Journal of Hydrology 517 (2014) 187–200

Contents lists available at ScienceDirect

Journal of Hydrology

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

A regional neural network ensemble for predicting mean daily river water temperature

Jefferson Tyrell DeWeber [a,*], Tyler Wagner [b]

[a] Pennsylvania Cooperative Fish and Wildlife Research Unit, The Pennsylvania State University, 413 Forest Resources Building, University Park, PA 16802, USA
[b] U.S. Geological Survey, Pennsylvania Cooperative Fish and Wildlife Research Unit, The Pennsylvania State University, 402 Forest Resources Building, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 2 October 2013
Received in revised form 17 March 2014
Accepted 5 May 2014
Available online 22 May 2014
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Fi-John Chang, Associate Editor

Keywords:
Water temperature
Neural network
River
Stream
Ensemble
Regional prediction

SUMMARY

Water temperature is a fundamental property of river habitat and often a key aspect of river resource management, but measurements to characterize thermal regimes are not available for most streams and rivers. As such, we developed an artificial neural network (ANN) ensemble model to predict mean daily water temperature in 197,402 individual stream reaches during the warm season (May–October) throughout the native range of brook trout *Salvelinus fontinalis* in the eastern U.S. We compared four models with different groups of predictors to determine how well water temperature could be predicted by climatic, landform, and land cover attributes, and used the median prediction from an ensemble of 100 ANNs as our final prediction for each model. The final model included air temperature, landform attributes and forested land cover and predicted mean daily water temperatures with moderate accuracy as determined by root mean squared error (RMSE) at 886 training sites with data from 1980 to 2009 (RMSE = 1.91 °C). Based on validation at 96 sites (RMSE = 1.82) and separately for data from 2010 (RMSE = 1.93), a year with relatively warmer conditions, the model was able to generalize to new stream reaches and years. The most important predictors were mean daily air temperature, prior 7 day mean air temperature, and network catchment area according to sensitivity analyses. Forest land cover at both riparian and catchment extents had relatively weak but clear negative effects. Predicted daily water temperature averaged for the month of July matched expected spatial trends with cooler temperatures in headwaters and at higher elevations and latitudes. Our ANN ensemble is unique in predicting daily temperatures throughout a large region, while other regional efforts have predicted at relatively coarse time steps. The model may prove a useful tool for predicting water temperatures in sampled and unsampled rivers under current conditions and future projections of climate and land use changes, thereby providing information that is valuable to management of river ecosystems and biota such as brook trout.

1. Introduction

Water temperature is a fundamental property of river habitat that shapes biological communities and determines ecosystem services. Water temperature can limit the distribution of species through physiological constraints and thus is an important factor in structuring aquatic assemblages (Caissie, 2006; Magnuson et al., 1979). River water temperature also places constraints on river metabolism and ecosystem services that depend upon energy transfers (Demars et al., 2011). Human activities that alter rivers directly (e.g., dams; reviewed in Olden and Naiman, 2010) or indirectly through changes to the landscape (e.g., land use; reviewed in Poole and Berman, 2001) can alter water temperatures. Global climate change is also expected to result in warmer river water temperatures (e.g., Mohseni et al., 1999; Nelson and Palmer, 2007; van Vliet et al., 2013) primarily as a result of increased air temperatures, and reduced summer flows may further exacerbate water temperature increases (Isaak et al., 2010; van Vliet et al., 2013). These changes are likely to affect riverine biota and may act independently or in conjunction with other abiotic or biotic factors to render river habitat unsuitable for some species (Ficke et al., 2007; Rahel and Olden, 2008). For example, stream warming due to climate change is predicted to have negative effects on coldwater fish species, such as Pacific salmon (*Oncorhynchus* spp.; Ruesch et al., 2012), but may also result in the upstream expansion of an introduced predator (smallmouth bass *Micropterus dolomieu*; Lawrence et al., 2012). Thus, the combined effects of physiological stress and expanding ranges of introduced predators could interact to have large negative effects on native coldwater fish populations. Because of its importance to biota and susceptibility to human

* Corresponding author. Tel.: +1 814 317 6591.
E-mail addresses: jtdeweber@gmail.com (J.T. DeWeber), txw19@psu.edu (T. Wagner).

http://dx.doi.org/10.1016/j.jhydrol.2014.05.035
0022-1694/

activities and climate change, river water temperature and anticipated changes resulting from climate and land use changes are of great interest for resource management and biodiversity conservation.

Although technological advances have made monitoring river water temperature comparatively feasible and inexpensive in recent years (Webb et al., 2008), it is still logistically infeasible to measure, and difficult to obtain existing data, for a significant portion of river reaches across large basins or regions due to limited fiscal resources for monitoring and a lack of coordination among various research programs (Isaak, 2011). As a result, models predicting river water temperature characteristics for unsampled time periods, in unsampled rivers or under alternative management or environmental scenarios have become common in recent years (e.g., Hill et al., 2013; Isaak et al., 2010; Mohseni et al., 1998; Nelson and Palmer, 2007; Wehrly et al., 2009). For example, models are useful for making predictions of water temperature under future climate (Isaak et al., 2010; Mohseni et al., 1999), alternative land use scenarios (Hill et al., 2013; Nelson and Palmer, 2007; Sugimoto et al., 1997), or various water release scenarios from impoundments (Olden and Naiman, 2010; Wright et al., 2009). Models are also useful for understanding the processes that control river water temperature (e.g., Johnson, 2004; Story et al., 2003). Models predicting river water temperature range from deterministic models that require detailed meteorological and hydrological data used to solve heat budget equations (e.g., Johnson, 2004; Story et al., 2003) to empirical models with varying degrees of spatial complexity (e.g., Ruesch et al., 2012) that rely upon relationships between water temperature observations and relatively easy to collect climatic and landscape variables (e.g., Chenard and Caissie, 2008; Hill et al., 2013; Isaak et al., 2010; Mohseni et al., 1998). Although deterministic models can perform well and are physically based, the detailed data on river-specific energy transfers that are required to develop these models makes transferability to other rivers difficult. By contrast, empirical models are often more easily transferable and thus more useful for predicting river water temperatures at unmonitored locations throughout large watersheds or regions to support local and transboundary management efforts (Caissie, 2006).

Hourly or daily variation in water temperature can be important for stream ecosystem functioning, and some models have predicted daily water temperature with moderate accuracy in individual streams using only air temperature (e.g., Caissie et al., 2001). However, because water temperature variability generally increases with the number of streams, empirical models for predicting in multiple streams and across regions usually predict at weekly, monthly or seasonal time steps to achieve reasonable accuracy (Caissie, 2006). The loss of temporal variation in predictions is undesirable because daily predictions could provide more information and can be summarized to yield weekly, monthly or seasonal metrics as needed. Prediction in geographically diverse basins and over large spatial extents is also improved by including landform, geological, and stream attributes that are directly or indirectly related to water temperature as predictors (e.g., Hill et al., 2013; Isaak et al., 2010; Wehrly et al., 2009). There are a growing number of empirical modeling techniques that allow for multiple predictors and have been used for predicting water temperature (e.g., regression, stochastic models with time series decomposition, geospatial models, machine learning). Artificial neural networks (ANNs) are a particularly promising machine learning method because they are able to model nonlinear relationships, handle interactions among predictors, and often have high predictive power (Lek and Guégan, 1999; Olden et al., 2008). ANNs have been used widely and often outperformed other methods for predicting streamflow (e.g., Besaw et al., 2010; Chen et al., 2013; Huo et al., 2012), dissolved oxygen (e.g., Antanasijević et al., 2013; Wen et al., 2013), fish species distributions (Olden and Jackson, 2002) and richness (Chang et al., 2013), and water temperature (e.g., Chenard and Caissie, 2008; Risley et al., 2003; Westenbroek et al., 2010).

Although predicting river water temperature is of importance for the management and conservation of many aquatic species (Domisch et al., 2011; Xenopoulos et al., 2005), it is of particular importance for the conservation of cold-water salmonids (Almodóvar et al., 2012; Jones et al., 2006; Isaak et al., 2010; McKenna et al., 2010; Ruesch et al., 2012,), including brook trout *Salvelinus fontinalis*. Brook trout is a species of management concern throughout much of its native range in the eastern U.S., and the Eastern Brook Trout Joint Venture (EBTJV, http://easternbrooktrout.org/) was formed to promote regional, transboundary management and conservation. Brook trout are limited physiologically to coldwater (mean July water temperature $<\sim 22$ °C) streams, rivers and lakes and are sensitive to habitat and biotic disturbances (MacCrimmon and Campbell, 1969). An EBTJV assessment concluded that brook trout populations were extirpated or reduced (>50% of previously suitable habitat lost) in >71% of subwatersheds, and these losses were attributed to human activities, which include historical forestry practices, habitat alterations, nonnative species introductions and recent land use changes (Hudy et al., 2008). Future water temperature increases as a result of global climate warming are expected to result in further losses of brook trout habitat throughout their native range in eastern North America (Clark et al., 2001; Flebbe et al., 2006; Meisner, 1990). Even where temperatures rise but remain suitable, brook trout growth could be reduced unless food availability and consumption increase with temperature (Ries and Perry, 1995). Past predictions of brook trout range shifts in the eastern U.S. due to climate change were made by identifying thermally suitable habitat based on surrogates of river water temperature (e.g., elevation, groundwater temperature as determined by mean annual air temperature), and overlaying projected air temperature changes to determine potential habitat losses (Flebbe et al., 2006; Meisner, 1990). Combining predicted river water temperature with thermal limits represents a more direct route for characterizing current thermally suitable habitat and future changes due to climate change.

To assist in the management of rivers and brook trout in the eastern U.S., we developed an ensemble model of 100 ANNs to predict mean daily river water temperature for the majority of streams throughout the brook trout's native range in the eastern U.S. We first compared four models of increasing complexity to determine how well daily water temperatures could be predicted by the following sets of predictors: (1) air temperature only, (2) air temperature and landform attributes, (3) air temperature, landform attributes and forested land cover, and (4) air temperature, landform attributes, and forest, agricultural and developed land covers. We then select a final model and demonstrate its utility by mapping predicted water temperatures averaged for the month of July across the 1980–2010 modeling period. Our ensemble approach proves useful for understanding the importance of predictor variables and we are not aware of other models described in the peer-reviewed literature that predict daily water temperatures in individual stream reaches throughout a similarly large region.

