A Review of Statistical Water Temperature Models

Loubna Benyahya, Daniel Caissie, André St-Hilaire, Taha B.M.J. Ouarda and Bernard Bobée

Abstract: The use of statistical models to simulate or to predict stream water temperature is becoming an increasingly important tool in water resources and aquatic habitat management. This article provides an overview of the existing statistical water temperature models. Different models have been developed and used to analyze water temperature-environmental variables relationship. These are grouped into two major categories: deterministic and statistical/stochastic models. Generally, deterministic models require numerous input data (e.g., depth, amount of shading, wind velocity). Hence, they are more appropriate for analyzing different impact scenarios due to anthropogenic effects (e.g., presence of reservoirs, thermal pollution and deforestation). In contrast to the deterministic models, the main advantage of the statistical models is their relative simplicity and relative minimal data requirement. Parametric models such as linear and non-linear regression are popular methods often used for shorter time scales (e.g., daily, weekly). Ridge regression presents an advantage when the independent variables are highly correlated. The periodic models present advantages in dealing with seasonality that often exists in periodic time series. Non-parametric models (e.g., k-nearest neighbours, artificial neural networks) are better suited for analysis of nonlinear relationships between water temperature and environmental variables. Finally, advantages and disadvantages of existing models and studies are discussed.

Résumé : Étant donné l'importance de la température dans l'habitat aquatique, et étant donné les impacts humains actuels et potentiels sur le régime thermique des rivières, il s'avère nécessaire de développer des outils de gestion des ressources hydriques. Cet article propose une synthèse bibliographique des différents modèles utilisables pour la simulation ou la prévision de la température de l'eau. De nombreux modèles existent pour prédire la température de l'eau en fonction des conditions environnementales. Ils se regroupent essentiellement en deux grandes catégories: les modèles déterministes et les modèles statistiques/stochastiques. Les modèles déterministes de la température de l'eau nécessitent souvent un grand nombre d'intrants (e.g., couvert végétal, vitesse du vent, profondeur de la rivière) qui ne sont pas toujours disponibles. Alternativement, les modèles statistiques offrent l'avantage de nécessiter moins de données et un temps de développement habituellement moins long que les modèles déterministes. L'article réalise une étude comparative de

Loubna Benyahya[1], Daniel Caissie[1], André St-Hilaire[2], Taha B.M.J. Ouarda[2] and Bernard Bobée[2]

[1] Fisheries and Oceans Canada, Gulf Fisheries Centre, Moncton, NB E1C 9B6
[2] INRS-ETE, Québec City, QC G1K 9A9

Submitted February 2007; accepted July 2007. Written comments on this paper will be accepted until March 2008.

© 2007 Canadian Water Resources Association

l'application des différents modèles statistiques existant à savoir les modèles paramétriques (régression linéaire, non-lineaire, régression « Ridge », modèles périodiques, etc.) et les modèles non-paramétriques (Réseaux de Neurones Artificiels et les k-Voisins les plus proches). Tous ces modèles ont donné des résultats encourageants voire satisfaisants et il est difficile d'exhiber la supériorité d'un modèle particulier. On conclue en présentant les avantages et les limites des différents modèles.

Introduction

Importance of Water Temperature for Fish Habitat

Water temperature is one of the most important parameters in ecosystem studies. Temperature can influence both chemical and biological processes such as dissolved oxygen concentrations, fish growth and even mortality. Many biological conditions are linked to river thermal regime. For instance, Hodgson and Quinn (2002) demonstrated that the triggering of the spawning period for Sockeye salmon (*Onchorhyncus nerka*) on the northwest coast of the United States was strongly influenced by water temperature. They found that when a threshold of 19°C was reached, spawning was interrupted as individual fish sought thermal refuge. Bjornn and Reiser (1991) showed that high stream temperatures in the range of 23-25°C can increase the mortality rate of salmonid fish. Johnson (1997) reported on the relation between the timing of emergence of Atlantic salmon (*Salmo salar*) fry and water temperature. Many studies have addressed human influences or anthropogenic perturbations of river thermal regime in order to better understand their impacts on fish habitat. For example, Beschta *et al*. (1987) have studied the impact of forestry operations on river water temperature whereas Webb and Walling (1993) looked at the impact of reservoirs on downstream temperatures.

River and stream temperatures can also be impacted globally, by climate change. Notably, Eaton and Scheller (1996) suggested that under a doubling of the current atmospheric carbon dioxide concentration, densities of a number of cold water fish species could decline by as much as 50 percent in the United States. Mohseni *et al*. (2003) reported that, under this scenario, suitable thermal habitat for cold water fish will suffer a 36 percent decrease.

These and other studies show the critical role of river water temperature and the importance of understanding the thermal regime of rivers for effective aquatic habitat management.

Temporal, Spatial Variability and the Need for Modelling Tools

The thermal regime of rivers is affected by heat exchanges in which both meteorological factors and physical characteristics are important. Meteorological factors affecting the temporal variability of energy exchanges include the annual cycle of incoming solar radiation as well as short-term conditions of wind speed, humidity and air temperature among others. The natural processes of heating and cooling highly depend on meteorological conditions as well as stream physical and hydrological characteristics. Stream characteristics affecting energy exchange include riparian vegetation (especially in small streams), stream aspect, channel geomorphology, valley topography, location of tributaries and groundwater inflows. Spatial and temporal variations in water temperature are important for aquatic resources (Vannote *et al*., 1980).

The natural spatiotemporal variability is often disturbed by human activities such as deforestation and regulation by dams (Webb and Walling, 1993). For instance, riparian vegetation, which can be affected by forestry operations, can have a major effect on various components of the stream heat budget (Brown and Krygier, 1967; St-Hilaire *et al*., 2000). Flow modification and/or water withdrawal can also impact on spatial and temporal variability of water temperature (Sinokrot and Gulliver, 2000). In a recent study, correlations between air and water temperature were shown to be stronger when discharge was below the annual median (Gu *et al*., 1998; Webb *et al*., 2003).

