A Computerized Method of Precipitation Data Quality Control

Ronald H. Campbell
A COMPUTERIZED METHOD OF PRECIPITATION
DATA QUALITY CONTROL

By

Ronald H. Campbell

Technical Report

Contract No. 14-06-D-6820
Office of Atmospheric Resources
Bureau of Reclamation
U.S. Department of the Interior

Utah Water Research Laboratory
College of Engineering
Utah State University
Logan, Utah

July 1971
ABSTRACT

A computerized data quality check and data edit program has been developed to aid in the processing of telemetered precipitation data. The logic considerations and program development are discussed and results are displayed. This effort has resulted in successful editing of large amounts of precipitation data gathered by automated remote data acquisition techniques. The application of this processing routine has resulted in reducing data editing costs to about three percent of that required to do the work manually.
LIST OF TABLES

Table                                                                                       Page
1  Relative frequency distribution of hourly precipitation values for Silver Lake-Brighton 3
   1957-66 November through April

LIST OF FIGURES

Figure                                                                                       Page
1  Types of questionable data points                                                        2
2  Overshoot correction procedure                                                            4
3  Sample output of a data set with data quality parameters                                 5
4  Final (solid curve) and original (dashed curve) data from Tony Grove Ranger Station gages 6
   602 and 603                                                                            
5  Final (solid curve) and original (dashed curve) data for Gold Hill gage 1001 and Cinnamon Creek gage 1203 7
6  Data sets for Franklin Basin gage 1603 and Hell Canyon gage 1701                        8
7  Final (solid curve) and original (dashed curve) data sets for Dry Bread Pond gage 5004 and Porcupine gage 12501 9
A COMPUTERIZED METHOD OF PRECIPITATION
DATA QUALITY CONTROL

Introduction

The experimental design of the Wasatch Weather Modification Project provides for the evaluation of cloud seeding activities through the comparison of precipitation amounts for seeded and unseeded periods. The present experimental design requires simultaneous half-hourly precipitation data from gages in and near the primary target area for each 8-hour experimental period. These data requirements for meteorological evaluation of the experiment have led to the development of new acquisition and processing techniques of precipitation data.

The requirements for evaluating cloud seeding experiments utilizing precipitation data brought about the construction of the Utah State University telemetering precipitation gage network (Israelsen and Griffin, 1969; Chadwick, 1968, 1969). The readout system for the network can provide as many as three readings per hour for each of 46 stations. Readouts are sequential, and errors are found in the data. Thus, a need arose to develop an objective method for interpolation of simultaneous data points from quality checked data.

The data quality problem results from a combination of several factors concerning the type of gage used, the nature of the transducer, and problems peculiar to radio telemetry, i.e. wind forces on the gage transmitted to the transducer, vibratory motions of the gage caused by the wind, snow and ice sticking or riming developing temporarily on the sides of the collection cans, radio interference, friction between the cans and guides.

With as many as 75,000 data points per year being read out from the network, the computerization of the entire precipitation data processing procedure was highly desirable. Half-hourly precipitation data were obtained in two steps. The first step was the relatively simple conversion of the electronic readout period to cumulative precipitation in the storage gage. The inches water equivalent is obtained through individual calibration curves which are determined each season for each station's transducer.

The second, and more complicated step, was the data quality check, editing, and data interpolation. Before a satisfactory data quality control program was developed, much editing was done by hand. Manual editing served to categorize anomalous data points and provided the basis for development of subsequent objective criteria by which data could be evaluated. These criteria and the computer program logic are discussed as follows.

Categorization of Questionable Data Points

Data points are examined for one station at a time and in sets of at least eight precipitation values. Typical data sets consist of from 40 to 200 points.

Data points considered questionable are categorized into three groups using the criteria outlined as follows. Figure 1 is a graphical display of these types.

1. Precipitation values may be so unreasonable that they are obviously unacceptable. These points are called “outliers” and seem to occur randomly. They occur above and below the cumulative precipitation curve. Radio interference or other electronic problems appear to be the most likely causes for these outliers.

2. Precipitation values may exhibit relatively small departures from the apparent cumulative precipitation curve. These departures also seem to occur randomly and are referred to as “noise” points. The most suspected sources of this type error are wind and friction. The wind may cause a bouncing or vibration of the can and friction may cause a sticking of the can guide. These noise points are usually within .03 inch of a five-point moving average of the cumulative precipitation curve.

