UNIVERSITY OF MINNESOTA
ST. ANTHONY FALLS LABORATORY
Engineering, Environmental and Geophysical Fluid Dynamics

Project Report No. 419

The Effect of Record Length on a Nonlinear Regression Model for Weekly Stream Temperatures

by

Troy R. Erickson, Omid Mohseni and Heinz G. Stefan

Prepared for

GRAZING LANDS RESEARCH LABORATORY
Agricultural Research Service, US Department of Agriculture
El Reno, Oklahoma

In cooperation with

MID-CONTINENT ECOLOGY DIVISION
US Environmental Protection Agency
Duluth, Minnesota

July 1998
Minneapolis, Minnesota

UNIVERSITY OF MINNESOTA
ST. ANTHONY FALLS LABORATORY
Engineering, Environmental and Geophysical Fluid Dynamics

Project Report No. 419

The Effect of Record Length on a Nonlinear Regression Model for Weekly Stream Temperatures

by

Troy R. Erickson, Omid Mohseni and Heinz G. Stefan

Prepared for

GRAZING LANDS RESEARCH LABORATORY
Agricultural Research Service, US Department of Agriculture
El Reno, Oklahoma

In cooperation with

MID-CONTINENT ECOLOGY DIVISION
US Environmental Protection Agency
Duluth, Minnesota

July 1998
Minneapolis, Minnesota

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, religion, color, sex, national origin, handicap, age or veteran status.

Prepared for: Metropolitan Council
Last Revised: 4/22/98
Disk Locators: PR419cov.doc; REPT419r.doc;
(Zip Disk #9/Stefan Reports

ABSTRACT

A four parameter, logistic stream temperature model using weekly air temperature as the predictor of weekly stream temperature was fitted by least squares regression to records varying in length from 12 to 32 years. The records were from four streams in Minnesota and three streams in Oklahoma. The purpose of the study was to test if stream temperature models formulated from 3-year samples were representative of stream temperature models developed from the seven, full-length records. This test was done because the model had previously been applied to 3-year records from 585 streams and associated weather stations in the US (Mohseni et al., 1997). Each full-length record was divided into 3-year samples containing up to 156 weekly air temperature and stream temperature data. The logistic stream temperature model was then fitted to the 3-year samples, as well as the full-length records. The models formulated from the full-length records were assumed to represent the "true" weekly air temperature/stream temperature relationships or "population" relationships. F-tests were used to determine whether statistical similarity between the 3-year sample and the full-length models existed. The results showed that approximately 33% of the 3-year sample relationships were not statistically similar to their respective population models. Further analysis of the 3-year sample and population regression parameters revealed notable discrepancies, especially for the parameter representing upper bound stream temperature. Twenty-six of thirty-one 3-year samples produced estimates of this parameter less than their respective population model. In addition, the 3-year sample estimates of upper bound stream temperature demonstrated a large variance. Nonlinear, least squares parameter estimates were found to be inherently biased. The bias of nonlinear regression parameters is reduced with increasing sample length. Three-year weekly air temperature and stream temperature records can not exhibit the natural variance found in longer records. Records of more than 3-year duration are therefore necessary for the consistent representation of long-term weekly air temperature/stream temperature relationships.

ACKNOWLEDGMENT

The work reported herein was supported by the Agricultural Research Service, USDA, Grazing Lands Research Laboratory, El Reno, Oklahoma in cooperation with the Mid-Continent Ecology Division, US Environmental Protection Agency, Duluth, Minnesota. Dr. Robert Williams and Dr. Virginia Snarski were the project officers. The stream temperature data were provided by Robert Scheller. Dr. Sanford Weisberg provided valuable suggestions for the statistical analysis presented herein.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Empirical stream temperature models are proposed to study the effects of projected global climate warming on freshwater ecosystems. Eaton and Scheller (1996) utilized a linear model relating air temperature to stream temperature to determine the effect of climate warming on the thermal habitat of several fish species. However, at the highest and lowest stream temperatures there can be a significant deviation from the linear relationship. Therefore, Mohseni et al. (1997) developed a nonlinear regression model between weekly air temperatures and weekly stream temperatures. An accurate representation of the weekly air temperature/stream temperature relationship is essential because of the assumption that the relationship remains unchanged under 2xCO_2 climate conditions (doubling of atmospheric CO_2) despite marked changes in climate parameters (air temperature, humidity, wind speed, solar radiation, cloud cover and precipitation). It is therefore of interest to investigate whether empirical stream temperature models typically formulated from a few years of data are representative of true weekly air temperature/stream temperature relationships.

Empirical air temperature/stream temperature relationships, linear or nonlinear, are generally formulated using a few years of data because extended stream temperature records are rare. Webb (1987) developed a linear air temperature/stream temperature relationship for the River Clyst in England using 6 years of daily temperature data. Similarly, Stefan and Preud'homme (1993) developed linear air temperature/stream temperature relationships for 11 midwestern streams of the US using 1 to 8 years of daily and weekly data exclusively from the open water season. Crisp and Howson (1982) developed linear relationships for 3 streams using 1 to 5 years of weekly data. Mohseni et al. (1997) fitted a nonlinear function to 3-year, weekly air temperature and stream temperature records from 585 streams located throughout the US. The regression models fitted in each of the previous studies produced good results (standard errors of estimate less than 2.0 °C). However, it is not known whether these stream temperature models and corresponding data are representative of the true, long-term weekly air temperature/stream temperature relationship of each stream.

Because least squares parameter estimates of linear stream temperature models are inherently unbiased, the number of data required for representative estimates of the true regression parameters is not restrictive, even though the quality of the estimates from sample to sample is directly affected by the number of data. The variance (spread) of these sample parameter distributions tends to be large when the number of data from which the parameters are determined is small. A sample parameter distribution with large variance implies that there is a greater probability of under- or overestimating the true regression parameter. In general, linear stream temperature models require few data. Pilgrim et al. (1996) found that linear, least squares regression parameters, as well as the quality of correlation in terms of the coefficient of determination, varied by an insignificant amount for weekly records equal to or larger than two years length.

The data requirements for representative estimates of true nonlinear regression parameters fitted by least squares regression are, in general, unknown. Least squares parameter estimates of nonlinear models are inherently biased (Ratkowsky, 1983). The degree of bias associated with these parameter estimates is related to the number of data used to formulate a specific model, as well as the character of the data in relation to the model being fitted. As the number of data is increased, the bias of the model's parameter(s) tends to decrease. Various nonlinear models, e.g. the Gompertz, Logistic, Richards, Morgan-Mercer-Flodin and Weibull-type, require differing, yet unspecified numbers of data for unbiased parameter estimates (Ratkowsky, 1983).

