LONDON, METEOROLOGICAL OFFICE.

Met.O.3 Technical Note No. 6.

Areal quality control of daily climatological data using station factor scores. By SPACKMAN, E.A.

London, Met.Off., Met.O.3 Tech.Note No.6, 1980, 31cm. Pp.iv+6, 8 pls.+App.17.4 Refs. Abs.p.ii.

An unofficial document - restriction on first page to be observed.

MET O 3 TECHNICAL NOTE NO 6

Areal Quality Control of Daily Climatological Data

using

Station Factor Scores

by

E A Spackman

November 1980

Met O 3
(Climatological Services Branch)
Meteorological Office
London Road
Bracknell
Berkshire

Summary

This paper describes a method for the detection of errors in daily climatological data which is based on the use of a set of empirical orthogonal functions. It also gives technical information on the computer programs used and gives a summary indicating how well data are 'fitted'.

Contents

3. Factor Score Data Sets

4. Query Data Sets

5. Statistics Data Set

6. Station position Data Set

I. Program software

1. The technique

1.1. Introduction

A new method has been developed to detect observations of daily climatological data in the UK which may be in error. An estimate is made at each station using a set of factors determined over all stations in the network. This estimate is based on the inter-relations between observations at all stations and, implicitly, should have taken into account the geographical location of the station, its altitude and any special characteristics that will consistently affect the observations. Since early in 1980, this method has replaced (for England and Wales) that described by Bryant, 1979. His method obtained an estimate by taking a mean over all the observations, after they had been normalised, for an appropriate quality control area. Each of these areas contained about 10 stations and thus did not take into account the variation of the element over those stations for the particular day being considered except insofar as it was reflected in the monthly mean.

1.2. Method

A set of factors (usually 15) have been determined, for each element to be checked, using the method described by Spackman, 1979. The form of Factor Analysis is in fact a Principal Components Analysis with Varimax rotation which is performed in such a way as to provide a set of 15 Factor Scores for each station. In order to detect suspect data the residuals are studied after fitting the factors to the data by linear regression. The model is given by

$$e^*(s,d) = a_0(d) + a_1(d).f_1(s) + a_2(d).f_2(s) + \ldots\ldots a_n(d).f_n(s)$$

where e^* is the estimated value of the parameter e at station s on day d and

$a_i(d)$ is the loading of factor i on day d

$f_i(s)$ is the score for factor i at station s

The residual R(s,d) is given by

$$R(s,d) = e(s,d) - e^*(s,d)$$

An observation is defined as suspect if

$$S = \left| \frac{R(s,d) - \bar{R}(s,d)}{\sigma(R(s,d))} \right| > 3.25$$

where S is the 'standardised deviation'

where $\bar{R}(s,d)$ is the mean residual

and $\sigma(R(s,d))$ is the standard deviation of the residuals.

For this test the mean and standard deviation are computed for the climatological area of the UK in which the station lies. The areas consist of

- Orkney, Shetland and Caithness
- Isle of Man
- Channel Islands
- the 10 Climatological districts (excluding stations in the above areas) shown in figure 1.

Suspect data are plotted on a pseudo-map on the line printer with 6 plots per page. Each plot covers an area of about 100 km x 100 km and contains all stations in the area. An example of the output is given in figure 2.

At the time of writing (November 1980) six elements are checked for quality by this method - maximum screen temperature, minimum screen temperature, grass minimum temperature, dry bulb temperature at 0900, wet bulb temperature at 0900 and daily sunshine duration. Appropriate sets of factors have been determined from observations taken at 3-day intervals during 1973-77. Stations with more than about 10 months observed data in one of the years 1972, 1978 as 1979 have also had factors estimated. Stations with no factors (because they had little or no data between 1972 and 1979) have them estimated (prior to the residual analysis) from whatever data are available for the month being processed.

1.3. Performance

The performance of the method can be judged, to some extent, by studying summary statistics of the residuals. The mean and standard deviation of the daily residuals at each station for each month and element have been computed. In most cases the mean residual is near zero usually being less than 0.5 degC for temperature and 0.5 hrs for sunshine. When the mean residual is much larger it can be used to identify stations where the calibration of the instruments should be checked. Values of the standard deviation appear to change with some consistency across the country and in some elements through the year. In order to present these values in a suitably condensed form the standard deviation of the daily residuals at each station has been averaged over all the stations in each climatological district. The values for 1978 are presented for 5 districts in figure 3. It can be seen that in general the standard deviation increases from south (district 5) to north (districts 0 and 1). It is also noted that there appears to be an annual variation - particularly for maximum temperature and sunshine. The annual variation for sunshine virtually disappears if the monthly values are scaled by dividing by the maximum possible hours of sunshine for that month.

There is sometimes considerable variability in the standard deviations from month to month - note for instance the relatively low values for minimum temperatures in March 1978 and the generally high values in January 1979 (a snowy, cold month). Summaries of these data for one year give only a general idea of the annual variation, if any, and more years need to be studied. Figure 4 gives the standard deviation of the daily residuals at each station averaged over all stations for 1978 and several months in 1979 and 1980. In this diagram the values for sunshine have been divided by the maximum possible hours of sunshine at 55N in each month. A further summary of the variation of the standard deviation from station to station is given in figure 5 by plotting the standard deviation of the station standard deviations. In interpreting these graphs it is reasonable to assume that the standard deviation of the daily values of the residuals at each station are approximately normally distributed and also increase from north to south. However it must also be noted that there is considerable inter-station variability and values at each station need to be studied for detailed assessment of the quality of fitting the data by this 'factor' method. It is beyond the scope of this paper to present such detailed information.

Some further general observations on the performance of the method are given in Appendix A.

As a by-product, the performance figures also give an assessment of the maximum values to be expected in the spatial variability of each parameter. Hopkins, 1977, has already addressed this problem in a study of maximum and minimum temperature, and sunshine for the topographically simple area of East Anglia. Values from his paper of the rms error of interpolation for a typical network with station spacing of d km are indicated by $T_W(d)$ for winter and $T_S(d)$ for summer. It can be seen that both techniques give broadly similar results taking into account the variation in the station spacing across the country and the variation in topography. The approximate distribution of climatological stations is given in figures 6a and 6b.

2. The 'suspect data' plotted map

For each suspect observation a map is plotted on the line printer; six maps can be plotted on each page and an example is given in figure 2. The suspect data are plotted at the centre of each map and identified by '≠' and all other observations nearby are plotted in the approximate relative geographical location to the suspect. All data are also output to the right of the chart and preceded by station DCNN, the first four characters of the station name and the altitude (in metres). If there is no space to plot an observation it is identified by a letter; the letter may be plotted on the map if there is room. Each map covers an area of about 100 km x 100 km but the scale is adjusted so that there are usually between 7 and 26 observations on each map.

The suspect observation is identified on the top line with

- station DCNN and first 8 characters of station name

- date (if no day is given the chart is for the monthly mean)

- name of element

- observed value of element (OB:)

- residual (R:)

- probable reason why data suspect (eg INDEX, SIGN, +/-5?, +/-10?, RESET) if any

- estimated value (E:)

The second line contains

- the standardised deviation (SD:)

- '(AMENDED DATA)' if the plotted data taken from the MWR - series periodic data is different from what it was (as printed on the top line) at the time of the quality control run

The bottom line contains

- width of chart (S: - for 'scale')

- 'DCNN NOT FOUND' if DCNN is not in back-up data set (MØ3.RCLIMOPN) which contains station positions.

- 'POSITION NOT KNOWN' if back-up data set (MØ3.RCLIMOPN) contains no position (i.e position is given as 'missing data')

(If the station location is not known, the suspect is printed together with suspect data from all other stations for which the location is not known)

Data are plotted as integers; most values are in tenths (viz tenths of degrees Celsius, tenths of hours of sunshine etc).

3. The computer programs

Processing and identification of the suspect data is accomplished in 2 stages:

Stage 1: A 'submitter' job is run which sets up the JCL for stage 2 and submits it. The 'submitter' job takes in commands in a special control language and effectively modifies a basic set of JCL.

Stage 2: A 'quality control' job is run for each element and each month to be processed.

Stage 1 provides a very flexible method of choosing which element and month to run. It also gives complete freedom over the choice of several arbitrary parameters including those which govern the criteria for identifying suspect data. One run of the 'submitter' job can submit several 'quality control' jobs.

An example is given in Appendix B of the standard card deck currently used for quality control of six elements. In normal operation it is only necessary to change the parameters for year (YY or YYYY) and month (MM). It is possible to select which elements to run by removing the control statements referring to those elements not required or inserting control statements for other elements.

A second example of a card deck that might be submitted is given in Appendix C; this processes five elements for data held on offline disks and is currently used to identify 'gross' errors in past data.

It should be noted that use of the command ACR= (or ACRONYM=) automatically invokes a set of default control statements which are scanned in order to choose default parameters that are to be used. The default control statements in current use are given in Appendices D and E.

A description of the control language and a summary of the commands used by the 'submitter' job is given in Appendix F. The strategy of the 'submitter' and 'quality control' jobs is given in section 4.