3. Several precipitation values in succession may indicate a negative trend in precipitation. These trends cannot be rationalized to be the result of evaporative losses or leaking cans. One explanation is that ice or snow stick temporarily on the outside of the gage. Since this is a weighing gage system, both this accumulation and subsequent melting or sublimation are reflected in the measurements. Temperatures observed during periods in which these negative trends occur support this concept. The data points describing such a trend are termed “overshoot” and seem to occur under a limited set of meteorological conditions.
Figure 1. Types of questionable data points. (Dashed line indicates questionable points.)
A Reasonable Cumulative Precipitation Curve

Objective criteria by which data points can be judged were developed from the standpoint of a reasonable cumulative precipitation curve. All data points making up the curve must indicate zero or a positive amount of precipitation. The value of $\Delta P/\Delta t$, change of precipitation with respect to time, must also have an upper bound. This bound is a function of the geographical location, elevation, and nature of the precipitation process. A value of $\Delta P/\Delta t > 0.5$ inch per hour is fairly uncommon for this region in the wintertime as indicated by the relative frequency of cold season hourly precipitation amounts (Table 1) at the Silver Lake-Brighton station. This is a fairly high-yield high-altitude station near the experimental area.

These results indicate that .5 inch per hour can be set as a general upper bound on $\Delta P/\Delta t$ for the data editing scheme. However, precipitation data points indicating rates greater than .5 inch per hour are accepted as valid if the curve continues a monotonic, non-decreasing trend. The main purpose in defining a maximum $\Delta P/\Delta t$ value is to insure that the first point of a data set is valid. $\Delta t$ values generally range from 20 to 40 minutes for the data being considered.

Table 1. Relative frequency distribution of hourly precipitation values for Silver Lake-Brighton 1957-66 November through April.

<table>
<thead>
<tr>
<th>Class (Inches)</th>
<th>Number of Occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>.01-.05</td>
<td>4271</td>
</tr>
<tr>
<td>.06-.10</td>
<td>578</td>
</tr>
<tr>
<td>.11-.15</td>
<td>153</td>
</tr>
<tr>
<td>.16-.20</td>
<td>44</td>
</tr>
<tr>
<td>.21-.25</td>
<td>8</td>
</tr>
<tr>
<td>.26-.30</td>
<td>6</td>
</tr>
<tr>
<td>above .30</td>
<td>0</td>
</tr>
</tbody>
</table>

The Data Quality Criteria

The preceding considerations are the basis for the complete set of data quality control criteria. These in combination with factors peculiar to equipment and data processing methods, provide enough information to successfully edit telemetered precipitation data.

An attempt was made to keep input data requirements of the computer program to a minimum and still obtain good results. A review of the manually edited data showed that satisfactory results could be obtained without using the supplementary data originally employed. This supplementary data included concurrent facsimile charts, temperature observations, upper air sounding data, and precipitation data from adjacent gages.

The first consideration in data quality control for the system is to check the time parameters for obvious keypunch errors or Automatic Readout Console errors. The number of data point errors in this category is very small, on the order of 0.1 percent. This check is accomplished by requiring that all data points lie within specified time bounds and that only one data point per station be accepted for a given time.

The next criterion is set up to discard outlying data points, those mentioned in category (1) above. The most useful test for outliers is the requirement that a data point be not more than .03 inch below the average of the preceding four data points and not more than .03 inch above the average of the following four data points. This is the requirement that, neglecting noise, a cumulative precipitation value has to lie on a monotonic, non-decreasing curve. Good results were obtained when only points yielding an absolute precipitation rate greater than .2 inch per hour were subjected to the test. Data points failing the test are deleted from the data set.

The effect of testing only points associated with $\Delta P/\Delta t > 0.2$ inch per hour was to leave in the negative trends mentioned previously in category (3).

Overshoot seems to occur under a rather limited set of meteorological conditions. Most of the time the amount of overshoot is less than .1 inch and was identified on about 10 percent of the total data read out from the network during the 1970-71 season. The criteria used to define overshoot is that each of three successive data points be at least .03 inch less than each of the two data points at the beginning of the negative trend.

After identifying the overshoot, the problem is to ascertain its beginning and ending times. It seemed reasonable to place the end of the overshoot at the point where the curve begins once again a monotonic, non-decreasing trend. The identification of the beginning of overshoot is more difficult. The only certain thing is that it occurred when the gage indicated a precipitation event. For the present, the beginning of overshoot is arbitrarily placed two data points prior to the point where the cumulative curve exceeds the precipitation value at the end of the overshoot. The peak value of the overshoot is adjusted downward to the precipitation value corresponding to the end of the overshoot and no precipitation is assumed during the time interval between the peak overshoot value and the end of the overshoot. The precipitation curve from the time of the start of the overshoot to the time of the peak overshoot value is adjusted to keep within the monotonic, non-decreasing curve criteria, and yet maintain as much of its original shape as possible. An example is shown in Figure 2.

