8-2017

A Comparison of Five Statistical Methods for Predicting Stream Temperature Across Stream Networks

Maike F. Holthuijzen
Utah State University

A COMPARISON OF FIVE STATISTICAL METHODS FOR PREDICTING STREAM TEMPERATURE ACROSS STREAM NETWORKS

by

Maike F. Holthuijzen

A thesis submitted in partial fulfillment
of the requirements for the degree

of

MASTER OF SCIENCE

in

Statistics

Approved:

_____	_____
Richard Cutler, Ph.D Major Professor	Adele Cutler, Ph.D Committee Member
_____	_____
Charles Hawkins, Ph.D Committee Member	Mark R. McLellan, Ph.D Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2017

ABSTRACT

A Comparison of Five Statistical Methods for Predicting Stream Temperature Across Stream Networks

by

Maike Holthuijzen, Master of Science

Utah State University, 2017

Major Professor: Dr. Richard Cutler
Department: Mathematics and Statistics

The integrity of freshwater ecosystems, particularly stream networks, is strongly influenced by water temperature, which controls biological processes and influences species distributions and aquatic biodiversity. The thermal regimes of streams and rivers are likely to change in the future due to climate change and other anthropogenic impacts, and our ability to model stream temperatures will be critical in understanding and predicting distribution shifts of aquatic biota. Recently developed spatial statistical network (SSN) models, which explicitly account for spatial autocorrelation and hydrological distance, can have high predictive accuracy. However, SSN models have can have high computation times and data pre-processing requirements, which may compromise their routine use under some circumstances. Other modeling approaches, such as machine learning techniques and generalized additive models (GAM), are promising alternatives to SSN models in that they are typically more computationally efficient and are subject to fewer assumptions

than SSN models. In particular, machine learning methods such as gradient boosting machines (GBM) and Random Forests (RF) are both computationally efficient and can automatically model high-order interactions and non-linear responses. GAMs also can fit highly non-linear relationships, which may produce prediction error in SSN models, which assume linear relationships between response and predictor variables. However, we cannot yet generalize regarding the relative strengths and weaknesses of different modeling approaches because a direct comparison of prediction accuracy has not yet been conducted across a variety of methods.

My objectives were to 1) compare the accuracies of linear (LM), SSN, GAM, RF, and GBM models in predicting stream temperature from field observations, 2) conduct simulations to determine the effect of autocorrelation strength on prediction accuracies among all methods, and 3) provide guidelines in choosing a prediction method for ecologists and other practitioners. Through simulations, I compared prediction accuracy of all methods on datasets with varying degrees of linearity, spatial autocorrelation, and error structure. Prediction accuracies were quantified as the test-set root mean square error (RMSE) for all methods. For the field data, SSN had the highest predictive accuracy overall, followed closely by GBM and GAM. LM performed poorly overall. Simulations showed that for linearly-structured, spatially autocorrelated data, SSN achieved the most accurate prediction accuracy of all methods. However, GAM had the best performance on non-linearly structure data in simulations, regardless of the degree of spatial autocorrelation. This study shows that machine learning methods and GAM may provide suitable alternatives to SSN models for many stream temperature prediction applications,

especially when modeling 1,000's of data points and when the assumption of linear relationships is suspect.

(75 pages)

PUBLIC ABSTRACT

A Comparison of Five Statistical Methods for Predicting Stream Temperature Across Stream Networks

Maike Holthuijzen

The health of freshwater aquatic systems, particularly stream networks, is mainly influenced by water temperature, which controls biological processes and influences species distributions and aquatic biodiversity. Thermal regimes of rivers are likely to change in the future, due to climate change and other anthropogenic impacts, and our ability to predict stream temperatures will be critical in understanding distribution shifts of aquatic biota. Spatial statistical network models take into account spatial relationships but have drawbacks, including high computation times and data pre-processing requirements. Machine learning techniques and generalized additive models (GAM) are promising alternatives to the SSN model. Two machine learning methods, gradient boosting machines (GBM) and Random Forests (RF), are computationally efficient and can automatically model complex data structures. However, a study comparing the predictive accuracy among a variety of widely-used statistical modeling techniques has not yet been conducted.

My objectives for this study were to 1) compare the accuracy among linear models (LM), SSN, GAM, RF, and GBM in predicting stream temperature over two stream networks and 2) provide guidelines in choosing a prediction method for practitioners and ecologists. Stream temperature prediction accuracies were compared with the test-set root mean square error (RMSE) for all methods. For the actual data, SSN

had the highest predictive accuracy overall, which was followed closely by GBM and GAM. LM had the poorest performance overall. This study shows that although SSN appears to be the most accurate method for stream temperature prediction, machine learning methods and GAM may be suitable alternatives.

ACKNOWLEDGMENTS

I would like to thank my committee members, Dr. Richard Cutler, Dr. Adele Cutler, and Dr. Charles Hawkins for their thoughtful advice and mentoring. I would like to thank the School of Graduate Studies and the Utah State University Department of Mathematics and Statistics for funding. I would especially like to thank my family members and friends for their moral and emotional support throughout my degree program.

Maike F. Holthuijzen

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

INTRODUCTION

The integrity of freshwater ecosystems, particularly stream networks, is strongly influenced by water temperature (Caissie 2006). Water temperature controls biological processes and influences species distributions and aquatic biodiversity (Hawkins et al. 1997, Hill and Hawkins 2014). In particular, maintaining stream temperatures within acceptable limits is crucial to the fitness of ichthyofauna and other aquatic biota (Caissie 2006, Isaak et al. 2012, Turschwell et al. 2016). Climate change and other anthropogenic impacts (e.g. dams, riparian vegetation removal, livestock grazing) can alter the thermal regime of rivers and streams, resulting in direct and indirect impacts on aquatic biota (Isaak et al. 2010, Piccolroaz et al. 2016) and potential losses of aquatic biodiversity (Heino et al. 2009, Isaak et al. 2012). Due to climate change and other anthropogenic impacts, the thermal regimes of rivers may change in the future, and our ability to accurately model stream temperatures will be critical in understanding and predicting distribution shifts and dynamics of aquatic biota as stream temperatures change (Gardner et al. 2003). Furthermore, accurate methods of stream temperature prediction will also save monitoring effort and time (Yuan 2004, Hawkins et al. 2010, Hill et al. 2013), and aid in creating restoration plans for aquatic ecosystems (Isaak et al. 2012).

It is imperative that managers and freshwater biologists have access to accurate and accessible methods of stream temperature prediction that perform well at local (individual stream reaches) to regional (entire stream networks spanning 100's of km^2) scales. Several types of modeling approaches (stochastic, deterministic, and mechanistic)

have been used to predict stream temperatures. From a statistical point of view, modeling stream temperature presents a challenge because 1) stream temperatures along a stream network are often autocorrelated (Peterson et al. 2006), 2) the relationships between stream temperature and predictor variables may not be linear (Cressie et al. 2006), and 3) factors may interact to influence stream temperature (Caissie 2006). Gardner et al. (2003) highlighted the use of kriging, a geostatistical method for spatial interpolation that accounts for spatial autocorrelation, for predicting stream temperatures over a stream network. However, the implementations were not conducted with known software packages, and the predictions were unreliable due to the use of invalid autocovariance functions (e.g. positive definite covariance matrices were not obtained) (see Ver Hoef et al. 2006). Non-spatial linear regression, which does not account for temporal or spatial autocorrelation, has also been used to model stream temperatures (Crisp and Howson 1982, Mohseni et al. 1998); however the temperature predictions based on non-spatial regression have been shown to be less accurate than spatial methods (Isaak et al. 2010, Turschwell et al. 2016). Finally, two machine learning methods, artificial neural networks (Chenard and Caissie 2008, DeWeber and Wagner 2014) and Random Forests (Hill et al. 2013, Turschwell et al. 2016) have also been used for stream temperature prediction over the entire United States and over several stream networks in Australia.

More recently, spatial statistical network (SSN) models (Ver Hoef et al. 2006) were developed to incorporate hydrological distance and spatial autocorrelation among observed temperature sites over a stream network. Although SSN models use kriging methods, the predictions are thought to be reliable because valid autocovariance functions

were obtained for these models (Ver Hoef et al. 2006). SSN models have been shown to be superior predictors than Random Forests (Turschwell et al. 2016) and non-spatial linear regression (Isaak et al. 2010) under the few circumstances compared to date. SSN models have been used to predict daily, weekly, and monthly mean temperatures over watershed (Isaak et al. 2010, Turschwell et al. 2016) and state- and region-wide scales (Isaak et al. 2013, Detenbeck et al. 2016). SSN models work best when data are spatially autcorrelated over a stream network and there are enough observations for distance calculations (for kriging) (Peterson et al. 2013). Ver hoef and Peterson et al. (2014) developed a GIS toolbox for preprocessing stream network data and the SSN package in R (R Core Team 2016) for modeling the data. However, the pre-processing step can take a considerable amount of computation time and requires users to have strong GIS skills in ArcMAP (ESRI 2011) (Peterson et al. 2013, Isaak et al. 2014). Additionally, model-fitting with the SSN package in R becomes computationally prohibitive for datasets with more than 2000 data points (Ver Hoef et al. 2014), which is problematic for prediction of stream temperature on larger scales and on watersheds that have been intensely sampled. Furthermore, SSN models cannot inherently model non-linear associations, and the application of SSN models and interpretation of results requires a sophisticated understanding of statistical theory.

It is also not yet clear how much more accurate SSN models are than other predictive methods, such as machine learning and general additive models. A few studies have used machine learning methods for stream temperature prediction (Chenard and Caissie 2008, Hill et al. 2013, DeWeber and Wagner 2014), but only one compared

prediction accuracy with SSN models (Turschwell et al. 2016). Thus, a comparative study among a variety of statistical prediction methods including machine learning methods, general additive models, and SSN models is warranted. In this study, we provide comparisons of stream temperature prediction accuracy among SSN models and four other modeling methods: multiple linear regression, generalized additive models, Random Forests, and gradient boosting machines. We chose to include linear models because they are the non-spatial counterpart of SSN models, are familiar to many users, and are routinely used for a variety if modeling applications. Generalized additive models, Random Forests, and gradient boosting machines represent classes of predictive methods that are increasingly used in ecology and natural resource disciplines (Cutler et al. 2007, Olden et al. 2008). We did not include artificial neural networks, because we wanted to evaluate models that could be interpreted in terms of predictor variable strength and behavior (Olden and Jackson 2002). Random Forests and gradient boosting machines can handle complex interactions and non-linear data structure (Cutler et al. 2007, Olden et al. 2008), and generalized additive models excel in modeling non-linear associations (Drexler and Ainsworth 2013). Because these three methods have strengths in modeling complex data structures, they are promising for use in stream temperature prediction and could potentially rival the predictive accuracy of SSN models, which are currently considered to be the state-of-the-science for statistically-based stream temperature prediction (Ver Hoef et al. 2006).

We had three main objectives:

1. Quantify the accuracy (using root mean square error) of linear models (LM), spatial statistical models (SSN), Random Forests (RF), gradient boosting machines (GBM), and general additive models (GAM) in predicting observed maximum weekly maximum temperature, mean summer stream temperature, and mean August stream temperature.
2. Conduct simulations to determine how all five methods perform with known linear and nonlinear spatial data.
3. Develop guidelines for choosing predictive methods for stream temperature data based on data attributes and user expertise.

METHODS

Study sites and data

We used data from two stream networks in the Boise and Clearwater National Forests in Idaho, USA to model stream temperature. The Boise dataset was originally used by Isaak et al. (2010), whereas the Clearwater data had not been previously modelled.

Response variables for both datasets included measures of mean monthly and weekly stream temperatures. Specifically, response variables for the Boise dataset were maximum weekly maximum stream temperature (Mwmt) and summer mean stream temperature (SummerMean). The response variable for the Clearwater dataset was August mean stream temperature (Stream_Aug). Mwmt was obtained by calculating the highest seven-day moving average of daily maximum stream temperatures. In both stream networks, stream temperatures were recorded hourly with digital thermographs (Tidbit™ devices) that were placed in streams mid-July and retrieved in mid-September (the summer period) (Isaak et al. 2010).

