

NON LINEAR PROGRAMMING OF TIME SERIES DATA TO MINIMIZE EUTROPHICATION IN TRUCKEE RIVER, NEVADA

ANPALAKI J. RAGAVAN, Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557

ABSTRACT

In Truckee River, Nevada high total phosphorus concentrations (TP) lead to Eutrophication and subsequent depletion of dissolved oxygen, increase in dissolved organic carbon and poor water quality. Identifying the exact pattern of relationship among the multiple independent variables that result in low TP is important to implement remediation methods. In this data mining study a non-linear model was developed to identify the relationship of multiple independent variables to TP in Truckee River, Nevada, sampled monthly (from January 1997 to October 2004) over six sites, which was minimized non-linearly with respect to TP. Independent variables included were alkalinity, total soluble phosphorus, stream flow, seasonality, man-made intervention, water pH, water temperature, dissolved organic carbon, and dissolved oxygen. SAS® procedure NLP was used to find the pattern of independent variables that minimize TP non-linearly below the compliance level (0.075mg/L) using least squares (LSQ) minimization. Fitted model predicted data closely explaining 96.7% of total variation. Residuals did not show specific pattern. All independent variables influenced TP significantly at 1% level. Overall LSQ minimization solution (0.0694 mg/L) to objective function was below the compliance level and observed (0.117 mg/L) and model predicted (0.113 mg/L) values for mean TP. Solutions to LSQ minimization were below observed mean TP at all sites.

KEYWORDS

Non-linear programming; PROC NLP; PROC NLMIXED; Truckee River; Least Squares Minimization

INTRODUCTION

Water quality management involves issue related to municipal, industrial and amenity irrigation practices. Due to increasing population and urbanization in Nevada in the past few years increased concentrations of total phosphorus in the Truckee River have been recorded. Due to river diversions and increased agricultural practices, heavy growths of aquatic weeds and benthic algae, caused by high nutrient loads and low flows, have plagued the river. Subsequently dissolved oxygen (DO) levels in the river have decreased due to plant respiration and decaying biomass. The low DO levels can impair the river's ability to support populations of Lahontan cutthroat trout, a threatened species, and cui-ui (kwee-wee), a national endangered species. Increased fertilizer use and sewage have modified the natural cycle of phosphorus, the relationships of which to soil use and agricultural, domestic and industrial activities are expected to rise in the future. There is also spatial variability in different catchments in phosphorus loading into Truckee River which imposes tremendous uncertainty in pollution load estimation. For example Steamboat Creek is currently the largest contributor of phosphorus to the Truckee River. Water management practices must be improved in Nevada, to guarantee improved quality of water of sustainable water bodies affected by development of urban and suburban areas. Determination of factors affecting or causing variation of phosphorus concentrations can provide a robust solution to quantify total phosphorus pollution in urban areas in Nevada.

Total phosphorus concentration (TP) in Truckee River varies both temporally and spatially and has been reported to be a function of several factors such as soluble phosphorus concentration (STP), stream flow (SF), seasonality (Summer), man made intervention (X1), dissolved organic carbon (DOC), DO, alkalinity (ALK), water pH (pH), and water temperature (Temp). The degree and spatial variation of influence of the different factors on TP need to be predicted to reflect different sources of the phosphorus loading into the river. The relationship of many of the factors to TP in the river has been found to be non-linear (Ragavan, 2008). Non-linear time series programming and modeling is an appropriate approach to analyze such data. Statistical models in which both fixed and random effects enter nonlinearly are becoming increasingly popular. Perhaps the greatest theoretical progress in time series analysis in the last ten years has been in the understanding of testing and modeling for nonlinearity. Nonlinear time series analysis raises the possibility of improving the power of parameter estimation and forecasting techniques. For any time series Y_t that is normal (and therefore linear) $\rho_k(Y_t) = \{\rho_k(Y_t)\}^2$ (where $\rho_k(\cdot)$ denotes the lag k autocorrelation). Any departure from this result indicates a degree of non-linearity.

The major focus of this data mining research is to apply non-linear programming to identify and model the combination of sites and factors that minimize TP in the River subject to previously identified non-linear constraints, which will enable designers to target and manage TP concentration in the Truckee River accurately as close as possible to their source of origination. The distribution of TP at the different sampling sites is modeled as a function of the independent variables mentioned above. The influence of TP on STP, ALK, SF, Summer, X1, pH, Temp, DOC, and DO at the time of monitoring on TP at six monitoring sites (McCarran Bridge (MC), Wordsworth Bridge (WB), Derby Dam (DD), Steamboat Creek (SC), Lockwood (LW), and North Truckee Drain (NTD)) along Truckee River, Nevada was programmed non-linearly as time-staged and as a location indicator that seek the combination of sites and variables that minimize TP in Truckee River during the study period. The SAS/OR® procedure NLP was used for this purpose. Although monthly data were obtained from Truckee Meadows Water Reclamation Facility (TMWRF, www.tmwrf.com) for the period from 1995 through 2004, only the data from 1997 through 2004 were used in the model due to too many missing observations. The objective function to be minimized for TP was fitted as a non-linear mixed model of the dependent and the independent variables. Data were corrected for missing values, and tested for non-stationarity using the MI procedure and the Augmented Dicky and Fuller test with the ARIMA procedure respectively in SAS® before fitting non-linear mixed model. The non-linear relationships among the independent and the dependent variables were identified and the non-linear mixed model was fitted using the NLMIXED procedure. The objective function for the non-linear programming was thus built using the NLMIXED procedure. The initial values for the parameters of the non-linear model were obtained from a linear mixed model fitted using the MIXED procedure in SAS®. The best model diagnostics, and parameter estimates of the objective function were obtained from PROC NLMIXED and residuals were tested for normality through the standard tests provided with PROC UNIVARIATE. Finally the value of the objective function and the parameters that minimizes TP was obtained through non-linear programming using PROC NLP by least squares minimization. The developed model can provide a guide to probable range and type of TP load generated and deposited into the Truckee River.