2. Study area

The study region included the native range of brook trout in the eastern U.S. as defined by the EBTJV, and represents approximately 30% of the worldwide native range of brook trout and 70% of its range in the U.S. (Fig. 1; Hudy et al., 2008). We modified the EBTJV region slightly to align with the boundaries of local catchments from the National Hydrography Dataset Plus Version 1.0

(NHDPlusV1; USEPA and USGS, 2005). The 197,402 NHDPlusV1 stream reaches that formed a topologically connected river network (i.e., canals, pipelines and other non-river reaches were not included) with all predictors of water temperature available was the base layer of the Geographic Information System (GIS) environment upon which our temperature model was built. The region reflects the distribution of brook trout, which are limited to higher elevations (>200 m, Flebbe 2006) in the southern portion of the region but are found at all elevations at higher latitudes, corresponding to suitably cold water temperatures. The predominant land cover in the region is second-growth forest, but several urban centers support a large human population, agricultural land use is widespread in lower elevations, and forest management is common throughout much of the range.

3. Methods

3.1. Water temperature data

We compiled water temperature data from state agency personnel, watershed organizations, authors of published studies, and publicly available data from the USGS National Water Information System (NWIS; http://waterdata.usgs.gov/nwis). One-time 'snapshot' temperature recordings were not used because continuous (e.g., hourly) measurements were needed to calculate mean daily river water temperatures. To include only sites located on NHDPlusV1 stream reaches, reduce the effects of dams on water temperature observations, and remove likely measurement errors, we screened water temperature data as follows. First, we assigned all water temperature sampling sites to the nearest NHDPlusV1 stream reach, and removed 67 sites that were not within 250 m of any stream reach. We used 250 m because coordinate accuracy was unknown and some coordinates likely originated from topographic maps with relatively poor accuracy, but most sites (85%) were within 50 m of a stream reach (mean = 30.9 m). Nevertheless, some sampling sites were located near confluences and could be attributed to the wrong stream reach using this criterion, which would associate the wrong landscape attributes with water temperature observations and affect the model. We would have ideally verified that sites were correctly attributed to stream reaches by comparing stream names, but this was not possible because 41% (80,932) of NHDPlusV1 stream reaches did not have names. Since

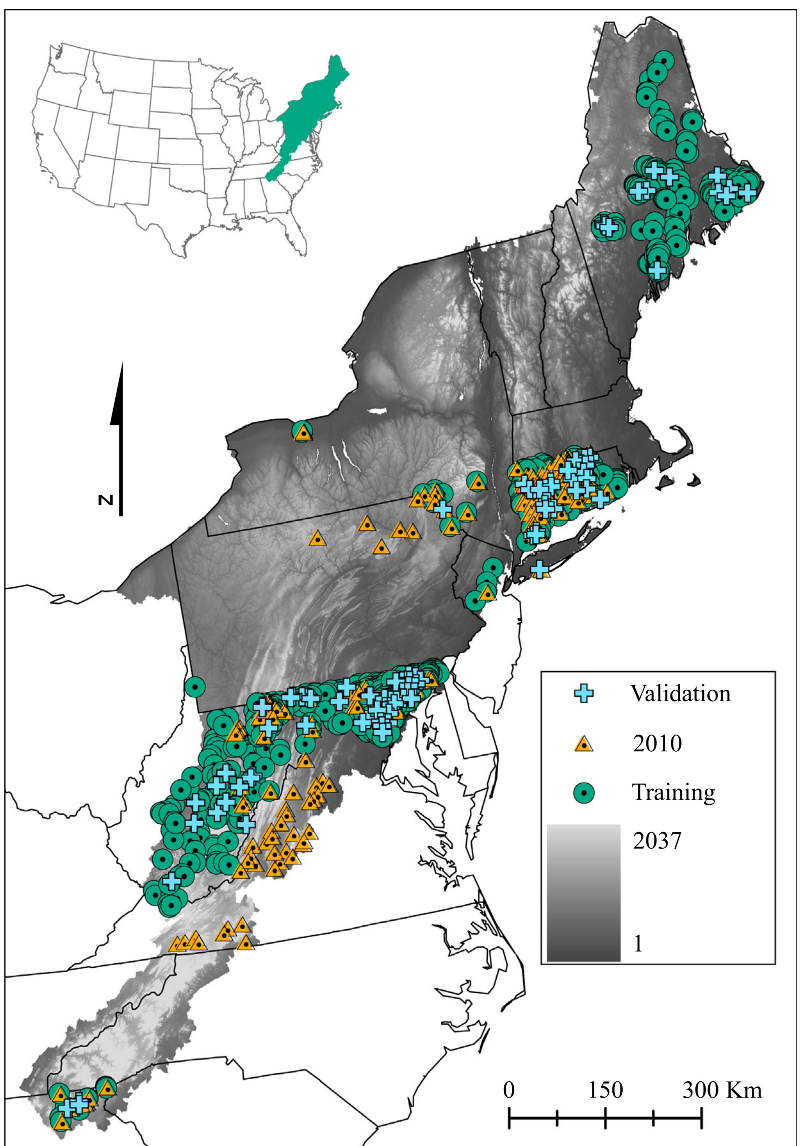

Fig. 1. Study region showing the location of stream reaches with water temperature observations used in training, validation and 2010 datasets. The background shows elevation (m), and the inset map shows the location of the study region relative to the continental U.S.

we were unable to match names, we fit a model excluding sites that were within 50 m of a confluence to determine if including sites that were most likely to have been attributed to the wrong stream reach (i.e., within 50 m of other stream reaches) negatively affected model performance. However, model accuracy and modeled relationships were nearly identical to the model including all sites (J.T. Deweber, unpublished data), suggesting that sites near confluences were most likely attributed to the correct stream each and did not affect the model. Thus, we linked water temperature sites to the nearest stream within 250 m, regardless of the proximity of confluences or other streams. We averaged water temperature in rare cases (6.1% of stream reaches) when two observations were available for the same stream reach and date. With the exception of data from the West Virginia Department of Natural Resources (WVDNR), we removed all sites where the nearest upstream dam (as determined by the National Inventory of Dams; USACE, 2005) was >100 ft in height or within 5 km. We chose these cutoffs because very large dams (i.e., >100 ft in height) can influence river water temperatures far downstream (Lehmkuhl, 1972; Lowney, 2000) and even relatively small dams can alter temperatures for short distances downstream (i.e., <5 km; Lessard and Hayes, 2003). We acquired data from the WVDNR after beginning model development and chose to use a more conservative criterion by removing all sites with a dam upstream regardless of size or distance because the other criteria would have involved significant resources to correct hydrologic errors in the NHDplusV1 network in West Virginia.

We selected all data from the screened sites collected after 1980 during the May to October period when water temperatures likely reach their maxima and are limiting for brook trout and other stenothermic biota. Daily water temperature observations were removed if any individual observation >35 °C or <0 °C was reported during the day or if the daily range (maximum minus minimum) was <0 °C or >30 °C. A few of the data sources provided only mean daily temperatures and we assumed that these had been previously screened for such errors. We then removed mean daily observations that exceeded 3 standard deviations of the annual mean temperature at a site because such extreme values were potentially air temperature measurements when loggers were exposed to the air due to reduced stream flows. Lastly, we removed a small number (212) of mean daily water temperature observations from 7 stream reaches during model development that were obvious errors. All included sites were required to have $\geqslant$30 mean daily records during the month of July because this is a critical period for brook trout due to high water temperatures.

3.2. Climatic and landscape predictors

We downloaded daily, empirical air temperature records for the time period 1980–2010 from the U.S. Historical Climatology Network from the National Climate Data Center (http://www.ncdc.noaa.gov/) for all sites that were within the study region plus a 10 km buffer to reduce artificial boundary effects. There were 1086 sites that recorded air temperature within the study region, but few sites recorded air temperature for all days during the 31 year period. To ensure that all stream reaches had records for all days during the study period, we selected the nearest 10 climate stations and calculated mean air temperature (from the current day). Because recent air temperature is likely to be important for determining water temperature, we also calculated prior 7 day mean air temperature – the moving average of air temperature from the previous 7 days inclusive of the current day. The average distance separation between stream reaches and the 10 nearest climate sites ranged from 12.0 to 90.1 km, with a mean of 30.7 km; the furthest distance separation between a stream reach and a climate station was 121.5 km. Although there could be substantial differences in observed air temperatures and elevations among the 10 nearest climate sites, including 10 sites was necessary because of gaps in air temperature records. In fact, a few stream reaches did not have available air temperature data for all days even with the inclusion of 10 air temperature sites. However, because increasing the number of sites would also increase distances between stream reaches and climate stations, we chose not to include the 961 (<0.4%) daily water temperature observations for which air temperature data was not available.

We compiled data for landform and land cover attributes that we expected to be important predictors of water temperature. All attributes except riparian forest were used in a national assessment of fish habitat condition and are described in Esselman et al. (2011). We calculated riparian forest as the % cover of forest within a 30 m buffer on each side of the NHDplusV1 stream reaches (60 m total width). Although elevation is closely related to water temperature, we did not include it in our models because it is indirectly related to water temperature primarily through effects on climate, and its inclusion could underestimate the effects of air temperature, especially under warming scenarios (Stanton et al., 2012). All physiographic attributes were summarized within the local (i.e., the portion of the catchment directly adjacent to each stream reach) and network (i.e., all areas upstream including the local catchment) catchments of each stream reach, which gave a total of 17 potential predictors. We refer to attributes as local or network depending upon the scale of measurement. To select a final set of predictors with minimal collinearity, we selected the attribute that was more strongly correlated with water temperature when two or more were correlated ($|r| > 0.5$). We used a different approach for selecting among highly correlated land cover types because we were specifically interested in modeling relationships between water temperature and each land cover type, especially local riparian forest cover, as several studies have shown the importance of shade from riparian vegetation on nearby temperatures (Johnson, 2004; Jones et al., 2006; Rutherford et al., 2004). Therefore, we selected a set of moderately uncorrelated ($|r| < 0.6$) land cover predictors that included at least one measure of riparian forest cover as well as forest, agricultural and developed land cover within either the local or network catchment. We recognized that the 2001 National Land Cover Dataset (NLCD) was unlikely to adequately reflect land cover during the 1980–2010 study period. However, we compared and found no discernible differences in model performance or relationships between models trained using data from 1980 to 2010 and models trained with data from only 1999 to 2003, which is centered on the year of land cover collection (J.T. Deweber, unpublished data). Therefore, we included 2001 NLCD as a metric of natural land cover and human activity. This process resulted in 7 potential landscape predictor variables (Table 1). Prior to model fitting, land cover variables (proportions) were logit transformed (0 and 1 were changed to 0.025 and 0.975 prior to transformation, respectively) and network catchment area was $\log_{10}$ transformed. All predictor variables were then standardized to mean 0 and standard deviation of 1.