The Boise data were collected between 1993 and 2006, whereas Clearwater data were collected between 1993 and 2011. In both stream networks, the majority of data were collected after 1999. Both datasets contained several years' worth of data for almost all monitoring sites, although the number of years of observation was not consistent across sites. To avoid weighting some sites more than others and temporal pseudoreplication, we eliminated duplicate observations at a site, similar to Detenbeck et

al. (2016). For sites with multiple observations, we randomly selected one year to model from each site. Doing so reduced the size of the Boise dataset from 780 to 506 observations and the Clearwater dataset from 4487 to 746 observations. Validation sets were created by randomly selecting 1/3 of the observations from each original dataset. The final Boise data thus contained 340 training and 146 test observations, and the Clearwater dataset contained 526 training and 220 test observations. Prior to modeling, we inspected histograms and boxplots of all variables for testing and training sets of both datasets to ensure that distributions were similar for training and testing datasets.

GIS pre-processing

To implement the SSN model for a stream network, the network must be continuous,, as this is a requirement for the calculation of the moving average function necessary for tail-up SSN models (Ver Hoef et al. 2014). We used ArcMap 10.2 (ESRI 2011) to construct the stream network. Stream network shapefiles were downloaded from the National Stream Internet (NSI), and waterbody polygons were downloaded from NHDPlusV2. Reach catchment areas (RCAs) were delineated from the 30m National Elevation Dataset (NED) digital elevation model (DEM). Calculations of stream distances (e.g. separation measured along the path of the stream (Rushworth et al. 2015) were also conducted in ArcMap 10.2 with the STAR extension (Peterson and Ver Hoef 2014). The resulting dataset was exported as a Spatial Stream Network (.SSN) object for analysis in R. The spatial weights needed to produce a positive-definite covariance matrix were based on watershed contributing area, a proxy for stream size (Ver Hoef et al.

2014). Hydrologic distance and spatial weights matrices were calculated with the SSN package in R (Ver Hoef et al. 2014).

Covariate descriptions

Boise data

Covariates for the Boise dataset consisted of 11 geomorphic and climatic predictors that had been previously shown or were hypothesized to influence stream temperature (Isaak et al. 2010). We provide general descriptions of them here, but note in some cases precise definitions (including units of measure) were not given. According to Isaak et al. (2010), geomorphic predictors were quantified from digital map layers in ArcMap 9.2. Isaak et al. (2010) previously analyzed a digital elevation model (DEM) with TauDEM software to summarize the six geomorphic predictor variables: watershed contributing area (Carea) was used as a proxy for stream size, network drainage density (Draind), elevation in meters (Elevation), percent of the catchment as glaciated stream valley (Gvalley), channel slope (Slope) as percent, percent catchment area classified as open water (Lake), and the percent of the catchment classified as alluviated valley bottom (Valleyb) (Isaak et al. 2010). We interpret the variable Valleyb to refer to the percent of the catchment that consisted of alluvial deposits in valley bottoms. Higher values would imply higher likelihood of ground water inputs to the stream. Drainage density is the length of stream in a watershed divided by the area of the watershed. Isaak et al. (2010) interpret this variable as quantifying the amount of stream channel exposed to solar radiation per unit area. As the amount of open water increases, larger areas of standing

water are exposed to solar radiation, which would imply a positive impact on stream temperature. Increasing values of Gvalley imply that more area of the catchment receives heavy snowfall and implies a negative relationship with stream temperature. All geomorphic variables were calculated within 1-km of monitoring sites on the stream network.

The four climatic predictor variables were: solar radiation (Rad), summer stream discharge (MnSummerFl), maximum weekly maximum air temperature (airMwmt), and summer mean air temperature (AirSummerMn). Solar radiation (Rad) was quantified by combining satellite imagery of vegetation and above-stream canopy photos. Specifically, Thematic Mapper satellite imagery was used to classify riparian vegetation, and later results were linked to field measurements of radiation at the stream surface (Isaak et al. 2010). Solar radiation was measured with hemispherical canopy photography, and the resulting photo film was used to calculate an index of total solar radiation. Estimates were matched with vegetation classifications from satellite imagery, and power-law relationships were used to predicted total radiation from vegetation type and catchment area (Isaak et al. 2010). The variables maximum weekly maximum air temperature (AirMwmt) and summer mean air temperature (AirsummerMn) were calculated from data obtained from NOAA weather stations near the study area (Arrowrock, Idaho City, and Ketchum, Idaho). Estimates of summer stream discharge (MnSummerFl) were derived from two USGS gages near the study area (Twin Springs and Featherville). Geographic location (Easting and Northing, UTM 11N, NAD 83) were also included as

variables. More detailed information on covariates is available in Appendix A5. For clarity, we chose to change variable names from those used by Isaak et al. (2010).

Clearwater data

Geomorphic and climatic predictor variables were similar to that of the Boise dataset and were chosen based on their likely influence on stream heat budgets for the Clearwater data (Isaak et al. 2016). Extraction of the covariates via NHDPlus V1, NSI, and other national databases was previously completed. The six geomorphic predictors included elevation (m) (Elevation), cumulative drainage area (km^2) (Cumdrainag), stream slope (percent) (Slope), base flow index (Bfi), percent glaciated valley (percentage of the catchment area classified as glacier) (Glacier), and percent lake (percentage of catchment area classified as open water) (Lake). Elevation values were obtained from a 30-m DEM associated with NHDPlus V1. Values for cumulative drainage and stream slope were also downloaded from NHDPlus V1. Glacier percentages were calculated with a standard flow accumulation routine in a GIS (data were downloaded from http://glaciers.research.pdx.edu/Downloads). One additional categorical variable coded as 0/1 (Dam_effect) was included to indicate whether a stream monitoring site was downstream from a reservoir, possibly creating anomalously cold tailwater. Canopy shade (percent) (Canopy) was used as a measure of stream shading and was compiled from the 2001 version of the National Land Cover Database (Homer et al. 2015). The three climatic variables included mean annual precipitation in mm (Precip), mean August air temperature (Air_Aug), and mean August stream discharge (Flow_Aug). Precipitation

was downloaded from NHDPlus V1 as the area weighted mean annual precipitation at bottom of flowline in mm. Mean August stream discharge was calculated by averaging across USGS flow gages with long-term records (data was downloaded from the National Water Information System website (USGS 2016)). Mean August air temperature for a river basin was obtained from the dynamically downscaled NCEP RegCM3 reanalysis (Hostetler et al. 2011) (data were downloaded from the USGS Regional Climate Downscaling website: http://regclim.coas.oregonstate.edu/). Measures of latitude and longitude in decimal degrees were included as Y_coord and X_coord, respectively. Values for all spatial covariates were assigned to 1-km reaches throughout the NSI network (Isaak et al. 2017). More detailed information on covariates is given in Table 1.

TABLE 1. Meanings and definitions for predictor variables in Boise and Clearwater datasets.

Boise		
Variable	**Definition**	**Meaning**
Elevation	Elevation in meters	Cooler air temperature and snowpack at high elevations have a negative impact on stream temperature
Draind	Drainage density	Provides a measure of the portion of stream exposed to solar radiation. Larger values have positive influence on stream temperature.
Carea	Contributing area	A proxy for stream size. Larger values have a positive influence on stream temperature
Gvalley	Percent glaciated valley	Percent of catchment area defined as glacier. Heavy snowfalls in these valleys should have a negative influence on stream temperature. On a scale of 0 to 1.
Valleyb	Alluviated valley bottom	Cool recharge water from aquifers has a negative impact on stream temperature. On a scale of 0 to 1.
Lake	Percent lake	Percentage of catchment area classified as open water; positively influences stream temperature. On a scale of 0 to 1.
SummerMnFl	Summer stream flow	Determines the volume of water available for heating. Larger values have a negative impact on stream temperature
AirSummerMean	Mean summer air temperature	Larger values have a positive impact on stream temperature
Slope	Channel slope	Steeper slopes result in fast-flowing stream. Larger values have a negative impact on stream temperature
AirMwmt	Maximum weekly maximum air temperature	Larger values have a positive impact on stream temperature
Rad	Solar radiation	Larger values have a positive impact on stream temperature
Easting		Measure of longitude

(Table continues)

Northing		Measure of latitude
Clearwater		
Variable	**Definition**	**Meaning**
Elev	Elevation at a stream site (m)	Air temperatures are cooler at higher elevations, representing a cooling effect on stream temperature.
Cumdrainag	Cumulative drainage area (km^2)	Represents the size of a stream; larger streams are likely to be exposed to greater amounts of solar radiation, representing a warming effect.
Slope	stream slope (percent)	the steeper a stream is, the faster the water flows, and the less it is able to absorb radiation. Higher values have a negative impact on streams.
Lake	Percent lake near a site	Larger values indicate greater portions of the stream are classified as open water; represents a warming effect. On a scale of 0 to 1.
Glacier	Percent catchment classified as glacier near a site	Larger values indicate more portion of the stream catchment is glaciated, which may cool stream temperatures. On a scale of 0 to 1.
Bfi	Base flow index	Streams with larger baseflows may be colder than other streams and less sensitive to climate warming.
Canopy	Percent of river shaded near a site	Streams with more shade receive less solar radiation; larger values negatively impact stream temperatures.
Flow_Aug	Mean August stream discharge	May or may not have an effect on stream temperature.
Air_Aug	Mean August air temperature, °C	Larger values have a warming effect on stream temperature.
Y_Coord	Latitude	
X_Coord	Longitude	
(Table continues)		
Precip	Annual precipitation (m)	Higher values have a negative impact on stream temperature.

Dam_effect	Effect of cold tailwater	Takes a value of 1 or 0 depending if a site is downstream from a dam or reservoir

STATISTICAL METHODS

We used five statistical methods for modeling stream temperature data: linear regression models (hereafter LM), spatial stream network (SSN) models, generalized additive models (GAM), Random Forests (RF), and gradient boosting machines (GBM). All analyses were conducted in R 3.3.1 (R Core Team 2016). Non-spatial models were fitted twice: once with and once without spatial covariates (e.g., Easting/Northings, Latitude/Longitude). Thus, for the two response variables for the Boise data (Mwmt and SummerMean), we obtained 2 final models for each of the non-spatial methods (2 response variables x 2 rounds of analyses x 4 methods = 16 final models). Similarly, there were 8 final non-spatial models for the Clearwater dataset, which only had one response variable (Stream_Aug). Since SSN models account for spatial attributes, each SSN models were fitted once for each response variable.

We considered interactions between variables previously used by Isaak et al. (2010) for LM, SSN, and GAM methods (Carea*Rad, AirMwmt*Rad, AirMwmt*MeanSummerFl, Elevation*Gvalley, and AirMwmt*Rad*MeanSummerFl), as well as the interaction between measures of latitude and longitude. We used the root mean square error (RMSE) of the test data sets to quantify the performance of each method:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}$$

where Y_i is the observed value of the response variable for the i^{th} observation, and $\hat{Y}_i$ is the predicted value for the i^{th} observation.

Linear Regression Models

In the multiple linear regression model (see, for example Seber and Lee, 2003), stream temperatures are related to a number of predictor variables through a linear formula with an error (residual) term. Mathematically, we write

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots \beta_p x_{ip} + \epsilon_i$$

for $i = 1,2, \ldots, n$ where Y_i is the stream temperature at the i^{th}location, x_{ij} is the value of the j^{th} predictor variable for the i^{th} observation, and ϵ_i is the error term for the i^{th} observation.

LM model fitting

We fit linear regression models using the `lm` function that is part of the base distribution of R. After inspecting scatter plots of predictor versus response variables for linearity,we applied the arcsin-square root transformation to Valleyb, Gvalley, and Lake. To avoid bias and overfitting, we used the lasso technique with 10-fold cross validation on the training set to perform variable selection for LM. The lasso technique is a shrinkage method that also performs variable selection and often results in sparse models (James et al. 2013). We performed the lasso technique using the GLMSELECT procedure in SAS (Version 9.4 of the SAS System for Windows). We set the SELECTION argument to 'LASSO' and specified external 10-fold cross validation with the optional

parameter CHOOSE = CVEX (SAS Institute 2015). After obtaining final predictor subsets for both response variables Mwmt and SummerMean, we calculated the RMSE for each model on the test dataset.