STUDY SITE

The Truckee River can be best described as a river in northern California and northern Nevada, that is 140 mi (225 km) long, originates from the mountains, south of Lake Tahoe, flows into the Lake Tahoe at its south end, drains part of the high Sierra Nevada, and empties into Pyramid Lake in the Great Basin (USEPA, 1991). The river passes through the Reno-Sparks metropolitan area, located in Nevada's Truckee Meadows. It flows generally northwest through the mountains to Truckee, California, and then turns sharply to the east and flows into Nevada, past Reno (Figure 1) and Sparks and along the northern end of the Carson Range. East of the Truckee Meadows, fourteen ditches remove water for irrigation. The most significant diversion is Derby Dam, where at least 32% of the river's water is diverted annually (Peternel and Laurel, 2005).

Truckee River's waters are an important source of drinking and irrigation along its valley and adjacent valleys. Increased urbanization and the prevalence of water diversions have caused a decline in water quality, and the resulting detrimental effects on habitat have brought about the need to restore the river to a more natural condition to improve habitat and the river's overall health. The water is quite clear near Lake Tahoe, but as it descends, the water turns muddy and concentrated in nutrients and other toxic elements by the time it passes Reno, Nevada. The California State Water Resources Control Board (State board) has classified under Section 330(d) of the Clean Water Act the middle reach of the Truckee River as "impaired". Because of the endangered species present and due to the fact that Lake Tahoe Basin comprises the headwaters of the Truckee River, the river has been the focus of several water quality investigations, the most detailed starting in the mid-1980s. Under the direction of the U.S. Environmental Protection Agency, comprehensive dynamic studies have been undertaken to study the impacts of a variety of land use and wastewater management decisions throughout the 3120 square mile Truckee River Basin and also to provide guidance to other U.S. river basins (USEPA, 1991). Analytes mostly addressed include nitrogen, phosphate, dissolved oxygen, and total dissolved solids. Impacts upon, the receiving waters of Pyramid Lake has also been analyzed (Source: Truckee River Geographic Response Plan, 2005).

TMWRF currently maintains 11 continuous monitoring stations within the Truckee water system. These stations are located at: Mogul, Steamboat Creek, McCarran Bridge, North Truckee Drain, Lockwood, Patrick, Waltham, Tracy, Painted Rock, Wadsworth and Marble Bluff Dam. Lockwood monitoring site is currently chosen as the compliance site for assessing total phosphorus maximum daily loads (TMDL) into the Truckee River because most controllable sources are thought to be upstream. Lockwood monitoring site is located in the lower Truckee River basin 65.6 river miles from Lake Tahoe located down stream, of McCarran Bridge, North Truckee Drain, and Steamboat Creek monitoring sites and Vista (www.tmwrf.com) (Figure 2). The TMDL compliance level for total phosphorus concentration for Truckee River is currently at 0.075 mg/L (214 lb/day) at Lockwood monitoring site. Existing data indicate that approximately 80 lb/day are attributable to non-point sources and background. The remaining 134 lb/day were set as the total phosphorus waste-load from the TMWRF. The total phosphorus concentration (TP), as classified by the Environmental Protection Agency (NDEP, 1994) is a conservative pollutant (conservative pollutants persist in the water segment of the aquatic environment over time remaining essentially constant in concentrations), hence not perturbed by seasonal variations or other short term cyclical and non-cyclical variations in the system. Concentration of conservative pollutant varies directly with the volumes of flows of dischargers of the receiving water body.



Figure 1: Truckee River in Reno, Nevada



Figure 2: TMWRF Monitoring Stations

NON LINEAR PROGRAMMING

Non Linear Programming (NLP) involves optimizing a continuous non-linear objective function (minimizing or maximizing) $f(x)$ with n independent (decision) variables, $x = (x_1, \dots, x_n)^T$ subject to constraints. Constraints include: i) linear and nonlinear, ii) equality and inequality, and iii) lower and upper bound. For example the optimization (minimization) of objective function $f(x)$ can be expressed as solving: $\min_{x \in R^n} f(x)$ subject to the following constraints:

$$\begin{aligned} c_i(x) &= 0 & i &= 1, \dots, m_e \\ c_i(x) &\geq 0 & i &= m_e + 1, \dots, m \\ u_i &\geq x_i \geq l_i & i &= 1, \dots, n \end{aligned}$$

where c_i 's are the constraint functions, and u_i, l_i 's are the upper and lower bounds. The above setting can be applied to real world problems to find optimal control values and/or maximum likelihood estimates as solutions.

NON LINEAR PROGRAMMING WITH SAS®

The NLP procedure in SAS® hereafter referred to as NLP procedure can be used to handle the following problems: i) quadratic programming, ii) constrained optimization (minimization/maximization), iii) unconstrained optimization (minimization/maximization), and iv) linear complementarity. SAS® provides a number of algorithms for solving the above optimization problems. The quadratic (non-linear) programming problem with the objective function

$$f(x) \text{ to be optimized (minimized or maximized) can be described as: } \min(\max) f(x) = \frac{1}{2} x^T G x + g^T x + b$$

subject to constraints: $c_i(x) = 0 \quad i = 1, \dots, m_e$; where $c_i(x)$'s are linear functions. In the above

example $g = (g_1, \dots, g_n)^T$ is a vector and b is a scalar of parameters, and G is a $n \times n$ symmetric matrix. The least

-squares non-linear programming problem can be described as: $\min f(x) = \frac{1}{2} \{f_1^2(x) + \dots + f_n^2(x)\}$ subject to

constraints: $c_i(x) = 0 \quad i = 1, \dots, m_e$; where the $c_i(x)$'s are linear functions, and $f_1(x), \dots, f_n(x)$ are nonlinear functions of x . In this study the LSQ statement was used instead of the MIN statement to tell the program that it is a least squares problem. The LSQ statement with PROC NLP identifies the objective function as the symbol specified with the LSQ statement. This increases the computing performance and numerical stability of solution.

PROC NLP in SAS® offers the following optimization techniques: i) Quadratic Active Set Technique, ii) Trust-Region Method, iii) Newton-Raphson Method With Line Search, iv) Newton-Raphson Method With Ridging (NRMR), v) Quasi-Newton Methods, vi) Double-Dogleg Method, vii) Conjugate Gradient Methods, viii) Nelder-Mead Simplex Method (NMSIMP), ix) Levenberg-Marquardt Method, x) Hybrid Quasi-Newton Methods. The NRNMR is the default optimization technique when there are no constraints. All the above optimization techniques except NMSIMP support continuous first-order derivatives of the objective function f . Derivatives can be computed in PROC NLP in the following three ways: i) analytically (using a special derivative compiler), which is the default method, ii) using finite difference approximations, iii) through user-supplied exact or approximate numerical functions. The particular

optimization technique must be selected with the TECH= option with PROC NLP. The Quasi-Newton optimization method was used in this study, since it is the only optimization method which allows non-linear constraints of program variables.