The logistic stream temperature model developed by Mohseni et al. (1997) is given by the equation

$$T_s = \mu + \frac{\alpha - \mu}{1 + \exp\left[\gamma\left(\beta - T_a\right)\right]} \tag{1}$$

where weekly air and weekly stream temperature are represented by T_a and T_s, respectively. The parameters α and μ represent upper bound and lower bound stream temperatures, respectively. β and γ define the geometry of the S-shaped logistic function (Figure 1). Each record used in Mohseni et al.'s (1997) study contained, at most, 156 data for a three year period (water years from October 1977 to September 1980). The model was successfully applied to 98% of 585 records. Because of its ability to predict stream temperatures year-round and its wide applicability, the logistic stream temperature model (Equation 1) will be used as the basis for a statistical comparison of weekly air temperature and stream temperature relationships derived from 3-year and 12-year or longer records.

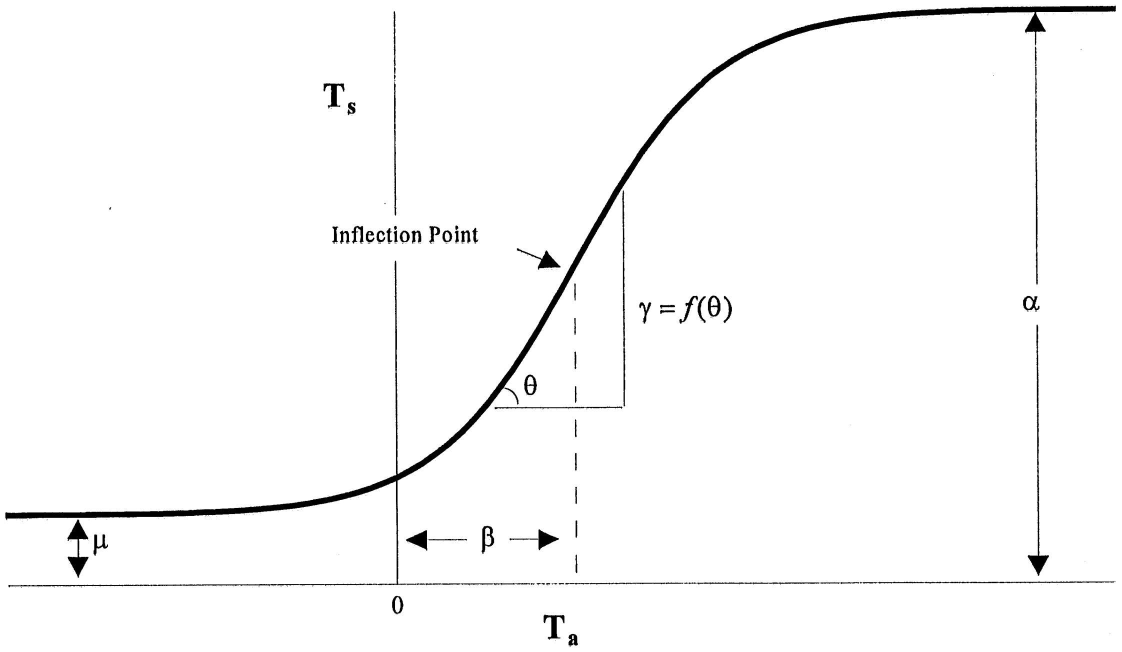

Figure 1.1 Schematic representation of the logistic stream temperature model and its parameters.

2. METHOD OF ANALYSIS

The method by which the logistic stream temperature model is fitted to 3-year samples and full-length records, and the statistical comparison of the results is presented in this section. Seven daily stream temperature records ranging from 12 to 32 years in length were analyzed. The daily stream temperatures were first paired with air temperatures from nearby first order weather monitoring stations and then averaged over weekly time periods. Each full-length, weekly air temperature and stream temperature record was then divided into 3-year samples. Using nonlinear, least squares regression, the logistic stream temperature model (Equation 1) was fitted to each 3-year sample and full-length record. A statistical comparison of the 3-year sample models and their respective full-length models was made by an F-test. Depending upon the magnitude of the resulting F-test statistics (observed significance levels) in relation to a prescribed significance level, the statistical similarity of the 3-year sample models was determined. A presentation of the preparation of 3-year sample data sets, determination of sample and population model parameters, formulation of the F-test statistics and interpretation of the F-test statistics follows.

2.1 Preparation of Data

The data used in previous studies of streams in Minnesota and Oklahoma (Pilgrim et al., 1995; Erickson and Stefan, 1996) were scanned to find stream temperature records of at least 12 years length and without large segments of missing data. Both studies used maximum, minimum or mean daily stream temperature data from the US Geological Survey (USGS) stream database. Hourly air temperature data were obtained from the Midwest Climate Center of the Illinois State Water Survey in Champaign, Illinois, for the Minnesota study. A CD-ROM created by the National Climatic Data Center in Asheville, North Carolina, was used to obtain hourly air temperature data for the Oklahoma study. Mean daily stream temperatures were either extracted directly from the stream temperature record or calculated by taking the average of the recorded maximum and minimum daily values. Mean daily air temperatures were calculated by determining the average of 24 hourly measurements. Daily air temperatures and stream temperatures were paired. Mean weekly air temperatures and stream temperatures were then calculated by averaging the mean daily values. Data from four streams in Minnesota and three streams in Oklahoma were retained for analysis. Table 1 provides each stream's name, location and length of recorded stream temperatures. Although air temperature data were never a limiting factor, the weather station names and locations are also provided in Table 1. The seven stream temperature records described in Table 1 were chosen because of their length, which ranged from 12 to 32 years, and their relative completeness. Other details concerning the first order weather monitoring stations and water quality monitoring stations used for this study were previously given by Pilgrim et al. (1995) and Erickson and Stefan (1996).