Whereas these corrective procedures for overshoot are quite intuitive, the results match favorably those done by hand with supporting meteorological data. It is questionable whether more effort is justified in deriving a better editing procedure. A new type gage is presently
being tested which could eliminate the overshoot problem.

The cumulative precipitation curve, having been subjected to the above treatments, is now ready to be checked for noise points. The requirement for the first data point of the set is that it be less than the average of the next four data points. The remainder of the data points are required to be greater than or equal to the preceding data point, but less than the average of the following four data points. This average is taken so that the effect of a noise point in the four will be negligible. If a point is outside this given range, but within 0.03 inch of the limits, it is corrected the least amount possible to bring it within the limits and retained. Otherwise, it is discarded.

On a few occasions the average of the four testing points is less than both the point being tested and the preceding test point. In this case data are processed as follows.

Let $P_i$ be the $i^{th}$ precipitation value. This is the point being tested. Let $P_{i+1}$ be the point which has just been tested and now is used to test $P_i$. Let $P_j, j = i+1, i+4$ be the remainder of the test points and let $A = \frac{1}{4} \sum_{j=i+1}^{i+4} P_j$.

The test for a noise point described above was the requirement that $P_{i+1} < P_i < A$. The problem arises when $P_{i+1} > A$. If this happens, the most often occurring value of $P_j, j=i, i+4$ is assigned to $A$ and the test is completed. If no value occurs more than once, the mid-point of the smallest region in which two or more of the $P_j, j=i, i+4$ lie is assigned to $A$.

These procedures have been programmed on a computer with various small additions and special provisions for handling the first few and last few data points of each set. Following the data editing procedures of the program, a tabulation for data quality control is accomplished by machine which includes the following: number of data points input, the number of outliers rejected, the number of noise points rejected, the number of noise points adjusted, the number of overshoot points adjusted, the number of data points left untouched, the total number of data points accepted, the total number of data points rejected and the percent of total input accepted. Figure 3 is an example of the original and final data sets. The computer program then interpolates half-hourly precipitation values for the entire time period covered by the data set—usually one or two days of data.

Results

All of the telemetered precipitation data for the 1969-70 and 1970-71 experimental seasons have been quality checked and edited by this program. The results agree well with manual editing.
The cost of manually editing data for a single experimental period was about $200. This compares to a computer cost of about $6. Thus, the initiation of computer editing of precipitation data has resulted in over a thirtyfold savings in the cost of this particular data processing.

The computer program is set up to edit a period of data which includes one or more experimental periods. It then outputs half-hourly data in individual sets for each experimental period. This process is completed one station at a time. A few data sets from the telemetry network are discussed and graphed for the experimental period of February 15, 1971, 0930 to 1730 MST.

Figure 4 shows data and graphs from Tony Grove Ranger Station gages 602 and 603. Both cases show noise points which were adjusted by the computer program.
Questionable points are indicated by a dashed line on the graphs.

Data for Gold Hill gage 1001 and Cinnamon Creek gage 1203 are graphed and displayed in Figure 5. These curves appear to contain overshoot as determined from a qualitative definition of the term. However, the quantitative criteria for overshoot is not met. That is, each of three successive data points must be at least .03 inch less than each of two data points at the beginning of the negative trend. The data for Cinnamon Creek gage 1203 would have met the overshoot criteria except that the highest point of the set was rejected as an outlier.

Figure 6 shows data for Franklin Basin gage 1603 and Hell Canyon gage 1701. These data sets were quality checked by the machine and required no adjustment.

Note that one outlier was rejected for the Franklin Basin gage. This was a duplicate time point which occurred at day 46,402.

The overshoot correction is shown in Figure 7 for Dry Bread Pond gage 5004 and Porcupine gage 12501. The first appearance of overshoot on the graph for Dry Bread Pond was not overshoot as defined quantitatively. The second appearance of overshoot on this graph and on the one following was identified as overshoot and treated as such.

The overshoot occurring on Porcupine gage 12501 is typical of snow and ice accumulation and subsequent melting off. The overshoot occurrence on Dry Bread Pond gage is more difficult to explain.
The data processing steps shown and discussed are typical of those taken on most of the data from the telemetry network. The actual handling of questionable data points for the cases of outliers and noise points is in accord with theory. However, the problem of where to place the beginning time of overshoot in a data set is still open to question.

**Conclusion**

The development of this data quality control program is a considerable achievement in automated precipitation data processing. Its development has reduced the cost of acquisition of reliable precipitation data to 3 percent of the original cost, and makes it possible to have the final edited precipitation data within a few days of the operational event, instead of the few weeks required by the manual editing. With the completion of a planned paper tape punch output, this time will be further reduced through the development and use of this program.
Figure 6. Data sets for Franklin Basin gage 1603 and Hell Canyon gage 1701. No corrections were made on data points in these sets. However, one time outlier at day 46.402 was found and discarded from the data set of station 1603.
Figure 7. Final (solid curve) and original (dashed curve) data sets for Dry Bread Pond gage 5004 and Porcupine gage 12501. Overshoot was detected and treated in these data sets.
References

Chadwick, Duane G., July 1968: USU Remote Total Precipitation Telemetry Station. PRWG30-6, Utah Water Research Laboratory, Utah State University, Logan, Utah.