Regardless of the optimization technique the objective function and the constraints to satisfy the optimal solution must be specified. All the above optimization techniques require a continuous objective function f . First and second order derivatives of the objective function can also be used. All optimization problems are solved iteratively in PROC NLP. Objective function and the constraints are specified algebraically using SAS programming statements in PROC NLP. A SAS data set can also be used to generate constraints and objectives. Solution to the non-linear programming problem can be saved in output data sets using an OUT= option with PROC NLP. PARMs statement allows parameters and decision variables be declared and be initialized. BOUNDS statement allows specification of boundary values for the parameters. LINCON and NLINCON statements with PROC NLP can be used to specify linear and non-linear constraints respectively for variables used in the objective function statement.

DATA INTEGRITY

NORMALITY AND STATIONARITY

The original TP time series in Truckee River, Nevada was not normal (Figure 3) and non stationary. There were significant spikes in the series autocorrelation function plot (Figure 4). First differencing of the TP series was found adequate to correct the series for non-stationarity (Figure 5) which also converted the data normal (Figure 6). The Augmented Dickey and Fuller test with PROC ARIMA was used to test the data for non-stationarity was used (SAS® code 1).

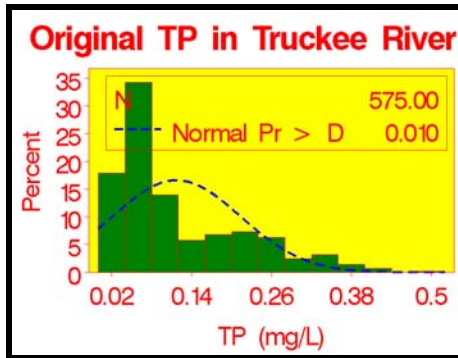


Figure 3: Histogram of the original TP series
With a normal curve superposed

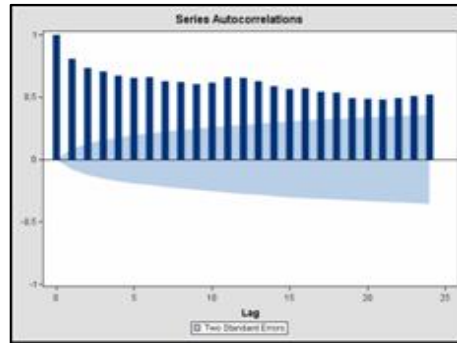


Figure 4: Autocorrelation function plot
of the original TP time series.

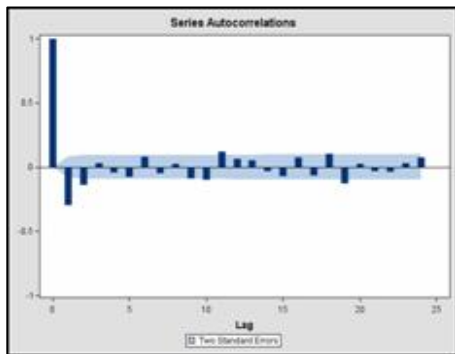


Figure 5: Autocorrelation function plot of
first differenced TP series.

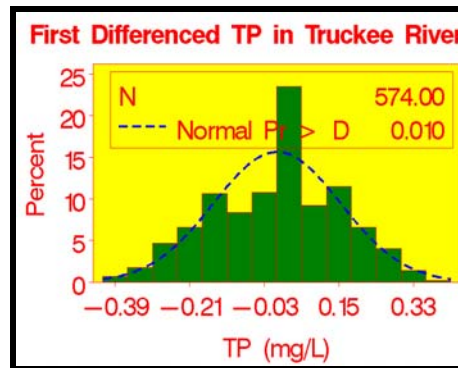


Figure 6: Histogram of the first differenced
TP time series with a normal curve superposed

MISSING VALUES AND OUTLIERS

No outliers and influential observations were encountered in the data. All observations were used in the analysis. Missing values for any decision variables for any observations in the data used can lead to a missing value of the objective function. This will hinder processing of corresponding by group of data. The NOMISS= option with PROC NLP can be used to skip the observations with missing values for any the decision variables. Since initial values for decision variables were used in this study no problem with missing values were encountered during non

linear programming. PROC MI was used to correct data for missing values for linear and non-linear mixed model building.

EXPLORATORY ANALYSIS OF ORIGINAL DATA

DISTRIBUTION OF VARIABLES BY SITE AND TIME

The dependent and independent variable distribution among sites are shown in Figure 7 (a-g). PROC BOXPLOT was used to generate box and whisker plots of distribution after counting the missing values. PROC MEANS with the NMIS= option was used to count the missing values for each variable separately at each site (SAS® Code 2). The distribution of overall TP by year is shown in Figure 7h. Site SC shows the largest mean TP value > 0.2 mg/L followed by site NTD (Figure 7a). Site MB has the lowest mean TP value below the compliance level (0.075mg/L). Mean STP was also the largest in SC followed by NTD (Figure 7e).