4. Operation of the jobs

As was noted in section 3, the 'quality control' job is submitted by a 'submitter' program controlled by commands in a simple control language. A summary of the software used by these programs is given in Appendix I.

4.1. The 'submitter' job

This program is executed by the FORTGLG procedure and uses a basic object module on MØ3.OBJLIB(RFAQCS). It operates as follows:

a. A 'model' JCL stream is input (on FT1ØFØØ1)

b. 'Override' control statements are input (from FTØ5FØØ1) and are processed - usually specifying particular modifications to the 'model' JCL. However when the ACRONYM command is found, 'default' control statements are input (from FT11FØØ1). The 'default' statements are scanned until a matching ACRONYM command is detected and then the subsequent statements (until the next ACRONYM command) are used to modify the 'model' JCL. When all 'default' statements have been scanned processing continues with the 'override' statements.

c. Each time the 'SUBMIT' command is encountered the modified JCL stream is submitted (to the internal computer reader), using the Met O 12 SUBJOB subroutine, together with a copy of the BIMED procedure (from FT12FØØ1).

4.2. The 'Quality Control' Job

The quality control job inputs the required data from the appropriate 'daily' or '09' MWR periodic data set, matches it with the required set of factor scores, estimates missing data, obtains residuals for the observations and analyses them. Observations with residuals falling outside limits which are determined statistically are regarded as suspect and output.

The structure of the job and a summary of the output is given in Appendix G. Several steps in the job use BMDP programs (Dixon, 1975).

4.3. Data Sets

Both the 'submitter' and 'quality control' stages use a number of data sets which contain either program code, data, or lists of various sorts. Details are given in Appendix H.

References

BRYANT, G W (1979). Met. Mag. 108

Archiving and quality control of climatological data.

HOPKINS, J S (1977). Met. Mag. 106

The spatial variability of daily temperatures and sunshine over uniform terrain.

SPACKMAN, E A (1979). Meteorological Office, Met O 3

Technical Note No. 3. A Multivariate Analysis of Temperature within the UK Climatological Network.

DIXON, W.J (editor) (1975). University of California Press

Biomedical Computer Programs.

Appendix A

General observations on the performance of the quality control method

- value of standardised deviation

  $\lesssim$ 3.5: data are usually acceptable

  $\gtrsim$ 4.5: data are usually regarded as in error

- it seems that most (but not all) unacceptable observations are detected by taking the limit of standardised deviation at 3.25, however up to 50% of the suspects may seem acceptable.

- it must not be assumed that all suspects have a reason (even subtle and hidden) to be wrong - statistically one can never definitively identify the erroneous and only the erroneous data.

- the technique seems to work better for temperature than for sunshine; in the case of sunshine it seems that small scale areas of sunshine are sometimes poorly fitted resulting in several nearby stations with similar values all being queried.

- some regions seem to be fitted relatively badly more often than others eg - coastal regions of East Anglia, outlying islands (although attempts have been made to deal specifically with the Orkneys and Shetlands, Isle of Man and Channel Islands) and sparse data areas in mountainous terrain .

Appendix B

'Submitter' card deck for current data

```
//*              THIS JOB CREATES JOBS FOR THE FACTOR METHOD QUALITY
//*              CONTROL.
//*              SET UP BY EDDIE SPACKMAN ON 5TH FEB 1979
//*              ONE JOB IS SET UP FOR EACH ELEMENT/PARAMETER
//SUBMIT EXEC FORTGLG,PARM.LKED='MAP,LET,LIST',COND=EVEN,TIME.GO=(,15)
//LKED.SYSPRINT DD SYSOUT=Z
//LKED.PROGLIB DD DISP=SHR,DSN=MET.PROGLIB
//LKED.MO3OBJ DD DISP=SHR,DSN=MO3.OBJLIB
//LKED.SYSIN DD DISP=SHR,DSN=MO3.CNTLDATA(RFAQCL)
//GO.FT06F001 DD SYSOUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=133)
//IJPDD DD DSN=ASP.VERSON32,DISP=SHR
//IJPWTR DD SYSOUT=Q
//IJPDALIB DD UNIT=DISK,VOL=SER=SYS006,DISP=OLD
//GO.INTOUT DD SYSOUT=(A,INTRDR)
//GO.INTINP DD UNIT=DISK,VOL=SER=SYS006,DISP=OLD
//GO.FT10F001 DD LABEL=(,,,IN),DISP=SHR,DSN=MO3.CNTLDATA(RFAQCJ)
//GO.FT11F001 DD LABEL=(,,,IN),DISP=SHR,DSN=MO3.CNTLDATA(RFAQCD)
//GO.FT12F001 DD LABEL=(,,,IN),DISP=SHR,DSN=MO3.CNTLDATA(RFAQCPR)
//GO.FT20F001 DD DISP=(NEW,DELETE),DSN=MO3.RFAQCJOB,LABEL=(,,,OUT),
// SPACE=(TRK,(4,1)),DCB=(RECFM=FB,LRECL=80,BLKSIZE=0960),
// UNIT=DISK,VOL=SER=SYS006
//GO.SYSIN DD DATA FOR OVERRIDE STATEMENTS
ACR;
ACCOUNT=(MO3201,L),WARDER.2271,; PRTY=2;
YY=80; MM=08; COMMENT=SPECIFY ON THIS CARD YEAR & MONTH;
ACR=TX;
SUBMIT=MO3LMWTX;
ACR=W9;
SUBMIT=MO3LMWW9;
ACR=TN;
SUBMIT=MO3LMWTN;
ACR=GN;
SUBMIT=MO3LMWGN;
ACR=T9;
SUBMIT=MO3LMWT9;
/*              THIS CARD MUST BE INCLUDED (TO END DD DATA I/P)
```

Appendix C

'Submitter' card deck for past data

```
//*              THIS JOB CREATES JOBS FOR THE FACTOR METHOD QUALITY
//*              CONTROL.
//*              SET UP BY EDDIE SPACKMAN ON 5TH FEB 1979
//*              ONE JOB IS SET UP FOR EACH ELEMENT/PARAMETER
//SUBMIT EXEC FORTGLG,PARM.LKED='MAP,LET,LIST',COND=EVEN,TIME.GO=(,30)
//LKED.SYSPRINT DD SYSOUT=Z
//LKED.PROGLIB DD DISP=SHR,DSN=MET.PROGLIB
//LKED.MO3OBJ DD DISP=SHR,DSN=MO3.OBJLIB
//LKED.SYSIN DD DISP=SHR,DSN=MO3.CNTLDATA(RFAQCL)
//GO.FT06F001 DD SYSOUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=133)
//IJPDD DD DSN=ASP.VERSON32,DISP=SHR
//IJPWTR DD SYSOUT=Q
//IJPDALIB DD UNIT=DISK,VOL=SER=SYS006,DISP=OLD
//GO.INTOUT DD SYSOUT=(A,INTRDR)
//GO.INTINP DD UNIT=DISK,VOL=SER=SYS006,DISP=OLD
//GO.FT10F001 DD LABEL=(,,,IN),DISP=SHR,DSN=MO3.CNTLDATA(RFAQCJ)
//GO.FT11F001 DD LABEL=(,,,IN),DISP=SHR,DSN=MO3.CNTLDATA(RFAQCDGR)
//GO.FT12F001 DD LABEL=(,,,IN),DISP=SHR,DSN=MO3.CNTLDATA(RFAQCPR)
//GO.FT20F001 DD DISP=(NEW,DELETE),DSN=MO3.RFAQCJOB,LABEL=(,,,OUT),
// SPACE=(TRK,(4,1)),DCB=(RECFM=FB,LRECL=80,BLKSIZE=0960),
// UNIT=DISK,VOL=SER=SYS006
//GO.SYSIN DD DATA FOR OVERRIDE STATEMENTS
ACCOUNT=(MO3201,L),R.WARDER.2271,; PRTY=1;
REPLACE;
//GO.FT31F001 DD DISP=(OLD,KEEP),DSN=MO3.RQC.QUERY.**GROSS,          00000187
//GO.FT10F001 DD DISP=(SHR,KEEP),DSN=MO3.RQC.QUERY.**GROSS,          00000232
REPLACEEND;
```

```
YY=79;
 COMMENT= INSERT CARDS TO USE DAILY PERIODIC DATA SET;
REPLACE;
//GO.ARCHIV90 DD DISP=SHR,LABEL=(,,,IN),DSN=MP.DMWR.Y79,              00000038
//  UNIT=DISK,VOL=(,RETAIN,SER=PAGE02)                                00000039
//GO.ARCHIV11 DD DISP=SHR,DSN=MP.DMWR.Y79,LABEL=(,,,IN),              00000220
//  UNIT=DISK,VOL=(,RETAIN,SER=PAGE02)                                00000221
REPLACEEND;
 COMMENT= INSERT CARDS FOR TX,TN,GN,SN;
MM=05; ACR=TX; SDPRN=4.00; SUB=M03LEGTX;
MM=05; ACR=TN; SDPRN=4.00; SUB=M03LEGTN;
MM=05; ACR=GN; SDPRN=4.00; SUB=M03LEGGN;
 COMMENT= INSERT CARDS TO USE 0900Z PERIODIC DATA SET;
REPLACE;
//GO.ARCHIV91 DD DISP=SHR,LABEL=(,,,IN),DSN=MP.H9MWR.Y79,             00000038
//  UNIT=DISK,VOL=(,RETAIN,SER=PAGE02)                                00000039
//GO.ARCHIV12 DD DISP=SHR,DSN=MP.H9MWR.Y79,LABEL=(,,,IN),             00000220
//  UNIT=DISK,VOL=(,RETAIN,SER=PAGE02)                                00000221
REPLACEEND;
 COMMENT= INSERT CARDS FOR T9,W9;
MM=05; ACR=T9; SDPRN=4.00; SUB=M03LEGT9;
MM=05; ACR=W9; SDPRN=4.00; SUB=M03LEGW9;
/*              THIS CARD ENDS DD DATA I/P
```

