Wishart prior to describe the covariance matrix,

$$\Sigma_{B_h} = \begin{pmatrix} \sigma^2_{[h0]} & \rho_1\sigma_{[h0]}\sigma_{[h1]} & \rho_2\sigma_{[h0]}\sigma_{[h2]} \\ \rho_1\sigma_{[h0]}\sigma_{[h1]} & \sigma^2_{[h1]} & \rho_3\sigma_{[h1]}\sigma_{[h2]} \\ \rho_2\sigma_{[h0]}\sigma_{[h2]} & \rho_3\sigma_{[h1]}\sigma_{[h2]} & \sigma^2_{[h2]} \end{pmatrix} \sim \text{Inv-Wishart}(diag(L), L+1)$$

where $\sigma_{[h1]}$, $\sigma_{[h1]}$ and $\sigma_{[h2]}$ are the standard deviations of the random HUC8 effects and $\rho_{1:3}$ are the correlation coefficients. In addition to random reach and HUC effects, we also allowed for the intercept to vary randomly by year (Figure 2E). We assumed a prior distribution of

$$B_y \sim \mathcal{N}(0, \sigma_{[y]})$$

for the random year effects with the standard deviation following a vague uniform distribution,

$$\sigma_y \sim uniform(0,100)$$

To estimate all the parameters and their uncertainties, we used a Bayesian analysis with a Gibbs sampler implemented in JAGS [http://mcmc-jags.sourceforge.net/] through R [*R Development Core Team*, 2016] using the rjags package [*Plummer*, 2016]. This approach was beneficial for hierarchical model flexibility and tractability for large datasets. We used vague priors for all parameters so all inferences would be based on the data. We ran 13,000 iterations on each of three chains with independent random starting values. We discarded the first 10,000 iterations, then thinned; saving every third iteration for a total of 3,000 iterations across three chains to use for inference. We visually inspected the MCMC iterations for convergence and autocorrelation, and mixing. We further used potential scale reduction factors (PSRF, $\hat{R}$) to assess mixing of the MCMC chains [*Brooks and Gelman*, 1998].

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

2.5 Model validation

To validate our model, we held out 10% stream reaches at random along with 10% of remaining reach-year combinations. Additionally, we held out all 2010 data because it was an especially warm summer across the northeastern U.S. based on the mean summer Daymet air temperatures. This approach was also used by [*DeWeber and Wagner*, 2014] and helps to assess the model's predictive ability under future warming conditions. In total, 26.4% of observations (100,909) and 33.3% of reaches (723) were held out for validation. From these validation data, we were able to generate a variety of *post hoc* validations to examine the predictive ability in reaches without data in a given year, reaches without data in any years, and all reaches in years without any data (as in future scenarios), and HUC8 without data. The most challenging validation scenario was for reaches in HUC8s without any data (Table 3).

2.6 Derived metrics

We used the meteorological data (air temperature, precipitation) from Daymet to predict daily stream temperatures for all stream reaches (<200 km^2) in the region for the synchronized period of the year from 1980-2015. The predictions are conditional on the specific random effects where available and receive the mean effect for reaches, HUC8s, and years when no data was collected. From these daily predictions, we derive a variety of metrics to characterize the stream thermal regime. These include mean (over the 36 years) July temperature, mean summer temperature, mean number of days per year above a thermal threshold (such as 18, 22 C), frequency of years that the mean daily temperature exceeds each of these thresholds, and the maximum 30-day moving means averaged across all years. We also calculated the resistance of water temperature to changes in air temperature during peak air temperature (summer) based on the cumulative difference between the daily temperatures. Finally, we assess the thermal sensitivity for each stream reach as the change in daily stream temperature per 1 C change in daily air temperature. This is essentially the reach-specific air temperature coefficient converted back to the original scale from the standardized scale. These metrics were selected for comparison to other studies [*DeWeber and Wagner*, 2014; *Jackson et al.,* 2018] and based on the interests of various biologists and state regulatory agencies expressed in discussions and workshops.

3 Results

3.1 Model fitting and MCMC convergence

To fit the model, we used 248,517 daily temperature observations from 1,352 stream reaches within 116 HUC8 subbasins over a 21-year period between 1995 and 2015, excluding all records from 2010 for validation. The iterations of the three MCMC chains converged on a single area of high posterior probability while exhibiting minimal autocorrelation, based on visual inspection of the iteration traceplots, partial vs. full density plots, and autocorrelation (ACF) plots. We chose not to thin chains more than necessary for computational memory purposes and processing of derived parameters because thinning in inefficient and can reduce the precision of summary estimates [*Link and Eaton* 2012]. The potential scale reduction factors (PSRF, $\hat{R}$) for all parameters and the multivariate PSRF were < 1.1, further indicating good convergence of the MCMC chains [*Brooks and Gelman*, 1998].

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

3.2 Coefficient estimates and variability

Most variables and their interactions had 95% Credible Intervals (CRI) that did not overlap zero (Table 2), which indicates a 95% probability that the true value of the effect differs from zero. The only parameters with CRI overlapping zero were the interactions of air temperature and forest cover and air temperature and Impounded Area. Drainage area alone was not significant but it was significant in its interactions with all combinations of air temperature and precipitation (Table 2). Air temperature (1-day and 7-day) had the largest effect on daily water temperature. The effect of air temperature was slightly dampened by interactions with precipitation and drainage area (negative 3-way interactions; Table 1). There was also a large autocorrelation coefficient (AR1 = 0.77).

There was much more unexplained random variation among sites (SD = 1.03) than among HUC8 (SD = 0.59), but the effects of air temperature on water temperature were only slightly more variable among sites (SD = 0.29) compared with HUC8 (SD = 0.27). There was very little random variability among years not explained by other parameters (SD = 0.28; Table 2).

From past experience in the northeast, we found that 2-day precipitation multipled by drainage area was a reasonable index of flow response to precipitation, whereas 30-day precipitation multiplied by drainage area was a good index of baseflow in small streams. We multiplied these terms by daily air temperature to serve as an index of how flow can moderate the effects of air temperature on stream temperature as seen in other studies [*Smith and Lavis*, 1975; *Webb et al.*, 2003; *van Vliet et al.*, 2011]. We also used interaction effects of impounded area with air temperature and forest cover with air temperature. This is because a forest might not just cool down a stream compared to other habitat but might moderate the fluctuations with air temperature. Similarly, relatively small impoundments could generally result in warming of streams but also could moderate the effect of air temperature on stream temperature. See the discussion for more regarding impoundment effects.