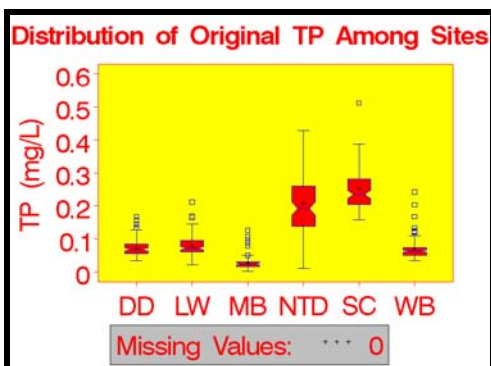


Figure 7a: Distribution of original TP by site

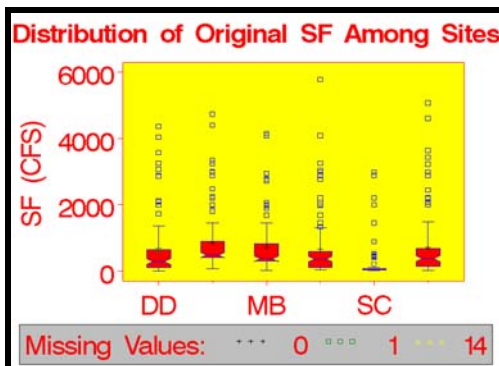


Figure 7b: Distribution of original SF by site

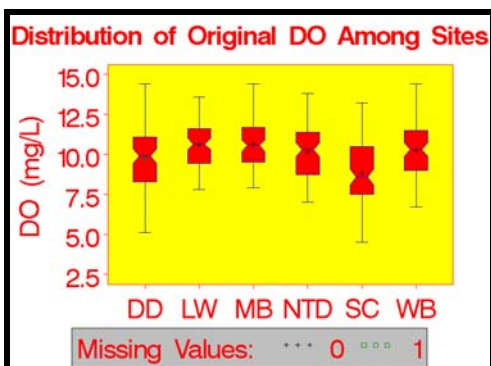


Figure 7c: Distribution of original DO by site

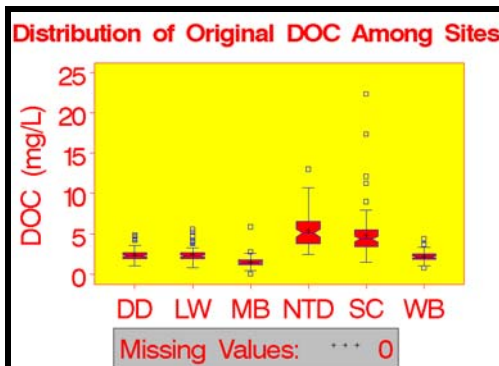


Figure 7d: Distribution of original DOC by site

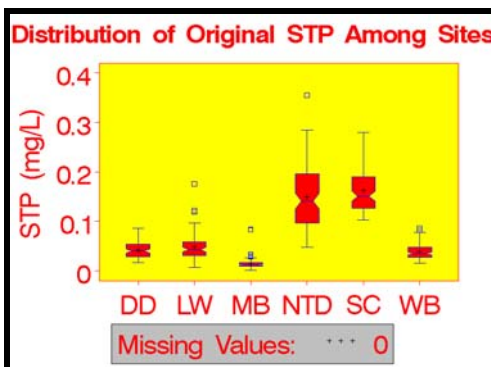


Figure 7e: Distribution of original STP by site

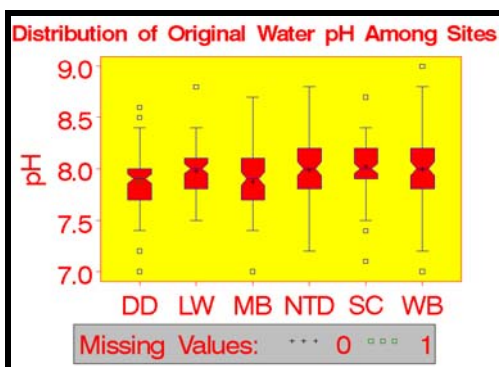


Figure 7f: Distribution of original pH by site

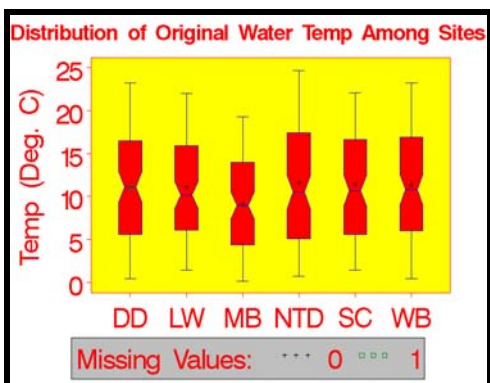


Figure 7g: Distribution of original water Temperature by site

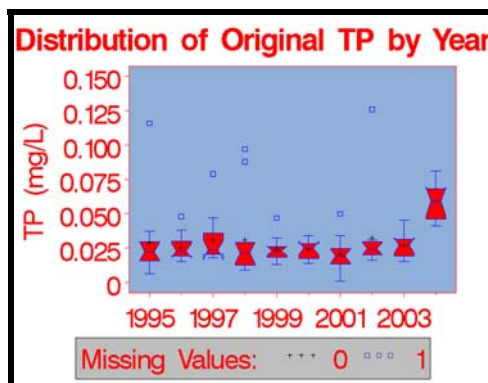


Figure 7h: Distribution of original TP by year

Mean DO was high at all sites despite the high TP levels above the compliance level especially at sites SC and NTD (Figure 7c). Mean water Temperature and pH were almost the same at all sites (Figures 7f and g). Mean of overall TP was below the compliance level during years 1995 through 2003 with a sudden jump during 2004 (Figure 7h). Significance of contribution of individual sites to overall TP in Truckee River at 5% level of significance was studied. Time series cross sectional regression procedure TSCSREG was used to fit a cross sectional regression model. All sites contributes significantly to overall TP ($p < 0.0001$) (Table 1).

OBSERVED RELATIONSHIP OF TP TO DECISION VARIABLES

The following relationships were observed among the dependent variable, TP and the independent (decision) variables. The relationship of TP to DOC, STP and Alkalinity were linear and positive. The relationship of TP to DO, SF, Temp and pH were non-linear (Figures 8a through g). PROC GPLOT was used to obtain the following relationship plots in SAS® (SAS® Code 3).

