BioOne COMPLETE

Predicting thermal reference conditions for USA streams and rivers

Authors: Hill, Ryan A., and Hawkins, Charles P.

Predicting thermal reference conditions for USA streams and rivers

Ryan A. Hill[1] AND Charles P. Hawkins[2]

Department of Watershed Sciences, Western Center for Monitoring and Assessment of Freshwater Ecosystems, and the Ecology Center, Utah State University, Logan, Utah 84322-5210, USA

Daren M. Carlisle[3]

National Water-Quality Assessment Program, US Geological Survey, 12201 Sunrise Valley Drive, Mississippi 413, Reston, Virginia 20192, USA

Abstract. Temperature is a primary driver of the structure and function of stream ecosystems. However, the lack of stream temperature (ST) data for the vast majority of streams and rivers severely compromises our ability to describe patterns of thermal variation among streams, test hypotheses regarding the effects of temperature on macroecological patterns, and assess the effects of altered STs on ecological resources. Our goal was to develop empirical models that could: 1) quantify the effects of stream and watershed alteration (SWA) on STs, and 2) accurately and precisely predict natural (i.e., reference condition) STs in conterminous USA streams and rivers. We modeled 3 ecologically important elements of the thermal regime: mean summer, mean winter, and mean annual ST. To build reference-condition models (RCMs), we used daily mean ST data obtained from several thousand US Geological Survey temperature sites distributed across the conterminous USA and iteratively modeled ST with Random Forests to identify sites in reference condition. We first created a set of dirty models (DMs) that related STs to both natural factors (e.g., climate, watershed area, topography) and measures of SWA, i.e., reservoirs, urbanization, and agriculture. The 3 models performed well ($r^2 = 0.84$–0.94, residual mean square error [RMSE] = 1.2–$2.0°C$). For each DM, we used partial dependence plots to identify SWA thresholds below which response in ST was minimal. We then used data from only the sites with upstream SWA below these thresholds to build RCMs with only natural factors as predictors ($r^2 = 0.87$–0.95, RMSE = 1.1–$1.9°C$). Use of only reference-quality sites caused RCMs to suffer modest loss of predictor space and spatial coverage, but this loss was associated with parts of ST response curves that were flat and, therefore, not responsive to further variation in predictor space. We then compared predictions made with the RCMs to predictions made with the DMs with SWA set to 0. For most DMs, setting SWAs to 0 resulted in biased estimates of thermal reference condition.

Key words: agriculture, dams, land use, modeling, prediction, Random Forests, reference condition, reservoirs, rivers, streams, temperature, urbanization.

Quantifying the thermal regime may be key to understanding the structure and function of all ecosystems (Brown et al. 2004). In lotic ecosystems, spatial and temporal variation in stream temperatures (STs) (see Table 1 for definitions of acronyms used in this paper) affects the distributions of individual species (Vannote and Sweeney 1980) and, hence, geographic variation in entire communities (Hawkins et al. 1997). Life-history patterns, individual growth and production, and ecosystem metabolism are also temperature dependent (Benke et al. 1988, Acuña et al. 2008). As a consequence, any natural or human-induced change in thermal regime probably will affect stream ecosystem structure and function.

Because of their ecological importance, STs are extensively monitored by local, state, and federal agencies (Haag and Luce 2008), and millions of dollars are spent annually in thermal remediation efforts (Wu et al. 2003, Seedang et al. 2008). However, determining whether the thermal condition of a stream has been altered requires that we compare observed STs to those expected under natural conditions (Hawkins et al. 2010). To make such assessments in the absence of historical data, reference-condition ST (RCST) must be

[1] E-mail addresses: ryan.hill@usu.edu

[2] chuck.hawkins@usu.edu

[3] dcarlisle@usgs.gov

TABLE 1. Definition of acronyms used in this paper.

Acronym	Definition
BFI	Base-flow index
CFD	Cumulative frequency distribution
DM	Dirty model
E	Expected
LOWESS	Locally weighted regression and smoothing scatterplots
MAST	Mean annual stream temperature
MSE	Mean squared error
MSST	Mean summer stream temperature
MWST	Mean winter stream temperature
NID	National Inventory of Dams
NLCD	National Land Cover Dataset
NSE	Nash–Sutcliffe coefficient of model efficiency
O	Observed
PBIAS	% bias
PDP	Partial dependence plot
RCM	Reference-condition model
RCST	Reference-condition stream temperature
RF	Random Forest
RMSE	Root mean squared error
RMSE/SD	Model RMSE/standard deviation of observed stream temperatures
ST	Stream temperature
SWA	Stream and watershed alteration
USGS	US Geological Survey

predicted. Useful RCST predictive models should account for the effects of naturally occurring stream and watershed features on water temperatures. Alternatively, if reference-condition streams are rare or unavailable, predictive models must account for the effects of human-caused stream or watershed alteration (SWA) on STs in a way that natural STs can be inferred.

The natural and anthropogenic factors that can affect STs are well known and vary spatially and temporally within and among watersheds (Ward 1985, Poole and Berman 2001, Allan 2004, Caissie 2006, Webb et al. 2008). Incoming solar radiation and its attenuation by streamside shading, incoming and outgoing long-wave radiation, evaporative cooling, and the stream surface area available on which these heat-exchange processes occur all play critical roles in determining STs. Other important factors include spatial variation in groundwater inputs and local climatic conditions, such as air temperature and precipitation. Human activities that affect STs include removal of streamside vegetation (Brown 1970, Bartholow 2000, Hagen et al. 2006, McTammany et al. 2007), dam operations, such as hypolimnetic vs epilimnetic release (Sinokrot et al. 1995, Preece and Jones 2002, Lessard and Hayes 2003, Olden and Naiman 2010, Risley et al. 2010), power generation and release of wastewater effluent (Stefan and Chau 1976, Kinouchi et al. 2007), runoff from urbanized areas (Klein 1979, Kinouchi et al. 2007, Nelson and Palmer 2007, Kaushal et al. 2010), and agricultural irrigation extraction and return flows.

A variety of models have been developed to predict STs. Most published ST models can be classified as single-site physical, single-site empirical, or multisite empirical models (see Hawkins et al. 2010). Both single-site physical and empirical models have limitations for use in regional ST assessments because they are parameterized for individual stream reaches or watersheds, and therefore, predictions at new, unmeasured locations probably would be inaccurate. In addition, application of single-site physical models to assess many streams in a large region would be cost and time prohibitive because they require measurement and parameterization of heat-exchange processes at each reach (Edinger et al. 1968, Brown 1969, Theurer et al. 1984, Morin et al. 1987, Caissie et al. 2007). Single-site empirical models require long-term time-series measurements of stream and air temperatures that are related through regression (Cluis 1972, Mohseni et al. 1998, van Vliet et al. 2011, Kelleher et al. 2012) or other empirical techniques (Chenard and Caissie 2008), and such data are available for few streams.

Multisite, empirical models hold the best potential for use in regional assessments. These models can make predictions at unmeasured locations (Hawkins et al. 2010), are often based on easily obtained geographical information system (GIS) predictors, and do not require long ST records. These models relate STs observed at multiple sites to local stream and watershed attributes, such as air temperature, watershed area, channel slope, elevation, and latitude (Miyake and Takeuchi 1951, Vannote and Sweeney 1980, Donato 2002, Risley et al. 2003, Jones et al. 2006, Wehrly et al. 2006, Isaak et al. 2010, McKenna et al. 2010). Such models should be able to predict RCSTs at new locations if they are developed with data from reference-condition sites. These models often use predictor variables, such as elevation and latitude, that are known to be correlated with ST but are not necessarily causative. These models typically have been focused on summer STs (Wehrly et al. 2009). However, Allan and Castillo (2007) noted that streams with similar summer STs can have different overall thermal regimes resulting from differences in winter STs, which could have substantial ecological effects (Haidekker and Hering 2008), and suggested characterizing the thermal regime to capture these differences.

When predicting RCST, models ideally would be based on data collected at sites in thermal reference

2013]

condition. However, the number of reference-quality sites present in a region may be limited, and these sites may not represent the full range of naturally occurring environments that need to be assessed. This issue is especially problematic in regions with substantial SWA (Kilgour and Stanfield 2006). However, if the effects of SWA can be accounted for in models (Soranno et al. 2011), it is theoretically possible to predict RCST by setting SWA to 0 (e.g., Baker et al. 2005). Such an approach would maximize the range of natural conditions (environmental space) to which models apply and should result in more robust models than those derived from data collected only at reference-quality sites. However, we do not yet know if such models adequately account for the effects of SWAs and, thus, produce unbiased estimates of RCST. Our general goal was to develop spatially explicit empirical models to predict reference-condition mean summer, mean winter, and mean annual STs (MSST, MWST, and MAST, respectively) at unmeasured locations across the conterminous USA. Our specific objectives were to: 1) develop models that included both natural factors and measures of SWA as predictor variables (henceforth dirty models [DM] because they contain the full range of SWA values), 2) use these initial DMs to identify stream reaches in thermal reference condition, 3) build reference-condition models (RCMs) with data from just those streams in thermal reference condition, and 4) compare general performance of both DMs and RCMs and determine if DMs provided similar estimates of RCST as RCMs when SWAs were set to 0 in the DMs.

Methods

Overview of RCM development

We used an iterative process to identify US Geological Survey (USGS) temperature sites in reference condition to develop models of RCST. We used an extensive database of STs to first build DMs that empirically related estimates of MSST, MWST, and MAST to spatial variation in natural factors and SWA. We then examined the relationship between STs and each of the SWAs to identify thresholds in SWA below which STs showed little or no association with SWAs. We used these thresholds to identify sites in thermal reference condition. Next, we built RCMs with data from just those sites identified as being in thermal reference condition. Last, to examine whether RCSTs can be predicted with DMs, we compared predictions made by setting SWA to 0 in DMs and predictions from RCMs with known RCSTs.