For the Clearwater data, we applied natural log transformations to both SLOPE and BFI. Clearwater models were constructed in the same way as described above, except we only considered the interaction between latitude and longitude for the Clearwater models.

Three important assumptions of the linear regression model are:

1. *Independence*: the error terms, ε_i, are statistically independent.
2. *Linearity*: the relationship between the predictor variables is linear as opposed to non-linear (but possibly additive).
3. *Limited interactions*: there are limited interactions among the predictor variables in terms of their effects on the response variable.

Spatial Stream Network Models

Spatial stream network models (Ver Hoef et al. 2006) address the independence assumption of LM issue by relaxing this assumption. SSN models take into account covariance structure of temperature data on a river network, allowing for the unique properties of stream networks such as branching structure, longitudinal connectivity, directed flow, and abrupt temperature changes at stream junctions (Isaak et al. 2014). Specifically, the vector of error terms in the multiple linear regression model, $\boldsymbol{\varepsilon}$, is decomposed into

$$\varepsilon = \boldsymbol{v}_{\mathrm{TU}} + \boldsymbol{v}_{\mathrm{TD}} + \boldsymbol{v}_{\mathrm{EU}} + v_{\mathrm{NUG}}$$

where $\boldsymbol{v}_{\mathrm{TU}}$ captures upstream ("tail up") autocorrelation, $\boldsymbol{v}_{\mathrm{TD}}$ captures downstream ("tail down") autocorrelation, $\boldsymbol{v}_{\mathrm{EU}}$ characterized autocorrelation structures, and v_{NUG} is a "nugget" effect due to, for example, the confluence of streams. The nugget effect may be thought of as sampling error for spatial data (Ver Hoef et al. 2014). More detailed information about SSN can be found in Appendix B.

SSN model fitting

We used the `SSN` package (Ver Hoef et al. 2014) in R v. 3.3.1 for analysis. First, we calculated distance matrices as specified in (Ver Hoef et al. 2014) and imported the prediction (test set) observations. For both data sets, we used the same transformations of predictor variables for SSN models as we did for LM. For each of the two response variables, Mwmt and SummerMean, we constructed models containing the subset of variables used for the linear models (without spatial covariates). We considered models with all combinations of Euclidian, tail up, and tail down covariance structures. Tail-up and tail-down covariance structures within the SSN package in R include the Spherical, Mariah, Empanovich, Linear-with-Sill, Gaussian, and Cauchy, while the options for Euclidean covariance structures are Gaussian, Spherical, Exponential, and Cauchy (Ver Hoef et al. 2014),. To determine which combination of covariance structures was best, we used the compareSSN function, which computes AIC scores and 10-fold cross validated prediction errors for a user-defined subset of candidate models (Ver Hoef et al. 2014). We chose the model with the lowest cross-validated prediction errors and used it to make

predictions on the test set. We followed the same methods for modeling the two response variables in the Boise dataset and the response variable in the Clearwater dataset.

Generalized Additive Models

GAM (Hastie and Tibshirani 1986) assume independent observations but relax the assumption of linearity in individual predictor variables to forming a sum of continuous functions of individual predictor variables. For the situation with normal errors, the GAM model may be written as:

$$Y_i = s_0 + s_1(x_{i1}) + s_2(x_{i2}) + \cdots s_p(x_{ip}) + \varepsilon_i$$

where s_0 is a constant and $s_1(), s_2(), \cdots, s_p()$ are smooth but unspecified functions of the respective predictor variables. Choices for smoothers to estimate $s_1, s_2, \cdots, s_p$ include local regression (loess), smoothing splines and regression splines (B-splines, P-splines, and thin-plate splines) (Wood 2006). Unsmoothed, linear effects may also be included in a GAM, resulting in a semi-parametric model. Prespecified interactions among two or, possibly, even three variables may be incorporated as well. The estimation of the unknown functions (s_j) in GAMs is by penalized maximum likelihood through the P-IRLS (penalized iteratively reweighted least squares) algorithm (Wood 2006). However because the model fitting process involves estimating the degree of "wiggliness" of the smoothed functions, the *generalized cross validation score (GCV)* or *Un-Biased Risk Estimator (UBRE),* is used to determine the optimal degree of smoothing (Wood 2006).

The main limitation of GAMs is their strain on computational resources, which limits their use on very large datasets (Friedman et al. 2001). Also, GAMs do not model interactions automatically in contrasts to tree-based methods. The user must determine which interactions to include in the model. Finally, extrapolation with GAMs may be problematic, especially if data are scarce at the endpoints (Wood 2006).Since their introduction, GAMS have been used extensively in ecology (Guisan et al. 2002), often for modelling species distributions of both plants and animals (Fewster et al. 2000, Austin 2002, Leathwick et al. 2006b, Moisen et al. 2006, Drexler and Ainsworth 2013). GAMs can be implemented in R with the package mgcv (Wood 2013) or the package gam (Hastie and Tibshirani 2013).

GAM model fitting

We used the mgcv (Wood 2006) package in R for all GAM analyses. For the Boise data, we first fit additive models (e.g. without interaction terms) with all predictor variables (with and without spatial covariates). We assessed residual versus individual predictor plots, residual q-q plots, and histograms of residuals to ensure normality and homogeneity of residual variance. We fit scale invariant tensor product interactions (Wood 2006) to model interaction terms used by Isaak et al. (2010). All predictor variables (except interaction terms) were modeled with penalized regression splines. To perform variable selection for GAM models, we used a shrinkage approach developed by Marra and Wood (2011), which can be invoked in the mgcv package by setting select = TRUE within the gam function. Smoothed effects that do not influence the response

variable are effectively set to zero (Marra and Wood 2011). The shrinkage approach was used with 10-fold cross validation. Because the shrinkage method resulted in all variables being retained in the model, we used the summary results from the 10-cross-validation folds to determine which, if any, predictors were consistently not significant (P > 0.1). Predictors that were consistently not significant were eliminated from the model, and cross-validation was again performed. This process was repeated until the cross-validated error began to increase.

Random Forests

Random Forests (Breiman (2001a) is a machine learning method for regression and classification that fits an ensemble of decision trees to data and combines the predictions from the trees to produce more accurate predictions.

The basic algorithm is as follows. Many samples are drawn from the original dataset. Observations that are in the original dataset but not in a given sample are said to be out-of-bag for that dataset. Decision trees are fit to each sample, usually without pruning, and with only a randomly selected subset of variables available for splitting at each node. Predictions are made for every observation that is out-of-bag for the sample a given tree is fit on, and then, in the case of regression, averaged over all the trees for each observation to produce a more accurate prediction. In the randomForest package in R (Liaw and Wiener 2002) the default number of samples (and hence fitted trees) is 500. The number of randomly selected variables available for splitting at each node in regression is $p/3$, where p is the total number of predictor variables, and in classification

it is $\sqrt{p}$, although these parameters should be tuned for optimal performance (Friedman et al. 2001). A detailed explanation of the algorithm is given by Breiman (2001a) and (Cutler et al. 2007). The impact of individual predictor variables on the response variable may be visualized using partial dependence plots (Friedman et al. 2001) which show the relationship between a predictor variable with the response variable, averaged over all predictor variables.

RF has received considerable attention as a predictive methodology for a wide range of ecological applications, from mapping tree distributions (Prasad et al. 2006) to predicting and classifying plant and animal species (Cutler et al. 2007) as well as predicting wildfire occurrence (Oliveira et al. 2012). Hill et al. (2014) used RF to model the vulnerability of stream networks to climate change, and Hill et al. (2013) created a temperature prediction model for the entire continental United States. Although relatively few studies have explicitly used RF in the context of stream temperature prediction (Hawkins et al. 2010, Hill et al. 2013, Hill and Hawkins 2014, Turschwell et al. 2016), the ability of RF to fit data with many predictor variables, nonlinear effects, and high order interactions (Cutler et al. 2007) make them an excellent candidate for more extensive use in stream temperature prediction.

RF model fitting

All RF regressions were fit using the randomForest package in R 3.3.1 (Liaw and Wiener 2002) with the default number of trees/samples (500) and the default numbers of variables available for splitting at each node ($p/3$). After fitting full models for each

dataset with and without measures of latitude and longitude, we inspected variable importance plots and refit models with subsets of important predictor variables. There are no P-values associated with RF, and variable selection via inspection of variable importance plot is subjective. To choose the best models, we tried fitting models with successively smaller variable subsets and eliminating variables based on the variable importance plots. We inspected the OOB error rate in each successive model and chose the one with the lowest OOB error rate as the final model.

Gradient boosting machines

Gradient boosting machines (Friedman 2001) is another machine learning ensemble classification method. Although the GBM algorithm is quite general it is frequently implemented with regression and classification trees. The algorithm works by sequentially fitting trees to the residuals from previous fits. In many applications, GBM has been found to be one of the most accurate classifiers (see, for example, Friedman, 2001). For more details of the GBM algorithm, see (Friedman 2001) or (Friedman et al. 2001). As with RF, the relationships between the response and individual predictor variables may be characterized using partial dependence plots (Friedman 2001). As with RF, GBM is able to deal with complex, high-order interactions among predictor variables. Finally, boosting supports the use of different loss functions, which, for regression includes the Huber loss function, MSE, and others (Friedman 2001).

GBM has been used in a variety of ecological studies for predicting species richness (Leathwick et al. 2006a), abundance (De'Ath 2007) and classification (Cappo et

al. 2005). In the R implementation, the parameters for gradient boosting machines are the number of trees, the shrinkage value, the interaction depth, and the minimum number of observations in terminal nodes. Optimal predictive power can be achieved by tuning these parameters, which is commonly done via a grid search. In a grid search, a statistical model is run with combinations of parameters and the cross-validated error rate is computed. The combination of parameter values resulting in the lowest cross-validated error rate is either chosen for the final model, or the grid search can be performed several times with successively finer-tuned parameter values.

GBM model fitting

All GBM models were fit using the gbm package in R (Ridgeway 2015), and the model fitting process was very similar to that for RF. After fitting full models with and without measures of latitude and longitude, we inspected variable importance plots. For GBM, we chose to use all variables for models with and without spatial covariates. By default, the parameters of the gbm function (number of trees, shrinkage value, the interaction depth, and the minimum number of observations in terminal nodes) were set at 500, 0.001, 1, and 10, respectively (Ridgeway 2013). To achieve optimal performance, we tuned these four regularization parameters with a grid search with the caret package (Kuhn 2016). Optimal tuning parameters are shown in Table 2.

TABLE 2. Tuning parameters for final GBM models for all response variables.

Response variable	No. trees	Shrinkage	Interaction depth
	Without Lat/Long		
Boise: Mwmt	1000	0.01	10
Boise: SummerMn	3000	0.01	8
Clearwater: Stream_Aug	10000	0.01	14
	With Lat/Long		
Boise: Mwmt	1000	0.01	12
Boise: SummerMn	3000	0.01	6
Clearwater: Stream_Aug	8000	0.1	18

SIMULATIONS

Non-spatial simulations

Non-spatial simulations were conducted with datasets representing all combinations of 1 and 2 variables, nonlinear and linear structure, and autocorrelated or independent error structure. Datasets had either independent error structure (with $\sigma^2 =$ 4 or 9) or autocorrelated error structure (with ρ = 0.7 or 0.8, with ρ being the autocorrelation parameter). Each dataset contained 200 observations. To create linear data for the one variable model, we generated 200 values of x_i from a $Uniform(-3, 3)$ distribution with the runif function in R. To create independent error terms, we generated 200 values from a $Normal(0, \sigma^2)$ distribution (with $\sigma^2 = 4$ or 9) using rnorm. Autocorrelated errors $\varepsilon_i$were generated sequentially using the following algorithm:

For i in 2 through the number of observations:

$$\varepsilon_i = \rho * \varepsilon_{i-1} + \sqrt{1 - \rho^2} * V_i,$$

where V_i is a vector of values from a $Normal(0, \sigma^2)$ distribution.

