NONPARAMETRIC TESTS FOR TRENDS IN WATER-QUALITY DATA USING THE STATISTICAL ANALYSIS SYSTEM

by Charles G. Crawford, James R. Slack,
and Robert M. Hirsch

U.S. GEOLOGICAL SURVEY
Open-File Report 83-550

1983

U.S. DEPARTMENT OF THE INTERIOR

JAMES G. WATT, Secretary

GEOLOGICAL SURVEY

Dallas L. Peck, Director

For additional information write to:

Chief Hydrologist
U.S. Geological Survey, WRD
410 National Center
Restion, Virginia 22092

Copies of this report can be purchased from:

Open-File Services Section
Western Distribution Branch
U.S. Geological Survey
Box 25425, Federal Center
Lakewood, Colorado 80225
(Telephone: [303] 234-5888)

TABLE OF CONTENTS

ILLUSTRATIONS

CONTENTS (continued)

TESTING FOR TRENDS IN WATER-QUALITY DATA USING THE STATISTICAL ANALYSIS SYSTEM

By Charles G. Crawford, James R. Slack, and Robert M. Hirsch

ABSTRACT

Two nonparametric procedures to test for trends in water-quality data (SEASKEN AND SEASRS) have been developed for the Statistical Analysis System* (SAS). The procedure SEASKEN tests for a monotonic trend in time by a modified form of Kendall's tau, the Seasonal Kendall test. The procedure SEASRS tests for a step trend between two different periods in a time series using a modified form of the Wilcoxon (Mann-Whitney) rank sum test, the Mann-Whitney-Wilcoxon rank sum test for seasonal data. Examples are presented using the two procedures. The source code and user's guide for each of the two procedures are also presented.

Procedures for flow adjusting water-quality data by the SAS procedures REG and SYSREG and techniques for plotting water-quality data as a time series by the SAS procedure PLOT are presented.

Additionally, examples are presented to demonstrate the use of the U.S. Geological Survey procedures QWRETR, DVRETR, QWSAS, and DVINPUT to retrieve data from the Geological Survey WATSTORE system and make it available to SAS.

INTRODUCTION

Increased public concern over the quality of the Nation's rivers in the past several decades has led Federal, State, and local officials to implement or greatly expand existing water-quality monitoring programs. Examples of such programs are the Geological Survey's Benchmark and NASQAN networks (see Briggs, 1978). Most of these monitoring programs have as a goal the detection of trends in water quality.

---

* Use of brand and firm trade names in this report is for identification purposes only and does not constitute endorsement by the U.S. Geological Survey.

Techniques of trend detection in water-quality data have correspondingly received much attention recently in the literature (see, for example, Fuller and Tsokos, 1971; Lettenmaier, 1976; and Hirsch and others, 1982). Most of the trend detection methods presented to date are based on classical (parametric) hypothesis testing. However, because of the nature of water-quality data (typically skewed, serially correlated, and showing seasonality), many of the assumptions underlying classical hypothesis tests are not met, rendering them inappropriate. (For a discussion of the problems associated with these tests, the reader is referred to Smith and others, 1982.) Recently, however, thinking has begun to shift toward distribution-free (nonparametric) tests that have less restrictive assumptions than their classical counterparts and are therefore less sensitive to the distribution of the water-quality time series.

This report describes procedures in the Statistical Analysis System (SAS) that can appropriately be used to detect trends in water-quality data. The report describes the use of both the standard procedures provided with SAS and two additional procedures, SEASKEN and SEASRS. A working knowledge of SAS by the reader is assumed. The report was written primarily for user's of the U.S. Geological Survey's WATSTORE data base and Amdahl computer system; however, the source code for the SAS macros and procedures are included in the appendices. The user's guide for the SEASKEN and SEASRS procedures are also included in the appendices.

RETRIEVING DATA FROM THE WATSTORE DAILY VALUES AND WATER QUALITY FILES

In order to use SAS on water-quality data, it is first necessary to get the data into a SAS data set. A SAS data set is a collection of data observations addressable by SAS procedures. Each observation has one or more variables associated with it. For more information about SAS data sets, see SAS Institute, Inc.

(1979) or SAS Institute, Inc. (1982a). Data can be entered into a SAS data set from data cards in the job stream or by reading data stored on disk or tape files. When using data from the WATSTORE file, one of the standard Survey retrieval procedures must first be used. One of two Survey applications programs (PROC QWSAS or the DVINPUT macro) is then called to convert the standard retrieval output into a SAS data set. PROC QWSAS is a SAS procedure that produces a SAS data set from the standard water quality file retrieval procedure QWRETR. PROC QWSAS is described in detail in the WATSTORE user's guide, volume 3, chapter IV, section R. DVINPUT is a SAS macro that converts output from the WATSTORE daily values file retrieval, DVRETR, to a SAS data set. Use of the DVINPUT macro is described in the WATSTORE message SAS documentation section (member WRD06). The standard retrieval procedures, DVRETR and QWRETR, are discussed in the WATSTORE user's guide, volumes 1 and 3, respectively.

Figure 1 shows example input using the QWRETR and QWSAS procedures. This example retrieves twelve parameters - two streamflow parameters (00060 & 00061) and ten water-quality constituents - for three stations from the water quality file. Both daily mean streamflow (parameter code 00060) and instantaneous streamflow (00061) should be retrieved. For purposes of this test, the mean streamflow may be used if the instantaneous streamflow is missing. PROC QWSAS creates a SAS data set named DATA1 containing the station identification number and the eleven parameters requested in the QWRETR procedure. The parameter values are stored in variables named Pnnnnn where nnnnn is the WATSTORE parameter codes given in the retrieval list. In addition, the variables YEAR, MONTH, DAY, DATE, DECTIME, and SNAME were requested in the QWSAS statement. The variable DATE is the day on which the sample was collected, in days since January 1, 1960. DECTIME is a decimal number representing the time the sample was collected, in

```
//F1 JOB
// CLASS
/*SETUP     118545/H
//PROCLIB DD DSN=WRD.PROCLIB,DISP=SHR
// EXEC QWRETR,VOL1=118545
//HDR.SYSIN DD *
M3  1968100119810930
XQWSASX
R000600006100410006300066500915009250093000940009450095070300
000060000610041000630006650091500925009300094000945009507030O
D       03276500
D       03374100
D       03378500
/*
// EXEC WRDSAS
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
PROC QWSAS YEAR MONTH DAY DATE DECTIME SNAME;
DATA;SET;IF P00061=. THEN P00061=P00060;DROP P00060;
PROC SORT;BY STATION YEAR MONTH DAY;
PROC PRINT;BY STATION;
VAR YEAR MONTH DAY DECTIME DATE   P00061 P00410 P00630 P00665 P00915 P00925 P009
30 P00940 P00945 P00950 P70300;
DATA TREND.MONTHLY;SET;
/*
//
```

Figure 1.--Example input for WATSTORE retrieval of water-quality data by the QWRETR and QWSAS procedures.

years. These two forms of sample collection time are useful in plotting the data as time series. Additionally, the variable DATE may be formatted in several different ways (see SAS Institute, Inc., 1982a, p. 409). SNAME is the variable containing the station name.

The variable DECTIME is required to use the SAS procedures SEASKEN and SEASRS. For data sets that do not already include the variable DECTIME, it can be easily generated using the following statement in a SAS data step:

DECTIME = YEAR(DT)+(JULDATE(DT)-YEAR(DT)*1000)/365;, (1)

where DT is the date the observation was made.

An example use of this is shown in figure 2.

The example in figure 1 also sorts and prints the information retrieved by station. Finally, the data is stored as file TREND.MONTHLY in the data set USERID.FILENAME on a direct access device for later analysis. Line 16 describes the existing file to be used for data storage, and line 24 copies the data to the disk file. For information on creating disk data sets on the USGS Amdahl computer system, see the USGS Computer Users Manual, Chapter 5.

To adapt this example input to a specific application, the user will need to (1) substitute the six digit volume number of the appropriate water quality back file tape for 118545 in lines 3 and 5; (2) substitute the appropriate retrieval dates on the master control card on line 7; (3) change the retrieval and output list as desired on lines 9 and 10; (4) substitute the desired station numbers in lines 11 through 13 (and delete or add D cards as required); (5) substitute an appropriate data set name in the DSN = field on line 16; and (6) change the variables list in lines 22 and 23 to agree with the parameters listed in the retrieval list and PROC QWSAS optional variables.

```
//F2 JOB
// CLASS
/*SETUP    115620/H
//PROCLIB DD DSN=WRD.PROCLIB,DISP=SHR
// EXEC DVRETR,AGENCY=USGS,VOL1=115620
//HDR.SYSIN DD *
M3  197010019800930
R000608015480155
F00003
D        03365500
/*
// EXEC WRDSAS,MACRO=DV,DSN='&&BKREC',DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
DVINPUT _SNAME
DATA DATAA(RENAME=(VALUE=P00060)) DATAB(RENAME=(VALUE=P80154)) DATAC(RENAME=(VAL
UE=P80155));SET;
IF PARMCODE=60 THEN OUTPUT DATAA;
IF PARMCODE=80154 THEN OUTPUT DATAB;
IF PARMCODE=80155 THEN OUTPUT DATAC;
DROP PARMCODE;
PROC SORT DATA=DATAA;BY STATION DATE;
PROC SORT DATA=DATAB;BY STATION DATE;
PROC SORT DATA=DATAC;BY STATION DATE;
DATA TEMP;MERGE DATAA DATAB DATAC;BY STATION DATE;
FORMAT DATE YYMMDD8.;
J=JULDATE(DATE);Y=YEAR(DATE);
DECTIME=Y+(J-Y*1000)/365;
LABEL P80154=SUSPENDED SEDIMENT CONCENTRATION (MG/L)
      P80155=SUSPENDED SEDIMENT DISCHARGE (TONS/DAY)
      P00060=STREAMFLOW (CFS);
PROC SORT;BY STATION DATE;
PROC PRINT;BY STATION;VAR DATE DECTIME P00060 P80154 P80155;
DATA TREND.DAILY;SET TEMP;
/*
//
```

Figure 2.--Example input for WATSTORE retrieval of streamflow data by the DVRETR and DVINPUT procedures.

The lines in the QWRETR example in figure 1 do the following:

Line 1. Job control.

2. Class (extension of line 1).

3. Instructs the computer operator to mount the specified backfile tape.

4. Invokes the WRD cataloged procedure library.

5. Invokes the procedure QWRETR and specifies the mounted backfile tape to be used in the retrieval.

6. Establishes that data for QWRETR follow.

7. Specifies that data from both the current and backfile are to be retrieved for the period October 1, 1968, to September 30, 1981.

8. Specifies that QWSAS will be invoked as an application program.

9. Restricts the retrieval to the listed parameters only.

10. Restricts the output list to the listed parameters only.

11-13. Specifies the stations to be included in the retrieval.

14. Job control language step separator card (step delimiter).

15. Invokes the procedure WRDSAS (WRD modified version of SAS).

16. Defines existing direct access data set to be used for storage of the retrieved data.

17. Establishes that SAS instructions follow.

18. Invokes the procedure QWSAS and requests the variables YEAR, MONTH, DAY, DATE, DECTIME, and SNAME to be included in the data set created in line 23 below.

19. Uses mean streamflow (parameter code 00060) if instantaneous streamflow (00061) is missing, and discards the mean streamflow parameter.

20. Sorts data set in order of variables listed in BY statement.

21. Prints data set.

22-23. Selects and orders variables to be included in print of data set.

24. Creates file TREND.MONTHLY containing the retrieved data and stores it in the data set USERID.FILENAME.

25. Step delimiter.

26. End of job.

Figure 2 shows example input using the DVRETR and DVINPUT macro procedures. This example retrieves streamflow, suspended-sediment concentration, and suspended-sediment discharge from the daily values file for one station. The DVINPUT macro creates a SAS data set containing the station identification number, parameter code, value, date, and the optional variable requested on the DVINPUT statement, station name. The variable PARMCODE contains the values of the WATSTORE parameter codes in the retrieval list. This format of the data set is awkward since all the variables requested in the retrieval list are combined into different observations of the variable PARMCODE. The statements in lines 16 through 31 convert the initial SAS data set into one containing the station identification number, station name, date, P00060, P80154, and P80155. In addition, lines 27 and 28 add the variables Y (year), J (Julian date), and DECTIME for each observation, and lines 29 through 31 add variable labels for suspended-sediment concentration, suspended-sediment discharge, and streamflow.

This example also sorts and prints the data contained in the modified data set named TEMP. Finally, the data set is stored as the file TREND.DAILY on the data set USERID.FILENAME on a direct access device.

Note that DVINPUT is a SAS macro and not a SAS procedure. It is not preceded by the statement PROC or followed by a semicolon.

To adapt this example setup to a specific application, the user will need to 1) substitute the six-digit volume number of the appropriate daily values backfile tape for 115620 in lines 3 and 5; 2) substitute the appropriate retrieval dates on the master control card on line 7; 3) change the retrieval and output list as desired on lines 8 and 9; 4) substitute the desired station numbers in line 10 (add or delete D cards as required); 5) change lines 18-20 and 29-31 appropriately; 6) substitute an appropriate data set name in the DSN = field on line 13; and 7) change the variable list in line 33 to agree with the parameters listed in the retrieval list and the DVINPUT optional variables.

The lines in the DVRETR example input of figure 2 do the following:

Line 1. Job control.

2. Class (extension of line 1).
3. Instructs the computer operator to mount the specified backfile tape.
4. Invokes the WRD cataloged procedure library.
5. Invokes the procedure DVRETR with AGENCY=USGS being the agency code and specifies the mounted backfile tape to be used in the retrieval.
6. Establishes that data cards for DVRETR follow.
7. Establishes that data from both the current and backfile are to be retrieved for the period October 1, 1970, to September 30, 1980.
8. Restricts the retrieval to the listed parameters only.
9. Restricts retrieval to listed statistics codes only.
10. Specifies the station to be included in the retrieval.
11. Step delimiter.
12. Invokes the procedure WRDSAS and defines the temporary data set containing retrieval data and supplies the macro DV.

13. Defines existing direct access data set to be used for storage of the retrieved data.

14. Establishes that SAS instructions follow.

15. Invokes the macro DVINPUT and requests that the variable SNAME be included in the created SAS data set.

16-31. Converts the SAS data set.

32. Sorts data set in order of variables listed in BY statement.

33. Prints the variables in the VAR statement in the order they are listed in the VAR statement.

34. Creates file TREND.DAILY containing the retrieved data and stores it in the data set USERID.FILENAME.

35. Step delimiter.

36. End of job.

DETERMINING THE RELATIONSHIP BETWEEN WATER-QUALITY CONSTITUENTS AND STREAMFLOW

Quite frequently, concentrations of water-quality constituents are related to streamflow. When a water-quality constituent and streamflow are related, apparent trends in water quality may be due only to fluctuations in streamflow rather than to changes in the processes that affect the introduction and fate of a given constituent in the stream. For example, consider a stream where dissolved solids and streamflow are negatively correlated. That is, as streamflow increases dissolved solids decrease and vice versa. During a period of drought, high dissolved solids concentrations would be expected. If this period of drought was followed by a period of wet weather, a decrease in dissolved solids concentrations would be expected. If such a time series was tested for trend in dissolved solids concentration, a significant downtrend would be indicated. However, such a trend could be entirely attributable to the fluctuation in streamflow during the period. In order to test for trends in the processes affecting dissolved solids during the period, it would be necessary to remove the effect of streamflow. Flow adjustment is an attempt to remove a major source of variation in water quality (streamflow) which may be masking those variations attributable to changes in the constituent inputs to the stream or in the processes occurring in the stream.

Smith and others (1982) described a flow-adjustment procedure suitable for this purpose. Their approach is to develop a time series of flow-adjusted concentrations (FAC) and to test that series for trend. FAC is defined as the actual concentration (C) minus the expected concentration ($\hat{C}$) predicted from the discharge (Q) relationship. The FAC should be randomly distributed with a mean of zero over the period of record if no change in the processes that affect the water-quality constituent have occurred. Of course, for some constituents - e.g., biological - adjustment by some other variable - e.g., solar radiation or air temperature - might be appropriate. Where some of the data are reported as "less than" a detection limit, these approaches to flow adjustment are not valid. If only a very few values are reported as "less than," then flow adjustment could be used provided some sensitivity analysis were done to check that the choice of values to use in place of the "less than" was not very influential in the overall results.

Some common models used for flow adjustment include the following:

(2) $\hat{C} = a+b\ Q$ linear

(3) $\hat{C} = a+b \ln (Q)$ log-linear

(4) $\hat{C} = a+b \frac{1}{1+\beta Q}$ hyperbolic, β a constant typically in the range $10^{-3}\ \overline{Q}^{-1} \leq \beta \leq 10^{2}\ \overline{Q}^{-1}$, where $\overline{Q}$ is mean discharge

(5) $\hat{C} = a+b \frac{1}{Q}$ inverse

(6) $\hat{C} = a+b_1Q+b_2Q^2$ quadratic

A good guide to selecting a flow adjustment equation is the R^2 value, but one should check plots of the predicted $(\hat{C})$ and observed (C) values versus Q (a log Q scale is usually desirable), and plots of the residual versus the predicted (Q) to confirm that the relationship fits well and is not excessively heteroscedastic (e.g., variance increases as predicted concentrated increases). For many constituents (particularly suspended constituents or biological ones like bacteria or plankton), these models may be inadequate, because the constituents are very heteroscedastic. In these cases, models based on fitting the log concentrations may be preferred.

Candidate models include the following:

(7) $\ln \hat{C} = a+b \ln Q$ log-log

(8) $\ln \hat{C} = a+b_1 \ln Q+b_2 (\ln Q)^2$ log-quadratic log

Deciding between a model based on concentration (equations 2-6) and one based on log concentration (equation 7 or 8) should <u>not</u> be based on R^2 values. Rather, the decision should be based on examination of residuals plots.

If none of the models considered results in a significant fit, as determined from the probability values for the t statistics in the cases with one explanatory variable (equations 2-5 or 7) or the probability value of the F statistic (equations 6 or 8), then no flow adjustment should be performed.

If one of the linear models (equations 2-6) is used, the residuals (Flow Adjusted Concentrations) are defined as $C - \hat{C}$. If one of the log models (equations 7 - 8) is used, the residuals are $\ln C - \widehat{\ln C}$. Note that in the former case the residuals have the dimensions of C (typically mg/L), but in the latter case they are dimensionless. This has important implications for the interpretation of trend test results. If log models are used and a slope is estimated in Proc SEASKEN, it must be transformed as follows: If B is the slope value reported by Proc SEASKEN, then $(e^B-1)\cdot100$ is the change in percent per year. If log models are used and a step change is being considered and D is the difference (mean FAC of period 2 minus mean FAC of period 1), then the percentage change from period 1 to period 2 would be $(e^D-1)\cdot100$.

The following section provides two examples of SAS jobs to search for appropriate flow adjustment models. The first, using the macro ABREGMAC, simply estimates the regression equation and computes the usual summary statistics; the second, using the macro REGMAC, provides extensive diagnostics on the results. These macros are intended to be illustrative of the kinds of analyses one may want to run. Knowledge gained by working with a particular data set should dictate variations of these that one may choose to pursue.

The SAS job listing and output in figure 3 follow the procedures used by Smith and others (1982) but includes several other models as well. It considers all of the models defined by equations 2-8. Plots of C versus Q and log C versus log Q are produced to aid in model selection.

Simple regression is used to estimate the coefficients a and b of each model, to calculate R^2 (the fraction of the variance of the dependent variable explained by the given function of Q) and p (the probability of erroneously rejecting the null hypothesis that b = 0). β is equal to $10^{(-2.5\ -\ \beta^*)}$ where β^* is the interger part of $\log_{10}\overline{Q}$, where $\overline{Q}$ is the mean streamflow. Eight different hyperbolic models are generated by incrementing the initial value of β by a factor of $10^{0.5}$ seven times.

The macro REGDATA (lines 6-43) calculates the necessary functions of Q needed for the linear regressions. The macro ABREGMAC (lines 44-64) does the linear regressions using the SAS regression procedure SYSREG (SAS Institute, Inc., 1979, p. 403). Lines 65-67 create a data set named DATAA with only sulfate and streamflow data for station 03374100. Line 68 sorts this newly

```
//F3 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
MACRO REGDATA
OPTIONS NOMACROGEN;
PROC MEANS DATA=FILENAME NOPRINT MEAN;BY STATION;VAR P00061;
OUTPUT OUT=MEANQ MEAN=M00061;
PROC SORT DATA=MEANQ;BY STATION;
DATA INPDATA;MERGE MEANQ FILENAME;BY STATION;
*GET LOG OF DEPENDENT VARIABLE;
LNDEPVAR=LOG(DEPVAR);
*GET LOG OF THE FLOW;
L00061=LOG(P00061);
*BETA = THE INTEGER PORTION OF THE BASE 10 LOG OF THE MEAN FLOW;
BETA=INT(LOG10(M00061));
*BN=DIFFERENT VALUES USED IN THE HYPERBOLIC FUNCTIONS;
B1=10**(-2.5-BETA);
B2=B1*10**0.5;
B3=B2*10**0.5;
B4=B3*10**0.5;
B5=B4*10**0.5;
B6=B5*10**0.5;
B7=B6*10**0.5;
B8=B7*10**0.5;
*HYPERBNQ=DIFFERENT HYPERBOLIC FUNCTIONS;
HYPERB1Q=1/(1+B1*P00061);
HYPERB2Q=1/(1+B2*P00061);
HYPERB3Q=1/(1+B3*P00061);
HYPERB4Q=1/(1+B4*P00061);
HYPERB5Q=1/(1+B5*P00061);
HYPERB6Q=1/(1+B6*P00061);
HYPERB7Q=1/(1+B7*P00061);
HYPERB8Q=1/(1+B8*P00061);
*INVQ= THE INVERSE OF THE FLOW;
INVQ=1/P00061;
*Q2=SQUARE OF THE FLOW;
Q2=P00061**2;
*LQ2=SQUARE OF THE LOG OF THE FLOW;
LQ2=L00061**2;
PROC SORT DATA=INPDATA;BY STATION;
%
MACRO ABREGMAC
OPTIONS NOMACROGEN;
PROC SYSREG DATA=INPDATA;BY STATION;
LINEAR:MODEL DEPVAR=P00061;
LOGLIN:MODEL DEPVAR=L00061;
```

Figure 3.--Example input and output of a SAS procedure to do flow adjustment regressions-abbreviated version.

```
HYPER2:MODEL DEPVAR=HYPERB2Q;
HYPER3:MODEL DEPVAR=HYPERB3Q;
HYPER4:MODEL DEPVAR=HYPERB4Q;
HYPER5:MODEL DEPVAR=HYPERB5Q;
HYPER6:MODEL DEPVAR=HYPERB6Q;
HYPER7:MODEL DEPVAR=HYPERB7Q;
HYPER8:MODEL DEPVAR=HYPERB8Q;
INVERS:MODEL DEPVAR=INVQ;
QUAD:  MODEL DEPVAR=P00061 Q2;
LOGLOG:MODEL LNDEPVAR=L00061;
LOGQAD:MODEL LNDEPVAR=L00061 LQ2;
PROC PLOT DATA=INPDATA;BY STATION;
PLOT DEPVAR*P00061;
PLOT LNDEPVAR*L00061;
%
DATA DATAA;SET TREND.MONTHLY;
IF STATION=' 03374100       ';
KEEP STATION SNAME P00945 P00061;
PROC SORT;BY STATION;
MACRO FILENAME DATAA%
MACRO DEPVAR P00945%
MACRO LNDEPVAR L00945%
REGDATA
ABREGMAC
/*
//
```

STATION IDENTIFICATION NUMBER=03374100

MODEL: LINEAR	SSE	9981.852	F RATIO	71.46	
	DFE	96	PROB>F	0.0001	
DEP VAR: P00945	MSE	103.977629	R-SQUARE	0.4267	

SULFATE DISSOLVED (MG/L AS SO4)

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	60.288824	1.418811	42.4925	0.0001	
P00061	1	-0.0005675	.000006713385	-8.4533	0.0001	STREAMFLOW, INSTANTANEOUS (CFS)

MODEL: LOGLIN	SSE	7440.682	F RATIO	128.65	
	DFE	96	PROB>F	0.0001	
DEP VAR: P00945	MSE	77.507101	R-SQUARE	0.5727	

SULFATE DISSOLVED (MG/L AS SO4)

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	136.455261	7.495376	18.2053	0.0001	
L00061	1	-9.317384	0.821471	-11.3423	0.0001	

MODEL: HYPER1	SSE	9943.372	F RATIO	72.11	
	DFE	96	PROB>F	0.0001	
DEP VAR: P00945	MSE	103.576787	R-SQUARE	0.4289	

SULFATE DISSOLVED (MG/L AS SO4)

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-1771.47	214.748414	-8.2491	0.0001	
HYPERB1Q	1	1831.854	215.727945	8.4915	0.0001	

MODEL: HYPER2	SSE	9864.229	F RATIO	73.45	
	DFE	96	PROB>F	0.0001	
DEP VAR: P00945	MSE	102.752388	R-SQUARE	0.4335	

SULFATE DISSOLVED (MG/L AS SO4)