Because of the ecological importance of stream water temperature and the potential impact of human activities on such temperatures, it is important to provide water resource and fisheries managers with efficient assessment and modelling tools. There are a number of different applications of water temperature

© 2007 Canadian Water Resources Association

models. They include water temperature forecasts for short (e.g., next day) or longer time scales (from weeks to years). Hindcast can be done when available time series are short in order to extend the records in the past. Water temperature simulations can also be done to investigate the potential impact of anthropogenic effects on the thermal regime of rivers. For instance, water temperature models have been included in decision support systems to assist managers in determining optimum outflows that maintain adequate temperature ranges for biota (e.g., Gu *et al.,* 1999; Krause *et al.,* 2005).

We describe a number of statistical models for predicting water temperature based on a literature review. The emphasis of the review will be on statistical models as they are widely used in many fields of study, mainly due to their relatively low data requirement and simplicity in application.

Water Temperature Model Categories

Different models have been developed and used to predict river water temperature. These models have been classified into two major categories: deterministic models and statistical models. In specific application, each type of model has advantages and drawbacks. Although the aim of this review is to compare available statistical models, deterministic models will also be briefly described within the next section.

Deterministic Models

Deterministic models are based on mathematical representation of the underlying physics of heat exchange between the river and the surrounding environment. Such models are generally carried out using an energy budget approach (Morin and Couillard, 1990; Sinokrot and Stefan, 1993; St-Hilaire *et al*., 2000). They require numerous input data including stream geometry, hydrology and meteorology. For example, physical characteristics of the stream such as the depth of water as well as the amount of shading and wind sheltering are often essential components of deterministic models in order to estimate the total energy exchanged within a river reach.

Deterministic models are efficient tools when the users want to simulate modifications to some components of the heat budget (St-Hilaire *et al*., 2000). Consequently, they are very useful for analyzing and comparing different impact scenarios due to anthropogenic effects such as presence of reservoirs, thermal pollution, deforestation and others. For instance, the SHADE model simulates water temperature and includes shading effect (Chen *et al*., 1998). More recently, St-Hilaire *et al*. (2000) modified the CEQUEAU hydrological and water temperature model (Morin *et al*., 1981) to include soil temperature and crown closure in its calculation of local advective terms in the heat budget. Other deterministic models include the U.S. Fish and Wildlife SSTEMP model (Bartholow, 1999), as well as a number of simpler models (e.g., Sinokrot and Stefan, 1993; Gu *et al*., 1998; Gu and Li, 2002; Younus *et al*., 2000). One potential limitation of this modelling approach is that these tools can be quite demanding in terms of data requirement and their implementation can therefore be complex.

Statistical/Stochastic Models

An alternative approach to deterministic models in predicting or simulating water temperatures is the use of statistical or stochastic models. In contrast to the deterministic models, the main advantage of the statistical models is their relative simplicity and minimal data requirement. Statistical models can generally be classified in two categories, namely parametric and non-parametric models. Among parametric models, a further classification can be made into regression models and stochastic models. Regression models are usually applied for predicting or simulating water temperature at weekly, monthly and annual time steps, relying mainly on the relatively high correlation between air and water temperature at those time scales. This correlation is always significant, because of the joint dependence of these two variables on solar radiation. In some cases, non-linear regression models have been used to better capture the leveling off of water temperature at both high and low air temperature (Mohseni *et al*., 1998).

When water temperatures are modelled for a shorter time step than weekly, linear and non-linear regression models are generally more difficult to apply due to the autocorrelation within the water temperature time series. Hence, for shorter than weekly

© 2007 Canadian Water Resources Association

time scales, e.g., daily, stochastic models are often used as well as other types of models that account for the autocorrelation.

Although stochastic models linking water to air temperatures offer a simple means of predicting or simulating water temperature, other statistical models, such as parametric (e.g., Box Jenkins, ARMA, etc.) and non-parametric models (k-Nearest Neighbours, Artificial Neural Networks, etc.) are "statistically faithful" to the type of time series and adequately represent water temperature variability. Non-parametric models differ from parametric models in that the model structure is not specified a priori, but is instead determined from data. The following sections provide a more detailed review of the statistical approaches available to model water temperature.

Parametric Statistical Models

Regression Models

Many studies have used a statistical approach to predict water temperature. Simple regression-based models have been successfully used to model water temperature as a function of one (usually air temperature) or more independent variables. The structure of these simple models can be depicted as

$$Tw(t) = a_0 + a_1 T_a(t) + \varepsilon(t) \quad (1)$$

where $Tw(t)$ is water temperature for a given time period; $T_a(t)$ is air temperature for the same time period as water temperature; a_0, a_1 are regression coefficients and $\varepsilon(t)$ is an error term.

Such linear regression models have been applied by many authors, including Smith (1981); Crisp and Howson (1982); Mackey and Berrie (1991); and Stefan and Preud'homme (1993). The latter demonstrated that the water-air temperature relationship becomes less scattered as the time interval of the data increases from two hours, through daily averages to weekly means. Pilgrim *et al*. (1998) also used linear regression to relate stream water temperature to air temperature for a number of sites in Minnesota (United States). They showed that the slope of the regression increases with time scale (daily, weekly and monthly). Erickson and Stefan (2000) showed that both the slope and intercept are a function of the time scale. They concluded that during open water periods, streams in Minnesota had a good linear air/water temperature relationship; however the warmer Oklahoma sites showed a non-linear structure when air temperature exceeded 25°C. This was most likely due to evaporative cooling.

Equation (1) specifies air temperature as the only independent variable, but the model has been generalized using multiple regression. Webb *et al*. (2003) noted that flow is another important variable that should be considered in water temperature models. Their study showed that air and water temperatures are more strongly correlated when flows are below median levels.