ST data

The USGS provided daily mean ST measurements for 3714 sites distributed across the conterminous USA (Fig. 1). A long period of record was available for some sites (e.g., 30 y), but we chose to analyze data from a 10-y period that spanned 1999 to 2008 to match years for which we had reliable landuse information (agriculture and urbanization). Daily records were often not continuous within or across the years of record at all sites, but this 10-y analysis window contained 2,766,369 daily records. We screened for and removed outliers from the data by visually examining plots of daily mean STs vs year, month, and calendar day for each USGS site to identify observations that were the result of instrument malfunctions, did not fit typical seasonal patterns of STs in the conterminous USA, or had values outside those generally expected within the conterminous USA ($-0.1°C \leq ST \leq 35°C$). We retained winter ST values as low as $-0.1°C$ because streams can become super-cooled to this temperature when air temperatures are $<0°C$ for several days (Martin 1981), and this value is within the reported range of accuracy of USGS temperature measurements (Wilde 2006). After quality-control screening, we excluded 98 sites from further analyses. We used the retained data to calculate MSST (July and August), MWST (January and February), and MAST for each site–year combination. We required that a monthly record used in analyses have recorded temperatures for $\geq \frac{2}{3}$ of its days. After these data manipulations, each USGS site had from 1 to 10 y of site–year observations. We randomly selected 1 site–year observation from each site for modeling (Table 2). For the 10-y analysis window, we identified 2136 MSST, 1580 MWST, and 996 MAST observations for modeling (Fig. 1).

Natural predictor variables

We used the Multi-Watershed Delineation Tool (Chinnayakanahalli et al. 2006) to delineate the upstream watershed boundaries for each site from 30-m USGS digital elevation models. For each predictor, we calculated the mean values within a watershed, the mean values within a 100-m-wide riparian buffer within the watershed, and the point-level measurement at the site (Appendix S1; available online from: http://dx.doi.org/10.1899/12-009.1.s1). The natural predictors included incoming solar radiation (Kumar et al. 1997), streamside vegetation height and density (Rollins and Frame 2006), Parameter-elevation Regressions on Independent Slopes Model (PRISM) air temperature and precipitation (Daly et al. 2008), dominant surficial geology type and
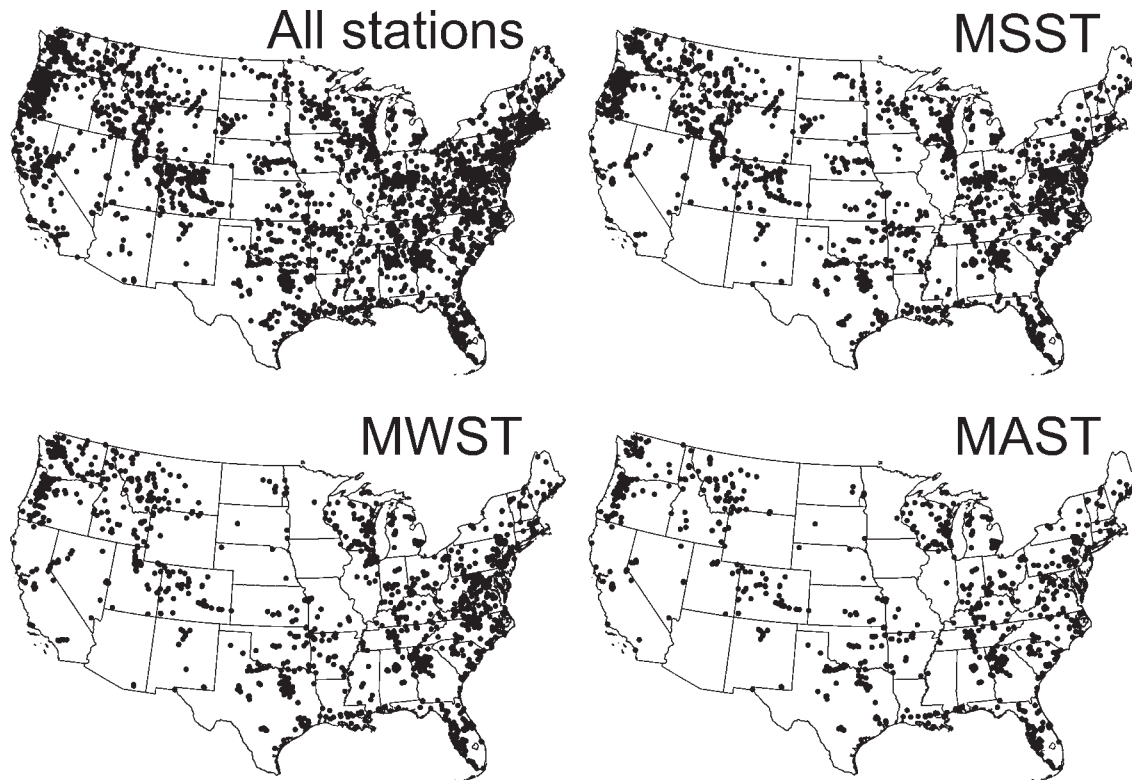
FIG. 1. Distribution of US Geological Survey sites with temperature data in the conterminous USA, and sites for which mean summer (MSST), winter (MWST), and annual (MAST) stream temperatures were calculated.

% watershed in each geology type (Reed and Bush 2001), soil characteristics, such as permeability, water-table depth, and bulk density (Wolock 1997), watershed shape and area, elevation range, channel slope, runoff (McCabe and Wolock 2010), base-flow index (BFI) (Wolock 2003), a stream flow-stability index (Appendix S1), the enhanced vegetation index (Huete et al. 2002), and the % area of each watershed in lake and wetland land cover (Homer et al. 2007) (see Appendix S1 for details). We based the selection of these potential predictors on an extensive literature review of the physical processes and stream and watershed characteristics previously shown to be important in either empirical or deterministic models. Solar radiation was computationally intensive to estimate for each watershed, so we tested the predictive value of this factor in a preliminary analysis of data obtained from 22 states west of the Mississippi River before developing models for the entire conterminous USA. Including solar radiation estimates failed to improve the western USA models, so we excluded solar radiation as a potential predictor for the conterminous USA models (see *Excluded Predictors* in Discussion). We did all spatial analyses with ArcGIS 9.3.1 Spatial Analyst (Environmental Systems Research Institute, Redlands, California). We also used the method published by Isaak et al. (2010) and applied inverse-distance weighting schemes to watershed and riparian-buffer averages for several predictors to place greater emphasis on values of the predictor that were spatially closer to each ST site. We used the weighting,

$$w_i = e^{-\left(\frac{D_f}{D_e}\right)}, \qquad [1]$$

TABLE 2. Summary statistics for mean summer (MSST), winter (MWST) and annual (MAST) stream temperature data.

Model	Sites	Minimum °C	Maximum °C	Mean °C
MSST	2136	4.5	33.7	21.3
MWST	1580	−0.1	23.4	5.6
MAST	996	3.2	26	13.8

where D_f represents the flow distance from any upstream pixel to the ST site and D_e represents an e-folding distance, i.e., the distance over which the weight decreases exponentially. We averaged the inversely weighted upstream pixels within the watershed or riparian buffer.

Indices of SWA

Reservoirs.—Release of water impounded by large, hypolimnetic-release dams results in cooler summer and warmer winter STs than in unregulated streams (Ward 1963, 1985). We used the georeferenced National Inventory of Dams (NID) (USACE 2009) to quantify the presence and size of dams and associated reservoirs in each watershed. The NID provides dam attributes, such as year of construction, structural height, and volume of each reservoir. However, examination of the NID revealed errors in the geographic locations of many dams. Important attributes, such as the year of completion and dam height, were incomplete for many records. In addition, some critical features, such as reservoir volume, were repeated in the database if a reservoir had multiple dikes or locks. Therefore, we screened 53,041 NID records to ensure they represented unique dam structures and had complete and accurate records of year of completion and reservoir volume (Appendix S2; available online from: http://dx.doi.org/10.1899/12-009.1.s2).

Dam height may be a better indicator of hypo- vs epilimnetic release, but we had to characterize reservoirs within each watershed by the total, mean, and maximum volumes of water they impounded. We used reservoir volume because numerous NID records lacked dam height information and, therefore, could not be used to model STs. For each dam in each watershed, we applied the exponentially decaying inverse-distance weighting with D_e = 50, 100, 150, and 200 km to account for the downstream attenuation of reservoir effects in our models. These distances were based on literature values (Preece and Jones 2002) and our own examination of sites below large reservoirs in which we found that thermal effects of reservoirs decreased exponentially with distance downstream and sometimes extended to ~75 to 150 km. In addition, we normalized these values by the watershed areas above each temperature site. We did these calculations only if a dam was constructed before the year temperatures were recorded at a site, e.g., a dam completed in 2005 was not counted for a ST recorded in 2000.

Agriculture and urbanization.—We estimated the total and percentage of each watershed in agricultural (row crop) and urban land uses (medium and high intensities) from the 2001 (version 2.0) and 2006 National Land Cover Dataset (NLCD) (Homer et al. 2007; http://www.mrlc.gov/). We matched ST data from 1999 to 2003 and 2004 to 2008 with the 2001 and 2006 NLCD layers, respectively, to ensure the estimated SWA was within 2 y of their respective temperature measurements. We also estimated the total area of riparian buffers composed of agricultural and urban land uses with the area of each land use pixel inversely weighted with D_e = 1, 4, 15, and 25 km above the ST sites. We normalized riparian estimates of each SWA by upstream watershed area.