Chadwick, Duane G., 1969: Telemetry System Modifications and 1968-69 Operation. PRWG30-8, Utah Water Research Laboratory, Utah State University, Logan, Utah.

APPENDIX A
The Computer Program

In this appendix, the computer program is described in separate parts related to the specific types of data quality checks and data editing procedures. Reference is made to the variable list in Appendix B and the computer program listing in the latter part of this appendix. The computer program is referred to by line number of the listing.

Lines 5-6
The starting day and ending day of the data set are read in. The variable KEROR is normally read in as a zero. If any other number is read in for KEROR, the program prints out intermediate data sets during the edit process. This is mainly for program debugging. The number of sets of operational event precipitation data to be output is the fourth input variable.

Lines 7-8
Starting and ending times for individual operational seeding events and their identification are read in.

Line 9
This is a data statement to define the last day of each month of the year in terms of day of year. It must be changed when data for a leap year is being edited.

Lines 10-12
The read, write, and punch channels for the computer are defined.

Lines 13-30
A data set for one station is read in from cards. There are seven cumulative precipitation values per card in the format explained in Appendix C. After each card is read, a test is made for change in station number. This change indicates the end of the data set for one station and a transfer is made to the processing portion of the program. A short routine is included here to make a reverse search on the last seven values read in for a station. Its purpose is to find the last data point of the set since the number of data points in a set is usually not an even multiple of seven.

Lines 31-42
The station identification is printed out and a check is made here to insure that more than seven data points are available for the editing procedure. The original set of data points is then transferred to another array to be kept for later printing and manual comparison with the edited data set.

Lines 43-77
A time check is made on the data points by first making sure that all values are stored in chronological order. If more than one point exists for a given time, the first of the two values is deleted. This is based on the assumption that a keypunch error was made and corrected on the second of the two cards. The deletion of the bad data point is accomplished by decreasing the total number of points in the set by one and setting the i\textsuperscript{th} storage location to the value of the \((i+1)^{\text{th}}\) point of the set. This is done from the bad point on to the end of the data set. A check is then made for negative cumulative precipitation values and data points lying outside of the data set region timewise.

Lines 78-99
This routine finds a reasonable starting data point for the set, i.e., reduces the possibility that an outlier be selected as the initial point. This is done in the following manner. Let \(b_1\) and \(b_2\) denote precipitation rates in inches/hour between points one and two and points two and three respectively. The first data point is regarded as valid if \(b_1 < -.5\) and \(b_2 > +.5\), or \(-.5 < b_1 < .5\) with no restrictions on \(b_2\), or \(b_1 > .5\) and \(b_2 < -.5\). The additional requirement that the absolute value of the difference in cumulative precipitation between points one and two be less than one inch must also be met. Justification of this criteria is somewhat lengthy and is not included in this report. If the first data point is rejected, the procedure is repeated until a suitable starting point is found.

Lines 100-137
Outlying data points are discarded here by the procedure outlined in the data quality criteria section of this report.
This section of the program checks the data set for overshoot, and if detected, corrects it as outlined previously in the Data Quality Criteria section.

Precipitation rates are calculated point by point for each pair of points in the data set and counting variables are initialized. If the debugging variable, KEROR, is other than zero, the data set and rates are printed out before the overshoot check starts. Lines 152 and 167 check for overshoot as described previously. Lines 168-176 find the end of the overshoot, and lines 177-191 correct the overshoot as explained in the previous section of this report. The adjusted and the original precipitation values are printed out for future reference.

Noise points are checked for in this section of the program. The checking and editing procedures are programmed as described in the Data Quality Criteria section of this report.

The edited and unedited data sets are printed. This allows for subsequent examination of both data sets simultaneously. An example of this printout is shown in Figure 3.

When the variable, KEROR, is set to a value other than zero, this section punches the edited curve in the same format as the input data. This allows the graphing of the two curves on a hybrid computer enabling evaluation of the editing procedure.

A printout of the data quality parameters is done at this stage. Figure 3 shows an example.