Appendix D

Default control statements for current data

(on MØ3.CNTLDATA(RFAQCD))

```
DSNJOB=MO3.RFAQCJOB;
CONV=0.10; RESMN=-20.; RESMX=+20.; RESCR=1.0; SDPRN=3.25; RSPRN=2.5;
SC=100.0; ORD=STN; NF=15;
REPLACE; GENERAL OVERRIDE CARDS
//*FORMAT PR,DDNAME=QUERIES.GO.FT62F001,OVFL=OFF,COPIES=2              00000010
//GO.FT07F001 DD SYSOUT=Z,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)            00000152
REPLACEEND;
ACR=SN; IEL=13; EL=SUN DURN; JOB=MO3RRQAS;
ACR=TX; IEL=03; EL=TMAX 9-9; JOB=MO3RRQAX;
ACR=TN; IEL=04; EL=TMIN 9-9; JOB=MO3RRQAN;
ACR=GN; IEL=14; EL=GMIN     ; JOB=MO3RRQAG;
ACR=T9; IEL=59; EL=DRY BULB; JOB=MO3RRQAT;
ACR=W9; IEL=60; EL=WET BULB; JOB=MO3RRQAW;
ACR=D9; IEL=61; EL=DEW PT   ; JOB=MO3RRQAD;
ACR=RN; IEL=09; EL=RAIN     ; CONV=0.10; JOB=MO3RRQAR;
ACR=SL; IEL=27; EL=SOIL 30  ; JOB=MO3RRQAE;
ACR=F9; IEL=55; EL=WIND 09  ; CONV=1.00; JOB=MO3RRQAF;
ACR=SN; VARMN=-.01; VARMX=24.0; RESMN=-10.; RESMX=+10.;
ACR=TX; VARMN=-20.; VARMX=+40.;
ACR=TN; VARMN=-40.; VARMX=+30.;
ACR=GN; VARMN=-40.; VARMX=+30.;
ACR=T9; VARMN=-30.; VARMX=+30.;
ACR=W9; VARMN=-30.; VARMX=+30.;
ACR=D9; VARMN=-30.; VARMX=+30.;
ACR=RN; VARMN=-.01; VARMX=125.; RESMN=-50.; RESMX=+50.;
ACR=SL; VARMN=-20.; VARMX=+30.;
ACR=F9; VARMN=-.01; VARMX=150.;
```

```
ACR=SN; RESSC=2.1; RESCR=0.0; RSPRN=0.0; SC=150.0; ORD=DAY;
ACR=TX; RESSC=1.3;
ACR=TN; RESSC=1.8;
ACR=GN; RESSC=2.4; RSPRN=3.0;
ACR=T9; RESSC=1.5;
ACR=W9; RESSC=1.3;
ACR=D9; RESSC=1.3;
ACR=RN; RESSC=6.0; RESCR=0.0; RSPRN=0.0;
ACR=SL; RESSC=1.0; RSPRN=0.5;
ACR=F9; RESSC=10.; RESCR=10.; RSPRN=00.;
ACR=TX; DSNFACT=M03.RVB.TX7377.E7279;
ACR=TN; DSNFACT=M03.RVB.TN7377.E7279;
ACR=GN; DSNFACT=M03.RVB.G777NOK;
ACR=T9; DSNFACT=M03.RVB.CD7377.E7279;
ACR=W9; DSNFACT=M03.RVB.EP7377.E7279;
ACR=D9; DSNFACT=M03.RVB.EP7377.E7279;
ACR=RN; DSNFACT=M03.RVB.RN7377.E7279;
ACR=SL; DSNFACT=M03.RVB.ER7377.E7279; NFACT=05;
ACR=F9; DSNFACT=M03.RVB.FF7377.E7279;
ACR=SN; DSNFACT=M03.RVB.SN7377.E7279;
ACR=TX; REPLACE; CARDS REQUIRED FOR T.MAX
//GO.FT07F001 DD SYSOUT=B,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)            00000152
REPLACEEND;
ACR=TN; REPLACE; CARDS REQUIRED FOR T.MIN
//GO.FT07F001 DD SYSOUT=B,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)            00000152
REPLACEEND;
ACR=SN; REPLACE; ONE COPY ONLY FOR SUN
//*FORMAT PR,DDNAME=QUERIES.GO.FT62F001,OVFL=OFF,COPIES=1              00000010
REPLACEEND;
```

Appendix E

Default control statements for past data

(on MØ3.CNTLDATA(RFAQCDGR))

```
DSNJOB=MO3.RFAQCJOB;
CONV=0.10; RESMN=-20.; RESMX=+20.; RESCR=1.0; SDPRN=4.00; RSPRN=2.5;
SC=100.0; ORD=STN; NF=15;
REPLACE; GENERAL OVERRIDE CARDS
//*FORMAT PR,DDNAME=QUERIES.GO.FT62F001,OVFL=OFF,COPIES=1                    00000010
//GO.FT07F001 DD SYSOUT=Z,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)                  00000152
REPLACEEND;
ACR=SN; IEL=13; EL=SUN DURN; JOB=MO3RRQAS;
ACR=TX; IEL=03; EL=TMAX 9-9; JOB=MO3RRQAX;
ACR=TN; IEL=04; EL=TMIN 9-9; JOB=MO3RRQAN;
ACR=GN; IEL=14; EL=GMIN     ; JOB=MO3RRQAG;
ACR=T9; IEL=59; EL=DRY BULB; JOB=MO3RRQAT;
ACR=W9; IEL=60; EL=WET BULB; JOB=MO3RRQAW;
ACR=D9; IEL=61; EL=DEW PT   ; JOB=MO3RRQAD;
ACR=RN; IEL=09; EL=RAIN     ; CONV=0.10; JOB=MO3RRQAR;
ACR=SL; IEL=27; EL=SOIL 30  ; JOB=MO3RRQAE;
ACR=F9; IEL=55; EL=WIND 09  ; CONV=1.00; JOB=MO3RRQAF;
ACR=SN; VARMN=-.01; VARMX=24.0; RESMN=-10.; RESMX=+10.;
ACR=TX; VARMN=-20.; VARMX=+40.;
ACR=TN; VARMN=-40.; VARMX=+30.;
ACR=GN; VARMN=-40.; VARMX=+30.;
ACR=T9; VARMN=-30.; VARMX=+30.;
ACR=W9; VARMN=-30.; VARMX=+30.;
ACR=D9; VARMN=-30.; VARMX=+30.;
ACR=RN; VARMN=-.01; VARMX=125.; RESMN=-50.; RESMX=+50.;
ACR=SL; VARMN=-20.; VARMX=+30.;
ACR=F9; VARMN=-.01; VARMX=150.;
```

```
ACR=SN; RESSC=2.1; RESCR=0.0; SDPRN=3.15; RSPRN=0.0; SC=150.0; ORD=DAY;
ACR=TX; RESSC=1.3;
ACR=TN; RESSC=1.8;
ACR=GN; RESSC=2.4; RSPRN=3.0;
ACR=T9; RESSC=1.5;
ACR=W9; RESSC=1.3;
ACR=D9; RESSC=1.5;
ACR=RN; RESSC=6.0; RESCR=0.0; RSPRN=0.0;
ACR=SL; RESSC=1.0; RSPRN=0.5;
ACR=F9; RESSC=10.; RESCR=10.; RSPRN=00.;
ACR=TX; DSNFACT=M03.RVB.TX7377.E7279;
ACR=TN; DSNFACT=M03.RVB.TN7377.E7279;
ACR=GN; DSNFACT=M03.RVB.G777NJK;
ACR=T9; DSNFACT=M03.RVB.CJ7377.E7279;
ACR=W9; DSNFACT=M03.RVB.DP7377.E7279;
ACR=D9; DSNFACT=M03.RVB.DP7377.E7279;
ACR=RN; DSNFACT=M03.RVB.RN7377.E7279;
ACR=SL; DSNFACT=M03.RVB.ER7377.E7279; NFACT=05;
ACR=F9; DSNFACT=M03.RVB.FF7377.E7279;
ACR=SN; DSNFACT=M03.RVB.SN7377.E7279;
ACR=TX; REPLACE; CARDS REQUIRED FOR T.MAX
//GO.FT07F001 DD SYSOUT=B,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)               00000152
REPLACEEND;
ACR=TN; REPLACE; CARDS REQUIRED FOR T.MIN
//GO.FT07F001 DD SYSOUT=B,DCB=(RECFM=F,LRECL=80,BLKSIZE=80)               00000152
REPLACEEND;
ACR=SN; REPLACE; ONE COPY ONLY FOR SUN
//*FORMAT PR,DDNAME=QUERIES.GO.FT62F001,OVFL=OFF,COPIES=1                 00000010
REPLACEEND;
```

Appendix F

The 'Submitter'

1. Control Language

This control language is common to both 'default' and 'override' inputs.