To facilitate the comparison, the seven full-length records were divided into nonoverlapping, 3-year samples. If a full-length air temperature and stream temperature record was complete, consecutive 3-year samples were created. If the total record length was not a multiple of 3 years the data found at the end of the full-length record composing the remainder were excluded. If data gaps existed in the long-term records, samples were formed that consisted of the most complete 3-year periods. Although the samples remained in sequence, the aggregate of the samples was not continuous. With exception of the Mississippi River at St.Paul, MN (USGS water quality monitoring station 05331000), all of the full-length stream temperature records had segments of missing data. The total number of sample data sets for each stream varied from three to ten. The total number of data within each long-term record, as well as the number of sample data sets is also provided in Table 1.

2.2 Determination of Sample and Population Regression Parameters

The logistic stream temperature model was fitted by nonlinear, least squares regression to the seven full-length records listed in Table 1, as well as the corresponding 3-year sample records. The fitted lines (without the data) are shown in Figures 2.1 through 2.7. Regression parameters obtained from the full-length records were assumed to provide the population parameters of Equation 1. Technically, the term "population" refers to an infinitely large collection of data. Although the seven full-length records are a finite collection of weekly air temperatures and stream temperatures, it will be assumed that they represent the true air temperature/stream temperature relationship. The regression parameters obtained from the 3-year samples were sample estimates of the population parameters. Population parameters, as well as sample parameters of the nonlinear regression analyses can be found in Table 2. In addition, Table 2 provides the number of data used, the goodness of fit or Nash-Sutcliffe coefficient (NSC) (Nash and Sutcliffe, 1970) obtained and the root mean squared error (RMSE) calculated for each nonlinear regression analysis (population or sample). The computational definitions of the Nash-Sutcliffe coefficient and root mean squared error are as follows:

Table 1. USGS water quality and first order weather monitoring station information.

USGS Station ID	Stream Name	Latitude	Longitude	Weather Station Location	Latitude	Longitude	Years of Record	Samples	Total Data
05330000	Minnesota River near Jordan, MN	44° 42’	93° 39’	Jordan, MN	44° 39’	93° 37’	14 (1973-1986)	3	504
05331000	Mississippi River at St.Paul, MN	44° 57’	93° 05’	St.Paul, MN	44° 58’	93° 05’	32 (1956-1987)	10	1522
05344980	Mississippi River near Red Wing, MN	44° 37’	92° 37’	St.Paul, MN	44° 58’	93° 05’	14 (1978-1982)	3	322
05345000	Vermillion River near Empire, MN	44° 40’	93° 03’	Farmington, MN	44° 40’	93° 11’	15 (1974-1988)	4	602
07150500	Salt Fork of the Arkansas River near Jet, OK	36° 45’	98° 07’	Wichita, KS	37° 39’	97° 25’	21 (1969-1989)	5	726
07152500	Arkansas River at Ralston, OK	36° 30’	96° 43’	Tulsa, OK	36° 12’	95° 54’	12 (1969-1980)	3	487
07161000	Cimarron River at Perkins, OK	35° 57’	97° 01’	Oklahoma City, OK	35° 24’	97° 36’	12 (1969-1980)	3	464

$$\mathrm{NSC} = 1 - \frac{\sum_{i=1}^{n}\left(T_{\mathrm{sim}_i} - T_{\mathrm{obs}_i}\right)^2}{\sum_{i=1}^{n}\left(\overline{T}_{\mathrm{obs}} - T_{\mathrm{obs}_i}\right)^2} \tag{2}$$

and

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(T_{\mathrm{sim}_i} - T_{\mathrm{obs}_i}\right)^2}{n-4}} \tag{3}$$

T_{sim} and T_{obs} represent simulated and observed stream temperatures, respectively; $\overline{T}_{\mathrm{obs}}$ is the mean of observed stream temperatures. Mean annual air temperatures and stream temperatures have also been included in Table 2 to provide an indication of the variability in the 3-year data in relation to the long-term means.

Stream temperature data that have seasonal hysteresis (Mohseni et al. 1997) were fitted with two functions: one function representing the rising limb and the second function representing the falling limb of the weekly air temperature/stream temperature relationship. Interestingly, streams whose population data did not have hysteresis often showed hysteresis in the corresponding 3-year samples. To maintain the similarity of models (3-year sample and population) and thus the integrity of the statistical comparison, hysteresis was neglected.

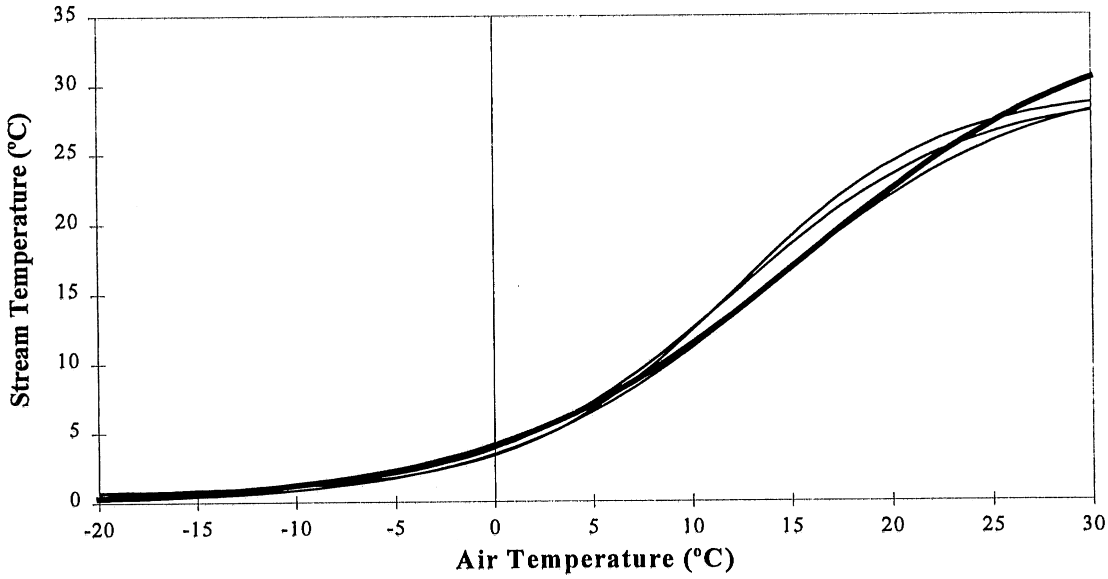

Figure 2.1 3-year sample (thin lines) and population (bold line) air and stream temperature relationships for USGS station 05330000, Minnesota River near Jordan, MN.