In situations where the predictor variables are highly cross-correlated amongst themselves (collinearity), the challenge is to minimize the possibility of including redundant variables in the model. Ridge regression is an attempt to deal with collinearity through the use of a form of biased estimation in place of ordinary least squares (OLS) regression. The basic justification for ridge regression is that a slightly biased estimator with smaller variance may be more advantageous than an unbiased estimator having large variance. The ridge regression is a regression approach under constraint (Hoerl and Kennard, 1970) for which the equation is similar to the multiple regression model. In matrix form, the multiple regression can be written as

$$Y = \beta X + \varepsilon \quad (2)$$

where Y is the vector representing the dependent variable; β is the vector of coefficients to be adjusted; X is the matrix of independent variables and ε is the error term.

The estimator β can be given by the equation

$$\beta = (X'X)^{-1} X'Y \quad (3)$$

In the ridge regression, a ridge constant K is included in order to avoid ill-conditioning of X

$$\beta = (X'X + KI)^{-1} X'Y \quad (4)$$

where I is the identity matrix.

The selected value for the ridge constant is the lowest value for which β is stabilized. Ahmadi-Nedushan *et al*. (2007b) have shown that a ridge

© 2007 Canadian Water Resources Association

regression model can be used to simulate daily water temperature with two lagged water temperature terms, air temperature and streamflow as exogenous variables. Simulations on the Moisie River (Québec) showed good results with a root mean square error (RMSE) <0.65˚C, an error that was similar in value to that of the linear regression model.

Most regression studies acknowledge that an error term exists (e.g., Equation (1)) but few discuss the importance and the characteristics of this error term. Given the high seasonality of water temperature, autocorrelation is likely to occur in the error term, especially when linear regression is used to model water temperature at short time steps (e.g., daily or weekly). The effect of seasonality was emphasized by Langan *et al.* (2001) who showed that the linear relationship can be partitioned on a seasonal basis. They found that the best fit between air and water temperature occurred in the summer. Seasonality can also be reflected by a hysteretic behaviour of the air-water relationship. Webb and Nobilis (1997) stated that the clockwise hysteresis in this relationship was caused by changes in seasonal regression slopes with warmer air temperature in the fall yielding the same water temperature as colder air temperature in the spring.

The assumption that the water–air temperature relationship is linear has been questioned. Indeed, Mohseni *et al*. (1998) also observed a non-linear behaviour between air and water temperatures at weekly intervals. Accordingly, these authors developed a model based on the logistic S-shaped function to predict average weekly stream temperatures at different locations in the United States. The logistic function used by Mohseni *et al*. (1998) to determine the air to water relation is given by

$$Tw = \frac{\alpha}{1 + e^{\Upsilon(\beta - T_a)}} \qquad (5)$$

where Tw and T_a represent water and air temperatures, α is a coefficient which estimates the highest water temperature, β is the air temperature at the inflexion point and Υ represents the steepest slope of the logistic function. The advantage of this model over the linear regression is that it can better represent the tendency of water temperature in some water bodies to level off at higher air temperatures (Mohseni and Stefan, 1999). This change in slope has mainly been attributed to groundwater inputs and the effects of freezing at low temperatures and to evaporative cooling at high temperatures.

Equation (5) was also used by Webb *et al*. (2003) to model water-air temperature relationship at different time steps on the Exe River, United Kingdom. They found significant non-linear relations between these two variables at the hourly time step.

Autoregressive Models

Among the autoregressive methods used to simulate and predict water temperature, some of models are labelled as "stochastic". In this approach, the water temperature time series are generally divided into two components, namely the long-term annual component (seasonal variation) and the short-term variations or departure from the annual component (residuals). Then, time series models (e.g., Box Jenkins, ARMA, etc.) are fitted to water temperature residuals (short-term variations), once the seasonal component of the signal has been removed. Applications of this approach include those of Kothandaraman (1971), Cluis (1972), and Caissie *et al*. (1998; 2001).

Autoregressive (AR) models take into account the autocorrelation structure within the stream water temperature time series and can also account for the correlation with external variables (e.g., air temperature, streamflow) at various lag periods. The seasonal variation can be modelled by a Fourier series analysis (Kothandaraman, 1971) or even a simple sinusoidal function (Caissie *et al*., 1998). For example, the sinusoidal function may be written in the following form

$$Tw_{seasonal}(t) = a + b\ Sin\left[\frac{2\pi}{365}(t + t_0)\right] + \varepsilon(t) \qquad (6)$$

where $Tw_{seasonal}(t)$ is the seasonal component of a temperature time series; a, b and t_0 are fitted coefficients that can be estimated using a nonlinear regression approach. It should be noted that Equation (6) needs to be modified when the sinusoidal function is fitted to a period shorter than one year (e.g., Tasker and Burns, 1974).

The short-term component is generally modelled using a variety of approaches ranging from multiple regression analysis (Kothandaraman, 1971), to a

© 2007 Canadian Water Resources Association

second order Markov chain (Cluis, 1972), or the use of Box-Jenkins time series analysis (Marceau *et al*., 1986). Caissie *et al*. (1998) compared these three different "stochastic" approaches to model mean and maximum daily water temperatures in a relatively small stream (Catamaran Brook, New Brunswick, 50 km^2 drainage area) using air temperature as the independent variable. Their preferred methodology, based on the comparison of goodness of fit (e.g., Nash coefficient) and error statistics (e.g., root mean square error) involved estimating an annual component in stream temperatures by fitting a Fourier series to the data and a second order Markov process model for the short term component. A similar study was also carried out by Caissie *et al*. (2001) in which they modelled maximum daily stream temperatures at the same study site (Catamaran Brook) using a stochastic model. Annual RMSE values for the six modelled years varied between 1 and 1.88°C.

The stochastic approach is a method which requires relatively few parameters and thus its application is simpler. This approach can provide very good results. For instance, Caissie *et al*. (1998) have obtained RMSE less than 0.9°C. Despite the good RMSE, this approach has some potential difficulties. For instance, a fixed sinusoidal function needs to be fitted to the time series. Thus, it can be argued that this may result in non-stationary residuals from year to year. Stationarity (i.e., no seasonality in the data) is one of the underlying hypotheses of a number of time series models, including AR and Box Jenkins. As such, users should verify that residuals are indeed stationary prior to implementing such approaches.