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	-544.097240	69.564112	-7.8215	0.0001	
HYPERB2Q	1	604.667673	70.551800	8.5705	0.0001	

```
MODEL:    HYPER3                 SSE     9645.029      F RATIO        77.31
                                 DFE           96      PROB>F        0.0001
DEP VAR: P00945                  MSE   100.469055      R-SQUARE      0.4461
         SULFATE DISSOLVED (MG/L AS SO4)

                           PARAMETER      STANDARD                                 VARIABLE
VARIABLE           DF       ESTIMATE         ERROR       T RATIO    PROB>|T|         LABEL

INTERCEPT           1    -155.241972     23.597084       -6.5789      0.0001
HYPERB3Q            1     216.378395     24.609842        8.7924      0.0001
-------------------------------------------------------------------------------
MODEL:    HYPER4                 SSE     9156.534      F RATIO        86.55
                                 DFE           96      PROB>F        0.0001
DEP VAR: P00945                  MSE    95.380563      R-SQUARE      0.4741
         SULFATE DISSOLVED (MG/L AS SO4)

                           PARAMETER      STANDARD                                 VARIABLE
VARIABLE           DF       ESTIMATE         ERROR       T RATIO    PROB>|T|         LABEL

INTERCEPT           1     -30.563720      8.933707       -3.4212      0.0009
HYPERB4Q            1      93.218221     10.019921        9.3033      0.0001
-------------------------------------------------------------------------------
MODEL:    HYPER5                 SSE     8437.529      F RATIO       102.11
                                 DFE           96      PROB>F        0.0001
DEP VAR: P00945                  MSE    87.890932      R-SQUARE      0.5154
         SULFATE DISSOLVED (MG/L AS SO4)

                           PARAMETER      STANDARD                                 VARIABLE
VARIABLE           DF       ESTIMATE         ERROR       T RATIO    PROB>|T|         LABEL

INTERCEPT           1      11.803547      4.093053        2.8838      0.0049
HYPERB5Q            1      54.320665      5.375720       10.1048      0.0001
-------------------------------------------------------------------------------
MODEL:    HYPER6                 SSE     7884.815      F RATIO       115.99
                                 DFE           96      PROB>F        0.0001
DEP VAR: P00945                  MSE    82.133486      R-SQUARE      0.5472
         SULFATE DISSOLVED (MG/L AS SO4)

                           PARAMETER      STANDARD                                 VARIABLE
VARIABLE           DF       ESTIMATE         ERROR       T RATIO    PROB>|T|         LABEL

INTERCEPT           1      28.670546      2.355140       12.1736      0.0001
HYPERB6Q            1      44.303562      4.113583       10.7701      0.0001
```

STATION IDENTIFICATION NUMBER=03374100

MODEL: HYPER7	SSE 7889.606	F RATIO	115.87
	DFE 96	PROB>F	0.0001
DEP VAR: P00945	MSE 82.183394	R-SQUARE	0.5469
SULFATE DISSOLVED (MG/L AS SO4)			

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	36.901791	1.678295	21.9877	0.0001	
HYPER87Q	1	49.306521	4.580650	10.7641	0.0001	

MODEL: HYPER8	SSE 8576.029	F RATIO	98.91
	DFE 96	PROB>F	0.0001
DEP VAR: P00945	MSE 89.333636	R-SQUARE	0.5075
SULFATE DISSOLVED (MG/L AS SO4)			

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	41.562157	1.421870	29.2306	0.0001	
HYPER88Q	1	72.286630	7.268460	9.9452	0.0001	

MODEL: INVERS	SSE 10538.97	F RATIO	62.61
	DFE 96	PROB>F	0.0001
DEP VAR: P00945	MSE 109.780918	R-SQUARE	0.3947
SULFATE DISSOLVED (MG/L AS SO4)			

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	46.117272	1.296412	35.5730	0.0001	
INVQ	1	27804.92	3514.116	7.9124	0.0001	

MODEL: QUAD	SSE 8825.482	F RATIO	46.21
	DFE 95	PROB>F	0.0001
DEP VAR: P00945	MSE 92.899808	R-SQUARE	0.4931
SULFATE DISSOLVED (MG/L AS SO4)			

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	64.405530	1.777655	36.2306	0.0001	
P00061	1	-0.00117688	0.000184011	-6.3957	0.0001	STREAMFLOW, INSTANTANEOUS (CFS)
Q2	1	1.06125E-08	3.00799E-09	3.5281	0.0006	

STATION IDENTIFICATION NUMBER=03374100

MODEL: LOGLOG	SSE 3.577098	F RATIO	104.24
	DFE 96	PROB>F	0.0001
DEP VAR: L00945	MSE 0.037261	R-SQUARE	0.5206

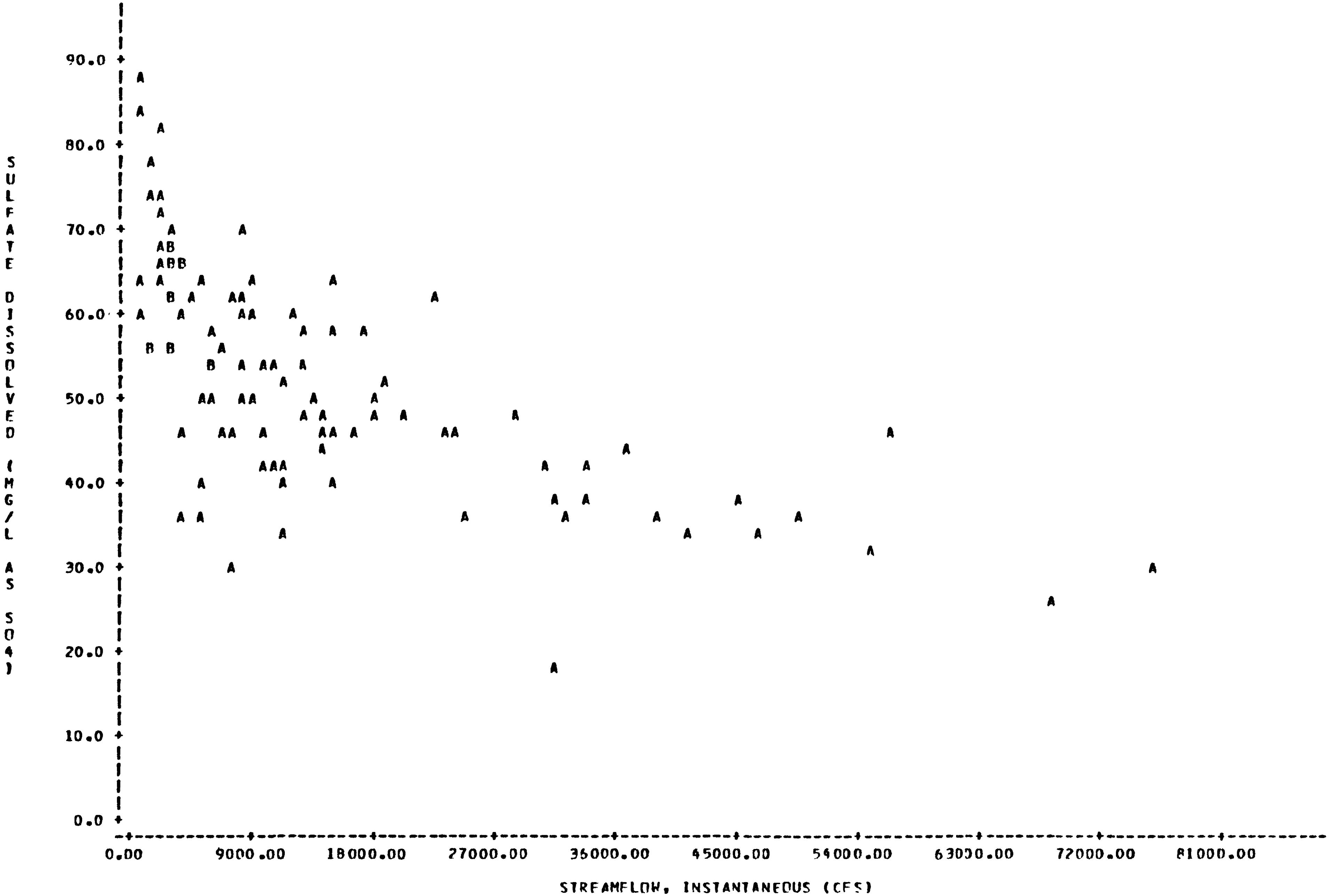
VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	5.582358	0.164343	33.9676	0.0001	
L00061	1	-0.183890	0.018012	-10.2096	0.0001	

MODEL: LOGQAD	SSE 3.530233	F RATIO	52.89
	DFE 95	PROB>F	0.0001
DEP VAR: L00945	MSE 0.037160	R-SQUARE	0.5268

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T RATIO	PROB>|T|	VARIABLE LABEL
INTERCEPT	1	4.407203	1.059222	4.1608	0.0001	
L00061	1	0.082749	0.238112	0.3475	0.7290	
L02	1	-0.014901	0.013269	-1.1230	0.2643	

STATION IDENTIFICATION NUMBER=03374100

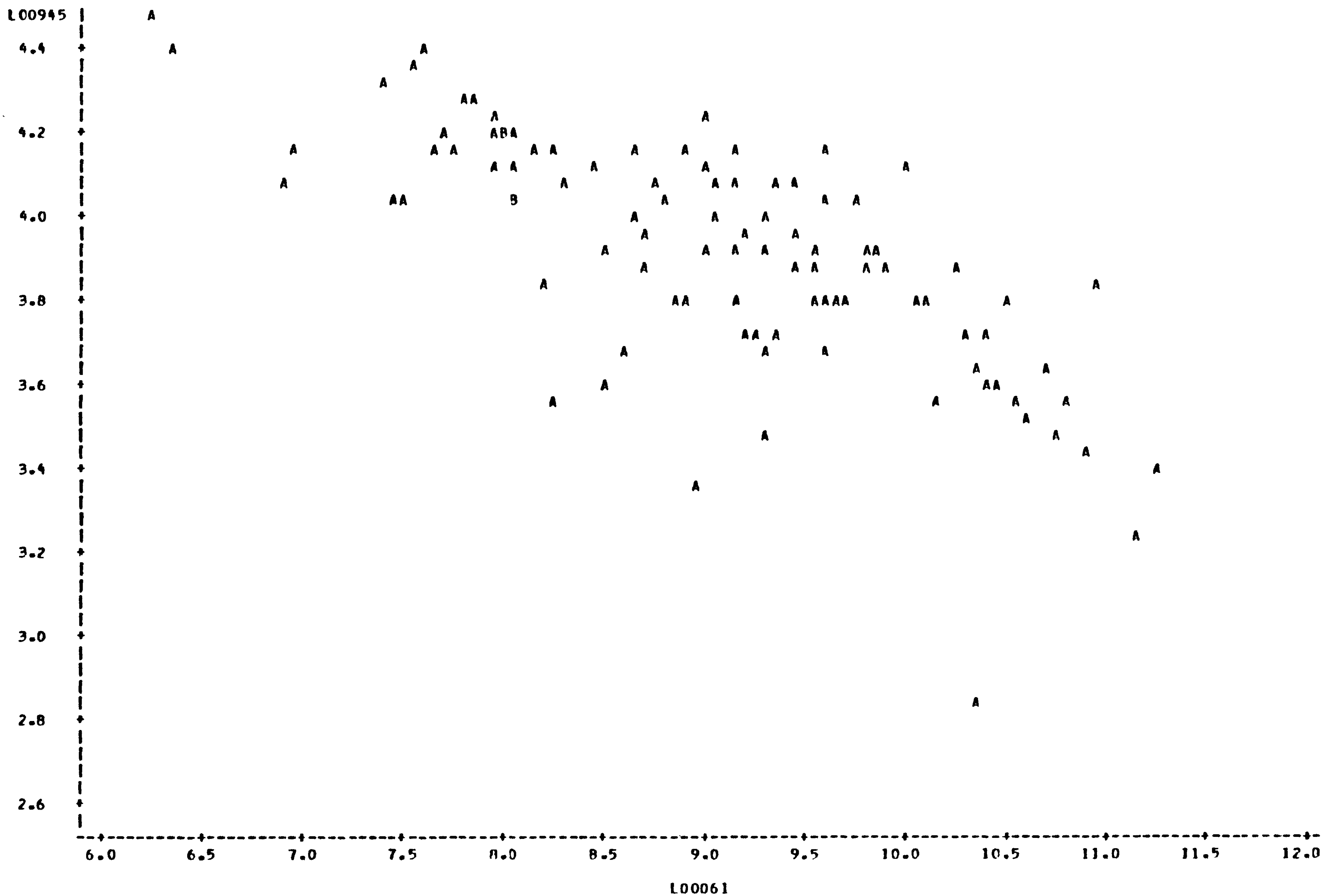
PLOT OF P00945*P00061   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 5 OBS HAD MISSING VALUES

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   15:41 THURSDAY, AUGUST 18, 1983   6
STATION IDENTIFICATION NUMBER=03374100

PLOT OF L00945*L00061   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 5 OBS HAD MISSING VALUES

created data set by station and SNAME (sorting by station and SNAME are required by REGDATA and ABREGMAC). The macros REGDATA and ABREGMAC are written with dummy names for the data set and dependent variables. Lines 69-71 substitute the desired data set and variable names into these macros. In this example, the flow-adjustment procedure will be done on the data set DATAA for the variable P00945 (sulfate). The macro LNDEPVAR inserts a name for the natural logarithm of the dependent variable. The argument of this macro may be any unused variable name the user chooses for the log of the variable given in line 70. The procedures could have been done on any number of variables by repeating the statements in lines 69-73 and making changes in the data set and dependent variables. (For more information on the use of a macro to substitute a variable for dummy names, see SAS Institute, Inc., 1979, p. 12.) The statement in line 72 executes the macro REGDATA and the statement in line 73 executes the macro ABREGMAC.

SYSREG (line 46) outputs for each of the models: (1) the error sum of squares (SSE); (2) the error degrees of freedom (DFE); (3) the error mean square (MSE); (4) the model F ratio and Prob >F (test and significance probability that all parameters except the intercept are zero); (5) the model R-square (R^2); and (6) the degrees of freedom, estimate, standard error, and T ratio and Prob >|T| for parameters in the model. The R-square and Prob >|T| for the function of Q in the model correspond to the R^2 and p value in the Smith and others (1982) flow-adjustment procedure. Plots of streamflow versus sulfate, in both arithmetic and log form, are shown in figure 3.

Figure 4 shows a listing and output of a sample SAS job that does the flow-adjustment procedure using the macro REGMAC (line 44). In addition to providing R^2 and p for each model, the program provides several diagnostic

```
//F4 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
MACRO REGDATA
OPTIONS NOMACROGEN;
PROC MEANS DATA=FILENAME NOPRINT MEAN;BY STATION;VAR P00061;
OUTPUT OUT=MEANQ MEAN=M00061;
PROC SORT DATA=MEANQ;BY STATION;
DATA INPDATA;MERGE MEANQ FILENAME;BY STATION;
*GET LOG OF DEPENDENT VARIABLE;
LNDEPVAR=LOG(DEPVAR);
*GET LOG OF THE FLOW;
L00061=LOG(P00061);
*BETA = THE INTEGER PORTION OF THE BASE 10 LOG OF THE MEAN FLOW;
BETA=INT(LOG10(M00061));
*BN=DIFFERENT VALUES USED IN THE HYPERBOLIC FUNCTIONS;
B1=10**(-2.5-BETA);
B2=B1*10**0.5;
B3=B2*10**0.5;
B4=B3*10**0.5;
B5=B4*10**0.5;
B6=B5*10**0.5;
B7=B6*10**0.5;
B8=B7*10**0.5;
*HYPERBNQ=DIFFERENT HYPERBOLIC FUNCTIONS;
HYPERB1Q=1/(1+B1*P00061);
HYPERB2Q=1/(1+B2*P00061);
HYPERB3Q=1/(1+B3*P00061);
HYPERB4Q=1/(1+B4*P00061);
HYPERB5Q=1/(1+B5*P00061);
HYPERB6Q=1/(1+B6*P00061);
HYPERB7Q=1/(1+B7*P00061);
HYPERB8Q=1/(1+B8*P00061);
*INVQ= THE INVERSE OF THE FLOW;
INVQ=1/P00061;
*Q2=SQUARE OF THE FLOW;
Q2=P00061**2;
*LQ2=SQUARE OF THE LOG OF THE FLOW;
LQ2=L00061**2;
PROC SORT DATA=INPDATA;BY STATION;
%
MACRO REGMAC
OPTIONS NOMACROGEN NOOVP;
PROC REG DATA=INPDATA;BY STATION;
ID P00061;
TITLE OUTPUT FROM MODLABEL: Y=FLOW;
MODLABEL:MODEL Y = FLOW / R;
```

Figure 4.--Example input and output of a SAS procedure to do flow adjustment regressions-extended version.

```
OUTPUT OUT=REGDAT1  P=P R=R STUDENT=SR COOKD=COOKD;
PROC SORT DATA=REGDAT1 ;BY STATION;
PROC PLOT DATA=REGDAT1 ;BY STATION;
PLOT P*L00061='P' Y*L00061='0' / OVERLAY;
PLOT R*P / VREF=0;
PROC UNIVARIATE DATA=REGDAT1 PLOT NORMAL;VAR R;
%
DATA DATAA;SET TREND.MONTHLY;
IF STATION=' 03374100        ';
KEEP STATION SNAME P00945 P00061;
PROC SORT;BY STATION;
MACRO FILENAME DATAA%
MACRO DEPVAR P00945%
MACRO LNDEPVAR L00945%
REGDATA
MACRO MODLABEL LINEAR%
MACRO Y DEPVAR%
MACRO FLOW P00061%
REGMAC
MACRO MODLABEL LOGLIN%
MACRO Y DEPVAR%
MACRO FLOW L00061%
REGMAC
MACRO MODLABEL LOGLOG%
MACRO Y LNDEPVAR%
MACRO FLOW L00061%
REGMAC
/*
//
```

MODEL: LINEAR
DEP VARIABLE: P00945 SULFATE DISSOLVED (MG/L AS SO4)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	1	7429.983	7429.983	71.459	0.0001
ERROR	96	9981.852	103.978		
C TOTAL	97	17411.835			

ROOT MSE	10.196942	R-SQUARE	0.4267
DEP MEAN	52.040815	ADJ R-SQ	0.4207
C.V.	19.59412		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|	VARIABLE LABEL
INTERCEP	1	60.288824	1.418811	42.492	0.0001	INTERCEPT
P00061	1	-0.0005675	.00006713385	-8.453	0.0001	STREAMFLOW, INSTANTANEOUS (CFS)

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0	1 2	COOK'S D
1	12799.97	58.000	53.025	1.037	4.975	10.144	0.490			0.001
2	55899.89	46.000	28.566	2.962	17.434	9.757	1.787		***	0.147
3	28399.94	49.000	44.172	1.388	4.828	10.102	0.478			0.002
4	9299.98	59.000	55.011	1.088	3.989	10.139	0.393			0.001
5	15299.97	45.000	51.606	1.031	-6.606	10.145	-0.651	*		0.002
6	31399.93	17.000	42.469	1.531	-25.469	10.081	-2.526	*****		0.074
7	5989.99	53.000	56.890	1.179	-3.890	10.129	-0.384			0.001
8	3029.99	68.000	58.569	1.287	9.431	10.115	0.932		*	0.007
9	2250.00	67.000	59.012	1.319	7.988	10.111	0.790		*	0.005
10	2600.00	73.000	58.813	1.305	14.187	10.113	1.403		**	0.016
11	6360.00	58.000	56.680	1.167	1.320	10.130	0.130			0.000
12	14799.96	57.000	51.890	1.030	5.110	10.145	0.504		*	0.001
13	22500.00	61.000	47.520	1.161	13.480	10.131	1.331		**	0.012
14	33599.93	42.000	41.221	1.643	0.779120	10.064	0.077			0.000
15	37000.00	44.000	39.291	1.826	4.709	10.032	0.469			0.004
16	14100.00	48.000	52.287	1.030	-4.287	10.145	-0.423			0.001
17	11700.00	59.000	53.649	1.047	5.351	10.143	0.528		*	0.001
18	4740.00	61.000	57.599	1.222	3.401	10.123	0.336			0.001
19	7979.98	.	55.760	1.120	.	.	.			.
20	11300.00	41.000	53.876	1.053	-12.876	10.142	-1.270	**		0.009
21	14299.96	45.000	52.174	1.030	-7.174	10.145	-0.707	*		0.003
22	23199.95	45.000	47.123	1.183	-2.123	10.128	-0.210			0.000
23	20299.95	48.000	48.769	1.100	-.768611	10.137	-0.076			0.000
24	38999.92	35.000	38.156	1.939	-3.156	10.011	-0.315			0.002
25	18099.96	50.000	50.017	1.058	-.017103	10.142	-0.002			0.000
26	45299.91	38.000	34.581	2.308	3.419	9.932	0.344			0.003
27	14899.97	64.000	51.833	1.030	12.167	10.145	1.199		**	0.007
28	9709.98	53.000	54.778	1.080	-1.778	10.140	-0.175			0.000
29	6549.99	56.000	56.572	1.161	-.571709	10.131	-0.056			0.000
30	980.00	60.000	59.733	1.374	0.267325	10.104	0.026			0.000

OUTPUT FROM LINEAR: P00945= P00061
STATION IDENTIFICATION NUMBER=03374100

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0 1 2	COOK'S D
31	2320.00	65.000	58.972	1.317	6.028	10.112	0.596	|*	0.003
32	1860.00	77.000	59.233	1.336	17.767	10.109	1.758	|***	0.027
33	2930.00	66.000	58.626	1.291	7.374	10.115	0.729	|*	0.004
34	8379.98	60.000	55.533	1.110	4.467	10.136	0.441	|	0.001
35	11099.97	51.000	53.990	1.056	-2.990	10.142	-0.295	|	0.000
36	46600.00	33.000	33.843	2.386	-.843347	9.914	-0.085	|	0.000
37	46599.91	.	33.843	2.386	.	.	.		.
38	12699.97	53.000	53.082	1.037	-.081596	10.144	-0.008	|	0.000
39	3759.99	65.000	58.155	1.259	6.845	10.119	0.676	|*	0.004
40	4899.99	50.000	57.508	1.216	-7.508	10.124	-0.742	*|	0.004
41	3709.99	46.000	58.183	1.261	-12.183	10.119	-1.204	**|	0.011
42	4099.99	60.000	57.962	1.246	2.038	10.121	0.201	|	0.000
43	1040.00	63.000	59.699	1.372	3.301	10.104	0.327	|	0.001
44	2790.00	70.000	58.706	1.297	11.294	10.114	1.117	|**	0.010
45	1980.00	82.000	59.165	1.331	22.835	10.110	2.259	|****	0.044
46	514.00	88.000	59.997	1.395	28.003	10.101	2.772	|*****	0.073
47	574.00	83.000	59.963	1.393	23.037	10.101	2.281	|****	0.049
48	25000.00	35.000	46.101	1.247	-11.101	10.120	-1.097	**|	0.009
49	31200.00	38.000	42.583	1.521	-4.583	10.083	-0.455	|	0.002
50	8090.00	50.000	55.698	1.117	-5.698	10.136	-0.562	*|	0.002
51	3100.00	56.000	58.530	1.285	-2.530	10.116	-0.250	|	0.001
52	3099.99	.	58.530	1.285	.	.	.		.
53	3440.00	65.000	58.337	1.271	6.663	10.117	0.659	|*	0.003
54	1680.00	56.000	59.335	1.344	-3.335	10.108	-0.330	|	0.001
55	1810.00	56.000	59.262	1.338	-3.262	10.109	-0.323	|	0.001
56	9860.00	42.000	54.693	1.077	-12.693	10.140	-1.252	**|	0.009
57	2840.00	66.000	58.677	1.295	7.323	10.114	0.724	|*	0.004
58	33800.00	37.000	41.107	1.653	-4.107	10.062	-0.408	|	0.002
59	10700.00	54.000	54.217	1.062	-.216579	10.142	-0.021	|	0.000
60	5620.00	64.000	57.099	1.191	6.901	10.127	0.681	|*	0.003
61	9220.00	63.000	55.056	1.090	7.944	10.139	0.783	|*	0.004
62	19000.00	51.000	49.506	1.073	1.494	10.140	0.147	|	0.000
63	17000.00	57.000	50.641	1.043	6.359	10.143	0.627	|*	0.002
64	7320.00	63.000	56.135	1.138	6.865	10.133	0.677	|*	0.003
65	8280.00	69.000	55.590	1.112	13.410	10.136	1.323	|**	0.011
66	11000.00	33.000	54.046	1.057	-21.046	10.142	-2.075	****|	0.023
67	3770.00	35.000	58.149	1.258	-23.149	10.119	-2.288	****|	0.040
68	3180.00	62.000	58.484	1.281	3.516	10.116	0.348	|	0.001
69	31600.00	.	42.356	1.541	.	.	.		.
70	32200.00	36.000	42.015	1.571	-6.015	10.075	-0.597	*|	0.004
71	8600.00	54.000	55.408	1.104	-1.408	10.137	-0.139	|	0.000
72	76000.00	30.000	17.159	4.253	12.841	9.268	1.386	|**	0.202
73	30400.00	41.000	43.037	1.482	-2.037	10.089	-0.202	|	0.000
74	13700.00	50.000	52.514	1.032	-2.514	10.145	-0.248	|	0.000
75	9350.00	50.000	54.983	1.087	-4.983	10.139	-0.491	|	0.001
76	6820.00	45.000	56.418	1.153	-11.418	10.132	-1.127	**|	0.008
77	6000.00	49.000	56.884	1.179	-7.884	10.129	-0.778	*|	0.004
78	68400.00	26.000	21.472	3.760	4.528	9.478	0.478	|	0.018
79	41100.00	34.000	36.965	2.060	-2.965	9.987	-0.297	|	0.002
80	7410.00	45.000	56.084	1.136	-11.084	10.134	-1.094	**|	0.008
81	10200.00	42.000	54.500	1.070	-12.500	10.141	-1.233	**|	0.008
82	54900.00	31.000	29.133	2.899	1.867	9.776	0.191	|	0.002
83	5440.00	40.000	57.202	1.197	-17.202	10.126	-1.699	***|	0.020