The syntax of the language (a sentence) is

< command > = < value > ;   (eg ACRONYM = TN;)

and is assumed to be contained in the first 72 positions of a punched card or card image of 80 positions. Spaces before the command field or after the 'semi-colon' are ignored. Spaces after the 'equal sign' and before the 'semi-colon' are assumed to be part of the value field. The value field may be empty and in this case the 'equal sign' may also (but need not) be omitted. A sentence must be complete on one card and not spill on to a second card. However each card may contain more than one sentence.

2. Commands

The following table describes the commands

Notes:

1. Commands: letters in capitals are obligatory whilst those in lower case are optional.

2. Format: Fortran conventions are used. (Note that for 'F format' the format on card input overrides that in the program if the decimal point is present.) F - floating point, I - integer and A - character.

3. The only commands in common use in 'override' statements are: ACRONYM, MM, YYYY, ACCOUNT, PRTY, REPLACE, DATA, JOBNAME and SUBMIT. The others are usually only found in 'default' statements or are used for special purposes (eg program development, debugging etc).

Command	Format of Value Field	Explanation
ACRônym	A2 (or empty)	Value identifies element to process - TX Maximum temperature (09-09) TN Minimum temperature (09-09) GN Grass Minimum Temperature (mixed 09-09 and dusk --09) T9 Dry Bulb at 0900 W9 Wet Bulb at 0900 SN Sunshine F9 Wind speed at 09 RN Rainfall total (09-09) SL Soil temperature (30 cm) D9 Dew Point at 0900 The value may be empty and then applies generally.

Command	Format of Value Field	Explanation
MM	I2	Month of year
YYyy	I2 or I4	Year Values read as I2 have 19ØØ added. Resulting values which are less than 1970 have 1000 added which should put them in the range 1970-2069.
ACCount	A12 to A56	Account field of job card: eg: (MØ32Ø5,R),EDDIE.X25Ø4,
PRTY	A1 or A2	Job priority
REPLACE	empty	Following cards are used as replacements for job stream. Position is identified by sequence number. The last card must be REPLACEEND;
DATA	empty	Following cards contain data in multiples of 3 cards per station. Format: IX,A4,A1/5X,15F5.1/16F5.1 for DCNN, region (district), data. The last card must be DATAEND;
JOBname	A8	Jobname
SUBmit	empty or A8	Submit JCL stream. Value identifies jobname.
SDPRN	F4.0	Lower limit of standardised deviation for plotting suspect data
IEL	I2	Element number (as in periodic MWR series write-up). 50 is added for 09 data set.
ELement	A8	Name to be used for element.
VARMN	F4.0	Minimum value of variable) Values treated
VARMX	F4.0	Maximum value of variable) as missing if
RESMN	F4.0	Minimum value of residual) out of these
RESMX	F4.0	Maximum value of residual) ranges.
RSPRN	F3.1	Lower limit of modulus of residual for plotting suspect data
RESSC	F3.1	Upper limit of standardised deviation in 1st scan of residual analysis beyond which station will not be included in statistics for second scan. (residual standardised critical).
RESCR	F3.1	Lower limit of modulus of residual beyond which station will not be considered suspect. (residual critical)

F2

Command	Format of Value Field	Explanation
SCale	F5.0	Width for chart plot. (in km)
DSNJOB	A1 to A44	DSN of data set on which JCL stream is output for submission by subroutine SUBJOB
DSNFact	A1 to A40	DSN of data set of factor scores.
NFact	I1 or I2	Number of factors
CONV	F4.2	Coefficient to convert MWR-series value to real meteorological unit (usually Ø.1Ø)
ORDer	A3	<value>= STN for station order or DAY for day order of plots.
PRint	empty	Print JCL stream (on FTØ6FØØ1) when job submitted.
MONitor	empty	Monitor control language statements by output to FTØ6FØØ1.
COMment	not checked	Value may be used for comment.
END / STOP	empty	Terminate all control language and submit JCL

3. Failures

There are 4 classes of failure that may occur with the control language:

- Invalid command
- Invalid syntax
- Wrong length value field
- Program malfunction.

In most cases a message should be printed which identifies the failure, the command causing it and the FT number of the DD-statement from which the command was read. After most failures the control language continues to be scanned but no further JCL is submitted. Program malfuntion will need to be referred to programmer support.

Appendix G

The structure of the Factor Analysis Quality Control Job

The job to perform the quality control consists of seven steps

1. CONTROL - to set control information for use by subsequent steps. (uses IEBGENER)

2. DATA - to extract data from periodic (MWR series) data set and match with factor scores for station.

3. DATASET - to identify and estimate missing (or unacceptable) values of the data or factor scores and to estimate values for these missing elements. (uses BMDPAM)

4. REGRESS - to obtain residuals by multiple linear regression of data for each day or factor scores. (uses BMDP6R)

5. COEFPRNT - to print and punch regression coefficients (factor loadings) and to print standard errors on regression coefficients

6. RESDANYL - to analyse residuals and print statistics for each station and suspect data. (uses BMDP1S)

7. QUERIES - to plot suspect data at 6 stations per page.

Output from the job is as follows:

- fiche
  - SYSMSG etc of system output, JCL etc.
  - program control statements and other output from BMBP programs
  - miscellaneous statistical output
  - duplicates of printer and some punch output
  - plots around stations with standardised deviation > 2.75
- printer
  - regression coefficients of factor loadings for each day and mean
  - mean, standard deviation and number of observations used in each region and for all regions for each day
  - estimates of all missing data and missing factor scores
  - plots of suspect data at six per line printer page

- punched cards

  - regression coefficients of factor loadings for each day and mean (maximum and minimum screen temperatures only)

- disk data sets

  - observations which may be suspect

  - statistics of the fitting of the observations

Appendix H. Data Sets.

1. 'Submitter' job

MØ3.CNTLDATA(RFAQCJ79)	input on FT1ØFØØ1	JCL stream (unmodified)
MØ3.CNTLDATA(RF AQCD) or MØ3.CNTLDATA(RFAQCDGR)	input on FT11FØØ1	default control statements for 'submitter' program
MØ3.CNTLDATA(RFAQCPR)	input on FT12FØØ1	procedure stream containing BIMED
MØ3.RFAQCJOB	output on FT2ØFØØ1	JCL stream (modified) deleted at end of run
MØ3.OBJLIB(RFAQCS)		'submitter' program
MØ3.CNTLDATA(RFAQCL)		link-edit control statements
MET.PROGLIB(SUBJOB)		subroutine SUBJOB

2. 'Quality control' Job

a. DATA step

MØ3.ROBJLIB(RRDDATA) MØ3.OBJLIB(MDDA)		object modules to read data from MWR periodic series data sets
MET.PROGLIB(GPACCESS)		GPACCESS routines
MET.PROGLIB(MOVECH)		subroutine to move characters
MP.DMWR.Yyy or MP.H9MWR.Yyy	input on ARCHIVyØ or ARCHIVy1	Data sets containing daily climatological data (periodic series for year 19yy)
MØ3.RVB.ee7377.E7279 or alternatives on METØ38	input on FT1ØFØØ1	Data set containing station factor scores for element ee

b. COEFPRNT step

MØ3.ROBJLIB(RRDRCSE)		S/R to read regression coefficients and standard errors from BMDP printer file

c. RESDANYL step

MØ3.OBJLIB.METØ38(RQCANALH)		S/R to analyse residuals
MØ3.OBJLIB(RSTAT)		S/R to compute statistics
MØ3.RQC.STAT.eemonyy	output on FT3ØFØØ1	D/S containing statistics of fitting for element(ee), month(mon) and year(yy)
MØ3.RQC.QUERY.ee	output on FT31FØØ1	D/S containing queries and suspects for element(ee)

d. QUERIES step

MØ3.ROBJLIB(RFAQCP)		Program to plot suspects on line-printer
MØ3.OBJLIB(MDDA)		Routine to input data from MWR periodic series D/S
MET.PROGLIB(GPACCESS)		GPaccess routines
MET.PROGLIB(MOVECH)		S/R to move characters
MP.DMWR.Yyy or MP.H9MWR.Yyy	input on ARCHIV11 or ARCHIV12	D/S containing daily climatological data (periodic series) for year 19yy
MØ3.RCLIMOPN	input on FT11FØØ1	D/S containing station locations
MØ3.RQC.QUERY.ee	input on FT1ØFØØ1	D/S containing queries and suspects for element(ee)

3. Factor Score Data Sets

MØ3.RVB.ee7377.E7279 (usually)	D/S containing factor scores for element(ee)

(Sophisticated techniques are required to create these which cannot be described here. It is unlikely that the need will exist to update these for a few years but a job is available which will estimate factors for a new station with a complete year of data. Otherwise the factor scores are estimated from the data available for the month of the quality control run for those stations with no factors.