Then, we have

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ \ for \ \ i = 1, 2, \cdots, n.$$

Constructing a two variable, linear dataset was identical to the construction of the one-variable dataset, except that two variables, X_1 and X_2 were included.

The nonlinear, one variable dataset was a log-linear model, and the nonlinear, two variable dataset was an additive model with a log-linear component and squared term (Table 3). The choice of non-linear data structure was based on known relationships between stream temperature and predictor variables at weekly or monthly levels of temperature aggregation (relationships between predictor variables and stream temperature can differ depending on the time period over which stream temperature is expressed (Caissie 2006, Turschwell et al. 2016). The association between stream temperature and air temperature, in particular, are generally not linear and have been previously modeled via logistic regression (Caissie 2006, Mayer 2012, Arismendi et al. 2014). Hill et al. (2013) also found logistic relationships best described relationships between air temperature and mean summer, winter, and annual stream temperatures. Others have observed an exponential relationship between maximum weekly average stream temperature and catchment area (Friele et al. 2016).

TABLE 3. Linear and nonlinear data models for non-spatial simulations

	Nonspatial Simulation Models	
	Linear	*Nonlinear*
1 var.	$Y = \beta_0 + \beta_1 x_1 + \varepsilon$	$Y = \beta_0 \left(e^{\beta_1 x_1}\right) + \varepsilon$
2 var.	$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$	$Y = \beta_0 \left(e^{\beta_1 x_1}\right) + \beta_2 (x_2)^2 + \varepsilon$

For all models, $\beta_0 = 1$, $\beta_1 = -1$, and $\beta_2 = 1.9$. Test datasets consisting of 200 observations each were created for computing RMSE values for each type of training dataset and predictive method. All non-spatial methods (LM, RF, GBM, GAMs) were fit to 50 datasets several times. Each of the 50 datasets represented a unique combination of linear/nonlinear structure and independent/autocorrelated error structure (where ρ equal to 0.7 or 0.8 and $\sigma^2 = 4$ or 9), and number of predictor variables. Next, predictions were made for each of the 50 training datasets onto the 50 test datasets for each method. The resulting 50 predictions were used to calculate 50 RMSEs for each method. Finally, the 50 MSEs were averaged to obtain one mean RMSE per method per data structure combination. We also calculated the mean resubstitution error and 10-fold cross validation error to compare with the test-set RMSE. Model fitting and calculations of RMSEs were done with a custom function in R. Additionally, we made use of the packages randomForest (Liaw and Wiener 2002), gbm (Ridgeway 2013), mgcv (Wood 2013), purr (Wickham 2016b), and dplyr (Wickham 2016a). GBM models were tuned using one sample training dataset.

Spatial simulations

Spatial simulations were conducted in the SSN package (Ver Hoef et al. 2014) in R. The createSSN function creates the .SSN object necessary for use in the function glmssn, which fits generalized linear models with spatially autocorrelated errors (Isaak et al. 2014, Ver Hoef et al. 2014). The createSSN function generates an artificial network system, placing training and testing observations on the branches of the artificial network.

To make the simulations realistic, we simulated large artificial networks with 200 branches, ~240 training sites and ~170 testing sites. The algorithm for creating the artificial network was kept at the default (igraphKamadaKawai), and training/testing observations were distributed across the network branches using the hardcoredesign algorithm with parameters 300 (number of maximum training observations), 200 (number of maximum testing observations), and 0.2 (the inhibition region). We performed all simulations with two predictor variables, X_1 and X_2. Although more observations would have been more realistic, adding more sites was not feasible given the large computation times required to produce two variable models.

We used the function simulateonSSN to simulate temperatures on the artificial stream network based on the type and strength of autocorrelation, model formula (which specifies a linear or nonlinear data structure), and coefficients (β_0, β_1 and β_2). The coefficients β_0, β_1, and β_2 were set at 1, -1, and 1.9, respectively. We used the exponential autocovariance model for both tail-up and tail-down autocorrelation. As in the non-spatial simulations, we wanted to use known non-linear relationships between stream temperature and predictors for the non-linear data structures. The relationship between air temperature and stream temperature can be modeled with a logistic function (Webb et al. 2008), which is why we choose the “non-linear logistic” equation in Table 4. The second non-linear data structure was constructed with the “non-linear exponential” equation in Table 4. We also constructed a linear data structure with two predictor variables as comparison.

We simulated each type of data structure (“logistic”, “exponential”, and “linear”) with strong and weak autocorrelation. The strength of autocorrelation depends on the partial sill and range, while a third parameter, the nugget, controls random variation. The partial sill was set at 2 for all simulations, the range varied from 2 (weak) to 10 (strong) autocorrelation, and the nugget was set at 0.01.

After we obtained the temperatures on the simulated stream networks, we used glmssn to fit an SSN model to each of the linear/nonlinear - autocorrelation combinations of data structure. We created a custom function to generate and fit 100 spatial datasets. Finally, after fitting models for each of the 100 datasets, we predicted onto the corresponding test datasets and calculated the average RMSE. Next, we used the training datasets to fit models using the other four methods (LM, RF, GBM, GAM), made predictions onto the corresponding test datasets, and computed the average RMSE for the 100 datasets. Simulations for fitting models with the four other methods were conducted with a custom function in R. Additional R packages used for fitting RF, GBM and GAM were randomForest, purrr, gbm, and mgcv.

TABLE 4. Linear and non-linear data models for spatial simulations. All models contained two variables.

Spatial simulation models
Linear $Y = \beta 0 + \beta 1 x_1 + \beta 2 x_2$
Nonlinear: "logistic" $Y = \beta 0 + \frac{25}{1 + e^{-2x_1}} + 2e^{-x_2}$
Nonlinear: "exponential" $Y = \beta 0 + e^{-x1} - 2{x_2}^2$

Timing of computations

Because we were interested in the amount of time the methods took to make predictions, we obtain estimates for the time each model took to compute a 50,000 observation dataset. We simulated a simple, linearly structured dataset with 50,000 observations and 5 predictor variables and applied RF to 5, 10, and 20 tree models. Then, we extrapolated to 500 tree models, which is the default number of trees. Computation time is linearly associated with the number of trees (Cutler 2017, pers. comm.). Next, we applied GAM to the dataset. We then applied SSN to datasets with 100, 500, and 1000 observations and then extrapolated these results to 50,000 observations. Computation times in SSN is linearly associated with the log of the number of observations. We used the microbenchmark function from the microbenchmark package (Mersmann 2015) in R

to time simulations. Computation time to predict stream temperatures for 50,000 sites ranged from 14 seconds (GAM) to 10 days (SSN). Both RF and GBM took 14 minutes to complete computations.

RESULTS

All methods except SSN/LM performed better with spatial covariates included in the models, so we report results only for those models. Overall, the models with the lowest test-set RMSE, regardless of dataset, were SSN and GAM (Table 5).

Boise River basin: Maximum weekly maximum stream temperature (Mwmt)

Overall, the SSN model with spherical tail-up, spherical tail-down and spherical Euclidean covariance structures (RMSE = 1.68) had the lowest test-set error. GAM and GBM also performed well with RMSEs of 1.78 and 1.85, respectively. RF and LM had the highest RMSE values (2.28 and 2.54, respectively). The LM/SSN models contained the fewest covariates, while GAM/GBM had the most; covariates for all models, including those without spatial covariates, are shown in Table A1.

Boise River basin: Summer mean stream temperature (SummerMn)

The SSN model achieved the lowest test-set RMSE (0.83) with Mariah tail up, Mariah tail down, and spherical Euclidean covariance structures. GBM and GAM performed nearly as well, with RMSEs of 0.85 and 0.91, respectively. RF and LM had the largest RMSE values (1.03 and 1.39, respectively). Test-set RMSEs for SummerMean were generally lower than for Mwmt. Similar to the Mwmt models, The LM/SSN models contained the fewest covariates, while GAM/GBM included the most; covariates for all models are shown in Table A2.

Clearwater River basin: August mean stream temperature (Stream_Aug)

The SSN model with linear-sill tail-up, linear-sill tail-down, and exponential Euclidean covariance structures outperformed all other models with an RMSE of 1.09. However, in contrast to the data from the Boise River basin, all methods achieved relatively good predictive performance for August mean stream temperature. RMSEs ranged from 1.16 (RF) to 1.79 (LM). GAM and LM had the largest RMSE values: 1.22 and 1.31, respectively. The LM model including spatial covariates contained only five covariates, while all other models included at least eight; covariates for all models are shown in Table A3.

TABLE 5. RMSE values for all final models for Boise and Clearwater response variables.

Method	Boise: Mwmt	Boise: SummerMn	Clearwater: STREAM_Aug
LM lat long	2.54	1.39	1.31
LM	2.6	1.41	1.4
RF lat long	2.03	1.03	1.16
RF	2.22	1.17	1.19
SSN	1.68	0.83	1.1
GAM lat long	1.78	0.91	1.22
GAM	2.34	1.26	1.34
GBM lat long	1.84	0.85	1.17
GBM	1.85	1.06	1.19

SIMULATION RESULTS

All simulations were conducted in under 10 hours on a laptop PC with an Intel™ i74810MQ CPU at 2.80 GHz.

Non-spatial simulations

Overall, GAM outperformed all other methods, regardless of error variance, degree of autocorrelation, and data structure. LM performed nearly as well as GAM, but only when data structure was linear and error variance was not autocorrelated. LM performed very poorly for all nonlinear data structures. GAM performed especially well when data structure was nonlinear and errors were autocorrelated.

RF and GBM outperformed LM only when data structure was nonlinear. RF almost always outperformed GBM, especially when data had a nonlinear structure. Several trends held true for all methods. First, methods performed better on one-variable datasets when errors were independent than when errors were autocorrelated. However, models performed slightly better on two-variable datasets when errors were autocorrelated than when errors were independent.

For linear datasets with independent errors, the test-set RMSE was lower for LM and GAM for both one variable models (2.03 and 2.03, for $\varepsilon = 4$, respectively) and two variable models (1.89 and 1.91 for $\varepsilon = 4$, respectively) than GBM and RF models (one variable RMSE = 2.28 and 2.36, for $\varepsilon = 4$, respectively and two variable RMSE = 2.08, 2.36 for $\varepsilon = 4$, respectively). However, for nonlinear data, RMSE values were lower for

one-variable RF and GAM models (2.24 and 2.02, for $\varepsilon = 4$, respectively) and two-variable RF and GAM models (3.03 and 2.09 for $\varepsilon = 4$, respectively) than LM (3.43 and 6.14 for $\varepsilon = 4$, respectively). RMSEs for GBM were slightly higher than those for RF (Table A1).

For one and two variable, linearly structure data with autocorrelated errors ($\rho = 0.7$), test-set RMSE values were lower for both one-variable (2.53 and 2.54, respectively) and two-variable LM and GAM models (1.90 and 1.92, respectively) than either one-variable (2.77 and 2.84 , respectively) or two-variable RF and GBM models (2.13 and 2.30, respectively).

For one and two-variable, non-linearly structured data with autocorrelated errors ($\rho = 0.7$), RMSEs for one variable GAM models were lowest (1.99), followed by RF, GBM, and LM (2.16, 2.58, and 3.43, respectively). The same trend was apparent for two variable GAM, RF, GBM and LM models (RMSE = 2.03, 2.66, 3.12, and 5.65, respectively). The results presented here are a summary of the main results, but all RMSE values can be found in (Table A4).

Spatial simulations

SSN models performed better than any other method only if the data structure was purely linear, regardless of the strength of autocorrelation. However, for nonlinear data structures (“logistic” or “exponential”, Table 3), GAM achieved the lowest RMSE values. RMSE values for RF and GBM were consistently greater than those for GAM, especially for non-linearly structure data. RMSE values for GBM and RF were very

similar. GBM outperformed RF by a small margin for the linear and exponential model, but the reverse was true for the logistic model (Table 6).