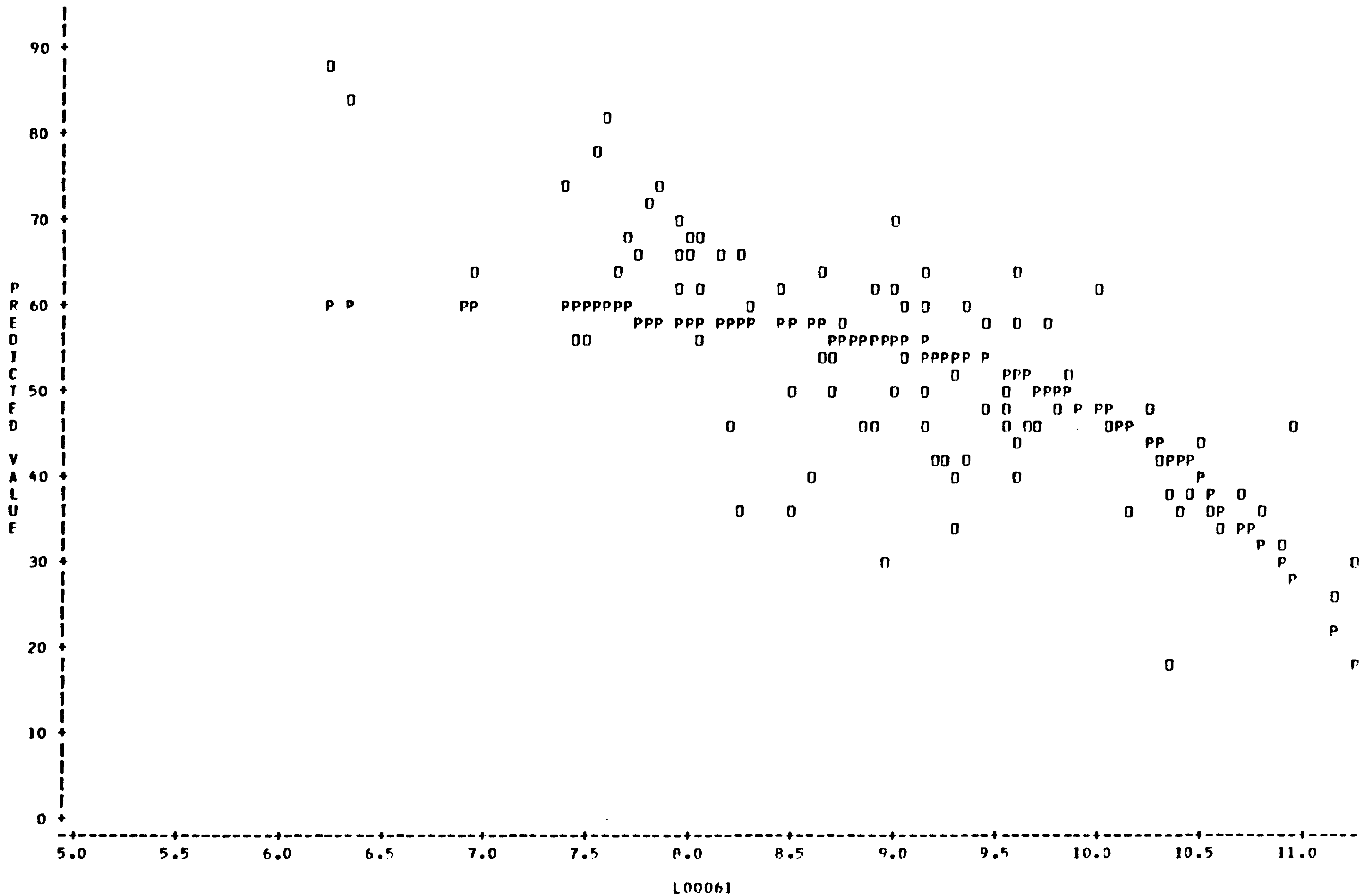
STATION IDENTIFICATION NUMBER=03374100

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0 1 2	COOK'S D
84	18100.00	48.000	50.017	1.058	-2.017	10.142	-0.199					0.000
85	49800.00	35.000	32.027	2.582	2.973	9.865	0.301					0.003
86	16500.00	45.000	50.925	1.038	-5.925	10.144	-0.584		*			0.002
87	14800.00	39.000	51.890	1.030	-12.890	10.145	-1.271		**			0.008
88	9500.00	45.000	54.898	1.084	-9.898	10.139	-0.976		*			0.005
89	6000.00	.	56.884	1.179	.	.	.		.
90	7580.00	29.000	55.987	1.131	-26.987	10.134	-2.663		*****			0.044
91	3200.00	56.000	58.473	1.281	-2.473	10.116	-0.244					0.000
92	2500.00	71.000	58.870	1.309	12.130	10.113	1.199			**		0.012
93	2050.00	64.000	59.125	1.328	4.875	10.110	0.482					0.002
94	3170.00	67.000	58.490	1.282	8.510	10.116	0.841			*		0.006
95	1650.00	74.000	59.352	1.345	14.648	10.108	1.449			**		0.019
96	14600.00	44.000	52.003	1.030	-8.003	10.145	-0.789		*			0.003
97	5780.00	54.000	57.009	1.186	-3.009	10.128	-0.297					0.001
98	12800.00	48.000	53.025	1.037	-5.025	10.144	-0.495					0.001
99	24100.00	45.000	46.612	1.214	-1.612	10.124	-0.159					0.000
100	11200.00	40.000	53.933	1.054	-13.933	10.142	-1.374		**			0.010
101	8160.00	62.000	55.658	1.115	6.342	10.136	0.626			*		0.002
102	4880.00	36.000	57.519	1.217	-21.519	10.124	-2.126		****			0.033
103	2860.00	61.000	58.666	1.294	2.334	10.114	0.231					0.000

SUM OF RESIDUALS 1.01608E-12
SUM OF SQUARED RESIDUALS 9981.852

OUTPUT FROM LINEAR: P00945= P00061
STATION IDENTIFICATION NUMBER=03374100

PLOT OF P*L00061 SYMBOL USED IS P
PLOT OF P00945*L00061 SYMBOL USED IS O

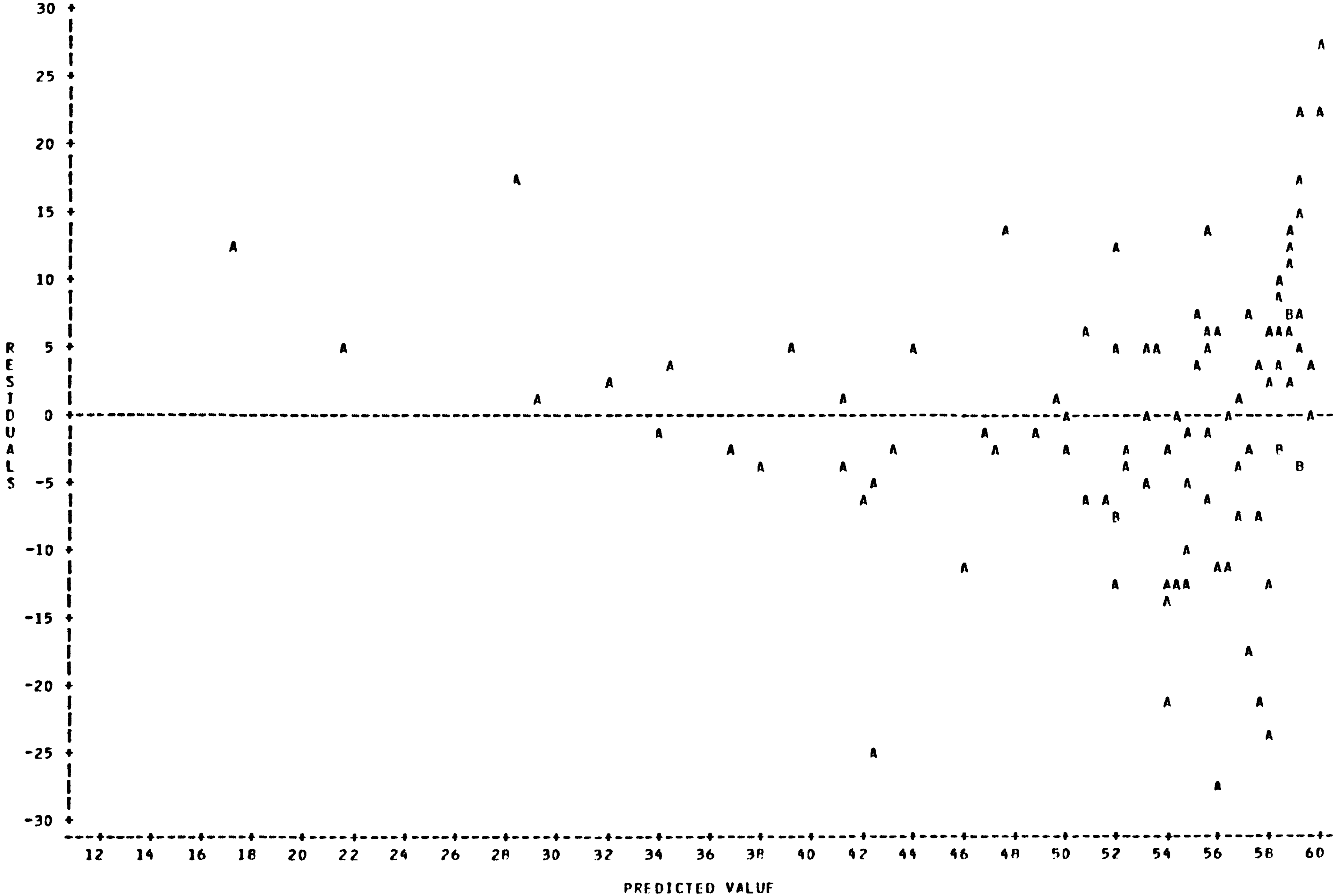


NOTE: 5 OBS HAD MISSING VALUES 47 OBS HIDDEN

OUTPUT FROM LINEAR: P00345= P00061
STATION IDENTIFICATION NUMBER=03374100

16:06 THURSDAY, AUGUST 18, 1983 5

PLOT OF R*P LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 5 OBS HAD MISSING VALUES

```
                                   UNIVARIATE

VARIABLE=R         RESIDUALS

              MOMENTS                                  QUANTILES(DEF=4)                          EXTREMES

N               98   SUM WGTS          98      100% MAX   28.0029      99%    28.0029       LOWEST     HIGHEST
MEAN     1.037E-14   SUM        1.016E-12       75% Q3    6.34615      95%    17.4509     -26.9872     17.4343
STD DEV    10.1442   VARIANCE     102.906       50% MED -0.149088      90%     12.898     -25.4694     17.7667
SKEWNESS -0.130921   KURTOSIS    0.660651       25% Q1   -5.19306      10%   -12.7116     -23.1493     22.8348
USS        9981.85   CSS          9981.85        0% MIN  -26.9872       5%     -21.07     -21.5194     23.0369
CV       9.784E+16   STD MEAN     1.02472                               1%   -26.9872     -21.0463     28.0029
T:MEAN=0 1.012E-14   PROB>|T|           1      RANGE       54.99
SGN RANK      69.5   PROB>|S|    0.806838      Q3-Q1     11.5392
NUM ^= 0        98                             MODE     -26.9872
D:NORMAL 0.0652829   PROB>D         >0.15

                                               MISSING VALUE        .
                                                       COUNT        5
                                                % COUNT/NOBS     4.85

    STEM LEAF                         #     BOXPLOT                         NORMAL PROBABILITY PLOT
       2 8                            1        0           27.5+                                                  *
       2 33                           2        |               |                                             * *
       1 578                          3        |               |                                          +++++
       1 1223334                      7        |               |                                     *******
       0 55555556667777778899        20     +-----+            |                                *******
       0 1112223333444               13     | + |              |                           ++++++
      -0 4443333333322222211110000   25     *-----*            |                     *******
      -0 88877666555                 11     +-----+            |                 ******
      -1 4333321110                  10        |               |            ******
      -1 7                            1        |               |        ++++**
      -2 321                          3        0               |     + ***
      -2 75                           2        0          -27.5+*    *
         ----+----+----+----+----+                             +----+----+----+----+----+----+----+----+----+----+
         MULTIPLY STEM.LEAF BY 10**+01                              -2        -1        +0        +1        +2
```

MODEL: LOGLIN
DEP VARIABLE: P00945 SULFATE DISSOLVED (MG/L AS SO4)

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB>F
MODEL	1	9971.154	9971.154	128.648	0.0001
ERROR	96	7440.682	77.507101		
C TOTAL	97	17411.835			

ROOT MSE	8.803812	R-SQUARE	0.5727
DEP MEAN	52.040815	ADJ R-SQ	0.5682
C.V.	16.91713		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR H0: PARAMETER=0	PROB > |T|	VARIABLE LABEL
INTERCEP	1	136.455	7.495376	18.205	0.0001	INTERCEPT
L00061	1	-9.317384	0.821471	-11.342	0.0001	

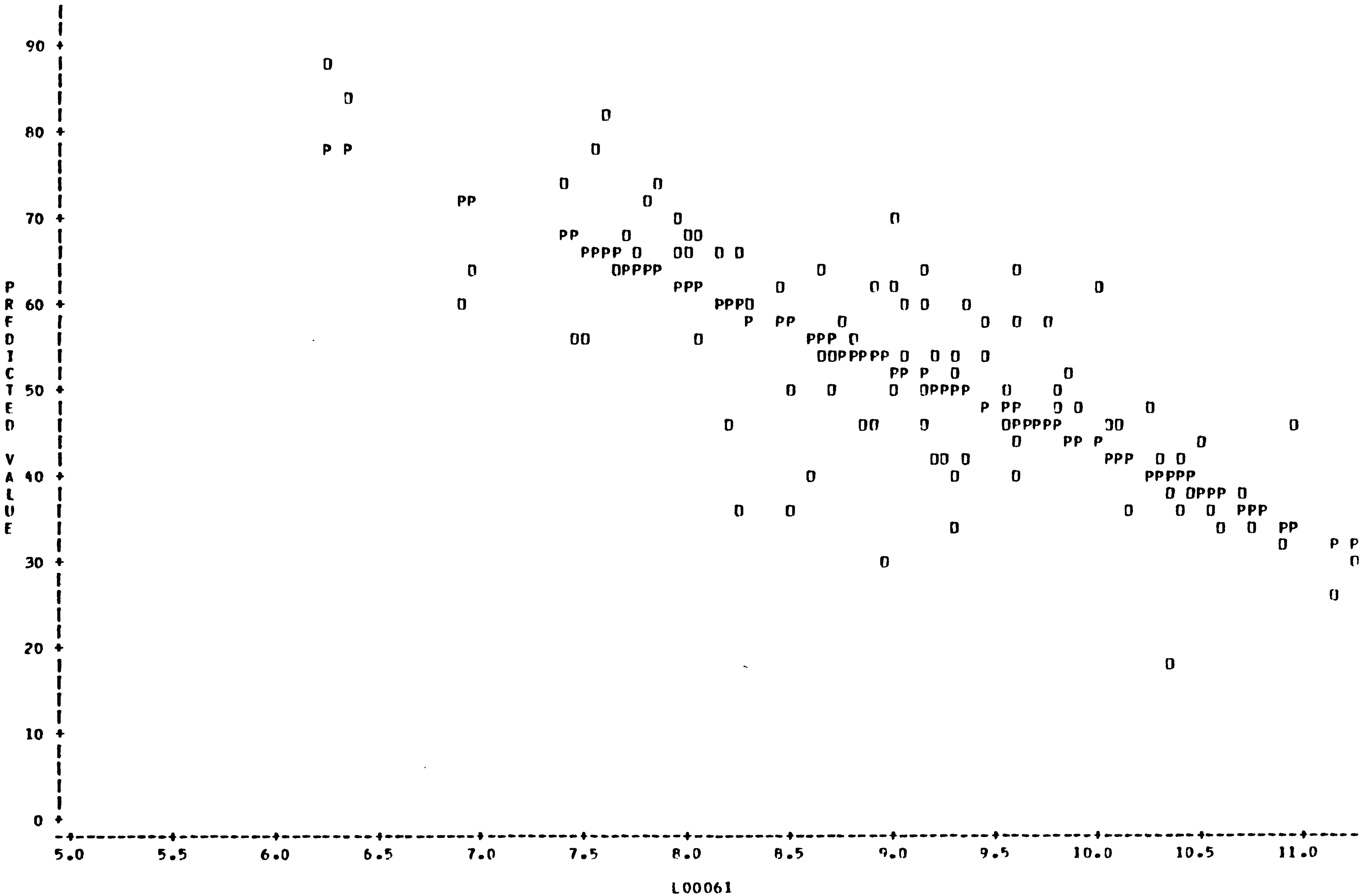
OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0 1 2	COOK'S D
1	12799.97	58.000	48.339	0.947318	9.661	8.753	1.104	| |** |	0.007
2	55899.89	46.000	34.604	1.776	11.396	8.623	1.322	| |** |	0.037
3	28399.94	49.000	40.913	1.324	8.086	8.704	0.929	| |* |	0.010
4	9299.98	59.000	51.315	0.891618	7.685	8.759	0.877	| |* |	0.004
5	15299.97	45.000	46.677	1.007	-1.677	8.746	-0.192	| | |	0.000
6	31399.93	17.000	39.978	1.386	-22.978	8.694	-2.643	| *****| |	0.089
7	5989.99	53.000	55.414	0.937731	-2.414	8.754	-0.276	| | |	0.000
8	3029.99	68.000	61.764	1.235	6.236	8.717	0.715	| |* |	0.005
9	2250.00	67.000	64.537	1.416	2.463	8.689	0.283	| | |	0.001
10	2600.00	73.000	63.190	1.326	9.810	8.703	1.127	| |** |	0.015
11	6360.00	58.000	54.856	0.923297	3.144	8.755	0.359	| | |	0.001
12	14799.96	57.000	46.986	0.994729	10.014	8.747	1.145	| |** |	0.008
13	22500.00	61.000	43.083	1.189	17.917	8.723	2.054	| |**** |	0.039
14	33599.93	42.000	39.347	1.429	2.653	8.687	0.305	| | |	0.001
15	37000.00	44.000	38.449	1.492	5.551	8.676	0.640	| |* |	0.006
16	14100.00	48.000	47.438	0.977546	0.562374	8.749	0.064	| | |	0.000
17	11700.00	59.000	49.176	0.924488	9.824	8.755	1.122	| |** |	0.007
18	4740.00	61.000	57.595	1.015	3.405	8.745	0.389	| | |	0.001
19	7979.98	.	52.741	0.891462	.	.	.		.
20	11300.00	41.000	49.500	0.917094	-8.500	8.756	-0.971	| *| |	0.005
21	14299.96	45.000	47.306	0.982405	-2.306	8.749	-0.264	| | |	0.000
22	23199.95	45.000	42.798	1.206	2.202	8.721	0.253	| | |	0.001
23	20299.95	48.000	44.042	1.135	3.958	8.730	0.453	| | |	0.002
24	38999.92	35.000	37.958	1.527	-2.958	8.670	-0.341	| | |	0.002
25	18099.96	50.000	45.111	1.079	4.889	8.737	0.560	| |* |	0.002
26	45299.91	38.000	36.563	1.629	1.437	8.652	0.166	| | |	0.000
27	14899.97	64.000	46.923	0.997220	17.077	8.747	1.952	| |*** |	0.025
28	9709.98	53.000	50.913	0.894859	2.087	8.758	0.238	| | |	0.000
29	6549.99	56.000	54.581	0.917093	1.419	8.756	0.162	| | |	0.000
30	980.00	60.000	72.281	1.994	-12.281	8.575	-1.432	| **| |	0.055

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0 1 2	COOK'S D
31	2320.00	65.000	64.252	1.396	0.748136	8.692	0.086	|	0.000
32	1860.00	77.000	66.311	1.541	10.689	8.668	1.233	|**	0.024
33	2930.00	66.000	67.077	1.255	3.923	8.714	0.450	|	0.002
34	8379.98	60.000	52.286	0.889581	7.714	8.759	0.881	|*	0.004
35	11099.97	51.000	49.667	0.913621	1.333	8.756	0.152	|	0.000
36	46600.00	33.000	36.299	1.648	-3.299	8.648	-0.382	|	0.003
37	46599.91	.	36.299	1.648	.	.	.		.
38	12699.97	53.000	48.412	0.945118	4.508	8.753	0.524	|*	0.002
39	3759.99	65.000	59.753	1.119	5.247	8.732	0.601	|*	0.003
40	4899.99	50.000	57.286	1.002	-7.286	8.747	-0.833	*|	0.005
41	3709.99	46.000	59.878	1.126	-13.878	8.731	-1.589	***|	0.021
42	4099.99	60.000	58.946	1.078	1.054	8.738	0.121	|	0.000
43	1040.00	63.000	71.728	1.950	-8.728	8.585	-1.017	**|	0.027
44	2790.00	70.000	62.533	1.283	7.467	8.710	0.857	|*	0.008
45	1980.00	82.000	65.728	1.499	16.272	8.675	1.876	|***	0.053
46	514.00	88.000	78.294	2.480	9.706	8.447	1.149	|**	0.057
47	574.00	83.000	77.265	2.395	5.735	8.472	0.677	|*	0.018
48	25000.00	35.000	42.102	1.249	-7.102	8.715	-0.815	*|	0.007
49	31200.00	38.000	40.037	1.382	-2.037	8.695	-0.234	|	0.001
50	8090.00	50.000	52.614	0.890753	-2.614	8.759	-0.298	|	0.000
51	3100.00	56.000	61.551	1.222	-5.551	8.719	-0.637	*|	0.004
52	3099.99	.	61.551	1.222	.	.	.		.
53	3440.00	65.000	60.582	1.165	4.418	8.726	0.506	|*	0.002
54	1680.00	56.000	67.259	1.610	-11.259	8.655	-1.301	**|	0.029
55	1810.00	56.000	66.565	1.559	-10.565	8.665	-1.219	**|	0.024
56	9860.00	42.000	50.770	0.896346	-8.770	8.758	-1.001	**|	0.005
57	2840.00	66.000	62.368	1.273	3.632	8.711	0.417	|	0.002
58	33800.00	37.000	39.292	1.433	-2.292	8.686	-0.264	|	0.001
59	10700.00	54.000	50.009	0.907189	3.991	8.757	0.456	|	0.001
60	5620.00	64.000	56.008	0.955634	7.992	8.752	0.913	|*	0.005
61	9220.00	63.000	51.396	0.891137	11.604	8.759	1.325	|**	0.009
62	19000.00	51.000	44.659	1.102	6.341	8.735	0.726	|*	0.004
63	17000.00	57.000	45.695	1.051	11.305	8.741	1.293	|**	0.012
64	7320.00	63.000	53.546	0.899163	9.454	8.758	1.080	|**	0.006
65	8280.00	69.000	52.398	0.889875	16.602	8.759	1.896	|***	0.019
66	11000.00	33.000	49.751	0.911947	-16.751	8.756	-1.913	***|	0.020
67	3770.00	35.000	59.728	1.118	-24.728	8.733	-2.832	*****|	0.066
68	3180.00	62.000	61.314	1.208	0.686054	8.721	0.079	|	0.000
69	31600.00	.	39.919	1.390	.	.	.		.
70	32200.00	36.000	39.743	1.402	-3.743	8.691	-0.431	|	0.002
71	8600.00	54.000	52.044	0.889319	1.956	8.759	0.223	|	0.000
72	76000.00	30.000	31.742	1.998	-1.742	8.574	-0.203	|	0.001
73	30400.00	41.000	40.279	1.366	0.720620	8.697	0.083	|	0.000
74	13700.00	50.000	47.706	0.967970	2.294	8.750	0.262	|	0.000
75	9350.00	50.000	51.265	0.891945	-1.265	8.759	-0.144	|	0.000
76	6820.00	45.000	54.205	0.909558	-9.205	8.757	-1.051	**|	0.006
77	6000.00	49.000	55.399	0.937297	-6.399	8.754	-0.731	*|	0.003
78	68400.00	26.000	32.724	1.921	-6.724	8.592	-0.783	*|	0.015
79	41100.00	34.000	37.470	1.562	-3.470	8.664	-0.400	|	0.003
80	7410.00	45.000	53.432	0.897737	-8.432	8.758	-0.963	*|	0.005
81	10200.00	42.000	50.454	0.900250	-8.454	8.758	-0.965	*|	0.005
82	54900.00	31.000	34.772	1.763	-3.772	8.625	-0.437	|	0.004
83	5440.00	40.000	56.311	0.965742	-16.311	8.751	-1.864	***|	0.021