Modeling approach

Random Forests.—We used Random Forest modeling (RF) (Breiman 2001) to empirically model STs. RF is a nonparametric, nonlinear modeling technique based on the well known classification and regression tree algorithm. However, an RF model is produced by building hundreds of regression trees from randomized subsets of the data, and predictions to new sites are simply the average of the predictions made by all trees in the resulting forest (see Cutler et al. 2007). We used the *randomForest* (Liaw and Wiener 2002) function in the R statistical software package (version 2.15.1; R Development Core Team, Vienna, Austria) to fit our models.

RF has been increasingly used in diverse natural-science applications, including meteorology (Holden et al. 2011), hydrology (Ordoyne and Friedl 2008), geomorphology (Francke et al. 2008, Snelder et al. 2011), ecology (Cutler et al. 2007, Peters et al. 2007, Chinnayakanahalli et al. 2011), and water-quality monitoring (Carlisle et al. 2009, 2010, Catherine et al. 2010). RF has generally superior predictive performance when compared with other modeling techniques (Prasad et al. 2006, Banfield et al. 2007, Cutler et al. 2007, Peters et al. 2007), and the RF algorithm is easy to understand conceptually (Cutler et al. 2007). RF models make no assumptions about normality of data and are resistant to over-fitting and multi-collinearity of predictor variables (Breiman 2001). In addition, spatial and temporal autocorrelations in the data do not affect RF predictions to new samples (Karpievitch et al. 2009). RF produces validation statistics by calculating the mean squared error (MSE) and pseudo-R^2 from the randomized subsets of data that are withheld (out-of-bag samples) during model development.

Variable selection.—We sought to produce RF models that were both interpretable and parsimonious in terms of the number of predictor variables used.

However, little guidance exists for variable selection with RF (Genuer et al. 2010). Therefore, we selected predictors that maximized the physical interpretability of the model, reduced redundancy among predictor variables, and maximized model performance. We developed the RF models by iteratively adding predictors that produced the greatest improvement in the RF performance metrics, were physically interpretable, and had low correlation with other predictors. We stopped the selection processes when additional predictors failed to decrease the square root of the MSE by ~0.1°C or were redundant with predictors already in the model.

Model performances.—We compared observed STs with their out-of-bag predictions to calculate several model-performance metrics (Moriasi et al. 2007): the Nash–Sutcliffe coefficient of model efficiency (NSE), % bias (PBIAS), and root mean squared error (RMSE) normalized by the observed standard deviation (RMSE/SD). NSE measures the total residual error relative to the total variance within the data. Models that perform well and have little bias have NSE values that are similar to the squared correlation coefficient (r^2), but NSE is more sensitive to deviation from the 1:1 line. We report both NSE and r^2. PBIAS estimates the tendency of a model to over predict (PBIAS $<$ 0) or under predict (PBIAS $>$ 0). RMSE measures the absolute error associated with each model and is in the units for which predictions are made (°C), whereas RMSE/SD allows comparison between models. Smaller values of RMSE and RMSE/SD indicate better model performance. In addition, we plotted observed vs predicted STs and visually examined the plots for outliers and biases.

Reference-site identification

To identify reference-quality sites, we used partial dependence plots (PDPs) (Hastie et al. 2001) to examine associations between ST and measures of SWA. A PDP is a plot of the average of the response variable (ST) vs a predictor variable and accounts for the effects of other predictor variables within the model (Hastie et al. 2001). We visually selected thresholds for each SWA below which the response in ST was minimized, while maximizing the number of sites retained for modeling.

Two important considerations are the range of natural conditions within which each model can be applied and whether environmental space was lost through reference-site selection. To compare the predictor space associated with the RCMs and DMs, we plotted the cumulative frequency distribution (CFD) of each natural predictor used in each model.

In addition, we overlaid these plots onto the CFDs of each predictor for all USGS sites with available ST data. Although probably not representative of all environments within the conterminous USA, the CFD plots of each predictor at all USGS ST sites encompass a large range of conditions. Thus, they allow comparison between the predictor space of each model and the predictor space of all ST sites in the conterminous USA. When we observed a difference between the RCM and DM in a predictor's CFD, we noted the point beyond which the reference-condition and dirty predictors did not overlap. We then examined the response of ST in the PDP beyond that point to determine how the RCMs might be affected by the lost predictor space. In addition, we compared maps of reference and nonreference site locations to identify regions where reference-site selection resulted in geographic underrepresentation.

RCMs vs DMs

We examined whether the DMs could be used to predict RCSTs by comparing SWA-zeroed predictions with RCM predictions and observed RCSTs. To make the SWA-zeroed predictions, we used a leave-one-out procedure that removed 1 site from the data, developed a DM on remaining sites, and predicted reference-condition ST at the withheld site by setting its SWA to 0. This procedure was repeated for each site across the full range of SWAs, i.e., true reference to the highest levels of alteration. The out-of-bag predictions can be obtained directly from the RF models, but we used the leave-one-out procedure in the RCMs to ensure comparability of predictions made with the DMs and RCMs. At nonreference sites, we simply applied the RCMs because these sites were not used in model development.

Environmental and ecological assessments are often conducted by comparing observed (O) conditions to those expected (E) in the absence of human alteration, computed as the deviation of E from O (e.g., O − E). For an assessment to be effective, O − E should be near 0 when sites are in reference condition and should depart measurably from 0 at thermally altered sites. We first compared RCM and SWA-zeroed DM predictions made at reference-condition sites to assess whether biases were present in RCMs or DMs when predicting to sites of known thermal condition. To estimate biases in predictions, we calculated the mean O − E at reference condition sites for both RCMs and SWA-zeroed DMs. We also quantified the precision of predictions as the standard deviation of O − E values at known reference sites. To assess if the relationship between O − E and SWA depended on whether RCMs

or SWA-zeroed DMs were used to predict E, we isolated the effects of each SWA by selecting sites that failed the reference screening for the particular SWA of interest, but passed the reference screening for the other SWAs (e.g., failed agriculture but passed the dam and urbanization screens). We then plotted O – E values against the full range of each SWA and fit locally weighted regression and smoothing scatter-plots (LOWESS) lines to the data (Cleveland 1979). We plotted a vertical line at the point for each SWA that we had previously defined as the boundary between reference and nonreference conditions. For streams to the left of the boundary, i.e., streams in reference condition, LOWESS lines should be near O – E = 0. As SWA increases, the LOWESS lines should deviate from O – E = 0. A LOWESS trend above O – E = 0 represents warming and below 0 represents cooling in response to a particular SWA. If predictions made by setting SWA to 0 perform similarly to predictions from RCMs, the LOWESS lines of the 2 models should show similar trends and overlap with each other. We log(x)-transformed all SWA measures to aid in interpretation of the plots.

Results

DMs

Mean summer stream temperature (MSST).—Nine predictors were selected to model MSSTs (Fig. 2, Appendix S3; available online from: http://dx.doi.org/10.1899/12-009.1.s3), including 6 natural predictors (Fig. 3) and 3 measures of SWA (Fig. 4). MSSTs warmed with increasing values of 5 predictors: mean summer air temperature, watershed area, soil bulk density, and 2 measures of SWA: % watershed in agricultural and urban land uses (henceforth agriculture and urban indices, respectively). Factors negatively associated with MSST, in rank order of importance, were BFI, maximum upstream reservoir volume (inversely weighted by a $D_e = 50$ km and normalized by watershed area; reservoir index), average channel slopes within the watershed, and elevation ranges within watersheds (Figs 3, 4).

Mean winter stream temperature (MWST).—As in the MSST model, mean winter air temperature was the most important predictor of MWSTs (Figs 2, 3). In addition to air temperature, 5 natural predictors (Fig. 3) and 3 measures of SWA (Fig. 4) were selected to model MWSTs (Fig. 2, Appendix S3). Two measures of SWA (the reservoir and urban indices) were positively associated with MWSTs, whereas the agricultural index was negatively associated with MWSTs (Fig. 4). Compared with the MSST model, the direction of the relationships between MWST and the
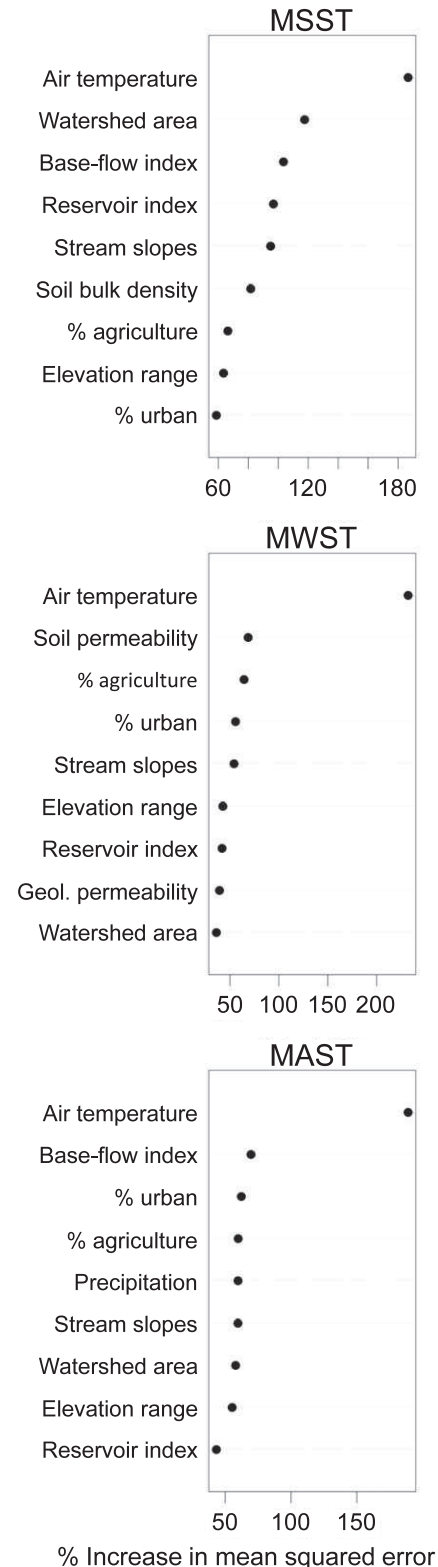