TABLE 6. RMSE values for all combinations of linear/nonlinear and strong/weak spatial autocorrelation for spatial simulations. The model with the smallest RMSE for each data structure is highlighted in bold.

Method	Linear, strong autocov.	Linear, weak autocov.	Exponential, strong autocov.	Exponential, weak autocov.	Logistic, strong autocov.	Logistic, weak autocov.
SSN	**1.57**	**1.81**	3.64	3.74	3.81	3.90
LM	1.93	1.96	3.79	3.80	3.87	3.89
RF	2.11	2.14	2.55	2.56	2.71	2.74
GBM	2.02	2.04	2.38	2.40	2.94	2.96
GAM	1.94	1.98	**2.12**	**2.14**	**2.09**	**2.14**

DISCUSSION

Our ability to accurately predict temperatures over stream networks is critical for studying the distributional shifts of aquatic organisms (Isaak and Rieman 2013) and for making conservation management decisions (Hill et al. 2013, Jones et al. 2014, Westhoff and Rosenberger 2016). Research that uses stream temperature predictions to investigate aquatic species' dynamics is urgently needed, as climate change related to human activity will continue to alter river thermal regimes (Welsh Jr et al. 2001). Predicting stream temperature presents a challenge because stream temperatures along a stream network are often autocorrelated (Peterson et al. 2006), the relationships between stream temperature and predictor variables may not be linear (Cressie et al. 2006), and factors may interact to influence stream temperature (Caissie 2006). The SSN method, especially, is promoted for use primarily because it is the most accurate method (Isaak et al. 2010). However, the degree of prediction accuracy necessary to achieve study objectives is not often considered. As biological phenomena are notoriously noisy, the accuracy of predictions may not help us understand general ecological trends in the biological entities we are studying. As we have shown in this study, machine learning methods and GAM are excellent alternatives to SSN because they provide accurate predictions and are more accessible, interpretable, and computationally efficient than SSN.

Computational demands and interpretability

In this study, we found that SSN models were most accurate among the methods evaluated for predicting stream temperature for actual data, but two machine learning

methods, GBM and GAM, also achieved high predictive accuracy. GAM performed particularly well in spatial and non-spatial simulations, when the relationships between the response and predictor variables were non-linear. The drawbacks of SSN include substantial data pre-processing requirements, high computational times, and limited ability to fit nonlinear data and interactions. Taken together, our results suggest that some machine learning methods, such as GAM, are viable alternatives to SSN that are accurate, computationally efficient, and are easy to automate for large datasets.

It has been argued that non-spatial methods that do not explicitly account for spatial autocorrelation cannot be used to derive valid statistical inference (Rushworth et al. 2015). However, we contend that, to some degree, apparent spatial autocorrelation may be due to non-linear associations between the response variable and the predictor variables. The high accuracies achieved by GAM and GBM could be due to the ability of both methods to fit nonlinear data structure or the ability of GBM to fit high-order interactions. Ultimately, the goal is to make accurate predictions (rather than parameter estimation). Our results suggest that machine learning and other non-spatial methods can provide predictions of stream temperature that are sufficiently accurate for many ecological purposes.

SSN models are computationally more intensive than methods described in this study, due to the calculation of covariance matrices (an n^2 operation) (Isaak et al. 2014, Rushworth et al. 2015). Computation efficiency is an important consideration as large amounts of stream temperature data is available online. In fact, datasets with more than 2000 observations would likely have to be processed on high-powered computing

facilities (Ver Hoef et al. 2014, Turschwell et al. 2016). Moreover, SSN models require data to be pre-processed in a GIS, a task which requires advanced GIS skills (Isaak et al. 2014). In contrast, RF, GBM and GAM are much more computationally efficient and do not require data to be pre-processed in a GIS.

Interpretability of models is important to model users wishing to infer causality from analyses. Machine learning methods, as well as GAM, have advantages over SSN in terms of interpretability. Non-machine learning methods (SSN, LM and GAM) are interpreted in terms of the sign and value of coefficients and p-values of coefficients. This traditional method of model interpretation is appealing because it is simple and relatively easy to explain to a wider audience. The SSN provides a slight advantage over LM in that information about the autocorrelation present in the data can be inspected visually with a Torgegram (similar to a variogram plot but specialized for stream network data) (Ver Hoef et al. 2014). However, the sole use of p-values and coefficients in interpretation is problematic, especially as p-values decrease as the number of observations increase (Breiman 2001b). Interpretability in SSN and LM is also hindered by the use of high order interaction terms, which nearly impossible to interpret in the context of the study (Wood 2006). GAM produces smoothing plots, which can be useful for determining the shape of relationships response and predictor variables (Wood 2006). Moreover, the additivity of GAM means these models include few, if any, interactions, adding to their interpretability (Hastie and Tibshirani 1986).

Although RF and GBM are usually described as being uninterpretable, “black box” methods (Lipton 2016), there are graphical methods applicable to both techniques

that assist with interpretation. Variable importance plots shed light on the number of variable to include in a GAM or RF. One- and two-variable partial dependence plots (Friedman et al. 2001, Friedman 2002) may be used to characterize the relationships between individual predictors or pairs of predictors and the response variable. These plots can be very useful for inferring ecological relationships (Cutler et al. 2007).

Accuracy

Our study shows that machine learning methods and GAM can, in fact, approach the accuracy of the SSN model. More accurate temperature predictions obtained with the SSN may be necessary to minimize the error in distribution models of thermally sensitive species. The thermal tolerance ranges of ectothermic aquatic species is important because it can be used to predict species' absence or presence and distributional shifts in response to climatic changes (Eaton and Scheller 1996). Although using species' thermal tolerance as a guide for determining the necessary prediction accuracy seems convenient, aquatic species do not always respond consistently to thermal changes (Isaak and Rieman 2013) either within species (Pörtner 2001) or within genera (Hildrew and Edington 1979). Thus, other methods from this study that are close in accuracy to the SSN model would be well suited for studies that rely on temperature predictions to make inferences about species' distributions.

The level of temperature aggregation is an important aspect in the study of river networks, as the thermal regime of rivers and aquatic species' distributions can be characterized by daily, weekly, monthly, or annual temperature metrics (Caissie 2006).

Temperature measures can affect biological inferences, because more aggregated measures (e.g. yearly, monthly) can be predicted with less error compared with less aggregated measures (weekly, daily) (Turschwell et al. 2016). Indeed, we found that RMSE values were lowest and very similar for all statistical methods in predicting mean monthly temperature measures compared to weekly measures. For studies investigating species' distributions and dynamics, more aggregated measures may be sufficient in predicting temperatures. The presence/absence of most freshwater species, including thermally sensitive salmonids, can be accurately predicted from mean monthly and even annual stream temperature (Buisson et al. 2008) (Keleher and Rahel 1996). Distributional patterns of freshwater trout species were closely related to mean July-August stream temperatures (Isaak and Hubert 2004). Finally, Hill and Hawkins (2014) used RF to predict mean summer stream temperature and model the presence of aquatic invertebrate species in 92 sites over the entire U.S with an RMSE of 1.9 °C. A large body of evidence shows that species' distributions can adequately be modeled by more aggregated stream temperatures, which means that non-SSN methods examined in this study (RF, GBM, and GAM) could be reliable prediction methods for most studies involving the prediction of aggregated temperature metrics.

Conclusions

Predicting stream temperatures accurately is important for gaining a better understanding of species' distributions as the global climate continues to change (Isaak et al. 2012), testing hypotheses about species' distributions (Hill 2013), and guiding

restoration and conservation decision-making (Hill and Hawkins 2014). Although we found that the SSN model was the most accurate predictor of stream temperature for two river networks in Idaho, two machine learning methods RF and GBM, as well as GAM, closely rivaled the accuracy of the SSN model. Predictions were especially close in predicting monthly or summer mean stream temperature measures. We also found that GAM (and machine learning methods, to a lesser extent) attained higher prediction accuracies than SSN when data had nonlinear structure, which suggests that these techniques would be better suited to most real-world data which rarely has linear structure.

Stream temperature prediction will become more important in the future, as species shifts due to climate change and other anthropogenic impacts will have long-lasting implications on the ecological functioning of aquatic environments. Thus, understanding the types of methods for stream prediction, their advantages, and disadvantages will be useful for scientists engaged in this research. The choice of stream temperature prediction method should depend on specific data characteristics, such as the presence of linear/nonlinear structure, number of observations and variables, and the availability of statistical and computational resources. Overall, machine learning methods and GAM provide advantages over the SSN model due to their prediction accuracy, computational efficiency, accessibility, and interpretability.

LITERATURE CITED

Arismendi, I., M. Safeeq, J. B. Dunham, and S. L. Johnson. 2014. Can air temperature be used to project influences of climate change on stream temperature? Environmental Research Letters **9**:084015.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological modelling **157**:101-118.

Breiman, L. 2001a. Random forests. Machine learning **45**:5-32.

Breiman, L. 2001b. Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science **16**:199-231.

Buisson, L., L. Blanc, and G. Grenouillet. 2008. Modelling stream fish species distribution in a river network: the relative effects of temperature versus physical factors. Ecology of Freshwater Fish **17**:244-257.

Caissie, D. 2006. The thermal regime of rivers: a review. Freshwater biology **51**:1389-1406.

Cappo, M., G. De'ath, S. Boyle, J. Aumend, R. Olbrich, F. Hoedt, C. Perna, and G. Brunskill. 2005. Development of a robust classifier of freshwater residence in barramundi (Lates calcarifer) life histories using elemental ratios in scales and boosted regression trees. Marine and freshwater research **56**:713-723.

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.

Cressie, N., J. Frey, B. Harch, and M. Smith. 2006. Spatial prediction on a river network. Journal of Agricultural, Biological, and Environmental Statistics **11**:127-150.

Crisp, D., and G. Howson. 1982. Effect of air temperature upon mean water temperature in streams in the north Pennines and English Lake District. Freshwater biology **12**:359-367.

Cutler, A. 2017. Comments on random forests. Logan, Utah.

Cutler, D. R., T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. Ecology **88**:2783-2792.

De'Ath, G. 2007. Boosted trees for ecological modeling and prediction. Ecology **88**:243-251.

Detenbeck, N. E., A. Morrison, R. W. Abele, and D. Kopp. 2016. Spatial statistical network models for stream and river temperature in New England, USA. Water Resources Research.

DeWeber, J. T., and T. Wagner. 2014. A regional neural network ensemble for predicting mean daily river water temperature. Journal of Hydrology **517**:187-200.

Drexler, M., and C. H. Ainsworth. 2013. Generalized additive models used to predict species abundance in the Gulf of Mexico: an ecosystem modeling tool. PloS one **8**:e64458.

Eaton, J. G., and R. M. Scheller. 1996. Effects of climate warming on fish thermal habitat in streams of the United States. Limnology and oceanography **41**:1109-1115.

ESRI. 2011. ArcGIS Desktop: Release 10. Environmental Systems Research Institute, Redlands, CA.

Fewster, R. M., S. T. Buckland, G. M. Siriwardena, S. R. Baillie, and J. D. Wilson. 2000. Analysis of population trends for farmland birds using generalized additive models. Ecology **81**:1970-1984.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. The elements of statistical learning. Springer series in statistics Springer, Berlin.

Friedman, J. H. 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics:1189-1232.

Friedman, J. H. 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis **38**:367-378.

Friele, P. A., K. Paige, and R. D. Moore. 2016. Stream Temperature Regimes and the Distribution of the Rocky Mountain Tailed Frog at Its Northern Range Limit, Southeastern British Columbia. Northwest Science **90**:159-175.

Gardner, B., P. J. Sullivan, and J. Lembo, Arthur J. 2003. Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. Canadian Journal of Fisheries and Aquatic Sciences **60**:344-351.

Guisan, A., T. C. Edwards, and T. Hastie. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological modelling **157**:89-100.

Hastie, T., and R. Tibshirani. 1986. Generalized additive models. Statistical science:297-310.

Hastie, T., and R. Tibshirani. 2013. GAM: generalized additive models. R package version 1.03 (2010).

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., 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.