Periodic Autoregressive Models

Stream temperatures and air temperatures are better correlated at the weekly and monthly timescale than at the hourly or daily scale because of the thermal inertia of water bodies (Pilgrim *et al*., 1998). Weekly stream temperature values have been used as criteria to characterize some fish habitat in a number of studies (e.g., Eaton and Scheller, 1996; Stefan *et al*., 2001), although it may be important in some studies to predict water temperature for shorter time scales. As stated in the previous section, the common procedure in modelling such periodic series is to deseasonalize the series prior to applying the stationary models (Salas *et al*., 1980; Vecchia, 1985; Salas, 1993; Chen and Rao, 2002). However, filtering time series may not yield stationary residuals due to periodic autocorrelations. In such cases, the resulting model may be misspecified for series in which periodic properties are present (Tiao and Grupe, 1980). To model periodicity in autocorrelations, periodic models can be advantageous. In such situations, an important class of periodic models consists of Periodic AutoRegressive (PAR) and Periodic AutoRegressive Moving Average (PARMA) models, which are extensions of commonly used ARMA models (Box and Jenkins, 1976) with the difference that the former use periodic parameters. ARMA models assume that the data are stationary (i.e., no seasonality in the data). PARMA models are widely used for prediction of economic time series (Osborn and Smith, 1989; Novales and de Frutto, 1997) as well as in the field of hydrology (Salas *et al*., 1980; Vecchia, 1985; Bartolini *et al*., 1988; Ula and Smadi, 1997). PAR and PARMA models are usually applied to time series at monthly time steps or more, which limits the number of periods to 12 and hence the number of parameters.

Recently, to model the average weekly maximum temperatures (*Tw*), Benyahya *et al*. (2007a) compared the performance of PAR and AR models. The PAR model of lag-*i* used was of the form

$$Tw_{v,\tau} = \mu_\tau + \sum_{i=1}^{p} \phi_{i,\tau}(Tw_{v,\tau-i} - \mu_{\tau-i}) + \varepsilon_{v,\tau} \quad (7)$$

where v is the year; τ is the season (or period); μ_τ is the mean of the water temperature process in season τ; and $\phi_{i,\tau}$ is the autoregressive parameter, which is estimated for each season by using the least square method. The error $\varepsilon_{v,\tau}$ is assumed to be normally distributed with mean zero and variance one. This model was calibrated using 18 years of average weekly maximum temperature series on the Deschutes River (Oregon, United States) and good modelling results have been obtained with a PAR(1) model with an average error (RMSE) less than 1°C, which is similar to the RMSE obtained from an AR(1) model. Benyahya *et al.* (2007a) commented that PAR models would likely perform better with longer time series.

PAR models rely strictly on the autocorrelation structure of the variable of interest (water temperature). For water resource and fisheries managers, it may be important to include other variables that may have

© 2007 Canadian Water Resources Association

an impact on the thermal regime of rivers, such as air temperature and flow. For this reason, the PAR method was extended by Benyahya *et al*. (2007b) to incorporate other input variables (e.g., air temperature, streamflow) called exogenous variables and therefore the PAR model become a PARX model. A PARX model representing the water temperature series may be written in the following form

$$Tw_{v,\tau} = \sum_{i_1=1}^{p_1} \phi 1_{i_1,\tau} Tw_{v,\tau-i_1} + \sum_{i_2=1}^{p_2} \phi 2_{i_2,\tau} Ta_{v,\tau-i_2} + \sum_{i_3=1}^{p_3} \phi 3_{i_3,\tau} Q_{v,\tau-i_3} + \varepsilon_{v,\tau} \quad (8)$$

where $\phi 1_{i_1,\tau}$, $\phi 2_{i_2,\tau}$ and $\phi 3_{i_3,\tau}$ are periodic parameters, p_1, p_2 and p_3 are the lags of water temperature, air temperature and streamflow, respectively, $\varepsilon_{v,\tau}$ is the error term and $\tau \rangle i_1, i_2, i_3$. This model was calibrated using 21 years of weekly water temperatures of the Nivelle River (France) and results indicated that the PARX model performed relatively well with RMSE <1.60°C.

The periodic models are particularly well adapted for weekly data with a good level of performance; however, these models require the estimation of a large number of parameters (i.e., one set of parameters for each period). Depending on the length of the time series available to calibrate the model this can violate the principle of parsimony (e.g., select a model with as few parameters as possible).

Non-Parametric Statistical Models

The second main category of statistical models use the so-called non-parametric approaches. The structure of these models is highly dependent on available data and there is generally no judgement made by the modeller on the statistical structure of the model. Non-parametric models are considered to be good "data-learners" and their use should be limited to the range of values encountered in the past, i.e., their performance in the extrapolation range can be less reliable. Nonetheless, recent advances in computational algorithms and computing power of modern computers have made it possible to implement these models with relative ease. Artificial Neural Networks as well as the k-Nearest Neighbours are examples of such approaches that have been adapted to water temperature modelling.

Artificial Neural Networks

An Artificial Neural Network (ANN) model is a mathematical structure capable of describing complex nonlinear relations between input and output data. The architecture of ANN model is inspired by biological nervous systems (Figure 1). As in nature, independent variables (or predictors) are fed as inputs in the input layer through nodes (the *n* neurons shown on the left of Figure 1) used during neural network training. The hidden layer, represented by the middle circles in Figure 1, is the location where the neural network is "trained," i.e., all outputs from the input layer are fed to each node, and weights are assigned to non-linear functions that combine the inputs (Ahmadi-Nedushan *et al.*, 2007a). The network connection weights are adjusted in order to minimize the error between the ANN outputs and the training set of the variable to be modelled. The weights of each node in the layers need to be adjusted. This can be done using several learning algorithms. One of the most popular learning algorithms is back propagation. In back propagation, a gradient descent is implemented to ensure that the direction of learning and rate of learning is appropriate.