FIG. 2.   Ranked importance (% increase in mean squared error) of the predictor variables for the mean summer (MSST), winter (MWST), and annual (MAST) stream temperature models. Geol.= geological.
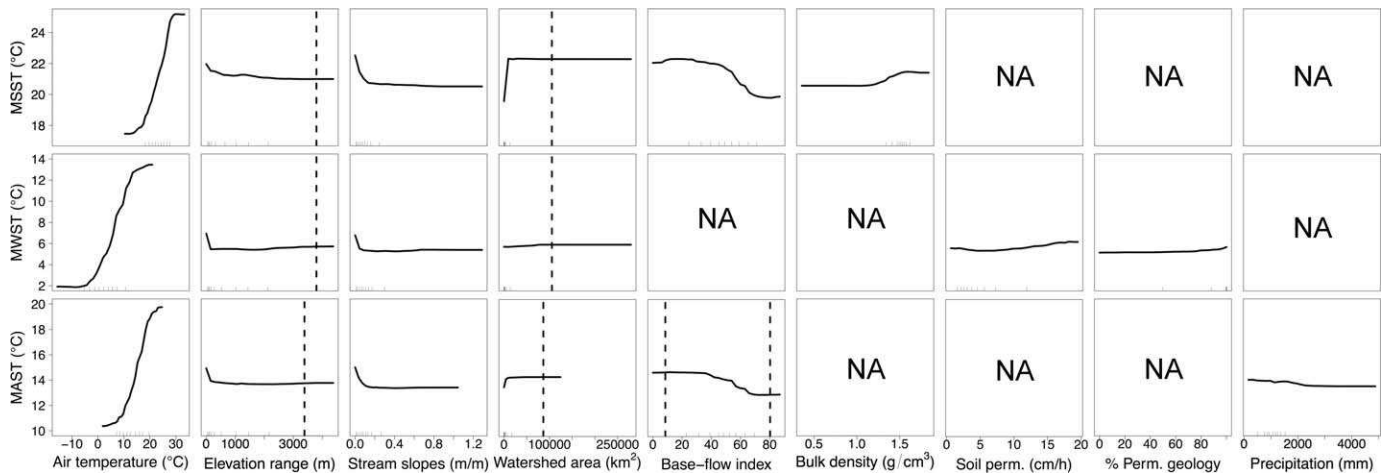
FIG. 3. Partial dependence plots showing how stream temperature responded to the individual natural predictors selected for the mean summer (MSST), winter (MWST), and annual (MAST) stream temperature dirty models. The vertical dashed lines represent the extremes of values observed at reference sites, if different from observations used in dirty models (DMs). NA = not applicable.
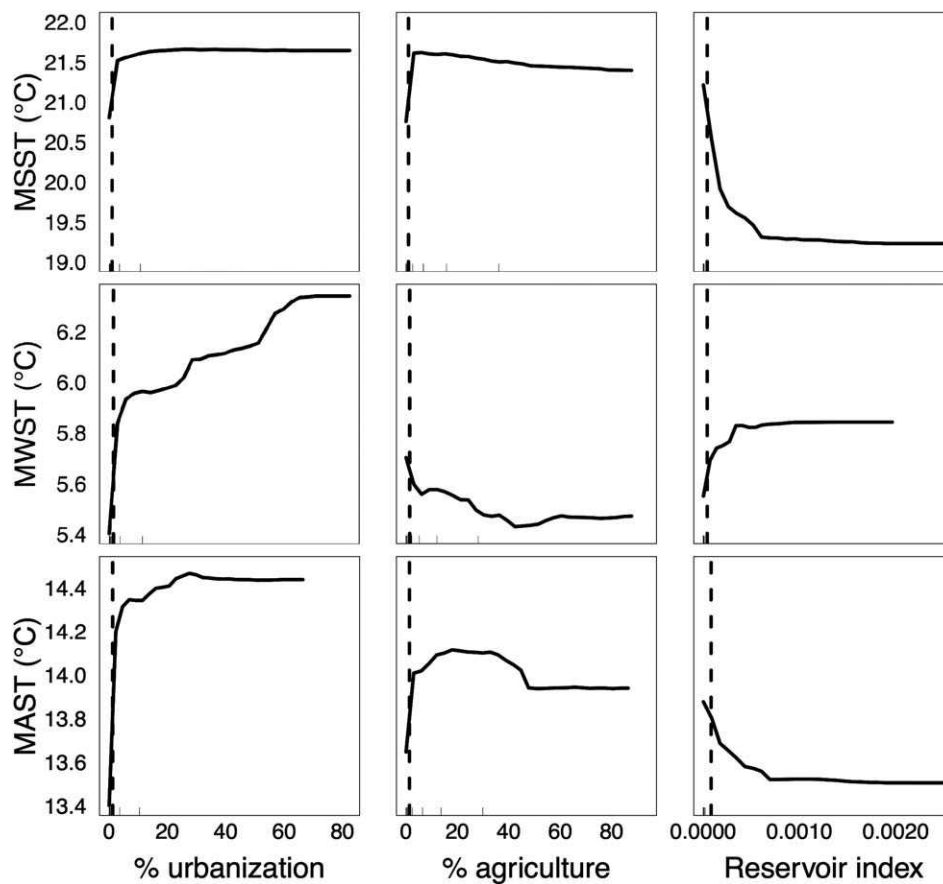

FIG. 4. Partial dependence plots showing how mean summer (MSST), winter (MWST), and annual (MAST) stream temperature responded to individual measures of stream and watershed alteration. The vertical dashed lines represent values of alteration below which we considered US Geological Survey stream temperature sites to be in thermal reference condition. NA = not applicable.

agricultural and reservoir indices were reversed (cf. MSST and MWST PDPs in Fig. 4). Slightly warmer MWSTs were associated with higher values of soil and geologic permeability (Fig. 3). These factors may be associated with the amount of shallow and deep groundwater flow within the watershed. Cooler MWSTs were associated with greater elevation range and steeper average channel slopes within the watershed. PDPs for watershed area and geologic permeability showed little response in MWSTs but both contributed to the overall performance of the model. Most watersheds with large areas were associated with slightly cooler MWSTs. Warmer MWST values occurred at the largest watershed areas, but the scarcity of data for large watersheds limited the reliability of trend lines in this part of the PDP (Fig. 3) (Hastie et al. 2001).

Mean annual stream temperature.—The predictor variables (Fig. 2, Appendix S3) selected for the MAST model and the directions of their relationships with MAST were very similar to those observed for the MSST model (cf. MSST and MAST; Figs 2, 3, 4). However, the order and relative magnitude of associations between MAST and its predictors differed. For example, the urban and agriculture indices were the 3rd and 4th most important predictors in the MAST model, whereas these predictors were ranked lower for the MSST model (cf. MSST and MAST; Fig. 2). In contrast, the reservoir index was ranked higher for the MSST model, compared with the MAST model (Fig. 2). Mean annual air temperatures, watershed area, and the urban and agriculture indices were positively associated with MASTs (Figs 3, 4). Increasing values of BFI, elevation range, average stream slopes within the watershed, long-term precipitation, and the reservoir index were all associated with cooler MASTs (Figs 3, 4).

Reference-site selection and models

We used conservative thresholds to select reference-condition sites (e.g., ≤ 1% agriculture and urbanization within the MSST watersheds). Applying the SWA thresholds (Fig. 4) to identify reference-condition sites for each model period identified 570 MSST, 481 MWST, and 273 MAST sites. The same natural predictors that were selected in the DMs were selected in the RCMs. The direction and pattern of ST responses to the natural predictors were very similar in the RCMs and DMs. PDPs are not shown here.

Reference screening decreased the geographic representativeness of the data, especially in midwestern states where agriculture is ubiquitous (cf. Figs 1 and 5). Despite the loss of geographic coverage of the reference
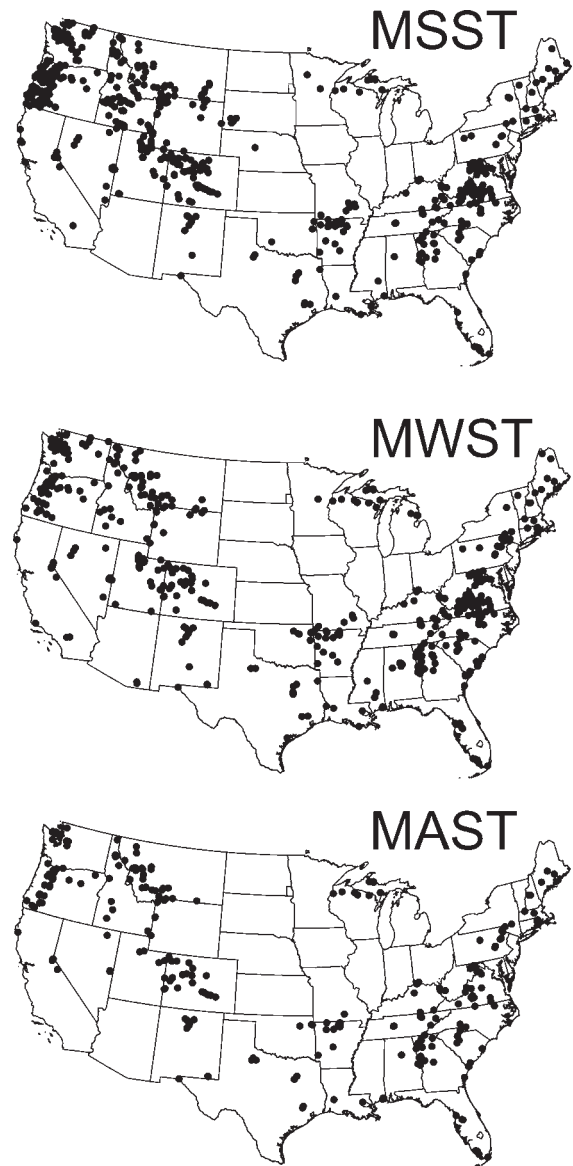