4. Query Data Set

MØ3.RQC.QUERY.ee	on METØ38	D/S containing queries and suspects (current runs)
or MØ3.RQC.QUERY.eeGROSS	on METØ38	" (gross error check)

5. Statistics Data Set

MØ3.RQC.STAT.eemonyy	on to METØ38	D/S containing statistics relating to fitting of the observations at each station

(These data need to be deleted from time to time - after being used to update annual data-set).

142

6. Station position Data Set

MØ3.RCLIMOPN	update by TSO or access CLIMASTER	D/S containing station positions and altitudes 80 byte record of all stations that have been open since 1971 Records in ascending order of DCNN

Record format is

Characters	Content	
1 to 4	DCNN	
5 to 12	NAME	
13 to 19	Easting	National Grid
20 to 26	Northing	(not Irish, not UTM)
27 to 30	Altitude	
31 to 34	(not used for quality control purposes)	
35 to 38	Year opened (or blank)	
39 to 42	Year closed (or blank)	
43 to 80	(not used for quality control purposes)	

Key to lower-case letters in D/S names

ee	acronym for element
mon	month in characters
yy	year

Appendix I Program Software describing object or load modules and source software if known.

1. 'Submitter' job

Object	Source	
MØ3.OBJLIB(RFAQCS)	Cards containing:	
	MAIN	controlling routine
	NAME	process control statements
	DATA	process special data
	REPLAC	process replacement cards
	SUBMIT	submit and print JCL
	ERROR	O/P error messages
	OUT	O/P data
	PARAM	decode control statement
	STRING	perform string operations compare (CLC) and move (MOVECH)
	RDJCL	input basic JCL and check correct
	MMYY	set month and year

MET.PROGLIB(SUBJOB)

2. 'Quality Control' job

Step	Object	Source
CONTROL	IEBGENER	
DATA	MØ3.ROBJLIB(RRDDATA) MØ3.OBJLIB(MDDA) MET.PROGLIB(GPACCESS) MET.PROGLIB(MOVECH)	MØ3.DEADSRCE(RRDDATA) MØ3.DEADSRCE(MDDA)
DATAEST	MØ3.LOADLIB.METØ38(BMDPAM)	Copy from BMDP-1979
REGRESS	MØ3.LOADLIB.METØ38(BMDP6R)	Copy from BMDP-1979
COEFPRNT	MØ3.ROBJLIB(RRDRCSE)	Card deck
RESDANYL	MØ3.LOADLIB.METØ38(BMDPIS)	Copy from BMDP-1979 which
includes	MØ3.OBJLIB.METØ38(RQCANALH)	METMERGE of MØ3.DEADSRCE (RQCANAL) and card deck
	MØ3.OBJLIB(RSTAT)	MØ3.DEADSRCE(RSTAT)
QUERIES	MØ3.ROBJLIB(FRAQCP) MØ3.OBJLIB(MDDA) MET.PROGLIB(GPACCESS) MET.PROGLIB(MOVECH)	MØ3.DEADSRCE(RFAQCP) MØ3.DEADSRCE(MDDA)

Special note: Object modules are usually produced using the Fortran X-compiler. However, in order to be compatible with the IHC-routines contained in the BMDP load modules MØ3.OBJLIB.METØ38(RQCANALH) must be produced using the Fortran H-compiler.

Fig 1:

Climatological District Map

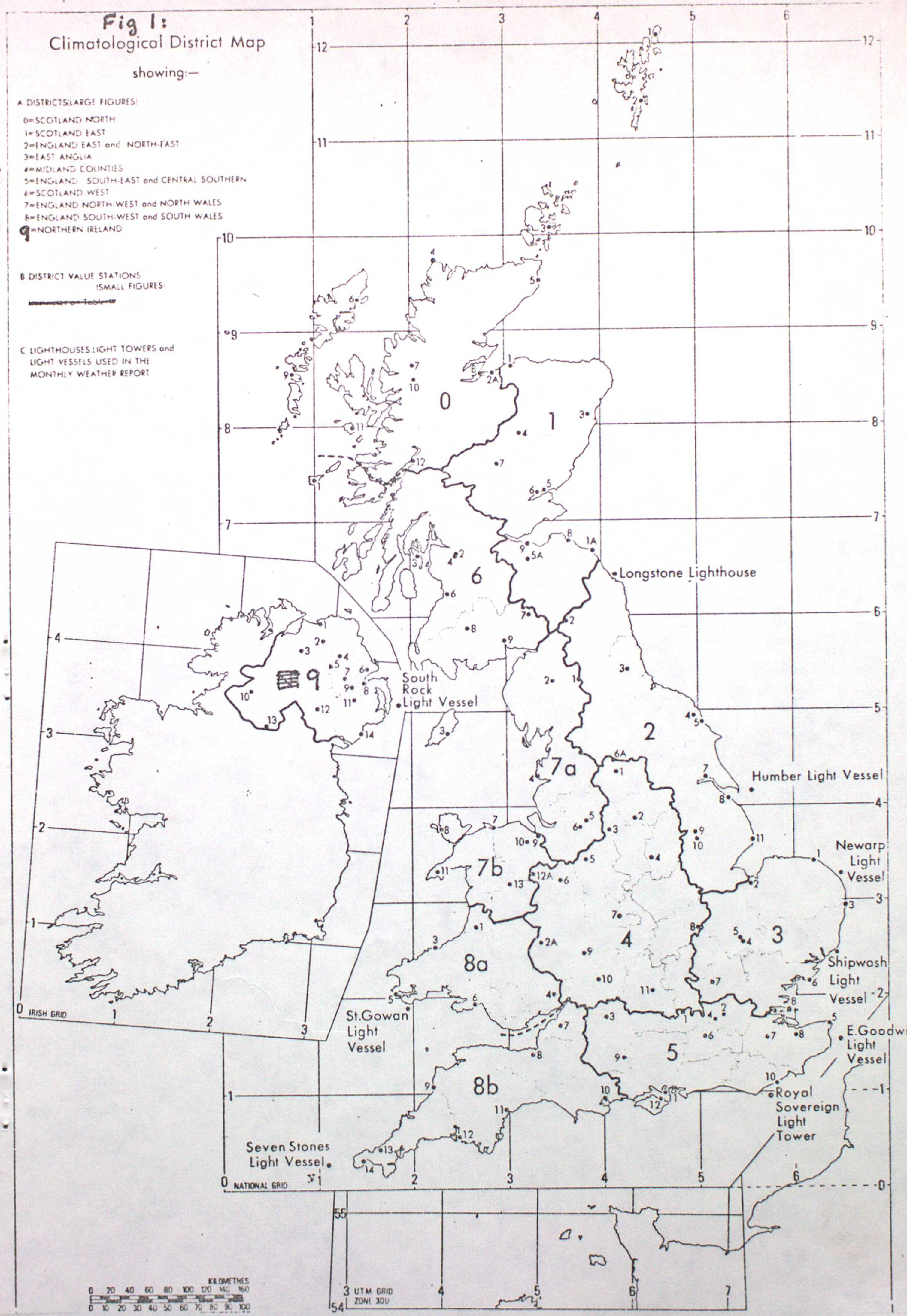
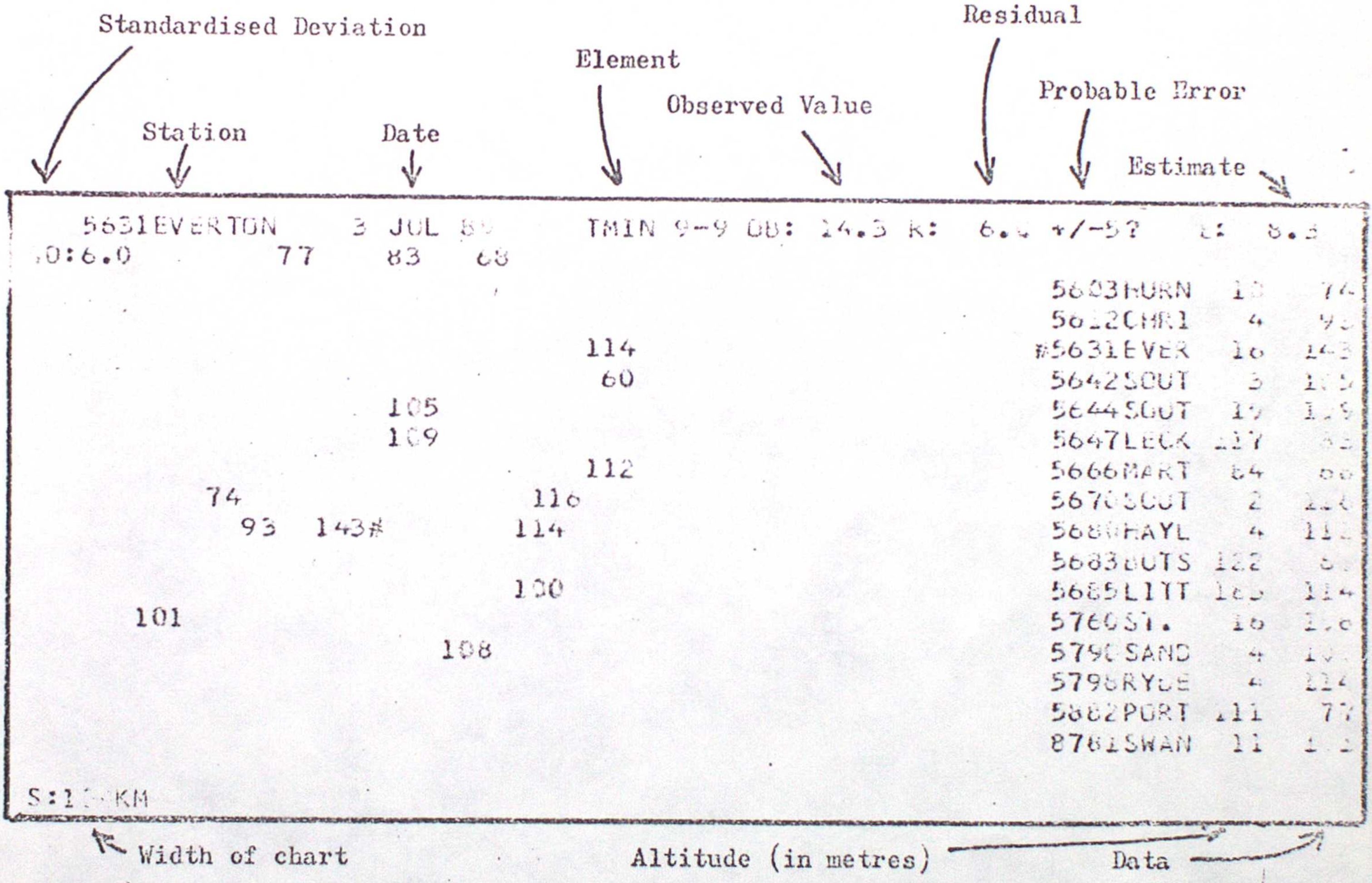