3.3. Model comparisons

We compared the following four models of increasing complexity to determine how well water temperature could be predicted by different sets of predictor variables:

(1) *An air temperature model:* This model included only mean air temperature from the current day and prior 7 day mean air temperature based on the prediction that air temperature is the best available predictor of predict water temperature regionally. Air temperature is closely related to climatic factors that determine water temperature (Caissie, 2006) and

Table 1
Names and sources of all natural and human disturbance landscape attributes that were used in analyses. The land cover code column lists the reference numbers from the source dataset used to calculate land cover types used in our analyses.

Attribute	Resolution	Units	Source	Land cover code
Network area	1:100,000	km[b]	Calculated using NHDPlusV1[a]	NA
Network mean aspect	30 m	Degree	National Elevation Dataset[b]	NA
Network mean baseflow index	1:100,000	% Groundwater contribution to baseflow	Wolock (2003)	NA
Local riparian forest	30 m	% Cover	NLCD 2001 version 13	41 + 42 + 43
Network forest	30 m	% Cover	NLCD 2001 version 13	41 + 42 + 43
Network developed land	30 m	% Cover	NLCD 2001 version 1[c]	21 + 22 + 23 + 24
Local agriculture	30 m	% Cover	NLCD 2001 version 1[c]	81 + 82

[a] USEPA and USGS (2005).
[b] Available at <http://ned.usgs.gov/>.
[c] Homer et al. (2004).

several models have predicted water temperature based off of air temperature alone (e.g., Chenard and Caissie, 2008; Mohseni et al., 1998).

(2) *A landform model:* This model included landform attributes (rows 1 – 3 in Table 1) in addition to air temperature predictors because these describe static watershed and stream characteristics that may affect water temperature and improve predictions. We selected three landform attributes that might affect water temperature through effects on river size and network position (network area), solar radiation (network mean aspect), or groundwater interactions (network baseflow index).

(3) *A forest landscape model:* This model included all predictors of model 2 plus measures of local riparian forest and network forest land cover. Forest land cover in the riparian zone and network catchment was expected to be related to lower mean daily water temperatures in the summer through effects on shading, ground temperature, and exposure to atmospheric energy transfers (Caissie, 2006).

(4) *An anthropogenic landscape model:* This model included local catchment agriculture and network catchment developed land covers in addition to all predictors in model 3 except network forest cover, which was not included due to high correlations with agriculture and developed land covers. We expected that measures of anthropogenic land cover may improve predictions as agriculture and developed land cover have been related to water temperature alterations, including increased summer temperatures (Hill et al., 2013; Poole and Berman, 2001).

3.4. Neural networks

Feed forward neural networks are widely used in ecology (e.g., Lek and Guégan, 1999; Olden and Jackson, 2002) and have been used for predicting river water temperatures (e.g., Chenard and Caissie, 2008; McKenna et al., 2010; Risley et al., 2003; Westenbroek et al., 2010). We briefly discuss the basics of ANNs that we used in this study. A single hidden-layer feed forward ANN (also referred to as a multi-layer perceptron) is a nonlinear model that consists of input neurons (predictor variables) connected to any number of hidden neurons in a single hidden layer, which are in turn connected to one output neuron (response variable). ANNs may also include skip-layer connections, which are direct connections between input neurons and output neurons that allow for linear relationships between predictors and the response variable. An ANN with skip-layer connections but no hidden neurons is analogous to a linear model, whereas an increasing number of hidden neurons allows for increasing nonlinearities in modeled relationships (Cheng and Titterington, 1994). Our models included skip-layer connections in addition to hidden neurons because preliminary comparisons showed that models with skip-layers achieved better performance with fewer weights (i.e., were more parsimonious) and models without skip-layer connections tended to underpredict warmer (>25 °C) water temperatures (J.T. Deweber, unpublished data). The learning process proceeds by assigning randomly selected (or pre-assigned) weights to the input-hidden, hidden-output, and input–output (i.e., skip-layer) connections, and iteratively adjusting the weights through a learning algorithm based on the difference between predicted and observed responses until a convergence criterion is met (Lek and Guégan, 1999).

We developed ANNs using the R package nnet (Venables and Ripley, 2002) using the conjugant gradient Broyden–Fletcher–Goldfarb–Shanno (BFGS) learning algorithm, which is recommended because it is more likely to find global optima compared to gradient descent methods (Dreyfus, 2005). One potential drawback of ANNs is that models can become overfit as a result of too many predictors, weights or training iterations (Dreyfus, 2005). A second drawback is that identical ANNs fit using the same dataset but different starting weights can have very different modeled relationships because they find locally optimal weights in complex datasets (Hansen and Salamon, 1990). While many studies develop several ANNs using different starting weights and select the best model based on model performance, an ensemble can improve predictions by combining information from multiple models (Hansen and Salamon, 1990). An ensemble approach can also be used to better understand the effects of predictor variables and their relative importance in ANNs analogous to variable importance measures in random forests (Breiman, 2001). Because our goal was a model with accurate predictor effects that could generalize to unsampled rivers in the EBTJV region, we selected an optimal ANN architecture using cross validation and used an ensemble of ANNs to make predictions, as described below.

Prior to model development, we withheld two validation sets to assess model performance in a different, relatively warmer year and at new sites. The first validation dataset included all data from 2010 to determine model performance under warmer conditions and at some new streams (133 of the 223 stream reaches with 2010 data did not have data from other years). We chose 2010 because mean July air temperature averaged across all climate stations in our region was at least 0.3 °C warmer and there were 26 more observations of extremely warm air temperatures (>32 °C) than any other year except 1999 (Table 2). We did not use 1999 data for validation because available water temperature modeling data was limited and did not reflect regionally warmer conditions. Mean July air temperatures were 1.3 °C warmer in the 2010 dataset than in the training dataset, partially due to regionally warmer conditions but also because many sites were in the southern portion of the study region (Fig. 1). After removing data for 2010, we obtained a second validation dataset for determining how well the model could generalize to new stream reaches across multiple years by randomly selecting 10% of stream reaches and withholding all associated data. We refer to these two validation datasets as

Table 2
Interannual comparison of three air temperature metrics for the 10 years with the warmest mean July air temperature for the entire region (Regional) or water temperature modeling dataset (Modeling Dataset). Metrics are sorted in decreasing order of July air temperature, and the standard deviation is shown in parentheses. July and Season were calculated as the regional average air temperature in July and throughout the May–October modeling season, respectively. $N > 32$ °C was the total number of mean daily air temperature observations exceeding 32 °C in each year across all temperature sites. These metrics show that 2010 air temperatures were warmer than almost all other years, especially for the modeling dataset.

Year	July	Season	N > 32 °C
Regional			
1999	23.3 (3.8)	17.9 (5.9)	74
2010	23.3 (3.7)	18.5 (5.9)	71
1988	22.9 (3.9)	17.1 (6.8)	43
2006	22.8 (3.1)	17.3 (6.0)	45
1995	22.7 (3.4)	17.8 (5.6)	27
1994	22.7 (3.0)	17.0 (5.8)	5
2005	22.7 (3.1)	18.1 (6.1)	15
1987	22.6 (3.7)	17.4 (6.2)	20
1993	22.6 (3.7)	17.4 (6.1)	36
2002	22.5 (3.8)	17.8 (6.5)	31
1980–2010	21.8 (3.5)	17.4 (5.8)	535
Modeling dataset			
2010	24.1 (3.1)	21.1 (4.7)	12
2008	22.8 (2.5)	19.2 (5.1)	0
2006	22.9 (2.4)	18.1 (5.2)	3
2002	22.8 (4.0)	19.9 (5.7)	17
1988	22.3 (3.6)	15.9 (6.7)	0
1994	22.0 (1.8)	16.9 (5.2)	0
1983	22.0 (2.7)	17.2 (5.8)	0
1999	21.8 (3.2)	17.3 (5.6)	0
2007	21.8 (3.2)	19.5 (4.9)	6
1980	21.8 (3.5)	17.1 (5.9)	0
1980–2010	21.8 (3.2)	18.8 (5.3)	38

the 2010 dataset and the validation dataset, respectively. The remaining data were used to select the ANN architecture and to train the final model. Withholding these two validation datasets meant that a large amount of data from approximately 30% of stream reaches could not be used for model training, which could have negatively affected model performance. To determine if additional training data might improve or otherwise alter modeled relationships, we also developed models using a much larger training dataset and a much smaller validation dataset comprised of data from 5% of stream reaches. The two different sized training datasets produced nearly identical models based on performance and modeled relationships (J.T. Deweber, unpublished data). This comparison suggested that the modeling approach is fairly invariant to the size of dataset used for model training. Thus, in this paper we present only the results of model development that used the two validation datasets described above because it provided a more rigorous validation of model performance under a larger range of conditions.

We used 10-fold cross validation with 3 repeats via the train function of the R package caret (Kuhn, 2008) to compare ANN architectures based on root mean square error (RMSE). The standard cross validation procedure split data randomly and resulted in a preliminary model with all predictors that performed exceptionally at locations included in the training data (RMSE = 1.26 °C) but could not generalize well to validation sites (RMSE = 2.68; J.T. Deweber, unpublished data). To ensure that we had a more accurate measure of generalizability to new sites, we employed an approach that we refer to as site-based cross validation: 90% of sites and respective data were randomly selected for training while the remaining 10% of sites and respective data were used to calculate generalization error in each iteration. Site-based cross validation ensured that the reported cross validation error was representative of prediction accuracy at new sites. For each of the four models described in Section 3.3, we used site-based cross validation to select the simplest ANN architecture with the best performance from a range of model architectures by varying the weight decay coefficient, number of hidden neurons and the number of training iterations. The weight decay coefficient penalizes unnecessarily large weights, which helps avoid overfitting and aids generalization (Krogh and Hertz, 1992). All four models described in Section 3.3 had optimal predictive ability with the same architecture: a decay coefficient of 0.1, 5 hidden neurons plus skip-layer connections, and 100 training iterations. Finally, because individual ANNs had different modeled relationships, we trained an ensemble of 100 ANNs for each model with the selected model architecture but different random starting weights. We calculated the median predicted water temperature from the ensemble of models as our final prediction because it was representative of most modeled relationships and was not sensitive to outliers.