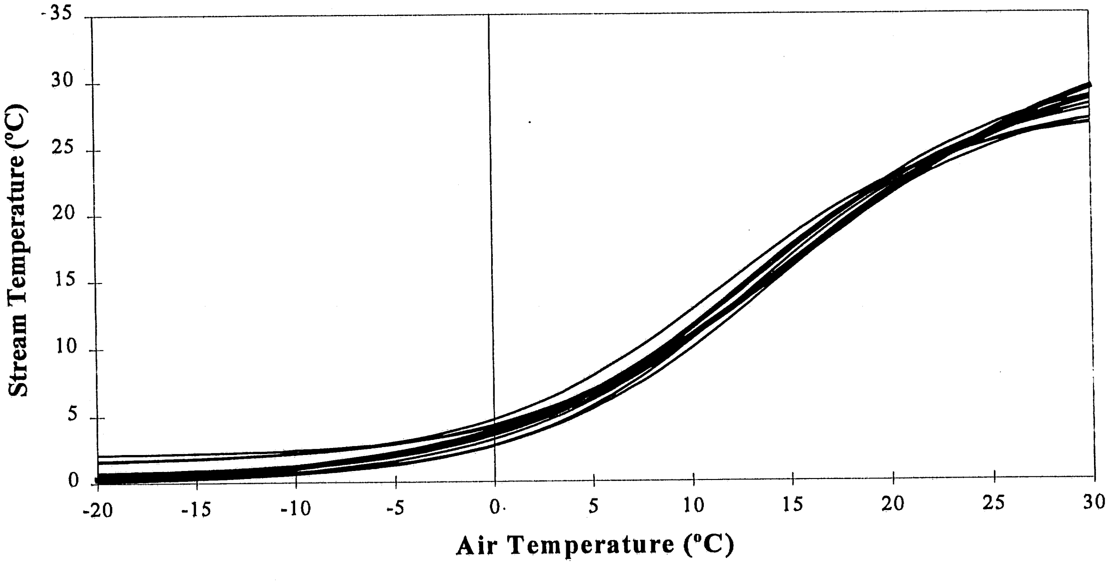

Figure 2.2 3-year sample (thin lines) and population (bold line) air and stream temperature relationships for USGS station 05331000, Mississippi River at St.Paul, MN.

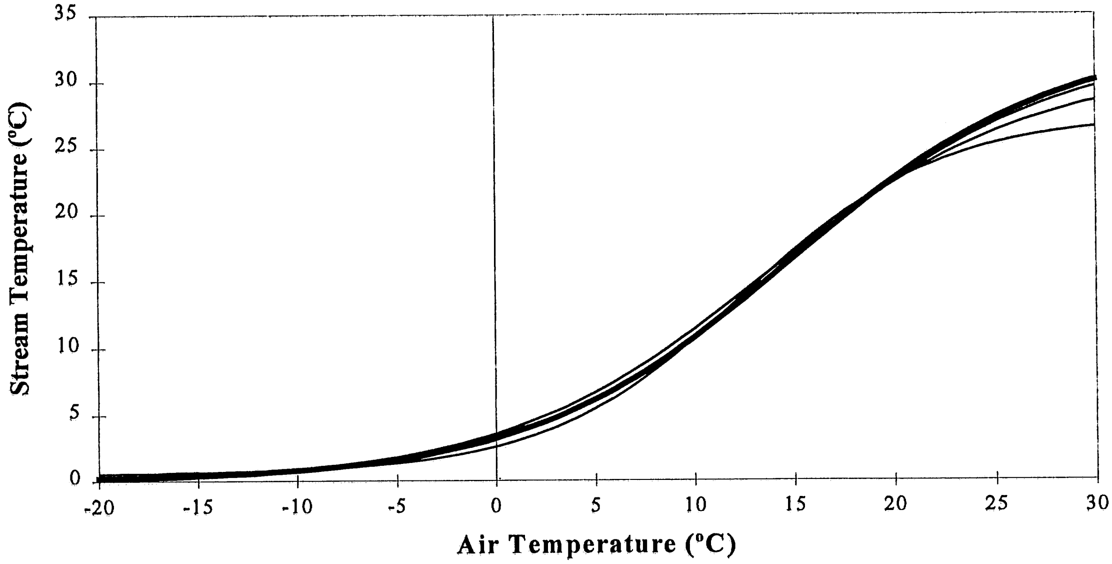

Figure 2.3 3-year sample (thin lines) and population (bold line) air and stream temperature relationships for USGS station 05344980, Mississippi River at Lock and Dam #3 near Red Wing, MN.

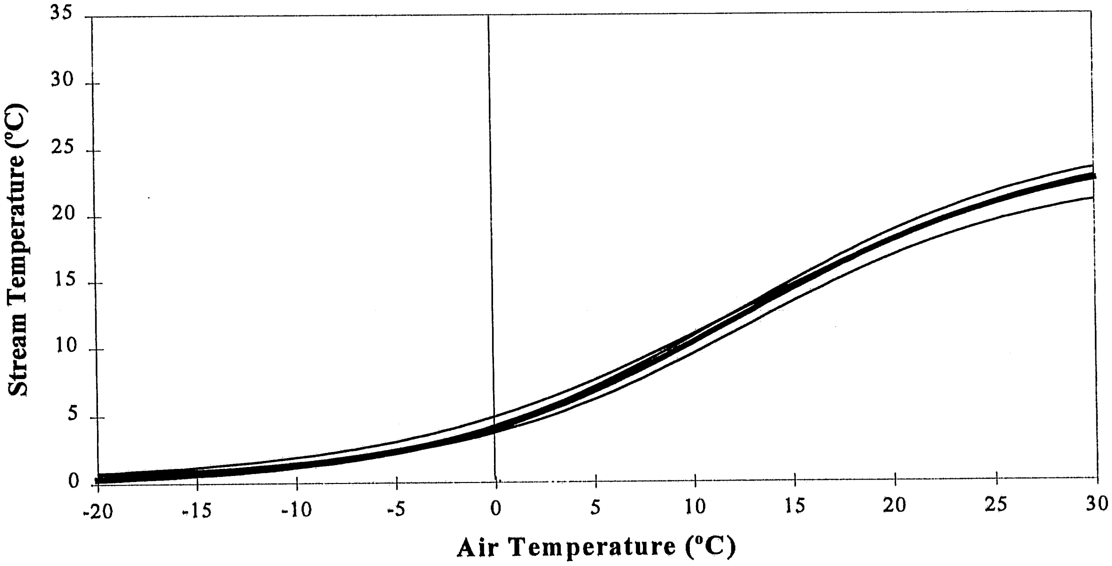

Figure 2.4 3-year sample (thin lines) and population (bold line) air and stream temperature relationships for USGS station 05345000, Vermillion River near Empire, MN.

