Regression Model for Daily Maximum Stream Temperature

David W. Neumann[1]; Balaji Rajagopalan[2]; and Edith A. Zagona[3]

Abstract: An empirical model is developed to predict daily maximum stream temperatures for the summer period. The model is created using a stepwise linear regression procedure to select significant predictors. The predictive model includes a prediction confidence interval to quantify the uncertainty. The methodology is applied to the Truckee River in California and Nevada. The stepwise procedure selects daily maximum air temperature and average daily flow as the variables to predict maximum daily stream temperature at Reno, Nev. The model is shown to work in a predictive mode by validation using three years of historical data. Using the uncertainty quantification, the amount of required additional flow to meet a target stream temperature with a desired level of confidence is determined.

DOI: 10.1061/(ASCE)0733-9372(2003)129:7(667)

CE Database subject headings: Decision support systems; Regression models; California; Nevada; Streams; Water quality; Water temperature.

Introduction

An increasingly common problem in western U.S. river basins and elsewhere in the world is that water storage and use for municipal, industrial, agricultural, and power production purposes leaves river biota with insufficient flow to maintain populations. Low flows threaten biota by deteriorating habitat and/or water quality. One of the most common summer water quality problems associated with low flows is high stream temperatures—low flows warm up more rapidly than higher flows. High stream temperatures reduce cold water fish populations by inhibiting growth and by killing fish at extremely high temperatures. For this reason, the impact of low flows and high stream temperatures on fish is an issue in many operations studies and National Environmental Policy Act Environmental Impact Statement analyses such as those on the Rio Grande, Colo., and Columbia basins (Bonneville 1995; U.S. Bureau of Reclamation 1995, 2000).

Resource managers use computer models to simulate river and reservoir operations. Computer simulations are useful to allow water managers to investigate the effects of varying inflows, legal policies, and operating strategies. To address the problem of warm stream temperatures, resource managers need to incorporate stream temperature objectives in their operations models and management decisions. This requires the ability to predict stream temperature. Because the prediction will be used in daily operating decisions, the prediction must meet the following specific requirements: it must be quick, accurate, easy to use, and spatially and temporally consistent with the operations models. To incorporate stream temperature in the operations model, the normal operating policies are simulated and the stream temperature is predicted. Based on the prediction, decisions can be made to release additional water, if necessary, to improve the stream temperature. As various researchers explain (Beck 1987; Reckhow 1994; Varis et al. 1994), the uncertainty of any prediction should be quantified for decision making purposes. Thus, the temperature prediction should also include a quantification of the uncertainty.

Two types of models have been developed in the past to predict stream temperatures: empirical or regression models and physical process models. Regression models have been developed to quantify and predict stream temperatures at various time scales. Mohseni et al. (1998) developed an S-shaped regression model to predict average weekly stream temperatures at different locations in the United States that account for hysteresis throughout a year. Mohseni et al. (2002) also developed statistical upper boundaries for weekly stream temperatures, noting that in the upper part of the S curve, increasing air temperature results in constant stream temperatures due to back radiation and evaporation. They showed that for an arid western U.S. desert region, the maximum weekly stream temperature is as high as 33°C. Hockey et al. (1982) developed a daily regression model relating spot mid-day stream temperature to flow rate and daily maximum air temperature. They concluded that their regression was not adequate because of lack of data. Gu et al. (1999) produced stream temperature regression models for various weather conditions. They found that correlation of flow to river temperature is possible and useful when weather parameters are decoupled from the model.

In contrast to regression based models, many physical process models have been developed. Physical process models attempt to model the underlying processes that affect stream temperature such as channel geometry, conduction, radiation, advection, and dispersion. Among various work, Taylor (1998); Carron and Rajaram (2001); and Brock and Caupp (1996) developed stream temperature models using mechanistic one- or two-dimensional heat advection/dispersion transport equations. Although a mechanistic temperature model could, in theory, give very accurate results, this type of model requires numerous detailed input data, is

[1]Water Resource Engineer, Stetson Engineers Inc., 2171 F. Francisco Blvd., Suite 11, San Rafael, CA 94965. E-mail: davidn@stetsonengineers.com

[2]Assistant Professor, Dept. of Civil, Environmental, and Architectural Engineering, Univ. of Colorado, UCB 426, Boulder, CO 80309-0426. E-mail: rajagopalan.balaji@colorado.edu

[3]Director, Center for Advanced Decision Support for Water and Environmental Systems (CADSWES), Univ. of Colorado, UCB 421, Boulder, CO 80309-0421. E-mail: zagona@colorado.edu

Note. Associate Editor: Robert G. Arnold. Discussion open until December 1, 2003. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on February 26, 2002; approved on August 13, 2002. This paper is part of the ***Journal of Environmental Engineering***, Vol. 129, No. 7, July 1, 2003. ©ASCE, ISSN 0733-9372/2003/7-667–674/$18.00.