Heino, J., R. Virkkala, and H. Toivonen. 2009. Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. Biological Reviews **84**:39-54.

Hildrew, A., and J. Edington. 1979. Factors facilitating the coexistence of hydropsychid caddis larvae (Trichoptera) in the same river system. The Journal of Animal Ecology:557-576.

Hill, R. A., and C. P. Hawkins. 2014. Using modelled stream temperatures to predict macro-spatial patterns of stream invertebrate biodiversity. Freshwater biology **59**:2632-2644.

Hill, R. A., C. P. Hawkins, and D. M. Carlisle. 2013. Predicting thermal reference conditions for USA streams and rivers. Freshwater science **32**:16.

Hill, R. A., C. P. Hawkins, and J. Jin. 2014. Predicting thermal vulnerability of stream and river ecosystems to climate change. Climatic change **125**:399-412.

Homer, C. G., J. A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. D. Herold, J. Wickham, and K. Megown. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. Photogramm. Eng. Remote Sens **81**:345-354.

Hostetler, S., J. Alder, and A. Allan. 2011. Dynamically downscaled climate simulations over North America: Methods, evaluation, and supporting documentation for users. 2331-1258, US Geological Survey.

Isaak, D., S. Wollrab, D. Horan, and G. Chandler. 2012. Climate change effects on stream and river temperatures across the northwest US from 1980–2009 and implications for salmonid fishes. Climatic change **113**:499-524.

Isaak, D. J., and W. A. Hubert. 2004. Nonlinear response of trout abundance to summer stream temperatures across a thermally diverse montane landscape. Transactions of the American Fisheries Society **133**:1254-1259.

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.

Isaak, D. J., E. E. Peterson, D. Nagel, J. Ver Hoef, and J. Kershner. 2013. A national stream internet to facilitate accurate, high-resolution status and trend assessments for water quality parameters and aquatic biotas, National Landscape Conservation Cooperative grant. Page US Fish and Wildlife Service.

Isaak, D. J., E. E. Peterson, J. Ver Hoef, D. Horan, and D. Nagel. 2016. Scalable population estimates using spatial-stream network (SSN) models, fish density surveys, and national geospatial frameworks for stream data. Canadian Journal of Fisheries and Aquatic Sciences **999** 1:10.

Isaak, D. J., E. E. Peterson, J. M. Ver Hoef, S. J. Wenger, J. A. Falke, C. E. Torgersen, C. Sowder, E. A. Steel, M. J. Fortin, and C. E. Jordan. 2014. Applications of spatial statistical network models to stream data. Wiley Interdisciplinary Reviews: Water **1**:277-294.

Isaak, D. J., and B. E. Rieman. 2013. Stream isotherm shifts from climate change and implications for distributions of ectothermic organisms. Global Change Biology **19**:742-751.

Isaak, D. J., S. J. Wenger, E. E. Peterson, J. Ver Hoef, D. Nagel, C. H. Luce, S. Hostetler, J. B. Dunham, B. B. Roper, S. Wollrab, G. Chandler, D. Horan, and S. Parkes. 2017. The NorWeST database and modeled summer temperature scenarios: Massive crowd-sourcing and new geospatial tools reveal broad climate warming of rivers and streams in the western U.S (in review). AGU Publications.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. An introduction to statistical learning. Springer.

Jones, L. A., C. C. Muhlfeld, L. A. Marshall, B. L. McGlynn, and J. L. Kershner. 2014. Estimating thermal regimes of bull trout and assessing the potential effects of climate warming on critical habitats. River Research and Applications **30**:204-216.

Keleher, C. J., and F. J. Rahel. 1996. Thermal limits to salmonid distributions in the Rocky Mountain region and potential habitat loss due to global warming: a geographic information system (GIS) approach. Transactions of the American Fisheries Society **125**:1-13.

Kuhn, M. 2016. caret: Classification and Regression Training.

Leathwick, J., J. Elith, M. Francis, T. Hastie, and P. Taylor. 2006a. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series **321**:267-281.

Leathwick, J., J. Elith, and T. Hastie. 2006b. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological modelling **199**:188-196.

Liaw, A., and M. Wiener. 2002. Classification and Regression by randomForest. R News **2**:18-22.

Lipton, Z. C. 2016. The mythos of model interpretability. arXiv preprint arXiv:1606.03490.

Marra, G., and S. N. Wood. 2011. Practical variable selection for generalized additive models. Computational Statistics & Data Analysis **55**:2372-2387.

Mayer, T. D. 2012. Controls of summer stream temperature in the Pacific Northwest. Journal of Hydrology **475**:323-335.

Mersmann, O. 2015. microbenchmark: Accurate Timing Functions.

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

Moisen, G. G., E. A. Freeman, J. A. Blackard, T. S. Frescino, N. E. Zimmermann, and T. C. Edwards. 2006. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. Ecological modelling **199**:176-187.

Olden, J. D., and D. A. Jackson. 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. Ecological modelling **154**:135-150.

Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: a primer for ecologists. The Quarterly review of biology **83**:171-193.

Oliveira, S., F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. Pereira. 2012. Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. Forest Ecology and Management **275**:117-129.

Peterson, E. E., A. A. Merton, D. M. Theobald, and N. S. Urquhart. 2006. Patterns of spatial autocorrelation in stream water chemistry. Environmental Monitoring and Assessment **121**:571-596.

Peterson, E. E., and J. M. Ver Hoef. 2014. STARS: An ArcGIS toolset used to calculate the spatial information needed to fit spatial statistical models to stream network data. J Stat Softw **56**:1-17.

Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. A. Falke, M. J. Fortin, C. E. Jordan, K. McNyset, P. Monestiez, A. S. Ruesch, and A. Sengupta. 2013. Modelling

dendritic ecological networks in space: an integrated network perspective. Ecology Letters **16**:707-719.
Piccolroaz, S., E. Calamita, B. Majone, A. Gallice, A. Siviglia, and M. Toffolon. 2016. Prediction of river water temperature: a comparison between a new family of hybrid models and statistical approaches. Hydrological Processes.
Pörtner, H. 2001. Climate change and temperature-dependent biogeography: oxygen limitation of thermal tolerance in animals. Naturwissenschaften **88**:137-146.
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.
Ridgeway, G. 2013. Generalized boosted models: a guide to the gbm package. 512.
Ridgeway, G. 2015. gbm: Generalized boosted regression models.
Rushworth, A., E. Peterson, J. Ver Hoef, and A. Bowman. 2015. Validation and comparison of geostatistical and spline models for spatial stream networks. Environmetrics **26**:327-338.
SAS Institute, I. 2015. Cary, North Carolina.
Seber, G. A., and A. J. Lee. 2003. Linear Regression: Estimation and Distribution Theory. Linear Regression Analysis, Second Edition:35-95.
Team, R. D. C. 2016. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria.
Turschwell, M. P., E. E. Peterson, S. R. Balcombe, and F. Sheldon. 2016. To aggregate or not? Capturing the spatio-temporal complexity of the thermal regime. Ecological Indicators **67**:39-48.
USGS. 2016. National Water Information System. United States Department of the Interior.
Ver Hoef, J. M., E. Peterson, and D. Theobald. 2006. Spatial statistical models that use flow and stream distance. Environmental and Ecological statistics **13**:449-464.
Ver Hoef, J. M., E. E. Peterson, D. Clifford, and R. Shah. 2014. SSN: An R package for spatial statistical modeling on stream networks. submitted to Journal of Statistical Software.
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.
Welsh Jr, H. H., G. R. Hodgson, B. C. Harvey, and M. F. Roche. 2001. Distribution of juvenile coho salmon in relation to water temperatures in tributaries of the Mattole River, California. North American Journal of Fisheries Management **21**:464-470.
Westhoff, J., and A. Rosenberger. 2016. A global review of freshwater crayfish temperature tolerance, preference, and optimal growth. Reviews in Fish Biology and Fisheries **26**:329-349.
Wickham, H. 2016a. dplyr: A Grammar of Data Manipulation.
Wickham, H. 2016b. purrr: Functional Programming Tools.
Wood, S. 2006. Generalized additive models: an introduction with R. CRC press.
Wood, S. 2013. The mgcv package. 215 p.

Yuan, L. L. 2004. Using spatial interpolation to estimate stressor levels in unsampled streams. Environmental Monitoring and Assessment **94**:23-38.

APPENDICES

APPENDIX A

MODEL COVARIATES

TABLE A1. Covariates for Mwmt models

LM/SSN	LM lat long	GAM	GAM lat long	RF	RF lat long	GBM	GBM lat long
Elevation	Easting	AirMwmt	AirMwmt	AirMwmt	AirSummerM	AirMwmt	AirMwmt
Gvalley*	Elevation	Carea	AirSummerM	Carea	Carea	AirSummerM	AirSummerM
Rad	Gvalley*	Draind	Carea	Draind	Draind	Carea	Carea
Rad x AirMwmt	Northing	Elevation	Draind	Elevation	Easting	Draind	Draind
	Rad	Gvalley	Easting x Northing (tensor product)	Gvalley	Elevation	Elevation	Easting
	Rad*AirMwmt	Rad	Elevation	Rad	Gvalley	Gvalley	Elevation
		Slope	Gvalley	Slope	Northing	Lake	Gvalley
		SummerMnFl	Rad	SummerMnFl	Rad	Rad	Lake
		Valleyb	Slope	Valleyb	Slope	SlopeE	Northing
			SummerMnFl		SummerMn-Fl	SummerMnFl	Rad
			Valleyb		Valleyb	Valleyb	Slope
							SummerMnFl
							Valleyb

*indicates variables were arcsin-transformed

TABLE A2. Covariates for SummerMean models

LM/SSN	LM lat long	GAM	GAM lat long	RF	RF lat long	GBM	GBM lat long
Carea	Carea	Elevation	AirMwmt	AirMwmt	AirSummer M	AirMwmt	AirMwmt
Elevation	Easting	AirMwmt	AirSummer-M	Carea	Carea	AirSummer-M	AirSummer-M
Gvalley*	Elevation	Carea	Carea	Draind	Draind	Carea	Carea
Rad x AirMwmt	Gvalley*		Draind	Elevation	Easting	Draind	Draind
SummerMnFl	Northing	Draind	Easting x Northing (tensor product)	Gvalley	Elevation	Elevation	Easting
Valleyb*	Rad x AirMwmt	Gvalley	Elevation	Rad	Gvalley	Gvalley	Elevation
	Summer-MnFl		Gvalley	SLOPE	Northing	Lake	Gvalley
	Valleyb*	Rad	Rad	Summer-MnFl	Rad	Rad	Lake
		Slope	Slope	Valleyb	Slope	Slope	Northing
		Summer-MnFl	Summer-MnFl		Summer-MnFl	Summer-MnFl	Rad
		Valleyb	Valleyb		Valleyb	Valleyb	Slope
							Summer-MnFl
							Valleyb

*indicates variables were arcsin-transformed

TABLE A3. Covariates for Stream_Aug models

LM/SSN	LM lat long	GAM	GAM lat long	RF	RF lat long	GBM	GBM lat long
Air_Aug	Cumdrainag*	Air_Aug	Air_Aug	Air_Aug	Air_Aug	Air_Aug	Air_Aug
Bfi	Elev	Bfi	Bri	Bfi	Bfi	Bfi	Bfi
Canopy	Precip	Canopy	Canopy	Canopy	Canopy	Canopy	Canopy
Cumdrainag*	X_Coord	Cumdrainag	Cumdrainag	Cumdrainag	Cumdrainag	Cumdrainag	Cumdrainag
Dam_effect	Y_Coord	Dam_effect	Dam_effect	Dam_effect	Dam_effect	Dam_effect	Dam_effect
Elev		Elev	Elev	Elev	Elev	Elev	Elev
Flow_Aug		Precip	Flow_Aug	Flow_Aug	Flow_Aug	Flow_Aug	Flow_Aug
Precip		Slope	X_Coord x Y_Coord (tensor product)	Precip	Precip	Precip	Precip
Slope*			Precip		X_Coord		X_Coord
			Slope		Y_Coord		Y_Coord

*indicates variables were log-transformed

APPENDIX B

NON-SPATIAL SIMULATION RESULTS

TABLE A4. RMSE values for non-spatial simulations. “LIN” and “NON” indicate linear and nonlinear data structure, respectively; “ind” and “auto” indicate independent and autocorrelated errors, respectively. “Number of vars” indicates the number of variables. “Resub RMSE”, “Test RMSE” and “CV RMSE” denote the resubstitution RMSE, test-set RMSE, and 10-fold cross-validated RMSE, respectively. The last column “ρ” is the amount of autocorrelation in the errors (for data sets with autocorrelated error structures only).