Interpolation, printing, and punching of half-hourly precipitation values is handled by this portion of the program. The day and month are defined in terms of the day of the year associated with the first data point of the set. An incremental parameter A is defined as 1/48. A time of day parameter ST is defined in decimal equivalence of integral numbers of half hours for the times at which the individual interpolated precipitation values are calculated. The proper interval for interpolation is determined from the parameter ST and the half-hourly precipitation value is computed. The time parameter is incremented by 1/48 of a day and the process is repeated for the time interval of the entire data set. The data are output on cards with one card for each 12 hours of data making two cards output per day. A card image is shown in Appendix C. Each card lists the station number, the year, month, and day of the data, 24 half-hourly values, maximum interpolation interval, a start time and end time for the data of each individual card, the operational event identification, and a number indicating whether a card is for a.m. or p.m.

If at one of the check points in the program, an indication is given that too much data has been discarded, the program branches to this point and prints out the input data set with a message to the user.

This routine transfers the first data points of the next set into the first seven locations of the data set array. Control is then transferred back to line 15 where the process is carried out on the next data set. A blank card is placed at the end of the data which causes zero to be read in as the station number. When this condition is sensed, the stop command in line 529 is executed.
C*******************************************************************************
C WASATCH WEATHER MODIFICATION PROJECT PRECIPITATION DATA QUALITY CONTROL
C PROGRAM DEVELOPED AND WRITTEN BY
C RONALD H. CAMPBELL
C RESEARCH ASSISTANT
C DEVELOPED THROUGH FUNDING BY THE BUREAU OF RECLAMATION
C OFFICE OF ATMOSPHERIC WATER RESOURCES
C*******************************************************************************
C DIMENSION DATA ARRAYS
0001 DIMENSION STS(25),ETS(25),OE(25)
0002 DIMENSION T(500),P(500),TC(500),PC(500)
0003 DIMENSION FMT(15),NUM(13),S(500),KT(7),KP(7),KS(7)
0004 INTEGER PRI(800),PT(800)
C READ IN DAY TIME LIMITS AND DAYS OF MONTH FOR OUTPUT
0005 READ (5,841) SCAY,ECAY,KERDR,NOSET
0006 841 FORMAT (2F6.3,11,I3)
0007 READ (5,2842) (OE(I),STS(I),ETS(I),I=1,NSET)
0008 2842 FORMAT (4(A4,2X,2F6.3,2X)
C THE FOLLOWING DATA CARD MUST BE CHANGED FOR LEAP YEAR
C THE NUMBERS ARE THE DAY OF THE END OF EACH MONTH OF THE YEAR WITH A ZERO
C AS THE FIRST VALUE
0009 DATA NUM/0,31,9,90,120,151,181,212,243,273,304,334,365/ 0010 NR=5
0011 NW=6
0012 NPUN=1
C READ IN DATA TO BE EDITED
0013 READ (NR,1) KSTN,KYR,(T(I),P(I),I=1,7)
0014 1 FORMAT (15,I2,3X,7(F6.3,F4.2))
0015 DO 20 I=1,71
0016 J=I+171
0017 K=J+6
0018 READ (NR,1) LSTN,KYR,(TL(I),PL(I),I=J,K)
0019 IF (LSTN.NE.KSTN) GO TO 21
0020 LYR=KYR
0021 20 CONTINUE
0022 21 NP=J-1
0023 NPP=NP