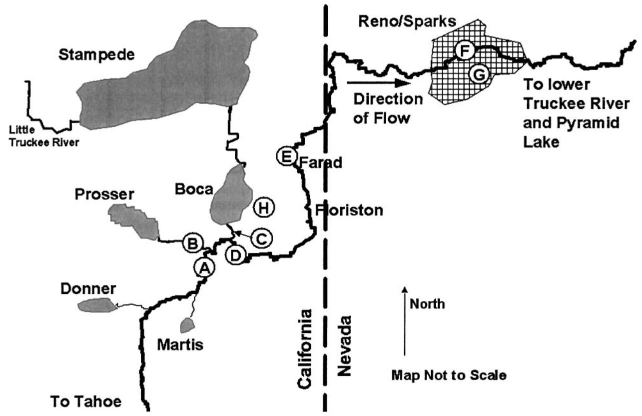

Fig. 1. Diagram of study section

computationally intensive and is, therefore, difficult to incorporate in a river and reservoir operations model. Empirical, regression based models can be computationally less intensive, therefore quick to implement and easy to validate. With regression models it is possible to easily quantify the uncertainty.

In this paper, we develop a regression model to predict low flow summer stream temperatures on the Truckee River at Reno. The model is developed using a stepwise linear regression procedure that selects the significant predictors. The regression model provides uncertainty estimates using standard linear regression theory. We develop a strategy to use the uncertainty information to determine the additional flow required to meet a temperature target with a given confidence level.

This paper is organized as follows. We present the water quality issues on the Truckee River. Next, we describe the development of the regression model and present statistical model diagnostics. We validate the model using historical data and present strategies to use the uncertainty of the prediction. Finally, we discuss the results and summarize the findings.

Truckee River Background

The methodology developed is applied to the Truckee River in California and Nevada. The Truckee River, like other basins in the western U.S., does not have the water resources to meet agricultural, municipal, and industrial purposes and still provide adequate habitat for fish. The Truckee River flows 187 km from Lake Tahoe in California's Sierra Nevada mountains through an arid desert before terminating in Nevada's Pyramid Lake. The upstream reservoirs, shown in Fig. 1, are operated to meet the Floriston rates, a flow target measured at the Farad gage on the California and Nevada border. The flow target, which dictates many of the release and storage decisions in the basin, varies between 8.5 and 14.2 m^3/s (300–500 cfs) depending on the time of year and the reservoir levels. The rates were established in the 1935 Truckee River Agreement to meet quantity requirements for irrigation and power production purposes. On average, the Floriston rates are higher than natural flows in the summer and fall and lower in the winter and spring. Sometimes, stream flows are lower than Floriston rates because of lack of unregulated inflows and available stored water. The low flows result in temperatures in the lower river that are too warm during the summer months for threatened cold water fish. In accordance with the 1996 Water Quality Settlement Agreement (WQSA), the federal government will purchase water rights that will be used to improve the water quality of the Truckee River, particularly in the lower reaches where the river flattens out in the desert between Reno and Pyramid Lake. The water acquired by the WQSA will be stored in upstream reservoirs and released as necessary to mitigate downstream water quality problems. In particular, this WQSA water will be released on a daily basis to meet a target daily maximum stream temperature. The stream temperature of the Truckee River between the confluence with the Little Truckee River and Reno is influenced mainly by natural warming, that includes solar radiation and conduction. Downstream of Reno, wastewater effluent and irrigation return flows enter the river, making accurate temperature predictions much more complex and uncertain. As a first step to improve Truckee River water quality, this paper investigates the temperature at Reno. A diagram of the study section is shown in Fig. 1.

Stream Temperature Model

The goal of regression models is to fit a set of data with an equation, the simplest being a linear equation. The linear regression model takes the form

$$\hat{T}=a_0+a_1x_1+a_2x_2+\ldots+a_nx_n \tag{1}$$

where $\hat{T}$=stream temperature; $a_0,a_1,a_2,\ldots,a_n$=coefficients; and $x_1,x_2,\ldots,x_n$=independent predictors.

The available data are summarized in Table 1 with the locations of the gaging sites shown in Fig. 1. Most of the temperature data were collected after 1993. Since 1993 and 1994 were dry years with low flows and high river temperatures, the same conditions that the prediction will be used, these are the most appropriate years to use in the empirical relationships. In addition, only data from June, July, and August will be used. We did not include September because the river cools in the latter half of the month. It is likely that the model developed will be applicable to the first half of September. We chose to look at data for which the flow at Farad was less than 14.2 m^3/s (500 cfs) because at flows above this threshold, there is rarely a temperature problem in the study reach. Also, 14.2 m^3/s (500 cfs) is a logical cutoff because, according to U.S. Bureau of Reclamation water managers (Scott, personal communication, 2001), additional water to mitigate temperature problems will not be released when the flow at Farad is above the legal flow target of 14.2 m^3/s (500 cfs).

Candidate predictors for the stream temperature prediction at Reno include:

1. Previous day's daily maximum stream temperature at Reno (location F);
2. Daily maximum stream temperature at the Truckee River below the confluence with Little Truckee River (location D);
3. Daily maximum air temperature at Reno (location G);
4. Daily maximum air temperature at Boca (location H);
5. Average daily flow at Reno (location F);
6. Average daily flow at Farad gage (location E); and
7. Daily maximum release temperature from Boca (location C).

The first predictor variable is not useful for the daily operations purposes. Although historically the stream temperature on any day is closely related to the stream temperature on the previous day, once water is released to affect the temperature, that relationship will be changed. For example, the previous day's temperature may be below the target but only because additional water was released. This corrected temperature is not related to the current day's temperature unless an equivalent flow is released. Therefore, the previous day's stream temperature cannot be used in the predictive model.