Linear/Nonlinear structure	Auto/Ind Error structure	Number of vars	Method	Resub RMSE	Test RMSE	CV RMSE	Error variance	Data model	ρ
LIN	ind	1	LM	1.94	2.03	1.98	4	$y = b0 + b_1*x_1$	0.7
LIN	ind	1	RF	1.12	2.28	2.28	4		0.7
LIN	ind	1	GBM	1.40	2.36	2.23	4		0.7
LIN	ind	1	GAM	1.94	2.03	1.99	4		0.7
NON	ind	1	LM	3.35	3.43	3.44	4	$y = b0 * \exp(-b_1*x_1)$	0.7
NON	ind	1	RF	1.16	2.24	2.32	4		0.7
NON	ind	1	GBM	1.75	2.67	2.62	4		0.7
NON	ind	1	GAM	3.35	2.02	2.09	4		0.7
LIN	ind	2	LM	1.97	1.89	2.03	4	$y = b0 + b_1*x_1 + b_2*x_2$	0.7

(Table continues)

LIN	ind	2	RF	1.12	2.08	2.25	4		0.7
LIN	ind	2	GBM	0.81	2.36	2.08	4		0.7
LIN	ind	2	GAM	1.97	1.91	2.07	4		0.7
NON	ind	2	LM	6.01	6.14	6.24	4	$y = b0 * \exp(-b_1*x_1) + b2*x_2^2$	0.7
NON	ind	2	RF	1.49	3.03	3.09	4		0.7
NON	ind	2	GBM	1.22	3.52	3.39	4		0.7
NON	ind	2	GAM	6.01	2.09	2.18	4		0.7
LIN	auto	1	LM	1.92	2.53	1.96	4	$y = b_0 + b_1*x_1$	0.7
LIN	auto	1	RF	1.10	2.77	2.24	4		0.7
LIN	auto	1	GBM	1.40	2.84	2.20	4		0.7
LIN	auto	1	GAM	1.92	2.54	1.98	4		0.7
NON	auto	1	LM	3.26	3.43	3.36	4	$y = b_0 * \exp(-b_1*x_1)$	0.7
NON	auto	1	RF	1.14	2.16	2.28	4		0.7
NON	auto	1	GBM	1.82	2.58	2.64	4		0.7
NON	auto	1	GAM	3.26	1.99	2.04	4		0.7
LIN	auto	2	LM	1.92	1.90	1.98	4	$y = b_0 + b_1*x_1 + b_2*x_2$	0.7
LIN	auto	2	RF	1.09	2.13	2.22	4		0.7
LIN	auto	2	GBM	0.77	2.30	2.32	4		0.7
LIN	auto	2	GAM	1.92	1.92	2.00	4		0.7
NON	auto	2	LM	5.96	5.65	6.20	4	$y = b0 * \exp(-b_1*x_1) + b_2*x_2^2$	0.7
NON	auto	2	RF	1.45	2.66	3.01	4		0.7
NON	auto	2	GBM	1.23	3.12	3.37	4		0.7

(Table continues)

NON	auto	2	GAM	5.96	2.03	2.08	4		0.7
LIN	ind	1	LM	3.02	3.25	3.09	9	$y = b_0 + b_1*x_1$	0.7
LIN	ind	1	RF	1.74	3.72	3.57	9		0.7
LIN	ind	1	GBM	2.17	3.82	3.48	9		0.7
LIN	ind	1	GAM	3.02	3.27	3.11	9		0.7
NON	ind	1	LM	3.92	4.37	4.01	9	$y = b_0 * exp(-b_1*x_1)$	0.7
NON	ind	1	RF	1.69	3.64	3.42	9		0.7
NON	ind	1	GBM	2.40	3.91	3.68	9		0.7
NON	ind	1	GAM	3.92	3.24	3.08	9		0.7
LIN	ind	2	LM	2.99	3.25	3.09	9	$y = b_0 + b_1*x_1 + b_2*x_2$	0.7
LIN	ind	2	RF	1.64	3.59	3.36	9		0.7
LIN	ind	2	GBM	1.21	3.91	3.58	9		0.7
LIN	ind	2	GAM	2.99	3.30	3.13	9		0.7
NON	ind	2	LM	6.36	7.00	6.61	9	$y = b_0 * exp(-b_1*x) + b_2*x_2^2$	0.7
NON	ind	2	RF	1.92	3.63	3.94	9		0.7
NON	ind	2	GBM	1.51	4.20	4.35	9		0.7
NON	ind	2	GAM	6.35	3.04	3.26	9		0.7
LIN	auto	1	LM	2.89	3.69	2.96	9	$y = b_0 + b_1*x_1$	0.7
LIN	auto	1	RF	1.66	3.99	3.39	9		0.7
LIN	auto	1	GBM	2.07	3.99	3.32	9		0.7
LIN	auto	1	GAM	2.89	3.70	2.97	9		0.7
NON	auto	1	LM	3.92	4.69	4.03	9	$y = b_0 * exp(-b_1*x_1)$	0.7

(Table continues)

NON	auto	1	RF	1.63	3.93	3.27	9		0.7
NON	auto	1	GBM	2.35	4.36	3.53	9		0.7
NON	auto	1	GAM	3.92	3.62	2.97	9		0.7
LIN	auto	2	LM	2.98	3.29	3.08	9	$y = b_0 + b_1*x_1 + b_2*x_2$	0.7
LIN	auto	2	RF	1.63	3.53	3.34	9		0.7
LIN	auto	2	GBM	1.17	3.85	3.52	9		0.7
LIN	auto	2	GAM	2.98	3.31	3.12	9		0.7
NON	auto	2	LM	6.39	6.52	6.63	9	$y = b_0 * \exp(-b_1*x_1) + b_2*x_2^2$	0.7
NON	auto	2	RF	1.87	3.67	3.86	9		0.7
NON	auto	2	GBM	1.49	4.17	4.22	9		0.7
NON	auto	2	GAM	6.39	3.02	3.20	9		0.7
LIN	auto	1	LM	2.00	2.58	2.04	4	$y = b_0 + b_1*x_1$	0.8
LIN	auto	1	RF	1.16	2.82	2.35	4		0.8
LIN	auto	1	GBM	1.45	2.92	2.31	4		0.8
LIN	auto	1	GAM	2.00	2.54	2.05	4		0.8
NON	auto	1	LM	3.18	2.84	3.27	4	$y = b_0 * \exp(-b_1*x_1)$	0.8
NON	auto	1	RF	1.09	2.10	2.17	4		0.8
NON	auto	1	GBM	1.70	2.43	2.49	4		0.8
NON	auto	1	GAM	2.60	1.85	1.94	4		0.8
LIN	auto	2	LM	1.91	1.96	1.97	4		0.8
LIN	auto	2	RF	1.09	2.19	2.20	4		0.8
LIN	auto	2	GBM	0.77	2.40	2.32	4		0.8
LIN	auto	2	GAM	1.91	1.98	1.99	4		0.8
NON	auto	2	LM	5.97	6.26	6.21	4		0.8

(Table continues)

NON	auto	2	RF	1.44	3.00	2.97	4		0.8
NON	auto	2	GBM	1.16	2.97	3.31	4		0.8
NON	auto	2	GAM	5.97	2.13	3.31	4		0.8
LIN	auto	1	LM	2.83	3.20	2.89	9	$y = b0 + b_1*x_1$	0.8
LIN	auto	1	RF	1.62	3.57	3.30	9		0.8
LIN	auto	1	GBM	2.02	3.64	3.21	9		0.8
LIN	auto	1	GAM	2.83	3.20	2.90	9		0.8
NON	auto	1	LM	3.95	3.57	4.06	9	$y = b_0 * exp(-b_1*x_1)$	0.8
NON	auto	1	RF	1.71	3.12	3.43	9		0.8
NON	auto	1	GBM	2.38	3.41	3.63	9		0.8
NON	auto	1	GAM	3.95	2.68	3.08	9		0.8
LIN	auto	2	LM	2.77	2.67	2.85	9	$y = b_0 + b_1*x_1 + b_2*x_2$	0.8
LIN	auto	2	RF	1.54	2.96	3.06	9		0.8
LIN	auto	2	GBM	1.12	2.69	2.89	9		0.8
LIN	auto	2	GAM	2.77	2.69	2.89	9		0.8
NON	auto	2	LM	6.34	6.38	6.58	9	$y = b_0 * exp(-b_1*x_1) + b_2*x_2^2$	0.8
NON	auto	2	RF	1.90	3.69	3.91	9		0.8
NON	auto	2	GBM	1.54	4.31	4.30	9		0.8
NON	auto	2	GAM	6.34	3.10	3.25	9		0.8

APPENDIX C

DETAILS OF SSN

SSN models as described by Ver Hoef et al. (2006) take into account covariance structure of temperature data on a river network, which allow for the unique properties of stream networks such as branching structure, longitudinal connectivity, directed flow, and abrupt temperature changes at stream junctions (Isaak et al. 2014). The basic form of the model is similar to that of a linear model

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\epsilon}$$

where **X** is a matrix of fixed effects predictor variables and **β** is a parameter vector for fixed effects; the mean, **Xβ**, is modeled using predictor variables known to influence the response (**Y**). In contrast to a linear model with an independent N ~ (0, 1) error structure (**ε**), SSN models account for upstream (tail up) and downstream (tail down) autocorrelation using a weighted moving average function (Ver Hoef and Peterson 2006). Tail up and tail down models are derived using hydrologic distance. A general form of the SSN model is:

$$\mathbf{Y} = \mathbf{X\beta} + v_{\mathrm{TU}} + v_{\mathrm{TD}} + v_{\mathrm{EU}} + v_{\mathrm{NUG}}$$

Where the v components correspond to tail up, tail down, and Euclidian autocorrelation structures and a nugget effect.

Another important property of SSN models is that they account for the important spatial relationships among flow-connected and flow-unconnected sites. In flow-connected sites, water flows from an upstream site past a downstream site. In flow-unconnected sites, water from one site cannot reach another site via normal stream flow (i.e. water must move upstream to reach the other site).

