NATURAL ENVIRONMENT RESEARCH COUNCIL

INSTITUTE of HYDROLOGY

REPORT No 14 MARCH 1972

A SYSTEM FOR PROCESSING DATA FROM AUTOGRAPHIC RECORDERS

by

J. R. Douglas

INSTITUTE OF HYDROLOGY
HOWBERY PARK
WALLINGFORD
BERKSHIRE

CONTENTS

FIGURES

INSTITUTE OF HYDROLOGY

Report No. 14 March 1972

A SYSTEM FOR PROCESSING DATA FROM AUTOGRAPHIC RECORDERS

by J. R. Douglas

ABSTRACT

Autographic recorders are widely used for the continuous measurement of many physical processes. Completely manual methods of data extraction from charts can lead to several types of error; they are also extremely tedious. A semi automatic system, which eliminates many of the errors and much of the labour, is described both in principle and in full operational detail.

1. AUTOGRAPHIC RECORDERS

1.1 The Institute of Hydrology uses autographic recorders for measuring many hydrometeorological variables such as streamflow, rainfall, temperature and atmospheric pressure. These measurements are collected from well instrumented catchment areas with the intention of monitoring water flux within those areas.

One value of the autographic chart compared, say, with a punched paper tape, or a magnetic recording, is that the measured data sequence is continuous, rather than having readings at discrete intervals. Also, the sequence is visible and any irregularities which are apparent on the recorded trace can be explained or removed before the data is used. Many such irregularities could remain undetected by even sophisticated quality control, and could therefore pass through processing into the archives.

These irregularities can be caused by:

a) malfunctioning of the sensing equipment
b) malfunctioning of the linkage equipment (friction in pulleys etc)
c) malfunctioning of the recording equipment (clock speed, paper expansion etc).

1.2 A thorough examination by an experienced chart analyst can normally locate and correct faults in the first two categories. In the absence of any recorder malfunctioning, the data could be extracted manually, written on to forms for inspection or for punching on to computer cards for further processing. The lack of recorder malfunctioning must first be established. In some cases the simple data extraction technique outlined cannot be used.

Normally autographic charts are marked with a time and a vertical scale. Ideally, it should be possible to take a vertical reading at any point in time simply by reading the scale coordinate at that time point.

however, in the event of the recorder clock speed being inaccurate, the actual time does not correspond exactly with the appropriate grid line. Therefore, corrections have to be made to the time scale so that the vertical coordinates are read at the correct times on the chart. For several reasons such as inaccurate loading of the recorder, expansion or shrinkage of the chart, the vertical scale grid may also be in error. When corrections have been made to both time and vertical scales, it is no longer a simple job to extract data from the charts manually. The scale corrections and data extraction can be simplified by using a digitising table.

2. D-MAC PENCIL FOLLOWER

2.1 The equipment

a) reading table
b) reading pencil
c) electronic console
d) paper tape punch
e) portable keyboard

The reading table has a surface of about 1 metre square, which is covered by a coordinate system with X and Y axes parallel to the table edges. Both axes are divided into 10^4, 0.1mm units, giving a total of 10^8 locations on the table surface.

The reading pencil is the sensing device whose movement over the table is followed by an automatic mechanism housed under the table. Coordinate position of the pencil is transmitted when required to the electronic console.

The electronic console converts signals from the reading pencil into digital form, which are displayed on a screen, and transmitted to the paper tape punch in the required format.

The portable keyboard allows additional information to be transmitted to the paper tape punch, so that output can be labelled and coded as required.

2.2 Correction of scale errors

X and Y coordinates are taken off the chart with reference to the datum of the table. Thus if the chart datum and scale coordinates are taken off first, the table coordinates can be interpreted in real (chart) units by a subsequent computer program. Clearly, once these datum points and scales are established, the grid lines on the chart are of no further use. Thus the following errors can be overcome:

a) inaccurate clock speed
b) vertical scale errors
c) wrong vertical units printed on the chart.

The clock speed cannot be checked unless the following information is available:

a) the time when the chart was put on to the recorder, plus a clear mark on the chart trace at this time

b) the time when the chart was removed plus a clear mark on the chart trace at this time.

Annotation of the charts is greatly simplified if the chart grid is printed with timing labels. Otherwise tedious counting of squares is unavoidable if the actual time scale of the chart is to be established. In the absence of the above information, the theoretical scale must be assumed. If there are additional time checks, a far more detailed assessment of the chart speed and its variation can be obtained. Any error, when found, is corrected by digitising calibration points the correct distance apart, rather than the theoretical distance for a given time base, as shown by Figure 2.1.

Vertical scale errors can be caused by inaccurate loading of the instrument, or by expansion or shrinking of the chart on the instrument. Their existence can be established only if independent observations of the measured parameter are available at marked times on the chart, in sufficient number, and through a wide enough vertical range. The independent observations might be:

a) reversal of trace or tipping of raingauge, indicating full scale deflection; second reversal, or resumption of tipped raingauge trace, indicating the baseline. Normally, the accuracy of the recorder itself can be assumed and any inaccuracy can be ascribed to the chart
b) manual observations with auxiliary equipment.

In the absence of sufficient information, the theoretical vertical scale must be assumed correct. A constant error in one direction indicates bad alignment of the paper, as in Figure 2.2. An actual scale smaller than the theoretical indicates chart expansion and vice versa, Figure 2.3.

In these cases, the error can be corrected by taking baseline calibration points on the true baseline, rather than the theoretical (chart) baseline, and full scale calibration at the true full scale deflection. The vertical range used in conjunction with these calibration points will give the properly positioned and correctly scaled vertical axis when the data is processed by the computer.

The units of the required output might for any reason be different from those of the chart. Provided that there is a constant conversion factor, the vertical range can be punched as the number of the required units, between the actual baseline and full scale deflection, irrespective of the chart grid.

2.3 *Digitising the Charts*

This is described in full detail in section 4. Briefly, each chart is positioned within the coordinate system on the d-mac table; the output paper tape is labelled with a tape sequence number, gauge number and date; calibration points are taken to define the time and vertical axes, and time base and vertical range are keyed in to scale these axes. The recorded trace is followed with the reading pencil, points being digitised at intervals along the trace. If a chart will not fit within the d-mac coordinate grid, it can be treated in sections. It is convenient to annotate continuous charts in such sections before digitisation is started.

A typical paper tape output from the digitiser would contain the following sequence of information.

a) tape sequence number; gauge number; data type code
b) chart sequence number
c) date, month and time at beginning of chart
d) and e) X and Y d-mac coordinates at either end of baseline
f) X and Y d-mac coordinates at full scale deflection
g) time base between baseline calibration points
h) vertical range between baseline and full scale deflection
i) a series of X and Y d-mac coordinates taken along the chart trace
j) a symbol denoting the end of the chart

2.4 Extending the output

The normal output resulting from a scan of recorder charts would, when processed by the program described in section 3, be in a form suitable for use as input to the Institute's catchment data processing programs. These data are long continuous sequences and consequently they do not contain all the detailed information which could be abstracted from the charts. Other requirements could include:

a) data at a much shorter interval for detailed studies of short events
b) frequency analyses, duration curves
c) analysis of rate of change of stage, rainfall intensity etc.

An advantage of the d-mac pencil follower is that all the basic information necessary to determine these are obtained from a single scan of the chart; only further additional subroutines to the computer program will be required.