3.5. Model characteristics and performance

For each model, we examined residual plots to compare predictions and observations, and to check for potential biases in model performance across the ranges of each predictor. To explore modeled relationships, we performed a sensitivity analysis and calculated sensitivity weights as a measure of variable importance using the procedure described by Olden et al. (2004). For the sensitivity analysis, at each of 24 evenly spaced values covering the range of each predictor variable, we predicted nine values of water temperature by varying all other predictors concurrently across nine evenly spaced quantiles from 0.1 to 0.9. We then calculated the median prediction at each of the 24 values of each predictor. A sensitivity weight for each predictor was calculated as the range in these 24 median predicted values, which approximates the maximum change in water temperature due to changes in the values of a predictor. We plotted the median responses for the predictions from each ANN individually and from the ensemble prediction. Lastly, for the ensemble predictions we also plotted water temperature responses to each predictor while holding all other variables at five selected quantiles (0.1, 0.3, 0.5, 0.7, and 0.9) to determine if the magnitude or direction of responses to a given predictor varied with the values of other predictors.

We selected a final model from models 1–4 described in Section 3.3 by comparing performance based on RMSE, but we also report three additional metrics that provide a more complete overview of model performance: RMSE divided by the standard deviation in the observed data (RMSE/SD), the Nash–Sutcliffe efficiency (NSE), and percent bias. RMSE/SD is a measure of prediction accuracy relative to the variability in the observed dataset that can use to compare the performance of different models and datasets (Hill et al., 2013; Moriasi et al., 2007). NSE is a measure of explained variation ranging from 0 (no variation explained) to 1 (all variation explained), and percent bias reflects whether a model tends to over-predict (negative values) or under-predict (positive values), with 0 representing no overall prediction bias (Moriasi et al., 2007). Values less than 0.5 for RMSE/SD, greater than 0.75 for NSE and less than ±10% for percent bias were suggested to represent very good performance for stream flow, nutrient and sediment models (Moriasi et al., 2007). To demonstrate the utility of this model, we summarized daily predictions to mean July water temperature from 1980 to 2010 and mapped these predicted mean July water temperatures in all 197,402 stream reaches in our study region. Predictive models in general, and neural network models in particular, may not perform well when new predictor values outside of the range of the training data are encountered. Thus, we also identified stream reaches with landscape predictor values outside the range of the training data and shaded these portions of the map to represent a 'map of ignorance' to identify portions of the study region where model performance may be especially limited (Rocchini et al., 2011).

4. Results

Of over 1 million records from 2565 sites that we compiled, 269,608 observations of mean daily water temperature from 1080 stream reaches met our criteria for inclusion in the model. We set aside 26,194 observations from 96 stream reaches for the validation dataset and 27,126 observations from 223 stream reaches for the 2010 validation dataset (Fig. 1). The training dataset included the remaining 216,288 observations from 866 stream reaches. Stream reaches in the training dataset had network catchment areas ranging from 0.568 to 19,208 km^2, and spanned a large range of other natural and anthropogenic landscape attributes (Fig. 2). The training, validation and 2010 datasets had similar variation in landscape attributes (Fig. 2). Of the 197,402 stream reaches in the population, 189,419 (96.0%) had landscape attribute values that were within the range of those in the training dataset. Most unrepresented streams had extreme values of network area (3750 stream reaches) or network mean aspect (3209 stream reaches), but some streams were not represented by network mean baseflow index or network forest.

The air temperature model predicted water temperature with only moderate accuracy (RMSE = 2.50, 2.55 and 2.46 °C for training, validation, and 2010 datasets, respectively) compared to other models, and the inclusion of landform predictors, especially network area, reduced RMSE by roughly 0.5 °C (Table 3). The addition of forested land cover further improved accuracy, but the anthropogenic landscape model performed poorer than the forest landscape model despite added complexity (Table 3). Further, we did not consider the anthropogenic landscape model to be suitable for predictions as the predicted negative effects of agriculture and developed land covers on water temperature (Fig. 3) were the opposite of previously published relationships and of the positive correlations with water temperature in our dataset (the correlation between water temperature and both local agriculture and network developed land covers was 0.13). We chose the forest landscape model as our final model because accuracy was highest (RSME = 1.91, 1.82, and 1.93 for training, validation, and 2010 datasets, respectively; Table 3) and the predicted negative effects of local riparian forest and network forest matched expectations from the literature. Relationships between predicted and observed mean daily water temperatures for the forest ensemble model were generally unbiased (i.e., followed a one-to-one relationship), but slight tendencies to over predict low temperatures and under predict high temperatures were evident (Fig. 4). For training and validation datasets combined, the average accuracy of daily predictions was very good (RMSE < 1.0 °C) at 22.5% of stream reaches and good (RMSE < 2.0 °C) at 74.3% of stream reaches, but was poor (RMSE > 4 °C) at a small number of stream reaches (2.3%). The model also performed reasonably well based on maximum deviation, as all predictions were within 2.0, 3.0, and 4.0 °C of observed water temperature at 39.3%, 59.4% and 75.5% of stream reaches.

Although we selected our final model using only RMSE, the other three performance metrics were also generally optimized for the forest landscape model (Table 3). NSE suggested that the forest landscape model explained at least as much variation for

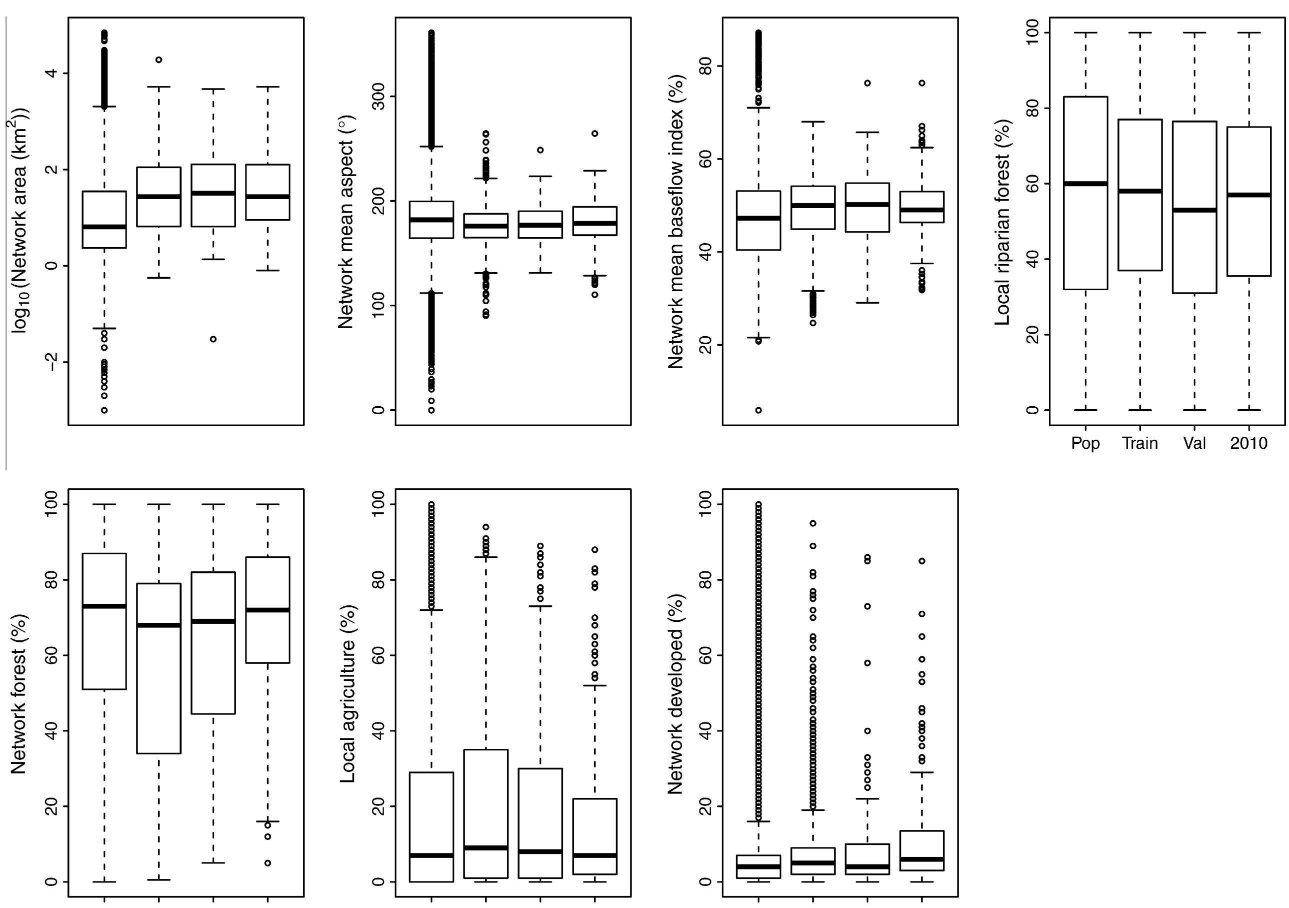

Fig. 2. Distributions of landscape predictor variables across the population of stream reaches (Pop), training (Train), validation (Val), and 2010 validation (2010) datasets used in the development and assessment of the river water temperature model. The predictors and their sources are listed in Table 1.

Table 3
Performance metrics and the number of weights (N weights) for 4 different river water temperature models (see Section 3.4 for model details). The performance metrics are root mean square error (RMSE, °C), RMSE divided by the observed standard deviation (RMSE/SD), the Nash–Sutcliffe efficiency (NSE), and percent bias (% bias).