Table 2. Regression summary table with coefficient estimates including the mean, standard deviation (SD), and 95% credible intervals (LCRI = 2.5%, UCRI = 97.5%).

Fixed effects:

Parameter	Mean	SD	LCRI	UCRI
Intercept	16.69	0.135	16.42	16.950
AirT	1.91	0.022	1.862	1.950
7-day AirT	1.36	0.029	1.302	1.417
2-day Precip	0.06	0.002	0.055	0.063
30-day Precip	0.01	0.006	0.001	0.026
Drainage Area	0.04	0.096	-0.145	0.232
Impounded Area	0.50	0.095	0.318	0.691
Forest Cover	-0.15	0.047	-0.246	-0.059
AirT x 2-day Precip	0.02	0.002	0.020	0.028
AirT x 30-day Precip	-0.01	0.004	-0.022	-0.007
AirT x Drainage	-0.06	0.029	-0.117	-0.006
AirT x Impounded Area	0.02	0.029	-0.035	0.077
AirT x Forest	-0.02	0.015	-0.051	0.009
2-day Precip x Drainage	-0.04	0.002	-0.042	-0.034
30-day Precip x Drainage	-0.06	0.006	-0.071	-0.046

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

PeerJ Preprints NOT PEER-REVIEWED

AirT x 2-day Precip x Drainage	-0.01	0.002	-0.016	-0.008
AirT x 30-day Precip x Drainage	-0.01	0.004	-0.019	-0.004
AR1	0.77	0.002	0.768	0.776

Random effects:

Group	Coef	SD	LCRI	UCRI
Site	Intercept	1.03	1.01	1.05
	AirT	0.29	0.27	0.31
	7-day AirT	0.35	0.33	0.37
HUC8	Intercept	0.59	0.48	0.72
	AirT	0.27	0.24	0.30
	7-day AirT	0.26	0.22	0.31
Year	Intercept	0.28	0.23	0.36

HUC8 coefficient correlations:

	Intercept	AirT	7-day AirT
Intercept			
AirT	0.640		
7-day AirT	0.338	0.234	

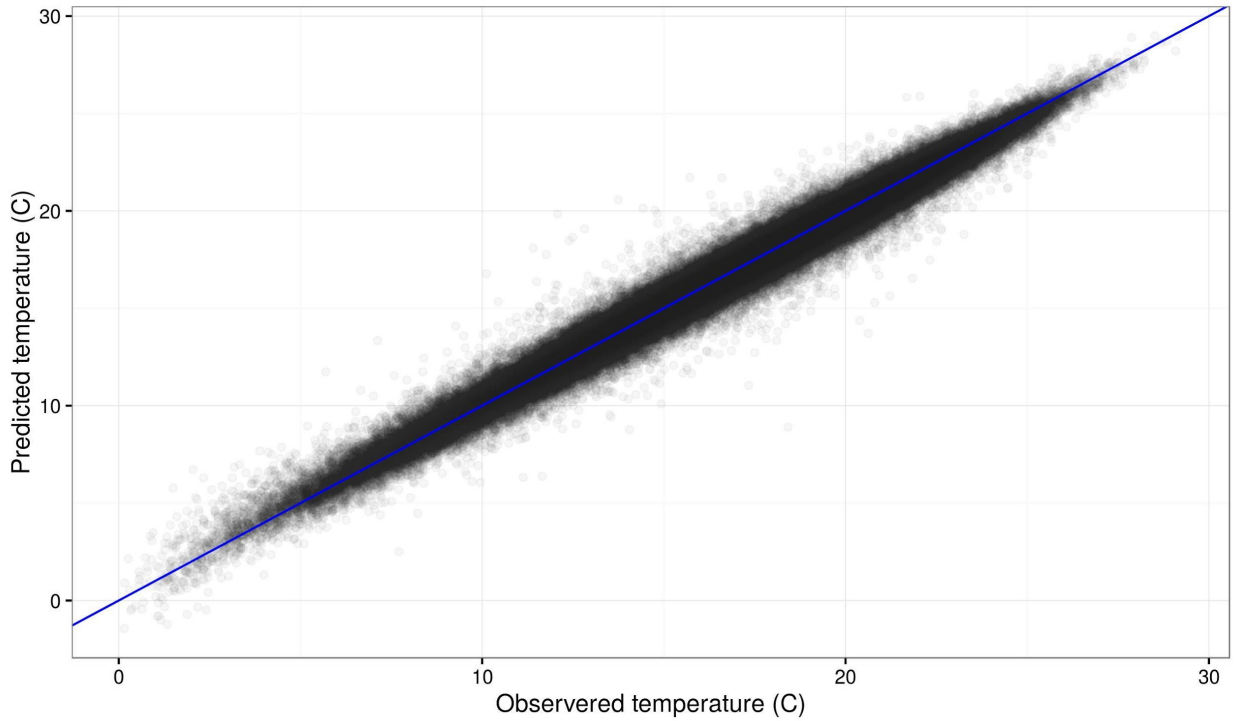
3.3 Evaluation of model fit and prediction