Fig. 2. Plot of 'suspect data'.

The map below is an example of the output. It shows that the Minimum Temperature for Everton (DCNN 5631) in Hampshire on 3rd July 1980 is estimated to have been 8.3 deg C. rather than the archived value of 14.3 deg C. (i.e. a probable error of about 5 deg C.) - the values are plotted in tenths of deg C.



The Standardised Deviation of 6.0 indicates that the observation is 6.0 standard deviations from the Estimate and almost certainly wrong. This is very quickly confirmed by visually inspecting the map and an amended value of 9.3 deg C. would be suggested. In some places it is also very necessary to consider the topography in the region of the suspect observation as this may have an important influence in deciding whether the observed value may in fact be correct.

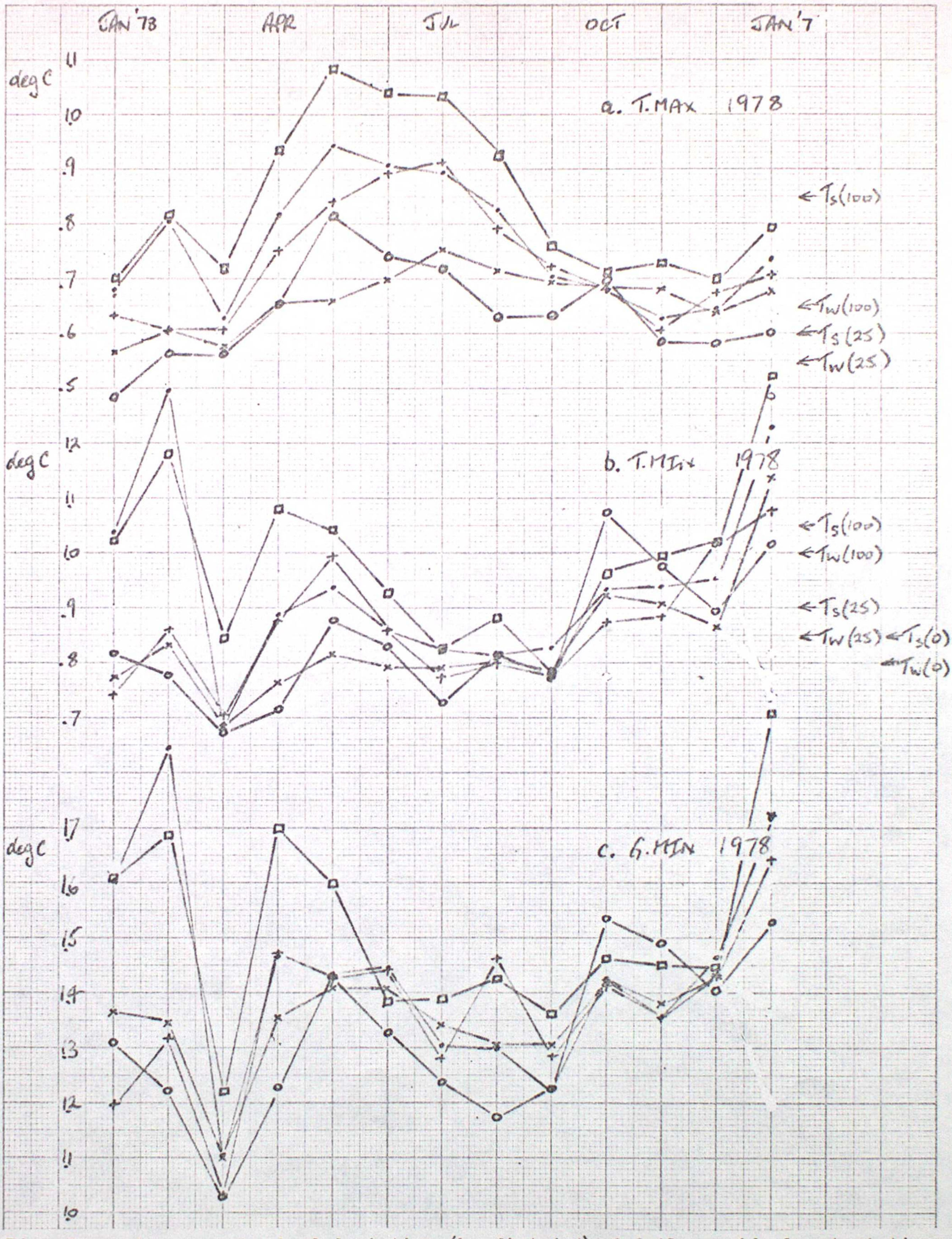


Fig. 3. Average standard deviation (by district) of daily residuals at stations - 1978

Districts: □ - 0, . - 1, x - 4, o - 5, + - 9.

(note: T identifies values from Hopkins, 1977 - see text for details)