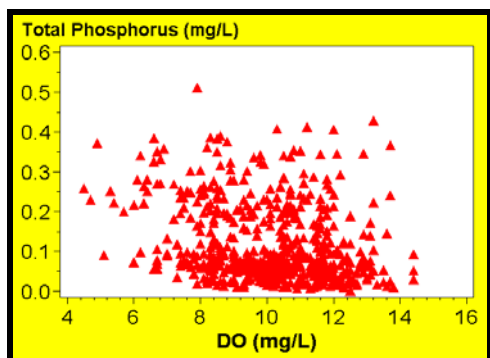


Figure 8a: Original TP versus DO

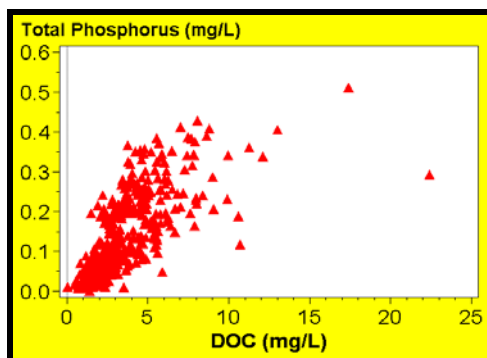


Figure 8b: Original TP versus DOC

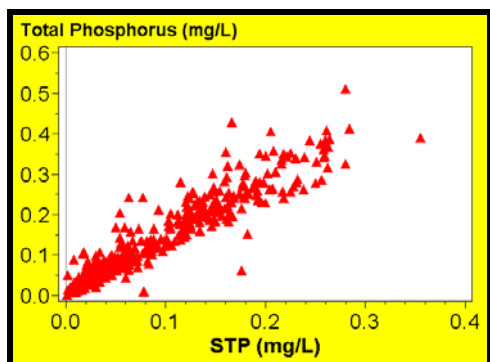


Figure 8c: Original TP versus STP

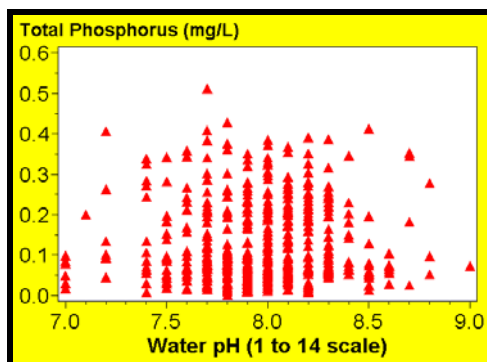


Figure 8d: Original TP versus pH

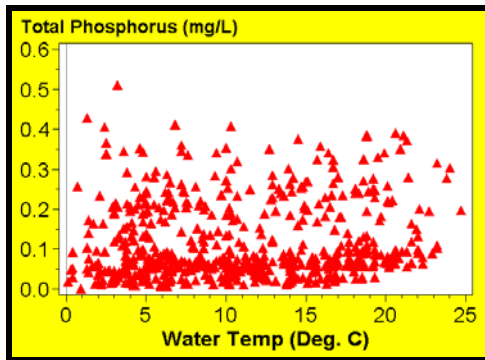


Figure 8e: Original TP versus Temperature

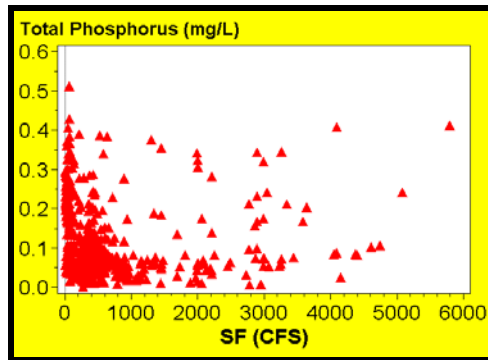


Figure 8f: Original TP versus SF

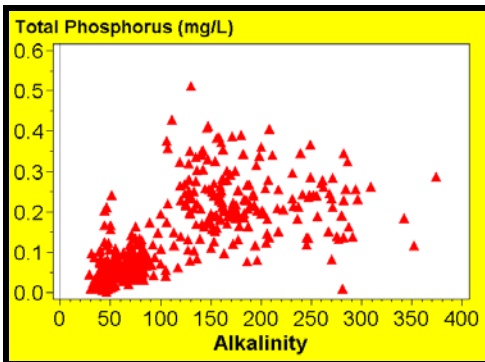


Figure 8g: Original TP versus Alkalinity

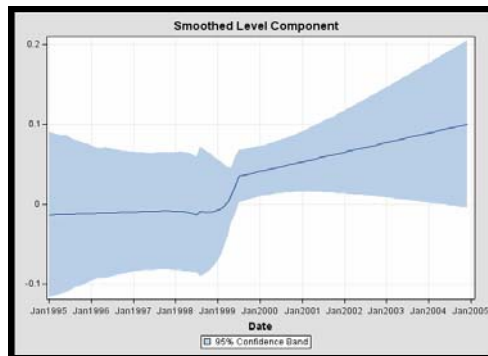


Figure 9: Trend component of original TP

UNOBSERVED COMPONENTS OF ORIGINAL TP

The unobserved variation over the study period from January 1995 through December 2004 of the original TP series were decomposed into trend (Figure 9), seasonal (Figure 10), cyclical (Figure 11), and irregular (Figure 12) components. Variation of TP over the years was seasonal with a period of 12 months. The trend over time of TP in Truckee River was slowly increasing up to 1998 below the overall mean value, with a sudden jump in 1999 and thereafter increasing at an increasing rate above the overall mean value until December of 2004. The cyclical component decreased from 1995 until December of 2004 with a sudden major fall during 1999 which could be due to man-made intervention. All three (trend, seasonal, cyclical) components were significant at the 5% level ($p < 0.05$) and was included as decision variables in the objective function. An intervention analysis was performed to include the cyclical variation in the objective function. PROC UCM was used to decompose the unobserved components of original TP time series (SAS® Code 4).

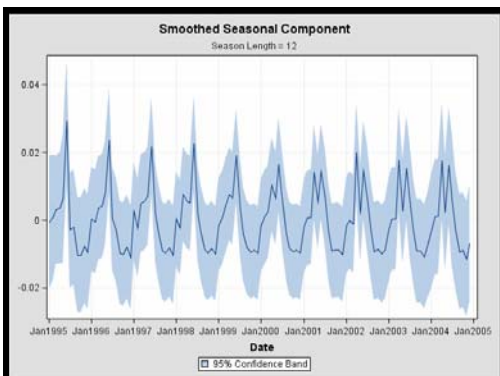


Figure 10: Seasonal component of original TP

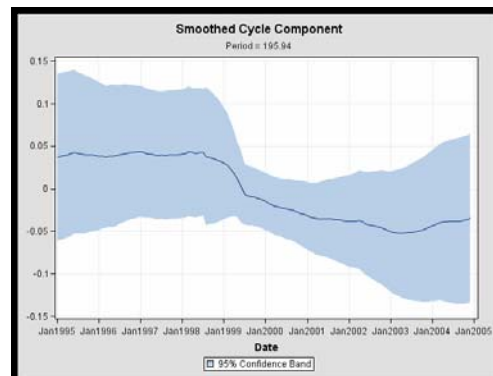


Figure 11: Cyclical component of original TP

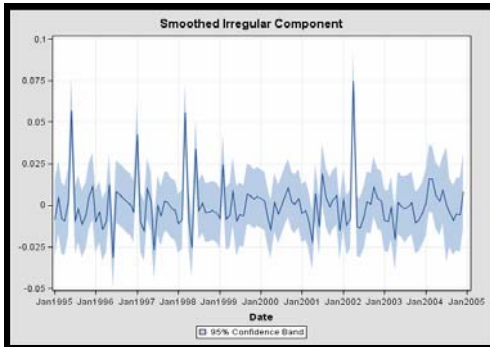


Figure 12: Irregular component of original TP

Table 1: Contribution of individual sites to overall TP

Site	Estimate	Standard Error	DF	t Value	Pr > t
MB	0.02696	0.00514	569	5.25	<.0001
WB	0.06892	0.00514	569	13.41	<.0001
SC	0.2517	0.00517	569	48.72	<.0001
LW	0.07804	0.00514	569	15.19	<.0001
DD	0.07227	0.00514	569	14.06	<.0001
NTD	0.2075	0.00514	569	40.38	<.0001