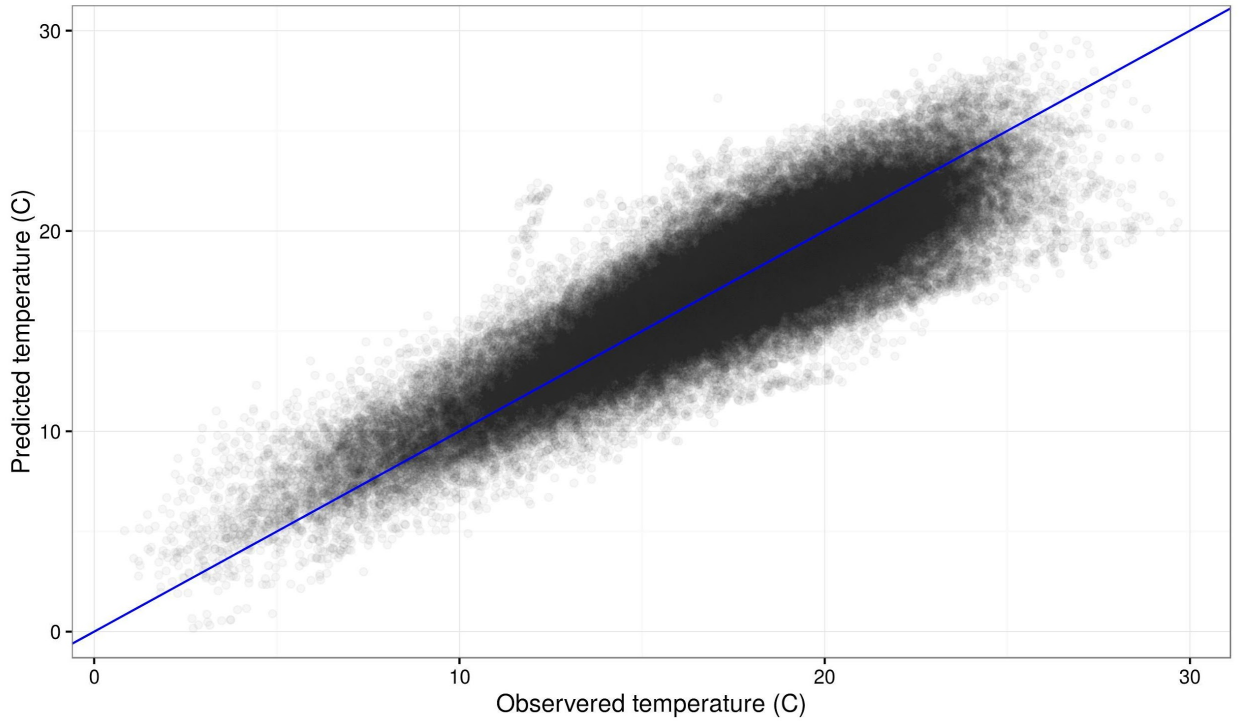
The overall Root Mean Squared Error (RMSE) was 0.61 C and the residuals were normally distributed and unbiased (exhibiting no visual heterogeneity), indicating that the model was a good approximation of the process generating the data. These predicted values are adjusted for residual error, but it is better to use the predictions prior to adjusting with the residual AR1 term to understand how well the model predicts temperatures when the previous day's observed temperature is unknown. The RMSE for the fitted data using unadjusted predictions was 1.08 C. All following predictions and summaries use the unadjusted values to focus on the predictive abilities of the model.

Specifically, to evaluate the spatial and temporal predictive power of our model, we used independent validation data consisting of 100,909 daily temperature observations from 723 stream reaches within 101 HUC8 subbasins over 20 years from 1996 to 2015. The overall unadjusted RMSE for all validation data was 2.03 C. Similar to the fitted data, there was generally no bias in the predictions of the validation data, with the potential exception of slight over-prediction at very low temperatures and possible slight under-prediction at temperatures over 25 C (Figure 3).

Figure 3. Relationship between observed and predicted water temperature. The top plot shows the observed temperature versus the predicted temperature for the fitted data and the bottom plot shows the predictions for the validation data withheld from modeling fitting.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018





Predicting to unsampled reaches in HUC8 with data from other reaches and in years with observed data elsewhere resulted in a RMSE of 1.96. Prediction for reaches in HUC8 with no data was considerably worse (Table 3). To assess predictive accuracy in warm years without data (potential for forecasting under climate change), we calculated the RMSE for all reaches in 2010 (excluded from model fitting) to be 2.13 C.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

Table 3. The root-mean-squared error (RMSE) based on the data used and excluded from different subsets of the validation data. N is the number of daily temperature observations used in each subset of the data. Validation data was completely withheld at random from the data used in model fitting (calibration). Only one set of validation data was used and the additional evaluations represent *post hoc* comparisons to examine the effects of different levels of data on the predictive ability.

Data Used	Reach	HUC8	Year	Reach-Year	RMSE	N
Fitted RMSE	X	X	X	X	0.59	248517
Overall validation RMSE	X	X	X	X	2.03	100909
Missing reach-year but reach, HUC8, and year with data	X	X	X		1.90	18401
Missing reaches but HUC8 and year with data		X	X		1.96	42602
Missing HUC8 but year with data			X		2.52	1081
Missing year but reaches and HUC8 with data	X	X		X	2.06	19090
2010 excluded but all other data available	Mixed	Mixed			2.13	38825
No data for reach, HUC8, or year					1.83	2644

We used the Daymet air temperature and precipitation along with landscape covariates to predict daily stream temperatures in each reach then calculated derived metrics of potential interest to biologists, managers, and policy makers (Table 4).

Table 4. Summary and description of derived metrics for each stream reach summarized for predictions from 1980-2015. The mean number of days over 18 and 22 C were only calculated for predictions in the middle 194 days of the year to avoid problems outside the synchronized period of the year while keeping the length consistent among reaches across the region.