STATION IDENTIFICATION NUMBER=03374100

```
                             PREDICT   STD ERR               STD ERR  STUDENT                        COOK'S
 OBS        ID      ACTUAL     VALUE   PREDICT  RESIDUAL  RESIDUAL  RESIDUAL     -2-1-0 1 2              D

  84  18100.00      48.000    45.111     1.079     2.889     8.737     0.331  |       |       |      0.001
  85  49800.00      35.000    35.681     1.695  -.680576     8.639    -0.079  |       |       |      0.000
  86  16500.00      45.000    45.973     1.038  -.973067     8.742    -0.111  |       |       |      0.000
  87  14800.00      39.000    46.986  0.994730    -7.986     8.747    -0.913  |      *|       |      0.005
  88   9500.00      45.000    51.117  0.893042    -6.117     8.758    -0.698  |      *|       |      0.003
  89   6000.00           .    55.399  0.937297         .         .         .                           .
  90   7580.00      29.000    53.221  0.895381   -24.221     8.758    -2.765  | *****|       |      0.040
  91   3200.00      56.000    61.256     1.205    -5.256     8.721    -0.603  |      *|       |      0.003
  92   2500.00      71.000    63.556     1.350     7.444     8.700     0.856  |       |*      |      0.009
  93   2050.00      64.000    65.405     1.476    -1.405     8.679    -0.162  |       |       |      0.000
  94   3170.00      67.000    61.343     1.210     5.657     8.720     0.649  |       |*      |      0.004
  95   1650.00      74.000    67.427     1.622     6.573     8.653     0.760  |       |*      |      0.010
  96  14600.00      44.000    47.113  0.989773    -3.113     8.748    -0.356  |       |       |      0.001
  97   5780.00      54.000    55.747  0.947437    -1.747     8.753    -0.200  |       |       |      0.000
  98  12800.00      48.000    48.339  0.947319  -.338893     8.753    -0.039  |       |       |      0.000
  99  24100.00      45.000    42.443     1.228     2.557     8.718     0.293  |       |       |      0.001
 100  11200.00      40.000    49.583  0.915338    -9.583     8.756    -1.094  |     **|       |      0.007
 101   8160.00      62.000    52.534  0.890380     9.466     8.759     1.081  |       |**     |      0.006
 102   4880.00      36.000    57.324     1.004   -21.324     8.746    -2.438  |   ****|       |      0.039
 103   2860.00      61.000    62.302     1.269    -1.302     8.712    -0.149  |       |       |      0.000

SUM OF RESIDUALS            2.07905E-11
SUM OF SQUARED RESIDUALS       7440.682
```

PLOT OF P*L00061 SYMBOL USED IS P
PLOT OF P00945*L00061 SYMBOL USED IS O

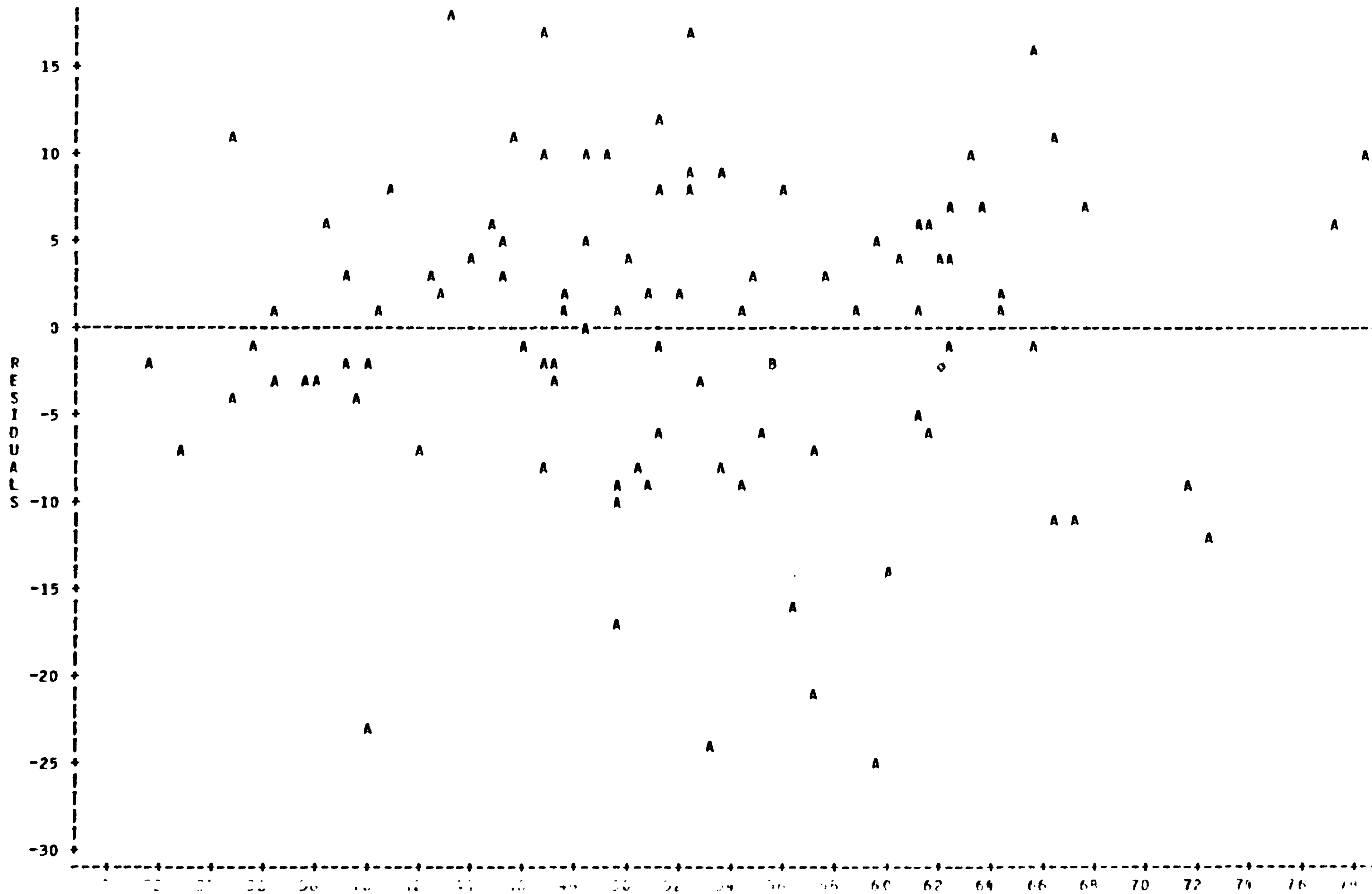


NOTE: 5 OBS HAD MISSING VALUES 44 OBS HIDDEN

OUTPUT FROM LOGLIN: P00945= L00061
STATION IDENTIFICATION NUMBER=03374100

16:06 THURSDAY, AUGUST 18, 1983 11

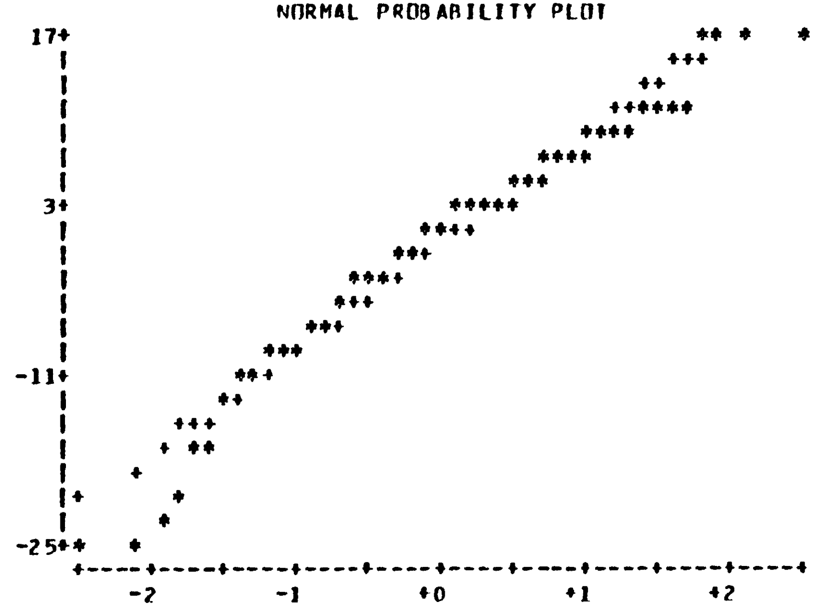
PLOT OF R*P LEGEND: A = 1 OBS, B = 2 OBS, ETC.



UNIVARIATE

```
VARIABLE=R            RESIDUALS

                  MOMENTS                                          QUANTILES(DEF=4)                               EXTREMES

N                 98     SUM WGTS          98       100% MAX   17.9168      99%     17.9169          LOWEST       HIGHEST
MEAN       2.121E-13     SUM        2.079E-11        75% Q3    5.85993      95%     11.8377        -24.7282       11.6044
STD DEV      8.75831     VARIANCE     76.7081        50% MED  0.900886      90%     9.84287        -24.2206       16.2716
SKEWNESS   -0.597348     KURTOSIS    0.624174        25% Q1   -5.32948      10%    -10.6342        -22.9778       16.6024
USS          7440.68     CSS          7440.68         0% MIN  -24.7282       5%    -16.9796        -21.3236       17.0765
CV         4.128E+15     STD MEAN    0.884723                                1%    -24.7282        -16.7509       17.9168
T:MEAN=0   2.398E-13     PROB>|T|           1       RANGE      42.645
SGN RANK       167.5     PROB>|S|       0.554       Q3-Q1     11.1894
NUM ^= 0          98                                MODE     -24.7282
D:NORMAL   0.0782438     PROB>D         0.143

                                                    MISSING VALUE      .
                                                            COUNT      5
                                                     % COUNT/NOBS   4.85

   STEM LEAF                      #       BOXPLOT
     16 3619                      4          |
     14                                      |
     12                                      |
     10 07346                     5          |
      8 01557788                  8          |
      6 2364577                   7          |
      4 004692677                 9       +-----+
      2 012356791469             12       |     |
      0 67771344                  8       *--+--*
     -0 777433073                 9       |     |
     -2 87531064330              11       |     |
     -4 63                        2       +-----+
     -6 31741                     5          |
     -8 62875540                  8          |
    -10 36                        2          |
    -12 93                        2          |
    -14                                      |
    -16 83                        2          |
    -18                                      |
    -20 3                         1          |
    -22 0                         1          0
    -24 72                        2          0
        ----+----+----+----+
```

NORMAL PROBABILITY PLOT



```
                         OUTPUT FROM LOGLOG: L00945= L00061
                      STATION IDENTIFICATION NUMBER=03374100

MODEL: LOGLOG
DEP VARIABLE: L00945

                     SUM OF         MEAN
SOURCE      DF      SQUARES       SQUARE      F VALUE       PROB>F

MODEL        1     3.883952     3.883952      104.235       0.0001
ERROR       96     3.577098     0.037261
C TOTAL     97     7.461050

    ROOT MSE       0.193032     R-SQUARE       0.5206
    DEP MEAN       3.916335     ADJ R-SQ       0.5155
    C.V.           4.928899

                  PARAMETER     STANDARD   T FOR HO:
VARIABLE    DF     ESTIMATE        ERROR  PARAMETER=0    PROB > |T|

INTERCEP     1     5.582358     0.164343       33.968        0.0001
L00061       1    -0.183890     0.018012      -10.219        0.0001

                             PREDICT  STD ERR           STD ERR  STUDENT                      COOK'S
 OBS          ID    ACTUAL     VALUE  PREDICT RESIDUAL RESIDUAL RESIDUAL   -2-1-0 1 2              D

   1 12799.97     4.060     3.843 0.020771 0.217170 0.191911    1.132 |      |**     |      0.008
   2 55899.89     3.829     3.572 0.038941 0.256444 0.189064    1.356 |      |**     |      0.039
   3 28399.94     3.892     3.697 0.029033 0.195096 0.190836    1.022 |      |**     |      0.012
   4  9299.98     4.078     3.902 0.019550 0.175524 0.192040    0.914 |      |*      |      0.004
   5 15299.97     3.807     3.810 0.022085 -.003804 0.191765   -0.020 |      |       |      0.000
   6 31399.93     2.833     3.678 0.030397 -.845044 0.190624   -4.433 |******|       |      0.250
   7  5989.99     3.970     3.983 0.020561 -.012619 0.191934   -0.066 |      |       |      0.000
   8  3029.99     4.220     4.108 0.027084 0.111270 0.191123    0.582 |      |*      |      0.003
   9  2250.00     4.205     4.163 0.031045 0.041724 0.190519    0.219 |      |       |      0.001
  10  2600.00     4.290     4.136 0.029065 0.154078 0.190832    0.807 |      |*      |      0.008
  11  6360.00     4.060     3.972 0.020244 0.088554 0.191968    0.461 |      |       |      0.001
  12 14799.96     4.043     3.817 0.021810 0.226475 0.191796    1.181 |      |**     |      0.009
  13 22500.00     4.111     3.740 0.026078 0.371328 0.191263    1.941 |      |***    |      0.035
  14 33599.93     3.738     3.666 0.031343 0.071865 0.190471    0.377 |      |       |      0.002
  15 37000.00     3.784     3.648 0.032720 0.136111 0.190239    0.715 |      |*      |      0.008
  16 14100.00     3.871     3.825 0.021434 0.045716 0.191839    0.238 |      |       |      0.000
  17 11700.00     4.078     3.860 0.020270 0.217741 0.191965    1.134 |      |**     |      0.007
  18  4740.00     4.111     4.026 0.022260 0.084923 0.191744    0.443 |      |       |      0.001
  19  7979.98         .     3.930 0.019546        .        .        .                             .
  20 11300.00     3.714     3.866 0.020108 -.152621 0.191982   -0.795 |     *|       |      0.003
  21 14299.96     3.807     3.823 0.021540 -.016233 0.191827   -0.085 |      |       |      0.000
  22 23199.95     3.807     3.734 0.026448 0.072750 0.191212    0.380 |      |       |      0.001
  23 20299.95     3.871     3.758 0.024886 0.112733 0.191421    0.589 |      |*      |      0.003
  24 38999.92     3.555     3.638 0.033486 -.083050 0.190106   -0.437 |      |       |      0.003
  25 18099.96     3.912     3.780 0.023658 0.132461 0.191577    0.691 |      |*      |      0.004
  26 45299.91     3.638     3.611 0.035713 0.026724 0.189700    0.141 |      |       |      0.000
  27 14899.97     4.159     3.815 0.021865 0.343546 0.191790    1.791 |      |***    |      0.021
  28  9709.98     3.970     3.894 0.019621 0.076212 0.192032    0.397 |      |       |      0.001
  29  6549.99     4.025     3.966 0.020108 0.058876 0.191932    0.307 |      |       |      0.001
  30   980.00     4.094     4.316 0.043717 -.221461 0.188017   -1.178 |    **|       |      0.038
```

STATION IDENTIFICATION NUMBER=03374100

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0 1 2	COOK'S D
31	2320.00	4.174	4.157	0.030617	0.017052	0.190589	0.089	| | |	0.000
32	1860.00	4.344	4.198	0.033782	0.145832	0.190053	0.767	| |* |	0.009
33	2930.00	4.190	4.114	0.027507	0.075246	0.191052	0.394	| | |	0.002
34	8379.98	4.094	3.921	0.019505	0.173176	0.192044	0.902	| |* |	0.004
35	11099.97	3.932	3.869	0.020032	0.062348	0.191990	0.325	| | |	0.001
36	46600.00	3.497	3.606	0.036141	-.109151	0.189619	-0.576	| *| |	0.006
37	46599.91	.	3.606	0.036141	.	.	.		.
38	12699.97	3.970	3.845	0.020723	0.125576	0.191917	0.654	| |* |	0.002
39	3759.99	4.174	4.069	0.024545	0.105844	0.191465	0.553	| |* |	0.003
40	4899.99	3.912	4.020	0.021978	-.107824	0.191777	-0.562	| *| |	0.002
41	3709.99	3.829	4.071	0.024693	-.242364	0.191446	-1.266	| **| |	0.013
42	4099.99	4.094	4.053	0.023631	0.041720	0.191580	0.218	| | |	0.000
43	1040.00	4.143	4.305	0.042761	-.161744	0.188236	-0.859	| *| |	0.019
44	2790.00	4.248	4.123	0.028135	0.125083	0.190971	0.655	| |* |	0.005
45	1980.00	4.407	4.186	0.032868	0.220243	0.190213	1.158	| |** |	0.020
46	514.00	4.477	4.434	0.054368	0.042861	0.185218	0.231	| | |	0.002
47	574.00	4.419	4.414	0.052516	0.004668	0.185751	0.025	| | |	0.000
48	25000.00	3.555	3.720	0.027375	-.164823	0.191081	-0.863	| *| |	0.008
49	31200.00	3.638	3.679	0.030309	-.041846	0.190638	-0.220	| | |	0.001
50	8090.00	3.912	3.928	0.019531	-.015622	0.192042	-0.081	| | |	0.000
51	3100.00	4.025	4.104	0.026800	-.078685	0.191163	-0.412	| | |	0.002
52	3099.99	.	4.104	0.026800	.	.	.		.
53	3440.00	4.174	4.085	0.025550	0.089488	0.191334	0.468	| | |	0.002
54	1680.00	4.025	4.217	0.035294	-.191338	0.189778	-1.008	| **| |	0.018
55	1810.00	4.025	4.203	0.034183	-.177632	0.189981	-0.935	| *| |	0.014
56	9860.00	3.738	3.891	0.019653	-.153591	0.192029	-0.800	| *| |	0.003
57	2840.00	4.190	4.120	0.027906	0.069509	0.191004	0.364	| | |	0.001
58	33800.00	3.611	3.665	0.031427	-.053795	0.190457	-0.282	| | |	0.001
59	10700.00	3.989	3.876	0.019891	0.112758	0.192005	0.587	| |* |	0.002
60	5620.00	4.159	3.995	0.020953	0.164248	0.191892	0.856	| |* |	0.004
61	9220.00	4.143	3.904	0.019539	0.239533	0.192041	1.247	| |** |	0.008
62	19000.00	3.932	3.771	0.024163	0.161188	0.191514	0.842	| |* |	0.006
63	17000.00	4.043	3.791	0.023037	0.251960	0.191653	1.315	| |** |	0.012
64	7320.00	4.143	3.946	0.019715	0.197096	0.192023	1.026	| |** |	0.006
65	8280.00	4.234	3.923	0.019511	0.310731	0.192044	1.618	| |*** |	0.014
66	11000.00	3.497	3.871	0.019995	-.374634	0.191994	-1.951	| ***| |	0.021
67	3770.00	3.555	4.068	0.024516	-.512707	0.191469	-2.678	| *****| |	0.059
68	3180.00	4.127	4.099	0.026487	0.027783	0.191206	0.145	| | |	0.000
69	31600.00	.	3.677	0.030485	.	.	.		.
70	32200.00	3.584	3.674	0.030746	-.090112	0.190568	-0.473	| | |	0.003
71	8600.00	3.989	3.916	0.019499	0.072581	0.192045	0.378	| | |	0.001
72	76000.00	3.401	3.516	0.043818	-.114514	0.187993	-0.609	| *| |	0.010
73	30400.00	3.714	3.684	0.029953	0.029363	0.190694	0.154	| | |	0.000
74	13700.00	3.912	3.831	0.021224	0.081245	0.191852	0.423	| | |	0.001
75	9350.00	3.912	3.901	0.019557	0.010996	0.192039	0.057	| | |	0.000
76	6820.00	3.807	3.959	0.019943	-.152385	0.191999	-0.794	| *| |	0.003
77	6000.00	3.892	3.983	0.020551	-.090784	0.191935	-0.473	| | |	0.001
78	68400.00	3.258	3.535	0.042127	-.276990	0.188379	-1.470	| **| |	0.054
79	41100.00	3.526	3.629	0.034259	-.102393	0.189968	-0.539	| *| |	0.005
80	7410.00	3.807	3.944	0.019684	-.137128	0.192026	-0.714	| *| |	0.003
81	10200.00	3.738	3.885	0.019739	-.147357	0.192020	-0.767	| *| |	0.003
82	54900.00	3.434	3.576	0.038660	-.141529	0.189121	-0.748	| *| |	0.012
83	5440.00	3.689	4.001	0.071175	-.311742	0.191867	-1.625	| ***| |	0.016

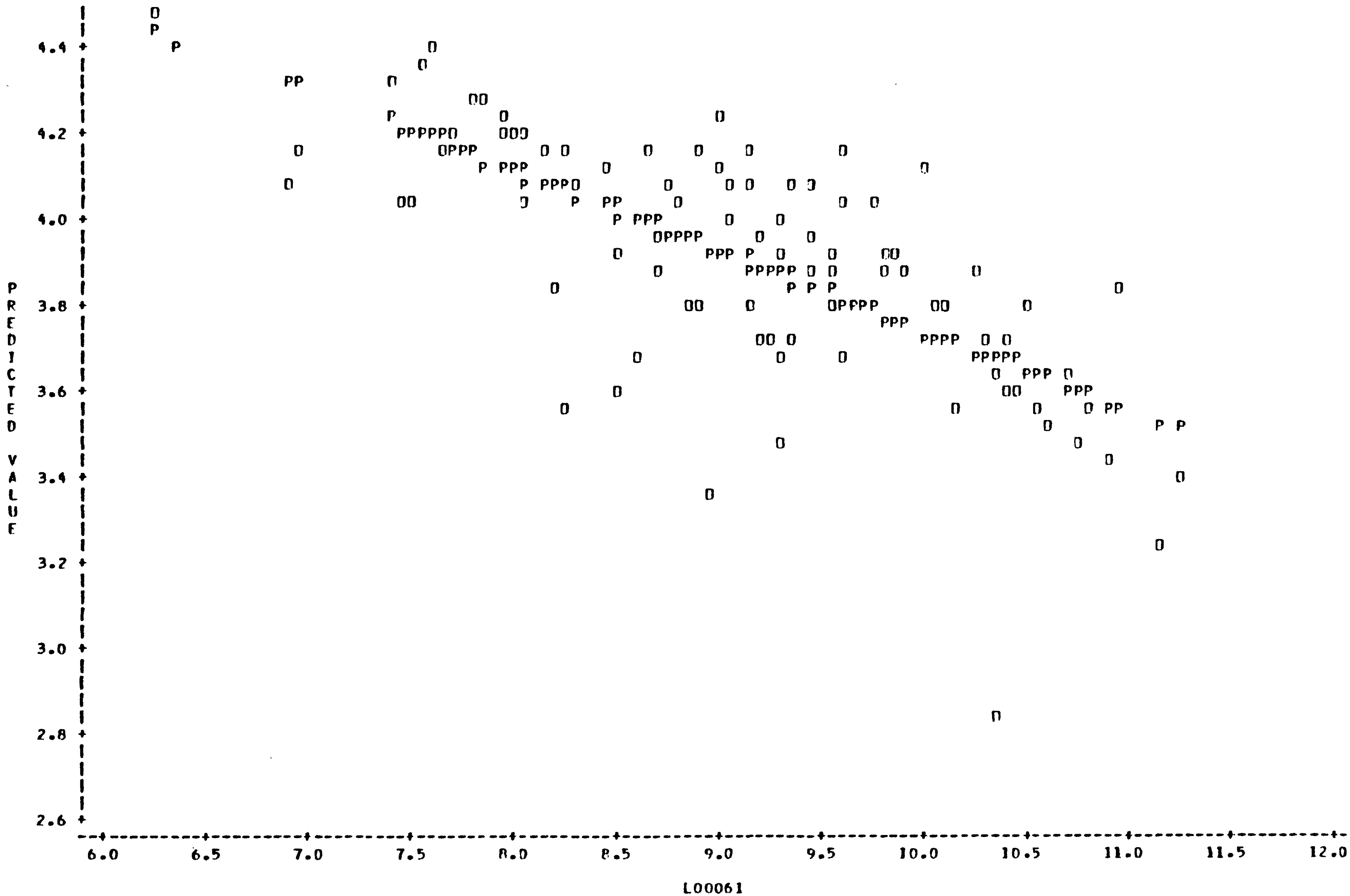
STATION IDENTIFICATION NUMBER=03374100

OBS	ID	ACTUAL	PREDICT VALUE	STD ERR PREDICT	RESIDUAL	STD ERR RESIDUAL	STUDENT RESIDUAL	-2-1-0 1 2	COOK'S D
84	18100.00	3.871	3.780	0.023658	0.091640	0.191577	0.478		0.002
85	49800.00	3.555	3.593	0.037154	-.038097	0.189423	-0.201		0.001
86	16500.00	3.807	3.797	0.022755	0.010082	0.191686	0.053		0.000
87	14800.00	3.664	3.817	0.021810	-.153014	0.191796	-0.798	*	0.004
88	9500.00	3.807	3.898	0.019581	-.091438	0.192037	-0.476		0.001
89	6000.00	.	3.983	0.020551	.	.	.		.
90	7580.00	3.367	3.940	0.019632	-.572323	0.192031	-2.980	*****	0.046
91	3200.00	4.025	4.098	0.026411	-.072847	0.191217	-0.381		0.001
92	2500.00	4.263	4.144	0.029592	0.119086	0.190750	0.624	*	0.005
93	2050.00	4.159	4.180	0.032367	-.021204	0.190299	-0.111		0.000
94	3170.00	4.205	4.100	0.026525	0.104762	0.191201	0.548	*	0.003
95	1650.00	4.304	4.220	0.035565	0.084062	0.189728	0.443		0.003
96	14600.00	3.784	3.819	0.021702	-.034888	0.191808	-0.182		0.000
97	5780.00	3.989	3.989	0.020773	-4.9E-04	0.191911	-0.003		0.000
98	12800.00	3.871	3.843	0.020771	0.027928	0.191911	0.146		0.000
99	24100.00	3.807	3.727	0.026916	0.079749	0.191147	0.417		0.002
100	11200.00	3.689	3.868	0.020070	-.178949	0.191986	-0.932	*	0.005
101	8160.00	4.127	3.926	0.019522	0.201074	0.192042	1.047	**	0.006
102	4880.00	3.584	4.021	0.022012	-.437079	0.191773	-2.279	****	0.034
103	2860.00	4.111	4.119	0.027815	-.007981	0.191018	-0.042		0.000