BUILDING THE OBJECTIVE FUNCTION

The following objective function for TP in Truckee River using the decision variables was built. The non linear relationship among the variables that best fit a non-linear mixed model was identified and the model was built using the NLMIXED procedure (SAS® Code 5). The initial values for the non-linear and linear parameters were obtained using the best parameter estimates from a linear mixed model fitted to the same data in a previous study by the author (Ragavan, 2008). The parameter estimates for all the decision variables were all highly significant ($p < 0.0001$) at the 1% level of significance except DOC ($p = 0.0148$) and the non-linear variable of Temp ($1/(1-\text{Temp})$, $p = 0.0155$) that too were almost significant (Table 2). Hence all the variables were included in the objective function as decision variables. Diagnostic statistics from the best model is shown in Table 3. Although, the interaction effects were not included in the model the unexplained variation was small with 96.7 % (Table 4) of the total variation explained by the model. Mean TP values at SC were slightly underestimated and that at NTD were slightly overestimated (Table 4). A plot of the overall original TP versus predicted TP showed R^2 above 0.90 (Figure 13). The Residuals did not show any particular relationship to the observed or predicted TP and did not have any significant pattern (Figure 14). The residuals were normal. The objective function thus built through non-linear mixed modeling is shown in Figure 15.

Table 2: Parameter estimates and probabilities from the non-linear mixed model

Variable	Parameter	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper	Gradient
Intercept	beta1	-0.3973	0.000306	559	-1297.3	<.0001	0.01	-0.3981	-0.3965	0.000652
DOC	beta2	0.005812	0.002377	559	2.45	0.0148	0.01	-0.00033	0.01195	0.001869
EXP(DO)	beta3	-3.2E-8	0	559	Infy	<.0001	0.01	-Infy	Infy	388.295
LOG(SF)	beta4	-0.00883	0.003100	559	-2.85	0.0046	0.01	-0.01684	-0.00081	0.004489
pH	beta5	0.03741	0.002664	559	14.05	<.0001	0.01	0.03053	0.04430	0.005111
STP	beta6	1.2065	0.000043	559	27993.9	<.0001	0.01	1.2064	1.2066	0.000054
EXP(Summer-12)	beta7	0.10000	0	559	Infy	<.0001	0.01	-Infy	Infy	0.001876
EXP(X1-12)	beta8	0.03674	0	559	Infy	<.0001	0.01	-Infy	Infy	-0.00116
1/(1-Temp)	beta9	-0.01616	0.006658	559	-2.43	0.0155	0.01	-0.03337	0.001048	-0.00279
Alkalinity	beta10	0.000390	0.000078	559	5.02	<.0001	0.01	0.000189	0.000590	0.083531
	g11	-0.00143	0	559	Infy	<.0001	0.01	-Infy	Infy	0
	g12	0.008910	0.000582	559	15.32	<.0001	0.01	0.007406	0.01041	-0.00583
	DDFP	1.0000	0	559	Infy	<.0001	0.01	-Infy	Infy	0

Table 3: Diagnostic statistics of the best Non-linear mixed model

Statistic	Value
-2 Log Likelihood	-1052
AIC (smaller is better)	-1026
AICC (smaller is better)	-1026
BIC (smaller is better)	-970.2

Table 4: Mean of observed and model predicted TP by site and overall

Site	Mean Observed	Mean Predicted
DD	0.072	0.072
LW	0.078	0.081
MB	0.027	0.034
NTD	0.208	0.220
SC	0.252	0.223
WB	0.069	0.067
Overall	0.117	0.113

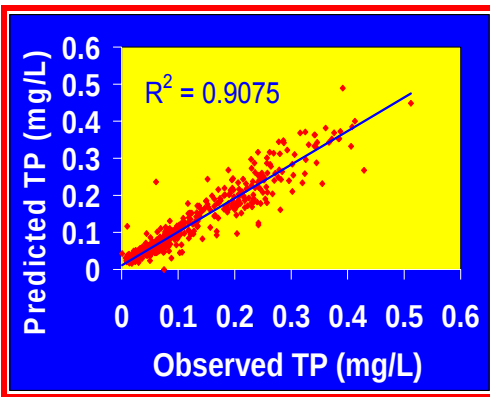


Figure 13: Original versus model predicted TP

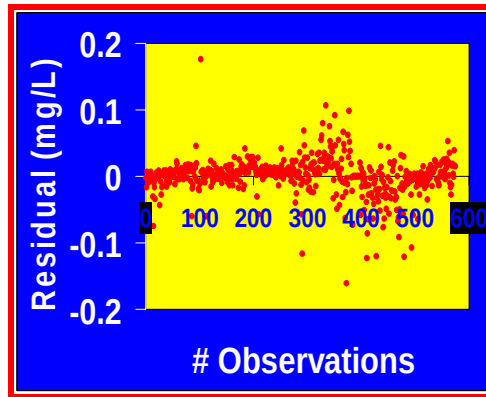


Figure 14: Non-linear mixed model residuals

LEAST SQUARES MINIMIZATION OF OBJECTIVE FUNCTION

Solution to the objective function was obtained through non-linear minimization using PROC NLP (SAS® Code 6). Quasi-Newton optimization method was used. Gradient and the Jacobian of non-linear constraints were computed by Finite Differences. Average values of the decision variables were used as initial values. Appropriate boundary and non-linear constraints were used. LSQ option was used to obtain non-linear least squares minimization solution with non-linear constraints. The parameter estimates from the non-linear mixed model was used as the Initial values for the parameter. This way the solution to the objective function was obtained very fast within few seconds. Value of the objective function was 0.0694 (0.0694223467 rounded to 4 decimals) which is below the compliance value. Value of the Lagrange function was -0.0684 (-0.068472587 rounded to 4 decimals). Value of the objective function is also less than the overall mean TP obtained from the non-linear mixed model and close to the lowest mean value by site. Determinant of the cross product Jacobian matrix was zero with 14 zero Eigen values. Estimates of the parameters with gradient objective function and gradient Lagrange function are shown in Table 5. Summary of optimization procedure (Table 6) and the values for Lagrange multipliers (Table 7) are shown as well as the objective and Lagrange functions by individual sites (Table 8). Objective function value at SC and NTD are above the compliance level of 0.075 mg/L. These two sites require appropriate management practices towards reducing TP loading into Truckee River.