Model	*N* weights	Subset	RMSE	RMSE/SD	NSE	% Bias
Air temperature	23	Training	2.50	0.62	0.62	0.00
		Validation	2.55	0.64	0.59	1.16
		2010	2.46	0.69	0.53	−1.99
Landform	41	Training	2.00	0.49	0.76	−0.03
		Validation	1.83	0.46	0.79	0.71
		2010	1.95	0.55	0.70	−1.42
Forest landscape	53	Training	1.91	0.47	0.78	−0.05
		Validation	1.82	0.46	0.79	0.26
		2010	1.93	0.54	0.71	−1.75
Anthropogenic landscape	65	Training	1.92	0.47	0.78	−0.07
		Validation	1.87	0.47	0.78	0.19
		2010	1.98	0.55	0.69	−1.74

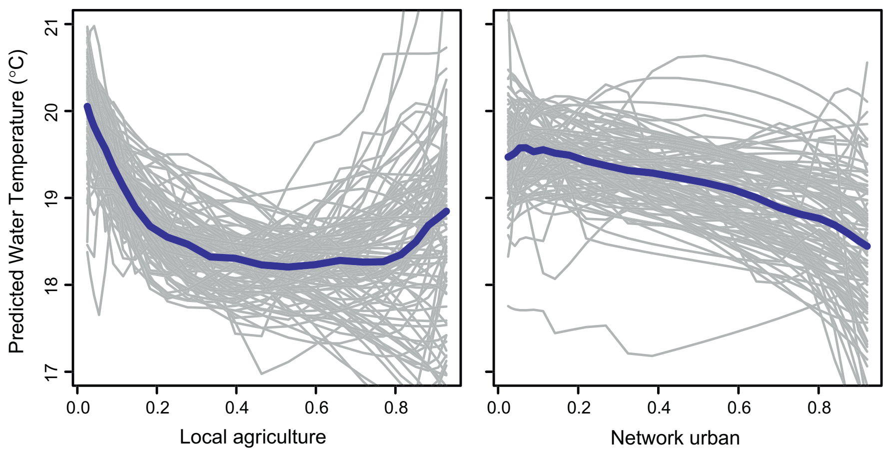

Fig. 3. Sensitivity analysis plots showing negative predicted mean daily river water temperature responses for the anthropogenic landscape ANN ensemble to increasing levels of local agriculture and network developed land covers. Predicted water temperature at each of 24 values of the predictor were calculated as the median of nine predictions obtained by varying all other predictors concurrently across nine evenly spaced quantiles from 0.1 to 0.9. The grey lines represent the predicted responses for each of the 100 ANNs in the ensemble, and the blue line represents the median prediction from all of the 100 ANNs and the final ensemble prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the training (0.78) and validation (0.79) datasets as other models, and explained more variation for the 2010 validation dataset (0.71) than all other models. Percent bias was relatively low for all models and well below the 10% cutoff for very good model performance suggested for streamflow data by Moriasi et al. (2007). However, percent bias values between −1.42 and −1.99 showed that water temperatures were slightly overpredicted on average for the 2010 validation dataset by all models (Table 3). All models had poorer performance for the 2010 validation datasets based on all metrics, but performance was still very good or good according to the guidelines of Moriasi et al. (2007).

Air temperature from the current day at the nearest 10 climate stations was the strongest predictor based on sensitivity analyses (sensitivity weight = 14.4, Table 4), but prior 7 day mean air temperature was also important (sensitivity weight = 10.2, Table 4). Sensitivity analysis plots revealed a strong linear relationship between air and water temperature (~0.4 °C water temperature increase per °C) that was consistent across individual ANNs in the ensemble (Fig. 5). The effect of prior 7 day air temperature was similar (~0.4 °C water temperature increase per °C) for air temperatures below 20 °C, but then gradually decreased to a ~0.2 °C water temperature increase per °C. The effects of landscape predictors generally varied greatly among individual ANNs, suggesting that effects were less certain compared to air temperature. Network area was the most important landscape attribute (sensitivity weight = 6.4; Table 4) and had a strong positive effect for very small watersheds (<1 km^2) that decreased until little effect was present for larger watersheds (>1000 km^2; Fig. 6). Ensemble predictions from the sensitivity analysis increased from ~16 °C for the smallest catchments to approximately 22 °C for the largest. Network mean aspect had a relatively weak positive effect on water temperature as it shifted from east (90°) to west (260°), whereas the effect of network baseflow index was nonlinear and produced a very small net effect (Fig. 6). Local riparian forest had a negative effect on water temperature across all values, and was strongest at extreme low and high values. In contrast, the effect of network forest was expected to decrease water temperature overall, but was weakly positive when forest was less than 75% and then strongly negative for forest cover >80%.

Daily predictions were slightly more accurate when summarized to calculate mean July water temperature (RMSE = 1.76 °C) compared to daily predictions for the training dataset, but were slightly lower for the validation dataset (RMSE = 1.89 °C) and 2010 datasets (RMSE = 1.97). Mean July prediction accuracy was very good (<1.0 °C) at 47.0% and good (<2.0 °C) at 77.8% of all stream reaches, but was poor (RMSE > 4 °C) at 3.4%. The spatial distribution of predicted mean July water temperature for 1980–2010 showed an expected transition from coldwater to warmwater stream reaches along a gradient from high to low elevation and upstream to downstream within the region (Fig. 7). Fig. 7 also shows the 7983 unrepresented stream reaches that had values of one or more landscape characteristics outside the range of the training data, where predictions may be most uncertain (Rocchini et al., 2011). Because the spatial detail of predictions

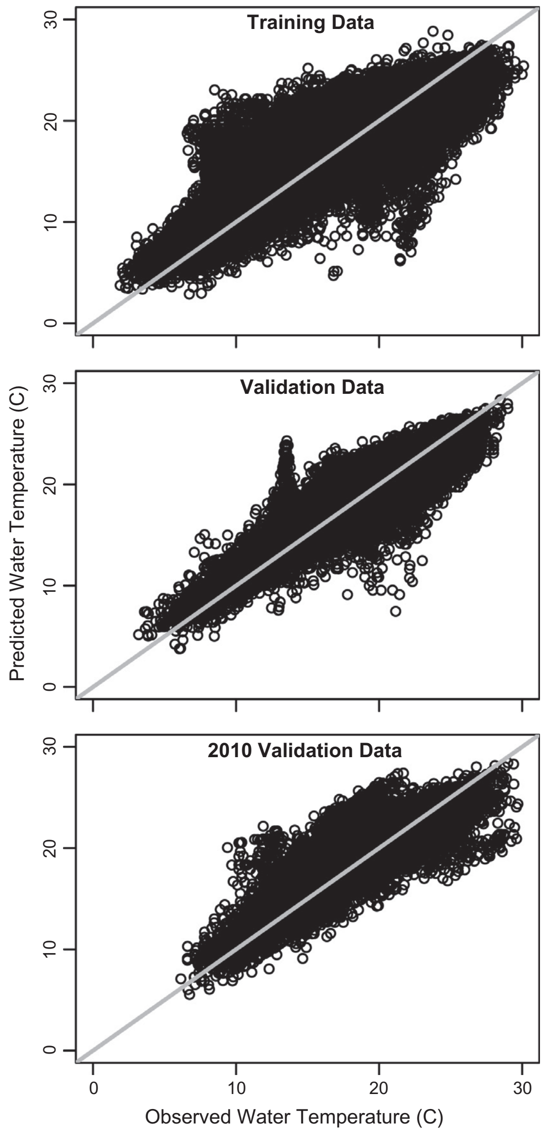

Fig. 4. Predicted and observed mean daily river water temperatures for the training, validation and 2010 validation datasets. The grey line represents a 1:1 line.

Table 4
Sensitivity weights and the direction of effect on water temperature of each predictor for the forest landscape ensemble model. The sensitivity weight is an approximation of the maximum change in water temperature due to a predictor and was calculated as part of sensitivity analyses described in Section 3.5 of the text.

Predictor Name	Sensitivity weight	Direction
Mean air temperature	14.4	+
Prior 7 day mean air temperature	10.2	+
Network area	6.4	+
Network forest	2.2	–
Network mean aspect	1.8	+
Network mean baseflow index	1.9	Unclear
Local riparian forest	1.4	–

within a river network is difficult to see at the regional extent, we also mapped predicted temperatures for a subset of the Penn's Creek watershed in central Pennsylvania (Fig. 8).

5. Discussion

Our results demonstrated that an ensemble of ANNs can accurately predict river water temperature at a daily time step within individual stream reaches throughout a large and geographically diverse region. Daily predictions have especially great value for management because they can capture relatively short-term temperature variation, which can drive system dynamics, and also be summarized to provide accurate metrics of thermal habitat (e.g., mean weekly or mean July water temperature). Further, predictions for individual, relatively short (~2 km) small stream reaches capture high spatial variability and can also be summarized to larger spatial extents as needed. Most models of water temperature at a daily temporal resolution have focused on single streams or relatively small basins and have typically achieved accuracies equivalent to 1–2 °C RMSE (Caissie et al., 2001; Chenard and Caissie, 2008; Gardner et al., 2003; Isaak and Hubert, 2001; Marcé and Armengol, 2008). Within moderate to large basins or regions, temperature modeling efforts have focused on predicting weekly, monthly, seasonal or annual river water temperature (e.g., Hill et al., 2013; Isaak et al., 2010; Mohseni et al., 1998; Wehrly et al., 2009). For example, Hill et al. (2013) used machine learning to model seasonal and annual water temperatures throughout the conterminous U.S. with good accuracy (RMSE = 1.2–2.0 °C).

It is difficult to compare our model performance to previous efforts because models predicting water temperature at daily

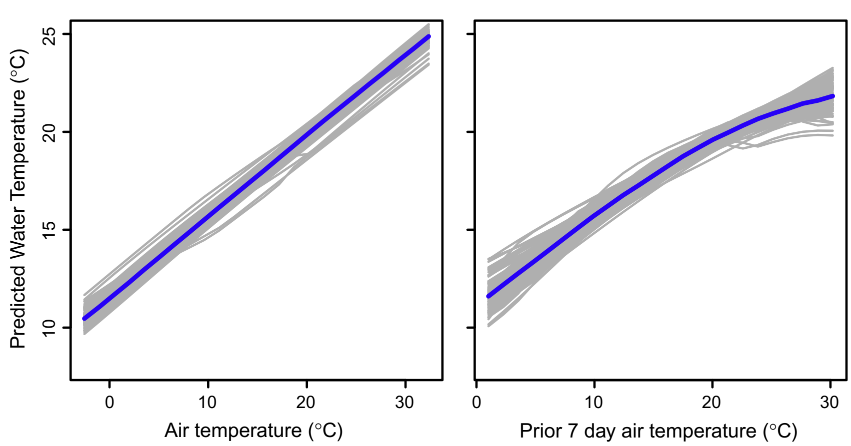

Fig. 5. Selected sensitivity analysis plots showing predicted mean daily river water temperature responses for the final selected ANN ensemble model to mean air temperature and prior 7 day mean air temperature. Predicted water temperature at each of 24 values of the predictor were calculated as the median of nine predictions obtained by varying all other predictors concurrently across nine evenly spaced quantiles from 0.1 to 0.9. The grey lines represent the predicted responses for each of the 100 ANNs in the ensemble, and the blue line represents the median prediction from all of the 100 ANNs and the final ensemble prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