SUM OF RESIDUALS 1.35425E-12
SUM OF SQUARED RESIDUALS 3.577098

OUTPUT FROM LOGLOG: L00945= L00061
STATION IDENTIFICATION NUMBER=03374100

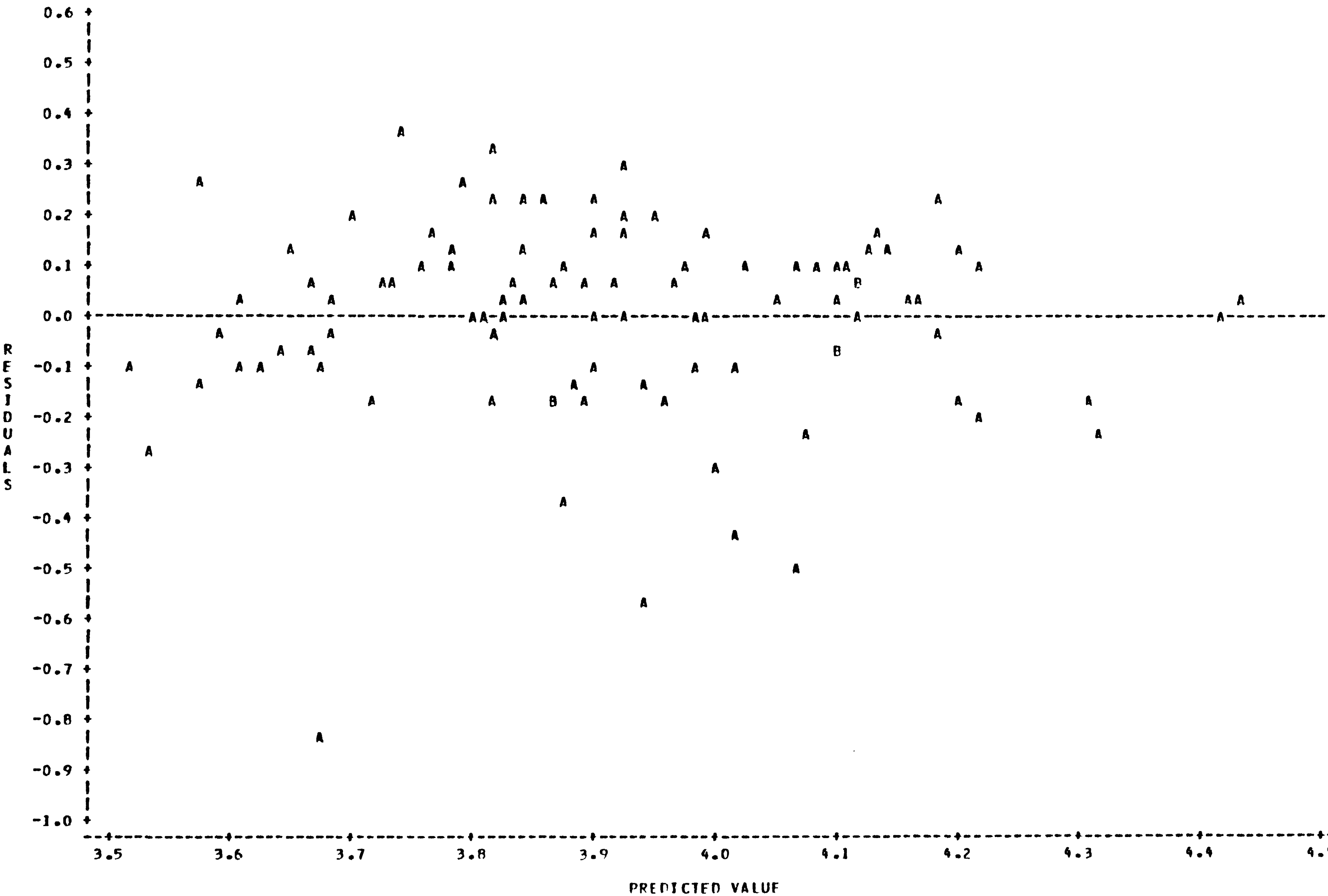
PLOT OF P*L00061 SYMBOL USED IS P
PLOT OF L00945*L00061 SYMBOL USED IS O



NOTE: 5 OBS HAD MISSING VALUES 45 OBS HIDDEN

STATION IDENTIFICATION NUMBER=03374100

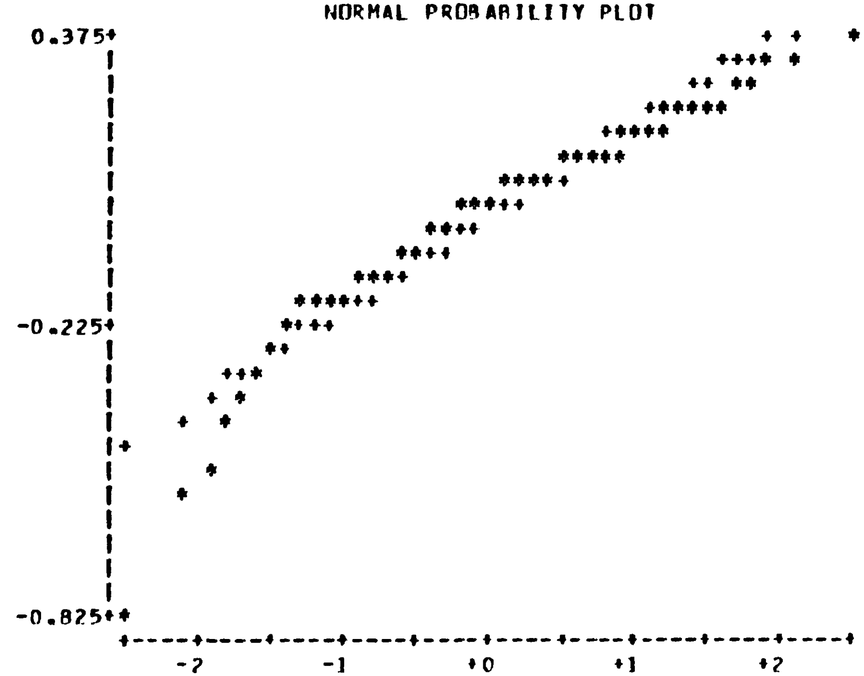
PLOT OF R*P LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 5 OBS HAD MISSING VALUES

```
                                              UNIVARIATE

VARIABLE=R             RESIDUALS

               MOMENTS                                     QUANTILES(DEF=4)                           EXTREMES

N                  98    SUM WGTS          98     100% MAX  0.371328       99%     0.371328       LOWEST       HIGHEST
MEAN        1.382E-14    SUM        1.354E-12      75% Q3    0.11434       95%     0.252185    -0.845044       0.25196
STD DEV      0.192035    VARIANCE   0.0368773      50% MED 0.0286457       90%     0.217227    -0.572323      0.256444
SKEWNESS     -1.35597    KURTOSIS     3.74112      25% Q1  -0.103751       10%     -0.19435    -0.512707      0.310731
USS            3.5771    CSS           3.5771       0% MIN -0.845044        5%    -0.377756    -0.437079      0.343545
CV          1.390E+15    STD MEAN   0.0193984                               1%    -0.845044    -0.374634      0.371328
T:MEAN=0    7.124E-13    PROB>|T|           1     RANGE      1.21637
SGN RANK        254.5    PROB>|S|    0.368084     Q3-Q1      0.21809
NUM ¬= 0           98                             MODE     -0.845044
D:NORMAL    0.0989695    PROB>D         0.018

                                                  MISSING VALUE       .
                                                          COUNT       5
                                                   % COUNT/NOBS    4.85

      STEM LEAF                      #         BOXPLOT
         3 7                         1            |
         3 14                        2            |
         2 56                        2            |
         2 00022234                  8            |
         1 556678                    6            |
         1 0111123334               10         +-----+
         0 5667777888888999         16         |     |
         0 1123333444               10         *--+--*
        -0 44322211000              11         |     |
        -0 9998875                   7         |     |
        -1 441110                    6         +-----+
        -1 9886655555               10            |
        -2 42                        2            |
        -2 8                         1            |
        -3 1                         1            |
        -3 7                         1            |
        -4 4                         1            0
        -4
        -5 1                         1            0
        -5 7                         1            0
        -6
        -6
        -7
        -7
        -8 5                         1            *
           ----+----+----+----+
           MULTIPLY STEM.LEAF BY 10**-01
```

NORMAL PROBABILITY PLOT



tools which aid the user in selecting the most appropriate regression model. The program uses the macro REGDATA to establish the data. The macro REGMAC uses the REG procedure (SAS Institute, Inc., 1982b, p. 39) rather than the SYSREG procedure to generate regression results. In addition to the information provided by SYSREG, the REG procedure in this application prints (1) the observed and predicted value for each observation; (2) the residual (FAC), the standard error of the residual and the studentized residual (the residual divided by its standard error); and (3) Cook's D, a measure of the change to the parameter estimates that would result from deleting each observation. This program also plots for each model the observed and predicted values of the selected dependent variables against the log of streamflow, and the FAC (residuals) against the predicted value. In these plots, the log of the flow is used on the x-axis to provide better resolution in the low discharge range. Finally, the macro does a residuals analysis with the Univariate procedure (SAS Institute, Inc., 1979, p. 427). This procedure does univariate statistics on the residuals, tests for normality by the Shapiro-Wilk W statistic (for sample sizes ≤50) or a modified version of the Kolmogorov-Smirnov D Statistic (for sample sizes >50) and provides a stemleaf, box plot, and probability plot of the residuals. For a discussion of regression analysis and the use of these dianostic tools, the reader is referred to Chatterjee and Price (1977), Belsley and others (1980), Daniel and Wood (1980), and Draper and Smith (1981).

The procedure is invoked three times in the example, shown in figure 4, once for the linear model of equation 2 (lines 65-68), once for the log-linear model of equation 3 (lines 69-72), and for the log-log model of equation 7 (lines 73-76). In each instance three variables are set - the model label (MODLABEL) in lines 65, 69, and 73, the independent variable (Y) in lines 66, 70, and 74, and the explanatory variables (FLOW) in lines 67, 71, and 75 -

then the macro REGMAC is called (lines 68, 72, and 76) to do the computation.

In the example shown in figure 4, the log-linear (equation 3) was selected to flow adjust the sulfate concentrations. The linear model (equation 2) is clearly deficient. Both plots demonstrate this quite clearly. However, the log-log model (equation 7) is not manifestly better or worse than the log-linear model. The pattern seen on the residuals plots for these two models are very similar, both showing some modest lack of fit at the low end and a slight amount of heteroscedasticity. The choice betwen these two models, in this case, is largely a matter of preference. The log-linear model has the advantage that the residuals are in units of mg/L.

PLOTTING WATER-QUALITY DATA AS A TIME SERIES

It is now possible to examine the flow-adjusted concentrations for trend. A good exploratory method for detecting trends is to plot the data as a time series. Figure 5 shows an example SAS program that generates time series plots for streamflow, sulfate concentration, and the flow-adjusted sulfate concentration for the data used in figure 3. Note that PROC REG (lines 12-14) is used to calculate the FAC and generate the data set used by PROC PLOT (lines 16-19) (SAS Institute, Inc., 1979, p. 343). Note also that the variable DECTIME (lines 17-19) is used as the time variable in these plots.

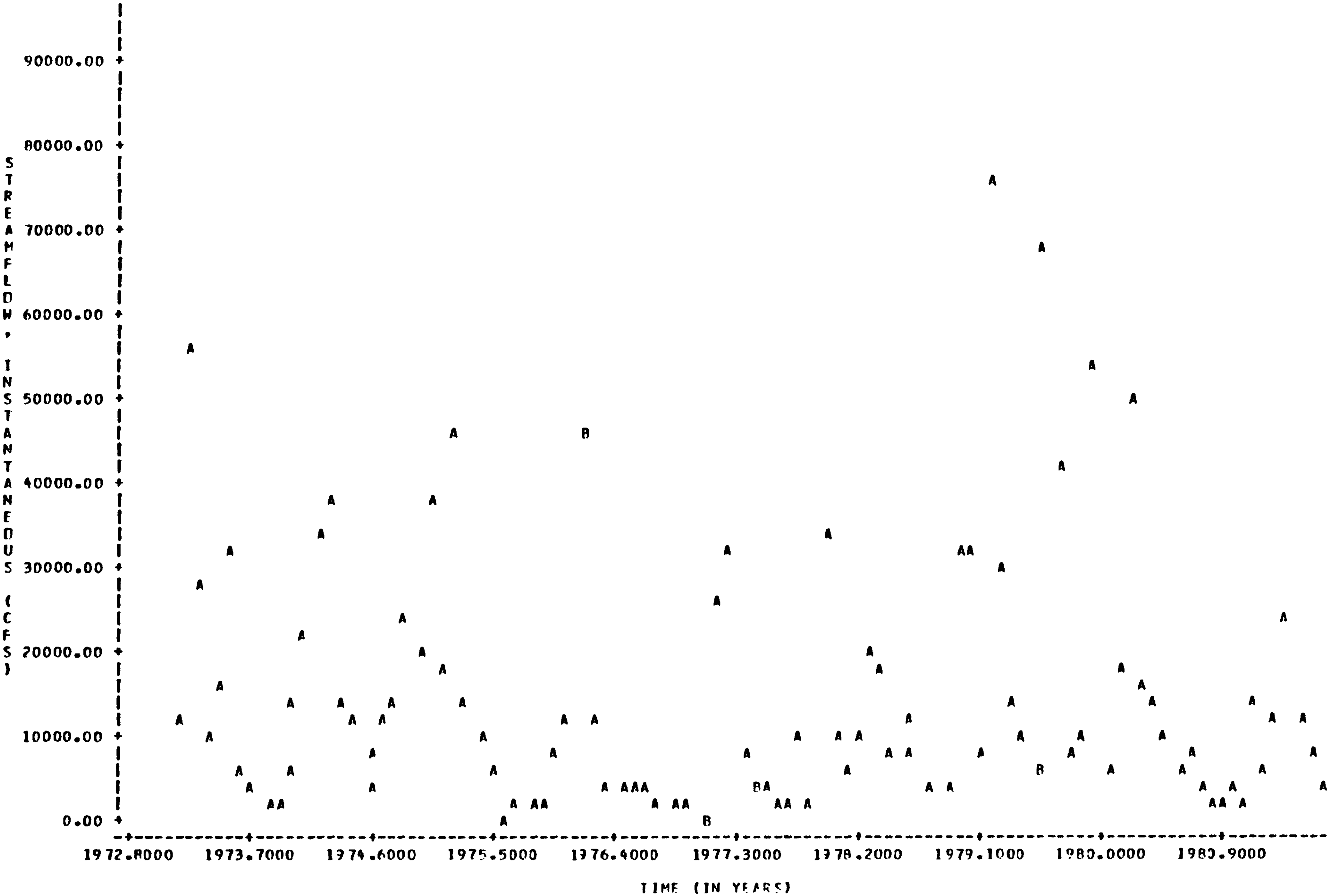
PROC PLOT produces line printer plots as shown in figure 5. More elaborate plots of SAS data can be made using PROC GPLOT (SAS Institute Inc., 1981b, p. 69). The procedure GPLOT is a part of the SAS/GRAPH system which is a computer graphics system for producing figures on terminal screens and plotters. A list of compatible graphics devices is given in the SAS/GRAPH User's Guide (SAS Institute, Inc., 1981b, p. 2).

```
//F5 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
OPTIONS NOOVP;
DATA DATAA;SET TREND.MONTHLY;
IF STATION=' 03374100      ';
L00061=LOG(P00061);
KEEP STATION SNAME P00945 P00061 L00061 DECTIME;
PROC SORT DATA=DATAA;BY STATION;
PROC REG DATA=DATAA NOPRINT;BY STATION;
MODEL P00945=L00061;
OUTPUT OUT=PLOTDATA P=P R=FAC;
PROC SORT DATA=PLOTDATA;BY STATION;
PROC PLOT DATA=PLOTDATA;BY STATION;
PLOT P00061*DECTIME;
PLOT P00945*DECTIME;
PLOT FAC*DECTIME;
/*
//
```

Figure 5.--Example input and output of SAS statements to plot water-quality data and regression residuals as time series.

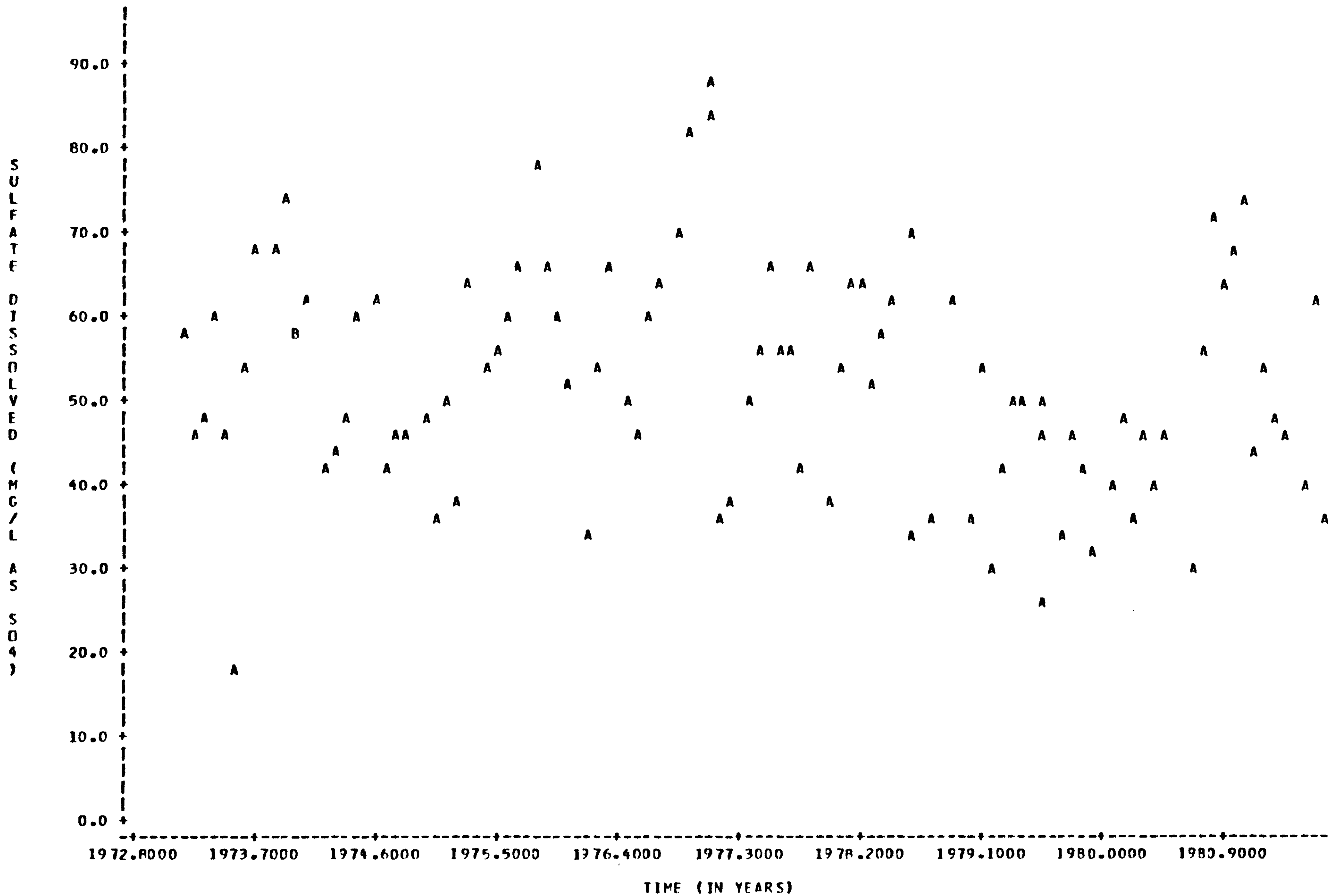
S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   16:16 THURSDAY, AUGUST 18, 1983   2
STATION IDENTIFICATION NUMBER=03374100

PLOT OF P00061*DECTIME   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   16:16 THURSDAY, AUGUST 18, 1983   3
STATION IDENTIFICATION NUMBER=03374100

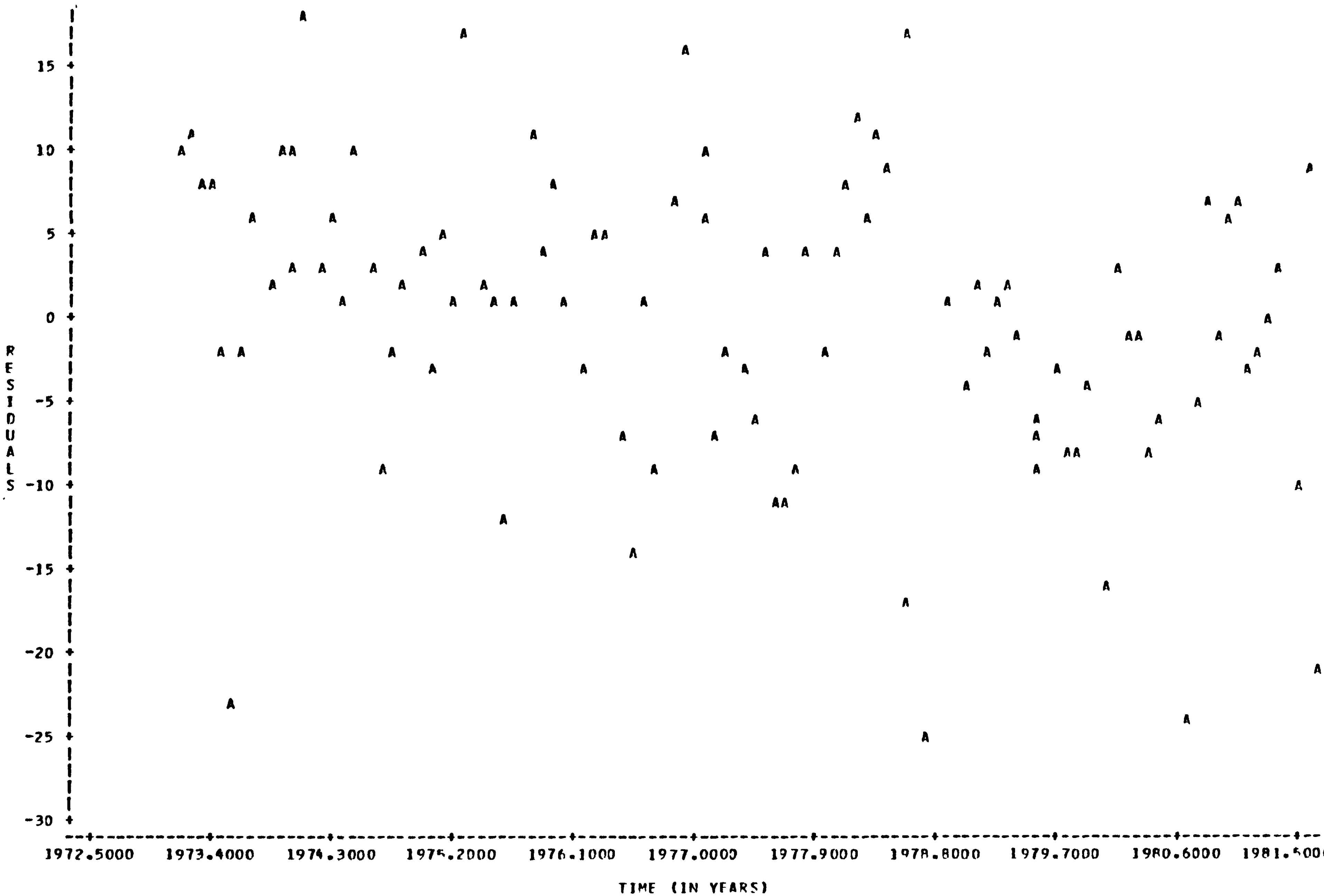
PLOT OF P00945*DECTIME   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 5 OBS HAD MISSING VALUES

STATION IDENTIFICATION NUMBER=03374100

PLOT OF FAC*DECTIME   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



NOTE: 5 OBS HAD MISSING VALUES

STATISTICAL PROCEDURES TO TEST FOR TRENDS IN WATER-QUALITY TIME SERIES

Two nonparametric tests for detecting trends in time series have been added to SAS at the U.S. Geological Survey Amdahl computer facility. PROC SEASKEN does the Seasonal Kendall test and slope estimator developed by Hirsch and others (1982). This procedure is suitable for detecting monotonic trends in time series with seasonality, missing values, or values reported as "less than." The procedure is not, however, robust against serial correlation. Additional information about the Seasonal Kendall test and slope estimator is given in the User's Guide presented in Appendix A.

PROC SEASRS tests for differences in the location parameters (mean, median, etc.) of two separate periods in a time series. The procedure uses a version of the Wilcoxon rank-sum test (Mann-Whitney test) modified to handle seasonality. This test is appropriate for detecting step trends (changes) in water quality before and after events such as construction of a dam or sewage treatment plant or an abrupt land-use change (construction, mining, clear cutting). Additional information about the modified Wilcoxon rank-sum test is given in the User's Guide presented in Appendix B.