In the field of hydrology, ANN modelling has been used for a variety of purposes, particularly in water quality applications, Conrads and Roehl (1999) used ANN models to simulate salinity, temperature, and dissolved oxygen. Hsu *et al*. (1998) used ANN for streamflow forecasting. Coulibaly *et al*. (2001) applied the temporal neural networks to hydropower reservoir inflow forecasting. Risley *et al*. (2003) estimated water temperatures in small streams in western Oregon using an ANN model. More recently, Belanger *et al*. (2005) compared two models of water temperature: artificial neural networks and multiple linear regression using air temperature and discharge as independent variables. Of these two models, results indicated that both approaches were equally good in predicting daily stream water temperature with RMSE of 1.06°C for the regression model and 1.15°C for the ANN model.

ANN has often been criticized based on the fact that the contribution of the input variables in predicting the output is difficult to disentangle within the network, and explanations regarding the relative importance of each independent variable are not as straightforward as in the case of linear regression methods (Olden and Jackson, 2002). However, trained

© 2007 Canadian Water Resources Association

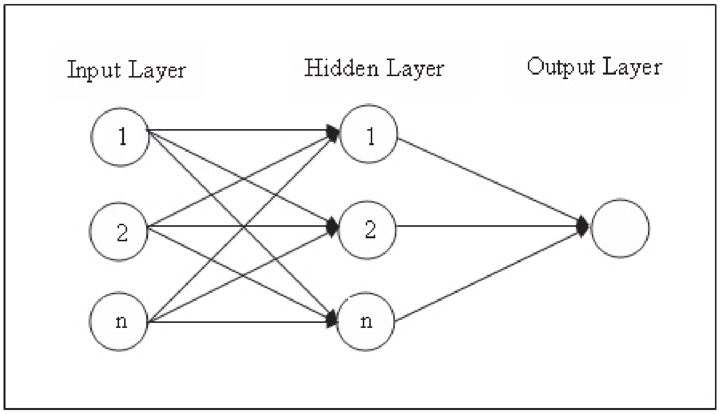

Figure 1. Artificial Neural Network representation.

networks have been shown to perform very well within the interpolation range of the training data set and a number of packages exist that allow for relatively easy implementation of the approach (e.g., Neural Networks toolbox, The Mathworks, 2007; Artificial Neuronal Network 1.00, logiware).

k-Nearest Neighbours (k-NN)

A k-NN is a method that consists of finding, for a given point in time, a small number of neighbours nearest to this value, and the prediction is estimated based on these neighbours. The key steps in the k-NN algorithm are:

Step 1. Compile a feature vector: the vector X consists of values of the selected input attributes (e.g., lagged water temperature, lagged air temperature and streamflow data) for which we are trying to find the k-Nearest Neighbours (usually k<4);

Step 2. Find the weighted sum of the attributes: since the scales of water and air temperature units are different than that of streamflow, the weighted attributes can be generalized as a weighted standardized norm (N)

$$\mathrm{N} = \sum_{i=1}^{N_{attributes}} w_i X_i \qquad (9)$$

where w_i are the weights and X_i are the vectors of standardized (i.e., subtract mean value and divide by the standard deviation) values of the selected attributes.

Step 3. Calculate the Euclidean distance between the norm of the day of interest and the norm of all other available data: For two norms (N_{j1},N_{j2}) calculated using vectors X_{j1} for the day of interest and X_{j2} (j_2 = 1…,j_1 −1, j_1 +1,…,m) for the m other days in the data base, the Euclidean distance (δ) is defined as

$$\delta(\mathrm{N}_{j1}, \mathrm{N}_{j2}) = \left| \mathrm{N}_{j1} - \mathrm{N}_{j2} \right| = \sum_{i=1}^{N_{attributes}} w_i \bullet \left| X_{i,j1} - X_{i,j2} \right| \qquad (10)$$

Step 4. Sort the distances δ in ascending order and retain only the first k nearest neighbours: the strategy for choosing the optimal k is to try several successive values of k (e.g., 2, 3 and 4) and to select the combination for which the model gives the best prediction.

Step 5. Assign weight to each of the k neighbours, thus the predicted value of the final output is computed as a weighted sum of the values of neighbouring observations.

The k-NN is an approach that has been used in hydrology to model rainfall-runoff processes and has been compared with autoregressive moving average models with exogenous inputs (ARMAX) (Karlsson and Yakowitz, 1987; Yakowitz and Karlsson, 1987). In agreement with the conclusions reported by the aforementioned authors, Galeati (1990) showed that

© 2007 Canadian Water Resources Association

the k-NN method provides lower mean square error predictions of daily mean flow than an autoregressive model with exogenous inputs (ARX). Recently, Benyahya *et al*. (2007a) compared the predictive capability of periodic autoregressive model (PARX) and k-NN to model weekly water temperature of the observation period (1984-2004) in the Nivelle River, France. It was concluded that PARX is better suited to model the periodicity in autocorrelations; nevertheless, k-NN is equally an interesting statistical water temperature model. The simulations on the Nivelle River yielded a relatively small root mean square error of 1.20°C. However, one potential drawback of k-NN, and other non-parametric methods such as ANN, is that they do not produce a parametric function of the model and in such case, any given condition not previously observed in the historic record cannot be predicted or simulated.

Conclusions

This paper provided an overview of the existing methods used in water temperature modelling and compared their relative advantages and drawbacks. Both the deterministic and statistical models are relevant depending on the problem under investigation and data availability. Statistical models, which are the main focus of this review, are important tools for the prediction of water temperature based on few input variables. It was found that most of the statistical models reported in the literature are based on simple and multiple linear regression as well as logistic function. These models have been effective in predicting river water temperature for longer time scales, i.e., weekly, monthly and using annual means. For shorter time scales, stochastic models have been more effective in the prediction of river water temperatures. As noted, most statistical methods reviewed in this study present advantages and disadvantages in different contexts (Table 1). The selection of a particular statistical model depends on the modelling objective as well as the type of data available. In situations when the water temperature modelling is carried out at daily time steps and when air temperatures are the only available data, the so-called “stochastic” models have been shown to perform well. AR models applied on residuals (i.e., the non-seasonal component, often called the stochastic component in the literature) are powerful tools that often produce root mean square errors (RMSE) that are less than 1°C. Users should be reminded that stationarity is assumed to exist in the time series of residuals when using this approach.