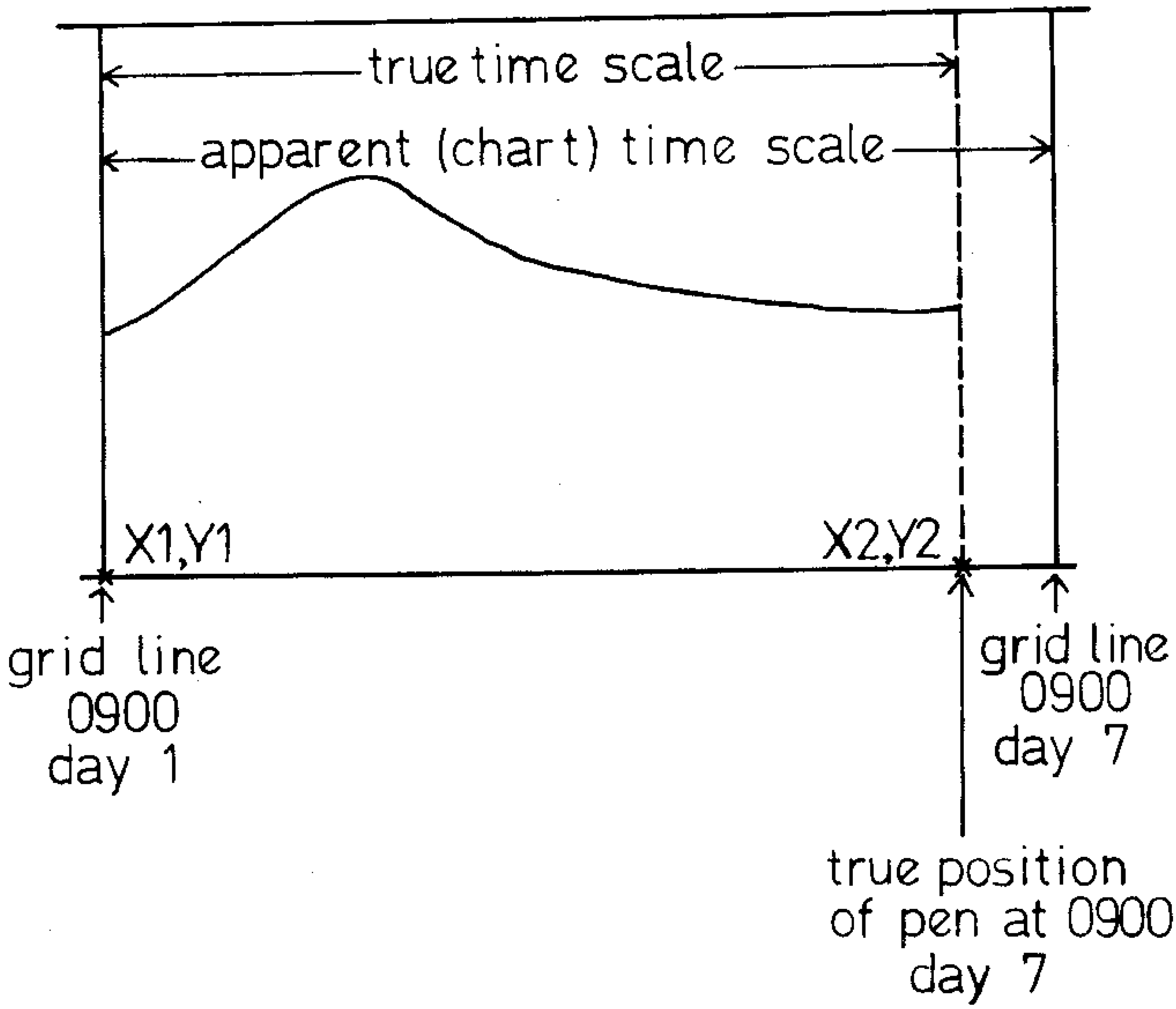


Figure 2.1 Correction of bad recorder clock speed

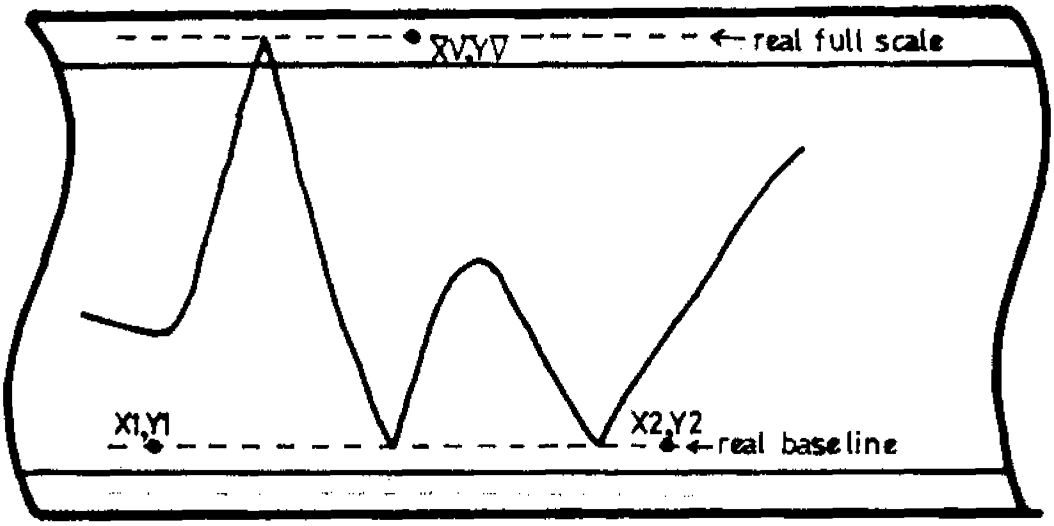


Figure 2.2 Bad positioning of chart identified by position of reversals

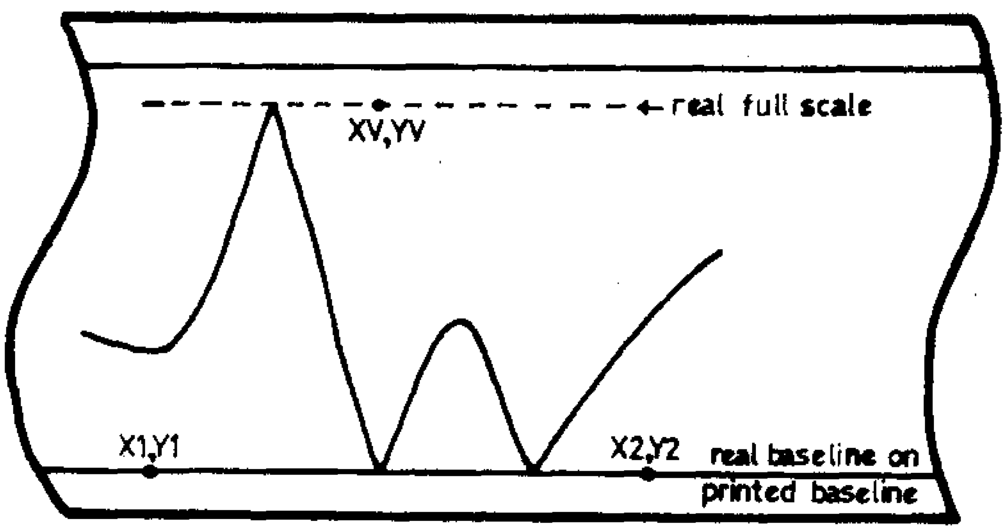


Figure 2.3 Expansion of chart identified by position of reversals

3. PROCESSING OF THE DIGITISED TRACE

A computer program has been written to interpret a sequence of d-mac coordinate points in terms of times and vertical scale readings. It has been written in a very general way, so that practically all types of autographic recorder charts can be processed. These can include continuous roll charts, short strip charts and charts with a curved vertical axis (but not circular charts); the charts may be from recorders with a reversing mechanism, or a revolving drum for extending the range of the instrument without reducing the chart scale. For periods when autographic charts are unavailable, the sequence of times and vertical scale readings are acceptable on punched cards. The program takes the set of discrete points, and from them regenerates the continuous record.

The sequence of operations performed by the program is:

a) the control of the flow of data, the count of days and months, the selection of the correct source of the current data - all according to the run control information
b) the reading of data from the appropriate source
c) if the data is from the d-mac pencil follower, the conversion of d-mac coordinates into pairs of time and vertical scale readings
d) the fitting of mathematical functions to the data points to regenerate the autographic trace
e) interpolation on the regenerated curves, to obtain vertical scale readings at the required points in time
f) the output of these readings in a form which is suitable for input to subsequent data processing programs.

3.1 Supervising the flow of data

A small number of computer cards provide information necessary to ensure the correct processing sequence. The cards, described in detail in section 4, provide the following information required by the program.

a) type of data
b) gauge number
c) number of months to be processed
d) date of first month

and then, for each month,

a) the number of days in the month
b) the number of readings per day required in the output
c) the timing of the first reading.

As the data can come from a number of sources, and not only from a d-mac digitisation, further data cards are required to establish the origin of each batch of data within a month; these contain,

a) the number of days of data in the current batch
b) the source of the batch.

3.2 Reading data from the chosen source

Data from the d-mac pencil follower are punched on eight track 1900 coded paper tape, which is input to the computer and listed. The list is checked for obvious errors, particularly in the keyboard input data, and its format. Editing should be carried out before further processing. Several monthly batches of data can be merged into a single large file for input to the program.

Data from punched cards, such as those produced by translation of the sixteen track tapes punched by Fischer and Porter stage recorders, must be introduced by a header card, as there is no other means of identification. This card should contain:

a) gauge number
b) starting date and month.

Normally the data cards have stage readings only, at a constant interval, but, as there can be no check of time keeping, they should be used only as a last resort, when autographic recorder charts are unavailable.

Pairs of readings of time and stage (or other variables) may be presented on manually punched cards; each card should contain up to seven data points, and be labelled with the gauge number, date and month.

3.3 Location, alignment and interpretation of the digital trace

The first two points taken off the chart are X1, Y1 and X2, Y2 shown on Figure 3.1. The first of these is taken as the datum for interpretation in time, and the line joining the points is taken as the baseline for interpretation of the data ordinates.

Obviously the chart cannot be positioned with this baseline coincident with the baseline of the table (which is unknown anyway) and thus a simple geometric form of interpretation is necessary. This is described by the following equations, to locate any point XX, YY, and they follow from the geometry of Figure 3.1.

$$L2 = XX - X1$$
$$L3 = YY - Y1$$

The angle of inclination of the chart to the table, α, is given by

$$\alpha = \arctan \frac{Y2-Y1}{X2-X1}$$

and the adjustment for rotation of the axes then follows, giving

$$L1 = L3 \cos \alpha - L2 \sin \alpha$$
$$L4 = L2 \cos \alpha + L3 \sin \alpha$$

The time scale for the chart section can be calculated as

$$\left(\frac{L4_{X2}}{\text{time base in hours}}\right) \text{ d-mac units per hour}$$

and the time of any point XX, YY from the beginning of the chart section can, therefore, be expressed as

$$(L4_{XX} * \frac{\text{time base}}{L4_{X2}}) \text{ hours}$$

A similar transformation can be made on the ordinate YY, given the vertical range.

If the vertical axis is curved, a correction must be made to all time calculations. This can be treated as a correction to the results obtained from following the interpretation described in the previous section and it is necessary when interpreting charts from autographic recorders which use a pivotted pen. It is assumed that the pen traces a circular arc on the chart as illustrated in Figure 3.2. This assumption can be shown to be invalid at the edges of the chart because of the curvature of the drum itself. However, the error involved is usually very small but could become significant if both the drum diameter and length of pivot arm are small. Only the time values require correction.

Calibration readings are required at three points on one constant time line. The upper one is the vertical calibration point XV, YV; the lower one must be on the chart baseline, and a further point should be approximately midway between the two. The centre, XC, YC, of the circle passing through these three calibration points is found as the intersection point of the perpendicular bisectors of the two chords BL and LV on Figure 3.2A. The radius, RA, can then be calculated geometrically (details of this calculation can be found in the computer program listing, lines 386-396). A further constant, KC, which is the horizontal distance from the centre of the circle to the vertical line passing through the intersection point of the trace arc and the baseline, can be calculated as

$$KC = XC - XB$$

Thus, to correct an XX value for curvature, it is necessary to add an amount, XD, which is, by convention, measured positive to the right of the vertical through the point where the trace arc meets the baseline. XD is calculated in terms of YD which, by convention, is measured positive above the horizontal line through the centre of the assumed circle.