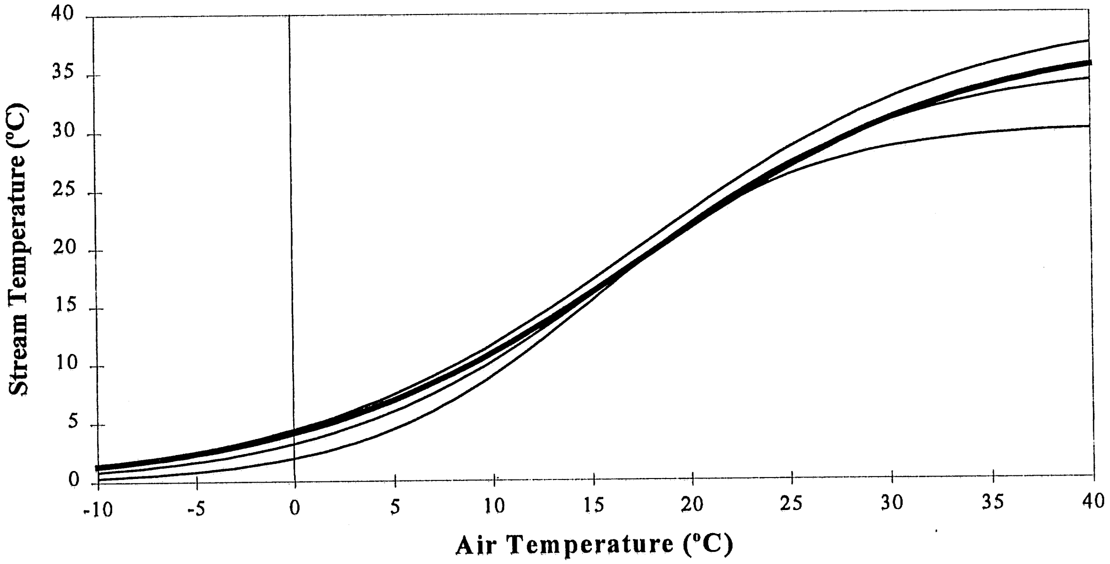

Figure 2.7 3-year sample (thin lines) and population (bold line) air and stream temperature relationships for USGS station 07161000, Cimarron River at Perkins, OK.

Table 2. Population and 3-year sample nonlinear regression results.

USGS ID. 05330000 - Minnesota River near Jordan, MN

Sample	*Mean Air Temp.*	*Mean Stream Temp.*	*NSC*	*RMSE*	*n*	α	β	γ	μ
Population	6.67	11.65	0.95	2.09	504	35.20	15.74	0.13	0.00
1	5.60	11.42	0.96	1.92	132	29.54	12.09	0.19	0.69
2	6.31	11.62	0.95	1.98	148	29.29	12.05	0.17	0.51
3	7.49	11.57	0.96	1.79	136	30.55	13.78	0.15	0.00

USGS ID. 05331000 - Mississippi River at St.Paul, MN

Sample	*Mean Air Temp.*	*Mean Stream Temp.*	*NSC*	*RMSE*	*n*	α	β	γ	μ
Population	7.67	11.75	0.93	2.49	1522	34.12	15.93	0.13	0.00
1	8.13	12.01	0.93	2.52	150	30.10	13.44	0.16	0.40
2	7.49	12.44	0.92	2.56	155	27.96	11.87	0.17	1.55
3	7.59	11.99	0.93	2.58	151	29.69	13.17	0.16	0.34
4	6.95	11.36	0.91	2.61	153	31.40	15.09	0.15	0.43
5	6.29	11.56	0.92	2.58	148	27.63	12.80	0.19	2.06
6	8.02	11.39	0.92	2.54	156	29.05	13.76	0.16	0.64
7	7.74	12.46	0.94	2.43	155	30.91	13.76	0.16	0.50
8	7.78	11.29	0.96	1.89	156	31.52	14.58	0.15	0.00
9	8.04	11.45	0.97	1.75	155	28.37	13.12	0.18	0.32
10	9.06	11.87	0.96	1.96	139	30.59	14.64	0.16	0.00

USGS ID. 05344980 - Mississippi River at Lock and Dam #3 near Red Wing, MN

Sample	*Mean Air Temp.*	*Mean Stream Temp.*	*NSC*	*RMSE*	*n*	α	β	γ	μ
Population	7.81	11.82	0.96	1.85	322	33.19	14.95	0.15	0.00
1	6.55	11.17	0.96	2.01	80	30.98	13.67	0.15	0.00
2	7.95	11.62	0.97	1.69	77	32.59	14.75	0.15	0.00
3	8.91	12.04	0.97	1.58	75	27.31	12.50	0.20	0.52

Table 2. (Continued)

USGS ID. 05345000 - Vermillion River near Empire, MN

Sample	*Mean Air Temp.*	*Mean Stream Temp.*	*NSC*	*RMSE*	*n*	α	β	γ	μ
Population	7.56	10.19	0.97	1.27	602	25.00	12.70	0.13	0.00
1	8.97	11.27	0.97	1.26	108	25.89	12.60	0.13	0.00
2	6.90	10.20	0.97	1.11	155	25.34	12.54	0.12	0.30
3	7.44	9.37	0.96	1.26	119	22.92	12.92	0.14	0.52
4	7.61	10.07	0.97	1.12	136	25.41	13.29	0.13	0.33

USGS ID. 07150500 - Salt Fork of the Arkansas River near Jet, OK

Sample	*Mean Air Temp.*	*Mean Stream Temp.*	*NSC*	*RMSE*	*n*	α	β	γ	μ
Population	12.84	15.70	0.96	1.79	726	36.11	15.95	0.12	0.00
1	11.81	14.92	0.96	1.70	107	33.15	13.68	0.13	0.00
2	12.57	15.81	0.96	1.81	137	37.37	15.83	0.12	0.00
3	12.64	15.03	0.96	1.86	146	33.67	15.16	0.13	0.00
4	13.49	16.31	0.97	1.39	141	29.74	12.23	0.14	0.77
5	13.26	15.98	0.97	1.49	156	34.80	15.00	0.12	0.00