Table 1. Available Relevant Data

Schematic locator	Location	Data available	Collection period
A	Truckee River above Prosser Creek (USGS 10348000)	Average daily flow	3/1993–9/1998
		Daily maximum stream temperature	3/1993–9/1998
		Hourly stream temperature	6/1993–10/1994
B	Prosser Creek below Prosser (10340500)	Average daily flow	1/1942–current
		Daily maximum stream temperature	3/1993–9/1998
		Hourly stream temperature	6/1993–10/1994
C	Little Truckee River below Boca (10344500)	Average daily flow	6/1980–current
		Daily maximum stream temperature	4/1993–9/1998
		Hourly stream temperature	6/1993–10/1994
D	Truckee River below Little Truckee River confluence	Average daily flow	7/1994–10/1994
		Hourly stream temperature	7/1994–10/1994
E	Truckee River at Farad (10346000)	Average daily flow	1/1909–current
		Daily maximum stream temperature	4/1980–9/1998
		Hourly stream temperature	7/1993–10/1994
F	Truckee River at Reno (10348000)	Average daily flow	7/1906–current
		Daily maximum stream temperature	8/1989–9/1998
		Hourly stream temperature	1/1994–11/1994
G	Reno Airport	Daily maximum air temperature	1/1986–12/1996
H	Near Boca Reservoir	Daily maximum air temperature	1/1986–12/1996

Predictor 2 is not an observed quantity; rather, it is a flow-weighted average of historical temperature observations at A, B, and C in Fig. 1. It is computed as

$$T_D=\frac{T_AQ_A+T_BQ_B+T_CQ_C}{Q_A+Q_B+Q_C} \tag{2}$$

where T_i = temperature of the water at location i; and Q_i = flow at location i. Eq. (2) represents a conservation of heat assuming there are no additional heat sources or sinks.

Fig. 2 shows scatter plots of the predictors and the daily maximum stream temperature at Reno along with a locally weighted regression curve (Loader 1999) through the scatter. The figure shows there is a strong positive correlation between air temperature and stream temperature, and a negative correlation between flow and stream temperature. These results are as expected. Higher flow leads to lower stream temperatures and warm air temperatures lead to warmer water temperatures. Also, there is a strong correlation between upstream stream temperatures (Boca release and location D) and stream temperatures at Reno. Since it appears that all of these predictors are related to Reno water temperatures, the goal is to select the best subset of predictors that explain the most variability in the stream temperature.

A stepwise regression procedure is used to select the best subset of predictors from the candidate predictors. The stepwise procedure selects the subset of predictors optimizing on one the following indicator statistics: Mallow's *Cp*, Akaike's Information Criteria (AIC), R^2, or adjusted R^2. The AIC and *Cp* statistics are widely used because they try to achieve a good compromise between the desire to explain as much variance in the predictor variable as possible (minimize bias) by including all relevant predictor variables, and to minimize the variance of the resulting

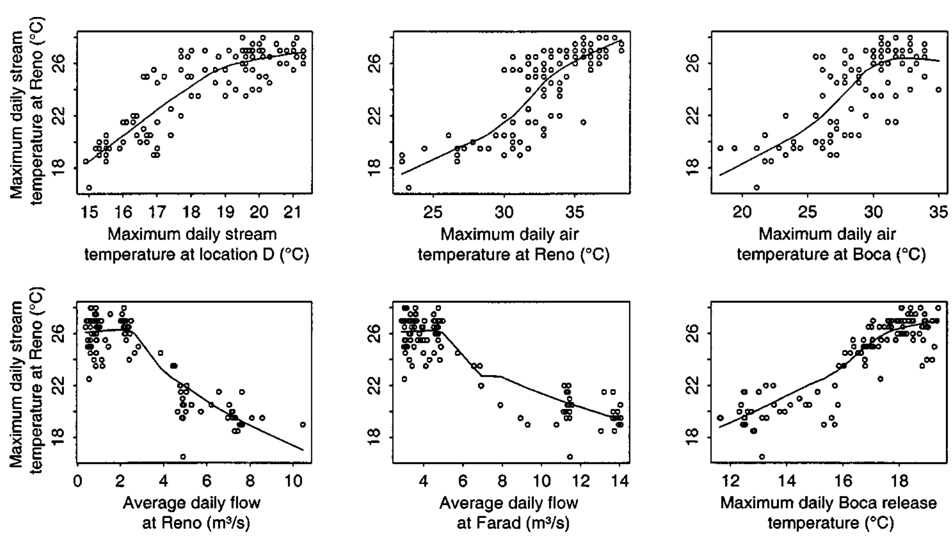

Fig. 2. Data used in regression relationships

Table 2. Stepwise Selection to Find Daily Maximum Stream Temperature at Reno

Stream temperature at Reno =	AIC value		
	f(variable in column 1)	*f*(flow at Farad, variable in column 1)	*f*(flow at Farad, Reno air *T*, variable in column 1)
Constant	1,016	239	140
Stream temperature at location D	309	198	153
Air temperature at Reno	379	**140**	—
Air temperature at Boca	500	190	159
Flow at Reno	278	250	158
Flow at Farad	**239**	—	—
Boca release temperature	244	225	155

estimates (minimize the standard error) by keeping the number of coefficients small. The stepwise regression procedure fits all possible combinations of predictors and selects the model that results in the most optimal indicator statistic.