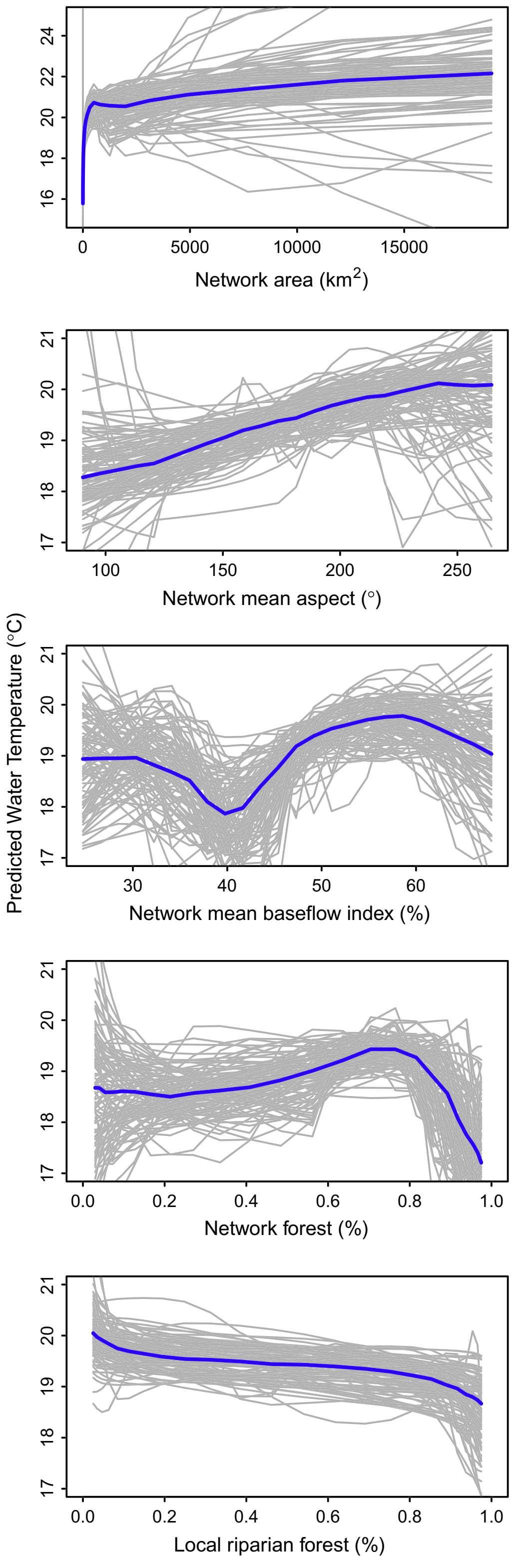

Fig. 6. Selected sensitivity analysis plots show predicted mean daily river water temperature responses for the final selected ANN ensemble model to network area, network mean aspect, network mean baseflow index, network forest and local riparian forest. Note that the vertical axes vary among plots. Predicted water temperature at each of 24 values of the predictor were calculated as the median of predictions obtained by varying all other predictors concurrently across nine evenly space quantiles from 0.1 to 0.9. The grey lines represent the predicted responses for each of the 100 ANNs in the ensemble, and the blue line represents the median prediction from all of the ANNs, which was the final ensemble prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

temporal resolution throughout a large basin or study region are rare in published literature, likely because of limited data availability, modeling difficulty, or study objectives that did not require predicting water temperatures at a high temporal (i.e., daily) resolution. We are aware of a model that was used by Lyons et al. (2009) to predict and summarize daily water temperatures in NHDPlus stream reaches of Wisconsin, but this model is only partially described in conference proceedings (Westenbroek et al., 2010). One exception is a recent model predicting mean daily water temperatures with moderate accuracy (RMSE = 2.8 °C) for global rivers (van Vliet et al., 2012), but comparisons are difficult because predictions were made at a much lower resolution (0.5° grid cells, ~50 km) than our stream reach (~2 km) based predictions. We can more directly compare our model performance to previous studies by summarizing our predictions to calculate mean July water temperature, a commonly used metric. For example, our model had marginally better accuracy for predicting mean July temperatures than models developed for the Upper Midwest (RMSE = 2.0–2.3 °C) using four different statistical methods (Wehrly et al., 2009). Better accuracy of our model may reflect the superior predictive performance of ANNs, as found by Chenard and Caissie (2008) for predicting daily temperatures in a small stream in New Brunswick, Canada. However, such comparisons are limited because accuracy measures such as RMSE should be scaled by observed variation to compare across models and datasets (Moriasi et al., 2007), but we do not know the variation in dataset of Wehrly et al. (2009). For this reason, seemingly superior accuracy of our model compared could be due to a difference in the observed variability in water temperatures between the two study regions, which means that we cannot compare performance in terms of explained variation. Reporting the observed standard deviation in future studies would benefit future model comparisons.

The site-based cross validation procedure was similar to using cross validation based on environmental gradients to select models as suggested by Wenger and Olden (2012), except sites were selected randomly without respect to any environmental gradient. Using this method effectively limited overfitting and increased generalizability, as the final model generalized to validation sites much better (RMSE = 1.82 °C) than a preliminary model developed based on a strictly random cross validation procedure (RMSE = 2.69 °C; J.T. Deweber, unpublished data). We also chose to take the median prediction from an ensemble of 100 ANNs because individual ANNs varied in their predicted effects and we wanted robust estimates of predictor effects (e.g., Figs. 5 and 6). As discussed by Hansen and Salamon (1990), using consensus based off of multiple ANNs is more likely to be correct because any single ANN could become fixed on local optima and make 'wrong decisions'. Although ensembles of neural networks may not be familiar, this approach is similar to more commonly used random forests, where predictions from a large number of classification and regression trees are combined because of the instability of any single tree (Breiman, 2001). Most importantly, our ANN ensemble had good accuracy and limited bias when applied to two validation datasets, which included a large number of new stream reaches with different combinations of landscape and land cover conditions, as well as a relatively warmer year. This suggests that the model can be used to reasonably predict water temperatures at unsampled stream reaches throughout the region under present conditions and future scenarios of climate and land use change. Combining these predictions with additional knowledge of stream systems may help ensure that such predictions reflect actual thermal conditions and changes.

Air temperature was the strongest predictor of river water temperature in our model, which was expected since climatic factors related to air temperature are the primary sources of thermal

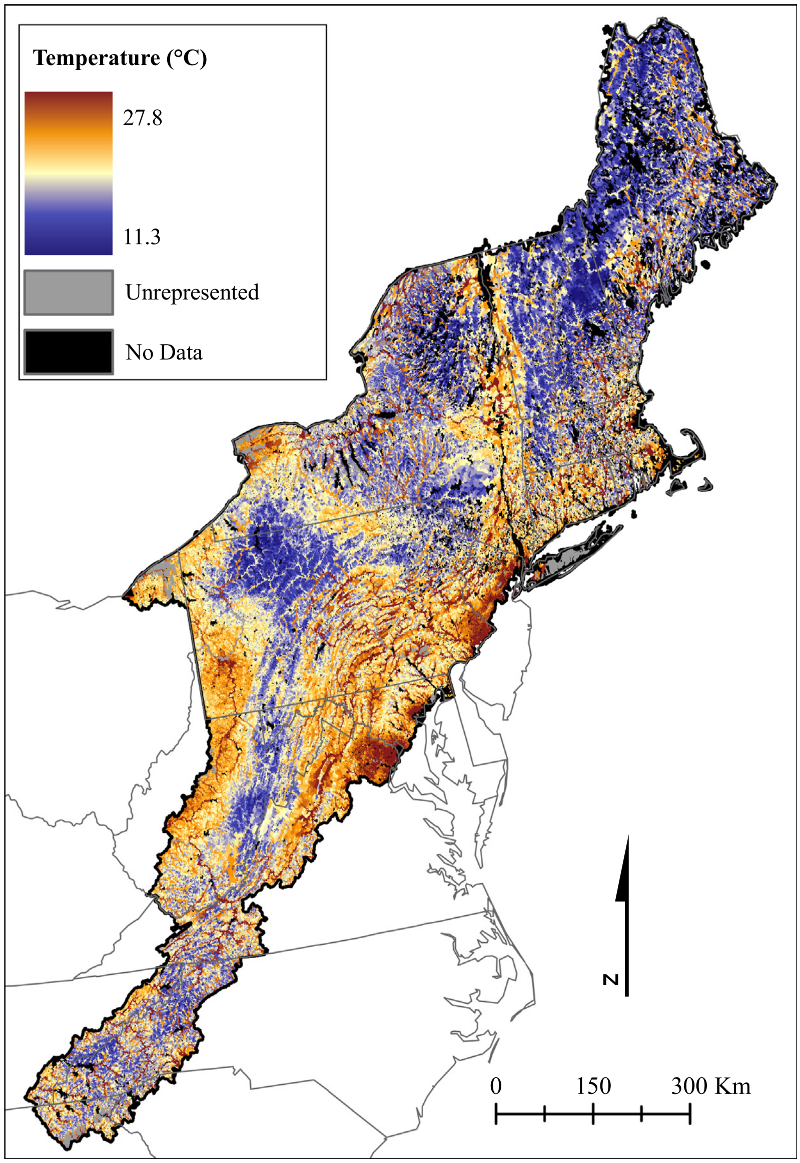

Fig. 7. Map of the regional spatial patterns of mean daily water temperatures (°C) predictions summarized as mean July temperature over the 1980–2010 modeling period. The areas shaded black (No Data) represent stream reaches that have no temperature predictions, while grey areas (Unrepresented) are stream reaches with predicted temperatures that are considered uncertain because one or more landscape attribute value is outside of the range of those in the training dataset.

energy transfers in streams (e.g., Johnson, 2004; Story et al., 2003), and empirical models have often used only air temperature to predict water temperatures (e.g., Caissie et al., 2001; Mohseni et al., 1998). Our results also suggest that including prior air temperature patterns can improve performance in empirical water temperature models. Similarly, Chenard and Caissie (2008) found that one day prior air temperature improved predictions of mean and maximum daily water temperature in a neural network model for a small stream in New Brunswick. Air temperature had a linear effect throughout its range and a 1 °C increase resulted in a ~0.4 °C water temperature increase (Fig. 5). The strength of the relationship did not diminish when air temperature exceeded 20 °C as suggested by Mohseni et al. (1998), but the dampened effect of higher prior 7 day air temperatures (Fig. 6) likely results in an overall diminishing effect.