FIG. 5. Distribution of US Geological Survey sites with temperature data within the conterminous USA, and sites for which mean summer (MSST), winter (MWST), and annual (MAST) stream temperatures were used to develop the reference-condition models.

data sets, CFD plots for the predictor variables showed that most of the predictor space was retained (cf. RCM, DM, and all USGS ST site CFD plots in Appendix S4; available online from: http://dx.doi.org/10.1899/12-009.1.s4), except for the largest watershed areas and elevation ranges (Fig. 6). The largest watersheds were not geographically concentrated, but the largest elevation ranges were concentrated in the Rocky and Appalachian mountains. The reference MAST data set lost additional predictor space at the lowest and highest values of BFI (Fig. 6). Sites with the lowest BFI
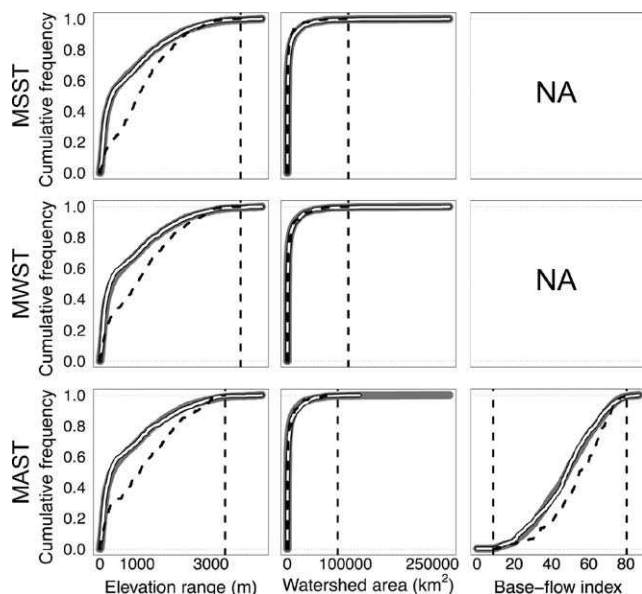
FIG. 6. Cumulative frequency distribution (CFD) plots of natural predictors that had truncated ranges (vertical black dashed lines) in the reference-condition models (black dashed) compared with dirty models (solid white) for mean summer (MSST), winter (MWST), and annual (MAST) stream temperatures. Solid grey lines represent the CFDs of all available US Geological Survey stream temperature sites. NA = not applicable.

values were spatially concentrated in the Southwestern and Central Plains States, such as Arizona, New Mexico, Texas, Oklahoma, Kansas, and Missouri. Sites with the highest BFI values occurred in the Rocky Mountains and northern Michigan. For most predictors, the reference-condition and SWA-influenced sites covered the same range of predictor values as the full set of USGS temperature sites. Only the highest stream

TABLE 3. The squared correlation coefficient (r^2), Nash–Sutcliffe coefficient (NSE), % bias (PBIAS), root mean squared error (RMSE), and RMSE/observed standard deviation (RMSE/SD) for the mean summer (MSST), winter (MWST), and annual (MAST) stream temperature models.

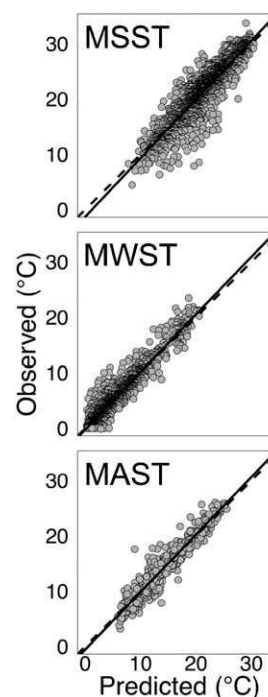| Model	r^2	NSE	PBIAS	RMSE (°C)	RMSE/ SD
DM					
MSST	0.84	0.84	0.07	2.0	0.40
MWST	0.92	0.92	−0.42	1.4	0.28
MAST	0.94	0.94	−0.05	1.2	0.25
RCM					
MSST	0.87	0.87	0.07	1.9	0.36
MWST	0.89	0.88	−0.70	1.4	0.34
MAST	0.95	0.95	−0.06	1.1	0.23

FIG. 7. Observed vs predicted mean summer (MSST), winter (MWST), and annual (MAST) stream temperatures with the least-squares fitted lines (dashes) and 1:1 lines (solid).

slopes and largest watershed areas were not included in our models. However, DM PDPs showed that STs were probably not sensitive to increased values of these predictors (see vertical lines in Fig. 3), i.e., response scope was similar in RCMs and DMs.

Model performances

Both the DMs and RCMs explained a large proportion of the variance in STs (r^2 values = 0.84–0.95, Table 3). The performance metrics and observed-vs-predicted plots were similar between the DMs and RCMs (Table 3), and only the DM observed-vs-predicted plots are presented here (Fig. 7). PBIAS values ranged between −0.7 (slight over-prediction of MWST RCM) and 0.07 (slight under-prediction of MSST RCM). These PBIAS values indicate little bias in the models and were well below the values Moriasi et al. (2007) suggested as indicative of good performance for stream characteristics modeled at monthly time steps with simulation models (i.e., stream flow PBIAS < ±10, sediment PBIAS < ±15, and N and P PBIAS < ±25). The PBIAS values associated with both the RCM and DM for MAST models were very small (−0.06 and −0.05, respectively), and observed and predicted values were in good agreement (Fig. 7). The NSE and RMSE/SD values also indicated good model
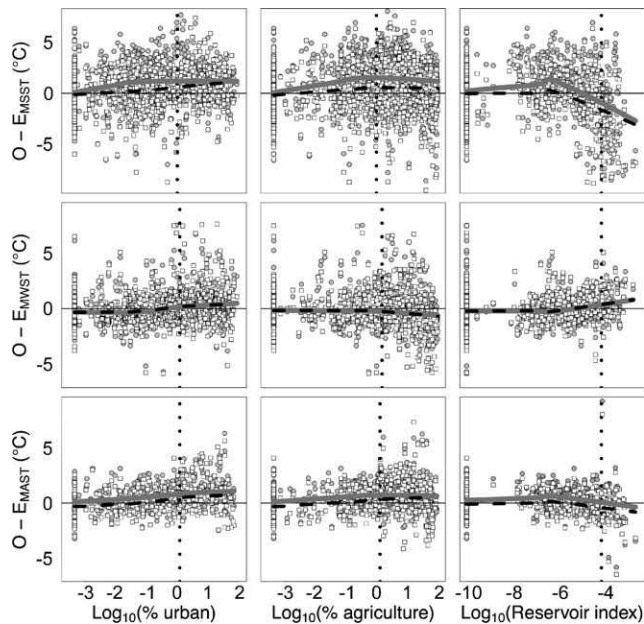
FIG. 8. Bias in model predictions of mean summer (MSST), winter (MWST), and annual (MAST) stream temperatures as a function of urbanization, agriculture, and reservoir alteration. Bias is measured as the difference between observed (O) and expected (E) reference-condition stream temperatures. Expected values for MSST, MWST, and MAST were derived from both reference-condition models (dashed line) and dirty models (solid grey line) for which stream and watershed alterations were set to 0. Vertical dashed lines represent thresholds used to define reference condition for each stream or watershed alteration measure.

performance based on values suggested by Moriasi et al. (2007) (i.e., NSE $\geq$ 0.75 and RMSE/SD $\leq$ 0.5; Table 3). The MWST RCMs and DMs had absolute RMSE values of 1.4°C. The MAST and MSST RMSE values for the RCM was slightly lower than that for the DM (MAST = 1.1 vs 1.2°C, MSST = 1.9 vs 2.0°C).

Predicting reference-condition ST with DMs

When applied to sites in reference condition, the SWA-zeroed DMs produced biased predictions of MSST and MAST (cf. LOWESS lines in Fig. 8; mean O – E values in Table 4). In contrast, the RCM predictions were unbiased. The MSST RCM was also more precise than the MSST DM (Table 4). The biases produced by the SWA-zeroed MSST and MAST DMs carried over to predictions made at nonreference sites (plotted to the right of the vertical dashed lines in Fig. 8). For nonreference sites, the DMs overestimated the effects of urbanization and agriculture relative to the RCMs. Conversely, the DMs underestimated cooling at nonreference sites below reservoirs. For

TABLE 4. Mean and standard deviation (SD) of mean summer (MSST), winter (MWST), and annual (MAST) stream temperature differences between observed (O) conditions and those expected (E) in the absence of human alteration (O – E) for dirty models (DM) and reference-condition models (RCM).

Model	Mean O – E	SD O – E
DM		
MSST	0.67	2.2
MWST	−0.07	1.4
MAST	0.42	1.1
RCM		
MSST	0.02	2.0
MWST	−0.04	1.4
MAST	−0.002	1.1

MWST, DM and RCM predictions agreed well (Fig. 8). Both the DM and RCM slightly overestimated MWST at reference-condition sites (LOWESS lines below 0), but these biases were small (mean O – E in Table 4).