USGS ID. 07152500 - Arkansas River at Ralston, OK

Sample	*Mean Air Temp.*	*Mean Stream Temp.*	*NSC*	*RMSE*	*n*	α	β	γ	μ
Population	15.98	16.40	0.96	1.74	487	33.99	16.81	0.13	0.00
1	16.12	16.71	0.96	1.60	134	32.31	15.79	0.14	0.00
2	15.79	15.53	0.97	1.34	147	29.84	15.21	0.16	0.00
3	15.55	16.24	0.97	1.60	141	31.70	15.38	0.15	0.00

USGS ID. 07161000 - Cimarron River at Perkins, OK

Sample	*Mean Air Temp.*	*Mean Stream Temp.*	*NSC*	*RMSE*	*n*	α	β	γ	μ
Population	15.42	17.18	0.95	2.06	464	37.83	17.54	0.12	0.00
1	14.85	17.59	0.97	1.71	140	39.79	17.36	0.12	0.00
2	13.75	15.33	0.96	1.79	138	35.41	16.49	0.14	0.00
3	15.91	16.79	0.96	1.96	101	30.36	14.87	0.18	0.00

2.3 Formulation of Test Statistics

The discrepancies that exist between 3-year sample models and population models, as shown in Figures 2.1 through 2.7 (Table 2 gives the fitted regression parameters), can be quantified in terms of an F-test statistic. In the F-test, the residual sum of squared errors of both sample and population models are compared instead of the individual model parameters α, β, γ and μ. While the resulting analysis does not provide information on any specific discrepancy, it does provide an effective, overall means of comparison.

F-test statistics were calculated for two types of comparisons. The first type was the grouped comparison. F-test statistics for this comparison were formulated with the residual sum of squared errors of a population model and the summation of the residual sum of squared errors of the corresponding 3-year sample models. The grouped comparison provided an overall quantification of the similarity between 3-year sample models and the corresponding population model. In addition, it showed whether all 3-year sample models were statistically similar or if at least one 3-year sample model was not statistically similar. The grouped comparison could not, however, distinguish the specific 3-year sample model(s) not statistically similar to its/their respective population model. Therefore, a second, individual comparison was made. F-test statistics for this comparison were formulated with the residual sum of squared errors of a population model and the residual sum of squared errors of a 3-year sample model. With this comparison, those sample models not statistically similar to their corresponding population model could be identified.

The F-test statistic for grouped or individual 3-year data comparisons is computed as follows:

$$F = \frac{\mathrm{RSS_p} - \sum_{j=1}^{m} RSS_j}{\sum_{j=1}^{m} RSS_j} \cdot \frac{df_2}{df_1} \tag{4}$$

where

$$\mathrm{RSS} = \sum_{\mathrm{i=1}}^{\mathrm{n}} (T_{\mathrm{sim}_i} - T_{\mathrm{obs}_i})^2 \tag{5}$$

and

$$\begin{aligned} df_1 &= \left(n_p - 4\right) - \sum_{i=1}^{m}\left(n_i - 4\right) \\ df_2 &= \sum_{i=1}^{m}\left(n_i - 4\right) \end{aligned} \tag{6}$$

The subscripts p and j identify parameters associated with a population relationship or a 3-year sample relationship, respectively. RSS is the residual sum of squared errors. Numerator degrees of freedom and denominator degrees of freedom are denoted by df1 and df2, respectively. The total number of data and 3-year samples is denoted by n and m, respectively. Because the output of Mohseni et al.'s (1997) logistic stream temperature model fitting algorithm provided only the root mean squared error (RMSE), RSS was determined by squaring RMSE and then multiplying this value by the total number of degrees of freedom, i.e. n-4 (Equation 3).

While the F-test statistic conveys some information, a more precise interpretation follows from the determination of the corresponding p-value. Generally, the p-value corresponding to a specific F-test statistic can be determined from tabulated results or computationally. In either case, three input parameters are required: the F-test statistic, the numerator degrees of freedom, df1, and the denominator degrees of freedom, df2. For this study, p-values were determined using an inverse F-distribution function provided on Microsoft Excel. Table 3 contains all information relevant to the determination of the F-test statistics and p-values.

Table 3. Comparative statistical analysis results.

USGS ID. 05330000 - Minnesota River near Jordan, MN

Comparison	RSS_p	*RSS*	n	df_1	df_2	*F-Test Statistic*	*P-Value*
Grouped	2184.05	1452.93	504	96	404	2.1176	0.0000
1	2184.05	471.86	132	372	128	1.2486	0.0697
2	2184.05	564.54	148	356	144	1.1604	0.1509
3	2184.05	416.53	136	370	130	1.4909	0.0040

USGS ID. 05331000 - Mississippi River at St.Paul, MN

Comparison	RSS_p	*RSS*	n	df_1	df_2	*F-Test Statistic*	*P-Value*
Grouped	9411.75	8265.06	1522	40	1478	5.1264	0.0000
1	9411.75	927.16	150	1372	146	0.9738	0.5985
2	9411.75	989.59	155	1367	151	0.9401	0.7073
3	9411.75	978.49	151	1371	147	0.9241	0.7526
4	9411.75	1015.00	153	1369	149	0.9004	0.8171
5	9411.75	958.52	148	1374	144	0.9243	0.7504
6	9411.75	980.64	156	1366	152	0.9567	0.6556
7	9411.75	891.64	155	1367	151	1.0556	0.3410
8	9411.75	542.96	156	1366	152	1.8176	0.0000
9	9411.75	462.44	155	1367	151	2.1377	0.0000
10	9411.75	518.62	139	1383	135	1.6739	0.0001

USGS ID. 05344980 -Mississippi River at Lock and Dam #3 near Red Wing, MN

Comparison	RSS_p	*RSS*	n	df_1	df_2	*F-Test Statistic*	*P-Value*
Grouped	1088.36	692.79	322	98	220	1.2181	0.0686
1	1088.36	307.05	80	252	76	0.7991	0.8960
2	1088.36	208.50	77	245	73	1.2574	0.1245
3	1088.36	177.24	75	247	71	1.4776	0.0264

Table 3. (Continued)