The AIC statistic, the likelihood version of the Cp statistic (*S-Plus 5 for UNIX Guide to Statistics* 1998, p. 153), is calculated as

$$\mathrm{AIC}=\hat{\sigma}^2(Cp+n) \tag{3}$$

and the Cp statistic is

$$Cp=p+\frac{(n-p)(s_p^2-\hat{\sigma}^2)}{\hat{\sigma}^2} \tag{4}$$

where n = number of observations; p = number of explanatory variables plus one (for the constant in the regression equation a_0); s_p^2 = mean square error of each p coefficient model; and $\hat{\sigma}^2$ = best estimate of the true error (Helsel and Hirsch 1992, p. 312). The adjusted R^2 is calculated as

$$\text{adjusted } R^2=1-\frac{s_p^2}{[(\mathrm{SS}_y)/(n-1)]} \tag{5}$$

where SS_y = total sum of squares.

The AIC statistic is used because it further rewards for having a low mean square error while penalizing for including too many variables. We performed a stepwise procedure on the set of predictor variables listed above, optimizing on AIC. Table 2 shows the AIC values for the stepwise procedure which indicate that air temperature at Reno and flow at Farad are the significant predictors.

A linear regression using the predictors selected has the following equation:

$$\hat{T}=a_0+a_1T_{\mathrm{Air}}+a_2Q \tag{6}$$

where T_{Air} = air temperature at Reno; and Q = flow at Farad. The regression coefficients are $a_0=14.4°\mathrm{C}$, $a_1=0.40$, and $a_2=-0.49°\mathrm{C/m^3/s}$. The adjusted R^2 for this regression is 0.91. Fig. 3 shows the estimated values of maximum daily Truckee River temperature at Reno from the regression equation plotted against the historical observations. The dotted line represents the best fit.

We also performed a stepwise selection procedure optimizing on the adjusted R^2 and Cp statistic as the indicator statistic. In addition to flow at Farad and air temperature at Reno, the stepwise procedure selected the flow at Reno and the stream temperature at the Truckee River below the Little Truckee River confluence (location D). This model has an adjusted R^2 of 0.92 which is not significantly different than the R^2 in the regression described by Eq. (6). Because the R^2 values are similar, it is more efficient to use the model with the smallest number of predictor variables. Therefore, the predictive temperature model described in Eq. (6) is used.

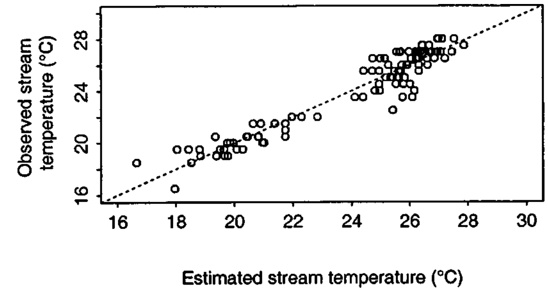

Fig. 3. Estimated versus observed daily maximum stream temperature for Truckee River at Reno

Although Boca's release temperature does have an impact on the Truckee River, the stepwise regression did not select this variable. This indicates that the prediction site at Reno is far enough downstream from the reservoir that air temperature and flow are the dominating factors. This assumes that the reservoirs are deep enough such that water released out of the bottom of the reservoir is cold. If the reservoir depth becomes too low, the regression developed is not valid.

The model is consistent with earlier work by Brock and Caupp (1996) in which they used air temperature and river flow to predict Truckee River temperatures to get the upstream boundary condition at Reno for their Dynamic Stream Simulation and Assessment Model; temperature (DSSAMt).

A local nonlinear regression model (Loader 1999) was also fit to the data using the predictors selected in the linear stepwise procedure. We tried local spans ranging from 0.05 to 0.95. The span that produced the highest R^2 value (0.96) was 0.95. The R^2 is very similar to the R^2 found from the linear model. Because the linear model is simpler and allows for easy uncertainty computations, we use the linear model. For other basins or predictors, a nonlinear local regression fit may be necessary to produce a reasonable fit.

Model Diagnostics

To investigate the performance of the regression model, we look at the following diagnostics: normality of the residuals, autocorrelation of the residuals, and cross validation of the data. Linear

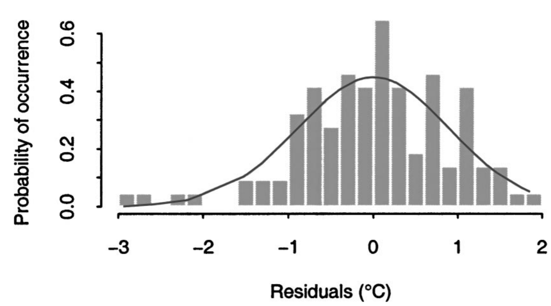

Fig. 4. Reno water temperature regression residuals histogram

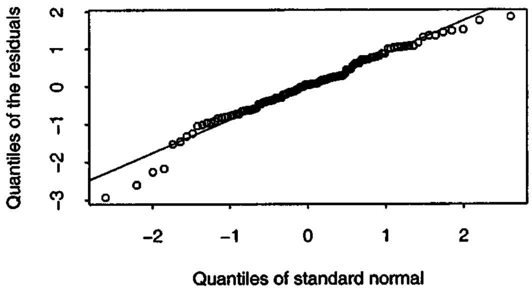

Fig. 5. Quantile versus quantile plot to test for normality

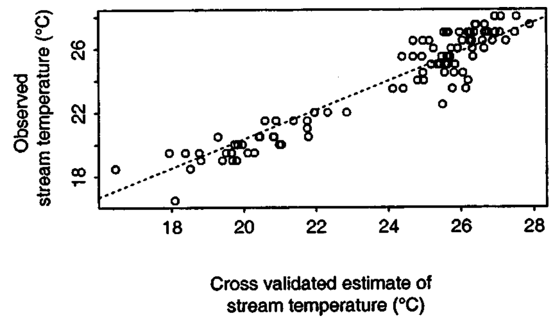

Fig. 7. Cross validation of daily maximum stream temperature regression, Truckee River at Reno