In many cases, it may be appropriate to apply one of these PROC's to the concentration data as well as to the flow-adjusted concentration (FAC) data. Trends or changes in FAC may be interpreted as an indication that the streamflow concentration relationship has changed; that is, for a given discharge one may expect different concentration values today versus some time in the past. In cases where there have been substantial changes in the flow frequency distribution (due to changing regulation, diversion, or consumption), it may be very useful to look for trends or changes in raw concentration data. This will be indicative of the combined effects of changes in human inputs to the stream and changes in the flow frequency distribution on the frequency distribution of concentrations.

The presence of a large number of missing values or "less than" values (censoring) in the records being tested does not substantially affect the significance of the tests, although their power (ability to detect actual trends) may be reduced. PROC SEASKEN has been tested with as much as 50 percent of the seasonal values missing or 50 percent of them censored without any problem in significance.

The consequences of using these tests when data are serially correlated is that the actual significance of the test becomes higher than the nominal significance level. For monthly data, serial correlations are generally in a range such that actual significance may be about twice as high as the nominal. In other words, a p-level reported as 0.05 may more accurately be as high as 0.10. The problem becomes severe where the measurements are from a large body of water with a long residence time (by comparison with the sampling frequ y). Aquifers, large lakes, or reservoirs, or the streams draining large lakes or reservoirs, may present problems. Keeping the parameters SEASON (see appendicies A and B) small, typically no larger than 12, will help avoid serious problems due to serial correlation. When the correlation problem is thought to be severe, reducing SEASON (reducing the number of values per year used in the test) should prove to be helpful.

Figure 6 shows a program listing and output of a SAS job that uses PROC SEASKEN to test for time trends in the streamflow, sulfate concentration, and the flow-adjusted sulfate concentrations used in the regression example shown in figure 3. The flow-adjusted concentrations were calculated by the REG procedure using the LOGLIN model shown in figure 3 and are included in the output data set generated by PROC REG.

```
//F6 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
DATA DATAA;SET TREND.MONTHLY;
IF STATION=' 03374100       ';
L00061=LOG(P00061);
KEEP STATION SNAME P00945 P00061 L00061 DECTIME;
PROC SORT DATA=DATAA;BY STATION;
PROC REG DATA=DATAA NOPRINT;BY STATION;
MODEL P00945=L00061;
OUTPUT OUT=PLOTDATA P=P R=FAC;
PROC SORT DATA=PLOTDATA;BY STATION;
PROC SEASKEN DATA=PLOTDATA SEASON=12;BY STATION;
VAR DECTIME P00061 P00945 FAC;
/*
//
```

Figure 6.--Example input and output of SAS statements to do the Seasonal Kendall test and slope estimator procedure on a water-quality data time series.

STATION IDENTIFICATION NUMBER=03374100

SEASONAL KENDALL TEST AND SLOPE ESTIMATOR FOR TREND MAGNITUDE

WATER YEARS 1973-1981

VARIABLE		NOBS	NVALS	TAU	P LEVEL	SLOPE
P00061	STREAMFLOW, INSTANTANEOUS (CFS)	103	98	-0.020	0.837	-60.00
P00945	SULFATE DISSOLVED (MG/L AS SO4)	98	96	-0.182	0.032	-1.333
FAC	RESIDUALS	98	96	-0.279	0.001	-1.066

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

The results of the SEASKEN test show that no significant trend is evident in streamflow but that both the sulfate and flow-adjusted sulfate concentrations exhibited highly significant decreasing trends (p-level(p)<0.05) during the period of record used in the test.

This trend can be further examined using PROC SEASRS. The data used in this example are from a station draining a watershed containing considerable surface coal mining. The regulations promulgated under the Surface Mine Control and Reclamation Act of 1977 (PL 95-87) were implemented toward the latter third of the period of record for this station (May 1979). Figure 7 shows a program listing and output of a SAS job that uses PROC SEASRS to determine if differences exist in streamflow, sulfate concentration, and the flow-adjusted sulfate concentration (using the LOGLIN model of figure 3) before and after this date. The results of the SEASRS test show that no significant step trend is evident in streamflow (p-level>0.05) but that sulfate concentration and the flow-adjusted sulfate concentration exhibit a highly significant trend (p-level<0.05).

The reader should be aware that both the Seasonal Kendall test and the Seasonal Wilcoxon rank-sum test are exploratory in nature. The fact that the Seasonal Wilcoxon rank-sum test indicates a difference between the periods in the time series prior to and following the implementation of the surface mine regulations does not necessarily imply causality, nor does it imply that it will continue into the future.

```
//F7 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
DATA DATAA;SET TREND.MONTHLY;
IF STATION=' 03374100       ';
L00061=LOG(P00061);
KEEP STATION SNAME P00945 P00061 L00061 DECTIME;
PROC SORT DATA=DATAA;BY STATION;
PROC REG DATA=DATAA NOPRINT;BY STATION;
MODEL P00945=L00061;
OUTPUT OUT=FACDATA P=P R=FAC;
PROC SORT DATA=FACDATA;BY STATION;
PROC SEASRS DATA=FACDATA SEASON=12 DATE=1979.33;BY STATION;
VAR DECTIME P00061 P00945 FAC;
/*
//
```

Figure 7.--Example input and output of SAS statements to do the Seasonal Mann-Whitney-Wilcoxon rank sum and slope estimator procedure on a water-quality data time series.

STATION IDENTIFICATION NUMBER=03374100

MANN-WHITNEY-WILCOXAN RANK SUM TEST FOR SEASONAL DATA

VARIABLE		FIRST SERIES (1973-1979) NOBS	NVALS	CUTOFF DATE	SECOND SERIES (1979-1981) NOBS	NVALS	P LEVEL	STEP
P00061	STREAMFLOW, INSTANTANEOUS (CFS)	73	69	1979.33	30	29	0.507	1050.
P00945	SULFATE DISSOLVED (MG/L AS SO4)	69	68	1979.33	29	28	0.006	-7.000
FAC	RESIDUALS	69	68	1979.33	29	28	0.006	-5.350

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

APPENDIX A: PROC SEASKEN user's guide with examples

This procedure tests for monotonic trends in time series using a modified form of Kendall's tau (Kendall, 1975) derived by Hirsch and others (1982). The null hypothesis for this test is that the probability distribution of the random variable does not change over time. In the Kendall's tau test, all possible pairs of data values are compared. If a later value (in time) is higher, a plus is recorded, if the later value is lower, a minus is recorded. If no trend exists in the data, the probability of a later value being higher or lower than any previous value is 0.50. In this case, the number of pluses should approximately equal the number of minuses. If the number of pluses greatly exceeds the number of minuses, the values later in the series are more often higher than those earlier in the series, indicating an uptrend. If the number of minuses greatly exceeds the number of pluses, a downtrend is indicated. In this modified Kendall's tau, the problem of seasonality is avoided by comparing only observations from the same season of the year. Thus, for monthly data with seasonality, January data is compared only with January data, and so on.

Trend magnitude is determined using the seasonal Kendall slope estimator (Hirsch and others, 1982). The slope estimate is taken to be the median of the slopes of the ordered pairs of data values compared in the Seasonal Kendall test:

A more complete discussion of the Seasonal Kendall test and Seasonal Kendall slope estimator and its use are given in Smith and others (1982).

SPECIFICATIONS

The following statements may be used in the SEASKEN procedure:

```
PROC SEASKEN SEASON = n options;
BY variables;
VAR variables;
```

The PROC SEASKEN statement invokes the procedure. The VAR statement is used to specify the numeric variables for the analysis. The BY statement is optional. The observations must be in chronological order. This can be done by using PROC SORT and sorting by the variables in the BY statement followed by DECTIME. For example:

```
PROC SORT; BY variable DECTIME;
```

PROC SEASKEN statement

The PROC SEASKEN statement has the form:

```
PROC SEASKEN SEASON = n options;
```

SEASON = n   gives the number of seasonal values per year. For example, SEASON = 12 would be used for data collected monthly and SEASON = 52 would be used for data collected weekly. SEASON = 1 may be used for annual summary data, such as average annual streamflow. Each water year will be divided into n equal length seasons. Note that for SEASON=12, for example, the 12 seasons will not correspond exactly to the 12 calendar months since months are not of equal length. For each season of each year, there may be no observations (a missing value will be assumed as the seasonal value), one observation (that value will be used as the seasonal value), or several (not more than 400) observations (the median of the values will be used as the seasonal value).

This statement is not optional and must be included.

The options that may be used in the PROC SEASKEN statement follow.

DATA = SASdataset   gives the name of the SAS dataset to be used by PROC SEASKEN. If it is omitted, the most recently created SAS data set is used.

DL1 - DL15 = detection limit

gives the detection limit of the analytical procedures used to determine the constituent values in the time series. The number associated with each call of the detection limit option refers to the order of the constituents on the VAR statement. The use of the detection limit option is most important when several different analytical procedures with different degrees of sensitivity were used to determine values of the same constituent in the time series. In this case, the highest detection limit of all the procedures should be used.

The detection limit option sets all values less than or equal to the detection limit equal to one-half the value of the detection limit.

This option would not be used when the test is applied to Flow-Adjusted Concentrations (FAC).

VAR statement

VAR dectime variables;

The names of the variables to be tested for trends are listed in the VAR statement. The first variable must contain the decimal time values corresponding to the times the observations in the time series were made. Dectime is in the form of decimal time; for example, 12 noon, June 1, 1975, will be 1975.4178. Up to 15 additional numeric variables may be included. The total number of seasonal values (number of seasons times the number of water years) may not exceed 1200. The VAR statement is not optional and must be included.

BY statement

BY variables;

A BY statement may be used with PROC SEASKEN to obtain separate analyses on observations in groups defined by the BY variables. The input data set must be sorted in order of the BY variables.

DETAILS

Missing Values

Missing values may appear in a time series used in SEASKEN. The existence of missing values presents no theoretical problem for applying the seasonal Kendall test for trend. Comparisons of data pairs where one member is missing are not included in the calculation of the test statistic. (Internally, missing values are represented as -1E31.)

One or more seasons may be devoid of data; for example, one may use SEASON=12 and have data only for the summer. In this case, all other months will be set to missing values.

Formulas

The test statistic, S_i, is:

$$S_i = \sum_{k=1}^{n_i-1} \sum_{j=k+1}^{n_i} \text{sgn}\,(X_{ij}-X_{ik})$$

Where n_i are the number of annual values for season i,

X_{ij} the seasonal value for season i and year j,

X_{ik} the seasonal value for season i and year k,

and

$$\text{sgn}\,(\Theta) = \begin{cases} 1 \text{ if } \Theta > 0 \\ 0 \text{ if } \Theta = 0 \\ -1 \text{ if } \Theta < 0 \end{cases}$$

The expected value of S_i is 0, $E[S_i] = 0$ and its variance is:

$$\text{Var}[S_i] = \frac{n_i(n_i-1)(2n_i+5) - \sum_{t_i} t_i(t_i-1)(2t_i+5)}{18}$$

where t_i is the extent of a given tie (number of X's involved in tie) for season i,

and $\sum_{t_i}$ denotes the summation over all ties.

The composite statistic of the seasonal statistics, S_i, is S':

$$E[S'] = \sum_{i=1}^{season} E[S_i] = 0$$

and the variance of S' is:

$$\text{Var}[S'] = \sum_{i=1}^{season} \text{Var}[S_i] = \sum_{i=1}^{season} \frac{n_i(n_i-1)(2n_i+5)}{18}$$

The normal approximation with a continuity correction of 1 (toward zero) is used to estimate $p = \text{Prob}[|S'| \geq s]$ (the probability that S' will depart from zero by the amount s or more). The standard normal deviate, Z, is calculated by:

$$Z = \begin{cases} \dfrac{S' - 1}{(\text{Var } S')^{1/2}} & \text{if } S' > 0 \\ 0 & \text{if } S' = 0 \\ \dfrac{S' + 1}{(\text{Var } S')^{1/2}} & \text{if } S' < 0 \end{cases}$$

The p-level of the test is determined from Z using the standard normal distribution. The p-level is the probability of obtaining a Z value that is as large or larger in absolute value than the one obtained if the null hypothesis were true (i.e., there was actually no trend). If one pre-selects a significance level α (the risk of rejecting the null hypothesis when it is actually true), then one should reject whenever the p-level is less than α. A typical value for α is 0.05.

The statistic τ (tau) is:

$$\tau = \sum_{i=1}^{season} \frac{S_i}{n_i(n_i-1)/2}$$

Note that τ may only take on values between -1 and +1. Negative values indicate downwards trend, positive values indicate upwards trend. It is a type of rank correlaton coefficient between the variable and time.

The seasonal Kendall slope estimator is the median value of

$$d_{ijk} = (X_{ij} - X_{ik}) / (j-k) \text{ for all } (X_{ij}, X_{ik}) \text{ pairs}$$

where d_{ijk} is the slope between seasonal values for season i, year j and season i, year k with j>k.

The slope estimate is valid only when all of the data are reported as above the limit of detection. If there are only a few "less thans" it can be viewed as a reasonable approximation.

Printed output

For each variable SEASKEN prints (see figures A1 and A2)

1. the variable name (VARIABLE)
2. the variable label, if any
3. the total number of non-missing observations in the input data (NOBS)
4. the number of seasonal values constructed from the observations (NVALS)
5. the statistic tau (τ) (TAU)
6. The significance probability (p-level) of the trend, two-sided (P LEVEL)
7. The estimate of trend magnitude, in units per year (SLOPE)

EXAMPLES

Example 1: Annual Mean Streamflow

In this example (figure A-1), PROC SEASKEN is called to test for a trend in the annual mean streamflow observed at a U.S. Geological Survey streamflow gaging station. The data was entered from card input (lines 9-19).

Example 2: NASQAN Data

In this example (figure A-2), PROC SEASKEN is called to test for trends in 10 water quality constituents at three U.S. Geological Survey NASQAN stations. The data was previously stored as a SAS data set on a direct access storage device. Note that the SAS data set, TREND.MONTHLY was sorted by STATION (line 6) so that the BY statement option (line 8) could be used with SEASKEN.

```
//A1 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//SYSIN DD *
DATA SFLOW;
 INPUT WYEAR AMQ @@;LIST;
 DECTIME=WYEAR-0.25;
 LABEL AMQ=ANNUAL MEAN STREAMFLOW;
 CARDS;
1925  8010 1926  4760 1927 10800 1928 10300 1929  4950 1930 10500 1931  2200
1932  4660 1933  5930 1934  5630 1935  3430 1936  5230 1937  6960 1938  7140
1939  7720 1940  3570 1941  2410 1942  3650 1943  5950 1944  6880 1945  4910
1946  7660 1947  3430 1948  6930 1949  8370 1950 10400 1951  9440 1952  8290
1953  6300 1954  2840 1955  5100 1956  5410 1957  3520 1958  8690 1959 10200
1960  6340 1961  4650 1962  8330 1963  3880 1964  2700 1965  5160 1966  4280
1967  6050 1968  7450 1969  7280 1970  6840 1971  7130 1972  4450 1973 11200
1974  9930 1975  7180 1976  8180 1977  3080 1978  6440 1979  6870 1980  6470
1981  5540
;
PROC SEASKEN SEASON=1;
VAR DECTIME AMQ;
/*
//
```


Figure A1.--Example input and output using PROC SEASKEN to test for a trend in annual mean streamflow.

S T A T I S T I C A L   A N A L Y S I S   S Y S T E M   15:38 THURSDAY, AUGUST 18, 1983   1

SEASONAL KENDALL TEST AND SLOPE ESTIMATOR FOR TREND MAGNITUDE

WATER YEARS 1925-1981

VARIABLE		NOBS	NVALS	TAU	P LEVEL	SLOPE
AMQ	ANNUAL MEAN STREAMFLOW	57	57	0.099	0.670	9.839

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

```
//A2 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
PROC SORT DATA=TREND.MONTHLY;BY STATION;
PROC SEASKEN DATA=TREND.MONTHLY SEASON=12 DL3=0.1 DL4=0.01 DL7=0.1 DL10=0.1;
BY STATION;
VAR DECTIME P70300 P00410 P00630 P00665 P00915 P00925 P00930 P00940 P00945 P0095
0;
/*
//
```

Figure A2.--Example input and output using PROC SEASKEN to test for trends in water-quality constituents.

STATION IDENTIFICATION NUMBER=03276500

SEASONAL KENDALL TEST AND SLOPE ESTIMATOR FOR TREND MAGNITUDE

WATER YEARS 1975-1981

VARIABLE		NOBS	NVALS	TAU	P LEVEL	SLOPE
P70300	SOLIDS, RESIDUE AT 180 DEG. C DISSOLVED	77	74	0.119	0.259	4.000
P00410	ALKALINITY FIELD (MG/L AS CACO3)	66	63	-0.072	0.565	-1.292
P00630	NITROGEN, NO2+NO3 TOTAL (MG/L AS N)	76	73	0.233	0.024	.1000
P00665	PHOSPHORUS, TOTAL (MG/L AS P)	76	73	-0.063	0.559	-.1000E-29
P00915	CALCIUM DISSOLVED (MG/L AS CA)	77	74	-0.077	0.469	-.5000
P00925	MAGNESIUM, DISSOLVED (MG/L AS MG)	77	74	0.031	0.794	.1000E-29
P00930	SODIUM, DISSOLVED (MG/L AS NA)	77	74	-0.021	0.877	-.1000E-29
P00940	CHLORIDE, DISSOLVED (MG/L AS CL)	77	74	0.093	0.376	.1833
P00945	SULFATE DISSOLVED (MG/L AS SO4)	77	74	-0.175	0.087	-.6458
P00950	FLUORIDE, DISSOLVED (MG/L AS F)	77	74	-0.046	0.598	-.1000E-29

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

STATION IDENTIFICATION NUMBER=03374100

SEASONAL KENDALL TEST AND SLOPE ESTIMATOR FOR TREND MAGNITUDE

WATER YEARS 1973-1981

VARIABLE		NOBS	NVALS	TAU	P LEVEL	SLOPE
P70300	SOLIDS, RESIDUE AT 180 DEG. C DISSOLVED	98	96	-0.073	0.400	-2.000
P00410	ALKALINITY FIELD (MG/L AS CACO3)	86	84	-0.214	0.022	-4.667
P00630	NITROGEN, NO2+NO3 TOTAL (MG/L AS N)	93	91	0.236	0.007	.8000E-01
P00665	PHOSPHORUS, TOTAL (MG/L AS P)	92	90	0.107	0.235	.5000E-02
P00915	CALCIUM DISSOLVED (MG/L AS CA)	98	96	-0.085	0.323	-.6667
P00925	MAGNESIUM, DISSOLVED (MG/L AS MG)	98	96	-0.065	0.456	-.2000
P00930	SODIUM, DISSOLVED (MG/L AS NA)	98	96	0.117	0.168	.1667
P00940	CHLORIDE, DISSOLVED (MG/L AS CL)	98	96	0.220	0.009	.5000
P00945	SULFATE DISSOLVED (MG/L AS SO4)	98	96	-0.182	0.032	-1.333
P00950	FLUORIDE, DISSOLVED (MG/L AS F)	98	96	-0.126	0.078	-.1000E-29

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

STATION IDENTIFICATION NUMBER=03376500

SEASONAL KENDALL TEST AND SLOPE ESTIMATOR FOR TREND MAGNITUDE

WATER YEARS 1975-1981

VARIABLE		NOBS	NVALS	TAU	P LEVEL	SLOPE
P70300	SOLIDS, RESIDUE AT 180 DEG. C DISSOLVED	72	70	0.011	0.957	.5835
P00410	ALKALINITY FIELD (MG/L AS CACO3)	63	62	-0.061	0.643	-.9750
P00630	NITROGEN, NO2+NO3 TOTAL (MG/L AS N)	75	72	0.174	0.098	.7550E-01
P00665	PHOSPHORUS, TOTAL (MG/L AS P)	75	72	0.130	0.222	.6333E-02
P00915	CALCIUM DISSOLVED (MG/L AS CA)	72	71	0.056	0.628	.7500
P00925	MAGNESIUM, DISSOLVED (MG/L AS MG)	73	71	-0.006	1.000	-.1000E-29
P00930	SODIUM, DISSOLVED (MG/L AS NA)	73	71	0.050	0.664	.6000E-01
P00940	CHLORIDE, DISSOLVED (MG/L AS CL)	76	72	0.174	0.098	.6333
P00945	SULFATE DISSOLVED (MG/L AS SO4)	74	72	-0.147	0.167	-1.717
P00950	FLUORIDE, DISSOLVED (MG/L AS F)	74	72	-0.027	0.800	-.1000E-29

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

Appendix B: PROC SEASRS user's guide with examples

The SEASRS procedure tests for differences in the location parameters of two separate periods in a time series using a modified version of the Wilcoxon rank-sum test (Bradley, 1968, p. 105). This test is equivalent to the Mann-Whitney test described in Conover (1971, p. 224). The null hypothesis for this test is that two populations comprised by data from two separate periods in a time series are identical. This test assumes that sampling was random.

If the null hypothesis is true, no distinction can be made between the n observations in the first period and the m observations in the second period of the time series. In effect, all were taken from a common population. Therefore, each of the possible combinations of the n+m observations taken from the common population were equally likely to have become the two samples actually collected. For each of these possible combinations exists a value of the test statistic, W. This statistic is the sum of the ranks of the n observations within the combined (n+m observations) sample. (The smallest value in the combined sample receives a rank of 1, the next smallest a rank of 2, etc.). The null hypothesis is rejected if the value of the test statistic W differs from the expected value of W by a preselected value corresponding to a desired probability.

For a more thorough discussion of the Wilcoxon rank-sum test, and examples of its use, the reader is referred to Bradley (1968, p. 105) and Hollander and Wolfe (1973, p. 68).

SPECIFICATIONS

The following statements may be used in the SEASRS procedure:

```
PROC SEASRS   SEASON = n   DATE = x options;
BY variables;
VAR variables.
```

The PROC SEASRS statement invokes the procedure. The VAR statement is used to specify the numeric variables for the analysis. The BY statement is optional. The observations must be in chronological order. This can be done by using PROC SORT and sorting by the variables in the BY statement followed by DECTIME. For example:

```
PROC SORT; BY variable DECTIME;
```

<u>PROC SEASRS statement</u>

The PROC SEASRS statement has the form:

```
PROC SEASRS   SEASON = n   DATE = x options;
```

SEASON = n — gives the number of seasonal values per year. For example, SEASON = 12 would be used for data collected monthly and SEASON = 52 would be used for data collected weekly. SEASON = 1 may be used for annual summary data, such as average annual streamflow. Each water year will be divided into n equal length seasons. Note that for SEASON=12, for example, the 12 seasons will not correspond exactly to the 12 calendar months, since months are not of equal length. For each season of each year, there may be no observations (a missing value will be assumed as the seasonal value), one observation (that value will be used as the seasonal value), or several (not more than 400) observations (the median of the observations will be used as the seasonal value).

This statement is not optional and must be included.

DATE = xxxx — gives the time separating the two periods of the time series. DATE is in the form of decimal time, for example, 12 noon, June 1, 1975, will be 1975.4178. If the two periods in the time series are widely separated, any date in the gap will suffice. An observation occurring at precisely time DATE is placed in the second period.

This statement is not optional and must be included.

The options that may be used in the PROC SEASRS statement follow.

DATA = SASdataset — gives the name of the SAS dataset to be used by PROC SEASRS. If it is omitted, the most recently created SAS data set is used.

DL1 - DL15 = detection limit

gives the detection limit of the analytical procedures used to determine the constituent values in the time series. The number associated with each call of the detection limit option refers to the order of the constituents on the VAR statement. The use of the detection limit option is most important when several different analytical procedures with different degrees of sensitivity were used to determine values of the same constituent in the time series. In this case, the highest detection limit of all the procedures should be used.

The detection limit option sets all values less than or equal to the detection limit equal to one-half the value of the detection limit.

This option would not be used when the test is applied to Flow-Adjusted Concentration (FAC).

VAR statement

VAR dectime variables;

The names of the variables to be tested are listed in the VAR statement. The first variable must contain the decimal time values corresponding to the times the observations in the time series were made. Dectime is in the form of decimal time; for example, 12 noon, June 1, 1975, will be 1975.4178. Up to 15 additional numeric variables may be included. The total number of seasonal values (number of seasons times the sum of the number of water years in the first period plus the number of water years in the second period) may not exceed 1200. The VAR statement is not optional and must be included.

BY statement

BY variables;

A BY statement may be used with PROC SEASRS to obtain separate analyses on observations in groups defined by the BY variables. The input data set must be sorted in order of the BY variables.

DETAILS

Missing values

Missing values may be included in time series used in SEASRS. The existance of missing values presents no theoretical problem for applying the Mann-Whitney-Wilcoxon rank-sum test for seasonal data. Ranks of observations with missing values are not included in the calculation of the test statistic. (Internally, missing values are represented as -1E31.)

One or more seasons may be devoid of data; for example, one may use SEASON=12 and have data only for the summer. In this case all other months will be set to missing values.

Formulas

The test statistic, W_i, is:

$$W_i = \sum_{j=1}^{n_i} R_n$$

where n_i is the number of annual values for season i in the first period in the time series,

and R_n are the ranks of the seasonal values for season i in the first period of the time series.