Features of SSN models, such as tail-up versus tail-down and flow-connected versus flow-unconnected may be understood after the introduction of some notation (also, see Fig. A1). Let x_i be the distance upstream on the i^{th} stream segment and l_i be the most downstream location on the i^{th} stream segment and u_i be the most upstream location on the i^{th} stream segment. Next, let I be the total set of stream indices. The index set of stream segments upstream of x_i, excluding x_i, is $\cup_{X_i} \subset I$. Next, let D_{x_i} be the index set of all stream segments downstream of x_i into which x_i flows, including the stream segment x_i. Now, let s_i and t_j be two stream segments. S_i and t_j are flow connected if $D_{s_i} \cap D_{t_j} = D_{s_i}$ or D_{t_j} . Two segments, s_i and t_j are flow unconnected if $D_{s_i} \cap D_{t_j} \neq D_{s_i}$ or D_{t_j}.Similarly, distance between stream segments (including the upstream segment but excluding the downstream segment) may be denoted as

$$B(s_i, t_j) = \begin{cases} (D_{s_i} \cap D_{t_j})^c \cap \left(D_{s_i} \cup D_{t_j}\right), & \text{if } s_i \text{ and } t_j \text{ are flow connected} \\ \emptyset, & \text{otherwise.} \end{cases}$$

Now, stream distance, which can be thought of as the shortest distance between two sites on a stream network, can be defined as follows :

$$d(s_i, t_j) = \begin{cases} |s_i - t_j|, \text{ if } s_i \text{ and } t_j \text{ are flow-connected} \\ (s_i - u) + (t_j - u), \quad \text{otherwise,} \end{cases}$$

where $d(s_i, t_j)$ is stream distance between two points s_i and t_j and, $u = \max\{u: k \in D_{s_i} \cap D_{t_j}\}$ (Ver Hoef et al. 2006). A valid autocovariance function for tail up models is shown in equation 1:

1

$$C(h|\theta) = \begin{cases} \int_{-\infty}^{\infty} g[(x|\theta)]^2 dx + v_j^2, \text{ if } h = 0 \\ \int_{-\infty}^{\infty} g(x|\theta) g(x - h|\theta) dx, \text{ if } h > 0, \end{cases}$$

where h is Euclidean distance, and $g(x|\theta)$ is the moving average function. Note that there is a discontinuity v_j^2 at $h = 0$ (the "nugget" effect) (Ver Hoef et al. 2006). Moving average functions work on the real line, which is defined from $-\infty$ to $+\infty$. On a stream network, however, stream segments split in two, so the moving average function is also split into two parts. To ensure stationary variances along stream segments, weights (ω), are assigned to stream segments based on stream flow volume (Ver Hoef et al. 2006). The addition of weighting prevents the inflation of variances for stream segments that have more upstream branching compared to other stream segments. In tail up models, the moving average function points in the upstream direction and correlation is calculated only between flow-connected sites, while in tail-down models, the moving average function points in the downstream direction, and correlation is calculated between flow-connected as well as flow-unconnected sites (Isaak et al. 2014, Ver Hoef et al. 2014). The

weights necessary for tail up models are not required to ensure stationarity of tail-down models.

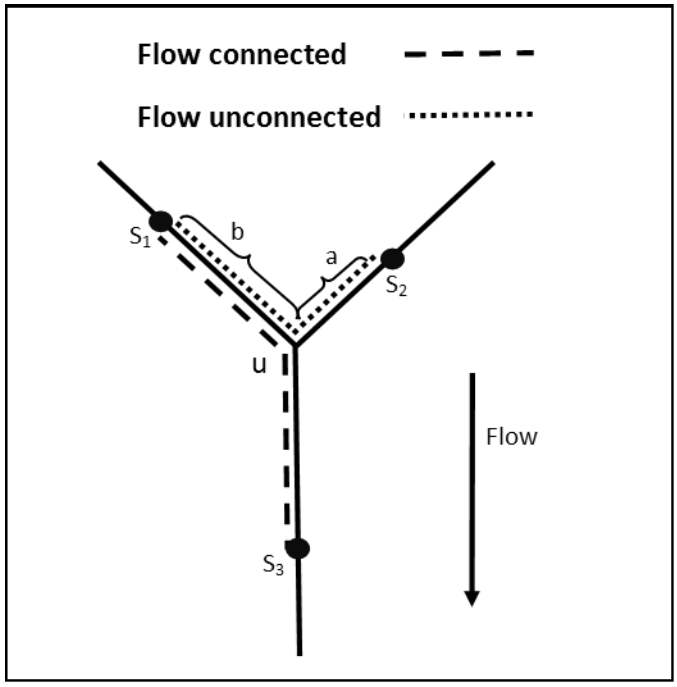

FIG. A1. A simple representation of flow-connected and flow-unconnected sites.

The tail up model of the covariance between two sites, s_i and t_j , is defined as

2

$$C_u(s_i, t_j|\theta) = \begin{cases} 0\,, & \text{if } s_i \text{ and } t_j \text{ are not flow-connected} \\ C_1(0) + v_{j\,,} & \text{if s = t} \\ \prod_{k \in B_{s_i,t_j}} \sqrt{\omega} C_1\left(d(s_i, t_j)\right)\,, & \text{otherwise,} \end{cases}$$

where $C_i(h) = \int_{-\infty}^{\infty} g(x|\theta)g(x-h|\theta)dx$ (Ver Hoef et al. 2006) , and $g(x|\theta)$ may be defined by valid autocovariance functions: linear with sill, spherical, Mariah, exponential, Cauchy, Empanovich, and Gaussian (Ver Hoef et al. 2006, Isaak et al. 2014).

Now, for example, the exponential covariance function for the tail-up model can be expressed as

3

$$C_u(h|\theta_u) = \sigma_u^2 e^{-\frac{3h}{\alpha_u}} .$$

A valid autocovariance function for tail-down models is

$$C_d(h|\theta) = \int_{-\infty}^{h} g(-x|\theta)g(-x-h|\theta)dx,$$

if sites are flow-connected, and for $b > a$ (Fig. A1), a valid autocovariance function is

$$C_d(a,b|\theta) = \int_{-\infty}^{-b} g(-x|\theta)g(-x-(b-a)|\theta)dx,$$

if sites are flow-unconnected.

Similarly, the exponential covariance function for the tail down model, which distinguishes between flow-connected and flow-unconnected sites, is expressed as

4

$$C_d(h|\theta_d) = \begin{cases} \sigma_d^2 e^{-\frac{3h}{\alpha_d}}, & \text{flow} - \text{connected} \\ \sigma_d^2 e^{-\frac{3(a+b)}{\alpha_d}}, & \text{flow} - \text{unconnected} . \end{cases}$$

In equation 4, a and b are defined as in Figure A1. In equations 3 and 4, σ_u^2 and σ_d^2 (> 0) are the "partial sills" (variance parameters and values of the covariance functions when the distance is 0) for the tail up and tail down models, respectively; α_u and α_d are the range parameters for tail up and tail down models, respectively, and h is the distance between two flow-connected sites (Ver Hoef and Peterson 2012).

Although autocovariance functions based on stream distance are useful for stream networks, climatic and/or geographic variables may be best modeled by Euclidean distance. SSN models can take this into account, resulting in a models that may potentially have tail-up, tail-down, and Euclidean covariance structures (Isaak et al. 2014). Another useful tool developed along with SSN models to visually inspect autocorrelation is the Torgegram (Peterson et al. 2013). Like a semivariogram, a Torgegram depicts semivariance over distance between sites; however, a Torgegram splits the semivariance between flow-connected and flow-unconnected sites (Peterson et al. 2013).

Since their introduction in 2006, SSN models have received some attention in the field of stream ecology, but due to the limited awareness of their potential, implementation of SSN models remains limited (Isaak et al. 2014). Falke et al. (2016) used SSN models to predict stream temperatures in habitat modelling for trout in the Great Basin; such a model could be used by managers to prioritize conservation management of streams. SSN models have been used numerous times to predict stream temperatures under climate change scenarios (Isaak et al. 2010). SSN models have also been used to predict the influence of geographically isolated wetlands on stream flow

(Golden et al. 2016). These models were successfully used to predict stream temperatures in different regions, including the western (Isaak et al. 2010), midwestern (Golden et al. 2016), and eastern United States (Detenbeck et al. 2016) and Australia (Turschwell et al. 2016). Isaak et al. (2016) estimated salt trout populations in the Salt River watershed in Utah, while Brennan et al. (2016) used SSN models to quantify the movement of strontium isotopes over a river network and geographic features. These two latter examples show that SSN models can be generalized to predict other ecological phenomena besides stream temperatures.

SSN models are more accurate in predicting stream temperatures than linear models (Isaak et al. 2010) and random forests (Turschwell et al. 2016). They can also be used to predict stream temperatures at unsampled locations along a stream network (Isaak et al. 2014). Random and mixed effects can also be used in SSN models, adding to their versatility (Ver Hoef et al. 2014). However, SSN models suffer from a considerable GIS data pre-processing requirement; furthermore, model implementation and interpretation requires advanced background in statistical theory and R computing skills (Isaak et al. 2014). These two barriers may be prohibitive in allowing practitioners or other stream scientists to analyze data using SSN models. In addition, SSN models are much slower computationally than other methods discussed in this paper, and for large datasets (> 2000 observations), high – power computing facilities are necessary for analysis (Ver Hoef et al. 2014). Finally, since SSN models are an extension of general linear models (GLMs), the assumptions of linearity, normality and heteroscedasticity of residuals, must also be met.

LITERATURE CITED

Beck, N., and S. Jackman. 1998. Beyond linearity by default: Generalized additive models. American Journal of Political Science **42**:596-627.

Brennan, S. R., C. E. Torgersen, J. P. Hollenbeck, D. P. Fernandez, C. K. Jensen, and D. E. Schindler. 2016. Dendritic network models: Improving isoscapes and quantifying influence of landscape and in-stream processes on strontium isotopes in rivers. Geophysical Research Letters.

Cappo, M., G. De'ath, S. Boyle, J. Aumend, R. Olbrich, F. Hoedt, C. Perna, and G. Brunskill. 2005. Development of a robust classifier of freshwater residence in barramundi (Lates calcarifer) life histories using elemental ratios in scales and boosted regression trees. Marine and freshwater research **56**:713-723.

De'Ath, G. 2007. Boosted trees for ecological modeling and prediction. Ecology **88**:243-251.

Detenbeck, N. E., A. Morrison, R. W. Abele, and D. Kopp. 2016. Spatial statistical network models for stream and river temperature in New England, USA. Water Resources Research.

Falke, J. A., J. B. Dunham, D. Hockman-Wert, and R. Pahl. 2016. A Simple Prioritization Tool to Diagnose Impairment of Stream Temperature for Coldwater Fishes in the Great Basin. North American Journal of Fisheries Management **36**:147-160.

Freund, Y., and R. E. Schapire. 1996. Experiments with a new boosting algorithm. Pages 148-156 *in* Icml.

Friedman, J., T. Hastie, and R. Tibshirani. 2001. The elements of statistical learning. Springer series in statistics Springer, Berlin.

Friedman, J. H. 2002. Stochastic gradient boosting. Computational Statistics & Data Analysis **38**:367-378.

Golden, H. E., H. A. Sander, C. R. Lane, C. Zhao, K. Price, E. D'Amico, and J. R. Christensen. 2016. Relative effects of geographically isolated wetlands on streamflow: a watershed-scale analysis. Ecohydrology **9**:21-38.

Hastie, T., and R. Tibshirani. 1986. Generalized additive models. Statistical science:297-310.

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.

Isaak, D. J., E. E. Peterson, J. Ver Hoef, D. Horan, and D. Nagel. 2016. Scalable population estimates using spatial-stream network (SSN) models, fish density surveys, and national geospatial frameworks for stream data. Canadian Journal of Fisheries and Aquatic Sciences **999** 1:10.

Isaak, D. J., E. E. Peterson, J. M. Ver Hoef, S. J. Wenger, J. A. Falke, C. E. Torgersen, C. Sowder, E. A. Steel, M. J. Fortin, and C. E. Jordan. 2014. Applications of spatial statistical network models to stream data. Wiley Interdisciplinary Reviews: Water **1**:277-294.

Leathwick, J., J. Elith, M. Francis, T. Hastie, and P. Taylor. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series **321**:267-281.

Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. A. Falke, M. J. Fortin, C. E. Jordan, K. McNyset, P. Monestiez, A. S. Ruesch, and A. Sengupta. 2013. Modelling dendritic ecological networks in space: an integrated network perspective. Ecology Letters **16**:707-719.

Ridgeway, G. 2013. Generalized boosted models: a guide to the gbm package. 512.

Turschwell, M. P., E. E. Peterson, S. R. Balcombe, and F. Sheldon. 2016. To aggregate or not? Capturing the spatio-temporal complexity of the thermal regime. Ecological Indicators **67**:39-48.

Ver Hoef, J. M., E. Peterson, and D. Theobald. 2006. Spatial statistical models that use flow and stream distance. Environmental and Ecological statistics **13**:449-464.

Ver Hoef, J. M., and E. E. Peterson. 2012. A moving average approach for spatial statistical models of stream networks. Journal of the American Statistical Association.

Ver Hoef, J. M., E. E. Peterson, D. Clifford, and R. Shah. 2014. SSN: An R package for spatial statistical modeling on stream networks. submitted to Journal of Statistical Software.

Wood, S. 2006. Generalized additive models: an introduction with R. CRC press.