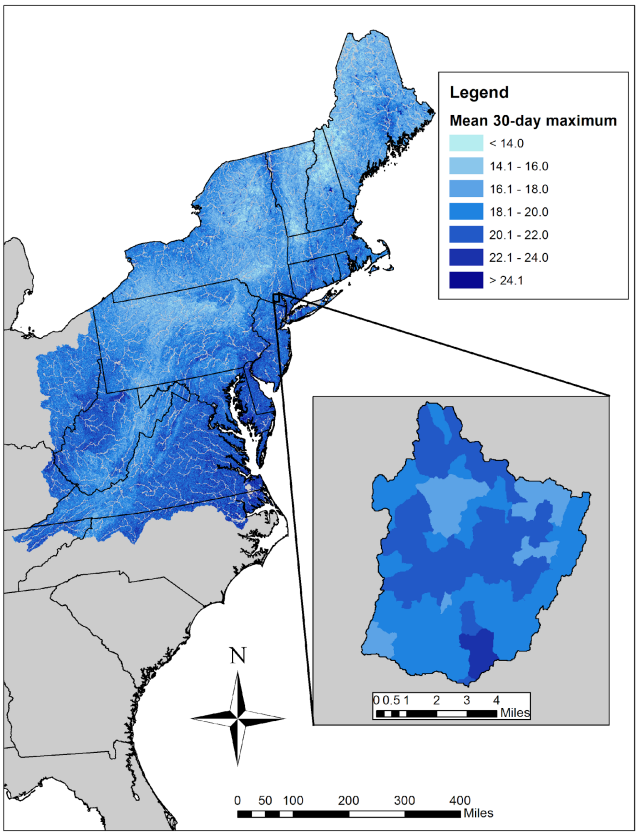
Metric	Mean	Min	Max	Description
Mean maximum temperature	20.57	12.61	34.11	Maximum daily mean water temperature (C) averaged over 36 years (1980 - 2015)
Max maximum temperature	22.30	14.05	35.25	Maximum over years of the maximum daily mean temperature
Mean July temperature	18.25	8.83	32.34	Mean daily July temperature over years
Mean August temperature	17.74	8.52	31.76	Mean daily August temperature over years
Mean summer temperature	17.49	7.92	31.77	Mean daily summer temperature over years
Mean 30-day maximum temperature	18.76	9.68	32.71	Maximum 30-day temperature for each year averaged over years
Mean number of days over 18 C	47.73	0.00	194.00	Mean number of days per year the mean daily temperature exceeds 18 C
Mean number of days over 22 C	5.17	0.00	194.00	Mean number of days per year the mean daily temperature exceeds 22 C
Annual frequency of exceeding 18 C	0.86	0.00	1.00	Frequency of years the mean daily temperature ever exceeds 18 C
Annual frequency of exceeding 22 C	0.28	0.00	1.00	Frequency of years the mean daily temperature ever exceeds 22 C
Mean annual resistance	311.95	69.96	789.80	Mean annual resistance of water temperature to peak (summer) air temperature
Thermal sensitivity	0.61	0.35	0.98	Thermal sensitivity of water temperature to changes in air temperature

We generated maps of mean derived metrics from temperatures predicted over the Daymet record (1980-2015). Land cover was assumed to be static over this period. Future

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

models could incorporate more details regarding land-use change over this time period. When scaled to view the entire region the patterns generally follow air temperature patterns with cooler temperatures at higher elevations and latitudes and warmer temperatures in urban, coastal, and lowland areas. An example of this can be seen on the mean 30-day maximum of the mean daily stream temperature map. However, when zoomed in to view individual catchments on the HUC8 or HUC10 scale, it is clear that there is considerable local variation in water temperatures (Figure 4) based on forest cover, drainage area, and local reach effects (unaccounted for local conditions including ground-surface water interactions), as expected based on the model coefficients and past research [*Kanno et al.*, 2013].

Figure 4. Map of the mean annual maximum 30-day mean stream temperature (mean temperature during the warmest 30-day period each year). The inset shows how much local variation there is that is not clearly visible on the regional map. Gray areas have no predictions, usually because they are in larger streams, outside the bounds of the data used in the model (>200 km^2 drainage area).



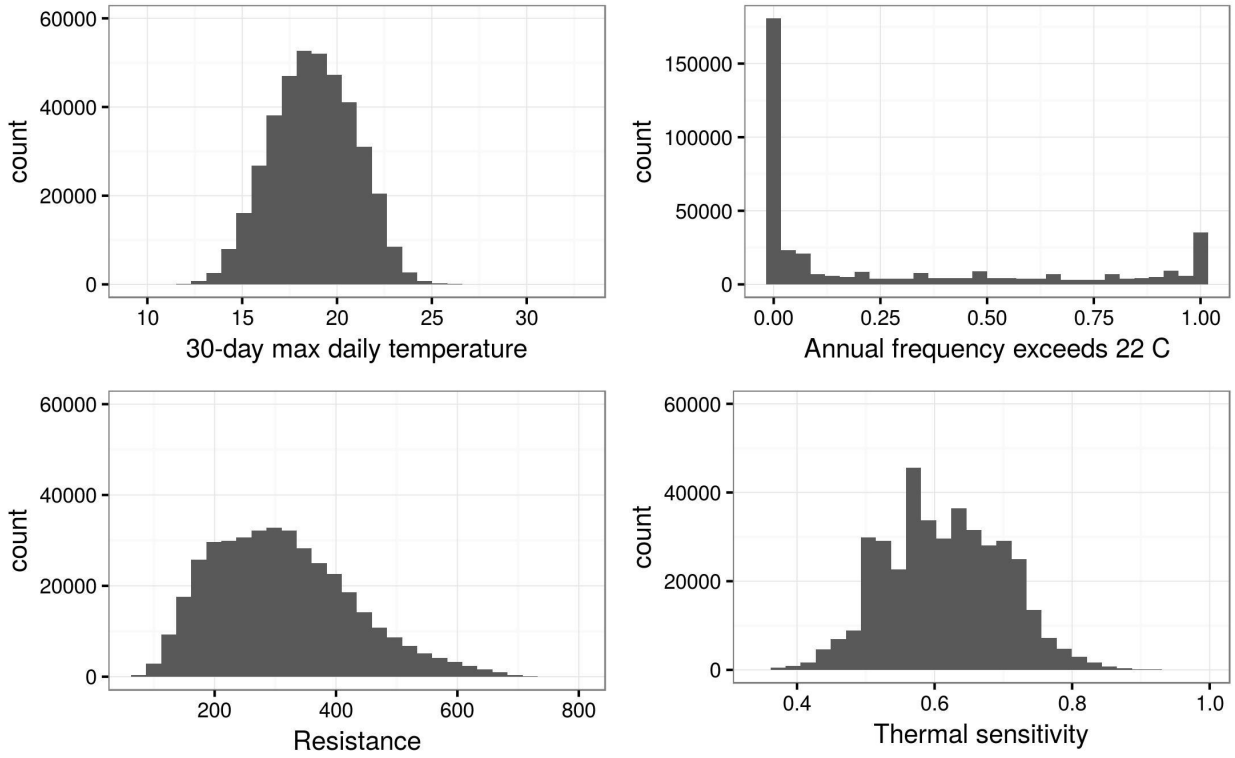
In lieu of presenting many small static maps, many of which would look somewhat similar when zoomed out to the regional scale, we added maps of the derived metrics to our web application which can be found at http://ice.ecosheds.org/stm. Users can zoom to specific areas and view information about individual stream reaches and associated catchments. There is also the ability to filter to locate areas with specific conditions. Our general Interactive Catchment Explorer (ICE) for the northeastern and mid-Atlantic regions of the U.S. will be regularly updated as new data become available and can be found at http://ice.ecosheds.org. It includes information about the landscape conditions and Brook Trout occupancy in addition to stream

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

temperatures. This is part of our web platform for Spatial Hydro-Ecological Decision Systems (SHEDS; http://ecosheds.org) where we present visualizations linking datasets, statistical models, and decision support tools to help improve natural resource management decisions. Interested users can contribute, view, and download (if user-designated as publicly available) data at http://db.ecosheds.org. As noted above, these data will be used to further improve model estimates and predictions, which will be presented in ICE.