Thus from Figure 3.2B

$$YD = YY - YC$$
$$XD = KC - \sqrt{RA^2 - YD^2}$$

A correction of the Y coordinate for pen reversals is necessary for charts from equipment fitted with a reversing mechanism where the width of the chart does not limit the range or clarity of the recording. Figure 3.3 illustrates a rising stage with a reversal at B, the real peak value at C, and a reversal at D on the falling stage. Clearly the correction affects only the ordinates and not the time scale.

The true value of the ordinate YY to which any scaling factor should be applied is thus

$$YY_{true} = YV + (YV - YY)$$

In the program, a general correction is made thus,

$$YY_{true} = 2*IN*YV + IO*YY$$

where the values of IN and IO are altered at each reversal according to information keyed in during the digitisation. Initially they have values 0 and 1 respectively and subsequently IO changes sign at each reversal, and IN is increased (ascending stage) or decreased (descending stage) by 1 after every 2 reversals in the same direction.

It is necessary to maintain a continuous record of time through a chart sequence, and this is facilitated by indicating that a subsequent chart is a continuation of a previous one. The final data point on the previous chart is assumed to coincide with the first point on the following chart.

3.4 The fitting of mathematical functions to the data points

The autographic trace is now regenerated within the computer by fitting suitable mathematical functions to the points, using SUBROUTINE CURFIT. In the case of streamflow, the natural shape of the autographic trace has been found to be well suited to the fitting of cubic functions. Using three data points at a time, and ensuring continuity of gradient between data intervals, a best fit equation of the form,

$$Y = a + bx + cx^2 + dx^3$$

is evaluated for each data interval. The values of the coefficients, a, b, c and d, and the times during which each set of coefficients are applicable, are stored by the computer.

In the case of rainfall, the natural trace is discontinuous and cubic curve fitting is not applicable. It is assumed that a straight line is the best approximation to the autographic trace between any pair of data points, and the curve fitting segment of the computer program is omitted. As a consequence of this, it is essential that data points are digitised wherever there is a change of slope of the trace.

3.5 Extraction of point data at regular intervals

Using the computer-regenerated trace, and the control information described above, the program carries out the following sequence of steps:

a) the time of the next required reading from the start of the data batch is calculated
b) the data interval in which this time occurs is found
c) an interpolation is made between the data points according to the mathematical function operating in the interval, to find the vertical coordinate at the given time.

Occasionally, when fitting a cubic curve to a set of points the "best fit" curve is unrealistic, (often because of some inaccuracy or omission in the pencil following). An example of this is shown in Figure 3.4. However, a check is built into the program to ensure that the interpolated value lies between the two end point values (± a small tolerance). If it does not, it is assumed that the cubic curve is not a good fit, and a straight line is used for interpolation between the points concerned.

Since some readings may already have been interpolated in the particular data interval, before the error was detected, these are corrected automatically in retrospect.

d) In the case of rainfall, it is the increment to the accumulated rainfall in a period which is required, rather than the actual vertical scale reading, which represents the accumulated total. The increment is calculated, and subsequently stored.

e) the value calculated is stored, and the program proceeds to the next time point required.

The values are thus extracted for regular intervals for the required period and are then held within the computer.

3.6 Data output and storage

The output and display of results differ according to the type of data.

Streamflow data:

a) a card image file (suitable for input to quality control and processing programs) is created containing the stage values, and all necessary control information

b) a lineprinter listing of the card image file; an example is shown in Figure 3.5A

c) a lineprinter monitor of the program run, giving warnings of missing data, inaccurate time scaling, and occurrences of poor interpolation; an example is shown in Figure 3.5B.

Rainfall data:

a) a card image file is created with regular rainfall totals on days with rain.
This is not suitable for input to subsequent programs until it is merged with control files and other recording raingauge and period gauge files, see Section 3.7

b) a lineprinter listing of the card image file; an example is shown in Figure 3.6A

c) a monthly raingauge summary sheet; an example is shown in Figure 3.6B

d) a lineprinter monitor of program run giving warnings of possible errors as in c) for streamflow data.

3.7 Sorting and merging of rainfall files

An ICL sort/merge computer library package is utilised to combine the various rainfall data files into one complete monthly rainfall file:

Originally there is a) one file for each recording raingauge
b) one file for period gauge totals, control cards and all other manually punched cards.

Cards or card images are arranged in the following order for each catchment month so that the resulting file is suitable for the quality control and processing programs.

a) LEAD CARD
b) CONTROL CARD
c) AREA CARD
d) period gauge card
e) recording gauge cards for the duration of the period gauge card, in chronological order and in ascending order of gauges within each day
f) period gauge card etc. until a month's batch is complete.

Complete details of this specific requirement can be found in Institute of Hydrology Report No. 15.

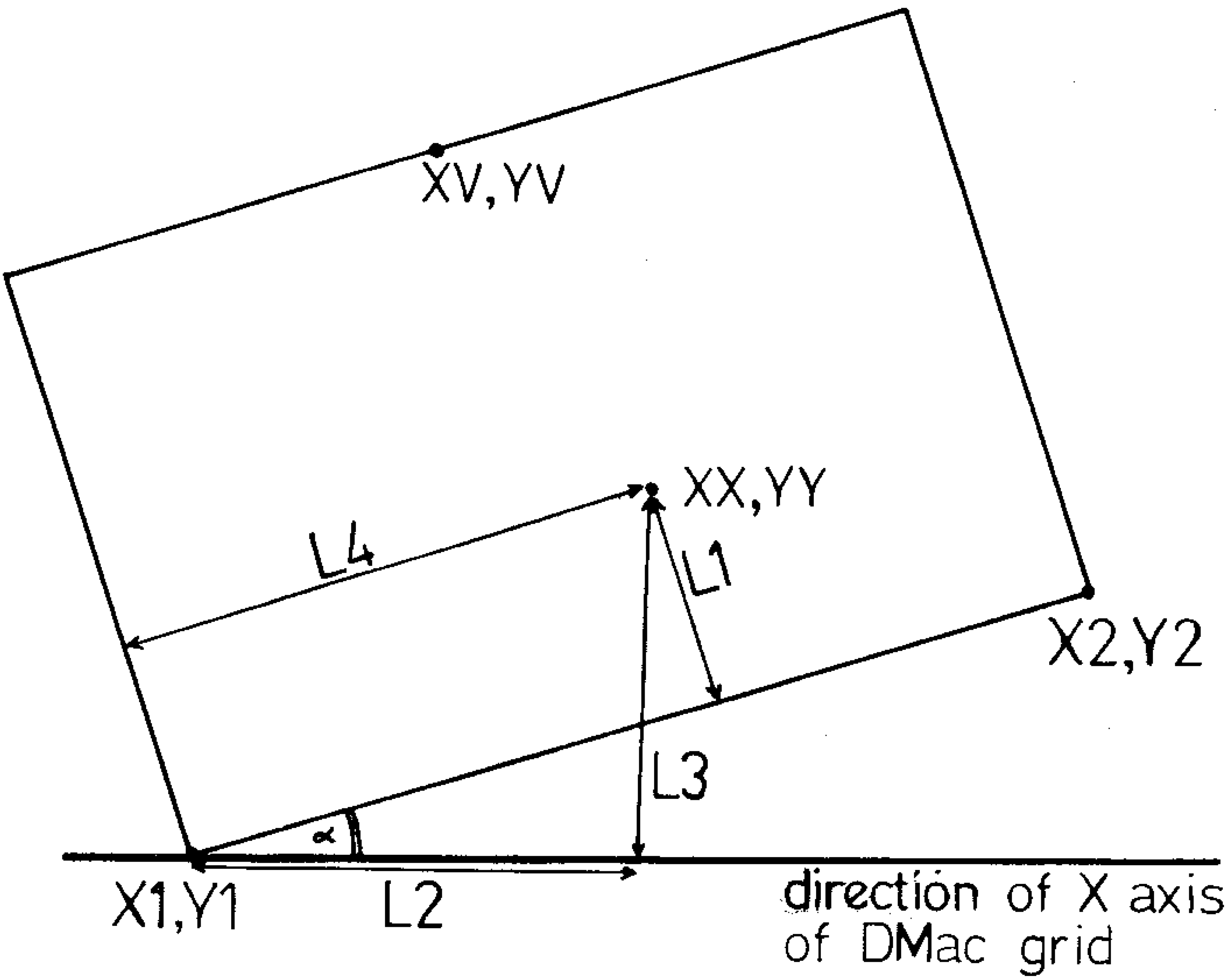


Figure 3.1 Correction for inclination of the chart on the DMac table

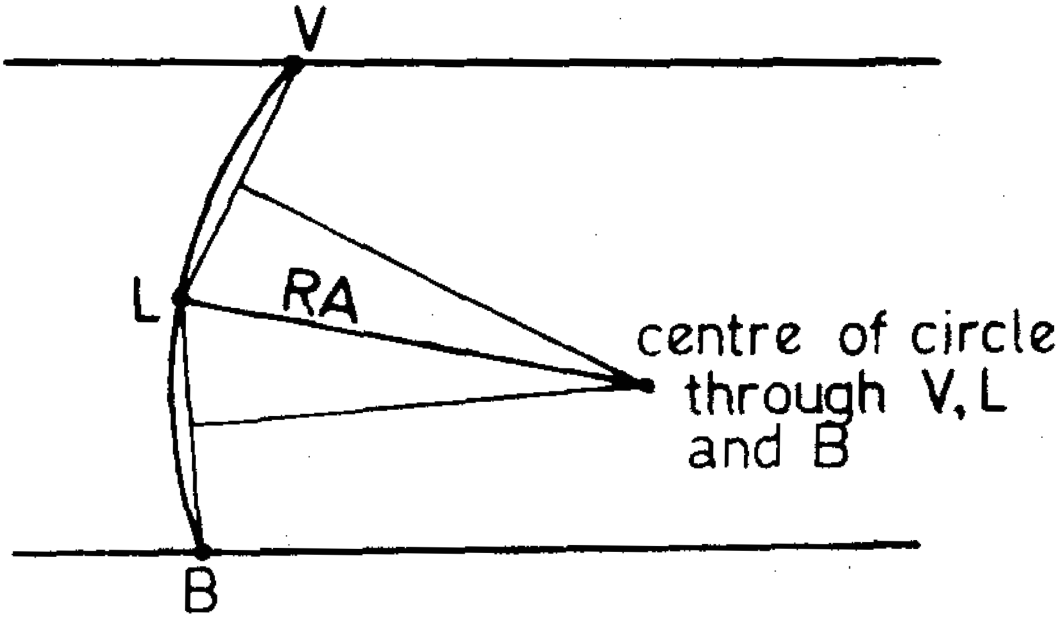


Figure 3.2A Calculation of radius of curvature for charts with curved vertical axis