0024 DO 25 I=1,7
0025 J=NP+I-1
0026 IF (T(I).LT.-0.1) GO TO 22
0027 NP=NP-I+1
0028 GO TO 25
0029 5 CONTINUE
0030 6 CONTINUE
C WRITE OUT STATION NUMBER
0031 WRITE (NW,666) KSTN
0032 666 FORMAT (1$I,I7,7I1,STATION,16)
0033 NPI=NP
0034 IF (NP.GT.7) GO TO 261
C CHECK FOR SUFFICIENT NUMBER OF DATA POINTS FOR EDITING
0035 WRITE (NW,262) (TI(I),PI(I),I=1,NP)
0036 262 FORMAT(14(F9.3,F6.2))
0037 GO TO 127
0038 261 CONTINUE
C TRANSFER ORIGINAL DATA FOR PRINTOUT LATER
0039 DO 25 I=1,NP
0040 TC(I)=TI(I)
0041 PC(I)=PI(I)
0042 NLCR=0
C**********************************************************
C TIME CHECK
0043 400 I=1
0044 401 I=I+1
0045 431 CONTINUE
0046 IF (TI(I).LT.T(I-1)) 402,402,402
0048 402 K=I-1
406 J=1
A=T(I)
T(I)+T(I-1)
A=P(I)
P(I)=P(I-1)
P(I-1)=A
GO TO 404
403 NP=NP-1
DO 407 K=1,NP
T(K)=T(K+1)
P(K)=P(K+1)
NLOR=NLOR+1
IF (J.LE.NP) GO TO 431
404 IF (J.LE.NP) GO TO 4C1
406 C******************************************************************************
C CHECK FOR NEGATIVE CUMULATIVE PRECIPITATION VALUES AND OUTLYING TIME POINTS
461 I=C
462 I=I+1
463 IF (P(I).LT.0.) GO TO 465
468 IF (T(I).LT.SDAY) GO TO 465
469 IF (T(I).GT.EDAY) GO TO 465
470 GO TO 466
465 NP=NP-1
DO 471 J=1,NP
P(J)=P(J+1)
471 T(J)=T(J+1)
NLOR=NLOR+1
476 IF (J.LE.NP) GO TO 463
466 IF (J.LE.NP) GO TO 462
C******************************************************************************
C FIND STARTING POINT FOR GOOD DATA
810 IF (NP.LT.7) GO TO 126
811 I=1
83 I=I+1
812 B1=(P(I)-P(I-1))/((T(I)-T(I-1))*24.)
813 B2=(P(I+1)-P(I))/((T(I+1)-T(I))*24.)
814 IF (B1+.5) GO TO 82,
815 A=ABS(B1)
816 IF (A.GT.1.) GO TO 81,
817 CONTINUE
818 A=P(I)-P(I-1)
819 A=ABS(A)
820 IF (A.GT.1.) GO TO 83,
821 J=I-1
822 DO 27 I=J,NP
823 K=I-J+1
824 T(K)=T(I)
825 P(K)=P(I)
826 NP=NP-J+1
827 NLOR=NLOR+J-1
828 CONTINUE
810 IF (I.EQ.2) GO TO 30
829 C CHECK FOR OUTLYING DATA POINTS
840 B=P(I)-P(I-1)
841 C=ABS(B)/((T(I)-T(I-1))*24.)
842 IF (C.LT.2) GO TO 443
843 IF (I.EQ.NP) GO TO 42C
845 C=P(I+1)-P(I)
846 D=B+C
847 D=ABS(D)
848 IF (D.LT.0.2) GO TO 420
C CHECK FOR ABOVE AVERAGE OF FOLLOWING POINTS
850 STO,
870 NB=I+5
871 JJ=I+1
872 IF (NB.GT.NP) NB=NP
IF (JJ.GT.NB) JJ=NB
DO 150 J=JJ,N8
ST=ST+P(J)
A=NB-JJ+1
ST=ST/A+.03
IF (P(I).GT.ST) GO TO 42C
GO TO 443
NP=NP-1
IF (NP.LE.NP) GO TO 126
DO 425 J=1,NP
T(I)=T(J+1)
P(I)=P(J+1)
NLR=NLR+1
IF (I.LE.NP) GO TO 428
GO TO 443
IF (I.LT.NP) GO TO 411
C OVERSHOOT ROUTINE
NV=NP-4
PR(I)=P(I)*10C.+.1
DO 201 I=2,NP
PR(I)=P(I)*10C.+.1
201 S(I)=P(I)-P(I-1))/((T(I)-T(I-1))*.24.)
NCNT=0
NO=0
J=1
NO=0
NP=NP-1
NAD=0
IF (KERGR.EQ.C) GO TO 814
WRITE (6,230) (I,T(I),P(I),S(I),I=1,NP)
814 1=1
202 I=I+1
IF (NO.EQ.1) GO TO 225
IF (NO.GE.1) GO TO 223
IF (S(I)) 204,205,205
204 NCNT=NCNT+1
205 NO=NO+1
IF (NCNT-3) 263,206,206
C OVERSHOOT IS INDICATED
C RECHECK OVERSHOOT IDENTIFICATION
L=I-2
LL=L-1
A=P(LL)-.02
DO 224 LA=L+1
IF (A.LE.P(LA+1)) GO TO 265
224 CONTINUE
IK=0
NO=1
I=I-1
IF (S(I)+.01) 207,208,208
207 1=I-2
208 CONTINUE
IK=IK+1
IF (IK-3) 211,209,209
C END OF OVERSHOOT IDENTIFICATION
C L IS POINT OF MAX PGPTA
C K IS LAST POINT OF OVERSHOOT
C IDENTIFY START TO ADJUST TIME