Although many of the derived metrics relating to peak temperatures have relatively similar broad-scale spatial patterns, there are some metrics that quantify other aspects of the thermal regime. For example, we calculated the resistance of water temperature to changes in air temperature during peak air temperature (summer) based on the cumulative difference between the daily temperatures. The distribution of resistance values was more right-skewed than the annual mean 30-day maximum temperature (Figure 5).

Figure 5. Histograms of the distribution of values for various derived metrics across all stream reaches. Clockwise from upper left: annual 30-day maximum of mean daily temperature, frequency of years the maximum daily temperature exceeds 22 C, thermal sensitivity of water temperature to changes in air temperature, and resistance to changes in summer air temperature (degree-days: C days^{-1}).



4 Discussion

Our approach handles most of the challenges described by *Letcher et al.* [2016] for modeling stream temperatures including the non-linear relationship between air and water temperature at high and low air temperatures, thermal inertia resulting in lagged responses of water temperature to changes in air temperature, incomplete time series data and locations with large differences in the amount of available data, spatial and temporal autocorrelation, and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

important predictors of stream water temperature other than air temperature. We expanded on their initial work to model daily stream temperatures of broad, regional spatial scales. Our model performed well for explaining landscape and climate effects on stream temperature and predicting temperatures to unsampled space and time. The model does especially well if a reach has been sampled in at least one previous period and if there are data from other reaches within the same HUC8 sub-basin. Our model also performed well when predicting validation data for a warm year, suggesting that it may be reliable for predictions under future climate change.

Rather than explicitly model the non-linear relationship between air and water at low and high temperatures, we took the approach of avoiding the low temperature issues associated with phase change and ice-cover by restricting data to the synchronized portion of the year. Using the synchronized period is advantageous because it is empirically-estimated for each location and year, rather than setting a broad subjective cut off or only using summer temperatures [e.g. *Garreta et al.*, 2009; *Mayer*, 2012; *Detenbeck et al.*, 2016]. This is most effective when there are data for that reach but we used the synchronized period from the HUC8 or nearest HUC8 when reach data were not available.

Our approach also partially accounted for the thermal inertial and lagged response of water temperature to changes in air temperature [*Letcher et al.*, 2016; challenge 3]. In our analysis, 7-day air temperature had nearly as much influence as same-day air temperature (Table 2), indicating the importance of accounting for delayed and averaged temperature in daily stream temperature models. We initially explored the inclusion of 1-day and 2-day lagged air temperature in place of the 7-day mean air temperature but the daily lags tended to be excessively correlated with same-day air temperature for inclusion (Pearson correlation $r > 0.7$).

A considerable advantage of our approach compared with models of weekly or monthly metrics (e.g. mean temperature) is the inclusion of short and incomplete time series [*Letcher et al.*, 2016; challenge 4]. Models that use data consolidated to the weekly, monthly, or seasonal scale require complete, continuous water temperature time series to avoid statistical bias [e.g. *Garreta et al.*, 2009; *Mayer*, 2012; *Detenbeck et al.*, 2016]. If a day or few days of data are missing or unreliable due to sensor malfunction, then entire time series must be excluded from the analysis or risk biased inference. This limits the number of locations that can be monitored and used for a regional analysis. There are now a large number of state and federal agencies, individual research scientists, and nongovernmental organizations measuring stream temperature because new sensors are relatively cheap and easy to use. As of 09 February 2017, there were 58 different organizations in the northeastern United States alone that have contributed stream temperature records to just our database (http://db.ecosheds.org/). However, these groups record data for varying lengths of time at a given location, from four days to more than a year, but even these longer time series often have missing or corrupted data, causing effective breaks in the time series. Our model can incorporate locations with time series of any length. Reaches will little data contribute less information to the model but do provide some local and spatial information. Locations with more data contribute more to the likelihood so there is less shrinkage of parameter estimates towards the mean values [*Gelman and Hill*, 2007]. One potential shortcoming of this approach is the propagation of uncertainty from the daily values to the derived metrics. The effects of this propagation and influence on risk assessment in natural management decision making is an area that warrants future study.

Our model also accounted for spatial and temporal correlation [*Letcher et al.*, 2016; challenge 5]. We allowed the overall intercept and effects of air temperature and 7-day mean air temperature to vary randomly among reaches and HUC8 sub-basins. This allows for

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

measurements and effects within a particular reach over time to be more correlated to each other than would be expected by random chance. Similarly, reaches within the same HUC8 were expected to be more similar to each other on average than to a random stream reach. For example, a reach or reach within a HUC8 that is typically cooler than expected based on the covariates in the model would have a lower random intercept. A reach that is less responsive to daily fluctuations in air temperature would have a lower coefficient for air temperature than the average stream. This later result could potentially be a coarse indicator of relative ground water inputs. We explicitly accounted for local air temperature, size of the stream (drainage area), forest cover, upstream impounded area, and indices of flow (2-day and 30-day precipitation x drainage area). Ground water temperatures tend to be more stable than surface temperatures, thereby effectively buffering stream temperature from rapid changes in air temperature [e.g. *Caissie*, 2006; *Kaushal et al.*, 2010; *Zhou et al.*, 2014]. As such, our reach-level random effect of air temperature may indicate relative groundwater input, but would reflect a combination of the amount of ground water and the depth or temperature of that ground water relative to the surface temperature. We did not test this hypothesis, but future work would benefit from comparing our reach-specific air temperature coefficient estimates to stream reaches with measured ground water influence. This is an active area of interest with other ground-water indices being explored and evaluated [*Figura et al.*, 2015; *Kurylyk et al.*, 2014; *Snyder et al.*, 2015].