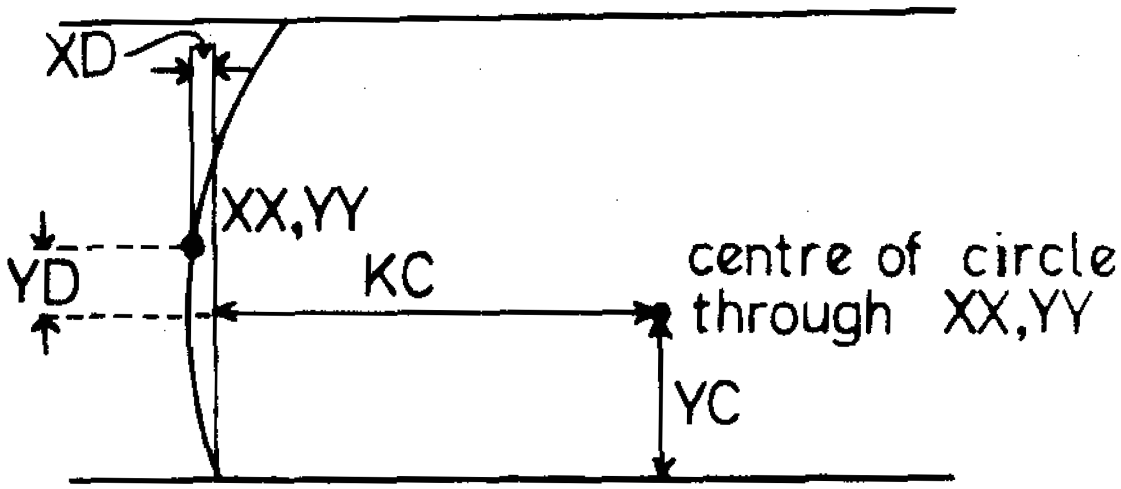


Figure 3.2B Correction of coordinates for curved vertical axis

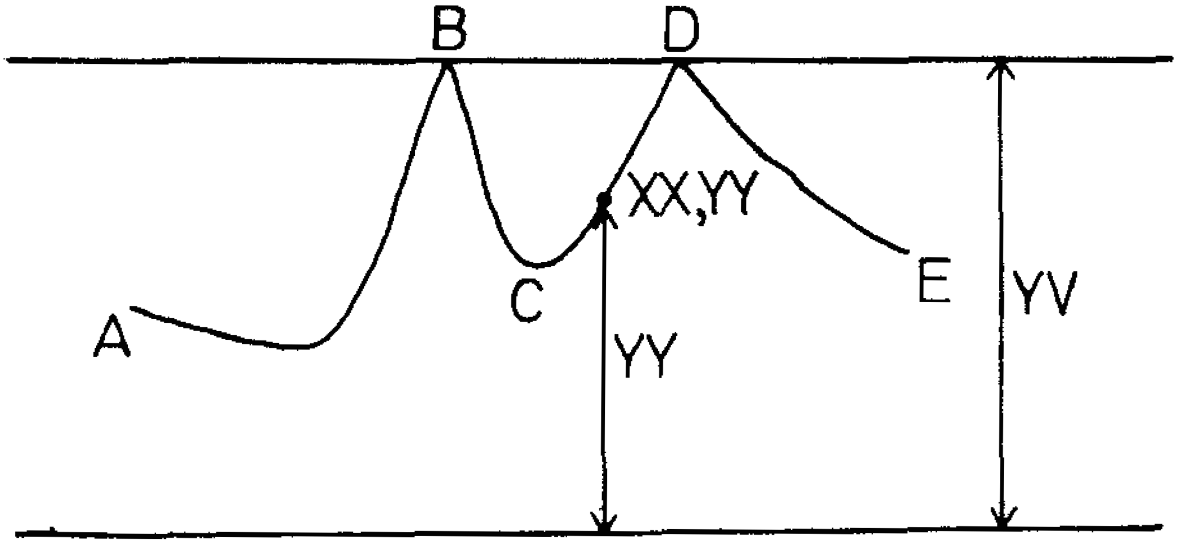


Figure 3.3 Correction of vertical scale readings for reversals

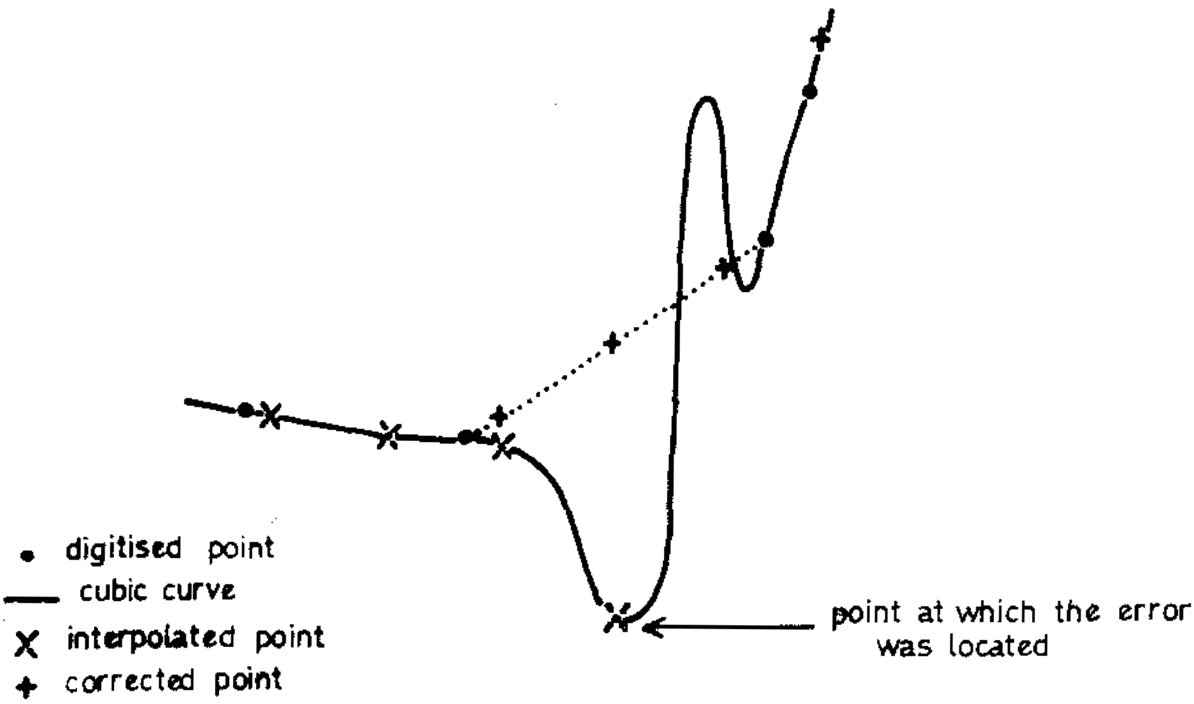


Figure 3.4 Location and correction of unrealistic cubic curve fitting

Figure 3.5 A An example of the results output by the streamflow preprocessing program

										91		6		170
	1521	0	370							1	2 1 3 2	631		170
309	309	308	306	305	303	301	300	298	297	1	1	6	1	170
295	293	291	289	287	285	282	279	276	273	1	2	6	1	170
270	265	260	252	246	237	229	222	214	207	1	3	6	1	170
200	194	190	184	179	174	170	165	160	155	1	4	6	1	170
151	148	146	144	143	145	149	155	164	174	1	5	6	1	170
186	199	212	229	237	243	242	241	239	235	1	6	6	1	170
232	228	225	221	217	211	205	198	191	186	1	7	6	1	170
182	177	172	166	161	156	153	149	146	143	1	8	6	1	170
140	138	136	133	131	129	127	125	124	123	1	1	6	2	170
122	121	120	119	118	117	116	115	115	117	1	2	6	2	170
120	127	136	147	160	190	228	252	264	274	1	3	6	2	170
279	282	283	282	280	277	275	271	267	261	1	4	6	2	170
254	246	240	234	229	224	219	214	210	206	1	5	6	2	170
202	198	195	192	189	186	184	181	179	176	1	6	6	2	170
174	172	170	168	166	164	162	160	159	157	1	7	6	2	170
156	155	153	152	150	149	148	146	145	144	1	8	6	2	170
143	142	140	139	138	136	135	134	133	131	1	1	6	3	170