The expected value of W_i is:

$$E\,[W_i] = [n_i\,(n_i + m_i + 1)\,] \,/\, 2$$

where m_i is the number of annual values for season i in the second period in the time series,

and its variance is:

$$Var\,[W_i] = [n_i\, m_i\,(n_i + m_i + 1)\,] \,/\, 12.$$

The composite statistic of the seasonal statistic, W_i, is W and is defined as:

$$W = \sum_{i=1}^{season} W_i.$$

The expectation of W is:

$$E\,[W] = \sum_{i=1}^{season} E\,[W_i]$$

and its variance is:

$$Var\,[W] = \sum_{i=1}^{season} [W_i].$$

The normal approximation is used to estimate $p = \text{Prob}\ [|W - E\ [W]| \geq w]$ (the probability that W will depart from its expected value by the amount w or more). The standard normal deviate, Z, is calculated by:

$$Z = \frac{W - E[W]}{(\text{Var}[W])^{1/2}} .$$

The p-level of the test is determined from Z using the standard normal distri-bution. The p-level is the probability of obtaining a Z value that is as large or larger in absolute value than the one obtained if the null hypothesis were true (i.e., there was actually no trend). If one pre-selects a significance level α (th risk of rejecting the null hypothesis when it is actually true), then one should reject whenever the p-level is less than α. A typical value for α is 0.05.

An estimate of the magnitude of the step trend is taken as the median of the difference between all pairs of seasonal values, one from each period but of the same season.

The step trend estimate is valid only when all of the data are reported as above the limit of detection. If there are only a few "less thans" it can be viewed as a reasonable approximation.

Printed Output

For each variable SEASRS prints (see figures B1 and B2):

1. the variable name (VARIABLE)
2. the variable label, if any
3. the number of original observations (NOBS) and number of constructed seasonal values (NVALS) in the first series
4. the CUTOFF DATE value
5. the number of original observations and number of constructed seasonal values in the second series
6. the significance probability of the difference, two-sided (P LEVEL)
7. the estimate of the step trend (STEP)

EXAMPLES

Example 1: Annual Mean Streamflow

In this example (figure B-1), PROC SEASRS is called to test for differences in annual mean streamflows observed at a U.S. Geological Survey streamflow gaging station before and after construction of a series of reservoirs in the drainage basin. The data was entered from card input (lines 9-19).

Example 2: NASQAN Data

In this example (figure B-2), PROC SEASRS is called to test for differences in 10 water-quality constituents at three U.S. Geological Survey NASQAN stations before and after implementation of the Surface Mining Control and Reclamation Act of 1977. The second and third stations are located in watersheds mined for coal. The first station is not and serves as a control. Note that the SAS data set, TREND.MONTHLY was sorted by STATION (line 6) so that the BY statement option (line 8) could be used with SEASRS.

```
//B1 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//SYSIN DD *
DATA SFLOW;
 INPUT WYEAR AMQ @@;LIST;
 DECTIME=WYEAR-0.25;
 LABEL AMQ=ANNUAL MEAN STREAMFLOW;
 CARDS;
1925  8010 1926  4760 1927 10800 1928 10300 1929  4950 1930 10500 1931  2200
1932  4660 1933  5930 1934  5630 1935  3430 1936  5230 1937  6960 1938  7140
1939  7720 1940  3570 1941  2410 1942  3650 1943  5950 1944  6880 1945  4910
1946  7660 1947  3430 1948  6930 1949  8370 1950 10400 1951  9440 1952  8290
1953  6300 1954  2840 1955  5100 1956  5410 1957  3520 1958  8690 1959 10200
1960  6340 1961  4650 1962  8330 1963  3880 1964  2700 1965  5160 1966  4280
1967  6050 1968  7450 1969  7280 1970  6840 1971  7130 1972  4450 1973 11200
1974  9930 1975  7180 1976  8180 1977  3080 1978  6440 1979  6870 1980  6470
1981  5540
;
PROC SEASRS SEASON=1 DATE=1963.75;
VAR DECTIME AMQ;
/*
//
```

Figure B1.--Example input and output using PROC SEASRS to test for a step in annual mean streamflow.

MANN-WHITNEY-WILCOXAN RANK SUM TEST FOR SEASONAL DATA

VARIABLE		FIRST SERIES (1925-1963) NOBS	FIRST SERIES (1925-1963) NVALS	CUTOFF DATE	SECOND SERIES (1964-1981) NOBS	SECOND SERIES (1964-1981) NVALS	P LEVEL	STEP
AMQ	ANNUAL MEAN STREAMFLOW	39	39	1963.75	18	18	0.770	250.0

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

```
//B2 JOB
// CLASS
//PROCLIB  DD DSN=WRD.PROCLIB,DISP=SHR
//A EXEC WRDSAS,DSN1=NULLFILE,DSN2=NULLFILE
//TREND DD DSN=USERID.FILENAME,DISP=OLD
//SYSIN DD *
PROC SORT DATA=TREND.MONTHLY;BY STATION;
PROC SEASRS DATA=TREND.MONTHLY SEASON=12 DATE=1979.4 DL3=0.1 DL4=0.01 DL7=0.1 DL
10=0.1;BY STATION;
VAR DECTIME P70300 P00410 P00630 P00665 P00915 P00925 P00930 P00940 P00945 P0095
0;
/*
//
```

Figure B2.--Example input and output using PROC SEASRS to test for step trends in water-quality constituents.

STATION IDENTIFICATION NUMBER=03276500

MANN-WHITNEY-WILCOXAN RANK SUM TEST FOR SEASONAL DATA

VARIABLE		FIRST SERIES (1975-1979) NOBS	FIRST SERIES (1975-1979) NVALS	CUTOFF DATE	SECOND SERIES (1979-1981) NOBS	SECOND SERIES (1979-1981) NVALS	P LEVEL	STEP
P70300	SOLIDS, RESIDUE AT 180 DEG. C DISSOLVED	49	47	1979.40	28	27	0.526	9.000
P00410	ALKALINITY FIELD (MG/L AS CACO3)	49	47	1979.40	17	16	0.662	2.000
P00630	NITROGEN, NO2+NO3 TOTAL (MG/L AS N)	49	47	1979.40	27	25	0.009	.3000
P00665	PHOSPHORUS, TOTAL (MG/L AS P)	49	47	1979.40	27	26	0.430	.1000E-01
P00915	CALCIUM DISSOLVED (MG/L AS CA)	49	47	1979.40	28	27	0.566	-1.000
P00925	MAGNESIUM, DISSOLVED (MG/L AS MG)	49	47	1979.40	28	27	0.652	.0
P00930	SODIUM, DISSOLVED (MG/L AS NA)	49	47	1979.40	28	27	0.030	-1.100
P00940	CHLORIDE, DISSOLVED (MG/L AS CL)	49	47	1979.40	28	27	0.275	-1.000
P00945	SULFATE DISSOLVED (MG/L AS SO4)	49	47	1979.40	28	27	0.000	-5.000
P00950	FLUORIDE, DISSOLVED (MG/L AS F)	49	47	1979.40	28	27	0.260	.0

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

STATION IDENTIFICATION NUMBER=03374100

MANN-WHITNEY-WILCOXAN RANK SUM TEST FOR SEASONAL DATA

VARIABLE		FIRST SERIES (1973-1979) NOBS	NVALS	CUTOFF DATE	SECOND SERIES (1979-1981) NOBS	NVALS	P LEVEL	STEP
P70300	SOLIDS, RESIDUE AT 180 DEG. C DISSOLVED	70	69	1979.40	28	27	0.166	-14.50
P00410	ALKALINITY FIELD (MG/L AS CACO3)	70	69	1979.40	16	15	0.211	-14.00
P00630	NITROGEN, NO2+NO3 TOTAL (MG/L AS N)	65	64	1979.40	28	27	0.001	.5000
P00665	PHOSPHORUS, TOTAL (MG/L AS P)	64	63	1979.40	28	27	0.654	.1000E-01
P00915	CALCIUM DISSOLVED (MG/L AS CA)	70	69	1979.40	28	27	0.578	-2.000
P00925	MAGNESIUM, DISSOLVED (MG/L AS MG)	70	69	1979.40	28	27	0.245	-2.000
P00930	SODIUM, DISSOLVED (MG/L AS NA)	70	69	1979.40	28	27	0.578	.6000
P00940	CHLORIDE, DISSOLVED (MG/L AS CL)	70	69	1979.40	28	27	0.608	1.000
P00945	SULFATE DISSOLVED (MG/L AS SO4)	70	69	1979.40	28	27	0.006	-7.000
P00950	FLUORIDE, DISSOLVED (MG/L AS F)	70	69	1979.40	28	27	0.746	.0

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

STATION IDENTIFICATION NUMBER=03378500

MANN-WHITNEY-WILCOXAN RANK SUM TEST FOR SEASONAL DATA

VARIABLE		FIRST SERIES (1975-1979) NOBS	NVALS	CUTOFF DATE	SECOND SERIES (1979-1981) NOBS	NVALS	P LEVEL	STEP
P70300	SOLIDS, RESIDUE AT 180 DEG. C DISSOLVED	47	46	1979.40	25	24	0.788	-3.500
P00410	ALKALINITY FIELD (MG/L AS CACO3)	47	46	1979.40	16	16	0.279	-10.00
P00630	NITROGEN, NO2+NO3 TOTAL (MG/L AS N)	46	46	1979.40	29	26	0.168	.3000
P00665	PHOSPHORUS, TOTAL (MG/L AS P)	46	46	1979.40	29	25	0.133	.3000E-01
P00915	CALCIUM DISSOLVED (MG/L AS CA)	45	45	1979.40	27	26	0.842	-1.000
P00925	MAGNESIUM, DISSOLVED (MG/L AS MG)	46	45	1979.40	27	26	0.139	-2.000
P00930	SODIUM, DISSOLVED (MG/L AS NA)	46	45	1979.40	27	26	0.547	-1.000
P00940	CHLORIDE, DISSOLVED (MG/L AS CL)	47	46	1979.40	29	26	0.512	-1.000
P00945	SULFATE DISSOLVED (MG/L AS SO4)	47	46	1979.40	27	26	0.089	-8.000
P00950	FLUORIDE, DISSOLVED (MG/L AS F)	47	46	1979.40	27	26	0.352	.0

NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS IN THE ORIGINAL DATA.
NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES CONSTRUCTED.

Appendix C: Source code for SAS procedures

Two separate programs are required to add a procedure to the Statistical Analysis System at a user's installation, a parsing module and a procedure module. The parsing module acts as a control program for the procedure and defines permissible options and parameters for the procedure. The procedure module inputs the appropriate data and computes the test statistic. The parsing module for Proc SEASKEN is given in figure C-1. The procedure module is given in figure C-2. The parsing module for Proc SEASRS is given in figure C-3. The procedure module is given in figure C-4. Information about SAS parsing and procedure modules can be found in the SAS Programmer's Guide (SAS Institute, Inc., 1981a). Only minimal JCL is shown since JCL is highly installation dependent.

```
//C1 JOB
// CLASS
// EXEC ASMSAS
//ASM.SYSIN DD *
         PRINT   NOGEN
PROC      SASPROC NAME=SEASKEN,LOADMOD=SEASKEN2,DEFLIST=1,
               DEFMODE=NUMERIC
PARMS     SASLIST  DL1,1,DL2,2,DL3,3,DL4,4,DL5,5,MODE=NUMERIC
PARMS     SASLIST  DL6,6,DL7,7,DL8,8,DL9,9,DL10,10,MODE=NUMERIC
PARMS     SASLIST  DL11,11,DL12,12,DL13,13,DL14,14,MODE=NUMERIC
PARMS     SASLIST  DL15,15,SEASON,16,MODE=NUMERIC
LISTS     SASLIST  VARIABLES,1,VAR,1,MODE=NUMERIC
          SASEND
          END
//LKED.SYSIN DD *
 SETSSI AC000003
NAME SEASKEN(R)
/*
//
```

Figure C1.--Parsing module for the Seasonal Kendall test and slope estimator procedure.

```
//C2 JOB
// CLASS
//EXEC FORTSAS
//C.SYSIN DD *
      COMMON /STATS/ NV,N,NOBS(15),NVALS(15),NDATES(2),TAU(15),
     &   ALPHA(15),SLOPE(15)
C     NOTIFY SAS.
      CALL SASFHX
C     PROCESS THE DATA.
C     GET NUMBER OF VARIABLES FROM SAS.
      NV1=NOVAR(1)
C     NV IS THE NUMBER OF VARIABLES EXCLUSIVE OF THE TIME VARIABLE.
      NV=NV1-1
      IF (NV.GT.0) GO TO 10
      CALL VARERR
      STOP
   10 CONTINUE
C     READ IN THE DATA.
      CALL READ
C     COMPUTE THE TEST STATISTICS.
      CALL SEAS
C     WRITE OUT RESULTS.
      CALL OUTPUT
      STOP
      END
C     SUBROUTINE TO READ IN THE OBSERVATIONS AND FORM THE SEASONAL VALUES.
      SUBROUTINE READ
      COMMON /STATS/ NV,N,NOBS(15),NVALS(15),NDATES(2),TAU(15),
     &   ALPHA(15),SLOPE(15)
      COMMON /DATA/ X(1200,15)
      COMMON /IOUNIT/ IPRINT
      LOGICAL*1 SETUP,HAVE(1200),EVEN
      REAL*8 PARM,BASETM,PERYR,Y(16)
      INTEGER L(15)
      REAL SORT(400,15)
      COMMON /LABCM/ NAME,LABEL
      COMMON /NAMECM/ NTYPE,NPOS,NLNG,NVARO,NNAME,NLABEL,NFORM,NIFORM,
     XNFL,NFD,NF,NJUST
      INTEGER*2 NTYPE,NPOS,NLNG,NVARO,NFL,NFD,NF,NJUST
      REAL*8 NNAME,NLABEL(5),NFORM,NIFORM
      REAL*8 NAME(15),LABEL(5,15)
      REAL XCHEK/-1.0E30/,XMISS/-1.0E31/
      M=0
      KOBS=0
C     GET SPECIFICATION OF EACH VARIABLE.
      DO 10 I=1,NV
C     CALL SAS TO LOAD COMMON BLOCK.
      CALL NAMEV(1,I+1,NTYPE)
C     SAVE NAME.
      NAME(I)=NNAME
```

Figure C2.--Procedure module for the Seasonal Kendall test and slope estimator procedure.

```
C       SAVE LABEL.
        DO 11 J=1,5
        LABEL(J,I)=NLABEL(J)
     11 CONTINUE
        NOBS(I)=0
        NVALS(I)=0
     10 CONTINUE
        SETUP=.FALSE.
C       CYCLE TO GET EACH OBSERVATION.
     20 CALL INPUT(IEND)
C       WRAP UP IF ALL OBSERVATIONS HAVE BEEN PROCESSED.
        IF(IEND.EQ.1)GO TO 40
C       GET AN OBSERVATION.
        KOBS=KOBS+1
        CALL VARX(1,Y)
        IF(SETUP) GO TO 80
C       FIRST OBSERVATION, GET READY.
C       CALCULATE TIME OF OCT. 1 OF WATER YEAR OF FIRST OBSERVATION.
        YEAR=Y(1)+0.25
        BASETM=IFIX(YEAR)-0.25
        NDATES(1)=YEAR
C       GET NUMBER OF SEASONS.
        PERYR=PARM(16)
        DO 70 I=1,1200
        HAVE(I)=.FALSE.
     70 CONTINUE
        LAST=1
        SETUP=.TRUE.
     80 CONTINUE
C       PROCESS OBSERVATION.
        M=(Y(1)-BASETM)*PERYR+1.0
        IF (M.GE.LAST) GO TO 200
        WRITE (IPRINT,9001) KOBS
   9001 FORMAT (' OBSERVATION NUMBER',I10,
      &    ' IS OUT OF CHRONOLOGICAL ORDER.')
        STOP
    200 CONTINUE
        IF (M.LE.1200) GO TO 210
        WRITE (IPRINT,9002) KOBS
   9002 FORMAT (' OBSERVATION NUMBER',I10,' IS LATER THAN SEASON 1200.')
        STOP
    210 CONTINUE
C       MULTIPLE OBSERVATIONS?
        IF (HAVE(M)) GO TO 90
C       NO. SAVE.
        HAVE(M)=.TRUE.
        K=2
        LAST=M
        DO 30 I=1,NV
C       PREPARE FOR MULTIPLE OBSERVATIONS.
```

```
      VAL=FIXMIS(Y(I+1))
      L(I)=0
      IF (VAL.GT.XCHEK) L(I)=1
      IF (VAL.GT.XCHEK) NOBS(I)=NOBS(I)+1
      IF (VAL.GT.XCHEK) NVALS(I)=NVALS(I)+1
      X(M,I)=VAL
      SORT(1,I)=VAL
   30 CONTINUE
      GO TO 20
C     MULTIPLE OBSERVATIONS
   90 CONTINUE
      DO 110 I=1,NV
      IF (K.EQ.2.AND.L(I).EQ.1) NVALS(I)=NVALS(I)-1
      VAL=FIXMIS(Y(I+1))
      IF (VAL.LT.XCHEK) GO TO 110
      NOBS(I)=NOBS(I)+1
      L(I)=L(I)+1
      SORT(L(I),I)=VAL
  110 CONTINUE
      CALL INPUT(IEND)
C     CYCLE UNTIL NEW SEASON.
      IF (IEND.EQ.1) GO TO 120
      KOBS=KOBS+1
      CALL VARX(1,Y)
      NEWM=(Y(1)-BASETM)*PERYR+1.0
      IF (NEWM.NE.M) GO TO 120
      K=K+1
C     LIMIT TO 400 OBSERVATIONS, MISSING OR NOT, IN ONE SEASON.
      IF (K.LE.400) GO TO 90
      WRITE (IPRINT,9005) KOBS
 9005 FORMAT (' AT OBSERVATION NUMBER,'I10,
     &   ' THERE WERE MORE THAN 400 OBSERVATIONS IN THAT SEASON.')
      STOP
C     GET MEDIANS.
  120 CONTINUE
      DO 130 I=1,NV
C     ANY NON-MISSING OBSERVATIONS?
      IF (L(I).EQ.0) GO TO 130
C     YES.
      NVALS(I)=NVALS(I)+1
      IH=(L(I)+1)/2
      EVEN=2*IH.EQ.L(I)
C     FIND MEDIAN.
      CALL VSRTA(SORT(1,I),L(I))
      IF (.NOT.EVEN) X(M,I)= SORT(IH,I)
      IF (EVEN)      X(M,I)=(SORT(IH,I)+SORT(IH+1,I))/2.0
  130 CONTINUE
      LAST=M
C     CONTINUE PROCESSING IF END-OF-FILE NOT ENCOUNTERED.
      IF (IEND.NE.1) GO TO 80
```

```
   40 IF (M.GT.1) GO TO 50
      WRITE (IPRINT,9003)
 9003 FORMAT (' THERE ARE FEWER THAN 2 SEASONS.')
      STOP
   50 CONTINUE
      N=LAST
      NDATES(2)=Y(1)+0.25
      IF (LAST*(LAST/PERYR+1.0).LE.60000.0) GO TO 60
      WRITE (IPRINT,9004)
 9004 FORMAT (' THE COMBINATION OF YEARS AND SEASONS EXCEEDS',
     &   ' AN INTERNAL LIMIT.')
      STOP
   60 CONTINUE
      DO 150 M=1,LAST
      IF (HAVE(M)) GO TO 150
      DO 140 I=1,NV
      X(M,I)=XMISS
  140 CONTINUE
  150 CONTINUE
      RETURN
      END
C     PERFORM KENDALL TEST.
      SUBROUTINE SEAS
      COMMON /STATS/ NV,N,NOBS(15),NVALS(15),NDATES(2),TAU(15),
     &   ALPHA(15),SLOPE(15)
      COMMON /DATA/ X(1200,15)
      REAL Y(60000)
      REAL*8 PARM
      REAL XCHEK/-1.0E30/
      LOGICAL ODD
      LOGICAL*1 WASTIE(1200)
      NPER=PARM(16)
      PERYR=NPER
      DO 1 J=1,NV
C     BYPASS DETECTION LIMIT PROCESSING IF NONE SPECIFIED.
      IF (PARM(J).EQ.0.0D0) GO TO 50
      DLIMIT=PARM(J)
      DO 100 I=1,N
      IF (X(I,J).LE.DLIMIT.AND.X(I,J).GT.0.0) X(I,J)=0.5*DLIMIT
  100 CONTINUE
   50 CONTINUE
C     ZERO OUT THE COUNTERS.
      DO 60 I=1,N
      WASTIE(I)=.FALSE.
   60 CONTINUE
      NPLUS=0
      NMINUS=0
      NCOMPT=0
      VARTOT=0.0
      INDEX=0
```

```
      FIXVAR=0.0
C     LOOP THROUGH THE SEASONS.
      DO 10 ISEAS=1,NPER
      NCOMP=0
      N1=N-NPER
C     LOOP THROUGH YEARS FOR VALUES IN SEASON ISEAS.
      DO 20 ISTART=ISEAS,N1,NPER
C     VALID VALUES?
      IF  (X(ISTART,J).LE.XCHEK) GO TO 20
C     VALUE IS ALWAYS TIED WITH ITSELF.
      NTIE=1
      N2=ISTART+NPER
C     TRY EACH LATER SEASON.
      DO 30 IEND=N2,N,NPER
C     VALID VALUE?
      IF (X(IEND,J).LE.XCHEK) GO TO 30
C     COMPARE.
      NCOMP=NCOMP+1
      INDEX=INDEX+1
      YY=(X(IEND,J)-X(ISTART,J))/((IEND-ISTART)/PERYR)
      IF (YY.GT.0.0) NPLUS=NPLUS+1
      IF (YY.LT.0.0) NMINUS=NMINUS+1
      IF (YY.EQ.0.0) NTIE=NTIE+1
C     MARK VALUES THAT ARE TIED.
      IF (YY.EQ.0.0) WASTIE(IEND)=.TRUE.
C     SAVE ADJUSTED DIFFERENCES.
      Y(INDEX)=YY
   30 CONTINUE
C     UPDATE VARIANCE CORRECTION IF TIES OCCURED AND TIES
C        WERE NOT COUNTED BEFORE.
      IF (NTIE.NE.1.AND..NOT.WASTIE(ISTART)) FIXVAR=FIXVAR+NTIE*(NTIE-
     &   1.0)*(2.0*NTIE+5.0)/18.0
   20 CONTINUE
C     ACCUMULATE THIS MONTH'S RESULTS.
      NCOMPT=NCOMPT+NCOMP
      NMONTH=(1.0+SQRT(1.0+8.0*NCOMP))/2.0
      VARTOT=VARTOT+NMONTH*(NMONTH-1.0)*(2.0*NMONTH+5.0)/18.0
   10 CONTINUE
C     DONE COMPARING.
      S=NPLUS-NMINUS
      VARTOT=VARTOT-FIXVAR
C     WERE THERE ANY VALID COMPARISONS AND AT LEAST TWO DIFFERENT VALUES?
      IF (NCOMPT.GT.0.AND.VARTOT.GT.0.0) GO TO 40
C     NONE. GO HOME EMPTY.
      TAU(J)=0.0
      ALPHA(J)=1.0
      SLOPE(J)=0.0
      GO TO 1
   40 CONTINUE
C     CALCULATE THE STATISTICS.
```

```
      TAU(J)=S/NCOMPT
C     CONTINUITY CORRECTION.
      IF (S.GT.0.0) S=S-1
      IF (S.LT.0.0) S=S+1
      Z=S/SQRT(VARTOT)
C     COMPARE TO THE STANDARD NORMAL DISTRIBUTION.
      IF (Z.LE.0.0) ALPHA(J)=2.0*CDFN(Z)
      IF (Z.GT.0.0) ALPHA(J)=2.0*(1.0-CDFN(Z))
C     SORT THE DIFFERENCES. VSRTA IS AN IMSL ROUTINE.
      CALL VSRTA(Y,INDEX)
C     PICK THE MEDIAN.
      ODD=MOD(INDEX,2).EQ.1
      IF (ODD) YMED=Y((INDEX+1)/2)
      IF (.NOT.ODD) YMED=0.5*(Y(INDEX/2)+Y((INDEX/2)+1))
      SLOPE(J)=YMED
      IF (SLOPE(J).NE.0.0) GO TO 1
C     ADJUST FOR THE FACT THAT TAU AND ALPHA MAY SAY THERE IS A
C        SIGNIFICANT TREND BUT THE ESTIMATE OF THE SLOPE IS
C        ZERO DUE TO TIES.
      IF (NMINUS.GT.NPLUS) SLOPE( J )=-1.0E-30
      IF (NMINUS.LT.NPLUS) SLOPE( J )=1.0E-30
    1 CONTINUE
      RETURN
      END
C     REPORT RESULTS.
      SUBROUTINE OUTPUT
      COMMON /LABCM/ NAME,LABEL
      REAL*8 NAME(15),LABEL(5,15)
      COMMON /STATS/ NV,N,NOBS(15),NVALS(15),NDATES(2),TAU(15),
     &   ALPHA(15),SLOPE(15)
      COMMON /IOUNIT/ IPRINT
C     STANDARD SAS HEADING.
      CALL STITLE(0,LINES)
      WRITE (IPRINT,6001) NDATES
 6001 FORMAT (/26X,'SEASONAL KENDALL TEST AND SLOPE ESTIMATOR FOR',
     &    ' TREND MAGNITUDE'//40X,'WATER YEARS ',I4,'-',I4//)
      WRITE (IPRINT,6002)
 6002 FORMAT (' VARIABLE',50X,'NOBS      NVALS      TAU      P',
     &    ' LEVEL      SLOPE'/1X,107('-')/)
      DO 10 I=1,NV
      WRITE (IPRINT,6003)  NAME(I),(LABEL(J,I),J=1,5),NOBS(I),NVALS(I),
     &   TAU(I),ALPHA(I),SLOPE(I)
 6003 FORMAT (1X,A8,3X,5A8,2I10,2F10.3,G15.4/)
   10 CONTINUE
      WRITE (IPRINT,6004)
 6004 FORMAT (/' NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS',
     &    ' IN THE ORIGINAL DATA.'/
     &    ' NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES',
     &    ' CONSTRUCTED.')
      RETURN
```

```
      END
C     CUMULATIVE DISTRIBUTION FUNCTION FOR THE NORMAL DISTRIBUTION.
C        PRIMARY REFERENCE IS ABRAMOWITZ AND STEGUN,
C        NBS HANDBOOK OF MATHEMATICAL FUNCTIONS, EQ. 26.2.19.
      FUNCTION CDFN(X)
      IF (X) 10,20,30
   10 CONTINUE
      IF (X.LT.-6.0) GO TO 40
      T=-X
      CDFN=0.5/(1.0+0.0498673470*T+0.0211410061*T**2+0.0032776263*T**3
     X+0.380036E-4*T**4+0.488906E-4*T**5+0.53830E-5*T**6)**16
      RETURN
   20 CONTINUE
      CDFN=0.5000
      RETURN
   30 CONTINUE
      IF (X.GT.6.0) GO TO 50
      CDFN=1.0-0.5/(1.0+0.0498673470*X+0.0211410061*X**2+0.0032776263*X
     X**3+0.380036E-4*X**4+0.488906E-4*X**5+0.53830E-5*X**6)**16
      RETURN
   40 CONTINUE
      CDFN=0.0000
      RETURN
   50 CONTINUE
      CDFN=1.0000
      RETURN
      END
C     CONVERT A SAS MISSING VALUE TO ONE RECOGNIZED BY THESE ROUTINES.
      FUNCTION FIXMIS(X)
      REAL*8 X,VALUE
      EQUIVALENCE(VALUE,MISSX)
      VALUE=X
      IF (VALUE.EQ.0.0D0.AND.MISSX.NE.0) VALUE=-1.0D31
      FIXMIS=VALUE
      RETURN
      END
//LKED.SYSIN DD *
 ENTRY ENTRY    THIS STATEMENT MUST BE PRESENT
 SETSSI AD000000
 NAME SEASKEN2(R)
/*
//
```

```
//C3 JOB
// CLASS
// EXEC ASMSAS
//ASM.SYSIN DD *
         PRINT   NOGEN
PROC      SASPROC NAME=SEASRS,LOADMOD=SEASRS2,DEFLIST=1,
               DEFMODE=NUMERIC
PARMS     SASLIST  DL1,1,DL2,2,DL3,3,DL4,4,DL5,5,MODE=NUMERIC
PARMS     SASLIST  DL6,6,DL7,7,DL8,8,DL9,9,DL10,10,MODE=NUMERIC
PARMS     SASLIST  DL11,11,DL12,12,DL13,13,DL14,14,MODE=NUMERIC
PARMS     SASLIST  DL15,15,SEASON,16,DATE,17,MODE=NUMERIC
LISTS     SASLIST  VARIABLES,1,VAR,1,MODE=NUMERIC
          SASEND
          END
//LKED.SYSIN DD *
 SETSSI AC000003
 NAME SEASRS(R)
/*
//
```