Similar to the spatial correlation, we accounted for temporal autocorrelation in the data. Temporal autocorrelation occurs when the data from the adjacent time-steps are not independent, such as yesterday's and today's stream temperatures being more similar to each other than expected compared to more distantly spaced temperature measurements. This is a common issue in modeling time series and causes estimation problems and bias if not accounted for in the model [*Caissie*, 2006; *Benyahya et al.*, 2007; *Ver Hoef and Peterson*, 2010]. We assumed a first-order autoregressive process (AR1) on the residuals of the linear model. Our estimated AR1parameter was 0.77 (Table 2), indicating a high level of temporal autocorrelation in the data and the coefficient estimates would have likely be biased if we did not account for this process. We also allowed the daily stream temperature (overall intercept) to vary randomly by year to account for similarities within years not accounted for by the other temporal effects. The random variability attributed to year was moderately low (SD = 0.28) and contributed considerably less to the variance than reach or HUC8 spatial random effects (Table 2).

Finally, our model includes predictors of stream temperature other than just air temperature [*Letcher et al.*, 2016; challenge 6]. Of the parameters modeled, the current day's air temperature and the mean air temperature over the previous seven days had the largest effect on daily stream water temperature. This is not surprising as we limited our analysis to small streams and to the synchronized period of the year when air and water temperature are most correlated. Past studies of small streams have also found air temperature to be the main predictor of stream temperature [e.g. *Mohseni and Stefan*, 1999; *DeWeber and Wagner*, 2014; *Detenbeck et al.*, 2016]. However, the effects of air temperature and 7-day air temperature were not identical across space. These effects varied moderately across sites and HUC8 (Table 2), with similar variance for both temperature effects although the daily air temperature had a slightly larger mean effect (Table 2). Additionally, air temperature had significant 3-way interactions with precipitation and drainage area. We used 2-day precipitation multiplied by drainage area as an index of flow associated with storms and 30-day precipitation multiplied by drainage area as an index of baseflow in these small headwater streams (A. Rosner *personal communication*). Therefore, the negative 3-way interactions with air temperature are what we would expect,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

indicating that at high flows the effect of air temperature on water temperature is dampened. The effect sizes of these interactions are extremely small, likely in part because of the coarseness of using precipitation multiplied by drainage area as an index of flow and not being able to account for local ground-surface water interactions at this broad spatial scale. Air temperature did not interact significantly with percent forest cover or impounded stream area. Alone, forest cover had a significant, but small, negative effect on stream temperature during the synchronized period, whereas impounded area had a significant, moderately large positive effect on temperature (Table 2). However, the effect of impoundments is more complicated that we were able to model in a broad spatial model. Different types of impoundments may have different effects and these may vary seasonally depending on flow, directed releases, and stratification of impounded water.

Although there were many variables that influenced daily stream temperature in our model (95% credible intervals not overlapping zero), the primary drivers were mean daily air temperature and 7-day previous air temperature (Table 2). These both corresponded strongly with stream temperature. There were very slight moderations of the air temperature effect due to our flow indices but they had limited influence on stream temperature. This may be due to the course nature of the flow index necessary for modeling over such a large range. Air temperature and flow have been important factors influencing stream temperature in other stream temperature models *Smith and Lavis*, 1975; *Webb et al.*, 2003; *van Vliet et al.*, 2011]. Additionally, impoundments had a moderate warming effect, whereas forest cover modest general cooling effect on daily ice-free stream temperatures similar to what was found in other studies [*DeWeber and Wagner*, 2014; *Jackson et al.*, 2018]. Neither impounded area or forest cover moderated the effect of daily air temperature on stream temperature (Table 2). In addition to explaining factors influencing stream temperatures, we were interested in predicting temperatures in unsampled time and space. As such, we examined the model fit as well as its predictive ability with various types of validation data. Our model fit the data well as indicated by the RMSE < 1 C. It also had a good ability to predict daily stream temperatures across space and time. With regards to predicting temperatures in warm years without fitted data, such as 2010, the model predicted temperatures well even in reaches with no other data (RMSE = 2.13 C). The predictions were better at reaches with data from other years (RMSE = 1.90 C), indicating that reach-specific data can improve predictions in future years but this improvement is not dramatic. The lack of dramatic improvement is likely due to multiple factors. Some of the reach-level variability is probably accounted for by other nearby reaches within the same HUC8 (influence of HUC8 random effects). We did not have sufficient data from combinations of reaches, HUC8, and years to compare the RMSE for HUC8 with single versus multiple observed reaches, but based on similar levels of variability explained at the reach and HUC8 levels it is likely that having data from other reaches in a HUC8 improves the predictions for unmonitored reaches in the same HUC8. Therefore, predictions will generally be worse at reaches within HUC8 with no data.

There are a few limitations in the current iteration of our modeling approach. The first is that ground-water influence is not explicitly accounted for, although the model may be useful in identifying ground water at sampled reaches. If region-scale ground water indices were available, it would be straightforward to add these into our modeling framework. The other limitation is that our model does not include spatially-explicit correlation following the downstream flow of the network [*Peterson et al.*, 2013; *Isaak et al.*, 2014]. inclusion of continuous spatial autocorrelation, but for now we found that reach and HUC8 random effects handled the spatial

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

correlations sufficiently. Other future advances could include detailed effects of impoundments using an exponential decay with distance, using a nonlinear model to include winter temperatures and larger streams [*Mohseni et al.*, 1998], or a dynamic model that allows the effect of air temperature to vary over time (seasonally) to account for hysteresis [*Letcher et al.*, 2016; challenge 2].

Our measure of resistance of water temperature to changes in air temperature during the summer was based on the cumulative difference between the daily air and water temperatures. This metric was intended as a potential index of ground water influence on stream temperature, similar to the reach-specific air temperature effect but focusing on the buffering during summer when ground water inputs are likely to have the largest influence. Streams with larger resistance values would be expected to have higher ground water influence because they would essentially be buffered from changes in air temperature during the warmest part of the year. This value could be adjusted for drainage area or flow since it is possible that larger streams always fluctuate less and it could be divided by mean water temperature during the summer to make it reflect the relative resistance. We anticipate future efforts to quantify the influence of ground water in summer stream temperature and explore how well this metric is able to predict those values.