NN=L-J+1
JA=J+1
DO 213 MM=JA,NN
M=NN-MM+1
IF (P(M).GT.P(K)) GC TO 213
GO TO 214
CONTINUE

CONTINUE

KM=M-1
IF (KM.LT.1) KM=1
IF (KM.EQ.1) GC TO 215
A=T(KM+1)-T(KM)
IF (A.GT.0.08) KM=KM+1
CONTINUE

WRITE (NW,980) KM,L,K
980 FORMAT (26H START, MID, AND END POINT,3I5)
CONTINUE

KM IS THE POINT PRECEEDING THE START OF THE OVERSHOOT
C=P(L)
WRITE (6,593)
593 FORMAT (32H OVERSHOOT ADJUSTED DATA FCLASSW )
WRITE (6,595)
595 FORMAT (1X,36H POINT TIME NEW VALUE OLD VALUE)
KKB=L-1
DO 222 M=KM,KKB
B=(P(KM)+(C-P(KM))*((T(M)-T(KM))/(T(L)-T(KM)))
A=P(M)
IF (P(M).GT.P(K)) P(M)=P(K)
WRITE (6,594) M,T(M),P(M),A
594 FORMAT (1X,14,F8.3,3X,F9.2,3X,F9.2)
CONTINUE

DO 221 M=L,K
A=P(M)
P(M)=P(K)
WRITE (6,594) M,T(M),P(M),A
226 CONTINUE

NAOS=NAOS+1
227 CONTINUE

GO TO 203
207 IK=0
208 CONTINUE
209 K=I-2
210 CONTINUE
211 IF (I.LT.NPS) GO TO 203
212 K=1
213 GO TO 210
214 CONTINUE
215 IF (I.LT.NPS) GC TO 202
216 NAOS=NAOS+1
217 CONTINUE
218 NCNT=0
219 NQ=0
220 GO TO 203
221 CONTINUE
222 CONTINUE

IF (I.LT.NPS) GO TO 227
223 IF (I.LT.NPS) GC TO 202
224 GO TO 210
225 CONTINUE
226 IF (I.LT.NPS) GC TO 202
227 IF (NAOS.EQ.0) GO TO 227
228 NAOS=0
229 DO 226 I=1,NP
A=PR(I)
A=PR(I)-A*.01
232 A=ABS(A)
233 IF (A.GT.0.005) NAOS=NAOS+1
234 CONTINUE
235 CONTINUE
236 CONTINUE

C*******************************************************************************
C NOISE POINT CHECKING ROUTINE

I=1
DO 238 J=1,NP
239 PR(J)=P(J)*100.*1
240 CONTINUE
241 M=0
242 CONTINUE
243 38 I=I+1
244 31 CONTINUE
A=0
K=I+1
L=I+4-M
DO 480 J=K,L
480 A=A+P(J)
B=L+K+1
A=A+B+.0001
IF (P(I-1).GT.A) GO TO 719
23 IF (P(I).LT.P(I-1)) GC TO 18
254 IF (P(I).GT.A) GO TO 22
255 GO TO 40
256 IF ( (P(I)+.03).LT.P(I-1)) GC TO 35
257 P(I)=P(I-1)
258 GO TO 40
259 IF ( (P(I)-.03).GT.A) GO TO 35
260 P(I)=A
261 GO TO 40
719 J=I+4
262 IF (J.GT.NP) J=NP
263 IF (J.GT.NP) J=NP
264 IF (J.IE.K+2) GO TO 726
265 A=.005
266
722 K=C
267 DO 720 MM=I,J
268 L=0
269 DO 721 N=I,J
270 B=P(MM)-P(N)
271 B=ABS(B)
272 IF (B.GT.A) GO TO 721
273 L=L+1
274 721 CONTINUE
275 IF (L.LE.K) GO TO 720
276 K=L
277 NAS=MM
278 720 CONTINUE
279 IF (K.GT.1) GC TO 723
280 A=A+.01
281 GO TO 722
282 723 IF (P(I-1).LE.P(NAS)) GO TO 724
283 K=I-1
284 725 P(K)=P(NAS)
285 K=K+1
286 IF (K.LT.1) GC TO 724
287 IF (P(K).GT.P(NAS)) GC TO 725
288 724 A=P(NAS)
289 GO TO 23
290 726 IF ( (P(I-1)-P(I+1)).LT..03) GO TO 35
291 P(I+1)=P(I-1)
292 A=P(I+1)
293 GO TO 23
294 35 NP=NP-1
295 NV=NV-1
296 IF (NP.LT.7) GO TO 126
297 DO 36 K=I,NP
298 T(K)=T(K+1)
299 PR(K)=PR(K+1)
300 36 P(K)=P(K+1)
301 NN=NP-1
302 IF (I.LE.NV) GO TO 31
303 M=M+1
304 IF (I.LE.NN) GO TO 31
305 I=I-1
306 GO TO 71
307 40 CONTINUE
308 IF (I.LT.NV) GO TO 38
309 M=M+1
310 IF (I.LT.NNP) GO TO 38
311 71 B=(P(I+1)-P(I))/((T(I+1)-T(I))*24.)
312 IF (B.LT.5) GO TO 74
313 GO TO 73
314 74 IF (B.GE.0.) GO TO 75
315 IF ( (P(I)-P(I+1)).GT.03) GO TO 73
316 P(I+1)=P(I)
### APPENDIX B