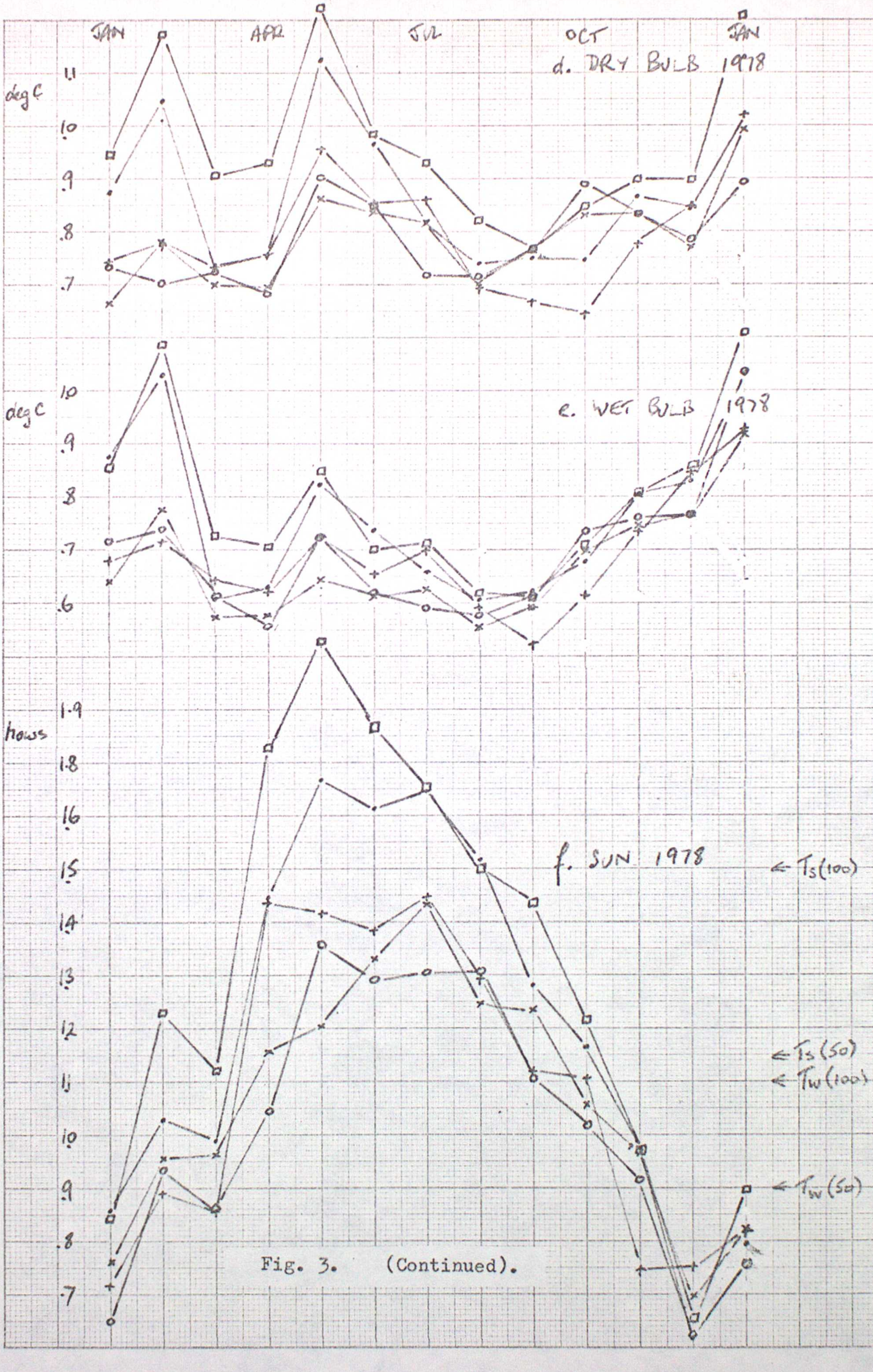


Fig. 3. (Continued).

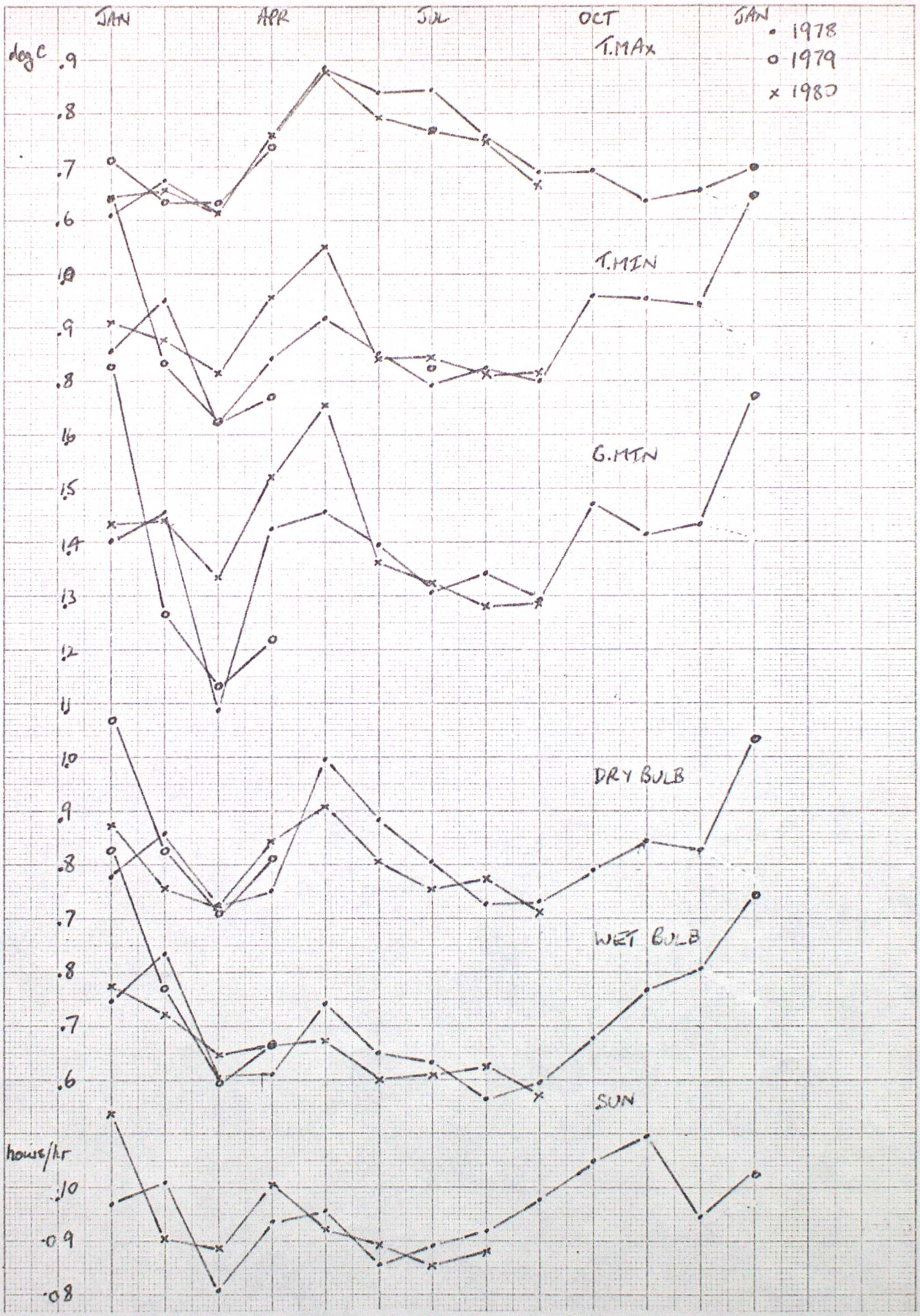


Fig. 4. Average standard deviation of daily residuals over all stations

Year: . 1978 o 1979 x 1980. Data are not complete

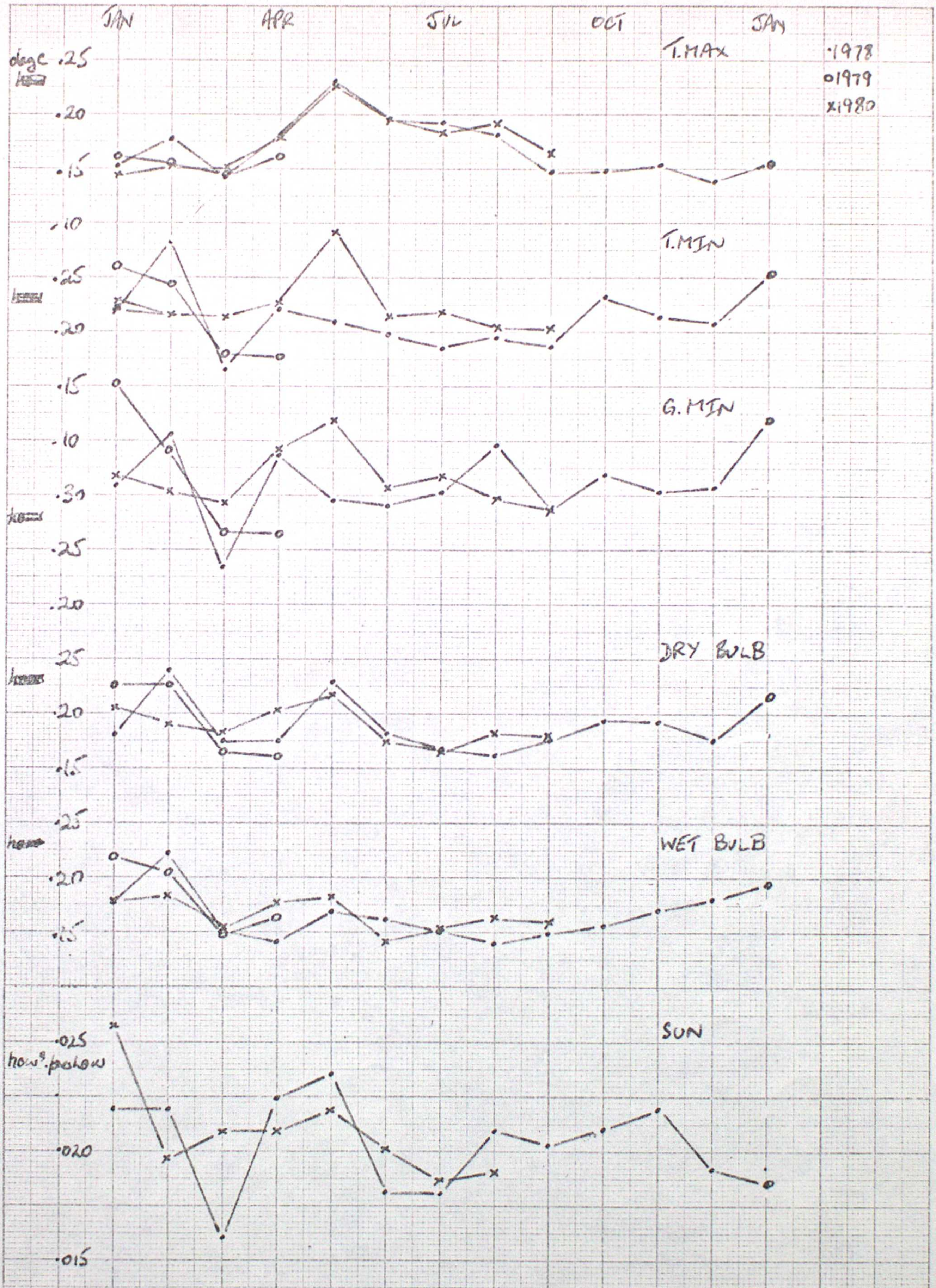


Fig. 5. Standard deviation of the standard deviation of daily station residuals over all stations. Year . 1978 o 1979 x 1980.