Similarly, thermal sensitivity (Figure 5) or the size of the specific reach random effect could serve as indicators of ground water influence. In particular, the specific reach slope of air temperature suggests that reaches with larger coefficients are highly responsive to changes in air temperature (little ground water buffering) and reaches with small coefficients are insensitive to changes in air temperature and therefore likely to have significant ground water influence. These metrics are hypothesized to indicate ground water influence but remain to be tested. Given the differences in the distributions of these metrics (Figure 5), it is likely that some will be considerably more effective as ground water indices than other metrics. A similar effort has recently shown promise in creating a ground water influence index from stream temperature data [*Snyder et al.*, 2015]. These indices would currently only apply to reaches with observed data, so the next step would be to find landscape and geological parameters that could predict the best ground water index across the region.

Acknowledgments

The authors thank A. Rosner for thoughtful discussions related to the analysis and inference and give special thanks to J. Walker for database creation and management, development of the Interactive Catchment Explorer, along with general discussions of the model and its uses. We also would like to thank all of the groups who kindly provided data to make this project possible. This project was funded by the USFWS North Atlantic Conservation Cooperative and the USGS Northeast Climate Science Center. Not all data used in the model are designated as “public”. Public data are available at http://db.ecosheds.org/ and all data (public and private) used in this manuscript can be provided by B. Letcher (bletcher@usgs.gov) in consultation with the agencies that collected the data.

References

Ahmadi-Nedushan, B., A. St-Hilaire, T. B. M. J. Ouarda, L. Bilodeau, E. Robichaud, N. Thiemonge, and B. Bobee (2007), Predicting river water temperatures using stochastic models : case study of the Moisie River (Quebec , Canada), *Hydrol. Process.*, *34*, 21–34,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

doi:10.1002/hyp.

Al-Chokhachy, R., S. J. Wenger, D. J. Isaak, and J. L. Kershner (2013a), Characterizing the Thermal Suitability of Instream Habitat for Salmonids: A Cautionary Example from the Rocky Mountains, *Trans. Am. Fish. Soc.*, *142*(3), 793–801, doi:10.1080/00028487.2013.778900.

Al-Chokhachy, R., J. Alder, S. Hostetler, R. Gresswell, and B. Shepard (2013b), Thermal controls of Yellowstone cutthroat trout and invasive fishes under climate change, *Glob. Chang. Biol.*, *19*(10), 3069–81, doi:10.1111/gcb.12262.

Beauchene, M., M. Becker, C. J. Bellucci, N. Hagstrom, and Y. Kanno (2014), Summer Thermal Thresholds of Fish Community Transitions in Connecticut Streams, *North Am. J. Fish. Manag.*, *34*(1), 119–131, doi:10.1080/02755947.2013.855280.

Benyahya, L. et al. (2007), A review of statistical water temperature models, *Can. Water Resour. J.*, *32*(3), 179–192.

Benyahya L., A. St-Hilaire, T. B. M. J. Ouarda, B. Bobée, and J. Dumas (2008), Comparison of non-parametric and parametric water temperature models on the Nivelle River, France. *Hydrol. Sci. J. 53*(3), 640-655, doi:10.1623/hysj.53.3.640.

Brooks, S. P. B., and A. G. Gelman (1998), General methods for monitoring convergence of iterative simulations, *J. Comput. Graph. Stat.*, *7*(4), 434–455, doi:10.2307/1390675.

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

Caissie, D., N. El-jabi, and M. G. Satish (2001), Modelling of maximum daily water temperatures in a small stream, *J. Hydrol.*, *251*(2001), 14–28.

Chang, H., and M. Psaris (2013), Local landscape predictors of maximum stream temperature and thermal sensitivity in the Columbia River Basin, USA., *Sci. Total Environ.*, *461–462*, 587–600, doi:10.1016/j.scitotenv.2013.05.033.

Detenbeck, N. E., A. C. Morrison, R. W. Abele, and D. A. Kopp (2016), Spatial statistical network models for stream and river temperature in New England, USA, *Water Resour. Res.*, *52*, 6018–6040, doi:10.1029/2008WR006912.M.

DeWeber, J. T., and T. Wagner (2014), A regional neural network ensemble for predicting mean daily river water temperature, *J. Hydrol.*, *517*, 187–200, doi:10.1016/j.jhydrol.2014.05.035.

Figura, S., D. M. Livingstone, and R. Kipfer (2015), Forecasting Groundwater Temperature with Linear Regression Models Using Historical Data, *Groundwater*, *23*(6), 943–954, doi:10.1111/gwat.12289.

Garreta, V., P. Monestiez, and J. M. Ver Hoef (2009), Spatial modelling and prediction on river networks: up model, down model or hybrid?, *Environmetrics*, *21*(5), 439–456,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

doi:10.1002/env.995.

Gelman, A., and J. Hill (2007), *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press, New York.

Hocking, D. J., K. J. Babbitt, and M. Yamasaki (2013), Comparison of silvicultural and natural disturbance effects on terrestrial salamanders in northern hardwood forests, *Biol. Conserv.*, *167*(1985), 194–202, doi:10.1016/j.biocon.2013.08.006.

Ver Hoef, J. M., and E. E. Peterson (2010), A Moving Average Approach for Spatial Statistical Models of Stream Networks, *J. Am. Stat. Assoc.*, *105*(489), 6–18, doi:10.1198/jasa.2009.ap08248.

Hurlbert, S. H. (1984), Pseudoreplication and the design of ecological field experiments, *Ecol. Monogr.*, *54*(2), 187–211, doi:10.2307/1942661.

Isaak, D. J. et al. (2014), Applications of spatial statistical network models to stream data, *WIREs Water*, *1*(June), 277–294, doi:10.1002/wat2.1023.

Isaak, D. J. D., C. H. Luce, B. E. B. Rieman, D. E. Nagel, E. E. Peterson, D. L. Horan, S. Parkes, and G. L. Chandler (2010), Effects of climate change and wildfire on stream temperatures and salmonid thermal habitat in a mountain river network., *Ecol. Appl.*, *20*(5), 1350–1371.

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

Johnson, S. L. (2004), Factors influencing stream temperatures in small streams: substrate effects and a shading experiment, *Can. J. Fish. Aquat. Sci.*, *61*(6), 913–923, doi:10.1139/f04-040.

Jones, L. A., C. C. Muhlfeld, L. A. Marshall, B. L. Mcglynn, and J. L. Kershner (2014), Estimating thermal regimes of bull trout and assessing the potential effects of climate warming on critical habitats, *River Res. Appl.*, *30*, 204–216, doi:10.1002/rra.