Figure C3.--Parsing module for the Seasonal Mann-Whitney-Wilcoxon rank sum test and slope estimator procedure.

```
//C4 JOB
// CLASS
// EXEC FORTSAS
//C.SYSIN DD *
C      NOTIFY SAS.
       CALL SASFHX
C      PROCESS THE DATA.
C      GET NUMBER OF VARIABLES FROM SAS.
       NV1=NOVAR(1)
C      NV IS THE NUMBER OF VARIABLES EXCLUSIVE OF THE TIME VARIABLE.
       NV=NV1-1
       IF (NV.GT.0) GO TO 10
       CALL VARERR
       STOP
   10 CONTINUE
C      READ IN DATA.
       CALL READ (NV1)
       DO 20 J=1,NV
C      COMPUTE RANK SUM FOR EACH VARIABLE.
       CALL SEASRS (J)
   20 CONTINUE
C      WRITE OUT RESULTS.
       CALL OUTPUT (NV)
       STOP
       END
C      SUBROUTINE TO READ IN THE OBSERVATIONS AND FORM THE SEASONAL VALUES.
       SUBROUTINE READ (NV1)
       COMMON /DATA/ X(1200,15),NX1,NX2,ALPHA(15),DF(15),CUTOFF,
     &   NOBS1(15),NOBS2(15),NVALS1(15),NVALS2(15),NDATES(4)
       REAL*8 CUTOFF
       COMMON /LABINF/ NAME(15),LABEL(5,15)
       INTEGER NOBS(15),NVALS(15)
       DIMENSION L(15)
       COMMON /IOUNIT/ IPRINT
       REAL*8 NAME,LABEL,Y(16),PARM,BASETM,PERYR
       REAL SORT(400,15)
       LOGICAL*1 HAVE(1200),SETUP,SECOND,EVEN
       COMMON /NAMECM/ NTYPE,NPOS,NLNG,NVARO,NNAME,NLABEL,NFORM,NIFORM,
     &   NFL,NFD,NF,NJUST
       INTEGER*2 NTYPE,NPOS,NLNG,NVARO,NFL,NFD,NF,NJUST
       REAL*8 NNAME,NLABEL(5),NFORM,NIFORM
       REAL XCHEK/-1.0E30/,XMISS/-1.0E31/
       M=0
       KOBS=0
       NV=NV1-1
C      GET SPECIFICTION OF EACH VARIABLE.
       DO 10 I=1,NV
C      CALL SAS TO LOAD COMMON BLOCK.
       CALL NAMEV(1,I+1,NTYPE)
C      SAVE NAME.
```

Figure C4.--Procedure module for the Seasonal Mann-Whitney-Wilcoxon rank sum test and slope estimator procedure.

```
      NAME(I)=NNAME
C     SAVE LABEL.
      DO 11 J=1,5
      LABEL(J,I)=NLABEL(J)
   11 CONTINUE
      NOBS(I)=0
      NVALS(I)=0
   10 CONTINUE
      SETUP=.FALSE.
      SECOND=.FALSE.
      MBASE=1
C     CYCLE TO GET EACH OBSERVATION.
   20 CALL INPUT(IEND)
C     WRAP UP IF ALL OBSERVATIONS HAVE BEEN PROCESSED.
      IF (IEND.EQ.1) GO TO 40
C     GET AN OBSERVATION.
      KOBS=KOBS+1
      CALL VARX(1,Y)
      IF (SETUP) GO TO 70
C     FIRST OBSERVATION, GET READY.
C     CALCULATE TIME OF OCT. 1 OF WATER YEAR OF FIRST OBSERVATION.
      YEAR=Y(1)+0.25
      NDATES(1)=YEAR
      BASETM=IFIX(YEAR)-0.25
C     GET NUMBER OF SEASONS.
      PERYR=PARM(16)
C     GET CUTOFF DATE.
      CUTOFF=PARM(17)
      DO 60 I=1,1200
      HAVE(I)=.FALSE.
   60 CONTINUE
      LAST=1
      SETUP=.TRUE.
C     PROCESS OBSERVATION.
   70 CONTINUE
      IF (SECOND.OR.Y(1).LT.CUTOFF) GO TO 80
C     FIRST OBSERVATION AFTER CUTOFF DATE.
      NDATES(2)=YEAR
      SECOND=.TRUE.
      MBASE=LAST+1
      NX1=LAST
      DO 75 I=1,NV
      NOBS1(I)=NOBS(I)
      NOBS(I)=0
      NVALS1(I)=NVALS(I)
      NVALS(I)=0
   75 CONTINUE
      YEAR=Y(1)+0.25
      BASETM=IFIX(YEAR)-0.25
      NDATES(3)=YEAR
```

```
   80 CONTINUE
      YEAR=Y(1)+0.25
      M=(Y(1)-BASETM)*PERYR+MBASE
      IF (M.GE.LAST) GO TO 90
      WRITE (IPRINT,9001) KOBS
 9001 FORMAT (' OBSERVATION NUMBER',I10,
     &   ' IS OUT OF CHRONOLOGICAL ORDER.')
      STOP
   90 CONTINUE
      IF (M.LE.1200) GO TO 100
      WRITE (IPRINT,9002) KOBS
 9002 FORMAT (' OBSERVATION NUMBER',I10,' IS LATER THAN SEASON 1200.')
      STOP
  100 CONTINUE
C     MULTIPLE OBSERVATIONS?
      IF (HAVE(M)) GO TO 110
C     NO. SAVE.
      HAVE(M)=.TRUE.
      K=2
      LAST=M
      DO 30 I=1,NV
C     PREPARE FOR POSSIBLE MULTIPLE OBSERVATIONS.
      VAL=FIXMIS(Y(I+1))
      L(I)=0
      IF (VAL.GT.XCHEK) L(I)=1
      IF (VAL.GT.XCHEK) NOBS(I)=NOBS(I)+1
      IF (VAL.GT.XCHEK) NVALS(I)=NVALS(I)+1
      SORT(1,I)=VAL
   30 X(M,I)=VAL
      GO TO 20
C     MULTIPLE OBSERVATIONS
  110 CONTINUE
      DO 130 I=1,NV
      IF (K.EQ.2.AND.L(I).EQ.1) NVALS(I)=NVALS(I)-1
      VAL=FIXMIS(Y(I+1))
      IF (VAL.LT.XCHEK) GO TO 130
      NOBS(I)=NOBS(I)+1
      L(I)=L(I)+1
      SORT(L(I),I)=VAL
  130 CONTINUE
      CALL INPUT(IEND)
C     CYCLE UNTIL NEW SEASON.
      IF (IEND.EQ.1) GO TO 140
      KOBS=KOBS+1
      CALL VARX(1,Y)
      NEWM=(Y(1)-BASETM)*PERYR+MBASE
      IF (NEWM.NE.M) GO TO 140
      K=K+1
C     LIMIT TO 400 OBSERVATION, MISSING OR NOT, IN ONE SEASON.
      IF (K.LE.400) GO TO 110
```

```
          WRITE (IPRINT,9005) KOBS
     9005 FORMAT (' AT OBSERVATION NUMBER',I10,
         &   ' THERE WERE MORE THAN 400 OBSERVATIONS IN THAT SEASON.')
          STOP
C         GET MEDIANS
      140 CONTINUE
          DO 150 I=1,NV
C         ANY NON-MISSING OBSERVATIONS?
          IF (L(I).EQ.0) GO TO 150
C         YES.
          NVALS(I)=NVALS(I)+1
          IH=(L(I)+1)/2
          EVEN=2*IH.EQ.L(I)
C         FIND MEDIAN.
          CALL VSRTA(SORT(1,I),L(I))
          IF (.NOT.EVEN) X(M,I)= SORT(IH,I)
          IF (EVEN)      X(M,I)=(SORT(IH,I)+SORT(IH+1,I))/2.0
      150 CONTINUE
          LAST=M
C         CONTINUE PROCESSING IF END-OF-FILE NOT ENCOUNTERED.
          IF (IEND.NE.1) GO TO 70
       40 IF (M.GT.1) GO TO 50
          WRITE (IPRINT,9003)
     9003 FORMAT (' THERE ARE FEWER THAN 2 SEASONS.')
          STOP
       50 CONTINUE
          IF (SECOND) GO TO 160
          WRITE (IPRINT,9004)
     9004 FORMAT (' THERE ARE NO OBSERVATIONS AFTER TJHE CUTOFF DATE.')
          STOP
      160 CONTINUE
          NDATES(4)=Y(1)+0.25
          NX2=LAST-NX1
          DO 170 I=1,NV
          NOBS2(I)=NOBS(I)
          NVALS2(I)=NVALS(I)
      170 CONTINUE
C         SET UNUSED CELLS TO MISSING.
          DO 210 M=1,LAST
          IF (HAVE(M)) GO TO 210
          DO 200 I=1,NV
          X(M,I)=XMISS
      200 CONTINUE
      210 CONTINUE
          RETURN
          END
C         PERFORM RANK SUM TEST.
          SUBROUTINE SEASRS (J)
          COMMON /DATA/ X(1200,15),NX1,NX2,ALPHA(15),DF(15),CUTOFF,
         &   NOBS1(15),NOBS2(15),NVALS1(15),NVALS2(15),NDATES(4)
```

```
      REAL*8 CUTOFF
      DIMENSION Y(1200),R(1200)
      REAL*8 PARM
      REAL XCHEK/-1.0E30/
      IF (NVALS1(J).EQ.0.OR.NVALS2(J).EQ.0) GO TO 700
      NPER=PARM(16)
      WS=0.0
      EW=0.0
      VW=0.0
      LAST=NX1+NX2
C     BYPASS DETECTION LIMIT PROCESSING IF NONE SPECIFIED.
      IF (PARM(J).EQ.0.0D0) GO TO 20
      DLIMIT=PARM(J)
      DO 10 I=1,LAST
      IF ( X(I,J).LE.DLIMIT.AND.X(I,J).GT.0.0) X(I,J)=0.5*DLIMIT
   10 CONTINUE
   20 CONTINUE
C     LOOP THROUGH THE SEASONS.
      DO 500 IM=1,NPER
C     LOOP THROUGH YEARS FOR VALUES IN SEASON IM.
C     FIRST PERIOD.
      NI = 0
      DO 60 I=IM,NX1,NPER
      XX=X(I,J)
      IF(XX.LT.XCHEK) GO TO 60
      NI = NI + 1
      Y(NI) = XX
   60 CONTINUE
C     SKIP SEASONS IF NO VALUES.
      IF(NI.EQ.0) GO TO 500
C     SECOND PERIOD.
      MI = 0
      DO 70 I=IM,NX2,NPER
      XX=X(I+NX1,J)
      IF(XX.LT.XCHEK) GO TO 70
      MI = MI + 1
      Y(MI+NI) = XX
   70 CONTINUE
C     SKIP SEASON IF NO VALUES.
      IF (MI.EQ.0) GO TO 500
      NN = MI + NI
      CALL RANK(Y,R,NN)
      RSUM = 0.0
      DO 130 I = 1, NI
      RSUM = RSUM + R(I)
  130 CONTINUE
      RSSQ=0.0
      DO 140 I=1,NN
      RSSQ=RSSQ+R(I)**2
  140 CONTINUE
```

```
C      ACCUMULATE STATISTICS.
       WS=WS+RSUM
       EW=EW+NI*(NN+1)/2.0
       VW=VW+MI*NI*(RSSQ/NN-(NN+1)**2/4.0)/(NN-1)
  500 CONTINUE
C      CLACULATE FINAL STATISTICS.
       IF(VW.LE.0.0) GO TO 700
       Z=(-(WS-EW)/SQRT(VW))
       IF (Z.LE.0.0) ALPHA(J)=2.0*CDFN(Z)
       IF (Z.GT.0.0) ALPHA(J)=2.0*(1.0-CDFN(Z))
C      GET ESTIMATE OF JUMP.
       CALL SEARSD(X(1,J),NX1,X(NX1+1,J),NX2,NPER,DF(J))
       RETURN
  700 CONTINUE
       ALPHA(J)=1.0
       DF(J)=0.0
       RETURN
       END
C      CALCULATE AN ESTIMATE OF THE STEP TREND.
       SUBROUTINE SEARSD(X1,NX1,X2,NX2,NPER,D)
       COMMON /IOUNIT/ IPRINT
       REAL X1(NX1),X2(NX2)
       REAL Y(12000)
       REAL XCHEK/-1.0E30/
       LOGICAL EVEN
       NY1=((NX1-1)/NPER)+1
       NY2=((NX2-1)/NPER)+1
       NMAX=NY1*NY2*12
       IF (NMAX.LE.12000) GO TO 10
       WRITE (IPRINT,9001) NX1,NX2
 9001 FORMAT (' THE Y ARRAY IN SEARSD IS TOO SMALL.'/
     &    'NX1=',I6,'NX2=',I6)
       STOP
   10 CONTINUE
C      CALCULATE THE DIFFERENCE BETWEEN EACH PAIR OF VALUES - ONE IN
C         THE FIRST SERIES, THE OTHER OF THE SECOND, AND BOTH OF THE
C         SAME SEASON.
       K=0
       DO 40 IM=1,NPER
       DO 30 I=IM,NX1,NPER
       XF=X1(I)
       IF (XF.LT.XCHEK) GO TO 30
       DO 20 J=IM,NX2,NPER
       XL=X2(J)
       IF (XL.LT.XCHEK) GO TO 20
       K=K+1
       Y(K)=XL-XF
   20 CONTINUE
   30 CONTINUE
   40 CONTINUE
```

```
C       GET MEDIAN.
        CALL VSRTA(Y,K)
        IH=(K+1)/2
        EVEN=2*IH.EQ.K
        IF (EVEN)       D=(Y(IH)+Y(IH+1))/2.0
        IF (.NOT.EVEN)  D=Y(IH)
        RETURN
        END
C       REPORT RESULTS.
        SUBROUTINE OUTPUT (NV)
        COMMON /DATA/ X(1200,15),NX1,NX2,ALPHA(15),DF(15),CUTOFF,
     &    NOBS1(15),NOBS2(15),NVALS1(15),NVALS2(15),NDATES(4)
        REAL*8 CUTOFF
        COMMON /LABINF/ NAME(15),LABEL(5,15)
        REAL*8 NAME,LABEL
        COMMON /IOUNIT/ IPRINT
C       STANDARD SAS TITLE.
        CALL STITLE (0,LINES)
        WRITE (IPRINT,6001)
   6001 FORMAT (/39X,'MANN-WHITNEY-WILCOXAN RANK SUM TEST FOR SEASONAL',
     &    ' DATA'///)
        WRITE (IPRINT,6002) NDATES
   6002 FORMAT (53X,'  FIRST SERIES  ',16X,' SECOND SERIES  '/
     &          52X,'   (',I4,'-',I4,')  ',16X,'  (',I4,'-',I4,')'/
     &    ' VARIABLE',
     &          43X,'   NOBS   NVALS ',' CUTOFF DATE  ',
     &    '   NOBS   NVALS',
     &    5X,'P LEVEL',5X,'STEP'/1X,123('-')/)
        DO 10 I=1,NV
        WRITE (IPRINT,6003) NAME(I),(LABEL(J,I),J=1,5),NOBS1(I),NVALS1(I),
     &    CUTOFF,NOBS2(I),NVALS2(I),ALPHA(I),DF(I)
   6003 FORMAT (1X,A8,3X,5A8,I6,I8,F13.2,I10,I8,F12.3,G15.4/)
     10 CONTINUE
        WRITE (IPRINT,6004)
   6004 FORMAT (//' NOBS IS THE NUMBER OF NON-MISSING OBSERVATIONS ',
     &    'IN THE ORIGINAL DATA.'/
     &    ' NVALS IS THE NUMBER OF NON-MISSING SEASONAL VALUES',
     &    ' CONSTRUCTED.')
        RETURN
        END
C       CUMULATIVE DISTRIBUTION FUNCTION FOR THE NORMAL DISTRIBUTION.
C          PRIMARY REFERENCE IS ABRAMOWITZ AND STEGUN,
C          NBS HANDBOOK OF MATHEMATICAL FUNCTIONS, EQ. 26.2.19.
        FUNCTION CDFN(X)
        IF (X) 10,20,30
     10 CONTINUE
        IF (X.LT.-6.0) GO TO 40
        T=-X
        CDFN=0.5/(1.0+0.0498673470*T+0.0211410061*T**2+0.0032776263*T**3
     &    +0.380036E-4*T**4+0.488906E-4*T**5+0.53830E-5*T**6)**16
```

```
      RETURN
   20 CONTINUE
      CDFN=0.500
      RETURN
   30 CONTINUE
      IF (X.GT.6.0) GO TO 50
      CDFN=1.0-0.5/(1.0+0.0498673470*X+0.0211410061*X**2+0.0032776263*X
     &   **3+0.380036E-4*X**4+0.488906E-4*X**5+0.53830E-5*X**6)**16
      RETURN
   40 CONTINUE
      CDFN=0.000
      RETURN
   50 CONTINUE
      CDFN=1.000
      RETURN
      END
C     CONVERT A SAS MISSING VALUE TO ONE RECOGNIZED BY THESE ROUTINES.
      FUNCTION FIXMIS(X)
      REAL*8 X,VALUE
      EQUIVALENCE(VALUE,MISSX)
      VALUE=X
      IF (VALUE.EQ.0.0D0.AND.MISSX.NE.0) VALUE=-1.0D31
      FIXMIS=VALUE
      RETURN
      END
      SUBROUTINE RANK(X,R,N)
C     COMPUTES THE RANKS OF THE VALUES IN VECTOR X
C     X IS NOT RE-ARRANGED ON RETURN
C     IN CASES OF TIES, MID-RANKS ARE USED
C     MISSING VALUES ARE NOT ALLOWED
      REAL X(N), R(N), Y(1200)
      INTEGER ORD(1200)
C     INITIALIZE ORD AND Y
      DO 10 I = 1, N
      ORD(I) = I
      Y(I) = X(I)
   10 CONTINUE
C     REARRANGE Y IN ASCENDING ORDER
      CALL VSRTR(Y,N,ORD)
C     INITIAL RANKING
      DO 100 I = 1,N
      R(ORD(I)) = I
  100 CONTINUE
C     ADJUST RANKS FOR TIES
      I = 1
C     VALUE ALWAYS TIED WITH ITSELF
      IEND = I
  150 IEND = IEND + 1
      IF (IEND.LE.N.AND.Y(IEND).EQ.Y(I)) GO TO 150
C     COMPUTE AVERAGE RANK
```

```
      AVE = (IEND*(IEND-1)-I*(I-1))/(2.0*(IEND-I))
      IENDM1 = IEND - 1
      DO 190 J = I, IENDM1
      R(ORD(J)) = AVE
  190 CONTINUE
      I = IEND
      IF(I.LT.N) GO TO 150
      RETURN
      END
//LKED.SYSIN DD *
 ENTRY ENTRY    THIS STATEMENT MUST BE PRESENT
 SETSSI AD000000
 NAME SEASRS2(R)
/*
//
```

REFERENCES

Belsley, D. A., Kuh, Edwin, and Welsch, R. E., 1980, Regression diagnostics: New York, John Wiley and Sons, 292 p.

Bradley, J. V., 1968, Distribution-free statistical tests: Englewood Cliffs, N. J., Prentice-Hall, 388 p.

Briggs, J. C., 1978, Nationwide surface water quality monitoring networks of the U.S. Geological Survey, in Establishment of Water-Quality Monitoring Programs, San Francisco, June 1978, Proceedings: Minneapolis, American Water Resources Association, p. 49-57.

Chatterjee, Samprit, and Price, Bertram, 1977, Regression analysis by example: New York, John Wiley and Sons, 462 p.

Conover, W. J., 1971, Practical nonparametric statistics: New York, John Wiley and Sons, 462 p.

Cook, R. D., 1977, Detection of influential observations in linear regression: Technometrics, v. 19, p. 15-18.

Daniel, Cuthbert, and Wood, F. S., 1980, Fitting equations to data, (2nd ed.): New York, John Wiley and Sons, 458 p.

Draper, N. R., and Smith, Harry, 1981, Applied Regression Analysis (2nd ed.) New York, John Wiley and Sons, 709 p.

Fuller, F. C., Jr., and Tsokos, C. P., 1971, Time series analysis of water pollution data: Biometrics, v. 27, p. 1017-1034.

Hirsch, R. M., Slack, J. R., and Smith, R. A., 1982, Techniques of trend analysis for monthly water quality data: Water Resources Research, v. 18, no. 1, p. 107-121.

Hollander, Myles, and Wolfe, D. A., 1973, Nonparametric statistical methods: New York, John Wiley and Sons, 503 p.

Kendall, Maurice, 1975, Rank correlation methods: London, Charles Griffin and Co., Ltd., 202 p.

Lettenmaier, D. P., 1976, Detection of trends in water quality data from records with dependent observations: Water Resources Research, v. 12, no. 5, p. 1037-1046.

SAS Institute, Inc., 1979, SAS users guide, 1979 edition: Cary, N. C., SAS Institute, Inc., 494 p.

SAS Institute, Inc., 1981a, SAS Programmer's guide, 1981 edition, Cary, North Carolina, SAS Institute, Inc. 184 p.

SAS Institute, Inc., 1981b, SAS/GRAPH users guide, 1981 edition: Cary, North Carolina, SAS Institute, Inc., 126 p.

SAS Institute, Inc., 1982a, SAS users guide - basics, 1982 edition: Cary, North Carolina, SAS Institute, Inc., 924 p.

SAS Institute, Inc, 1982b, SAS users guide - statistics, 1982 edition: Cary, North Carolina, SAS Institute, Inc., 584 p.

Smith, R. A., Hirsch, R. M., and Slack, J. R., 1982, A study of trends in total phosphorus measurements at NASQAN stations: U.S. Geological Survey Water-Supply Paper 2190, 34 p.

U.S. Geological Survey - WATSTORE User's Guide

U.S. Geological Survey Computer User's Manual

☆U.S. GOVERNMENT PRINTING OFFICE:1983-421-614/588