Data are not complete

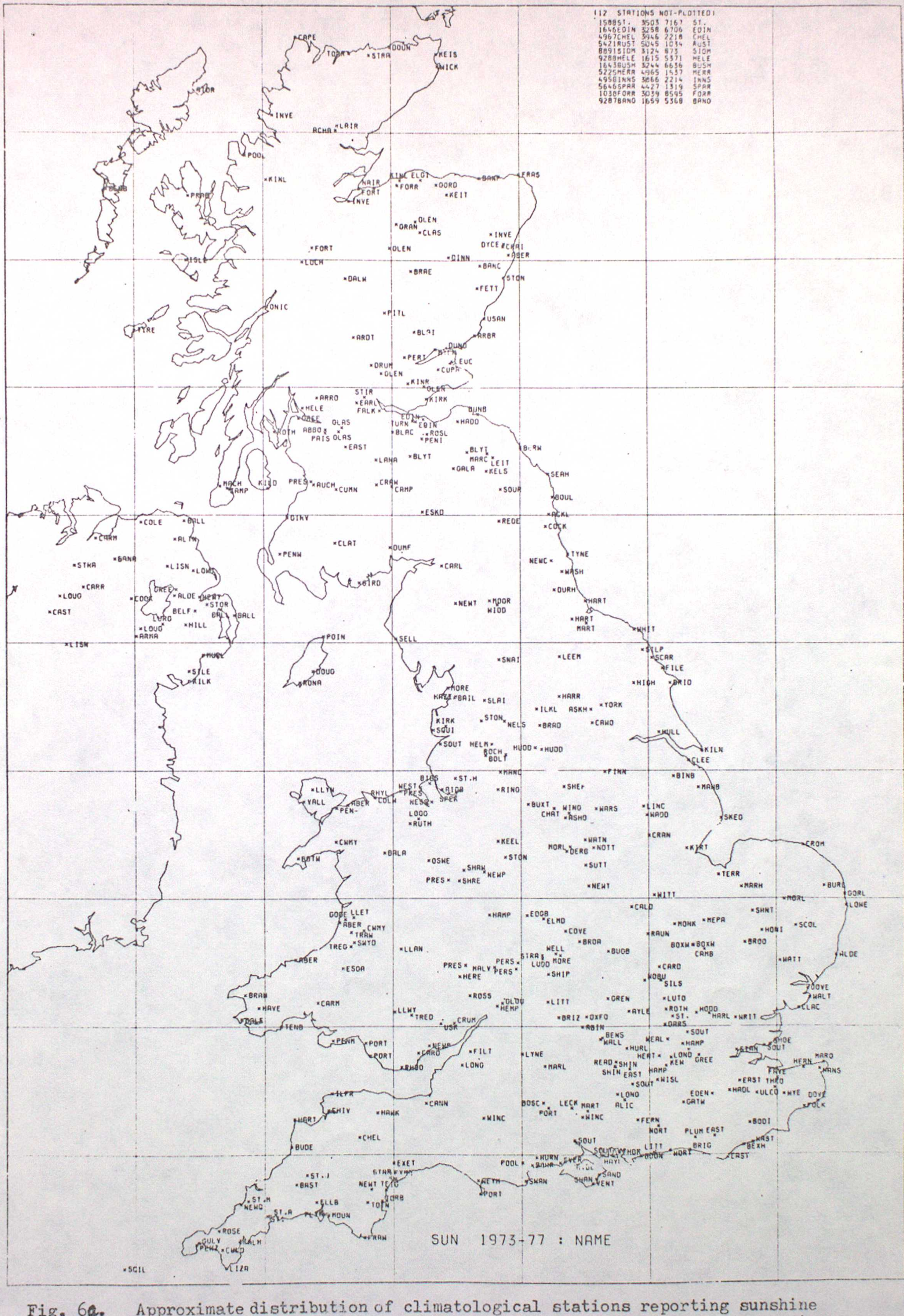


Fig. 6a. Approximate distribution of climatological stations reporting sunshine duration (the first four characters of the station name are plotted).

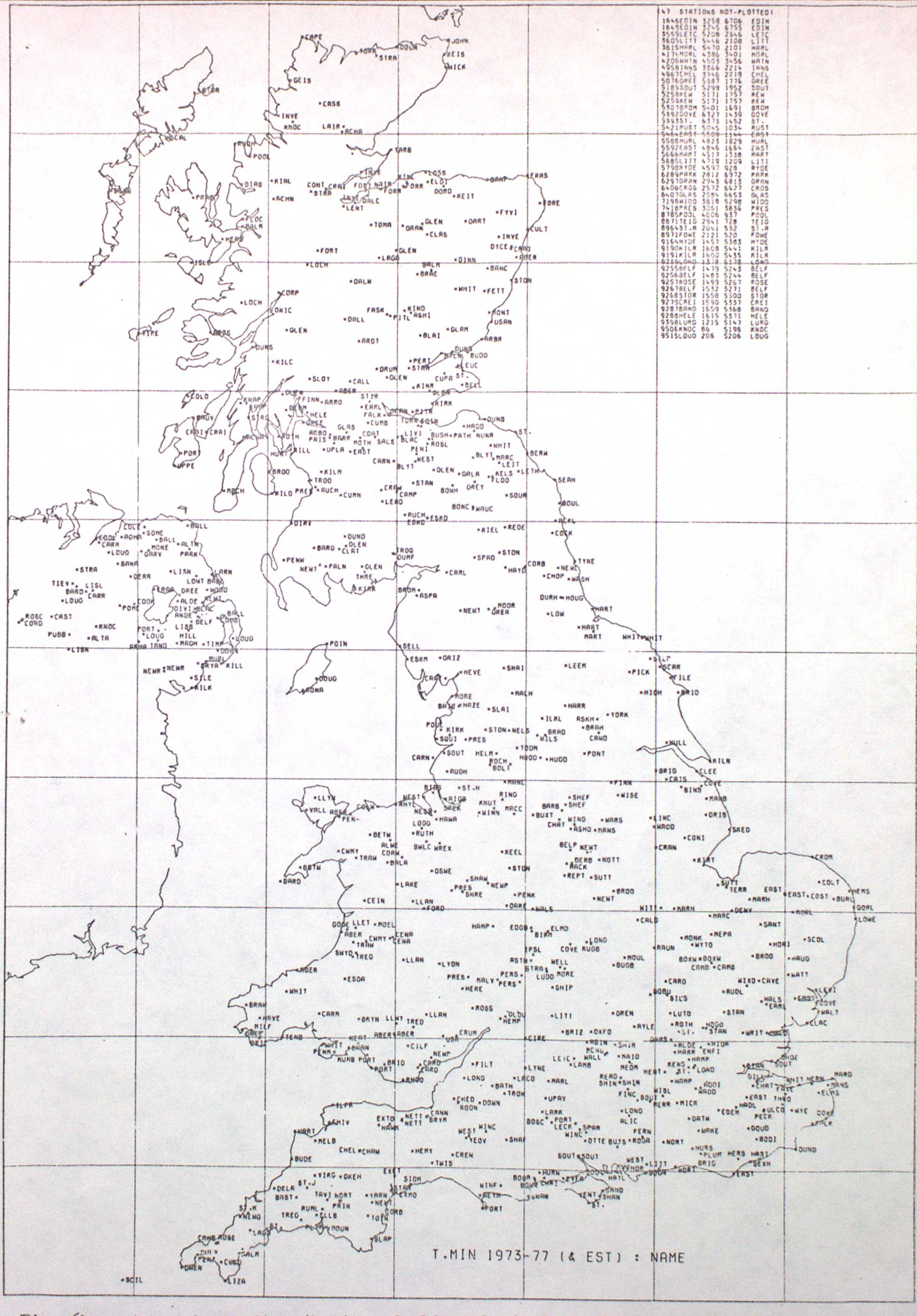


Fig. 6b. Approximate distribution of climatological stations reporting temperature. (The first four characters of the station name are plotted).