The O – E LOWESS trends were consistent with the PDP plots (cf. Figs 4 and 8). The MSST and MAST models showed warming in response to increasing values of agriculture within the watershed and cooling in association with the reservoir index. In contrast, the winter model showed the reverse relationship with these measures of SWA. All models displayed warming associated with greater urbanization within the watershed. In addition, most of the O – E LOWESS lines began to deviate from 0 at SWA values that were lower than the thresholds we used to define reference condition (vertical dashed lines in Fig. 8), implying a response in ST to SWA below the thresholds used to select reference-condition sites.

Discussion

Assessments of our models suggest they accurately and precisely estimate STs across a large geographic extent with varied environments, but several factors must be considered. First, our models must be placed in context with other published empirical ST models. A favorable comparison of the performance of our models with that of other published models should provide additional confidence in their potential use for: 1) assessing the thermal conditions of USA streams, 2) providing a mechanistic understanding of macroecological patterns in streams and rivers, and 3) exploring historical and future responses of streams to climate change. In addition, we can gain insight into the relative influence of certain landscape features on STs by comparing the selected and

excluded predictors of published empirical models that were developed at different geographic scales. Last, we briefly consider the use of DMs and RCMs to infer RCST and the implications of our findings for hindcasting of water-quality variables.

Model performance

Spatially explicit models that relate landscape features to stream characteristics, such as STs, are gaining popularity (Wang et al. 2006), but most previous work has not reported performance statistics that would allow objective comparison with our models. Isaak et al. (2010) modeled summer STs (15 July–15 September) with data from 780 ST sites within the Boise River, Idaho. Based on leave-one-out cross validation, they reported an RMSE of 0.74°C and an SD of observed STs of 2.7°C, resulting in an RMSE/SD of 0.27. This value is smaller than the RMSE/SD values of our MSST models but similar to those of our MWST and MAST models (Table 3). Wehrly et al. (2006) modeled mean July STs in lower Michigan, and reported an SD of residual errors of 1.9°C. However, Wehrly et al. (2006) did not report the SD of observed STs. To compare the performance of their model with ours, we used their reported range of observed July STs (9.2–26.7°C) to calculate a normalized SD of residual errors of 11%, which is higher than our normalized SD of residual errors of 7% for MSST. These values suggest similar or better performance of our models but at a spatial scale several orders of magnitude larger than was used in the 2 previous studies. Our models are an important advance in characterizing regional variation in STs, especially given the spatial scale at which they can be applied.

Model applications

Assessments of the ecological condition of streams are routinely conducted in the USA and elsewhere, and researchers have expended substantial effort on developing statistical tools to objectively assess the biological condition of streams (reviewed by Hawkins et al. 2010). Similar approaches could be applied with the models presented here to assess the thermal condition of streams. We used natural landscape predictors that allow accurate predictions of STs at unmeasured locations, and these site-specific predictions of reference-condition STs can be used as benchmarks to infer whether an assessed stream reach is thermally impaired. Furthermore, ST models could be used in support of ecological assessments because ST is a major determinant of the distribution of aquatic species within a landscape (Vannote and Sweeney 1980, Haidekker and Hering 2008). Many

ecological assessments compare observed biota with the biota predicted to occur under reference environmental conditions (Moss et al. 1987, Hawkins et al. 2000, Simpson and Norris 2000). The species distribution models used to predict reference-condition biota typically use surrogates of natural ST, such as latitude, elevation, or drainage area. These surrogates are imperfect predictors of thermal reference conditions in streams. Inclusion of well predicted STs in species distribution models such as River InVertebrate Prediction and Classification System (RIVPACS; Moss et al. 1987, Hawkins et al. 2000) should improve the precision and accuracy of ecological assessments and their interpretation. In addition, conducting a thermal assessment in conjunction with a biological assessment should aid in diagnosing whether altered temperature is a likely cause of observed biological impairment.

ST models will be essential tools in establishing a more comprehensive understanding of ST changes that have already occurred and probably will occur in response to climate warming. For example, Isaak et al. (2010) used a multisite empirical model in the Boise River basin, Idaho, to account for variation in observed STs between 1993 and 2006. They found that the effects of climate change on thermal habitats depend on landscape context and that the loss of available Bull Trout (*Salvelinus confluentus*) thermal habitat was greatest in headwater streams. However, most empirical studies of the potential effects of climate change on STs were based on empirical stream–air temperature relationships at individual sites (e.g., Mohseni et al. 1999, 2003), and thus, the landscape context associated with differing vulnerabilities of STs to predicted changes in climate could not be considered. Empirical models derived from data that cover the range of conditions found within a region of interest will have much greater utility in assessing the potential region-specific effects of climate change on STs and identifying individual streams and regions that may be especially vulnerable to climate change.

Excluded predictors

Those predictors that were excluded from the models during calibration were as notable as the predictors that were selected. We expected estimates of solar radiation to be strongly associated with variation in STs among sites, especially in summer. However, solar radiation was not a significant predictor in any model. When we included solar radiation in the pilot western USA MSST model, RMSE decreased by only <0.1°C. If we substituted

solar radiation for air temperature, MSST and MWST RMSEs increased by 17 and 80%, respectively. The observed lack of strong association between ST and solar radiation may have been the result of inaccurate estimates of solar radiation striking each stream. However, Wehrly et al. (2006) also noted a weak association between STs and solar radiation in a multisite empirical model of STs in Michigan. Conversely, Isaak et al. (2010) found that radiation was an important predictor of STs in the Boise River basin, Idaho. Whether solar radiation is an important predictor of STs in empirical models may be related to the scales at which models are developed, the effects of cloud cover on solar radiation (not measured in this analysis), and the spatial variability of radiation relative to other predictors within the model. Wehrly et al. (2006) suggested that studies in which solar radiation is a good predictor of STs are generally conducted in single watersheds where other environmental predictors vary little relative to canopy cover and, thus, the solar radiation striking the stream. In short, at large spatial scales, air temperature may integrate the multiple heat-exchange processes that influence ST.

We also included several short- and long-term measures of precipitation as potential predictors (Appendix S1) and expected them to be strong predictors of STs because of their relationship with stream flow. However, long-term precipitation was only moderately important as a predictor in the MAST model. Additional research may be needed to better characterize precipitation (e.g., timing of precipitation events) for predicting MSST and MWST or to conclude that precipitation is a weak predictor of STs at a large geographic scale. Last, in contrast to the observation of Wehrly et al. (2009), who found that mean July STs in Michigan were positively related to the amount of upstream lentic waterbodies, lakes and wetlands were not selected in any of the models. The importance of lentic waterbodies to July STs in Michigan and Wisconsin may reflect the prominence of this landscape feature in these states and its role in influencing STs at that scale relative to the conterminous USA.

RCMs vs DMs

Stream assessments must be precise and unbiased to be useful. If a management goal were to maintain or restore naturally occurring thermal reference conditions, on average the SWA-zeroed MSST and MAST O – E models would underprotect (Type I error) sites with upstream reservoirs and overprotect (Type II error) sites with urban and agricultural land uses within the watersheds. For these thermal attributes, the RCMs would provide more accurate and defensible assessments. However, for MWST, use of either the RCM or the DM would allow reasonably precise and unbiased assessments. These results have important implications for hindcasting of historical conditions. The DMs we developed included both reference and nonreference sites and, therefore, did not extrapolate beyond the range of the data. However, even with the benefit of a full spectrum of SWA information, the MSST and MAST DMs produced biased predictions of reference-condition STs. Models calibrated without data from sites in reference condition would have to extrapolate predictions of thermal reference conditions, which would almost certainly result in larger biases than observed in our DMs.

Our analyses also illustrate a specific challenge associated with establishing reference-condition expectations from a network of reference sites that vary in their quality (i.e., the amount of SWA potentially affecting them). The most liberal land-cover thresholds we defined were 1.5% of the watershed in agriculture or urbanization in the MWST models. The MSST and MAST thresholds were more conservative (agriculture and urban indices $\leq 1\%$ in MSST watersheds, and ≤ 1.2 and 1.3%, respectively, in MAST watersheds). Yet several of the RCM O – E LOWESS lines showed systematic deviation from 0 in response to these SWAs below these thresholds (Fig. 8). The deviations were small enough for urbanization and agriculture that use of the thresholds we selected would not seriously compromise predictions of true RCSTs. However, the deviations in O – E values associated with the reservoir index were larger, a result implying that we should consider adjusting the reservoir threshold when selecting reference sites. For example, if the reference-condition threshold were adjusted to a $\log_{10}$(reservoir index) value of -5, biases in the O – E values at reference sites could be minimized (Fig. 8). However, doing so would reduce the MAST reference observations from 273 to 224 for the conterminous USA and further reduce the spatial and environmental representativeness of the model. The addition of nonUSGS ST sites could increase the number and environmental representativeness of reference-condition sites (e.g., http://greatnorthernlcc.org/technical/stream-temp-maps). However, additional reference-quality streams are not likely to be identified in regions with nearly ubiquitous SWA, such as agriculture in the midwestern USA (Fig. 5). Selecting sites that are "reference enough", while maintaining a sufficient number of sites to be representative of the environments within a region, is a major challenge in all environmental assessments.

The inability of the SWA-zeroed MSST and MAST DMs to produce unbiased O – E values could be caused by the coarseness of the SWA measures, such as the reservoir index. First, because of incomplete NID records, we were forced to use reservoir volumes as a predictor. Reservoir volume is only weakly associated with reservoir depth within the NID (r^2 = 0.27), and the temperature of the water released by a dam is a function of the depth at which it is released (Bonnet et al. 2000, Lindim et al. 2011). The addition of information to the NID that specifies the depth or type of water release (e.g., hypolimnetic or epilimnetic) might improve the accuracy of our models. Alternatively, correcting and completing NID structure-height information could improve results because this attribute is probably better correlated with the likelihood of thermal stratification in reservoirs than volume and, thus, the temperature of released water. Second, we expended considerable effort to screen 53,041 NID records, but errors still exist within the data. We noted several outliers within the calibration data sets while developing the models. These outliers often were associated with inaccurate reservoir location information, and correction improved predictions. However, missing or inaccurate information may not always result in obvious outliers, but rather noise within the models. Additional screening of the NID could improve confidence in predictions.