Although air temperature was important, our results show that landform and land cover attributes can greatly improve predictions. The positive effects of watershed area were expected because river temperature generally increases with river size for a number of reasons, including reduced groundwater cooling, increased exposure to atmospheric exchanges as river width increases, and temperatures reaching equilibrium downstream (Caissie, 2006). Mean aspect in the network had a positive linear effect in our study, which is not surprising because the shift from eastern to western facing catchments likely results in increased solar radiation, which is a primary energy input for rivers (Johnson, 2004; Story et al., 2003). Groundwater interactions are an important determinant of water temperatures (Caissie, 2006; Poole and Berman, 2001) and can improve performance in regional models (e.g., Wehrly et al., 2009; Morrill et al., 2005), but detailed datasets to represent these interactions were lacking in the study region. We included the baseflow index, an interpolated measure of groundwater contribution to baseflow estimated at USGS gaged streams (Wolock, 2003), but the model did not suggest that water temperatures decrease with increasing values as expected during the warm season. Its limited performance in the model likely reflects the limited ability of this metric to account for groundwater interactions. Other studies in mountainous areas have found

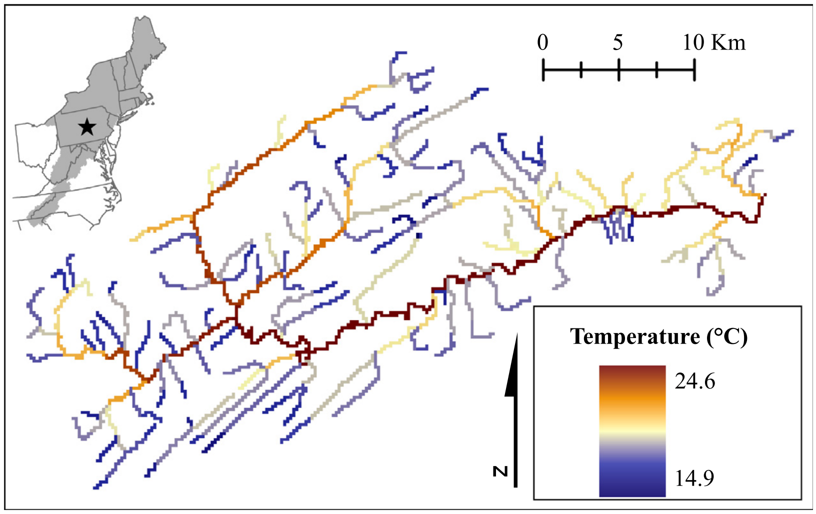

Fig. 8. Map showing detailed spatial patterns of mean daily water temperature (°C) predictions summarized as mean July temperature over the 1980–2010 modeling period for a portion of Penn's Creek watershed in central Pennsylvania. The star in the inset map shows the location of Penn's Creek in the study region.

that elevation can be an effective predictor of water temperature (Isaak et al., 2010; Ruesch et al., 2012) and this would most likely be true in our region as well. However, we did not include elevation because its presence in the model could reduce air temperature effects and downplay the impacts of increasing air temperatures under climate change (Stanton et al., 2012).

We expected positive effects of agricultural and developed land cover in our model and negative effects of forest land cover because these relationships have been well documented in the literature (for reviews see Caissie, 2006; Poole and Berman, 2001). For example, Hill et al. (2013) showed that mean summer river water temperatures in rivers with small amounts of agricultural and urban (medium and high developed) land cover (>1%) were slightly higher (~0.5–1 °C) than in rivers with essentially no agricultural and urban land cover. In contrast to our expectations, the model with agricultural and developed land covers predicted counterintuitive, negative effects and we thus chose the simpler model with only forest land cover. However, the predicted decrease in water temperature as network forest cover increased from 80% to 100% can be interpreted as an increase in water temperature as human land cover increases and was consistent with previously published reports of water temperature increases due to forestry activities (e.g., Beschta and Taylor, 1988). The predicted increases in water temperature as network forest increased from 0% to 60% is not consistent with expected trends and is likely due to relationships with other landscape attributes that control water temperature. For example, stream reaches with relatively little network forest (<40%) had smaller watersheds (mean network area = 37.1 km^2) than reaches with more forest cover (mean network area = 152.2 km^2) in our training dataset, which could lead to a modeled positive relationship because streams with small watershed are predicted to be cooler due to the strong effect of network area. The effect of local riparian forest cover was negative and suggested that the greatest cooling could be expected as riparian forests increased from 0% to 10% and from 90% to 100%. Prior studies have shown warming water temperatures in response to the full and partial removal of riparian forest vegetation (e.g., Rutherford et al., 2004) so it is not surprising that temperatures are predicted to be cooler for stream reaches with 100% riparian forest cover.

Our modeling approach predicted daily water temperature as single events and did not specify spatial or temporal links between predictions. Since river water temperatures are more likely to be similar on consecutive days and in connected river reaches, modeling spatiotemporal autocorrelation could potentially improve model performance. A couple of methods that could be used include recurrent or dynamic ANNs that account for temporal autocorrelation and have been used to successfully forecast streamflow (e.g., Besaw et al., 2010; Chen et al., 2013), and spatial regression models that incorporate downstream connectivity of river systems and have been used to predict weekly water temperatures in relatively large basins (Isaak et al., 2010; Ver Hoef et al., 2004). Although autocorrelation was not directly modeled in this study, some degree of realistic spatial and temporal structure was reflected in model predictions because it was embedded in predictors.

6. Conclusions

We developed an ensemble of ANNs that predict mean daily water temperature with good accuracy (RMSE = ~1.9 °C) and low overall bias (percent bias <±2 °C) for two large validation datasets during the warm season throughout a large and physiographically diverse study region. This is the first publication that we are aware of to describe a model predicting daily water temperatures within individual stream reaches (~2 km in length) throughout a large region, as most previous regional efforts have focused on weekly, monthly or seasonal predictions. Our results demonstrate how combining predictions from an ensemble of ANNs can improve model accuracy and the estimation of predictor effects. Predictor effects as revealed by a sensitivity analysis varied widely among ANNs, and we had more confidence in the relationships between water temperature and climatic, landform and land cover predictors from the median ensemble prediction than from any single ANN. Daily water temperature predictions in individual stream reaches can be used directly or summarized spatially or temporally to yield water temperature metrics for a number of applications, including the management and conservation of aquatic organisms, including mussels, macroinvertebrates and fish species such as brook trout. For example, thermally suitable habitat under present and future conditions can be mapped to determine potential habitat for brook trout or other target species. Although our model had reasonable accuracy and represented most rivers throughout the region, combining local knowledge with model predictions may help ensure that management decisions more accurately reflect actual thermal conditions.