Figure 3.5 B An example of the monitoring output from the streamflow preprocessing program

```
***********************************************
*RUN OF FLOW PREPROCESSING PROGRAM DATED JAN 1972*
***********************************************

              THE FOLLOWING FLOW DATA IS TO BE PROCESSED

                     DATA FROM SITE   6

                     6 MONTHS OF DATA,
                     COMMENCING AT 170

**************************************************************

    TIME SCALE NOT SUFFICIENTLY ACCURATE: TRAVERSE FROM  4  170

          DAY  1 HAS BEEN INTERPOLATED
          DAY  2 HAS BEEN INTERPOLATED
                 BAD-FIT CURVE AROUND READING 95
                 BAD-FIT CURVE AROUND READING 96
          DAY  3 HAS BEEN INTERPOLATED
                 BAD-FIT CURVE AROUND READING  1
          DAY  4 HAS BEEN INTERPOLATED
          DAY  5 HAS BEEN INTERPOLATED
          DAY  6 HAS BEEN INTERPOLATED
          DAY  7 HAS BEEN INTERPOLATED
          DAY  8 HAS BEEN INTERPOLATED
          DAY  9 HAS BEEN INTERPOLATED
          DAY 10 HAS BEEN INTERPOLATED
          DAY 11 HAS BEEN INTERPOLATED
          DAY 12 HAS BEEN INTERPOLATED
                 BAD-FIT CURVE AROUND READING  2
                 BAD-FIT CURVE AROUND READING  3
                 BAD-FIT CURVE AROUND READING  4
          DAY 13 HAS BEEN INTERPOLATED
          DAY 14 HAS BEEN INTERPOLATED
          DAY 15 HAS BEEN INTERPOLATED
          DAY 16 **NO DATA
          DAY 17 **NO DATA
          DAY 18 **NO DATA
          DAY 19 **NO DATA
          DAY 20 F & P CARDS USED
          DAY 21 F & P CARDS USED
          DAY 22 F & P CARDS USED
          DAY 23 F & P CARDS USED
          DAY 24 HAS BEEN INTERPOLATED
          DAY 25 HAS ONLY 95 READINGS

          NO MORE DATA THIS MONTH

     PROCESSING OF DATA FOR  170 COMPLETED
```

Figure 3.6 A An example of the results output by the rainfall preprocessing program

0	0	0	0	0	0	0	0	0	0	0	0	2	1	1	2	1	468
0	0	0	0	0	0	0	0	0	4	4	2	2	2	1	2	1	468
0	1	2	1	1	0	0	0	0	0	0	0	2	1	1	2	2	468
0	0	0	0	0	0	0	0	0	0	0	0	2	2	1	2	2	468
0	0	0	0	0	2	0	0	0	0	0	0	2	1	1	2	3	468
0	1	1	2	0	2	0	0	0	0	0	0	2	2	1	2	3	468
												2	1	0	2	4	468
												2	1	0	2	5	468
												2	1	0	2	6	468
												2	1	0	2	7	468
0	0	0	0	0	0	0	0	0	0	0	0	2	1	2	2	1	468
0	0	0	0	0	0	0	0	0	8	2	2	2	2	2	2	1	468
0	0	4	3	1	0	0	0	0	0	0	0	2	1	4	2	2	468
0	0	0	0	0	0	0	0	0	0	0	0	2	2	4	2	2	468
0	0	0	0	0	3	0	0	0	0	0	0	2	1	4	2	3	468
0	0	1	0	0	1	1	2	0	0	0	0	2	2	4	2	3	468
												2	1	0	2	4	468
												2	1	0	2	5	468
												2	1	0	2	6	468
												2	1	0	2	7	468

MONTHLY RECORDING RAINGAUGE SUMMARY
****

RAINGAUGE ; CATCHMENT NO. 2, GAUGE NO. 1

RAINFALL RECORDED IN 468

UNITS ; MM/10

DATE ****	TIME - O'CLOCK 09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	DAILY TOTAL
1 468		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4	2	10
2 468		0	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
3 468		0	0	0	0	0	2	0	0	0	0	0	0	0	1	1	2	0	2	0	0	0	0	0	0	8
4 468	NO RAIN																									
5 468	NO RAIN																									
6 468	NO RAIN																									
7 468	NO RAIN																									
8 468	NO RAIN																									
9 468	NO RAIN																									
10 468	GAUGE NOT WORKING																									
11 468	GAUGE NOT WORKING																									
12 468	GAUGE NOT WORKING																									
13 468	NO RAIN																									
14 468	NO RAIN																									
15 468		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	4
16 468		4	2	0	0	0	2	4	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	15

Figure 3.6 B An example of the monthly raingauge summary sheet output by the rainfall preprocessing program

4. REFERENCE SECTION

4.1 Examination of charts

In order to use the d-mac system to the full by eliminating the errors mentioned in earlier sections, a certain amount of preparatory work must be undertaken. A standard form (shown as Figure 4.1 for continuous charts such as Leupold and Stevens stage charts and Figure 4.2 for Dines raingauge charts, where only short portions of the chart will be digitised) should be completed as outlined. Office annotation should be readily distinguishable from remarks written on the chart in the field.

At the beginning of the chart and on the inspection form write:

a) the name and/or number of the gauge
b) the date of the beginning of the chart
c) the data of inspection of the chart
d) signature of inspector.

Working through the chart from beginning to end, write on the inspection form:

a) - in column 5 - the time, in G.M.T., and date of each visit to the recorder, provided that the chart trace is marked at this time.
b) - in 'comments' column - vertical scale information such as the position on the chart scale of reversals, raingauge syphonings, and any independent measurements.
c) mark the chart with the following symbols, which will be keyed in during digitisation and used in the interpretation program:

   " - 1" for raingauge syphoning
   " - 2" for a reversal during rising stage
   " - 3" for a reversal during falling stage.

d) (only for charts which have days and hours printed) check whether the times of any visits and that at which the chart was removed agreed with the times on the chart grid. If they do, the time scale correction procedures, section 4.2, can be omitted.

4.2 Time scale correction procedures

Calculate the distance between the time marks. Working back through the chart, count the number of grid squares between each of the previously noted time marks. Charts are normally divided by bold lines at, say, three hourly intervals, and faint lines at say quarter-hourly; count the small squares at either end and the number of blocks between these. Write these totals in column 7 of the form.

Calculate the scales between these time marks. First calculate the time intervals between the entries in column 5, and enter it in column 6. Then calculate the time scale for each interval (in grid squares per hour) by dividing the figure in column 7 by that in column 6. The result should be entered in column 8.

The entries in column 8, that is, the time scale between each visit, should now be examined. If the time scale in any interval is outside the tolerable range given in Figure 4.3 for various charts, it should be considered suspect, and unless covered by a comment such as "clock stopped" should be checked by recounting the squares, and repeating the arithmetic. If the error is not corrected and cannot be located by a more thorough examination of the chart, expert advice should be sought.

If the error is found to be a machine fault, the interval in which it occurs should be ignored in subsequent calculations of mean time scale and a comment to this effect should be made on the inspection form. If a change in chart time scale is apparent during the chart length, it may be appropriate to divide the chart into sections, the mean time scale being evaluated for each.

The mean time scale should now be calculated for each section of the chart, using all intervals except those eliminated during the examination above. Sum the duration, column 6, keeping a running total in column 9. Sum the number of grid squares, column 7, keeping a running total in column 10.

Calculate:

$$\text{Mean time scale for chart section} = \frac{(\text{Final total in column 10})}{(\text{Final total in column 9})} \text{ grid squares/hr}$$

The mean time scale is the figure which should be used for annotation of the chart section to which it refers.

4.3 Chart annotation

It should first be decided whether a continuous digitisation is required, or whether digitisation of selected portions of a chart is all that is necessary. The latter is the case with raingauge charts, which need only be digitised for periods of rainfall. For a continuous digitisation, annotation should be made according to the flow diagram, Figure 4.4 while for a selective digitisation, the flow diagram, Figure 4.5, should be used.

4.4 Digitisation

Having completed the above preparation, the chart can now be digitised using the d-mac pencil follower. The basic method of digitisation is to place the autographic chart on to the d-mac table and position it such that a section (delimited by 'x's during annotation) is central on the table. Using the portable keyboard, a header label is punched on to the output paper tape and the location of the chart on the d-mac coordinate system is established by punching points at the left hand and then the right hand 'x' on the chart base line. The upper calibration point on the full scale line should then be digitised. These points are shown on Figure 4.6(a) and they locate the chart on the table and are used in all subsequent calibrations.

If the stage axis of the chart is curved, two curvature calibration points are taken on the same stage axis as the upper calibration point, as shown in Fig. 4.6(b); (XL, YL) is taken first, then (XB, YB). Now the time and vertical scales are set by punching on the keyboard the number of hours (x 100) between the two baseline calibration points and the true range for the full scale deflection on the vertical axis. Having located and scaled the chart, it is now possible to follow the autographic trace, digitising sufficient points along the trace to avoid any ambiguity in the reconstruction of the curve. It should be remembered that subsequent curve fitting relies on the existence of detailed information at points where the gradient of the trace is changing rapidly. The trace is followed across the chart; pen reversals, raingauge syphonings, etc., are noted by punching coding from the keyboard as specified in the appropriate flow diagram, Figure 4.7 or 4.8. Having reached the right hand end of the chart section, the traverse is terminated by a symbol punched in from the keyboard and the tape is then either terminated or, if more traverses are to follow, they are introduced by new header coding. Figure 4.7 for stream stage charts and Figure 4.8 for recording raingauge charts are provided for use as a reference during chart digitisation and should be followed carefully.

4.5 Errors: their location and correction

In preparing a tape, particularly a long tape with several chart traverses, it is likely that errors will be made. They will be:

a) operator unaware	:	data points incorrectly taken
b) operator unaware	:	simple omission or mispunching of control data, spaces or new-line characters
c) operator immediately aware	:	errors noticed should be noted in the operator's log to facilitate correction.