Concluding remarks

Our RCMs accurately and precisely predicted reference STs at unmeasured streams across a broad range of environments in the conterminous USA. We think these models represent a significant step toward a more comprehensive assessment of the environmental and ecological conditions of USA rivers and streams. Thermal assessments would complement previous and ongoing assessments of the biological (Paulsen et al. 2008) and hydrologic condition (Carlisle et al. 2009) of USA streams and rivers. In addition, these models provide a tool for understanding how specific SWAs have affected STs and how other alterations, such as climate change, might further alter them in the future.

To our knowledge, no investigators have compared RCM predictions and DM hindcastings of reference condition. Relative to RCMs, the DMs produced biased estimates of reference-condition STs. These predictions potentially could be improved with better landuse information that accounts for more specific alterations, such as reservoir-release temperatures, wastewater treatment facilities in urban areas, irrigation withdrawals, and return flows associated with agricultural and mining activities. However, these types of data are not readily available everywhere and will take time to develop. Unless a high degree of confidence exists that the available measures of SWA account for nearly all of the thermal alteration that occurs at different sites, we recommend caution in using DMs to predict reference-condition water quality.

Literature Cited

ACUÑA, V., A. WOLF, U. UEHLINGER, AND K. TOCKNER. 2008. Temperature dependence of stream benthic respiration in an Alpine river network under global warming. Freshwater Biology 53:2076–2088.

ALLAN, J. D. 2004. Landscapes and riverscapes: the influence of land use on stream ecosystems. Annual Review of Ecology, Evolution, and Systematics 35:257–284.

ALLAN, J. D., AND M. M. CASTILLO. 2007. Stream ecology: structure and function of running waters. 2nd edition. Springer, Dordrecht, The Netherlands.

BAKER, E. A., K. E. WEHRLY, P. W. SEELBACH, L. WANG, M. J. WILEY, AND T. SIMON. 2005. A multimetric assessment of stream condition in the Northern Lakes and Forests ecoregion using spatially explicit statistical modeling and regional normalization. Transactions of the American Fisheries Society 134:697–710.

BANFIELD, R. E., L. O. HALL, K. W. BOWYER, AND W. P. KEGELMEYER. 2007. A comparison of decision tree ensemble creation techniques. Transactions on Pattern Analysis and Machine Intelligence 29:173–180.

BARTHOLOW, J. M. 2000. Estimating cumulative effects of clearcutting on stream temperatures. Rivers 7:284–297.

BENKE, A. C., C. A. S. HALL, C. P. HAWKINS, R. H. LOWE-MCCONNELL, J. A. STANFORD, K. SUBERKROPP, AND J. V. WARD. 1988. Bioenergetic considerations in the analysis of stream ecosystems. Journal of the North American Benthological Society 7:480–502.

BONNET, M. P., M. POULIN, AND J. DEVAUX. 2000. Numerical modeling of thermal stratification in a lake reservoir. Methodology and case study. Aquatic Sciences 62:105–124.

BREIMAN, L. 2001. Random forests. Machine Learning 45:5–32.

BROWN, G. W. 1969. Predicting temperatures of small streams. Water Resources Research 5:68–75.

BROWN, G. W. 1970. Predicting the effect of clearcutting on stream temperature. Journal of Soil and Water Conservation 25:11–13.

BROWN, J. H., J. F. GILLOOLY, A. P. ALLEN, V. M. SAVAGE, AND G. B. WEST. 2004. Toward a metabolic theory of ecology. Ecology 85:1771–1789.

CAISSIE, D. 2006. The thermal regime of rivers: a review. Freshwater Biology 51:1389–1406.

CAISSIE, D., M. G. SATISH, AND N. EL-JABI. 2007. Predicting water temperatures using a deterministic model: application on Miramichi River catchments (New Brunswick, Canada). Journal of Hydrology 336:303–315.

CARLISLE, D. M., J. FALCONE, AND M. R. MEADOR. 2009. Predicting the biological condition of streams: Use of geospatial indicators of natural and anthropogenic characteristics of watersheds. Environmental Monitoring and Assessment 151:143–160.

CARLISLE, D. M., J. FALCONE, D. WOLOCK, M. MEADOR, AND R. NORRIS. 2010. Predicting the natural flow regime: Models for assessing hydrological alteration in streams. River Research and Applications 26:118–136.

CATHERINE, A., D. MOUILLOT, N. ESCOFFIER, C. BERNARD, AND M. TROUSSELLIER. 2010. Cost effective prediction of the eutrophication status of lakes and reservoirs. Freshwater Biology 55:2425–2435.

CHENARD, J. F., AND D. CAISSIE. 2008. Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. Hydrological Processes 22:3361–3372.

CHINNAYAKANAHALLI, K., C. KROEBER, R. A. HILL, D. G. TARBOTON, J. R. OLSON, AND C. P. HAWKINS. 2006. The multi-watershed delineation tool: GIS software in support of regional watershed analyses. Utah State University, Logan, Utah.

CHINNAYAKANAHALLI, K. J., C. P. HAWKINS, D. G. TARBOTON, AND R. A. HILL. 2011. Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States. Freshwater Biology 56:1248–1265.

CLEVELAND, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association 74:829–836.

CLUIS, D. A. 1972. Relationship between stream water temperature and ambient air temperature. Nordic Hydrology 3:65–71.

CUTLER, D. R., T. C. EDWARDS, K. H. BEARD, A. CUTLER, AND K. T. HESS. 2007. Random forests for classification in ecology. Ecology 88:2783–2792.

DALY, C., M. HALBLEIB, J. I. SMITH, W. P. GIBSON, M. K. DOGGETT, G. H. TAYLOR, J. CURTIS, AND P. P. PASTERIS. 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. International Journal of Climatology 28:2031–2064.

DONATO, M. M. 2002. A statistical model for estimating stream temperatures in the Salmon and Clearwater River Basins, Central Idaho. U.S. Geological Survey Water Resources Investigations Report 02-4195. US Geological Survey, Boise, Idaho.

EDINGER, J. E., D. W. DUTTWEILER, AND J. C. GEYER. 1968. The response of water temperatures to meteorological conditions. Water Resources Research 4:1137–1143.

FRANCKE, T., J. A. LÓPEZ-TARAZÓN, AND B. SCHRÖDER. 2008. Estimation of suspended sediment concentration and yield using linear models, random forests and quantile regression forests. Hydrological Processes 22:4892–4904.

GENUER, R., J.-M. POGGI, AND C. TULEAU-MALOT. 2010. Variable selection using random forests. Pattern Recognition Letters 31:2225–2236.

HAAG, I., AND A. LUCE. 2008. The integrated water balance and water temperature model LARSIM-WT. Hydrological Processes 22:1046–1056.

HAGEN, E. M., J. R. WEBSTER, AND E. F. BENFIELD. 2006. Are leaf breakdown rates a useful measure of stream integrity along an agricultural landuse gradient? Journal of the North American Benthological Society 25:330–343.

HAIDEKKER, A., AND D. HERING. 2008. Relationship between benthic insects (Ephemeroptera, Plecoptera, Coleoptera, Trichoptera) and temperature in small and medium-sized streams in Germany: a multivariate study. Aquatic Ecology 42:463–481.

HASTIE, T. J., R. J. TIBSHIRANI, AND J. H. FRIEDMAN. 2001. The elements of statistical learning: Data mining, inference, and prediction. Springer, New York.

HAWKINS, C. P., J. N. HOGUE, L. M. DECKER, AND J. W. FEMINELLA. 1997. Channel morphology, water temperature, and assemblage structure of stream insects. Journal of the North American Benthological Society 16:728–749.

HAWKINS, C. P., R. H. NORRIS, J. N. HOGUE, AND J. W. FEMINELLA. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. Ecological Applications 10:1456–1477.

HAWKINS, C. P., J. R. OLSON, AND R. A. HILL. 2010. The reference condition: predicting benchmarks for ecological and water-quality assessments. Journal of the North American Benthological Society 29:312–343.

HOLDEN, Z. A., M. A. CRIMMINS, S. A. CUSHMAN, AND J. S. LITTELL. 2011. Empirical modeling of spatial and temporal variation in warm season nocturnal air temperatures in two North Idaho mountain ranges, USA. Agricultural and Forest Meteorology 151:261–269.

HOMER, C., J. DEWITZ, J. FRY, M. COAN, N. HOSSAIN, C. LARSON, N. HEROLD, A. MCKERROW, J. N. VANDRIEL, AND J. WICKHAM. 2007. Completion of the 2001 National Land Cover Database for the conterminous United States. Photogrammetric Engineering and Remote Sensing 73:337–341.

HUETE, A., K. DIDAN, T. MIURA, E. P. RODRIGUEZ, X. GAO, AND L. G. FERREIRA. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment 83:195–213.

ISAAK, D. J., C. H. LUCE, B. E. 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. Ecological Applications 20:1350–1371.

Jones, K. L., G. C. Poole, J. L. Meyer, W. Bumback, and E. A. Kramer. 2006. Quantifying expected ecological response to natural resource legislation: a case study of riparian buffers, aquatic habitat, and trout populations. Ecology and Society 11:15.

Karpievitch, Y. V., E. G. Hill, A. P. Leclerc, A. R. Dabney, and J. S. Almeida. 2009. An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. PLoS One 4: e7087.