Table 5: Estimates of LSQ Minimization Parameters

Variable	Parameter	Estimate	Gradient Objective Function	Gradient Lagrange Function
Intercept	beta1	-0.3973	0.372619	0.372619
DOC	beta2	0.005812	1.173749	1.173749
EXP(DO)	beta3	-3.20E-10	8542.42631	0
LOG(SF)	beta4	-0.00883	2.207488	2.207488
pH	beta5	0.03741	2.98095	2.98095
STP	beta6	1.2065	0.028319	0.028319
EXP(Summer-12)	beta7	0.1	1.012883	1.012883
EXP(X1-12)	beta8	0.03674	1.012883	1.012883
1/(1/Temp)	beta9	-0.01616	-0.037487	-0.037487
Alkalinity	beta10	0.00039	37.865512	37.865512
Value of Objective Function = 0.0694223467				
Value of Lagrange Function = -0.068472587				

Table 6: Quasi Newton LSQ Minimization Results

Quasi-Newton Parameter	Value
Minimum Iterations	0
Maximum Iterations	200
Maximum Function Calls	500
Iterations Reducing Constraint Violation	20
ABSGCONV Gradient Criterion	0.00001
GCONV Gradient Criterion	1.00E-08
ABSFCONV Function Criterion	0
FCONV Function Criterion	2.22E-16
FCONV2 Function Criterion	1.00E-06
FSIZE Parameter	0
ABSXCONV Parameter Change Criterion	0
XCONV Parameter Change Criterion	0
XSIZE Parameter	0
ABSCONV Function Criterion	-1.34E+154
Line Search Method	2
Starting Alpha for Line Search	1
Line Search Precision LSPRECISION	0.4
DAMPSTEP Parameter for Line Search	.
FD Derivatives: Accurate Digits in Obj.F	15.6535598
FD Derivatives: Accurate Digits in NLCon	15.6535598
Singularity Tolerance (SINGULAR)	1.00E-08
Constraint Precision (LCEPS)	1.00E-08
Linearly Dependent Constraints (LCSING)	1.00E-08
Releasing Active Constraints (LCDEACT)	.

OPTIMIZED MODEL FOR TP IN TRUCKEE RIVER

$$\begin{aligned}
 &SF1=LOG(SF) \\
 &Summer1=EXP(Summer-12) \\
 &X11=EXP(X1-12) \\
 &Temp1=(1/(1-Temp)) \\
 &DO1=EXP(DO) \\
 &TP = -0.3973+0.005812*DOC-3.2E-10*DO1- \\
 &\quad 0.00883*SF1 + 0.03741*pH+ 1.2065*STP+ \\
 &\quad 0.1*Summer1+0.03674*X11-0.01616*Temp1+ \\
 &\quad 0.00039*Alkalinity
 \end{aligned}$$

Figure 15: Objective function used for least squares minimization of TP

Table 7: First order Lagrange multipliers

Active Constraint	Lagrange Multiplier
Equal BC	beta3
Nonlin IC	[2]
Nonlin IC	[4]
Nonlin IC	[6]

Table 8: Objective and Lagrange Functions by site

Site	Objective Function	Lagrange Function
MB	0.0327	-0.0618
WB	0.0460	-0.0663
SC	0.1347	-0.0574
LW	0.0487	-0.0668
DD	0.0463	-0.0663
NTD	0.1288	-0.0591
Overall	0.0694	-0.0685

SAS® CODE

SAS® CODE 1

```
PROC ARIMA DATA=monthly;
IDENTIFY VAR=tp(1,12)
STATIONARITY=(adf=(1,2,4,6,12));

RUN;
```

SAS® CODE 3

```
PROC GPLOT DATA=monthly;

PLOT tp * Alkalinity / overlay
legend=legend1 href= 0
haxis=axis1 HMINOR=4 vaxis=axis2
vminor=1;
LABEL tp='Total Phosphorus (mg/L)'
Alkalinity='Alkalinity';

RUN;
```

SAS® CODE 4

```
PROC UCM DATA=monthly PRINTALL;
ID Date INTERVAL=Month;
MODEL tp;
IRREGULAR plot=smooth;
LEVEL variance=0 noest plot=smooth;
SLOPE variance=0 noest plot=smooth;
CYCLE rho=1 noest=rho plot=smooth;
SEASON length=12 plot=smooth;

RUN;
```

SAS® CODE 2

```
PROC MEANS DATA=monthly NOPRINT;
VAR tp;
BY site;

OUTPUT OUT=Cancel NMISS=ncancel;

DATA Comp;
MERGE monthly Cancel;
by site;

RUN;

symbol1 v=plus c=black;
symbol2 v=square c= red;
symbol3 v=triangle c=yellow;
TITLE 'Distribution of
Original TP Among Sites';

PROC BOXPLOT DATA=Comp;
PLOT tp*site = ncancel /
boxstyle = schematicid
cboxes=blue cboxfill = red
cframe=vligb nohlabel
symbollegend = legend1
notches;

legend1 label=('Missing Values:')
cborder = black cframe=ligr;
label tp ='TP (mg/L)';

RUN;
```