The prepared paper tape has all the necessary headings and can now be input to the computer. At the end of the tape a short program has been punched to list the contents of the file into which the tape was read. The tape should, therefore, be returned from the computer with a lineprinter listing of its contents. It should be noted that the specific instructions given at the end of Figures 4.7 and 4.8 are applicable only when using an ICL 1900 series computer with the GEORGE III operating system, and they should be amended if a different operating system is in use. The listing should be examined; the errors noted in the operator's log should be noted and marked and any other errors should be found by a careful inspection of the listing, although type a) errors (which will occur only very rarely) may not be at all obvious.

After this examination, all the errors that have been found should be corrected using the editing facilities of the computer. This will not physically correct the data tape, but will correct the contents of the computer file into which the tape was read. Editing techniques vary from computer to computer, but those in use on the ICL 1903A (or any 1900 series machine using the GEORGE III operating system) are summarised in the ICL manual, Operating Systems GEORGE III and IV pp 110-113.

The corrected data files should now be prepared for use with the computer 'preprocessing' programs. It is convenient to process the data in batches of several months and this will require merging of several monthly data files. The method of batching will depend on the computer system; on the 1903A it is normally done by using the 'MERGE' facility in the GEORGE EDITOR and it is described in the above reference.

4.6 Assembly of card data for streamflow

Card files of two types may be necessary for a preprocessing program run:

a) control cards, which are always required
b) supplementary hydrological data when necessary for filling gaps in the paper tape record.

A file must always be input to the computer filestore, containing the following cards.

INPUT :CATCH,CONTROLCARDS

RUN CARD

ITEM	SYMBOL	FORMAT	COLUMNS	REMARKS
Flow symbol	MTYPE	I1	5	always 1
Catchment Number	MSITE	I3	8 - 10	
Number of months to be processed	MONTHS	I2	14 - 15	
Starting month and year	NOMONTH	I4	17 - 20	

then, for each month

LEAD CARD

ITEM	SYMBOL	FORMAT	COLUMNS	REMARKS
Lead symbol	NINE	I1	61	always 9
Flow symbol	LTYPE	I1	62	always 1
Catchment Number	LSITE	I3	72 - 74	
Month and year	LMONTH	I4	77 - 80	

and,

CONTROL CARD

ITEM	SYMBOL	FORMAT	COLUMNS	REMARKS
Catchment area	AREA	F10.1	1 - 10	units hectares
Zero correction	ZERO	F5.0	11 - 15	units same as stage
Max stage in month	SMAX	F5.0	16 - 20	"
Flow symbol	ITYPE	I1	62	always 1
Interpolation index	INTERP	I1	65	1 = on interval 2 = mid interval;
Input frequency index	IFREQ	I1	67	1 = 15 mins 2 = 30 mins 3 = 1 hour 4 = 2 hours 5 = 3 hours6 = 6 hours 7 = 12 hours8 = daily
Output frequency index	JFREQ	I1	69	same as IFREQ
Units index	INDEX	I1	71	1 = stage x 10^{-3} ft 2 = stage mm 3 = flow x 10^{-5} ins/hr 4 = stage x 10^{-2} ft
Catchment number	ISITE	I3	72 - 74	
Number of days in month	IDAYS	I2	75 - 76	
Month and year	IMONTH	I4	77 - 80	

and, for each data batch within the month

TYPE CARD

ITEM	SYMBOL	FORMAT	COLUMNS	REMARKS
Number of days in batch	NDAYS	I2	4 - 5	
Source of data	NDATA	I1	10	1 = d-mac tape 2 = F and P cards 3 = time/stage cards 4 = data missing

then, at the end of the file, a single card terminating the input,

****

When supplementary data cards are being used to fill gaps in the autographic chart record, they should be presented as follows.

a) Stream stages translated by Fischer and Porter from 16-hole punched paper tape. These cards each contain 12 stream stage values, but no other information such as times, dates, gauge numbers etc. Batches of these cards may be used to fill in gaps in the autographic record, provided that each batch is identified separately with this missing information. The file should be input to the filestore with the following cards:

```
INPUT :CATCH,CARDDATAFILE
```

then, for each batch

HEADER CARD

ITEM	SYMBOL	FORMAT	COLUMNS	REMARKS
Type of cards	KDATA	I1	5	2 = Fischer and Porter
Catchment Number	KSITE	I3	8 - 10	
Date of first card	KDATE	I2	14 - 15	
Month and year	KMONTH	I4	17 - 20	
Time (x 100 hours of start of useful data from start of first card	KTIME	I4	22 - 25	
Frequency of data on these cards	KFREQ	I1	30	as IFREQ (see Control card)

then the batch of data cards, in their standard format; at the end of the batch there must be one blank card, and after the blank card at the end of the final batch, a file terminator card.

****

b) If the Fischer and Porter cards are not available for filling gaps in the autographic record, cards containing pairs of time and stage readings can be assembled to form a batch within "CARDDATAFILE" The batch needs no header, as each card carries identification. Card format is as follows:

ITEM	SYMBOL	FORMAT	COLUMNS	REMARKS
Data points, time and stage (7 pairs)	(KX(N),KY (N), N = 1, 7)	7(2I5)	1 - 70	
Catchment number	KSITE	I3	72 - 74	
Date	KDATE	I2	75 - 76	
Month and year	KMONTH	I4	77 - 80	

Each batch should be terminated with a blank card.

Obviously, there are many other ways of accepting alternative stage data and any change must be consistent with a change in the appropriate format in the program.

4.7 Running the flow preprocessing program under GEORGE III

The program is run with a custom-built macro, "FLOWPPRUN", which must be supplied with the following as parameters:

a) d-mac paper tape file name (blank if not present)

b) hydrological data card file name (blank if not present)

c) control cards file name

d) preprocessed output file name.

The job would have the form:

```
JOB MYJOB,:CATCH
FLOWPPRUN DMACFILENAME,CARDDATAFILE,CONTROLCARDS,OUTPUTFILE
ENDJOB
****
```

The macro reads the preprocessing program from magnetic tape into core, assigns peripherals to the data and output files, and enters the program. If the program terminates in the normal way, the input data files are erased. Otherwise they are left in core, to enable a rerun of the program to be made with appropriate corrections. The program itself is removed from core and all outputs are listed on the line-printer. The format of the output file is suitable for input to the Institute's data processing package.

The name of the output file is made up as follows:

$$CCC\ t\ mm_s\ mm_f\ yy$$

where CCC are the first three letters of the catchment name

- t is the data type, in this case 'F' for flow
- mm_s is the starting month
- mm_f is the final month
- yy is the year

The preprocessing program is stored in file FLOWPPPROG, and the compiled program (ready for running) in FLOWPPBIN. The run macro FLOWPPRUN and the program FLOWPPPROG are listed in the appendix to this manual.

4.8 Assembly of card data for recorded rainfall

A file must be input to the filestore containing the following cards:

```
INPUT :CATCH,CONTROLRFALL
```

and then, to describe each batch of data

RUN CARD

ITEM	SYMBOL	FORMAT	COLUMNS	REMARKS
Rainfall Symbol	ITYPE	I1	5	always 2
Gauge number	IRG	I2	9 - 10	
Catchment number	ISITE	I3	13 - 15	
Number of months to be processed	MONTHS	I2	19 - 20	
Starting month and year	NOMONTH	I4	22 - 25	
Frequency of output	IFREQ	I1	30	

then a blank card and a file terminator card

****

4.9 Running the rain preprocessing program under GEORGE III

The program is run with a custom-built macro, "RAINPPRUN", which must be supplied with the following as parameters:

a) d-mac paper tape file name

b) run control cards file name

c) preprocessed output file name.

The job would have the form:

```
JOB MYJOB,:CATCH
RAINPPRUN DMACFILENAME,CONTROLRFALL,OUTPUTFILE
ENDJOB
****
```

The macro reads the preprocessing program from magnetic tape into core, assigns input and output peripherals and enters the program. If the program terminates normally, input files are erased and output listed and retained; otherwise the output files are listed and erased and the input files retained to enable a rerun of the program to be made. The program is removed from core.

The output filename is constructed as follows:

$$R\ nnn\ gg\ mm_s\ mm_f\ yy$$

where	R	indicates rainfall
	nnn	is the catchment number
	gg	is the raingauge number
	mm_s	is the starting month
	mm_f	is the final month
	yy	is the year

for example, gauge 2 at Coalburn, January to June 1972, would be output to a file: R00602010672

When data from all the recording raingauges within a catchment have been processed, the output files can be merged into one large file, which will contain all the recording gauge information for the catchment. The gauge number in the original filename can be replaced by 'XX' to indicate that this is the completed file.

4.10 Sorting the output cards into the correct order

The order of cards in the output file will not be suitable for input to the rainfall processing programs, as a chronological sorting, rather than just a sorting by gauge number, is required. Period gauge cards and control cards must also be incorporated into the data file at this stage. To rearrange the data cards, a SORTING program must be run on the computer.

This program reads each card and slots it in the correct position in the output file according to various indices, such as date, gauge number and card number etc.

A custom built macro is used to run the sort program, and the following must be given as parameters:

a) recorded rainfall filename
b) period gauge and control card filename
c) output filename.