Acknowledgements

We thank the following individuals and agencies for generous contributions of water temperature data: Andy Dollof, Anthony Raburn of the Georgia DNR, Joan Trial of the Maine DMR, John Sweka of the USFWS Northeast Fishery Center, Mark Hudy, the Maryland DNR Monitoring and Non-tidal Assessment Division, Neal Hagstrom of the Connecticut DEEP, the New Hamphsire FGC, Rich Kern of the Vermont FWD, Roy Martin, Steve Means of the Pennsylvania DEP, the Susquehana River Basin Commission, Tamara Smith, and the Wood-Pawcatuck Watershed Association. We would also like to thank J. Olden and three anonymous reviewers for constructive comments on a previous version of this manuscript, as well as all members of the Fish Habitat, Climate, and Land Use Change group for support and ideas leading to this manuscript. Data included in this document were provided by the Maryland Department of Natural Resources Monitoring and Non-tidal Assessment Division. Funding for this research was provided by the U.S. Geological Survey, National Climate Change and Wildlife Science Center. Any use of trade, firm, or product names if for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Almodóvar, A., Nicola, G.G., Ayllón, D., Elvira, B., 2012. Global warming threatens the persistence of Mediterranean brown trout. Glob. Change Biol. 18 (5), 1549–1560.
Antanasijević, D., Pocajt, V., Povrenović, D., Perić-Grujić, A., Ristić, M., 2013. Modelling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study. Environ. Sci. Pollut. Res. 2012, 9006–9013.
Besaw, L.E., Rizzo, D.M., Bierman, P.R., Hackett, W.R., 2010. Advances in ungauged streamflow prediction using artificial neural networks. J. Hydrol. 386 (1–4), 27–37.
Beschta, R.L., Taylor, R.L., 1988. Stream temperature increases and land use in a forested Oregon watershed. Water Resour. Bull. 24 (1), 19–25.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Caissie, D., 2006. The thermal regime of rivers: a review. Freshw. Biol. 51 (8), 1389–1406.
Caissie, D., El-Jabi, N., Satish, M.G., 2001. Modelling of maximum daily water temperatures in a small stream using air temperatures. J. Hydrol. 251 (1–2), 14–28.
Chang, F.-J., Tsai, W.-P., Chen, H.-K., Yam, R.S.-W., Herricks, E.E., 2013. A self-organizing radial basis network for estimating riverine fish diversity. J. Hydrol. 476, 280–289.
Chen, P.-A., Chang, L.-C., Chang, F.-J., 2013. Reinforced recurrent neural networks for multi-step-ahead flood forecasts. J. Hydrol. 497, 71–79.
Chenard, J.F., Caissie, D., 2008. Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. Hydrol. Process. 22 (17), 3361–3372.
Cheng, B., Titterington, D.M., 1994. Neural networks: a review from a statistical perspective. Stat. Sci. 9 (1), 2–30.
Clark, M.E., Rose, K.A., Levine, D.A., Hargrove, W.W., 2001. Predicting climate change effects on Appalachian trout: combining GIS and individual-based modeling. Ecol. Appl. 11 (1), 161–178.
Demars, B.O.L., Manson, J.R., Olafsson, J.S., Gislason, G.M., Gudmundsdottir, R., Woodward, G., Reiss, J., Pichler, D.E., Rasmussen, J.J., Friberg, N., 2011. Temperature and the metabolic balance of streams. Freshw. Biol. 56 (6), 1106–1121.
Domisch, S., Jähnig, S.C., Haase, P., 2011. Climate-change winners and losers: stream macroinvertebrates of a submontane region in Central Europe. Freshw. Biol. 56 (10), 2009–2020.
Dreyfus, G., 2005. Neural Networks: Methodology and Applications. Springer, New York.
Esselman, P.C. et al., 2011. An index of cumulative disturbance to river fish habitats of the conterminous united states from landscape anthropogenic activities. Ecol. Restor. 29 (1/2), 133–151.
Ficke, A.D., Myrick, C.A., Hansen, L.J., 2007. Potential impacts of global climate change on freshwater fisheries. Rev. Fish Biol. Fish. 17 (4), 581–613.
Flebbe, P.A., Roghair, L.D., Bruggink, J.L., 2006. Spatial modeling to project southern Appalachian trout distribution in a warmer climate. Trans. Am. Fish. Soc. 135 (5), 1371–1382.
Gardner, B., Sullivan, P.J., Lembo, A.J., 2003. Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. Can. J. Fish. Aquat. Sci. 60 (3), 344–351.
Hansen, L.K., Salamon, P., 1990. Neural network ensembles. IEEE Trans. Pattern Anal. Mach. Intell. 12 (10), 993–1001.
Hill, R.A., Hawkins, C.P., Carlisle, D.M., 2013. Predicting thermal reference conditions for USA streams and rivers. Freshw. Sci. 32 (1), 39–55.
Homer, C., Huang, C.Q., Yang, L.M., Wylie, B., Coan, M., 2004. Development of a 2001 national land-cover database for the United States. Photogramm. Eng. Remote Sens. 70 (7), 829–840.
Hudy, M., Thieling, T.M., Gillespie, N., Smith, E.P., 2008. Distribution, status, and land use characteristics of subwatersheds within the native range of brook trout in the Eastern United States. N. Am. J. Fish. Manage. 28 (4), 1069–1085.
Huo, Z. et al., 2012. Integrated neural networks for monthly river flow estimation in arid inland basin of Northwest China. J. Hydrol. 420–421, 159–170.
Isaak, D.J., 2011. Stream Temperature Monitoring and Modeling: Recent Advances and New Tools for Managers. Stream Notes. Stream Systems Technology Center, July 2011. 7p.
Isaak, D.J., Hubert, W.A., 2001. A hypothesis about factors that affect maximum summer stream temperatures across montane landscapes. J. Am. Water Resour. Assoc. 37 (2), 351–366.
Isaak, D.J., Luce, C.H., Rieman, B.E., Nagel, D.E., Peterson, E.E., Horan, D.L., Parkes, S., Chandler, G.L., 2010. Effects of climate change and wildfire on stream temperatures and salmonid thermal habitat in a mountain river network. Ecol. Appl. 20 (5), 1350–1371.
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.
Jones, K.L., Poole, G.C., Meyer, J.L., Bumback, W., Kramer, E.A., 2006. Quantifying expected ecological response to natural resource legislation: a case study of riparian buffers, aquatic habitat, and trout populations. Ecol. Soc. 11 (2), 15.
Krogh, A., Hertz, J.A., 1992. A simple weight decay can improve generalization. In: Moody, J.E., Hanson, S.J., Lippmann, R.P. (Eds.), Advances in Neural Information Processing Systems 4. Morgan Kaufmann Publishers, Inc., San Mateo, California, pp. 950–957.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28 (5), 1–26.
Lawrence, D.J., Olden, J.D., Torgersen, C.E., 2012. Spatiotemporal patterns and habitat associations of smallmouth bass (*Micropterus dolomieu*) invading salmon-rearing habitat. Freshw. Biol. 57, 1929–1946.
Lehmkuhl, D.M., 1972. Change in thermal regime as a cause of reduction of benthic fauna downstream of a reservoir. J. Fish. Res. Board Can. 29 (9), 1329–1332.
Lek, S., Guégan, J.F., 1999. Artificial neural networks as a tool in ecological modelling, an introduction. Ecol. Modell. 120 (2–3), 65–73.
Lessard, J.L., Hayes, D.B., 2003. Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. River Res. Appl. 19 (7), 721–732.
Lowney, C.L., 2000. Stream temperature variation in regulated rivers: evidence for a spatial pattern in daily minimum and maximum magnitudes. Water Resour. Res. 36 (10), 2947–2955.
Lyons, J. et al., 2009. Defining and characterizing coolwater streams and their fish assemblages in Michigan and Wisconsin, USA. N. Am. J. Fish. Manage. 29 (4), 1130–1151.
MacCrimmon, H.R., Campbell, J.S., 1969. World distribution of brook trout, *Salvelinus fontinalis*. J. Fish. Res. Board Can. 26 (7), 1699–1725.
Magnuson, J.J., Crowder, L.B., Medvick, P.A., 1979. Temperature as an ecological resource. Am. Zool. 19 (1), 331–343.
Marcé, R., Armengol, J., 2008. Modelling river water temperature using deterministic, empirical, and hybrid formulations in a Mediterranean stream. Hydrol. Process. 22 (17), 3418–3430.
McKenna, J.E., Butryn, R.S., McDonald, R.P., 2010. Summer stream water temperature models for Great Lakes streams: New York. Trans. Am. Fish. Soc. 139 (5), 1399–1414.
Meisner, J.D., 1990. Effect of climatic warming on the southern margins of the native range of brook trout, *Salvelinus fontinalis*. Can. J. Fish. Aquat. Sci. 47 (6), 1065–1070.
Mohseni, O., Stefan, H.G., Erickson, T.R., 1998. A nonlinear regression model for weekly stream temperatures. Water Resour. Res. 34 (10), 2685–2692.
Mohseni, O., Erickson, T.R., Stefan, H.G., 1999. Sensitivity of stream temperatures in the United States to air temperatures projected under a global warming scenario. Water Resour. Res. 35 (12), 3723–3733.
Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., Veith, T., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.
Morrill, J.C., Bales, R.C., Conklin, M.H., 2005. Estimating stream temperature from air temperature: implications for future water quality. J. Environ. Eng. – Asce 131 (1), 139–146.
Nelson, K.C., Palmer, M.A., 2007. Stream temperature surges under urbanization and climate change: data, models, and responses. J. Am. Water Resour. Assoc. 43 (2), 440–452.
Olden, J.D., Jackson, D.A., 2002. A comparison of statistical approaches for modelling fish species distributions. Freshw. Biol. 47 (10), 1976–1995.
Olden, J.D., Naiman, R.J., 2010. Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. Freshw. Biol. 55 (1), 86–107.
Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Modell. 178 (3–4), 389–397.
Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. 83 (2), 171.

Poole, G.C., Berman, C.H., 2001. An ecological perspective on in-stream temperature: natural heat dynamics and mechanisms of human-caused thermal degradation. Environ. Manage. 27 (6), 787–802.

Rahel, F.J., Olden, J.D., 2008. Assessing the effects of climate change on aquatic invasive species. Conserv. Biol. 22 (3), 521–533.

Ries, R.D., Perry, S.A., 1995. Potential effects of global climate warming on brook trout growth and prey consumption in central Appalachian streams, USA. Clim. Res. 5 (3), 197–206.

Risley, J.C., Roehl Jr., E.A., Conrads, P.A., 2003. Estimating Water Temperatures in Small Streams in Western Oregon using Neural Network Models. USGS Water-Resources Investigation Report 02-4218. <http://pubs.er.usgs.gov/publication/wri024218>.

Rocchini, D., Joaquìn, H., Lengyel, S., Lobo, J.M., Jimènez-Valverde, A., Ricotta, C., Bacaro, G., Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Prog. Phys. Geogr. 35 (2), 211–226.

Ruesch, A.S., Torgersen, C.E., Lawler, J.J., Olden, J.D., Peterson, E.E., Volk, C.J., Lawrence, D.J., 2012. Projected climate-induced habitat loss for salmonids in the John Day River Network, Oregon, USA. Conserv. Biol. 26 (5), 873–882.

Rutherford, J.C., Marsh, N.A., Davies, P.M., Bunn, S.E., 2004. Effects of patchy shade on stream water temperature: how quickly do small streams heat and cool? Mar. Freshw. Res. 55 (8), 737–748.

Stanton, J.C., Pearson, R.G., Horning, N., Ersts, P., Reşit Akçakaya, H., 2012. Combining static and dynamic variables in species distribution models under climate change. Methods Ecol. Evol. 3 (2), 349–357.

Story, A., Moore, R.D., Macdonald, J.S., 2003. Stream temperatures in two shaded reaches below cutblocks and logging roads: downstream cooling linked to subsurface hydrology. Can. J. For. Res. 33 (8), 1383–1396.

Sugimoto, S., Nakamura, F., Ito, A., 1997. Heat budget and statistical analysis of the relationship between stream temperature and riparian forest in the Toikanbetsu River basin, northern Japan. J. For. Res. 2 (2), 103–107.

USACE, 2005. National Inventory of Dams. <http://crunch.tec.army.mil/nid/webpages/nid.cfm> (accessed May 2011).

USEPA and USGS, 2005. National Hydrography Dataset Plus Version 1.0, NHDPlusV1. <www.horizon-systems.com/nhdplus/>.

van Vliet, M.T.H., Yearsley, J.R., Franssen, W.H.P., Ludwig, F., Haddeland, I., Lattenmaier, D.P., Kabat, P., 2012. Coupled daily streamflow and water temperature modelling in large river basins. Hydrol. Earth Syst. Sci. 16 (11), 4303–4321.

van Vliet, M.T.H., Franssen, W.H.P., Yearsley, J.R., Fulco, L., Haddeland, I., Lattenmaier, D.P., Kabat, P., 2013. Global river discharge and water temperature under climate change. Global Environ. Change 23 (2), 450–464.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S, fourth ed. Springer, New York.

Ver Hoef, J.M., Peterson, E., Theobald, D., 2004. Spatial statistical models that use flow and stream distance. Environ. Ecol. Stat. 13, 449–464.

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

Wehrly, K.E., Brenden, T.O., Wang, L.Z., 2009. A comparison of statistical approaches for predicting stream temperatures across heterogeneous landscapes. J. Am. Water Resour. Assoc. 45 (4), 986–997.

Wen, X., Fang, J., Diao, M., Zhang, C., 2013. Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China. Environ. Mon. Assess. 185 (5), 4361–4371.

Wenger, S.J., Olden, J.D., 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods Ecol. Evol. 3 (2), 260–267.

Westenbroek, S., Stewart, J.S., Buchwald, C.A., Mitro, M., Lyons, J.D., Greb, S., 2010. A Model for Evaluating Stream Temperature Response to Climate Change Scenarios in Wisconsin. In: Potter, K.W., Frevert, D.K. (Eds.), Watershed Management: Innovations in watershed management under land use and climate change. American Society of Civile Engineers, Reston, Virginia, pp. 1–12. http://dx.doi.org/10.1061/41143(394)1.

Wolock, D.M., 2003. Base-Flow Index Grid for the Conterminous United States. USGS Open-File Report 03-263, <http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml> (accessed May 2010).

Wright, S.A., Anderson, C.R., Voichick, N., 2009. A simplified water temperature model for the Colorado River below Glen Canyon Dam. River Res. Appl. 25 (6), 675–686.

Xenopoulos, M.A., Lodge, D.M., Alcamo, J., Marker, M., Schulze, K., Van Vuuren, D.P., 2005. Scenarios of freshwater fish extinctions from climate change and water withdrawal. Global Change Biol. 11 (10), 1557–1564.