regression theory assumes residuals are normally distributed and symmetric about the mean. A histogram of the residuals, Fig. 4, shows that the residuals of the Reno water temperature estimates appear to be normally distributed, centered around zero. We can quantify whether or not this distribution is Gaussian by looking at Fig. 5 which shows the quantiles of the residuals versus the quantiles of a normal distribution. If the points fall on the line, the distribution is normal. To formally test for normality, a correlation is computed between the residual and normal quantiles. For the distribution to be normal, the correlation must be greater than or equal to the 95% confidence level, critical probability plot correlation coefficient in Helsel and Hirsch (1992). The correlation for our data is 0.987 and the critical value for a 95% confidence level and 108 observations is 0.987. Therefore, the residuals are significantly normal.

One of the assumptions of linear regression theory is that the residuals have no autocorrelation. Fig. 6 shows the autocorrelation function (ACF) plot of the residuals. The dotted lines are the 95% confidence lines. If no ACF estimates fall outside the 95% confidence limit, one can safely assume there is no autocorrelation. The autocorrelation plot in Fig. 6 shows that there is some autocorrelation between the residuals at lag 1 but shows no clear trends. A lack of autocorrelation in the residuals together with the normality of their distribution indicates that the model fits the data well.

To further test the regression, a cross validation technique is used. In cross validation, one historical observation is dropped from the fitting process and is predicted using the regression fit based on the remaining observations. This is repeated for all observations. The cross validated estimates are plotted against the actual observations in Fig. 7. The R^2 value between the cross validated estimates and observed values is 0.91, which is quite good. This further shows that the relationship fits the data well. This R^2 value is slightly less than the regression fitting R^2 because the cross validation is more of a predictive mode.

Model Verification

An empirically developed multiple linear regression model may fit the data used to estimate the regression coefficients very well, but its ability to predict new data is not certain. We validate the model using observations not used in fitting the regression to assess the ability of the model to predict future events.

Fig. 8 shows the predicted and observed daily maximum stream temperature at Reno for June, July, and August of 1990, 1991, and 1992. The predicted temperatures are from Eq. (6). Missing predictions indicate that the Farad flow was greater than 14.2 m^3/s (500 cfs). The R^2 value for each year is also shown in Fig. 8. The R^2 values found in this validation process are lower than the fitting procedure which is consistent with linear regression theory. Fig. 3 shows that there are two regions in the fitting procedure, the range below 23°C has less scatter than the range above 23°C. In other words, the regression is better at explaining variance below 23°C than above. As a result, the skill in predicting temperatures below 23°C is better. The temperature prediction model will be used to try to meet a temperature target at Reno of approximately 22°C by releasing additional water. Therefore, the regression and confidence intervals are valid in the range in which the prediction will be used.

Uncertainty of Predicted Temperatures

Now that we have created a stream temperature model, we need to quantify the uncertainty. Helsel and Hirsch (1992, p. 300) define the *confidence interval* as the range ($\pm$ the mean) of values in which the mean of estimates by regression will lie. For example, the 95% confidence interval indicates that 95% of the time, the

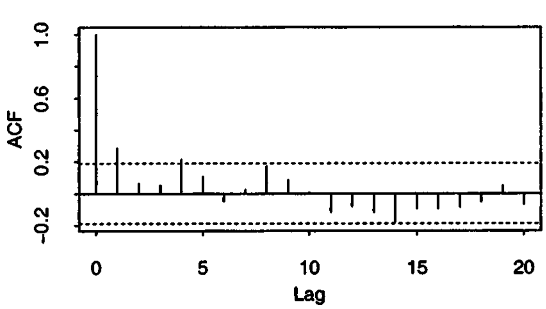

Fig. 6. Reno water temperature regression residuals autocorrelation

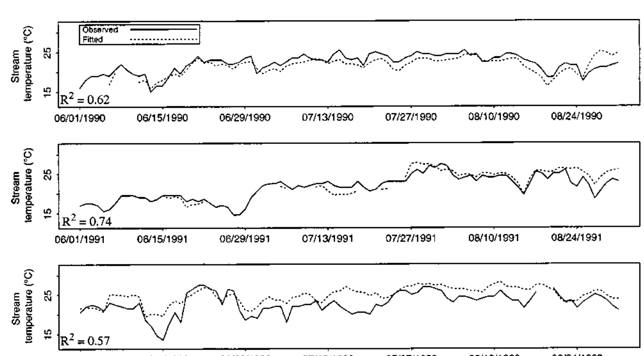

Fig. 8. Validation of daily maximum stream temperatures

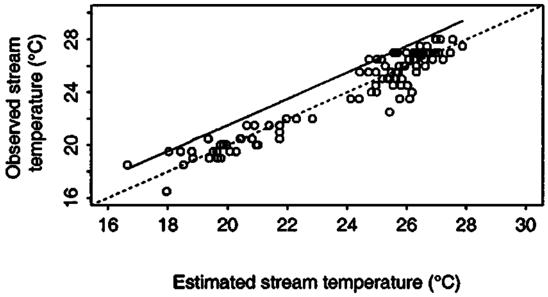

Fig. 9. Estimated versus observed daily maximum stream temperature for Truckee River at Reno