The name of the output file is made up as follows:

$$CCCtmm_s mm_f yy$$

where
CCC are the first three letters of the catchment name
t is the data type, in this case 'R' for rainfall
mm_s is the starting month
mm_f is the final month
yy is the year

The SORT job would have the form:

```
JOB MYJOB,:CATCH
SORTRAIN RECDRAINFALL,PERDRAINFALL,OUTPUTFILE
ENDJOB
****
```

Figure 4.1 Inspection form for continuous stage charts

1. Site of the recorder ______________

2. Chart put on at _____ hours GMT on ___/___/19___

3. Chart taken off at _____ hrs GMT on ___/___/19___

4. Chart examined by ______________

Detail of chart time scale:

5. (a) (b) time/date GMT of visit	6. hours since last visit	7. grid lines since last visit a1+B*12+a2	8. interval time scale = 7./6.	9. hours from start of chart = 6.	10. grid lines from start of chart = 7.	11. comments on trace; position of reversals,etc

12. Mean time scale _______ grid lines per hour (=TS)

13. a) Grid lines in 1 day (TS * 24) __________

b) Grid lines in 2 days (TS * 48) __________

c) Grid lines in 3 days (TS * 72) __________

Figure 4.2 Inspection form for recording raingauge charts

1. a) Catchment number ________

   b) Gauge number ______

2. Chart put on at ___ hours GMT on ___/___/19___

3. Chart taken off at ___ hrs GMT on ___/___/19___

4. Chart examined by __________

Detail of chart time scale:

5. (a) time GMT of visit	5. (b) date of visit	5. (c) time GMT on chart	6. hours since last visit	7. grid lines since last visit a1+B*10+a2	8. interval time scale = 7./6.	9. hours from start of chart = 6.	10. grid lines from start of chart = 7.

11. Comments on trace; location of baseline and full scale from gauge syphonings.

12. Mean time scale ______ grid lines per hour

FIGURE 4.3 Tolerable ranges in time scale for various types of autographic chart.

Type of recorder	Speed of chart	Tolerable range of time scale (grid squares per hour)
Leupold and Stevens stage recorder	9.6 inches per day	3.96 - 4.04
Leupold and Stevens stage recorder	4.8 inches per day	1.98 - 2.02
Dines weekly recording raingauge	6 inches per hour	59.4 - 60.6
Dines *monthly* recording raingauge	1 inch per hour	9.9 - 10.1

Figure 4.4 STEP DIAGRAM FOR CHART ANNOTATION

(for continuous digitisation)

STEP 1 : Calculate MAXDAYS = the maximum number of days which will fit within the d-mac coordinate system; given the approximate time scale of the chart.

STEP 2 : a) find the start of the chart;

b) locate the start of the first whole day, as follows,

: calculate the time from the start of the chart to 0900 hours,

: using the current time scale, calculate the number of grid squares,

: count these grid squares on the chart,

: mark the true baseline with a 'X' at this time and rule a line from the baseline to full scale.

c) set N to MAXDAYS - 1

STEP 3 : If there is a break in the chart, or the end of the chart occurs in the next N days, reset N to the number of whole days before the discontinuity.

STEP 4 : If the end of the month occurs within the next N days reset N to the number of days up to the end of the month, and set ENDSWITCH to ON.

STEP 5 : a) Locate the end of day N, as follows:

: find the number of grid squares in N days, using the current time scale

: count these grid squares on the chart,

: mark the true baseline at this time with an 'X', and rule a line from the baseline to full scale.

: find the position of the true full scale from the form, and mark the true full scale with an 'X' midway between this and the previous baseline mark. See Figure 4.6 (a).

: If the chart has a curved vertical axis, mark the chart with the additional calibration points shown in Figure 4.6 (b).

b) If the ENDSWITCH is ON, go to STEP 9. If there is a break in the chart, go to STEP 6.

c) Set N to MAXDAYS, and go to STEP 3.

STEP 6 : a) Write BREAK IN CHART on the chart.

b) Find the next time mark on the chart.

c) Locate the start of the day on which this time mark was written in the same way as in STEP 2 (b)

d) Set N to MAXDAYS - 1, and continue with STEP 7.

STEP 7 : a) If the break in the chart occurred within these N days, reset N to the whole number of days after the break, and set BRKSWITCH to ON.

b) Locate the start of the day N days prior to the last mark, and mark the true baseline and full scale as described in STEP 5 (a).

c) If the BRKSWITCH is ON, go to STEP 8.

d) Set N to MAXDAYS and repeat STEP 7.

STEP 8 : a) Return to 0900 hours of the day of the last time mark, and reset BRKSWITCH to OFF.

b) Set N to MAXDAYS, and return to STEP 3.

STEP 9 : Write END OF MONTH on the chart, reset ENDSWITCH to OFF, set N to MAXDAYS, and return to STEP 3.

Figure 4.5 STEP DIAGRAM FOR CHART ANNOTATION

(for selective digitisation)

STEP 1 : Calculate MAXHOURS = the maximum number of hours which will fit within the d-mac coordinate system; given the approximate time scale of the chart.

STEP 2 : a) Find the start of the chart;
b) Locate the start of the first whole hour, as follows,
: calculate the time from the start of the chart to the hour,
: using the current time scale, calculate the number of grid squares,
: count these grid squares on the chart,
: mark the hour on the chart with a vertical bar, time and date;
c) Set N = MAXHOURS

STEP 3 : If there is no rainfall in the next N hours, go to STEP 4.
a) Set N = number of whole hours before rainfall;
b) Set RAINSWITCH ON.

STEP 4 : If there is no break in the continuity of the chart in the next N hours, go to STEP 5.
There is no useful data for this day.
a) Delete all previous rain periods since 0900 of this day;
b) Search for the next time mark;
c) If there was no rain between the day of the break and the time mark, go to STEP 2 (b).
d) Locate the start of the hour with the time mark, as follows,
: calculate the time from the hour to the time mark,
: using the current time scale, calculate the number of grid squares,
: count these grid squares on the chart,
: mark the start of the hour with a vertical bar, time and date:

e) Set N = MAXHOURS

f) If the break occurred within N hours before the bar, reset N to the highest whole number of hours after the break. Set BREAKSWITCH to ON.

g) Find the start of the hour N hours prior to the last mark, as follows,

: find the number of grid squares in N hours from the form,

: count these grid squares on the chart,

: mark the hour on the chart with a vertical bar, time and date;

h) If BREAKSWITCH is ON, find the hour mark after the time mark,

: set N to MAXHOURS, reset BREAKSWITCH to OFF, and go to STEP 3.

i) Go to STEP 4 (e)

STEP 5 : a) Locate the end of hour N, as follows,

: find the number of grid squares in N hours from the form,

: count these grid squares on the chart,

: mark the hour on the chart with a vertical bar, time and date;

b) If RAINSWITCH is OFF, set N to MAXHOURS and to STEP 3.

c) Mark the true base line at the start of the first hour with rainfall with an 'X'.

d) Search for the end of the rain

: Set N to the number of whole hours containing rainfall

: If N is greater than MAXHOURS, reset N to MAXHOURS

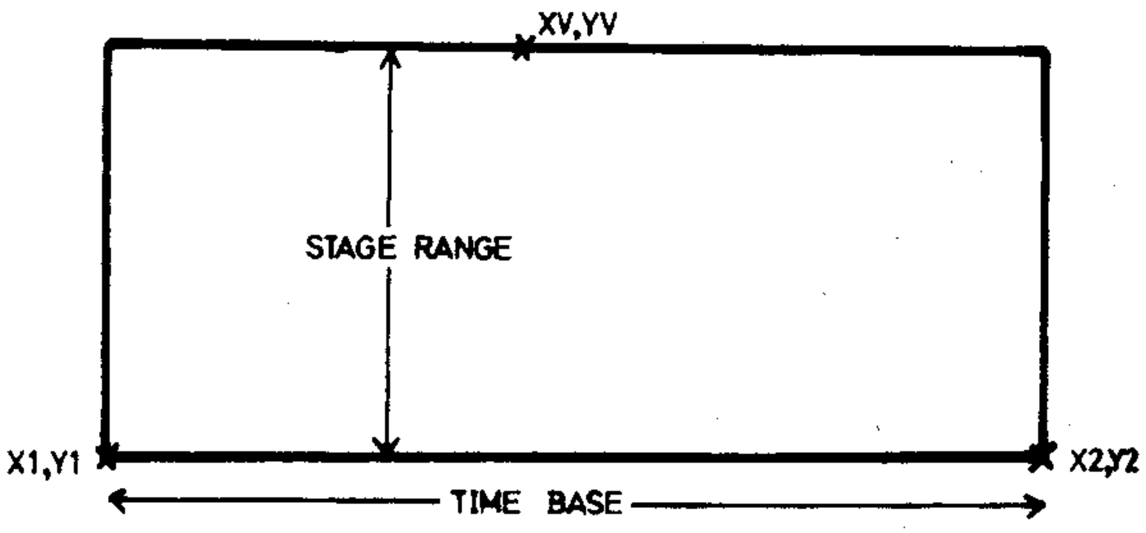
e) Locate the end of hour N, as before (STEP 5 a)).

f) Mark the true baseline with an 'X' at this time.

g) Midway between this and the previous mark, mark the true full scale with an 'X'.