More recently other statistical approaches have been applied and these statistical models show promising results. For example, when periodicity is observed in the correlation structure, periodic models have been proposed. These models, such as PAR and PARX models, are extensions of commonly used ARMA models with the difference that the former use periodic parameters. Despite of their parameterization, the periodic models were shown to be an interesting tool for modelling weekly water temperatures based on their capability to model periodicity in autocorrelations (Benyahya *et al*., 2007a; b).

The literature shows that non-parametric models are also a promising area of predictive statistical water temperature modelling owing to their capability in simulating the complex nonlinear relationships between response variable and environmental variables. Although non-parametric methods do not provide users with a conventional mathematical function, many readily applicable algorithms exist and their implementation has become easier with the advent of powerful computing technology. Users should be reminded that these approaches are site specific, as is the case in many statistical modelling approaches and their performance capability in the extrapolation domain is usually limited.

Aside from the statistical structure of the data, the selection of a model may be driven by regional considerations. For instance certain statistical approaches may perform better in specific climatic contexts (e.g., rivers with freezing periods versus warmer rivers). A comparative study of water temperature models using a range of sites with different hydroclimatic conditions would be useful.

In conclusion, this literature review shows that the selection of an appropriate statistical water temperature model depends on the comparison of advantages and disadvantages of different methods in a particular context. The context is defined by the following factors:

- Time step required (e.g., daily, weekly, monthly, etc.);
- Length of time series;

© 2007 Canadian Water Resources Association

- Statistical properties of time series (e.g., seasonality, normality of residuals, etc.);
- The need to formalize the relationship between independent variables and water temperature; and
- Hydroclimatic specificities.

Although most of the literature describes statistical model implementations in a simulation mode, these approaches could be used in forecasting as well, especially where forecasts of input variables exist (i.e., air temperature and/or flow). By developing a predictive relationship between water temperature and environmental variables, users could estimate how stream temperatures are likely to respond to these variables, and therefore how the health of many aquatic species could be threatened. Statistical models have played an important role, in the past, in studying water resource and fisheries management issues. It is believed that with the new generation of statistical models this approach will remain important, mainly because water temperature can be predicted with fewer input parameters than with deterministic models.

Table 1. Advantages and disadvantages of statistical models based on approach and identification of related studies.

Modelling Approach	Advantages	Disadvantages	Examples of Water Temperature Modelling Applications
Linear regression	Straightforward in application	Less appropriate when the assumption of a linear relationship cannot be verified.	Pilgrim *et al.* (1998) Erickson and Stefan (2000) Ahmadi-Nedushan *et al.* (2007b)
Logistic function	Appropriate when nonlinear relationship is observed in the data	Performed poorly in some cases when studying daily water temperature time series	Mohseni *et al.* (1998) Caissie *et al.* (2001) Webb *et al.* (2003).
Stochastic model	Appropriate when studying daily water temperature time series. Focuses on adequate modelling of residuals.	Appropriate when residuals are stationary	Cluis (1972) Caissie *et al.* (1998) Caissie *et al.* (2001)
Periodic Autoregressive models	Can capture the periodic autocorrelation in periodic time series	Necessitates the fitting of a greater number of parameters, which may violate the principle of parsimony.	Benyahya *et al.* (2007a; b)
Artificial Neural Network k-Nearest Neighbours	No assumptions concerning statistical distributions and relationships are required; Capability in picking up the complex nonlinear relationships between input and output variables Completely data-driven.	Relatively more costly in computational time; Does not provide users with a conventional mathematical function and physical interpretation. Cannot predict conditions outside of historic range. Requires relatively long time series. Dependency on the selection of a "good value" for k	Risley *et al.* (2003) Belanger *et al.* (2005) Galeati (1990) Benyahya *et al.* (2007a)

© 2007 Canadian Water Resources Association

References

Ahmadi-Nedushan, B., A. St-Hilaire, M. Bérubé, É. Robichaud, N. Thiémonge and B. Bobée. 2007a. "A Review of Statistical Methods for the Evaluation of Aquatic Habitat Suitability for Instream Flow Assessment." *River Research and Applications,* 22(5): 503-523.

Ahmadi-Nedushan, B., A. St-Hilaire, T.B.M.J. Ouarda, L. Bilodeau, É. Robichaud, N. Thiémonge and B. Bobée. 2007b. "Predicting River Water Temperatures Using Stochastic Models: Case Study of the Moisie River (Québec, Canada)." *Hydrological Processes,* 21(1): 21-34.

Bartholow, J.M. 1999. *SSTEMP, Version 1.1.0 USGS Midcontinent Ecological Science Center Geological Survey.* Report 4041. http://www.mesc.gov/rsem/rsm_software.html.

Bartolini, P., J.D. Salas and J.T.B. Obeysekera. 1988. "Multivariate Periodic ARMA(1,1) Processes." *Water Resources Research,* 24: 1237-1246.

Bélanger, M., N. El-Jabi, D. Caissie, F. Ashkar and J.M. Ribi. 2005. "Estimation de la température de l'eau en rivière en utilisant les réseaux de neurones et la régression linéaire multiple." *Revue des sciences de l'Eau,* 18(3): 403-421.

Benyahya, L., A. Saint-Hilaire, T.B.M.J. Ouarda, B. Bobée and B. Nedushan. 2007a. "Modeling of Water Temperatures Based on Stochastic Approaches: Case Study of Deschutes River." *Journal of Environmental Engineering and Science,* 6(4): 437-448.