SAS® CODE 5

```
PROC NLMIXED DATA=Monthly QPOINTS=10 ALPHA=0.01 TECH=QUANEW; UPDATE=DDFP;

PARMS
beta1=0.000246 beta2=0.006199 beta3=-0.000000548 beta4=-0.000001 beta5=-
0.0000002455 beta6=-0.7867 beta7=0.1 beta8=0.03665 beta9=-0.00066
beta10=-0.002 g11=-0.001428 to 0.02 by 0.001 g12=-0.001 to 0.01 by 0.001;

eta = beta1+ beta2*DOC+beta3*(EXP(D0))+beta4*(LOG(ABS(SF)))+beta5*
pH+beta6*STP+beta7*(EXP(Summer-1))+beta8*(EXP(X1-1)+ beta9*(1/(1-
Temp)))+ beta10*(Alkalinity)+g12*b1;
num = eta; mu= num;

MODEL tp ~ NORMAL(mu,g12);
RANDOM b ~ NORMAL(0,g12) SUBJECT=Site
PREDICT mu OUT=cdf;

RUN;
```

SAS® CODE 6

```

PROC NLP PALL TECH=quanew CLPARM=BOTH BEST=10 FD=Forward OUTMOD=model;
LSQ tp;

PARMS
    beta1= -0.3973, beta2= 0.005812, beta3 = -0.0000000032, beta4=-0.00883,
    beta5= 0.03741, beta6= 1.2065, beta7= 0.1, beta8= 0.03674, beta9=-
    0.01616, beta10= 0.00039, sf =374.0, Temp=10.94, summer1=1, x11=1,
    do2=10.04, pH=8.0, doc=3.15, stp=0.076, Alkalinity=101.62;
BOUNDS
    -3.2E-10 <= beta3 <= 3.2E-10, 0 < beta7 <= 1, 0 < beta8 <= 1;
NLINCON
    nlc1-nlc2 > 0., 6.14421E-06 <= nlc3 <= 1.67017E-05,
    6.14421E-06 <= nlc4 <= 1.67017E-05, 1 <= nlc5 <= 5.18471E+21,
    1 <= nlc6 <= 14, nlc7-nlc10 >= 0.;
    nlc1 = Log(sf); nlc2= (1-Temp); nlc3=EXP(summer1);
    nlc4 = EXP(x11); nlc5=EXP(do2); nlc6=pH; nlc7=doc;
    nlc8=stp; nlc9=Alkalinity; nlc10=do2; nlc11=(1/nlc2);

tp = ((beta1 + beta2*nlc7 + beta3*nlc5 + beta4*nlc1 + beta5*nlc6+
    beta6*nlc8 + beta7*nlc3 + beta8*nlc4 + beta9*nlc11
    + beta10*nlc9));

RUN;

```

CONCLUSIONS

Tend of observed TP in Truckee River is significantly on the increase. All the six sites studied contribute significantly towards TP in Truckee River. The non-linear objective function built predicted the original TP in the Truckee River closely accurately explaining 96.7 percent of the total variation. TP in Truckee River can be predicted accurately as a function of these seven decision variables using the given objective function. All the independent variables were highly significant at 1% significance level ($p < 0.01$) except the non-linear variation of water Temperature ($p = 0.0155$). The relationship of DO to TP is exponential and that of SF is logarithmic. The relationship of DOC and STP to TP is linear and positive. The relationship of Alkalinity to TP is almost linear and positive. The relationship of SF and DO to TP is non-linear and negative. The TP is also seasonal with a 12 month period and affected by cyclical man-made intervention events. The least squares minimization value of the objective function (0.069 mg/L) for TP is almost close to the compliance level for TP (0.075 mg/L) in Truckee River. Sites NTD and SC contribute significantly to TP loading into the River. TP loading from these two sites require careful monitoring to reduce build up of TP and subsequent Eutrophication in Truckee River.

REFERENCES

- Akaike, H. (1974) A new look at the statistical model identification, IEEE trans.: Autom. Control, AC-19, 716-723.
- Box, G.E.P., and Jenkins, G.M. (1976) Time series Analysis Forecasting and Control, (2nd ed.): Holden-Day, San Francisco, Ca.
- Buse, A. (1973) Goodness of Fit in Generalized Least Squares Estimation, American Statistician, 27, 106-108.
- DaSilva, J.G.C. (1975) The Analysis of Cross-Sectional Time Series Data, Ph.D. dissertation, Department of Statistics, North Carolina State University.
- Fuller, W. (1978) Introduction to Time Series, New York: John Wiley & Sons, Inc.
- Harville, D.A. (1988) Mixed-Model Methodology: Theoretical Justifications and Future Directions, Proceedings of the Statistical Computing Section, American Statistical Association, New Orleans, 41-49.
- NDEP (1994) Truckee River final total maximum daily loads and waste load allocations. Nevada Division of Environmental Protection, Carson City, Nevada.
- Parks, R.W. (1967) Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated, Journal of the American Statistical Association, 62, 500-509.
- Peternel, K., and Laurel, S. (May 15-May 19, 2005) Truckee River Restoration Modeling, World Water and Environmental Resources Congress. Anchorage, Alaska, USA.
- Ragavan, A. (2008) Data Mining Application of Non-Linear Mixed Modeling in Water Quality Analysis, Proceedings of the Data Mining and Predictive Modeling Section, SAS® Global Forum, San Antonio, TX, Paper 140-2008.
- Schafer, J.L. (1999) Multiple Imputation: A Primer, Statistical Methods in Medical Research, 8, 3-15.
- Schafer, J.L. (1997) Analysis of Incomplete Multivariate Data, New York: Chapman and Hall.
- Searle, S. R. (1988) Mixed Models and Unbalanced Data: Wherefrom, Whereat, and Whereto?, Communications in Statistics - Theory and Methods, 17(4), 935-968.
- Searle, S. R., Casella, G., and McCulloch, C.E. (1992) Variance Components, New York: John Wiley & Sons, Inc.
- Truckee Meadows Water Reclamation Facility: www.tmwrf.com
- Truckee River Geographic Response Plan, 2005: http://ndep.nv.gov/bca/emergency/truckee_river_plan05.pdf
- USEPA (1991) Guidance for water quality-based decisions: The TMDL process. EPA 440/4-91-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Name: Anpalaki J. Ragavan M.S.
Enterprise: Department of Mathematics and Statistics, University of Nevada, Reno
Address: 3925, Clear Acre Lane, # 188, Reno, NV 89512, USA.
Work phone: (775)-327-5260
Home phone: (775)-674-0397
Email: ragavan@unr.edu

SAS and all other SAS institute Inc. product or service names are registered trade marks or trade marks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trade marks of their respective companies.