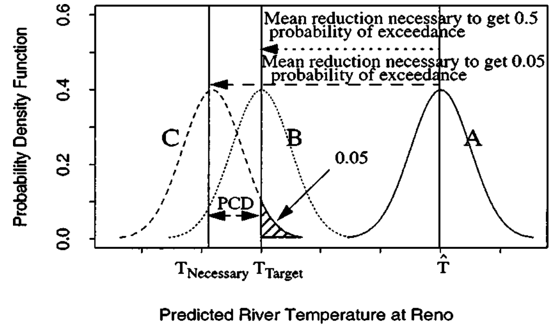

Fig. 10. Temperature reduction to meet desired exceedance probability

mean estimated response variable will be within the interval. A similar concept called the *prediction interval* is used in a predictive model. The prediction interval is defined as "the confidence interval for prediction of an estimate of an individual response variable." For example, the 95% prediction interval indicates that 95% of the time the predicted value will be within the interval.

Linear regression theory provides the prediction interval to be (Helsel and Hirsch 1992, p. 300)

$$\text{Prediction Interval}=[\hat{y}-t(\alpha/2,n-p)\sigma\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0},\quad \hat{y}+t(\alpha/2,n-p)\sigma\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}] \tag{7}$$

where $t(\alpha/2,n-p)$ = quantile given by the $100(\alpha/2)$ percentile on the student's t-distribution having $n-p$ degrees of freedom (Ang and Tang 1975, p. 237). At large degrees of freedom $(n-p)$ the students t distribution is identical to a Gaussian distribution. The desired confidence level is $1-\alpha$. There are n observations used to create the regression and p explanatory variables plus one (for the intercept term). The standard deviation of the residuals is σ; $\mathbf{x}_0$ is the vector $\{1,x_1,x_2,\ldots,x_p\}$ where $x_1,x_2,\ldots,x_p$ are the predictor variables. The matrix $\mathbf{X}$ consists of a column of ones and the matrix of the new observations of predictor variables:

$$\mathbf{X}=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{8}$$

Because the prediction is for summer only, we are only concerned with an upper boundary. By evaluating the student's t distribution at α instead of $\alpha/2$, we get the upper limit to be

$$\text{Prediction Upper Limit}=\hat{y}+t(\alpha,n-p)\sigma\sqrt{1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0} \tag{9}$$

This means that with $100(\alpha)$% confidence, Eq. (9) is the upper limit for the predicted value at $\mathbf{x}_0$. Using historical data, an upper prediction interval can be computed for the full range of predictor variables. Fig. 9 shows the dotted regression line from Eq. (6) and the solid 95% confidence upper prediction interval line from Eq. (9). The upper prediction interval is approximately 1.5°C from the dotted, best fit line. Most of the observations are below the upper prediction interval line as expected. Lowering the prediction confidence below 95% would move the upper prediction interval closer to the fitted regression line (i.e., the dotted line).

Like a confidence interval, the prediction interval is smaller near the center of the data and larger toward the edges. However, we can assume that the prediction interval is linear. This assumption is valid because the second term under the square root, $\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0$, in Eq. (7) is small compared to the first term, 1, provided the sample size is large (Helsel and Hirsch 1992, p. 242). This leads to an approximation of the prediction interval as

$$\text{Prediction Interval}=[\hat{y}-t(\alpha,n-p)\sigma,\quad \hat{y}+t(\alpha,n-p)\sigma] \tag{10}$$

Prediction Confidence Distance

As the stream temperature model in Eq. (6) includes flow as a predictor, we can release additional water to lower warm stream temperatures. The operations approach is as follows: determine reservoir releases based on baseline operating policies, predict the stream temperature using Eq. (6). If the predicted stream temperature is too high, release additional water to meet a target temperature. The regression and the prediction upper interval can be used to determine a strategy to determine how much additional water to release.

To this end, we develop the variable called the prediction confidence distance (PCD). Fig. 10 illustrates this concept. Using the regression model, Eq. (6), we predict a stream temperature $\hat{T}$ and its associated Gaussian distribution denoted by curve A. $\hat{T}$ is too warm and may adversely affect fish. By releasing more water, we can shift the distribution to the left. If the expected value of the distribution is shifted to the target temperature, T_{Target}, as shown by curve B, the probability of exceeding that target is 0.5. Shifting the distribution to the left of the target temperature, a distance defined as PCD, such that the distribution gives a specified probability of exceeding the target temperature. Curve C shows the distribution that results by shifting the distribution to $T_{\text{Necessary}}$, which is the target minus the PCD such that the distribution gives 0.05 probability of exceeding T_{Target}. The PCD is defined as the distance from the mean to the prediction interval as in Eq. (10):

$$\text{PCD}=t(\alpha,n-p)\sigma \tag{11}$$

The empirical regression formula to predict stream temperature from flow and air temperature, Eq. (6), is used to determine the additional water required to lower the temperature such that the probability of exceeding the target is as specified. The predicted daily maximum air temperature is given; thus, the only controlling variable that can influence Truckee River temperature is flow. Rearranging Eq. (6) to solve for flow gives

$$Q=\frac{\hat{T}-a_1T_{\text{Air}}-a_3}{a_2} \tag{12}$$

Table 3. Additional Flow Required at Farad to Reduce Daily Maximum River Temperature to Target of 22°C