Benyahya, L., A. St-Hilaire, T.B.M.J. Ouarda, B. Bobée and J. Dumas. 2007b. "A Comparison of Non-Parametric and Parametric Water Temperature Models: Case Study of the Nivelle River (St-Pée, France)." Submitted to *Hydrological Sciences Journal*.

Beschta, R.L., R.E. Bilby, G.W. Brown, L.B Holtby and T.D. Hofstra. 1987. "Stream Temperature and Aquatic Habitat: Fisheries and Forestry Interactions." In *Streamside Management: Forestry and Fishery Interactions*, Salo, E.O. and T.W. Cundy (Eds.). University of Washington, Institute of Forest Resources. Contribution No. 57, 191-232.

Bjornn, T.C. and D.W. Reiser. 1991. "Habitat Requirements of Salmonids in Streams." In Meehan, W.R. (Ed.) *Influences of Forest and Rangeland Management on Salmonid Fishes and their Habitats.* Special Publication 19. American Fisheries Society, Bethesda, MD.

Box, G.E.P. and G.M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. Revised Edition. (2nd Edn.). Holden-Day, San Francisco, CA.

Brown, G.W and J.T. Krygier. 1967. "Changing Water Temperatures in Small Mountain Streams." *Journal of Soil and Water Conservation*, 22: 242–244.

Caissie, D., N. El-Jabi and A. St-Hilaire. 1998. "Stochastic Modelling of Water Temperatures in a Small Stream Using Air to Water Relations." *Canadian Journal of Civil Engineering,* 25: 250–260.

Caissie, D., N. El-Jabi and M. Satish. 2001. "Modeling of Maximum Daily Water Temperatures in a Small Stream Using Air Temperatures." *Journal of Hydrology,* 251: 14-28.

Chen, H.L and A.R. Rao. 2002. "Testing Hydrologic Time Series for Stationarity." *ASCE, Journal of Hydrologic Engineering,* 7(2): 129-136.

Chen, Y.D., R.F. Carsel, S.C. McCutcheon and W.L. Nutter. 1998. "Stream Temperature Simulation of Forested Riparian Areas: 1. Watershed-Scale Model Development." *ASCE, Journal of Environmental Engineering,* 124: 304–315.

Cluis, D.A. 1972. "Relationship Between Stream Water Temperature and Ambient Air Temperature – A Simple Autoregressive Model for Mean Daily Stream Water Temperature Fluctuations." *Nordic Hydrology,* 3(2): 65-71.

© 2007 Canadian Water Resources Association

Conrads, P.A. and E.A. Roehl. 1999. "Comparing Physics-Based and Neural Network Models for Simulating Salinity, Temperature, and Dissolved Oxygen in a Complex Tidally Affected River Basin." In *Proceedings of the 1999 South Carolina Environmental Conference*, Myrtle Beach, March 15-16, 1999: Columbia, South Carolina, Water Environmental Association of South Carolina.

Coulibaly, P., F. Anctil and B. Bobée. 2001. "Multivariate Reservoir Inflow Forecasting Using Temporal Neural Networks." *ASCE, Journal of Hydrologic Engineering,* 6(5): 367-376.

Crisp, D.T. 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.

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(5): 1109-1115.

Erickson, T.R. and H.G. Stefan. 2000. "Linear Air/Water Temperature Correlations for Streams during Open Water Periods." *ASCE, Journal of Hydrologic Engineering,* 5(3): 317-321.

Galeati, G. 1990. "A Comparison of Parametric and Non-Parametric Methods for Runoff Forecasting." *Hydrological Sciences Journal,* 35(1): 79-94.

Gu, R. and Y. Li. 2002. "River Temperature Sensitivity to Hydraulic and Meteorological Parameters." *Journal of Environmental Management,* 66: 43-56.

Gu, R., S. McCutcheon and C.G. Chen. 1999. "Development of Weather Dependent Requirements for River Temperature Controls." *Environmental Management*, 24(4): 529-540.

Gu, R., S. Montgomery and T. Austin. 1998. "Quantifying the Effects of Stream Discharge of Summer River Temperature." *Hydrological Sciences Journal,* 43(6): 885-894.

Hodgson, S. and T.P. Quinn. 2002. "The Timing of Adult Sockeye Salmon Migration into Fresh Water: Adaptations by Populations to Prevailing Thermal Regimes." *Canadian Journal of Zoology,* 80: 542-555.

Horel, A. and R. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics,* 12: 55-67.

Hsu, K., H.V. Gupta and S. Sorooshian. 1998. "Streamflow Forecasting Using Artificial Neural Networks." In *Proceedings of the International Water Resources Engineering Conference*, August 3-7, 1998, Memphis, Tennessee: Reston, Virginia, ASCE, 2: 1362-1367.

Johnson, T.A. 1997. "Downstream Movement of Young-of-the-Year Fishes in Catamaran Brook and the Little Southwest Miramichi River, New Brunswick." *Journal of Fish Biology,* 51: 1047–1062.

Karlsson, M. and S. Yakowitz. 1987. "Nearest-Neighbor Methods for Non-Parametric Rainfall-Runoff Forecasting." *Water Resources Research,* 23: 1300–1308.

Kothandaraman, V. 1971. "Analysis of Water Temperature Variations in Large Rivers." ASCE, *Journal of the Sanitary Engineering Division*, 97(SA1): 19-31.

Krause, C., T.J. Newcomb and D. Orth. 2005. "Thermal Habitat Assessment of Alternative Flow Scenarios in a Tailwater Fishery." *River Research and Applications*, 21: 581-593.

Langan, S.J., L. Johnston, M.J. Donaghy, A.F. Younson, D.W. Hay and C. Soulsby. 2001. "Variation in River Water Temperature in an Upland Stream over a 30-Year Period." *The Science of the Total Environment*, 265: 195-207.

Mackey, A.P. and A.D. Berrie. 1991. "The Prediction of Water Temperatures in Chalk Streams from Air Temperature." *Hydrobiologia*, 210: 183-189.