**Description of Program Symbols**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Arrays</th>
<th>Indexes for the outlier routine</th>
<th>Description</th>
<th>Arrays</th>
<th>Parameters used in the half-hourly interpolations</th>
<th>Description</th>
<th>Arrays</th>
<th>Months and day of output data</th>
<th>Description</th>
<th>Arrays</th>
<th>Year</th>
<th>Description</th>
<th>Arrays</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETS</td>
<td>End time of operational event</td>
<td>JJ, NB</td>
<td>JJJ, KKK, MN</td>
<td>Parameters used in the half-hourly interpolations</td>
<td>KK, KDAY</td>
<td>Data set debug parameter</td>
<td>Month and day of output data</td>
<td>KSTAR, KSTART</td>
<td>Beginning time of interpolated data</td>
<td>Beginning time of interpolated data</td>
<td>KSTO, KSTOP</td>
<td>End times of interpolated data</td>
<td>Description</td>
<td>Arrays</td>
</tr>
<tr>
<td>KP, KS, KT</td>
<td>Output parameters to be punched for plotting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>NUM</td>
<td>Day of year which ends each month</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>OE</td>
<td>Operational event identification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>P</td>
<td>Cumulative precipitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>PC</td>
<td>Cumulative precipitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>PR</td>
<td>Cumulative half-hourly precipitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>PT</td>
<td>Non-cumulative half-hourly precipitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>S</td>
<td>Change of precipitation with time</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>STS</td>
<td>Beginning time of operational event</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>T</td>
<td>Time of data point</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>TC</td>
<td>Time of data point</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td><strong>Single Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>A, B, C, D, ST</td>
<td>Storage parameters for sorting and miscellaneous computation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>AA, MB</td>
<td>Largest time interval of interpolation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>B1, B2</td>
<td>$\Delta P/\Delta t$ for points one through three</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>BBB</td>
<td>Beginning time for interpolated data set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>DA, DB</td>
<td>Difference between successive cumulative precipitation values</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>EDAY</td>
<td>Ending day and time of data set</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
<tr>
<td>I, J, K, L, M</td>
<td>Control parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Arrays</td>
</tr>
</tbody>
</table>
### Operational Event Information

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-4</td>
<td>Operational event number</td>
</tr>
<tr>
<td>5-6</td>
<td>Blank</td>
</tr>
<tr>
<td>7-12</td>
<td>Beginning time of operational event in thousandths of days of the year</td>
</tr>
<tr>
<td>13-18</td>
<td>Ending time of operational event in thousandths of days of the year</td>
</tr>
<tr>
<td>19-20</td>
<td>Blank</td>
</tr>
</tbody>
</table>

This pattern continues for each operational event data subset contained in the main data set. The program provides for up to 25 such data subsets making a possibility of 9 cards for this information. The example has two data subsets.
### Precipitation Data Card

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>Station number</td>
</tr>
<tr>
<td>6-7</td>
<td>Year</td>
</tr>
<tr>
<td>8-10</td>
<td>Blank</td>
</tr>
<tr>
<td>11-16</td>
<td>Time of precipitation point in thousandths of days of the year</td>
</tr>
<tr>
<td>17-20</td>
<td>Cumulative precipitation in hundredths of inches</td>
</tr>
<tr>
<td>21-26</td>
<td>Same as 11-16 for the next data point</td>
</tr>
<tr>
<td>27-30</td>
<td>Same as 17-20 for the next data point</td>
</tr>
</tbody>
</table>

This pattern continues for the remainder of the data set. If the number of points in the data set is not an even multiple of 7, the remaining fields on the last card are left blank.

20171 467921743 468251743 468651743 0 0 0 0 0 0 0 0 0
20171 466081743 466741743 466891743 467101743 467311743 467501743 467711743
20171 464631743 464861743 465071743 465261743 465471743 465681743 465891743
20171 463191743 463391743 463631743 463841743 464021743 464231743 464421743

This pattern continues for the remainder of the data set.
## Card Output

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>Station number</td>
</tr>
<tr>
<td>6-7</td>
<td>Year</td>
</tr>
<tr>
<td>8-9</td>
<td>Month</td>
</tr>
<tr>
<td>10-11</td>
<td>Day</td>
</tr>
<tr>
<td>12-59</td>
<td>Half-hourly precipitation (2 columns per half hour)</td>
</tr>
<tr>
<td>60-62</td>
<td>Blank</td>
</tr>
</tbody>
</table>

63 Maximum length of interpolation period

64-65 Blank

66-69 Beginning time of data

70-73 Ending time of data

74 Blank

75-78 Operational event ID

79 Blank

80 First or second half of day