Predicted temperature (°C)	Probability of exceedance					
	0.05	0.10	0.20	0.30	0.40	0.50
22	3.2	2.5	1.6	1.0	0.5	0
23	5.4	4.7	3.8	3.2	2.7	2.2
24	7.6	6.9	6.0	5.4	4.8	4.4
25	9.8	9.1	8.2	7.6	7.1	6.5
26	11.9	11.2	10.4	9.7	9.2	8.7
27	14.1	13.4	12.5	11.9	11.4	10.9
28	16.3	15.6	14.7	14.1	13.6	13.1

Note: Values in table are additional flow required (m^3/s).

where T_{Air} = predicted air temperature at Reno; Q = flow at Farad; and $\hat{T}$ = target water temperature at Reno. Evaluating Eq. (12) with $T_{\text{Necessary}}$ as $\hat{T}$, we get the required flow at Farad

$$Q_{\text{Required}}=\frac{T_{\text{Necessary}}-a_1T_{\text{Air}}-a_3}{a_2} \quad (13)$$

Rearranging Eq. (13), the necessary temperature at Reno, can be expressed as

$$T_{\text{Necessary}}=a_0+a_1T_{\text{Air}}+a_2Q_{\text{Required}} \quad (14)$$

Subtracting Eq. (14) from Eq. (6) gives

$$\hat{T}-T_{\text{Necessary}}=a_2(Q-Q_{\text{Required}}) \quad (15)$$

Rearranging, we get the additional flow required at Farad

$$(Q_{\text{Required}}-Q)=\frac{\hat{T}-T_{\text{Necessary}}}{-a_2} \quad (16)$$

To make this more general, we can also define $T_{\text{Necessary}}$ as in Fig. 10

$$T_{\text{Necessary}}=T_{\text{Target}}-\text{PCD} \quad (17)$$

We can replace $T_{\text{Necessary}}$ in Eq. (16) with Eq. (17) to get

$$\Delta Q=\frac{\hat{T}-T_{\text{Target}}+\text{PCD}}{-a_2} \quad (18)$$

In the example illustrated in Fig. 10, the predicted stream temperature calculated from Eq. (6) based on baseline operations, at Reno is $\hat{T}$. We want to lower the temperature to a target, T_{Target}, with probability of exceedance P. The PCD for P exceedance is given by Eq. (11). To find the additional flow required at Farad we enter the predicted temperature $\hat{T}$, the target temperature T_{Target}, and the PCD into Eq. (18). The result ΔQ is the additional flow that must be released to reduce the stream temperature to the target with the specified probability of exceeding the target stream temperature. To use a different probability of exceedance, the confidence level in the PCD calculation can be modified.

A lookup table was developed for each target temperature for easy use in a decision support system. For a target temperature, the table has the initial predicted temperature on one axis and the probability of exceedance on the other axis. The values in the table are the additional flow necessary to reduce the temperature to the target as calculated by Eq. (18). Table 3 shows additional flows needed for a target temperature of 22°C. The table works as follows. The expected water temperature at Reno is predicted using the regression Eq. (6). This value is found in the first column, and the additional flow needed is found in the desired probability of exceedance column. Linear interpolation can be performed between rows if necessary. The additional flow required for a probability of exceedance of 0.5 at the predicted value equal to the target value of the table is zero because the mean predicted value is the target value. But, additional flow is required if a more confident prediction is required.

Discussion and Interpretation of Results

The stepwise selection procedure creates a standardized process to select the most relevant predictors. This is useful when there are large amounts of data that appear to be related to the stream temperature. For summer Truckee River stream temperatures, the most significant predictors are flow and air temperature. The stream temperature prediction model fits the historic data well ($R^2=0.9$) and fits the verification period relatively well. A more accurate, less simple model could be developed, particularly for the high temperature range. The relationships in this study were strongly linear, therefore linear regression is adequate. In other studies, nonlinear techniques that can capture the dependence structure are attractive and should be explored. Further data and monitoring will help to improve the relationship to make it more certain. Less water will be necessary to meet the temperature targets with the desired probability of exceedance allowing water to be saved for the future.

The structure of the linear prediction model lends itself to relatively easy computation of uncertainties of the prediction. Using the uncertainty, the additional flow required can be calculated such that the probability of exceeding a target temperature is as desired. This is useful as decision makers can use varying probability of exceedances to determine how much water to release. They might decide that on a given day they must meet the temperature target with a high degree of certainty and will set the probability of exceedance very low. Or, they might decide they only have minimal confidence in the prediction and will, therefore, not release as much water. The structure of the prediction leads to flexibility of operations.

There is another aspect of uncertainty that could be explored. In an operations mode, the prediction model will use a forecasted dialy maximum air temperature. Currently, the National Weather Service and other agencies provide a single value for the air temperature forecast. As a result, the uncertainty of the stream temperature regression does not take the uncertainty of the air temperature forecast into account. To include this information in the uncertainty, one could assume a distribution of predicted air temperatures and perform Monte Carlo analysis to simulate ensembles of predicted air temperatures. Consequently, ensembles of stream temperature predictions can be obtained from the regression.

The effect of different confidence levels, use of climate information, and the effect of using information about the previous days stream temperatures on future stream temperatures and volume of water necessary are further explored by Neumann (2001). The stream temperature model is used by the Decision Support System (DSS) to help determine how much stored water to release to try to meet stream temperature targets downstream. In this application, the stream temperature prediction works very well because of its speed in the operations DSS and the ability to easily quantify and use the uncertainty in the decisions making algorithm.