Kanno, Y., J. Vokoun, and B. Letcher (2013), Paired stream-air temperature measurements reveal fine-scale thermal heterogeneity within headwater Brook Trout stream networks, *River Res. Appl.*, *30*(6), 745–755, doi:10.1002/rra.

Kaushal, S. S., G. E. Likens, N. a Jaworski, M. L. Pace, A. M. Sides, D. Seekell, K. T. Belt, D. H. Secor, and R. L. Wingate (2010), Rising stream and river temperatures in the United States, *Front. Ecol. Environ.*, *8*(9), 461–466, doi:10.1890/090037.

Kery, M. (2010), *Introduction to WinBUGS for Ecologists: A Bayesian Approach to Regression, ANOVA, Mixed Models, and Related Analyses*, Academic Press, Boston.

Kiraly, A., and I. Janosi (2002), Stochastic modeling of daily temperature fluctuations, *Phys.*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

Rev. E, *65*(5), 1–6, doi:10.1103/PhysRevE.65.051102.

Kurylyk, B. L., K. T. B. MacQuarrie, and C. I. Voss (2014), Climate change impacts on the temperature and magnitude of groundwater discharge from shallow, unconfined aquifers, *Water Resour. Res.*, *50*, 3253–3274, doi:10.1002/2013WR014588.

Letcher, B. H., D. J. Hocking, K. O'Neill, A. R. Whiteley, K. H. Nislow, and M. J. O'Donnell (2016), A hierarchical model of daily stream temperature using air-water temperature synchronization, autocorrelation, and time lags, *PeerJ*, *4:e1727*, 1–26, doi:10.7717/peerj.1727.

Link W. and M. Eaton (2012), On thinning of chaings in MCMC. *Methods Ecol. Evol. 3*(1), 112-115, doi:10.1111/j.2041-210X.2011.00131.x

MassDEP (2013), *314 CMR 4.00 Massachusetts Surface Water Quality Standards*.

Mayer, T. D. (2012), Controls of summer stream temperature in the Pacific Northwest, *J. Hydrol.*, *475*, 323–335, doi:10.1016/j.jhydrol.2012.10.012.

Milanovich, J. R., D. J. Hocking, W. E. Peterman, and J. A. Crawford (2015), Effective Use of Trails for Assessing Terrestrial Salamander Abundance and Detection : A Case Study at Great Smoky Mountains National Park, *Nat. Areas J.*, *35*(4), 590–598, doi:10.3375/043.035.0412.

Mohseni, O., and H. G. Stefan (1999), Stream temperature/air temperature relationship: a physical interpretation, *J. Hydrol.*, *218*(3–4), 128–141, doi:10.1016/S0022-1694(99)00034-7.

Mohseni, O., H. G. Stefan, and T. R. Erickson (1998), A nonlinear regression model for weekay stream temperatures, *Water Resour. Res.*, *34*(10), 2685–2692.

Peterson, E. E., J. M. Ver Hoef, and M. Jay (2010), A mixed-model moving-average approach to geostatistical modeling in stream networks, *Ecology*, *91*(3), 644–651.

Peterson, E. E. et al. (2013), Modelling dendritic ecological networks in space: an integrated network perspective., *Ecol. Lett.*, *16*(5), 707–19, doi:10.1111/ele.12084.

Plummer, M. (2016), rjags: Bayesian graphical models using MCMC,

R Development Core Team (2016), R: A language and environment for statistical computing,

Sivri, N., N. Kilic, and O. N. Ucan (2007), Estimation of stream temperature in Firtina Creek (Rize-Turkiye) using artificial neural network model, *J. Environ. Biol.*, *28*(1), 67–72.

Sivri, N., H. K. Ozcan, O. N. Ucan, and O. Akincilar (2009), Estimation of Stream Temperature in Degirmendere River (Trabzon- Turkey) Using Artificial Neural Network Model, *Turkish J. Fish. Aquat. Sci.*, *9*, 145–150, doi:10.4194/trjfas.2009.0204.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018

Smith K., M. Lavis, (1975), Environmental influences on the temperature of a small upland stream. *Oikos 26*, 228-236, doi:10.2307/3543713.

Snyder, C. D., N. P. Hitt, and J. A. Young (2015), Accounting for groundwater in stream fish thermal habitat responses to climate change, *Ecol. Appl.*, *25*(5), 1397–1419, doi:10.5061/dryad.th6g8.

Sura, P., M. Newman, and M. A. Alexander (2006), Daily to Decadal Sea Surface Temperature Variability Driven by State-Dependent Stochastic Heat Fluxes, *J. Phys. Oceanogr.*, *36*, 1940–1958.

van Vliet, M. T. H., F. Ludwig, J. J. G. Zwolsman, G. P. Weedon, and P. Kabat (2011), Global river temperatures and sensitivity to atmospheric warming and changes in river flow, *Water Resour. Res.*, *47*, 1–19, doi:10.1029/2010WR009198.

Webb B. W., P. D. Clack, D. E. Walling, (2003), Water-air temperature relationships in a Devon river system and the role of flow. *Hydrol. Proc., 17*, 3069-3084, doi:10.1002/hyp.1280.

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

Webb J. A., M. J. Stewardson, and W. M. Koster (2010), Detecting ecological responses to flow variation using Bayesian hierarchical models. *Freshwater Biol. 55*, 108-126, doi:10.1111/j.1365-2427.2009.02205.x.

Xu, C., B. H. Letcher, and K. H. Nislow (2010a), Context-specific influence of water temperature on brook trout growth rates in the field, *Freshw. Biol.*, *55*(11), 2253–2264, doi:10.1111/j.1365-2427.2010.02430.x.

Xu, C. L., B. H. Letcher, and K. H. Nislow (2010b), Size-dependent survival of brook trout Salvelinus fontinalis in summer: effects of water temperature and stream flow, *J. Fish Biol.*, *76*(10), 2342–2369, doi:10.1111/j.1095-8649.2010.02619.x.

Zhou, S., X. Yuan, S. Peng, J. Yue, X. Wang, H. Liu, and D. D. Williams (2014), Groundwater-surface water interactions in the hyporheic zone under climate change scenarios, *Environ. Sci. Pollut. Res.*, 13943–13955, doi:10.1007/s11356-014-3255-3.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27069v1 | CC0 Open Access | rec: 30 Jul 2018, publ: 30 Jul 2018