USGS ID. 05345000 - Vermillion River near Empire, MN

Comparison	*RSS_p*	*RSS*	*n*	*df_1*	*df_2*	*F-Test Statistic*	*P-Value*
Grouped	964.51	699.31	602	96	502	1.9831	0.0000
1	964.51	165.11	108	494	104	1.0193	0.4635
2	964.51	186.05	155	447	151	1.4135	0.0062
3	964.51	182.57	119	483	115	1.0197	0.4591
4	964.51	165.58	136	466	132	1.3667	0.0159

USGS ID. 07150500 - Salt Fork of the Arkansas River near Jet, OK

Comparison	*RSS_p*	*RSS*	*n*	*df_1*	*df_2*	*F-Test Statistic*	*P-Value*
Grouped	2313.36	1826.81	726	55	667	3.2300	0.0000
1	2313.36	297.67	107	619	103	1.1268	0.2281
2	2313.36	435.72	137	589	133	0.9731	0.5910
3	2313.36	491.26	146	580	142	0.9081	0.7768
4	2313.36	264.70	141	585	137	1.8125	0.0000
5	2313.36	337.46	156	570	152	1.5614	0.0005

USGS ID. 07152500 - Arkansas River at Ralston, OK

Comparison	*RSS_p*	*RSS*	*n*	*df_1*	*df_2*	*F-Test Statistic*	*P-Value*
Grouped	1462.33	940.29	487	73	410	3.1182	0.0000
1	1462.33	332.80	134	353	130	1.2499	0.0686
2	1462.33	256.77	147	340	143	1.9747	0.0000
3	1462.33	350.72	141	346	137	1.2550	0.0616

USGS ID. 07161000 - Cimarron River at Perkins, OK

Comparison	*RSS_p*	*RSS*	*n*	*df_1*	*df_2*	*F-Test Statistic*	*P-Value*
Grouped	1952.06	1239.54	464	93	367	2.2684	0.0000
1	1952.06	409.37	140	324	136	1.6407	0.0005
2	1952.06	442.17	138	326	134	1.4578	0.0062
3	1952.06	388.00	101	370	90	1.1301	0.2445

2.4 Interpretation of Statistical Parameters

The comparison of 3-year sample and population regression models has been reduced to the interpretation of an F-test statistic and, more importantly, its corresponding p-value. According to Devore and Peck (1997), "The p-value (also sometimes called the observed significance level) is a measure of inconsistency between the hypothesized value for a population characteristic and the observed sample. It is the probability, assuming that the null hypothesis is true, of obtaining a test statistic value at least as contradictory to the null hypothesis as what actually resulted." For this analysis, the null hypothesis, H_o, assumes that all relationships, population and 3-year sample, are coincident. The alternative hypothesis, H_a, assumes that all relationships formulated from a common data set are nonequivalent. In terms of variables, the null hypothesis and alternative hypothesis can be represented as follows: for H_0: $M_{population} = M_i$, and for H_a: $M_{population} \neq M_i$, where $M_{population}$ and M_i designate a specific population model and its corresponding ith 3-year sample model, respectively. The p-value provides a measure of the inconsistency between the residual sum of squared errors of the population model and the grouped or individual 3-year sample models. It is evident from the previous definition that larger p-values correspond to a stronger agreement between the 3-year sample and population regression models. Alternatively, an F-test statistic close to unity or below represents good agreement between the population and 3-year sample models.

While large p-values indicate that there is strong agreement between 3-year sample and population models, a decision must be made regarding the minimum p-value that will constitute sufficient agreement between the two model types. The selection of an arbitrarily large significance level such as 0.10 would ensure statistical similarity between those 3-year sample models whose characteristics duplicate those of the population model. However, many 3-year sample models would not attain this value because the stream temperature model only accounts for air temperature as a predictor of stream temperature. Other influencing factors such as solar radiation, relative humidity, wind speed, water depth, groundwater inflow, artificial heat inputs and thermal conductivity of the sediments were neglected as a means of simplification. Because of the potential for large variations within each sample used for comparison, a smaller significance level is more appropriate for this study. The significance level chosen for this analysis was 0.01. This significance level preserves the integrity of the analysis while allowing for some leniency to account for the numerous extraneous factors of influence upon the weekly air temperature/stream temperature relationship.

Statistical similarity is determined by a simple comparison of the observed p-value and the significance level of 0.01. If the observed p-value for any statistical comparison is less than 0.01, then that particular 3-year sample or grouped 3-year sample relationship is considered not statistically similar to the population relationship. This also implies that the null hypothesis is rejected in favor of the alternative hypothesis. On the other hand, any grouped 3-year sample or individual 3-year sample relationship with a p-value greater than the significance level of 0.01 is classified as being statistically similar. In this instance, there is sufficient statistical evidence for the null hypothesis.

3. RESULTS

The results of the comparative statistical analyses will be presented first. A discussion of the grouped comparisons and individual comparisons are included in this section. The regression parameters α, β, γ and μ in relation to sample size are discussed in the following section. Special attention is paid to the variation of the parameter representing upper bound stream temperature in Equation 1.

3.1 Comparative Statistical Analyses

3.1.1 Grouped Comparisons

Most groups of 3-year sample models were found to be not statistically similar to their corresponding population models (Table 3). Only one of seven grouped comparisons had a p-value larger than 0.01, the selected limiting significance level. This indicates that at least one 3-year sample regression model within each of the six grouped data sets with p-values less than 0.01 was not statistically similar to its respective population model. p-values below 0.01 indicate that there is only a low chance of obtaining a group of sample regression relationships with an associated sum of residual squared errors equivalent to that observed in the corresponding population model. Data from the Mississippi River at Lock and Dam #3 near Red Wing, MN (USGS water quality monitoring station 05344980) were the exception. A grouped p-value of 0.069 was exhibited for this particular site indicating that there was relatively good agreement between the grouped 3-year sample and population models. Each of the individual 3-year sample models for this water quality monitoring station can be considered statistically similar to the population model.

The full-length records were not entirely represented within the 3-year samples for 6 of the 7 records studied. For example, the Vermillion River near Empire, MN (USGS water quality monitoring station 05345000) had 602 data in the population model, but only 518 data comprised the four 3-year samples (Table 2). The magnitude of the residual sum of squared errors (RSS) is dependent only upon predicted stream temperatures relative to those observed and the number of data in each sample or full-length record (Equation 5). If data are excluded, the effect upon the F-test statistic will be under- or overestimation. Under- or overestimation will result because the denominator of Equation 5 will be lower or higher depending upon the characteristics of the excluded data.