Summary

We presented a regression model to predict daily maximum stream temperatures. A stepwise procedure was used to select a parsimonious set of predictors that capture as much variance of the stream temperature as possible. The results of this study show that Truckee River stream temperatures at Reno can be predicted using a simple linear regression relationship based on flow and air temperature. A nonlinear relationship is also explored but does not improve the prediction significantly. Linear regression theory is used to quantify the prediction uncertainty. Using the uncertainty, a method is developed to determine the additional flow required to meet a target temperature with a desired level of confidence. This is useful because not only the prediction but the confidence level can be used in the decision-making procedure.

Acknowledgments

The writers would like to thank Merlynn Bender, Tom Scott, Gregg Reynolds, Jeff Boyer, and Jim Brock for help and advice. This work was funded in part by the U.S. Bureau of Reclamation and was conducted at the Center for Advanced Decision Support for Water and Environmental Systems (CADSWES), at the Univ. of Colorado, Boulder. In addition, we would like to thank the reviewers whose comments have helped to improve the manuscript.

Notation

The following symbols are used in this paper:

AIC = Akaike's information criteria;
a = coefficient;
adjusted R^2 = coefficient of determination adjusted for degrees of freedom;
Cp = Mallow's Cp statistic;
n = number of observations;
PCD = prediction confidence distance;
p = number of explanatory variables;
Q = stream flow;
Q_{Required} = flow necessary to have desired probability of stream temperature exceedance;
R^2 = coefficient of determination;
SS_y = total sum of squares;
s_p^2 = mean square error of p coefficient model;
$\hat{T}$ = predicted stream temperature;
T = stream temperature;
T_{Air} = air temperature;
T_{mixed} = completely mixed water temperature;
$T_{\text{Necessary}}$ = stream temperature required to have specified probability of exceedance;
T_{Target} = desired stream temperature;
$\mathbf{X}$ = matrix of a column of ones and each new observation;
x = independent predictor variable;
$\mathbf{x}_0$ = $\{1, x_1, x_2, \ldots, x_p\}$;
$\hat{y}$ = predicted response variable;
α = confidence level;
σ = standard deviation of residuals; and
$\hat{\sigma}^2$ = best estimate of true error.

References

Ang, A. H-S., and Tang, W. H. (1975). *Probability concepts in engineering planning and design*, Wiley, New York.

Beck, M. B. (1987). "Water quality modeling: a review of the analysis of uncertainty." *Water Resour. Res.*, (23)8, 1393–1442.

Bonneville Power Administration (BPA), Bureau of Reclamation and U.S. Army Corps of Engineers. (1995). "Columbia River System Operation Review: Final environmental impact statement." *Rep. No. DOE/EIS-0170*, Dept. of Energy, Portland, Ore.

Brock, J. T., and Caupp, C. L. (1996). "Application of DSSAMt water quality model—Truckee River, Nevada for Truckee River Operating Agreement (TROA) DEIS/DEIR: simulated river temperatures for TROA." *Technical Rep. No. RCR96-7.0*, Rapid Creek Research Inc., Boise, Id.

Carron, J. C., and Rajaram, H. (2001). "Impact of variable reservoir releases on management of downstream temperatures." *Water Resour. Res.*, 37(6), 1733–1743.

Gu, R., McCutcheon, R., and Chen, C. J. (1999). "Development of weather-dependent flow requirements for river temperature control." *Environ. Manage. (N.Y.)*, 24(4), 529–540.

Helsel, D. R., and Hirsch, R. M. (1992). *Statistical methods in water resources*, Elsevier Science B.V., Amsterdam, The Netherlands.

Hockey, J. B., Owens, I. F., and Tapper, N. J. (1982). "Empirical and theoretical models to isolate the effect of discharge on summer water temperature in the Hurunui River." *J. Hydrol. (New Zealand)*, 21(1), 1–12.

Loader, C. (1999). *Statistics and computing: local regression and likelihood*, Springer, New York.

Mohseni, O., Erickson, T. R., and Stefan, H. G. (2002). "Upper bounds for stream temperatures in the contiguous United States." *J. Environ. Eng.*, 128(1), 4–11.

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

Neumann, D.W. (2001) "An operations model for temperature management of the Truckee River above Reno, Nevada." Wasters thesis, Univ. of Colorado, Boulder, Colo.

Reckhow, K. H. (1994). "Water quality simulation modeling and uncertainty analysis for risk assessment and decision making." *Ecol. Modell.*, 72, 1–20.

S-Plus 5 for UNIX guide to statistics. (1998). MathSoft, Inc., Seattle.

Taylor, R. L. (1998). "Simulation of hourly stream temperature and daily dissolved solids for the Truckee River, California and Nevada." *Water-Resources Investigations Rep. No. WRI 98-4064*, U.S. Geologic Survey, Reston, Va.

U.S. Bureau of Reclamation. (1995). "Operation of Glen Canyon Dam." *Final EIS*, U.S. Department of the Interior, Washington, D.C.

U.S. Bureau of Reclamation. (2000). "Rio Grande and low flow conveyance channel modifications." *DEIS*, U.S. Department of the Interior, Washington, D.C.

Varis, O., Kuikka, S., and Taskinen, A. (1994). "Modeling for water quality decisions: uncertainty and subjectivity in information, in objectives, and in model structure." *Ecol. Modell.*, 74, 91–101.