© 2007 Canadian Water Resources Association

Marceau, P., D. Cluis and G. Morin. 1986. "Comparaison des performances relatives à un modèle déterministe et à un modèle stochastique de température de l'eau en rivière." *Canadian Journal of Civil Engineering,* 13(3): 352-364.

Mohseni, O. and H.G. Stefan. 1999. "Stream Temperature/Air Temperature Relationship: A Physical Interpretation." *Journal of Hydrology,* 218: 128-141.

Mohseni, O., H.G. Stefan and J.G. Eaton. 2003. "Global Warming and Potential Changes in Fish Habitat in U.S. Streams." *Climatic Change,* 59: 389–409.

Mohseni,, O., H.G. Stefan and T.R. Erickson. 1998. "A Nonlinear Regression Model for Weekly Stream Temperatures." *Water Resources Research,* 34(10): 2685–2692.

Morin, G., J.-P. Fortin, J.-P. Lardeau, W. Sochanska and S. Paquette. 1981. "Modèle CEQUEAU : Manuel d'utilisation." Rapport de recherche no R-93, INRS-Eau, Sainte-Foy.

Morin, G. and D. Couillard. 1990. "Predicting River Temperatures with a Hydrological Model." *Encyclopedia of Fluid Mechanics*, 10, Gulf Publishing Company, Hudson, Texas, 171-209.

Novales, A. and R.F. de Frutto. 1997. "Forecasting with Periodic Models: A Comparison with Time Invariant Coefficient Models." *International Journal of Forecasting,* 13: 393-405.

Olden, J.D. and D.A. Jackson. 2002. "Illuminating the 'Black Box': Understanding Variable Contributions in Artificial Neural Networks." *Ecological Modelling,* 154: 135-150.

Osborn, D. and J. Smith. 1989. "The Performance of Periodic Autoregressive Models in Forecasting Seasonal U.K. Consumption." *Journal of Business and Economic Statistics,* 7: 117-127.

Pilgrim, J.M., X. Fang and H.G. Stefan. 1998. "Stream Temperature Correlations with Air Temperatures in Minnesota: Implications for Climate Warming." *Journal of the American Water Resources Association,* 34(5): 1109–1121.

Risley, J.C., E.A. Roehl and P.A. Conrads. 2003. "Estimating Water Temperatures in Small Streams in Western Oregon Using Neural Network Models." *U.S Geological Survey* Report 02-4218.

Salas, J.D., J.W Delleur, V.Yevjevich and E.L. Lane. 1980. *Applied Modelling of Hydrological Time Series*. Water Resources Publications, Littleton, CO.

Salas, J.D. 1993. "Analysis and Modelling of Hydrologic Time Series." In Maidment, D.R., *Handbook of Hydrology.* McGraw-Hill.

Sinokrot, B.A. and H.G. Stefan. 1993. "Stream Temperature Dynamics: Measurements and Modeling." *Water Resources Research,* 29(7): 2299–2312.

Sinokrot, B.A. and J.S Gulliver. 2000. "In-Stream Flow Impact on River Water Temperatures." *Journal of Hydraulic Research,* 38: 339–349.

Smith, K. 1981. "The Prediction of River Water Temperatures." *Hydrological Sciences Bulletin*, 26: 19-32.

St-Hilaire, A., G. Morin, N. El-Jabi and D. Caissie. 2000. "Water Temperature Modelling in a Small Forested Stream: Implication of Forest Canopy and Soil Temperature." *Canadian Journal of Civil Engineering,* 27: 1095-1108.

Stefan, H.G. and E. Preud'homme. 1993. "Stream Temperature Estimation from Air Temperature." *Water Resources Bulletin,* 29(1): 27-45.

Stefan, H.G., X. Fang and J.G. Eaton. 2001. "Simulated Fish Habitat Changes in North American Lakes in Response to Projected Climate Warming." *Transactions of the American Fisheries Society,* 130: 459–477.

© 2007 Canadian Water Resources Association

Tasker, G.D. and W.A. Burns. 1974. "Mathematical Generalization of Stream Temperature in Central New England." *Water Resources Bulletin*, 10: 1133-1142.

Tiao, G.C. and M.R. Grupe. 1980. "Hidden Periodic Autoregressive Moving Average Models in Time Series Data." *Biometrika,* 67: 365-373.

Ula, T.A. and A.A. Smadi. 1997. "Periodic Stationarity Conditions for Periodic Autoregressive Moving Average Processes as Eigen-Value Problems." *Water Resources Research*, 33: 1929-1934.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell and C.E. Cushing. 1980. "The River Continuum Concept." *Canadian Journal of Fisheries and Aquatic Sciences,* 37: 130–137.

Vecchia, A.V. 1985. "Periodic Autoregressive-Moving Average (PARMA) Modeling with Applications to Water Resources." *Water Resources Bulletin,* 21: 721-730.

Webb, B.W. and D.E. Walling. 1993. "Temporal Variability in the Impact of River Regulation on Thermal Regime and Some Biological Implications." *Freshwater Biology,* 29: 167–182.

Webb, B.W. and F. Nobilis. 1997. "Long-Term Perspective on the Nature of the Air-Water Temperature Relationship: A Case Study." *Hydrological Processes,* 11(2): 137-147.

Webb, B.W., P.D. Clack and D.E. Walling. 2003. "Water-Air Temperature Relationships in a Devon River System and the Role of Flow." *Hydrological Processes,* 17: 3069-3084.

Yakowitz, S. and M. Karlsson. 1987. "Nearest-Neighbor Methods with Application to Rainfall/Runoff Prediction." In Macneill, I.B. and G.J. Umphrey (Eds.) *Stochastic Hydrology.* D. Reidel, Norwell, MA.

Younus, M., M. Hondzo and B.A. Engel. 2000. "Stream Temperature Dynamics in Upland Agricultural Watershed." *ASCE, Journal of Environmental Engineering,* 126: 518–526.

© 2007 Canadian Water Resources Association