3.1.2 Individual Comparisons

The F-test statistics and p-values calculated for individual 3-year sample and population relationships are found in Table 3. Ten of the thirty-one 3-year samples (32.3%) were not statistically similar to their respective population models. Further analysis showed that 5 of 20 (25.0%) and 5 of 11 (45.5%) of the 3-year sample models from Minnesota and Oklahoma streams, respectively, were not statistically similar to their respective population models.

The results of the F-test for the comparison of individual 3-year sample models to their corresponding population models showed similar characteristics for two records in Minnesota and Oklahoma. The p-values determined for the ten 3-year samples from the Mississippi River at St. Paul, MN (USGS water quality monitoring station 05331000) ranged from 0 to 0.817. Strangely, the first seven consecutive samples had p-values ranging from 0.341 to 0.817 and the final three samples had p-values approximately equal to zero. Similarly, the five 3-year sample models from the Salt Fork of the Arkansas River near Jet, OK (USGS water quality monitoring station 07150500) had p-values ranging from zero to 0.777. The first three consecutive p-values ranged from 0.228 to 0.777 while the final two were approximately equal to zero. In both instances, statistical similarity existed between the 3-year sample models and their respective population models when the p-values were high. However, after several years, i.e. approximately 21 years for the Mississippi River record and 9 years for the Salt Fork of the Arkansas River record, statistical similarity suddenly failed to exist. The point of interest is this sudden transition which is not yet fully explained.

A possible explanation is to attribute the sudden change in statistical similarity to operational changes of hydraulic structures upstream of the water quality monitoring stations and/or the addition or removal of cooling water inputs within the duration of each full-length stream temperature record. Stefan and Chau (1977) discussed the influence of cooling water inputs to the Mississippi River at USGS water quality monitoring station 05331000. However, the reported periods of cooling water inputs did not coincide with those of the statistical analysis. The stream temperature time series for both streams did not provide any insight into the observed trends either.

3.2 Relation Between Regression Parameters and Sample Size

The 3-year sample estimates of a, b, g and m of Equation 1 were found to differ from the parameters of their respective population models. The largest discrepancies occurred between 3-year sample estimates of a and b and the corresponding population parameters. The following discussion will be limited to the differences between the population value and 3-year sample estimates of the regression parameter representing upper bound stream temperature, a.

The population values of upper bound stream temperature, a, exceeded the corresponding 3-year sample estimates of a (Table 2) for most records. An exception was present, however, for the record of the Vermillion River near Empire, MN where three of four 3-year sample estimates of a exceeded the population value. In addition, records from the Salt Fork of the Arkansas River and Cimarron River in Oklahoma each had one 3-year sample estimate of a exceed the population value. Overall, twenty-six of thirty-one 3-year sample estimates of a were lower than their respective population values.

Small discrepancies between 3-year sample estimates and population values of a are expected due to inconsistent operation of hydraulic structures on gaged streams and the addition or removal of thermal inputs within the full-length stream temperature records. The differences between 3-year sample estimates and population values of a went beyond these expectations and were as large as 7.5 °C. In general, the variance of the 3-year sample estimates of a was quite large. The largest variance among the seven groups of 3-year sample estimates of a was 9.4 °C (USGS water quality monitoring station 07161000); whereas, the smallest variance was 1.01 °C (USGS water quality monitoring station 05330000). Based upon these observations, it is conceivable that the 3-year sample estimates of a are the primary cause for nonsimilarity between the two model types and that the parameter error for b, g and m have only a secondary influence.

Weekly stream temperature models formulated from 3-year records will not consistently represent "true" air temperature/stream temperature relationships. Air temperature and stream temperature records of short duration (e.g. 3 years) do not capture the extent of the natural variation found in long-term records. For example, 3-year samples had a hysteresis even when their respective full-length models had no hysteresis. Those 3-year records do not provide an adequate representation of long-term air temperature/stream temperature relationships because they are disproportionately influenced by rare events. An increased sample size will produce sample regression models more consistent with true weekly air temperature/stream temperature relationships.

REFERENCES

Crisp, D.T. and G. Howson, "Effect of Air Temperature upon Mean Water Temperature in Streams in the North Pennines and English Lake District, *Freshwater Biology* 12:359-367.

Devore, J.L., "*Probability and Statistics for Engineering and the Sciences*", Belmont: Duxbury Press, New York, NY, 1997.

Erickson, T.R. and H.G. Stefan, "Correlation of Oklahoma Stream Temperatures with Air Temperatures", University of Minnesota, St.Anthony Falls Laboratory, *Project Report No. 398*, October, 1996.

Mohseni, O., T.R. Erickson, and H.G. Stefan, "A Non-Linear Model for Weekly Stream Temperatures at 585 Gaging Stations in the US", University of Minnesota, St.Anthony Falls Laboratory, *Project Report No. 401*, August, 1997.

Nash, J.E. and J.V. Sutcliffe, "River Flow Forecasting Through Conceptual Models", I-A, Discussion of Principles, *Journal of Hydrology*, 10, 282-290, 1970.

Pilgrim, J.M. and H.G. Stefan, "Correlations of Minnesota Stream Temperatures with Air Temperatures", University of Minnesota, St.Anthony Falls Laboratory, *Project Report No. 382*, December, 1995.

Ratkowsky, D.A., "*Nonlinear Regression Modeling: a Unified Practical Approach*", Marcel Dekker, Inc., New York, NY, 1983.

Ross, G.J.S., "*Nonlinear Estimation*", Springer Verlag New York Inc., New York, NY, 1990.

Stefan, H.G. and V.N. Chau, "River and Cooling Water Temperature Fluctuations", *Journal of the Hydraulics Division*, ASCE, Vol. 102, No. HY11, Proc. Paper 12514, November, 1976 pp. 1613-1627.

Stefan, H.G. and E.B. Preud'homme, "Stream Temperature Estimation from Air Temperature", *Water Resources Bulletin*, AWRA, Vol. 29, No. 1, February, 1993 pp. 27-45.

Webb, B.W., "The Relationship between Air and Water Temperatures for a Devon River", *Rep. Trans. Devon. Ass. Advmt Sci.*, 119:197-222.

Weisberg, S., "*Applied Linear Regression*", John Wiley & Sons, Inc., New York, NY, 1985.