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. Frontiers in Ecology and the Environment 8:461–466.

Kelleher, C., T. Wagener, M. Gooseff, B. McGlynn, K. McGuire, and L. Marshall. 2012. Investigating controls on the thermal sensitivity of Pennsylvania streams. Hydrological Processes 26:771–785.

Kilgour, B. W., and L. W. Stanfield. 2006. Hindcasting reference condition in streams. Pages 623–639 in R. M. Hughes, L. Wang, and P. W. Seelbach (editors). Influences of landscapes on stream habitats and ecological assemblages. American Fisheries Society, Bethesda, Maryland.

Kinouchi, T., H. Yagi, and M. Miyamoto. 2007. Increase in stream temperature related to anthropogenic heat input from urban wastewater. Journal of Hydrology 335: 78–88.

Klein, R. D. 1979. Urbanization and stream quality impairment. Water Resources Bulletin 15:948–963.

Kumar, L., A. K. Skidmore, and E. Knowles. 1997. Modelling topographic variation in solar radiation in a GIS environment. International Journal of Geographical Information Science 11:475–497.

Lessard, J. L., and D. B. Hayes. 2003. Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. River Research and Applications 19:721–732.

Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2(3):18–22.

Lindim, C., J. L. Pinho, and J. M. P. Vieira. 2011. Analysis of spatial and temporal patterns in a large reservoir using water quality and hydrodynamic modeling. Ecological Modelling 222:2485–2494.

Martin, S. 1981. Frazil ice in rivers and oceans. Annual Review of Fluid Mechanics 13:379–397.

McCabe, G. J., and D. M. Wolock. 2010. Century-scale variability in global annual runoff examined using a water balance model. International Journal of Climatology 31:1739–1748.

McKenna, J. E., R. S. Butryn, and R. P. McDonald. 2010. Summer stream water temperature models for Great Lakes streams: New York. Transactions of the American Fisheries Society 139:1399–1414.

McTammany, M. E., E. F. Benfield, and J. R. Webster. 2007. Recovery of stream ecosystem metabolism from historical agriculture. Journal of the North American Benthological Society 26:532–545.

Miyake, Y., and U. Takeuchi. 1951. On the temperature of river waters of Japan. Japanese Journal of Limnology 15: 145–151.

Mohseni, O., T. R. Erickson, and H. G. Stefan. 1999. Sensitivity of stream temperatures in the United States to air temperatures projected under a global warming scenario. Water Resources Research 35:3723–3733.

Mohseni, O., H. G. Stefan, and J. G. Eaton. 2003. Global warming and potential changes in fish habitat in US streams. Climatic Change 59:389–409.

Mohseni, O., H. G. Stefan, and T. R. Erickson. 1998. A nonlinear regression model for weekly stream temperatures. Water Resources Research 34:2685–2692.

Moriasi, D. N., J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith. 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE 50: 885–900.

Morin, G., D. Couillard, D. Cluis, H. G. Jones, and J. M. Gauthier. 1987. Prediction of temperatures in rivers using a conceptual-model. Hydrological Sciences Journal – Journal des Sciences Hydrologiques 32:31–41.

Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17:41–52.

Nelson, K. C., and M. A. Palmer. 2007. Stream temperature surges under urbanization and climate change: data, models, and responses. Journal of the American Water Resources Association 43:440–452.

Olden, J. D., and R. J. Naiman. 2010. Incorporating thermal regimes into environmental flows assessments: modifying dam operations to restore freshwater ecosystem integrity. Freshwater Biology 55:86–107.

Ordoyne, C., and M. A. Friedl. 2008. Using MODIS data to characterize seasonal inundation patterns in the Florida Everglades. Remote Sensing of Environment 112: 4107–4119.

Paulsen, S. G., A. Mayio, D. V. Peck, J. L. Stoddard, E. Tarquinio, S. M. Holdsworth, J. Van Sickle, L. L. Yuan, C. P. Hawkins, A. T. Herlihy, P. R. Kaufmann, M. T. Barbour, D. P. Larsen, and A. R. Olsen. 2008. Condition of stream ecosystems in the US: An overview of the first national assessment. Journal of the North American Benthological Society 27:812–821.

Peters, J., B. De Baets, N. E. C. Verhoest, R. Samson, S. Degroeve, P. De Becker, and W. Huybrechts. 2007. Random forests as a tool for ecohydrological distribution modelling. Ecological Modelling 207:304–318.

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

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9:181–199.

PREECE, R. M., AND H. A. JONES. 2002. The effect of Keepit Dam on the temperature regime of the Namoi River, Australia. River Research and Applications 18:397–414.

REED, J. C., AND C. A. BUSH. 2001. Generalized geologic map of the conterminous United States. US Geological Survey, Denver, Colorado. (Available from: http://pubs.usgs.gov/atlas/geologic/48States/)

RISLEY, J. C., J. CONSTANTZ, H. ESSAID, AND S. ROUNDS. 2010. Effects of upstream dams versus groundwater pumping on stream temperature under varying climate conditions. Water Resources Research 46:W06517.

RISLEY, J. C., E. A. ROEHL, AND P. A. CONRADS. 2003. Estimating water temperatures in small streams in western Oregon using neural network models. US Geological Survey, Portland, Oregon.

ROLLINS, M. G., AND C. K. FRAME. 2006. The LANDFIRE prototype project: nationally consistent and locally relevant geospatial data for wildland fire management. USDA Forest Service General Technical Report RMRS-GTR-175. Rocky Mountain Research Station, US Department of Agriculture, Fort Collins, Colorado.

SEEDANG, S., A. G. FERNALD, R. M. ADAMS, AND D. H. LANDERS. 2008. Economic analysis of water temperature reduction practices in a large river floodplain: an exploratory study of the Willamette River, Oregon. River Research and Applications 24:941–959.

SIMPSON, J. C., AND R. H. NORRIS. 2000. Biological assessment of river quality: development of AUSRIVAS models and outputs. Pages 125–142 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, UK.

SINOKROT, B. A., H. G. STEFAN, J. H. MCCORMICK, AND J. G. EATON. 1995. Modeling of climate-change effect on stream temperatures and fish habitats below dams and near groundwater inputs. Climatic Change 30:181–200.

SNELDER, T. H., N. LAMOUROUX, AND H. PELLA. 2011. Empirical modelling of large scale patterns in river bed surface grain size. Geomorphology 127:189–197.

SORANNO, P. A., T. WAGNER, S. L. MARTIN, C. MCLEAN, L. N. NOVITSKI, C. D. PROVENCE, AND A. R. ROBER. 2011. Quantifying regional reference conditions for freshwater ecosystem management: a comparison of approaches and future research needs. Lake and Reservoir Management 27:138–148.

STEFAN, H. G., AND V. N. CHAU. 1976. River cooling water temperature fluctuations. Journal of Hydraulics Division, American Society of Civil Engineers 102:1613–1627.

THEURER, F. D., K. A. VOOS, AND W. J. MILLER. 1984. Instream water temperature model. Instream flow information paper 16. US Department of the Interior Fish and Wildlife Service, Fort Collins, Colorado.

USACE (US ARMY CORPS OF ENGINEERS). 2009. National inventory of dams. US Army Corps of Engineers, Washington, DC. (Available from: http://geo.usace.army.mil/pgis/f?p=397:1:0:)

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 Resources Research 47:W02544.

VANNOTE, R. L., AND B. W. SWEENEY. 1980. Geographic analysis of thermal equilibria: a conceptual model for evaluating the effect of natural and modified thermal regimes on aquatic insect communities. American Naturalist 115:667–695.

WANG, L. Z., P. W. SEELBACH, AND R. M. HUGHES. 2006. Introduction to landscape influences on stream habitats and biological assemblages. Pages 1–23 in R. M. Hughes, L. Wang, and P. W. Seelbach (editors). Landscape influences on stream habitat and biological assemblages. American Fisheries Society, Bethesda, Maryland.

WARD, J. C. 1963. Annual variation of stream water temperature. American Society of Civil Engineers, Journal of the Sanitary Engineering Division 89:3710–3732.

WARD, J. V. 1985. Thermal characteristics of running waters. Hydrobiologia 125:31–46.

WEBB, B. W., D. M. HANNAH, R. D. MOORE, L. E. BROWN, AND F. NOBILIS. 2008. Recent advances in stream and river temperature research. Hydrological Processes 22:902–918.

WEHRLY, K. E., T. O. BRENDEN, AND L. Z. WANG. 2009. A comparison of statistical approaches for predicting stream temperatures across heterogeneous landscapes. Journal of the American Water Resources Association 45:986–997.

WEHRLY, K. E., M. J. WILEY, AND P. W. SEELBACH. 2006. Influence of landscape features on summer water temperatures in Lower Michigan streams. Pages 113–128 in R. M. Hughes, L. Wang, and P. W. Seelbach (editors). Landscape influences on stream habitat and biological assemblages. American Fisheries Society, Bethesda, Maryland.

WILDE, F. D. 2006. Temperature (version 2). U.S. Geological Survey Techniques of Water-Resource Investigations. Book 9, Page 22. US Geological Survey, Reston, Virginia.

WOLOCK, D. M. 1997. STATSGO soil characteristics for the conterminous United States. U.S. Geological Survey Open-File Report 97-656. US Geological Survey, Lawrence, Kansas.

WOLOCK, D. M. 2003. Base-flow index grid for the conterminous United States. U.S. Geological Survey Open-File Report 03-263. US Geological Survey, Reston, Virginia.

WU, J. J., K. SKELTON-GROTH, W. G. BOGGESS, AND R. M. ADAMS. 2003. Pacific salmon restoration: trade-offs between economic efficiency and political acceptance. Contemporary Economic Policy 21:78–89.