h) Set the RAINSWITCH to OFF, and return to STEP 3

(a) Straight stage chart



(b) Curved stage chart

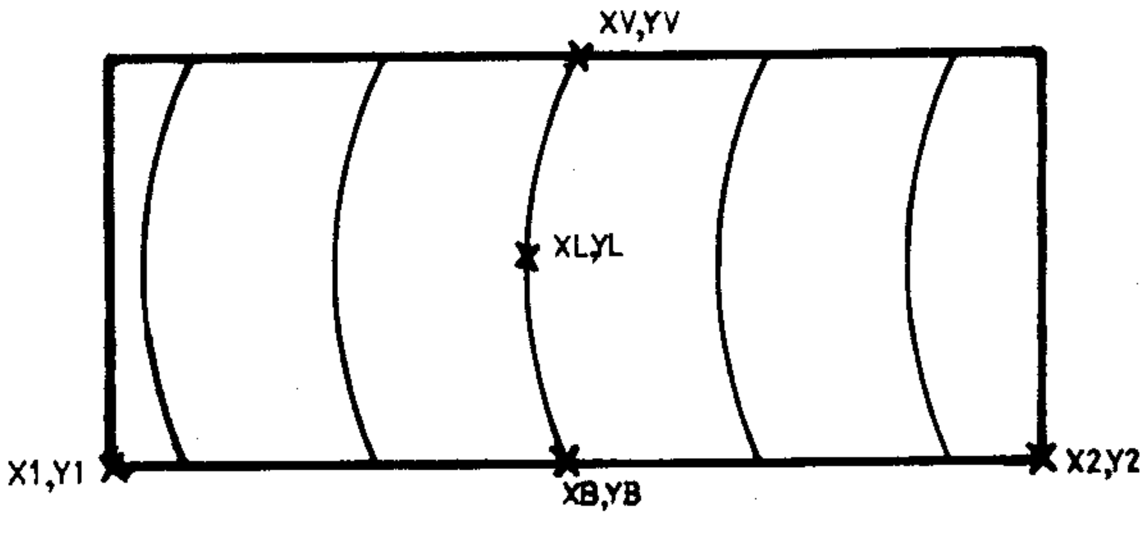
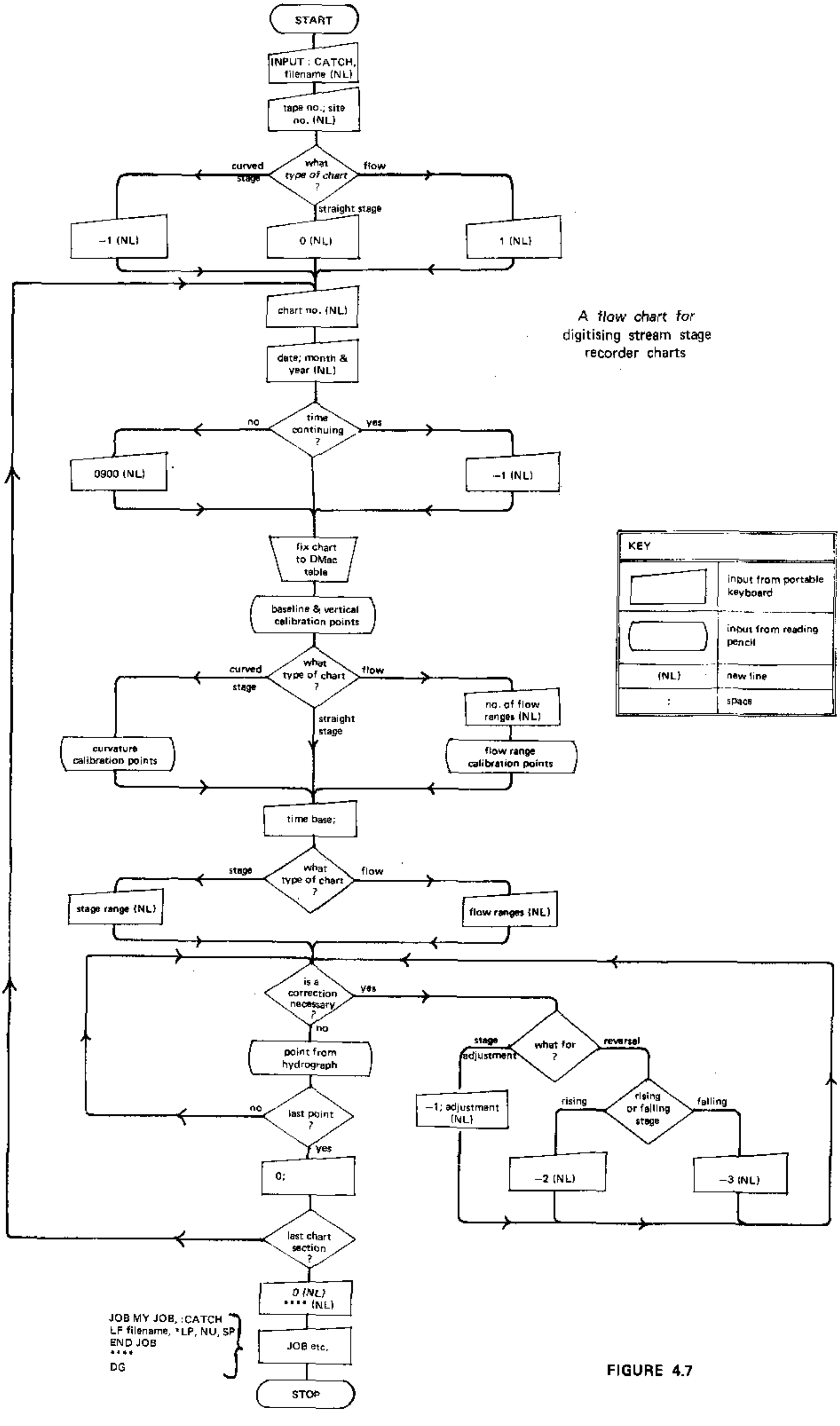


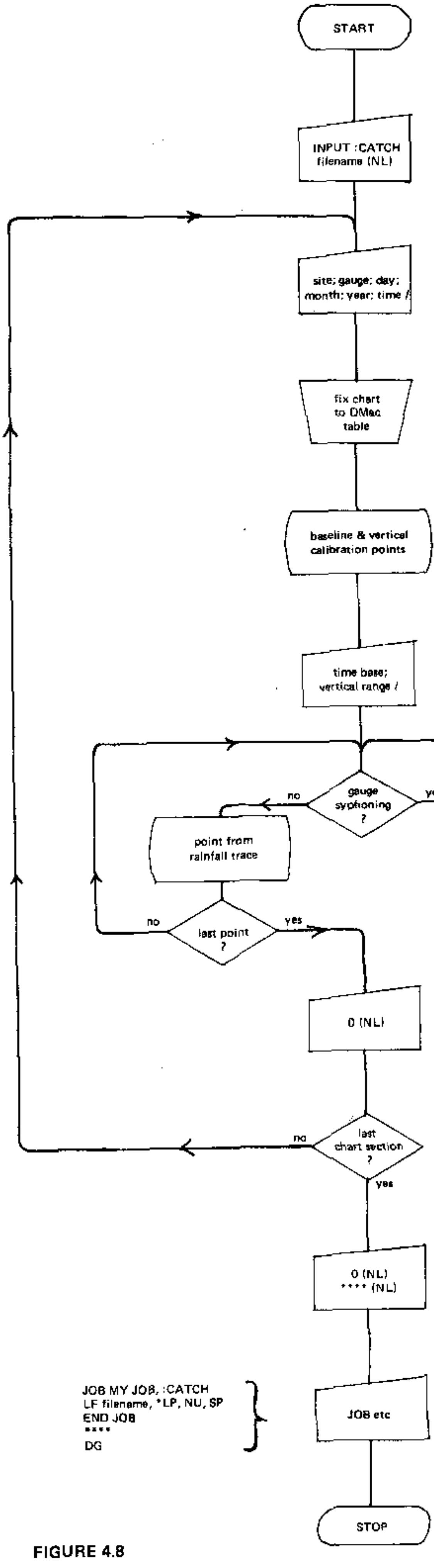
Figure 4.6 The position of calibration points



A flow chart for digitising stream stage recorder charts

FIGURE 4.7

A flow chart for digitising recording raingauge charts



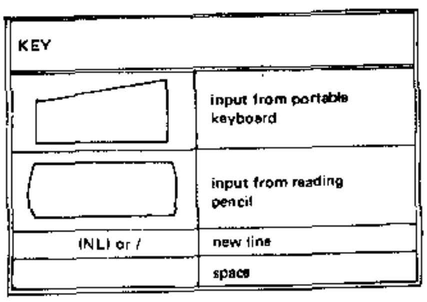


FIGURE 4.8

Report No. 6	1969	"A feasibility study and development programme for continuous dilution gauging", by P. H. Hosegood, M.K. Bridle and P. W. Herbertson
Report No. 7	1969	"Installation of access tubes and calibration of neutron moisture meters", by C. W. O. Eeles
Report No. 8	(in press)	"The soil hydrology of the Plynlimon experimental catchments", by J. P. Bell
Report No. 9	1971	"River level sampling periods", by P. W. Herbertson, J. R. Douglas and A. Hill
Report No. 10	1971	"User's testing schedule for the Wallingford Probe system", by P. M. Holdsworth
Report No. 11	1971	"Report on precipitation", by J. C. Rodda
Report No. 12	(in press)	"Estimating the T-year flood for a catchment with only a short period of record: some methods applied to mean daily discharge from the River Eden at Temple Sowerby", by J. Frost
Report No. 13	1971	"A computer program to use orthogonal polynomials in two variables for surface-fitting", by D. Richards
Report No. 14	1972	"A system for processing data from autographic recorders", by J. R. Douglas
Report No. 15	(in press)	"The quality control and processing of streamflow, rainfall and evaporation data", by D. T. Plinston and A. Hill

REPORTS OF INSTITUTE OF HYDROLOGY

INSTITUTE OF HYDROLOGY

Report No. 1	1970	"Publications Handbook. Guide to preparation of manuscripts for publication. Papers for submission to external journals. Institute of Hydrology Reports", by B. C. Kennedy

HYDROLOGICAL RESEARCH UNIT

Report No. 2*	1967	"Snow in Great Britain", by P. C. Rowe
Report No. 3*	1968	"Evaporation from forests", by J. B. Stewart
Report No. 4*	1968	"Design of steep stream gauging structures for the Plynlimon experimental catchments", by P. W. Herbertson

INSTITUTE OF HYDROLOGY

Report No. 5	1969	"Processing errors in the analysis of catchment data", by A. N. Mandeville, D. T. Plinston and A. Hill

* Out of print